"EXPLORATORY ANALYSIS OF TRANSBORDER DYNAMICS, TRANSFRONTERIDAD, AND PREDICTIVE VARIABLES USING MACHINE LEARNING TECHNIQUES TO BUILD PREDICTIVE REGRESSION MODELS"

TESIS PARA: **MAESTRÍA EN CIENCIAS DE LOS DATOS Y PROCESAMIENTO DE DATOS MASIVOS (BIG-DATA)**

PRESENTA(N): **JESÚS RIVAS ALFARO**

DIRECTOR(A) DE TESIS: **MTRO. CÉSAR ROBLEDO JIMÉNEZ**

01 de Marzo de 2022. Aguascalientes, México

Aguascalientes, Ags., 01 de marzo de 2022.

LIC. ROGELIO MARTÍNEZ BRIONES
UNIVERSIDAD CUAUHTÉMOC PLANTEL AGUASCALIENTES
RECTOR GENERAL

P R E S E N T E

Por medio de la presente, me permito informar a Usted que he asesorado y revisado el trabajo de tesis titulado:

"Exploratory Analysis of Transborder Dynamics, Transfronteridad, and Predictive Variables Using Machine Learning Techniques to Build Predictive Regression Models"

Elaborado por **Lic. Jesús Rivas Alfaro**, considerando que cubre los requisitos para poder ser presentado como trabajo recepcional para obtener el grado de **Maestría en Ciencia de los Datos y Procesamiento de Datos Masivos (Big Data).**

Agradeciendo de antemano la atención que se sirva a dar la presente, quedo a sus apreciables órdenes.

ATENTAMENTE

Mtro. César Robledo Jiménez

A Quien Corresponda
Presente

Yo, <u>Jesús Rivas Alfaro</u>, con matrícula <u>MMED19822</u>, egresado del programa *Maestría en Ciencia de los Datos y Procesamiento de Datos Masivos (Big Data)* de la Universidad Cuauhtémoc, plantel Aguascalientes, identificado con IFE-INE o CC, Nº *<u>IDMEX1561862320</u>*, pretendo titularme con el trabajo de tesis titulado:
"Exploratory Analysis of Transborder Dynamics, Transfronteridad, and Predictive Variables Using Machine Learning Techniques to Build Predictive Regression Models",

Por la presente Declaro que:

1.- Este trabajo de tesis, es de mi autoría.
2.- He respetado el Manual de Publicación APA para las citas, referencias de las fuentes consultadas. Por tanto, sus contenidos no han sido plagiados, ni ha sido publicado total ni parcialmente en fuente alguna. Además, las referencias utilizadas para el análisis de la información de este Trabajo de titulación están disponibles para su revisión en caso de que se requiera.
3.- El Trabajo de tesis, no ha sido auto-plagiado, es decir, no ha sido publicado ni presentado anteriormente para obtener algún grado académico previo o título profesional y se han contemplado las correcciones del Comité Tutorial.
4.- Los datos presentados en los resultados son reales, no han sido falseados, ni duplicados, ni copiados y por tanto los resultados que se presentan en el trabajo de tesis, constituirán aporte a la realidad investigada.
5.- De identificarse fraude, datos falsos, plagio información sin citar autores, autoplagio, piratería o falsificación, asumo las consecuencias y sanciones que de mi acción se deriven, sometiéndome a la normatividad vigente de la Universidad Cuauhtémoc, plantel Aguascalientes, Instituto de Educación de Aguascalientes, la Secretaria de Educación Pública, Ministerio de Educación Nacional y/o las autoridades legales correspondientes.
6. Autorizo publicar mi tesis en el repositorio de Educación a Distancia de la Universidad Cuauhtémoc, plantel Aguascalientes.

Jesús Rivas Alfaro
rivas251_alfaro@hotmail.com
+1619-940-8598

Index

Table Index

Figure Index

Resumen

México y los Estados Unidos comparten aproximadamente 1900 millas de frontera, lo que ha causado el desarrollo de regiones y metrópolis transfronterizas a lo largo de este espacio geográfico. Más de cuatro millones de personas se trasladan entre ambas fronteras de una forma relativamente sencilla para comprar productos y servicios varios, visitar familiares y amigos, y realizar otras actividades con el objetivo de obtener lo mejor de cada lado de la frontera. En esta tesis, el autor describe el panorama actual en la frontera México-Estados Unidos, compilando resultados de una revisión de literatura que incluye conceptos de teoría de la frontera de más de 30 autores. El objetivo principal de este documento es incrementar el cuerpo de conocimiento acerca sobre transfronteridad. Este concepto se refiere al nivel de dependencia e interacción que un individuo tiene con ambos lados de la frontera; específicamente, este estudio tiene como objetivo determinar si las habilidades multiculturales están estadísticamente correlacionadas con transfronteridad. Esta relación particular tiene importancia debido a los niveles de bienestar superiores que individuos con altos niveles de transfronteridad experimentan. Con el fin de lograr este objetivo, se utilizó un diseño de investigación cuantitativa por medio del desarrollo de una encuesta que respondieron más de 400 individuos transfronterizos. Adicionalmente, el trabajo de esta tesis requirió que el autor procesara más de 10 millones de registros de individuos viviendo en México y los Estados Unidos proporcionado por los buros de estadística de ambos países. Se generaron una serie de definiciones operacionales para los conceptos de habilidades multiculturales y transfronteridad. Las habilidades multiculturales se dividieron en tres variables: a) Fluencia de idioma, b) Multiculturalidad, y c) Nivel educativo. Los resultados demuestran

que la fluencia de idioma y la multiculturalidad están estadísticamente correlacionadas con transfronteridad mientras que el nivel educativo no. Finalmente, se construyó un modelo Machine Learning de Aprendizaje Supervisado para predecir los niveles de transfronteridad con base en más de diez variables. Los resultados de este ejercicio muestran que, a pesar de haber construido un modelo con significancia estadística, el poder de predicción no es lo suficientemente grande para obtener resultados relevantes o predecir valores potenciales de transfronteridad.

Keywords: Transfronteridad, transborder, machine learning, modelo de aprendizaje supervisado, modelo de regresión, frontera, transmigración.

Abstract

Mexico and the U.S. share approximately 1.9 thousand miles of border, which has caused the development of transborder regions and metropolis throughout this geographical space. More than four million people go back and forth between border seamlessly to purchase various goods and services, visit family members and friends, and do other activities looking to get the best from each side of the border. In this thesis, the author describes the current landscape at the Mexico-U.S. border, compiling results from a literature review that includes border theory concepts from more than 30 authors. The main objective from this document is to increase the body of knowledge regarding transfronteridad. This concept refers to the level of dependency and interaction an individual has with both sides of the border. More specifically, this study aims to determine if multicultural skills are statistically correlated to transfronteridad. This particular relationship has importance due to the increased well-being that individuals with high levels of transfronteridad experience. In order to accomplish this objective, a quantitative research design was followed through the development of a survey that more than 400 transborder responded. Additionally, the work for this thesis required the author to process more than 10 million records from individuals living in Mexico and the U.S. provided by both the American and Mexican census bureaus. Operational definitions for the concepts of multicultural skills and transfronteridad were developed. Multicultural skills were split into three variables: a) Language Proficiency, b) Multiculturality, and c) Educational Attainment. Results showed that Language Proficiency and Multiculturality are statistically correlated to transfronteridad while Educational Attainment was not. Finally, a Supervisor Machine Learning Regression model was built to predict levels of

transfronteridad based on more than ten variables. Results from this exercise showed that while a statistically significant model was built, the prediction power was not strong enough to obtain relevant results or predict potential values for transfronteridad.

Acknowledgments

Thanks to all the people who made this document possible. Thanks to my wife, Nallely Espitia, who was always there to bring me back to my happy place when I was having a hard time balancing the commitments from work, study, personal life, and more. Thanks to my cousin, Marco Olivas, who answered my questions and provided tutoring countless times. Thanks to my mentors, Agustín Sandéz, Nicolás Velázquez, Jorge Gaytán, and Gerardo Coutiño for all the lessons that you taught me and were critical to get me here. And special thanks to my thesis director, Mtro. César Robledo, who provided several hours of support to ensure I was able to complete this work and kept his door always open for me to reach out.

Dedication

A mi madre,
Velia Alfaro,
porque hasta el día de hoy sigo cosechando todo lo que ella sembró con tanto amor.

A mi esposa,
Nallely Espitia,
porque la energía que irradias es la que me permite vivir la vida al máximo.

Chapter 1. Introduction

1.1 Chapter one introductory comments

In this chapter, the author provides contextual information to the reader that is critical to understanding the importance of the Mexico-U.S. border dynamics. A review of the state of art is provided in conjunction with a short discussion for different term definitions that are important for this document. This chapter includes a section stating the specific problem identified and the justification to conduct this study. Moreover, a set of research questions are stablished and a small discussion regarding ethical aspects and feasibility for this study is included. Finally, a short description regarding the research design planned to be used is shown.

1.2 Framing the Research Problem

1.2.1. Context

Every day there are thousands of individuals crossing the Mexico – U.S. border for a variety of reasons that may include working, visiting family or friends, studying, and much more. In fact, in 2019, there were an average of more than 530 thousand daily crossings from Mexico to the U.S. border. According to (Ojeda, 2005), there is no other border in the world like the Mexico – U.S. one, the fact that the most powerful country in the world, a first world country is connected to a third world to whom it shares a complicated history has created an endless number of social, political and economic dynamics that have shaped the lives of millions of people at the border and prompted the development of very interesting lines of research.

Alegría (1990) defines international borders as the situation where two countries have an adjacent space with shared and unshared processes simultaneously happening.

Due to their double function of allowing and restricting the movement of people and goods, borders prompt the development of regions with cultural, social, economic, and political dynamics significantly different than the ones occurring at non-border regions (Escamilla, 2019; González, 1988; Herzog, 1990).

Mexico and the U.S. share almost two thousand miles of border and they are not the exception to this phenomenon (U.S. Census Bureau, 2011). Across their border, there are 15 major hubs or transborder regions, which are composed by pairs of border-contiguous cities spread across 97 border-adjacent counties and municipalities (Hernández, 2020; Orraca, 2015). Each of these regions experience cross-border interactions in different cultural, social, economic, and political levels (Herzog, 1990).

The individuals participating frequently in this kind of interactions are called transborders (Vélez-Ibañez 2010 cited by Falcón & Orta 2018). Schiller (2005, p.1) defines transborders as the "people who live their lives across the borders of two or more nation-states, participating in the normative regime, legal and institutional system and political practices of these various states". Public data available shows that as of 2020 the number of transborders residing in the Mexico-U.S. border can be estimated as 4,349,463 (Mexican National Institute of Statistics and Geography [INEGI], 2021 & U.S. Census Bureau, 2020).

Iglesias-Prieto (2004) states that cross-border interactions represent the existence of *transfronteridad*. Transfronteridad can be defined as the interaction, exchange, and dependency one transborder individual has with twos or more countries. Examples of transfronteridad are individuals crossing the border to visit relatives or attend social events (interactions); purchase products or services (exchanges); and work or study.

According to Iglesias-Prieto, it is possible to identify at least four levels of transfronteridad: sporadic, commercial and impersonal (e.g. crossing the border to do Christmas shopping); periodical but non-emotional (e.g., individual crossing the border to fill their car with gas); periodical and emotional (e.g., individual crossing the border weekly to visit their relatives); and daily and emotional (e.g., individual crossing the border daily to go back home where their family waits for them).

1.2.2. Problem Definition

Iglesias-Prieto (2008) states that individuals displaying high levels of transfronteridad will show cultural richness, a complex identity, and a nuanced understanding about the dynamics occurring at the border. In practical terms, this means that transborders with high levels of transfronteridad have a set of skills and knowledge that allows them to move seamlessly between Mexico and the U.S., benefitting from the cultural, social, and economic resources of each country (i.e., *getting the best from both worlds*; Falcón & Orta, 2019; Orraca-Romano, 2019). Examples of transborders enjoying the best from both worlds are a) U.S.-residing transborders crossing to Mexico to receive medical services at a lower cost than in the U.S. or b) Mexico-residing transborders crossing to the U.S. to work and earn higher wages than in Mexico. Fostering transfronteridad then becomes equivalent to fostering the well-being of individuals in the border regions; tapping into the potential of transborders regions as a mechanism to secure increased quality of life among border-region residents then relies on supporting and fostering the development on transborders with high levels transfronteridad (Iglesias-Prieto, 2008; Orraca, 2015; Vega, 2016). Conversely, neglecting transfronteridad,

represents a missed opportunity for governments developing public policy and organizations working with society members.

Existing research demonstrates the important role transfronteridad plays at increasing the quality of life for individuals at the border (Iglesias-Prieto, 2008; Orraca, 2015; Vega, 2016). Still, transfronteridad is a double-edged sword. Governments must be prepared to manage the dynamics that accompany transfronteridad. While transfronteridad increases the overall quality of life in border regions by allowing transborders to benefit from higher wages, lower product costs, higher quality in services and more, it also generates public problems such as macroeconomic imbalances (e.g., lower wages for certain groups, higher level of prices compared to non-border regions, unbalanced impact from public finance requirements on individuals), legal loops or lack of enforceability (e.g., inability to effectively enforce certain requirements such as holding car insurance), polarization in public opinion (e.g., different thoughts about people living and not living as transborders), public health challenges (challenges to stablish and enforce quarantine procedures under a binational region), and more. Governments must be able to acknowledge transfronteridad and develop public policy in two levels: coexistence and symbiosis. The first one implies developing public policy that does not harm transborder populations while also successfully managing the negative impacts from transborder dynamics so that other populations are also not affected negatively. The second level, symbiosis, is actively developing public policies that leverage opportunities derived from the transborder phenomenon; thus, fostering transfronteridad.

This thesis' author believes that these problems can be addressed in a more effective way by tapping into a bigger body of knowledge around border dynamics; hence,

more research about this topic is required. This research study will explore the transfronteridad phenomenon in general and also develop a machine learning supervised model to identify relevant correlation between transfronteridad and a set of variables. Specifically, this study will focus on exploring the concepts of transfronteridad and transborders and identifying whether or not there is a statistically significant correlation between transfronteridad and multicultural competency skills.

1.2.3. Research Questions

1.2.3.1. Main Research Question

To support the need for a bigger body of knowledge regarding transborders and transfronteridad, this research focuses on answering the following question:

What is the relationship between multicultural competency and levels of transfronteridad presented by transborders?

1.2.3.2. Secondary Research Questions

Additionally, this research will also allow the following secondary questions to be answered via a literature review, a statistical descriptive analysis, a statistical inferential analysis, and the development of a machine learning supervised regression model:

- How can we measure transfronteridad among transborders?

- Is it possible to build a supervised learning method based on regression to predict levels of transfronteridad using information related to transborder dynamics?

1.3 Justification

Whether if it is buying a product, selecting a school, or deciding which social events to attend, transborders will consider options in both sides of the border (i.e., Mexico and the U.S.). This situation creates a challenge to policy makers trying to address public

problems that relate to transborder populations or transborder dynamics. Even if policy makers have developed the best public policy or program that can be developed, within the constraints of a given country, it does not remove the option for transborders to opt for whatever it is available in the other country (e.g., a better-paid job in the U.S., a closer-to-the-family neighborhood in Mexico). On top of that, transborders show varied levels of education, language fluency, income, and political engagement, even when controlling for country of residence (González, 1988). This makes the transborder population a challenging one to reach with public policy.

Some of the reasons why the border regions and the transborder dynamics are so relevant are the following. 1) The great number of inhabitants living at the border region. In 2010, more than 19 million people were living in one of the 97 municipalities and counties across the Mexico-U.S. border (Orraca, 2015; U.S. Census Bureau, 2020; [Mexican] National Institute of Statistics and Geography [INEGI], 2020); 2) The magnitude of the cross-border activity. In 2019, more than 194 million crossings were made from Mexico to the U.S. -this more than 531 thousand crossings per day (Bureau of Transportation Statistics [BTS], 2021). 3) The connections they allow to happen. Borders allow people that historically have belonged to the same population to stay connected (Escamilla, 2019). 4) The impact they have on consumers in both borders. Transborder commerce support both border economies and transborders quality of life as they allow an exchange of goods and services between Mexico and U.S. residents (Mungaray-Moctezuma & Calderón, 2015). 5) The opportunity they provide to Mexico residents -with the proper immigration status and the required documentation- to work in the U.S. and earn a more competitive salary than what they could earn in Mexico (Orraca, 2015). And

6) the cultural exchange they allow by letting individuals to become fluent in two languages and more than that, in two cultures, which is common for Mexico-residing students that commute daily to attend classes in the U.S. (Rocha & Orraca, 2018).

1.3.1. Research Importance

This research will allow to understand better the main traits of the transborder population (including demographic, geographic, psychographic, and behavioral characteristics). Transborder dynamics impact considerably the border cities of both countries, shaping them in a different way than non-border cities, hence their importance.

Public policies in both sides of the border could be informed by the results from this research, allowing policy makers to understand the nuances that make this population different than the main population and why policies designed for non-transborder populations might not get traction among transborder ones. Additionally, private industries could get benefited from understanding better the consumption patterns from this population that differ from the non-transborder population in both borders. Increasing the literature available about this population will also allow private companies to build more effective marketing strategies to target this population accurately.

This work will increase the literature about border regions and cross-border dynamics which are different than the acculturation/enculturation existing body of literature, and different than the body of knowledge that has been developed for undocumented migration practices (González, 1988). Even though some principles from these frameworks can be applied to the transmigration phenomenon, it is quite differently since being a transborder identity development process is different than the acculturation

one. Not only that, but this work will also analyze the transborder population across the border as one, which has not been comprehensively addressed by literature in the past.

Finally, this research will introduce an instrument to measure the level of transfronteridad presented by transborder individuals. It is important to mention that this research attempts to be an exploratory one; therefore, more than trying to land in definitive factors to be included in transfronteridad instrument, it will identify factors of importance to be included in more formal iterations of this models.

1.3.2. Research Feasibility

This research project will involve launching a survey for transborder population. For this, the research author has previously identified more than 100 Facebook groups used by transborders that will be used as channels to advertise this project's survey. The author received authorization via Facebook messaging to post survey invites from more than 60 group administrators. In order to develop and launch the survey, a Survey Monkey license has been acquired, allowing the author to create an online survey instrument with several skip patterns and logic to enhance survey respondents user experience. In addition to this, the data analysis and machine learning model development will require access to a computer that is able to process data using R and SPSS, which is already available.

1.4. Method

This research will follow a quantitative approach with a survey that will be deployed to collect data. Even though different statistical methods will be implemented to describe and infer traits from the population, it will all be done under an exploratory approach. The author will pull data at a given point in time and will not give any kind of treatment to the

individuals involved in it; therefore, this study can be categorized as a non-experimental, cross-sectional (Hernández, 2014). The author will use the following methods to analyze the population of interest:

- Frequencies and descriptive statistics to characterize the transborder population.

- Regression analyses to identify correlations and incidence ratios for different traits in our population.

- Machine supervised learning models to build multivariable model to predict transfronteridad.

1.5. Chapters Description

The first chapter of this thesis provides the reader an overview of the current border landscape. The author provides context regarding border dynamics occurring on a daily basis at the Mexico-U.S. borderlands. Additionally, this document also specifies the problem definition and the research questions that were identified in the early stages of this projects and that were used to develop the research design for this project. Finally, a justification is provided, including narrative highlighting the importance and feasibility for this research project.

The second chapter of this thesis relates to the theoretical framework guiding this research. The writer conducted a comprehensive literature review that involved more than 30 authors and presents the results from these efforts under this section. In this chapter, the document includes provides a review of the state of art regarding border theory and Machine Learning methods available. Additionally, it provides information regarding border theory concepts used in this document; transborder groups definitions; an

assessment of the current border landscape; a historic review of the Mexico-U.S. borderscape and the relationships between these two countries; information related to transborder living at the border; a definition and description for the constructs of multicultural competency and transfronteridad, and a description of various machine learning methods relevant to this study.

In the third chapter, the author provides information related to the method used to pursue the research objectives from this thesis. The author includes information on the research objectives; the hypotheses identified; the methodological design; the population studied; some ethical considerations; an instrument description; a description of the field activities; and a description of the data process followed to analyze the information collected.

The fourth chapter relates to the results from the collection and processing stages in this research project; it provides descriptive tables, crosstabs, and regression output to present the results obtained after processing the data collected.

In the fifth chapter, the author discusses the results found, connecting them to different pieces of the literature reviewed. Additionally, he provides a set of conclusions, recommendations and also highlights a ser of new lines of research future project could look into. Finally, the author provides information related to the application of the results obtained, the contributions made to the society and the field, and strengths, weaknesses, opportunities, and threats identified during the research process.

In the remaining sections of this document, the writer provides samples of the instruments used to collect information from respondents, a glossary, a crosswalk showing how different concepts learned during the master program were applied in this

document, the results from the Turnitin Analysis and the syntax used for all the processing and analysis made.

1.6. Chapter one closing comments

In this chapter, the author provided information to learn the importance of the topic studied. While unknown to many individuals, the magnitude of the transborder dynamics impact the lives of millions of inhabitants living in the borderlands; thus, increasing the knowledge regarding this topic is in benefit of society. Moreover, based on recent literature, embracing or neglecting these dynamics at both the macro and micro levels can impact the levels of well-being for populations living in the border. Due to this situation, the author identified a set of research questions to be answered in relation to the transborder dynamics, specifically about transfronteridad. In order to answer these questions, the author will be using a quantitative research design composed by descriptive and inferential statistical analyses and also through the development of a set of supervisor machine learning models based on regression analysis to build predictive models around the concept of transfronteridad that can be useful to inform public policy development.

Chapter 2. Theoretical Framework

2.1 Chapter two introductory comments

The theoretical framework is a critical section from this document. This chapter illustrates various concepts related to transborder dynamics, including the terms transmigration, transfronteridad, border, and more. Additionally, in this concept, terms related to Machine Learning models and Big Data are discussed based on a literature review performed by the author. In the first section from this chapter, a quick review of the state of the art is provided with concepts from more than 10 authors and recent works. The objective of the theoretical framework is to allow the reader to understand better relevant concepts for this research as well as methods and techniques used to process the information needed to answer the research questions defined.

2.2. State of the Art

2.2.1. Overview

So far, the literature reviewed shows that past work on this topic has been focused on describing transborders either as a homogeneous group or an heterogenous one with subsets defined by the transborder's main reason to cross the border (e.g., work, school, shopping). The body of existing literature speaks about work transborders, which are individuals that cross the border frequently for work purposes (Arámburo, 1987; Fimbres, 2000; González, 1988; Orraca, 2015; Vega, 2016); student transborders, who cross the border almost daily to attend classes (Falcón & Orta, 2018; Rocha & Orraca, 2018; Vega, 2016); and consumer transborders, which are the ones crossing the U.S. border with the main purpose of purchasing lower priced or higher-quality goods and services (Mungaray-

Moctezuma & Calderón, 2015); thus, the existing research has analyzed the transborder population considering their demographic, geographic and behavioral traits (see table 1).

Recent literature regarding data science and big-data methods to analyze the information speaks about a great availability of techniques available to perform this type of analysis. Authors advise caution when selecting the best methods to follow; while there may be a variety of options to opt for when trying to build knowledge from a certain array of data, the analyst must take into consideration factor such as the nature of the data being analyzed, the expected outcome of the analysis, the applications, and more. In general, it possible to create different classifications to organize all the methods available, a simple one could consist of method related to classification, regression, and clustering (Vadim, 2018).

Table 1

State of Art

Reference	Type Quantitative	Qualitative	Relationship to thesis	Transborders	Border	Migration	Commuters	Transfronteridad	Regression	Big Data	Machine Learning	Superviser Learning	Predictive modeling
Falcón, V., & Orta, A. (2018).		X	Information needed to understand better transborder students' dynamics at the border	X	X		X	X					
Calzada, E., Covas, M., Ramirez, D., Miller, L., & Huan, K.-Y. (2016).	X		Information needed to understand applications that cultural adaptation theory has on transborder dynamics and relationship with border theory concepts.			X			X				
Rocha, D., & Orraca-Romano, P. (2018).	X		Information needed to understand better transborder students' dynamics at the border	X			X	X	X				
Orraca-Romano, P. (2019).	X		Information needed to understand better transborder workers' dynamics at the border	X			X	X	X				
Tapia, L. (2017).		X	Information needed to understand better transborder dynamics at the border	X	X			X					
Escamilla, E. (2019).		X	Information needed to develop an operational definition for the concept of Transfronteridad.	X		X		X					
Vega, G. (2016).	X		Information needed to understand better transborder workers' dynamics.	X			X						
Vargas, V., & Coubès, M. L. (2017).	X		Information regarding factors impacting transborder dynamics.	X	X								
Orraca, P., Rocha, D., & Vargas, E. (2017).	X		Information regarding factors impacting transborder dynamics.	X									
Hernández, H. A. (2020).		X	Information needed to understand better transborder dynamics at the border	X	X	X		X					
Vadim, K. (2018).	X		Information needed to understand better various data mining techniques useful to analyze big data.							X			X
Kosinski, M., & Behrend, T. (2017).	X		Information needed to understand better various data mining techniques useful to analyze big data.							X			X
Alzubi, J., Nayyar, A., & Kumar, A. (2018).	X		Information needed to understand better various techniques to build Machine Learning models.								X	X	X
Lu, C., Lin, G., Wu, T., Hu, I., & Chang, Y. (2021).	X		Information regarding using machine learning methods to analyze information from cross border dynamics.						X	X	X	X	X
Sharma, S., Kang, D., de Oca, J., & Mudgal, A. (2021).	X		Information regarding using machine learning methods to analyze information from cross border dynamics.						X		X	X	X

Note. For a detailed descriptive tables of the state of art, refer to appendix D.

Source. Author's elaboration.

2.2.2. Related Work Description

2.2.2.1. The Transborder Identity Formation Process: An Exploratory Grounded Theory Study of Transfronterizo College Students from the San Diego – Tijuana Border Region.

In this study from (Falcón & Orta, 2018), they used a grounded theory approach to explore the identity development process for undergraduate students living transborder lifestyles in the San Diego – Tijuana border region. Authors highlighted the increased prevalence of commuting students in this border in comparison to the past. In order to collect information, the authors went through a series of interviews with 12 transborder students living in Mexico while studying in the U.S. and found that much of the transborder identity is develop or influenced by factors associated with each person's obstacles faced in their day-to-day transborder activities. Results from this study show interview participants crossing the border every day using the pedestrian ports of entry that did not identify themselves with individuals living transborder lifestyles that did not involve this practice (e.g., students from their school visiting Tijuana often but only for fun or visiting family). The authors closed this research exercise providing a conceptual model that explains the transborder identity development process. The importance of this research for the author's study resides in the use of identity concepts that help to explain the development process and day-to-day dynamics of transborder individuals, specially studying commuters.

2.2.2.2. A longitudinal study of cultural adaptation among Mexican and Dominican immigrant women.

In this study, (Calzada, Covas, Ramirez, Miller, & Huan, 2016) explore the differences in the cultural adaptation process shown by immigrant women from different origins. Authors used the cultural adaptation framework as opposed to assimilation theories considering that immigrant individuals' cultural adaptation process is not a linear one that goes from unacculturated to acculturated but a bidimensional one in which the will experience different things and reconstruct their identity based on acculturation and enculturation process. In this context, authors understand acculturation as the process in which an immigrant individual gets exposed to a new country, usually the mainstream one in the country they have arrived; and enculturation as the identity individuals build around the first culture. Depending on their experiences -and several other factors, individuals may fall in one of four categories represented by the intersection of acculturation and enculturation: 1) Integration where individuals have successfully achieved acculturation while keeping their first culture -enculturation, 2) Assimilation where individuals successfully achieved acculturation while disengaging with their first culture, 3) Separation where individuals do not achieve acculturation while keeping their first culture intact, and 4) marginalization in which individuals disengage from both the mainstream culture in the new country and their first culture to build an identity that may take some aspects of none from the two cultures. The authors collected information using a longitudinal study design in which they examined variables related to cultural adaptions such as language competency, identity, and cultural knowledge. The one-year longitudinal study was conducted in the U.S., specifically in New Yor City, including

immigrant Latinx women from Mexico and Dominican Republic, and found that poverty and the condition of living in immigrant-dense neighborhood can be correlated with reduced acculturation levels for both Mexican and Dominican women. Additionally, the results also showed that cultural adaptation is a complex construct affected by individual and cultural contexts. The importance of this study relies in the interest from the author to identify the level of cultural adaptation transborder individuals phase when living transborder lifestyles, with results showing that this aspect is highly conditioned by the specific dynamics from each transborder regions; while some border cities from the U.S. are composed in its majority by Hispanic population (e.g., Calexico and San Luis A.Z.), others show a more diverse composition such as San Diego impacting the number of situations in which transborder are exposed to cultural shock experiences that form their identities. Additionally, it was important to understand that while immigrant individuals go through a cultural adaptation process in which there are two clear cultures playing a role, transborder cultural processes and identity formation are related but not exactly the same due to the role of "ni de aqui ni de alla" -or *otherness*- and other concepts that present a different experience compared to living outside of the borderlands.

2.2.2.3. Estudiantes de educación superior transfronterizos: Residir en México y estudiar en Estados Unidos.

In their study, (Rocha & Orraca-Romano, 2018) the authors explore the transborder dynamics from students studying in the U.S. while residing in Mexico. Authors collected student information from the Intercensal Survey 2015 from the Mexican Census Bureau (INEGI), specifically from 9 Mexican border municipalities (Tijuana, Mexicali, San Luis Rio Colorado, Nogales, Ciudad Juarez, Ciudad Acuña and Piedras Negras,

Matamoros, Nuevo Laredo, and Reynosa). Authors pulled their sociodemographic data and used descriptive methods to generate discussion about students commuting to school Mexico – U.S. Part of the highlights from the authors include information regarding a difference in the opportunities from students studying in Mexico vs in the US.; specifically, they mentioned that while the opportunity to study in the U.S. is consider a great one, individuals, even those with U.S. citizenship living in Mexico, may not be able to pursue due to their income level or the lack of family members and other support nets in the US. The authors also highlighted -just like (Falcón & Orta, 2018), the facts that while students living this situation may all be considered transborders, the experience will be completely different to the students crossing the border to get to school every day versus those crossing back and forth for pleasure. The important of this study for this paper relies in the discussion about struggles transborder can experience when it comes to accessing education in the U.S. while living in Mexico.

2.2.2.4. Orraca-Romano (2019). Cross-Border Earning of Mexican Workers Across the US-Mexico Border.

In this study, (Orraca-Romano, 2019), the author analyzed data from the Mexican Census Bureau, using a longitudinal approach and describing the changes in magnitude for the transborder population living in Mexico while working in the U.S. The author examined earning from cross-border and non-cross border workers highlighting that in general, there was a decline in the number of cross-border working from 2000 to 2010 and an increase from 20210 to 2015. Additionally, the author found that cross-border individuals earned nearly double that non-cross-border employees when controlling for skill levels with differences being accentuated by gender and among the low-educated.

The author concludes that human capital levels do not condition the earning from workers in general as much as the condition of being a cross-border worker due to the different in the productivity between the two labor markets. Another finding from the study relates to the cross-border workers skewing more male, married, and salaried, and also working shorter workweeks. The importance of this study for this paper relies on the descriptions of work commuters that are one of the main groups analyzed in this study.

2.2.2.5. Borders, mobility and the transborder space: Reflections for a discussion.

In this study, Tapia (2017), highlights the importance of following a transdisciplinary approach to analyze transborder dynamics; history is key to understand border-related concepts. The author also mentions the importance of understanding how any phenomenon happening in the borderlands cannot be automatically categorized as part of transborder dynamics (i.e., transfronteridad does not come automatically or by definition). The author advises to consider the magnitude of transborder exchange, geographical proximity and the population size before labeling something as part of transfronteridad. The author's study is based on a literature review and posterior reflections to provide conclusions that will add up to the debate. The importance of the study relies in the discussion or distinction regarding things happening in the borderlands not necessarily being considered part of transfronteridad which is key to the discussion of our study.

2.2.2.6. Transfronteridad and Everyday Commuter Negotiations of Tijuana's Borderscape.

The author, (Escamilla, 2019), goes through an exploration of personal narratives from individuals crossing the border frequently to explain transborder demographics and concept of transfronteridad -key to this study. The author conducted a series of ethnographies from individuals crossing the Tijuana-San Diego border to understand their day-to-day activities, challenges, and experiences. Part of this study provides information related to the border crossing process and what they call narratives of crossing from commuters and their borderland activities. The authors explain how day to day commuting allows individuals to develop transfronteridad which is described as the "cultural practices and expressions of identity that are adopted and maintained by means of routinized transborder movements" (37). The importance of this study relies on all the information provided about transfronteridad since it allowed to understand the direction required to build a construct for this concept.

2.2.2.7. Commuter Population on the Northern Border: The Case of Mexicali-Calexico and Tijuana-San Diego.

The central objective of this study, (Vega, 2016), was to provide a description of the sociodemographic traits from transborder population using information from non-cross-borders as a benchmark, combining specifically work commuters living in Mexicali and Tijuana while working in Calexico and San Diego. Additionally, the author crossed the findings from this analysis with the results from a qualitative study, interviews to complement the analysis. The information used for this analysis comes from the 2010 housing and people census from the Mexican census bureau. The importance of this

study relies on providing sociodemographic information about the population of interest for this study.

2.2.2.8. Working and giving birth in the United Sates: Changing Strategies of Transborder Life in the North of Mexico.

The study, (Vargas & Coubès, 2017), analyze changes in two common activities among transborders living in Mexico; working and giving birth. Using data from the Mexican Census, specifically from the housing and people censuses of 2000 and 2010, the authors found differences in the access to both activities that varied by level of education. Additionally, findings include an observed decrease in the number of cross border workers while an observed increase in the cross-border births reported. Authors conclude the study by stating that evidence allows to think that the border has become a more selective barriers allowing more access to certain groups that other based on age, educational level, and others. The importance of this study resides in their discussion related to two very common practices from the transborder lifestyle, working cross border and cross-border birth.

2.2.2.9. Estudiantes de educación superior transfronterizos: Residir en México y estudiar en Estados Unidos

In their paper, (Rocha & Orraca-Romano, 2018), the authors explored potentials factors associated with individuals opting for following transborder lifestyles focused on commuting to school in the U.S. from Mexico. Authors created and inferential statistical model to estimate the probability of individuals to adopt this practice based on different sociodemographic information available in the 2015 Intercensal Survey from INEGI. In this document it was highlighted that most of the students going through this process are

U.S. born and living in Tijuana or Ciudad Juarez. The results from the study showed that there is a significant statistical correlation between a higher probability of being a cross-border student with age, household income, and also having someone in the household living a work commuting lifestyle. Authors expressed that an individual may be interested to go back and forth between border to study due to an increased quality level in education and interest in eventually transitioning to the U.S. labor market. The important of this study has to do with the author's interest of understanding factors that impact the transborder lifestyle and practices that are recurrent among them.

2.2.2.10. The Mexico-U.S. borders: asymmetry and transgressions.

In their study, (Hernández H. A., 2020), the author provides important information related to the current status in the border, including historical events that have impacted the border relationships between both countries such as 9/11 and the COVID-19 pandemic. The author also describes some of the functioning mechanisms at the border such as the types of lanes individuals can use to cross and the distinction between security levels when crossing from Mexico to the U.S. and the other way around. The author also provides information related to the extension of the border and a brief description regarding the number of U.S. countries and Mexican municipalities participating in border dynamics. Additionally, the author highlights the existence of 15 pair of border-adjacent twin sisters such as Mexicali-Calexico and Tijuana-San Diego. The importance of this study for this document relies in the fact that provides an overview of the current state of border dynamics and a brief introduction to very common aspects of the border that individuals must be aware if interested in learning more about the topic.

2.2.2.11. Overview of different approaches to solving problems of data mining

In this study, (Vadim, 2018), provides and overview of different approaches available to solve problems of data mining. In its introduction, this paper highlights the existence of different methods to solve problems including classification, regression, and clustering. The analyst will need to select the best method to solve a particular problem based on the characteristics of the data, the phenomenon being studied and other factors. According to the document, data mining problems can be solved in two ways, using a supervised learning approach or and unsupervised one. Both approaches come from Machine Learning which is the concept that groups all these techniques into one. Supervised learning approaches consist in model where there is data to judge the quality of a model, the quality of a prediction, in this way , the user will be able to discern is a particular model provided reliable results or not. It is important to keep in mind that all the supervised methods are iterative since the user will need to go back and forth changing different parameters in the model, including, or excluding variables, until the model hits the maximum quality possible. Usually, the problems that can be resolved with supervised models can be divided in classification and regression problems. Conversely, unsupervised learning model are mostly focused in exploring datasets due to the inexistence of classification variables that would allow to test the accuracy of models. The importance of this paper to the study relies in providing concepts required to conduct machine learning analysis using the data obtained from the survey.

2.2.2.12. Big data in the behavioral sciences.

In their study, (Kosinski & Behrend, 2017), the authors describe different applications that Big Data has in the behavioral sciences. It is important to highlight that in terms of societal applications, such as the contents of this paper, the authors mentioned that approached based on big data methods allow to gather human behavior insights that can improve society, institutions, and people's lives. The authors warn that there is still more to improve in terms of privacy as many rules have not yet been defined leaving data in gray areas moving from two extremes in which on one side, user data may fall in the wrong hands of enterprises may be taking too far the expression "if you do not pay for a service, then you are the product, not the customer" (125). On the other extreme, tasks such as collecting informed consents and deidentifying data can make individuals struggle to get to a point in which they can use data. Finally, the authors brough up the important of being aware of the risk of overfitting and misapplications from big-data-driven models. Which can happen when there is not enough fundamental theory behind the analytical developments inside big data models. The importance from this paper relies in understanding cautions analysts must have when working with big data.

2.2.2.13. Machine learning from theory to algorithms - an overview.

This paper from (Alzubi, Nayyar, & Kumar, 2018), provides an overview of Machine Learning. Authors start with a descriptions of machine learning highlighting that is a multi-disciplinary field focused on solving problems such as classification, anomaly detection, regression, clustering, and reinforcement problems. The authors provide and historical overview of the developments on the fields from 1950 all the way to 2017. Part of the document includes a description of the generic Machine Learning model that consists of

six steps : 1) collecting and preparing data, 2) selecting features or picking variables, 3) choosing an algorithm, 4) selecting models and parameters, , 5) training the models, and 6) evaluation the performance of the model. It is important to emphasize the iterative aspect of these steps though the authors do not mention this in their document. The importance of this document relies of providing information related to different machine learning methods, including regression analysis which will be used in this document.

2.2.2.14. Influencing Factors of Cross-Border E-Commerce Consumer Purchase Intention Based on Wireless Network and Machine Learning.

In their paper, (Lu, Lin, Wu, Hu, & Chang, 2021) created a machine learning model to predict consumption frequency of e-commerce cross border consumers. In order to build this model, the authors created a survey with items related to factors theoretically associated with consumptions patters and fielded the instrument in China collecting one hundred responses. An inferential statistical approach was used for this study and the results showed that individuals see their cross-border consumption frequency affected by the products perceived usefulness, perceived ease of use, consumers' income level, education attainments, age, gender, service, and safety index. This paper was important to this study even though its data relates to transborder dynamics occurring in a different continent as the principles used regarding consumptions patterns are applicable to individuals crossing the Mexico -U.S. border to acquire products and services at a lower price or with higher quality.

2.2.2.15. Machine learning methods for commercial vehicle wait time prediction at a border crossing

In this study, (Sharma, Kang, de Oca, & Mudgal, 2021) present the results of a machine learning created to estimate the crossing border times for the Mexico – U.S. commercial borders. In order to do this, the authors used different Machine Leaning methods including Gradient Boosting Regression, and Random Forest regression methods. The results were not conclusive due to the high variability of the data creating non-reliable predictions. Authors concluded that will the results were not conclusive, the information gathered will help to understand the performance from this type of models better. This paper is important to the document as it presents one example of Machine Learning applications developed to explore and solve cross-border problems occurring in the Mexico-US. Border.

2.3. Border Theory Concepts

2.3.1. Border

While there are several definitions available for the concept of border (see table 2), most of them state that borders are, a) physical barriers, b) allow countries to exercise their sovereignty, c) accomplish a double function , and d) while being physical barriers, they will not stop politic, cultural, social, and economic exchange from happening.

The border, being a physical blockade, not only separates. Escamilla (2019) explains that this means borders are porous. Likewise, Duarte-Herrera, (2001), states that the border play a double role stopping what is foreign from what is not. Still, the border lets go through goods, services and cultural traits that would be beneficial for the nation (142). Borders allow the interactions and cultural exchange between the inhabitants of

both sides of them; accomplishing -as stated before- two opposite tasks, separation, and unification of regions (González, 1988). While the first task is developed to protect the sovereignty of countries -inevitably creating discontinuities, the second one allows cultural connections to be made while exchange and interaction processes are happening (Mungaray-Moctezuma & Calderón, 2015).

The existing literature provides different examples on how borders, particularly the Mexico-U.S. one, simultaneously separate and unify. Escamilla speaks about the high number of Mexican families distributed across both borders -which in this case gives the border a role of joining and separating families (2019). Goldberg (2001) describes in their paper life stories from people living at the U.S.-Mexico border highlighting how people might interpret the border in quite different ways. While some people see the border almost as invisible, -as if American border cities were just an extension of the Mexican border cities, other people find very hard to avoid noticing the differences between the two places.

Another question that authors reflect on is whether the border separates more than what it unifies. When describing the borderscape, Alegria (1990) states that the U.S. border moved from being a place that allowed almost-free crossing in the past century, to a place with a legal break or stop for these and other populations. As a result of this division, cities across the border stopped being similar to developing their own political, economic, social, cultural, and urban structures with some borders showing more

dissimilitude than others. Based on this, the author can say that at least compared to the past century, the border is now separating more than in the past[1].

Table 2 (1/2)

Border Definitions Available in Literature

Author	Definition
(Escamilla, 2019 , p. 3)	"The borders are popularly understood primarily as physical blockades that control the diffusion of goods and people from side to side. They are also gateways between the north and south through which people, goods, and cultural practices travel…Despite the securitization measures, the borders are extremely porous and the interactions of the inhabitants of both sides are an indication of the limits of a nation-state border".
(González, 1988, p. 281)	[The border] "constitutes a mean to instrument state policies of restriction in the economic, demographic, and political dimensions, adjusting said policies to the internal and international interest of the country" [adapted from Spanish].

---

[1] Martínez (1994) states that the relationship between Mexico and the U.S. at their border have moved through three stages "Alienation" (1560 – 1880), "Coexistence" (1880 – 1920), and "Interdependence" (1920 – Present). Following this line of thought, Alegria's statement (1990) should be interpreted paying special attention to **century**; this means that the border is more separating than in past century, but not beyond that.

Table 2 (2/2)

Border Definitions Available in Literature

Author	Definition
(Martínez, 1994, p. 5)	"A border is a line that separates one nation from another, or, in the case of internal entities, one province or locality from another. The essential functions of a border are to keep people in their own space and to prevent, control, or regulate interactions among them".
(Mungaray-Moctezuma & Calderón, 2015, p. 200)	"The border can be considered a physical or administrative barrier that breaks the continuity for goods and people influx, regulating their transit more or less, either making the influx happen in an easier way, filtering it, transforming it, pivoting it to other places or even preventing it from happening" [adapted from Spanish].

Source. Author's elaboration.

2.3.2. Levels of Borderland Interaction

Borderlands can be distinguished from the interior zones of other countries because of the borderlands milieu, which is described by Martinez (1994, pág. 10) as "the unique forces, processes, and characteristics that set borderland apart from interior

zones", this includes transnational interactions, international conflict and accommodation, ethnic conflict, and accommodation, and separateness".

Martínez (1994) says that there are four models of borderland interaction; 1) alienated borderlands, 2) coexistent borderlands, 3) interdependent borderlands, and 4) integrated borderland. The Mexico – U.S. borderland has gone through most of these stages throughout the years. An alienated borderland characterizes by absent or very limited interactions between the two countries sharing a border. There is tension and borderlander from each country will have few interactions if any. During this time, Martínez considers that the border was in a model of alienated borderland that would have been characterized with a very low to nonexistent interchange. The author considers this state started since 1560 but could consider from the birth of Mexico as a country and lasted until two generations after the Texas rebellion.

Coexistent borderlands show stability intermittently and allow very limited interaction and only borderlanders are able to develop relationships with their counterparts that go further that acquaintance. Martínez state that the Mexico and the U.S. border went through a coexistence level of interactions between 1880 and 1920; a period characterized with an increasing level of law enforcement in the border, a flow of U.S. capital into Mexico, the development of different borderland cities and Mexican migration to the United States. Additionally, the author mentions that during the Mexican Revolution, there was increased anti-Americanism in Mexico.

Interdependent borderlands show consistent stability. There is economic and social cross-border interaction with borderlander carrying meaningful relationships between borders. For Martínez, the Mexico – U.S. border has experienced

interdependence interactions from 1920 to the present; examples of these interdependence are the NAFTA agreement, border rectification agreements, water treaties, expansion of border interactions, including industrialization, trade, tourism, and migration. The growth of transborder regions such as the Tijuana – San Diego, El Paso – Cd Juarez are other examples of this interdependence.

Finally, integrated borderlands show permanent and strong interactions. For these types of borderlands, their economies are well synchronized and the movement between people and good is not restricted. Individuals that belong to either country see each other as part of only one social system.

2.3.3. Transmigration

According to Gonzalez (1988), transmigration is a migratory phenomenon where there is a constant migration flux between two countries. Duarte-Herrera (2001) states that transmigration is experienced everyday as a circular movement done by work and student commuters but also individuals crossing the border to visit family and friends, and also those looking for medical, religious, and recreational services. Transmigration then cannot only be linked to one specific activity such as working or studying. Still, the two major transmigration dynamics happening at the border do relate to work and study: 1) individuals crossing through the U.S. border during specific times of the year to participate

in economic activities (e.g., working as field workers during cropping seasons) and 2) individuals crossing through the U.S. border daily to work or study (i.e., commuters[2]).

According to Alegría (1990), a transmigrant is "the person working on one side of the border while living in the other, crossing the international border continuously in home-work trips. He also mentions that transmigration, understood only as transborder working at the other side of their respective border is something happening from Mexico to the U.S. considerably more than the other way around due to the wage salary differences; still, the author mentions that, while not happening that frequently, there are individuals earning higher incomes in the Mexican side as well, such as managers and specialized technicians working in the maquila industry may on the Mexican side that opt to live in the U.S. for different reasons.

2.3.4. Transborder.

Borders were originally created to prevent people -and things- going from one place to another without regulations; they were made in a way that they would create barriers between cities. However, the reality is that borders are more complex than that. Even though there are processes regulating the entry of people and things, the borders are more porous than they seem, and this can be observed on the thousands of people going back and forth between borders every day or in the impact that borders have on cities sharing them, creating a symbiotic relationship. In the middle of this dynamic are

---

[2] The commuter term has been discussed since the 60's by both Mexico and the U.S. The term by itself has not been recognized in the immigration laws yet is commonly used to describe individuals crossing the border on a daily basis (González, 1988).

the people interested in doing this country-to-country transition for different reasons including work, entertainment, and family. The literature calls this group of people transfronterizo or transborder. Withing this group, different subgroups can be identified such as commuter and cross-border consumers (Martínez, 1994). The first groups refer to the population that makes very frequent trips to the U.S. with the intent of working or receiving educations services. The second group refers to people from both countries that enjoy going back and forth between countries to buy products, receive services, visit family members, among other things.

Goldberg (2001), makes an important distinction between borderlanders and transborders. Borderlanders, which are people living in border cities or regions are not necessarily transborders; different traits should be there such as transnational interaction, ethnic conflict, and accommodation. This means, for example, that just because an individual lives in Tijuana, this borderlander is not necessarily a transborder. Additionally, just because an individual holds a visa or a U.S. residence card, that does not make them part of the group, they need to have the interactions across both borders, the conflict of joy of moving between cultures, the sentiments of confusion or enjoyment of belonging to two places of constantly changing your chip based on the situation (Escamilla, 2019).

On the distinction between borderlanders and transborder, Martínez (1994) states that borderlanders are people with "opportunities unavailable to people from the heartland areas... Through exposure to transnational interactions and transculturation, bordelanders are able to develop versatility in the human relationships, and access to a foreign economy increases employment possibilities and consumer choices" (cited by Escamilla 2019, p.25). It is important to highlight that there is literature stating borderlands

and transborder as synonyms and also other stating the latter is a subset of the first one. For the purposes of this study, transborders will be considered a subset from borderlanders.

One of the characteristics from people living in the borderlands in the separateness feeling. According to Martinez, individuals living in the borderlands have a complex and ambiguous identity due to constantly felling pulls from two directions. While the may get to experience the culture from their interior zones, there is no doubt that the country on the other side of the border puts them in a situation where a borderland identity gets developed. Aspects such as a weakened nationalism are part of their identity considering that they participate in the legal and day-to-day activities from two countries. This is commonly criticized by people from the interior zones of countries. Things like "being culturally and morally corrupt", *agringamiento*, addiction to foreign products, and an increased level of acceptation for both illegal activities and people and goods contraband are not well seen by people living in the interior of both countries (1994, p. 22) while they are potentially part of the daily interactions of a borderlander.

2.3.5 The Difference between Immigrants and Transborders.

While there are different frameworks used to analyze the distinct types of migration, this document focuses on the difference between one-time and circular migration. While one time migration could be the case for individuals who crossed through the U.S. border once and changed their residency to this country permanently, circular migration could be divided into two subgroups: infrequent and frequent. Infrequent circular migration is the ones that happens at specific times of the year with individuals crossing the border to participate in specific economic activities. Conversely, frequent migration is

the one that occurs almost on a daily basis with individuals going back and forth for work or study. Based on these characteristics, transmigration can be considered a subset of migration (see figure 1).

Figure 1

Types of Migration
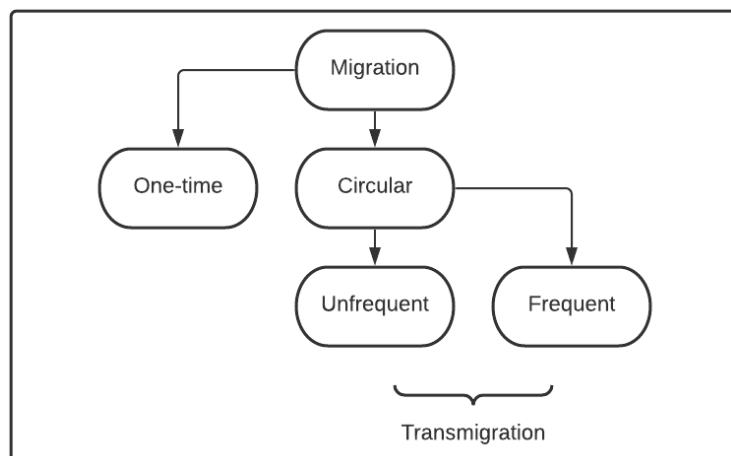
Source. Author's elaboration based on Gonzalez (1988).

Orraca (2015) points out important information that is useful to distinguish between one time and circular-infrequent immigrants and transborders. First, it is important to mention that while immigrants change their country of origin to live in the U.S. (in this case from Mexico), cross-border workers go back and forth between countries. The author mentions that sometimes the condition of staying as a cross-border worker or becoming an immigrate will depend on the degree of success that an individual achieves in the U.S. According to the results of their study, this author highlights that, compared to cross-border workers, immigrants are younger, more educated, have fewer children, have higher earnings, and work less hours per week. The data analyzed by the author shows that cross-border workers are more likely to be homeowners than their counterparts, probably due to the lower cost of housing in Mexico. The author also believes cross-

borders might decide to not immigrate permanently to cope with increased household's expenses due to having more children. In general, the author describes the cross-border group as privileged when compared to the immigrant group. The author points out how the increase of number of immigrants in the U.S. contrasts with the numbers of cross-border works which can be tied with circular migration.

2.3.6. Transfronteridad

Iglesias states that just because there is geographic vicinity, it does not mean that two border cities, counties, or municipalities will interact with each other at a high level. This will depend on the level of transfronteridad existent. The author defines transfronteridad as "the levels of exchange, dependency between one and the other side, to the quantity or frequency, the intensity, direction of the material and symbolic exchange scale, the sense of belonging to both sides as well as the social and cultural meaning the individuals makes of their border crossings and border interactions" (2010, p. 182). Similarly, Escamilla (2019 , p. 37) defines the concept of transfronteridad as "the cultural practices and expressions of identity that are adopted and maintained by means of routinized transborder movements".

Transfronteridad is experienced in different ways by transborders, for example, for some it might be experienced via code switching, a common practice among bicultural transborders consistent in switching the language, behavior, expressions to adapt to different situations that they face due to living in the borderscape -or borderlands. The author also mentions that transfronteridad can be either embraced or neglected. Several people decide to neglect it due to the complexity of balancing a transborder lifestyle -it could be seen as a privilege or as a burden, which means making a conscious or

unconscious decision to not cultivate the abilities, skills, knowledge, or willingness

needed to make the most of the daily negotiations between borders (Escamilla, 2019).

   2.3.6.1. Levels and determinants for transfronteridad.

Iglesias-Prieto (2004; 182) provides guidance to measure transfronteridad

exemplifying a series of day-to-day micro interactions but also binational relationships

between countries at the macro level. According to the author, the more transfronteridad

there is, the more interaction and compromise with the other side of the border. The

author establishes different levels of transfronteridad (see table 3). In the first and lower

level, individuals have cross-border interactions that are sporadic, transactional; these

interactions do not get further from a vendor-client relationship. Examples of this level of

interaction can be seen with individuals crossing to Tijuana to have a meal at *Plaza*

Chapultepec or to *Avenida Revolución* to visit bars. On the other side, it could mean

individuals crossing to Brownsville to buy clothes at *Ross* or getting a burger at

Whataburger. The second level of interaction is based on periodic cross border

interactions with dynamics that are not necessarily emotive or sentimental. Examples of

this level of interaction are individuals periodically crossing the border to receive medical

services, filling up fuel tanks or purchasing groceries. A third level of interaction happens

when there is sentiment behind the interactions -warm and emotive interactions as

Iglesias defines them. Examples of this level of interaction are visits for family events such

as weddings, funerals and more. It is important to highlight that individual at this level of

transfronteridad will also participate in interactions from the first two levels, meaning that

if an individual feels comfortable visiting their family at the other side of the border

periodically, more likely than not, the individuals will also participate in the activities

described for levels one and two -same for individuals engaged in level two. At this level or interactions, individuals can be considered knowledgeable of the urban space from both countries; however, this level of interaction is more common from Mexican American individuals than from Mexican or Anglo individuals. Finally, for the fourth level -the true transborder for Iglesias-Prieto, individuals, in addition to showing the interactions previously described, are also participating in emotive, intensive, and varied interactions between borders. Examples of interactions at this level are those held by individuals with properties or businesses in both countries, bilinguals, with family and social ties in both places fully engaging in social, economic, political, and other dynamics from both countries.

The author also explains that cultural richness is correlated with higher levels of transfronteridad. Usually, the individuals who are able to participate in interactions from the highest bracket of transfronteridad have double citizenship, are bilingual and bicultural, and have lived, studied, or worked in both countries. These individuals cross the border frequently, even on a daily basis; they have the knowledge and expertise required to move in both sides of the border seamlessly, and because of this, they are the ones with the best ability to understand and explain all the nuances in the border. Operationally speaking, this can be understood as a positive correlation between cultural richness, language proficiency, and educational levels to higher levels of transfronteridad.

Table 3

Levels of Transfronteridad

Level	Examples
Lower level Sporadic and transactional interactions	Mexican transborder crossing the border to buy fast food in the U.S.; American transborder crossing the border to buy prescribed drugs in Mexico; Mexican-American crossing the border to get a haircut in Mexico.
Lower middle level Periodic but not emotive interactions	Mexican transborder crossing the border to put gasoline every week; American transborder crossing the border every month for doctor check-in; Mexican-American crossing the border every two weeks to go to the movie theater.
Upper middle level Periodic and emotive interactions	Transborder crossing the border frequently to visit family members; transborder crossing the border daily to work.
Upper level Periodic, emotive, and varied interactions	Transborder crossing the border to visit their second house in either Mexico or the U.S.; transborder crossing the border to visit spouse and close family.

Source. Author's elaboration based on Iglesias-Prieto (2004).

2.3.5. Transborder regions

Transfronteridad is a concept useful for both micro and macro levels. At a macro level, transfronteridad "depends on the level of interdependence between the people of both sides and the relation of the people to the spaces in term of frequency of crossings, the meaning behind the crossing and even the direction of crossing". When there is strong interaction between border cities, there is transfronteridad (Escamilla, 2019 citing Igleasias-Prieto, p.3).

If transfronteridad is occurring at a macro level, between two or more border cities, a transborder region is likely to exist; thus, transfronteridad can be considered a unifying dynamic that fosters the development of transborder regions. On this note, Alegria (1990) states that transmigration -which in this case may be reinterpreted as transfronteridad, gets translated at the spatial dimension in the form of binational pairs of cities spread throughout the border -transborder regions. Each of these spaces will become a metropolitan binational urban center with people moving back and forth, participating in dynamics that are not only determined by the structural difference between both countries but by the particular characteristics of each metropolitan space. Depending on their level of interdependence -and the asymmetry in the relationship- transborder regions might show a variety of interactions such a "close cooperation among local authorities, cross-border living arrangements and employment patterns, the use of each other's educational and recreational facilities, and mutual celebration of holidays and festivals" (Martínez, 1994, p. 12).

Transborder cities -and regions- work in different ways than the cities in their respective countries. In many instances, this kind of cities share a common identity with

each other (Mungaray-Moctezuma & Calderón, 2015). Still, research shows that this phenomenon is more evident on the Mexican side of the border. Moreover, a community does not automatically become part of a transborder community; it is not a matter of just being a community, neighborhood, or city across the border (Goldberg, 2001). Despite this asymmetry, cities in transborder regions affect each other when it comes to urban development; specifically, guiding planning efforts to accomplish objectives such as building spaces appealing to the specific taste of individuals from U.S. crossing the border to buy products or participate in social activities at Mexico, or creating infrastructure to transport individuals crossing the border from Mexico to work at the U.S. (Mungaray-Moctezuma & Calderón, 2015).

2.4. **Transborder Groups at the Mexico-U.S. Border**

While the literature includes a comprehensive list of frameworks to group transborders based on their country of origin, language proficiency, language preference, cultural identity, and more. This text focuses in five of the groups identified by Martínez (1994); the worker, the student, the Mexican consumer, the Mexican-American consumer, and the Anglo consumer. It is important to highlight that the author´s work is not only limited to these non-mutually-exclusive group; there are more than 20 groups describe in their literature. The five groups described in this document are focused in analyzing transborder groups by their behavior and geographical characteristics (see figure 2).

Figure 2

Transborder subgroups
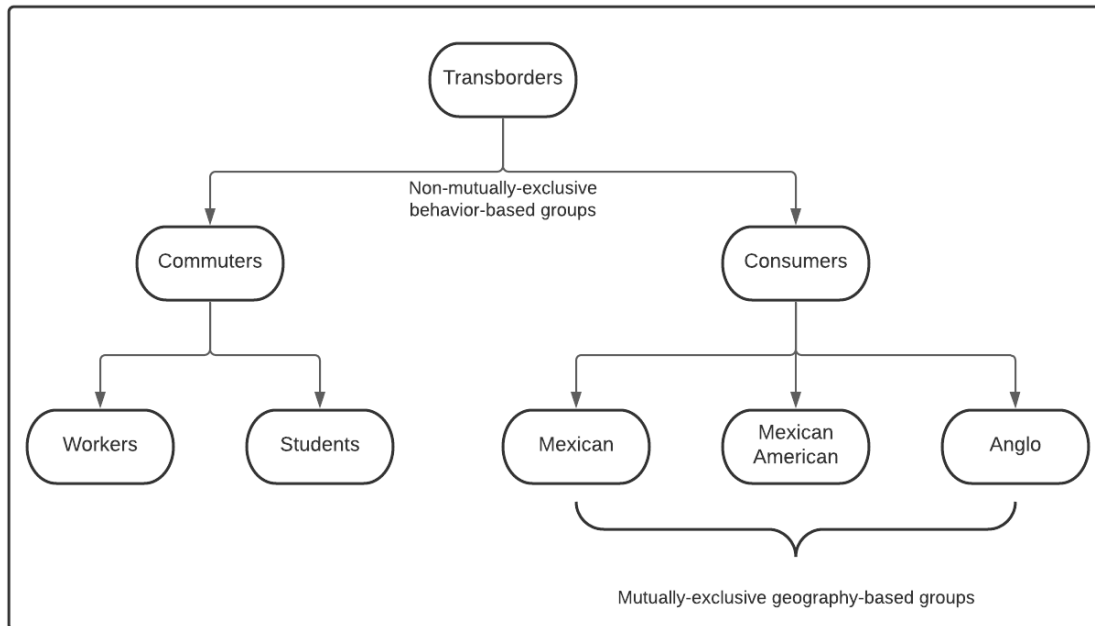
Source. Author's elaboration based on Martínez (1994).

2.4.1. The Transborder Worker

Transborder workers can be defined as individuals living in Mexico but crossing to the U.S. several times per week for work purposes (Orraca, 2015 citing Estrella, 1993; Alegría, 2000; and Martínez, 1994). The practice of going back and forth between borders may have started back in the 40's after the implementation of the bracero program between Mexico and the U.S. While this program enabled Mexican worked to live and work in the U.S., many participants opted for living in the Mexican borders states while continue working in the U.S.

One of the main reasons why individuals residing in Mexico opt for crossing the border on a daily basis to work in the U.S. is the wage difference between both countries. The minimum wage in the U.S. outweighs Mexico's considerably, allowing individuals that

opt for this lifestyle to have a privileged status among their local group of reference; this means that while the income gained by a transborder individuals may not necessarily put them in a high socioeconomic level when compared to individuals living in the U.S., it does it when compared to people from their residence area -Mexico in this case. Being a transborder commuting for work then becomes a very effective way to "improve a household's income position relative to others in its reference group" (Orraca-Romano, Cross-Border Earnings of Mexican Workers Across the US–Mexico Border, 2019 , p. 452).

The job characteristics including job wages, position, and benefits condition whether an individual will have a disadvantaged economic status in the U.S. An individual's job wages will be positively correlated to their educational level, their condition of having prior life experience in the U.S., and their language proficiency. In a nutshell, working commuters may have better job conditions or benefits if they are bilingual and have more years of experience working in the U.S. (Fimbres & Ortega, 2001). Additionally, the education level will also impact their ability to obtain better jobs. This particular group is expected to under index when it comes to having more than one individual working in the household due to their increased income (Orraca-Romano, 2019).

Even though there is not a formal calculation for the current number of transborder workers living in the Mexico – U.S. border, there are at least four major events that may have drove the size of this population down in the last two decades. The first one, 9/11; after the terrorist attack to the Twin Towers on September 9, 2001, border security was increased dramatically, impacting border-crossing times negatively and eventually

making people opt for staying to live in the U.S. The second event is the Great Recession in 2007; an economic downturn that impacted the levels of employment in the U.S., including the number of employment available for transborder workers. The third event refers to the increased level of violence in Mexican border towns, which has made several people to opt for living in the U.S. instead of crossing the border as transborder commuters. Finally, in 2016, the political environment after Trump's election as the President of the U.S. increased the likelihood for individuals to consider living in the U.S. permanently. It is important to highlight that while these events may have negatively impacted the size of this population, transborders continue to have a very strong importance in the region as their relatively higher earnings boost the economies of Mexican border cities while supporting job supply needs in the U.S. (Orraca-Romano, 2015).

2.4.2. The Transborder Student

The term *Transborder Student* refers to the individuals crossing the border frequently with the objective of attending school. This population is also known in the literature as commuter students.

Crossing the border to attend school can easily add up between 2 and 4 hours to the daily dynamics of students, impacting their well-being negatively; thus, a question arises, why do people become transborder students? According to the literature, the main reason why people participate in this practice is to gain tools to compete in the current economic environment; this becomes even more true when considering the fact that some of the best universities in the world are in the U.S and the fact that working in the U.S. while living in Mexico is a great to improve an individual's socioeconomic position (Orraca-

Romano, 2015). Being born in the U.S. and eventually starting to attend classes in that country while living in Mexico is part of a very common social dynamic happening at the border. Under this practice, there is the expectation that the student will get a competitive edge by becoming bicultural and will have the freedom -and tools needed- to eventually decide what is their preferred transborder life/work arrangement in the Mexico – U.S. border (Rocha & Orraca-Romano, 2018).

Rocha & Orraca (2018) state the Commuter students have been historically ignored by most of the educational system in different countries due to their small size[3] (citing Zuñiga & Hamman, 2009). Some of the challenges that students face when attending classes in the U.S. and crossing the border on a daily basis are the fact that Faculty might not understand the dynamics they go through daily and therefore, do not provide any kind of flexibility in terms of allowing them to join classes later when border crossing time were longer than usual. Other examples of struggles in the literature refer to the lack of positive or memorable life experiences from transborders during their High School period due to spending some much time going back and forth between borders (Rocha & Orraca-Romano, 2018).

---

[3] In the case for Mexican transborder students, while their percentage compared in the overall population is not as high as the percentage of cross-state students in the center of Mexico (in Ciudad de Mexico and Mexico state for example) there is a higher connotation of integration for the first group considering that involves two countries, two cultures and two languages (Rocha & Orraca-Romano, 2018).

Transborder students represent a small percentage of the overall student population in all border cities due to different factors conditioning the ability for an individual for opt for this life arrangement including income level, commuting practices, cost of education in the U.S., transportation availability, and more (Rocha & Orraca-Romano, 2018).

2.4.3. Transborder Consumers

In their study, Mungaray-Moctezuma & Calderon (2015), mention that there is a gap of information regarding mobility associated to consumption at the borderscape. They also introduce a concept called nonworking traveling which refers to the mobility that does not involve moving from the residence to the working place -or backwards. It is the kind of mobility done for purposes such as tourism, consumption, commerce, recreation, and medical services. All this is important to this study since that is the kind of mobility the transborder consumers do; thus, it can be considered a good definition for this group. Three subgroups of transborder consumers will be highlighted in this study: a) Mexico-resident transborder consumers, b) Mexican American transborder consumers, and c) Anglo transborder consumers (Martínez, 1994).

2.4.3.1. Mexico-resident transborder consumers

Most of the population residing in Mexico that are able to cross the border legally will do so with the purpose of purchasing different goods in the U.S. that are perceived of a higher quality or show a lower price (Mungaray-Moctezuma & Calderón, 2015). However, there are aspects conditioning the magnitude of frequency of crossing for this population, including the peso-dollar parity, the expected waiting times at the borders, the gas price, and the relationship between how easily it could be to purchase the same items

in Mexico compared to the quality/cost benefits consumers could expect for getting them in the U.S. The way in which Mexico-resident transborder consumers interact with the border is conditioned by their consumption dynamics. This group can be considered a very knowledgeable group in terms of assessing consumption options between Mexico and the U.S. Once important characteristic from this group is that while the constantly visit the U.S., the level of acculturation they will achieve in the American culture is very limited due to the nature of the exposure they have in the U.S. -only transactional interactions.

2.4.3.2. The U.S.-resident Mexican American transborder consumer

Mexican American consumers living in the U.S. are attracted to Mexico because of the bargains they can get (Martínez, 1994). This group can get items such as food, clothes, and services at lower prices than in the U.S. While this group shows intense consumption activities in Mexico like the prior group in the U.S. Mexican American transborder consumers distinguish themselves from the Mexican transborder ones due to an increased likelihood for the first group to have roots in the Mexican side, with family and friends being part of the regular interactions these individuals have when crossing the border. Mexican American transborders are more likely than Mexican transborders to display higher levels of transfronteridad as they do more than product and service consumption when crossing the border. In many instances, many of them are recent Mexican borderlanders that moved to the U.S. and are still going through an acculturation process. While the individuals from this group are more likely to absorb pieces from the American culture, the fact that they cross the border to visit Mexico so often makes this process to move slowly or something not at all.

2.4.3.3. The U.S.-resident Anglo cross-border consumer

While the Mexican American consumers tend to do a variety of activities when crossing the border, activities from Anglo transborder are more restricted. Example of these activities are shopping, expeditions, weekend trips, and vacations in Mexico (Martínez, 1994). Part of the reasons for these individuals to limit their transborder activities are their limited language proficiency, border crossing times, and negative perceptions they may have about the safety in Mexico.

2.5. Current Mexico-U.S. border landscape

2.5.1. The Mexico-U.S. border

While it has been established that borders not only control the traffic of goods and people but allow cities to fusion into transborder regions, this does not come naturally; borders make cultural and development differences evident (Mungaray-Moctezuma & Calderón, 2015). This is important to highlight considering the fact that there is no other border in the world with more economical disparity that the Mexico - U.S. border (Anderson 2003 & Mora 2006 cited by Orraca, 2015). Mexico and the United States share approximately 1,933 miles of border across the Mexican states of Baja California, Sonora, Chihuahua, Coahuila, Nuevo Leon, and Tamaulipas; and the U.S. states of California, Arizona, New Mexico, and Texas (U.S. Census Bureau, 2011; Hernández, 2020). To manage the movement of goods and people throughout this shared border, both countries have installed ports of entry in different cities (see figure 3)**.**

Borderlands are characterized by international conflict and accommodation. This can be understood as the situations in which borderlanders get caught in the middle of international problems (Martínez, 1994). For example, during the Texas rebellion, the

Texas-Mexico border became unstable and many borderlanders had to choose "between remaining in their war-torn land" or abandoning it for a safer ground -decisions that individual living in the interior zones from both borders did not need to make. Because of this type of situation transborders may develop internal or local mechanisms to offset international pressure; border-style diplomacy, as known by (Martínez, 1994). Borderlanders tend to push on their interior or central governments to seek recognition for borderlands problems; however, it is always challenging to get that acknowledgement. Due to this, borderlanders have developed a variety of local solutions to offset this situation, including smuggling and surreptitious trading practices that occur when individuals do not declare goods at customs, to save on taxes and then subsequently selling them as a way to earn an income.

An important distinction to make, particularly speaking about the border cities of Mexico, is their limited connectivity between border cities. Border cities are well connected to the center of the country and to their respective sister cities at the U.S. side. However, they are not connected across the Mexican border -they show weak links (Fernández, 1987).

Figure 3

Border cities in the Mexico-U.S. border
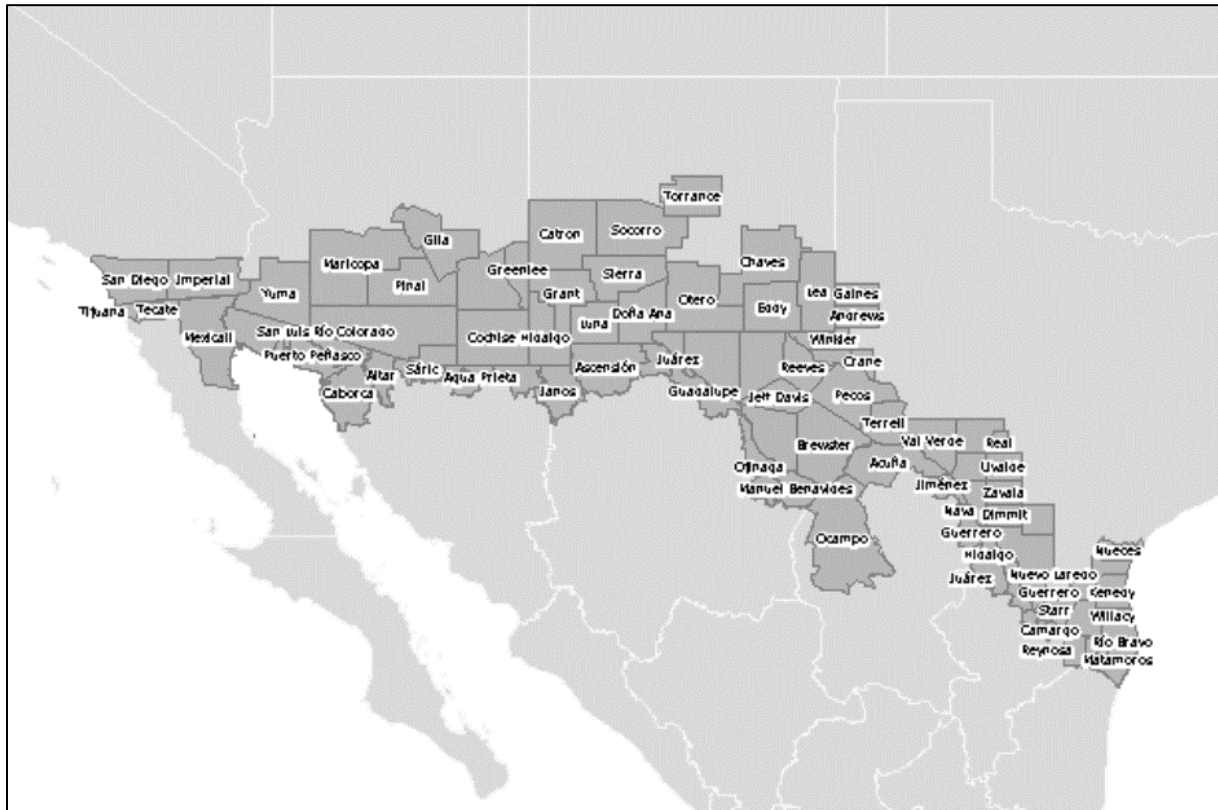
Source. Author's elaboration based on (Orraca-Romano, Immigrants and Cross-Border Workers in the U.S.-Mexico Border Region , 2015).

2.5.2. Crossing magnitude.

When trying to analyze the crossing magnitude at the Mexico – U.S. border individuals can use information from the Bureau of Transportation Statistics or the Mexican Census Bureau. However, the information related to border crossing from the Mexican side is aggregated in a way that users cannot distinct between crossing at the north or south borders from the country. Due to this situation, the author has decided to measure and analyze border crossing magnitude using information from the U.S. ports of entry considering that Mexico-residents crossing to the U.S. will cross via those ports just

like U.S.-residents will return back to their country using the same ports providing a good idea of the transborder dynamics occurring at the different ports of entry.

Table 4 (1/2)

Number of border crossings by port of entry, 2019

Port of Entry	n	%
Total	194, 669,176	100
Arizona	23,955,505	12%
Douglas	3,660,549	2%
Lukeville	1,172,503	1%
Naco	684,882	<1%
Nogales	10,745,214	6%
San Luis	7,636,476	4%
Sasabe	55,881	<1%
California	78,653,898	40%
Andrade	1,977,746	1%
Calexico	12,713,669	7%
Calexico East	6,780,146	3%
Cross Border Xpress	1,620,661	1%
Otay Mesa	15,899,396	8%
San Ysidro	36,724,706	19%
Tecate	2,937,574	2%
New Mexico	2,795,077	1%

Table 4 (2/2)

Number of border crossings by port of entry, 2019

Port of Entry	n	%
Columbus	1,118,586	1%
Santa Teresa	1,676,491	1%
Texas	89,264,696	46%
Boquillas	22,679	0%
Brownsville	12,875,059	7%
Del Rio	3,423,649	2%
Eagle Pass	7,089,984	4%
El Paso	27,436,807	14%
Hidalgo	11,542,936	6%
Laredo	17,544,905	9%
Presidio	1,764,236	1%
Progreso	3,918,677	2%
Rio Grande City	929,840	0%
Roma	1,804,381	1%
Tornillo	911,543	0%

Note. Data is sorted in descending order by number of crossings. Total number of crossings by port of entry are calculated as the sum of the number of crossings by bus passengers, pedestrians, personal vehicle passengers, train passengers, and trucks.

Source: Author´s elaboration with data from Bureau of Transportation Statistics, 2021.

The Mexico-U.S. border is a very active one with millions of crossing every year. According to data from the Bureau of Transportation Statistics, in 2019 there were more than 190 million crossing at one of the 27 ports of entry. This is relevant considering the size of both Mexico and the U.S. While there may not be much meaning in comparing them, it is interesting to see that the number of crossings in 2019 were equivalent to 42% of the total population in both countries (458.4 million inhabitants: World Bank, 2020). Results from 2019 were used for this analysis considering the impact of COVID-19 in the number of crossings for 2020 and 2021. Data available shows that in 2019, the Texas border was the most active with 89 million crossings, representing close to half of the total crossings (46%). The California border was the second most active border with close to 79 million crossings (40% of the total crossing in that year). In third place, we found the Arizona border with close to 24 million crossings (12%). Finally, the least active border was the New Mexico one with 2.8 million crossing that represented one percent of the total crossings. The results by port of entry show a concentration of border activity. From the 27 ports of entry in the U.S., six ports concentrate 64% of the total crossings . At the top of the list, it is San Ysidro one with close to 37 million crossings. This is aligned with information provided from different authors stating that it is the border most frequently cross in the world (Falcon & Orta; 2018). It is interesting to see that while California is not the most active border from the U.S., it is the one with half of the top port of entry in terms of border crossings. After the San Ysidro border, we see that the border of El Paso, Laredo, Otay Mesa, Brownsville, and Calexico are the most active ones with approximately 27.4 million, 18.5 million, 15.9 million, 12.9 million, and 12.7 million crossings respectively (Bureau of Transportation Statistics, 2021).

2.5.4. Mexico-U.S. Transborder Regions.

Commonly, people refer to borders as something beyond the actual physical divide between two countries; this is the geographical space surrounding them. However, this in reality refers to the borderlands. According to (Martínez, 1994), these are the regions adjacent to a border; these regions may be small when the interaction between two countries is reduced and large when the interaction is bigger. Different authors have defined these as transborder regions or metropolis. While in the Mexico-U.S. border it is very common to find twin sisters with increased binational interactions, this is not a common phenomenon in other border across the globe (Fimbres & Ortega, 2001 citing Alegria 1992).

Different authors state the number of transborder regions in the Mexico-U.S. region between seven and 15 (Orraca, 2015; Hernández, 2020). For this study, the author decided to use the definition from Orraca (2015) that states there are nine main transborder region in the border. Demographics traits of the transborder regions have important implications as they can be correlated with the economic vocation of each region, price levels, income levels, among other sociodemographic variables and dynamics. Table 5 shows that the nine transborder regions identified show very diverse traits. When it comes to combined population, the Tijuana – San Diego border is the largest one in terms of populations. Conversely, the Nogales-Santa Cruz is the smallest. In terms of the population differential, understood as the ratio between the population in the Mexican side and the U.S. side, the Ciudad Acuña y Piedras Negras – Maverick region is the one with the highest differential. This means that in proportional terms, Ciudad Acuña and Piedras Negras are relatively bigger than Maverick compared to this

difference for all the other regions. On the opposite side, the Tijuana-San Diego region shows the lowest differential. Since this differential is under 1, it means that San Diego's population is higher than Tijuana's. In terms of Density differential, which is the ratio between the density in both cities (Mexican and U.S.), it is possible to see that the Nuevo Laredo – Webb region shows the highest one; Nuevo Laredo's density is 11 times higher than Webb's which can speak to the phenomenon of transborder Mexican cities growing too close from the border due to the economic opportunities derived from this geographical location. The lowest density differential can be found in Matamoros – Cameron with .64, which means that Cameron has a higher density that Matamoros. However, the relative difference between both places is relatively and considerably smaller than the one seen for the top one. In terms of combined crossings, the Tijuana – San Diego region is the largest one in the border and according to many authors the largest in the world. The San Luis Rio Colorado – Yuma region was the one with the lowest number of crossings among the transborder regions in the border. Finally, in terms of crossings per inhabitant, the Nogales – Santa Cruz, which was actually the smallest region in terms of population, is the one with more crossing per inhabitant, which could mean two things or a combination of them; this is a very active border with a lot of transfronteridad between its member, or it's in the middle of the road to get to bigger cities in either Mexico or in the U.S. Finally, the Reynosa-Hidalgo pair is the one with the least crossings per inhabitant.

Table 5 (1/2)

Characteristics for Transborder Regions across the Mexico-U.S. border

Transborder Region	Combined Population	Population Differential	Density Differential
Tijuana – San Diego	5,221,157	.58	5.91
Mexicali – Imperial	1,229,494	5.84	4.35
San Luis Río Colorado – Yuma	402,902	.98	1.57
Nogales – Santa Cruz	312,451	5.55	10.13
Ciudad Juárez – El Paso	2,378,107	1.75	1.29
Ciudad Acuña y Piedras Negras – Maverick	397,272	5.86	1.63
Nuevo Laredo – Webb	692,172	1.59	11.32
Reynosa – Hidalgo	1,575,548	.81	1.05
Matamoros - Cameron	962,996	1.29	.64

Table 5 (2/2)

Characteristics for Transborder Regions across the Mexico-U.S. border

Transborder Region	Combined Crossings	Crossing per inhabitant
Tijuana – San Diego	54,244,763	10.4
Mexicali – Imperial	19,493,815	15.9
San Luis Río Colorado – Yuma	7,376,476	18.3
Nogales – Santa Cruz	10,745,214	34.4
Ciudad Juárez – El Paso	30,024,841	12.6
Ciudad Acuña y Piedras Negras – Maverick	10,513,633	26.5
Nuevo Laredo – Webb	17,544,905	25.3
Reynosa – Hidalgo	11,542,936	7.3
Matamoros - Cameron	12,875,059	13.4

Note. Data for Mexico correspond to municipalities. Data for U.S. corresponds to counties. Top value for each indicator has been underlined. Bottom value for each indicator has been double underlined.

Source. U.S. Census Bureau, Quick Facts, 2020 & Mexican Census Bureau, Housing and Populations Census, 2020.

2.5.5. The Mexico-U.S. borderscape from a macroeconomic point of view.

The nature of the consumption patterns happening at the border has to do much with the dollar-peso parity that created an increased purchasing power to those earning dollars and spending in pesos. According to (Fernández, 1987), the rapid devaluation in the value of the peso compared to the American dollar has impacted the border cities in different ways including a negative impact for U.S.-based commerce, positive for Mexico-based commerce, an increase in the output and employment for the manufacturing companies which with a weaker peso can pay for more employees and become more productive with the same investment, inflation for the border cities residents in Mexico since they many of their products come from the U.S. and also reduction in the offer they can afford in the US. to which Mexican businesses will increase prices, it also gives people live in Mexico but working in the U.S., and people living in the U.S. but constantly visiting Mexico, an increased purchase power with on one hand is beneficial for both consumers and vendors, but not that good for locals not working in the us which need to struggle with the inflation.

2.5.5.2. Business Landscape at the Border

One of the impacts the transborder phenomenon has for local Mexican markets is the increase in the offer of products with high price elasticity (i.e., products where even a small change in price may incur in a high but inverse change in consumption; Alegría, (1990). This fosters an environment in which industries such as entertainments, restaurants and similar can thrive if they successfully cater to the interests, needs and preferences from this population. There is interdependence between Mexico and the U.S. -though asymmetrical. Still, in the U.S. side, several shopping centers, malls, and other

commercial business depend heavily on Mexican consumers crossing the border to acquire different products are a lower price or with higher quality (Martinez, 1994; Mungaray-Moctezuma & Calderón, 2015). Some other sectors in the U.S. that get benefit from the binational interaction are banking receiving deposits from Mexican individuals, real state from wealthy Mexican purchasing hours, business investments owned by Mexican, labor-intensive industries that benefit from Mexican accepting low-wage payment that help them to keep cost low and break strikes when needed -or prevent unions from being created. From the Mexican side, maquiladoras employ thousands of Mexican employees that benefit from this industry arrangement; tourism that receive millions of individuals both Mexican Americans and Anglo to consume products in Mexico; services that they can get at a lower cost or with a higher quality; commuting worked from Mexico that are able to access better wages than the ones they could get in Mexico while still being able to remain in Mexico with their families. Back in 1990 transborders did not represent more than 8% of the total working population at any Mexican border city; still, they were contributing between 14 and 20 of the total wage income in those cities. This is due to the fact that transborders tend to use the majority of their income on the Mexican (Alegría, 1990).

2.5.5.2. Job Markets at the Border

There is a permanent dynamic or look for balance in the U.S. between ensuring industry and agriculture have a reliable cheap work source -illegal and legal Mexican workers-, but also speaking to the arguments of local Americans complaining about the impact in the labor market of foreign and cross-border employees (Fernández, 1987). Particularly the Agricultural Industry, according to the author, has been using the work

supply from Mexico as water faucet they open and close based to their needs so that they can ensure they have enough workers for short term needs. It is not uncommon to see foreign workers to be encouraged while the economy is booming, and expulsed when the economy is not going so well (González, 1988). Work supply and demand are the real conditioners of migration policies between Mexico and the U.S. (Fimbres D. , 2000). The advantages of transmigrants (not necessarily transborder) rely on the minimal cost the receiving country pays for a ready-to-use worker (González, 1988). This migration workers have been educated in Mexico, they are in a working productive age, which reduces the investment the U.S. need to make in education, and not only that but also social security benefits such as healthcare expenses just because younger employees are less likely to need them. In addition to this, the fact that these employees might live in Mexico allows the U.S. not to need to worry about them needing education or other services for themselves or their children. U.S. employers get benefited by situations where the other side shows overpopulation combined with a labor surplus and poverty, which let them pull qualified workers willing to take lower salaries and with a very low likelihood of unionizing while also affecting Mexican industries that could be benefitted from having more qualified workers (Duarte-Herrera, 2001 citing Martinez, 1994). From an economic standpoint, the international working dynamics are created by structural differences between countries, in this case, a first and a third world country sharing a geographic space; being adjacent (Alegría, 1990). It is this quality what allows individuals to take individual advantages of these structural differences (i.e., higher wages in the American side, higher purchase power due to a weakened currency). Another important thing to consider when it comes to international working dynamics is the factors impacting

work supply and demands; while in non-border markets individuals will see their work accessibility determined in its majority by the difference by the existent of inexistent job supply gap, the level of training of the worker that the cost of movement, in the binational regional, individuals will also be impacted by the characteristics of the sector they are interested in (i.e., some transborder regions then to skew to the primary sector such as Mexicali/Calexico, San Luis/Yuma vs others that skew more towards commerce, services and industry such as Tijuana/San Diego), the migratory status (i.e.,. individuals with citizen or resident immigration status may have an easier time getting a job or better conditions that individual with working visas or illegally working), and crossing times (i.e., individuals that are able to cross using SENTRI lanes or even ready lanes may be able to get jobs north of the border since they will not need to account for extended border crossing times while individuals crossing without a car or using the regular crossing lane will need to stay close to the border to compensate for this time.

2.6. History of the Mexico-U.S. borderscape

When looking at the history of the borderscape, it is useful to use the framework provided by (Martínez, 1994) that consists in a four-stages model for border regions. According to the author, all borderscape can be categorized under one of these categories: alienation, characterized by conflict or increased tension; coexistence, characterized by intermittent stability; interdependence; characterized by stability; and integration; characterized by unrestricted movement of people and goods. The borderland went through a period of alienated borderlands since the inception of Mexico until 1880, a period of coexistence from 1880 to 1920, and since 1920 there has been a period of

interdependence. So far, the border region has not achieved integration like other borders in the world (see figure 4).

Figure 4

Mexico-U.S. Borderland Interactions
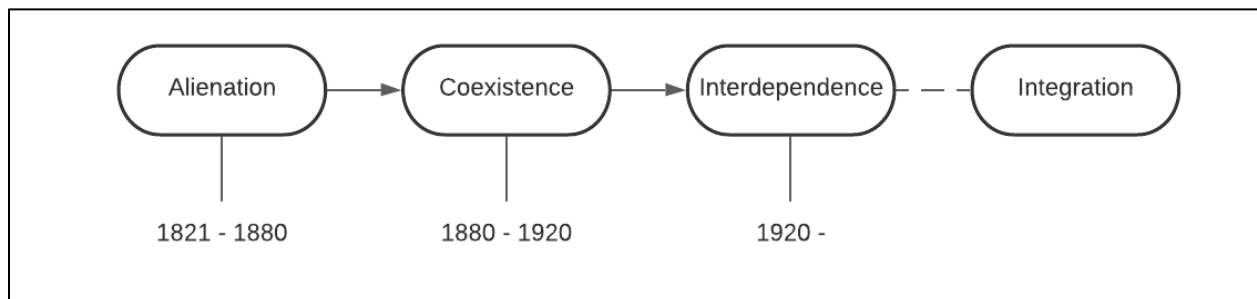
Source. Author's elaboration based on (Martínez, 1994)

These are some of the complicated points in the Mexico-U.S. relationships that impacted and crafted the borderland dynamic: a) Early 1920 – Debate about immigration quotas; b) Early 1930, repatriation of half-million Mexicans from the U.S. after the great depression; c) Late 1940, open door for Mexicans as work power replacement while there was war; there were incidents and disagreements about the bracero program; d) early 1950, massive deportation from Mexicans under the operation wetback with many former bracero program members becoming permanent citizens of the Mexican side of the borderland and the creation of the maquiladoras along all the Mexican border cities (Border Industrialization Program); e) 1960 to 1980, continued debate over legislation to curb undocumented immigration; f) 1980, Simpson-Mazzoli bill was approved giving millions of Mexicans in the U.S. legal immigration status – this was accompanied with new restrictive laws to enforce the border; g) 2000 – 2016, after 9/11, more restrictions to cross the border a before and after in terms of technology and costs to keep the border from the U.S.; h) 2016 – to present, the U.S. 2016 elections with the victory of Donald

Trump marked a hit in the Mexico-U.S. relationships due to a change in the tone in binational relationships and also the impact that COVID-19 pandemic had on the border with partial closures throughout both countries (Martínez, 1994; Hernández, 2020).

2.7. Transborders at the Mexico-U.S. Border

2.7.1. Transborder Culture and Identity

In order to fully understand the concepts of transborder culture and identity, it is important to first learn the general concepts of culture and identity. Moreover, the concept of community also becomes important when trying to understand these constructs. Community, for the purposes of this study, can be understood as a state of mind that gives people experiencing it a sense of belonging. Community is not a physical place, it could be, but not necessarily. Examples in the literature speak to communities identified by transborder individuals that can feel a sense of belonging with transborders in places far away from their location more than to individuals or places that are close to them (e.g., Hispanic communities in Nogales Arizona having a sense of community with people living in Mexico more than with people living in predominantly-white neighborhoods in the same city; (Goldberg, 2001). The second concept to highlight is culture. Culture can be understood as the social category system compromised by symbols, traditions, messages, and more that a particular group or society develops to cope through different challenges during their social evolution process -as a group. Culture is not necessarily tied to a physical location (i.e., there is more than culture based on the place where you live). Culture can be associated or tied back to different groups bases in the gender, ethnicity, sexual preference, age group, and many more aspects. Culture is an important concept when discussing transborder dynamics as the constant transborder interactions

throughout the year have come to create something known as transborder culture (Goldberg, 2001). Finally, identity can be understood as the aspects from a particular culture that an individual may follow. For example, while in the transborder culture, there may me messages already established regarding the important of knowing how to make the most of time while waiting in the line (e.g., do homework, call family members, sleep), individuals practicing transborder dynamics may not related to this situation if they don't need to wait in line to cross the border (i.e., SENTRI pass holder). In a nutshell, an individual crossing the border on a daily basis, may have adopted, or showing aspects from the transborder culture as they are part of this particular groups; however, considering that the individual does not share the common practice of waiting in line with other transborders, this individual may not identify as one of them and therefore, not feel a sense of community with them (Falcón & Orta, 2018).

It is valid to bring the classic assimilation theory that establishes that when two cultures meet through migration processes, the minority group will tend to lose its identity to avoid conflict with the mainstream group. There are two important things two be mentioned regarding this. One is that evidence has shown that assimilation is best achieved if both groups (majority and minority) converge into mutual acceptance. These ideas are important considering that transborder populations, especially those that commute daily to work or school, area faced with two cultures, and make conscious and unconscious decisions regarding which things of which cultures they take temporarily or permanently. Still, there has not been an explanation developed for the transborder phenomena. Are the transborder populations supposed to get to one of the two cultures after a time? Is theirs a third culture? Are they considered biculturals? (Calzada, Covas,

Ramirez, Miller & Huan, 2016 citing LaFromboise et al. 1993, Gordon, 1961, and Alba & Nee, 2003). It is important to understand two basic aspects of the theories related to the acculturation process. Acculturation is understood as the process of interacting in the US mainstream culture when that is not your culture of origin, while enculturation is one's culture of origin (Calzada, Covas, Ramirez, Miller, & Huan, 2016). Another cultural adaptation argument from Berry (2003; cited by Calzada et al, 2016) states that cultural adaptation can be explained in terms of a four-fold model in which two dimensions participate, culture of origin named as enculturation, and new culture named as acculturation. Depending on how the individual reacts to becoming or learning from a new culture, the individual will fall into four categories. Rejection of the culture of origin while participating in the new culture is called assimilation; rejection of the new culture while maintaining the original culture is called separation; maintenance of the original culture while also adapting to the new one is called integration; and rejection of both cultures is called marginalization. Can the transborder experience be analyzed in term of this four-fold model? If so, what happen to U.S. raised transborders, which would be considered their first culture? (see figure 5).

Figure 5

Cultural Adaptation Theory Four-fold Acculturation Model
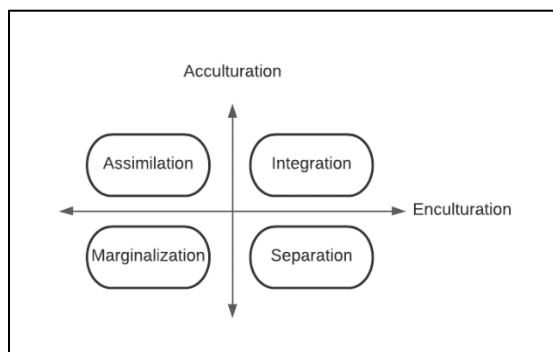
Source. Author's elaboration based on Berry, 2003.

Part of the purpose of this work is to reflect on how different transborder group live their transborder lives. A particular group of interest is confirmed by transborder populations that despite of crossing the border daily or very frequently, working in the U.S. and being exposed to the U.S. culture a lot, have decided to not adopt many traits of the American culture (language, traditions, ways of thinking, not even friends). According to Postes and Zhou's (1993) segmented assimilation model, this has to do with the fact that acculturation is affected by different characteristics that both the individual learning from a new culture/society, and the receiving culture/society bring. In their theory, they state that characteristics such as country of origin and socioeconomic status play a role from the potential person to learn from a new culture. Still, these characteristics play with some of the receiving society such as discrimination practices determining receptivity for the learner. Because of the play of these variables, it could happen that lower economic individuals remain in marginalized societies or subsocieties within the mainstream (Portes & Rumbaut, 2006). This would explain why even though some transborder workers might have worked their lives in the U.S. they do not participate in many activities besides their work in that country. If there is an interaction between migrating individual and receiving society in which both get something from each other, how does that play for transborders that are not migrating, or at least not in the general sense? Is being a transborder who lives in Tijuana but works/studies in the U.S. the equivalent of being marginalized in one of the most segmented societies in the U.S.? According to Calzada et al (2016), different factor can impact of the acculturation process for everyone. these factors include aspects related to the reasons why a particular individual might be migrating, how long is the individual expecting to stay in the new place

(it could be permanently or they might think it will not be permanent, despite ending up like that, this can really impact how much is someone willing to absorb the new place's culture...all these theories relate to individuals getting adapted to the US mainstream culture, but little is said about individuals that might have been raised in the US culture, or some kind of hybrid of the U.S.-Mexican culture, these individual would be the ones that despite living in the U.S. when little, decided to go back to the Mexico to take advantages of lower prices, it could also relate to individuals that live in the U.S. but decide to frequently go to Mexico...does these individuals change their culture, adapt to the Mexican culture even though they are not living in Mexico, but frequently visiting? These cross-border consumers from the U.S. might be visiting Mexico more often that the Mexico cross-border consumers, so is there any impact of this in their identity? It is important to mention that by blending we do not mean it in terms of the cultural adaptation theory. The acculturation and enculturation model refers to population that are migrating to the United States and follow a process to incorporate to the mainstream culture which could lead to different results (such as assimilation), in this case, blending is more of a back-and-forth phenomenon. People might have lived like this since they were little, therefore they have not followed a process of acculturation.

The border culture consists of borrowing aspects from each culture that accommodate to the unique reality of the borderlands; things like "language, religion, values, customs, traditions, holidays, food, clothing, and architecture" can be found mingled in both sides of the borderland (Martínez, 1994, p. 57). Unlike other communities, transborder communities -and people living in them- construct and reconstruct their identities frequently since there is a migration flow that continually happens (Goldberg,

2001). Identity is a complex concept. It involves variables such as gender, ethnicity, class, and power. Not only how a potential individual fall on each of these categories, but the ideas the individual has around them (Goldberg, 2001). As the author mentions, in the community they researched, Nogalenses defined themselves as American or Mexicans not based on what governmental policies would say but based on the cultural traits and social networks. In their paper, Falcón and Orta (2018) describe how students attending classes in the U.S. while living in Mexico develop a transborder identity that gets defined by their experiences and interactions in the U.S.-Mexico border what experiences and interactions? Falcón & Orta (2018) state that transborder identities are based on how each individual experience transborder interactions. This means that two people living transborder lives could have a totally different feeling respect to that. The authors include evidence of people stating that even though multiple people are part of this practice, they do not consider the experience the same as they do it in different ways (i.e., different frequency of border crossing, different crossing mechanism, different purpose of visit).

Two important concepts about transborder identity are brought up by Goldberg (2001: 59). The first one from Anzaldúa (1987) states that transborder culture is "a synergy of two cultures with various degrees of Mexicanness or Angloness" (Anzaldúa, 1987:63). On the other hand, Kearey provides concepts that helps to describe the individuals transborder that take the best from both worlds with his concept of polybian that he uses to define individuals who "adapt their being to different modes of existence Kearney, 1996:141 being cited by Goldberg on 59)). According to (Campos, 2012), while the border may be seen as a blockage or barrier preventing culture from being mixed, it has not prevented the development of a transborder identity. Differences such as

English/Spanish have been overcome by Spanglish, challenges associated with dollar/peso distinction have been blurred by the existence of currency exchange offices. Just like those examples, the transborder identity have been developed challenging dichotomic conceptions about them and us. The importance of identity and the fear of losing your identify is reflected with one of Goldberg's (2001) informants who states how her students struggle or have a lack of willingness to learn English in transborder communities since they consider it is somehow a signal of betraying your identity, losing your sense of being a Mexican...somehow denying that part of you. Language proficiency dimension it is complex to be combined with the transborder identity concept. On one hand, it is not hard to find examples of individuals having a hard time to become willing to learn the English language due to feeling that somehow, they are betraying their Mexican identity. At the same time Goldberg (2001) establishes that "bilingualism is associate with border identity", which takes us back to the idea that within the transborder population, there are different mindsets or groups that can be analyzed in terms of different information such as language proficiency, willingness to learn another language, and more. Still, the authors warns that language is not the only or the most important variable to understand the transborder identity. Additional factors such as length of residence at the border, conceptions of place are especially important. The transborder identity construct it is a nuanced picture. Just like Falcón & Orta (2018) informants mentioned, some of them did not identify with their counterparts living transborder lives since while some of them did not have a car to cross the border every day, others were SENTRI-pass holders which, according to them, drastically changes the number of challenges a particular individual faces to live a transborder life.

According to Falcon & Orta (2018), the process of developing a transborder identity gets informed by interacting, coping, and adapting to obstacles through the transborder context. The authors explain that the transborder identity can be described as unique (when compared to the Mexican and American experience and lifestyle), in between (when having to constantly move back and forth between two cultures), with a permanent needs of adaptation (when needing to adapt to their current reality if they are in Mexico or the U.S.), a mixture of cultures (when they cannot determine what music, traditions, they prefer; it's a mix of both), a definable culture (since transborders consider that being able to choose they belong to two nations without citizen, the experience crossing the border, their language are all aspects of this culture), and with cultural practices and beliefs such as getting the best from both worlds, not thinking of the border as separating two cities, but some kind of middle entry between them. They think of the separate nations and both places as one city.

2.7.2. The Transborder Lifestyle and Mindset

While there is not a unique transborder lifestyle or mindset due to the fact that the border and border interactions are not experienced in the same way by all individuals, it is possible to identify some commonalities among transborder individuals living this lifestyle in the Mexico-U.S. border. Perhaps, some of these commonalities will resonate better with one group or other. Still, it does not mean that exploring them is not a positive exercise during the process to increase the body of knowledge regarding transfronteridad (Duarte-Herrera, 2001). These commonalities are a) an interest for getting the best form both worlds, b) a mixed identity, c) a high level of consciousness regarding immigration-related aspects, and a d) lack of nationalism.

One of the biggest is not the biggest commonalities among transborder is a shared interest for getting the best from both worlds. A transborder individual may be getting breakfast in Mexico and then dinner at the U.S. for one or other reason. This is a very important phenomenon as it comes with several implications for governments and organizations trying to engage with this group. If these entities do not fully understand the shared interest transborders have on getting the best from both worlds, the will continue to fail in developing public policy or marketing strategies that can effectively reach these populations. The effects of this type of situation will be seen in reduced levels of usage for public services offered to these populations or reduced popularity for certain products that may have been developed thinking of these individuals. Transborder individuals will not spend all their income -or time- in one side of the border. Instead, they will look for the best option at each side. Examples of this can be seen in individuals opting for working in the U.S. both living in Mexico to receive higher wages while paying less rent. Other example are individuals living in the U.S. but getting most of their recreational time from Mexico to have a higher sense of safety from living in the U.S. while also saving money due to the lower prices of services in Mexico.

A second aspect tied to the transborder lifestyle, that may be more recurrent among Mexican American transborders is having a mixed identity. On one side, transborder experience influence from the Mexican side in the form of traditions, holidays, ways of doing things. Simultaneously, because of the back and forth, transborder will also be exposed to influence from the U.S. side creating a mixed identity that will vary based on the individual's unique circumstance.

A third aspect that needs to be highlighted is an increased consciousness about immigration aspects or things occurring at a given point in time. While non-transborders may not think much about the immigration status in either country, for transborder this is a recurrent topic in their head; whether in the form of actions they take on a day-to-day basis such as crossing the border, or in the form of daily struggles experienced by them or family members due to their specific migratory situation.

2.7. Multicultural Competency and Transfronteridad

According to the Kansas State University, multicultural competency is "the knowledge, skills, and personal attributes needed to live and work in a diverse world" (2001, Paragraph 1). This concept is important when studying transfronteridad since transborder experiences and dynamics are highly conditioned by the level of multicultural competency an individual has. While there is not an operational definition for multicultural competency available, it is possible to cross pieces from the literature with the definition provided by the Kansas University to stablish the following variables as potential metrics that could be used to measure the construct of multicultural competency: a) Language Proficiency, b) Educational Attainment, c) Multiculturality, and d) Life experiences in the U.S. (see figure 6).

2.7.1. Educational Attainment

According to Gonzalez (1988), the income levels for transborder populations is impacted by the educational attainment of workers. Educational attainment may determine whether or not a transborder will live in the U.S. or Mexico which will also condition the transborder dynamics they will hold.

Figure 6

Multicultural Competency Definition and Operationalization
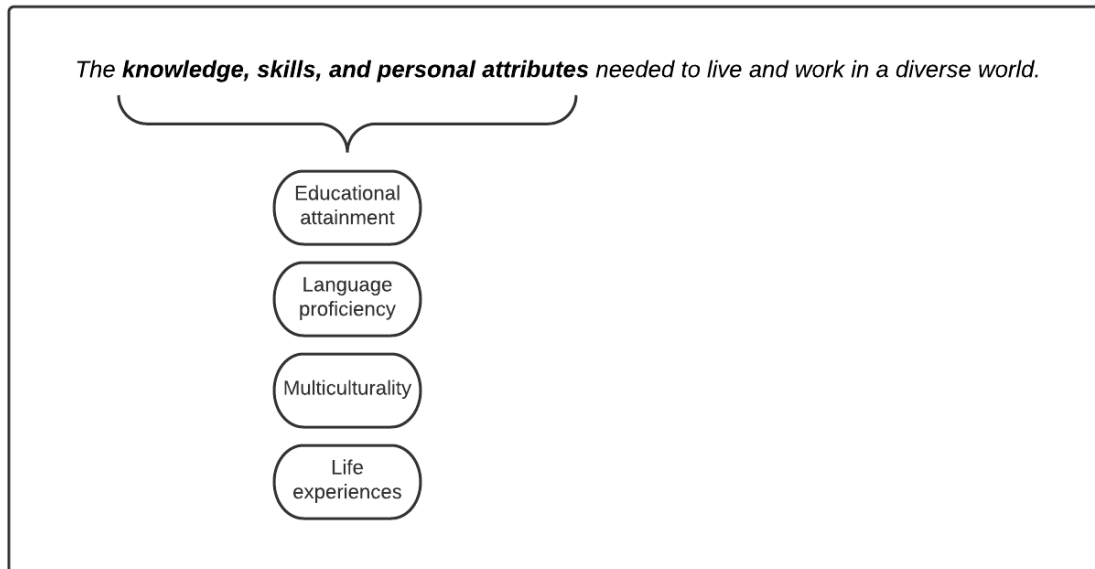
Source. Author's elaboration based on Kansas State University (2001, Parragraph 1).

2.7.2. Language Proficiency

Language is a gatekeeper for transborders that conditions the kind of job they can have, their income, their ability to mingle in either society, the level of comfort they will have performing different activities in each place, among other things. Transborders need to develop both English and Spanish proficiency to avoid being excluded from groups in both sides of the border. This means that transborder dynamics and the ability from individuals to participate on them will be conditioned by their language proficiency. Examples of the impact of limited language proficiency can be seen in the fact that Anglo and Mexican transborders show lower levels of transfronteridad than Mexican American transborder due to the fact that the first two groups are less likely to be bilingual (Falcón & Orta, 2018).

2.7.3. Multiculturality.

People that live transborder lifestyles make the choice, whether or not they want to be part of both societies. If they go for the first, being able to blend into both societies is critical (Falcón & Orta, 2018). Being able to connect with two cultures does not come automatically or easy for everyone, even if they are proficient in both languages; it goes further than languages.

2.7.4. Life Experiences at the Other Side of the Border.

Having experiences of life in the U.S. can be associated with higher levels of transfronteridad for commuters. This is due to an increased understanding of the U.S. work culture, their networks, a trajectory in the U.S. that employers may value more than one in Mexico, and potentially, proficiency in the English language. These concepts are important as they provide evidence to state that there is a correlation between the experiences the type of lifestyle a transborder can live and their conditions of having lived in the other side of the country. Mexican-American and Anglo individuals crossing the border may also be able to access a wider array of goods and services if they have previously lived in Mexico consequently displaying higher levels of transfronteridad (Orraca-Romano, 2019).

2.8. Big Data

The term *Big Data* is one of recent creation; thus, there is still a lot of discussion regarding what it really means. According to (Riahi & Riahi, 2018), Big Data -of course- refers to giant databases, working with arrays of data with millions of observations; however, the term means more than just that. The authors mention that *Big Data* should be understood not only as millions of data points but also as the various techniques used

to extract knowledge from these data points. Additionally, the authors, highlight the polymorphic nature of this term due to the exponential increase of information available in the world. In general, we can stablish that *Big Data* refers to the vast amount of data available nowadays and that can be processed using a variety of techniques.

There are five concepts that have acquired relevance when referring to Big Data, these are a) volume, b) variety, c) velocity, d) veracity, and e) value (see figure 7). These concepts refer, in the same order to the amount of data, its variety in terms of data types, the frequency at which these data grow, the accuracy of the data, and the potential value that can be generated from it. These concepts reflect the multidimensional nature of Big Data, meaning that individuals should not assess Big Data arrays solely on one of these.

Big Data is more than just data arrays; it is critical to use the proper methods to extract value from these. Because of this, big data analytics become a critical component of the value creation process for this product. According to (Riahi & Riahi, 2018), there are four different types of analytics for Big Data; these are a) descriptive, b) diagnostic, c) predictive, and d) prescriptive. These types will answer to the questions what is happening? Why is it happening? What may happen in the future? And what should be done about it?

According to (Fan, Han, & Liu, 2014), the main goals for Big Data are to provide the ability to predict the future for particular phenomena while also understanding the relationship between the features used to analyze it. With millions of data points available, Big Data Analytic methods allow individuals and organizations to understand the nuances in the dynamics occurring in our world; however, there are certain challenges expressed by the authors that may prevent reaching full potential. These include "*noise*

accumulation, spurious correlation and incidental homogeneity…heavy computational cost and algorithmic instability…heterogeneity, experimental variations and statistical biases" (p. 294).

The relationship between this project and Big Data concepts relies on the need for this research design to generate datasets containing close to 20 million records with sociodemographic information provided by both the Mexican and U.S. Census Bureaus. More information related to these activities is discussed in the method and result chapters.

Figure 7

Big Data's 5 V's
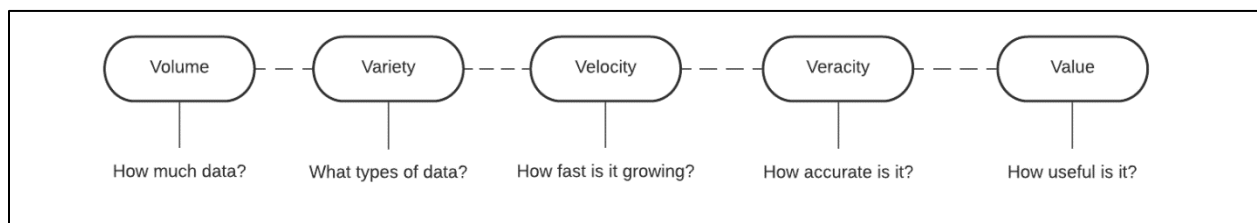
Source. Author's elaboration based on (Fan, Han, & Liu, 2014).

2.9. Regressions

The term regression was coined by Francis Galton in the 19th century during one of his studies in which he analyzed the relationship in height between parents and children. Galton found that children height tended to *regress* to the same levels than their parents. While the work from this author was not focused on creating this particular method or data analysis -regressions- he open the door for the posterior development of the science called econometrics that was based on regression models (Gujarati & Porter, 2010).

Statistically speaking, regression is a statistical technique that allows to analyze the relationship between one dependent variable and one of multiple independent ones. Moreover, after identifying the structure of this relationship, this technique allows to predict the value that the dependent variable will show if the independent variable(s) is set to a specific value.

Regression analysis is a very recurrent statistical technique with applications in all fields including economics, engineering, social science, medicine, and much more. Regression analysis, while not commonly advertised like that, is the base of many supervised machine learning model -what happens in the black box of many of these algorithms.

One important distinction made by (Gujarati & Porter, 2010), is the fact that regression analyses provide information to model different phenomena under statistical and stochastic principles. This is important since it means that the results of a regression analysis will always be subject to random errors making it impossible to predict the outcomes of a particular event with 100% accuracy. This is different that what is called a deterministic model in which relationships show constant magnitudes, directionality, and behavior. Examples of deterministic models are lays such as the Gravity Law or the Electricity Law. An additional limitation that regression users must be aware of is the fact that a regression analysis will never provide enough evidence to establish causality -these must come from theoretical knowledge.

While there is an endless number of variations and models available for regression analysis, the basic one is based on one dependent variable -commonly Y- and at least one independent variable -commonly X. These two -or more- variables will have some

type of interaction that will be described in a particular way through coefficients -beta values- and error values. Depending on how much of Y the combination of X and the coefficient(s) are able to explain, the smaller the error will become, which will elevate the fitness of the model. In regression, the fitness of a model or its ability to successfully predict the dependent variable value is captured using a metric called R-squared or R^2

Finally, it is important to mention that regression model can explain different type of relationships between variables, including linear, logarithmic, quadratic, or exponential relationships (see figure 8).

Figure 8

Common Regression Relationship Types
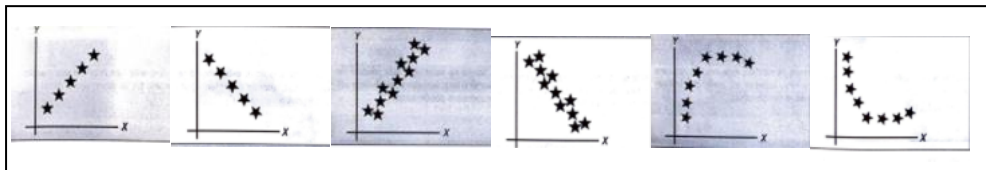
Source. (de la Garza, Morales, & González, 2013, p. 45)

2.10. Machine Learning

Machine Learning refers to the various techniques aimed to build problem-solving programs through learning. This learning process occurs using data of multiple types (records, videos, images, and much more) and allows the discovery of underlying patterns that may or may now be perceivable to the average human being (Lawrynowicz & Tresp, 2014). Based on these patterns, a Machine Learning model will be able to classify or predict information. On this note, it is important to mention that one key component of an effective learning model is the ability to generalize results based on a set of training data points (Shalev-Shwartz & Ben-David, 2014).

When determining whether or not a particular problem must be solver using Machine Learning, it is important to remember that Machine Learning models tend to be very effective in helping with problems that are too complex to solve and also those that require adaptivity. Both of these characteristics can be hardly addressed using standard programming since their implications is an always-changing scenario that may not be included in the output standard coding can generate (Shalev-Shwartz & Ben-David, 2014).

There are two major types of Machine Learning models, supervised and unsupervised (see figure 9). The supervised models are those where there are clear inputs and outputs. In this way, the machine will have a clear output to base its learning on. In a nutshell, a supervised learning model is based on a formula identified through machine learning that converts an input into the observed output. Conversely, unsupervised models are those where there is a clear input but no output. In this type of model, the machine would not like to mimic an output, but group data based on the received input.

Figure 9

Unsupervised vs Supervised Machine Learning Models
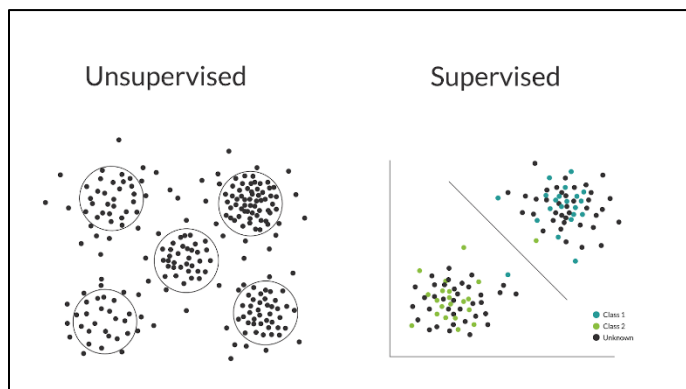
Source. (Wu, 2021)*.*

Machine Learning is considered a branch of artificial intelligence and has applications in many fields including economics, finances, medicine, and much more. In the last ten years, there have been a variety of developments supported by machine learning exercises such as self-driving cards, speech recognition, and web search algorithms. The applications for Machine Learning continue to grow and improve every day (Coursera, n.d.).

2.10.1 Supervised Machine Learning Models

Supervised Machine Learning models consist in algorithms aimed to generate predictive models to classify various information. According to Kotsiantis (2007), the development of supervised machine learning models generally follow an eight-step process: 1) stablishing the problem to solder, 2) identify data to use, 3) pre-process the data, 4) define a training set from the data, 5) select algorithm to use, 6) train model, 7) tune parameters, 8) evaluate model, 9) adjust as needed (see figure 10).

While there are several supervised models available, Kotsiantis (Kotsiantis, 2007) highlight five major types: 1) logic-based algorithms (such as decision trees), 2) perceptron-based algorithms, 3) statistical learning algorithms (such as Naïve Bayes), 4) instance-based learning algorithms, and 5) support vector machines. Considering all the options available, the user will need to carefully assess what model to use based on a series of performance metrics that include accuracy, precision, and recall. This author provides a set of recommendations based on their experience using different models:

- Support Vector Machines tend to perform better than other algorithms with multidimensional and continuous data sets.

- Naïve Bayes tend to require less space and therefore demand less resources from the computers being used to run this type of models.

- Logic-based algorithms tend to be relatively easy to be interpreted while neural networks tend to be the most complicated ones.

Figure 10

Process to Develop a Supervised Machine Learning Model
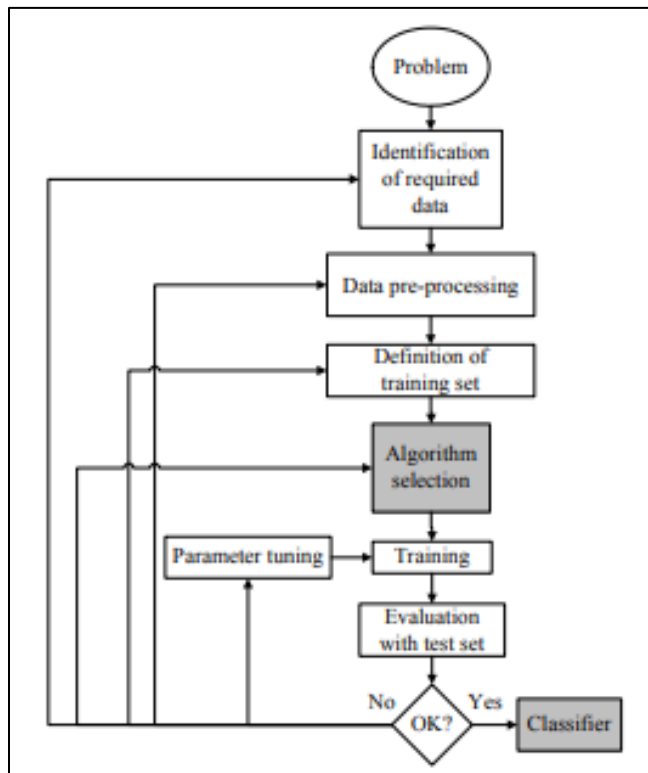
Source. (Kotsiantis, 2007).

2.11. Chapter two closing comments

In this chapter, the author provided concepts related to transborder dynamics, including transfronteridad, transmigration, border, and more. Additionally, concepts from Big Data, Machine Learning, Data Mining are briefly discussed in preparation for chapters three and four as they related to the research method and the results obtained.

Chapter 3. Method

3.1 Chapter three introductory comments

Chapter number three illustrates the methods used to answer the research questions identified and achieve the objectives established for this project. In addition to this information, the author includes the study hypotheses and an operationalization exercise used to determine the best way to measure the variables involved in this study. This chapter also includes a review of sociodemographic traits from the population of interests; this is considered a critical piece from the study since it required the manipulation of more than 10 million records provided by both the Mexican and U.S. Census Bureaus through the usage of Big Data processing techniques. The chapter also includes a review of the instrument used to collect data from transborder individuals and some highlights from the survey campaign process.

3.1. Objectives

3.1.1. Central Objective

Determine if multicultural competency, language proficiency, and educational attainment have a significant influence on the levels of transfronteridad by implementing supervised machine learning algorithms based on regression to generate knowledge that allows the development of public policies.

3.1.2. Specific Objectives

1. Collect public information about the transborder population using statistical data collection methods to generate information related to the transborder population.

2. Create a survey for transborder population using concepts from questionnaire design, research methodology, and sampling to generate information related to this group.

3. Analyze public information and the results from the survey using descriptive statistical and big data methods and techniques to characterize the transborder population.

4. Define the construct of multicultural competency by reviewing the literature related to border theory and determine the variables that must be to measure and include it the supervised machine learning model.

5. Analyze multicultural competency and the levels of transfronteridad using Random-Forest-type supervised machine learning algorithms to determine if there is a statistically-significative correlation between them.

6. Compare the scores assigned by the Random Forest algorithm for each of the multicultural-competency-relates attributes by interpreting results to determine the one with the highest impact on the levels of transfronteridad.

7. Generate a compendium of results by analyzing charts derived from the Random Forest algorithm to develop public policies.

3.2. Hypotheses

This research project operates under the system of hypotheses, with conceptual and operational definitions stated in table 6:

Null Hypotheses H_0 - There is not a statistically significant correlation between **multicultural competency** and **transfronteridad**. This means that the supervised

machine learning regression model will not show significant scores for each of the multicultural competency identified and for the overall model statistic F.

Alternative Hypotheses H_0 - There is a statistically significant correlation between **multicultural competency** and **transfronteridad**. This means that the supervised machine learning regression model will show significant scores for each of the multicultural competency skills identified and for the overall model statistic F.

3.3. Methodological Design

This research will follow a quantitative approach with a survey that will be deployed to collect data. Even though different statistical methods will be implemented to describe and infer traits from the population, it will all be done under an exploratory approach. The author will pull data at a given point in time and will not give any kind of treatment to the data from the individuals involved in it; therefore, this study can be categorized as a non-experimental, cross-sectional (Hernández, 2014). The author will use the following methods to analyze the population of interest:

- Frequencies and descriptive statistics to characterize the transborder population.

- Regression analyses to identify correlations and incidence ratios for different traits in our population.

- Machine supervised learning models to build multivariable model to predict transfronteridad.

Table 6

Conceptual and Operational Definitions for Variables Included in Hypotheses

Variable	Definition	Questions in survey	Role in model
Transfronteridad	The "level of exchange and dependency with one or the other side [Mexico and the U.S.]. The quantity, frequency, intensity, direction, and magnitude of the material and symbolic exchange, to both places" [Mexico and the U.S.] (Iglesias-Prieto, 2010; translated from Spanish).	Q5: Country or countries in which respondents do different activities	Dependent variable
Multicultural Competency	"Knowledge, skills, and personal attributes needed to live and work in a diverse world" (Kansas State University, 2001).	Q7: *Level of a*greement to culture-related statements Q10*: E*ducational attainment Q11 & Q12*: L*anguage proficiency	Independent variable

Source. Author´s elaboration.

3.4. Population

3.4.1. Population Description

The population of interest for this study are transborders 18 years old or older living in the Mexico-U.S. border. While there is no data nor methodology available that could guide the author to a solid estimate for the number of transborders in the Mexico-U.S. border, it is possible to get an approximate by using public data available from the Mexico's National Institute for Statistics and Geography (INEGI for its acronym in Spanish) and the U.S. Census Bureau. According to INEGI, there are 89,391 adult people currently living in Mexico while working in the U.S. A little more than eleven thousand adult living in Mexico (11,126) are also studying in the U.S. In addition to this, there are 84,865 with American citizenship which can also be considered transborders due to their *legal ability* to go back and forth between borders. Finally, data available from INEGI also show that 45,961 adult individuals currently residing in one of the border cities from Mexico, was living in the U.S. five years ago (see table 7). This particular group was also included in the estimate as long as the reason why the stopped leaving in the U.S. was not a deportation assuming that they continue to be able to go back and forth between borders -important to remember that having experiences living in both sides of the border have an important impact in the ability people has to live transborders lifestyles (Orraca-Romano, 2019). Combining all these groups, and considering overlapping (i.e., one individual may be a commuter worker while also being an American citizen), the estimate from the Mexican side ascends to 168,698. On the other side of the border, according to the U.S. Census Bureau, there are 4,180,765 people with Mexican ascendance. The author needs to point out that there is no data available to confirm that all the Mexican-

ascendent individuals living in a U.S. border are crossing the border; therefore, not all should be classified as transborders -factors such as immigration status and the lack of roots in Mexico make the author question this assumption. However, the author decided to move forward with it to generate a draft estimate considering that crossing the border is a recurring activity among Hispanics living in the U.S. border, and illegal immigration tends to happen at outside of the border since illegal immigrants try to stay as far as the can from the border. The number of transborder may also be underrepresented even making this assumption when considering that a number of non-Hispanic U.S. inhabitants participating in the border milieu (Martínez, 1994). In addition to that, it is important to also consider that the number of transborders calculated is not including the transborder living in Mexico crossing to the U.S. using Visas, which easily range in the millions considering the number of Visas the United States authorizes every year for Mexican citizens; for context, in average, more than 1.3 million Mexican people received a visa every year in the timeframe 2011 – 2019 (U.S. Bureau of Consular Affairs, 2020).

The following tables show different sociodemographic information of our populations including a breakdown by country of residence, level of education, age, gender, and state of residence.

According to table 7, individuals including in the Mexican transborder group skew male with 62%; a little more than half of them are young (less than 34 years; 53%). The majority of the individuals did not have college education (77%). Finally, transborders were distributed in their majority in Baja California with Chihuahua in the second place.

Regarding Mexican American individuals considered transborders, we see a very close split between male and female individuals. This group skewed a little older than the

Mexican one with only 39% falling under 35 years old. Like the Mexican group, the majority of individuals did not have college experience though the percentage was lower than with the Mexican group (57%). Regarding the state of residence, we see a that most of the Mexican American transborders were residing in Texas with Arizona as the second.

3.4.2. Big Data Processing Results

In order to build the tables above, the author required to work with information provided by both the Mexican and U.S. Census Bureaus. Both sources provided information to describe the population of interest in the form of csv files that were appended using SPSS. Two datasets were built; one to describe transborders residing in Mexico, and one to describe transborders in the U.S.

The dataset for transborders residing in Mexico were built by downloading the results from the Mexican Census of Housing and Population for 2020. This was a csv file that contained more than 15 million of records showing transborder sociodemographic traits. In addition to using the file provided by the Mexican Census Bureau, the author became familiar with the dataset structure using data dictionaries and a code catalog provided by the institution. As for the dataset for transborder residing in the U.S. side, the author downloaded a series of datasets from the 2019 American Community Survey. This was two datasets containing more than 3 million records from U.S. households. In order to combine and extract information from these datasets, the author became familiar with the datasets structure and other important information. Screenshots from the files required to build the datasets are shown in figure 11.

Table 7 (1/2)

Population Sociodemographic Traits (People Residing in the U.S.)

Variable	n	%
Residing in the U.S.	4,180,765	100
Sex		
Female	2,141,401	51
Male	2,039,364	49
Age Group		
18 - 24	717,137	17
25 – 34	916,005	22
35 – 44	789,703	19
45 - 49	358,252	9
50 – 54	318,929	8
55 – 64	486,109	12
65 or more	594,630	14
Educational Attainment		
Elementary or less	268,011	7
Middle School	313,417	8
High School	1,597,201	40
Less than two years in College	957,159	24
Two or more years in College	844,307	21
Educational Attainment		
State of Residency		
Arizona	1,312,542	31
California	823,213	20
New Mexico	210,814	5
Texas	1,834,196	44

Table 7 (2/2)

Population Sociodemographic Traits (People Residing in Mexico)

Variable	n	%
Residing in the Mexico	168,698	100
Sex		
Female	64,243	38
Male	104,455	62
Age Group		
18 - 24	42,140	25
25 – 34	47,449	28
35 – 44	26,479	16
45 - 49	14,763	9
50 – 54	12,804	8
55 – 64	15,309	9
65 or more	9,754	6
Educational Attainment		
Elementary or less	17,813	11
Middle School	36,720	22
High School	77,040	46
Less than two years in College	4,209	3
Two or more years in College	32,476	19
State of Residency		
Baja California	86,880	52
Coahuila	6,671	4
Chihuahua	29,556	18
Nuevo León	315	<1
Sonora	19,862	12
Tamaulipas	25,414	15

Note. Percentages may not add up to 100% due to rounding error.

Source. Author´s elaboration with data from INEGI, 2021 and the U.S. Bureau of Statistics, 2020.

3.5. Ethical Considerations

In terms of consequences, this research contemplates conducting one survey, which does not involve any kind of treatment on the population of interest; therefore, no negative effects are expected. More on this, information from the survey will be anonymous and presented as aggregates only to protect the participants' privacy. In general terms, there are no adverse consequences that may be expected from this survey considering that the instrument will be an confidential online form; thus, there are no extraordinary ethical considerations to assess.

Figure 11

Big Data Processing Results Screenshots
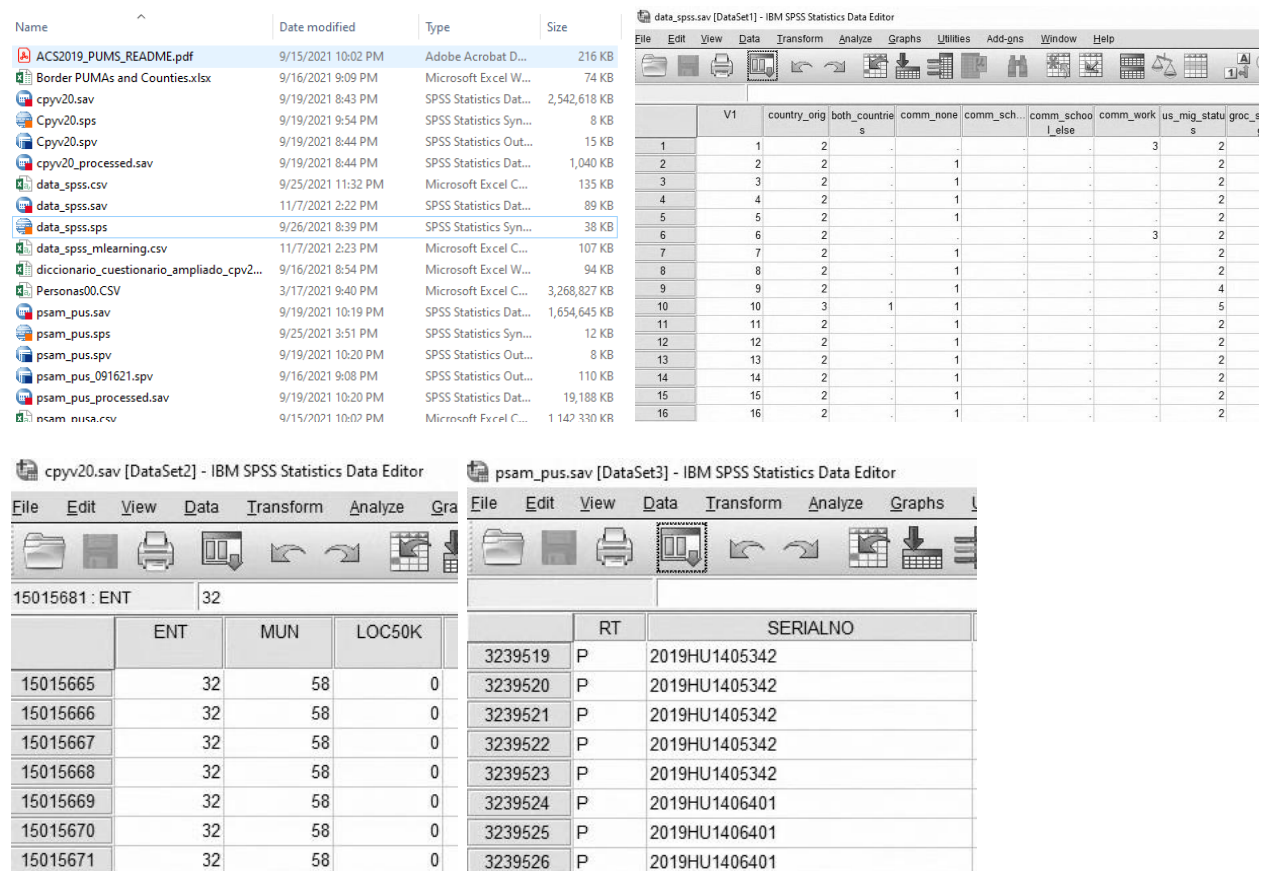
Source. Author's elaboration.

3.6. Instrument

The instrument consists of five sections (see appendices A and B); a) questions to identify were individual lives, b) questions to identify what kind of transborder the individual is, c) questions to identify which activities are done in either country, d) questions to identify what languages individual uses, e) questions to explore cultural identity, and f) demographic questions.

The following steps were followed to develop and test the quantitative instrument:

1. Operationalized hypotheses

2. Developed draft questionnaire

3. Reviewed and improved questionnaire based on developer's knowledge

4. Programmed questionnaire in survey platform -while programming was occurring, improvements were made

5. Tested questionnaire in survey platform -while testing was occurring, improvements were made

6. Conducted cognitive testing with 10 individuals to measure various aspects including, survey fatigue and question ambiguity. After every test run, the following questions were made to the respondents to inform improvements required:

   a. Timing of the survey [reviewed online]

   b. General impressions from the survey

      i. Was it long?

      ii. Was it complex?

      iii. Did you feel any of the questions did not make sense for you?

iv. Did you think any of the questions was not clear enough?

 v. Was it boring?

vi. Was it too personal?

vii. Was there any topic you thought we were going to ask about, but

we did not?

c. Quick discussion of the answers provided

d. Discussion regarding how this would be answered if they did not know

developer, also listen to suggestions about ways to improve response

rate and in general improve questionnaire

7. Developed English version and also conducted 3 cognitive testing sessions

8. Piloted with 40 individuals from five Facebook groups for transborders identified

to measure timing, ambiguity, sample variance, and questions that could harm

the response rate (i.e., making folks leave survey). In addition to that, the author

was also interested in measuring the levels of participation from Facebook

group members to ensure it was viable to collect a higher number of responses

during the live stage of the survey

9. Analyzed responses from piloting stage to measure for reliability using a

Cronbach Alpha analysis

3.6.1. Instrument Validity

In order to ensure the survey had reliability, The author performed a Cronbach

Alpha analysis using the results from the piloting strategy to determine is the scale for

transfronteridad was reliable -a key scale for this project's objective. The results (.843)

show a very good level of reliability considering that values above .7 are considered positive for Cronbach.

3.7 Sample Size

To determine the sample size, the author used a formula based on the following parameters: a) a deviation standard that was established at .5 to maximize the sample size, a confidence interval set at 95% that gets translated to 1.96 assuming a normal distribution for the variable of interest, a 5 percent margin of error, and a total population of 4,349,463 (Hernández, Fernández, & Baptista, 2014). By doing this, results will become representative for all the population's traits, including those that are more heterogeneous (see figure 8). Results show that from a population of 4,349,463, a total of 384 people should complete this project's survey to ensure results are representative. The author advises the reader to remember the central limit theorem concepts that establish that getting closer to infinite, population increases will not incur in sample size increases at the same level. This is important to consider since not all Mexican transborders were identified in the population, yet the author is comfortable using this sample size based on the concepts stablished (see figure 12).

Figure 12

Sample Calculation

$$n = (\frac{z_{a/2}\sigma}{E})^2 = (\frac{1.96 \; x \; 0.50}{0.05})^2 = (\frac{0.98}{0.05})^2 = (19.6)^2 = 384.16$$

Source. Author's elaboration based on (Anderson, Sweeney, & Williams, 2013).

It is important to mention that it was not possible to implement a practice known a sample fixation due to not have a solid or official estimate about the number and composition of the transborder population. The author preferred not to go through this

107

process using the estimate that it was developed to avoid intentionally skewing the sample. In addition to that, the method, as stated below, consists of multivariable statistical methods that may not be affected by potential skewness in the sample results considering that we are mostly interested in the internal correlation between variables for each individual not between them.

3.8. Field Activities

3.8.1. Field Activities Description

The field activities for the survey were done by the author in four stages. The first stage consisted in identifying all the groups in Facebook for transborders and submitted request to become a part of them. The Facebook groups covered all the Mexico-U.S. transborder regions and ranged from 300 to almost 400,000 members (see figure 13). More than 100 Facebook groups from transborders were identified, with more than 60 registration requests being accepted (see figure 13). The second step consisted in contacting a little less than a third group administrators (22 groups were contacted) to request their permission to post the survey link (see figure 13). Strategically, the Facebook groups contacted were the smallest ones in membership to test the acceptance rate from administrators and determine if changes in the messaging used to request permission needed to be done. Out of these 22 requests, five granted the permission necessary to post the survey (see figure 13). The thirds stage consisted in posting the survey link in these five groups only once. In addition to collecting data for piloting purposes, the author was interested in measuring the participation rate and potential problems from survey respondents. The fourth stage consisted in contacting all the remaining administrators to request the same permissions. Finally, the fifth stage

consisted in posting the link in all the groups where the author received permission to go

with the final data collection (see figure 13).

Figure 13

Facebook Groups and Survey Campaign Work Screenshots
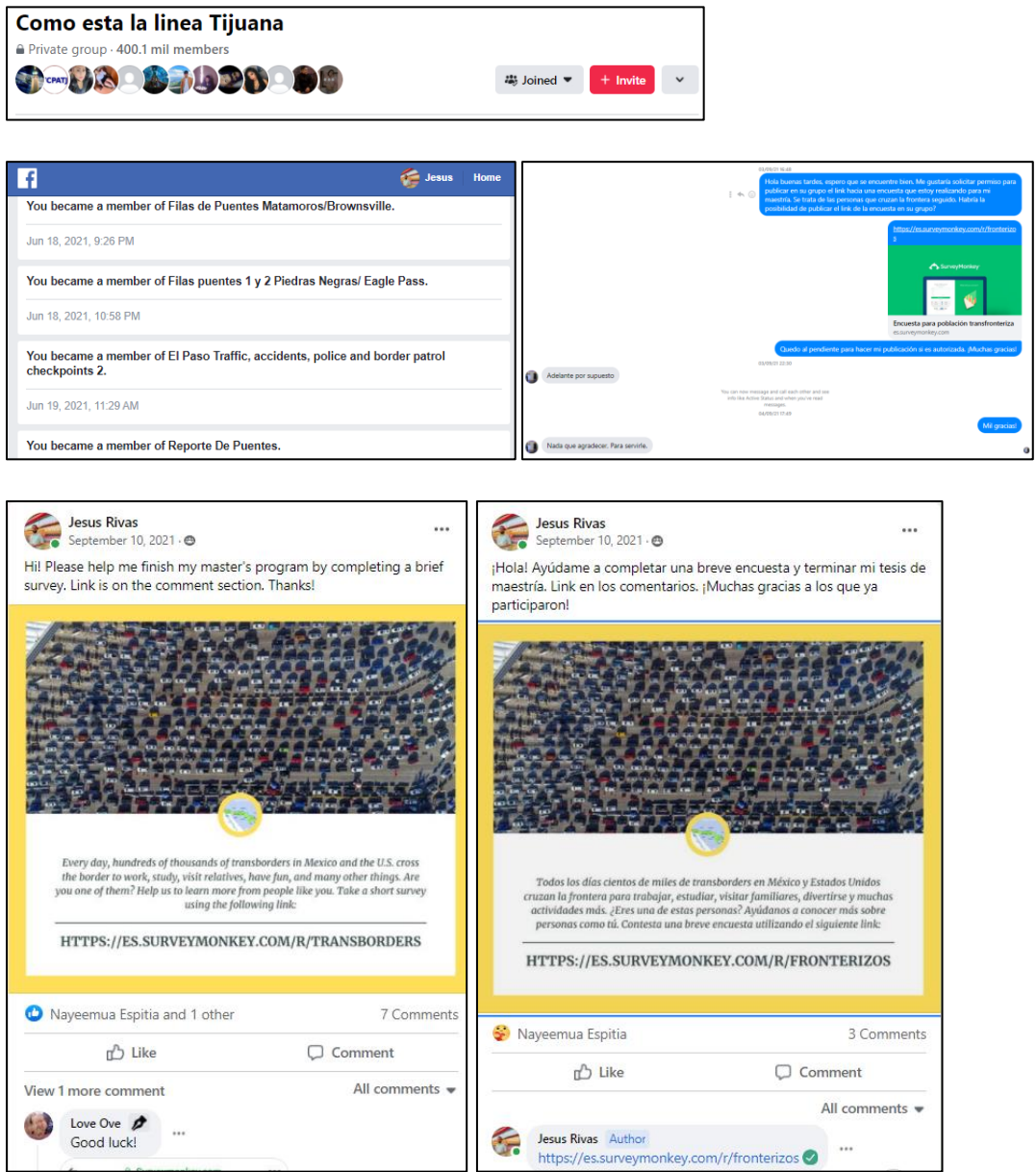
Source. Author's elaboration.

3.8.2. Survey Campaign Results

The piloting stage for the survey was done in the first week of July while the full launch and survey campaign occurred between August and September of 2021. At the end of this process 541 responses were collected (see Figure 14). However, after cleaning the data, 139 survey responses were eliminated due to quality issues; specifically, these responses were incomplete, inconsistent, or showed hints of straightlining.

Figure 14

Field Results Screenshots

Source. Author's elaboration.

3.9. Data Processing and Software Usage

The survey was programmed using survey monkey, a survey platform that allows individuals to create, launch, and manage survey projects with various survey features, including survey logic, formatting, response forwarding, and more. In order to create two versions for the survey -Spanish and English, two different survey projects were created in this platform and different survey links generated (see figure 15). Once the survey campaign was over, the results from both surveys, were pulled and processed using R

studio. In order to do this, the author used R packages such as psych, gmodels, expss, dplyr, and more (see figure 16). After the survey results were processed, SPSS was used to generate a set of descriptive tables, crosstabs and regressions. Additional processing was required to account for differences in the way R Studio and SPSS process information (see figure 16). Finally, a set of Machine Learning Supervised Techniques was use through R Studio to generate various predictive models to explore the relationship between transfronteridad and a set of variables. In order to do this, the author used R packages such as pshych, randomFores, ROCR, e1071, caTolls, glmnet, rept, ggplot2, and caret see figure 16).

Figure 15

Survey Monkey's Interface Screenshots

Source. Author's elaboration.

Figure 16

Data Processing Work and Software Usage in R Studio and SPSS

Source. Author's elaboration.

3.10. Chapter three closing comments

In a broad sense, this study aims to analyze the relationship between transfronteridad and a set of sociodemographic variables to understand better transborder dynamics occurring at the Mexico-U.S. border. To accomplish this, a quantitative method will be used based with emphasis on developing a supervised machine learning regression model to predict transfronteridad levels.

Chapter 4. Results

4.1. Chapter four introductory comments

In chapter four, the author presents the results from the analyses conducted based on the research design previously discussed. There are five major sections in this chapter: a) a review of the basic results from the descriptive analysis of the information collected, b) a review of the results from bivariable analysis made using crosstabs, c) a review of the results from the scale development exercise -used to develop scales for multiculturality and transfronteridad, d) a review of the results from the multivariable regressions created to determine if transfronteridad is statistically correlated to language, multiculturality and educational attainment, and e) a review of the results from the development of a machine supervised learning model to predict levels of transfronteridad using more than ten variables.

4.2. Basic Results

The results from the survey show that respondents used in majority the Spanish version of the survey (93%) while the rest picked the English version (see table 8). It is important to consider that this survey was posted in more than 60 Facebook groups composed by transborders, so there may be impact from the selection criteria being reflected here. Still, as stated before, this should be considered an exploratory exercise and future exercises will be pursued to stablish a valid sample framework that could guide other iterations of this survey to ensure representativity. From the respondents, 58% stated their country of residence in Mexico while the rest in the US. In terms of immigration status, 46% of the respondents were American citizens, 41% U.S. Visa holders, 11% American Legal Residents, 1% Work Visa Holders, and less than 1% did not have the

proper documents to enter the U.S. -though these individuals were terminated in the survey due to not qualifying for the study. Throughout this document, the author presented different categories under which is possible to classify transborder individuals -some of these were calculated using the results from this survey. According to them, 41% of the respondents could be consider American Consumers (the author did not create a distinction between Mexican-American and Anglo Consumers as some authors do), 41% were classified as Mexican Consumers, 14% Working Commuters, and 4% School Commuters. The majority of the survey respondents were bilingual with 58 percent. From the remainder, 41% were Spanish speakers and 1% English speakers. In terms of Mexican ascendance, 90% of the respondents indicated they had it. In general terms, results show that the survey provided a diverse picture of all the transborder population considering there is variations for all the variables included in the survey which will be positive for the regression analysis.

Table 8 includes sociodemographic information. The sample of respondents skewed female with 65% of respondents selecting that gender. The participants were mostly under 65 years old with two or more years in college (67%). Most of them had lived in both countries (51%), with an important proportion living only in Mexico (50%). Additionally, respondents indicated in their majority having only worked in one country (Mexico 41% of the total respondents and the U.S. 30%). From the remaining respondents, 29% indicated they had worked in both countries and less than 1% of the respondents had not worked in either country. Additional sociodemographic information collected in the survey show that respondents had student in only one country mostly (50% for Mexico and 17% for the use). From the remaining respondents, 34 percent of

individuals had studied in both countries. Respondents also indicated the frequency they were crossing the border with a very balanced distribution. According to the results of the survey, 12 percent of the respondents had been crossing the border with the frequency stablished for less than a year; 16% from one to three years; 10% from three to five years, and 63% more than five years. Eighty two percent of the respondents were living in a border city. From all the respondents, 53% were living in the California – Baja California border region; 14% in the Arizona – Sonora region; and 33% in the New Mexico/Texas – Chihuahua/Coahuila/Nuevo Leon/Tamaulipas region.

A different view for the prior question indicates that individuals tend to do some activities more than others; only in their countries, in the opposite one or in both. While the results for activities not done in either country or both of them will not change by the definition of this metric, with it is possible to see that individuals tend to vote only in their country of residence which can be expected considering the immigration implications voting requirements exist (65 percent; see table 9).

The survey included questions regarding the activities individuals were doing on each side of the border since that information was used to create an indicator to measure the construct for transfronteridad. According to the results of the survey we can see that the most popular activity to be done in the U.S. regardless of the country or residence was to buy clothes (73 percent; see figure 17); in Mexico was getting a haircut (75%); and in both countries it was doing something fun (70%).

Table 8

Sociodemographic and Other Characteristics of Survey Respondents (1/3)

Characteristic	n	%
Source of Information		
Survey Spanish Version	373	92.8
Survey English Version	29	7.2
Country of Residency		
United States	166	41.3
Mexico	236	58.7
Immigration Status in the U.S.		
No documents allowing legal entry to the U.S.[a]	3	.7
Tourist-visa-holder	166	41.3
Work-visa-holder	5	1.2
United States Resident	43	10.7
United States Citizen	185	46.0
Transborder Category[b]		
American Consumer	166	41.3
Mexican Consumer	164	40.8
Work Commuter	56	13.9
School Commuter	16	4.0
Language Proficiency		
Spanish-only Speaker	162	40.6
English-only Speaker	4	1.0
Bilingual	233	58.4
Mexican Ascendancy Status		
Yes	350	90.4
No	37	9.6
Gender		
Female	260	65.2
Male	137	34.3
Prefer not to answer	2	.5

Table 8

Sociodemographic and Other Characteristics of Survey Respondents (2/3)

Characteristic	n	%
Age Group		
Less than 18 years	1	.3
18 – 24	33	8.3
25 – 34	90	22.6
35 – 44	94	23.6
45 – 49	51	12.8
50 – 54	35	8.8
55 – 64	64	16.0
65 or more	31	7.8
Education Level		
Elementary of less	3	.8
Middle School	13	3.3
High School	56	14.0
Less than two years in college	60	15.0
Two or more years in college	267	66.9
Countries where respondent has lived		
Only in the United States	37	9.3
Only in Mexico	159	40.1
Both countries	201	50.6
Residency in border city		
Yes	73	18.3
No	326	81.7
Countries where respondent has worked		
Only in the United States	119	29.9
Only in Mexico	162	40.7
Both countries	115	28.9
None	2	.5

Table 8

Sociodemographic and Other Characteristics of Survey Respondents (3/3)

Characteristic	n	%
Countries where respondent has studied		
Only in the United States	67	16.8
Only in Mexico	196	49.2
Both countries	135	33.9
Border-crossing Frequency		
More than three times per week	79	19.8
2 – 3 times per week	81	20.3
1 time per week	90	22.6
Every 10 – 15 days	57	14.3
Once per month	50	12.5
Less than once per month	42	10.5
Number of years crossing the border		
Less than a year	46	11.6
From one to three years	62	15.6
From three to five years	41	10.3
More than five years	248	62.5
Border region		
California – Baja California	213	53.4
Arizona – Sonora	154	13.9
New Mexico and Texas – Chihuahua, Coahuila, Nuevo Leon, and Tamaulipas	131	32.9

Note. Missing values are excluded (402 total respondents).

[a] Respondents who answered this option were terminated in the survey.

[b] Respondents residing in Mexico who fell in more than one category were reclassified into one using the following hierarchy: 1) Work Commuter, 2) School Commuter, and 3) Mexican Consumer.

Figure 17

Summary Table. "In which country or countries do you typically do the following activities?

■U.S. ■México ▫Both ▫None

Activity	U.S.	México	Both	None
Buy clothes	73%	1%	26%	1%
Buy groceries	32%	11%	57%	<1%
Vote	28%	49%	14%	10%
Go on a road trip	23%	14%	60%	3%
Go to a concert	16%	18%	47%	19%
Go to the doctor	15%	61%	23%	1%
Eat at a fancy restaurant	13%	27%	54%	5%
Visit friends	13%	20%	66%	1%
Go on vacations	13%	20%	64%	3%
Go to movie theater	12%	54%	25%	9%
Go to church	12%	42%	14%	32%
Follow the news	11%	20%	65%	3%
Do something fun	11%	17%	70%	2%
Eat street/fast food	11%	21%	67%	2%
Get a haircut	10%	76%	10%	3%
Visit family	8%	23%	69%	<1%
Go to party	7%	29%	62%	3%

Percentage
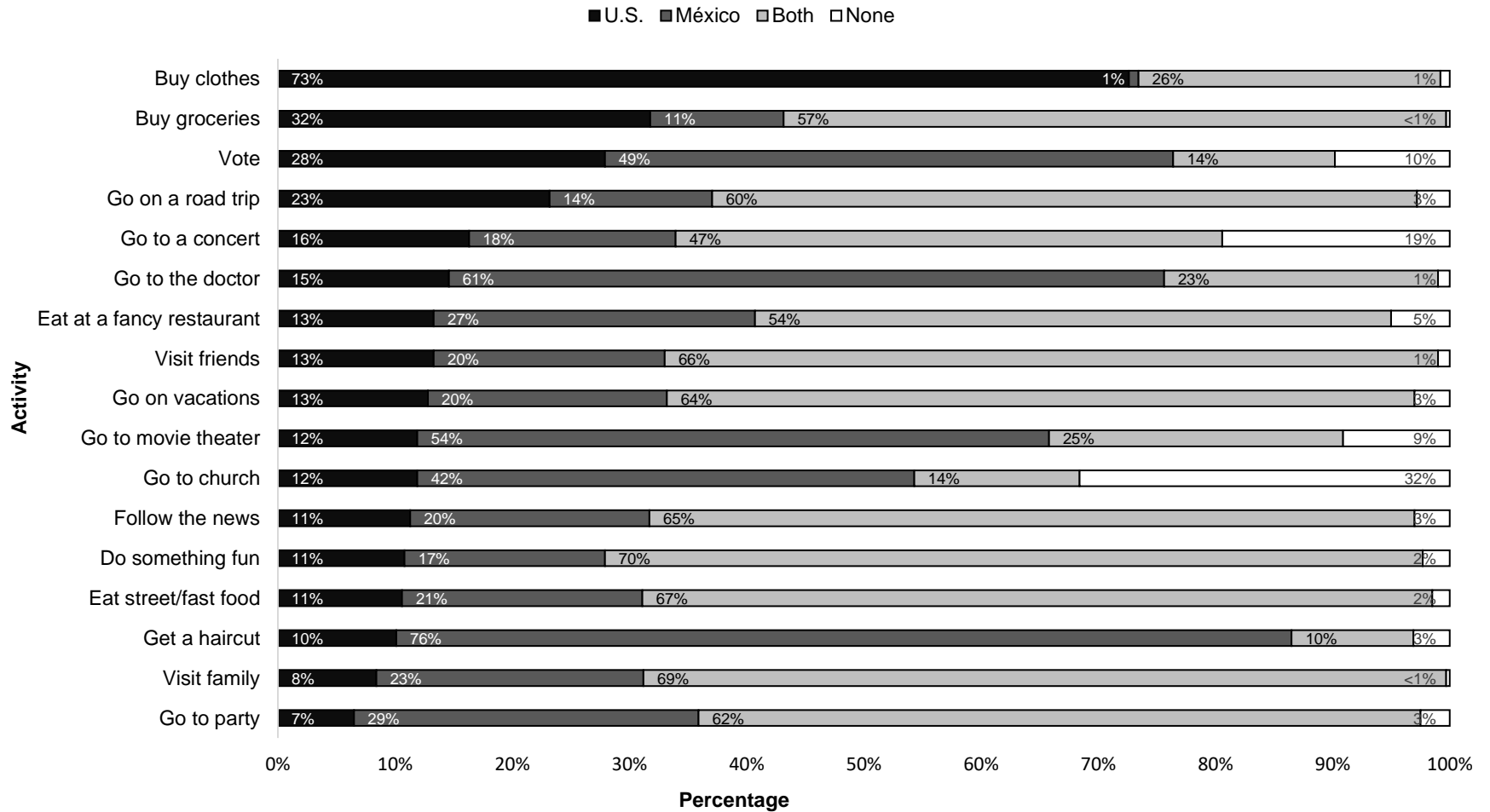
Note. Missing values are excluded (402 total respondents).

Figure 18

Summary Table. "In which country or countries do you typically do the following activities? (Recoded results)

■Country of residence　■Country where they do not reside　▨Both countries　□None

Activity	Country of residence	Country where they do not reside	Both countries	None
Vote	65%	12%	14%	10%
Get a haircut	64%	23%	10%	3%
Go to the doctor	58%	17%	23%	1%
Go to movie theater	54%	12%	25%	9%
Go to a church	45%	9%	14%	32%
Buy clothes	36%	37%	26%	1%
Buy groceries	32%	11%	57%	<1%
Go to a party	27%	9%	62%	3%
Follow the news	25%	7%	65%	3%
Go to a fancy restaurant	24%	17%	54%	5%
Visit friends	23%	9%	64%	4%
Go on a road trip	21%	16%	60%	3%
Go to a concert	20%	14%	47%	19%
Eat street/fast food	19%	12%	67%	2%
Do something fun	17%	11%	70%	2%
Go on vacations	16%	17%	64%	3%
Visit family	16%	14%	67%	3%

Note. Missing values are excluded (402 total respondents).

Figure 19

Summary Table. "Indicate the country or culture more adequate based on your way of thinking"
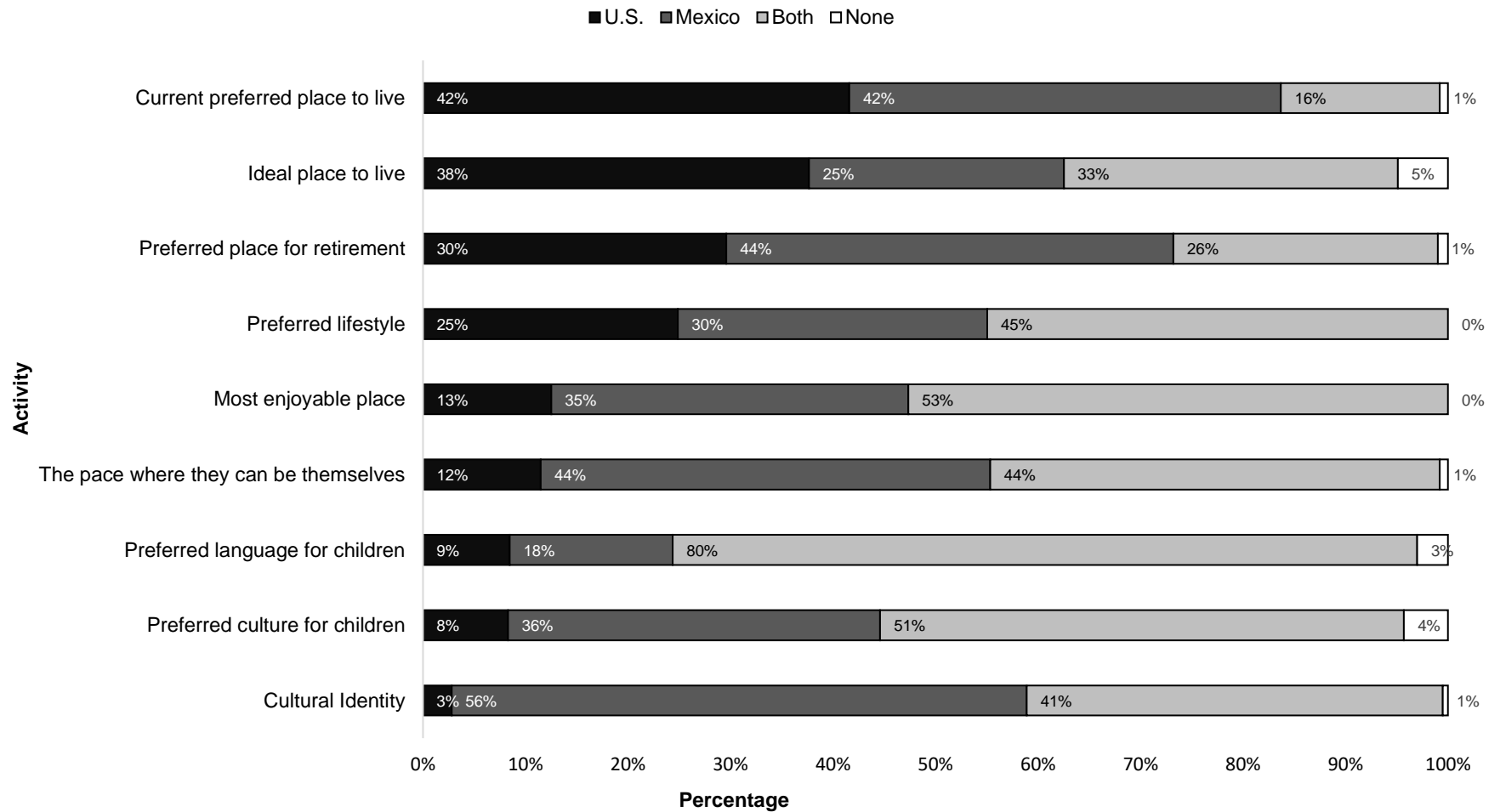
■U.S. ■Mexico ▨Both ▫None

Activity	U.S.	Mexico	Both	None
Current preferred place to live	42%	42%	16%	1%
Ideal place to live	38%	25%	33%	5%
Preferred place for retirement	30%	44%	26%	1%
Preferred lifestyle	25%	30%	45%	0%
Most enjoyable place	13%	35%	53%	0%
The pace where they can be themselves	12%	44%	44%	1%
Preferred language for children	9%	18%	80%	3%
Preferred culture for children	8%	36%	51%	4%
Cultural Identity	3%	56%	41%	1%

Percentage

Note. Missing values are excluded (402 total respondents). Results are sorted by percentages for the U.S.

121

In the survey, a section regarding cultural attitudes was included to measure the construct of multicultural skills. According to the results of the survey, respondents did not show a clear pattern. Perhaps, it is possible to note the fact that questions related to being open to other cultures, and languages ranked at the top in terms of respondents indicating they are interested in both cultures. On the other side we could see more concentration towards the Mexican culture when indicating the culture, the followed or identified with.

4.3. Crosstabs results

Considering that respondents could have lived in either side of the country and their differences may have important implications to the results we ran some crosstabs to control for this situation considering that the survey was not using a sample framework to weight results. Respondents were divided in three groups according to (Martínez, 1994). It is important to mention that this classification system does not include all the groups created but the author but only the ones we collected evidence for.

When crossing variables, we can tell that the American consumer tend to buy their clothes in the US (86 percent; see figure 20); buy groceries in either the U.S. or both countries but not in Mexico exclusively (99%); vote in the U.S. during the elections (54%), go on road trips in both countries (57%), and go to concerts in both countries with more frequency (48%); go to the doctor in both countries with more frequency (39.2%); go out to eat at a "fancy" restaurant in both countries (59.6%); visit friends in both countries (64%); go on vacations (64%) and go to the movie theaters in both countries with more frequency that in the other options (34%), not go to church (32%), follow the news from both countries (68%), do something fun in both countries (72%), go out to eat "on the

street" or fast food (72%), get a haircut in Mexico (53.5%), visit family members in both countries (71.2%), and go to a gathering or a party in both countries (68%).

For the "Mexican Consumers" group, results show that they tend to buy their clothes in the US (62%); buy groceries in both countries (61%); they vote in the Mexico during the elections (89%); they go on road trips in both countries (64%); go to concerts in both countries with more frequency (41%); they go to the doctor in Mexico (90%); they go out to eat at a "fancy" restaurant in both countries (51%); visit friends in both countries (67%); go on vacations in both countries (64%); go to the movie theaters in Mexico with more frequency that in the other options (75%); go to church in Mexico (63%); follow the news from both countries (60%); do something fun in both countries (69%); go out to eat "on the street" or fast food (62%); get a haircut in Mexico (94%); visit family members in both countries (66%); and go to a gathering or a party in both countries (59%).

Finally, for the "Work Commuters" group, results show that they tend to buy their clothes in the US (71%); buy groceries in both countries (66%); they don't vote as frequently as the other groups and those who do it were similarly represented in Mexico and the U.S. (29 and 27 percent respectively); go on road trips in both countries (61%); go to concerts in both countries with more frequency (48%); go to the doctor in Mexico (57%); go out to eat at a "fancy" restaurant in both countries (45%); visit friends in both countries (65%); go on vacations in both countries (57%); go to the movie theaters in Mexico with more frequency that in the other options (77%); not go to church (52%); follow the news from both countries (75%); do something fun in both countries (63%); go out to eat "on the street" or fast food (64%); get a haircut in Mexico (89%); visit family members in both countries (67%); and go to a gathering or a party in Mexico (54%)

123

Figure 20

Summary Table. "In which country or countries do you typically do the following activities?(Among American Consumers)

■U.S. ■México ▨Both ▢None

Activity	U.S.	México	Both	None
Buy clothes	86%	<1%	14%	1%
Vote	54%	15%	19%	13%
Buy groceries	50%	6%	49%	<1%
Go on a road trip	29%	12%	57%	3%
Go to the doctor	27%	34%	39%	<1%
Go to movie theater	27%	27%	34%	12%
Go to church	27%	21%	20%	32%
Get a haircut	23%	54%	20%	4%
Visit friends	22%	12%	64%	2%
Follow the news	20%	9%	68%	3%
Go to a concert	18%	12%	48%	21%
Eat at a fancy restaurant	14%	28%	60%	5%
Do something fun	13%	13%	72%	2%
Eat street/fast food	12%	15%	72%	1%
Go on vacations	11%	28%	64%	3%
Go to party	11%	18%	68%	3%
Visit family	7%	22%	71%	<1%

Note. Missing values are excluded (156 total respondents). Results are sorted by percentages for the U.S.

Figure 21

Summary Table. "In which country or countries do you typically do the following activities?(Among Mexican Consumers)

■U.S. ■México ▫Both ▫None
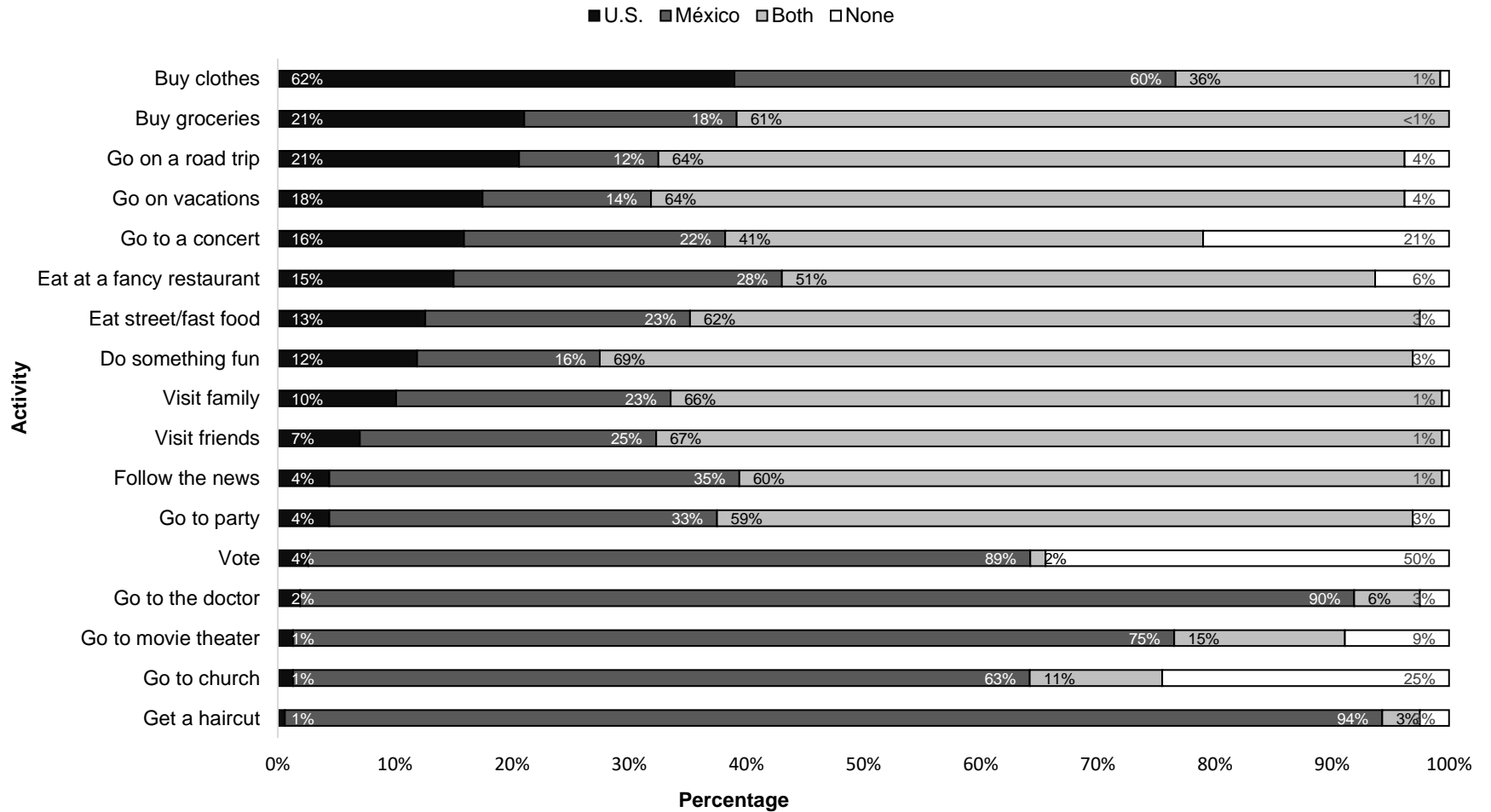
Note. Missing values are excluded (161 total respondents). Results are sorted by percentages for the U.S.

Figure 22

Summary Table. "In which country or countries do you typically do the following activities?(Among Work Commuters)

■U.S. ■México □Both □None

Activity	U.S.	México	Both	None
Buy clothes	71%	4%	25%	<1%
Vote	27%	29%	30%	14%
Go to the doctor	16%	57%	27%	<1%
Go to a concert	15%	24%	48%	13%
Go on a road trip	14%	25%	61%	<1%
Buy groceries	13%	21%	66%	1%
Visit family	9%	24%	67%	<1%
Eat at a fancy restaurant	9%	43%	45%	4%
Visit friends	7%	28%	65%	<1%
Follow the news	7%	9%	75%	9%
Go on vacations	5%	36%	57%	2%
Get a haircut	4%	89%	4%	4%
Do something fun	4%	34%	63%	<!%
Eat street/fast food	4%	30%	64%	2%
Go to movie theater	<1%	77%	20%	4%
Go to church	<1%	43%	5%	52%
Go to party	<1%	54%	46%	<1%

Percentage

Note. Missing values are excluded (66 total respondents). Results are sorted by percentages for the U.S.

In table 9, the author shows the activities done by transborder individuals sorted by country where the activity is done. According to the results, American Consumers' top three activities in the U.S. are buying clothes, voting during elections, and buying groceries (86, 54, and 50% respectively. Their top three in Mexico are getting haircuts, going to the doctor, going out to eat on "fancy:" restaurants and going on vacations (54, 34, and a tie for 28 percent for the last two respectively). Their top activities done in both countries were going out to eat "on the street" or fast food, doing something fun, and visiting family members (72, 71, and 71 percent respectively). Finally, their top activities not done were going to church, going to concerts, and voting during the elections (32, 21, and 13 percent respectively).

For the "Mexican Consumers" their top activities were as following. In the U.S. were buying clothes, going on road trips, and buying groceries (62, 21, and 20 percent respectively). In Mexico their top activities were getting a haircut, going to the doctor, and voting during the elections (94, 90, and 89 percent respectively). In both countries, their top activities were doing something fun, visiting friends, and visiting family members (69, 67, and 66 percent respectively). Finally, their top activities not done were going to church, going to a concert, and going to movie theaters (25, 21 and 9 percent respectively).

For "Work Commuters" their top activities in the U.S. were buying clothes, going to the doctor, and going to concerts (71, 16, and 15 percent respectively). In Mexico, their top activities were getting haircuts, going to the movie theater, and going to the doctor (89, 77, and 57 percent respectively). In both countries their top activities were following the news, visiting family members, and buying groceries (75, 67, and 66 percent respectively).

Table 9

Cross table. "In which country or countries do you typically do the following activities?

Results by Transborder Group (1/4)

Activity	American Consumers		Mexican Consumers		Work Commuters	
	n	%	n	%	n	%
United States						
Buy clothes	142	85.5	100	62.1	40	71.4
Vote during the elections from that country	89	53.6	6	3.8	15	26.8
Buy groceries	83	50.3	32	20.1	7	12.5
Go on a road trip	47	28.5	33	20.6	8	14.3
Go to the doctor	45	27.1	3	1.9	9	16.1
Go to the movie theater	45	27.1	2	1.3	0	0.0
Go to church	45	27.1	2	1.3	0	0.0
Get a haircut	36	22.6	1	0.6	2	3.7
Visit friends	36	22.1	11	7.0	4	7.4
Follow the news from that country	33	19.9	7	4.4	4	7.1
Go to a concert	29	18.2	25	15.9	8	14.8
Go out to eat at a "fancy" restaurant	23	13.9	24	15.0	5	8.9
Do something fun	21	12.7	19	11.9	2	3.6
Go out to eat "on the street" or fast food	20	12.1	20	12.6	2	3.6
Go on vacations	19	11.4	28	17.5	3	5.4
Go to a gathering or a party	19	11.4	7	4.4	0	0.0
Visit family members	11	6.7	16	10.1	5	9.3

Table 9

Cross table. "In which country or countries do you typically do the following activities?

Results by Transborder Group (2/4)

Activity	American Consumers		Mexican Consumers		Work Commuters	
	n	%	n	%	n	%
Mexico						
Get a haircut	85	53.5	147	93.6	48	88.9
Go to the doctor	56	33.7	144	90.0	32	57.1
Go out to eat at a "fancy" restaurant	36	27.7	45	28.1	24	42.9
Go on vacations	36	27.7	23	14.4	20	35.7
Go to the movie theater	44	26.5	119	75.3	43	76.8
Visit family members	36	22.1	37	23.4	13	24.1
Go to church	35	21.1	100	62.9	24	42.9
Go to a gathering or a party	30	18.1	53	33.1	30	53.6
Go out to eat "on the street" or fast food	25	15.2	36	22.6	17	30.4
Vote during the elections from that country	24	14.5	143	89.4	16	28.6
Do something fun	22	13.3	25	15.6	19	33.9
Visit friends	20	12.3	40	25.3	15	27.8
Go to a concert	19	11.9	35	22.3	13	24.1
Go on a road trip	19	11.5	19	11.9	14	25.0
Follow the news from that country	15	9.0	56	35.0	5	8.9
Buy groceries	1	.6	29	18.2	12	21.4
Buy clothes	0	0.0	1	0.6	2	3.6

Table 9

Cross table. "In which country or countries do you typically do the following activities?

Results by Transborder Group (3/4)

Activity	American Consumers		Mexican Consumers		Work Commuters	
	n	%	n	%	n	%
In both countries						
Go out to eat "on the street" or fast food	119	72.1	99	62.3	36	64.3
Do something fun	119	71.7	111	69.4	35	62.5
Visit family members	116	71.2	104	65.8	36	66.7
Follow the news from that country	113	68.1	96	60.0	42	75.0
Go to a gathering or a party	112	67.5	95	59.4	26	46.4
Go on vacations	106	63.9	103	64.4	32	57.1
Visit friends	104	63.8	106	67.1	35	64.8
Go out to eat at a "fancy" restaurant	99	59.6	81	50.6	25	44.6
Go on a road trip	94	57.0	102	63.7	34	60.7
Buy groceries	81	49.1	97	61.0	37	66.1
Go to a concert	77	48.4	64	40.8	26	48.1
Go to the doctor	65	39.2	9	5.6	15	26.8
Go to the movie theater	57	34.3	23	14.6	11	19.6
Get a haircut	32	20.1	5	3.2	2	3.7
Go to church	33	19.9	18	11.3	3	5.4
Vote during the elections in that country	31	18.7	3	1.9	17	30.4
Buy clothes	23	13.9	58	36.0	14	25.0

Table 9

Cross table. "In which country or countries do you typically do the following activities?

Results by Transborder Group (4/4)

Activity	American Consumers		Mexican Consumers		Work Commuters	
	n	%	n	%	n	%
I do not do that activity						
Go to church	53	31.9	39	24.5	29	51.8
Go to a concert	34	21.4	33	21.0	7	13.0
Vote during the elections from that country	22	13.3	8	5.0	8	14.3
Go to the movie theater	20	12.0	14	8.9	2	3.6
Go to out at a "fancy restaurant"	8	4.8	10	6.3	2	3.6
Get a haircut	6	3.8	4	2.5	2	3.7
Go on vacations	5	3.0	6	3.8	1	1.8
Go on a road trip	5	3.0	6	3.8	0	0.0
Go to a gathering or a party	5	3.0	5	3.1	0	0.0
Follow the news from that country	5	3.0	1	0.6	5	8.9
Do something fun	4	2.4	5	3.1	0	0.0
Visit friends	3	1.8	1	0.6	0	0.0
Go out to eat "on the street" or fast food	1	0.6	4	2.5	1	1.8
Buy clothes	1	0.6	2	1.2	0	0.0
Go to the doctor	0	0.0	4	2.5	0	0.0
Visit family members	0	0.0	1	0.6	0	0.0
Buy groceries	0	0.0	0	0.0	1	0.6

Note. Missing values are excluded (402 total respondents). Respondents classified as "School Commuters" and unclassified are excluded due to small sample size (less than 30). Top three items per activity for each group are bold.

4.4. Building Scales

The results from the scale show a distribution that concentrates between 40 and 70 percent concentrating more than half of the cases there. In addition to this, results shows that survey respondents averaged a .49 transfronteridad level in a scale of 0 to 1. In terms of multiculturality, respondents concentrated their results between 20 and 70 percent with an average of .44 in a 0 – 1 scale. The results from the transfronteridad were broken by different control groups to take a first stab at identifying important interactions. Most of the control items show interesting interactions that will be assessed for statistical significance through a regression analysis (see table figure 23).

In addition to presenting general results for the scales of transfronteridad and multiculturality, the author grouped the data by different categories available in the survey data. According to table 10, the average level of transfronteridad ranged between 40 and 60 points for most of all groups with a few exceptions such as English-only speakers who showed reduced levels of transfronteridad (lower than 20 points), individuals with low educational attainment who showed increased levels of transfronteridad (higher than 60 points), and individuals with high levels of multiculturality who showed similar transfronteridad levels than the prior group.

Figure 23

Scales for Transfronteridad and Multiculturality (Indexed to 100)
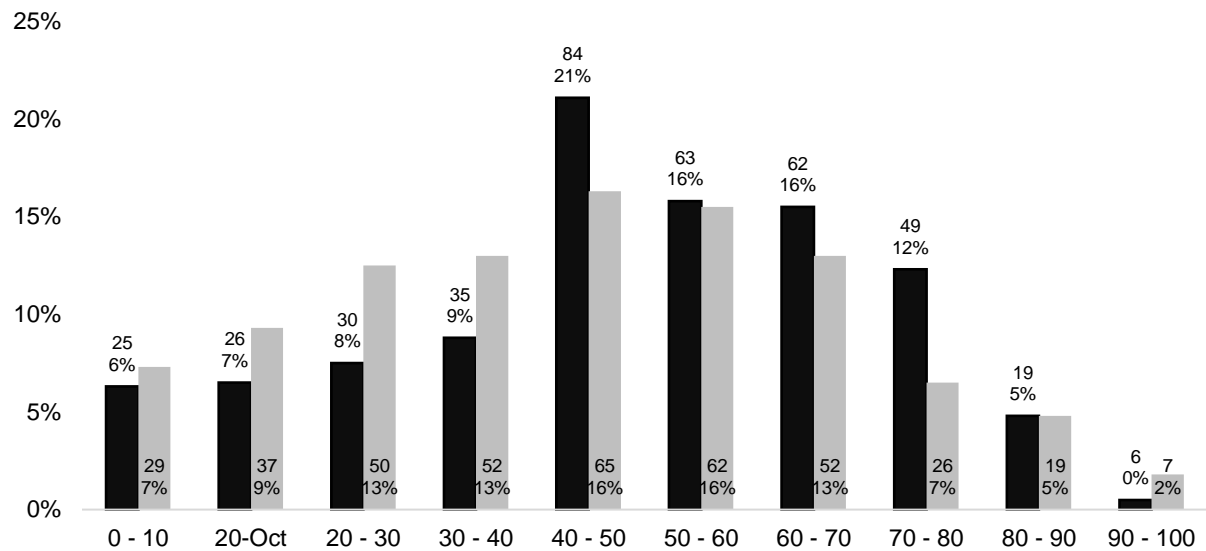
Note. Missing values are excluded (402 total respondents).

Source. Author's elaboration.

Table 10

Sociodemographic and Other Characteristics of Survey Respondents (1/4)

Characteristic	n	Average Transfronteridad
Country of Residency		
United States	166	53
Mexico	236	47
Immigration Status in the U.S.		
No documents allowing legal entry to the U.S.[a]	3	NA
Tourist-visa-holder	166	46
Work-visa-holder	5	47
United States Resident	43	52
United States Citizen	185	52

Table 10

Sociodemographic and Other Characteristics of Survey Respondents (2/4)

Characteristic	n	Average Transfronteridad
Transborder Category[b]		
American Consumer	166	53
Mexican Consumer	164	46
Work Commuter	56	48
School Commuter	16	58
Language Proficiency		
Spanish-only Speaker	162	45
English-only Speaker	4	21
Bilingual	233	53
Mexican Ascendancy Status		
Yes	350	50
No	37	42
Gender		
Female	260	50
Male	137	48
Prefer not to answer	2	48
Age Group		
Less than 18 years	1	47
18 – 24	33	48
25 – 34	90	49
35 – 44	94	49
45 – 49	51	52
50 – 54	35	48
55 – 64	64	48
65 or more	31	54
Education Level		
Elementary of less	3	67
Middle School	13	54
High School	56	40
Less than two years in college	60	52
Two or more years in college	267	51

Table 10

Sociodemographic and Other Characteristics of Survey Respondents (3/4)

Characteristic	n	Average Transfronteridad
Countries where respondent has lived		
Only in the United States	37	43
Only in Mexico	159	46
Both countries	201	54
Countries where respondent has worked		
Only in the United States	119	48
Only in Mexico	162	47
Both countries	115	55
None	2	44
Countries where respondent has studied		
Only in the United States	67	46
Only in Mexico	196	47
Both countries	135	55
Border-crossing Frequency		
More than three times per week	79	52
2 – 3 times per week	81	49
1 time per week	90	53
Every 10 – 15 days	57	52
Once per month	50	50
Less than once per month	42	34
Number of years crossing the border		
Less than a year	46	46
From one to three years	62	47
From three to five years	41	49
More than five years	248	51

Table 10

Sociodemographic and Other Characteristics of Survey Respondents (4/4)

Characteristic	n	Average Transfronteridad
Residency in border city		
Yes	73	47
No	326	50
Multiculturality Scale		
0 – 20	66	32
20 – 40	102	49
40 – 60	127	52
60 – 80	78	56
80 - 100	26	66

Note. Missing values are excluded (402 total respondents).

[a] Respondents who answered this option were terminated in the survey.

[b] Respondents residing in Mexico who fell in more than one category were reclassified into one using the following hierarchy: 1) Work Commuter, 2) School Commuter, and 3) Mexican Consumer.

Source. Author's elaboration.

4.5. Regression

The results from the regression in Table 11 show that we were able to reject the null hypothesis since the F score obtained was statistically significant at <1 percent. Specifically, the variables multiculturality and language show statistically significant coefficients but not education. According to the first row, both multiculturality and language have positive coefficients that prove that they are positively impacting the variable transfronteridad; this is, more multiculturality and language proficiency levels can be associated with higher levels of transfronteridad. Conversely, the education variable did not show the same behavior. Results from the model show that these three variables

had a 17% prediction power and were significant beyond the 1 percent level which is very positive.

In addition to a general regression analysis, the author ran a series of additional regression models to control for different variables that literature predicted to be correlated with levels of transfronteridad as well. According to the results from these models, it is possible to stablish that multiculturality affected transborders when controlling for all variables except for those were there was not enough evidence to run tests[4] and for age, education, border crossing frequency, experience crossing the border. For language, the results were less conclusive considering that the correlation was not significant for all subgroups when controlling for the previously stablished sociodemographic variables.

---

[4] Such as undocumented respondents, work-visa holders, school commuters, English-only speakers, individuals who prefer not to answer their gender, minors, respondents with education as the middle school level or below, and respondents who had not worked in any country.

Table 11

Regression Analysis Output (1/4)

Characteristic	n	R	R²	F	Sig.	Multiculturality			Language			Education		
						Beta	t	Sig.	Beta	t	Sig.	Beta	t	Sig.
Total	399	.413	.171	27.128	**.000**	.367	7.917	**.000**	.132	2.789	**.006**	.025	.531	.595
Country of Residency														
United States	166	.417	.174	11.361	**.000**	.70	5.176	**.000**	.182	2.405	**.017**	.004	.056	.955
Mexico	233	.394	.155	14.030	**.000**	.361	5.831	**.000**	.095	1.528	.128	.032	.525	.600
Immigration Status in the U.S.														
No documents allowing legal entry to the U.S [a]	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Tourist-visa-holder	166	165	.409	.167	**.000**	.372	5.073	**.000**	.104	1.399	.164	.021	.290	.772
Work-visa-holder	5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
United States Resident	43	.564	.318	6.052	**.002**	.486	3.520	**.001**	.153	1.030	.309	.096	.669	.507
United States Citizen	185	.360	.130	8.990	**.000**	.319	4.596	**.000**	.164	2.347	**.020**	.026	.365	.716
Mexican Ascendancy Status														
Yes	350	.423	.179	25.090	**.000**	.382	7.709	**.000**	.114	2.248	**.025**	.032	.646	.519
No	37	.406	.165	2.170	.110	.285	1.779	.085	.225	1.403	.170	-.107	-.667	.509
Transborder Category [b]														
American Consumer	166	.417	.174	11.361	**.000**	.370	5.176	**.000**	.182	2.405	**.017**	.004	.056	.955
Mexican Consumer	164	161	.395	.156	**.000**	.381	5.105	**.000**	.065	.857	.393	-.21	-.275	.784
Work Commuter	56	.401	.161	3.314	**.027**	.289	2.207	**.032**	.201	1.532	.132	.121	.948	.347
School Commuter	16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 11

Regression Analysis Output (2/4)

Characteristic	n	R	R²	F	Sig.	Multiculturality			Language			Education		
						Beta	t	Sig.	Beta	t	Sig.	Beta	t	Sig.
Language Proficiency														
Spanish-only Speaker	162	.393	.154	14.489	**.000**	.393	5.382	**.000**	NA	NA	NA	-.003	-.043	.966
English-only Speaker	4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Bilingual	233	.342	.117	15.232	**.000**	.333	5.363	**.000**	NA	NA	NA	.070	1.122	.263
Gender														
Female	260	.409	.167	17.128	**.000**	.362	6.267	**.000**	.115	1.950	.052	.069	1.182	.238
Male	137	.448	.201	11.136	**.000**	.395	5.006	**.000**	.171	2.138	**.034**	-.056	-.702	.484
Prefer not to answer	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Age Group														
Less than 18 years	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
18 – 24	33	.502	.252	3.252	**.036**	.133	.818	.420	.151	.851	.402	.392	2.202	**.036**
25 – 34	90	.400	.160	5.445	**.002**	.282	2.854	**.005**	.286	2.870	**.005**	-.052	-.522	.603
35 – 44	94	.612	.375	18.004	**.000**	.455	5.266	**.000**	.153	1.773	.080	.245	2.868	**.005**
45 – 49	51	.473	.223	4.506	**.007**	.403	3.125	**.003**	.204	1.475	.147	.043	.311	.757
50 – 54	35	.396	.157	1.926	.146	.395	2.366	**.024**	.099	.551	.586	-.018	-.098	.923
55 – 64	64	.477	.228	5.894	**.001**	.396	3.268	**.002**	.089	.736	.465	-.186	-1.628	.109
Countries where respondent has lived														
Only in the United States	37	.715	.511	11.479	**.000**	.407	2.945	**.006**	.306	2.215	**.034**	-.367	-3.002	**.005**
Only in Mexico	159	.427	.182	11.523	**.000**	.366	4.895	**.000**	.161	2.126	**.035**	-.001	-.009	.993
Both countries	201	.332	.110	8.112	**.000**	.284	4.230	**.000**	.069	.990	.324	.149	2.145	**.033**

Table 11

Regression Analysis Output (3/4)

Characteristic	n	R	R²	F	Sig.	Multiculturality			Language			Education		
						Beta	t	Sig.	Beta	t	Sig.	Beta	t	Sig.
Education Level														
Elementary of less	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Middle School	13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
High School	56	.532	.283	10.467	**.000**	.518	4.292	**.000**	.044	.367	.715	NA	NA	NA
Less than two years in college	60	.350	.123	3.991	**.024**	.214	1.727	.090	.277	2.232	**.030**	NA	NA	NA
Two or more years in college	267	.428	.183	29.641	**.000**	.395	6.999	**.000**	.114	2.018	**.045**	NA	NA	NA
Countries where respondent has worked														
Only in the United States	119	.417	.174	8.079	**.000**	.349	4.094	**.000**	.204	2.285	**.024**	-.054	-.606	.546
Only in Mexico	162	.398	.159	9.940	**.000**	.350	4.732	**.000**	.128	1.693	.092	.031	.415	.679
Both countries	115	.415	.172	7.707	**.000**	.378	4.372	**.000**	.180	2.012	**.047**	.015	.173	.863
None	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Countries where respondent has studied														
Only in the United States	67	.427	.182	4.684	**.005**	.309	2.548	**.013**	.135	1.111	.271	-.155	-1.345	.183
Only in Mexico	196	.400	.160	12.159	**.000**	.377	5.617	**.000**	.079	1.155	.250	.017	.252	.802
Both countries	135	.451	.204	11.177	**.000**	.340	4.307	**.000**	.208	2.590	**.011**	.140	1.722	.088
Border-crossing Frequency														
More than three times per week	79	.342	.117	3.306	**.025**	.230	2.119	**.037**	.199	1.827	.072	.157	1.439	.154
2 – 3 times per week	81	.478	.228	7.597	**.000**	.465	4.561	**.000**	.089	.882	.381	-.010	-.098	.922
1 time per week	90	.387	.150	5.049	**.003**	.152	1.474	.144	.336	3.071	**.003**	-.106	-.996	.322
Every 10 – 15 days	57	.518	.269	6.486	**.001**	.492	4.158	**.000**	.115	.963	.340	.023	.192	.848
Once per month	50	.298	.089	1.489	.230	.266	1.869	.068	.102	.680	.500	-.013	-.090	.929
Less than once per month	42	.528	.278	4.884	**.006**	.557	3.731	**.001**	-.107	-.693	.492	.059	.409	.685

Table 11

Regression Analysis Output (4/4)

Characteristic	n	R	R²	F	Sig.	Multiculturality			Language			Education		
						Beta	t	Sig.	Beta	t	Sig.	Beta	t	Sig.
Number of years crossing the border														
Less than a year	46	.460	.212	3.760	.018	.418	2.965	**.005**	.152	1.087	.283	-.021	-.146	.884
From one to three years	62	.500	.250	6.458	**.001**	.473	4.141	**.000**	.085	.709	.481	.065	.551	.584
From three to five years	41	.332	.110	1.529	.223	.065	.417	.679	.215	1.377	.177	.219	1.407	.168
More than five years	248	.433	.188	18.781	**.000**	.387	6.595	**.000**	.139	2.307	**.022**	-.041	-.699	.485
Residency in border city														
Yes	73	.382	.146	3.938	.012	.332	2.940	**.004**	.183	1.648	.104	.117	1.038	.303
No	326	.428	.183	24.075	**.000**	.386	7.514	**.000**	.123	2.333	**.020**	.002	.036	.971
Border region														
California – Baja California	213	.448	.201	17.540	**.000**	.369	5.833	**.000**	.195	3.040	**.003**	-.010	-.160	.873
Arizona – Sonora	55	.463	.214	4.627	**.006**	.448	3.591	**.001**	.123	.982	.331	.027	.216	.830
New Mexico and Texas – Chihuahua, Coahuila, Nuevo Leon, and Tamaulipas	131	.358	.128	6.240	**.001**	.322	3.756	**.000**	.066	.762	.447	.055	.648	.522

Note: Missing values are excluded (402 total respondents).
[a] Respondents who answered this option were terminated in the survey.
[b] Respondents residing in Mexico who fell in more than one category were reclassified into one using the following hierarchy:
   1) Work Commuter, 2) School Commuter, and 3) Mexican Consumer.

4.6. Machine Learning

Even though several regression models were run using the three main variables analyzed from this study, the author had interest to see if a higher prediction power could be achieved by using Machine Learning techniques, specifically, supervised Machine Learning models to create one that could incorporate more variables into the mix. Due to this, the author built four Supervised Machine Learning models based in regression techniques and 16 variables as opposed to the original three (see Figure 24). Results however, should not be considered definitive due to the exploratory character of this research. In addition to this, it is important to highlight that social science models are focused more towards identifying significant variables to a phenomenon more than achieving a prediction power close to 100 percent. In fact, social research and prediction rules of thumb establish that when doing regressions, prediction powers close to 100% could be a sign of auto collinearity or multicollinearity, which implies redundancy in the model (Gujarati & Porter, 2010). Figure 9 illustrates the variables included in this model. This selection was made based on the literature review and potential variables impacting the levels of transfronteridad per different authors.

In order to build this project's machine learning models, four different approaches will be used. Instructions to work on these models were provided by (Singh, 2019). In general terms the author states that machine learning is used in different way to solve business problems. Machine learning models can be organized into two big groups classification and regression.

Figure 24

Variables used to develop a Supervised Machine Learning Model to Predict Transfronteridad
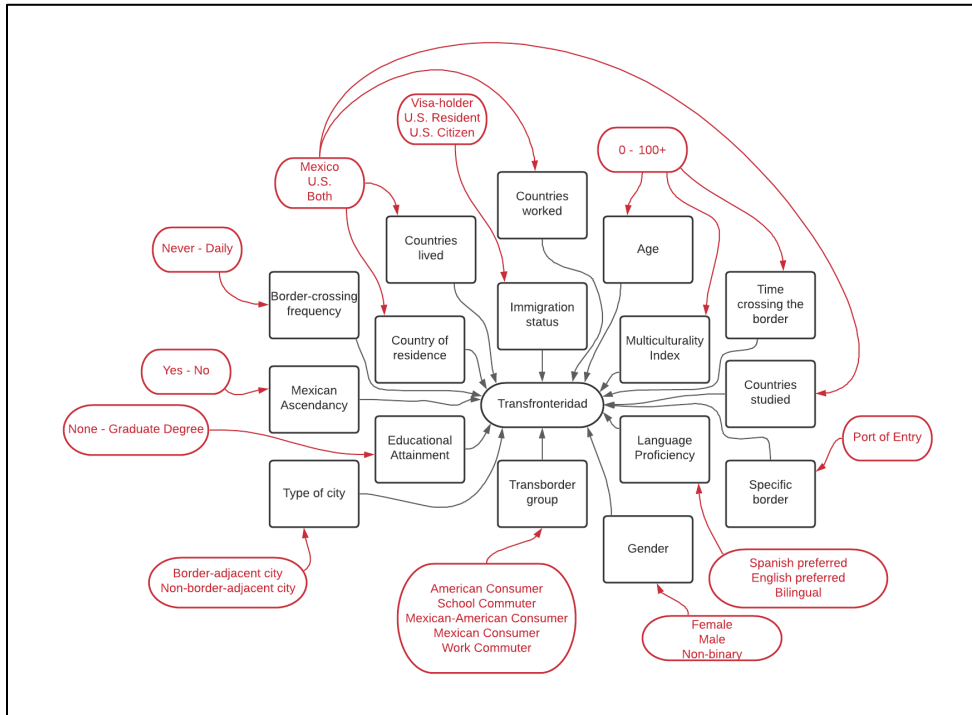
Source. Author's elaboration.

Within the regression model available, the author highlights four types: a) Linear Regression, b) Ridge Regression, c) LASSO Regression, and d) Elastic Net Regression. According to the author, the Linear Regression is the simplest regression model available. In order to run this model, it is assumed that the variables being used show a normal distribution and that are also independent from each other. The model assumes that all variables are linearly correlated to the output variable. The Ridge Regression is a variant from the linear model in which the loss function is modified to make a less complex model that the linear model. This may translate into better results, but not necessarily all the time. The LASSO Regression, with LASSO standing out for Least Absolute Shrinkage

and Selection Operator, is a variant from the linear regression model in which the sum of the coefficients is limited to improve results. Finally, the Elastic Net Regression is a combination of the ridge and LASSO models due to combining both model principals to improve results. The author specified the process that needs to be followed to build any of these models (see figure 25).

Figure 25

Steps to build regression machine learning supervised models
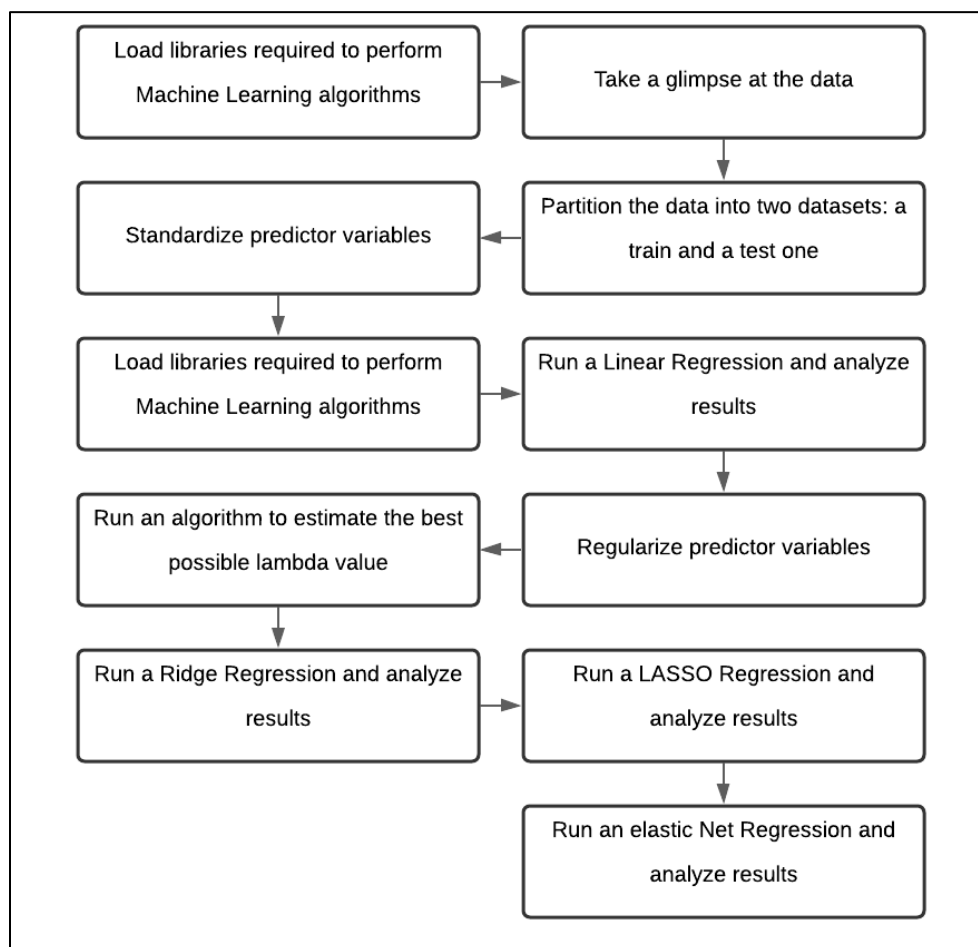
Source. Author's elaboration based on (Singh, 2019).

Figure 26

Screenshots from R Studio showing Development of Supervised Models

Source. Author's elaboration.

Results obtained from these four modules can be seen in Table 18; no model passed from a 30% prediction power which under these circumstances -considering that this is a social model and an exploratory analysis, is positive. On this note, it is important to highlight that in social sciences the Pareto Principle is often very used to talk about 80/20 situations in which organizations may need to prioritize their efforts to increase efficiency. The translation from these principles is that more likely than not the remaining 70% of explanation will reside in several numbers of variables that cannot be controlled by individuals or organizations. This means, there are several aspects from transfronteridad that will go outside of variables that government could control via public policies or private organization through marketing campaigns and others. When looking

to maximize levels of transfronteridad, therefore, there would not be much value in aiming to create a model with hundreds or thousands of variables to increase prediction levels if the variables used are outside of the developers control. Additionally, high R-squared results can indicate problems in the research design, sampling or data such as biased estimates, random correlations, or error patterns description instead of relationships between variables (Frost, n.d.).

Figure 27

Results from Supervised Machine Learning Regression Models
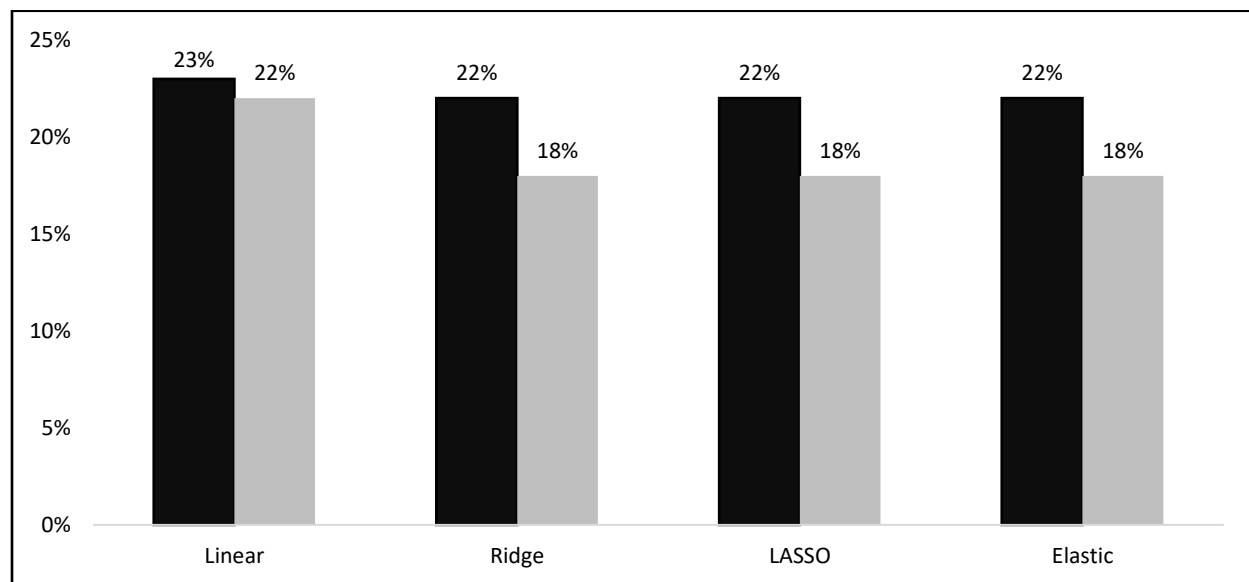
Source. Author's elaboration.

4.7. Chapter four closing comments

In this chapter the author confirmed the existence of a statistically significant correlation between transfronteridad and both language proficiency and multiculturality, achieving the main objective from this research. Additionally, the author learned from the results of the machine learning model that it is not possible to predict accurately the levels of transfronteridad using the selection of variables identified.

Chapter 5. Discussion and Conclusions

5.1. Chapter five introductory comments

Chapter five holds the conclusions from this study. Results in chapter four showed that while there is a statistically significant correlation between transfronteridad and both language proficiency and multiculturality, the relationship is not strong enough to enable predictions, not even after including additional variables to the model. In this chapter the author discusses these results and includes a series of reflections regarding implications from these results, future areas of research and benefits that the study will bring to society in general.

5.1. Discussion

The results showed that transborders tend to opt for the U.S. as their preferred destination to buy clothes which is related to the lower prices and higher quality perceived by consumers at the border. The implications from this are that an important proportion of the transborder movement from Mexico to the U.S. by Mexican consumers will include buying clothes as the purpose of the visit which resonates with the findings from (Mungaray-Moctezuma & Calderón, 2015).

In terms of activities being done in a certain country, Mexico showed a high percentage for haircuts, which has to do with the lower prices and higher quality that businesses across the border can offer. It is very often than Mexican-American individuals cross the border to receive this service in one of the many businesses offering it.

Social activities, according to the results, tend to not be concentrated in either side of the border; there is a history of family and social relationships that go beyond borders as Escamilla (2019) states. Because of this, it is no surprise that visiting family, friends

and doing something fun at the border went to the top of the activities done in both sides of the border.

According to the results from the survey, there are some activities that tend to only happen on one side of the border. Results show that voting during elections falls in this category. Based on the literature available, it is possible to guess that the immigration status prevent individual to vote in the country where they do not live. This has important implications considering that while the immigration status may prevent from people to both in either country which is interesting considering that while they may not have the immigration status needed to vote, their daily activities in other countries make them live in two countries but not being able to participate in the political dynamics of one of those places.

In terms of cultural statement, results show that while most of them showed distributed responses across all the options, a transborders being interested in having their children learn both Spanish and English went at the top of the list which may be caused due to the importance these individuals have seen speaking two languages represent for them in terms of cultural connections and accessing better opportunities. This information is important for public policy development as it allows to understand that for both sides of the border government would do a god job in investing in training people to learn two languages.

As part of analysis and processing of the results, three groups were created, and information was grouped by that. Results show that all three groups, American Consumers, Mexican Consumers and Commuters equally prefer buying their clothes in the U.S., which is due to the increase quality with lower prices in that side of the border -

a concept that resonates with (Mungaray-Moctezuma & Calderón, 2015). Buying groceries was an activity that while American Consumers are mostly doing in either only the U.S. or in both countries, for Mexican consumers and Commuters, it happens in both countries most of the time. This could be caused due to the fact that most of the interactions done by American consumers in Mexico relate to personal connections and entertainment, perhaps leaving too little time for activities such as buying groceries. Additionally, the marketing behind various U.S. products has been so effective that Mexican consumers and commuters may have the impression that there are better products in the U.S. side.

One interesting thing that was found is that commuters tend to vote considerably less than the other two groups. This is an important phenomenon since it implies that they will not be as politically engaged as other groups, which may have implications considering that reduced political participation reduces the lookout that groups do on their politicians which can ultimately lead to worst government outcomes. This could be causes due to the fact that commuters are the ones spending most of their times split in two countries, they are genuinely living in two places and may not feel that they lived in a certain place. However, just like commuters, American consumers also shows a similar level of lack of participation, at least higher than the Mexican consumers, which can be due to the fact that they are not from here nor there.

Results show that entertainment activities are mostly realized in both countries, which could relate to the fact that both sides of the border have much to offer in these areas; while a transborder can go and enjoy a night of party in one of the Mexican cities in the border, they can also go and enjoy a concert in the U.S. side. Results show that

going on a road trip, go to a concert, going to eat on restaurants or on the street, going on vacations, going with friends, and doing fun things in general are activities happening in both countries in most cases. The results showed that transborders are staying on top of the news from both countries in most cases which is not a complex task to accomplish considering that in the border there are many communication channels, programs and more focused in providing news happening in both sides of the border. It was interesting to see how commuters tend to stay on top of the news from both countries more often than the other two groups which may have to do with the fact that while the are crossing the border one of the recurrent activities they do is to listen the radio which may include news. Results also show that commuters do not use to go to the church which is different than Mexican consumers, we can again assign this to the fact that commuters tend to have limited time to participate in other activities due to crossing times.

Results from the scale from transfronteridad ranged from 0 to 100 showing that not all transborder individuals show the same level of interaction with either or both sides of the border. This aligns with this project's hypotheses that stablishes that transfronteridad is conditioned by a series of factor surrounding the transborder life. A second scale, for multiculturality also shows variance which is useful to test the hypotheses considering that in order to identify correlation with variables in linear models it is important to have variation in the sample.

Results from the regression show that the null hypotheses were rejected for Multiculturality and Language. This means that there is a significant correlation with the levels of transfronteridad shown by transborders and their multicultural skills and language proficiency. As authors explain, the ability to work with different cultures make

transborders to transition easily between culture, access services and products from different countries without struggling. On the other side, language proficiency allows individuals to go back and forth between two worlds without struggling to make this transition. On the other side, results were not significant for level of education which may be caused to the fact that while les educated transborders may be more likely to be part of lower socioeconomic groups, that does not prevent them from adjusting their consumer decisions to get the most from the border in their unique way.

The additional regressions that were run show that in general education will never be correlated with levels of transfronteridad. Multicultural skills, differently, they were consistently correlated with transfronteridad which indicates that there are no interactions to consider in this relationship based in demographic information. Finally, language proficiency, was not significant for individuals living in Mexico, tourists, U.S. Residents, Mexican consumers, Work Commuters, and non-Mexican individuals. This could either be caused for the small sample size that some of these groups have or the fact that for some of this group, the existence of Spanish offerings in the U.S. due to the increased Hispanic population in the U.S. border has created enough opportunities for non-English speaking individual to thrive in both sides of the border.

Finally, the results from the research, while enough to reject the null hypothesis for two variables, are not conclusive enough to stablish causality, which could be seen in the prediction power achieved after running several different machine learning models that were not able to predict more than 25% of the results.

5.2 Answering Research Questions

This research project was developed to answer a main question related to the correlation between multiculturality and transfronteridad. Specifically, the question to be answered was *What is the relationship between multicultural competency and levels of transfronteridad presented by transborders?* Based on the results obtained, the author can say that multiculturality and transfronteridad and correlated significantly as it was proven in the regressions performed in this research projects. Results, however, show that within the variables used to measure the construct of multiculturality, only language and multicultural skills were proven significantly correlated to transfronteridad. This means that changes observed in educational attainment among transborders were not accompanied by changes in the levels of transfronteridad. Moreover, while language and multiculturality skills showed statistically significant correlations, the results from the machine learning models show that level of transfronteridad cannot be predicted using these variables solely. Still, results continue to be important to increase the body of knowledge around transborder populations and dynamics, and as it was mentioned before, when it comes to social dynamic modeling, the current literature show that prediction powers close to 100% are not recurrent and in some times may be reason to doubt the validity of research designs or project results showing that behavior.

Additionally, this research also aimed to answer two secondary questions. Specifically, a) *how can we measure transfronteridad among transborders?* And b) *is it possible to build a supervised learning method based on regression to predict levels of transfronteridad using information related to transborder dynamics?* The results obtained showed that future research projects may measure transfronteridad among individuals

using a selection of recurrent day-to-day activities (social and personal) that can be fit into survey items able to identify in which country or countries transborder individuals are performing these activities. After collecting this information, researchers will need to do process the data like it was done in this project to create a transfronteridad index. As for the second questions, based on the results found, all four models presented similar results, but the linear regression showed slightly better results in terms of R-square (22% while all the other models showed 18%). Still, additional work will need to be done to continue increasing the prediction power of this type of models by adding other variables that may be correlated with transfronteridad.

5.2. Conclusions

1. Multiculturality is significantly correlated with the levels of transfronteridad, this means that more multicultural skills will allow individuals to participate in more activities in both sides of the border, which translates into increased well-being.

2. Language proficiency is significantly correlated with the levels of transfronteridad, this is, the condition of being bilingual, will allow individuals to participate in more activities in both sides of the border, which translates into increased well-being. Still, this variable is not correlated when controlled for individuals living in the Mexican side which could be due to the fact that while they cannot do all activities that a bilingual individual could, there is enough supply of activities in Spanish to allow them to do.

3. Educational attainment is not significantly correlated with the levels of transfronteridad. This could be due to the fact that individuals will make unique

choices to maximize how much they can't take from the border regardless their educational level; reduced education level will not condition which activities an individual will do in either side of the border.

5.3. Recommendations

1. Continue exploring the transborder population considering their size and the dynamics happening at the border. Avoiding this activity will not help the border region to continue progressing or increase transfronteridad at the macro level.

2. Continue working towards creating a formal sample framework to generate statistically representative sample projects that will lead to better results further than only exploratory models.

3. Ensure local border governments are aware of the transborder dynamics happening in their regions so that they can maximize benefits for the population but also prevent problematics associated with these dynamics.

4. Foster an environment of multiculturality in border regions to ensure transfronteridad is maximized to increase the overall well-being in these regions.

5. Continue building language proficiency in both sides of the border via bilingual educational programs to ensure that is not a barrier in building stronger transborder regions.

6. Invest in researching more this population to understand better how to develop public policy that tackles and prevent problems associate with this group. Additionally, this will increase the effectiveness of public policy. In

addition to this, the same principle must be applied to the private sector considering that a lot of opportunities are being missed by marketers due to a lack of knowledge of this population.

5.4. New Lines of Research

The work from this study has allowed the author to identify a series of potential new lines of research that may be pursued by the author and other individuals interested in border theory concepts. These potential new lines of research are a) Developing a sample framework to enable future probabilistic-sampling-based research projects, b) Strengthen the concept of Transfronteridad, and c) Develop additional iterations for the operational definitions for Transfronteridad and Multicultural Skills.

The lack of a formal sample framework to enable following probabilistic sampling may be the biggest opportunity for future research considering how little this topic has been researched. In order to develop this sample framework, additional research would need to be done to identify the proportion of borderlanders residing in Mexico that are able to cross the border. Additionally, information regarding their frequency of crossing would need to be collected. Similarly, information from transborders residing in the U.S. side would need to be collected, specially, for Anglo individuals crossing it, considering that no source of information for this phenomenon was identified. While the concept of Transfronteridad was key for the development of this study, the author identified gaps in the form of limited literature speaking to this concept; not more than 3 authors speaking to this concept were identified during the literature review. Additionally, it is important to continue exploring the redundancy between transmigration and transfronteridad, including determining if these two concepts are speaking to the same construct. Finally,

while for the purposes of this exploratory exercise it was enough to develop straightforward operational definitions for the concepts of transfronteridad and multicultural skills, more work will need to be done in the future to identify best ways to measure these constructs.

5.5. Results Applications and Contributions to Society and Field

The results from this research project will be helpful to increase the body of knowledge for border theory; specifically, the literature review collected, compiled, and presented in this paper will inform future iterations of research projects to explore concepts related to transfronteridad. Two of the major findings from this research project showed that transfronteridad can be increased among transborder populations by fostering multiculturality and bilingualism. Local government from border cities then should incorporate language proficiency and multiculturality campaigns into their public policy development so that they can allow people to get the benefit from being individuals with higher levels of transfronteridad. Additionally, this paper, provide an operational definition to count the number of transborder population in the Mexico – U.S. border. While the author expressed that this definition must be improved to take into consideration individuals that may be using Visas to cross from Mexico to the U.S., it is still a good mechanism to get a ballpark as to the number of individuals living under this lifestyle. This particular number (above 4 million people) can be used by marketers to calculate purchasing power among this population and use that information as a base for future market research projects and also determine optimal levels of marketing investments in the region.

5.6. Thesis Analysis (SWOT structure) and chapter five closing comments

Some of the strengths identified in this study are the comprehensive literature review conducted to incorporate relevant concepts regarding transborder dynamics. This was very useful to connect the findings from the statistical and machine learning modeling processing to the current literature. Conversely, one of the major weaknesses from this particular project is the lack of a sample frame that could allow for probabilistic sample selection, reducing the external validity of this project's results. Additionally, the reduced prediction power from the machine learning supervised regression model limits the applications from the model itself though it is important to consider that statistical and machine learning models analyzing social phenomena -like this- are rarely able to accomplish this task.

In terms of opportunities, both governments and organizations at border regions have a great one to foster transfronteridad at the micro and macro levels by investing in initiatives to increase bilingualism in both sides of the border and also multiculturality. By doing this, they are enabling their citizens to access a wider array of goods and services in both sides of the border which will increase their well-being. Finally, it is important to consider different threats that may impact the development of transfronteridad in the border regions such as challenges associated with the international political environment which could go back to tense conversations depending on the results of the 2022 and 2024 U.S. elections and the continuous impact of the COVID-19 pandemic that may causes additional border closures.

References

Alegría, T. (1990). Ciudad y transmigración en la frontera de México con Estados Unidos. *Frontera Norte*, 7-38.

Alegría, T. (2000). Juntos pero no revueltos: ciudades en la frontera México-Estados Unidos. *Revista Mexicana de Sociología*, 89-107.

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: an overview. *Journal of physics*, n.d.

Anderson, D., Sweeney, D., & Williams, T. (2013). *Estadística para Negocios y Economía.* Ciudad de México: Cengage Learning.

Bureau of Transportation Statistics [BTS]. (2021). *Monthly Border Crossing Statistics [Dataset].* Bureau of Transportation Statistics.

Calzada, E., Covas, M., Ramirez, D., Miller, L., & Huan, K.-Y. (2016). A Longitudinal Study of Cultural Adaptation among Mexican and Dominican Immigrant Women. *Journal of International Migration and Integration*, 1049–1063.

Campos, D. &. (2012). Crossing the border: Mobility as a resource in the Tijuana/San Diego and Tecún Umán/Tapachula regions. *Estudios fronterizos*, 9-32.

Coursera. (s.f.). *Aprendizaje Automático*. Obtenido de Coursera: https://www.coursera.org/learn/machine-learning

de la Garza, J., Morales, B., & González, B. (2013). *Análisis Estadístico Multivariante.* Ciudad de México: Mc Graw Hill.

Duarte-Herrera, C. A. (2001). Defining the US-Mexico border as hyperreality. *Estudios fronterizos*, 139-165.

Escamilla, E. (2019). Transfronteridad and Everyday Commuter Negotiations of Tijuana's Borderscape [Unpublished master's thesis]. Basel University.

Falcón, V., & Orta, A. (2018). The Transborder Identity Formation Process: An Exploratory Grounded Theory Study of Transfronterizo College Students from the San Diego-Tijuana Border Region. *Journal of Transborders Studies*, 1-26.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, 293 - 314.

Fernández, R. (1987). Fronteras paradójicas: contrastes de desarrollo en la frontera México-EUA. . *Estudios Fronterizos*, 113-122.

Fimbres, D. (2000). Emigración, inmigración y retorno: el ciclo de los inmigrantes mexicanos en Caléxico, California, EUA. *Estudios fronterizos*, 93-120.

Fimbres, D., & Ortega, V. (2001). Inmigración y movilidad laboral de trabajadores mexicanos residentes legales en una ciudad fronteriza de Estados Unidos: Caléxico, California. *Estudios fronterizos,*, 9 - 40.

Fimbres, D., & Ortega, V. (2001). Inmigración y movilidad laboral de trabajadores mexicanos residentes legales en una ciudad fronteriza de Estados Unidos: Caléxico, California. *Estudios fronterizos*, 9-40.

Frost, J. (s.f.). *Five Reasons Why Your R-squeared Can Be Too High*. Obtenido de Statistics By Jim.

Goldberg. (2001). Transnationalism and Borderlands: Concepts of Space on the US-Mexico Border and Beyond. . *Estudios Fronterizos*, 49-91.

González, B. A. (1988). Transmigración legal en la frontera México-Estados Unidos. *Revista Mexicana de Sociología*, 277-322.

Gujarati, D., & Porter, D. (2010). *Econometría.* Ciudad de México: Mc Graw Hill.

Hernández, H. A. (2020). La frontera México-Estados Unidos: asimetrías y transgresiones. *Nueva sociedad*, 59-69.

Hernández, S., Fernández, C., & Baptista, L. (2014). *Metodología de la Investigación.* Ciudad de México: McGraw Hill.

Iglesias-Prieto, N. (2004). En pocas palabras: Representaciones discursivas de la frontera México-Estados Unidos. . *Aztlán: A Journal of chicano studies*, 145-154.

Iglesias-Prieto, N. (2010). Trascendiendo límites: La frontera México-Estados Unidos en el cine. *Key Tropes in Inter-American Studies.*

Kansas State University. (2001). *Multicultural Competency Development: Preparing Students to Live and Work in a Diverse World.* Obtenido de https://tilford.k-state.edu/resources/multicultural-competency-development.html

Kosinski, M., & Behrend, T. (2017). Editorial overview: Big data in the behavioral sciences. *n.d.*

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 249 - 268.

Lawrynowicz, A., & Tresp, V. (2014). Introducing Machine Learning. *Institute of Computing Science, Poznan University of Technology*, 35 - 50.

Lu, C., Lin, G., Wu, T., Hu, I., & Chang, Y. (2021). Influencing Factors of Cross-Border E-Commerce Consumer Purchase Intention Based on Wireless Network and Machine Learning. *Security and Communication Networks.*

Martínez, O. J. (1994). *Border People; Life and Society in the U.S. - Mexico Borderlands.* Arizona.

Mexican National Institute of Statistics and Geography [INEGI]. (2021). *2020 Census of Housing and Population [Raw microdata].* Mexican National Institute of Statistics and Geography [INEGI].

Mungaray-Moctezuma, A., & Calderón, R. J. (2015). Infraestructura vial y movilidad de consumo en el espacio transfronterizo de Mexicali y Valle Imperial. *Estudios fronterizos*, 195-219.

Ojeda, N. (2005). Familias transfronterizas y familias transnacionales: algunas reflexiones. . *Migraciones internacionales*, 167-174.

Orraca, P., Rocha, D., & Vargas, E. (2017). Cross-border school enrollment: Associated factors in the US–Mexico borderlands. *The Social Science Journal*, 389-402.

Orraca-Romano, P. (2015). Immigrants and Cross-Border Workers in the U.S.-Mexico Border Region . *Frontera Norte*, 5-34.

Orraca-Romano, P. (2019). Cross-Border Earnings of Mexican Workers Across the US–Mexico Border. *Journal of Borderlands Studies*, 451 - 469.

Riahi, Y., & Riahi, S. (2018). Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering*, 524 - 528.

Rocha, D., & Orraca-Romano, P. (2018). Estudiantes de educación superior transfronterizos: Residir en México y estudiar en Estados Unidos. *Frontera norte*, 103-128.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning.* New York: Cambridge University Press.

Sharma, S., Kang, D., de Oca, J., & Mudgal, A. (2021). Machine learning methods for commercial vehicle wait time prediction at a border crossing. *Research in Transportation Economics*.

Singh, D. (12 de November de 2019). *Pluralsight*. Obtenido de Linear, Lasso, and Ridge Regression with R: https://www.pluralsight.com/guides/linear-lasso-and-ridge-regression-with-r

Tapia, L. (2017). Las fronteras, la movilidad y lo transfronterizo: Reflexiones para un debate. *Estudios fronterizos*, 61-80.

U.S. Census Bureau. (2020). *American Community Survey [Raw Microdata]*. U.S. Census Bureau.

Vadim, K. (2018). Overview of different approaches to solving problems of data mining. *Procedia computer science*, 234-239.

Vargas, V., & Coubès, M. L. (2017). Working and giving birth in the United States: Changing strategies of transborder life in the north of Mexico. *Frontera norte*, 57-82.

Vega, G. (2016). Población commuter de la frontera norte: el caso de Mexicali-Calexico y Tijuana-San Diego. *Estudios Demográficos y Urbanos,*, 207-238.

Wu, E. (29 de July de 2021). *Supervised vs. Unsupervised Machine Learning*. Obtenido de ExtraHop: https://www.extrahop.com/company/blog/2019/supervised-vs-unsupervised-machine-learning-for-network-threat-detection/

Appendices

Appendix A. Spanish Language Survey Instrument

Encuesta para población transfronteriza

Estimada/o participante,

Mi nombre es Jesús Rivas, soy un estudiante de la Universidad Cuauhtémoc. Estoy realizando una investigación como parte de mi programa de maestría. En mi investigación, busco aprender más acerca de la población transborder (personas cruzando la frontera México – Estados Unidos).

Esta encuesta te tomará 10 minutos. Sus preguntas se relacionan con tus actividades cotidianas y opiniones sobre algunos temas de la dinámica fronteriza entre México y Estados Unidos.

Tu honestidad a lo largo de la encuesta es muy importante. **No te preocupes; en esta encuesta no tendrás que proporcionar tu nombre, teléfono, o domicilio. Además, la información que proporciones es confidencial por lo que no se compartirá con ninguna autoridad; solo se utilizará para fines de investigación.**

Muchas gracias por tu tiempo. Si tienes cualquier pregunta acerca de este estudio, por favor contáctame en el correo rivas251_alfaro@hotmail.com.

Si el COVID-19 y el cierre parcial de la frontera México - Estados Unidos afectó drásticamente tus actividades al otro lado de la frontera o te ha impedido cruzar la frontera durante los últimos meses, **por favor contesta las siguientes preguntas pensando en tu comportamiento antes de esta situación.**

* 1. ¿En qué país vives?

◯ Estados Unidos

◯ México

◯ En ambos países

Encuesta para población transfronteriza

RECUERDA: Tus respuestas son confidenciales y no serán compartidas con ninguna autoridad.

* 2. ¿En que país acostumbras dormir más noches por semana?

◯ Estados Unidos

◯ México

* 3. ¿Cruzas a Estados Unidos por alguno de los siguientes motivos? (selecciona todos los que apliquen)

☐ Escuela (para mí)

☐ Escuela (llevar a alguien más)

☐ Trabajo

☐ Ninguna de las anteriores

161

* 4. ¿Cúal es tu estatus migratorio en los Estados Unidos?

○ No tengo visa ni ningún otro documento que me permita cruzar a Estados Unidos

○ Tengo visa de turista

○ Tengo visa de trabajo o employment authorization document (EAD)

○ Soy residente estadounidense (green card)

○ Soy ciudadano estadounidense

* 5. ¿En qué país o países realizas las siguientes actividades **normalmente**?

	Estados Unidos	México	En ambos países	No realizo esa actividad
Comprar mandado	○	○	○	○
Ir a una reunión o fiesta	○	○	○	○
Votar en las elecciones de ese país	○	○	○	○
Leer/Mira/Escuchar las noticias de ese país	○	○	○	○
Salir a comer a un restaurant "elegante"	○	○	○	○
Ir a la iglesia	○	○	○	○
Salir a comer "en la calle" o comida rápida	○	○	○	○
Ir al cine	○	○	○	○
Hacer algo divertido	○	○	○	○
Ir al doctor	○	○	○	○
Comprar ropa	○	○	○	○
Salir de vacaciones	○	○	○	○
Salir de viaje en auto	○	○	○	○
Cortarme el cabello	○	○	○	○
Asistir a un concierto	○	○	○	○

* 6. ¿Cuál es el idioma que utilizas **normalmente** para las siguientes actividades?

	Inglés	Español	Ambos	No realizó esa actividad
Hablando con mi supervisor(a) del trabajo	○	○	○	○
Hablando con mis compañeros(as) del trabajo	○	○	○	○
Hablando con mis vecinos(as)	○	○	○	○
Hablando con mis amigos(as)	○	○	○	○
Hablando con mis familiares cercanos	○	○	○	○
Mirando la televisión o videos	○	○	○	○
Escuchando la radio o música	○	○	○	○
Leyendo un libro, periódico, etc.	○	○	○	○
Hablando con mis hijos(as)	○	○	○	○
Hablando con mi pareja	○	○	○	○

* 7. Por favor completa las siguientes oraciones indicando el país o cultura más adecuado de acuerdo con tu forma de pensar.

	Estados Unidos/Estadounidense	México/Mexicana	Ambos/Ambas	Ninguno / No aplica
La cultura y/o tradiciones en la que deseo criar/estoy criando a mis hijos	○	○	○	○
En lugar donde prefiero vivir actualmente	○	○	○	○
El lugar en donde me gustaría vivir cuando me jubile/retire	○	○	○	○
El lugar en donde siento que puedo ser yo mismo	○	○	○	○
El lugar donde me la paso mejor	○	○	○	○
La cultura o tradiciones que sigo o con las que me identifico	○	○	○	○
El estilo de vida que prefiero vivir	○	○	○	○
El idioma del país que deseo que mis hijos hablen	○	○	○	○
El lugar en donde viviría si el dinero no fuera un problema	○	○	○	○

163

* 8. ¿Cuál es tu género?

○ Masculino

○ Femenino

○ No binario

○ Prefiero no responder

○ Otro (especifique):

[   ]

* 9. ¿Cuál es tu edad?

○ Menos de 18 años

○ 18 - 24

○ 25 - 34

○ 35 - 44

○ 45 - 49

○ 50 - 54

○ 55 - 64

○ 65 o más

* 10. ¿Cuál es tu nivel educativo?

○ Primaria o menor

○ Secundaria

○ Preparatoria

○ Menos de dos años en la universidad

○ Dos o más años en la universidad

* 11. ¿Cuál es tu nivel de inglés?

○ No sé inglés

○ Entiendo inglés cuando lo escucho, pero no lo hablo

○ Hablo inglés un poco

○ Hablo inglés bien

○ Hablo inglés muy bien

164

* 12. ¿Cuál es tu nivel de español?

○ No sé español

○ Entiendo español cuando lo escucho, pero no lo hablo

○ Hablo español un poco

○ Hablo español bien

○ Hablo español muy bien

* 13. ¿Eres Mexicano, Mexicoamericano, o Chicano?

○ Sí

○ No

* 14. ¿Qué tan frecuentemente cruzas la frontera?

○ Más de tres veces por semana   ○ Cada 10 - 15 días

○ 2 - 3 veces por semana   ○ Una vez por mes

○ 1 vez por semana   ○ Menos de una vez al mes

* 15. ¿Por cuánto tiempo has cruzado la frontera con esta frecuencia?

○ Menos de un año

○ De uno a tres años

○ De tres a cinco años

○ Más de cinco años

* 16. ¿Cuál es la frontera que usas para entrar/salir de tu país **normalmente**?

○ Tijuana – San Diego   ○ Ciudad Acuña - Del Rio   ○ Sonoyta - Lukeville

○ Tecate – San Diego   ○ Agua Prieta - Douglas   ○ Naco - Naco

○ Mexicali – Calexico   ○ Piedras Negras – Eagle Pass ○ Ojinaga - Presidio

○ Los Algodones – Andrade/Yuma   ○ Ciudad Juarez – El Paso   ○ Ciudad Río Bravo - Progreso

○ San Luis Río Colorado – San Luis ○ Chihuahua – Fabens/Tornillo ○ Ciudad Camargo/Los Fresnos – Rio Grande City

○ Nogales - Nogales   ○ El Porvenir – Fort Hancock   ○ Ciudad Miguel Alemán - Roma

○ Matamoros – Brownsville   ○ Reynosa – Higalgo Pharr   ○ San Jerónimo – Santa Teresa

○ Palomas - Columbus   ○ Nuevo Laredo – Laredo

* 17. Vives en una ciudad que esté a más de 100 millas de la frontera? (ejemplo: Phoenix AZ; Hermosillo, SON; Los Angeles, CA; Monterrey, NL; etc.)

○ Sí

○ No

* 18. Por favor selecciona el país o los países en donde has realizado las siguientes activities **a lo largo de tu vida**:

	Estados Unidos	México	Ambos países	Ninguno
País o países en donde he **vivido**	○	○	○	○
País o países en donde he **trabajado**	○	○	○	○
País o países en donde he **estudiado**	○	○	○	○

* 19. Por favor selecciona el país o los países en donde apliquen lo siguiente:

	En Estados Unidos	En México	En ambos países	En ninguno de los países
País o países en donde tengo **familia** que frecuento	○	○	○	○
País o países en donde tengo **amigos(as)** que frecuento	○	○	○	○

Encuesta para población transfronteriza

¡Muchas gracias por participar en esta encuesta!

Esta encuesta ha terminado. Puedes cerrar esta ventana/sitio web.

Si seleccionaste la opción "No tengo visa" has sido enviado al final de esta encuesta ya que no calificas para el resto de las preguntas. Te agradezco mucho tu interés.

Appendix B. English Language Survey Instrument

Transborders survey

Dear respondent,

My name is Jesus Rivas; I am a student from Cuauhtemoc University conducting a research study as part of my master's program. In my research, I am looking to learn more about the transborder population -the people crossing the border between Mexico and the United States.

This survey will take you 10 minutes to complete. Its questions relate to your day-to-day activities and opinions about some topics from the Mexico - United States border dynamics.

Your honesty through the survey is very important. **Do not worry; you will not need to provide your name, telephone, or address in this survey. Moreover, the information you provide is confidential, so it will not be shared with any organization; it will only be used for research purposes.**

Thank you very much for your time. If you have any questions about this study, please contact me at rivas251_alfaro@hotmail.com.

If COVID-19 and the partial closure of the Mexico - United States border impacted your activities on the other side of the border drastically, or if it has prevented you from crossing the border during the last couple of months, **please answer the following questions thinking on your behavior before this situation.**

* 1. In which country do you live?

◯ United States

◯ Mexico

◯ In both countries

REMEMBER: Your responses are confidential and will not be shared with any organization.

* 2. In which country do you usually sleep more nights per week?

◯ United States

◯ Mexico

* 3. Do you cross to the United States for any of the following reasons? (select all that apply)

☐ School (for me)

☐ School (take someone else)

☐ Work

☐ None of the above

* 4. What is your immigration status in the United States?

○ I do not have a visa nor any other document that would allow me to enter the United States

○ I have a tourist visa

○ I have a work visa or an employment authorization document (EAD)

○ I am a United States resident (green-card holder)

○ I am a United States citizen

* 5. In which country or countries do you **typically** do the following activities?

	United States	Mexico	In both countries	I do not do that activity
Buy groceries	○	○	○	○
Go to a gathering or a party	○	○	○	○
Vote during the elections from that country	○	○	○	○
Read/Watch/Listen to the news about that country	○	○	○	○
Go out to eat at a "fancy" restaurant	○	○	○	○
Go to church	○	○	○	○
Go out to eat "on the street" or fast food	○	○	○	○
Go to the movie theater	○	○	○	○
Do something fun	○	○	○	○
Go to the doctor	○	○	○	○
Buy clothes	○	○	○	○
Go on vacations	○	○	○	○
Go on a road trip	○	○	○	○
Get a haircut	○	○	○	○
Go to a concert	○	○	○	○

* 6. What is the language that you **typically** use for the following activities?

	English	Spanish	Both	I do not do that activity
Talking with my work supervisor	○	○	○	○
Talking with my coworkers	○	○	○	○
Talking with my neighbors	○	○	○	○
Talking with my friends	○	○	○	○
Talking with my close relatives	○	○	○	○
Watching TV or videos	○	○	○	○
Listening to the radio or music	○	○	○	○
Reading a book, newspaper, etc.	○	○	○	○
Talking with my children	○	○	○	○
Talking with my partner	○	○	○	○

* 7. Please complete the following statements indicating the country or culture more adequate based on your way of thinking.

	United States / American	Mexico / Mexican	Both	None / Not applicable
The culture and/or traditions in which I want to raise my children	○	○	○	○
The place where I would rather live currently	○	○	○	○
The place where I would rather live once I retire	○	○	○	○
The place where I feel I can be myself	○	○	○	○
The place where I have the best time	○	○	○	○
The culture or traditions I follow or that I identify with	○	○	○	○
The lifestyle I prefer to live	○	○	○	○
The language I want my children to speak	○	○	○	○
The place where I would live if money was not a problem	○	○	○	○

* 8. What is your gender?

○ Male

○ Female

○ Non-binary

○ I prefer not to answer

○ Other (specify):

[   ]

* 9. What is your age?

○ Less than 18 years

○ 18 - 24

○ 25 - 34

○ 35 - 44

○ 45 - 49

○ 50 - 54

○ 55 - 64

○ 65 or more

* 10. What is your education level?

○ Elementary or less

○ Middle school

○ HIgh school

○ Less than two years in college

○ Two or more years in college

* 11. What is your English proficiency level?

○ I do not know English

○ I understand English when I listen to it, but I cannot speak it

○ I speak English a little bit

○ I speak English well

○ I speak English very well

* 12. What is your Spanish proficiency level?

⚪ I do not know Spanish

⚪ I understand Spanish when I listen to it, but I cannot speak it

⚪ I speak Spanish a little bit

⚪ I speak Spanish well

⚪ I speak Spanish very well

* 13. Are you Mexican, Mexican American, or Chicano?

⚪ Yes

⚪ No

* 14. How often do you cross the border?

⚪ More than three times per week   ⚪ Every 10 - 15 days

⚪ 2 - 3 times per week   ⚪ Once per month

⚪ 1 time per week   ⚪ Less than once per month

* 15. For how long have you been crossing the border with this frequency?

⚪ Less than a year

⚪ From one to three years

⚪ From three to five years

⚪ More than five years

* 16. What is the border that you **typically** use to enter/exit your country?

⚪ Tijuana – San Diego   ⚪ Ciudad Acuña - Del Rio   ⚪ Sonoyta - Lukeville

⚪ Tecate – San Diego   ⚪ Agua Prieta - Douglas   ⚪ Naco - Naco

⚪ Mexicali – Calexico   ⚪ Piedras Negras – Eagle Pass   ⚪ Ojinaga - Presidio

⚪ Los Algodones – Andrade/Yuma   ⚪ Ciudad Juarez – El Paso   ⚪ Ciudad Río Bravo - Progreso

⚪ San Luis Río Colorado – San Luis ⚪ Chihuahua – Fabens/Tornillo   ⚪ Ciudad Camargo/Los Fresnos – Rio Grande City

⚪ Nogales - Nogales   ⚪ El Porvenir – Fort Hancock

⚪ Matamoros – Brownsville   ⚪ Reynosa – Higalgo Pharr   ⚪ Ciudad Miguel Alemán - Roma

⚪ Palomas - Columbus   ⚪ Nuevo Laredo – Laredo   ⚪ San Jerónimo – Santa Teresa

* 17. Do you live in a city that is more than 100 miles from the border? (example: Phoenix AZ; Hermosillo, SON; Los Angeles, CA; Monterrey, NL; etc.)

○ Yes

○ No

* 18. Please select the country or countries where you have done the following activities **throughout your life**:

	United States	Mexico	Both countries	None
Country or countries where I have **lived**	○	○	○	○
Country or countries where I have **worked**	○	○	○	○
Country or countries where I have **studied**	○	○	○	○

* 19. Please select the country or the countries in which the following is applicable:

	Unites States	Mexico	Both countries	None
Country or countries where I have **family** that I frequent	○	○	○	○
Country or countries where I have **friends** that I frequent	○	○	○	○

Thank you so much for participating in this survey!

This survey has ended. You may close this window/web site now.

If you selected the option "I do not have a visa" you have been sent to the end of this survey since you do not qualify for the rest of the questions. Thank you for your interest.

Appendix C. Glossary

Table 12

Term Glossary

Term	Definition
Border	Physical barrier separating two countries while also managing the entrance of goods and people.
Borderland	Adjacent regions to a border.
Machine Learning	Set of tools useful to identify patterns, analyze information and build exploratory and predictive models.
Regression Model	Statistical technique to identify correlations between a set of independent variables and one or more dependent variables.
Supervise Machine Learning Model	Set of algorithms within the Machine Learning techniques focused in identifying patterns based on an outcome variable.
Transborder	Individual with the ability of crossing the border back and forth with relative easiness.
Transfronteridad	The level of interaction and dependence towards both sides of the border.

Source. Author's elaboration.

Appendix D. State of the Art Summary

Table 13

State of Art (1/8)

Reference	Information
Falcón, V., & Orta, A. (2018). The Transborder Identity Formation Process: An Exploratory Grounded Theory Study of Transfronterizo College Students from the San Diego-Tijuana Border Region. Journal of Transborders Studies, 1-26.	Place or Population: Transborder College Student living at the Tijuana/San Diego border. Objective: Build an exploratory grounded theory to study transborder students. Method: Qualitative Interviews. Results and Conclusions: A series of challenges, coping mechanisms, experiences, and other insights from 12 students participating in daily transborder dynamics. Transborder students build their transborder identities based on their daily struggles crossing the border, their support networks, their socioeconomic position, and other factors that vary at the individual level. Relationship: Document provides information needed to understand better transborder students' dynamics at the border
Calzada, E., Covas, M., Ramirez, D., Miller, L., & Huan, K.-Y. (2016). A Longitudinal Study of Cultural Adaptation among Mexican and Dominican Immigrant Women. International Migration and Integration, 1049–1063.	Place or Population: Immigrant Women from Mexico and Dominican Republic living in Ney York City, U.S. Objective: Identify factors conditioning the Cultural Adaption process among Mexican and Dominican Immigrant Women. Method: Correlation Analysis using a Longitudinal Dataset. Results and Conclusions: Poverty and the condition of living in immigrant-dense neighborhood can be correlated with reduced acculturation levels for both Mexican and Dominican women. Cultural adaptation is a complex construct affected by individual and cultural contexts; thus, it is not possible to predict cultural adaptation indexes based on a few indicators. Relationship: Document provides information needed to understand applications that cultural adaptation theory has on transborder dynamics and relationship with border theory concepts.

174

Table 13

State of Art (2/8)

Reference	Information
Rocha, D., & Orraca-Romano, P. (2018). Estudiantes de educación superior transfronterizos: Residir en México y estudiar en Estados Unidos. Frontera norte, 103-128.	Place or Population: Transborder students living in the Mexican municipalities of Tijuana, Mexicali, San Luis Río Colorado, Nogales, Ciudad Juárez, Ciudad Acuña and Piedras Negras, Matamoros, Nuevo Laredo and Reynosa. Objective: Characterize students living in Mexico while attending classes in the U.S. Method: Descriptive Statistics using results from the 2010 Mexican Population Census and the 2015 Intercensal Survey Results and Conclusions: The condition of living in Mexico while studying in the U.S. is impacted by factors such as socioeconomic level, U.S. immigration status, condition of belonging to a transborder household, among others. Transborder students experience different challenges while living in Mexico and studying in the U.S. These challenges vary by their socioeconomic level and other conditions surrounding them. Relationship: Document provides information needed to understand better transborder students' dynamics at the border
Orraca-Romano, P. (2019). Cross-Border Earnings of Mexican Workers Across the US–Mexico Border. Journal of Borderlands Studies, 451 - 469.	Place or Population: Transborder worked living in Mexico while working in the U.S. Objective: Characterize the transborder population living in Mexico while working in the U.S. Method: Multivariable Regression Analysis using a Longitudinal Dataset. Results and Conclusions: A decline in the number of transborder workers was observer between 2010 and 2015, which was associated with the economic outlook at that time. Additionally, the author found various variables impacting transborder earning. The earning for Mexican transborder working in the U.S. are conditioned by their skills, educational attainment, condition of having lived and worked in the U.S., by gender, among other variables. Relationship: Document provides information needed to understand better transborder workers' dynamics at the border

Table 13

State of Art (3/8)

Reference	Information
Tapia, L. (2017). Las fronteras, la movilidad y lo transfronterizo: Reflexiones para un debate. Estudios fronterizos, 61-80.	Place or Population: Transborder population living in the Mexico-U.S. border. Objective: Highlight the importance of using transdisciplinary approached when analyzing transborder dynamics. Method: Literature review and case study methods. Results and Conclusions: Phenomena happening at the borderlands cannot be automatically categorized as part of the transborder dynamics. The authors provides evidence for a reflection regarding phenomena that should and should not be considered part of transborder dynamics. Relationship: Document provides information needed to understand better transborder dynamics at the border
Escamilla, E. (2019). Transfronteridad and Everyday Commuter Negotiations of Tijuana's Borderscape [Unpublished master's thesis]. Basel University.	Place or Population: Transborder population living in the Tijuana-San Diego border. Objective: Explore the personal narratives from individuals crossing the border frequently. Method: Qualitative ethnographies. Results and Conclusions: Day-to-day commuting allow individuals to develop transfronteridad which can be defined as interactions occurring at the border that require international crossings to be completed most of the time. Relationship: Document provides information needed to develop an operational definition for the concept of Transfronteridad.

Table 13

State of Art (4/8)

Reference	Information
Vega, G. (2016). Población commuter de la frontera norte: el caso de Mexicali-Calexico y Tijuana-San Diego. Estudios Demográficos y Urbanos,, 207-238.	Place or Population: Transborder worker living either in Mexicali or Tijuana while working in Calexico or San Diego Objective: Characterize transborder inhabitants living at the Mexico-U.S. border and compare this group against non-transborder inhabitants. Method: Descriptive Statistics using results from the 2010 Mexican Population Census. Results and Conclusions: Individuals crossing the border to work at the United States present different demographic traits when compared to their non-transborder counterparts living in Mexico. Relationship: Document provides information needed to understand better transborder workers' dynamics.
Vargas, V., & Coubès, M. L. (2017). Working and giving birth in the United States: Changing strategies of transborder life in the north of Mexico. Frontera norte, 57-82.	Place or Population: Transborder individuals living in the Mexico-U.S. border. Objective: Characterize transborder individuals living in Mexico who crosses the border to give birth in the U.S. Method: Descriptive Statistics using results from the 2000 and 2010 Mexican Population Census. Results and Conclusions: Finding include a decrease in the number of transborder workers while an increase in the transborder births reported. Authors concluded that evidence allows to think that border has developed a set of barriers that limit the ability from individuals living in the borderlands to participate in transborder dynamics, including age and educational levels. Relationship: Document provides information regarding factors impacting transborder dynamics.

Table 13

State of Art (5/8)

Reference	Information
Orraca, P., Rocha, D., & Vargas, E. (2017). Cross-border school enrollment: Associated factors in the US–Mexico borderlands. *The Social Science Journal*, 389-402.	Place or Population: Transborder individuals living in the Mexico-U.S. border. Objective: Create an inferential statistical model to estimate the probability of individuals to adopt a transborder lifestyle. Method: Logit Regression Models using Data from the 2015 Intercensal Survey. Results and Conclusions: There is a significant statistical correlation between a higher probability of being a cross-border student with their age, household income, and also having someone in the household living a work commuting lifestyle. Authors expressed that individuals may be interested to go back and forth between border to study due to an increased quality level in education and interest in eventually transitioning to the U.S. labor market Relationship: Document provides information regarding factors impacting transborder dynamics.
Hernández, H. A. (2020). La frontera México-Estados Unidos: asimetrías y transgresiones. Nueva sociedad, 59-69.	Place or Population: Transborder population living in the Mexico-U.S. border. Objective: Provide relevant information the current status of transborder relationship between Mexico and the U.S. Method: Literature review and case study methods. Results and Conclusions: The author describe various functioning mechanisms at the border such as the types of lanes individuals use to cross and the differences between levels of security in Mexican and U.S. ports of entry. Relationship: Document provides information needed to understand better transborder dynamics at the border

Table 13

State of Art (6/8)

Reference	Information
Vadim, K. (2018). Overview of different approaches to solving problems of data mining. Procedia computer science, 234-239.	Place or Population: General Reflection about Data Mining. Objective: Provide an overview of the different data mining techniques available to solve problems. Method: Literature review and case study methods. Results and Conclusions: Data Mining problems can be solved in two ways, using supervised learning or unsupervised learning approaches. Both approaches come from Machine Learning techniques. In general, supervised models are divided into classification and regression problems. Conversely, unsupervised models then to be exploratory and lack of classification variables to that allow testing model accuracy. Relationship: Document provides information needed to understand better various data mining techniques useful to analyze big data.
Kosinski, M., & Behrend, T. (2017). Editorial overview: Big data in the behavioral sciences. n.d.	Place or Population: General Reflection about applications for Big Data. Objective: Describe the applications Big Data may have in behavioral sciences. Method: Literature review and case study methods. Results and Conclusions: There is still more to improve in terms of privacy when it comes to Big Data, which gray areas that need to be addressed. Additionally, it is important for analyst to understand the importance of connecting results from data models to real applications. Analysts not actively avoiding pitfalls associated with this gap may be overfitting and misapplying models. Relationship: Document provides information needed to understand better various data mining techniques useful to analyze big data.

Table 13

State of Art (7/8)

Reference	Information
Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: an overview. Journal of physics, n.d.	Place or Population: General Reflection about applications for Machine Learning. Objective: Describe the applications for Machine Learning methods and describe the six steps for the development of Machine Learning methods. Method: Literature review and case study methods. Results and Conclusions: The authors highlight that Machine Learning is a multidisciplinary field that focuses in solving problems such as classification, anomaly detention, regression, clustering, and reinforcement problems. There are six steps in the general model for Machine Learning project development described by these authors in this document. Relationship: Document provides information needed to understand better various techniques to build Machine Learning models.
Lu, C., Lin, G., Wu, T., Hu, I., & Chang, Y. (2021). Influencing Factors of Cross-Border E-Commerce Consumer Purchase Intention Based on Wireless Network and Machine Learning. Security and Communication Networks.	Place or Population: Cross-borders participating in e-commerce activities in China. Objective: Describe the applications for Machine Learning methods and describe the six steps for the development of Machine Learning methods. Method: Machine Learning models using a survey dataset. Results and Conclusions: The authors created a machine learning model to predict consumption frequency of e-commerce among cross-border consumers. Results show that individuals see their cross-border consumption frequency affected by the products perceived usefulness, perceived ease of use, consumers'' income level, educational attainment, age, gender, service, and safety index. Relationship: Document provides information regarding using machine learning methods to analyze information from cross border dynamics.

Table 13

State of Art (8/8)

Reference	Information
Sharma, S., Kang, D., de Oca, J., & Mudgal, A. (2021). Machine learning methods for commercial vehicle wait time prediction at a border crossing. Research in Transportation Economics.	*Place or Population: Transborders using the commercial lanes to access the U.S. through the Mexico-U.S. border.*
	Objective: Develop a model to predict the crossing border times at the commercial lanes in the Mexico-U.S. border.
	Method: Machine Learning models using a survey dataset.
	Results and Conclusions: Machine Learning methods, including Gradient Boosting Regression and Random Forest methods were applied. The results were not conclusive due to the high variability of the data creating non-reliable predictions. Authors concluded that while results were not conclusive, the information gathered will be helpful to understand the performance of this type of models.
	Relationship: Document provides information regarding using machine learning methods to analyze information from cross border dynamics.

Source. Author's elaboration.

181

Ancillary

Ancillary A. Master's Program Concepts Applied in Research Project

Table 14 (1/2)

Concepts from master's Program Courses used in this Research Project

Course	Concepts Used
1st Quarter	
Introduction to Cloud Computing	Use of cloud-based systems for thesis work such as OneDrive, Google Drive, and Tableau Public
Fundamentals for Statistics	Use of Statistical Methods for study, including descriptive and inferential statistics
Data Analysis with R	Use of R packages for data processing, including *psych*, *gmodels*, *expss*, *dplyr*, *ggcorplot*, *ltm*, and *mirt*.
Machine Learning and Data Treatment	Use of Supervised Machine Learning methods for data analysis such as linear regression
2nd Quarter	
Knowledge Discovery with Databases	Use of Supervised Machine Learning methods for data analysis such as linear regression
Dynamic Statistical Models	Use of Statistical Methods for study, including descriptive and inferential statistics
Business Analytics	Use Strategic Business concepts to determine research method for study
Data Warehousing and Business Intelligence	Use of Database and Data Warehousing tools for thesis work

Table 14 (2/2)

Concepts from master's Program Courses used in this Research Project

Course	Concepts Used
3rd Quarter	
Advanced Technology and Databases	Use of Database and Data Warehousing tools for thesis work
Numerical and Optimization Methods	Use optimization-based methods for study, including descriptive and inferential statistics
Web Analytics	Use of web-analytics tools to manage survey
Big Data Workshop	Use of big data tools and methods to process more than 20 million records from the Mexican and American censuses datasets
4th Quarter	
Data Visualization Workshop	Use of data visualization tools such as Mapa Digital de Mexico and Tableau to develop maps for thesis
Advanced Technologies in Data Mining and Classification	Use of Supervised Machine Learning methods for data analysis such as linear regression
Big Data Cloud	Use of big data tools and methods to process more than 20 million records from the Mexican and American censuses datasets
Research Project and State of the Art on Big Data	Use of research methods for thesis work including method development and state of the art research

Source. Author's elaboration.

Ancillary B. Turnitin Analysis

Ancillary C. Syntax used to process and report on data

<u>Using SPSS to process 12+ million records from Mexican Census Data.</u>

```
* Encoding: UTF-8.
GET DATA  /TYPE=TXT
 /FILE="C:\Users\jrivas\Desktop\Census
Data\Personas00.CSV"
 /ENCODING='Locale'
 /DELCASE=LINE
 /DELIMITERS=","
 /ARRANGEMENT=DELIMITED
 /FIRSTCASE=2
 /IMPORTCASE=ALL
 /VARIABLES=
 ENT F2.0
 MUN F3.0
 LOC50K F4.0
 ID_VIV F12.0
 ID_PERSONA A17
 COBERTURA F1.0
 ESTRATO A14
 UPM F5.0
 FACTOR F2.0
 CLAVIVP F2.0
 NUMPER F2.0
 SEXO F1.0
 EDAD F2.0
 PARENTESCO F3.0
 IDENT_MADRE F2.0
 IDENT_PADRE F2.0
 ENT_PAIS_NAC F3.0
 NACIONALIDAD F1.0
 SERSALUD F2.0
 AFRODES F1.0
 REGIS_NAC F1.0
 DHSERSAL1 F2.0
 DHSERSAL2 F1.0
 RELIGION F4.0
 DIS_VER F1.0
 DIS_OIR F1.0
 DIS_CAMINAR F1.0
 DIS_RECORDAR F1.0
 DIS_BANARSE F1.0
 DIS_HABLAR F1.0
 DIS_MENTAL F1.0
 CAU_VER F1.0
 CAU_OIR F1.0
 CAU_CAMINAR F1.0
 CAU_RECORDAR F1.0
 CAU_BANARSE F1.0
 CAU_HABLAR F1.0
 CAU_MENTAL F1.0
 HLENGUA F1.0
 QDIALECT_INALI F1.0
 HESPANOL F1.0
 ELENGUA F1.0
 PERTE_INDIGENA F1.0
 ASISTEN F1.0
 MUN_ASI F3.0
 ENT_PAIS_ASI F3.0
 TIE_TRASLADO_ESCU F1.0
 MED_TRASLADO_ESC1 F2.0
 MED_TRASLADO_ESC2 F1.0
 MED_TRASLADO_ESC3 F1.0
 NIVACAD F2.0
 ESCOLARI F2.0
 NOMCAR_C F4.0
 ALFABET F1.0
 ESCOACUM F2.0
 ENT_PAIS_RES_5A F3.0
 MUN_RES_5A F3.0
 CAUSA_MIG F2.0
 CAUSA_MIG_V F4.0
 SITUA_CONYUGAL F1.0
 IDENT_PAREJA F2.0
 CONACT F2.0
 OCUPACION_C F3.0
 SITTRA F1.0
 AGUINALDO F1.0
 VACACIONES F1.0
 SERVICIO_MEDICO F1.0
 UTILIDADES F1.0
 INCAP_SUELDO F1.0
 SAR_AFORE F1.0
 CREDITO_VIVIENDA F1.0
 INGTRMEN F5.0
 HORTRA F2.0
 ACTIVIDADES_C F4.0
 MUN_TRAB F3.0
 ENT_PAIS_TRAB F3.0
 TIE_TRASLADO_TRAB F1.0
 MED_TRASLADO_TRAB1 F2.0
 MED_TRASLADO_TRAB2 F2.0
 MED_TRASLADO_TRAB3 F1.0
 HIJOS_NAC_VIVOS F1.0
 HIJOS_FALLECIDOS F1.0
 HIJOS_SOBREVIV F1.0
 FECHA_NAC_M F2.0
 FECHA_NAC_A F4.0
 SOBREVIVENCIA F1.0
 IDENT_HIJO F2.0
 EDAD_MORIR_D F2.0
 EDAD_MORIR_M F1.0
 EDAD_MORIR_A F1.0
 EDAD_MORIR_TD F3.0
 TAMLOC F1.0.
CACHE.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.

SAVE     OUTFILE='C:\Users\jrivas\Desktop\Census
Data\cpyv20.sav'
 /COMPRESSED.

DATASET ACTIVATE DataSet1.
IF (ENT =        2        & MUN =        2
        )BORDER = 1.
```

IF (ENT =   2   & MUN =   3
)BORDER = 1.
IF (ENT =   2   & MUN =   4
)BORDER = 1.
IF (ENT =   5   & MUN =   2
)BORDER = 1.
IF (ENT =   5   & MUN =   12
)BORDER = 1.
IF (ENT =   5   & MUN =   13
)BORDER = 1.
IF (ENT =   5   & MUN =   14
)BORDER = 1.
IF (ENT =   5   & MUN =   22
)BORDER = 1.
IF (ENT =   5   & MUN =   23
)BORDER = 1.
IF (ENT =   5   & MUN =   25
)BORDER = 1.
IF (ENT =   8   & MUN =   5
)BORDER = 1.
IF (ENT =   8   & MUN =   28
)BORDER = 1.
IF (ENT =   8   & MUN =   35
)BORDER = 1.
IF (ENT =   8   & MUN =   37
)BORDER = 1.
IF (ENT =   8   & MUN =   42
)BORDER = 1.
IF (ENT =   8   & MUN =   52
)BORDER = 1.
IF (ENT =   8   & MUN =   53
)BORDER = 1.
IF (ENT =   19   & MUN =   5
)BORDER = 1.
IF (ENT =   26   & MUN =   2
)BORDER = 1.
IF (ENT =   26   & MUN =   4
)BORDER = 1.
IF (ENT =   26   & MUN =   17
)BORDER = 1.
IF (ENT =   26   & MUN =   19
)BORDER = 1.
IF (ENT =   26   & MUN =   39
)BORDER = 1.
IF (ENT =   26   & MUN =   43
)BORDER = 1.
IF (ENT =   26   & MUN =   48
)BORDER = 1.
IF (ENT =   26   & MUN =   55
)BORDER = 1.
IF (ENT =   26   & MUN =   59
)BORDER = 1.
IF (ENT =   26   & MUN =   60
)BORDER = 1.
IF (ENT =   26   & MUN =   70
)BORDER = 1.
IF (ENT =   28   & MUN =   7
)BORDER = 1.
IF (ENT =   28   & MUN =   14
)BORDER = 1.
IF (ENT =   28   & MUN =   15
)BORDER = 1.

IF (ENT =   28   & MUN =   22
)BORDER = 1.
IF (ENT =   28   & MUN =   24
)BORDER = 1.
IF (ENT =   28   & MUN =   25
)BORDER = 1.
IF (ENT =   28   & MUN =   27
)BORDER = 1.
IF (ENT =   28   & MUN =   32
)BORDER = 1.
IF (ENT =   28   & MUN =   33
)BORDER = 1.
IF (ENT =   28   & MUN =   40
)BORDER = 1.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (BORDER = 1).
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (ENT =   2   & MUN =   2
)CITY = 22.
IF (ENT =   2   & MUN =   3
)CITY = 23.
IF (ENT =   2   & MUN =   4
)CITY = 24.
IF (ENT =   5   & MUN =   2
)CITY = 54.
IF (ENT =   5   & MUN =   12
)CITY = 512.
IF (ENT =   5   & MUN =   13
)CITY = 513.
IF (ENT =   5   & MUN =   14
)CITY = 514.
IF (ENT =   5   & MUN =   22
)CITY = 522.
IF (ENT =   5   & MUN =   23
)CITY = 523.
IF (ENT =   5   & MUN =   25
)CITY = 525.
IF (ENT =   8   & MUN =   5
)CITY = 85.
IF (ENT =   8   & MUN =   28
)CITY = 828.
IF (ENT =   8   & MUN =   35
)CITY = 835.
IF (ENT =   8   & MUN =   37
)CITY = 837.
IF (ENT =   8   & MUN =   42
)CITY = 842.
IF (ENT =   8   & MUN =   52
)CITY = 852.
IF (ENT =   8   & MUN =   53
)CITY = 853.
IF (ENT =   19   & MUN =   5
)CITY = 195.
IF (ENT =   26   & MUN =   2
)CITY = 262.
IF (ENT =   26   & MUN =   4
)CITY = 264.

```
IF (ENT =        26      & MUN =        17
      )CITY = 2617.
IF (ENT =        26      & MUN =        19
      )CITY = 2619.
IF (ENT =        26      & MUN =        39
      )CITY = 2639.
IF (ENT =        26      & MUN =        43
      )CITY = 2643.
IF (ENT =        26      & MUN =        48
      )CITY = 2648.
IF (ENT =        26      & MUN =        55
      )CITY = 2655.
IF (ENT =        26      & MUN =        59
      )CITY = 2659.
IF (ENT =        26      & MUN =        60
      )CITY = 2660.
IF (ENT =        26      & MUN =        70
      )CITY = 2670.
IF (ENT =        28      & MUN =        7
      )CITY = 287.
IF (ENT =        28      & MUN =        14
      )CITY = 2814.
IF (ENT =        28      & MUN =        15
      )CITY = 2815.
IF (ENT =        28      & MUN =        22
      )CITY = 2822.
IF (ENT =        28      & MUN =        24
      )CITY = 2824.
IF (ENT =        28      & MUN =        25
      )CITY = 2825.
IF (ENT =        28      & MUN =        27
      )CITY = 2827.
IF (ENT =        28      & MUN =        32
      )CITY = 2832.
IF (ENT =        28      & MUN =        33
      )CITY = 2833.
IF (ENT =        28      & MUN =        40
      )CITY = 2840.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (ENT_PAIS_TRAB = 221) TB_W = 1.
IF (ENT_PAIS_NAC = 221) TB_B = 1.
IF (ENT_PAIS_RES_5A = 221 & CAUSA_MIG_V <>
801) TB_E = 1.
IF (ENT_PAIS_ASI = 221) TB_S = 1.
IF (TB_W = 1 OR TB_B = 1 OR TB_E = 1 OR TB_S =
1) TB = 1.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (NIVACAD = 14) ESCOLARIDAD = 5.
IF (NIVACAD = 13) ESCOLARIDAD = 5.
IF (NIVACAD = 12) ESCOLARIDAD = 5.
IF (NIVACAD = 11 & ESCOLARI > 1) ESCOLARIDAD
= 5.
IF (NIVACAD = 11 & ESCOLARI = 1) ESCOLARIDAD
= 4.
IF (NIVACAD = 10 & ESCOLARI > 1) ESCOLARIDAD
= 5.
IF (NIVACAD = 10 & ESCOLARI = 1) ESCOLARIDAD
= 4.
```

```
IF (NIVACAD > 3 & NIVACAD < 10) ESCOLARIDAD =
3.
IF (NIVACAD = 3) ESCOLARIDAD = 2.
IF (NIVACAD < 3) ESCOLARIDAD = 1.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (EDAD < 18) EDADR = 0.
IF (EDAD >= 18 & EDAD <=24) EDADR = 1.
IF (EDAD >= 25 & EDAD <=34) EDADR = 2.
IF (EDAD >= 35 & EDAD <=44) EDADR = 3.
IF (EDAD >= 45 & EDAD <=49) EDADR = 4.
IF (EDAD >= 50 & EDAD <=54) EDADR = 5.
IF (EDAD >= 55 & EDAD <=64) EDADR = 6.
IF (EDAD >= 64 & EDAD <=999) EDADR = 7.
IF (EDAD =999) EDADR = 999.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (BORDER = 1 & TB = 1 & EDAD >= 18 &
EDAD ~= 999).
EXECUTE.

SAVE      OUTFILE='C:\Users\jrivas\Desktop\Census
Data\cpyv20_processed.sav'
 /COMPRESSED.

WEIGHT BY FACTOR.

* Custom Tables.

DATASET ACTIVATE DataSet1.
FREQUENCIES  VARIABLES=TB_W  TB_B  TB_E
TB_S TB
 /ORDER=ANALYSIS.

* Custom Tables.
CTABLES
 /VLABELS      VARIABLES=SEXO      EDADR
ESCOLARIDAD ENT DISPLAY=LABEL
 /TABLE SEXO [C][COUNT F40.0, COLPCT.COUNT
PCT40.1]    +    EDADR    [C][COUNT    F40.0,
COLPCT.COUNT PCT40.1]
   +    ESCOLARIDAD    [C][COUNT    F40.0,
COLPCT.COUNT PCT40.1] + ENT [C][COUNT F40.0,
COLPCT.COUNT PCT40.1]
 /CATEGORIES    VARIABLES=SEXO    EDADR
ESCOLARIDAD   ENT   ORDER=A   KEY=VALUE
EMPTY=EXCLUDE TOTAL=YES
   POSITION=AFTER.

OUTPUT SAVE NAME=Document1
 OUTFILE='C:\Users\jrivas\Desktop\Census
Data\Cpyv20.spv'
 LOCK=NO.
```

Using SPSS to process 2+ records from the American Community Survey

```
* Encoding: UTF-8.

GET DATA  /TYPE=TXT
 /FILE="C:\Users\jrivas\Desktop\Census
Data\Personas00.CSV"
 /ENCODING='Locale'
 /DELCASE=LINE
 /DELIMITERS=","
 /ARRANGEMENT=DELIMITED
 /FIRSTCASE=2
 /IMPORTCASE=ALL
 /VARIABLES=
 ENT F2.0
 MUN F3.0
 LOC50K F4.0
 ID_VIV F12.0
 ID_PERSONA A17
 COBERTURA F1.0
 ESTRATO A14
 UPM F5.0
 FACTOR F2.0
 CLAVIVP F2.0
 NUMPER F2.0
 SEXO F1.0
 EDAD F2.0
 PARENTESCO F3.0
 IDENT_MADRE F2.0
 IDENT_PADRE F2.0
 ENT_PAIS_NAC F3.0
 NACIONALIDAD F1.0
 SERSALUD F2.0
 AFRODES F1.0
 REGIS_NAC F1.0
 DHSERSAL1 F2.0
 DHSERSAL2 F1.0
 RELIGION F4.0
 DIS_VER F1.0
 DIS_OIR F1.0
 DIS_CAMINAR F1.0
 DIS_RECORDAR F1.0
 DIS_BANARSE F1.0
 DIS_HABLAR F1.0
 DIS_MENTAL F1.0
 CAU_VER F1.0
 CAU_OIR F1.0
 CAU_CAMINAR F1.0
 CAU_RECORDAR F1.0
 CAU_BANARSE F1.0
 CAU_HABLAR F1.0
 CAU_MENTAL F1.0
 HLENGUA F1.0
 QDIALECT_INALI F1.0
 HESPANOL F1.0
 ELENGUA F1.0
 PERTE_INDIGENA F1.0
 ASISTEN F1.0
 MUN_ASI F3.0
 ENT_PAIS_ASI F3.0
 TIE_TRASLADO_ESCU F1.0
 MED_TRASLADO_ESC1 F2.0
 MED_TRASLADO_ESC2 F1.0
 MED_TRASLADO_ESC3 F1.0
 NIVACAD F2.0
 ESCOLARI F2.0
 NOMCAR_C F4.0
 ALFABET F1.0
 ESCOACUM F2.0
 ENT_PAIS_RES_5A F3.0
 MUN_RES_5A F3.0
 CAUSA_MIG F2.0
 CAUSA_MIG_V F4.0
 SITUA_CONYUGAL F1.0
 IDENT_PAREJA F2.0
 CONACT F2.0
 OCUPACION_C F3.0
 SITTRA F1.0
 AGUINALDO F1.0
 VACACIONES F1.0
 SERVICIO_MEDICO F1.0
 UTILIDADES F1.0
 INCAP_SUELDO F1.0
 SAR_AFORE F1.0
 CREDITO_VIVIENDA F1.0
 INGTRMEN F5.0
 HORTRA F2.0
 ACTIVIDADES_C F4.0
 MUN_TRAB F3.0
 ENT_PAIS_TRAB F3.0
 TIE_TRASLADO_TRAB F1.0
 MED_TRASLADO_TRAB1 F2.0
 MED_TRASLADO_TRAB2 F2.0
 MED_TRASLADO_TRAB3 F1.0
 HIJOS_NAC_VIVOS F1.0
 HIJOS_FALLECIDOS F1.0
 HIJOS_SOBREVIV F1.0
 FECHA_NAC_M F2.0
 FECHA_NAC_A F4.0
 SOBREVIVENCIA F1.0
 IDENT_HIJO F2.0
 EDAD_MORIR_D F2.0
 EDAD_MORIR_M F1.0
 EDAD_MORIR_A F1.0
 EDAD_MORIR_TD F3.0
 TAMLOC F1.0.
CACHE.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.

SAVE      OUTFILE='C:\Users\jrivas\Desktop\Census
Data\cpyv20.sav'
 /COMPRESSED.

DATASET ACTIVATE DataSet1.
IF (ENT =          2        & MUN =          2
        )BORDER = 1.
IF (ENT =          2        & MUN =          3
        )BORDER = 1.
IF (ENT =          2        & MUN =          4
        )BORDER = 1.
IF (ENT =          5        & MUN =          2
        )BORDER = 1.
IF (ENT =          5        & MUN =          12
        )BORDER = 1.
```

IF (ENT =   5   & MUN =   13
)BORDER = 1.
IF (ENT =   5   & MUN =   14
)BORDER = 1.
IF (ENT =   5   & MUN =   22
)BORDER = 1.
IF (ENT =   5   & MUN =   23
)BORDER = 1.
IF (ENT =   5   & MUN =   25
)BORDER = 1.
IF (ENT =   8   & MUN =   5
)BORDER = 1.
IF (ENT =   8   & MUN =   28
)BORDER = 1.
IF (ENT =   8   & MUN =   35
)BORDER = 1.
IF (ENT =   8   & MUN =   37
)BORDER = 1.
IF (ENT =   8   & MUN =   42
)BORDER = 1.
IF (ENT =   8   & MUN =   52
)BORDER = 1.
IF (ENT =   8   & MUN =   53
)BORDER = 1.
IF (ENT =   19   & MUN =   5
)BORDER = 1.
IF (ENT =   26   & MUN =   2
)BORDER = 1.
IF (ENT =   26   & MUN =   4
)BORDER = 1.
IF (ENT =   26   & MUN =   17
)BORDER = 1.
IF (ENT =   26   & MUN =   19
)BORDER = 1.
IF (ENT =   26   & MUN =   39
)BORDER = 1.
IF (ENT =   26   & MUN =   43
)BORDER = 1.
IF (ENT =   26   & MUN =   48
)BORDER = 1.
IF (ENT =   26   & MUN =   55
)BORDER = 1.
IF (ENT =   26   & MUN =   59
)BORDER = 1.
IF (ENT =   26   & MUN =   60
)BORDER = 1.
IF (ENT =   26   & MUN =   70
)BORDER = 1.
IF (ENT =   28   & MUN =   7
)BORDER = 1.
IF (ENT =   28   & MUN =   14
)BORDER = 1.
IF (ENT =   28   & MUN =   15
)BORDER = 1.
IF (ENT =   28   & MUN =   22
)BORDER = 1.
IF (ENT =   28   & MUN =   24
)BORDER = 1.
IF (ENT =   28   & MUN =   25
)BORDER = 1.
IF (ENT =   28   & MUN =   27
)BORDER = 1.

IF (ENT =   28   & MUN =   32
)BORDER = 1.
IF (ENT =   28   & MUN =   33
)BORDER = 1.
IF (ENT =   28   & MUN =   40
)BORDER = 1.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (BORDER = 1).
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (ENT =   2   & MUN =   2
)CITY = 22.
IF (ENT =   2   & MUN =   3
)CITY = 23.
IF (ENT =   2   & MUN =   4
)CITY = 24.
IF (ENT =   5   & MUN =   2
)CITY = 54.
IF (ENT =   5   & MUN =   12
)CITY = 512.
IF (ENT =   5   & MUN =   13
)CITY = 513.
IF (ENT =   5   & MUN =   14
)CITY = 514.
IF (ENT =   5   & MUN =   22
)CITY = 522.
IF (ENT =   5   & MUN =   23
)CITY = 523.
IF (ENT =   5   & MUN =   25
)CITY = 525.
IF (ENT =   8   & MUN =   5
)CITY = 85.
IF (ENT =   8   & MUN =   28
)CITY = 828.
IF (ENT =   8   & MUN =   35
)CITY = 835.
IF (ENT =   8   & MUN =   37
)CITY = 837.
IF (ENT =   8   & MUN =   42
)CITY = 842.
IF (ENT =   8   & MUN =   52
)CITY = 852.
IF (ENT =   8   & MUN =   53
)CITY = 853.
IF (ENT =   19   & MUN =   5
)CITY = 195.
IF (ENT =   26   & MUN =   2
)CITY = 262.
IF (ENT =   26   & MUN =   4
)CITY = 264.
IF (ENT =   26   & MUN =   17
)CITY = 2617.
IF (ENT =   26   & MUN =   19
)CITY = 2619.
IF (ENT =   26   & MUN =   39
)CITY = 2639.
IF (ENT =   26   & MUN =   43
)CITY = 2643.

```
IF (ENT =        26      & MUN =        48
      )CITY = 2648.
IF (ENT =        26      & MUN =        55
      )CITY = 2655.
IF (ENT =        26      & MUN =        59
      )CITY = 2659.
IF (ENT =        26      & MUN =        60
      )CITY = 2660.
IF (ENT =        26      & MUN =        70
      )CITY = 2670.
IF (ENT =        28      & MUN =        7
      )CITY = 287.
IF (ENT =        28      & MUN =        14
      )CITY = 2814.
IF (ENT =        28      & MUN =        15
      )CITY = 2815.
IF (ENT =        28      & MUN =        22
      )CITY = 2822.
IF (ENT =        28      & MUN =        24
      )CITY = 2824.
IF (ENT =        28      & MUN =        25
      )CITY = 2825.
IF (ENT =        28      & MUN =        27
      )CITY = 2827.
IF (ENT =        28      & MUN =        32
      )CITY = 2832.
IF (ENT =        28      & MUN =        33
      )CITY = 2833.
IF (ENT =        28      & MUN =        40
      )CITY = 2840.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (ENT_PAIS_TRAB = 221) TB_W = 1.
IF (ENT_PAIS_NAC = 221) TB_B = 1.
IF (ENT_PAIS_RES_5A = 221 & CAUSA_MIG_V <>
801) TB_E = 1.
IF (ENT_PAIS_ASI = 221) TB_S = 1.
IF (TB_W = 1 OR TB_B = 1 OR TB_E = 1 OR TB_S =
1) TB = 1.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (NIVACAD = 14) ESCOLARIDAD = 5.
IF (NIVACAD = 13) ESCOLARIDAD = 5.
IF (NIVACAD = 12) ESCOLARIDAD = 5.
IF (NIVACAD = 11 & ESCOLARI > 1) ESCOLARIDAD
= 5.
IF (NIVACAD = 11 & ESCOLARI = 1) ESCOLARIDAD
= 4.
IF (NIVACAD = 10 & ESCOLARI > 1) ESCOLARIDAD
= 5.
IF (NIVACAD = 10 & ESCOLARI = 1) ESCOLARIDAD
= 4.
IF (NIVACAD > 3 & NIVACAD < 10) ESCOLARIDAD =
3.
```

Using R Studio to process survey data

```
# Load libraries
library(psych)
library(gmodels)
```

```
IF (NIVACAD = 3) ESCOLARIDAD = 2.
IF (NIVACAD < 3) ESCOLARIDAD = 1.
EXECUTE.

DATASET ACTIVATE DataSet1.
IF (EDAD < 18) EDADR = 0.
IF (EDAD >= 18 & EDAD <=24) EDADR = 1.
IF (EDAD >= 25 & EDAD <=34) EDADR = 2.
IF (EDAD >= 35 & EDAD <=44) EDADR = 3.
IF (EDAD >= 45 & EDAD <=49) EDADR = 4.
IF (EDAD >= 50 & EDAD <=54) EDADR = 5.
IF (EDAD >= 55 & EDAD <=64) EDADR = 6.
IF (EDAD >= 64 & EDAD <=999) EDADR = 7.
IF (EDAD =999) EDADR = 999.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (BORDER = 1 & TB = 1 & EDAD >= 18 &
EDAD ~= 999).
EXECUTE.

SAVE      OUTFILE='C:\Users\jrivas\Desktop\Census
Data\cpyv20_processed.sav'
 /COMPRESSED.

WEIGHT BY FACTOR.

* Custom Tables.

DATASET ACTIVATE DataSet1.
FREQUENCIES  VARIABLES=TB_W  TB_B  TB_E
TB_S TB
 /ORDER=ANALYSIS.

* Custom Tables.
CTABLES
 /VLABELS        VARIABLES=SEXO        EDADR
ESCOLARIDAD ENT DISPLAY=LABEL
 /TABLE SEXO [C][COUNT F40.0, COLPCT.COUNT
PCT40.1]    +    EDADR    [C][COUNT    F40.0,
COLPCT.COUNT PCT40.1]
    +    ESCOLARIDAD    [C][COUNT    F40.0,
COLPCT.COUNT PCT40.1] + ENT [C][COUNT F40.0,
COLPCT.COUNT PCT40.1]
 /CATEGORIES      VARIABLES=SEXO      EDADR
ESCOLARIDAD   ENT   ORDER=A   KEY=VALUE
EMPTY=EXCLUDE TOTAL=YES
   POSITION=AFTER.

OUTPUT SAVE NAME=Document1
 OUTFILE='C:\Users\jrivas\Desktop\Census
Data\Cpyv20.spv'
 LOCK=NO.
```

```r
library(expss)
library(dplyr)
library(ggcorrplot)
library(ltm)
library(mirt)

# Establish directory
setwd("C:/Users/rivas/Google          Drive/Maestría/5.
Thesis/Segmenting Border Users/Survey Processing")

#Import results
data_spanish                                    <-
read.csv("spanish_completes_092021.csv", header =
FALSE, dec = ",", sep = ",")
data_english                                    <-
read.csv("english_completes_092021.csv", header =
FALSE, dec = ",", sep = ",")
data_spanish$source <- 1
data_english$source <- 2

# Eliminate first two cases that are rubbish
data_spanish <- data_spanish %>% slice(-c(1:2))
data_english <- data_english %>% slice(-c(1:2))

#Merge both versions
data <- rbind(data_spanish,data_english)

# Dump variables that are not needed
data <- data[,-c(1:9)]

#Assign names to variables
var_names                                       <-
c("country_orig","both_countries","comm_none","com
m_school","comm_school_else","comm_work","us_mi
g_status","groc_shopping","reunion","fam_reunion","v
oting","news","restaurant","church","food_street","mov
ies","fun","doctor","clothes","vacations","road_trip","ha
ircut","concert","lang_sup","lang_coworker","lang_nei
ghbor","lang_shop","lang_friend","lang_family","lang_
watch","lang_listen","lang_read","lang_children","lang
_partner","culture_children","living_place","retirement"
,"myself","best_time","culture_myself","lifestyle","pref_
lang_chil","living_place_money","gender","gender_oth
er","age","educational_level","english_proficiency","sp
anish_proficiency","mexican","border_freq","time_cros
sing","border","non_border_city","countries_lived","co
untries_worked","countries_studied","countries_family
","countries_friends","source")
names(data)<- var_names

#Convert variables into numeric
for(var in var_names){
  data[[var]] <- as.numeric(as.character(data[[var]]))
}

#Correct NA's created after change in variable type
for(var in var_names){
  data[[var]][is.na(data[[var]])]<-0
}

# Create selected crosstabs tables to validate that data
matches to Survey Monkey

CrossTable(data$gender,   data$source,   format   =
"SPSS", digits = 0)
CrossTable(data$educational_level,        data$source,
format = "SPSS", digits = 0)

#Create a system of variables to measure
transfronteridad and related variables
  #Temporary extract of variables related to activity
questions
  temp_act <- data[-c(1:7, 10, 24:60)]
  #Calculate activities in the three transborder spaces
  temp_act_aux <- data.frame(matrix(ncol = 9, nrow =
402))
  x <- c("act_us", "act_mex", "act_both", "act_us_aux",
"act_mex_aux",    "act_both_aux",    "act_na_aux",
"act_missing_aux","act_total_aux")
  colnames(temp_act_aux) <- x
  temp_act_aux$act_us <- rowSums(temp_act == 1) /
(rowSums(temp_act == 1)+rowSums(temp_act ==
2)+rowSums(temp_act == 3))
  temp_act_aux$act_mex <- rowSums(temp_act == 2)
/ (rowSums(temp_act == 1)+rowSums(temp_act ==
2)+rowSums(temp_act == 3))
  temp_act_aux$act_both <- rowSums(temp_act == 3)
/(rowSums(temp_act == 1)+rowSums(temp_act ==
2)+rowSums(temp_act == 3))
    #Auxiliar
    temp_act_aux$act_us_aux <- rowSums(temp_act
== 1)
    temp_act_aux$act_mex_aux <- rowSums(temp_act
== 2)
    temp_act_aux$act_both_aux <- rowSums(temp_act
== 3)
    temp_act_aux$act_missing_aux                    <-
rowSums(temp_act == 0)
    temp_act_aux$act_na_aux <- rowSums(temp_act
== 4)
    temp_act_aux$act_total_aux                       <-
(rowSums(temp_act == 1)+rowSums(temp_act ==
2)+rowSums(temp_act == 3))
    temp_results         <-         data.frame(temp_act,
temp_act_aux)
    #Append results to main dataset
    data <- data.frame(data, temp_act_aux)

#Create a system of variables to measure language
usage
  #Temporary extract of variables related to language
questions
  temp_lang                                        <-
data.frame(data$lang_sup,data$lang_coworker,data$l
ang_neighbor,data$lang_shop,data$lang_friend,data
$lang_family,data$lang_watch,data$lang_listen,data$
lang_read,data$lang_children,data$lang_partner)
  #Calculate variables related to language
```

```r
    temp_lang_aux <- data.frame(matrix(ncol = 9, nrow
= 402))
    x    <-    c("lang_us",    "lang_mex",    "lang_both",
"lang_us_aux",    "lang_mex_aux",    "lang_both_aux",
"lang_na_aux", "lang_missing_aux","lang_total_aux")
    colnames(temp_lang_aux) <- x
    temp_lang_aux$lang_us <- rowSums(temp_lang ==
1) / (rowSums(temp_lang == 1)+rowSums(temp_lang
== 2)+rowSums(temp_lang == 3))
    temp_lang_aux$lang_mex <- rowSums(temp_lang
==    2)    /    (rowSums(temp_lang    ==
1)+rowSums(temp_lang  ==  2)+rowSums(temp_lang
== 3))
    temp_lang_aux$lang_both <- rowSums(temp_lang
==    3)    /(rowSums(temp_lang    ==
1)+rowSums(temp_lang  ==  2)+rowSums(temp_lang
== 3))
    #Auxiliar
    temp_lang_aux$lang_us_aux                  <-
rowSums(temp_lang == 1)
    temp_lang_aux$lang_mex_aux                 <-
rowSums(temp_lang == 2)
    temp_lang_aux$lang_both_aux                <-
rowSums(temp_lang == 3)
    temp_lang_aux$lang_missing_aux             <-
rowSums(temp_lang == 0)
    temp_lang_aux$lang_na_aux                  <-
rowSums(temp_lang == 4)
    temp_lang_aux$lang_total_aux               <-
(rowSums(temp_lang == 1)+rowSums(temp_lang ==
2)+rowSums(temp_lang == 3))
    temp_results_lang    <-    data.frame(temp_lang,
temp_lang_aux)
    #Append results to main dataset
    data <- data.frame(data, temp_lang_aux)

#Create a system of variables to measure cultural
identity
    #Temporary extract of variables related to cultural
questions
    temp_cult                                  <-
data.frame(data$culture_children,data$living_place,d
ata$retirement,data$myself,data$best_time,data$cult
ure_myself,data$lifestyle,data$pref_lang_chil,data$livi
ng_place_money)
    #Calculate variables related to cultural identity
    temp_cult_aux <- data.frame(matrix(ncol = 9, nrow
= 402))
    x    <-    c("cult_us",    "cult_mex",    "cult_both",
"cult_us_aux",    "cult_mex_aux",    "cult_both_aux",
"cult_na_aux", "cult_missing_aux","cult_total_aux")
    colnames(temp_cult_aux) <- x
    temp_cult_aux$cult_us <- rowSums(temp_cult == 1)
/ (rowSums(temp_cult == 1)+rowSums(temp_cult ==
2)+rowSums(temp_cult == 3))
    temp_cult_aux$cult_mex <- rowSums(temp_cult ==
2) / (rowSums(temp_cult == 1)+rowSums(temp_cult
== 2)+rowSums(temp_cult == 3))
    temp_cult_aux$cult_both <- rowSums(temp_cult ==
3) /(rowSums(temp_cult == 1)+rowSums(temp_cult ==
2)+rowSums(temp_cult == 3))
    #Auxiliar
    temp_cult_aux$cult_us_aux <- rowSums(temp_cult
== 1)
    temp_cult_aux$cult_mex_aux                 <-
rowSums(temp_cult == 2)
    temp_cult_aux$cult_both_aux                <-
rowSums(temp_cult == 3)
    temp_cult_aux$cult_missing_aux             <-
rowSums(temp_cult == 0)
    temp_cult_aux$cult_na_aux <- rowSums(temp_cult
== 4)
    temp_cult_aux$cult_total_aux               <-
(rowSums(temp_cult == 1)+rowSums(temp_cult ==
2)+rowSums(temp_cult == 3))
    temp_results_cult    <-    data.frame(temp_cult,
temp_cult_aux)
    #Append results to main dataset
    data <- data.frame(data, temp_cult_aux)

#Create a variable that merges "country_orig" and
"both_countries"
country_adjustment <- ifelse(data$country_orig == 3,
data$both_countries, data$country_orig)
  #Make sure the variable works as expected
  country_adjustment_val                     <-
data.frame(data$country_orig,    data$both_countries,
country_adjustment)
 #Convert variable into dataframe and append to the
main dataframe
  data <- data.frame(data, country_adjustment)

#Create a variable that creates groups for usual
transborder groups
data$group <- ifelse(data$country_adjustment == 1, 1,
ifelse((data$country_adjustment    ==    2    &
data$comm_work          ==          3),          2,
ifelse((data$country_adjustment    ==    2    &
data$comm_school          ==          1),          3,
ifelse((data$country_adjustment    ==    2    &
data$comm_school_else        ==        2),        3,
ifelse((data$country_adjustment    ==    2    &
data$comm_school != 1 & data$comm_school_else !=
2 & data$comm_work != 3), 4, 5)))))
 #Ensure variable was created correctly
 group_val   <-  data.frame(data$country_adjustment,
data$comm_school,          data$comm_school_else,
data$comm_work, data$comm_none, data$group)

#Create a variable for language proficiency
data$lang_prof <- ifelse(data$english_proficiency == 0
|    data$spanish_proficiency    ==    0,    0,
ifelse(data$english_proficiency        >=        1        &
data$english_proficiency          <=          3          &
data$spanish_proficiency        >=        4,        1,
ifelse(data$spanish_proficiency        >=        1        &
data$spanish_proficiency          <=          3          &
data$english_proficiency          >=          4,          2,
ifelse(data$english_proficiency        >=        4        &
data$spanish_proficiency >= 4, 3, 4))))
 #Ensure variable was created correctly
 lang_prof_val                              <-
data.frame(data$english_proficiency,
data$spanish_proficiency, data$lang_prof)
```

```
#Create a variable that creates groups for experience
living in either country
data$exp_us_aux <- ifelse(data$countries_lived == 1 |
data$countries_worked == 1 | data$countries_studied
== 1, 1, 2)
data$exp_mex_aux <- ifelse(data$countries_lived ==
2   |   data$countries_worked   ==   2   |
data$countries_studied == 2, 1, 2)
data$exp_both_aux <- ifelse(data$countries_lived ==
3   |   data$countries_worked   ==   3   |
data$countries_studied == 3, 1, 2)
data$exp_none_aux <- ifelse(data$countries_lived ==
4   &   data$countries_worked   ==   4   &
data$countries_studied == 4, 1, 2)
data$exp   <-   ifelse((data$exp_us_aux   ==   1   &
data$exp_mex_aux == 1) | data$exp_both_aux ==
1,3,ifelse(data$exp_us_aux   ==   1   &
data$exp_mex_aux == 2,1,ifelse(data$exp_us_aux ==
2 & data$exp_mex_aux == 1,2,4)))
  #Ensure variable was created correctly
  exp_val                                      <-
data.frame(data$countries_lived,data$countries_work
ed,data$countries_studied,data$exp_us_aux,data$ex
p_mex_aux,data$exp_both_aux,data$exp_none_aux,
data$exp)

#Assign labels to variables
data = apply_labels(data, country_orig = "In which
country do you live?",
            both_countries = "In which country do you
usually sleep more nights per week?",
            #country_mod = "Country of residence",
            comm_none = "Do you cross to the United
States for any of the following reasons? (select all that
apply) - None",
            comm_school = "Do you cross to the
United States for any of the following reasons? (select
all that apply) - School (for me)",
            comm_school_else = "Do you cross to the
United States for any of the following reasons? (select
all that apply) - School (take someone else)",
            comm_work = "Do you cross to the United
States for any of the following reasons? (select all that
apply) - Work",
            us_mig_status   =   "What   is   your
immigration status in the United States?",
            groc_shopping = "In which country or
countries do you typically do the following activities? -
Buy groceries",
            reunion = "In which country or countries
do you typically do the following activities? - Go to a
gathering or a party",
            fam_reunion = "In which country or
countries do you typically do the following activities? -
Go to a gathering or a party with family members",
            voting = "In which country or countries do
you typically do the following activities? - Vote during
the elections from that country",
            news = "In which country or countries do
you   typically   do   the   following   activities?   -
Read/Watch/Listen to the news about that country",

            restaurant = "In which country or countries
do you typically do the following activities? - Go out to
eat at a fancy restaurant",
            church = "In which country or countries do
you typically do the following activities? - Go to church",
            food_street   =   "In   which   country   or
countries do you typically do the following activities? -
Go out to eat on the street or fast food",
            movies = "In which country or countries do
you typically do the following activities? - Go to the
movie theater",
            fun = "In which country or countries do you
typically do the following activities? - Do something
fun",
            doctor = "In which country or countries do
you typically do the following activities? - Go to the
doctor",
            clothes = "In which country or countries do
you typically do the following activities? - Buy clothes",
            vacations = "In which country or countries
do you typically do the following activities? - Go on
vacations",
            road_trip = "In which country or countries
do you typically do the following activities? - Go on a
road trip",
            haircut = "In which country or countries do
you typically do the following activities? - Get a haircut",
            concert = "In which country or countries do
you typically do the following activities? - Go to a
concert",
            lang_sup = "What is the  language that
you typically use for the following activities? - Talking
with my work supervisor",
            lang_coworker = "What is the  language
that you typically use for the following activities? -
Talking with my coworkers",
            lang_neighbor = "What is the  language
that you typically use for the following activities? -
Talking with my neighbors",
            lang_shop = "What is the  language that
you typically use for the following activities? - Hablando
con los/las empleados(as) de una tienda",
            lang_friend = "What is the  language that
you typically use for the following activities? - Talking
with my friends",
            lang_family = "What is the  language that
you typically use for the following activities? - Talking
with my close relatives",
            lang_watch = "What is the  language that
you typically use for the following activities? - Watching
TV or videos",
            lang_listen = "What is the  language that
you typically use for the following activities? - Listening
to the radio or music",
            lang_read = "What is the  language that
you typically use for the following activities? - Reading
a book, newspaper, etc.",
            lang_children = "What is the  language
that you typically use for the following activities? -
Talking with my children",
```

lang_partner = "What is the language that you typically use for the following activities? - Talking with my partner",

culture_children = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The culture and/or traditions in which I want to raise my children",

living_place = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The place where I would rather live currently",

retirement = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The place where I would rather live once I retire",

myself = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The place where I feel I can be myself",

best_time = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The place where I have the best time",

culture_myself = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The culture or traditions I follow or that I identify with",

lifestyle = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The lifestyle I prefer to live",

pref_lang_chil = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The language I want my children to speak",

living_place_money = "Please complete the following statements indicating the country or culture more adequate based on your way of thinking. - The place where I would live if money was not a problem",

gender = "What is your gender?",

gender_other = "What is your gender? - Other (specify)",

age = "What is your age?",

educational_level = "What is your education level?",

english_proficiency = "What is your English proficiency level?",

spanish_proficiency = "What is your Spanish proficiency level?",

mexican = "Are you Mexican, Mexican American, or Chicano?",

border_freq = "How often do you cross the border?",

time_crossing = "For how long have you been crossing the border with this frequency?",

border = "What is the border that you typically use to enter/exit your country?",

non_border_city = "Do you live in a city that is more than 100 miles from the border? (example:

Phoenix AZ; Hermosillo, SON; Los Angeles, CA; Monterrey, NL; etc.)",

countries_lived = "Please select the country or countries where you have done the following activities throughout your life:
 - Country or countries where I have lived",

countries_worked = "Please select the country or countries where you have done the following activities throughout your life:
 - Country or countries where I have worked",

countries_studied = "Please select the country or countries where you have done the following activities throughout your life:
 - Country or countries where I have studied",

countries_family = "Please select the country or the countries in which the following is applicable: - Country or countries where I have family that I frequent",

countries_friends = "Please select the country or the countries in which the following is applicable: - Country or countries where I have friends that I frequent",

source = "Source of information",

exp ="Experience living in both countries")

```
#Calculate correlations between variables
 #Create a data frame that includes the variables we
are interested in analyzing
 alc                                    <-
data.frame(data$act_us,data$act_mex,data$act_both
,data$lang_us,data$lang_mex,data$lang_both,data$c
ult_us,data$cult_mex,data$cult_both)
 #Imputation of means when applicable
 for(i in 1:ncol( alc)) {
   alc[ , i][is.na( alc[ , i])] <- mean( alc[ , i], na.rm =
TRUE)
 }
 # Compute a correlation and p-values matrix
 corr <- round(cor(alc), 1)
 pvalue <- cor_pmat(alc)
 #Visualize correlation matrix
 ggcorrplot(corr, hc.order = TRUE, type = "lower",
p.mat = pvalue, lab = TRUE)

# Convert all control variables into factors
 ##Will not assign labels until all data has been
dumped to prevent assigning labels to the wrong
factor(s)
 data$country_adjustment                    <-
factor(data$country_adjustment)
 ##,labels = c("U.S.","Mexico"))
 data$group <- factor(data$group)
 ##,labels   =   c("U.S.-residing   consumer","Work
commuter",   "Study   commuter",   "Mexico-residing
consumer","Others"))
 data$gender <- factor(data$gender)
 ##,labels = c("Male","Female", "Non-binary", "I prefer
not to answer","Other"))
 data$age <- factor(data$age)
 ##,labels = c("Less than 18 years","18 - 24", "25 - 34",
"35 - 44","45 - 44", "45 - 49","50 - 54", "55 - 64","65 or
more"))
```

```r
data$educational_level                      <-
factor(data$educational_level)
 ##,labels = c("Elementary or less","Middle school",
"High school", "Less than two years in college","Two or
more years in college"))
 data$lang_prof <- factor(data$lang_prof)
 ##,labels   =    c("Spanish-proficient","English-
proficient", "Bilingual"))
 data$mexican <- factor(data$mexican)
 ##,labels = c("Mexican","Non-Mexican"))
 data$border_freq <- factor(data$border_freq)
 ##,labels = c("More than three times per week","2 - 3
times per week","1 time per week","Every 10 - 15
days","Once per month","Less than once per))
 data$time_crossing <- factor(data$time_crossing)
 ##,labels = c("Less than a year","From one to three
years","From three to five years","More than five
years"))
 data$border <- factor(data$border)
 ##,labels    =    c("Tijuana-San    Diego","Tecate-San
Diego","Mexicali-Calexico","Los        Algodones-
Andrade/Yuma","San     Luis    Rio    Colorado-San
Luis","Nogales-Nogales","Matamoros-
Brownsville","Palomas-Columbus","Ciuda  Acuña-Del
Rio","Agua   Prieta-Douglas","Piedras   Negras-Eagle
Pass","Ciudad        Juarez-El        Paso","Chihuahua-
Fabens/Tornillo","El              Porvenir"-Fort
Hancock","Reynosa-Hidalgo Pharr","Nuevo Laredo-
Laredo,"Sonoyta-Lukeville","Naco-Naco","Ojinaga-
Presidio","Ciuda      Rio     Bravo-Progreso","Ciudad
Camargo/Los  Fresnos-Rio  Grande  City","Ciudad
Miguel Alemán-Roma","San Jerónimo-Santa Teresa"))
 data$non_border_city                        <-
factor(data$non_border_city)
 ##,labels = c("Non-border city","Border city"))
 data$countries_family                       <-
factor(data$countries_family)
 ##,labels = c("Family in the U.S. only","Family  in
Mexico only", "Family in both countries", "No family"))
 data$countries_friends                      <-
factor(data$countries_friends)
 ##,labels = c("Friends in the U.S. only","Friends  in
Mexico only", "Friends in both countries", "No friends"))
 ####data$exp <- factor(data$exp)
 ##,labels = c("Friends in the U.S. only","Friends  in
Mexico only", "Friends in both countries", "No friends"))
 data$exp <- factor(data$exp)
 ##,labels  =  c("Only  experienced  the  U.S.","Only
experienced Mexico", "Inconsistent"))

    #Normalizing dataset with activities so that we can
run correlation and factor analysis with it
    #Values will be recoded in a way that 0 will continue
to be NA, 1 will mean that the respondent is doing the
activity in the place they live, 2 will mean that they are
doing it in the other country, and 3 will mean that they
are  doing  it  in  both  places  -the  ultimate  level  of
transfronteridad. It is important that while correlates,
transfronteridad is not exactly the same that the "best
from both worlds" construct when it comes to saving
money, that is a subset.

    data$groc_shopping_n                     <-
ifelse(data$groc_shopping      ==      4,     4,
ifelse(data$groc_shopping      ==      3,     3,
ifelse(data$country_adjustment               ==
data$groc_shopping,                          1,
ifelse(data$country_adjustment      ==      1     &
data$groc_shopping        ==      2,          2,
ifelse(data$country_adjustment      ==      2     &
data$groc_shopping == 1,2,0 )))))
    data$reunion_n <- ifelse(data$reunion == 4, 4,
ifelse(data$reunion           ==           3,     3,
ifelse(data$country_adjustment == data$reunion, 1,
ifelse(data$country_adjustment == 1 & data$reunion
== 2, 2, ifelse(data$country_adjustment == 2 &
data$reunion == 1,2,0 )))))
    data$fam_reunion_n  <- ifelse(data$fam_reunion
==    4,    4,   ifelse(data$fam_reunion   ==   3,   3,
ifelse(data$country_adjustment == data$fam_reunion,
1,    ifelse(data$country_adjustment     ==     1     &
data$fam_reunion        ==        2,           2,
ifelse(data$country_adjustment     ==      2      &
data$fam_reunion == 1,2,0 )))))
    data$voting_n  <-  ifelse(data$voting  ==  4,  4,
ifelse(data$voting           ==           3,     3,
ifelse(data$country_adjustment == data$voting, 1,
ifelse(data$country_adjustment == 1 & data$voting ==
2, 2, ifelse(data$country_adjustment == 2 &
data$voting == 1,2,0 )))))
    data$news_n  <-  ifelse(data$news  ==  4,  4,
ifelse(data$news           ==           3,     3,
ifelse(data$country_adjustment == data$news, 1,
ifelse(data$country_adjustment == 1 & data$news ==
2, 2, ifelse(data$country_adjustment == 2 &
data$news == 1,2,0 )))))
    data$restaurant_n <- ifelse(data$restaurant == 4,
4,    ifelse(data$restaurant     ==     3,     3,
ifelse(data$country_adjustment == data$restaurant, 1,
ifelse(data$country_adjustment      ==      1     &
data$restaurant          ==        2,          2,
ifelse(data$country_adjustment      ==      2     &
data$restaurant == 1,2,0 )))))
    data$church_n  <-  ifelse(data$church  ==  4,  4,
ifelse(data$church           ==           3,     3,
ifelse(data$country_adjustment == data$church, 1,
ifelse(data$country_adjustment == 1 & data$church
== 2, 2, ifelse(data$country_adjustment == 2 &
data$church == 1,2,0 )))))
    data$food_street_n <- ifelse(data$food_street ==
4,    4,   ifelse(data$food_street   ==   3,   3,
ifelse(data$country_adjustment == data$food_street,
1,    ifelse(data$country_adjustment     ==     1     &
data$food_street         ==        2,          2,
ifelse(data$country_adjustment      ==      2     &
data$food_street == 1,2,0 )))))
    data$movies_n  <-  ifelse(data$movies  ==  4,  4,
ifelse(data$movies           ==           3,     3,
ifelse(data$country_adjustment == data$movies, 1,
ifelse(data$country_adjustment == 1 & data$movies
== 2, 2, ifelse(data$country_adjustment == 2 &
data$movies == 1,2,0 )))))
    data$fun_n    <-    ifelse(data$fun    ==    4,    4,
ifelse(data$fun           ==           3,           3,
```

```
ifelse(data$country_adjustment == data$fun, 1,
ifelse(data$country_adjustment == 1 & data$fun == 2,
2, ifelse(data$country_adjustment == 2 & data$fun ==
1,2,0 )))))
     data$doctor_n <- ifelse(data$doctor == 4, 4,
ifelse(data$doctor         ==         3,        3,
ifelse(data$country_adjustment == data$doctor, 1,
ifelse(data$country_adjustment == 1 & data$doctor ==
2,  2,  ifelse(data$country_adjustment  ==  2  &
data$doctor == 1,2,0 )))))
     data$clothes_n <- ifelse(data$clothes == 4, 4,
ifelse(data$clothes        ==         3,        3,
ifelse(data$country_adjustment == data$clothes, 1,
ifelse(data$country_adjustment == 1 & data$clothes
== 2, 2, ifelse(data$country_adjustment == 2 &
data$clothes == 1,2,0 )))))
     data$vacations_n <- ifelse(data$vacations == 4, 4,
ifelse(data$vacations        ==         3,        3,
ifelse(data$country_adjustment == data$vacations, 1,
ifelse(data$country_adjustment == 1 & data$vacations
== 2, 2, ifelse(data$country_adjustment == 2 &
data$vacations == 1,2,0 )))))
     data$road_trip_n <- ifelse(data$road_trip == 4, 4,
ifelse(data$road_trip        ==         3,        3,
ifelse(data$country_adjustment == data$road_trip, 1,
ifelse(data$country_adjustment == 1 & data$road_trip
== 2, 2, ifelse(data$country_adjustment == 2 &
data$road_trip == 1,2,0 )))))
     data$haircut_n <- ifelse(data$haircut == 4, 4,
ifelse(data$haircut        ==         3,        3,
ifelse(data$country_adjustment == data$haircut, 1,
ifelse(data$country_adjustment == 1 & data$haircut
== 2, 2, ifelse(data$country_adjustment == 2 &
data$haircut == 1,2,0 )))))
     data$concert_n <- ifelse(data$concert == 4, 4,
ifelse(data$concert        ==         3,        3,
ifelse(data$country_adjustment == data$concert, 1,
ifelse(data$country_adjustment == 1 & data$concert
== 2, 2, ifelse(data$country_adjustment == 2 &
data$concert == 1,2,0 )))))

     write.csv(data,"C:/Users/rivas/Google
Drive/Maestría/5.    Thesis/Segmenting    Border
Users/Survey Processing/data_spss.csv", row.names
= TRUE)
```

Using SPSS to process survey results and run egressions

```
* Encoding: UTF-8.
* Encoding: .

GET DATA  /TYPE=TXT
 /FILE="C:\Users\jrivas\Desktop\Census
Data\data_spss.csv"
 /ENCODING='Locale'
 /DELCASE=LINE
 /DELIMITERS=","
 /QUALIFIER='"'
 /ARRANGEMENT=DELIMITED
 /FIRSTCASE=2
 /IMPORTCASE=ALL
 /VARIABLES=
V1 F3.0
country_orig F1.0
both_countries F1.0
comm_none F1.0
comm_school F1.0
comm_school_else F1.0
comm_work F1.0
us_mig_status F1.0
groc_shopping F1.0
reunion F1.0
fam_reunion F1.0
voting F1.0
news F1.0
restaurant F1.0
church F1.0
food_street F1.0
movies F1.0
fun F1.0
doctor F1.0
clothes F1.0
vacations F1.0
road_trip F1.0
haircut F1.0
concert F1.0
lang_sup F1.0
lang_coworker F1.0
lang_neighbor F1.0
lang_shop F1.0
lang_friend F1.0
lang_family F1.0
lang_watch F1.0
lang_listen F1.0
lang_read F1.0
lang_children F1.0
lang_partner F1.0
culture_children F1.0
living_place F1.0
retirement F1.0
myself F1.0
best_time F1.0
culture_myself F1.0
lifestyle F1.0
pref_lang_chil F1.0
living_place_money F1.0
gender F1.0
gender_other F1.0
age F1.0
educational_level F1.0
english_proficiency F1.0
spanish_proficiency F1.0
mexican F1.0
border_freq F1.0
time_crossing F1.0
border F2.0
non_border_city F1.0
countries_lived F1.0
countries_worked F1.0
countries_studied F1.0
countries_family F1.0
countries_friends F1.0
```

196

```
    source F1.0
    act_us A18
    act_mex A18
    act_both A18
    act_us_aux F2.0
    act_mex_aux F2.0
    act_both_aux F2.0
    act_na_aux F2.0
    act_missing_aux F2.0
    act_total_aux F2.0
    lang_us A17
    lang_mex A17
    lang_both A17
    lang_us_aux F1.0
    lang_mex_aux F2.0
    lang_both_aux F2.0
    lang_na_aux F1.0
    lang_missing_aux F2.0
    lang_total_aux F2.0
    cult_us A17
    cult_mex A17
    cult_both A17
    cult_us_aux F1.0
    cult_mex_aux F1.0
    cult_both_aux F1.0
    cult_na_aux F1.0
    cult_missing_aux F1.0
    cult_total_aux F1.0
    country_adjustment F1.0
    group F1.0
    lang_prof F1.0
    exp_us_aux F1.0
    exp_mex_aux F1.0
    exp_both_aux F1.0
    exp_none_aux F1.0
    exp F1.0
    groc_shopping_n F1.0
    reunion_n F1.0
    fam_reunion_n F1.0
    voting_n F1.0
    news_n F1.0
    restaurant_n F1.0
    church_n F1.0
    food_street_n F1.0
    movies_n F1.0
    fun_n F1.0
    doctor_n F1.0
    clothes_n F1.0
    vacations_n F1.0
    road_trip_n F1.0
    haircut_n F1.0
    concert_n F1.0.
CACHE.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.

COMPUTE act_us_numeric=0.
COMPUTE act_mex_numeric=0.
COMPUTE act_both_numeric=0.
COMPUTE lang_us_numeric=0.
COMPUTE lang_mex_numeric=0.
COMPUTE lang_both_numeric=0.
```

```
COMPUTE cult_us_numeric=0.
COMPUTE cult_mex_numeric=0.
COMPUTE cult_both_numeric=0.
EXECUTE.

IF (act_us = '1') act_us_numeric=1.
IF (act_mex = '1') act_mex_numeric=1.
IF (act_both = '1') act_both_numeric=1.
IF (lang_us = '1') lang_us_numeric=1.
IF (lang_mex = '1') lang_mex_numeric=1.
IF (lang_both = '1') lang_both_numeric=1.
IF (cult_us = '1') cult_us_numeric=1.
IF (cult_mex = '1') cult_mex_numeric=1.
IF (cult_both = '1') cult_both_numeric=1.
EXECUTE.

IF    (act_us   ~=   '1'   &   act_us   ~=   'NA')
act_us_numeric=NUMBER(act_us,f10.8).
IF    (act_mex   ~=   '1'   &   act_mex   ~=   'NA')
act_mex_numeric=NUMBER(act_mex,f10.8).
IF    (act_both   ~=   '1'   &   act_both   ~=   'NA')
act_both_numeric=NUMBER(act_both,f10.8).
IF    (lang_us   ~=   '1'   &   lang_us   ~=   'NA')
lang_us_numeric=NUMBER(lang_us,f10.8).
IF    (lang_mex   ~=   '1'   &   lang_mex   ~=   'NA')
lang_mex_numeric=NUMBER(lang_mex,f10.8).
IF    (lang_both   ~=   '1'   &   lang_both   ~=   'NA')
lang_both_numeric=NUMBER(lang_both,f10.8).
IF    (cult_us   ~=   '1'   &   cult_us   ~=   'NA')
cult_us_numeric=NUMBER(cult_us,f10.8).
IF    (cult_mex   ~=   '1'   &   cult_mex   ~=   'NA')
cult_mex_numeric=NUMBER(cult_mex,f10.8).
IF    (cult_both   ~=   '1'   &   cult_both   ~=   'NA')
cult_both_numeric=NUMBER(cult_both,f10.8).
EXECUTE.

IF (comm_school ~= 1 & comm_school_else ~=2 &
comm_work ~= 3) comm_none = 1.
EXECUTE.

DO IF (act_us = 'NA').
RECODE act_us_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (act_mex = 'NA').
RECODE act_mex_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (act_both = 'NA').
RECODE act_both_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (lang_us = 'NA').
RECODE lang_us_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (lang_mex = 'NA').
RECODE lang_mex_numeric (0=SYSMIS).
```

END IF.
EXECUTE.

DO IF (lang_both = 'NA').
RECODE lang_both_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (cult_us = 'NA').
RECODE cult_us_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (cult_mex = 'NA').
RECODE cult_mex_numeric (0=SYSMIS).
END IF.
EXECUTE.

DO IF (cult_both = 'NA').
RECODE cult_both_numeric (0=SYSMIS).
END IF.
EXECUTE.

COMPUTE countries_family_n=4.
COMPUTE countries_friends_n=4.
EXECUTE.

IF (countries_family = 4) countries_family_n=4.
IF (countries_family = 3) countries_family_n=3.
IF (country_adjustment = countries_family) countries_family_n=1.
IF (country_adjustment = 1 & countries_family = 2) countries_family_n=2.
IF (country_adjustment = 2 & countries_family = 1) countries_family_n=2.
IF (countries_friends = 4) countries_friends_n=4.
IF (countries_friends = 3) countries_friends_n=3.
IF (country_adjustment = countries_friends) countries_friends_n=1.
IF (country_adjustment = 1 & countries_friends = 2) countries_friends_n=2.
IF (country_adjustment = 2 & countries_friends = 1) countries_friends_n=2.
EXECUTE.

VARIABLE LABELS
country_orig 'In which country do you live? (Original Response)'
both_countries 'From the both countries in which you live, in which one do you usually sleep more nights per week?'
comm_none 'Do you cross to the United States for any of the following reasons? - None'
comm_school 'Do you cross to the United States for any of the following reasons? - School'
comm_school_else 'Do you cross to the United States for any of the following reasons? - School for someone else'
comm_work 'Do you cross to the United States for any of the following reasons? - Work'
us_mig_status 'What is your immigration status in the United States?'

groc_shopping 'In which country or countries do you typically do the following activities? (Buy groceries)'
reunion 'In which country or countries do you typically do the following activities? (Go to a gathering or a party)'
fam_reunion 'In which country or countries do you typically do the following activities? (Go to a family gathering or reunion)'
voting 'In which country or countries do you typically do the following activities? (Vote during the elections from that country)'
news 'In which country or countries do you typically do the following activities? (Read/Watch/Listen to the news about that country)'
restaurant 'In which country or countries do you typically do the following activities? (Go out to eat at a "fancy" restaurant)'
church 'In which country or countries do you typically do the following activities? (Go to church)'
food_street 'In which country or countries do you typically do the following activities? (Go out to eat "on the street" or fast food)'
movies 'In which country or countries do you typically do the following activities? (Go to the movie theater)'
fun 'In which country or countries do you typically do the following activities? (Do something fun)'
doctor 'In which country or countries do you typically do the following activities? (Go to the doctor)'
clothes 'In which country or countries do you typically do the following activities? (Buy clothes)'
vacations 'In which country or countries do you typically do the following activities? (Go on vacations)'
road_trip 'In which country or countries do you typically do the following activities? (Go on a road trip)'
haircut 'In which country or countries do you typically do the following activities? (Get a haircut)'
concert 'In which country or countries do you typically do the following activities? (Go to a concert)'
lang_sup 'What is the language that you typically use for the following activities? (Talking with my work supervisor)'
lang_coworker 'What is the language that you typically use for the following activities? (Talking with my coworkers)'
lang_neighbor 'What is the language that you typically use for the following activities? (Talking with my neighbors)'
lang_friend 'What is the language that you typically use for the following activities? (Talking with my friends)'
lang_family 'What is the language that you typically use for the following activities? (Talking with my close relatives)'
lang_watch 'What is the language that you typically use for the following activities? (Watching TV or videos)'
lang_listen 'What is the language that you typically use for the following activities? (Listening to the radio or music)'
lang_read 'What is the language that you typically use for the following activities? (Reading a book, newspaper, etc.)'

lang_children 'What is the language that you typically use for the following activities? (Talking with my children)'
lang_partner 'What is the language that you typically use for the following activities? (Talking with my partner)'
culture_children 'The culture and/or traditions in which I want to raise my children'
living_place 'The place where I would rather live currently'
retirement 'The place where I would rather live once I retire'
myself 'The place where I feel I can be myself'
best_time 'The place where I have the best time'
culture_myself 'The culture or traditions I follow or that I identify with'
lifestyle 'The lifestyle I prefer to live'
pref_lang_chil 'The language I want my children to speak'
living_place_money 'The place where I would live if money was not a problem'
gender 'What is your gender?'
gender_other 'Other gender (specifiy)'
age 'What is your age?'
educational_level 'What is your education level?'
english_proficiency 'What is your English proficiency level?'
spanish_proficiency 'What is your Spanish proficiency level?'
mexican 'Are you Mexican, Mexican American, or Chicano?'
border_freq 'How often do you cross the border?'
time_crossing 'For how long have you been crossing the border with this frequency?'
border 'What is the border that you typically use to enter/exit your country?'
non_border_city 'Do you live in a city that is more than 100 miles from the border?'
countries_lived 'Please select the country or countries where you have lived:'
countries_worked 'Please select the country or countries where you have worked:'
countries_studied 'Please select the country or countries where you have studied:'
countries_family 'Country or countries where I have family that I frequent:'
countries_friends 'Country or countries where I have friends that I frequent:'
source 'Data Source'
country_adjustment 'Country of Residence (adjusted)'
group 'Transborder Group'
lang_prof 'Language Proficiency'
exp 'Exposure'
groc_shopping_n 'Buy groceries (recoded)'
reunion_n 'Go to a gathering or a party (recoded)'
voting_n 'Vote during the elections from that country (recoded)'
news_n 'Read/Watch/Listen to the news about that country (recoded)'
restaurant_n 'Go out to eat at a "fancy" restaurant (recoded)'
church_n 'Go to church (recoded)'

food_street_n 'Go out to eat "on the street" or fast food (recoded)'
movies_n 'Go to the movie theater (recoded)'
fun_n 'Do something fun (recoded)'
doctor_n 'Go to the doctor (recoded)'
clothes_n 'Buy clothes (recoded)'
vacations_n 'Go on vacations (recoded)'
road_trip_n 'Go on a road trip (recoded)'
haircut_n 'Get a haircut (recoded)'
concert_n 'Go to a concert (recoded)'
act_us_numeric 'Percentage of activities done in the U.S. only'
act_mex_numeric 'Percentage of activities done in Mexico only'
act_both_numeric 'Percentage of activities done in both countries'
lang_us_numeric 'Percentage of situations in which English is used'
lang_mex_numeric 'Percentage of situations in which Spanish is used'
lang_both_numeric 'Percentage of situations in which both languages are used'
cult_us_numeric 'Percentage of statements matched to the American culture only'
cult_mex_numeric 'Percentage of statements matched to the Mexican culture only'
cult_both_numeric 'Percentage of statement matched to both cultures'.

DELETE VARIABLES
fam_reunion
lang_shop
act_us
act_mex
act_both
act_us_aux
act_mex_aux
act_both_aux
act_na_aux
act_missing_aux
act_total_aux
lang_us
lang_mex
lang_both
lang_us_aux
lang_mex_aux
lang_both_aux
lang_na_aux
lang_missing_aux
lang_total_aux
cult_us
cult_mex
cult_both
cult_us_aux
cult_mex_aux
cult_both_aux
cult_na_aux
cult_missing_aux
cult_total_aux
exp_us_aux
exp_mex_aux
exp_both_aux

exp_none_aux
fam_reunion_n.

VALUE LABELS
country_orig both_countries 1 'United States' 2 'Mexico' 3 'In both countries'.
EXECUTE.

VALUE LABELS
comm_none 1 'None'.
EXECUTE.

VALUE LABELS
comm_school 1 'School'.
EXECUTE.

VALUE LABELS
comm_school_else 2 'School for someone else'.
EXECUTE.

VALUE LABELS
comm_work 3 'Work'.
EXECUTE.

VALUE LABELS
us_mig_status
1 'I do not have a visa nor any other document that would allow me to enter the United States'
2 'I have a tourist visa'
3 'I have a work visa or an employment authorization document (EAD)'
4 'I am a United States resident (green-card holder)'
5 'I am a United States citizen'.
EXECUTE.

VALUE LABELS
groc_shopping reunion voting news restaurant church food_street movies fun doctor clothes vacations road_trip haircut concert 1 'United States' 2 'Mexico' 3 'In both countries' 4 'I do not do that activity'.
EXECUTE.

VALUE LABELS
lang_sup lang_coworker lang_neighbor lang_friend lang_family lang_watch lang_listen lang_read lang_children lang_partner 1 'English' 2 'Spanish' 3 'Both languages' 4 'I do not do that activity'.
EXECUTE.

VALUE LABELS
culture_children living_place retirement myself best_time culture_myself lifestyle pref_lang_chil living_place_money 1 'United States/American' 2 'Mexico/Mexican' 3 'Both' 4 'None/Not applicable'.
EXECUTE.

VALUE LABELS
gender 1 'Male' 2 'Female' 3 'Non-binary' 4 'I prefer not to answer' 5 'Other'.
EXECUTE.

VALUE LABELS

age
1 ' Less than 18 years'
2 '18 - 24'
3 '25 - 34'
4 '35 - 44'
5 '45 - 49'
6 '50 - 54'
7 '55 - 64'
8 '65 or more'.
EXECUTE.

VALUE LABELS
educational_level
1 'Elementary or less'
2 'Middle school'
3 'High school'
4 'Less than two years in college'
5 'Two or more years in college'.
EXECUTE.

VALUE LABELS
english_proficiency
1 'I do not know English'
2 'I understand English when I listen to it, but I cannot speak it'
3 'I speak English a little bit'
4 'I speak English well'
5 'I speak English very well'.
EXECUTE.

VALUE LABELS
spanish_proficiency
1 'I do not know Spanish'
2 'I understand Spanish when I listen to it, but I cannot speak it'
3 'I speak Spanish a little bit'
4 'I speak Spanish well'
5 'I speak Spanish very well'.
EXECUTE.

VALUE LABELS
mexican non_border_city 1 'Yes' 2 'No'.
EXECUTE.

VALUE LABELS
border_freq
1 'More than three times per week'
2 '2 - 3 times per week'
3 ' 1 time per week'
4 'Every 10-15 days'
5 'Once per month'
6 'Less than once per month'.
EXECUTE.

VALUE LABELS
time_crossing
1 'Less than a year'
2 'From one to three years'
3 'From three to five years'
4 'More than five years'.
EXECUTE.

VALUE LABELS
border
1 'Tijuana – San Diego'
2 'Tecate – San Diego'
3 'Mexicali – Calexico'
4 'Los Algodones – Andrade/Yuma'
5 'San Luis Río Colorado – San Luis'
6 'Nogales - Nogales'
7 'Matamoros – Brownsville'
8 'Palomas - Columbus'
9 'Ciudad Acuña - Del Rio'
10 'Agua Prieta - Douglas'
11 'Piedras Negras – Eagle Pass'
12 'Ciudad Juarez – El Paso'
13 'Chihuahua – Fabens/Tornillo'
14 'El Porvenir – Fort Hancock'
15 'Reynosa – Higalgo Pharr'
16 'Nuevo Laredo – Laredo'
17 'Sonoyta - Lukeville'
18 'Naco - Naco'
19 'Ojinaga - Presidio'
20 'Ciudad Río Bravo - Progreso'
21 'Ciudad Camargo/Los Fresnos – Rio Grande City'
22 'Ciudad Miguel Alemán - Roma'
23 'San Jerónimo – Santa Teresa'.
EXECUTE.

VALUE LABELS
countries_lived countries_worked countries_studied
countries_family countries_friends 1 'United States' 2
'Mexico' 3 'In both countries' 4 'None'.
EXECUTE.

VALUE LABELS
source 1 'Spanish version' 2 'English version'.
EXECUTE.

VALUE LABELS
country_adjustment 1 'United States' 2 'Mexico'.
EXECUTE.

VALUE LABELS
group 1 'American Consumer' 2 'Work Commuter' 3
'School Commuter' 4 'Mexican Consumer' 5 'Other'.
EXECUTE.

VALUE LABELS
lang_prof 1 'Proficient in Spanish' 2 'Proficient in
English' 3 'Bilingual' 4 'Other'.
EXECUTE.

VALUE LABELS
exp 1 'Experience only in the U.S.' 2 'Experiences in
Mexico' 3 'Experiences in both countries' 4 'Other'.
EXECUTE.

VALUE LABELS
groc_shopping_n reunion_n voting_n news_n
restaurant_n church_n food_street_n movies_n
countries_family_n countries_friends_n
fun_n doctor_n clothes_n vacations_n road_trip_n
haircut_n concert_n 1 'Only in their country of
residence' 2 'Only in the country they do not reside' 3
'In both countries' 4 'They do not do the activity'.
EXECUTE.

COUNT act_count_1=groc_shopping_n reunion_n
voting_n news_n restaurant_n church_n food_street_n
    movies_n fun_n doctor_n clothes_n vacations_n
road_trip_n haircut_n concert_n countries_family_n
    countries_friends_n(1).
VARIABLE LABELS act_count_1 'US_act'.
EXECUTE.

COUNT act_count_2=groc_shopping_n reunion_n
voting_n news_n restaurant_n church_n food_street_n
    movies_n fun_n doctor_n clothes_n vacations_n
road_trip_n haircut_n concert_n countries_family_n
    countries_friends_n(2).
VARIABLE LABELS act_count_2 'Mex_act'.
EXECUTE.

COUNT act_count_3=groc_shopping_n reunion_n
voting_n news_n restaurant_n church_n food_street_n
    movies_n fun_n doctor_n clothes_n vacations_n
road_trip_n haircut_n concert_n countries_family_n
    countries_friends_n(3).
VARIABLE LABELS act_count_3 'Transborder
activities'.
EXECUTE.

COMPUTE act_perc = act_count_3 / (act_count_1 +
act_count_2 + act_count_3).
VARIABLE LABELS act_perc 'Percentage for
Activities'.
EXECUTE.

RECODE act_perc (MISSING=SYSMIS) (0 thru .2=1)
(.2 thru .4=2) (.4 thru .6=3) (.6
    thru .8=4) (.8 thru 1=5) INTO act_perc_r.
EXECUTE.

COUNT cult_count_1=culture_children living_place
retirement myself best_time culture_myself
    lifestyle pref_lang_chil living_place_money(1).
VARIABLE LABELS cult_count_1 'US_Cult'.
EXECUTE.

COUNT cult_count_2=culture_children living_place
retirement myself best_time culture_myself
    lifestyle pref_lang_chil living_place_money(2).
VARIABLE LABELS cult_count_2 'Mex_Cult'.
EXECUTE.

COUNT cult_count_3=culture_children living_place
retirement myself best_time culture_myself
    lifestyle pref_lang_chil living_place_money(3).
VARIABLE LABELS cult_count_3 'Multiculturality'.
EXECUTE.

COMPUTE multiculturality_perc = cult_count_3 /
(cult_count_1 + cult_count_2 + cult_count_3).
VARIABLE LABELS multiculturality_perc 'Percentage
for Multiculturality'.

EXECUTE.

RECODE multiculturality_perc (MISSING=SYSMIS) (0 thru .2=1) (.2 thru .4=2) (.4 thru .6=3) (.6 thru .8=4) (.8 thru 1=5) INTO multiculturality_perc_r. EXECUTE.

VALUE LABELS
act_perc_r multiculturality_perc_r
1 '0% - 20%'
2 '20% - 40%'
3 '40% - 60%'
4 '60% - 80%'
5 '80% - 100%'.
EXECUTE.

RECODE lang_prof (0=SYSMIS) (1 thru 2=0) (3=1) INTO lang_prof_r.
VARIABLE LABELS lang_prof_r 'Language Proficiency (dummy)'.
EXECUTE.

RECODE border (1=1) (2=1) (3=1) (4=2) (5=2) (6=2) (7=3) (8=3) (9=3) (10=2) (11=3) (12 thru 16=3) (17 thru 18=2) (19 thru 23=3) INTO border_r.
EXECUTE.

RECODE educational_level (0=SYSMIS).
RECODE both_countries (0=SYSMIS).
RECODE comm_none (0=SYSMIS).
RECODE comm_school (0=SYSMIS).
RECODE comm_school_else (0=SYSMIS).
RECODE comm_work (0=SYSMIS).
RECODE groc_shopping (0=SYSMIS).
RECODE   reunion   (0=SYSMIS).
RECODE   voting   (0=SYSMIS).
RECODE   news   (0=SYSMIS).
RECODE   restaurant   (0=SYSMIS).
RECODE   church   (0=SYSMIS).
RECODE   food_street   (0=SYSMIS).
RECODE   movies   (0=SYSMIS).
RECODE   fun   (0=SYSMIS).
RECODE   doctor   (0=SYSMIS).
RECODE   clothes   (0=SYSMIS).
RECODE   vacations   (0=SYSMIS).
RECODE   road_trip (0=SYSMIS).
RECODE   haircut   (0=SYSMIS).
RECODE   concert   (0=SYSMIS).
RECODE   lang_sup   (0=SYSMIS).
RECODE   lang_coworker   (0=SYSMIS).
RECODE   lang_neighbor   (0=SYSMIS).
RECODE   lang_friend   (0=SYSMIS).
RECODE   lang_family   (0=SYSMIS).
RECODE   lang_watch   (0=SYSMIS).
RECODE   lang_listen   (0=SYSMIS).
RECODE   lang_read   (0=SYSMIS).
RECODE   lang_children   (0=SYSMIS).
RECODE   lang_partner   (0=SYSMIS).
RECODE   culture_children (0=SYSMIS).
RECODE   living_place   (0=SYSMIS).
RECODE   retirement   (0=SYSMIS).
RECODE   myself   (0=SYSMIS).

RECODE   best_time   (0=SYSMIS).
RECODE   culture_myself   (0=SYSMIS).
RECODE   lifestyle (0=SYSMIS).
RECODE   pref_lang_chil   (0=SYSMIS).
RECODE   living_place_money
(0=SYSMIS).
RECODE   gender   (0=SYSMIS).
RECODE   gender_other   (0=SYSMIS).
RECODE   age   (0=SYSMIS).
RECODE   english_proficiency
(0=SYSMIS).
RECODE   spanish_proficiency
(0=SYSMIS).
RECODE   mexican (0=SYSMIS).
RECODE   border_freq   (0=SYSMIS).
RECODE   time_crossing   (0=SYSMIS).
RECODE   border   (0=SYSMIS).
RECODE   non_border_city (0=SYSMIS).
RECODE   countries_lived (0=SYSMIS).
RECODE   countries_worked (0=SYSMIS).
RECODE   countries_studied (0=SYSMIS).
RECODE   countries_family (0=SYSMIS).
RECODE   countries_friends (0=SYSMIS).
RECODE   source   (0=SYSMIS).
RECODE   country_adjustment
(0=SYSMIS).
RECODE   group   (0=SYSMIS).
RECODE   lang_prof   (0=SYSMIS).
RECODE   exp   (0=SYSMIS).
RECODE   groc_shopping_n (0=SYSMIS).
RECODE   reunion_n   (0=SYSMIS).
RECODE   voting_n (0=SYSMIS).
RECODE   news_n   (0=SYSMIS).
RECODE   restaurant_n   (0=SYSMIS).
RECODE   church_n   (0=SYSMIS).
RECODE   food_street_n   (0=SYSMIS).
RECODE   movies_n   (0=SYSMIS).
RECODE   fun_n   (0=SYSMIS).
RECODE   doctor_n (0=SYSMIS).
RECODE   clothes_n   (0=SYSMIS).
RECODE   vacations_n   (0=SYSMIS).
RECODE   road_trip_n   (0=SYSMIS).
RECODE   haircut_n   (0=SYSMIS).
RECODE   concert_n   (0=SYSMIS).
RECODE   countries_family_n
(0=SYSMIS).
RECODE   countries_friends_n
(0=SYSMIS).
EXECUTE.

SAVE   OUTFILE='C:\Users\jrivas\Desktop\Census Data\data_spss.sav'
 /COMPRESSED.

DATASET ACTIVATE DataSet1.
FREQUENCIES   VARIABLES=   source country_adjustment us_mig_status group country_orig both_countries comm_none comm_school comm_school_else comm_work
   groc_shopping reunion voting news restaurant church food_street movies

fun doctor clothes vacations road_trip haircut concert countries_family countries_friends
   culture_children living_place retirement myself best_time culture_myself lifestyle pref_lang_chil
   living_place_money gender gender_other age educational_level english_proficiency
   spanish_proficiency mexican border_freq time_crossing border non_border_city countries_lived
   countries_worked countries_studied groc_shopping_n
   reunion_n voting_n news_n restaurant_n church_n food_street_n movies_n fun_n doctor_n
   clothes_n vacations_n road_trip_n haircut_n concert_n countries_family_n countries_friends_n act_perc_r multiculturality_perc_r
   /ORDER=ANALYSIS.

DESCRIPTIVES VARIABLES=act_perc multiculturality_perc
   /STATISTICS=MEAN STDDEV MIN MAX.

CROSSTABS
   /TABLES=groc_shopping reunion voting news restaurant church food_street movies fun doctor clothes
   vacations road_trip haircut concert countries_family countries_friends groc_shopping_n reunion_n
   voting_n news_n restaurant_n church_n food_street_n movies_n fun_n doctor_n clothes_n vacations_n
   road_trip_n haircut_n countries_family_n countries_friends_n BY group
   /FORMAT=AVALUE TABLES
   /CELLS=COUNT COLUMN
   /COUNT ROUND CELL.

DATASET ACTIVATE DataSet1.
* Custom Tables.
CTABLES
   /VLABELS VARIABLES=act_perc multiculturality_perc group DISPLAY=LABEL
   /TABLE act_perc [MEAN, MEDIAN] + multiculturality_perc [MEAN, MEDIAN] BY group
   /CATEGORIES VARIABLES=group ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES POSITION=AFTER.

* Custom Tables.
CTABLES
   /VLABELS VARIABLES=group gender age educational_level country_adjustment countries_lived
   countries_worked countries_studied border_freq time_crossing non_border_city
   multiculturality_perc_r lang_prof us_mig_status mexican act_perc
   DISPLAY=LABEL
   /TABLE group [C] + gender [C] + age [C] + educational_level [C] + country_adjustment [C] +
   countries_lived [C] + countries_worked [C] + countries_studied [C] + border_freq [C] +
   time_crossing [C] + non_border_city [C] + multiculturality_perc_r [C] + lang_prof [C] +

   us_mig_status [C] + mexican [C] BY act_perc [S][COUNT F40.0, MEAN]
   /CATEGORIES VARIABLES=group ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES POSITION=AFTER
   /CATEGORIES VARIABLES=gender age educational_level country_adjustment countries_lived
   countries_worked countries_studied border_freq time_crossing non_border_city
   multiculturality_perc_r lang_prof us_mig_status mexican ORDER=A KEY=VALUE EMPTY=INCLUDE.

REGRESSION
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT act_perc
   /METHOD=ENTER multiculturality_perc lang_prof_r educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
   /SELECT=country_adjustment EQ 1
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT act_perc
   /METHOD=ENTER multiculturality_perc lang_prof_r educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
   /SELECT=country_adjustment EQ 2
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT act_perc
   /METHOD=ENTER multiculturality_perc lang_prof_r educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
   /SELECT=us_mig_status EQ 2
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT act_perc
   /METHOD=ENTER multiculturality_perc lang_prof_r educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
   /SELECT=us_mig_status EQ 4
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT act_perc

203

```
   /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=us_mig_status EQ 5
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=group EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.


DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=group EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=group EQ 4
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=lang_prof EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=lang_prof EQ 3
```

```
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=mexican EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=mexican EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=gender EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=gender EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.
```

```
DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 4
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 5
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 6
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 7
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=age EQ 8
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)

 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=educational_level EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=educational_level EQ 4
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=educational_level EQ 5
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_lived EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_lived EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.


DATASET ACTIVATE DataSet1.
```

REGRESSION
 /SELECT=countries_lived EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_worked EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_worked EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_worked EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_studied EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_studied EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc

 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=countries_studied EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 4
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 5
 /MISSING LISTWISE

```
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_freq EQ 6
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=time_crossing EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=time_crossing EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=time_crossing EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=time_crossing EQ 4
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=non_border_city EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=non_border_city EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_r EQ 1
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_r EQ 2
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.

DATASET ACTIVATE DataSet1.
REGRESSION
 /SELECT=border_r EQ 3
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT act_perc
 /METHOD=ENTER multiculturality_perc lang_prof_r
educational_level.
```

Using R Studio to run(Machine Learning Supervised Algorithms

```
# Load libraries
library(psych)
```

```
library(gmodels)
library(expss)
library(dplyr)
library(ggcorrplot)
library(ltm)
library(mirt)
library(plyr)
library(readr)
library(caret)
library(ggplot2)
library(repr)
library(glmnet)
library(caTools)
library(e1071)
library(ROCR)
library(randomForest)

# Establish directory
setwd("C:/Users/rivas/Google          Drive/Maestría/5.
Thesis/Segmenting Border Users/Survey Processing")

#Import results
data_spss_mlearning                                   <-
read.csv("data_spss_mlearning.csv", header = TRUE,
dec = ",", sep = ",")

#Keep only the variables required for models
data_spss_mlearning <- subset(data_spss_mlearning,
select                                                =
c(act_perc,multiculturality_perc,lang_prof_r,education
al_level,country_adjustment,us_mig_status,group,me
xican,gender,age,countries_lived,countries_worked,c
ountries_studied,border_freq,time_crossing,non_bord
er_city,border_r))

#Convert variables into numeric
var_names<-names(data_spss_mlearning)
for(var in var_names){
  data_spss_mlearning[[var]]                          <-
as.numeric(as.character(data_spss_mlearning[[var]]))
}

#Correct NA's created after change in variable type
for(var in var_names){

data_spss_mlearning[[var]][is.na(data_spss_mlearnin
g[[var]])]<-0
}

#Partition data
set.seed(100)
index=sample(1:nrow(data_spss_mlearning),0.7*nrow
(data_spss_mlearning))
train = data_spss_mlearning[index,]
test = data_spss_mlearning[-index,]
dim(train)
dim(test)

#Standardizing variables
cols                                                  =
c('multiculturality_perc','lang_prof_r','educational_level
','country_adjustment','us_mig_status','group','mexica
```

```
n','gender','age','countries_lived','countries_worked','c
ountries_studied','border_freq','time_crossing','non_bo
rder_city','border_r')
pre_proc_val                                          <-
preProcess(train[,cols],method=c("center","scale"))
train[,cols]=predict(pre_proc_val,train[,cols])
test[,cols]=predict(pre_proc_val,test[,cols])
summary(train)

#Creating a linear regression model
lr = lm(act_perc ~ multiculturality_perc + lang_prof_r +
educational_level     +     country_adjustment     +
us_mig_status + group + mexican + gender + age +
countries_lived      +      countries_worked      +
countries_studied + border_freq + time_crossing +
non_border_city + border_r, data = train)
summary(lr)

#Creating evaluation metrics
  #Step 1 - create the evaluation metrics function

eval_metrics = function(model, df, predictions, target){
  resids = df[,target] - predictions
  resids2 = resids**2
  N = length(predictions)
  r2 = as.character(round(summary(model)$r.squared,
2))
  adj_r2                                              =
as.character(round(summary(model)$adj.r.squared,
2))
  print(adj_r2) #Adjusted R-squared
  print(as.character(round(sqrt(sum(resids2)/N),  2)))
#RMSE
}

  # Step 2 - predicting and evaluating the model on train
data
predictions = predict(lr, newdata = train)
eval_metrics(lr, train, predictions, target = 'act_perc')

  # Step 3 - predicting and evaluating the model on test
data
predictions = predict(lr, newdata = test)
eval_metrics(lr, test, predictions, target = 'act_perc')

#Regularize coeficients needed for Ridge regresion
cols_reg                                              =
c('act_perc','multiculturality_perc','lang_prof_r','educati
onal_level','country_adjustment','us_mig_status','grou
p','mexican','gender','age','countries_lived','countries_
worked','countries_studied','border_freq','time_crossin
g','non_border_city','border_r')

dummies <- dummyVars(act_perc  ~  .,  data  =
data_spss_mlearning[,cols_reg])

train_dummies  =  predict(dummies,  newdata  =
train[,cols_reg])

test_dummies  =  predict(dummies,  newdata  =
test[,cols_reg])
```

```
print(dim(train_dummies)); print(dim(test_dummies))

#Creating a Ridge regression model
x = as.matrix(train_dummies)
y_train = train$act_perc

x_test = as.matrix(test_dummies)
y_test = test$act_perc

lambdas <- 10^seq(2, -3, by = -.1)
ridge_reg = glmnet(x, y_train, nlambda = 25, alpha = 0,
family = 'gaussian', lambda = lambdas)

summary(ridge_reg)
cv_ridge <- cv.glmnet(x, y_train, alpha = 0, lambda =
lambdas)
optimal_lambda <- cv_ridge$lambda.min
optimal_lambda

# Compute R^2 from true and predicted values
eval_results <- function(true, predicted, df) {
  SSE <- sum((predicted - true)^2)
  SST <- sum((true - mean(true))^2)
  R_square <- 1 - SSE / SST
  RMSE = sqrt(SSE/nrow(df))


  # Model performance metrics
  data.frame(
    RMSE = RMSE,
    Rsquare = R_square
  )

}

# Prediction and evaluation on train data
predictions_train    <-    predict(ridge_reg,    s    =
optimal_lambda, newx = x)
eval_results(y_train, predictions_train, train)

# Prediction and evaluation on test data
predictions_test    <-    predict(ridge_reg,    s    =
optimal_lambda, newx = x_test)
eval_results(y_test, predictions_test, test)

# Run a LASSO Regression
lambdas <- 10^seq(2, -3, by = -.1)

# Setting alpha = 1 implements lasso regression
lasso_reg <- cv.glmnet(x, y_train, alpha = 1, lambda =
lambdas, standardize = TRUE, nfolds = 5)

# Best
lambda_best <- lasso_reg$lambda.min
lambda_best

lasso_model <- glmnet(x, y_train, alpha = 1, lambda =
lambda_best, standardize = TRUE)

predictions_train    <-    predict(lasso_model,    s    =
lambda_best, newx = x)
eval_results(y_train, predictions_train, train)
```

```
predictions_test    <-    predict(lasso_model,    s    =
lambda_best, newx = x_test)
eval_results(y_test, predictions_test, test)

#Run an Elastic Net Regression
# Set training control
train_cont <- trainControl(method = "repeatedcv",
                number = 10,
                repeats = 5,
                search = "random",
                verboseIter = TRUE)

# Train the model
elastic_reg <- train(act_perc ~ .,
            data = train,
            method = "glmnet",
            preProcess = c("center", "scale"),
            tuneLength = 10,
            trControl = train_cont)


# Best tuning parameter
elastic_reg$bestTune

# Make predictions on training set
predictions_train <- predict(elastic_reg, x)
eval_results(y_train, predictions_train, train)

# Make predictions on test set
predictions_test <- predict(elastic_reg, x_test)
eval_results(y_test, predictions_test, test)
```