THE POTENTIAL ROLE AND APPLICATION OF EMERGING DATA IN A BORDER-CROSSING CONTEXT

by

Okan Gurbuz, Erik Vargas, Ipek N. Sener, and Rafael M. Aldrete

Project performed by
Center for International Intelligent Transportation Research

185922-00010
The Potential Role and Application of Emerging Data in a Border-Crossing Context

October 2022

Report prepared by
Center for International Intelligent Transportation Research
Texas A&M Transportation Institute
4050 Rio Bravo, Suite 212
El Paso, Texas 79902

TEXAS A&M TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

DISCLAIMER AND ACKNOWLEDGMENTS

This research was performed by the Center for International Intelligent Transportation Research, a part of the Texas A&M Transportation Institute (TTI). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein.

The authors would like to thank TTI researcher David Salgado for his help in the early stage of INRIX data compilation. In addition, the authors would like to thank the Texas Department of Transportation and CIITR for providing free access to INRIX data for the time periods of January 20–March 19, 2020, and October–November 2019. The authors would also like to acknowledge SafeGraph Inc. for providing free access to Texas point-of-interest data during the execution of this study. Finally, the authors thank Dawn Herring for her editorial review of this report.

CHAPTER 1: INTRODUCTION

BACKGROUND

Millions of people cross the international border between the United States and Mexico, making it one of the busiest crossings in the world. Approximately $1.5 billion in exports flows through the border each day, and trade with Mexico supports over 6 million jobs in the United States (Figueroa et al., 2012). With 26 major land ports of entries (LPOEs) along the U.S.-Mexico border, the state of Texas shares the longest border with Mexico, and most of the crossings allow commercial, vehicular, and pedestrian crossings. According to the Bureau of Transportation Statistics (BTS), before the COVID-19 pandemic and its restrictions were in place, the daily average of people crossing into Texas was reported to be nearly 245,000, with nearly 90 million people crossing into the United States through Texas in 2019 (BTS, 2022). Although border crossings dropped significantly in response to pandemic-related temporary border restrictions, the number of crossings increased again in 2021 and were reported as nearly 60 million (BTS, 2022).

The El Paso–Ciudad Juárez international border serves as the busiest port of entry (POE) in Texas and the second busiest POE among the borders between the United States and Mexico. Each year, millions of people cross the border from Juárez to El Paso for various trip purposes, including work or education, shopping, health services, and visiting family and friends (Vargas et al., 2021; Vargas et al., 2022. According to BTS (2022), in 2019, almost 30 million people and 12 million vehicles crossed the border from Juárez to El Paso. Even under border restrictions in 2020, the total number of people crossed the border was around 12 million (BTS, 2022).

Despite being a critical business and leisure destination and offering direct access to interstates, intermodal facilities, and a large population (Sener et al., 2015), transportation agencies in the El Paso region have relatively limited funding to deliver the transportation improvements needed to support economic development and thus enhance the quality of life for residents. Local transportation authorities have difficulty justifying requests for transportation funding that reflects the overall usage of the transportation network due to the limited data on cross-border trips. As noted in a study by Zmud and Sener (2019), information on business and leisure travelers is not typically captured in travel demand studies, and mobility solutions have usually been "based on the needs of people who live and work in cities, without much thought to those who travel there." In addition to time and budget limitations, for binational metropolitan regions, additional challenges exist due to the involvement of multiple and international jurisdictions, and the data collected through scheduled surveys can only cover limited information on cross-border trips (Vargas et al., 2021).

Recent developments in technology and the mass adoption of smart mobile devices have provided an opportunity to passively record and track people's travel behaviors. Anonymized travel data are increasing and, at the same time, becoming cheaper to purchase. These data sources allow transportation professionals to gain insight into travel behavior and the mobility of people by providing valuable information about trips in a selected region. While transportation professionals generally view emerging data sources as a supplement to traditional survey data

(Lee et al., 2016), research has shown that these sources can be used for various purposes for different modes (Lee and Sener, 2020). Various studies have explored the use of emerging data in context; however, only a few research attempts have focused on the use of such data sources for cross-border mobility analysis (Vargas et al., 2021; Vargas et al., 2022).

STUDY OBJECTIVES

In recognition of the potential of crowdsourced data with detailed spatiotemporal information, this study aimed to perform a deep-dive analysis exploring northbound cross-border trips using location-based data sources. The research was built on a previous project by the research team (Vargas et al., 2021) that analyzed three months of cross-border INRIX data (INRIX, n.d.) for the time period between January 20 and March 19, 2020. In this follow-up project, researchers expanded the study by first examining different periods of INRIX data (October–November 2019), which allowed the researchers to demonstrate a typical traffic flow between the El Paso–Ciudad Juárez border cities. In other words, the research team explored four months of pre-pandemic data (from two different periods: October–November 2019 and January 20–March 19, 2020) and one month of data (March 20–April 19, 2020) that captured the initial period of border restrictions. The results of this additional analysis of INRIX data are available in Vargas et al. (2022).

While aimed to be complementary to the prior research of the authors, the primary objective of this current study was to explore the potential role and application of location-based services (LBS) data. This report is thus devoted to the discussion and analysis of the LBS data in a border-crossing context as another potential data source through a specific case using SafeGraph data. LBS data are cost-effective and considered an alternative way of collecting data, especially to understand visitation patterns. SafeGraph is a data company that provides point-of-interest (POI) data based on business listing and location information collected in partnership with third-party data providers such as mobile application developers (SafeGraph, n.d.). The study utilized the data corresponding to the period of January 2018 to December 2021 (four years) and focused on selected cities in Texas.

The remainder of this report includes four chapters:

- Chapter 2 positions the study by reviewing the related works on cross-border mobility and the use of POI data.
- Chapter 3 describes the methodological details of the research study, including data attributes and limitations, data collection, and processing.
- Chapter 4 provides the results of the analysis conducted, including findings related to cross-border trips at the selected locations.
- Chapter 5 concludes the report with a summary and final remarks, highlighting the contributions and limitations and suggesting several possible research directions.

CHAPTER 2:
CROSS-BORDER MOBILITY AND POINT-OF-INTEREST DATA

CROSS-BORDER MOBILITY BETWEEN THE UNITED STATES AND MEXICO

The United States and Mexico share a long history that includes social, cultural, and economic relations. The 1994 North American Free Trade Agreement set rules to stabilize commercial trade and help catalyze the growth of the United States and Mexico. Urbanized Texas border communities experience millions of monthly border crossings. The Paso del Norte (PdN) region, for example, is one of the most highly populated areas along the U.S.-Mexico border. The region, which includes the city of El Paso in Texas, the city of Las Cruces in New Mexico, and Ciudad Juárez in Chihuahua, Mexico, has a population close to 2.5 million (U.S. Census Bureau, 2020; INEGI, 2020). Thousands of residents commute frequently across the international border in the region. Located at the crossroads and the midpoint of the U.S.-Mexico border, the PdN region offers one of the most efficient trade routes from east to west (Houston to Los Angeles) as well as south to north (Mexico to Canada).

Residents travel between Mexico and the United States for various purposes. For instance, Sener et al. (2015) conducted a synthesis of cross-border travel decision analyses focused on the El Paso region and reported numerous reasons for border crossings, including education and work purposes, shopping activities, and visiting family and friends.

Figure 1 demonstrates the monthly crossings from Mexico to the United States at the Texas LPOEs from January 2018 to July 2022 (BTS, 2022). Total monthly crossings were calculated by adding the numbers of pedestrian, bus, train, and personal vehicle passenger crossings. This total was added to the number of trucks that crossed; it was assumed that one person crossed with each truck crossing. Northbound crossings were also calculated for the PdN region.

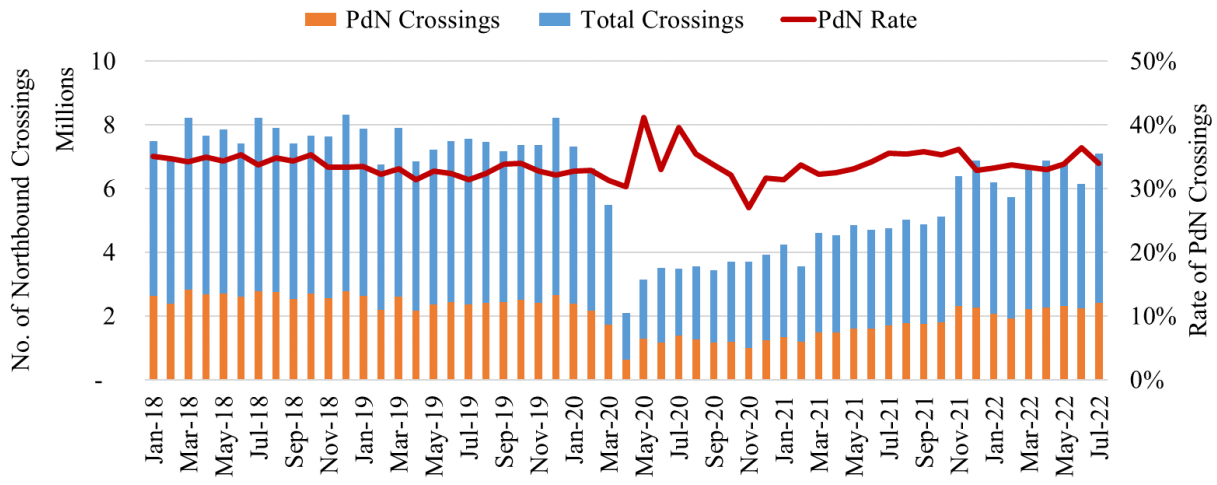


Figure 1. Monthly Northbound Crossings (Number of People) into Texas

As demonstrated in the figure, the pandemic-related border restrictions impacted the number of crossings significantly. The 7.5 million average monthly total crossings in Texas decreased to

2 million in April 2020. Although there has since been a significant increase in numbers—especially since COVID-19–related cross-border travel restrictions were lifted in early November 2021—pre-pandemic values have not yet been reached. The latest available number of crossings (July 2022) is still 6 percent lower compared to the same month before the pandemic (July 2019). Figure 1 also demonstrates the importance of the PdN region in terms of number of crossings; with one-third of total northbound crossings in Texas taking place in the PdN region.

Previous studies with a focus on binational regions have mainly tried to identify characteristics, behaviors, patterns, and preferences of cross-border travelers. Moreover, nearly all of the previous research used in-person surveys. Baruca and Zolfagharian (2013) examined both U.S. and Mexican residents and reported the shopping motivations for cross-border travelers on the opposite side of the border from where they lived. Using a survey conducted in a mall, Guo et al. (2006) investigated Mexican residents' motives for shopping in the United States. Similarly, Sullivan et al. (2012) used the mall intercept method at a large Texas mall during the Christmas holiday season to report Mexican residents' shopping expenditures and the impact of their visits on the local economy and the region on the U.S. side. Based on a small sample of survey participants, Yuan et al.'s (2013) study examined the motivations of Mexican shoppers in the United States. Bojanic (2011) surveyed Mexican residents in different malls to explore the impacts of age and family life experiences. Gutierrez et al. (2021) aimed for a smaller binational region (Mexicali–Imperial Valley) to explore the cross-border dynamics and whether those dynamics are related to the scale of the cities that make up binational regions.

The City of El Paso International Bridges Department (IBD, 2020) conducted a long-term survey (between October 2019 to March 2020) of cross-border travelers and received 8,623 survey responses at the El Paso international bridges. IBD aimed to document important social and economic activities and the magnitude of retail and service expenditures made by cross-border travelers. Over 60 percent of the survey participants reported their primary place of residence as Mexico. Study findings showed that the top reasons for Mexico-domiciled travelers to travel to the United States were shopping, social, work, and school activities, which aligned with results from the review conducted by Sener et al. (2015) and the crowdsourced-based data analysis conducted in the prior study of the researchers (Vargas et al., 2021), as seen in Table 1.

Table 1. El Paso Northbound Top Reasons for Crossing

Authors	Year of Publication	Top Reasons for Crossing to the U.S.		
		1st	2nd	3rd
Vargas et al.	2021	Shopping	School	Medical
City of El Paso IBD	2020	Shopping	Social (family)	Work/School
Sener et al.	2015	Shopping	Social (family)	Work

SAFEGRAPH AS POINT-OF-INTEREST DATA

After SafeGraph made its data available for academic research free of charge (SafeGraph, 2020), several studies were conducted to assess the comprehensiveness and compliance of the data in various states of the United States. Juhasz and Hochmair (2020) analyzed visitation patterns that were derived from the SafeGraph platform for three major cities in Florida: Miami, Orlando, and Jacksonville. They also conducted an event analysis for Hurricane Irma in the Miami metropolitan area. Lee et al. (2021) used SafeGraph data in the context of the Texas winter storm that occurred in 2021. Using the spatiotemporal data for Harris County, they reported the changes in visitation patterns to grocery stores and restaurants just before, during, and after the event. Prestby et al. (2020) used SafeGraph location data to explore neighborhood isolation in Milwaukee, Wisconsin.

When the World Health Organization declared COVID-19 a pandemic, researchers from different backgrounds shifted their focus and conducted pandemic-related studies. This was the case for mobility-focused researchers as well. Data-driven approaches were used to analyze the spatiotemporal characteristics of the disease spread. For example, researchers used SafeGraph data to assess the potential use of contact-tracing applications (Gurbuz et al., 2021b) and to link border mobility and vaccination rates with disease spread in the El Paso–Ciudad Juárez binational region (Gurbuz et al., 2021a). In their study, Killeen et al. (2020) looked at multiple variables to understand the U.S. response to COVID-19 and used SafeGraph data to understand the mobility of people at the county level. In another study using SafeGraph data, Gao et al. (2020) quantified the degree of residents' reactions to stay-at-home mandates throughout the United States. They also correlated the mobilities in selected regions with the spread of COVID-19. Meanwhile, Brzezinski et al. (2020) calculated the cost of lockdowns and benefits of the voluntary social distancing practices. They tracked mobility data from SafeGraph across the United States. Chang et al. (2020) developed a mobility network model using SafeGraph data to explore the higher infection rates among disadvantaged groups. Their model used visitor density and length of visit to identify super-spreader locations. Similarly, Kulkarni (2021) tried to address the impact of mobility and long-duration visits at bars or restaurants on disease spread. SafeGraph data from Minnesota allowed researchers to identify and analyze travel patterns and link them to the spread of COVID-19 (Sharma et al., 2022). Roy and Kar (2020) collected SafeGraph data on over 500,000 visits in over 2,800 census block groups in the city of Los Angeles before and during lockdowns as well as during different phases of reopening. They used machine learning algorithms to link the income levels of different census block groups with the disease spread to represent the vulnerability of the different income-level groups.

CHAPTER 3:
DESCRIPTION AND CHARACTERISTICS OF POI DATA USED

Various types of applications on mobile devices can serve as sources of location data, including navigation, weather, shopping, social connector, retailer, and other apps. Location data provide geographical information about a specific device associated with a time identifier. The device can then be attributed to an individual since such devices are mainly used by single individuals. Companies and developers collect location data in many different ways. Some collect data and operate in the integrated application when it is open, while others can run in the background to gain more insight into the movement and patterns of the device owner. Location-based apps collect data with the user's consent, which might be a one-time consent or consent each time the app collects data.

As an emerging source of LBS data, SafeGraph provides aggregated anonymized, high-frequency geolocation (trip) data from mobile device applications that have opted into location-sharing services. SafeGraph data include detailed information about POIs across the United States, Canada, the United Kingdom, and some other countries (SafeGraph, n.d.).

SafeGraph POI data include location information, brand and business attributes, and North American Industry Classification System (NAICS) categorical coding. POI data also provide information to recognize any trend in the number of visits to a particular POI. This includes how often people visit a location, the duration of the visit, the origin of the trip, and other trip destinations linked to the POI. Based on the latest release, SafeGraph reported having over 12 million POIs with 7,150 distinct brands in the United States (SafeGraph, 2022).

Although SafeGraph does not share traffic trajectories of visits to POIs, the data include information about the home of the visitors. This information allows general insight into the travel behavior patterns in a selected region. The datasets may also allow for the study of urban dynamics and customer preferences. SafeGraph has made its data available for academic research free of charge in a specified format that allows various researchers to explore the datasets. Potential example use of SafeGraph data in transportation includes:

- Development of origin-destination studies.
- Spatial isolation of neighborhoods.
- Event analysis.
- Pandemic response and customer reactions to policy changes.

DATA ATTRIBUTES

SafeGraph provides different sets of data sources. To capture the origin and destination of trips, the *places* and *patterns* datasets must be used together. The places dataset provides basic information on each POI, whereas the patterns dataset provides details on the trips to that POI. Table 2 lists the attributes of the places and patterns datasets. The "use case" column summarizes the key attributes of the cross-border mobility analysis considered in this study.

Table 2. Attributes of SafeGraph Datasets (SafeGraph, n.d.)

Attribute	Use Case	Places	Patterns	Description
Location ID	✓	✓	✓	Unique ID tied to the POI.
Location name	✓	✓	✓	Name of the place.
Brands	✓	✓	✓	If the POI is an instance of a larger brand, this will contain that brand name.
Top & sub-category		✓		Label associated with the first four and first six digits of the POI's NAICS category.
NAICS code	✓	✓		POI's NAICS category describing the business.
Latitude and longitude	✓	✓		Exact location of the POI.
Address		✓	✓	Detailed address of the POI including the phone number.
Phone number		✓		Phone number of the POI.
Open hours		✓		Opening and closing times in the POI's local time.
Open/closed date		✓		Year and month of the POI opened/closed.
Data collection period	✓	✓		Start and end time for the measurement period.
Visit counts	✓		✓	Number of visits to the POI during the date range.
Visits by day			✓	Number of visits to the POI each day (local time) over the covered period.
Location census block groups	✓		✓	Census block group the POI is located within.
Visitor counts	✓		✓	Number of unique visitors to the POI during the date range.
Visitor home census block groups			✓	Number of visitors to the POI from each census block group based on the visitor's home location.
Daytime visitor census block groups			✓	Number of visitors to the POI from each census block group based on primary daytime location between 9 a.m. to 5 p.m.
Visitor country of origin	✓		✓	Number of visitors to the POI from each country based on the visitor's home country code.
Distance from home			✓	Median distance from home traveled by visitors.
Median dwell time			✓	Median minimum dwell time in minutes.
Bucketed dwell times			✓	Distribution of visit dwell times based on specified buckets.
Related same-day brands			✓	Other brands that the visitors to this POI visited on the same day as the visit to this POI.
Related same-month brands			✓	Other brands that the visitors to this POI visited in the same month as the visit to this POI.
Popularity by hour			✓	Number of visits in each hour over the course of the date range in local time.
Popularity by day			✓	Number of visits in total on each day of the week over the course of the date range in local time.
Device type			✓	Number of visitors to the POI that are using Android vs. iOS.
Carrier name			✓	Number of visitors to the POI based on the carrier of the device.

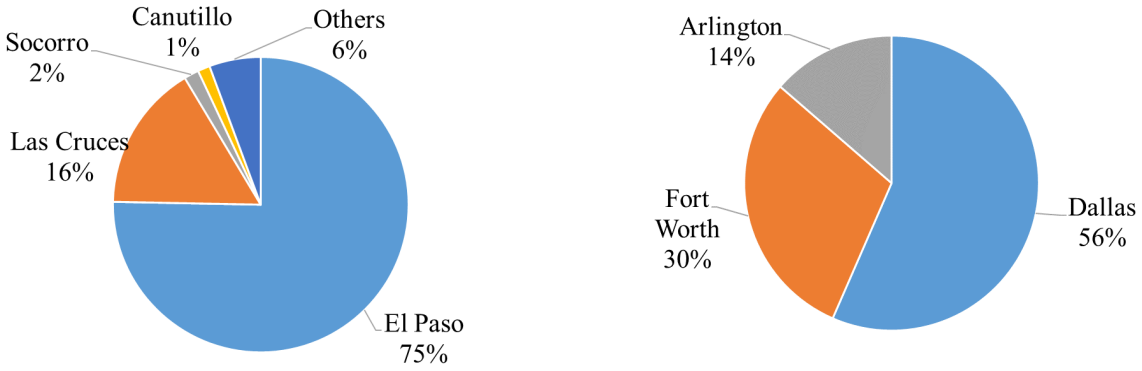
Location ID is a unique identification for the POI that was used to merge the datasets. The location name and the brand were used for verification of the business. The dataset provides six digits of the NAICS categories. This study followed the sector classification based on the first two digits of the NAICS categories. Latitude and longitude information provide the exact location of the business, which is useful in mapping and developing origin-destination studies. Moreover, the dataset provides the address of the POI, which is helpful for further verification of the location. The dataset mainly provides two counts: (a) visit counts, and (b) visitor counts. *Visits* are the number of visits to the POI during the selected time interval, which includes repeated visitors. On the other hand, *visitors* are the number of unique visitors to the POI during the date range. This information mainly helps to differentiate the employees and the customers who visited the POI. The dataset also provides information about the origin of the trips, including the visitor's home census block group, county of origin, and distance from home. Finally, the median dwelling time of visitors at the POI is given in minutes.

DATA LIMITATIONS

SafeGraph data account for individuals who have a cell phone with location services turned on. The sampling rate is calculated for each region based on the active users of data sources. Although a good source of information, the data have limitations due to sampling biases. Similar to various other crowdsourced data sources, data may not adequately represent trips taken by specific population groups who are less likely to have mobile devices, such as senior adults and young children as also reported by Chiou and Tucker (2020). Moreover, since the willingness to share personal information decreases with age, the decrease in data sharing with mobile devices may cause another bias (Goldfarb and Tucker, 2012). Another limitation is related to the accuracy of the GPS location data. Businesses sharing the same building or located in close proximity to each other could be mislabeled. SafeGraph has noted continually trying to include new POIs in its database, but it currently does not contain all POIs. Therefore, some trips may mislabel the trip's destination. One other major limitation relates to non-commercial POIs. Data coverage is low in non-commercial POIs, and due to privacy concerns, SafeGraph does not provide any information for residential destinations (Li et al., 2022).

DATA COLLECTION

During the execution of this research study, SafeGraph allowed the researchers access to monthly Texas and New Mexico datasets from 2018 to 2021. The researchers filtered the cities of the PdN region (El Paso, TX; Las Cruces, NM; Anthony, TX; Anthony, NM; Canutillo, TX; Clint, TX; Fabens, TX; Horizon City, TX; Mesilla, NM; San Elizario, TX; Santa Teresa, NM; Socorro, TX; Sunland Park, NM; Tornillo, TX, and Vinton TX). A single dataset was created by merging the places and patterns datasets of the selected cities and was named the PdN region. Similarly, Dallas, Fort Worth, and Arlington were merged and named the DFW region. The number of POIs for each multicity region is demonstrated in Figure 2. Figure 3 presents the spatial distribution of POIs in the PdN region. Other large Texas cities that attract Mexico-domiciled travelers, including Austin, Brownsville, Houston, Laredo, McAllen, and San Antonio, were also explored for comparison purposes.



(a) PdN region (b) DFW region

Figure 2. Distribution of POIs in Selected Multicity Regions in Texas and New Mexico

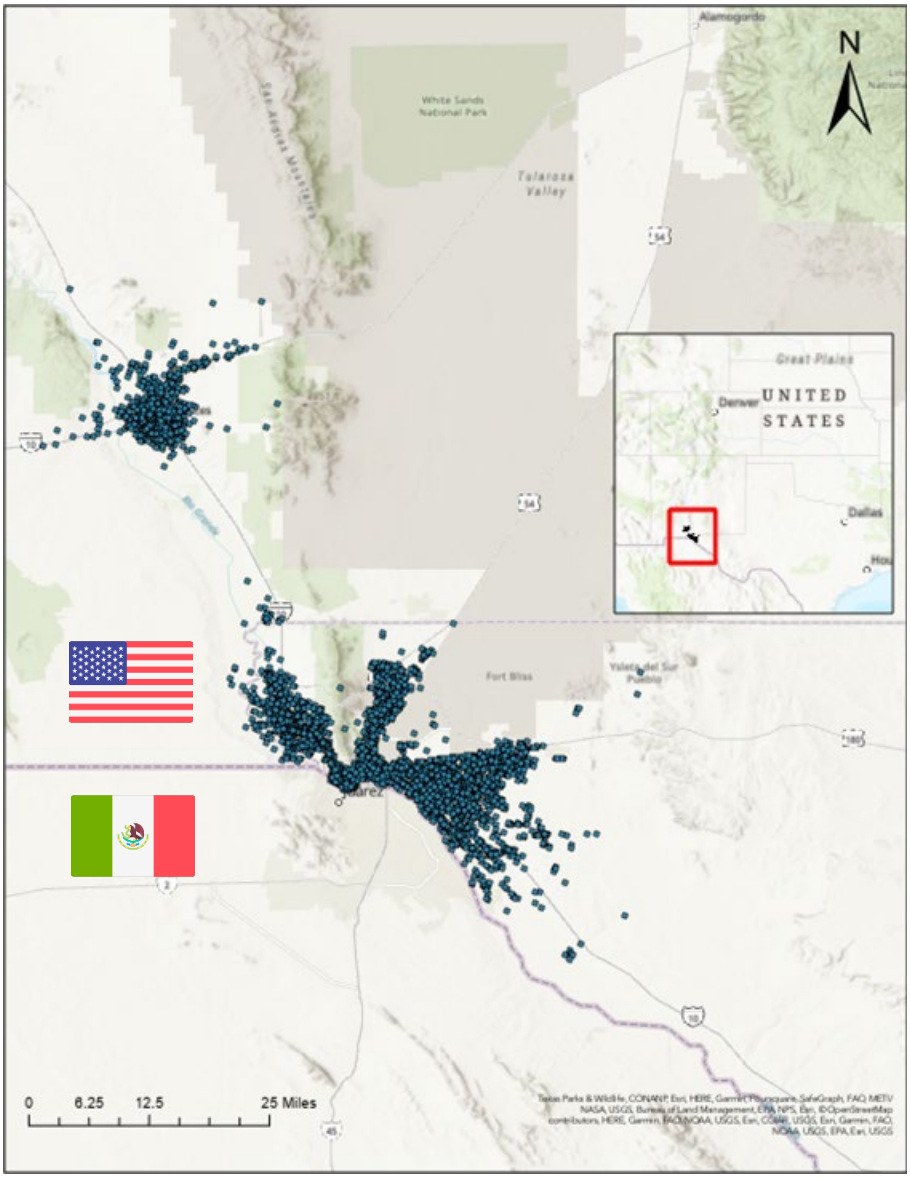


Figure 3. Spatial Distribution of POIs in the PdN Region

This study developed multiple approaches to assess the usability of the SafeGraph dataset for analyzing cross-border activities of Mexico-domiciled travelers in U.S. cities. The researchers collected four years of data (2018 to 2021) by using the monthly aggregated visits and monthly Mexico-domiciled visitors from the patterns dataset. To smooth monthly visitor counts and remove potential seasonal patterns, the researchers merged the data to get the annual counts for each POI for the selected locations. This study also calculated the number of POIs for each region that had at least one visit for a one-year period (Figure 4). SafeGraph increased its POI coverage in recent years. Although the pandemic led to lockdowns and business closures in nearly all of the selected locations, the number of POIs covered by the data has still been on the rise since 2018. Houston had the greatest number of POIs, over 50,000, in 2021, followed by the DFW region, San Antonio, and Austin.

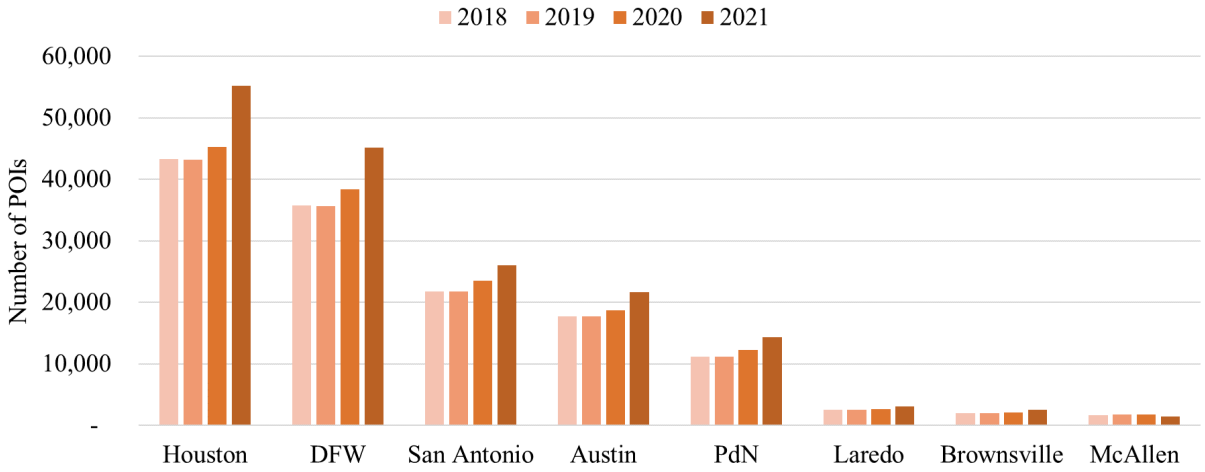


Figure 4. Number of POIs for the Selected Locations

SAMPLE SIZE

The adoption of smart devices has grown significantly in the last decade. The services provided by mobile devices offer a powerful tool for understanding the aggregated behavior of user travel patterns. However, the question among researchers is still whether the size of related datasets is large enough to conduct an exploratory analysis for the entire population.

The City of El Paso IBD (2020) conducted an international bridges cross-border survey at various entry points to the United States to understand the social and expenditure profile of cross-border travelers. In-person surveys were conducted all days of the week for a six-month period from October 2019 to March 2020 at three of the region's main ports of entry. Participants were asked to share their primary place of residence and grouped as Mexican residents and U.S. residents. Of the 8,263 survey participants, 60.3 percent declared that they were Mexican residents.

BTS reported the total number of crossings as 13,367,126 in all possible modes of transportation including bus passengers, pedestrians, personal vehicle passengers, and train passengers (BTS, 2022). Considering the output of the IBD survey mentioned above, 60.3 percent of those could be Mexican residents (8,060,377 crossings). By having the same period of

data in the SafeGraph dataset for the same region, the researchers were able to develop a sampling analysis.

Due to privacy considerations, SafeGraph reports four visitors when the number of visitors is less than four to a particular POI. In other words, POIs overrepresent the actual visitors if the number of visitors is reported as four. Sampling analysis was conducted considering this fact. For the aforementioned six-month period, SafeGraph reported 235,533 visitors whose home was labeled as Mexico. Considering the POIs reported four visitors, this value might be reduced to 208,269 visitors. In other words, 208,269 to 235,533 visitors represented the 8,060,377 crossings reported by BTS, which made the sample size between 2.6 and 2.9 percent for all cross-border visitors whose home was Mexico (Mexico-domiciled).

CHAPTER 4:
FINDINGS FROM POI-BASED DATA ANALYSIS

The research team conducted a step-by-step analysis of the emerging POI data from SafeGraph to develop a comprehensive picture of northbound cross-border trips, particularly for social and economic cross-border activities. The research team developed a methodology to explore the destinations of cross-border travelers with information on the distribution and characteristics of local trips that originated from Mexico. The time span allowed the research team to identify the degree and patterns of trips that took place before and after the implementation of pandemic-related border-crossing restrictions. In addition to exploring the role of utilizing location-based crowdsourced data sources to examine cross-border trips, this study developed an alternative approach to assist regional agencies seeking to demonstrate how external cross-border traffic impacts the community.

CROSS-BORDER VISITORS BY DESTINATION

Although most of the POIs are located in highly populated regions in Texas, as expected, most of the Mexico-domiciled visitors were captured at the border regions. Figure 5 demonstrates the number of visitors from Mexico to the selected locations in the Texas and New Mexico regions of the United States from 2018 to 2021. Laredo and the PdN region had the highest number of visitors, followed by McAllen and Brownsville, all of which share borders with Mexico and have ports of entry.

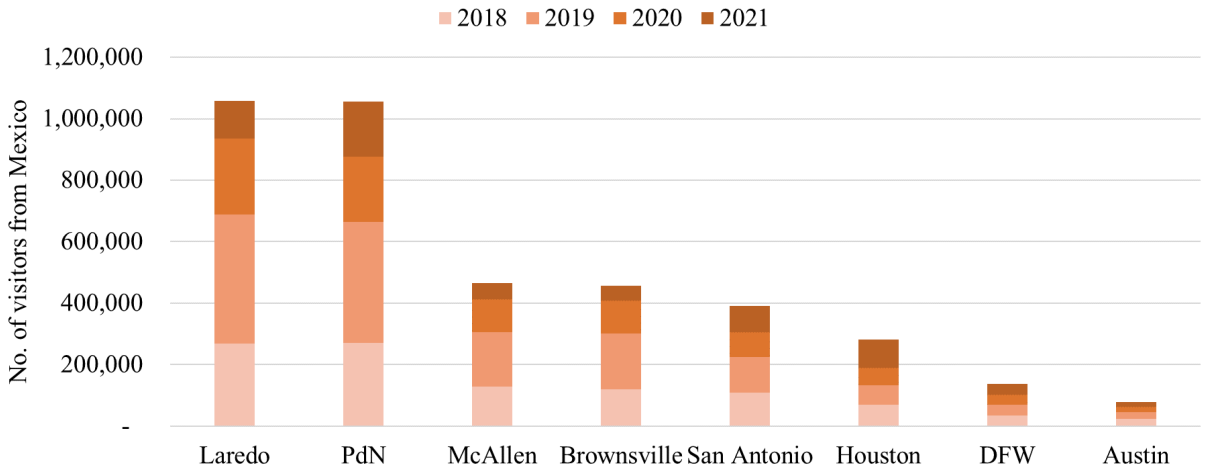


Figure 5. Number of Visitors from Mexico for the Selected Locations

Figure 6 provides visual insight by mapping the number of visitors to the selected regions. The locations of the POIs were captured from the places dataset, and the number of visitors gathered from the patterns dataset was attached to each POI. As illustrated with the gradual bubble map, the border locations where the bubbles are larger indicate locations where Mexico-domiciled travelers had more trips compared to the other locations selected, even though they are comparatively smaller in terms of the overall population.

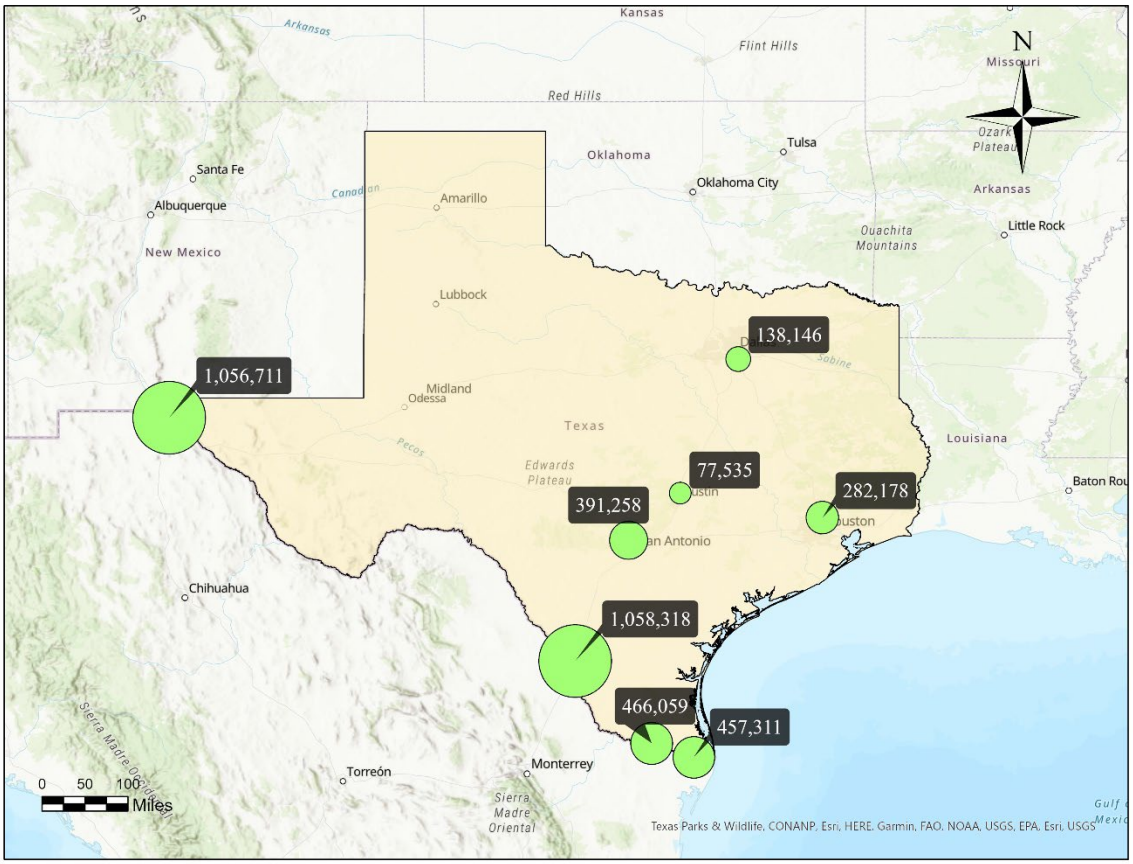


Figure 6. Mexican Trips to Selected Texas Cities

Figure 7 provides a detailed look into the number of visitors for each POI in El Paso by Mexico-domiciled travelers in October 2019. The downtown area of El Paso and various shopping malls along I-10 were the main destinations for the trips from Mexico. As part of this study, an online map with four layers—one for each year—was created to illustrate the most-visited locations by travelers from Mexico between 2018 and 2021. The online heat map* also helps visualize the changes in number of visits in the PdN region over the study period.

https://arcg.is/18znu4

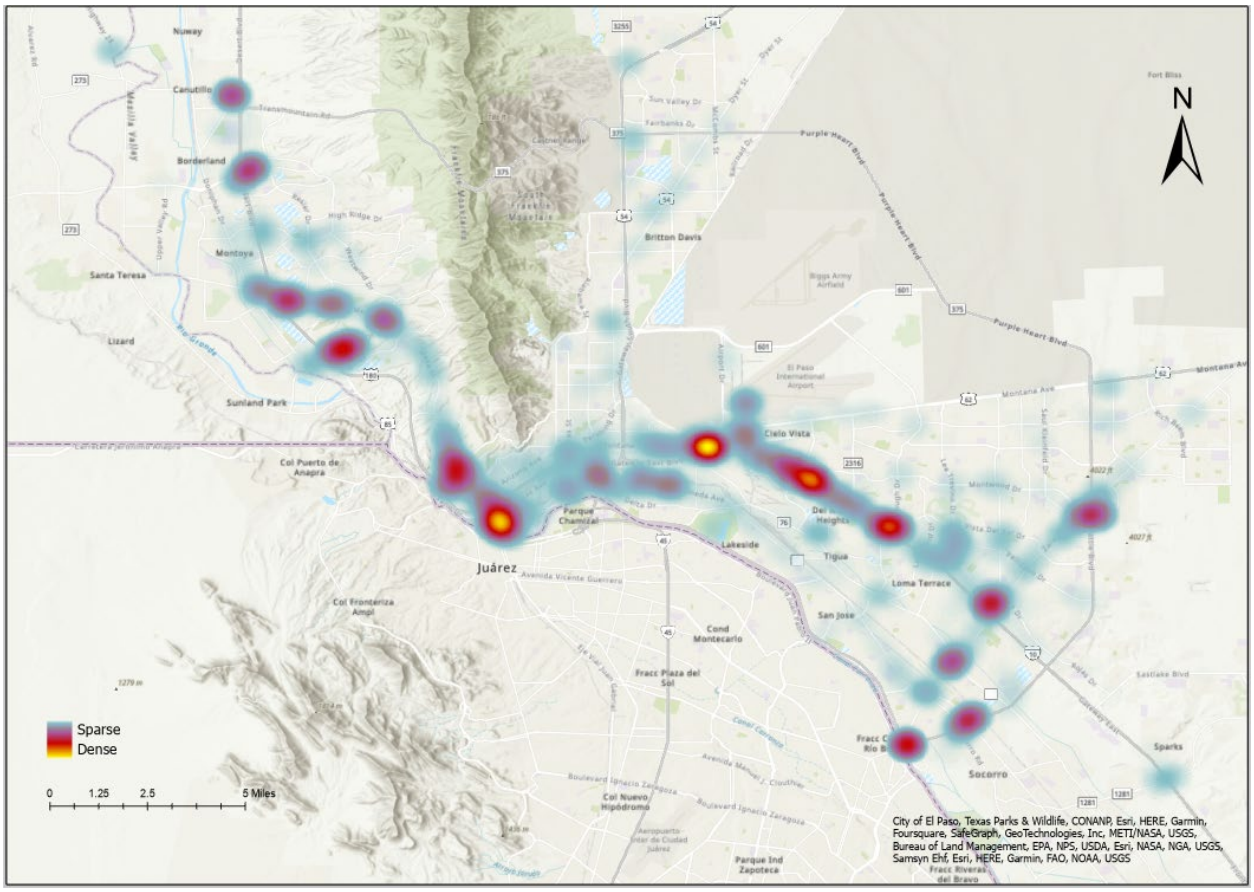


Figure 7. Number of Visitors from Mexico to the POIs in El Paso—October 2019

Finally, Table 3 lists the most-visited places by Mexico-domiciled travelers for each location (i.e., the list counts the categories recorded in the top 10 most-visited places). In summary, Mexico-domiciled travelers generally visited shopping malls and retail stores at the locations sharing the border. On the other hand, in other locations, categories including amusement parks, hotels, airports, automobile dealers, restaurants, coffee shops, healthcare offices, and transit stations were among the most-visited locations. For example, healthcare businesses were among the top 10 most-visited places in San Antonio, and transit stations made the list in Austin and the DFW region.

Table 3. Top 10 Most-Visited Categories by Locations

Categories	Austin	Brownsville	DFW	Houston	Laredo	McAllen	PdN	San Antonio
Shopping Malls	6	4	4	3	4	3	6	4
Retail Stores	2	6	1	1	6	6	4	4
Amusement Parks	—	—	1	1	—	—	—	—
Hotels	—	—	2	—	—	—	—	—
Airports	—	—	—	1	—	1	—	—
Automobile Dealers	—	—	—	1	—	—	—	—
Restaurants, Coffee Shops	—	—	1	2	—	—	—	1
Health Care	—	—	—	—	—	—	—	1
Museums and Parks	1	—	—	1	—	—	—	—
Transit Stations	1	—	1	—	—	—	—	—

Note: — = none. The values in the table should be interpreted as in the following example: In Austin, six out of the 10 most-visited places were shopping malls, while in the DFW region, two hotels were recorded in the top 10 most-visited places.

CROSS-BORDER VISITORS BY SECTOR

NAICS is a standard six-digit code developed by federal statistical agencies to classify businesses for the purpose of collecting and publishing statistical data related to the U.S. business economy. SafeGraph has defined the NAICS code for each POI. This study focused on the first two digits to classify the sector of the businesses and reported the cross-border activities by Mexico-domiciled travelers.

Table 4 lists the number of cross-border travelers for each two-digit NAICS code and includes the distribution to demonstrate the Mexico-domiciled travelers' preferences in terms of sectors. The retail trade sector, including businesses to sell merchandise in small quantities to the public, covered over 44 percent of Mexico-domiciled travelers' trips. General merchandise stores, clothing stores, liquor stores, electronics and appliance stores, and health and personal care stores are some examples of this sector. In addition, over 27 percent of Mexico-domiciled travelers preferred to visit POIs grouped under accommodation and food services. Restaurants, drinking places, and traveler accommodations (hotels) are the main categories of this sector. Nearly 13 percent of the travelers visited businesses in real estate and rental and leasing. Shopping malls are the major trip attraction POIs considered under this category. Moreover, amusement parks, museums, historical sites, and sports venues are grouped under the arts, entertainment, and recreation sector, and nearly 6 percent of the cross-border Mexico-domiciled travelers preferred to visit those places. Finally, healthcare-related visits accounted for 3 percent, followed by education-related visits with 2 percent.

Table 4. Cross-Border Mexico-Domiciled Travelers by Sector

NAICS Code	Definition	Total Mexican Visitors	Percent Distribution
22	Utilities	67	0.0%
23	Construction	2,113	0.1%
31–33	Manufacturing	12,530	0.3%
42	Wholesale Trade	8,004	0.2%
44–45	Retail Trade	1,733,605	44.2%
48–49	Transportation and Warehousing	41,449	1.1%
51	Information	21,377	0.5%
52	Finance and Insurance	35,184	0.9%
53	Real Estate and Rental and Leasing	497,376	12.7%
54	Professional, Scientific, and Technical Services	7,033	0.2%
55	Management of Companies and Enterprises	630	0.0%
56	Administrative and Support and Waste Management and Remediation Services	2,676	0.1%
61	Educational Services	76,889	2.0%
62	Health Care and Social Assistance	116,203	3.0%
71	Arts, Entertainment, and Recreation	227,479	5.8%
72	Accommodation and Food Services	1,067,159	27.2%
81	Other Services	63,009	1.6%
92	Public Administration	11,772	0.3%

In addition to total trips to the selected regions, the researchers also checked the differences among the study locations. Figure 8 shows the five most-visited sectors for each location. Interestingly, all locations display the same five sectors, with minor shifts in the order. Cross-border travelers preferred mostly retail trade stores, with higher values at border communities (Brownsville, McAllen, Laredo, and the PdN region). The distribution of the retail trade sector was lower for the other cities; the accommodation and food services and the arts, entertainment, and recreation sectors had higher rates in non-border locations (Austin, DFW, Houston, and San Antonio).

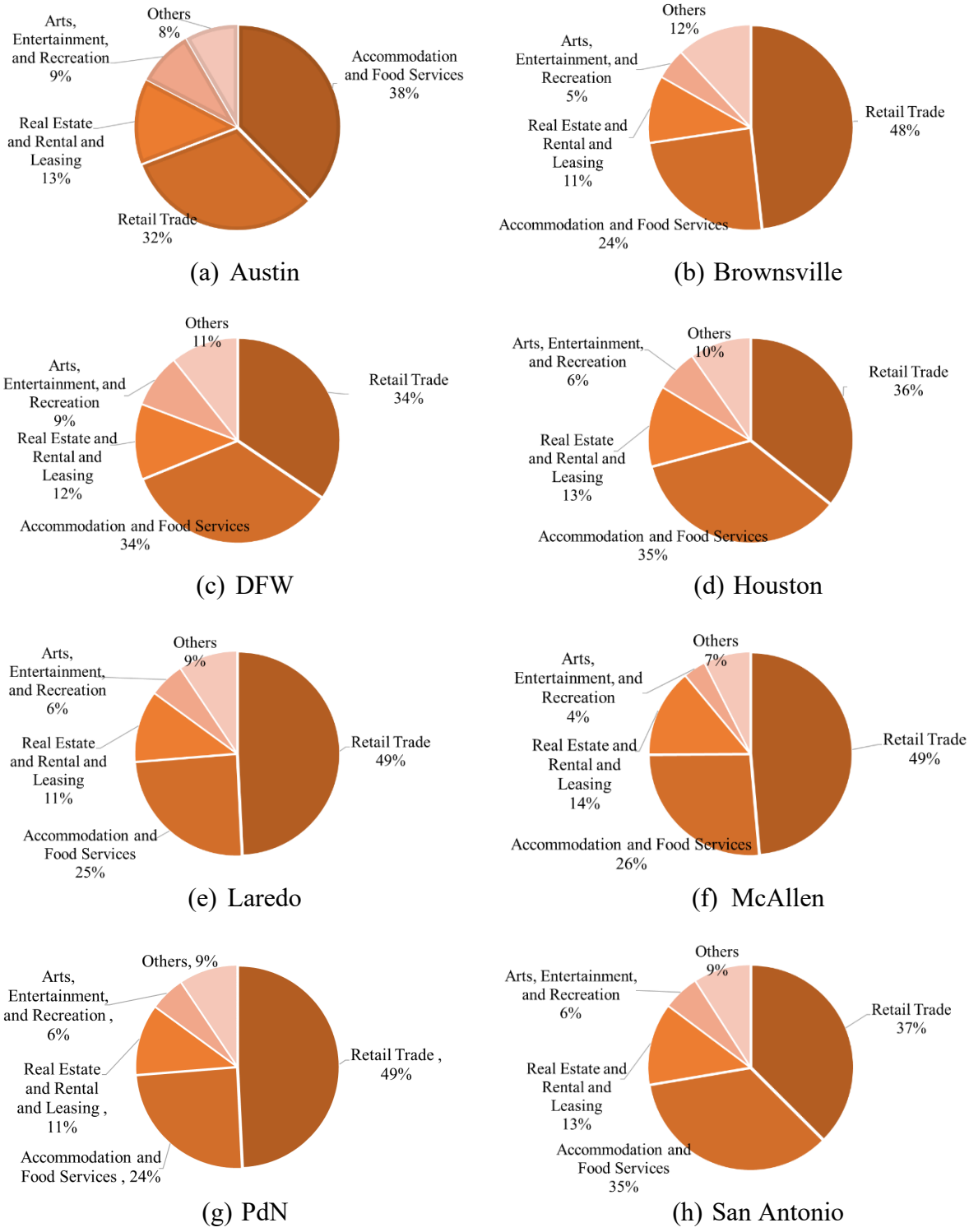


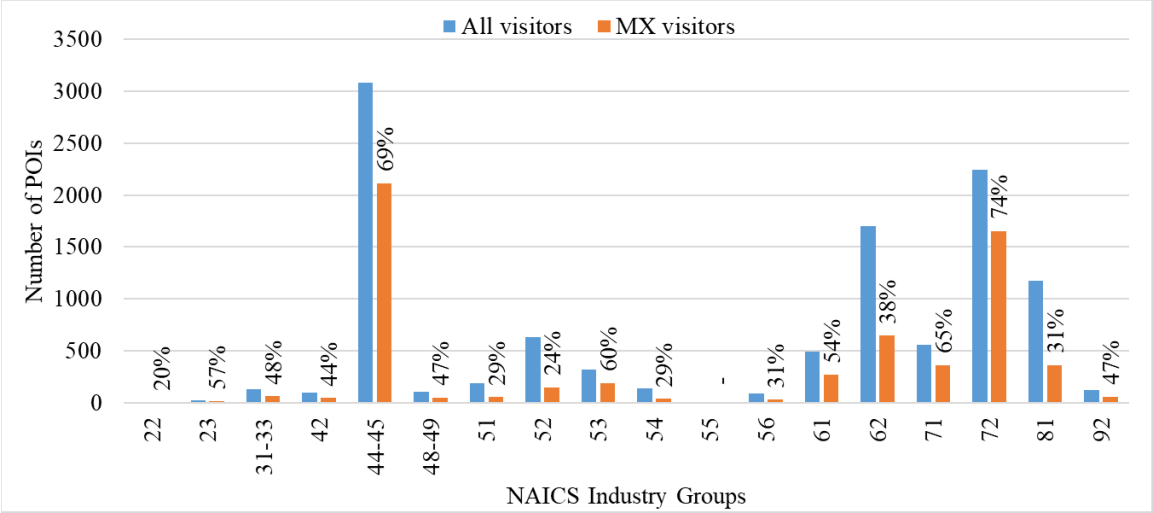
Figure 8. Cross-Border Mexico-Domiciled Travelers' Distribution by Category

Table 5 illustrates the distribution of POIs for the PdN region over time. As shown in the table, SafeGraph increased the number of POIs in the dataset from 11,095 in 2018 to 14,276 in 2021. The retail trade industry group captured the greatest number of POIs in the PdN region.

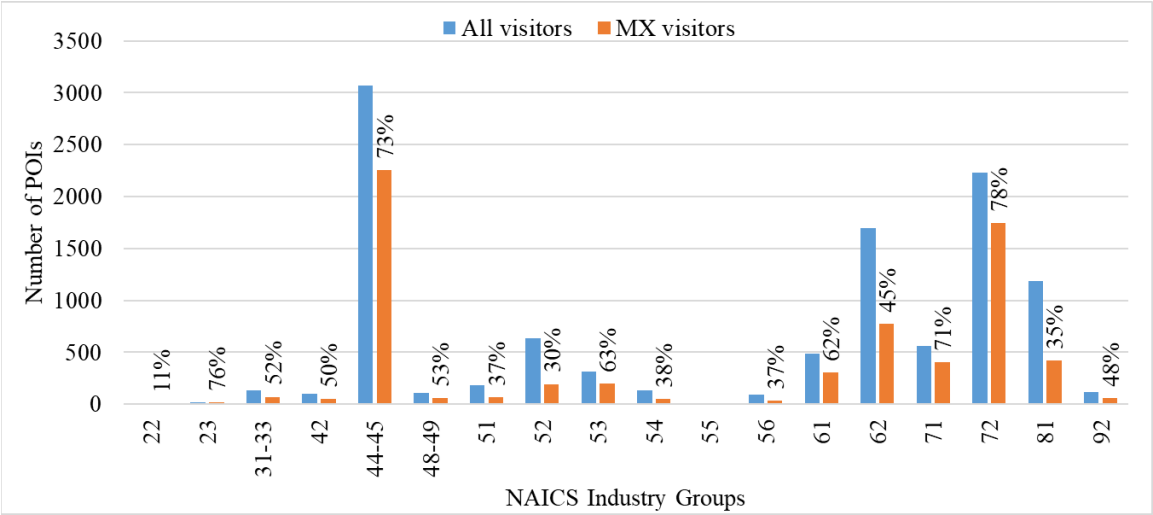
Table 5. Total Number of POIs in the PdN Region

NAICS Code	Industry Title	2018	2019	2020	2021
11	Agriculture, Forestry, Fishing and Hunting	—	—	—	—
21	Mining	—	—	—	—
22	Utilities	10	9	11	11
23	Construction	23	21	31	65
31–33	Manufacturing	128	132	168	232
42	Wholesale Trade	99	98	123	190
44–45	Retail Trade	3,079	3,073	3,353	3,777
48–49	Transportation and Warehousing	106	106	123	184
51	Information	184	182	166	190
52	Finance and Insurance	632	636	701	811
53	Real Estate Rental and Leasing	319	316	366	427
54	Professional, Scientific, and Technical Services	134	130	155	197
55	Management of Companies and Enterprises	—	—	—	25
56	Administrative and Support and Waste Management and Remediation Services	88	89	50	63
61	Educational Services	492	490	491	511
62	Health Care and Social Assistance	1,704	1,699	1,825	2,104
71	Arts, Entertainment, and Recreation	561	562	831	872
72	Accommodation and Food Services	2,243	2,229	2,346	2,637
81	Other Services (except Public Administration)	1,172	1,185	1,351	1,819
92	Public Administration	121	121	125	161
	Total	11,095	11,078	12,216	14,276

Figure 9 compares the number of POIs in the PdN region that had at least one visitor from Mexico by industry group for each study year. Since 2018, the POIs classified as accommodation and food services (Code 72) have consistently had the highest rate in the region. In addition, 74 percent of all POIs classified as accommodation and food services were visited by at least one visitor residing in Mexico in 2018. Not until 2021 did this dramatically change. In 2021, the rate of Mexico-domiciled visitors decreased across all industry sectors. This decrease could be due to a change in data coverage that the researchers were unaware of during the execution of this project.

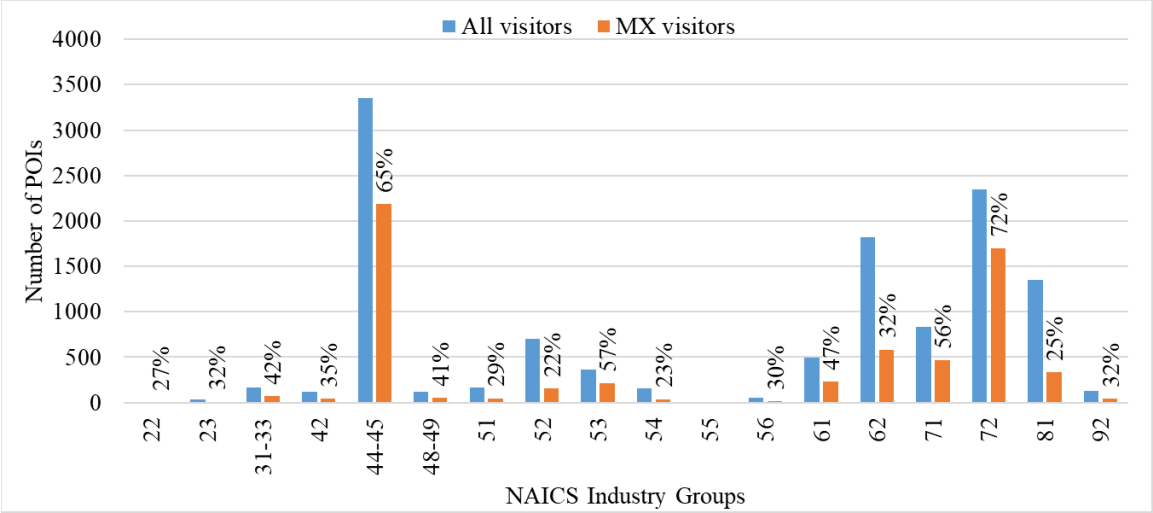


(a) 2018

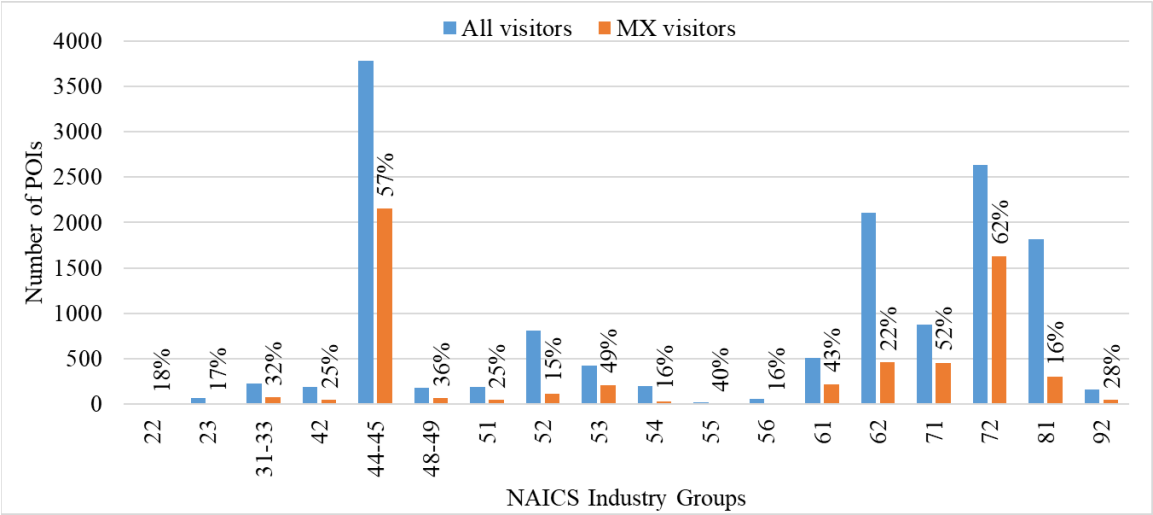


(b) 2019

Figure 9. Number of POIs in the PdN Region by Industry Sector



(c) 2020



(d) 2021

Figure 9. Number of POIs in the PdN Region by Industry Sector (Continued)

Although the locations had similar most-visited categories, one of the biggest differences among the selected areas was the rate of cross-border Mexico-domiciled travelers within the total number of trips. This value was calculated by dividing the total number of Mexico-domiciled travelers by the sum of Mexico- and U.S.-domiciled visitors captured in the SafeGraph dataset within the same period. There was a significant gap between border and non-border communities. Among the border communities, the lowest rate was experienced in the PdN region with 2.1 percent, which was nearly 10 times higher than the highest non-border city, San Antonio, which had 0.2 percent Mexico-domiciled visitors in the dataset (see Figure 10).

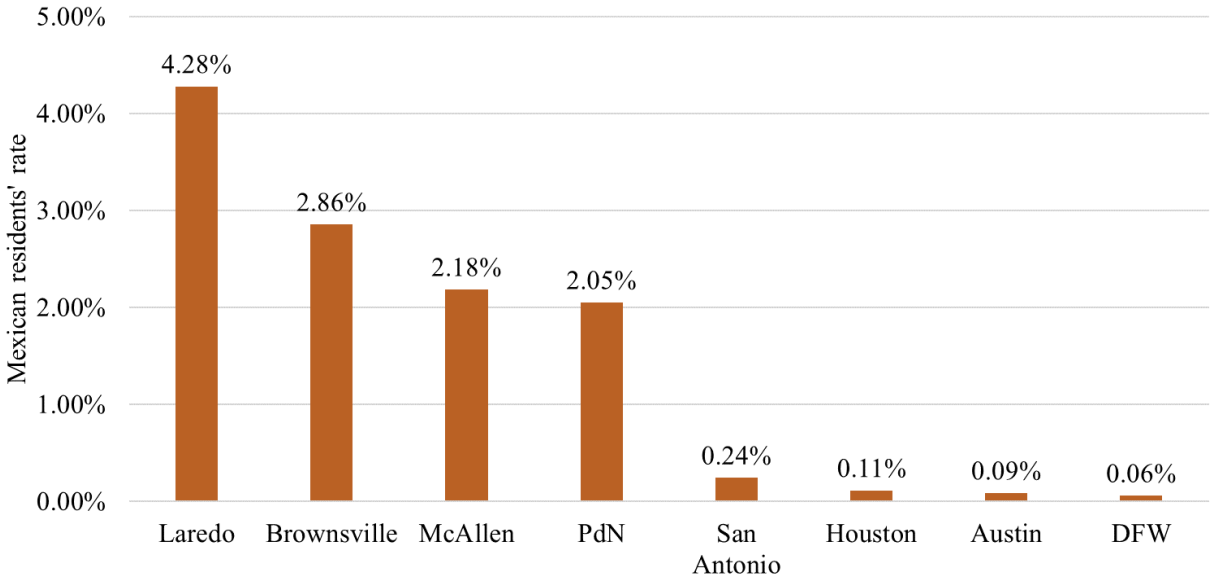
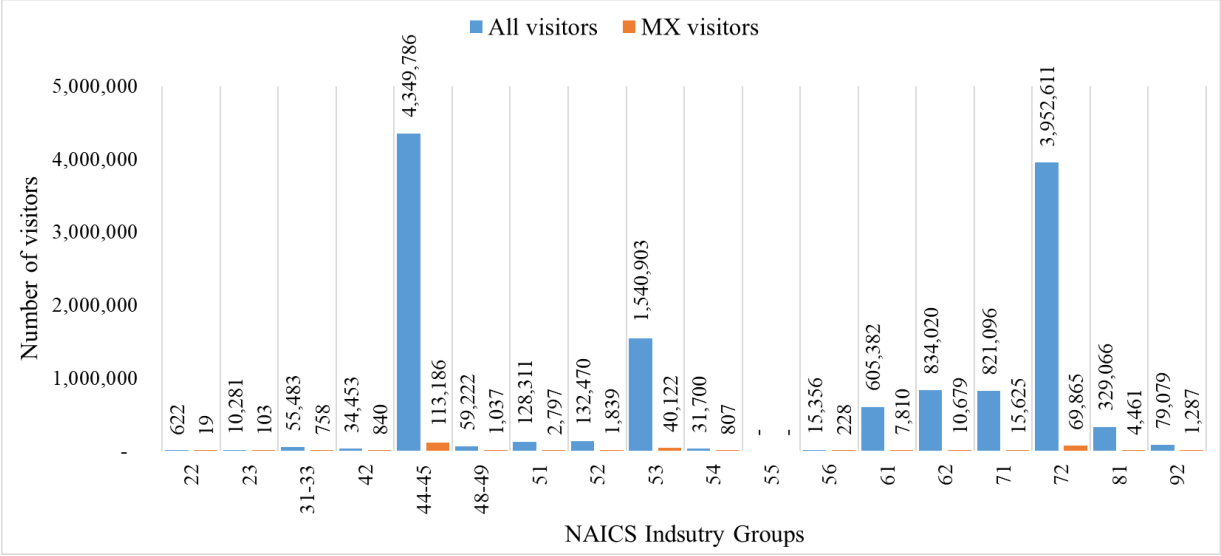
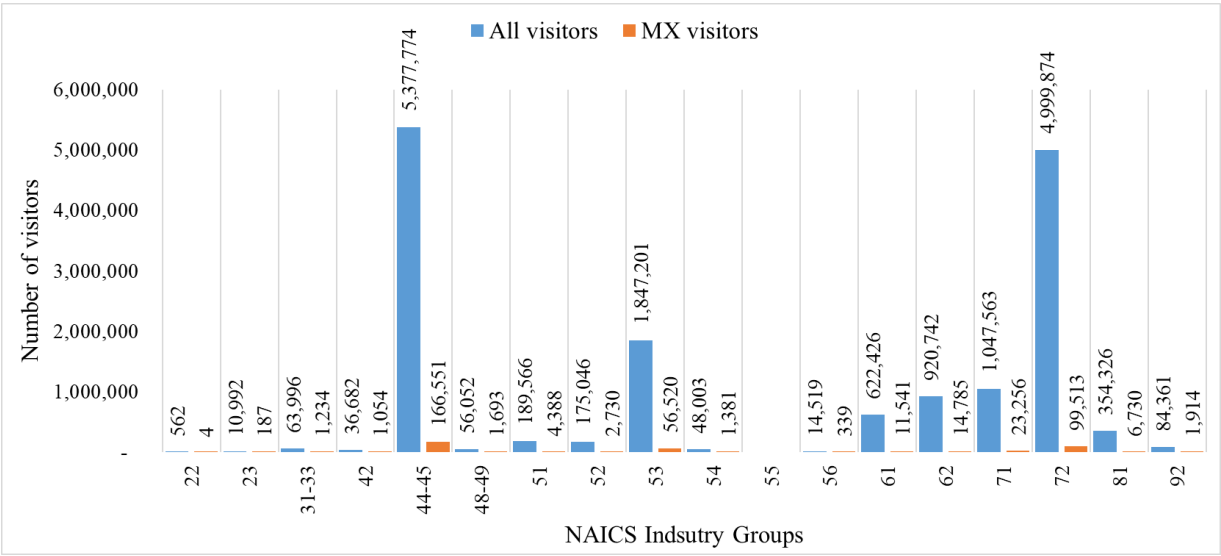


Figure 10. Mexico-Domiciled Visitor Rate among Total Visitors

Figure 11 depicts the number of Mexico-domiciled visitors in the PdN region alongside all visitors by industry over time. All visitors are shown using blue bars, while Mexico-domiciled visitors are indicated with orange bars. As shown in the figure, SafeGraph reported 113,186 Mexico-domiciled visitors to retail trade businesses in 2018. In the same year, 4,349,786 people total visited the same POIs. The potential reduction in SafeGraph data coverage in 2021 might lead to lower rates of Mexico-domiciled travelers (Figure 10d). Prior to the pandemic, the overall rate of Mexico-domiciled travelers in the PdN region was close to 3 percent (Figure 11a, 11b), but the rate dropped to less than 2 percent after border travel restrictions (Figure 11c).

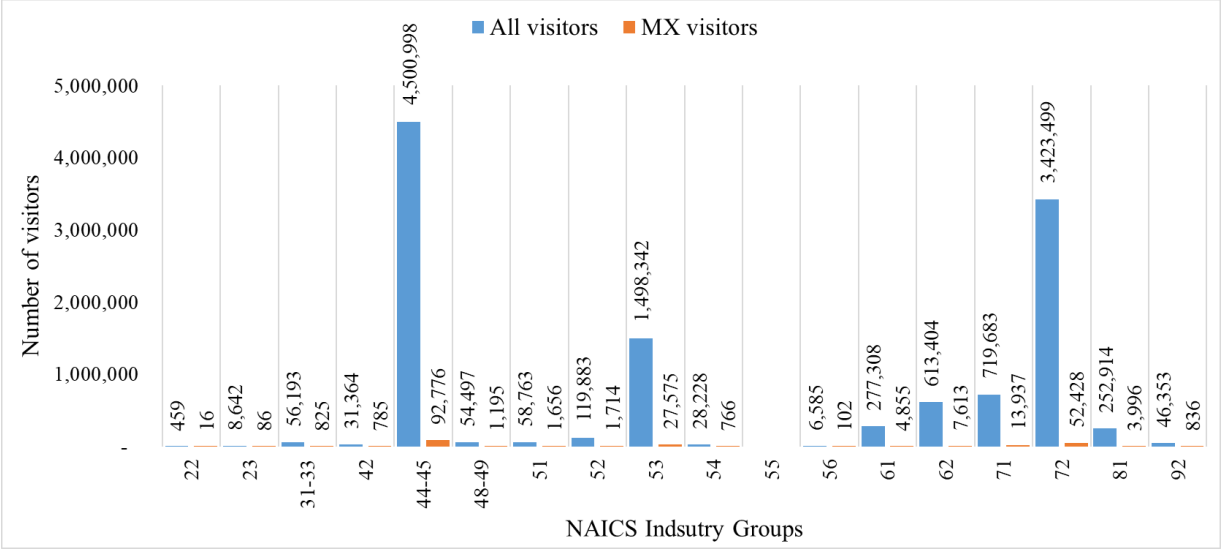


(a) 2018

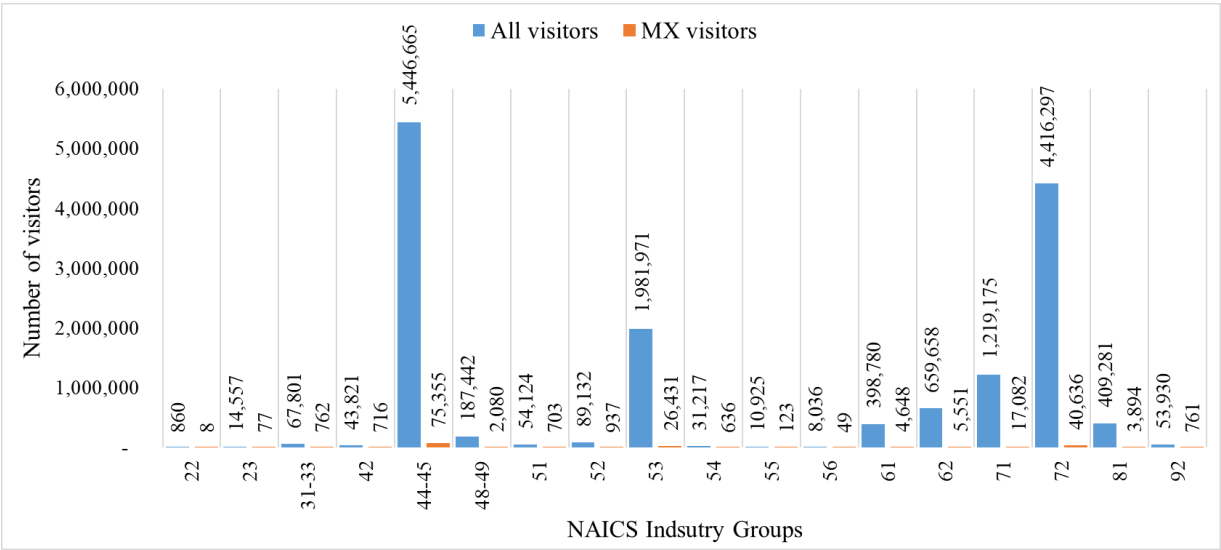


(b) 2019

Figure 11. Number of Mexico-Domiciled Visitors in the PdN Region by Industry Sector



(c) 2020



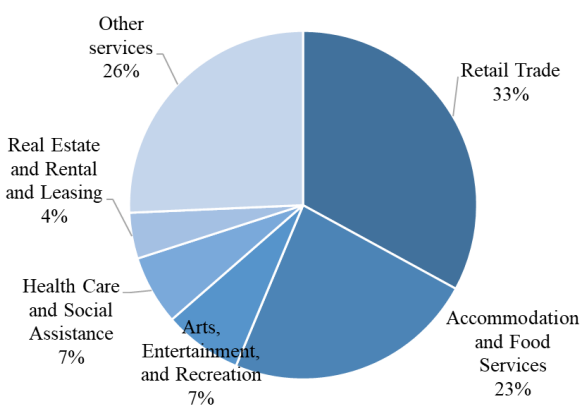
(d) 2021

Figure 11. Number of Mexico-Domiciled Visitors in the PdN Region by Industry Sector (Continued)

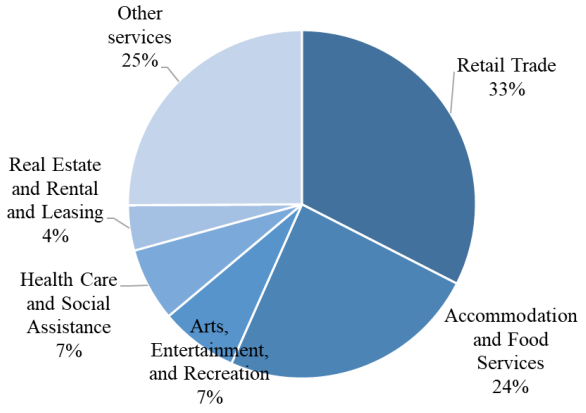
COMPARISON OF DESTINATIONS USING SAFEGRAPH AND INRIX DATA

INRIX and SafeGraph provide distinct data characteristics—with INRIX focusing on individual trip trajectories and SafeGraph focusing on POI visitations—with likely potential to integrate and expand their capabilities in various applications. Integrating or comparing these two datasets was outside the scope of this study. However, to garner a quick glimpse into the potential differences between the two data sources, the researchers examined the distribution of destinations using two months of pre-pandemic data (October to November 2019) and after cleaning the data from residential trips for a more closely matched comparison.

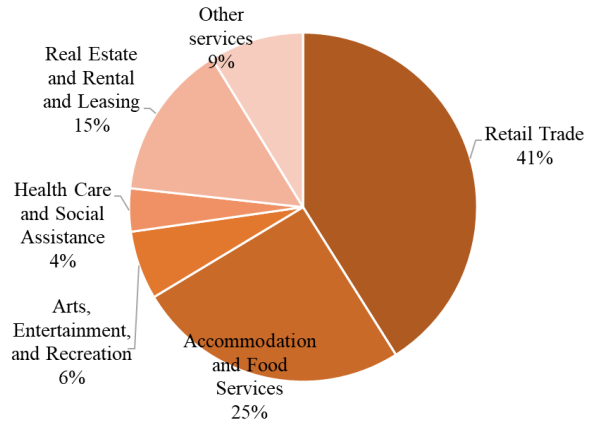
Figure 12 demonstrates the northbound trip destinations in October and November 2019 by industry sector in the PdN region. Although INRIX covers all northbound trips, including those made by U.S. residents, and SafeGraph may have captured multiple trips made by a traveler from Mexico, the distribution of trips in terms of industry categories revealed that the most-visited places were in roughly the same order in both datasets. As seen in the figure, the retail trade industry, followed by accommodation and food services, recorded the highest number of visits in both the INRIX and SafeGraph datasets. SafeGraph captured more visitors to real estate and rental and leasing businesses, in which all shopping malls are grouped. This was probably a result of SafeGraph providing data only for visitors whose home location is listed in Mexico. In other words, SafeGraph data represent more Mexico-domiciled travel patterns, whereas INRIX data provide information on all northbound crossings, including U.S. residents. According to the most recent survey conducted by the City of El Paso IBD (2020), 39.7 percent of northbound border-crossing travelers were U.S. residents.



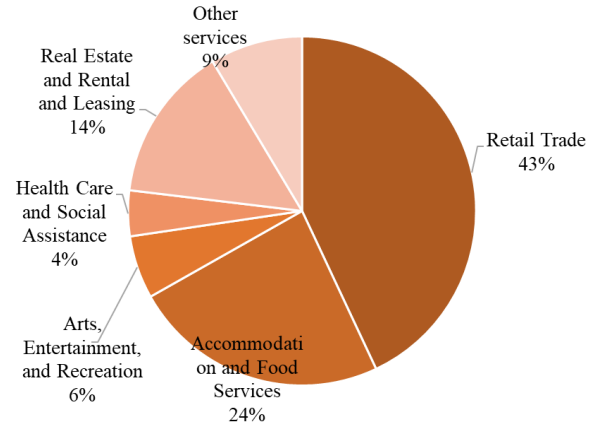
(a) October 2019 INRIX



(b) November 2019 INRIX



(c) October 2019 SafeGraph



(d) November 2019 SafeGraph

Figure 12. Northbound Trip Destinations by Industry Sector

CHAPTER 5: CONCLUSION

Border communities are crucial business and leisure destinations, serving a much larger population than their own. They are significantly impacted by cross-border traffic through local trips that are usually generated by commuters or visitors who reside across the international border. However, local transportation agencies have relatively limited funding to deliver the transportation improvements needed to support economic development in the region. The lack of reliable, continuous, and large-scale cross-border travel data makes it difficult for local transportation authorities to make the case for transportation funding that considers the overall usage of the transportation network.

Recent advancements in technologies and the proliferation of smartphones have created new data sources. In recognition of these advancements, this study conducted a step-by-step analysis to assess the potential role and use of emerging data sources with a primary focus on LBS data. The LBS crowdsourced data used in this study were drawn from SafeGraph, which is the data provider of POI data based on business listings and location information alongside the number of trips. Although SafeGraph does not share traffic trajectories of visits to POIs, the data include information about the home of the visitors. This information allows general insight into the travel behavior patterns in a selected region. The datasets may also allow for the study of urban dynamics and customer preferences, as well as various other studies such as origin-destination studies, event analysis or policy analysis focusing on emergencies.

The researchers focused on a specific use case by analyzing the spatial and temporal visitation patterns of POIs in select U.S. cities—including four locations to represent the border regions (Brownsville, McAllen, Laredo, and PdN) and four non-border locations (Austin, DFW, Houston, and San Antonio)—to yield a clear comparison of Mexico-domiciled travelers' choices. The study utilized data from January 2018 to December 2021 (four years), thus allowing researchers to identify the similarities or differences in mobility patterns before and after the COVID-19 pandemic and the pre-and post-implementation of border-crossing restrictions. The study extended the insights into the potential use of SafeGraph data explored in previous studies and provided a new approach by focusing on the home country (Mexico) of the visitors. Detailed analyses were conducted to identify the visitation patterns (e.g., most-visited places by Mexico-domiciled travelers) and the main sectors (based on NAICS sector codes for each POI) that attracted Mexico-domiciled travelers the most. The findings of this study can be particularly helpful for officials and local business owners to identify current and potential points of interest for visitors and evaluate future developments in the region.

The study also conducted an additional follow-up analysis using other crowdsourced data obtained from INRIX. The findings of these additional analyses were not the focus of this report (see Vargas et al. [2021] and Vargas et al. [2022] for details related to the INRIX-based border-crossing analyses conducted by the research team). Nonetheless, as part of this study, the researchers performed a quick comparison analysis and identified similar patterns of the most-visited places in both datasets. Future studies including a robust comparison between crowdsourced data obtained from different location-based data sources could strengthen the value of such data sources and potentially expand their usage. Such a comparison would also

help to better understand the limitations of use in certain applications, such as those related to sampling size and bias, representativeness of specific population groups, or identification of demographic groups.

The study findings highlight the critical role of emerging data sources in understanding the characteristics and impact of cross-border trips. While this study provided an overview of cross-border travel patterns within and outside of a specific region along the U.S.-Mexico border, the insights obtained as well as the overall approach followed in the study are applicable and transferable to examine cross-border trip characteristics in any border community. Given the difficulty in collecting reliable, continuous, large-scale cross-border travel data, crowdsourced data sources provide a critical means of gathering information, especially for communities with limited data or resources to collect and maintain the data. Although some limitations exist, LBS data or other crowdsourced data can serve as an additional and complementary source of information to traditional traveler surveys or count data. With increases in coverage, as well as improvement in sampling biases and representativeness, the data have the potential to help urban planners and business owners better understand the dynamics of specific regions. This understanding becomes particularly vital for cross-border transportation authorities who are seeking tools to demonstrate the need for funding for transportation improvements—based not only on the population but also on the number of users of the transportation network, whether they reside in the border regions or not, and their activity and destination patterns.

REFERENCES

Baruca, A., and Zolfagharian, M., 2013. Cross-border shopping: Mexican shoppers in the US and American shoppers in Mexico. *International Journal of Consumer Studies, 37*(4), 360–366. https://doi.org/10.1111/j.1470-6431.2012.01097.x

Bojanic, D. C., 2011. The impact of age and family life experiences on Mexican visitor shopping expenditures. *Tourism Management, 32*(2), 406–414. https://doi.org/10.1016/j.tourman.2010.03.012

Brzezinski, A., Kecht, V., and Van Dijcke, D., 2020. The cost of staying open: Voluntary social distancing and lockdowns in the US. Economics Series Working Papers 910, University of Oxford, Department of Economics. http://dx.doi.org/10.2139/ssrn.3614494

Bureau of Transportation Statistics (BTS), 2022. Border crossing/entry data. U.S. Department of Transportation. Accessed 8/4/2022. https://www.bts.gov/browse-statistical-products-and-data/border-crossing-data/border-crossingentry-data

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J., 2020. Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. *Nature, 589*. https://doi.org/10.1101/2020.06.15.20131979

Chiou, L., and Tucker, C., 2020. *Social distancing, internet access and inequality*. National Bureau of Economic Research (No. w26982). https://doi.org/10.3386/w26982

City of El Paso International Bridges Department (IBD), 2020. International Bridges Crossborder Survey: El Paso-Ciudad Juárez social and expenditure profile. https://pdnuno.com/data/ibcs

Figueroa, A., Lee, E., and Van Schoik, R., 2012. *Realizing the full value of cross-border trade with Mexico*. New Policy Institute and the North American Center for Transborder Studies, Arizona State University.

Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A. K., Reyes, J. F. M., Yandell, B. S., and Patz, J. A., 2020. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA Network Open, 3*(9). https://doi.org/10.1001/jamanetworkopen.2020.20485

Goldfarb, A., and Tucker, C., 2012. Shifts in privacy concerns. *American Economic Review: Papers and Proceedings, 102*(3), 349–353. https://doi.org/10.1257/aer.102.3.349

Guo, C., Vasquez-Parraga, A. A., and Wang, Y., 2006. An exploratory study of motives for Mexican nationals to shop in the US: More than meets the eye. *Journal of Retailing and Consumer Services, 13*(5), 351–362. https://doi.org/10.1016/j.jretconser.2005.11.002

Gurbuz, O., Aldrete, R. M., and Salgado, D., 2021a. *Border mobility and vaccination during a pandemic*. Center for International Intelligent Transportation Research. https://static.tti.tamu.edu/tti.tamu.edu/documents/185920-00015.pdf

Gurbuz, O., Aldrete, R. M., Salgado, D., and Vazquez, M., 2021b. *Contact tracing to maintain mobility at the border during a pandemic*. Center for International Intelligent Transportation Research. https://static.tti.tamu.edu/tti.tamu.edu/documents/185921-00008.pdf

Instituto Nacional de Estadística y Geografía (INEGI). 2020. Datos. https://www.inegi.org.mx/datos

INRIX. n.d. Products https://inrix.com/products. Accessed 8/4/2022.

Juhász, L., and Hochmair, H. H., 2020. *Studying spatial and temporal visitation patterns of points of interest using SafeGraph data in Florida*. Florida International University. https://doi.org/10.1553/giscience2020_01_s119

Killeen, B. D., Wu, J. Y., Shah, K., Zapaishchykova, A., Nikutta, P., Tamhane, A., Chakraborty, S., Wei, J., Gao, T., Thies, M., and Unberath, M., 2020. *A county-level dataset for informing the United States' response to COVID-19*. https://ciis.lcsr.jhu.edu/lib/exe/fetch.php?media=courses:456:2020:projects:456-2020-03:cis_final_report.pdf

Kulkarni, A., 2021. *Human mobility patterns linked to COVID-19 prone locations*. https://doi.org/10.7916/d8-1z8r-ns13

Lee, C. C., Maron, M., and Mostafavi, A., 2021. *Community-scale big data reveals disparate impacts of the Texas winter storm of 2021 and its managed power outage*. https://doi.org/10.48550/arXiv.2108.06046

Lee, K., and Sener, I. N., 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applications. *Transportation Research Interdisciplinary Perspectives, 4*, 100095. https://doi.org/10.1016/j.trip.2020.100095

Lee, R. J., Sener, I. N., and Mullins III, J. A., 2016. An evaluation of emerging data collection technologies for travel demand modeling: From research to practice. *Transportation Letters, 8*(4), 181–193. https://doi.org/10.1080/19427867.2015.1106787

Li, X., Huang, X., Li, D., and Xu, Y., 2022. Aggravated social segregation during the COVID-19 pandemic: Evidence from crowdsourced mobility data in twelve most populated US metropolitan areas. *Sustainable Cities and Society*, 81, 103869.

Prestby, T., App, J., Kang, Y., and Gao, S., 2020. Understanding neighborhood isolation through spatial interaction network analysis using location big data. *Environment and Planning A: Economy and Space, 52*(6), 1027–1031. https://doi.org/10.1177/0308518X19891911

Roy, A., and Kar, B., 2020. Characterizing the spread of COVID-19 from human mobility patterns and sociodemographic indicators. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities* (pp. 39–48). https://doi.org/10.1145/3423455.3430303

SafeGraph, n.d. Your partner in places data. Accessed 8/4/2022. https://www.safegraph.com/SafeGraph, 2020. SafeGraph provides CDC and 1000+ organizations with data to fight the COVID-19 crisis. Accessed 8/4/2022. https://www.safegraph.com/blog/safegraph-provides-cdc-fed-and-1000-organizations-with-data-to-fight-the-covid-19-crisis

SafeGraph, 2022. Places summary statistics. Accessed 10/16/2022. https://docs.safegraph.com/docs/places-summary-statistics

Sener, I. N., Lorenzini, K. M., and Aldrete, R. M., 2015. A synthesis on cross-border travel: Focus on El Paso, Texas, retail sales, and pedestrian travel. *Research in Transportation Business & Management, 16*, 102–111. https://doi.org/10.1016/j.rtbm.2015.05.002

Sharma, A., Farhadloo, M., Li, Y., Gupta, J., Kulkarni, A., and Shekhar, S., 2022. Understanding COVID-19 effects on mobility: A community-engaged approach. *AGILE: GIScience Series*, 3, 1-15.

Sullivan, P., Bonn, M. A., Bhardwaj, V., and DuPont, A., 2012. Mexican national cross-border shopping: Exploration of retail tourism. *Journal of Retailing and Consumer Services, 19*(6), 596–604. https://doi.org/10.1016/j.jretconser.2012.07.005

U.S. Census Bureau, 2020. QuickFacts. Accessed 8/4/2022. https://www.census.gov/quickfacts/fact/table/US/PST045219

Vargas, E., Gurbuz, O., Sener, I. N., and Aldrete, R. M., 2022. Examining pre- and post-pandemic cross-border trips using crowdsourced data at the second-busiest US-Mexico border community. *Findings*. https://doi.org/10.32866/001c.38429

Vargas, E., Sener, I. N., Gurbuz, O., Salgado, D., and Aldrete, R. M., 2021. *Exploration of cross-border trip characteristics using crowdsourced data*. Center for International Intelligent Transportation Research. http://tti.tamu.edu/documents/185921-00016.pdf

Yuan, J. J., Fowler, D. C., Goh, B. K., and Lauderdale, M. K., 2013. Mexican cross-border shoppers' motivations to the USA. *International Journal of Culture, Tourism and Hospitality Research, 7*(4), 394–410. https://doi.org/10.1108/IJCTHR-02-2013-0004

Zmud, J., and Sener, I. N., 2019. *Intersections: Business & leisure travel, mobility and the economic health of cities*. Technical report prepared for the Avis Budget Group.