

Unreported Crimes in Mexico: A Crowdsourcing Approach

By

José Alberto Dután

International Affairs Capstone

Vert 4-212

Professor Robert Smith

Baruch College

December 2023

Table of Contents:

I. Executive Summary.....	3
II. Introduction.....	4
<i>ENVIPE</i>	4-8
<i>IGI-MEX 2022</i>	8-10
III. Crowd-Source Approach Examples	10-11
IV. My Crowd-Source Approach.....	11-12
<i>First Approach</i>	11
<i>Second Approach</i>	12
V. Desired Outcomes.....	12-13
VI. Intended Purpose.....	13-14
VII. Crowdsourcing Data Disadvantages.....	14
VIII. Conclusion.....	15
IX. Bibliography.....	16-17
X. Appendix.....	18-25

Executive Summary:

Widespread crime in Mexico terrorizes its citizens, preventing them from living in a state with proper law and order. In order to help alleviate the situation, the Mexican government and academic researchers created innovative studies to track the phenomenon and report on possible causes. This led to the creation of the National Survey on Victimization and Perception of Public Safety (ENVIPE) and the Global Impunity Index for Mexico (IGI-MEX). While such studies are widely accepted, they still suffer from potential flaws and with the use of crowdsourced data it can potentially help confirm or improve the results from the studies.

The decision to employ a crowdsourcing approach using Twitter data to analyze crime in Mexico is grounded on my belief that Twitter has the potential to act as a database for crime records. Specifically, if a proper code is in place that can filter a desired set of tweets based on crime, then it is possible to extract and plot those tweets on a map. As a result, a high concentration of tweets in certain areas could indicate the level of criminal activity. While I was unable to properly produce that code, I do provide a solid framework on how others with greater resources can create a more comprehensive analysis of my approach.

Introduction

Mexico has for many years suffered insecurity at all levels of the government as corruption, crime, and violence accompanied with extremely high levels of impunity plague the nation. As a result, Mexican citizens are the principal victims of a crumbling criminal justice system. With limited, almost non-effective, federal, state, and municipal tools to properly report and track crime, official crime statistics do not reflect the reality of the situation. To uncover that reality, the Mexican government created the National Survey on Victimization and Perception of Public Safety (ENVIPE) to record the phenomenon of the Cifras Negras (Dark Figures), which is the underreporting of crime. Soon after, academic researchers used the cifras negras to help develop the Global Impunity Index for Mexico (IGI-MEX) to measure impunity levels across the country as well as providing insights on possible causes. While these alternative measures do provide a more accurate estimate, I suggest the implementation of a crowd-sourcing model as an additional effort to help analyze crime. I understand that crowd sourcing data is not perfect and has its limitations, but it can be useful to possibly record and predict real-time crime and public safety perception analyses. These outputs can be used in conjunction with IGI-MEX and ENVIPE, which also have their own faults, to make an overall better determination of the crime situation in Mexico.

The decision to use a crowd-sourcing model is based on the fact that Mexico does not have a proper crime recording database system, which distorts its crime statistics. However, crowd-sourced data from Twitter can help alleviate the problem by serving as a substitute platform. Twitter is an already well-established system, able to record massive amounts of text data that can be extracted and analyzed with a Natural Language Toolkit (NLTK) to filter tweets relating to criminal activities. The benefit is that it can assist in the underreporting of crime and add more certainty to official statistics by being another approach to verify them. Furthermore, Twitter is a private company headquartered in the United States (US) limiting possible Mexican government interference. Overall, I argue that using crowd-sourced data is a viable option for the Mexican government to produce more precise figures to study crime in the country.

National Survey on Victimization and Perception of Public Safety (ENVIPE)

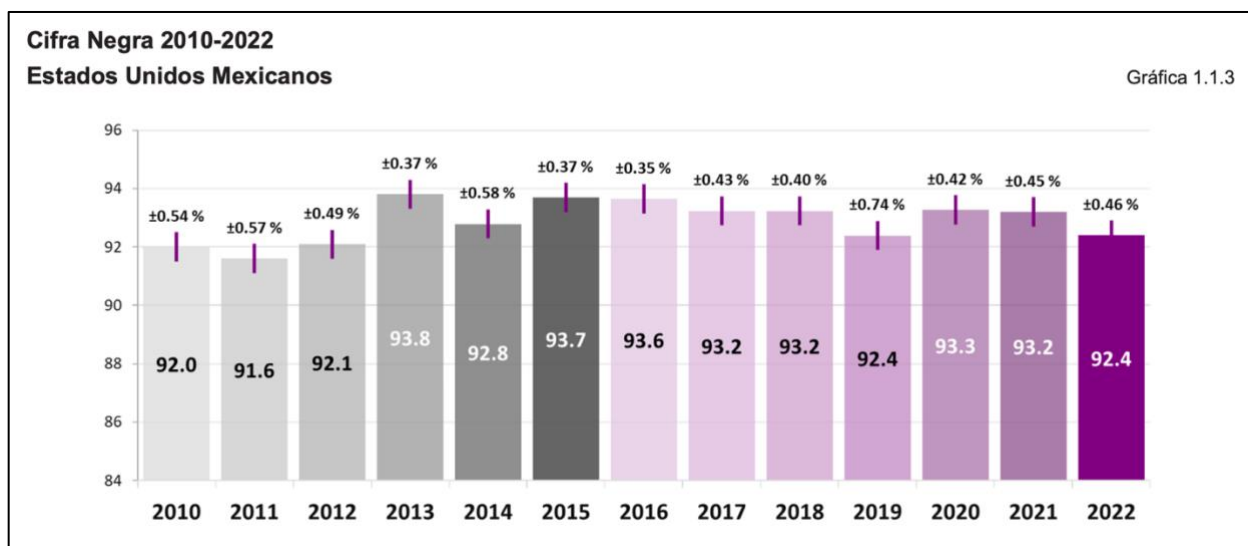
To help address the situation of crime in the country, the Mexican government utilized the National Institute of Statistics and Geography (INEGI) to create ENVIPE in 2010. To add credibility to the study, the framework for ENVIPE is built on decades worth of development in victimization and crime perception surveys conducted by the United Nations (UN) and US (ENVIPE Macro Conceptual, 2022: p3-7). These surveys aimed to collect unreported crimes or crimes not registered by law enforcement, including the public's view on public safety. While there is no exact standardization for these surveys as each country defines crime and crime perception differently, they lay solid foundations for survey design and implementation. For reliable and consistent data, ENVIPE annually conducts its surveys with face-to-face interviews, uses a substantial sample size of over 100,000 households spread across urban and rural areas, a 90 percent confidence level, employs a well-qualified and educated staff, and limits the maximum non-response rate to 15 percent (ENVIPE Informe Operativo, 2022: p9-10). As stated by ENVIPE, its objective is the following:

“To obtain information with representativeness at a national and state level to make estimates on the prevalence of crime that affected households in 2022, the crime incidence levels and the dark figure. Likewise, it seeks to obtain information on the perception of public safety and the performance of institutions in charge of public safety and justice. Information is also collected on the victimization context and economic and social impact of crime. This, in order to provide the general public with information and to generate elements for public policy decision-making in these matters.”

For the year 2022, ENVIPE achieved an effective response rate of 88.19 percent of completed interviews that produced the following results:

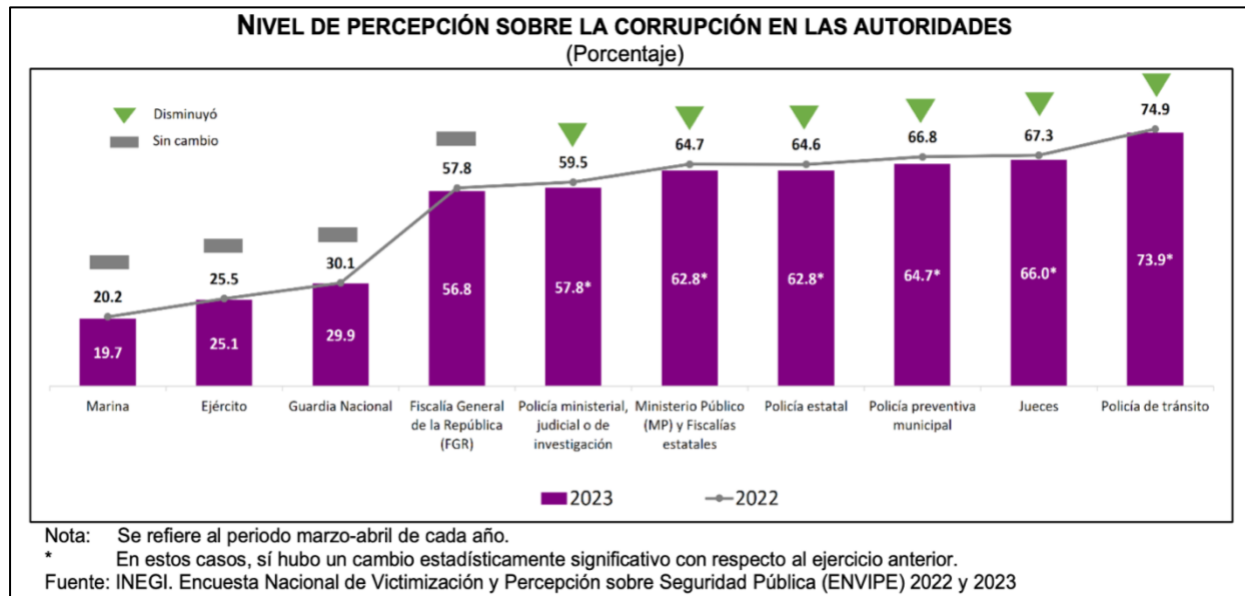
- In 2022, 27.4% of households in Mexico had at least one of their members as a victim of crime.
- In 2022 there were 21.1 million victims aged 18 and over. The crime prevalence rate was 22,587 victims per 100,000 inhabitants, a lower figure compared to 2021.
- The most frequent crimes were fraud, robbery, or assault on the street or public transport and extortion.
- In 2022, 26.8 million crimes occurred. Of these, 92.4% were not reported or the authority did not initiate an investigation file. This under-registration is called the black figure.

Out of these estimates, the most important figure is the cifras negras as shown in the last bullet point, which translates to 24.74 million unreported crimes. This number is calculated by considering the ratio of unreported crimes, including crimes not under investigation or lacking a case file, and cases where it is not specified as to whether a report was filed. This ratio is then divided by the total reported crimes per one hundred (ENVIPE Informe Operativo, 2022: p2). Basically, the cifras negras are the crimes not reported to the Public Ministry or State Prosecutor’s Office (ENVIPE Macro, 2022: p96). From the graph below, it is evident that the cifras negras presents a major and consistent problem for the state.

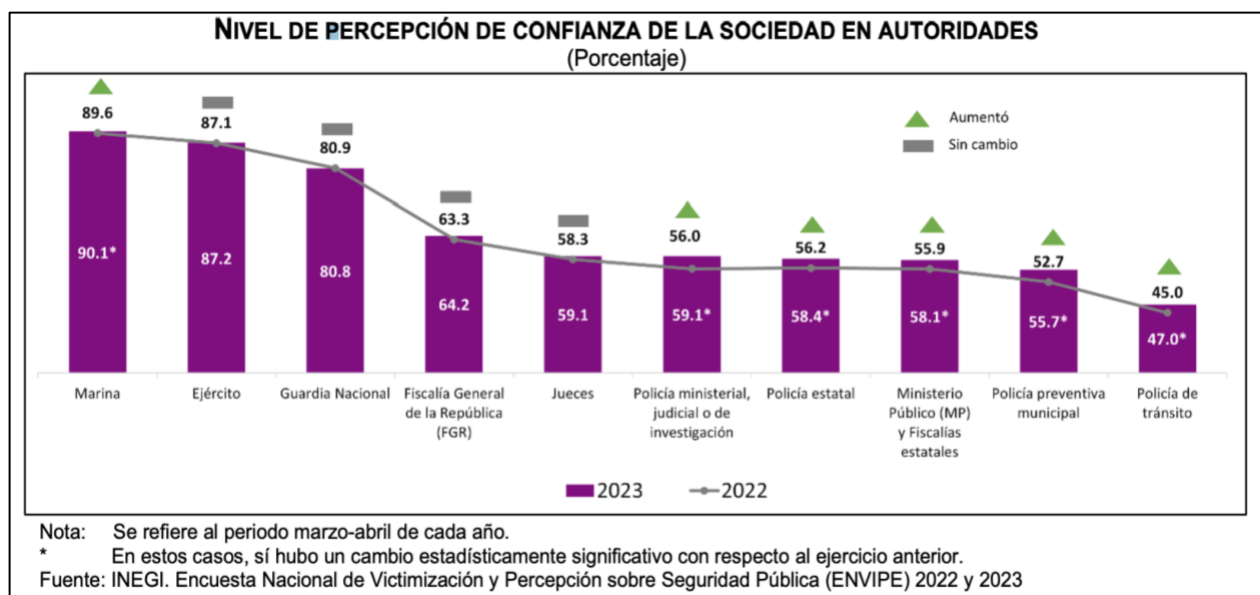


(ENVIPE Informe Operativo, 2022: p3)

This begs the question as to why crime is not reported, and based on ENVIPE it could be a result of the negative perceptions on civil authorities. From the first graph below, it shows that military components are perceived to be substantially less corrupt than the civil entities designed to ensure safety and justice in the community. These beliefs most likely translate to the level of confidence as demonstrated in the second graph below. For instance, the three military components with the lowest perception of corruption also have the highest perception of confidence from the public. The two graphs appear to show an inverse relationship with each other meaning that as perception of corruption increases, then the perception of confidence decreases.

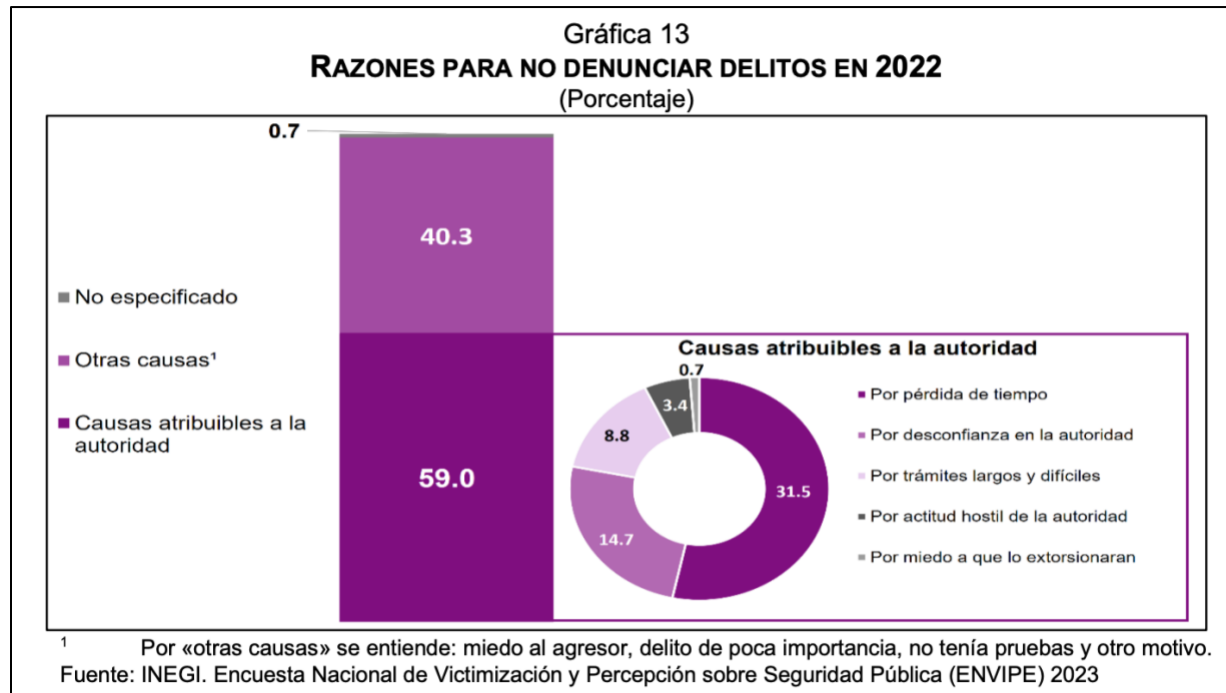


(ENVIPE PRENSA, 2023: p28)



(ENVIPE PRENSA, 2023: p27)

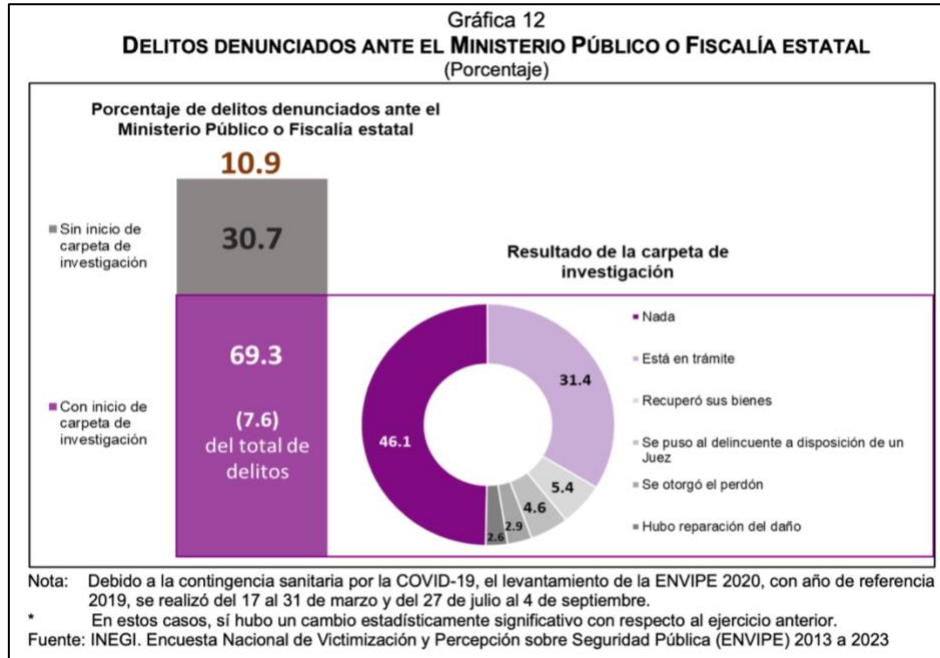
A more detailed analysis suggests that negative perceptions account for 14.7 percent of the reasons why people do not report crime. According to the graph below, other reasons are “A waste of time” (31.5%), “Due to long or difficult procedures” (8.8%), “Due to hostile attitudes from authorities” (3.4%), and “Due to fear of extortion” (0.7%). All of these percentages add up to a total of 59 percent indicating that over half of the reasons why people decide not to report crimes are causes related to the authorities.



(ENVIPE PRENSA, 2023: p27)

This lack of trust in government institutions is explored further by analyzing their low performance. Prime examples are the Public Ministry (Ministerio Público) and State Prosecutor’s Office (Fiscalía Estatal). The Public Ministry investigates and prosecutes criminal cases as it is the only entity authorized to exercise criminal action and request punishment/restitution on behalf of the state (ENVIPE Macro Conceptual, 2022: p99). Meanwhile, the State Prosecutor’s Office oversees criminal investigations and is responsible for providing the conditions for social peace in the state (ENVIPE Macro Conceptual, 2022: p97). These entities have a substantial involvement in people’s daily affairs since they are in charge of criminal matters in their respective jurisdictions. Unfortunately, as seen in the graph below, when crimes are reported to these systems, 30.7 percent of them never reach an open investigation, and for the 69.3 percent that do, the outcomes are not favorable. Crimes under open investigations face significant challenges as 46.1 percent result in “Nothing” and 31.4 percent are currently in “Progress”, which is a stage that can persist for any length of time. Such a lack of operational/functional ability severely taints the reputation of the justice system and law enforcement, two bodies that work closely together, and most likely leads to the positive views on the military, despite the crimes it commits as discussed in the appendix. To clarify, Mexican citizens do not report crimes to the Marines, Army, or National Guard because soldiers and their respective departments do

not have the investigative authority. Instead, the military fights crime and maintains order against DTOs.



(ENVIPE PRENSA 2023: p19)

Based on these statistics, it is apparent that the justice system does not effectively serve the public. The cifras negras exposes that a lot of the crimes are essentially lost, not reported, and a majority of those that are indeed recorded never reach a proper conclusion. In addition, crime statistics face substantial interference by the justice system itself. Despite, Mexico enacting changes to their criminal justice in 2016, the state still practices the notion of “guilty until proven innocent”. In the year 2020, about 85 percent of individuals incarcerated were never convicted of any crime (Human Rights Watch, 2023: p412). This is not surprising considering that judges and prosecutors can legally practice the use of “arraigo”, which gives them the ability to do pretrial detentions on individuals up to forty days without being formally charged (State Department Human Rights Report, 2022: p9). Lastly, there is substantial evidence that state actors from all levels of government often collude with criminals in covering up crimes and deter families from seeking retribution (Congressional Research Service (IN FOCUS), 2022: p2022). All of these actions both add more to the number of unreported crimes and manipulate official crime statistics.

Global Impunity Index for Mexico (IGI-MEX)

With the development of the cifras negras, Mexican researchers were able to create the Global Impunity Index for Mexico (IGI-MEX) as a means to better understand crime in the country. The index is based on the theoretical concept of impunity first developed by the UN Refugee Agency (UNHCR) which states:

“The index understands impunity, in this sense, as “the non-existence, whether in fact or in law, of criminal responsibility on the part of the authors of violations, as well as civil,

administrative, or disciplinary responsibility because they escape any investigation with a view to their indictment, detention, processing, and, if found guilty, appropriate sentencing, including compensation for the damage caused to their victims".

A higher impunity score indicates a higher level of repression due to its impact on a government's ability to provide its citizens the enjoyment of basic rights and accountability. To demonstrate this, researchers developed the "Chain of Impunity" consisting of three dimensions known as the structural, functional, and human rights dimension. Each dimension is made up of variables whose data comes from federal institutions that have been given substantial constitutional authority and legitimization to produce critical statistics for the development and evaluation of the country. These sources are the following:

1. National Census of State Justice Prosecution
2. National Census of State Law Enforcement
3. National Census of State Government, Public Security and Penitentiary System
4. Executive Secretariat of the Public Security National System
5. ENVIPE
6. National State Human Rights Census
7. National Federal Human Rights Census.

That being said, the impunity index is determined by calculating the average of the individual indexes across all three dimensions. For the year 2022, the national impunity index was 60.08 (IGI-MEX, 2022: p27), a fall from the previous index of 69.12. Researchers highlight that this decrease could be a result of the change in the methodology for the 2022 edition in order to account for new factors, but that the overall impunity indexes in each state did not change much from before. A deeper analysis on the dimensions demonstrates the causes for the high impunity index.

Structural Dimension (SD)

This SD is responsible for measuring the design and existence of institutional capacities for security policies, and is made up of two components, the Security System and Justice System. The former had an index of 70.93 and the latter had an index of 72.18. This gives the SD an overall index of 71.55 (IGI MEX, 2022: p37). See Figure 1 in the appendix.

Functional Dimension (FD)

The purpose of the FD is to measure the functionality of the components, which is the Security System and Justice System. The former had an index of 28.11 and the latter had an index of 35.25. This gives the FD an overall index of 31.68 (IGI MEX, 2022: p38). The researchers note this low index for FD does not necessarily mean that this dimension is effective for two possible reasons. First, the variables in FD have extreme outliers that could be balancing out the average index. Second, some of the variables get their data from registered crimes, not accounting the cifras negras, suggesting that the true effectiveness of authorities to investigate crimes is unknown as they will never have had the opportunity to work on unreported crimes. See Figure 2 in the appendix.

Human Rights Dimension (HRD)

The HRD measures the capabilities to address human rights violations and got an index of 62.16 (IGI MEX, 2022: p39). See Figure 3 in the appendix.

Cifras Negras

The cifras negras is not considered a dimension, but it is still part of the overall calculation for the impunity index. The figure is used in conjunction with the HRD. See Figure 3 in the appendix.

Findings

The study concludes that high impunity levels are prevalent across all Mexican states and that it is highly likely that the phenomenon relates to the deterioration of law and order. Researchers advocate for more investment in the structure and functionality of the security and justice system, including human rights.

Crowd-Source Approach Examples

My alternative measure is to use a crowd-sourcing model along with the use of Natural Language Processing (NLP) in order to capture the unreported crimes in Mexico. The rationale is that people will most likely report a crime on an online platform, such as Twitter, rather than go to the proper authorities. That being said, as social media platforms are registering millions of users daily, collecting vital information, the process of crowdsourcing data from online is increasingly becoming a more attractive method for data analysis and collection in finding interesting insights, patterns, and solutions to a wide array of challenges. One such challenge is the study of crime as presented in two studies.

The first study uses Twitter data to analyze crime in the ambient population, which refers to the number of people in a given outdoor area at a specific time, in Leeds, UK. The main purpose of this method is to improve crime rate cluster analysis. To clarify, crime rate reveals the “quantity and extent of criminal events”, and this calculation is usually derived from residential population data (Malleon and Andresen, 2015: p1). This information can be used to create a clustering map demonstrating the location of criminal activity. However, the main drawbacks are that on a granular level, analyses like that do not show you the progression of crime throughout the day or time in an exact location. For instance, I have attached a clustering map example from the mapping software Esri in the appendix called Figure 4. As you can see, the clustering map totals the given data points of certain features into a cluster whose size depends on the number of data points it contains. To improve this type of analysis, the researchers developed a spatio-temporal clustering method to detect significant clusters of events in space and time (Malleon and Andresen, 2015: p4). This method was used on reported crime data from the local police department and extracted Twitter data that contained GPS coordinates. The study was able to successfully estimate two clusters that showed the ambient population with spatio-temporal crime clustering. In other words, the researchers were able to show the density of the population in a certain area as well as the most frequent day and time a crime would occur (Malleon and Andresen, 2015: p7).

The second study uses data from the Place Pulse 2.0 app to analyze the spatial distribution of crowdsourced perceptions of space and safety in Atlanta. To get the perception of safety from users, the app presented two images in which the user was asked to select which image was “safer”. These images were chosen at random and covered 56 cities across 28 countries. With R software, the researcher extracted the app’s dataset and filtered images only from Atlanta, Georgia. By organizing the data, the researcher was able to put the geocoordinates into columns to successfully conduct a join that connected the images to their respective location on the map as seen in Figure 5. The researcher concluded that his safety perception map could be layered on top crime data as a means to possibly demonstrate how safety perception influences crime.

My Crowdsourc Overview

The purpose of discussing these two studies is to show that crowdsourcing data complemented with other analysis methods is a proven way to analyze crime. Similarly, my approach resembles the two studies in that I also use Twitter Data and clean it up for its use in mapping software. That being said, in the first approach I was unable to plot the data on a map due to difficulties in downloading the QGIS mapping software. Likewise, in the second approach, I confronted limitations in my coding abilities as I could not work out the proper code that would tally, or count, specific words like “violencia” and “asesinarion” in each row of the DataFrame. Despite my failures, I do provide a solid roadmap on how a researcher can possibly tackle a crowdsourcing approach with Twitter data to study crime in Mexico.

For my analysis, I conducted two approaches on how to use Twitter data due to API limitations that are discussed further in the `twitter_api_process` pdf file. To get started, I created a twitter developer account in order to gain access to their Basic Level API. This access will let you extract Twitter data.

First Approach

With the Basic Level API, I was able to filter tweets based on geolocation, however, these geo-located tweets had to be manually searched for on the actual twitter website using the advanced search settings. I only selected two tweets so I can easily demonstrate how one can extract and clean the data with Python. Once all the proper filters were selected in the Postman website, I ran the command to obtain my desired set of tweets. By default, the final output is in JSON file format. When this file is extracted into Python, I converted the file contents into a DataFrame to clean and organize the data. By doing so, I had a clean and organized table with columns labeled “text”, “date created”, and “geo coordinates”. After this, I converted the data into an excel file in order for it to be used in a map making software. If I were able to successfully download the QGIS mapping software, then the next step would have been to layer this file on top of a map of Mexico that contained relevant geo coordinates like latitude, longitude, state, and city. The last step would have been to simply do a table join on a common field between the data of the file and the data of the map. As a result, I would have had a point on a map of Mexico in whichever state the tweet originated from.

Second Approach

The Basic Level API was also used in this approach. For this one, I focused on demonstrating how to extract and clean text data for NLP analysis. Similar to the first approach, I also used the advanced search settings to create my desired twitter dataset. For each tweet I wanted to use, I copied its unique ID and added it into the Postman website. Once everything was successfully added, I ran the command. Again, the output was in JSON file format, and I extracted this into Python. To clean and reorganize the data, I had to put it into a DataFrame. When this part was done, I imported the NLTK library to show the steps how one would need to use NLP tools to properly conduct any sort of analysis. For instance, I tokenized each row of text into a list of words and then used the stop words method to get rid of all filler words. I also disregarded any tweets not in Spanish as these NLP methods, to my knowledge, can only be set to analyze one specific language at a time. At this stage, I was unable to construct the code to count specific target words relating to crime or violence. Also, I was unable to create the proper code to do sentiment analysis that would have shown the negative or positive rating for the way certain words were used in a context.

Desired Outcomes

Although the final result for both approaches can be viewed as unsuccessful attempts to analyze crime in Mexico, it was only unsuccessful due to my personal technical abilities and API restrictions. The approaches themselves are logical, sound, and most definitely possible to become another viable method to inspect the crime situation in Mexico. To clarify, in an ideal scenario, with no limitations or restrictions whatsoever, then I could have added all of my necessary inputs into the Postman website. For example, I could have instructed Postman to simply extract thousands of tweets, historical and current ones, that met the following criteria:

1. Tweets in the Spanish language.
2. Tweets from certain parts of Mexico. This is to compare crime statistics in states.
3. Tweets with specified targeted words relating to criminal activity.
4. Tweets with specified hashtags

**Examples of targeted words are the following: killed, kidnapped, stolen, drugs, gun shots, tortured, robbed, fraud, and the like.*

**All tweets must have geocoordinates to pinpoint in which region the tweet possibly originated from*

Again, these tweets would have been extracted almost simultaneously on Postman, not done manually, and most importantly not split up in two different approaches. Now, given a scenario where I successfully extracted 10,000 geo-located tweets based on the criteria above, then my next step would have been to combine my two approaches into one. The first step would be to review, clean, and organize my entire dataset into “text”, “date created, and “geo coordinates” columns. After that, my second step would be to use NLP techniques to tokenize each tweet and remove filler words. This would allow me to do sentiment analysis to determine if my specified targeted words in said tweet were associated with a more positive or negative context. Finally, a tweet that contains targeted words that were associated in a negative context would then be

plotted on a map of Mexico. With this, I could do a cluster map or a graduated map analysis as a way to show where there is a concentration of crime related tweets in Mexico. Keep in mind that my hypothesis in this is that a concentration of crime related tweets, specifically those expressed in a more negative context, could be representative of crime happening in that region.

In addition, for those who are more tech savvy, you are presented with more opportunities when it comes to using crowdsourced data. For example, my approaches and the approach I would have liked to do are essentially based on using primitive tools. In other words, an idea like mine that uses crowdsourced twitter data to help record or explain crime in Mexico would require the use of big data technologies, which is a skill I currently do not have. The idea behind big data technologies is that you can more easily handle a dataset that is in the millions, and this is extremely beneficial for predictive analytics. As a result, you can build a model to possibly see future outcomes or trends on criminal activities across Mexico. Also, by using big data tools it would then be possible to create a real-time dashboard meaning that you can filter, clean, run NLP analysis, and map the points basically simultaneously on its own.

Intended Purpose

I acknowledge that my approach is not perfect considering the limitations and flaws in crowdsourced data, however, it can serve as an additional layer to help add more validity to ENVIPE and IGI-MEX as both studies present a flaw. Firstly, is that the cifras negras figure produced by ENVIPE is not an accurate estimate of the true phenomenon of unreported crimes in Mexico. The reason is that it is extremely difficult and virtually impossible to garner the correct number of unreported crimes. Still, this does not discredit the efforts by ENVIPE, but in the end its cifras negras figure will most likely never be precise. Secondly, the inaccuracy of this figure presents a major problem because it influences two out of the three dimensions used by IGI-MEX to help calculate the impunity index for Mexico. For example, the Human Rights dimension uses the cifras negras figure from ENVIPE and because of the inaccuracy it is almost certain that the dimension score should be much higher. The same can be said for the Functional Dimension as it measures the functionality of the Security System and Justice System on reported crimes, not unreported crimes. This dimension scored the lowest with an overall index of 31.68. As previously mentioned, the researchers indicated that the low score does not necessarily translate to an overall effectiveness as the number could significantly change if those systems were presented with the handling of millions of unreported crimes. As a result, the functional dimension is probably much worse, thus increasing the impunity index for Mexico. In regard to my approach, these flaws can be softened if a more comprehensive analysis was done with NLP that calculated the number of tweets discussing about crimes in certain vicinities. The hope in this approach is that it can possibly record tweets talking about both reported and unreported crime in an area. If this approach created a model that was fully tested, trained, and executed then it can add confirmation or contradiction on results produced by ENVIPE and IGI-MEX. To clarify, if there is a significant number of tweets talking about crimes in a particular state, but that state was recorded by ENVIPE to have a low cifra negra and a low impunity index by IGI-MEX, then it is likely that there is inaccuracy in their data and that it should be revisited. Likewise, if a state is recorded to have a significant number of tweets talking about crime and is coincided with a high cifra negra and a high impunity index, then it is likely that the data is indeed accurate. In addition, with the right code a person can create a sentiment analysis model

that can determine whether a certain topic, like crime, is talked more on a positive or negative scale from 0 – 100, with zero being the most negative and a hundred being the most positive.

Crowdsourcing Data Disadvantages

While the use of crowdsourced data is increasing in popularity, it is important to recognize its imperfections. A notable example is that the data may not be representative of the entire population resulting in biases favoring specific demographics. For instance, in a forensic research case using crowdsourced data, participants in a two-part online survey exhibited an average age of 33.87. This highly skewed average was obtained despite an initial sample size of 1874 people, with ages ranging from 18 -78. This sample size was reduced by 69 percent after the first part of the online survey (Baker et al., 2016: p39-40). This massive drop in participation may be attributed to the fact that young people are more likely to contribute to websites (Malleon and Andresen, 2015: p3). Furthermore, evidence from the same study revealed that 68 percent of participants had at least some or a full college education background (Baker et al., 2016: p40). A separate study by the University of Nevada, which focused on representativeness of online surveys, discovered a disproportionate number of well-educated participants and a concentration of respondents residing in mid to high-level urban areas across all three of their online surveys (Heen and Meithe, 2014: p3-6). That being said, one can infer that individuals with an educated background might also happen to be more affluent members of society as backed by numerous studies linking education to wealth.

Given these strong notions, it is extremely probable that there is some sort of digital divide on the internet. Although, these studies were conducted in the US and not in Mexico, I argue that the findings from the studies can be easily applied to Mexican internet users. For example, it is almost universally agreed that education and wealth are highly correlated. Given that case, it is reasonable to assume that most college graduates would gravitate to areas that offer more opportunities, which are predominantly found in urban rather than rural areas. As for the young age in internet participants, this could be a result of the fact that the internet is also relatively young having gained popularity in the late 1990s to early 2000s. Such discoveries are extremely important when using crowdsourced data and is almost certain to be flaws in my approach. For instance, it is very likely that the tweets extracted from Twitter mainly come from young people, people with college education, and those in urban areas.

Another disadvantage of using crowdsourced data is that spam bots are prevalent across all social media platforms, especially Twitter. In my case, this runs the risk of data inaccuracy as an individual can easily create a spam bot and manipulate the data by mass posting false tweets about crime in Mexico. Moreover, crowdsourced data also suffers from user self-selection bias meaning that the data only records from people who voluntarily decide to participate online leaving an entire group of non-participants out of the pool (Buil-Guil and Solymosi, 2020: p3). Again, this bias would definitely play a role in my crowdsourcing approach. Lastly, I would also like to mention that the biggest obstacle in doing crowdsourced data is that it runs close to the lines of being unethical. This rings true especially if a government institution, like INEGI, were to engage in such activities as people may see it as an intrusion to their privacy.

Conclusion

Overall, I argue that the use of crowdsourced data is a valid approach to analyze crime in Mexico. Despite my shortcomings in my analyses, I do provide an effective framework on how my approaches can be done if given no limitations or API restrictions, and their effectiveness in helping to possibly verify results produced by ENVIPE and IGI-MEX. Still, it is important to acknowledge that crowdsourced data is still a relatively new concept, thus has significant challenges in the data. However, given adequate resources and time, I do believe that it is highly likely for a researcher to construct an accurate model to properly record crime and analyze it in Mexico. Lastly, crime in Mexico is such a complex and dense topic that simply reading reports about ENVIPE and IGI-MEX do not fully explain how the country reached such a tragic level of crime, so please feel free to review the appendix to read on the history of the approach of the Mexican government to fight drug trafficking organizations.

Bibliography

- Baker, Melissa & Fox, Paul & Wingrove, Twila. (2016). Crowdsourcing as a Forensic Psychology Research Tool. *American Journal of Forensic Psychology*. 34. 37-50.
- Beittel, J. S. (2022, June 7). Mexico: Organized Crime and Drug Trafficking Organizations. Congressional Research Service. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R41576/47>
- Buil-Gil, D., & Solymosi, R. (2023). Using crowdsourced data to study crime and place. In E. R. Groff, & C. P. Haberman (Eds.), *Understanding crime and place: A methods handbook* Temple University Press.
- Campos, M., Álvarez, L., & Cano, J. (2021, October 21). Seguridad Pública en el PEF 2022: más gasolina para la militarización. Retrieved from <https://www.mexicoevalua.org/seguridad-publica-en-el-pef-2022-mas-gasolina-para-la-militarizacion/>
- CIA World Factbook
- Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (2023). Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública 2023 : ENVIPE : informe operativo / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2023.
- Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (2023). Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública 2023 : ENVIPE : marco conceptual / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2023.
- Heen, Miliikeala & Lieberman, Joel & Miethe, Terance. (2014). A Comparison of Different Online Sampling Approaches for Generating National Samples. 10.13140/RG.2.2.24283.62243.
- House Foreign Affairs Committee & Tom Lantos Human Rights Commission. (2023, December 14). Organized Crime, Gangs, and Human Rights in Latin America: Hearing. Statement of Stephanie Brewer, Director for Mexico, Washington Office on Latin America (WOLA).
- House Foreign Affairs Committee & Tom Lantos Human Rights Commission. (2022, June 23). *Human Rights Challenges in Mexico*: Hearing. Statement submitted for the record by Daniella Burgi-Palomino, Co-director, Latin America Working Group (LAWG). Retrieved from <https://humanrightscommission.house.gov/sites/evo-subsites/humanrightscommission.house.gov/files/documents/LAWG%20Statement%20for%20the%20Record%20on%20Hearing%20on%20Human%20Rights%20Challenges%20in%20Mexico%206.30.22.pdf>
- Human Rights Watch. World Report. © 2023 by Human Rights Watch.

Instituto Nacional de Estadística y Geografía. (2023). *Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública 2023: ENVIPE:Prensa. Retrieved from https://en.www.inegi.org.mx/contenidos/saladeprensa/boletines/2023/ENVIPE/ENVIPE_23.pdf

Malleson, Nick & Andresen, Martin. (2015). Malleson, N., & Andresen, M.A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, 4, Article 10.. 4. Article 10.. 10.1186/s40163-015-0023-8.

Rodríguez Sánchez Lara, Gerardo & Chaidez, Azucena & Le Clercq, Juan. (2022). IGI-MEX 2022 Estructura y Función de la Impunidad en México.

Seelke, C. R., & Martin, R. L. (2022, July 25). Mexico: Addressing Missing and Disappeared Persons. Congressional Research Service IN FOCUS. Retrieved from <https://sgp.fas.org/crs/row/IF11669.pdf>

Seelke, C. R. (2023, May 17). Mexico: Background and U.S. Relations. Congressional Research Service. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R42917>

The World Justice Project. (2019). *Cuánta Tortura: Prevalencia de Violencia Ilegal en El Proceso Penal Mexicano 2006-2016*.

U.S. Department of State. (2023). *Mexico 2022 Human Rights Report*. Retrieved from https://www.state.gov/wp-content/uploads/2023/02/415610_MEXICO-2022-HUMAN-RIGHTS-REPORT.pdf

Appendix:

Figure 1

Dimensión	Sistema	Variable	Promedio
DIMENSIÓN ESTRUCTURAL	SISTEMA DE SEGURIDAD	Personal en las procuradurías por cada cien mil habitantes	44.43
		Agencias del Ministerio Público especializadas en delitos graves	2.02
		Agencias del Ministerio Público por cien mil habitantes	2.24
		Agentes del Ministerio Público por mil delitos registrados	8.58
		Policías judiciales por cada cien mil habitantes	0.86
		Personal destinado a funciones de seguridad pública por cada cien mil habitantes	83.35
	SISTEMA DE JUSTICIA	Magistrados y jueces por cada cien mil habitantes	4.36
		Personal total en Tribunal Superior de Justicia por cada cien mil habitantes	62.65
		Secretarios en el Tribunal Superior de Justicia por cien mil habitantes	10.29
		Personal penitenciario entre capacidad instalada	128.14
		Personal carcelario entre reclusos	0.25
		Personal de la Procuraduría General de Justicia o Fiscalía General	83.35
		Bienes inmuebles de la Procuraduría General de Justicia o Fiscalía General	2.32
		Cantidad de elementos ascendidos 2017-2019	8.61
		Presupuesto per cápita en procuración de justicia	14.29

Structural Dimension

Security System	Justice System
Staff in the attorney general's offices per one hundred thousand inhabitants	Magistrates and judges per one hundred thousand inhabitants
Specialized Public Prosecutor's Offices for Serious Crimes	Total personnel in the Supreme Court of Justice per one hundred thousand inhabitants
Public Prosecutor's Offices per one hundred thousand inhabitants	Secretaries in the Supreme Court of Justice per one hundred thousand inhabitants
Public Prosecutors per one thousand registered crimes	Penitentiary personnel among installed capacity
Judicial Police per one hundred thousand inhabitants	Prison personnel among inmates
Personnel assigned to public safety functions per one hundred thousand inhabitants	Personnel of the Attorney General's Office or General Prosecutor's Office

	Real estate properties of the Attorney General's Office or General Prosecutor's Office
	Number of promoted personnel 2017-2019
	Per capita budget in law enforcement

Figure 2

Dimensión	Sistema	Variable	Promedio
DIMENSIÓN FUNCIONAL	SISTEMA DE SEGURIDAD	Presuntos delitos registrados por cien mil habitantes	1,432.44
		Porcentaje de encarcelados por homicidio entre homicidios en averiguaciones previas iniciadas	39.41
		Porcentaje de encarcelados por robo entre averiguaciones previas iniciadas	67.53
	SISTEMA DE JUSTICIA	Causas penales en primera instancia entre carpetas de investigación totales	1.34
		Sentenciados entre ingresos penitenciarios	1.06
		Porcentaje de reclusos sin sentencia en primera instancia	0.39
		Víctimas registradas en las averiguaciones previas, investigaciones y carpetas de investigación abiertas por el Ministerio Público	62,549.59
		Causas penales y expedientes ingresados entre Estatal - Delitos - 2015 - 2021	39.37

Functional Dimension

Security System	Justice System
Presumed crimes registered per one hundred thousand inhabitants	Criminal cases in the first instance out of total investigation files
Percentage of incarcerated individuals for homicide out of homicide cases in initiated investigations	Sentenced individuals out of penitentiary admissions
Percentage of incarcerated individuals for robbery out of initiated preliminary investigations	Percentage of inmates without a sentence in the first instance
	Victims recorded in preliminary investigations, inquiries, and investigation files opened by the Public Prosecutor's Office
	Criminal cases and files entered between State - Crimes - 2015 - 2021

Figure 3

DERECHOS HUMANOS	Oficinas del Organismo de Protección de Derechos Humanos, por nivel de gobierno y entidad federativa según tipo	0.26
	Solicitudes de queja recibidas por el Organismo de Protección de Derechos Humanos	103.43
	Servicios de atención inmediata realizados por el Organismo de Protección de Derechos Humanos	260.66
	Expedientes de queja atendidos por el Organismo de Protección de Derechos Humanos	71.83
	Expedientes de queja calificados concluidos por el Organismo de Protección de Derechos Humanos	41.32
	Personal de las áreas operativas del Organismo de Protección de Derechos Humanos	0.06
	Personal de las visitadurías del Organismo de Protección de Derechos Humanos, por nivel de gobierno, entidad federativa y tipo de visitaduría según cargo y sexo 2019	1.37
	Titulares de las visitadurías del Organismo de Protección de Derechos Humanos, por nivel de gobierno y entidad federativa según sexo 2019	0.27
CIFRA NEGRA	Expedientes de queja atendidos entre solicitudes de queja recibidas	1.06
	Cifra negra	91.88

Human Rights

Offices of the Human Rights Protection Organization, by level of government and federal entity according to type
Complaint requests received by the Human Rights Protection Agency
Immediate actions services conducted by the Human Rights Protection Agency
Complaint cases handled by the Human Rights Protection Agency
Qualified complaint cases concluded by the Human Rights Protection Agency
Personnel in the operational areas of the Human Rights Protection Organization
Personnel in the branches of the Human Rights Protection Agency, by level of government, federal entity, and type of branch according to position and gender 2019
Heads of the branches of the Human Rights Protection Organization, by level of government and federal entity according to gender 2019
Complaint cases handled between complaint requests received
The Dark Figure (Unreported Crimes)

Legend

UK Supermarkets 2021 Q2

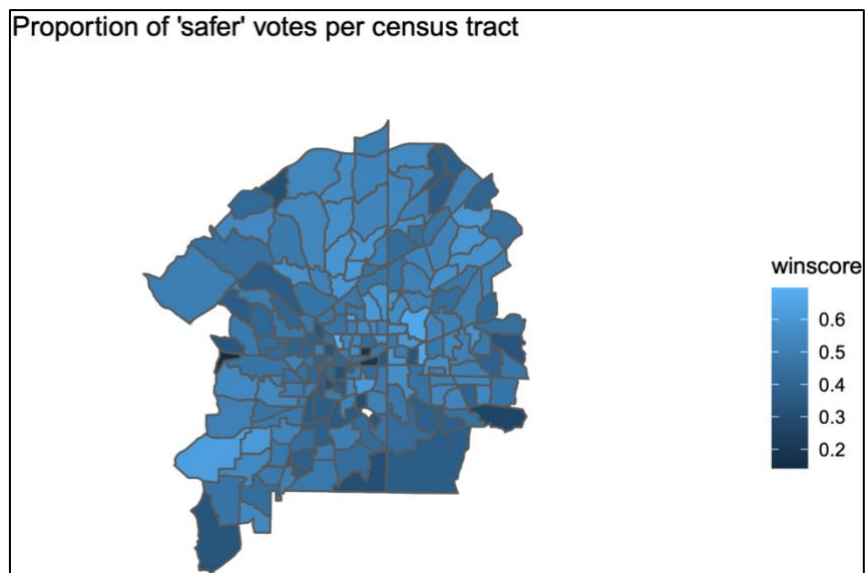
retailer

- Tesco
- The Co-operative Group
- Spar
- Sainsbury's
- Marks and Spencer
- Iceland
- Aldi
- Lidl
- Morrisons
- Asda
- Other

Number of features

- > 2,210
- 1,700
- 1,000
- 600
- < 2

Figure 5



21

Overview of Mexican Public Safety Policy and Effects:

There is a general consensus by Mexican, American, and international officials that the increasing militarization of public safety is a major obstacle to the stability of Mexico. To clarify, Mexico has had a long history of being a major transit route of heroin, marijuana, and synthetic drugs destined for the US, resulting in major problems for both countries. In order to combat the growing presence of powerful drug cartels and drug trafficking organizations (DTOs) in Mexico as well as the flow of drugs in the US, both countries developed the Mérida Initiative in 2007. During the next three years, Mexico received about 1.3 billion dollars, as seen in the chart below, in funds that were directed to invest heavily in Mexico's federal security forces and military leading the government to embrace a war-on-crime approach (Brewer, 2022: p6).

Table 1. FY2008-FY2010 Mérida Funding for Mexico by Aid Account						
(\$ in millions)						
Account	FY2008 Supp. (P.L. 110- 252)	FY2009 Bridge (P.L. 110- 252)	FY2009 (P.L. 111- 8)	FY2009 Supp. (P.L. 111- 32)	FY2010 P.L. 111- 117)	Account Totals
ESF	20.0	0.0	15.0	0.0	15.0	50.0
INCLE	215.5	48.0	246.0	160.0	190.0	859.5
FMF	116.5	0.0	39.0	260.0	5.3	420.8
Total	352.0	48.0	300.0	420.0	210.3	1,330.3

Sources: U.S. Department of State, FY2008 Supplemental Appropriations Spending Plan, FY2009 Appropriations Spending Plan, and FY2009 Supplemental Spending Plan. FY2010 Consolidated Appropriations Act (P.L. 111-117).

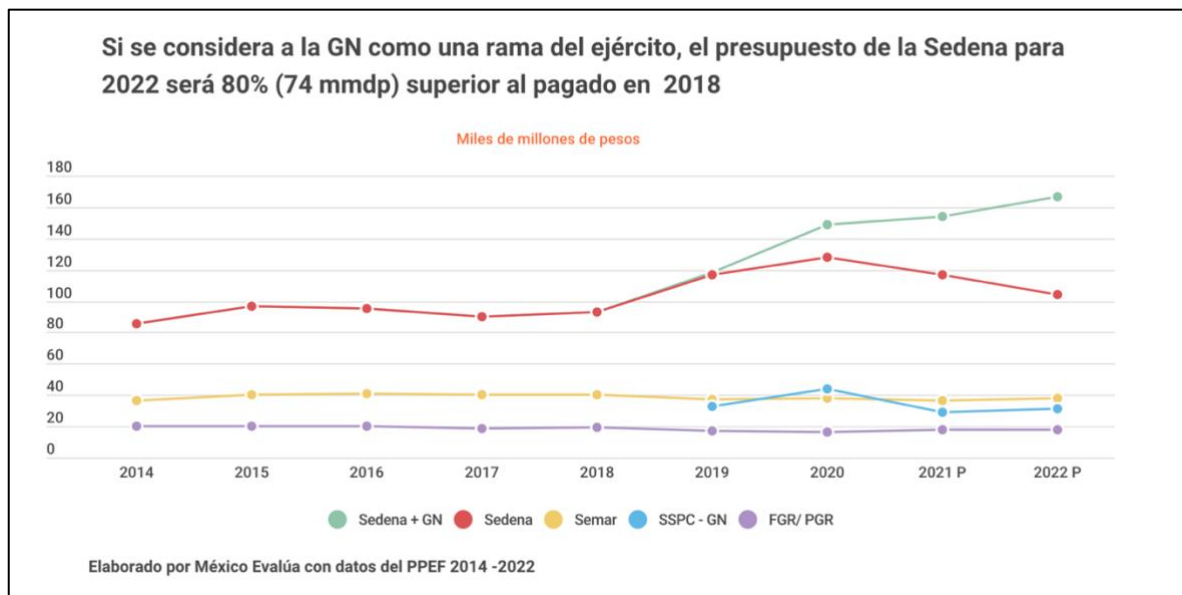
Notes: ESF=Economic Support Fund; FMF=Foreign Military Financing; INCLE=International Narcotics Control and Law Enforcement.

(Congressional Research Service, 2010: p6)

Although, the initiative soon changed its approach after 2010, Mexico on its own has continued a militaristic approach to combat DTOs and manage crime. Currently, the Mexican military is responsible for providing internal security, disaster response, humanitarian assistance, socio-economic development and defending the state (CIA World FactBook). The armed forces expansion in providing public security was furthered by the disbandment of the militarized federal police force in December 2019, resulting in its successor, the National Guard (GN). On September 2022, the GN was transferred over to the Secretariat of National Defense (SEDNA), and overwhelmingly consisted of military police units from the Army and Navy, including former federal police officers.

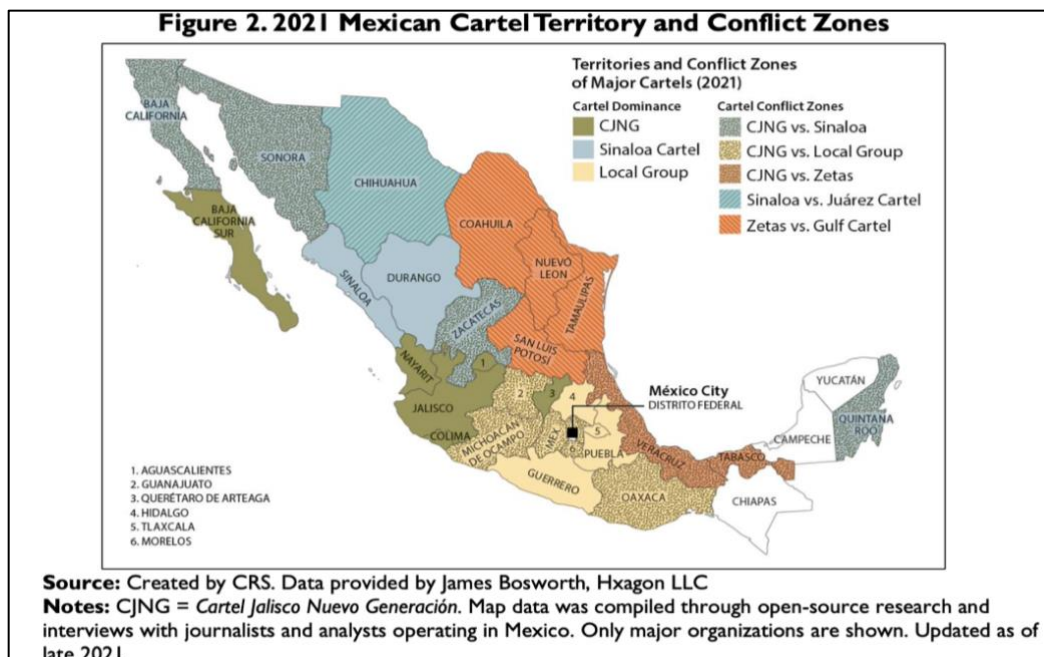
The expansion of the armed forces also coincided with greater interference in political affairs. For example, in the 2023 Tom Lantos Human Rights Commission, a bipartisan commission established by the US House of Representatives, the Director for Mexico and Migrants Rights expressed grave concerns of the Mexican government adding military officers to civilian positions. This practice is even supported by the Mexican President, López Obrador, who told incoming governors to consult with the heads of SEDENA and SEMAR, the Secretariat of the Navy, before appointing any state public security ministers.

With a military presence across the country, its budget has grown at an alarming rate even amid austerity measures for the very institutions that are in charge of civil justice. For instance, a 2021 report by the independent Mexican think tank, México Evalúa, exposes how the federal budget for the National Guard in 2021 was 62.8 billion pesos, which was a seventy percent increase in the original allotted amount of 26 billion pesos. This massive increase was agreed upon by the government even though funds for local public security remained almost unchanged at 8 billion pesos (México Evalúa, 2021).



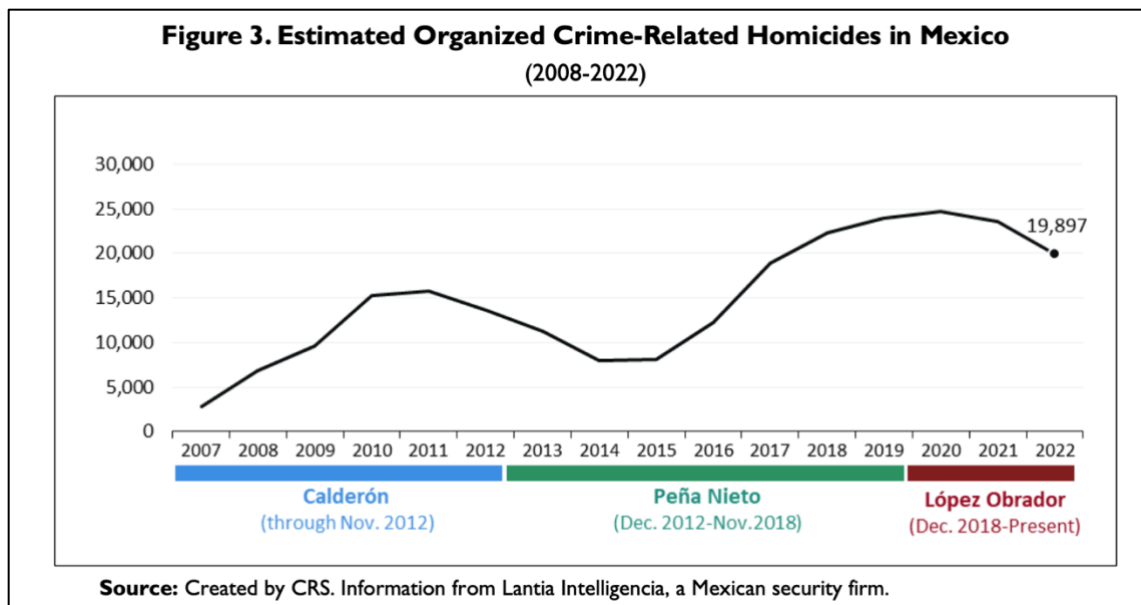
Even though, the chart was made in 2021, it exposes the sheer disparity in the distribution of federal funds between military and civilian entities. Civilian entities like the Attorney General's Office (FGR/PGR) and SSPC do not even come close to receive funds like SEDENA despite increasing crime and mass corruption. As a matter of fact, if the GN became an extension of the military, which it did, then in the year 2022 SEDENA and SEMAR would effectively control 149 billion pesos (México Evalúa, 2021).

Following the years of the militarization in public security, Mexico has and continues to experience a severe deterioration of law and order. prior to 2005, there were four dominant and relatively stable DTOs, but after the implementation of the Mérida Initiative they quickly fragmented between 2008 - 2010 into nine major groups along with hundreds of smaller crime factions and mafias (Congressional Research Service, 2022: p10). While there is evidence that these fragmentations were inevitable due to infighting and competition, many analysts agree that government operations targeting the kingpins of DTOs increased instability and greater violence (Congressional Research Service, 2022: p2-22).



(Congressional Research Service, 2022: p10)

The map above shows the extent of conflict zones that span the country. Since the fragmentation, homicide killings have more than tripled and disappearances are growing at an exponential rate. According to the Human Rights Watch, the homicide rate reached 28 per 100,000 people in the year 2021. That same year had estimates of over 52,000 unidentified bodies found largely in mass graves or cemeteries (Latin America Working Group, 2022: p2). Whereas in 2022, official cases of disappearances exceeded well over 100,000 with about ninety percent reported to have taken place since 2007 (Congressional Research Service, 2022: p11). From that same number, a quarter of the people are women and girls and tragically 16,000 are children. The chart below



(Congressional Research Service, 2023: p5)

demonstrates how much carnage happens in Mexico, and even more worrisome, is that this graph does not fully capture the extent of organized crime-related homicides since crime is severely underreported. Not to mention, it excludes the true fate of the individuals officially classified as disappeared. In other words, if half of the 100,000 disappearance cases, a number that is reported to be much higher as mentioned previously, ends up as an organized crime-related homicide then that would be 50,000. If that 50,000 were to be dispersed from 2008 to 2022, then it means that each year would be off by 3,571 homicides. Likewise, that number does not include the 52,000 unidentified bodies found in unmarked graves.

Similar to DTOs and other smaller criminal gangs, the armed forces also commit atrocities against the Mexican population. Since the military deployment to fight crime in 2007, the army has been responsible for the killing of 5,335 civilians while also using the spyware Pegasus, a controversial Israeli phone hacking software used to infiltrate iOS and Android devices, against human rights defenders, journalists, and opposition party members (World Report, 2023: p412–413). There have been numerous complaints filed against the military due to their incessive involvement in torture and arbitrary detentions of innocent people. The World Justice Project calculated that 88 percent of people detained by the navy and 85 percent of people detained by the army, from 2006 to 2016, experienced either torture or maltreatment. Within that same year, it was discovered that 41 percent of women detained by the army and 21 percent of women detained by the navy reported being raped while in detention (World Justice Project, 2019, p31-49). Unfortunately, leaks from the government, such as the “Guacamaya Leaks”, highly suggest that senior military officials have willingly thwarted investigations into crimes committed by their soldiers.