

Mechanical and Industrial Engineering Department IE 6600: Computation and Visualization. Spring 2025

# **Project 2: Border Crossing Entry Data**

Group 2

Members: Praneeta Mahentesh, Bhavya Kapadia, Soumya Gosala

#### Introduction:

The Border Crossing Entry Data dataset provides comprehensive insights into cross-border movements at U.S.-Canada and U.S.-Mexico land border ports. This dataset, collected by U.S. Customs and Border Protection (CBP), records various modes of transportation such as trucks, trains, buses, personal vehicles, passengers, and pedestrians crossing the border. The dataset is widely used for transportation analysis, border security assessments, trade policy formulation, and infrastructure planning.

# **Objective**

The primary objective of this study is to analyze border crossing trends over time, identify key locations with the highest volume of crossings, and examine seasonal and mode-specific variations. Using exploratory data analysis (EDA) and visualization techniques, we aim to:

- 1. Understand temporal trends in cross-border movements.
- 2. Identify the busiest ports and states for border crossings.
- 3. Analyze the distribution of crossings across different transportation modes.
- 4. Provide insights that can inform policy decisions on trade, transportation, and border management.

# **Data Processing and Analysis**

### 1. Data Acquisition:

The "Border\_Crossing\_Entry\_Data.csv" (<u>Link to dataset</u>) dataset, consists of 398,658 records and contains 10 fields, each representing critical attributes related to border crossings. Below is a summary of the data fields and their respective data types:

Column Name	Data Type	Description
Port Name	object	The name of the port of entry where the crossing occurred.
State	object	The U.S. state in which the port is located.
Port Code	int64	A unique numerical identifier for the port of entry.
Border	object	Specifies whether the crossing occurred at the U.SCanada or U.SMexico border.
Date	object	The date when the border crossing was recorded.
Measure	object	The mode of transportation used for crossing (e.g., trucks, buses, pedestrians).

Value	int64	The number of border crossings recorded for that entry.
Latitude	float64	The geographical latitude of the port.
Longitude	float64	The geographical longitude of the port.
Point	object	A combined representation of latitude and longitude in geographic coordinate format.

### 2. Data Inspection and Cleaning:

After obtaining the dataset, we carefully examined it to ensure that the data was clean, accurate, and ready for analysis. This process involved checking for missing values, duplicate records, and making necessary corrections to the data format.

### 2.1 Inspecting the Data

We first looked at the dataset's structure to understand what kind of data we were working with. This included checking:

- The number of rows and columns.
- The type of information stored in each column (e.g., numbers, text, dates).
- Whether any values were missing.
- If there were any duplicate entries.
- The dataset contained 398,658 rows and 10 columns with information about border crossings, including the port name, state, date, type of crossing (trucks, buses, pedestrians, etc.), and the number of crossings.
- One issue we found was that the date column was stored as text instead of an actual date format, which would make it harder to analyze trends over time.

# 2.2 Cleaning the Data

To ensure the data was reliable, we made several improvements:

- Fixing the Date Format: The "Date" column was originally stored as text (e.g., "Jan 2024"), so we converted it into a proper date format. This allows us to analyze trends over time more easily.
- Checking for Missing Values: We checked each column to see if any data was missing.
   Fortunately, there were no missing values, meaning we could use the dataset without needing to fill in any gaps.
- Removing Duplicate Entries: We found 10 duplicate rows, which we removed to ensure that every border crossing record was unique.
- Optimizing Data Storage: Some columns, like "Port Name," "State," "Border," and "Measure," were stored as plain text. Since these columns contain repeated categories (e.g., the same state name appearing many times), we converted them into a more efficient format, which reduced the dataset's size and made it faster to process.

#### 2.3 Final Cleaned Dataset

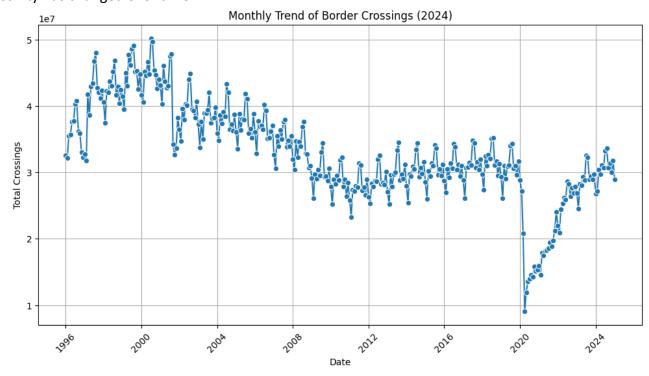
After cleaning, the dataset had 398,648 rows (after removing duplicates) and was well-structured:

- The date column was properly formatted.
- There were no missing values.
- The data was optimized for analysis.

## **Exploratory Data Analysis (EDA)**

### 1. Monthly Trend of Border Crossings (2024)

The line chart below shows the monthly trend of border crossings from the mid-1990s to 2024. Each point represents the total number of crossings for a given month, providing a clear view of how border activity has changed over time.



#### **Key Observations:**

- Growth in Crossings (1996 Early 2000s): The number of border crossings steadily increased, reaching a peak around 2000-2001. This could be attributed to economic growth, increased trade activities, and rising travel between the U.S. and its neighboring countries.
- Fluctuations and Slight Decline (Mid-2000s Late 2010s): After the peak, crossings showed some fluctuations but remained relatively stable until 2008-2010. A slight downward trend can be seen in the late 2010s, possibly due to policy changes, tighter border regulations, or economic shifts.
- Sharp Decline in 2020: A significant drop occurred in early 2020, which aligns with the COVID-19 pandemic. This period saw worldwide travel restrictions, temporary border closures, and

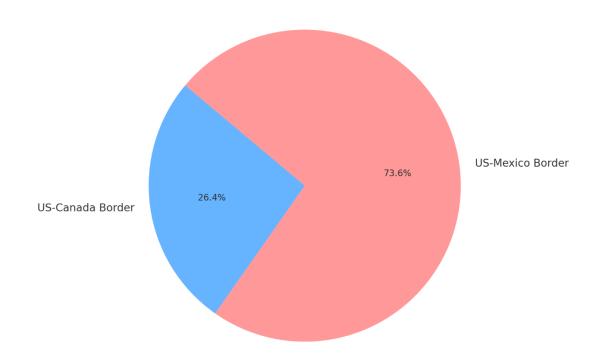
#### IE6600

- decreased cross-border trade and commuting.
- Recovery and Gradual Increase (2021 2024): post-pandemic, crossings gradually rebounded, indicating a return to normal travel and trade activities. However, the recovery is not yet at prepandemic levels, suggesting that border policies, economic changes, or altered travel behavior may still be influencing movement.

This graph highlights long-term trends in border activity, with notable impacts from economic and global events like 9/11 (2001), the 2008 financial crisis, and the COVID-19 pandemic (2020). Understanding these patterns can help policymakers and businesses prepare for future disruptions and optimize border management strategies.

# **Proportion of Border Crossings by Border Type (2024)**

Proportion of Border Crossings by Border Type (2024)



The pie chart above represents the distribution of border crossings between the U.S.-Canada and U.S.-Mexico borders in 2024.

#### **Key Observations:**

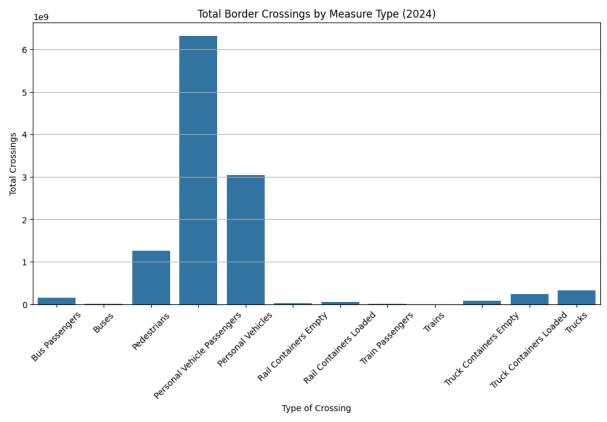
- US-Mexico Border Dominance: The majority of border crossings (73.6%) occur at the U.S.-Mexico border. This reflects the high level of trade, travel, and migration between the U.S. and Mexico, particularly in states like Texas, California, and Arizona.
- US-Canada Border Share: The U.S.-Canada border accounts for 26.4% of total crossings.
   Although lower than the U.S.-Mexico crossings, it still represents a significant volume of movement, largely driven by commercial trade and tourism between the U.S. and Canada.
- Possible Explanations for the Difference: The longer and more active commercial routes along the U.S.-Mexico border contribute to the higher crossings. Geopolitical factors, including stricter immigration policies and trade agreements, influence the border activity. Weather conditions also play a role, as the northern border experiences seasonal fluctuations in crossings, whereas the southern border sees more consistent year-round activity.

Higher suicide rates among elderly populations may be attributed to factors such as social isolation, health deterioration, and financial stress, making them particularly vulnerable. The gradual increase in suicide rates from adolescence to middle adulthood underscores the need for targeted mental health

interventions that support young and middle-aged adults during critical life transitions. Additionally, the sharp rise in suicide rates among individuals aged 75 and older highlights the urgent need for improved elderly care programs, accessible mental health services, and awareness initiatives to address the unique challenges faced by aging populations.

### 2. Total Border Crossings by Measure Type (2024)

The bar chart below displays the total number of border crossings in 2024, categorized by the type of crossing. Each bar represents a different mode of transportation or travel method used to cross the U.S.-Canada and U.S.-Mexico borders.



#### **Key Observations**

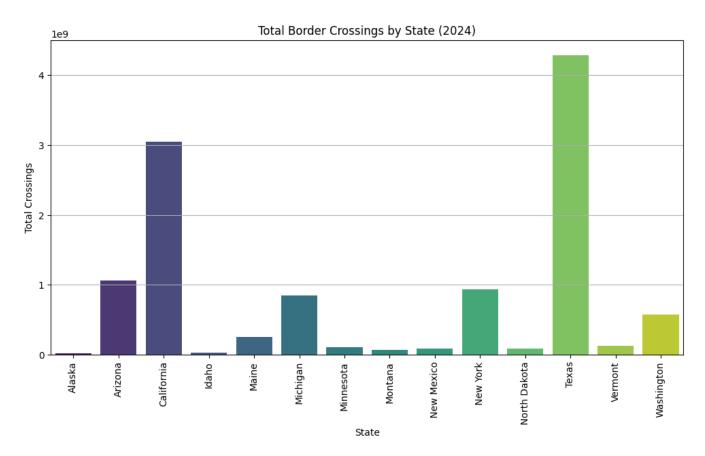
- Dominance of Personal Vehicles and Passengers: The largest category is personal vehicle
  passengers, accounting for the highest number of crossings. This is followed by personal
  vehicles, highlighting the significance of individual travel for work, tourism, and trade.
- Pedestrian Crossings as a Major Contributor: Pedestrian crossings also represent a notable portion, particularly at busy urban border crossings like those between California and Mexico.
- Lower Volumes for Commercial Transport: While trucks, rail containers, and buses contribute to cross-border traffic, their numbers are significantly lower than personal vehicle crossings.
- Truck crossings, including both loaded and empty containers, show moderate levels of activity, which reflects commercial trade movement.

Minimal Train and Bus Passenger Traffic: Train passengers and buses have the lowest number
of crossings, suggesting that these modes of transport are less frequently used for cross-border
travel compared to personal vehicles and trucks.

This visualization highlights the overwhelming reliance on personal vehicles for cross-border movement, with commercial transport playing a supporting role. The findings suggest that policies aimed at improving border efficiency should prioritize vehicle crossings while also considering infrastructure enhancements for pedestrian and commercial transportation.

### 3. Total Border crossings by State

The bar chart below illustrates the total number of border crossings by U.S. states in 2024, providing insight into which states experience the highest volume of cross-border activity.



#### **Key Observations**

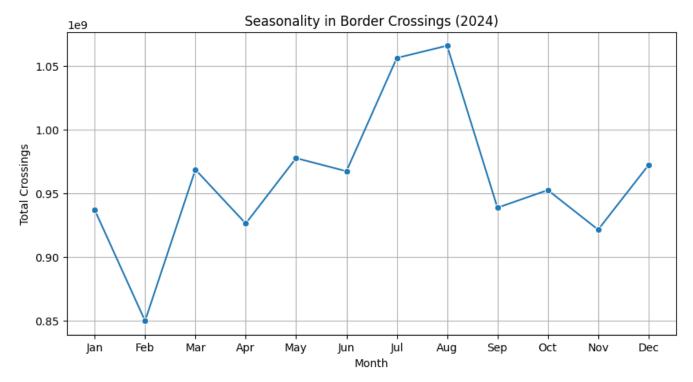
Texas as the Leading State: Texas records the highest number of border crossings, surpassing all
other states by a significant margin. This reflects the state's extensive border with Mexico and
major ports of entry such as Laredo, El Paso, and Brownsville, which facilitate both commercial
trade and personal travel.

- California and Arizona as Major Contributors: California ranks second, driven by major border crossings like San Ysidro and Otay Mesa, which experience high volumes of personal and commercial vehicle traffic. Arizona follows closely behind, with key crossings such as Nogales contributing to its substantial border activity.
- Michigan and New York Leading in the North: Michigan and New York stand out among U.S.-Canada border states, reflecting heavy cross-border trade and travel between the U.S. and Canada.
- Michigan's Detroit-Windsor corridor and New York's Buffalo-Niagara Falls crossing are among the busiest on the northern border.
- Lower Activity in Other States: States like North Dakota, Maine, and Washington show moderate border crossing volumes, largely influenced by trade and regional travel.
- Idaho, Montana, and Vermont have lower crossing numbers, likely due to their smaller population centers and fewer major border ports.

This visualization highlights Texas as the dominant state for border crossings, followed by California and Arizona, emphasizing the importance of the U.S.-Mexico border in overall cross-border movement. The significant activity at the U.S.-Canada border in Michigan and New York further reinforces the role of these states in trade and transportation between the U.S. and Canada. Understanding these patterns can help policymakers optimize border infrastructure, security, and trade facilitation efforts.

### 4. Seasonality in border crossings

The line chart above represents the **monthly variations in border crossings** for 2024, highlighting seasonal trends and fluctuations in cross-border movement.



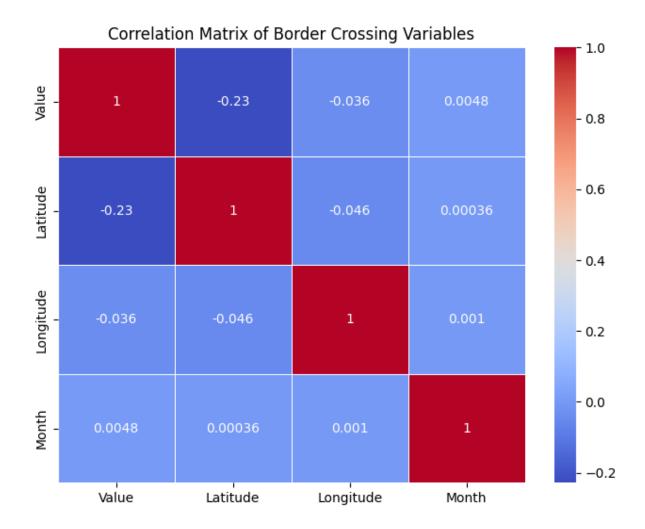
#### **Key Observations**

- Winter Decline (January February): A noticeable drop in crossings occurs in February, which could be attributed to colder weather conditions, reduced tourism, and fewer travel activities during this period.
- Spring and Early Summer Increase (March June): Crossings begin to increase in March, likely
  due to spring break travel, improved weather, and rising commercial activity. This trend
  continues into the summer months, with steady growth in May and June, reflecting higher
  travel demand.
- Peak Travel in Late Summer (July August): The highest number of crossings is observed in July and August, indicating increased tourism, vacation travel, and trade activities. This period aligns with summer holidays and seasonal employment opportunities, leading to a surge in movement across borders.
- Fall Decline (September November): Crossings drop significantly in September, likely due to the end of summer vacations and the return to regular work and school schedules.
- October and November show moderate activity but remain lower compared to peak summer months.
- Slight Rebound in December: Crossings increase again in December, which can be linked to holiday travel, festive season tourism, and cross-border shopping.

This visualization highlights clear seasonal patterns in border crossings, with peak activity in summer and lower volumes in winter. These trends are influenced by weather conditions, holiday seasons, and economic activities. Understanding these fluctuations helps in resource planning, staffing adjustments at border checkpoints, and policy-making for smoother cross-border operations.

#### 5. Correlation Matrix of Border Crossing Variables

The heatmap above displays the correlation matrix for selected numerical variables in the border crossing dataset, showing how different factors relate to each other.



#### **Key Observations:**

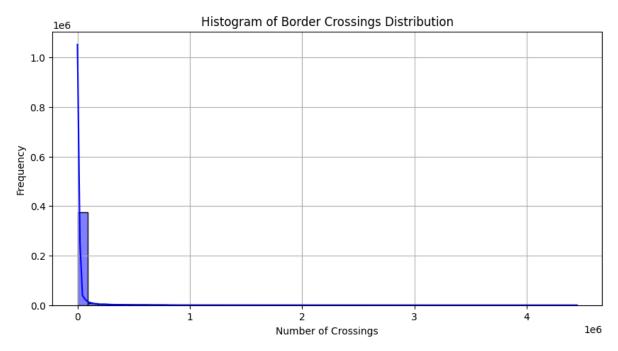
- Value vs. Latitude: A negative correlation (-0.23) suggests that border crossings tend to decrease as latitude increases. This indicates that crossings are more frequent at lower latitudes, aligning with the high activity along the U.S.-Mexico border compared to the U.S.-Canada border.
- Value vs. Longitude: A very weak correlation (-0.036) suggests that longitude has little impact on the number of crossings. This is expected as major crossing points are spread across both east-west directions of the U.S.
- Value vs. Month: The correlation (0.0048) is near zero, indicating that seasonality does not strongly influence total crossings across all months. However, this does not rule out seasonal variations for specific transportation modes, which may require further analysis.
- Latitude vs. Longitude: A very weak correlation (-0.046) indicates no significant relationship between these geographic variables, which is expected as border ports are distributed across different regions.

This correlation matrix suggests that geographical factors, particularly latitude, play a role in border crossing volumes, with higher crossings occurring in southern regions. However, other factors like

policy, trade agreements, and economic conditions likely have a stronger influence, which requires deeper analysis beyond simple correlations.

### 6. Histogram of Border Crossings Distribution

The histogram above illustrates the distribution of border crossings, showing how frequently different crossing counts occur across all recorded entries.



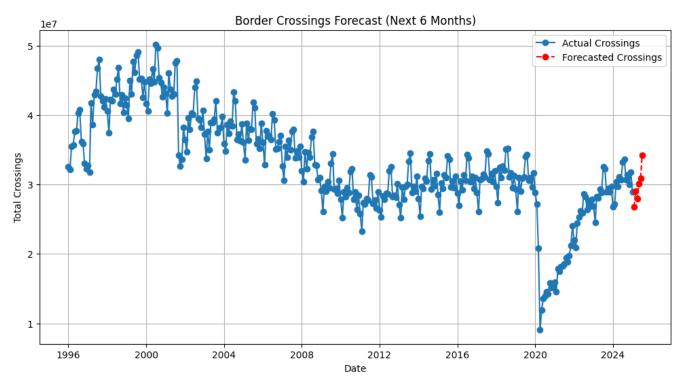
#### **Key Observations**

- Highly Skewed Distribution: The histogram shows an extreme right skew, meaning most border crossings are concentrated at lower values, with very few instances of extremely high crossings.
- This suggests that the majority of ports and crossing points handle relatively small numbers of entries, while only a few locations experience significantly high traffic.
- Presence of Outliers: A long tail extending towards higher values indicates the presence of hightraffic border crossings, such as those in Texas, California, and Michigan, where millions of crossings occur. These high values stand out compared to the majority of the dataset, confirming that border activity is highly concentrated at select locations.
- Most Common Crossing Counts: The majority of crossings fall within the lower range, reinforcing the idea that while a few major border ports handle a massive volume of traffic, many smaller ports experience low to moderate crossings.

This distribution confirms that border crossing activity is unevenly distributed, with a few high-traffic ports dominating the dataset. Understanding this imbalance can help authorities optimize border security, staffing, and infrastructure investments by focusing on high-traffic areas while ensuring efficient operations at lower-traffic ports.

### 7. Border Crossings Forecast (Next 6 Months)

The line chart above presents the forecasted trend for border crossings over the next six months, using historical data to predict future movement. The blue line represents actual past crossings, while the red dashed line indicates the forecasted values.



#### **Key Observations:**

- Gradual Recovery from the Pandemic Dip: The sharp drop observed in 2020 aligns with the COVID-19 pandemic, which caused severe disruptions in travel and trade. Since then, border crossings have shown a steady recovery, gradually approaching pre-pandemic levels.
- Forecasted Growth in the Short Term: The forecast predicts a continued increase in border crossings, suggesting that cross-border trade, travel, and migration are expected to rise. This growth trend aligns with historical seasonal patterns, economic recovery, and relaxed travel restrictions.
- Potential Fluctuations in Forecasted Values: The slight variations in the forecast indicate seasonal effects and policy-driven impacts on border activity. While the overall trend is upward, certain months may experience temporary dips due to external factors such as economic slowdowns, policy changes, or geopolitical influences.

The forecast suggests that border activity will continue increasing in the coming months, reinforcing the importance of efficient border management, infrastructure planning, and policy adjustments to handle the rising movement. Monitoring these trends will be essential for anticipating potential surges and ensuring smooth operations at key border ports.

### **Conclusion**

This analysis of border crossing data provides valuable insights into trends, seasonal variations, geographic disparities, and future projections of cross-border movement between the U.S.-Mexico and U.S.-Canada borders. Through a combination of exploratory data analysis, visualizations, and forecasting techniques, several key findings have emerged:

- Border crossing volumes are highly concentrated in a few key states, with Texas, California, and Arizona accounting for the majority of movement at the U.S.-Mexico border, while Michigan and New York dominate U.S.-Canada crossings.
- Personal vehicle passengers represent the most common mode of crossing, far surpassing commercial transport such as trucks and rail containers, highlighting the significance of personal travel in cross-border movement.
- Seasonal trends play a major role in fluctuations, with crossings peaking in the summer months
  due to increased tourism and trade activity, while winter months show a decline.
- A small number of ports handle an overwhelming majority of crossings, demonstrating the need for targeted investment in infrastructure, staffing, and security measures at high-traffic locations.
- The impact of global events such as the COVID-19 pandemic is evident, with a sharp decline in crossings during 2020 followed by a gradual recovery in the years since.
- Future forecasts suggest continued growth in border crossings, reinforcing the importance of efficient border management strategies, trade facilitation, and security enhancements.

These findings underscore the critical role of border infrastructure in sustaining economic activity, trade relations, and travel between the U.S., Canada, and Mexico. As border activity is expected to rise, it is crucial for policymakers to focus on optimizing traffic flow, enhancing security, and adapting to evolving travel and trade demands. A data-driven approach will be key in ensuring that border management remains efficient, secure, and adaptable to changing global conditions.