Final Paper
DS 3001

Abstract

This paper investigates seasonal patterns in cross-border traffic between the United States, Canada, and Mexico. Data from the Bureau of Transportation Statistics was used, which provided summary statistics for inbound crossings at the port level. The research question explores whether seasonal patterns can predict monthly crossing volumes at major U.S. border entry points. The analysis could be employed using regression tools to evaluate the predictive ability of seasonal and other factors. Results indicated a seasonal pattern, with peak crossings occurring during the summer months and dips in the winter months. External factors also played a role, such as economic conditions and policy changes. Most evident was global events, like the COVID-19 pandemic, which significantly impacted crossing volumes. EDA and visualizations demonstrated the dominance of personal vehicles and trucks in border crossings, the influence of seasonal trends, and the long-term effects of external factors. While seasonal trends provide a foundation for the prediction of border crossings, the study highlights the importance of accounting for other factors, especially in economic and policy variables for a more comprehensive analysis.

Introduction

Cross-border traffic plays a crucial role in the economic and social development between the United States and its neighboring countries, Canada and Mexico. From trade to tourism, understanding the factors that influence border crossing volumes is important for effective transportation planning and forecasting. This study focuses on the question: “Can seasonal patterns in border crossings predict monthly cross-border traffic at major U.S. borders of entry?”

After coming across a dataset that provided summary statistics on major border crossing ports, motivation stemmed from the idea of trying to understand and see what EDA could reveal about the data set. With a Commerce background, a global/international perspective has always been crucial. Today, businesses are increasingly practicing some level of international management, especially as organizations are becoming more diverse. On a greater scale, given the dataset, this can help optimize cross-border information and manage traffic flows, as well as forecast the impact of external factors.

Seasonal trends are especially relevant given their influence on travel. However, it is important to not solely rely on seasonal factors. This leads to overlooking other important influences, such as policy changes or global disruptions, as was seen with the COVID-19 pandemic.

The findings below demonstrate seasonal patterns do play a role in the volume of cross-border traffic. Summer months show higher crossing volumes while winter months see a decline, likely due to the harsher weather conditions and reduced demand. Visualizations also highlight the dominance of personal vehicles and trucks in cross-border travel, showing the role of personal and commercial transportation. However, it is important to note that the analysis holds limitations in focusing solely on seasonal patterns, external factors also hold influence.

Data

The dataset, titled “Border Crossing Entry Data,” is provided by the Bureau of Transportation Statistics (BTS) and accessed via data.gov. It contains historical monthly data on cross-border traffic at the U.S. ports of entry and is categorized by vehicle type.

As the dataset provides border crossings over the years, it can be analyzed to gauge potential patterns in seasonal increases in traffic. An understanding of this can be useful in predicting future border crossing volumes.

Sample of dataset

Port Name	State	Port Code	Border	Date	Measure	Value	Latitude	Longitude	Point
Roma	Texas	2310	US-Mexico Border	Dec 2023	Buses	46	26.404	-99.019	POINT (-99.018981 26.403928)
Del Rio	Texas	2302	US-Mexico Border	Dec 2023	Trucks	6552	29.327	-100.928	POINT (-100.927612 29.326784)
Roma	Texas	2310	US-Mexico Border	Nov 2023	Trucks	3753	26.404	-99.019	POINT (-99.018981 26.403928)
Douglas	Arizona	2601	US-Mexico Border	Oct 2023	Buses	13	31.334	-109.560	POINT (-109.560344 31.334043)
Beecher Falls	Vermont	206	US-Canada Border	Aug 2023	Trucks	422	45.013	-71.505	POINT (-71.505309 45.013411)
Laredo	Texas	2304	US-Mexico Border	Aug 2023	Buses	2843	27.500	-99.507	POINT (-99.507412 27.499561)
Morgan	Montana	3319	US-Canada Border	Aug 2023	Trucks	20	49.000	-107.832	POINT (-107.831819 48.999829)
Hidalgo	Texas	2305	US-Mexico Border	Aug 2023	Trucks	59677	26.095	-98.271	POINT (-98.271092 26.095032)
Raymond	Montana	3301	US-Canada Border	Jul 2023	Buses	3	48.999	-104.574	POINT (-104.574333 48.999194)
Roma	Texas	2310	US-Mexico Border	Jul 2023	Bus Passengers	949	26.404	-99.019	POINT (-99.018981 26.403928)
Wildhorse	Montana	3323	US-Canada Border	May 2023	Truck Containers Empty	1	48.999	-110.215	POINT (-110.215083 48.999361)

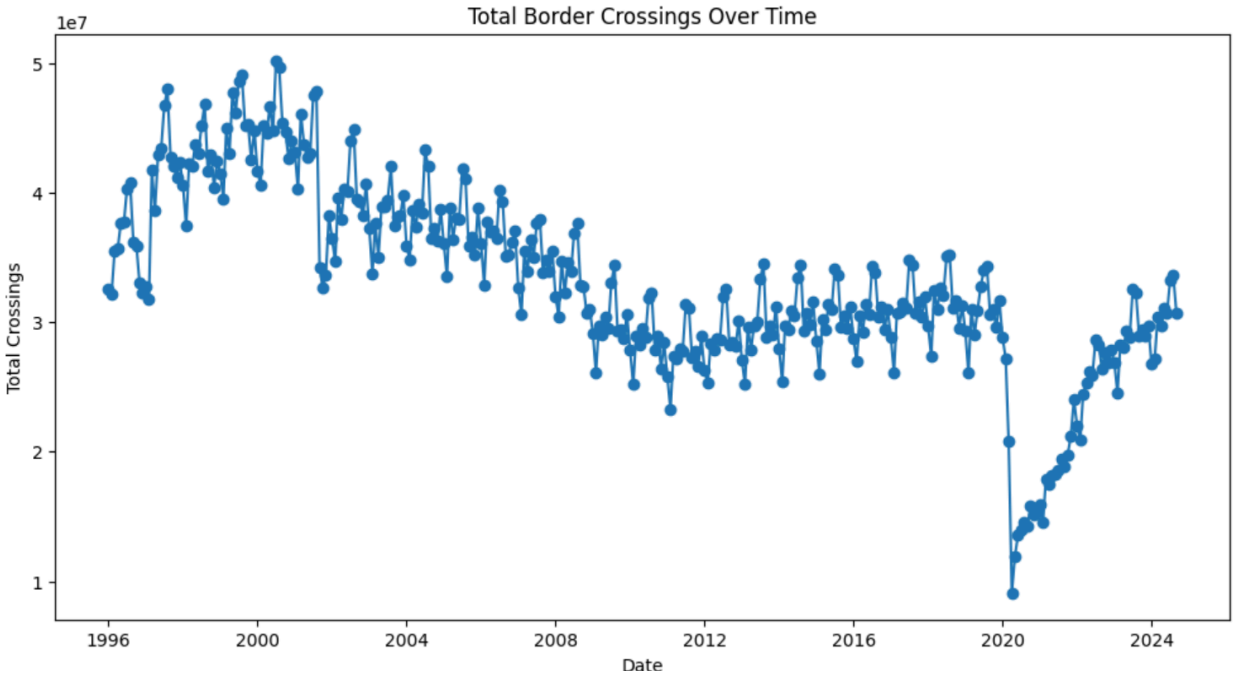
Sample of EDA/Data Wrangling

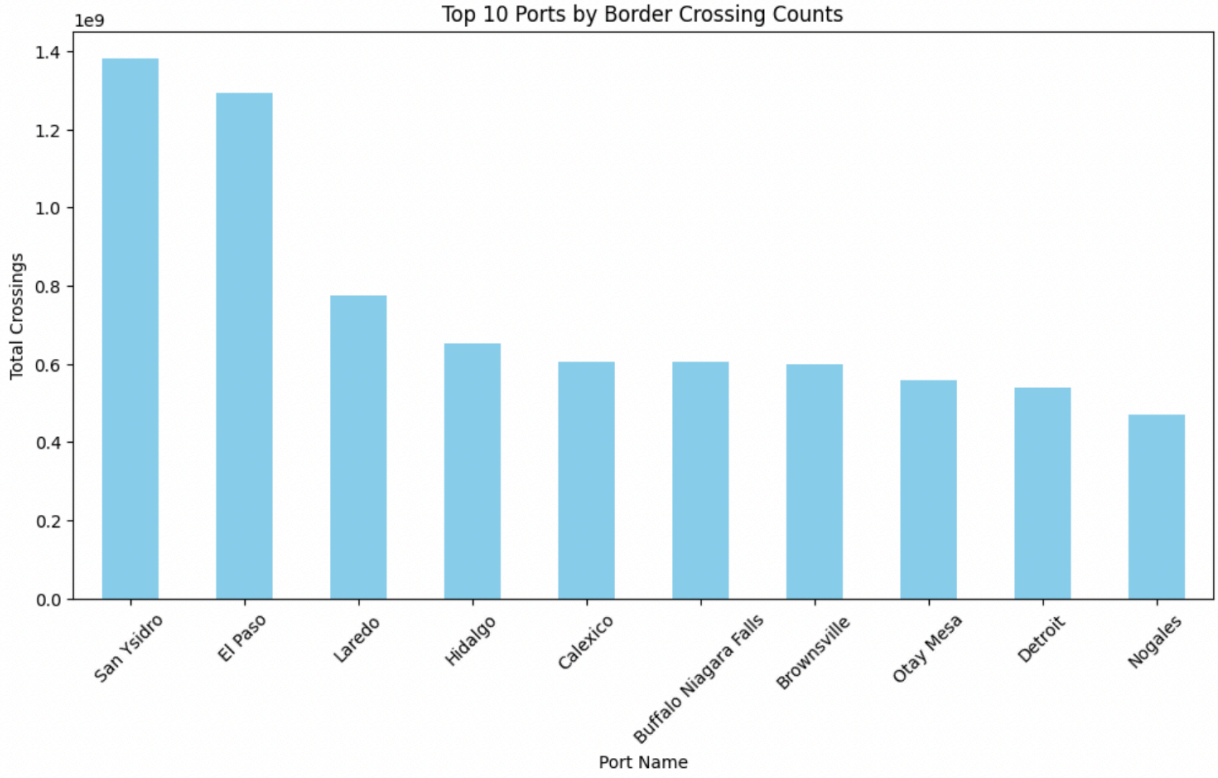
```
print(df.shape)
```

```
(395638, 10)
```

```
print(df.describe())
```

```
           Port Code         Value       Latitude      Longitude
count  395638.000000  3.956380e+05  395638.000000  395638.000000
mean     2446.496201  2.884933e+04      43.974092     -99.730139
std      1204.138442  1.495808e+05       8.151891      18.259105
min       101.000000  0.000000e+00      25.952000    -141.001000
25%      2304.000000  0.000000e+00      42.999000    -114.728000
50%      3012.000000  1.430000e+02      48.122000    -101.628000
75%      3401.000000  2.984000e+03      49.000000     -89.585000
max      3814.000000  4.447374e+06      62.615000     -66.980000
```


Total Border Crossings Over Time
1e7
5
4
3
2
1
Total Crossings
1996
2000
2004
2008
2012
2016
2020
2024
Date


Top 10 Ports by Border Crossing Counts
1e9
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Total Crossings
San Ysidro
El Paso
Laredo
Hidalgo
Calexico
Buffalo Niagara Falls
Brownsville
Otay Mesa
Detroit
Nogales
Port Name

Methods

The research question guiding this analysis is "Can seasonal patterns in border crossings predict monthly cross-border traffic at major US borders of entry." The data set from data.gov provides reliable and historical data that can be used to explain why traffic patterns may vary seasonally. It will also be able to go beyond seasonal patterns, exploring the relationship of other predictive factors.

An observation in this study is the monthly entry count at a specific location, in this case, port. This is categorized by vehicle type (truck, bus, personal vehicle, etc).

This would be an unsupervised learning analysis as I will not be using labeled trained data. It will also focus on regression as the goal of the analysis is to predict crossing volume.

Linear regression will be used as a baseline for the prediction model, as well as Lasso regression as it can help determine which factors contribute the most to crossing volume (while reducing overfitting).

To measure success, it will be determined if a model can predict crossing volume and go beyond by exploring other factors that may be impactful. Certain metrics can measure and answer the research question. For example, "Can R^2 measure the variance in crossing volume as shown by seasonal factors," and "Can RMSE help assess prediction accuracy."

To prepare this data set for the analysis, I can clean and filter the data by changing some to dummy variables. This way, I will be able to use categorical data in the model so that I can have qualitative factors, whether that be location or vehicle, represented.

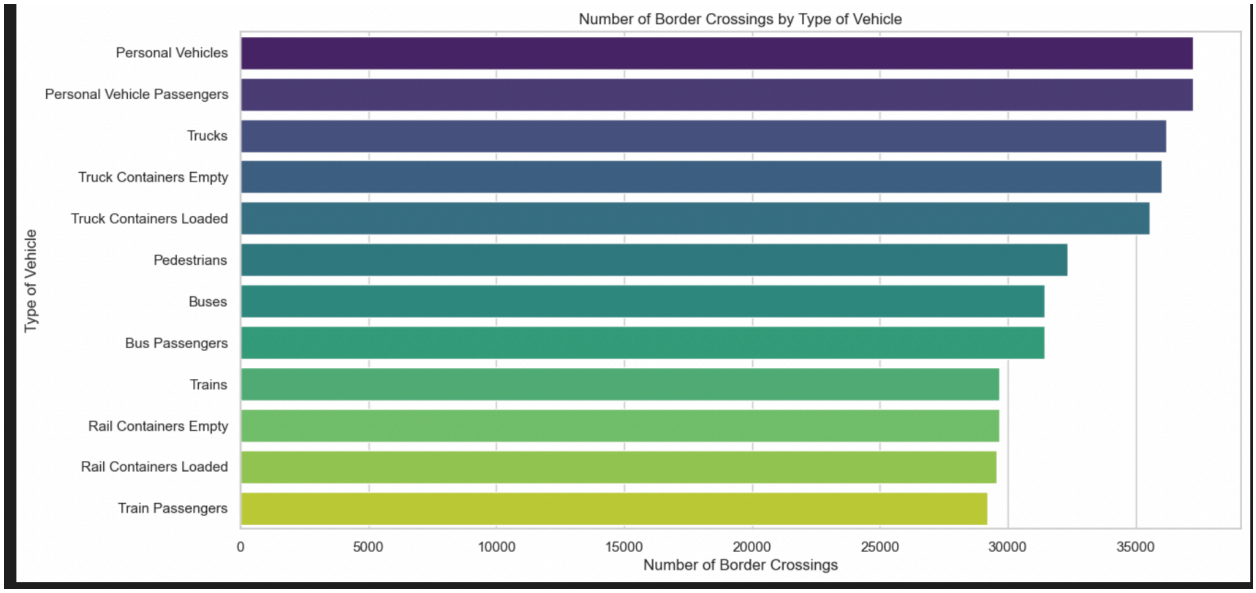
An anticipated weakness is that seasonal factors within the data will not be enough to influence cross-border traffic as there may be other external influences, such as economic trends or policy/administration changes, that could also create variations. As a result, LASSO may be useful in helping irrelevant features. Another weakness is that focusing on seasonal factors may lead to overfitting, so utilizing cross-validation will be important so that it is not centered on short-term patterns.

If this approach fails, an insight to be gained is “what other methods and metrics may help reach a more accurate prediction of the research question.” More specifically, it will be able to demonstrate if border crossing patterns are not as driven by seasonal factors as originally expected, showing the need to consider additional sources and models that may explain crossing volume.

To present results, I will create visualizations, such as bar charts for monthly crossing volumes and tables for comparing model metrics to show if there is predictive success and seasonal factors.

Results

Visual 1: Distribution of Border Crossing by Vehicle



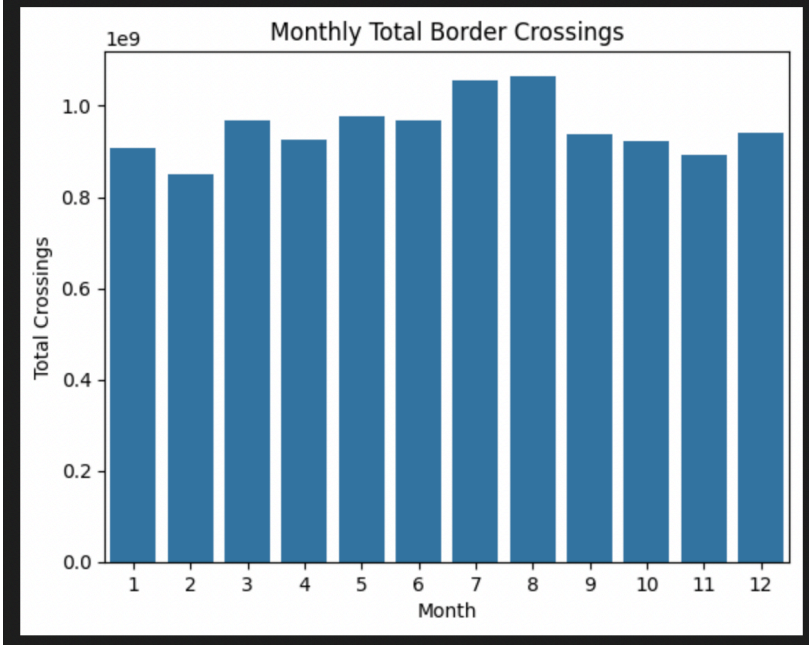
This visual provides an insightful breakdown of the dominant modes of cross-border traffic. Personal vehicles, followed by personal vehicle passengers, are the highest in volume, demonstrating the significant role individuals and families play in cross-border movement. This could reflect common travel, such as tourism, family visits, or commuting, which has an impact on overall traffic patterns. The substantial presence of trucks in teh data, both loaded and empty, also demonstrates the importance of commercial trade and logistics across borders, as there is reliance on efficient transportation for goods and raw materials.

The lower volume of rail containers, loaded and empty, and train passengers shows that there is a smaller proportion of cross-border traffic that is not as commonly transported by rail and other vehicle types. This could reflect infrastructure, cost-effectiveness, and usage patterns that favor personal vehicles and trucks for their flexibility in travel and higher capacity for rail as rail is for bulk and long-distance transportation.

Visual 2: Monthly Total Border Crossings

```
sns.barplot(x='Month', y='Value', data=monthly_counts)
plt.title('Monthly Total Border Crossings')
plt.xlabel('Month')
plt.ylabel('Total Crossings')
plt.show()
```

✓ 0.1s



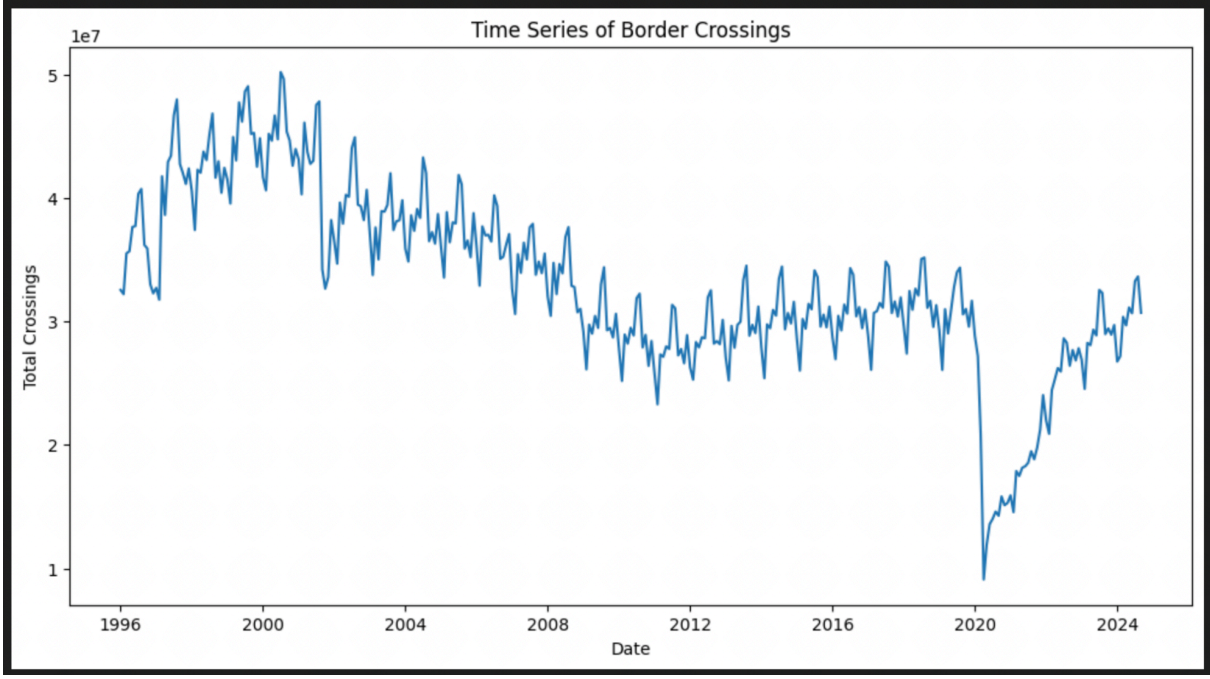
This shows a seasonal pattern coming into play in the volume of crossings. The months with the highest total crossings are June (6) and July (7). This corresponds to the summer season, a period that sees increased travel adn trade activity. This aligns with the expectation that warmer months produce higher border activity due to favorable weather conditions and some holiday seasons during this time.

Additionally, the lowest total crossings are in February (2) and November (11). This could be attributed to the winter weather conditions, which deter non-essential travel and impact the transportation of goods. Similarly, November reduced crossings reflect holiday slowdowns and before the end-of-the-year holiday surge, as well as possible impacts from unpredictable weather events.

Visual 3: Time Series of Border Crossings

```
plt.figure(figsize=(12, 6))
sns.lineplot(data=time_series, x='Date', y='Value')
plt.title('Time Series of Border Crossings')
plt.xlabel('Date')
plt.ylabel('Total Crossings')
plt.show()
```

✓ 0.3s

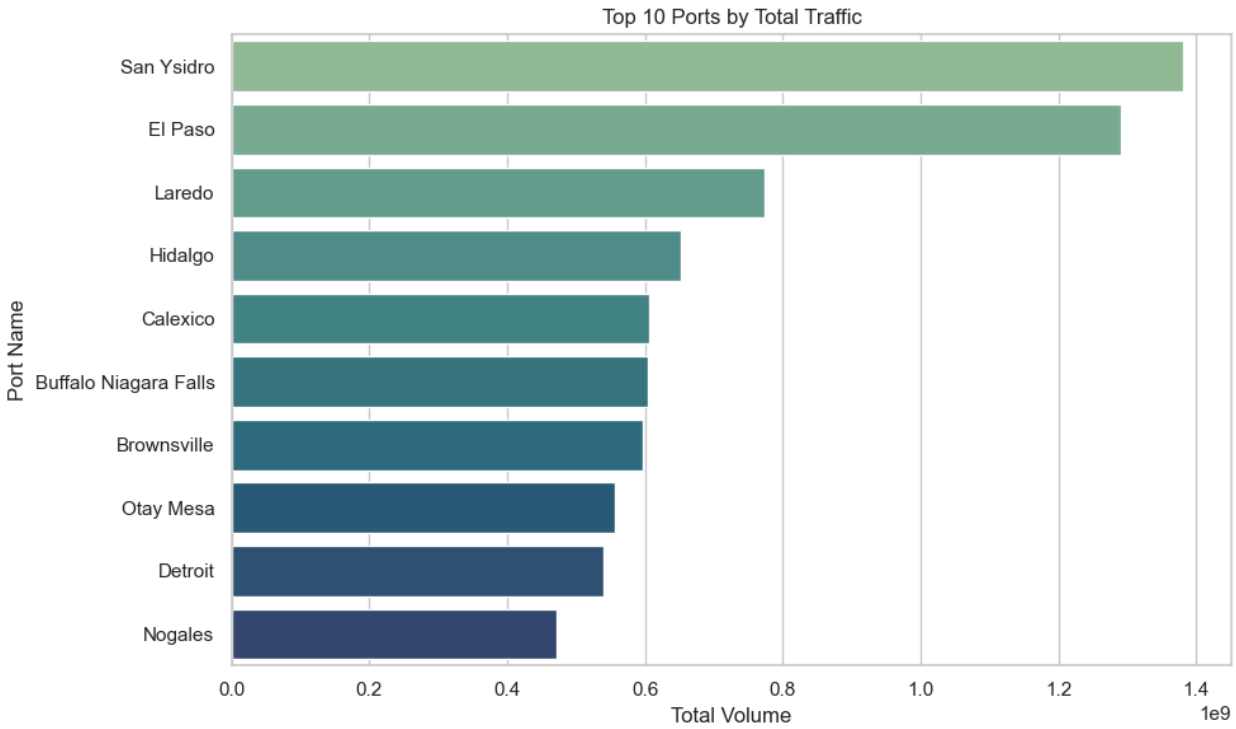


This visual shows a trend from 1996-2002 where there is a period of relatively stable growth and consistent increase in cross-border traffic. However, the notable decline after 2002 suggests there was an impact from external factors, possible policy changes or economic shifts, as well as increased security measures.

The dramatic dip in 2020 aligns with the global outbreak of COVID-19, demonstrating the disruption caused by the pandemic that led to reduced cross-border activity due to lockdowns, travel restrictions, and economic uncertainty as well as supply shortages.

The subsequent upward trend post-2020 indicates a recovery period as restrictions eased and policies began adapting to new norms, pointing to potential seasonal and economic factors resuming.

Visual 4: Top Ports by Total Traffic



The graph above displays the top 10 border ports ranked by traffic volume, listing San Ysidro as the busiest crossing. Located at the U.S.-Mexico border, San Ysidro is known as the “Gateway to the Americas” due to its international border. It is the most southern community in California, located between San Diego and Tijuana, and is valuable for its tourism. Additionally, the listing of ports such as Buffalo Niagara Falls, shows the diverse geographic contributions to cross-border activity that the United States faces.

Conclusion

Leveraging seasonal patterns to predict cross-border traffic volumes provides much value. Cross-border transportation enhances economic development between regions, and much goes into the facilitation of this. From safety to security and environment, governments need to address border challenges to improve mobility by adequately planning and forecasting.

It is important to also note limitations that may impact the analysis. While models offer useful predictions, they can be constrained by leaving out important contextual variables, assuming correlation or linearity, and incomplete data coverage. First, the dataset may not fully capture all border crossings, especially in minor/non-important points, leading to potential gaps in teh data. External factors, such as policy or economic shifts also influence border traffic, which is not reflected in the data. The data's monthly granularity also fails to account for daily or hourly patterns, which could provide more specific insights into traffic trends.

Overall, seasonal trends provide a reliable baseline for analysis, and it is important to go beyond and identify external influences for a better understanding of the full scope of border traffic dynamics. The dominance of ports like San Ysidro reflects the importance of personal and commercial exchanges. Exploring advanced machine learning techniques and time series forecasting can also improve the accuracy of predicting and accounting for non-linear patterns.

By addressing these areas, future studies and institutions can better support border-crossing and transportation planning, which in turn enhances policy decisions and economic growth. This provides a valuable foundation for understanding cross-border traffic trends, with the potential to inform policy decisions and improve management.

References

Border Crossing/Entry Data. Border Crossing/Entry Data | Bureau of Transportation Statistics. (n.d.). https://www.bts.gov/explore-topics-and-geography/geography/border-crossingentry-data#1

Improving cross-border mobility and Trade. (2023, August). https://www.transportation.gov/sites/dot.gov/files/2023-08/OSTX-Momentum-Toolkit-Border-Planning_August2023_0.pdf

Publisher Bureau of Transportation Statistics. (2024, November 21). *Department of Transportation - Border Crossing Entry Data*. Catalog. https://catalog.data.gov/dataset/border-crossing-entry-data-683ae