Julia Monson
Dec 20, 2025

# U.S Border Crossings 2019-2025

My project analyzes the U.S border crossing activity using public data from the U.S Bureau of Transportation Statistics (BTS). Recording and keeping up-to-date data on the fluctuation of U.S border crossing activity is important because it's an indicator of economic activity such as trade, travel behavior, and seasonal movement patterns. My project shows how border crossing volumes fluctuate depending on these factors: Season, border regions, different types of crossings, and the volumes year to year. These metrics and big data can help quantify the interrelationships from these variables based on the visualization models presented.

The description for this data set in the BTS states, "The Bureau of Transportation Statistics (BTS) Border Crossing Data provide summary statistics for inbound crossings at the U.S.-Canada and the U.S., Mexico border at the port level. Data are available for trucks, trains, containers, buses, personal vehicles, passengers, and pedestrians. Border crossing data are collected at ports of entry by U.S. Customs and Border Protection (CBP). The data reflect the number of vehicles, containers, passengers or pedestrians entering the United States. CBP does not collect comparable data on outbound crossings. Users seeking data on outbound counts may therefore want to review data from individual bridge operators, border state governments, or the Mexican and Canadian governments."

The research question addressed in Power BI are:

1. How have border crossing volumes changed over time between 2016 and 2025?
2. To what extent do seasonal patterns influence border crossing volumes?
3. Are there systematic differences in crossing volumes between the U.S.–Mexico and U.S.–Canada borders?
4. How do different types of crossings (personal travel, commercial freight, rail, and pedestrian) relate to overall crossing volumes?
5. What factors are associated with high-volume border crossing observations?

The dataset contains monthly counts of border crossings at U.S. ports of entry, by border location, port, state, and type of crossing. The time frame of the data captured went from 2016-2025, capturing recent/up-to-date data while keeping older data to use for analysis purposes. After data preparation, variables were added to be used in the analysis including, Year, Month, Date, Quarter, BorderRegion, MeasureCategory, Crossings, LogCrossings, HighVolumeFlag, and seasonality variables derived from the calendar month.

Variable Definitions

Several variables were created:

- Crossings: The original count of border crossings reported by BTS.
- LogCrossings: The natural logarithm of Crossings, used to reduce skewness and stabilize variance.
- BorderRegion: Indicates whether the crossing occurred at the U.S. Mexico or U.S. Canada

border.

• MeasureCategory: A classification of crossing types, including Personal Travel, Commercial Freight, Rail, and Pedestrian crossings.

• HighVolumeFlag: Shows as 1, if the crossing count is at or above the 75th percentile of the dataset, and 0 if criteria are not met.

• Month_Sin and Month_Cos: Sine and cosine transformations of the month variable used to represent seasonality in continuous form.

There are some interrelationships that I can predict are going to show between variables and the dataset. I can guess border crossing volumes are going to increase over time from population growth and economic expansion. Strong seasonal patterns will be driven by tourism and trade cycles. Border regions is another variable I'd be familiar with from the news. As the U.S. Mexico border typically supports higher volumes of both personal and commercial crossings than the U.S. Canada border.

Data preparation has several steps. First, I started with the raw BTS dataset that include only observations from 2016 through 2025. Date-related variables such as Year, Month, Quarter, and YearMonth were created to support time-based analysis.

The Crossings variable was log-transformed to create LogCrossings, this was for my regression model to prevent skewness. Measure values were grouped into broader MeasureCategory classifications to reduce dimensionality and improve interpretability. A HighVolumeFlag variable was created to highlight years and season that has a high-volume crossing rate.

Several quantitative methods came directly from the excel file. Correlation analysis was used to examine linear relationships among crossing volumes, time indicators, seasonality variables, and border indicators. This provided a starting point or baseline for future analysis.

A simple linear regression model was made with LogCrossings as the dependent variable and Year as the independent variable. This model shows numerical data and how long-term trends are affected through border crossing activity over the years.

A multiple regression model was made with LogCrossings as the dependent variable and predictors including Year, seasonal terms, border indicator, and MeasureCategory. This model shows the joint effects of time, seasons, border region, and crossing type on crossing volumes.

The logistic regression model was used to show the odds of observations representing a high-volume crossing. The dependent variable in this model is HighVolumeFlag, and the predictors correlate with those used in the multiple regression model. Logistic regression is used because the outcome variable is binary.

The correlation analysis indicates strong relationships between crossing volumes, time variables, and seasonal indicators, supporting the use of regression models in excel. The simple linear regression show a strong correlation between Year and LogCrossings, indicating that border crossing volumes have increased overtime. The multiple regression shows how the volumes of border crossing change depending on the seasons; this also accounts for long-term trends you can see in Power BI. The border indicator and measure category variables show that the volumes change depending on the geographical location of the border. The logistic regression

results indicate that time, seasonality, and border region indicate if an observation is “high-volume” or not. These results go back to the linear regression model.

Overall, this project demonstrates that U.S. border is influenced by a combination of long-term trends, seasonal patterns, border region differences, and crossing types. By combining regression and correlation models, you can see a descriptive analysis that shows there are interrelationships between all variables. The results show that border crossing volumes have increased overtime and vary during certain seasons. Some borders get more action than others. Different types of crossings seem to be dependent on the location of the border used. These findings can help with transportation, trade, and when to put more guards by the boarders.

## References

Publisher Bureau of Transportation Statistics. (2025, November 22). *Department of Transportation - Border Crossing Entry Data*. Catalog. https://catalog.data.gov/dataset/border-crossing-entry-data-683ae