Final Project on Data Driven Border Control Strategies and Insights

Agenda

- Real-World Problem
- Dataset Overview
- Reasons for selecting the dataset
- Methodologies Used
- Results of the Analysis
- Conclusion
- Recommendations

Real-World Problem

- Challenges in managing land border crossings along the US-Mexico and US-Canada borders.
- Increasing traffic volumes and diverse vehicle types pose logistical and security challenges.

Importance of Big Data:

- Big data analytics enable border management agencies to better understand, predict, and respond to the logistical and security challenges posed by increasing traffic volumes and diverse vehicle types at US land border crossings.

Border Crossing Entry Dataset

Dataset divided into 4 categories:

A.) String Variables:

- Port Name: This variable contains the names of the ports where border crossing activities occur.
- State: Indicates the state where each port is located.
- Border: Specifies the type of border, such as "US-Mexico Border" or "US-Canada Border."
- Date: Represents the date when each border crossing activity took place.

Border Crossing Entry Dataset

B.) Integer Variables:

- Port Code: Each port has a unique numerical code assigned to it.
- Value: Indicates the quantity or count associated with each measure. For example, the number of vehicles, passengers, or containers.

C.) Categorical Variables:

- Measure: This variable represents different categories or types of measures taken at the ports. Examples include Buses, Trucks, Pedestrians, Personal Vehicles, Truck Containers Loaded, Truck Containers Empty, Rail Containers Empty, Bus Passengers, Train Passengers.

Border Crossing Entry Dataset

D.) Geographical Variables:

- Latitude: It is a numerical variable used to pinpoint the north-south position of each border crossing location.
- Longitude: It is a numerical variable used to pinpoint the east-west position of each border crossing location.
- Point: This variable represents the exact geographic location of each port in a specific format called "POINT." It combines the latitude and longitude coordinates into a single data point representing the spatial location of the port.

Reason for Selecting the Dataset

Relevance: The dataset focuses on land border crossings, which are crucial points of entry for goods, people, and vehicles between the US, Mexico, and Canada. Understanding traffic patterns and trends at these borders is essential for trade, security, and logistics management.

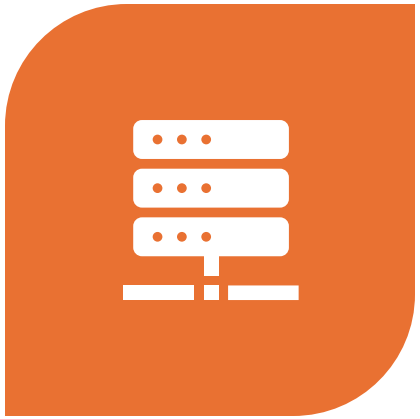
Scope: The dataset covers a wide range of variables, including port names, states, port codes, border types, dates, types of measures (e.g., vehicles, pedestrians), counts, and geographic coordinates. This comprehensive data allows for in-depth analysis and insights into border activities.

Real-world Impact: Analyzing this dataset can reveal valuable insights such as the busiest border crossings, types of vehicles or passengers frequenting these borders, seasonal variations in traffic, and geographic hotspots. These insights can inform policy decisions, resource allocation, security measures, and infrastructure planning at land border crossings.

Predictive Analysis: The dataset's historical data can be leveraged for predictive modeling, enabling the forecasting of future traffic patterns, identifying potential anomalies, and optimizing border management strategies proactively.

Big Data Potential: With its large volume of observations and diverse variables, this dataset is well-suited for big data analytics. Advanced analytics techniques can be applied to uncover hidden patterns, correlations, and trends that traditional methods may overlook.

Methodology used to analyze the data



K-MEANS CLUSTERING

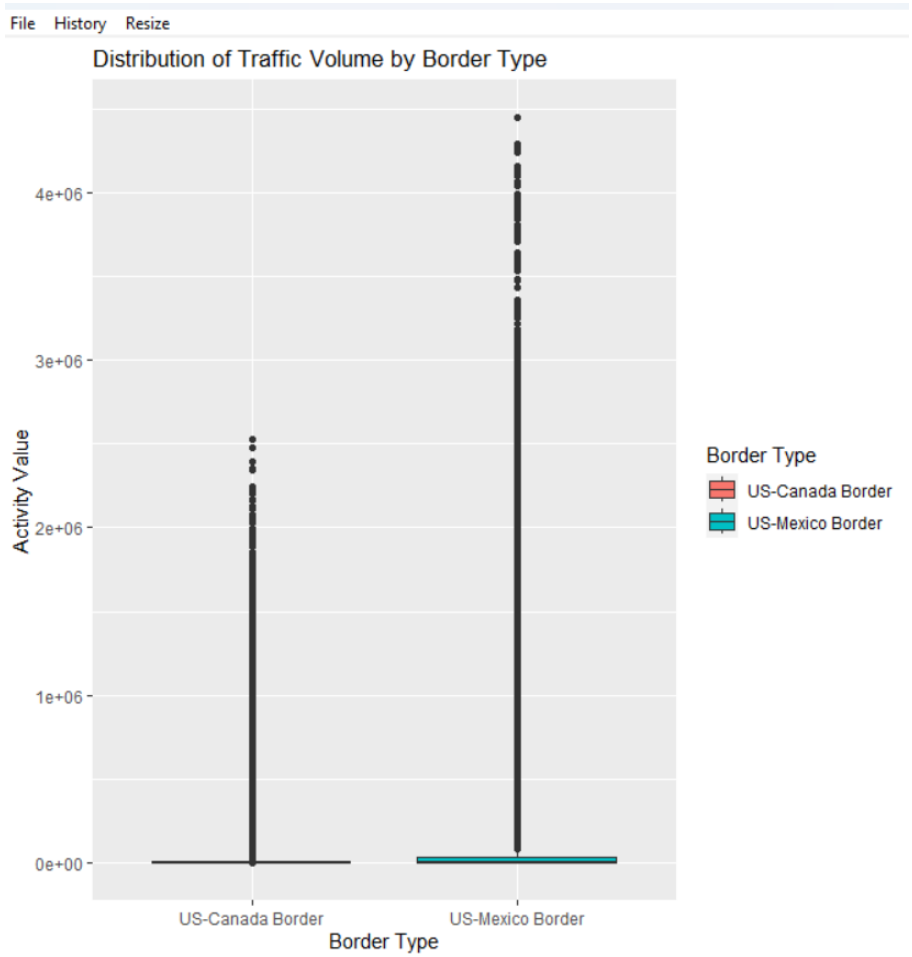
DECISION TREE

POWER BI

Results of the analysis

Insights Using Box plot in R:

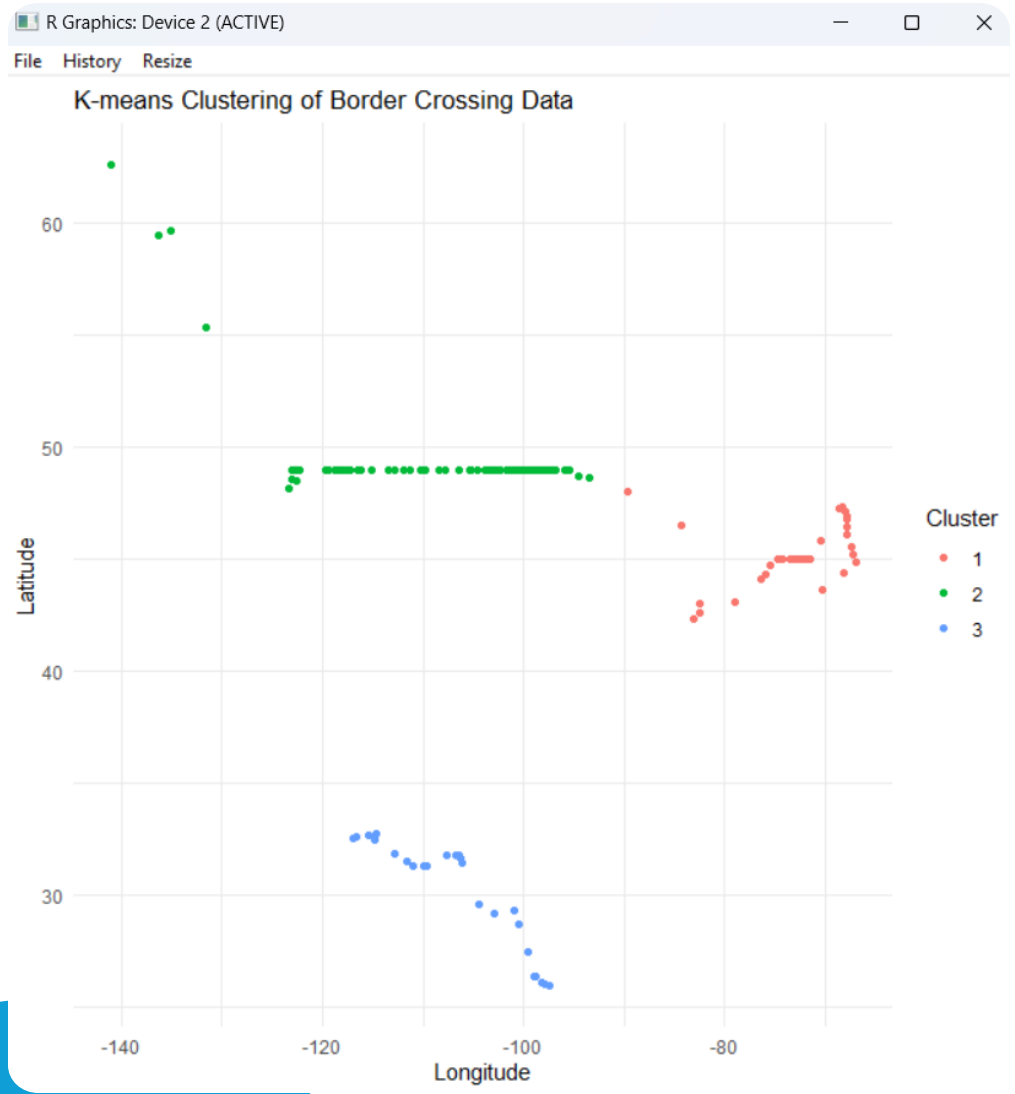
- Analysis revealed significant variations in traffic volumes.
- Ports along the US-Mexico border showed higher traffic volumes.
- US-Canada border ports exhibited lower traffic volumes in comparison.



Map Visualization using Power BI

Key Insights:

- The map chart effectively represents the geographical distribution of border activities, highlighting areas of high passenger traffic.
- Personal vehicle passengers exhibit the highest number of crossings at the borders based on the chart.

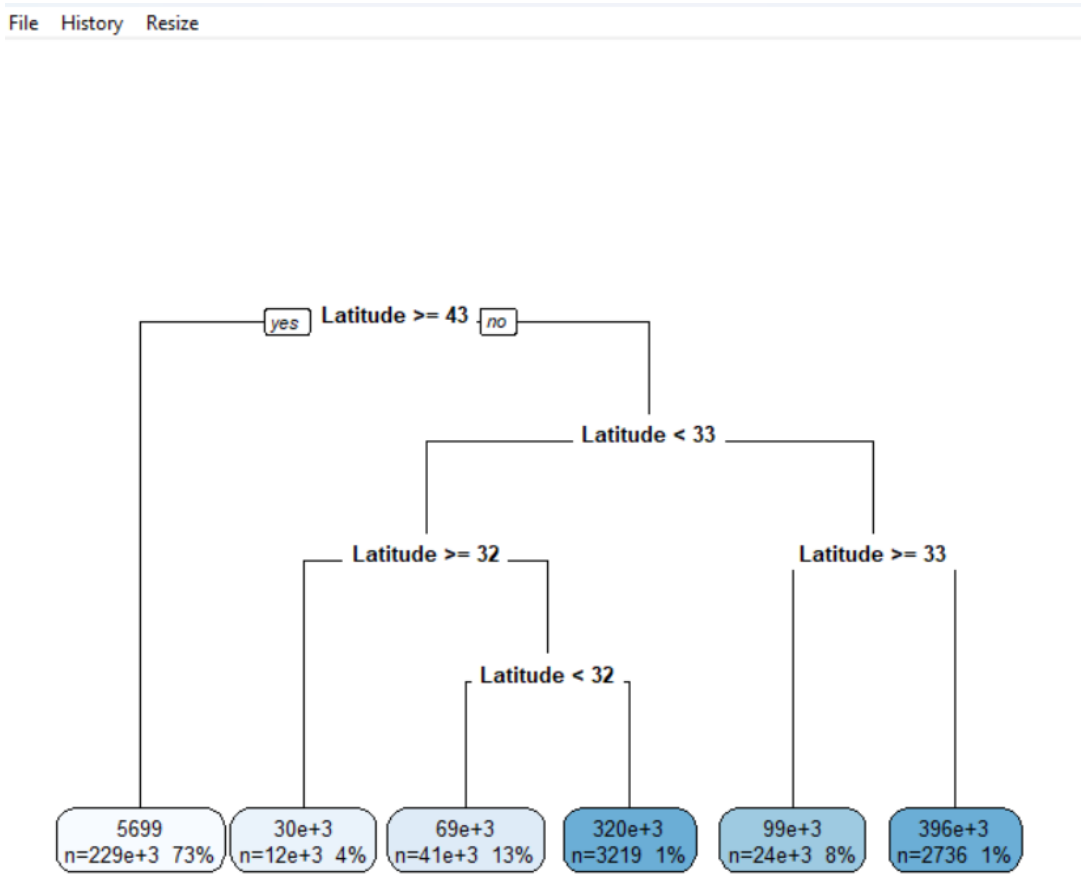


Analysis of Traffic Volume using K-means clustering

- Clusters indicate regions with high or low traffic volume or distinct traffic patterns based on geographical location.
- Clusters with higher traffic volume or clusters located in specific geographic regions can indicate hotspots or areas of significant activity.
- Spatial clustering of border crossing data using Latitude and Longitude can reveal geographical patterns, traffic behaviors, hotspot identification, anomaly detection, and offers a visual representation for informed decision-making.

Predicting Traffic Volume based on Latitude Using Decision Tree Model

- Traffic Prediction: Estimate traffic volume at latitudes for management and resource allocation.
- Hotspot Identification: Identify high/low traffic areas by latitude for planning and development.
- Geographic Trends: Analyze traffic trends with latitude to understand regional variations.
- Resource Optimization: Allocate resources based on predicted traffic levels at different latitudes.
- Decision Support: Aid decision-making in border control, transport, and tourism with geographical traffic insights.

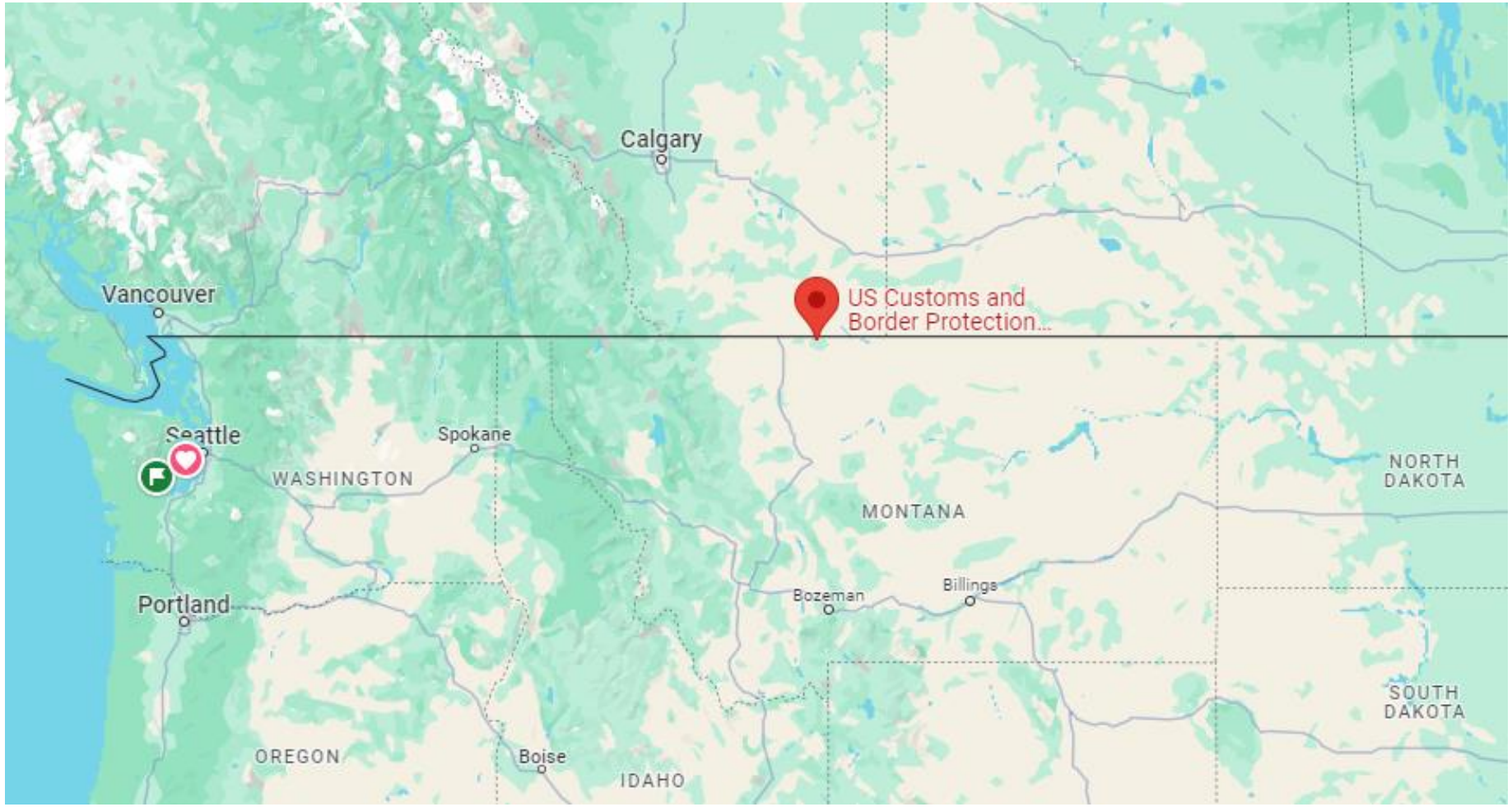


Border Comparisons

San Ysidro

Whitlash

Map of Whitlash Border Crossing Point



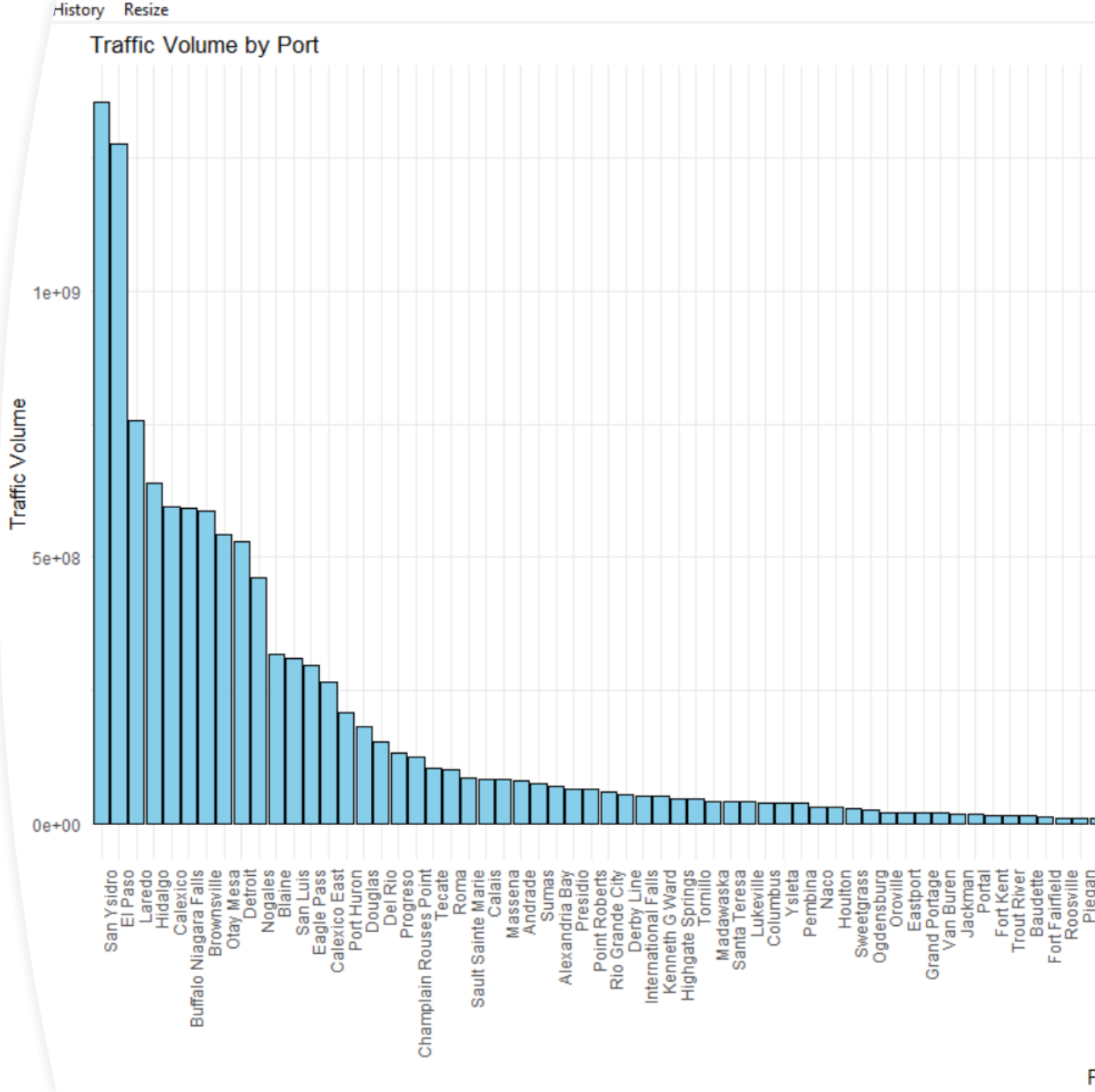
Traffic Volume by ports

Key Insights:

- Highest Volume Port: San Ysidro port shows the highest traffic volume among all ports.
- Lowest Volume Port: Whitlash port indicates the lowest traffic volume compared to others.

Significance:

- Understanding traffic volumes by ports is crucial for resource allocation, infrastructure planning, and policy decisions aimed at optimizing border management and economic impact.



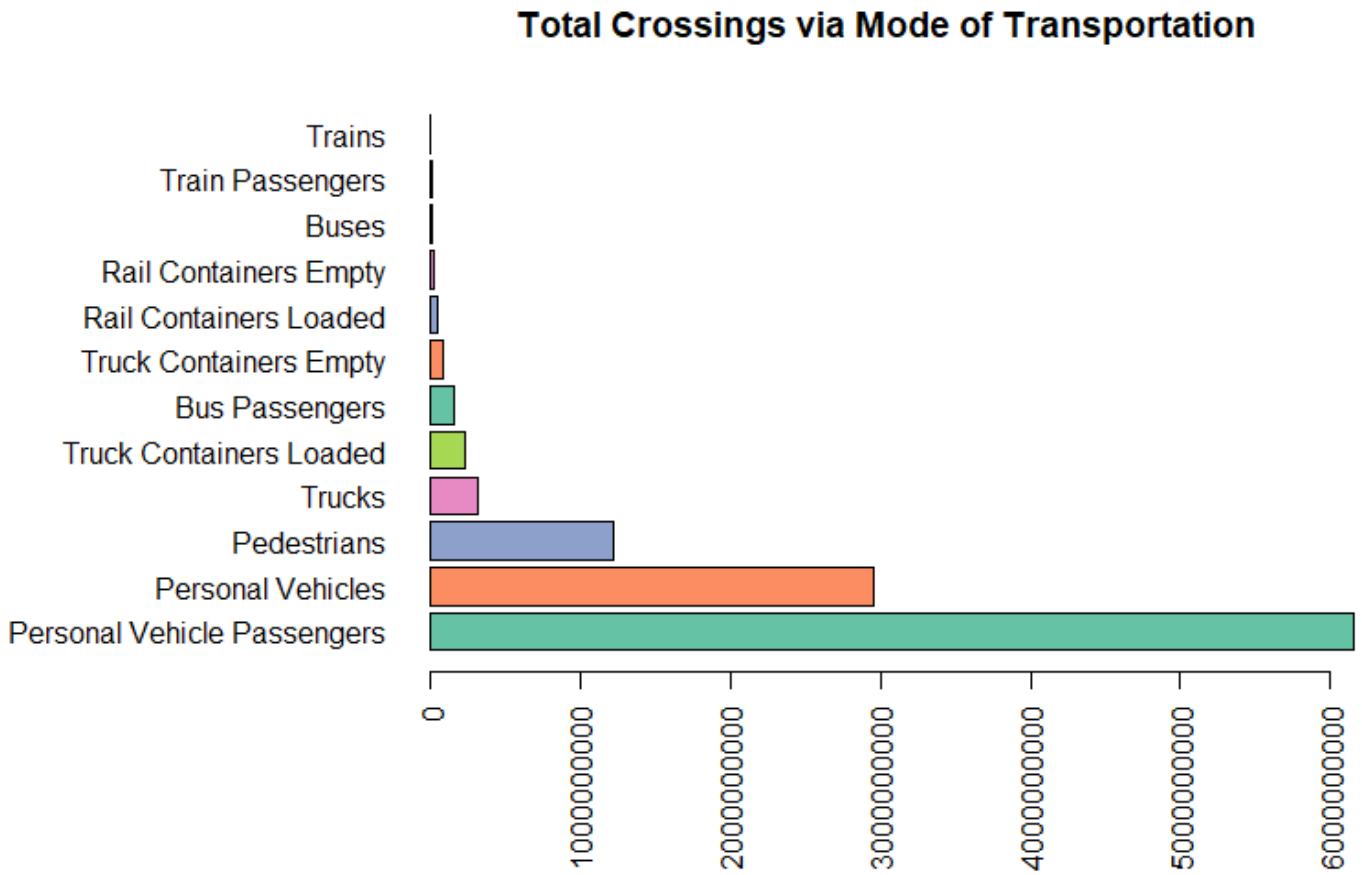
Count of Activities by type using bar chart

Key Insight:

- Personal vehicle passengers constituted the bulk of crossings.

Understanding traffic patterns is crucial for strategic planning:

- Border security measures can be tailored based on these patterns.
- Infrastructure development can be aligned with the needs of high-traffic ports.
- Policy decisions regarding trade and immigration can be influenced by these insights.



Border Crossing by State

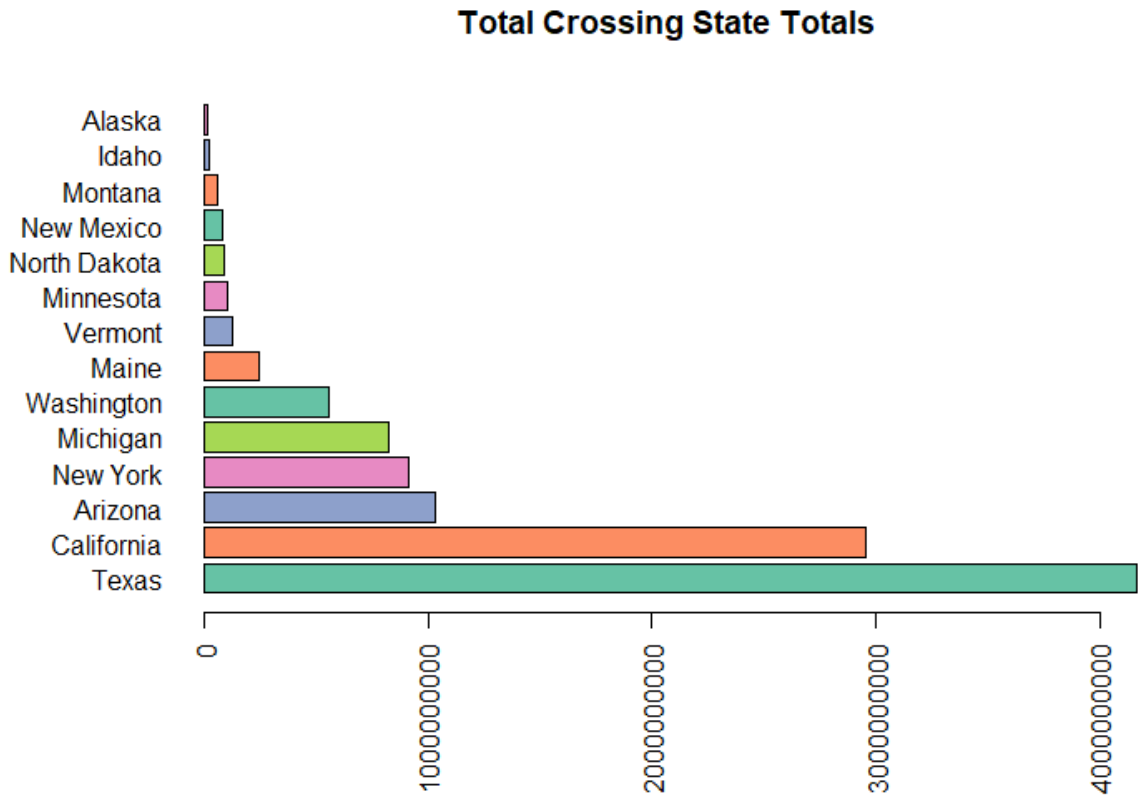
Findings from the Visualization:

- Texas has the most activity.

- Idaho and Alaska exhibit the lowest traffic activities
  - Idaho = tiny span of border
  - Alaska = remote areas

Significance:

- Understanding state-wise variations in border crossing activities helps in targeted resource allocation and strategic decision-making for optimized border management.



Conclusion

- In summary, this presentation delved into the complexities of managing land border crossings along the US-Mexico and US-Canada borders. Through extensive data analysis and visualization techniques, we uncovered key insights into traffic volumes, geographical patterns, and predictive trends.

- By leveraging advanced methodologies like K-Means Clustering, Decision Trees, and Power BI visualization, we gained valuable insights for informed decision-making. These insights are vital for optimizing resource allocation, enhancing security measures, and improving infrastructure planning at land border crossings.

- Ultimately, understanding traffic patterns and trends is critical for ensuring efficient trade, security, and transportation operations. With data-driven insights guiding our approach, we can better manage land border crossings to facilitate smoother cross-border movements and bolster national security.

Recommendations

Infrastructure Investment: Expand port facilities and improve road networks at high-traffic border crossings to facilitate smoother movement of goods and people.

Enhanced Security Measures: Utilize predictive analytics for real-time monitoring to enhance border security and prevent illegal crossings.

Collaborative Approach: Foster collaboration among border management agencies and international partners to improve coordination and response capabilities.

Regular Data Analysis: Establish a systematic approach to monitor traffic trends and adjust strategies accordingly.

Investment in Analytics: Strengthen data analytics capabilities within border agencies to leverage big data for improved decision-making.

Public Awareness Campaigns: Educate travelers and stakeholders about border regulations and safety measures to promote compliance and cooperation.

References

- Border Crossings by Mode, Border, and State | Tyler Data & Insights. (n.d.). https://data.bts.gov/Research-and-Statistics/Border-Crossings-by-Mode-Border-and-State/erjk-mneb
- https://data.transportation.gov/api/views/keg4-3bc2/rows.csv?accessType=DOWNLOAD
- GeeksforGeeks. (2024, March 11). K means Clustering Introduction. GeeksforGeeks. https://www.geeksforgeeks.org/k-means-clustering-introduction/
- GeeksforGeeks. (2023, August 20). *Decision tree*. GeeksforGeeks. https://www.geeksforgeeks.org/decision-tree/
- Yung, Z. (2024, May 2). How to create Power BI Maps: A comprehensive guide. Coupler.io Blog. https://blog.coupler.io/power-bi-maps/