# DATA SCIENCE MINOR PROJECT REPORT

(Project Semester January-April 2025)

Border Crossing Entry Data Analysis

[Dataset and PPT](https://drive.google.com/drive/folders/1eMoIGm2hout86wBZ-gI6B94KkRWApq_D?usp=drive_link)

(Please click on it CTRL + Click to view excel work and ppt)

# Submitted by:

Name: Aditya Singh

Registration No: 12323240

Program and Section: Bachelor Technology CSE – K23SH

Course Code: INT217

# Under the Guidance of:

Ashima Bansal, Assistant Professor

Discipline of CSE

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

# CERTIFICATE

This is to certify that Aditya Singh bearing Registration no. 12323240 has completed INT217 project titled, “Border Crossing Entry Data Analysis” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor

Ashima Bansal

Assistant Professor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab

Date:

# DECLARATION

I, Aditya Singh, student of B.Tech CSE under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 11th April 2025

Signature: Aditya Singh

Registration No: 12323240

Name: Aditya Singh

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project supervisor, Prof. Jane Smith, for her invaluable guidance, support, and encouragement throughout the project. I am also thankful to the faculty and staff of the School of Computer Science and Engineering, Lovely Professional University, for providing all the necessary resources and support.

# TABLE OF CONTENTS

1. Introduction v

2. Source of dataset v

3. Dataset Preprocessing v

4. Analysis on dataset: for each objective vi

i. Objective 1: Trend Analysis vi

ii. Objective 2: Comparison of US-Canada Border vii

iii. Objective 3: State wise Border Crossing viii

iv. Objective 4: Top Border Crossing ix

v. Objective 5: Geospatial Analysis x

5. Conclusion xi

6. Future scope xii

7. References xii

# 1. Introduction

Border crossing data represents the count of people, vehicles, and goods entering and exiting a country through authorized ports. Analyzing this data can help understand international traffic flow, port utilization, and trends in travel or trade. This project aims to visualize the trends and patterns in U.S. border crossings through an interactive dashboard in Excel.

# 2. Source of Dataset

The dataset used for this project is titled “Border Crossing Entry Data”, sourced from the U.S. government’s open data platform (https://data.transportation.gov/). It includes data on the number of crossings, categorized by state, port name, type of crossing (e.g., trucks, pedestrians), date, and geographical coordinates (latitude and longitude).

# 3. Dataset Preprocessing

- Removed duplicate records.

- Checked for and handled missing/null values.

- Standardized text formatting (proper case, trimmed spaces).

- Converted date format to standard yyyy-mm.

- Formatted the data as a table in Excel for dynamic referencing.

# 4. Analysis on Dataset (for Each Objective)

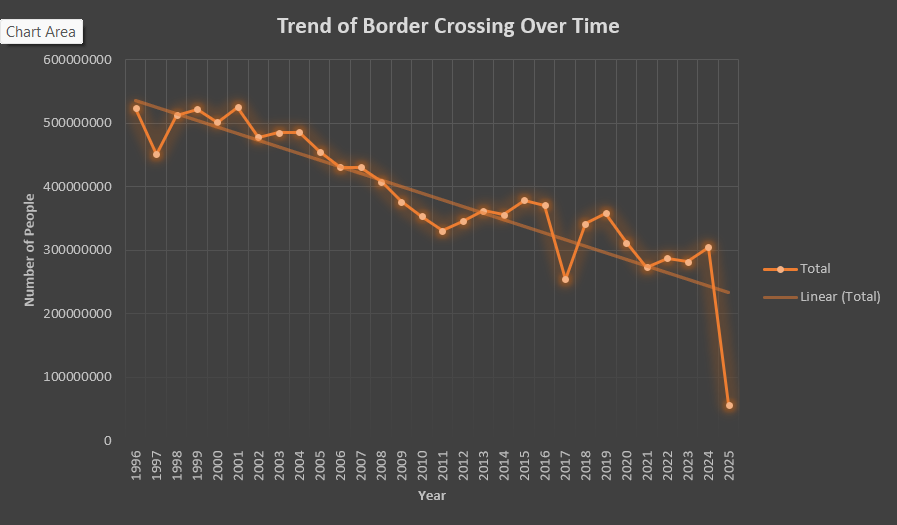
# Objective 1: Trend Analysis of Border Crossings

- General Description: In this objective, I focused on analyzing the overall trend of border crossings over time. The main goal was to observe how the number of people, vehicles, and goods crossing the U.S. borders has changed from the year **1996 to 2025**. This helps in understanding patterns like growth, decline, or seasonal fluctuations in border activity. Such insights are valuable for authorities to monitor international movements and make data-driven decisions.

- Specific Requirements: To perform this analysis, it was necessary to summarize the dataset based on **Date** and Value. I ensured that the dates were properly formatted and grouped them by **month and year** to make the trend more readable and meaningful.

- Analysis Results: I used Excel’s **Pivot Table** feature to group the border crossing values by **date**. I formatted the date column to show **years**, and then applied filters using **slicers** to dynamically view trends for individual crossing.

- Visualization: To present the trends clearly, I created a **Line Chart** using the Pivot Table output. The chart displays the number of crossings over time, and I added a **slicer** for the "Measure" field so that viewers can filter and view trends by specific crossing types. This interactive chart allowed me to explore how different categories behaved over the years and identify any major spikes or dips in activity.



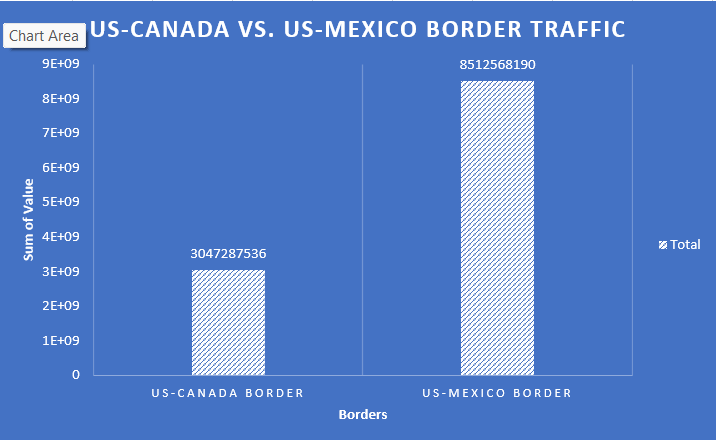
# Objective 2: Comparison of US-Canada vs. US-Mexico Border Traffic

- General Description: In this objective, my aim was to compare the total number of border crossings between the two major international borders of the United States — the **US-Canada** border and the **US-Mexico** border. This comparison helped me understand which border is more active in terms of people and goods movement. It also gave insights into cross-border relationships and trade flow patterns between the U.S. and its neighboring countries.

- Specific Requirements: For this analysis, I grouped the data by the **"Border"** column, which clearly indicates whether the crossing took place at the US-Canada or US-Mexico border. I then summarized the **"Value"** column, which represents the total number of crossings, to get the overall traffic volume for each border.

- Analysis Results: After creating a Pivot Table and grouping the data by Border, it was clear that the **US-Mexico border** had significantly higher crossing volumes compared to the US-Canada border. This pattern was visible across most crossing types and time periods, especially for commercial transport like trucks and personal vehicles.

- Visualization: I used a **Column Chart** to represent the comparison visually. The chart shows the total number of crossings for each border side by side, making the difference easy to understand at a glance.



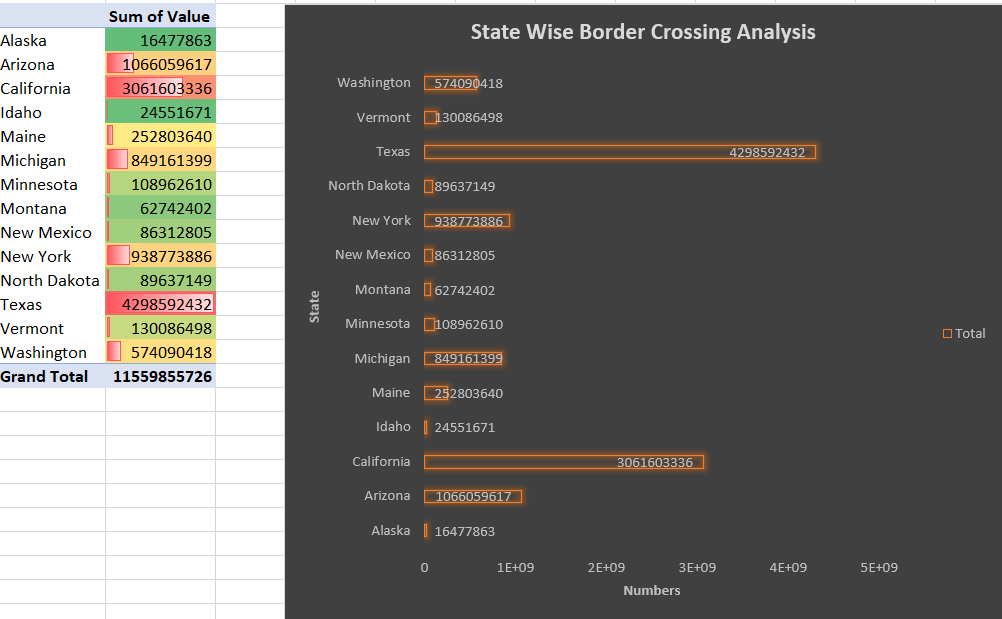
# Objective 3: State-wise Border Crossing Analysis

- General Description: The purpose of this objective was to analyze the dataset from a **state-wise perspective**. I wanted to identify which U.S. states have the **highest number of border crossings**. This analysis is important because some states, due to their geographical location, naturally experience more cross-border activity. By comparing states, I could see which ones are most active and might require more infrastructure, security, or transportation resources.

- Specific Requirements: To achieve this, I grouped the data by the **"State"** column and calculated the **total number of crossings** in each state using the **"Value"** column. This gave me a clear picture of which states are experiencing the highest and lowest traffic.

- Analysis Results: From the summarized data, it was observed that **Texas and California** had the highest number of crossings among all states. This result is expected since both these states have multiple border ports and lie along the US-Mexico border, which sees heavy commercial and personal movement.

- Visualization: I used **conditional formatting with a color scale** (heatmap) in Excel to visually highlight the traffic volumes across different states — states with higher values appeared in darker shades. Additionally, I created a **bar chart** to represent the total crossings per state. This chart made it easy to visually compare all states and instantly spot the busiest ones.



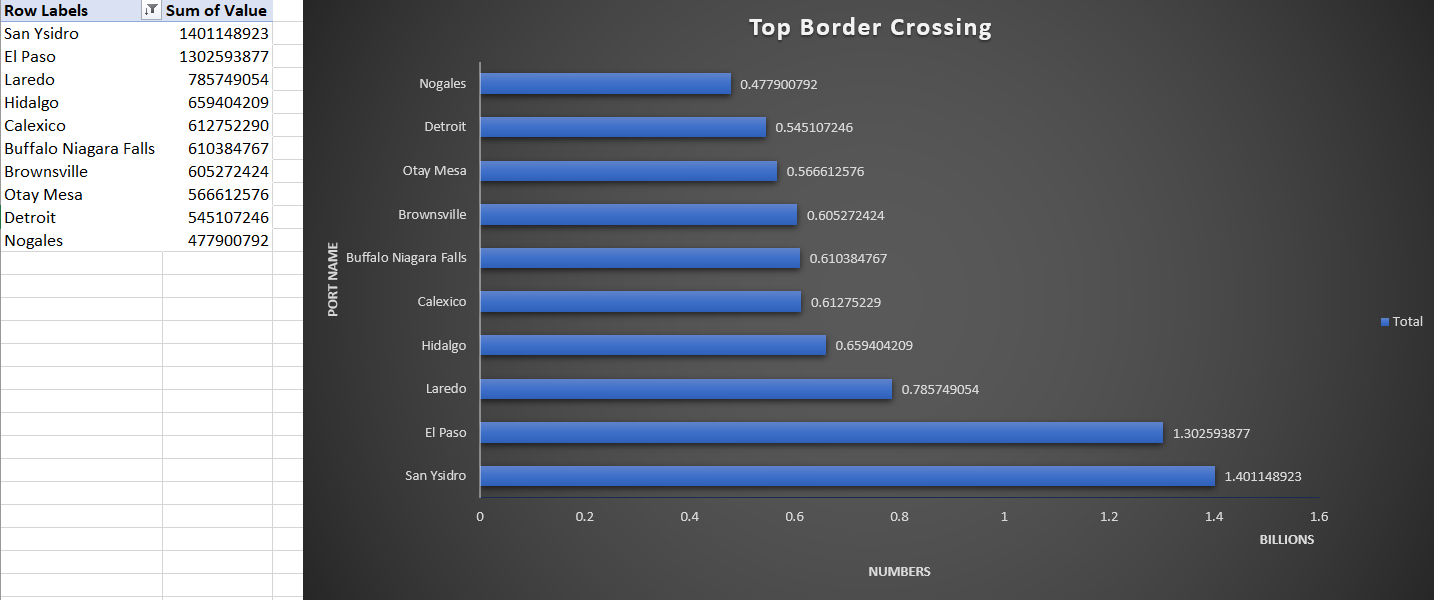
# Objective 4: Top Ports for Border Crossings

- General Description: In this objective, I focused on identifying the **busiest border ports** across the United States. The goal was to find out which specific ports handle the **highest volume of traffic**, regardless of the state they are in. This analysis is useful for understanding port-level operations and highlighting major entry and exit points used by travelers and commercial vehicles.

- Specific Requirements: To complete this analysis, I grouped the dataset by the **“Port Name”** column and calculated the total **“Value”** (number of crossings) for each port. Then, I **sorted** the results in descending order to bring the busiest ports to the top and applied a filter to show only the **Top 10** ports with the highest crossing values.

- Analysis Results: After sorting and analyzing the data, I found that ports like **San Ysidro**, **El Paso**, and **Laredo** consistently appeared at the top in terms of total border traffic. These ports are among the most important ones, especially for trade and large-scale movement of goods between the U.S. and its neighboring countries.

- Visualization: I created a **bar chart** that clearly displays the **top 10 ports** based on their total number of crossings. The chart was sorted from highest to lowest so that it’s easy to see which ports are the most active. This visual representation made the analysis very clear and impactful.



# Objective 5: Geospatial Visualization with 3D Map

- General Description: In this objective, my goal was to create a **geographical visualization** of the border crossing data to see how traffic is distributed **across different port locations**. Instead of just looking at numbers, I wanted to use a **map-based view** to show where the most active ports are located in the U.S. This gives a clear and visual understanding of which areas handle more border movement.

- Specific Requirements: For this, I used the **Latitude** and **Longitude** columns from the dataset to **plot each port on a map**. I also used the **"Value"** column to represent the volume of border crossings. This allowed me to display the intensity of traffic at each port in a more interactive and visual way using Excel’s **3D Maps** feature.

- Analysis Results: The result was a dynamic **3D map** that showed each port as a column on the map. The **height of each column** represented the total number of crossings for that port — the taller the column, the more traffic at that location. Ports like Laredo and El Paso stood out with significantly taller columns, showing that they handle more traffic than others.

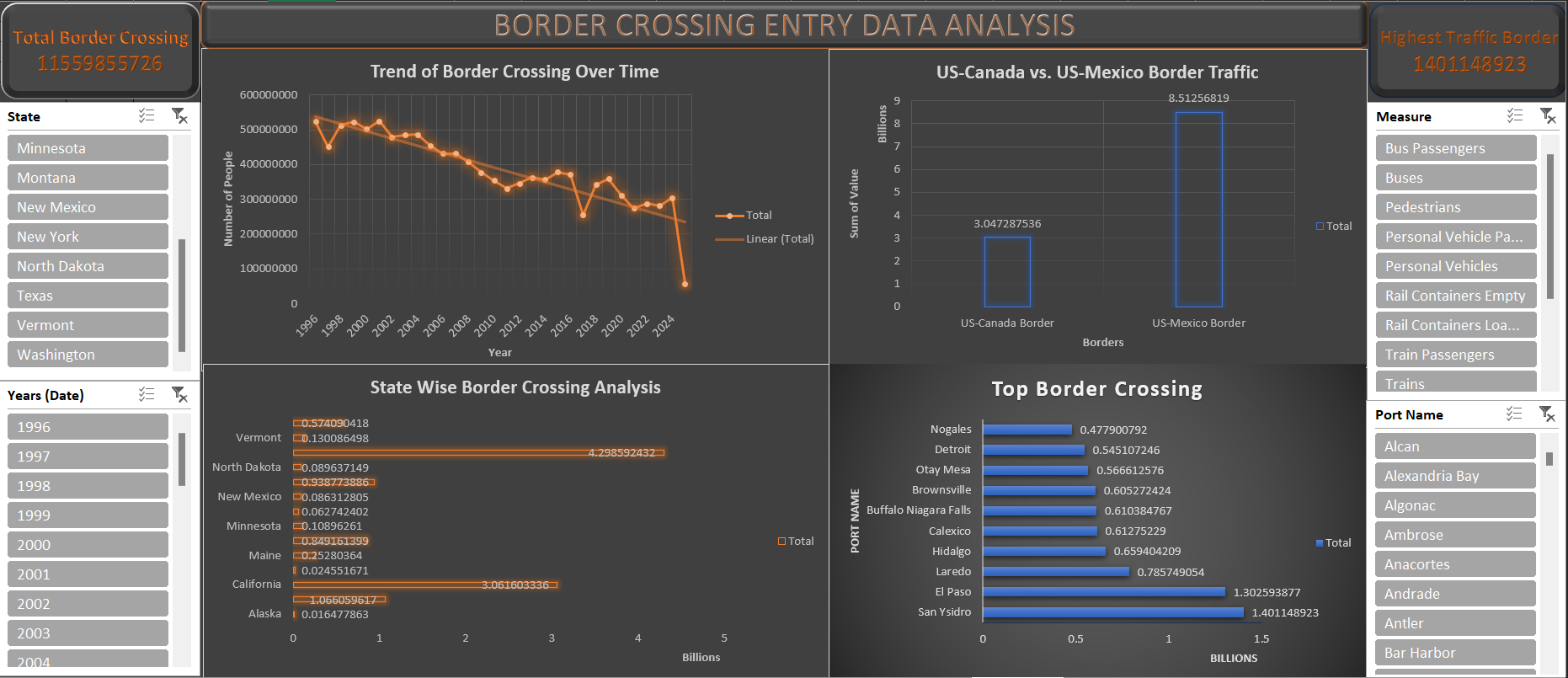
- Visualization: I used **Excel’s 3D Maps** tool to generate the visualization. Each port was plotted accurately using geographic coordinates, and the crossing volume was represented by column height. This made it very easy to spot traffic hotspots and understand the geographical distribution of border activity in a visually impressive way.





# 5. Conclusion

The dashboard offers a comprehensive overview of border crossing trends in the United States over time. It clearly shows a declining trend in total border crossings from 1996 to 2025, with a noticeable drop after 2020, likely due to the COVID-19 pandemic and related travel restrictions. When comparing the two major international borders, the US-Mexico border consistently recorded a much higher volume of crossings than the US-Canada border, highlighting its greater significance in terms of travel and trade. The state-wise analysis reveals that Texas and California are the leading states for border activity, indicating their strategic importance due to geography and infrastructure. Furthermore, port-level data emphasizes that San Ysidro, El Paso, and Laredo are the busiest ports, handling the largest share of cross-border movement. Overall, the dashboard successfully meets all objectives, offering valuable insights into historical patterns, regional distribution, and the most active entry points across the country.



# 6. Future Scope

- Automate data updates via Power Query.

- Integrate forecasting using Power BI or advanced Excel tools.

- Build a real-time dashboard using Python or web tools.

- Add anomaly detection for suspicious border crossing surges.

# 7. References

[1] Border Crossing Entry Data. U.S. Department of Transportation. [dataset link](https://catalog.data.gov/dataset/border-crossing-entry-data-683ae)

[2] Microsoft Support - Use 3D Maps in Excel. https://support.microsoft.com/3d-maps-excel

[3] How to use Pivot Tables in Excel - Microsoft Learn