Border Crossing Entry Data
Project 3 - Group 10

I. Introduction and Inspiration

In this project, our team: Ethan Tebbe, Gavin Plemon, Leon Harris, and Sarah Ruth, chose a dataset called Border Crossing Entry Data. We found it interesting since travel across the U.S. land border is a major driver of the U.S. economy and tourism sector. Cross border trade by land accounted for $1.27 trillion in 2022 (Department of Transportation 2023) and 199,771,070 people crossed the border on land through ports of entry in 2022 (CBP 2024). We initially found the dataset on Kaggle.com, but decided to use the actual dataset from the U.S. Department of Transportation website. The dataset can be found here:

https://data.bts.gov/Research-and-Statistics/Border-Crossing-Entry-Data/keg4-3bc2/data

What is the Border Crossing/Entry Data?

The Border Crossing/Entry Data provides information on vehicles/equipment, passengers, and pedestrians entering the United States through land ports on the U.S.-Canadian and U.S.-Mexican border. The Bureau of Transportation Statistics (BTS) obtains this data, once a quarter, from the U.S. Customs and Border Protection. BTS then assesses, analyzes, summarizes, and disseminates monthly and annual Border Crossing/Entry data. Most data elements are available beginning in 1994 (U.S. Department of Transportation, 2021).

II. Data Cleaning and Filtering

For data cleaning, we decided to use Python in a Jupyter Notebook. We read the CSV file into a Pandas dataframe and then performed a few operations to get the data prepared.

After an initial inspection, we made a list of changes to be made to the data: a few data types were off, we wanted to change the column names, and we wanted to reorganize some columns and drop some columns. We decided to change the column names to be more usable for coding first. We returned a list comprehension of the column names with all characters in lowercase and all spaces replaced with underscores.

Second, we split the date column into two separate columns, month and year. Our year column was not an integer, so we changed it to be one. We took the month column and transcribed it from English months (Jan, Feb, Mar) into numeric months (1, 2, 3). After clearing up the date column, we dropped that column as well as the point column–which we did not see ourselves using in the final product.

After those steps were complete, the data was clean but not filtered to our needs. Records in the DataFrame totalled over 100,000 entries–a gargantuan number to be shoving into a sqlite database for our project. We needed to trim the records down to a manageable amount. We decided to use a range of 10 years, from 2014 until 2024. To do this, we applied a filter to our DataFrame where the year is greater than or equal to 2014. We exported this DataFrame as a CSV and then used sqlalchemy to create a database from the DataFrame.

III. Backend

For the backend of the website, we decided to use Python's flask library. We defined five HTML routes: /, /dashboard, /map, /about_us, and /works_cited. For the route used to interact with the database, we took the approach of making the call into an application programming interface call. We read the additional specifiers on the route /requestData–this is done via the request package. After the initial URL, additional variables can be specified following a question mark. For the initial specifier, a question mark is used as a marker for the clause; but, for all specifiers past the first, an ampersand is used to mark the clause. All possible specifiers are:

- Select: a comma separated list of columns to return.
- Where: a comma separated list of conditions to filter data with.
- Order By: a comma separated list of value and direction pairs. Each pair is separated by a colon–the value is which column to order results with, and the direction can be ascending or descending (ASC / DESC).
- Group By: a comma separated list of values to group results by.
- Limit: an integer specifying the amount of results to return.

With all of these routes defined, the rest of the work to be done was in the HTML, CSS, and JavaScript for the site pages.

IV. Design Concepts

For our responsive web design we used Bootstrap with Bootswatch(LUX). We decided to use our dashboard to display the filterable data through a line chart and donut chart for visualizations. Using the latitude and longitude found in our data, we created a map page showing all the border crossings that can be filtered as well. For the color scheme, we agreed on a color palette chosen from the website coolors.co – the link to our specific palette is:

https://coolors.co/palette/177e89-084c61-db3a34-ffc857-323031

V. Guiding Research Questions

Our visualizations and maps were mainly meant to answer these three questions. What are the patterns and volumes of the various modes of transport at entry ports into the U.S.? Has the overall traffic at U.S. ports of entry increased or decreased over time? Are these changes localized to specific ports or regions? Some of the inspirations for these questions were driven

by the recent pandemic. We were curious to see how the pandemic affected the data, given so much of the data set pertained to passenger travel. We figured it would affect some areas and methods of travel differently. For example, was cargo traffic affected to the same magnitude as car passenger numbers? And were rural crossings as affected as urban crossings? Next, we will go over what information the website shows, and how it answers these questions.

VI. About Our Website and How it Answers Our Questions

The website is made up of 5 main parts. Our “Home page” summarizes the website and explains how to navigate. The “Works Cited” page lists the resources used in creating the website and visualizations, and the “About Us” page provides a brief bio on the collaborators. The final two and most important parts are the “Map” and “Dashboard”. They inform and allow users to interact with the data in a way that can teach them the trends throughout this set and answer the questions we had proposed when initially reviewing the data. The dynamic map gives a geographic visualization of the ports of entry from Canada and Mexico, with a marker that can be clicked to open a sunburst chart containing the individual port information since 2014. The interactive dashboard includes a line chart with crossings grouped over time and a donut chart with the percentage of crossings by measure. These measures can be filtered by port name, state, port code, border, vehicle type, values, dates, latitude, and longitude.

VII. Conclusion

We had several research questions regarding the Border Crossing Entry Data. The first question was to find what the most popular transport method is. Passenger travel by vehicle is the most popular and pedestrian travel comes in a far second. We also wanted to look at the entry volume over time. There were large drop-offs in passenger travel during the pandemic although the train cargo traffic stayed largely the same. July and August are the busiest months for passenger travel in non-pandemic years. Finally, we wanted to see the difference between the Canadian/U.S. border crossings and Mexican/U.S. border crossings. The Mexican border receives significantly more traffic than the Canadian border with San Ysidro, California being the busiest Mexican port of entry and Buffalo, New York being the busiest Canadian Port of entry.

Future work that could be done would be to add additional visualizations. A heatmap in addition to the current map would help identify major crossing areas at a glance. We could have a boxplot for each category of travel with each month a datapoint. A bar chart could show the differences between each border. It would also be useful to find additional data that includes sea and air travel to compare with land entry and data to include cargo by volume.

We have identified some biases and limitations to the dataset. The data that we worked with did not include entry into the U.S. by air or sea, only land. Also, the data does not account for unique vehicles that may travel back and forth over the border multiple times per day. This could skew the results. Since the data is collected from over one hundred ports, this could lead to incomplete data or different standards for recording. Finally, geopolitical and environmental factors can affect the flow of people across the border, such as the pandemic, immigrants escaping violence in their home country, and natural disasters.

The insights shown in this research are valuable for policy, security, migration management, as well as travel and tourism. Government agencies, researchers, advocacy groups, and the general public can all use this information to aid in policy-making, community support, and solutions to promote change.

Works Cited

U.S. Customs and Border Protection (2024, June 21). *On a Typical Day in Fiscal Year 2022, CBP...* U.S. Customs and Border Protection. Retrieved August 1, 2024, from https://www.cbp.gov/newsroom/stats/typical-day-fy2022

U.S. Department of Transportation (n.d.). *Border Crossing Entry Data*. Bureau of Transportation Statistics. Retrieved August 10, 2024, from https://data.bts.gov/Research-and-Statistics/Border-Crossing-Entry-Data/keg4-3bc2/data

U.S. Department of Transportation (2021, January 12). *Border Crossing/Entry Data*. Bureau of Transportation Statistics. Retrieved August 10, 2024, from https://www.bts.gov/explore-topics-and-geography/geography/border-crossingentry-data#1

U.S. Department of Transportation (2023, August 23). *Transborder Data Mid-Year Report 2023*. Bureau of Transportation Statistics. Retrieved August 1, 2024, from www.bts.gov/data-spotlight/transborder-data-mid-year-report-2023#:~:text=U.S.%20land%20borders%20with%20Canada,trade%20partners%20of%20agricultural%20goods

Wolfe, J. (2024, June 4). *San Diego is Once Again a Top Migrant Entry Point.* The New York Times. https://www.nytimes.com/2024/06/04/us/san-diego-migrants-california.html