

Project Assignment – BORDER CROSSING ENTRY DATA

Meghana Katta

Department of Data Analytics Engineering, George Mason University

AIT 580: Analytics Big data to Information

Dr. Moody Amakobe

April 30, 2023

ABSTRACT:

The number of people who enter the United States by land, air, and seaports is disclosed in the Border Crossing Entry Data dataset. The U.S. Bureau of Transportation Statistics compiles the dataset, which includes information on the method of entry, the location of the port of entry, and the nationality or country of origin of those entering.

This dataset's analysis can provide light on migration trends and patterns, which has crucial ramifications for public safety, resource allocation, economic analysis, and policymaking. The dataset can be used to identify possible problems or safety risks connected to forms of transportation or geographies as well as to study the patterns and trends of border crossings over time.

The dataset can also help with decisions on international trade, immigration, and border security. Policymakers may more effectively allocate resources to ensure they reach people who need them most by using the data to influence their decisions and better serve the public interest.

The Border Crossing Entry Data dataset is a useful tool for scholars, decision-makers, and other stakeholders who are interested in comprehending migration patterns and trends and informing choices on immigration and border security.

INTRODUCTION:

The movement of people, goods, and vehicles across international borders is referred to as a border crossing. Trade, tourism, and other economic and social activity depend on these crossings. However, they also pose issues about public health, immigration, and security. Understanding these issues and creating efficient policies and tactics to solve them depend on the collection of data on border crossings. Such information, which includes the quantity of crossings, the varieties of people and commodities moving across borders, as well as the routes and means of transportation taken, can shed light on trends and patterns in cross-border mobility. The utilization of this information can enhance border security, ease, commerce and tourism, and advance public health and safety.

Data on border crossing entries is important in the context of immigration and border security for several reasons. First and foremost, it is essential for spotting and stopping security threats like terrorism, drug trafficking, and people- or drug-smuggling. Second, it aids with the management of border crossings and the enforcement of immigration regulations by immigration authorities. Thirdly, it aids in locating and keeping an eye on those who might be a threat to national security or public safety.

By giving authorities access to fast and accurate information, effective border crossing entry data management helps enhance border security and immigration control. Additionally, it improves public safety and aids in the prevention of crime and terrorism. However, managing this data presents difficulties due to issues with data security, privacy, and sharing, among other things.

Therefore, it is essential to effectively manage these difficulties while maintaining the harmony between personal privacy and national security.

Research questions:

1. Which states and ports receives the highest number of inbound flows?
2. How does the border crossing volume at the different borders changed over time?
3. what are the frequently used measures at US-Mexico border for different States?
4. Can we develop a regression model to predict the volume of border crossings at a given port and date based on the type of border crossing measure, latitude, and longitude?

How information about border crossings is gathered and processed.

Depending on the country, the technology, and the resources available at the border crossing, entry data is gathered and processed using a variety of methods. Here are a few typical approaches:

- **Personal Information:** At the border, officials may inquire about a person's name, nationality, date of birth, and specifics of their travel documents. The border officer manually enters this data into the immigration system.
- **Biometric Information:** Additional information such as fingerprints, iris scans, and facial recognition may also be gathered. This calls for the use of specialist tools like biometric scanners, which record the person's distinctive biometric markers. Then, for later use, this information is kept in a database.
- **Automated Passport Control (APC):** Kiosks for Automated Passport which feature touchscreens where passengers can input their contact information and biometric information like fingerprints and facial recognition. The kiosk then checks the information and gives the traveler permission.
- **Electronic Travel Authorization (ETA):** Before entering some nations, passengers must get an ETA electronically. By uploading personal information and biometric data online, immigration authorities can process it before granting clearance.

Data is gathered, analyzed, and then safely stored in a database for later use. To ascertain whether the traveler poses any risk or is in conformity with immigration laws, this data is cross-referenced against several security and immigration databases. The information is also utilized statistically, such as to monitor travel trends. To protect traveler privacy while upholding national security, it is essential to make sure that this data is managed safely and responsibly.

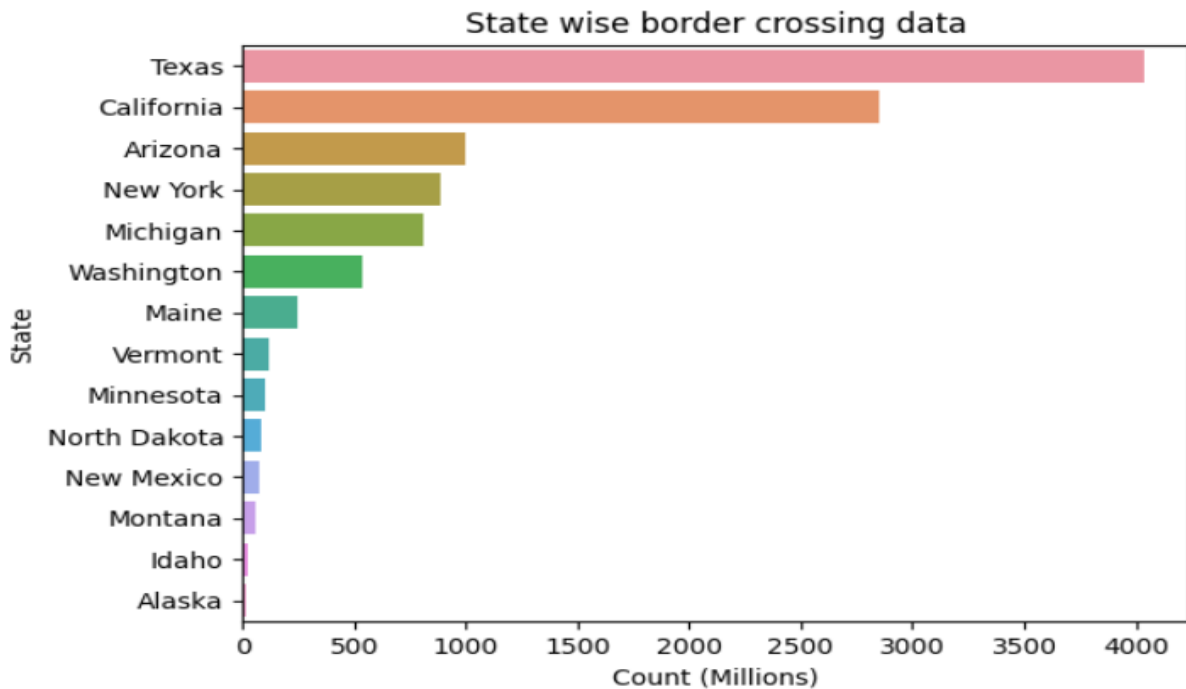
Data Fields: The dataset contains a variety of data fields that can be used to examine border crossing trends and patterns. Port of entrance, date, travel mode, number of crossings, state or province of the port of entry, and kind of crossing are a few of the crucial data variables present in the dataset.

Data set:

1	Port Name	State	Port Code	Border	Date	Measure	Value	Latitude	Longitude	Point		
2	Del Rio	Texas	2302	US-Mexico B	Feb-23	Trucks	5823	29.327	-100.928	POINT (-100.927612 29.326784)		
3	Roma	Texas	2310	US-Mexico B	Aug-09	Trains	0	26.404	-99.019	POINT (-99.018981 26.403928)		
4	Skagway	Alaska	3103	US-Canada B	Feb-23	Pedestrians	1	59.63	-135.164	POINT (-135.164444 59.629722)		
5	Brownsville	Texas	2301	US-Mexico B	Feb-23	Trains	70	25.952	-97.401	POINT (-97.40067 25.95155)		
6	Sault Sainte I	Michigan	3803	US-Canada B	Feb-23	Truck Contai	2322	46.509	-84.361	POINT (-84.360833 46.508611)		
7	Lukeville	Arizona	2602	US-Mexico B	Feb-23	Pedestrians	6383	31.88	-112.817	POINT (-112.817077 31.880195)		
8	Dalton Cache	Alaska	3106	US-Canada B	Feb-23	Personal Veh	294	59.451	-136.362	POINT (-136.361944 59.450556)		
9	Calexico East	California	2507	US-Mexico B	Feb-23	Pedestrians	39683	32.673	-115.388	POINT (-115.388018 32.673389)		
10	Pembina	North Dakot	3401	US-Canada B	Feb-23	Trains	89	49	-97.237	POINT (-97.237036 49.000453)		
11	Laurier	Washington	3016	US-Canada B	Feb-23	Bus Passenge	23	49	-118.224	POINT (-118.223777 49.000083)		
12	Lancaster	Minnesota	3430	US-Canada B	Feb-23	Truck Contai	12	49	-96.801	POINT (-96.800527 49.000194)		
13	Sault Sainte I	Michigan	3803	US-Canada B	Feb-23	Truck Contai	526	46.509	-84.361	POINT (-84.360833 46.508611)		
14	Fortuna	North Dakot	3417	US-Canada B	Feb-23	Truck Contai	21	49	-103.809	POINT (-103.80925 48.999555)		
15	Porthill	Idaho	3308	US-Canada B	Feb-23	Pedestrians	3	49	-116.499	POINT (-116.49925 48.999861)		
16	Otay Mesa	California	2506	US-Mexico B	Feb-23	Trains	9	32.55	-116.939	POINT (-116.938538 32.55033)		
17	Ferry	Washington	3013	US-Canada B	Feb-23	Trucks	31	49	-118.761	POINT (-118.761166 49.000083)		
18	Alcan	Alaska	3104	US-Canada B	Feb-23	Personal Veh	1144	62.615	-141.001	POINT (-141.001444 62.614961)		
19	Richford	Vermont	203	US-Canada B	Feb-23	Pedestrians	1	45.012	-72.589	POINT (-72.588559 45.01174)		

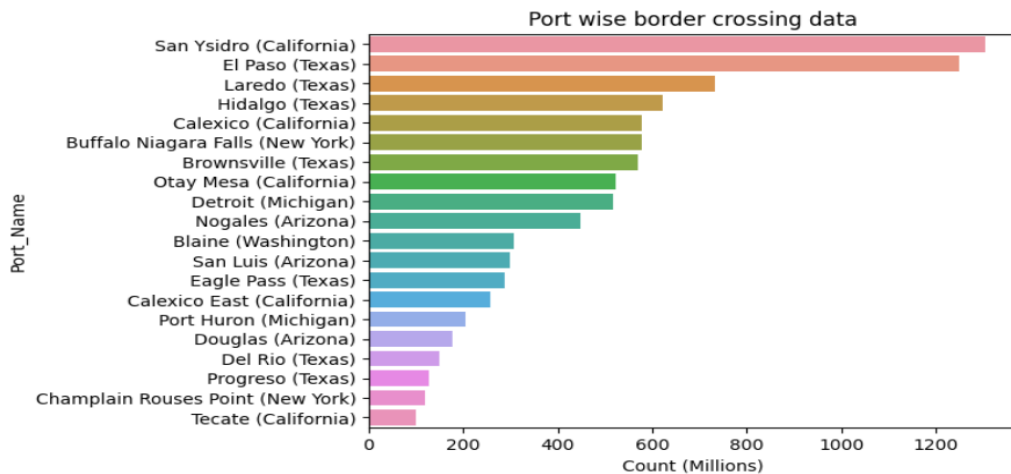
RESEARCH QUESTIONS:

1. Which states and ports receive the highest amount of inbound flow?



Interpretation:

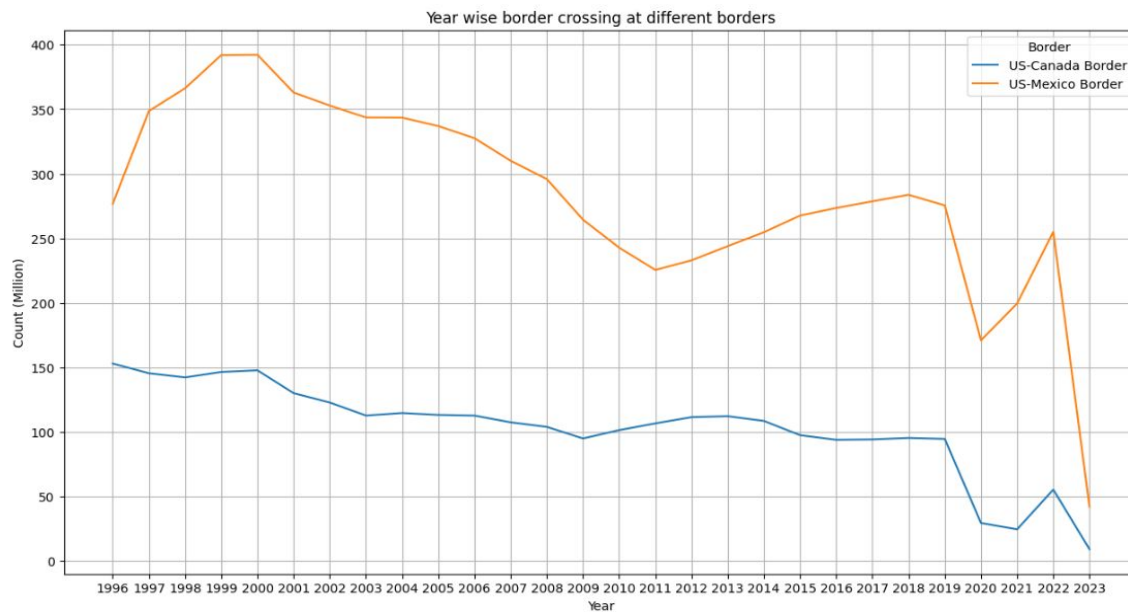
Texas has the highest volume of border crossings, followed by California and Arizona. Texas, California, and Arizona are the top three states along the US-Mexico border.



Interpretation:

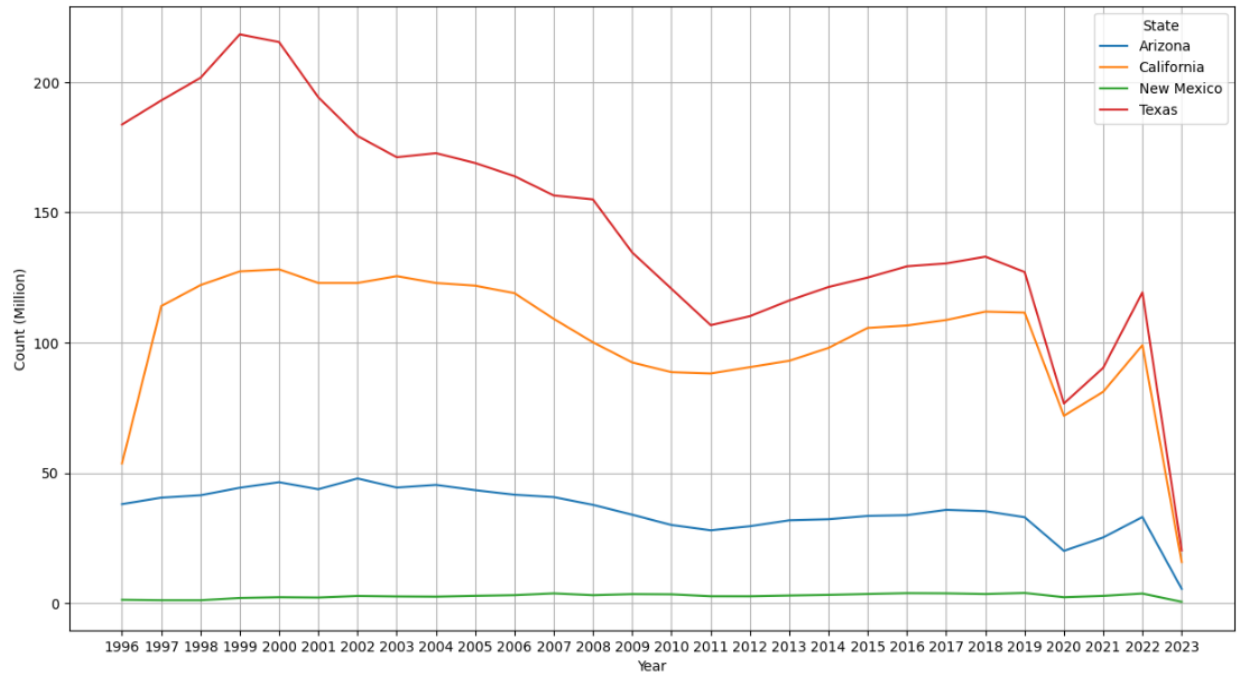
Out of 116 ports, the top 30 ports with the most border crossings are represented in the plot. The highest inbound traffic occurs at San Ysidro in California, followed by El Paso and Laredo in Texas.

2.How does the border crossing volume at the different borders changed over time?



Interpretation:

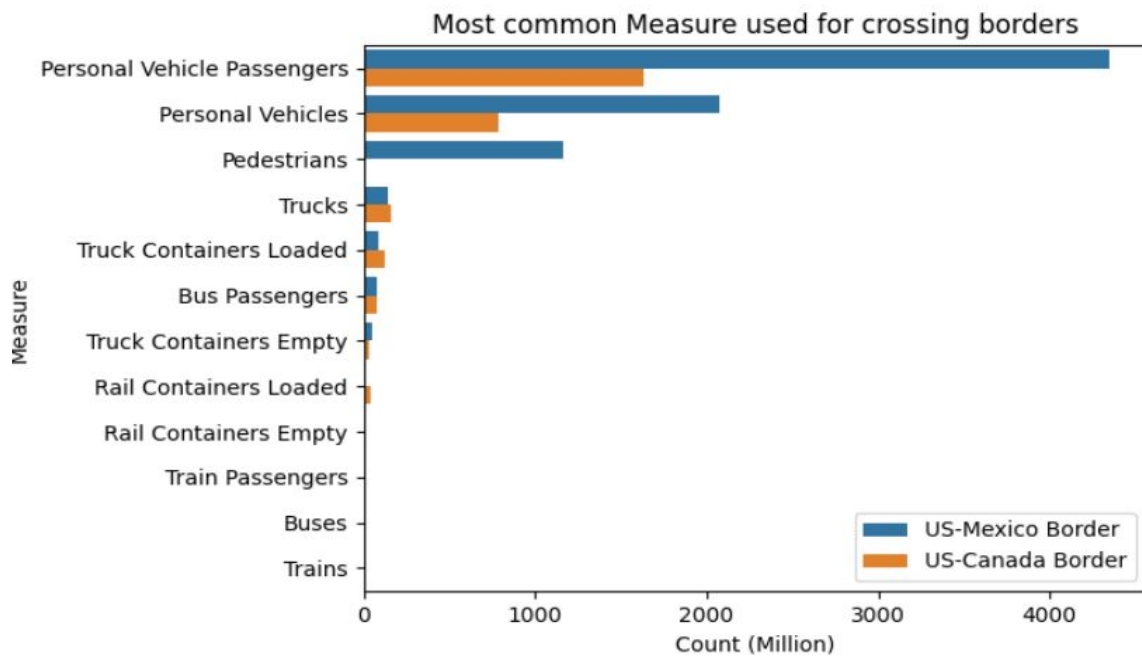
When compared to the US-Canada border, the border crossing volume at the US-Mexico border is higher every year. Volume spike can be seen in the years 1999 and 2000, while a dip in volume can be seen in 2020 due to pandemic.



Interpretation:

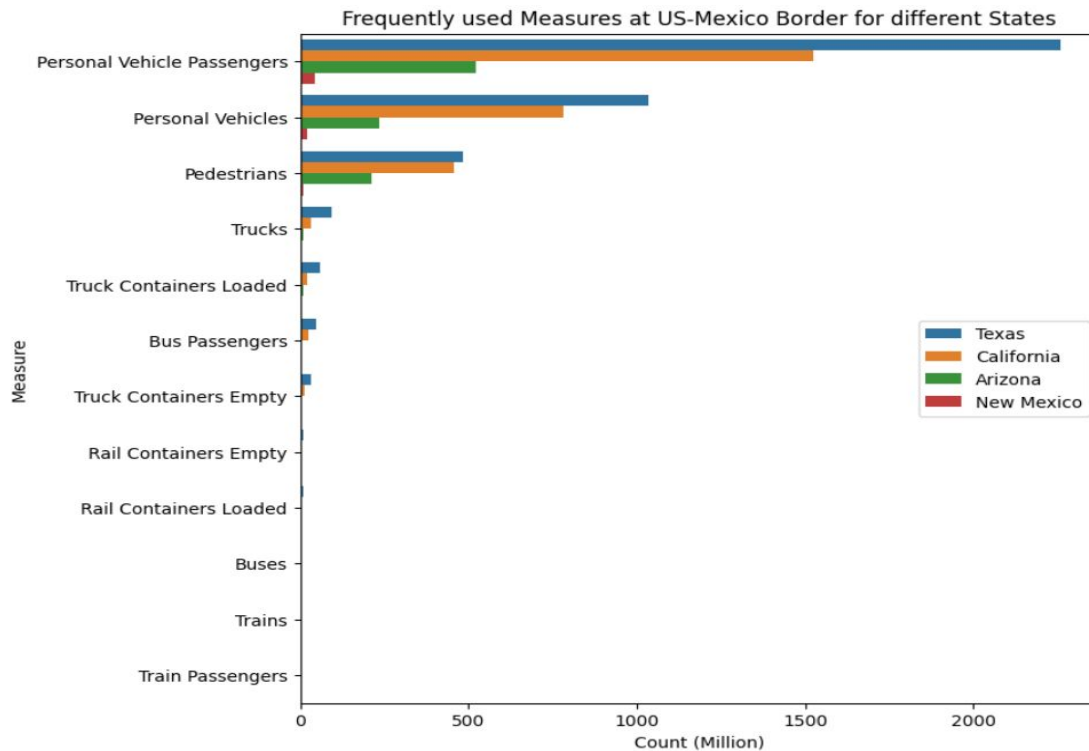
Peak volumes along the US-Mexico border in 1999 and 2000 can be attributed to Texas. And Texas state has the biggest volume of the four states near the US-Mexico border.

3.what are the frequently used measures at US-Mexico border for different States?



Interpretation:

At both borders, personal vehicle passengers is the most popular method of border crossing. Personal vehicles are the second-most popular border crossing method.



Interpretation:

Most border crossings in the states around US-Mexico border involve, personal vehicle passengers, followed by pedestrians and personal vehicles.

4.Can we develop a regression model to predict the volume of border crossings at a given port and date based on the type of border crossing measure, latitude, and longitude?

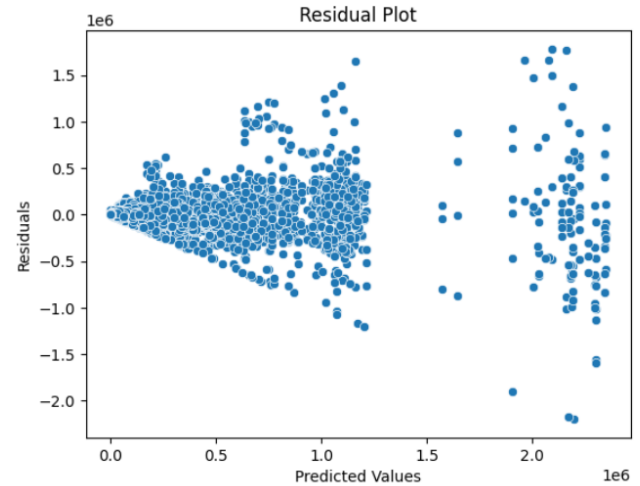
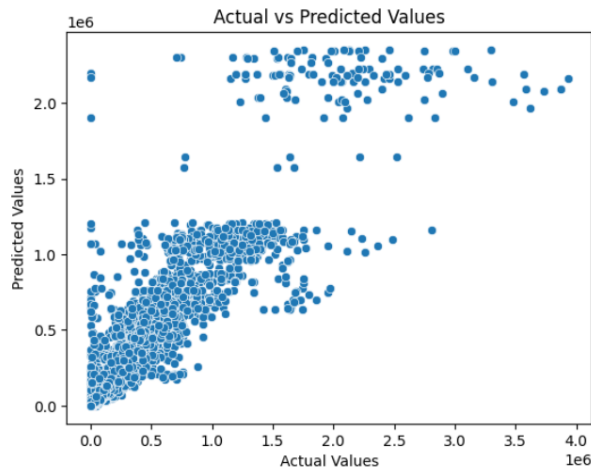
Model Used: Random Forest model (Regression)

Feature Variables: Month, Latitude, Longitude, Measure

Target Variable: Value

Result metrics of model

R-Squared: 0.87



Potential uses:

The dataset for border crossings has a variety of potential applications, including studying historical trends in border crossings, identifying trends in modes of travel, and guiding immigration and border security policy decisions. Additionally, it can be used to strengthen border security, ease commerce and tourism, and advance public security and safety.

Limitations:

The Border Crossing Entry Data dataset has certain drawbacks in addition to its many advantages. These include limitations in the scope of the data, incomplete or inaccurate data, and potential biases in the data collection process.

Conclusion:

In handling border security and immigration control, border crossing entrance data is essential. Border officials may confirm travelers' identities, follow their activities, and spot those who might be a security danger by gathering and analyzing personal, biometric, and travel data. Several advantages can come from effective administration of this data, including improved immigration control, simplified border processing, statistical analysis, and cross-border cooperation.

However, managing border crossing entry data comes with several difficulties, including balancing security and convenience while considering privacy issues, data protection laws, technical infrastructure, integration, and interoperability. It is crucial that border officials deal with these issues and make sure that data is gathered and managed securely and responsibly while safeguarding traveler rights and privacy.

Overall, effective administration of border crossing entry data necessitates considerable thought and preparation but can have a big impact on economic development, immigration control, and national security.

REFERENCES:

- [1] U.S. Bureau of Transportation Statistics. (2021, August 30). Border Crossing Entry Data. [Online]. Available: <https://catalog.data.gov/dataset/border-crossing-entry-data-683ae>
- [2] Python - <https://www.python.org/>
- [3] Jupyter Notebook - <https://jupyter.org/>
- [4] Seaborn - <https://seaborn.pydata.org/>
- [5] Scikit-learn - <https://scikit-learn.org/stable/index.html>