Project Assignment – 2 (BORDER CROSSING ENTRY DATA)

Santosh Vardhan Nayudu(G01414178)

Leela Sagar Gudhe(G01408516)

Meghana Katta(G01408184)

Department of Data Analytics Engineering, George Mason University

AIT 580: Analytics Big data to Information

Dr. Moody Amakobe

April 16, 2023

Project Initiation.

We have selected this data from the internet, we have selected this dataset because it helps us to identify the migration patterns between US -Mexico and US- Canada border and also the imports and exports ratio of can be determined because of which we can implement the new rules and also, we can provide the new resources allocation for services.

Problem Identification.

The data also does not include data on the number of border crossings by different types of people, such as citizens, non-citizens, and tourists.

The expected benefits of the project to the public good are:
• **Economic activity boost:** This project will improve economic activity by making it easier for things and people to pass through borders internationally.
• **Increased protection:** By making it tough for terrorists and criminals to cross borders, the project will increase protection for the country.
• **Enhanced environmental protection:** The project will enhance environmental protection by reducing the number of trucks on the roads.
• **Identifying migratory patterns:** The dataset can be used to identify historical border crossing patterns and trends. Policymakers can decide wisely on resource allocation and border security measures by studying this data.
• **Informing policy choices:** The dataset can be used to help make choices on immigration, trade, and border security. Policymakers can make decisions that better suit the needs of the public by knowing the data .

Expected benefits of the project to the 'public good,' outweigh potential risks to certain populations.

Data Discovery, Inventory, Screening, & Acquisition.

Since the data was obtained from official sources that themselves obtained their information from border sources, there is no disproportionate coverage of different regions in the data.

The geographic coverage of the data is sufficient. A sizable chunk of the United States' land border with Canada and Mexico, which spans 13 states, is represented in the data set by information on border crossings. In order to ensure that the data are representative of the entire nation, the data collection also contains states from both the northern and southern boundaries of the United States.

Data Wrangling.

According to study, the data quality is quite strong and includes considerable data from 13 different US states.

Due to the fact that the data only covers information from 13 states, it is not indicative of the full United States. The data also excludes information on the number of border crossings by other groups of persons, including tourists, non-citizens, and citizens. Because it lacks information on the number of border crossings by other groups of persons, the data is not representative of the populations of those groups. Additionally, the collection

includes current information on border crossings.

Fitness-for-Use Assessment.

The results are limited by the fact that neither the number of border crossings by different categories of persons nor the data on how many of the 18 states that border Canada and Mexico are included.
As a result, the findings cannot be used to determine the causes of the variations in border crossings among various states and groups of individuals. Given the goal of the study, the findings are valuable because they reveal the quantity of border crossings in 13 different states. It is possible to utilize this data to spot potential issue locations and follow trends in border crossings.

According to the data analysis, the prospective advantages are supported by increased economic activity and improved environmental protection. However, because the data does not contain information about the different categories of persons, improved security cannot be a benefit.

Statistical Modeling & Analysis.

Since we employed standard data analysis techniques and anyone can use the analysis methods, the data used and the data analysis methods used in this research are both open and transparent.

Communication and Dissemination.

The methods utilized are standard data analysis approaches, and the data is readily accessible.

References.

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

Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020, February 21). *Doing data science: framework and case study*. Harvard Data Science Review. Retrieved April 16, 2023, from https://hdsr.mitpress.mit.edu/pub/hnptx6lq/release/10