**Paul Nguyen  
Student ID: 2914916  
MSDE692\_X40  
Week 3 Project Update**

In this project, we will perform various exploratory analyses along with constructing an ARIMA for the US\_border\_crossing dataset from Kaggle. The dataset for this project was obtained from the link below. In contains 7 attributes and over 35,000 lines. The attributes are: Port\_name, State, Port\_code, Border (Canada or Mexico), Date, Measure, and Value.

**Dataset Link:**

<https://www.kaggle.com/divyansh22/us-border-crossing-data>

**Timeline:**

Week 2- Obtain and clean the dataset (Done)

Week 3- Perform exploratory analyses (Done)

Week 4- Create graphs/charts

Week 5- Construct the ARIMA model

Week 6- Create presentation

Week 7- Revise codes and finalize presentation

Week 8- Present project and provide feedbacks to classmates

**PROPOSED ‘TO DO’ FROM THE LAST WEEK**

1. Create bar graph on most frequently crossed states/ports
2. Closer examination of bordering states

**THIS WEEK’S PROGRESS**

Some of the preliminary exploratory analyses for this dataset are as follow: the US-Canada border has 27.7 billion occasions in crossings, while US-Mexico border has 7.3 billion crossings. The most popular methods of crossing the border are personal vehicles, followed by pedestrians, and trucks.

**ISSUES AND DISCUSSION**

Initially, it looks like there are far more crossings at the US-Canadian border than the US-Mexican border, which is a surprising sign. Hopefully further analyses will unveil underlying reasons.

**TODO**

1. create bar graphs and pie charts of crossing occurrences
2. calculate the mead, median, and mode of crossings at each port for bordering states.