**Paul Nguyen  
Student ID: 2914916  
MSDE692\_X40  
Week 4 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 (Done)

Week 5- Construct the ARIMA model

Week 6- Create presentation

Week 7- Revise codes and finalize presentation

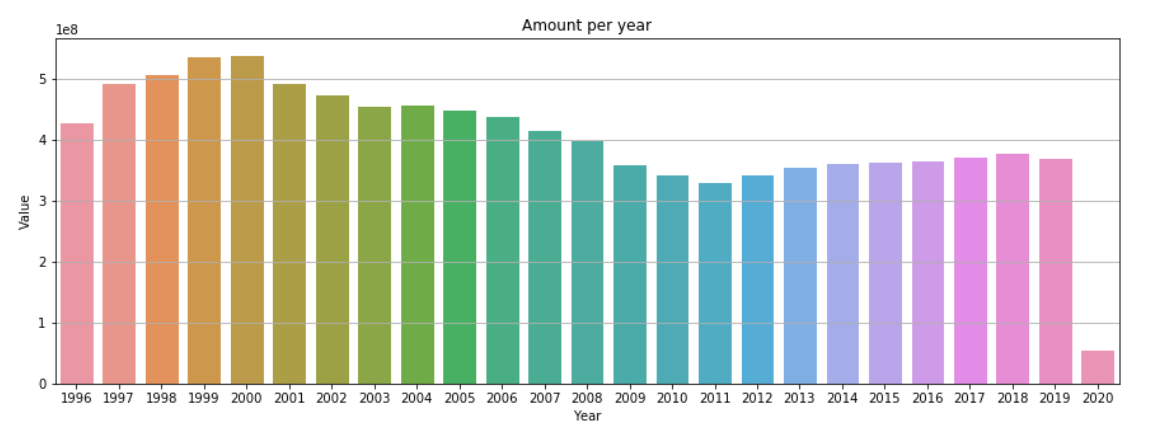
Week 8- Present project and provide feedbacks to classmates

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

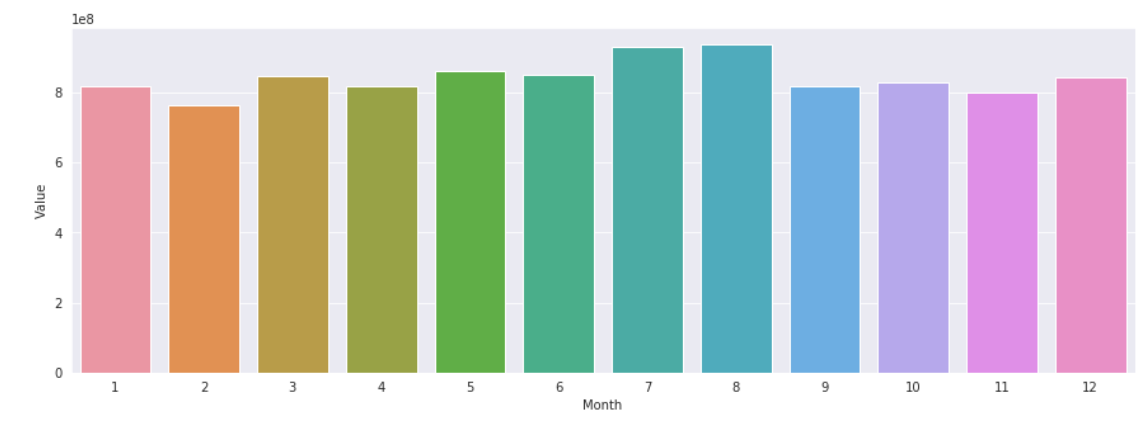
1. Perform descriptive analyses
2. Create visuals for various metrics

**THIS WEEK’S PROGRESS**

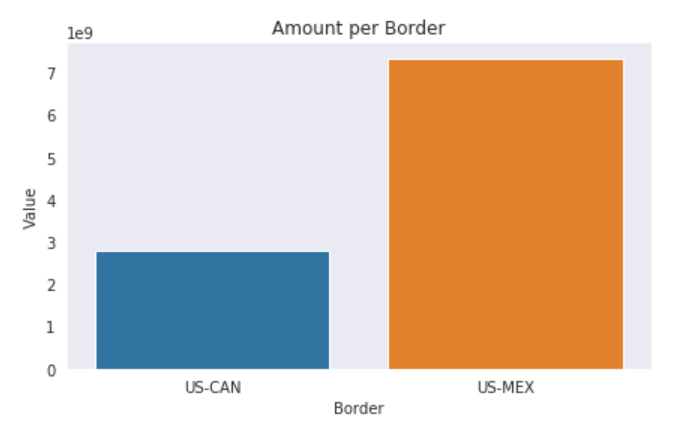
Below are screenshots of the border crossing occurrences on a yearly basis.



Here is a screenshot of the number of crossing grouped by month.



Comparing Canadian and Mexican borders.



**ISSUES AND DISCUSSION**

Some of the information revealed from the graphs above appear predictable such as the number of crossings at the Mexican border far outnumbers the number of crossings at the Canadian border. Furthermore, we can see that July and August have the highest historical number of crossings.

**TODO**

1. Begin construction of the ARIMA model