**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 (Partially done)

Week 6- Create presentation

Week 7- Revise codes and finalize presentation

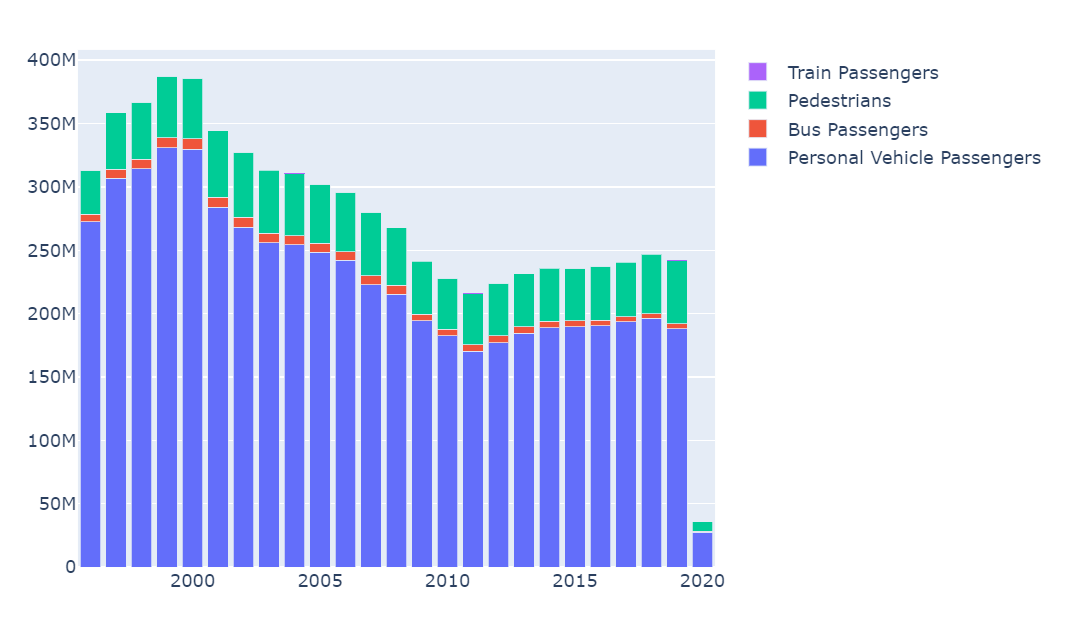
Week 8- Present project and provide feedbacks to classmates

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

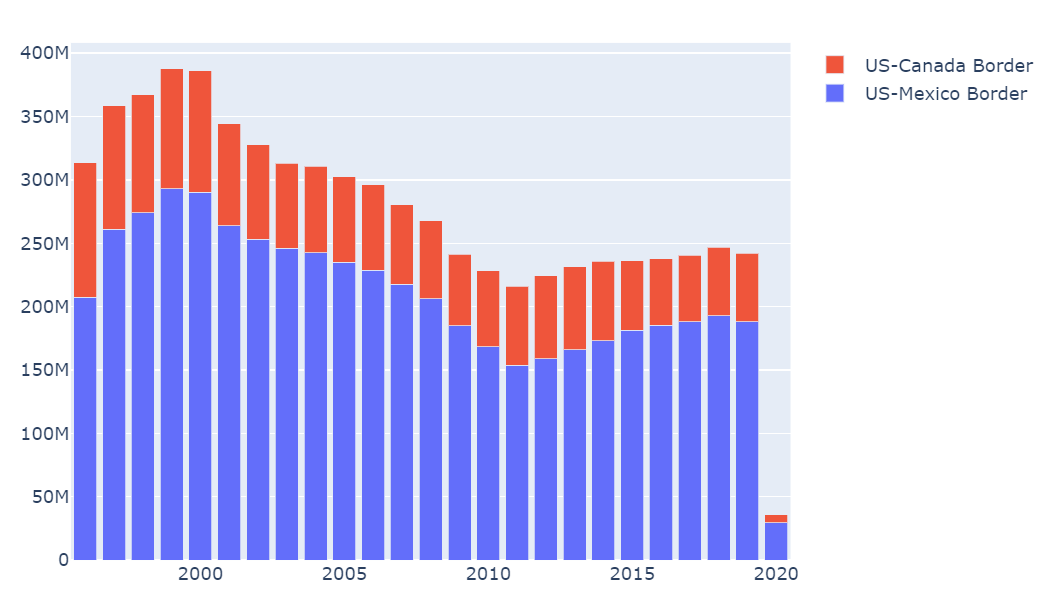
1. Begin construction of the ARIMA model

**THIS WEEK’S PROGRESS**

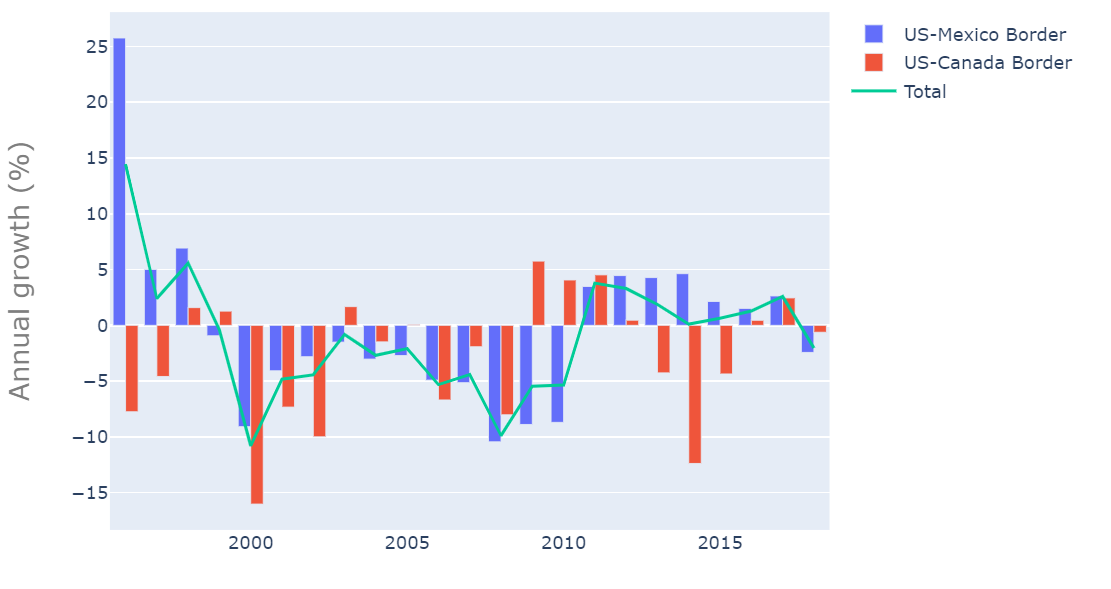
Below are a few more graphics that offer insights into this dataset. This indicates the crossing occurrences over the years, and broken down by methods of transportation:



This compares the ratio of crossing between the Mexico and Canadian borders by year:



Here is the fluctuation in growth for crossing occurrences by year:



**ISSUES AND DISCUSSION**

The descriptive analysis for this week proves to be a bit more inciteful. For example, we can see most border crossings are carried out by personal vehicles. Likewise, the vast majority of crossing occurrences take place at the US-Mexican border. One question to ponder is; why is there such a shrink in border crossings from 2000 to 2010?

**TO DO**

1. Finish the ARIMA model
2. Create presentation