**Paul Nguyen  
Student ID: 2914916  
MSDE692\_X40  
Week 6 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 (Done)

Week 6- Create presentation (Partially done)

Week 7- Revise codes and finalize presentation

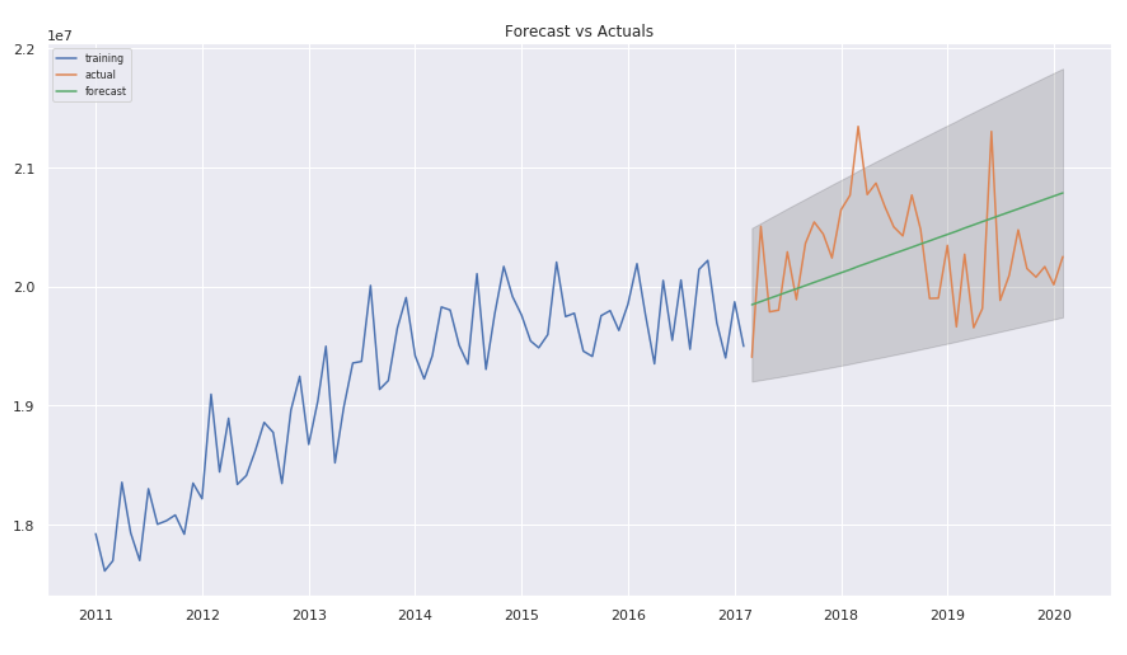
Week 8- Present project and provide feedbacks to classmates

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

1. Finish the ARIMA model
2. Create presentation

**THIS WEEK’S PROGRESS**

As we can see below, the ARIMA model used data prior to 2017 to train itself and predicted the border crossing occurrence trend for 2020. In the presentation, we will discuss the pattern of crossings based on weather, which makes sense when weather is the most ideal to make the journey. Therefore, in the ‘actual’ line below, we can see the highest-to-date spikes are in summer of 2018 and 2019. We can safely speculate that these higher spikes are due to the current political wrestle over immigration control methods. People are fearing that crossing the border will be increasingly difficult. Therefore, they are crossing sooner rather than later.



**ISSUES AND DISCUSSION**

The ARIMA model used 95% confidence level (alpha= 0.05)

**TO DO**

* Finish presentation and finetune graphics