Stat 343: Time Series Analysis

Junta Callahan and Ethan Norton

Professor Gong

30 October 2020

Final Project Proposal

**Part I: Group Members, Project Title, Presentation Date**

Junta Callahan and Ethan Norton will be presenting Time Series Data: U.S. Environmental Protection Agency S02 pollution on November 11th, 2020.

Retrieved from the U.S. Environmental Protection Agency, Junta Callahan and Ethan Norton will attempt to forecast sulfur emissions in California. Sulfur emissions can be defined as being released naturally by volcanic activity and is a product of copper extraction as well as fossil fuel contamination. The data is retrieved from 1987 to 2019 with four quarters listed as Winter, Spring, Summer, and Fall. In total we have 81 data points that record SO2 (sulfur dioxide). We are focusing more on the environmental science approach. We believe that possible variables are cars and wildfires. Cars are looking at improving overall efficiency, however there is still room for improvement. Additionally, there has been an increasing level of wildfires recently.

Based on the graph above, we can see that our data contains seasonality since our data contains quarterly data recorded by seasons. We also know this by observing the plotted data. We can identify a regular pattern over a one-year calendar period, where we see these dramatic “humps”. These humps are most likely during the summer since the sulfur content is highest during this time than any other seasons based on viewing our data points. We have also noticed that there is no trend in our plotted data since there is no increase or decrease of sulfur content over time. Lastly, we know that our data is nonstationary based on the ACF plot and by also conducting the Ljung-Box test. If we observe the ACF plot, we can see that some of the lags are not within the shaded region therefore concluding that our data is nonstationary. We also conducted a Ljung-Box test where we had a p-value less than 0.05. Therefore, we reject the null hypothesis and conclude that we have sufficient evidence that our data is nonstationary.

If approved, we are aiming to forecast about three years of sulfur emissions (split into four quarters). We will look into forecasting future sulfur emissions. In total, this data is crucial to study due to future levels of emissions. If these levels are projected to worsen, then we should be mindful of that.