Exploring US Economic Forecasts and other Insights

**INTRODUCTION**

I am looking at several economic factors to see how well the United States of America is doing.  I am specifically looking at the factors of GDP, health expenditure, life expectancy, internet usage, unemployment rates and the increase in population. I am doing this as an exploratory analysis, but my main goal is to see how much money we will be spending on health and if it is going up or down, especially in relation to the other factors. I think that we can learn a lot about predicting the health expenditure and other variables I have chosen and can cover a wide range of clients, those being in the government, health field and even in communications.

**Data**

The data that I am looking at came from a data repository called Quandl. Quandl houses multiple time series data based off the major stock markets and how businesses are doing in a day to day fashion. They also deal with macro level time series like country economies, country health statistics and demographics of the different countries. The data I collected came from multiple data sets. The data sets in question came from a combination of the IMF Cross Country Macroeconomic Statistics and the World Bank World Development Indicators database.

Some of the main limitations of this data only shows the values of the variables, it does not go into a deep dive of why the numbers are what they do, as in no historical information just the pure numbers. So, the only thing I can predict with these is how high or low the values can go in the coming years.

I did quite a few alterations to my different data sets. The first thing I tried to do was import all the datasets I got into R and create a single table with all the variables but that turned out to not work as different variables had different lengths, so instead I decided to create six different data frames. Once I created the data frames I changed their value names to what those values were, as an example I changed the Value in GDP to USA\_GDP. Once I created the separate data frames I then reversed the order of the table so that the dates were chronologically accurate from earliest to present. Next I created an extended time series object in R by using the as.xts () function so I could then preform the proper analysis on the data, since I could not do the predictions with a data frame.

**Exploratory Analysis**

I explored the data with graphs of the time series to see how the values fluctuated all of them were in a positive direction and seemed to be steadily increasing. I did not notice any real dramatic shifts in values. I also tested the trend stationarity of my data with a kpss test. The test showed that there is evidence that it is not a trend stationarity with a p value less than 5%.

**DATA ANALYSIS**

The data analysis had to be changed for this project. I was thinking I was going to try out a regression model to predict the changes in either health expenditure or using health expenditure to see about change in life expectancy, but I decided with help from my mentor to change that idea and go with forecast predictions of my data. I like this approach better as it serves as more of an exploratory sense and, I can use it as a benchmark to see how well things are either improving or in some cases to see a potential decrease in values. I will be using ARIMA models as the base for my forecast predictions. Specifically, I will be using auto ARIMA as that way R can help me do the preprocessing for the ARIMA model better.