Capstone Milestone Report

Connor Vaughan

The goal of this Capstone project is to examine the relationship between unemployment and variables such as race, gender, and age. (Over time) The research questions this project intends to answer is how has the differences in the unemployment rate between demographics changed over time as well as where are they expected to go in the future? The results of this project will provide useful information to policymakers and employers. The unemployment rate is a key indicator used by policymakers to measure the strength of the economy and when unemployment falls below potential real GDP also falls. If policymakers can focus on specific groups of people who tend to be unemployed disproportionally, they can have a more significant impact on the economy when it enters a recession. Policymakers in the past have created some programs to support groups of people who tend to be unemployed at a higher rate than the general population, such as minorities, so it will be interesting to see how the relationship between race and unemployment has changed over time. Another insight this project may provide is to see if certain groups are affected by recessions (ex. 2007-2009) more than others. If so, policymakers can focus on these groups more to help get the economy out of a recession.

All of the data I will be using for this project comes from the Bureau of Labor Statistics. It is currently in the form of an excel file, so I will need to import all the data sets into R. Each demographic has its own data set so I will need to find a way to combine all the data sets into one data frame. The format is also kind of strange because each data set has information at the top that I don’t want in my final data frame so I will need to find a way to remove all this information. One limitation of this data is that the Asian demographic only has data back to 2003 and I don’t know how much this will affect my project. All the other demographics have data all the way back to 1973 with no missing values. One of the things this dataset will not be able to explain is why there is such a disparity between the unemployment rates of the different demographics. A lot of research has been done into this question and there hasn’t been a concise answer. This is all the cleaning and wrangling I will probably have to do because the data is already relatively well organized.

The approach I am taking will be to look at descriptive statistics to understand the historical trends and differences of the unemployment rate amongst the demographics. Using the skills I’ve learned with GGplot I will be able to plot the data to better visualize these trends. From there I need to create a forecasting model and I’ve only just begun learning how to forecast with R. This area will be my primary focus going forward. I understand how regressions are used for data analysis, but I don’t think they will be that useful for this project. The approach I’m using hasn’t changed much since I began this project, but I still need to better learn the methodology I plan on using.