**Summary of Analysis**

**Outcome of EDA**

Ultimately, the EDA was very successful in terms of supporting the development of the linear model. In the EDA we saw the distributions and scatter plots of the variables and their relationships to Life Expectancy and it illustrated quite clearly which variables would be of use and which variables probably wouldn’t.

**What do you feel was missed during the analysis?**

In terms of the what I was going for an my initial objectives I don’t feel like I missed out on much. I suppose one of the major areas that I didn’t look at was how the data behave based off of different sub-groups I.E. GDP values below a certain threshold, HDI values above a threshold, the predictor variable distributions based off of Status, etc. However, despite not exploring these types of areas I feel that I accomplished my goal succinctly within my analysis.

**Were there any variables you felt could have help in the analysis?**

Yes actually, one of the big ones that I really wanted was average education level (in years of schooling) in each of the countries for the data. I don’t know why but I wanted to observe how the average schooling would have impacted total life expectancy. I feel like that variable alone could have acted as a proxy for other variables such as education spending, infrastructure, etc.

**Were there any assumptions you made you felt were incorrect?**

I’m not sure if the feeling is that they were incorrect but rather if I had any idea what I was doing. The one that comes to mind is the transformation on the GDP variable that I did within the analysis. I feel like I missed something in terms of my full understanding of how transformations worked. It seemed to be acceptable and provide value but I’m not sure if there is more to doing a transformation like that or not. Also, my assumption in regards with the outliers was another rocky one, and that is one that I feel like I was incorrect on. It made sense at the time and it still does, but I don’t feel like it’s a problem that is just that simple, and I don’t feel like I handle the outliers properly in this regard.

**What challenges did you face, what did you not fully understand?**

Most of the challenges in the analysis were definitely in the code of python. Oddly enough one of the things that I struggled with was just the basic data management like renaming data, slicing data, and even removing NA values from datasets. I think a lot of this had to do with the fact that I started this project right after completing a 30+ hour project in R and I found it extremely difficult to adjust on the little things. In contrast to this all of the other more complex items like the CDF, PMF and Linear Model were much more straight forward and simple to do.