**Write up Project**Rami Aldoush

Data Science Project

My project is about World Happiness and I extracted my data from Kaggle.com.

So, let me introduce my project, my dataset is 12 Columns and 151 rows total and data is a comparison between 151 countries around the world. And the features of the dataset are consisting of a relationship between happiness score with other variables such as Health life expectancy, freedom, Economy GDP.

Stating questions is there is a correlation between happiness score and any other variables?

Is GDP affect happiness score?

After asking my question now it’s time explore the data I found some outlier 12 records in Government corruption but the rest of the data, not outliers and there were some null values in GDP which it didn’t make any sense so I dropped those null values as part of data cleaning.

After cleaning the data from any null and possible outliers is the time to run some statistical analysis for the features of choice so for this case calculating the mean and median of (Health& life Expectancy, Freedom, Happiness Score, Freedom).

Then plotting all variables to see any possible connection between variables, and after looking at the plots there was a strong relationship between all variables with the exception of Freedom variable. After progressing and exploring in data now it’s time to make some decisions about what formal modeling I need to choose and since its labeled data and most of it numerical, so supervised learning is the best model for this dataset and specifically is Regression analysis.

And values to be predicted Health life expectancy, GDP, Happiness Score.

After going through the process did, I rephrase the question and did I have different expectation and the answer was no my expectations were the same with one difference of discard some variables and keep others.