Title: Trends in Happiness Levels Globally

Name: Max Kortebein

The World Happiness Report is a survey given to citizens of different countries every year to gauge general happiness levels about aspects of their society on a per country basis. The first report was published in 2012 and there have been yearly reports since then. In this project I have the goal of analyzing the data and trying to find key factors that make people universally happy, along with finding trends in regions around the world.

The first dataset I used was from Kaggle at https://www.kaggle.com/unsdsn/world-happiness and gave information from the World Happiness Report from years 2015 to 2019. There were 5 csv files that I had to combine, and after combining I made an average column for happiness score, freedom score and trust in government. The second dataset I used at https://www.kaggle.com/fernandol/countries-of-the-world contained information on general facts about each country, such as population, GDP, etc.. After combining the merged happiness dataset with the facts dataset, I went from 156 countries to 141 countries because of excess missing data in some places.

I start by looking at different regions and figuring out the average happiness scores of countries in those regions. In **Figure 1** I group the countries by region and then make a bar plot with the x axis being the average happiness score and the y axis being the region. This graph was surprising, because in the raw dataset Scandinavian countries like Switzerland, Norway, and Denmark were always rated top five in happiness, but in this graph Western Europe is rated third globally. I also learned that the maximum happiness by region is about 7.3 in Australia and New Zealand, while the minimum is around 4.1 in Sub-Saharan Africa.

I used K-Means for **Figure 2**, and plotted government trust versus freedom because these would seem to go hand in hand. I used 10 clusters, 1 for each region and tested to see if each country (represented as a point) could be grouped into regions. The inertia with 10 clusters is about 2.95, going to 11 clusters brings this number to 3.3, and going down to 9 clusters makes it 2.6. Through testing multiple cluster amounts and referring to the raw data, I concluded that government trust versus freedom can not be defined on a per region basis.

In **Figure 3** I experimented with multiple variables corresponding to average happiness score, and then decided on plotting average happiness versus GDP per capita. I tried implementing a pipeline including Polynomial Features and Linear Regression but was getting a prediction score of 0.540 versus 0.546 using simple linear regression, so I ultimately decided to plot with just linear regression. One interesting observation was that the number of phones per 1000 people was a better indicator of happiness score (0.561) than GDP (0.546). Other notable factors I tested were literacy rates and infant mortality rates, with literacy rates having a prediction score of 0.287 and infant mortality rates having a score of 0.399 (which has a slope of -0.0254, one of the only negative slopes from my tests).

To conclude, my analysis shows that there are some aspects of a country that impact general happiness more than others. For example, GDP and phones owned per 1000 people are better indicators of happiness than factors like literacy and infant mortality rates. I also show that besides grouping directly by region, there is not a great way to tell regions apart with quantitative data. Hopefully this analysis gave insight into trends in happiness around the world, along with some of the reasons for them.

Figure 1: Happiness Score by World Region

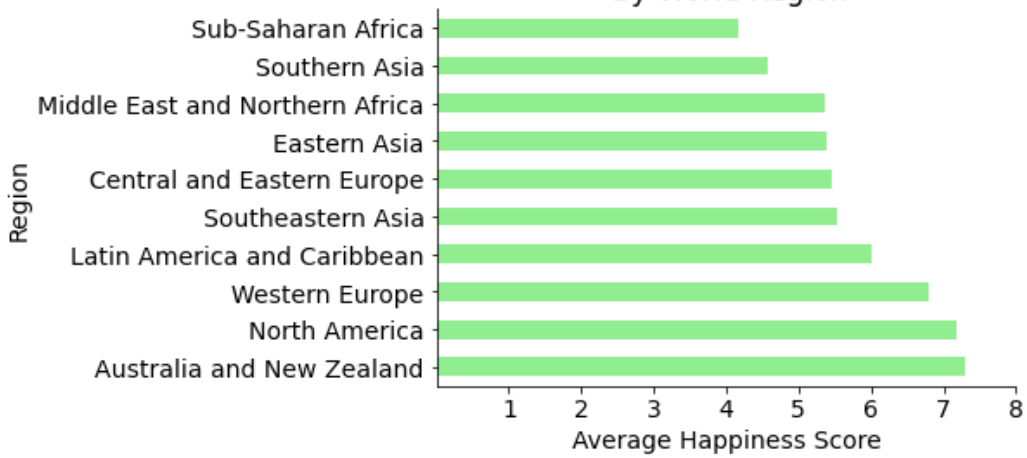

Figure 2: KMeans Freedom vs. Trust in Government

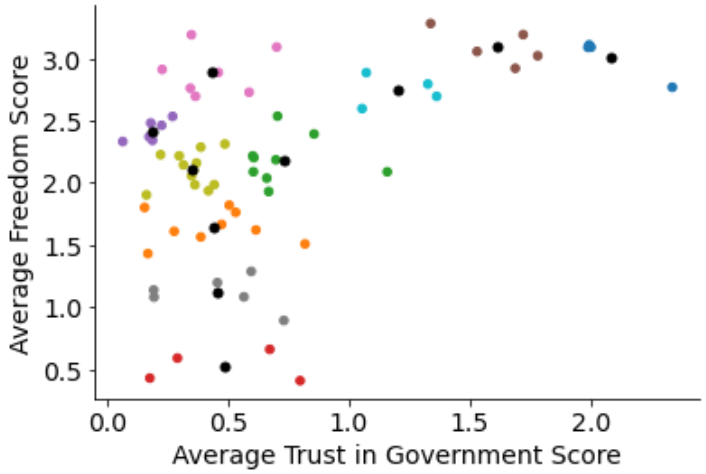

Figure 3: Linear Regression Happiness vs. GDP