**The Pursuit of Happiness**

**DSC520-Final Project**

**Timothy Robbins**

**11/11/2019**

**Summarize the problem statement you addressed.**

The problem statement was to determine which factors contribute the most to happiness around the world and then use those variables to create a model for predicting happiness.

**Summarize how you addressed this problem statement (the data used and the methodology employed).**

I used the 2018 World Happiness Report which I found at Kaggle.com (<https://www.kaggle.com/njlow1202/world-happiness-report-data-2018>). The report is a landmark survey of the state of global happiness.

I started out by printing histograms, the structure, and summary statistics of the main variables (Happiness, Economy, Family, Freedom, Generosity, and Trust) related to happiness to gain a better understanding of the data and to help identify any potential outliers. I also used the boxplot$out function to identify outliers within plus/minus 1.5 times IQ. I reviewed the outliers for errors and reasonableness and marked the erroneous one’s as NA. I then created a clean dataset with more useful column headings that excluded the erroneous data. I also created a variable for the continent of each country in the dataset, so I could break down and summarize the data by continent.

With a clean dataset, I ran a correlation matrix to identify the relationships between variables. Economy (GDP) had the highest correlation (.81) with Happiness, followed by Family (.78), Health (.77), Freedom (.57), Generosity (.23), and Trust (.41). I then created multiple linear regression models to find the combination of variables that would contribute most to the model. I found that the strongest model consisted of Happiness as the outcome variable, with Economy, Family, and Health, as predictor variables.

I then created scatter plots, with linear regression lines fitted, comparing the Happiness scores based on Economy, Family, and Health across continents.

**Summarize the interesting insights that your analysis provided.**

The scatter plots showed that there was a positive linear relationship between Happiness and Economy, Happiness and Family, and Happiness and Health, for all continents (except maybe for the continent of Australia which only included data for two countries: Australia and New Zealand).

Based on Economy, the highest happiness scores were in Europe, followed closely by North America and Australia, with Asia, South America, and Africa finishing on the low end. The same trend continued with Family, with the continents being ranked in the same order as with Economy. However, the slope of the line for South America was nearly flat, indicating that Family on that continent didn’t really affect the happiness score. Finally, based on Health, the results were about the same, which also included a nearly flat line for South America indicating that Health had little effect on the happiness score.

**Summarize the implications to the consumer (target audience) of your analysis.**

The results found that the happiest people were in Europe, especially in Scandinavia. Since happiness is such an important aspect of our lives, I believe it would be worthwhile to conduct additional research on the countries in Scandinavia to get a better understanding of why the people there are happier than in other places. Then we might consider applying some of those same practices here in the United States, or in other countries.

**Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.**

The data is limited, in that the categories used in this analysis are fairly broad, so it is more difficult to understand the specific reasons people may be happier. I believe someone else or me could improve the analysis by looking at the other end of the spectrum, as well. For example, what are the least happy countries and what are the factors contributing to their unhappiness. Then, if possible, changes could be made in those countries to reduce unhappiness and thus improve happiness. Also, we could use the regression model to try to predict how certain changes in those countries might improve their happiness score.