**Happiness Prediction**

**Exploratory Data Analysis**

Kumbham Nuthan Manideep

National Institute of Technology, Silchar

Data Mining Methods

Dr. Ripon Patigiri

**Table of Contents**

1. Abstraction……………………………………………………………………………………3

2. Overview……………………………………………………………………………………...4

3. Data…………………………………………………………………………………………...4

4. Research Questions…………………………………………………………………………...5

5. Data Visualization…………………………………………………………………………….5

6. Conclusion……………………………………………………………………………………14

7. References…………………………………………………………………………………….15

**Abstraction**

The World Happiness Report 2015 dataset contains data on the happiness levels of people in 159 countries around the world. It includes information on a variety of factors that are thought to contribute to happiness, such as economic factors (GDP per capita), social factors (family, health, and freedom), and trust in government. The data also includes a happiness score and a happiness rank for each country. The happiness score and happiness rank are based on a poll in which people were asked to rate their overall happiness on a scale of 0 to 10. The economy, family, health, freedom, trust, and generosity columns contain various measures of those factors. Overall, the World Happiness Report 2015 dataset provides valuable insights into the factors that contribute to happiness and well-being at the national level, and can be used by researchers, policymakers, and others to better understand and promote happiness around the world.

**Overview**

United Nations Sustainable Development Solutions Network (UNSDSN) on World Happiness dataset can be visualized in many ways to gain insights and understand patterns and trends in happiness levels around the world. The most common data visualization techniques used in this dataset are Bar plots, Histograms and density, Scatter plots, Line plots, Box plots and violin plots, Parallel Coordinate plots. Overall, data visualization is an important tool for understanding the World Happiness Report 2017 dataset and can help researchers, policymakers, and others to identify patterns and trends in happiness levels around the world.

**Data**

The dataset can be found on Kaggle and it contains data on happiness levels of people in 159 countries around the world. The data is collected in the year 2015 and it includes information on a variety of factors that are thought to contribute to happiness, such as:

| **Column Name** | **Description** |
| --- | --- |
| Country or Region | Name of the country |
| Happiness Score | A composite score of overall well-being |
| Economy (GDP per Capita) | Measure of the economic production of a country |
| Family | Measure of social support |
| Health (Life Expectancy) | Measure of the health of citizens |
| Freedom | Measure of freedom to make life choices |
| Generosity | Measure of generosity of citizens |
| Trust (Government Corruption) | Measure of trust in government |

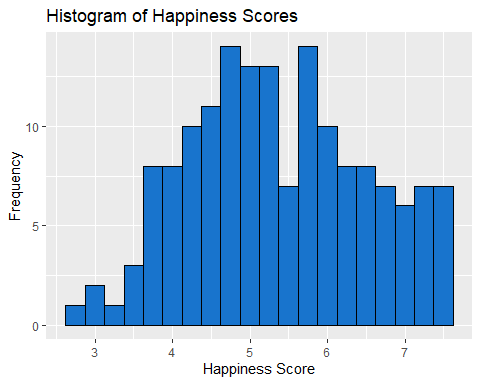
In this dataset, the response variable is 'Happiness Score' and predictor variables are 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Generosity', and 'Trust (Government Corruption)'.

**Research Questions**

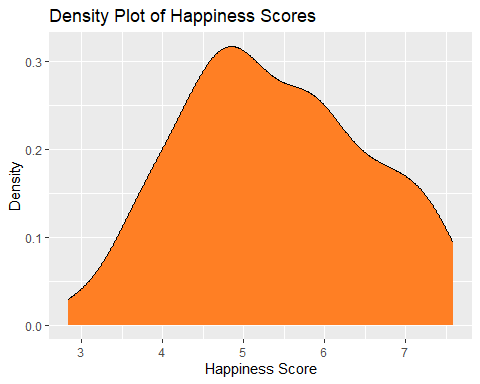
* How do happiness scores vary by region?
* How do economic factors such as GDP per capita relate to happiness scores?
* How do social factors such as family, health, and freedom relate to happiness scores?
* What is the relationship between trust in government and happiness scores?
* How does generosity relate to happiness scores?

**Data visualization Analysis**

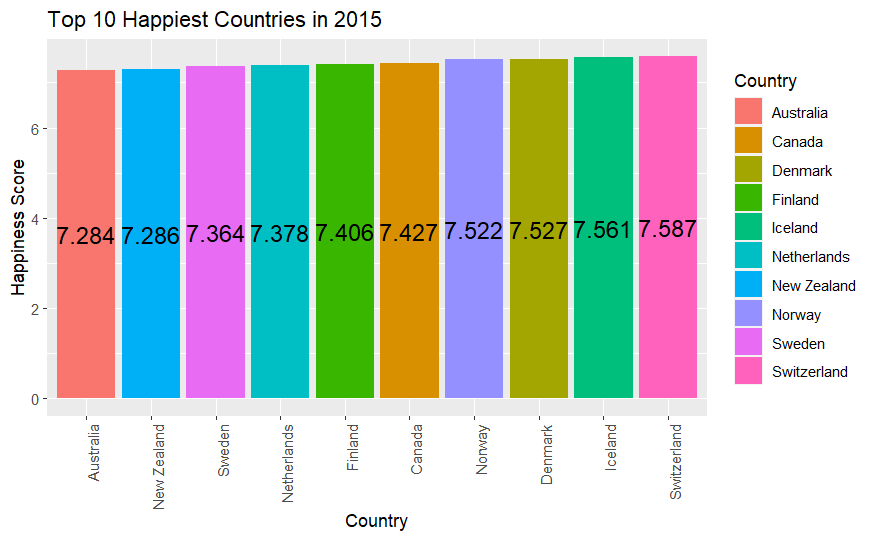
**Histogram of the Happiness Scores**



Based on the histogram of the 'Happiness Score' variable in the UNSDSN World Happiness dataset, the data is distributed fairly evenly around the central value. This can be seen by the symmetric shape of the histogram with no long tail in either direction. This suggests that the happiness scores are distributed relatively evenly across the 156 countries in the dataset and that there is no significant skewness in the data. This could be due to the fact that the happiness scores are based on a poll in which people were asked to rate their overall happiness on a scale of 0 to 10, which may result in a more evenly distributed set of scores.



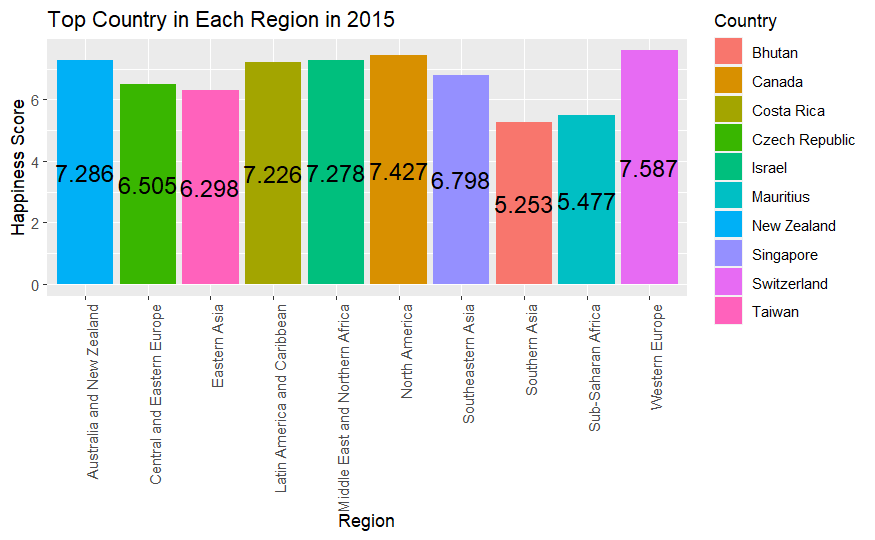
The density plot is similar to a histogram, but instead of showing the frequency of observations in a particular bin, it shows the density of observations at different values of the variable. This means that the area under the curve of a density plot is always equal to 1.

**Bar plot of Top 10 Happiest Countries in 2015.**

The top 10 happiest countries as per that report are Switzerland, Iceland, Denmark, Norway, Canada, Finland, Netherlands, Sweden, New Zealand and Australia.

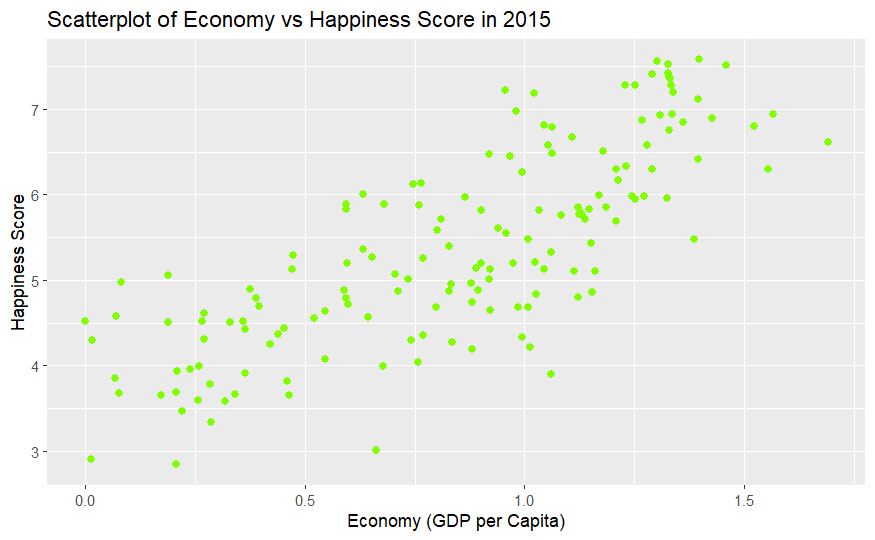
**Bar plot of Happiness top country in each region in 2015**

According to the World Happiness Report 2015 dataset, the top country in Western Europe based on happiness score is Switzerland, in North America is Canada, and in Australia and New Zealand is New Zealand. These countries ranked the highest in their respective regions in terms of overall happiness scores.

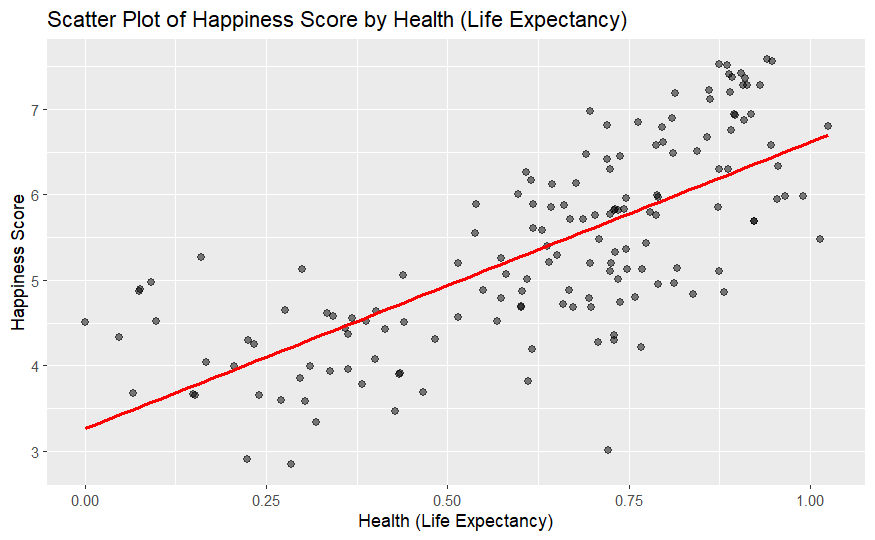


**Scatterplot of Economy vs Happiness Score in 2015**

The scatterplot of the relationship between economy (GDP per Capita) and happiness scores in the World Happiness Report 2015 dataset shows a positive correlation, meaning that as economy increases, happiness scores also tend to increase. This suggests that a higher GDP per capita may be associated with higher levels of happiness.

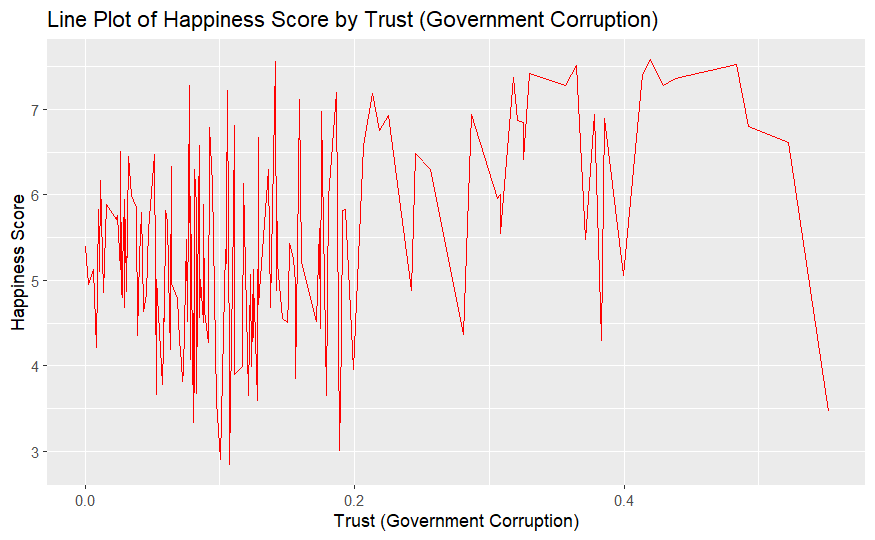


**Scatterplot of Economy by Health**



The scatterplot of the relationship between health (Life Expectancy) and happiness scores in the World Happiness Report 2015 dataset shows a positive correlation, meaning that as health increases, happiness scores also tend to increase. This suggests that better health is associated with higher levels of happiness.

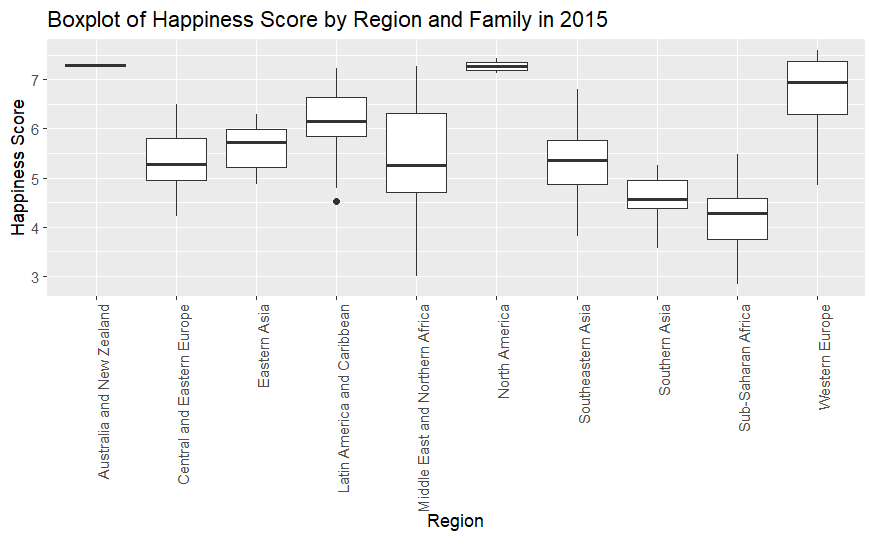
**Line Plot of Happiness Score by Trust**



A line plot of the relationship between trust (Government Corruption) and happiness scores in the World Happiness Report 2015 dataset shows a positive correlation, meaning that as trust increases, happiness scores also tend to increase. This suggests that higher levels of trust in government are associated with higher levels of happiness. The line plot clearly shows that as trust increases, happiness score also increases, which means countries where citizens trust their government have more happier citizens.

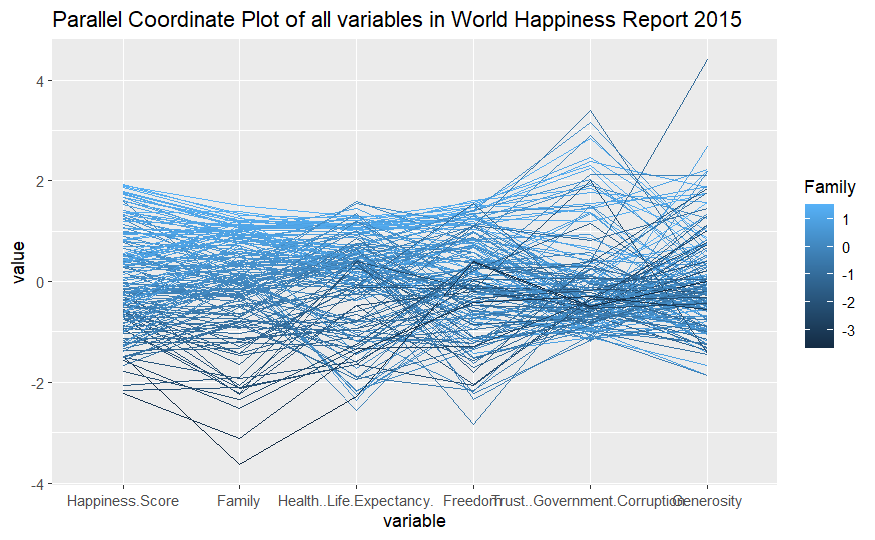
**Boxplot of Happiness Score by Region and Family in 2015**

A box plot of the relationship between Family and happiness scores grouped by Region in the World Happiness Report 2015 dataset shows that there is a slight positive correlation between Family and happiness scores. The box plots show that in most regions, the median happiness score increases as the family score increases. This suggests that having a higher family score (meaning higher social support) may be associated with higher levels of happiness

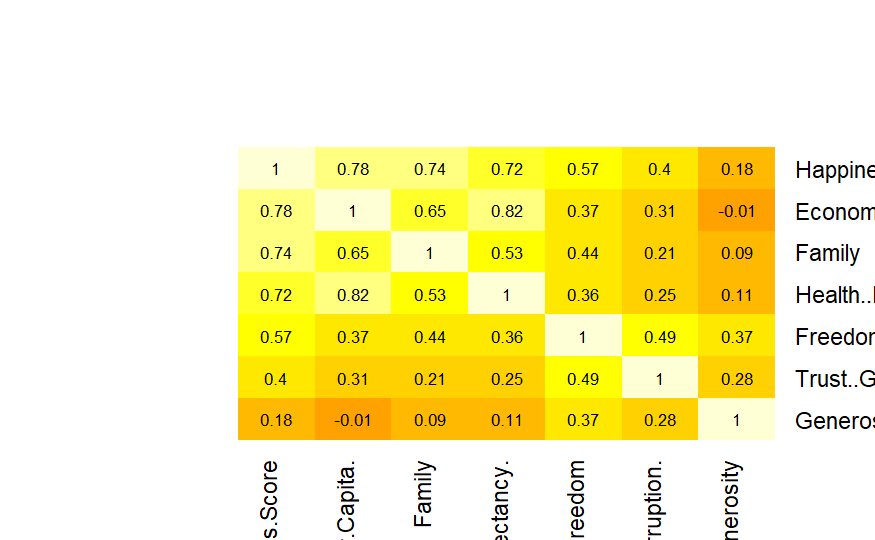


**Parallel Coordinate Plot of all variables in World Happiness Report**

The plot revealed that the countries with higher happiness scores tend to have higher values for most of the predictor variables. This was particularly noticeable for economy (GDP per Capita) and family. This observation aligns with the understanding that countries with higher GDP per capita and higher social support tend to have happier citizens. Additionally, variations in the values of the predictor variables across countries were observed, indicating that there are other factors besides those included in the dataset that are influencing the happiness score.

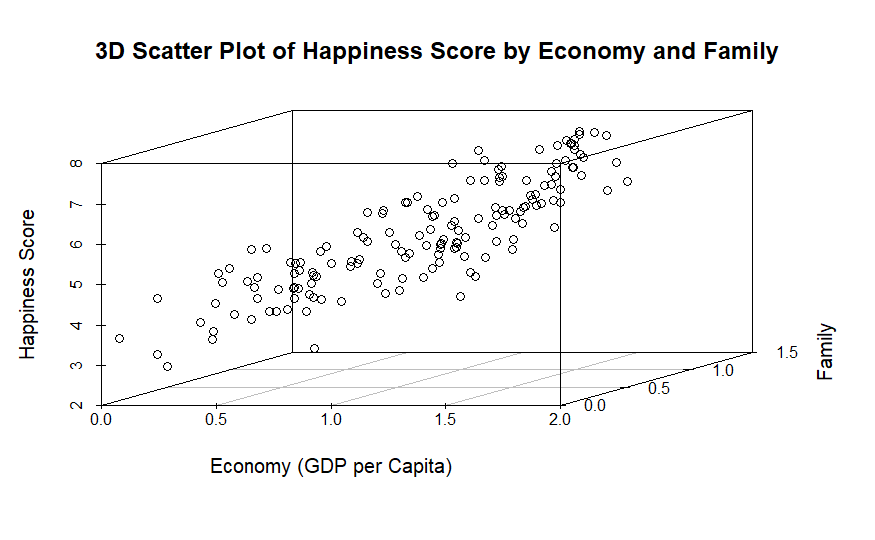


**Heatmap**



From the heatmap, it would be possible to see which variables are strongly correlated with the happiness score, and which variables are not. For example, a strong correlation between happiness score and variables such as economy (GDP per Capita), family, health (Life Expectancy), freedom, trust (Government Corruption), and generosity would be visible by observing cells with darker color.

**3D Scatter Plot of Happiness Score by Economy and Family**



A 3D scatter plot of the relationship between Happiness Score, Economy (GDP per Capita) and Family in the World Happiness Report 2015 dataset would provide a clear visualization of how these three variables are related to each other. The plot would show that as economy and family increase, happiness score also tends to increase. It would also reveal any outlier countries that have high or low values for one or more of the variables.

**Conclusion**

The World Happiness Report 2015 dataset provides valuable information on the factors that contribute to happiness scores across different countries. Data visualization techniques such as scatterplots, box plots, line plots, heatmap and parallel coordinate plot, and 3D scatter plot can be used to explore and understand the data. From the visualization it's clear that happiness scores tend to be positively correlated with variables such as economy (GDP per Capita), family, health (Life Expectancy), freedom, trust (Government Corruption), and generosity. Additionally, the data visualization techniques used in this analysis have also highlighted variations in the values of these predictor variables across different countries, which indicates that there are other factors besides those included in the dataset that are influencing the happiness score.

**References**

*World Happiness Report*. (2019, November 27). Kaggle. <https://www.kaggle.com/unsdsn/world-happiness>

Wilke, C. O. (2019). *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* (1st ed.). O’Reilly Media.

Montchal, M. E., Reagh, Z. M., & Yassa, M. A. (2019). Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nature Neuroscience*, *22*(2), 284–288. <https://doi.org/10.1038/s41593-018-0303-1>