World Happiness Report EDA

Jennifer Ruiz

DSC 530

Background

Background Information

- The World Happiness Report is a landmark measure of happiness around the globe.

- The survey is administered annually to 158 countries.

- This analysis focuses on the 2015 World Happiness Report

- A link to the full dataset can be found here:

    - https://www.kaggle.com/unsdsn/world-happiness#2015.csv

Research Questions

Are there any trends in Happiness rank based on Region?

What influence do Freedom and Trust have upon the Happiness Rank?

What impact does Health have on Happiness rank?

How does GDP impact the Happiness Score?

Which variables have the most significant impact upon the Happiness Score?
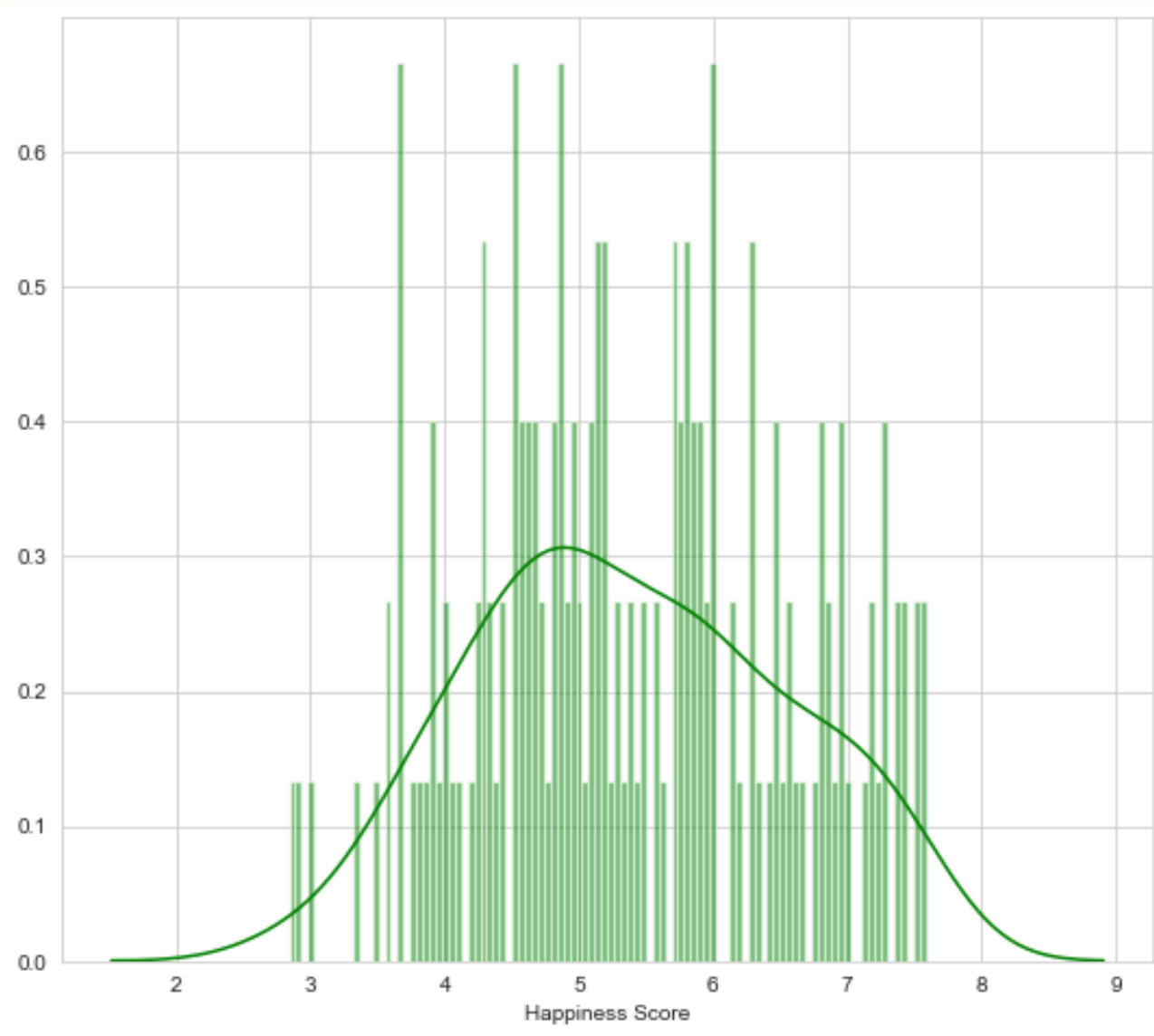
The Data

———

Variables of Interest

- Happiness Score: ranking of Happiness on scale of 1-10

- Happiness Rank: Final Rank of Each Country

- Country: Name of Each Participating Country

- Region: Geographical Region

- Economy (GDP): measures the economy of the country

- Health: Measure of Life Expectancy for each country

- Freedom: Level of Liberty for each country
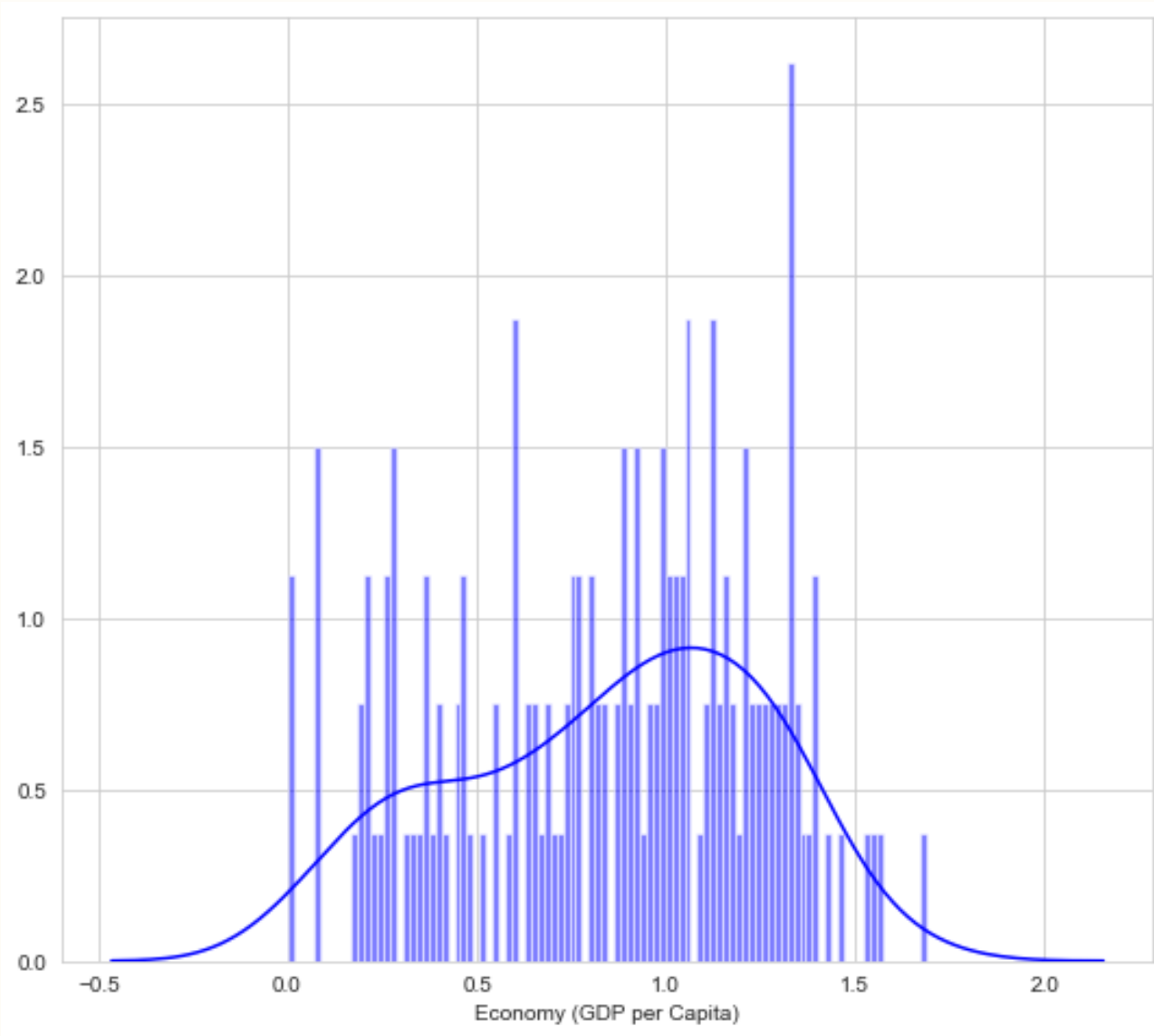
- Trust: measures government corruption

Outliers

Each of the five main variables chosen contain outliers. After reviewing the data and associated histograms, the decision was to keep outliers within the dataset as they are key to a full analysis.
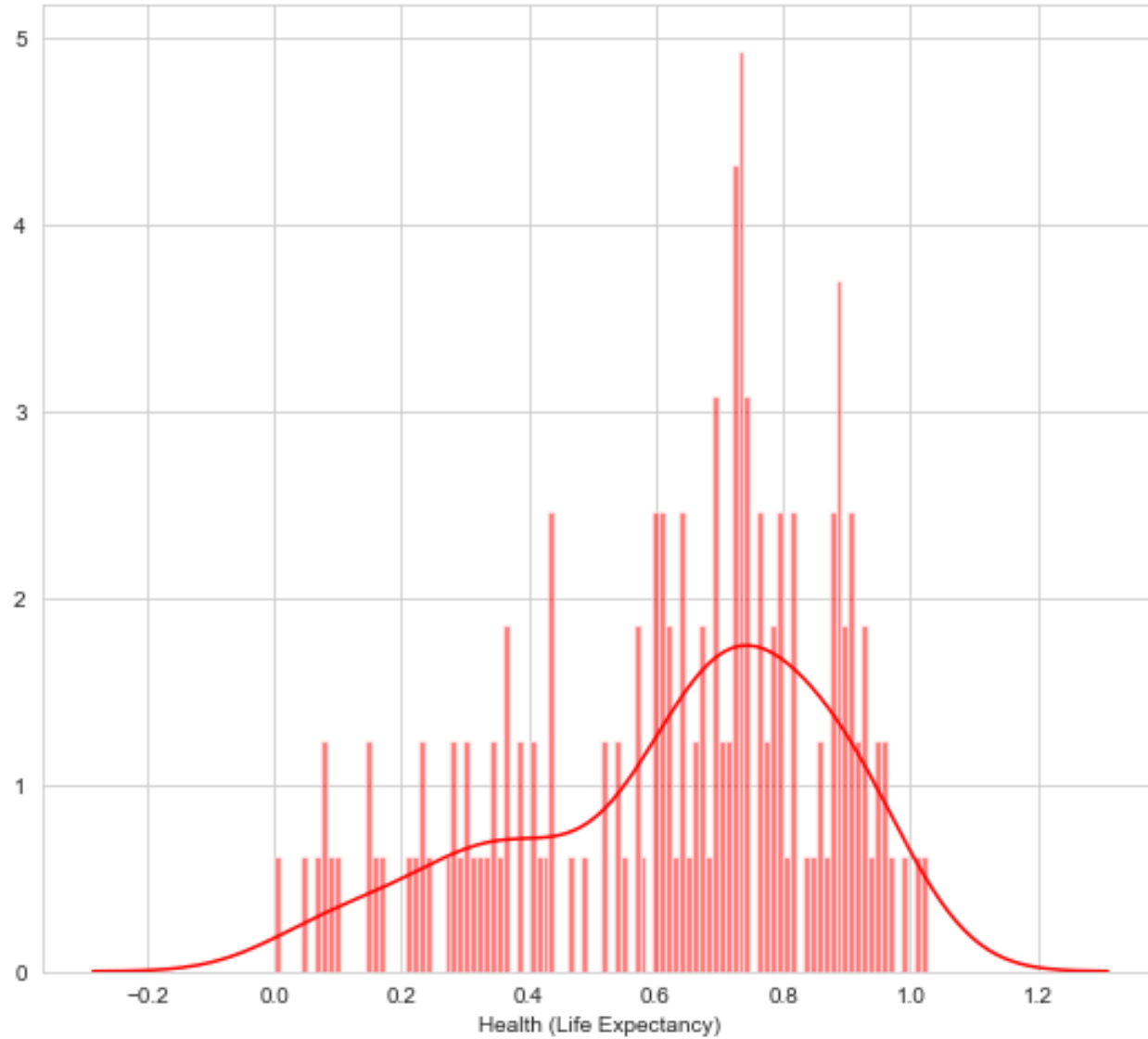
Happiness Score

– Mean: 5.38

– Mode: 5.19

– Std: 1.15

– Min value: 2.84

– Max value: 7.58

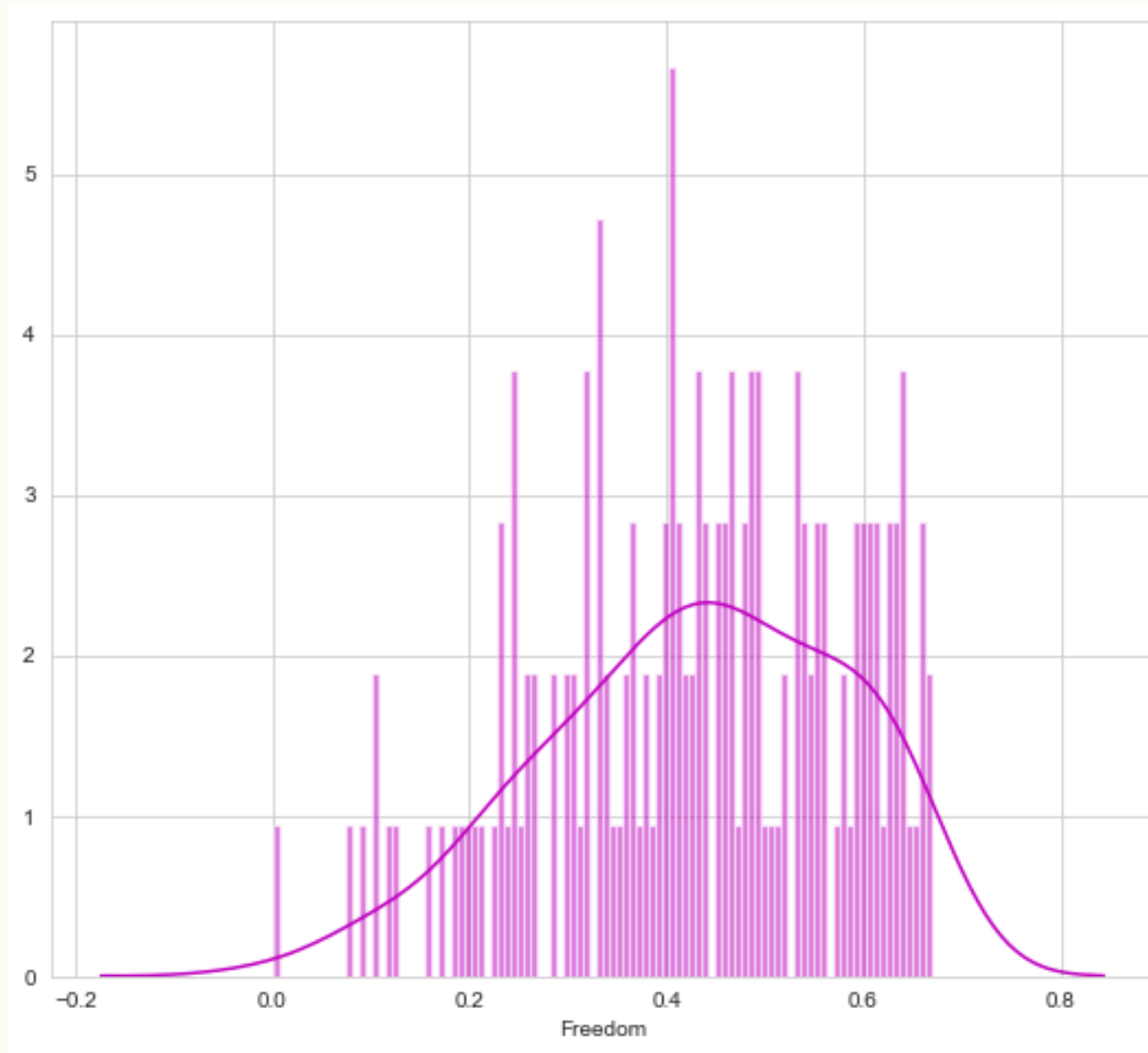– Distribution: the distribution is slightly positively skewed

Economy (GDP)

– Mean: 0.85

– Mode: multiple modes

– Std: 0.40

– Min Value: 0.00

– Max Value: 1.69

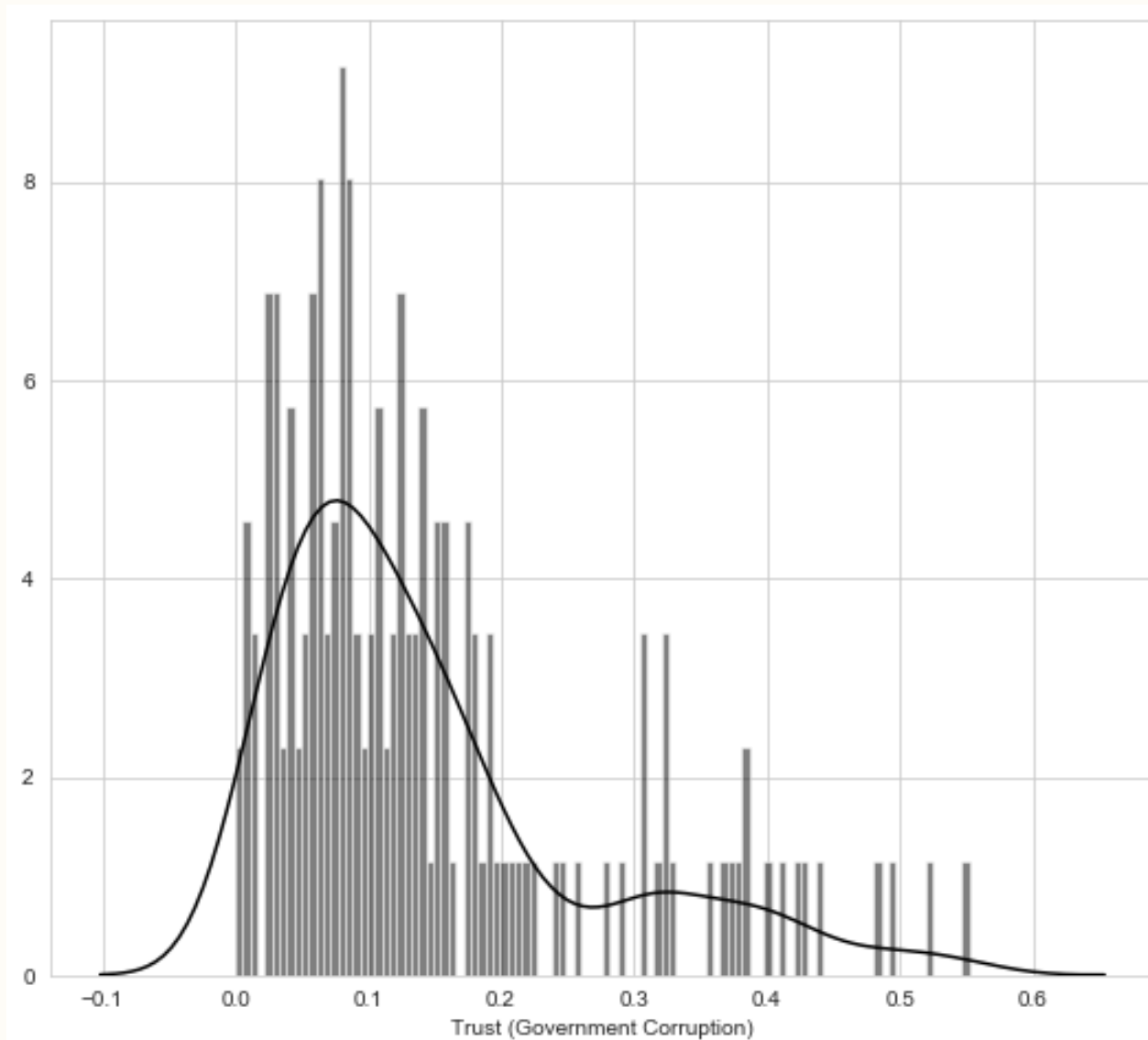– Distribution: the distribution has a slight negative skew

Health (Life Expectancy)

- Mean: 0.63
- Mode: 0.92
- Std: 0.25
- Min Value: 0.00
- Max Value: 1.02
- Distribution: the distribution is negatively skewed

Freedom

– Mean: 0.43

– Mode: Multiple Modes

– Std: 0.15

– Min Value: 0.00

– Max Value: 0.67

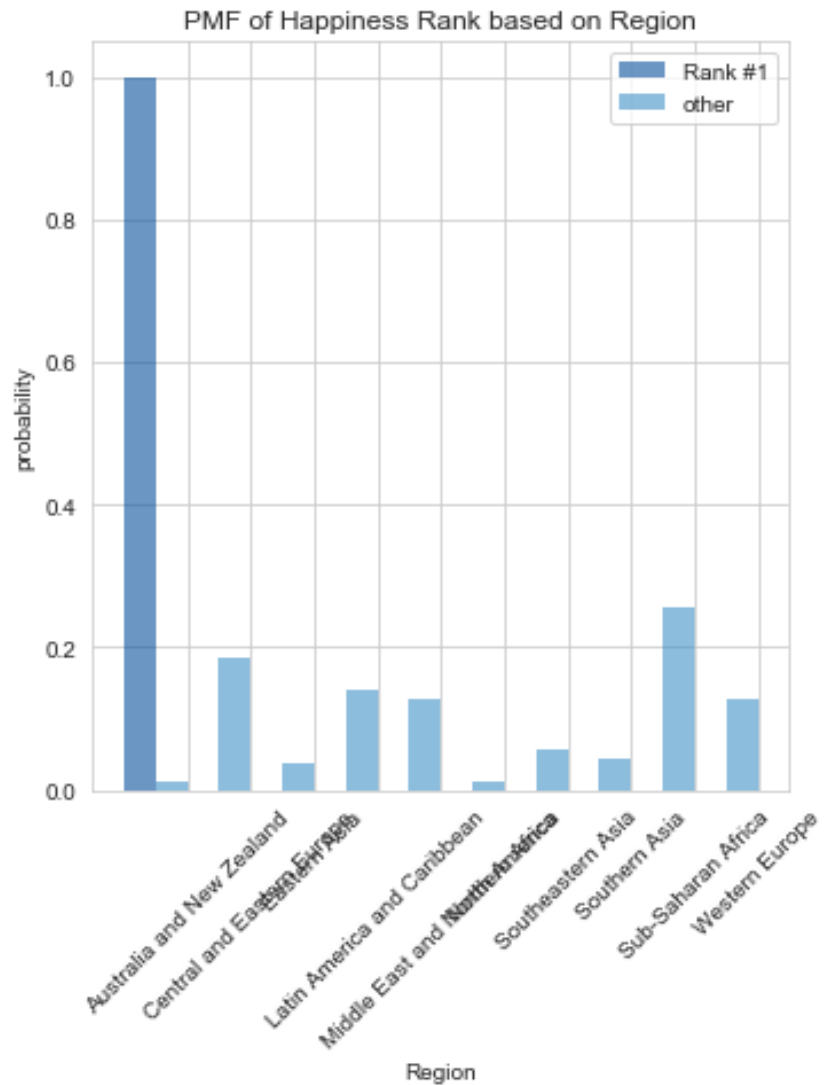– Distribution: the distribution is has a slight negative skew
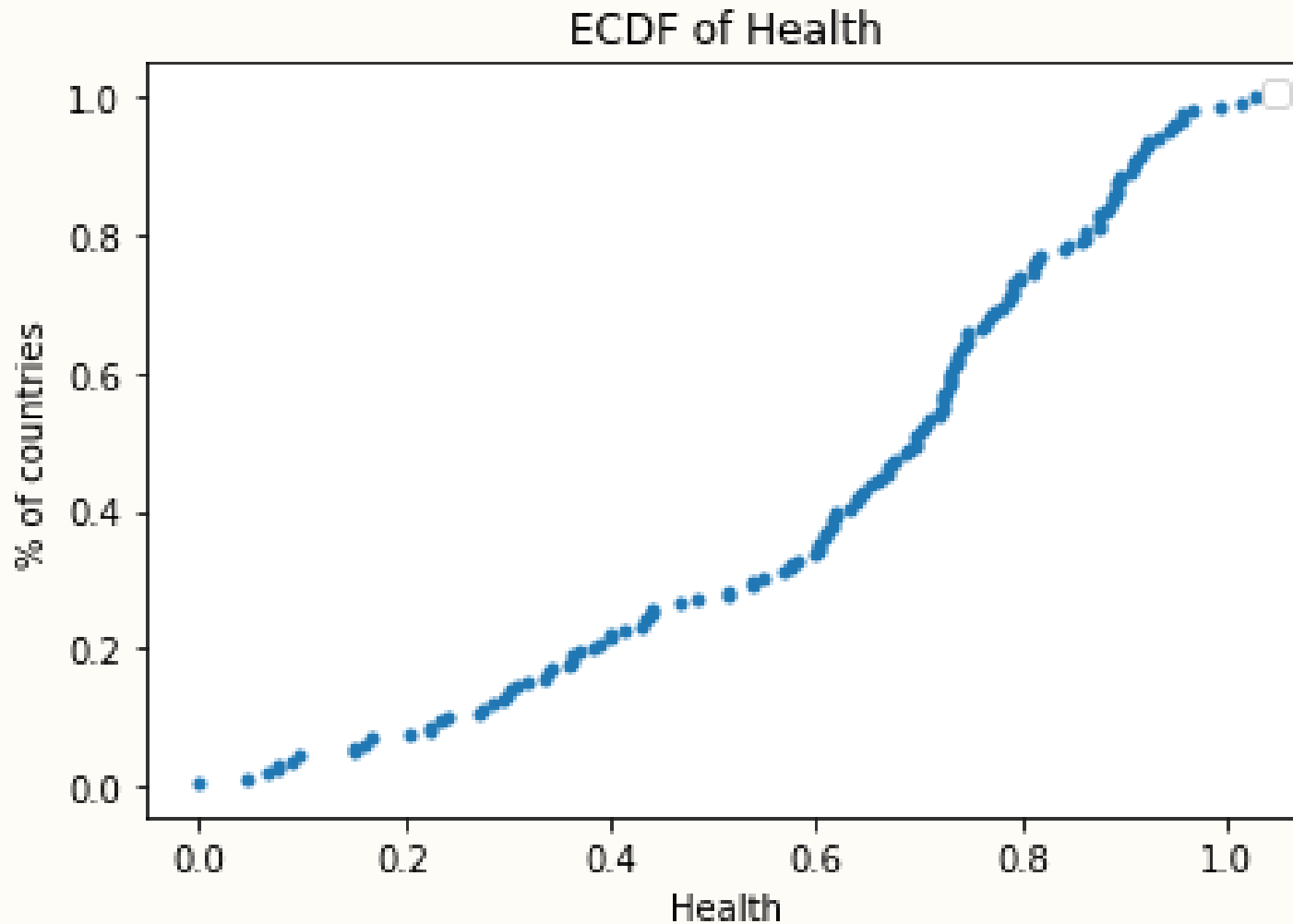
Trust (Government Corruption)

- Mean: 0.14
- Mode: 0.33
- Std: 0.12
- Min Value: 0.00
- Max Value: 0.55
- Tails: Distribution is positively skewed

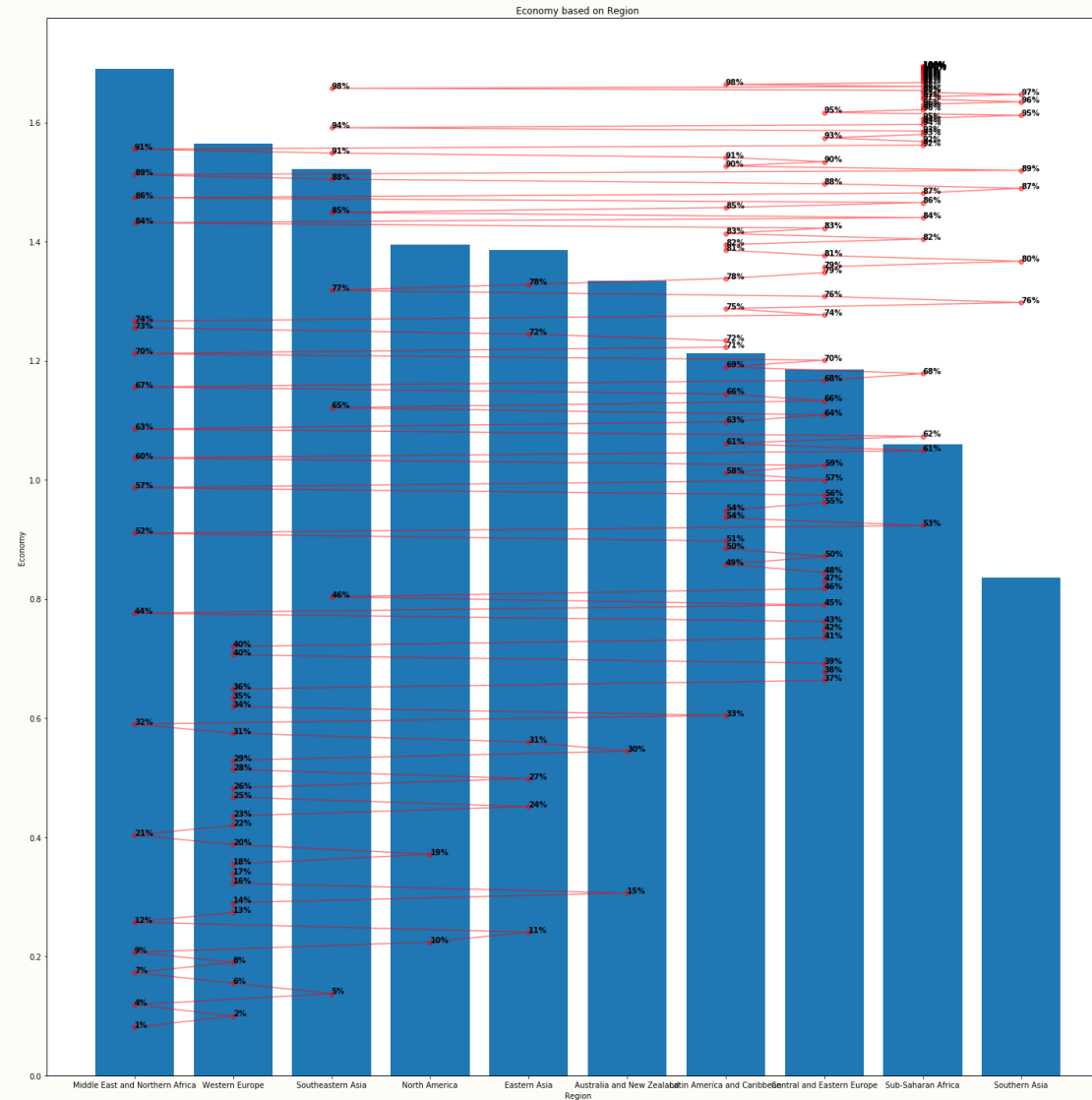# Analysis

PMF of Happiness Rank based on Region

Based on the PMF, Western Europe tends to have the highest happiness rankings, while the Middle East/ North Africa and Australia/New Zealand tend to have the lowest happiness rankings.

CDF of Health

- The CDF indicates that 40% of the countries surveyed have a Life Expectancy less than 60 years (0.6).

- 80% of countries have a life expectancy less than 80 years.

- 20% of countries have a life expectancy greater than 80 years.

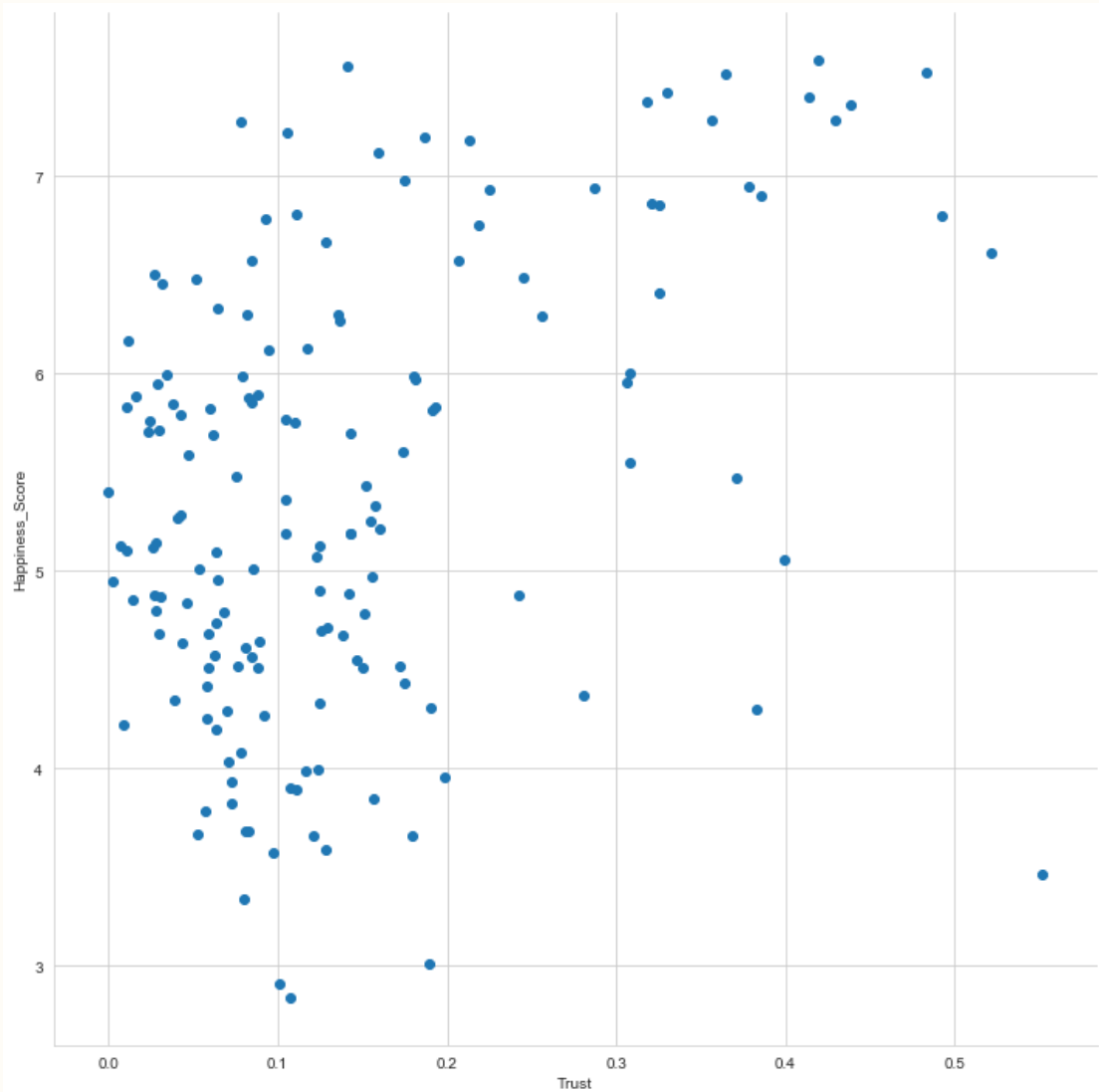- Less than 1% of countries have a life expectancy of more than 100 years.

Pareto Distribution of Economy (GDP)

— The pareto plot reveals that the Middle East and North Africa have the highest GDP of the countries surveyed.

— Southeast Asia has the lowest GDP of the countries surveyed.

— North America and Eastern Europe have similar GDP values.

Trust vs. Happiness Score

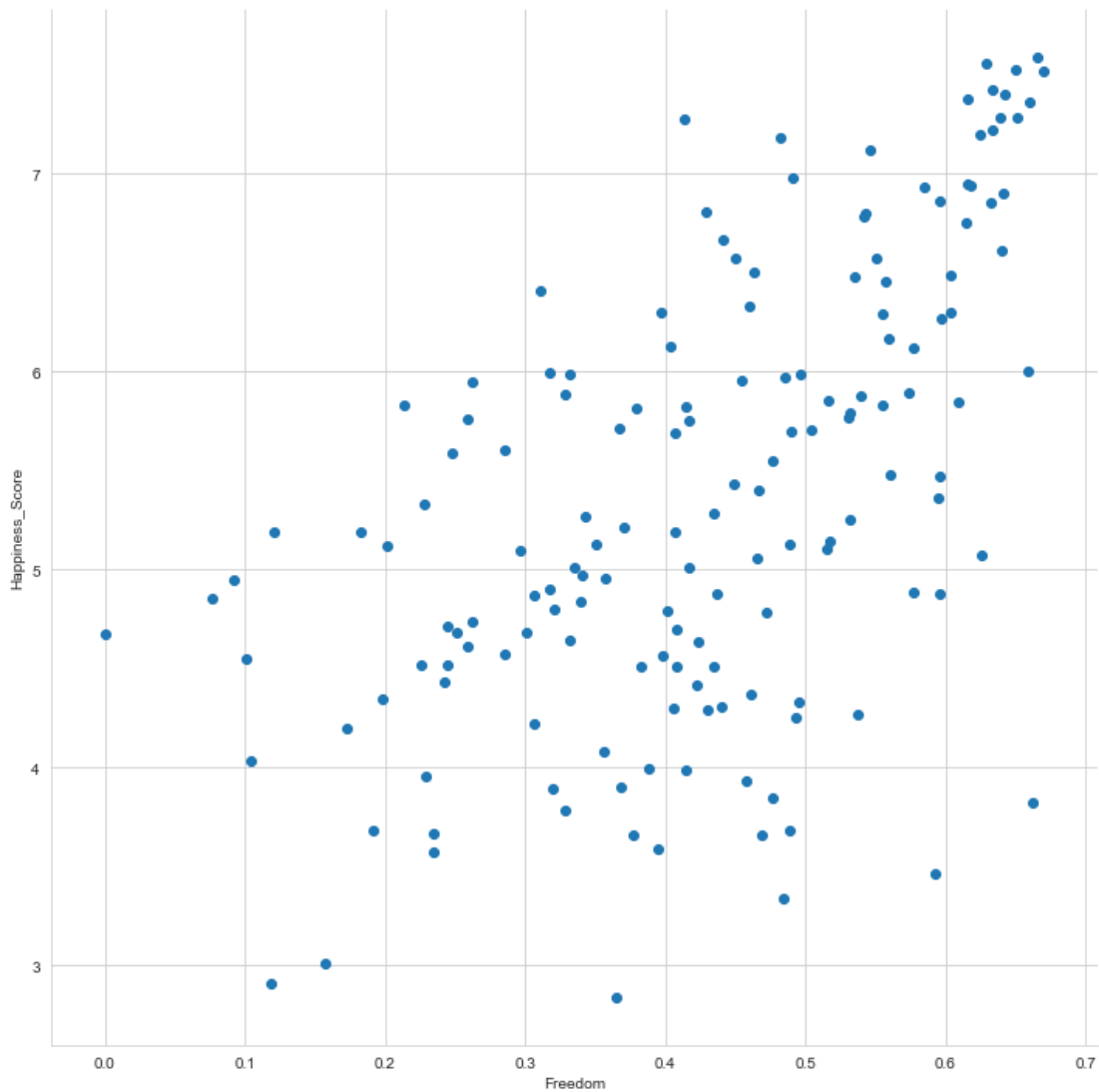It seems that Trust (the level of government Corruption) has a mild, positive effect on Happiness Score.

– Correlation: 0.40

– Covariance: 0.05

Freedom vs. Happiness

Freedom has a moderate positive correlation with the Happiness Score of countries surveyed.

– Correlation: 0.58

– Covariance: 0.10
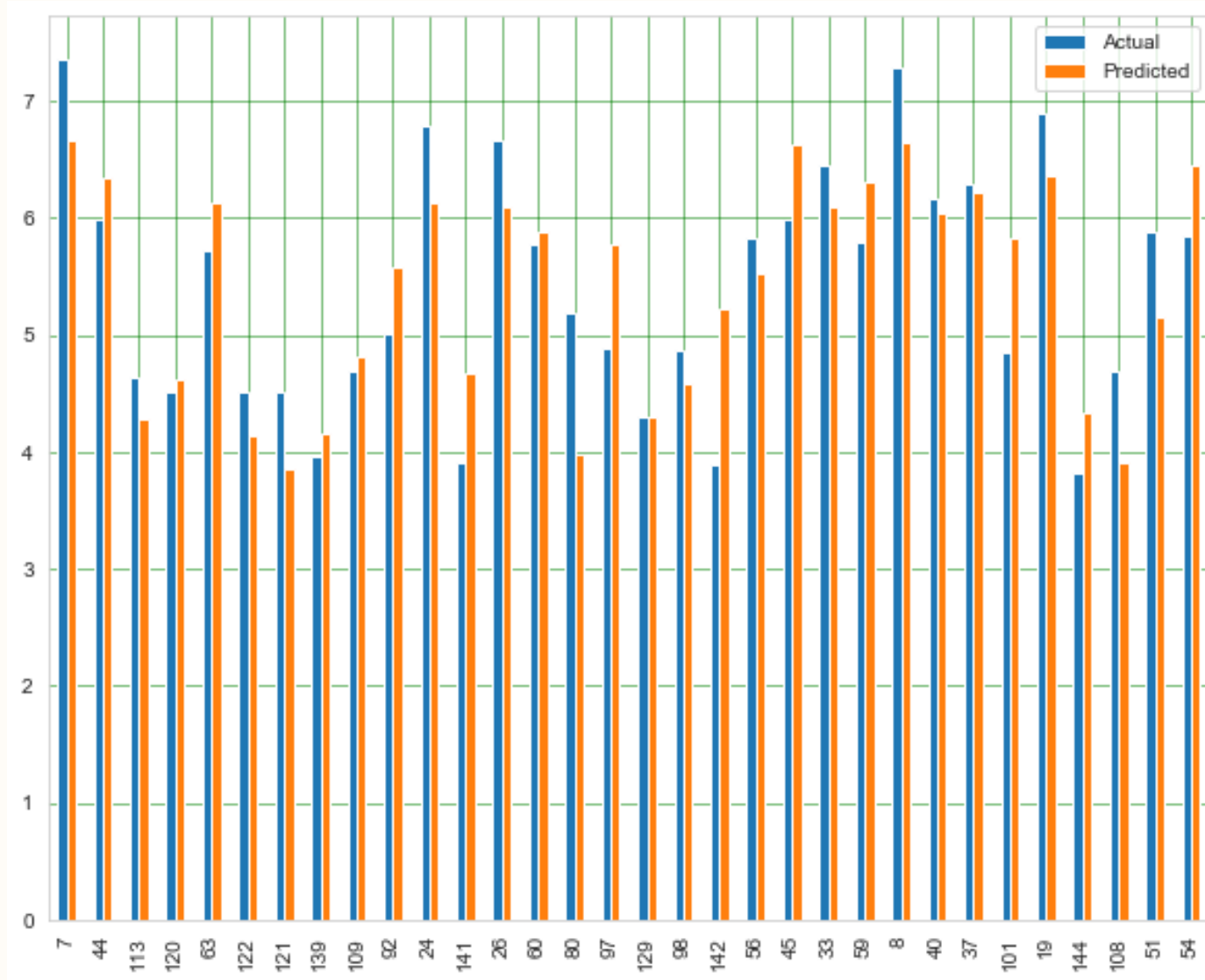
Chi Squared Test for Country and Region

A chi-squared test was conducted to identify any association between Country and Region.

- **Hypothesis:** Country and Region are associated with one another.

- **Result:** There is no association between country and region.

- Significance Level: 0.05

- df: 1413

- p-value: 1.0

- Critical value: 1501.56

Multiple Regression

– A multiple regression model was created to investigate the degree to which Economy, Family, and Health predict Happiness Score.

– R^2: 0.72

– The high R-squared value and low std error for each factor indicate the model is a good fit for the data.

Conclusion

Final Remarks

- Through this analysis, several factors seem to have a strong relationship with the Happiness Score. The top three were: Economy, Family, and Health.

- This analysis produced compelling insights and warrants future investigation with more in-depth analysis.