Exploring global happiness scores and its drivers

Image source: https://globalwellnessinstitute.org/

For Data Science Course: By Irina Nizhnik

About the Data Set

Target variable (Y)

- Ladder score:
  - Stands for overall level of happiness in each country
  - Continuous variable from 1-10
  - Mean: 5.47
  - Median: 5.15
  - Standard Dev: 1.11
  - Max: 7:80
  - Min: 2.56

Independent variables (Xs)

- Includes features that could impact Y, for instance
  - GDP/per capita
  - Social support
  - Healthy life expectancy
  - Freedom score
  - Corruption score
  - Generosity score
  - Dystopia score

The dataset contains 153 rows – one for each country in the dataset

Data Preprocessing

- Checks conducted
  - Missing value
  - Duplicate value
  - Inconsistency check
  - Irrelevant features
  - Outliers in Y and Xs
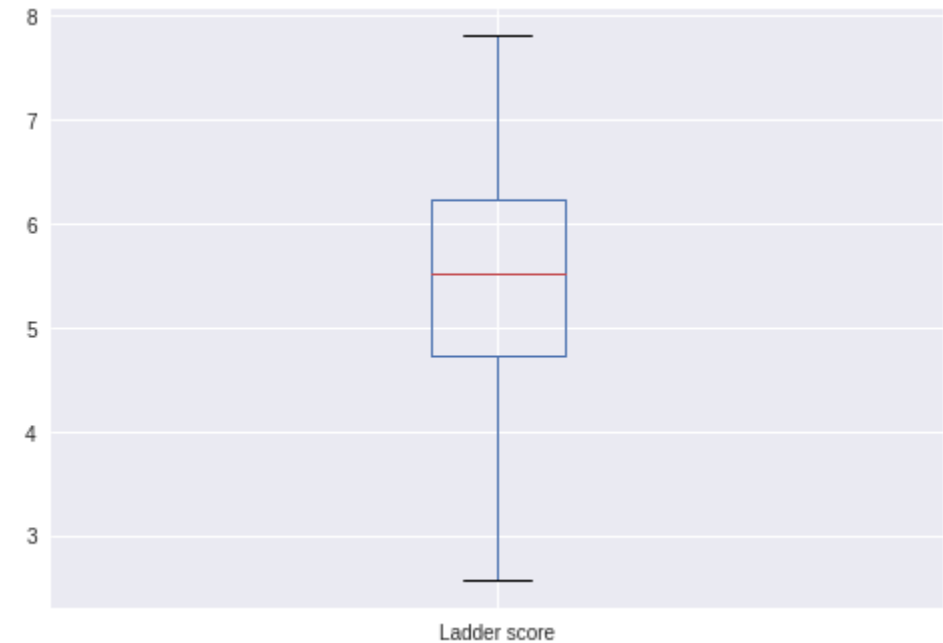 - Train/test dataset
  - Standard/scale dataset

- Results
  - No missing value
  - No duplicate value
  - Minor inconsistency
  - None
  - Nothing significant
  - Done
  - To be done

Exploring the target – Ladder Score (Happiness score)

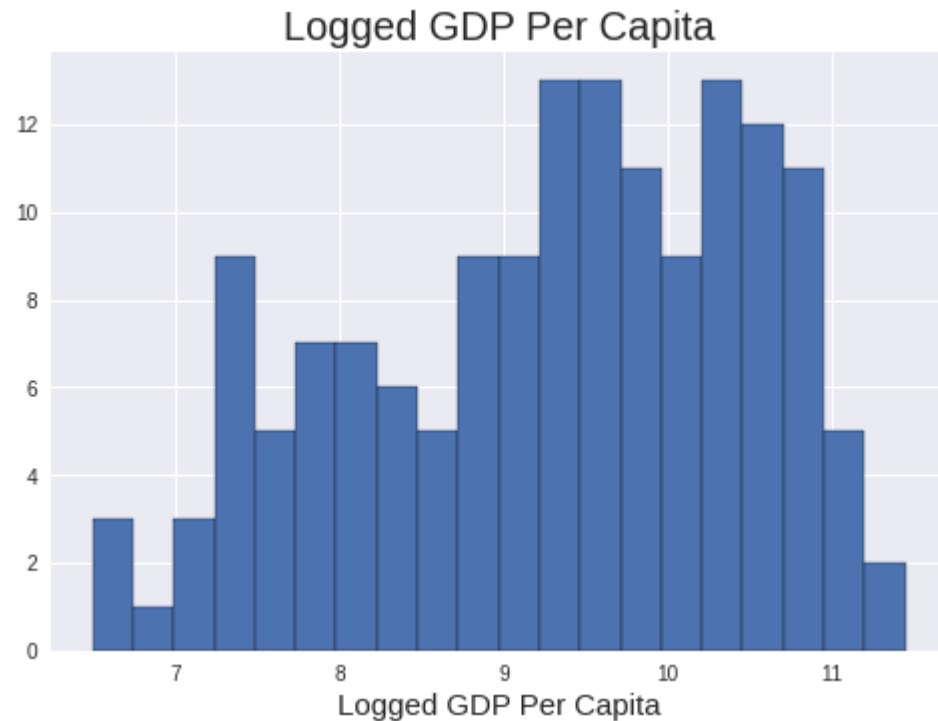Histogram of happiness levels - on scale of 1-10

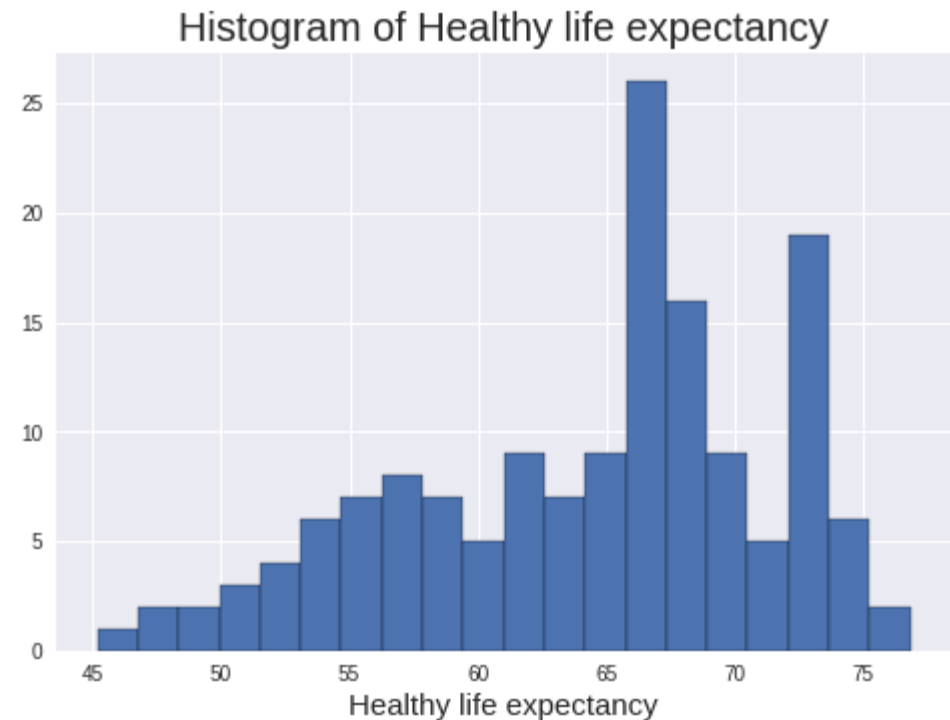Most countries fall within the 25-50% percentile of 4.72 to 6.22 happiness score

Mean happiness score of 5.47, with a range of 2.56-7.8 (Denmark) and standard deviation of 1.11

Exploring some of the Xs (1/2) – Logged GDP/Capita and Healthy Life Expectancy
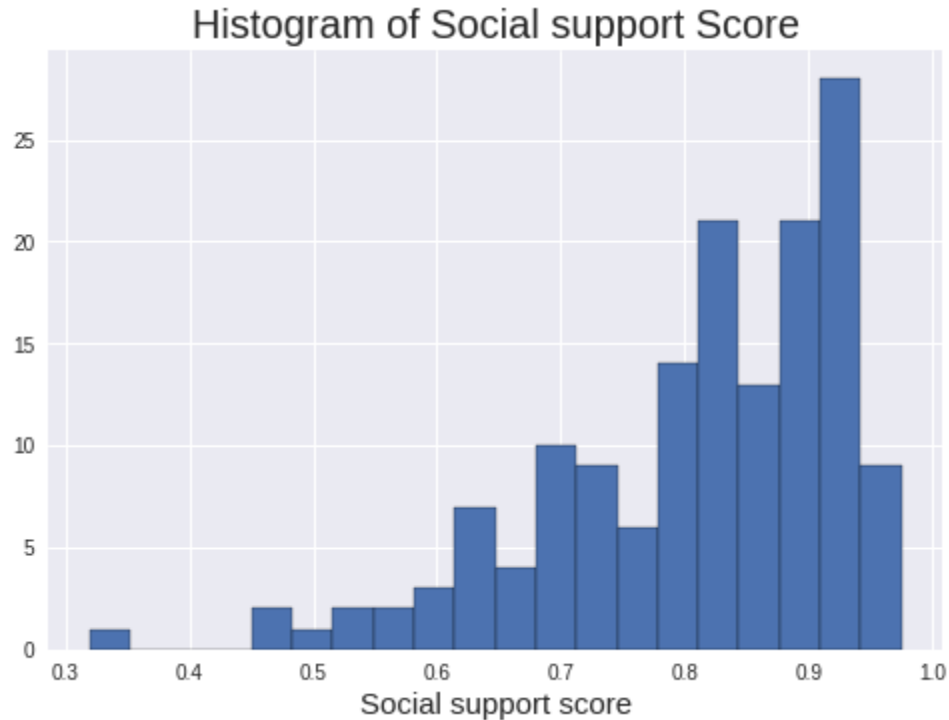

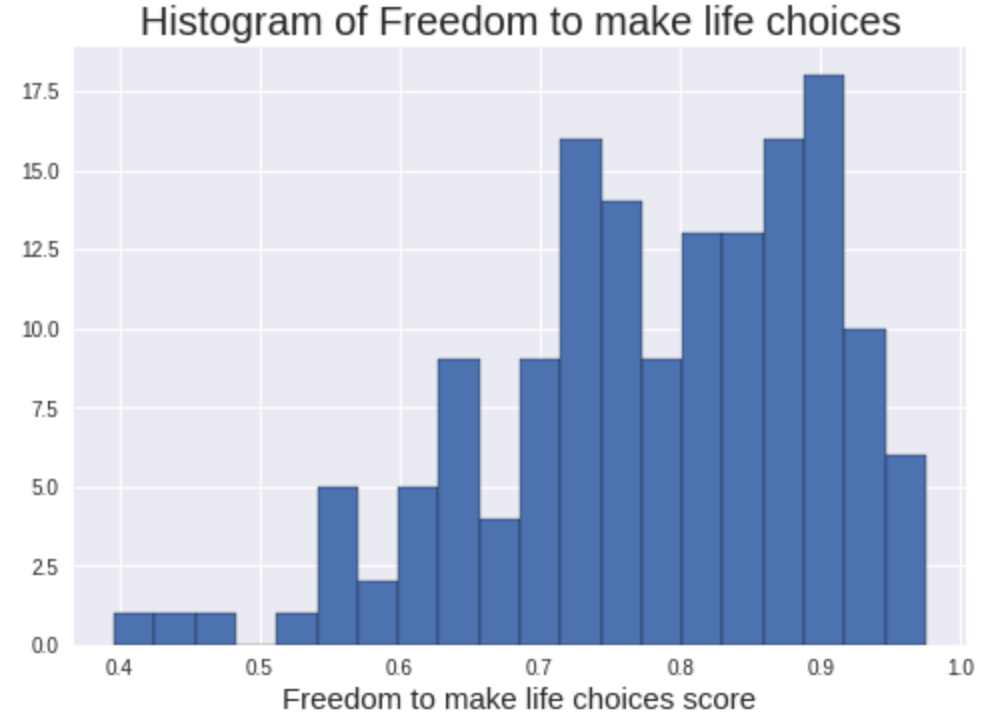Mean logged GDP/Capital of 9.29, with a range of 6.49 to 11.45

Mean healthy life expectancy of 64 years, with a range of 45 years to 76 years

Exploring some of the Xs (2/2) – Social support and freedom score

Histogram of Social support Score

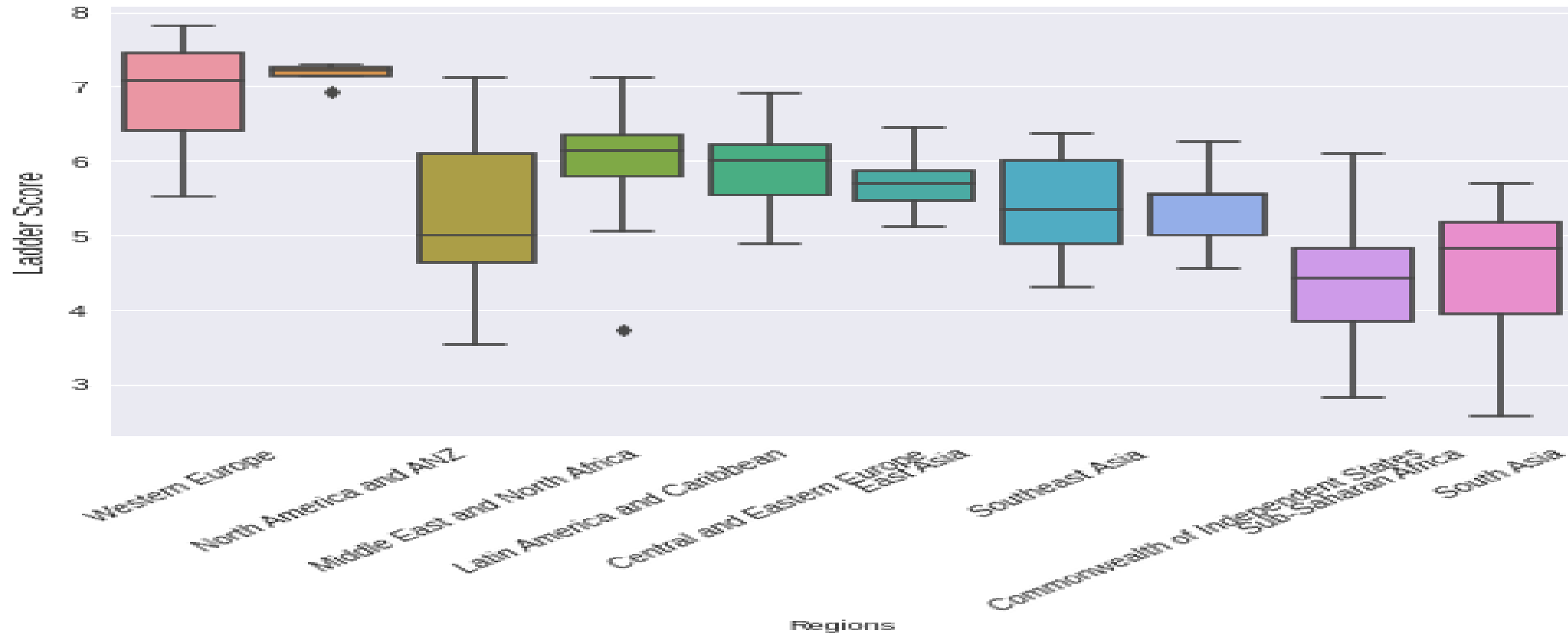Histogram of Freedom to make life choices

Mean social support score of 0.80 (on a scale of 0-1) with a range of 0.32 to 0.90

Mean freedom score of 0.78 (on a scale of 0-1) with a range of 0.39 to 0.97

Happiness score by region

As can be seen, there is a significant difference in happiness score by region. Western Europe and North America have the highest ranges, while Sub-Saharan Africa and South Asia lowest numbers

Exploring correlations

The correlation heatmap shows a number of important relationships:
• Very high correlation between happiness score (Ladder Score) and Social support levels, GDP Per Capita, Healthy life expectancy, Freedom to make life choices
• Not a very high correlation between Ladder Score and Generosity or Perception of Corruption
• Some of the features also show significant correlation with each other, for example, social support levels with GDP Per Capita