The Life Expectancy of a Poor Country will Enhance by Improving the GDP per Capita.

Assignment Name: Data Storytelling

Data Science Career Track

Springboard

Submitted By: Shahjahan Ahmed

01 November 2021

Storytelling topics:

The Life Expectancy of a Poor Country will Enhance by Improving the GDP per Capita.

Required Steps:

- ➢ Using visualizations to find the key attribute that affect the life expectancy of a poor country and evaluate the correlation.
- ➢ Conduct a hypothesis test to evaluate the significance of correlation.
- ➢ Summarize the findings and make appropriate suggestions.

Major findings:

- ➢ There is not a significant correlation between life expectancy and other attributes such as Corruption, Freedom, Generosity.
- ➢ There is a significant high correlation between GDP and social support. When GDP increases, social support also increases.
- ➢ Also, when social support increases, life expectancy also increases. So, high GDP is helping the life expectancy to increase.
- ➢ Finally, the Life expectancy is extremely correlated with GDP per capita
  a. Countries with high GDP have higher life expectancy
  b. Countries with low GDP have lower life expectancy
- ➢ To increase life expectancy, countries need to improve the GDP.

Target Audience:

Non-technical Audience

Data Sources:

https://www.kaggle.com/PromptCloudHQ/world-happiness-report-2019/code

Key attributes and their definition:

Life Expectancy:

Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age, and other demographic factors including sex.

GDP:

GDP stands for "Gross Domestic Product" and represents the total monetary value of all final goods and services produced (and sold on the market) within a country during a period.

Log of GDP:

If GDP grows at a constant rate, then the log of GDP, graphed against time t, is a straight line with slope equal to the growth rate g.

Exploratory data Analysis:

Dataset Summary:

The source of the dataset is Kaggle.com. The first data set contains the happiness ranking of 152 countries and contains the following attributes: 'Country', 'Ladder', 'SD_Ladder', 'Positive_Affect', 'Negative_Affect', 'Social_Support', 'Freedom', 'Corruption', 'Generosity', 'Log_GDP_Per_Capita', 'Life_Expectancy'.

Numerical distributions of data:

The histogram showing the distribution of numerical attributes. It is visible that there is no trend or skewness in the data. The dataset looks like uniformly distributed. There is also no outlier in the dataset and most of columns does not contain any null values.
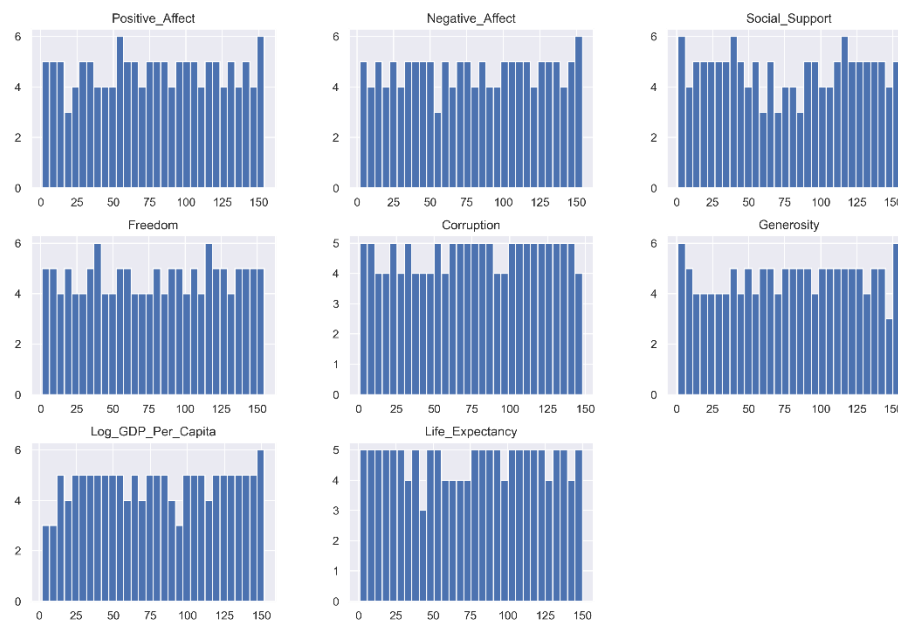

Figure: Histogram showing the distribution of numerical columns

Relationship between pair of attributes:

The pair plot showing the pairwise scatterplot among the attributes and the correlations among them. Among the pairwise scatterplots there is a correlation between 'Freedom' and 'Positive_Affect' columns, between 'Social_Support' and 'Negative_Affect columns, between 'Log_GDP_Per_Capita' and 'Social_Support' columns, between 'Life_Expectancy' and 'Social_Support' columns, and between 'Life_Expectancy' and 'Log_GDP_Per_Capita' columns. Among the correlations the 'Life_Expectancy' and 'Log_GDP_Per_Capita' columns showing significant positive correlation and the Pearson correlation coefficient is 0.87. Also, the heatmap showing the exact same pairwise correlations between columns.
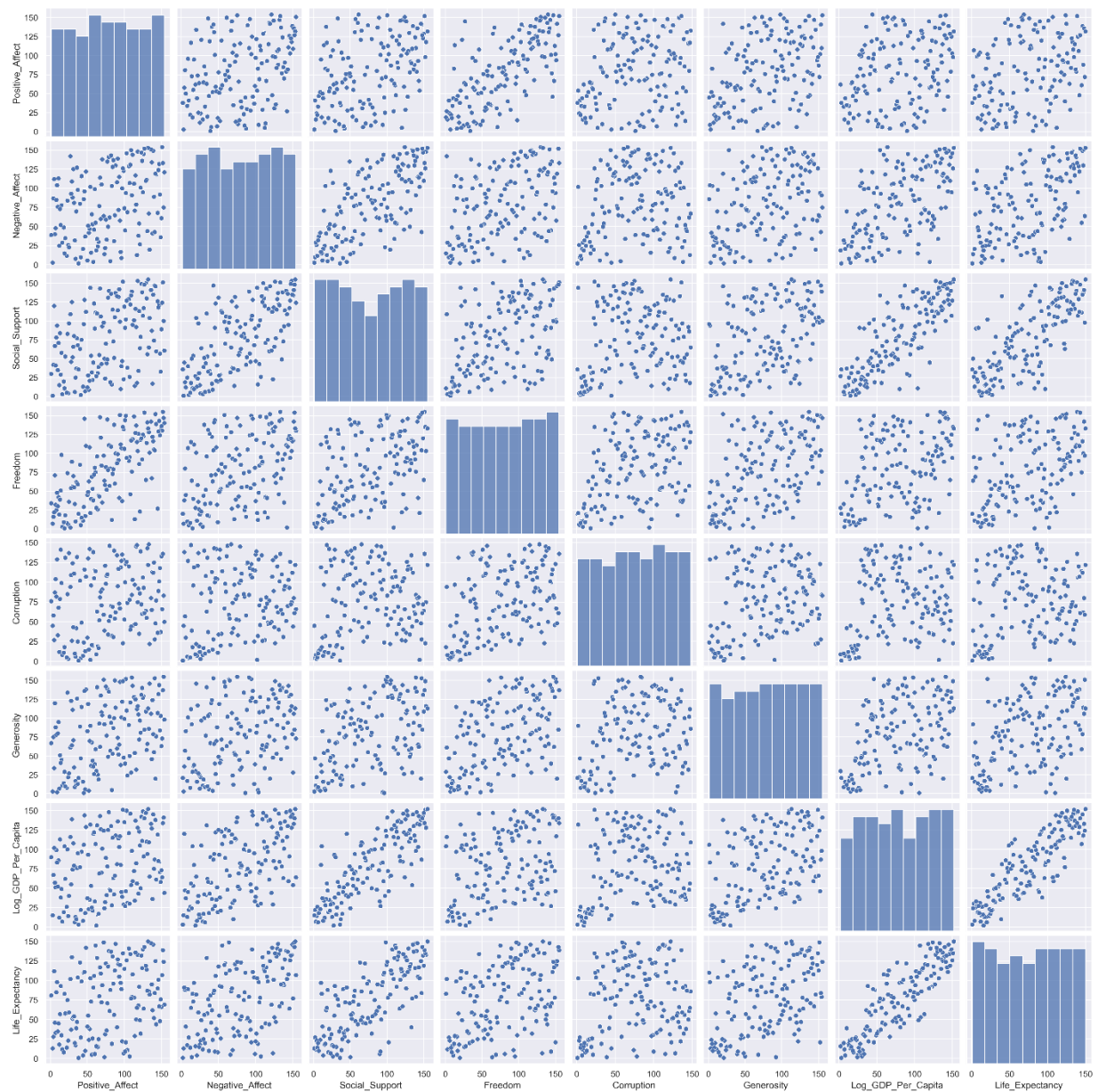

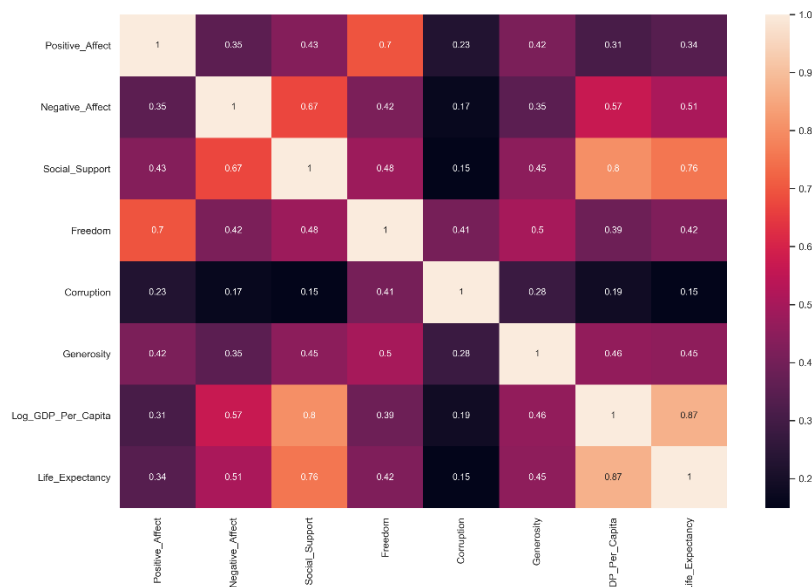Figure: The pairwise scatterplot of all the attributes

Figure: Heatmap showing the correlations

Visualizations to find correlation:

Non-significant correlations between 'Life_Expectancy' and other attributes:

➢ The regression plot is showing the scatterplot and the regression line between two attributes
➢ The regression plot between 'Freedom' and 'Log_GDP_Per_Capita' column is showing the non-significant correlation between the two attributes and the resulting Pearson Correlation Coefficient is 0.38.
➢ The regression plot between 'Life_Expectancy' and 'Corruption' column is also showing the non-significant correlation between the two attributes and the resulting Pearson Correlation Coefficient is 0.14.
➢ The regression plot between 'Life_Expectancy' and 'Positive_Affect' column is showing also the non-significant correlation between the two attributes and the resulting Pearson Correlation Coefficient of 0.33.
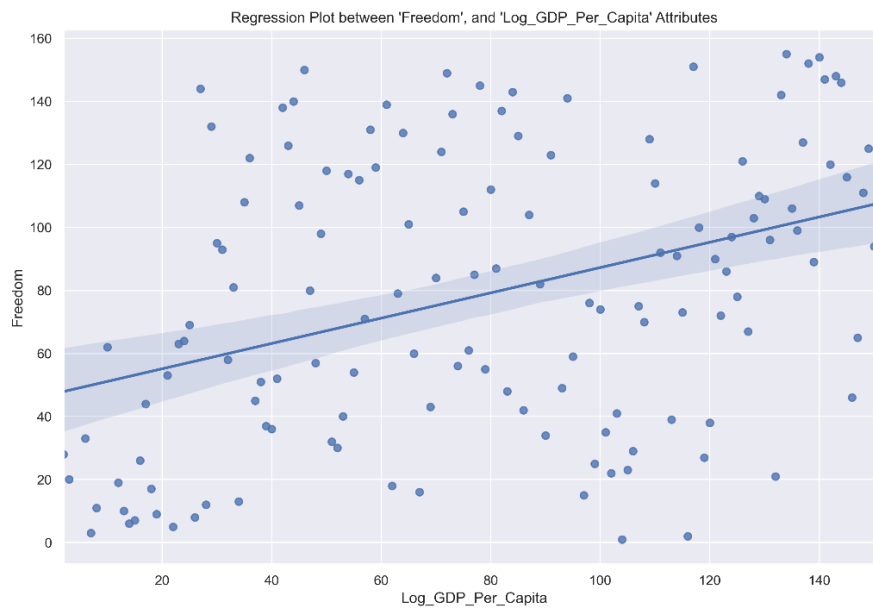
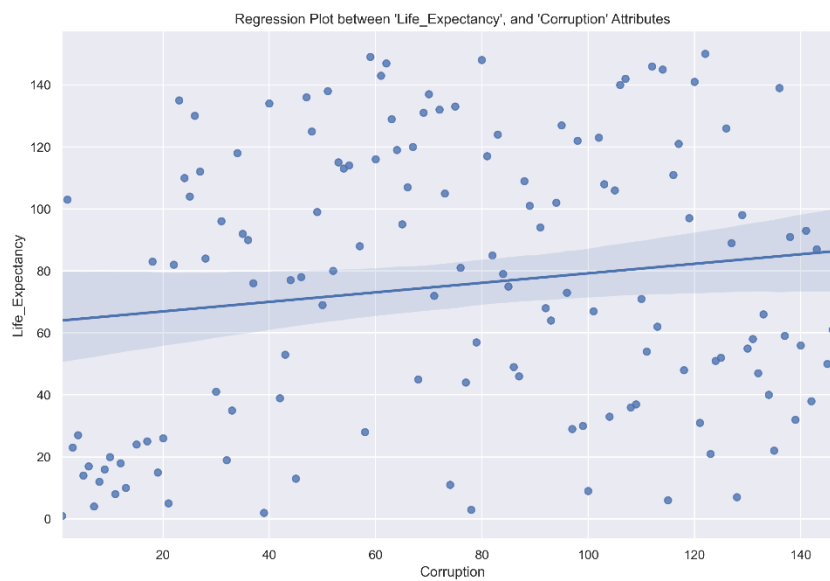Figure: Regression plot between 'Freedom' and 'Log_GDP_per_Capita' columns

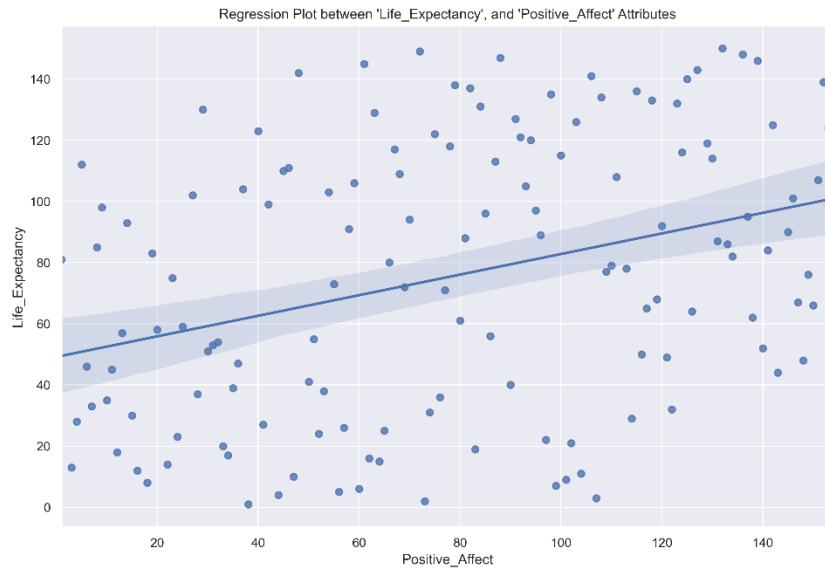Figure: Regression plot between 'Life_Expectancy' and 'Corruption' columns

Figure: Regression plot between 'Life_expectancy' and 'Positive_Affect' columns

Significant correlations between 'Life_Expectancy' and other attributes:

➢ The figures below shows that there is a significant positive correlation between 'Log_GDP_Per_Capita' and 'Social_Support' columns and the resulting Pearson correlation coefficient is 0.80.

➢ Improving the 'Log_GDP_Per_Capita' will improve the social support and vice versa.

➢ When the 'Social_Support' improves, the 'Life_Expectancy' will also improve. The Pearson correlation coefficient between the two attribute is 0.76.

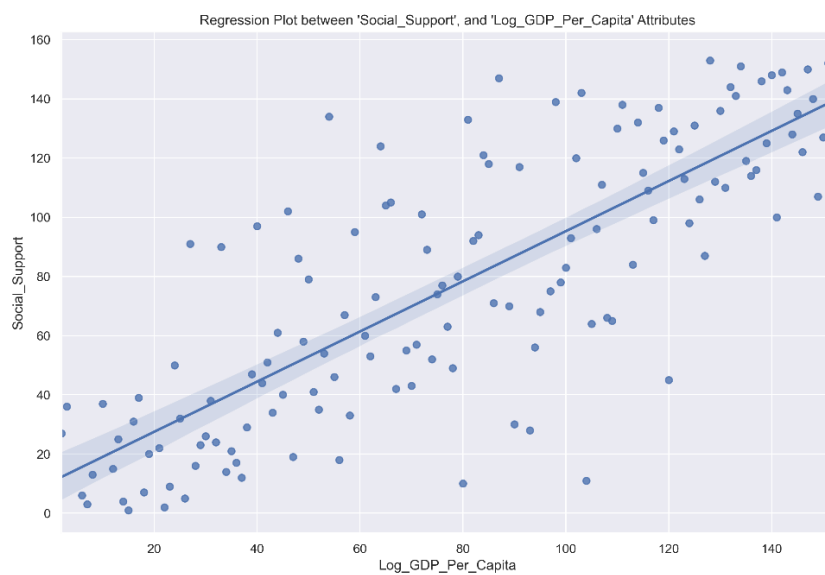➢ Finally, we can conclude that the 'Log_GDP_Per_Capita' can make a significant change in 'Life_Expectancy'.

Figure: Regression plot between 'Social_Support' and 'Log_GDP_per_Capita' columns

Figure: Regression plot between 'Life_Expectancy' and 'Social_Support' columns

Extreme Correlation between 'Life_Expectancy' and 'Log_GDP_Per_Capita':

Finally, by visualizing the regression plot between 'Life_Expectancy' and 'Log_GDP_Per_Capita' columns, it is evident that there is significantly high positive correlation between those columns. When the GDP increases the corresponding life expectancy also increases and vice versa. The Pearson correlation coefficient is 0.87, which is significantly high to make a conclusion that countries with high GDP will have high Life Expectancy.
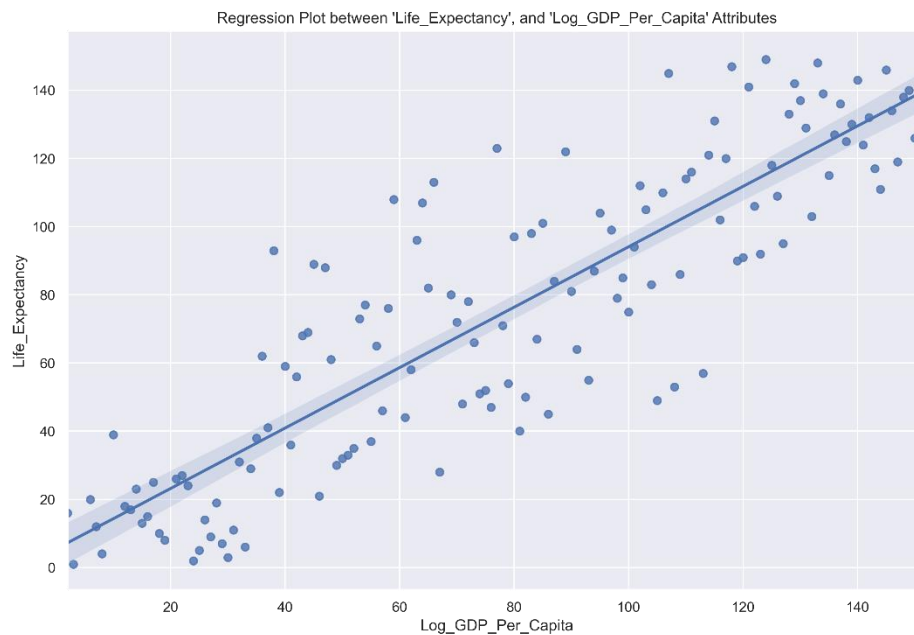

Figure: Correlation between 'Life_Expectancy' and 'Log_GDP_Per_Capita'

Hypothesis Formation:

Null Hypothesis; H_O: There is no correlation between life expectancy and GDP

Alternate Hypothesis; H_1: There is a correlation between life expectancy and GDP

Hypothesis Testing:

Firstly, I'm using the pairwise bootstrap for measuring Pearson correlation coefficient between life expectancy and GDP columns to compare with original Pearson correlation coefficient of 0.87. I have taken 1000 bootstrap samples and measuring the correlation coefficient between them. The regression plot is showing the visual distribution of regression lines and the histogram is showing the distribution of coefficients. From the histogram it visible that the Pearson correlation coefficient is significantly high. The 95% confidence interval for bootstrap correlation coefficient is [0.83106202, 0.90821219].
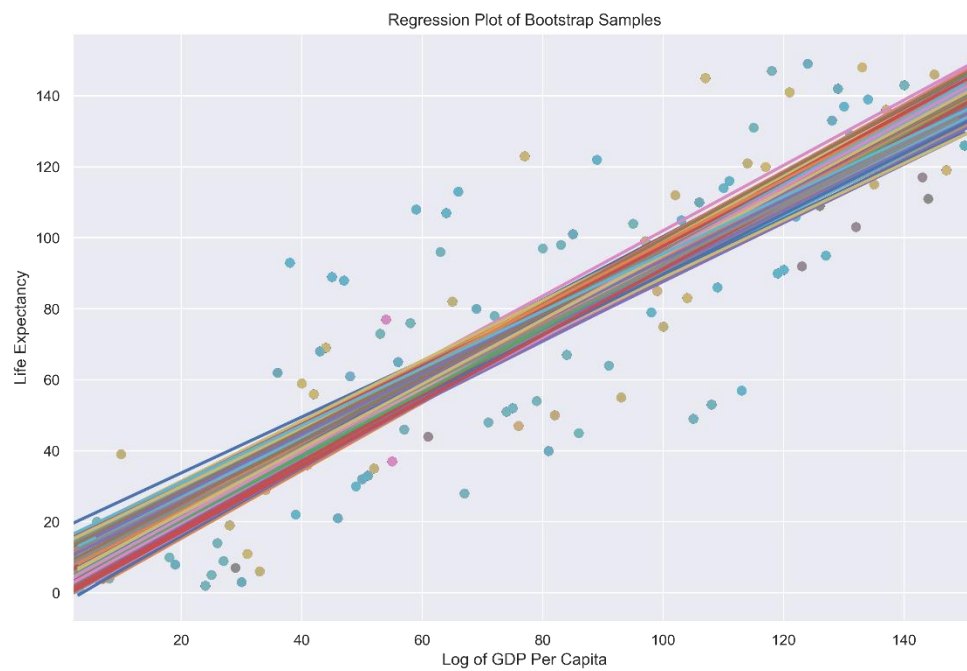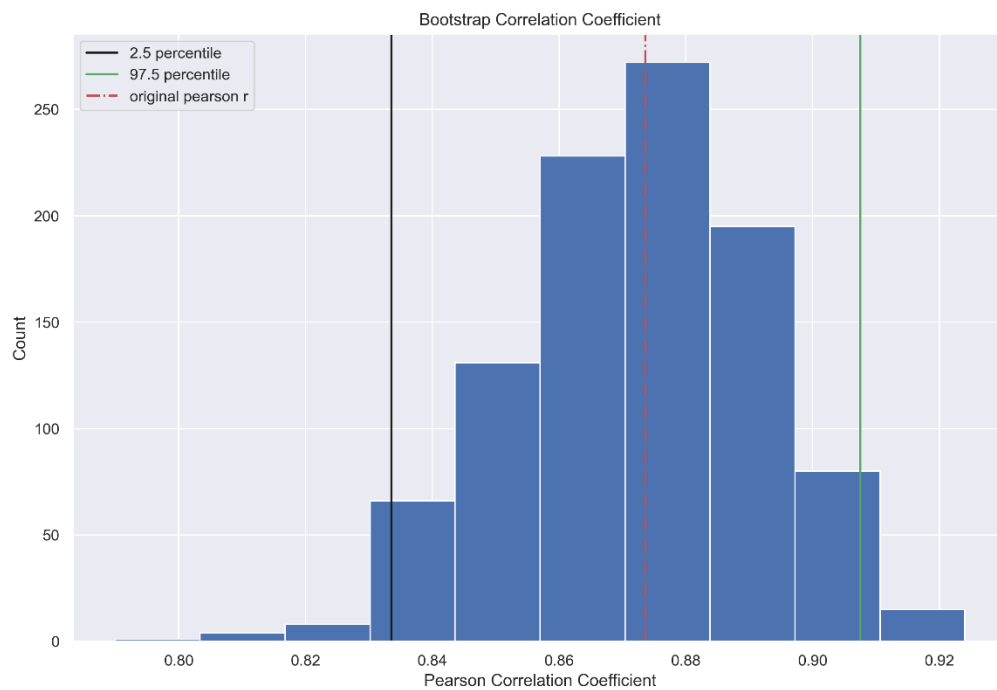
Figure: Regression plot of bootstrap samples

Figure: Bootstrap correlation coefficient

Secondly, I'm using the permutation test for measuring the significance of correlation. The histogram is showing that not a single observation is above the original Pearson correlation of 0.87, and the p value is almost 0. So, we are rejecting null hypothesis and accepting the alternate hypothesis that there is a significant correlation between life expectancy and GDP.
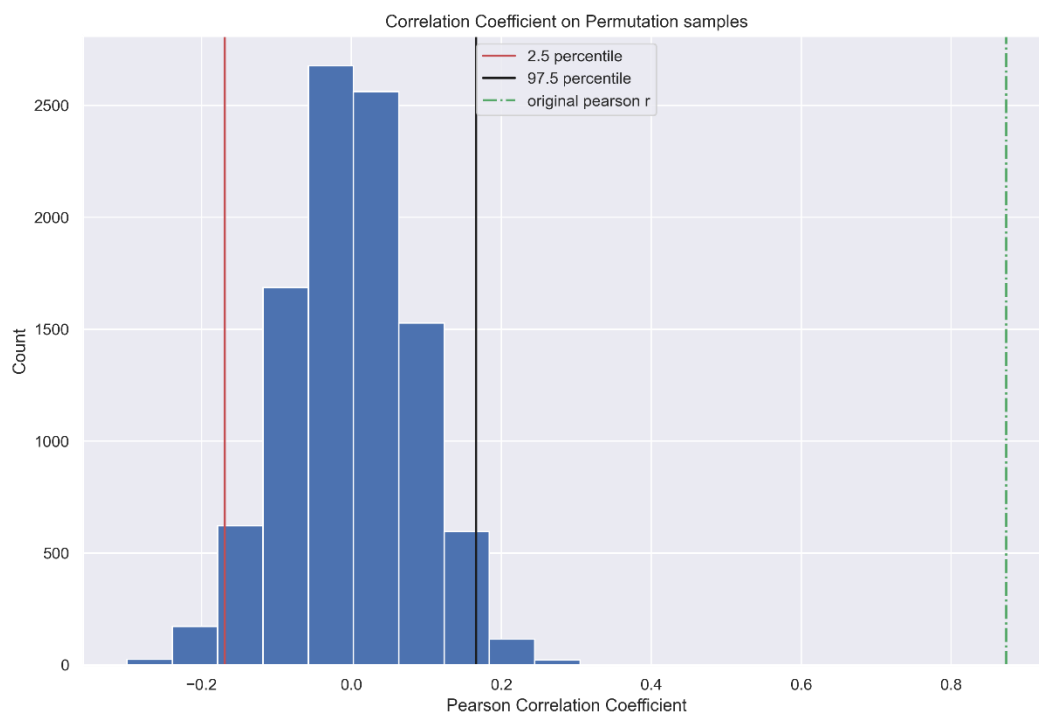

Figure: Hypothesis test on permutation samples

Conclusion:

By analyzing the data and using the hypothesis test I can conclude that compared with other attributes GDP is the main one which has extreme correlation with life expectancy. More broadly, countries with high GDP will have higher life expectancy. So, to enhance the life expectancy, poor countries need to improve the GDP per capita.