World Happiness Report

# Introduction

The World Happiness Report is a landmark survey of the state of global happiness. The report continues to gain global recognition as governments, organizations, and civil society increasingly uses happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy, and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

The happiness scores and rankings use data from the Gallup World Poll. The columns following the happiness score estimate the extent to which each of the six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world’s lowest national averages for each of the six factors. They do not impact the total score reported for each country, but they explain why some countries rank higher than others.

I have downloaded this dataset from the Kaggle source [[1]](#_References). I have selected it since it has a very high population among analysts and it also has great practical implications. I can learn many new things from its analysis. The objective is to analyze the effect of features on the happiness score. What are the top countries which have the highest happiness scores? What are those countries which have the highest social support values? We try to find out their answers.

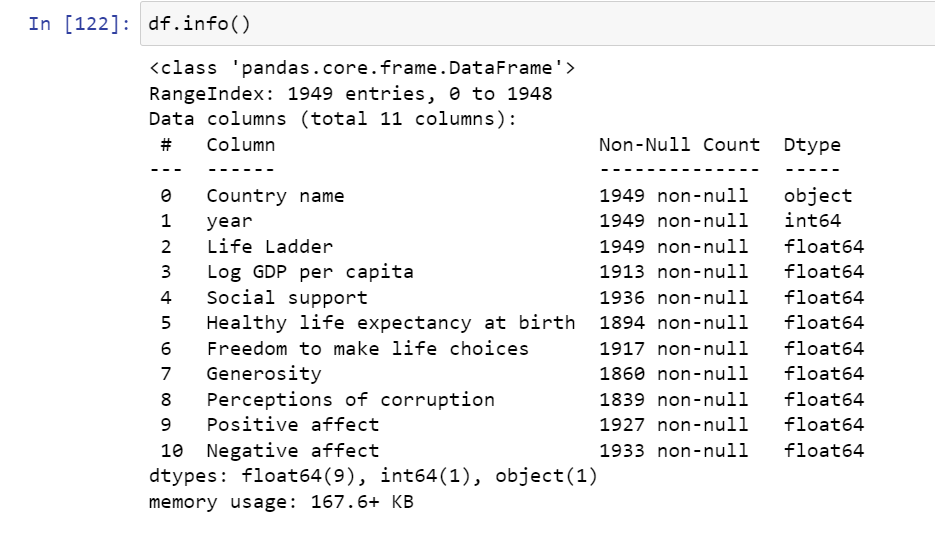
# Research

Xenia Kniazeva, a Kaggle user, has worked amazingly on this dataset [[2]](#_References). At first, she went through the whole data to understand its features as well as its statistics. She then extracted another feature, the happiness score feature, from existing features to have a single feature outcome that may explain the whole dataset. Thus, by looking at the correlation table she identified the features that are highly related to happiness score. So, she visualized them through graphs to understand them better. She also looked at the features data year-wise to have a better understanding.

# Processing

## Load the dataset

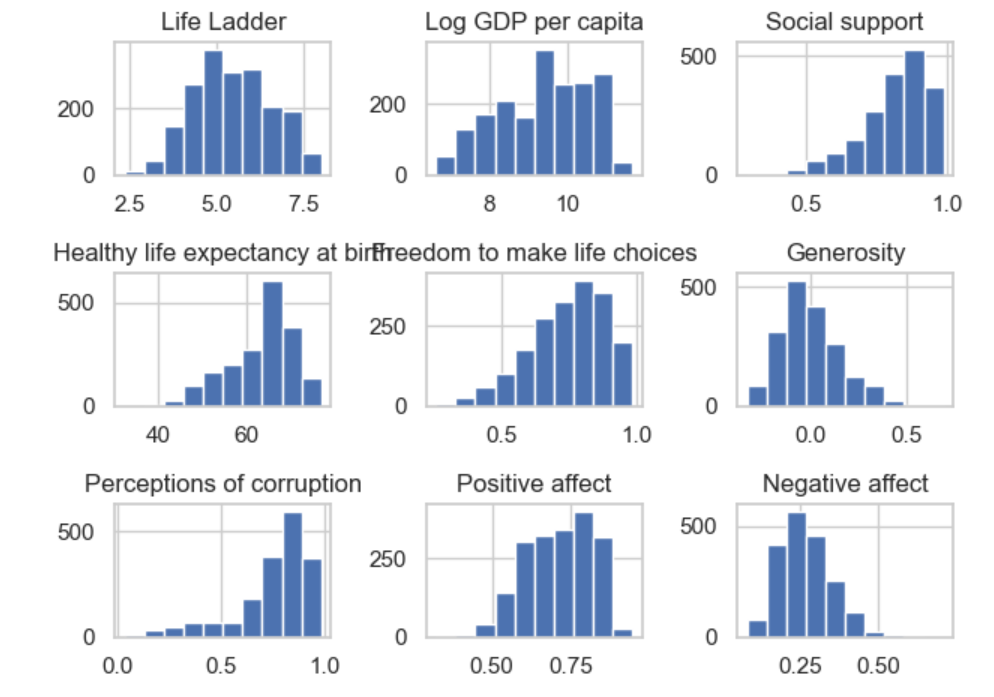




### Metadata about columns

1. Ladder score - means if you have the bigger score you can have better possibilities in life, than someone whose have less score
2. Log GDP per capita - means Gross Domestic Product (GDP) per capita (per person)
3. Social support - means that person has a family, friends, and other people, who can make support them in difficult life situations
4. Healthy life expectancy at birth - years of a healthy life that a newborn can expect to have
5. Freedom to make life choices - means that people can choose a job, family, friends, government, and place to live on their own choice
6. Generosity - means the quality of kindness and generous
7. Perceptions of corruption - means the degree of corruption in the country

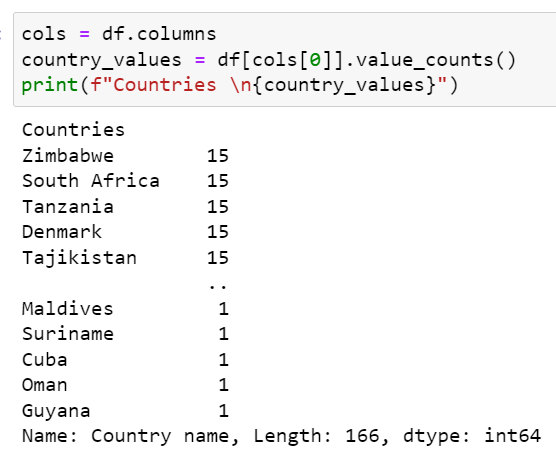
## Data Distribution

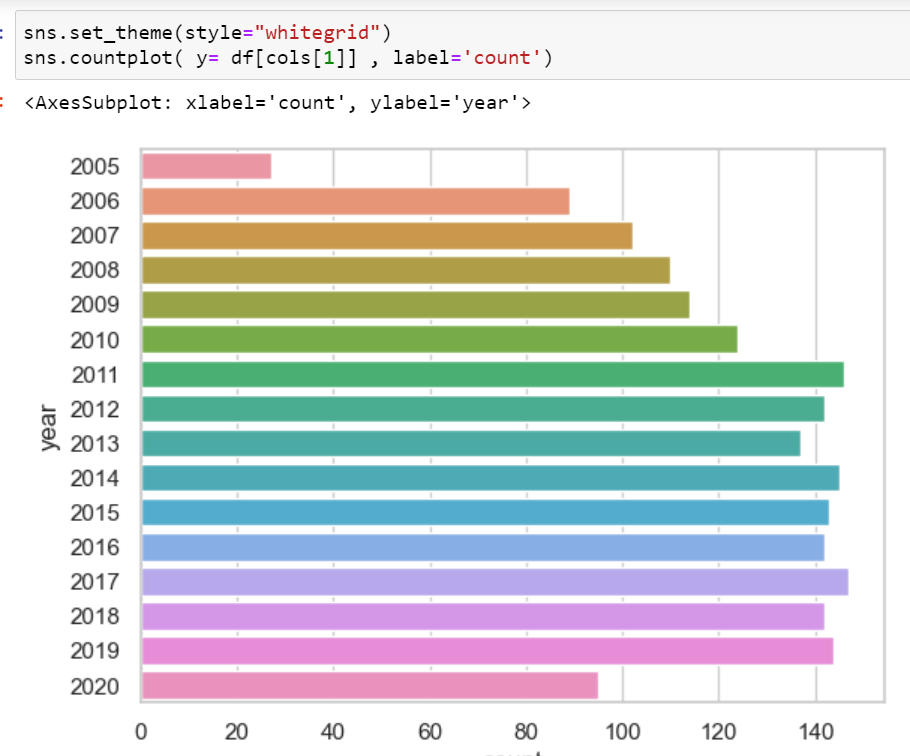


Most of the features have a normal distribution (bell curve distribution) except “Perceptions of corruption”, and “social support”. However, that does not affect so much on our analysis so we need to keep the data as it is.

## Data Frequency

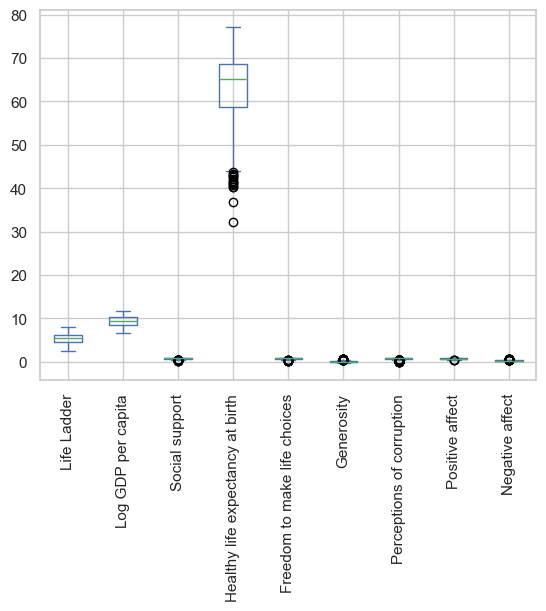
We have two categorical variables “Year” and “Country names”.





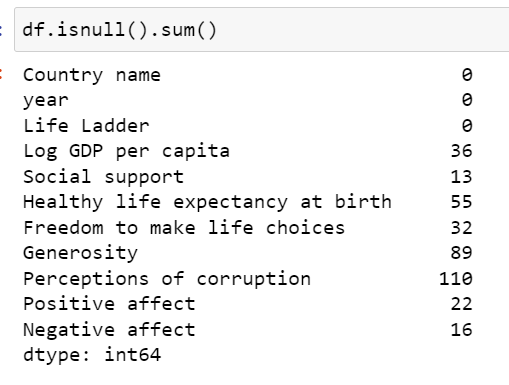
## Boxplots for Outlier detection

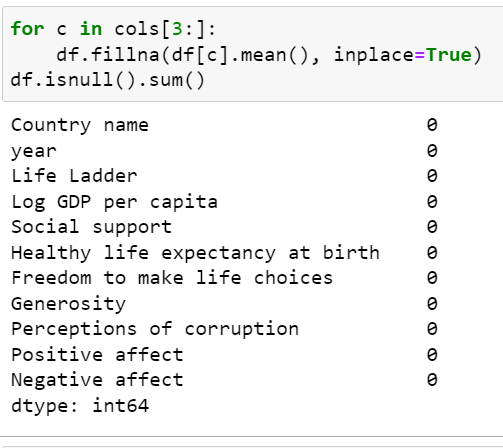
By visualizing the following boxplot, we can see that we do have some outliers in our data. However, these are not too many points and we also do not have a really large data set so we can leave them out. However, If they would have been much larger, they would have affected our analysis since they affect the mean values a lot. One thing we could do is scale the features that may resolve some outlier problems. However, by doing that we might lose the information from the feature that it is providing before.



## Missing data

Missing data points in any dataset are the biggest issue that needs to be carefully diagnosed. It is not always due to inaccurate data points. It can be because of any reason like someone forgetting to insert that value, it not getting properly processed while loading, etc. I used the technique of inserting the mean value to resolve this issue.



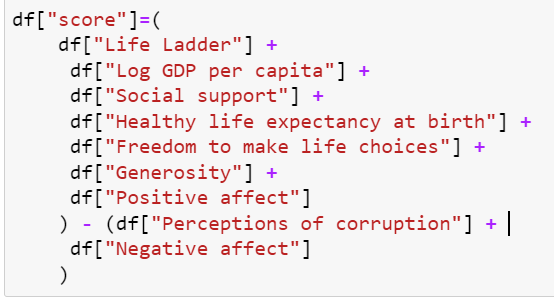


## Data Duplication

It can be a big problem for any dataset. However, there are no duplicate data in our dataset.

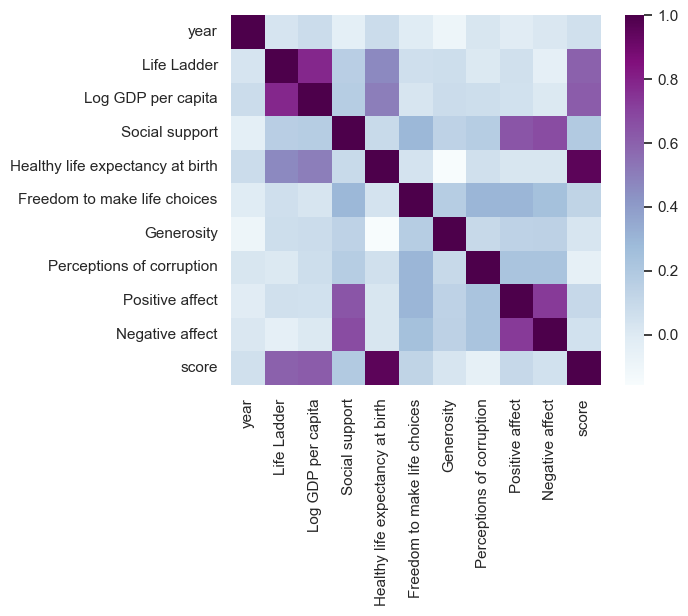
# Analysis

At first, I calculated the happiness score from the other features using the following formula. It is the main feature in which we are interested.



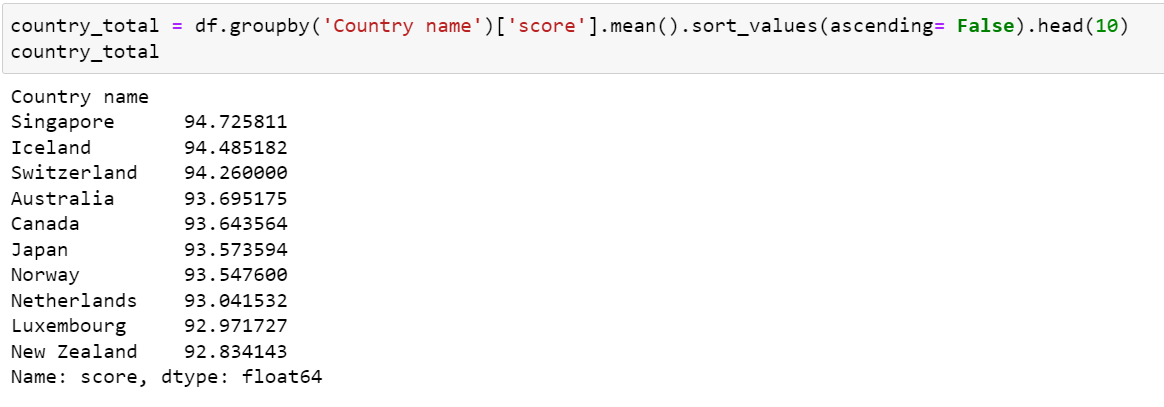
Next, there is a correlation graph using the following code of line

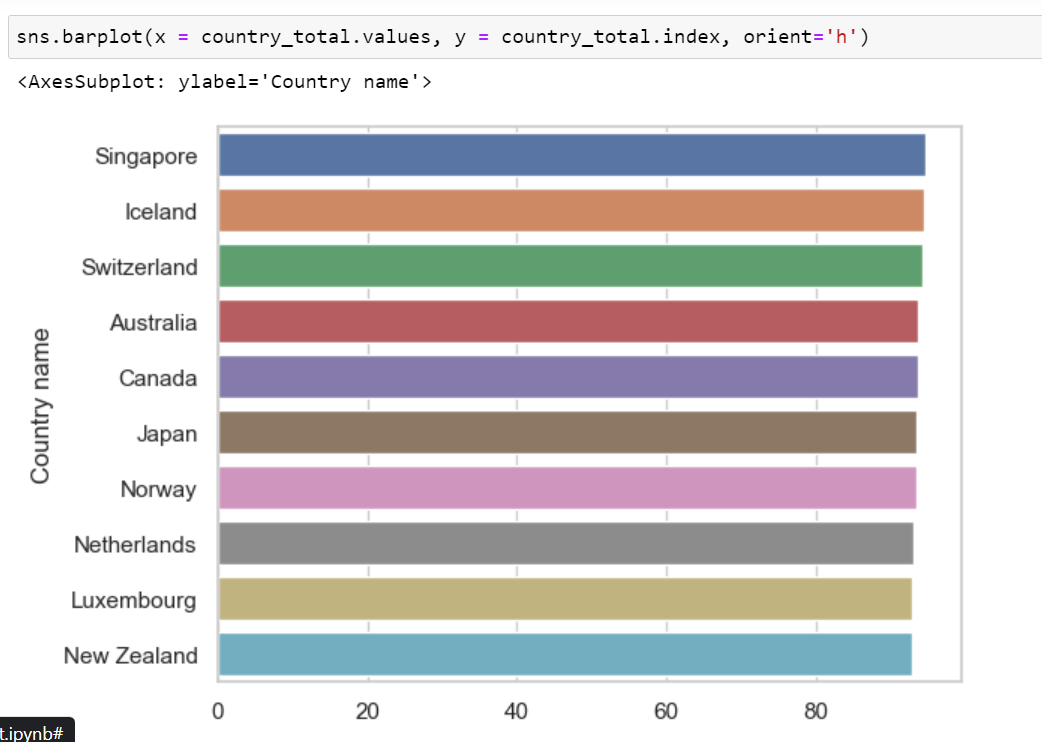




From this correlation graph, we can see that score (happiness score) is highly correlated with the feature “Health life expectancy at birth”. So, we can expect that countries with a high life expectancy at birth ratio would have better happiness scores based on our data.

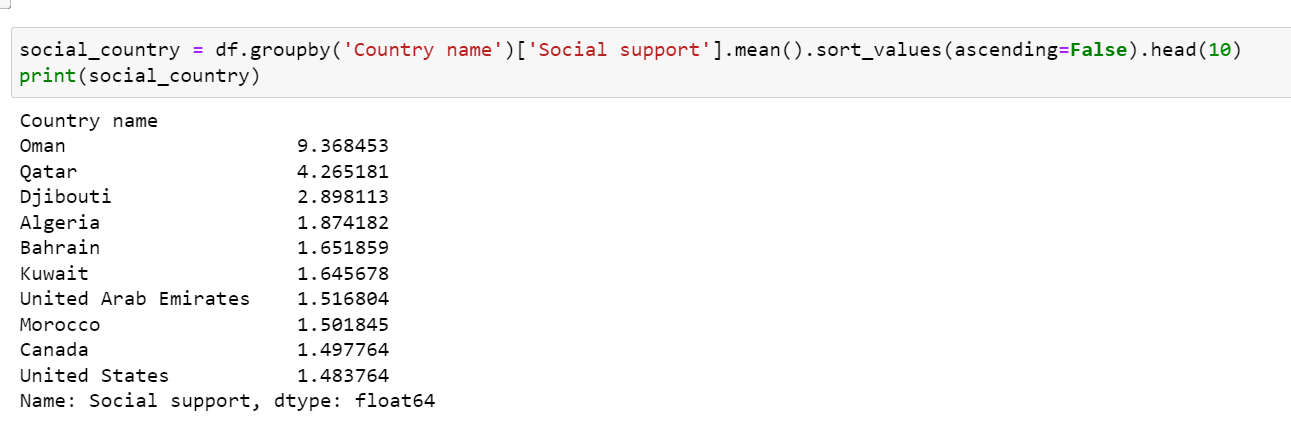
Let’s find out the top 10 countries which have the highest happiness scores:

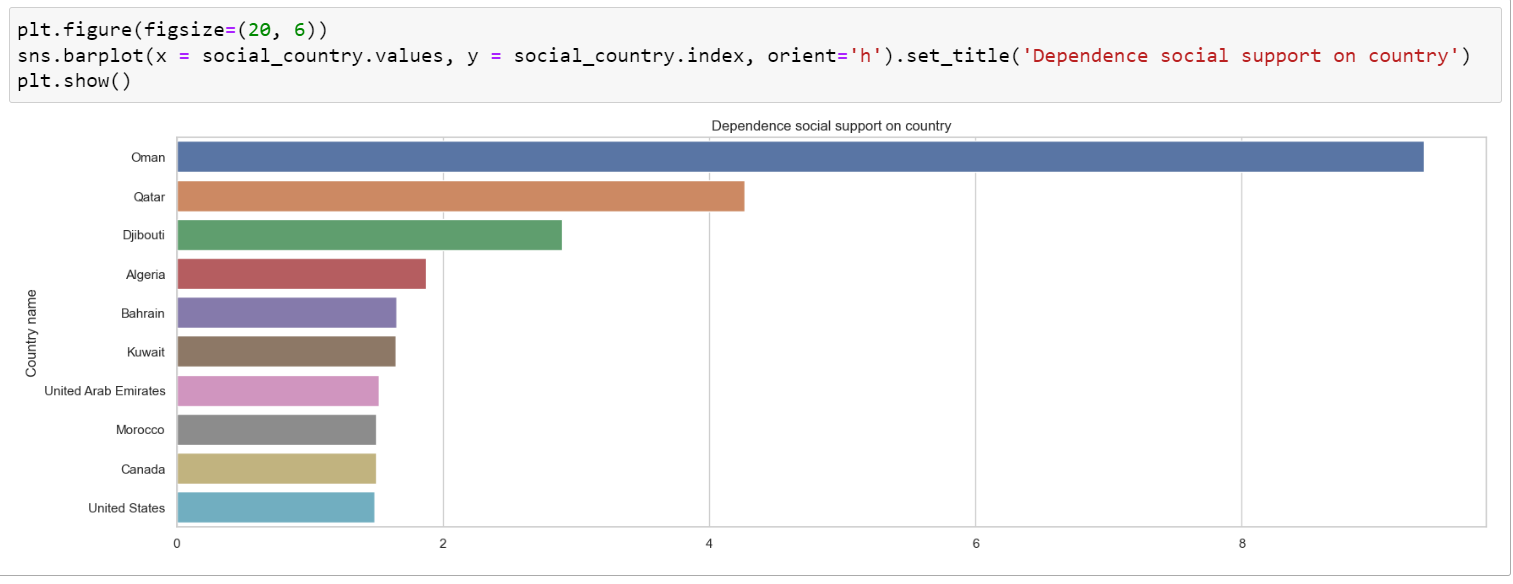




So, Singapore has the highest happiness score which is 94.

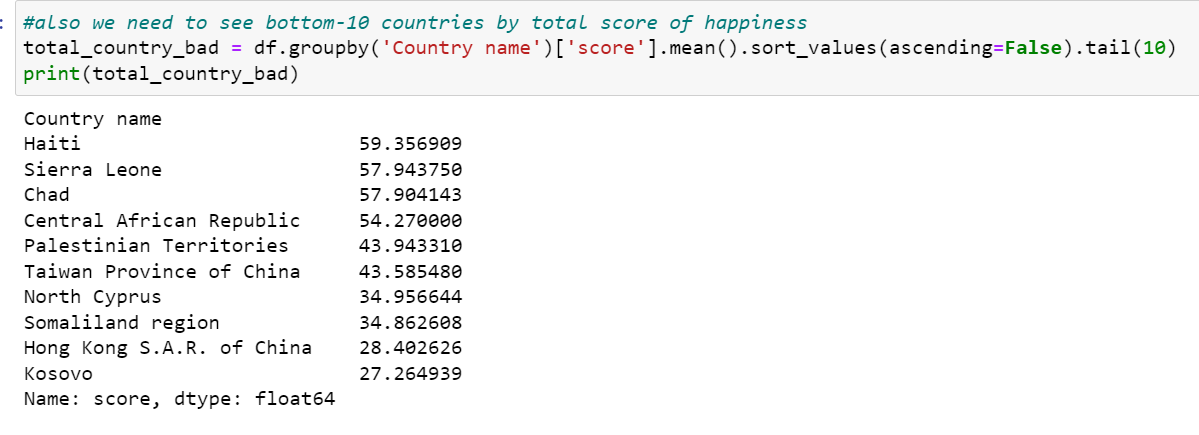
Now, Let us see what are the top social countries based on our dataset.

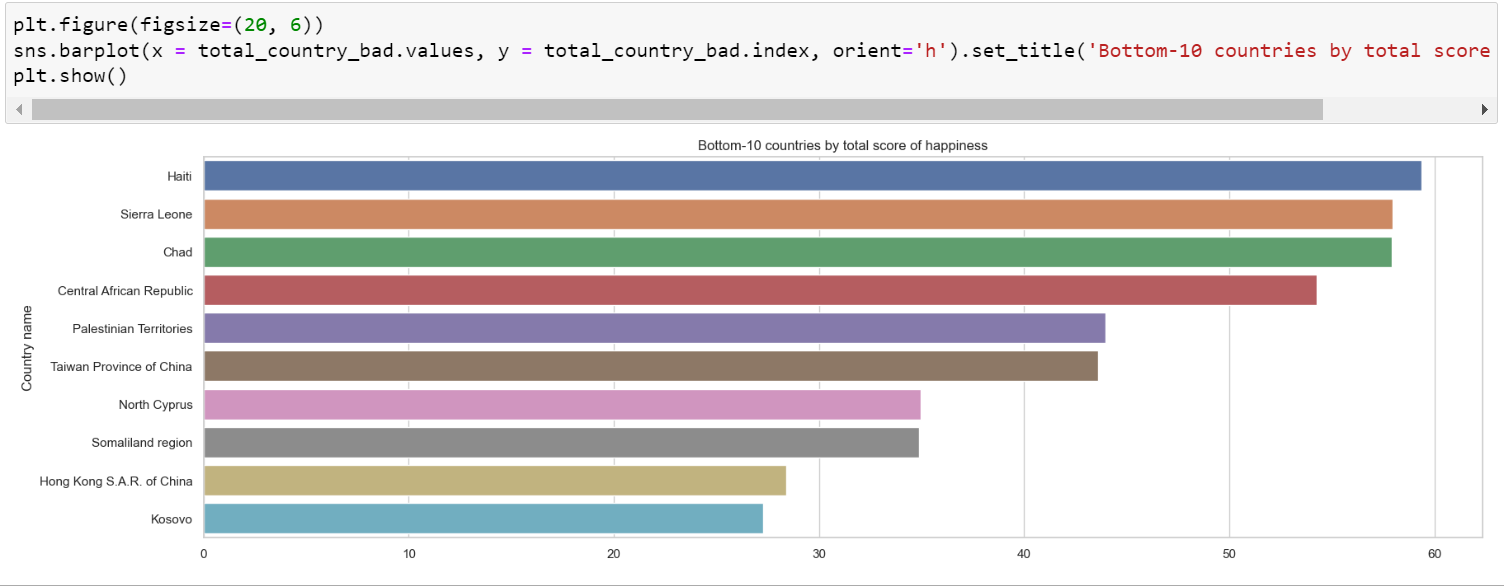




Oman is at first in this respective.

Some countries are in really a bad situation in terms of happiness scores.



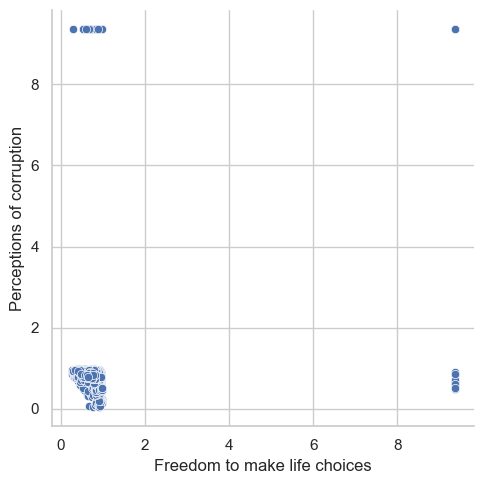


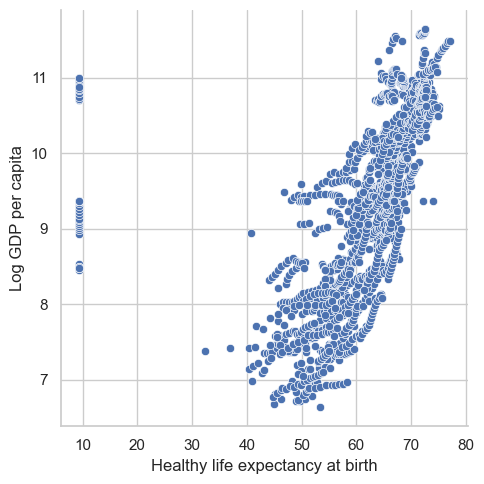
So, Haiti and some others are in really difficult situations in terms of happiness scores. As for the number five, we have Palestine which is a disputed territory right now and under the barbarism of Jews.

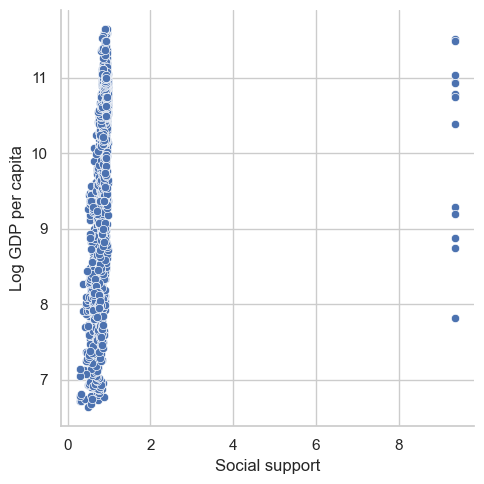
## Correlation by Scatter Plots

The following are the scatter plots between some features like score and life expectancy, log GDP vs life expectancy, etc. We can see that these are also highly correlated. However, the other two graphs on the right are not.









# References

[1]. [World Happiness Report 2021](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021?resource=download), Kaggle Data Source

[2]. [World Happiness Report EDA](https://www.kaggle.com/code/xeniakniazeva/world-happines-report-eda/data), Xenia Kniazeva