This project includes the analysis of a World Happiness Report retrieved from Kaggle. The site contains yearly datasets from 2015 to 2019, from which the dataset of the year 2019 has been considered for the analysis. The dataset is in the form of a CSV file containing the happiness scores of countries along with a few relevant factors. The goal of this project is to find the correlation of those factors with the world happiness score with the help of the provided dataset.

The World Happiness Report has gained recognition as governments and other organizations use this report while making and implementing their policies. It can be inferred that the analysis of this report to determine the significance of various factors affecting the world happiness score will be equally important for these policymakers.

The features provided in the dataset are GDP per capita, Social support, Healthy life expectancy, freedom of choice, Generosity, and Perceptions of corruption. According to intuition, it can be induced that GDP per capita, Social support, Healthy life expectancy and Freedom of choice can have a positive correlation with happiness score. Contrary to this, the perception of corruption will have a negative correlation with happiness scores. However, generosity does not seem to have a relation with the happiness score. **So, our hypothesis is that the feature GDP per capita, social support, and healthy life expectancy can affect positively for increasing the happiness score. Whereas the perception of corruption will have a negative impact on the happiness score. Furthermore, generosity has zero correlation with the happiness score.**

For the testing of the hypothesis, data will be analyzed by visualization techniques and correlation computation to test whether the given features are helpful for the prediction of happiness scores of different countries or not. The data will be passed through exploratory data analysis to detect if some information is missing or not so that the empty records can be handled before feature analysis. It is intended to group the data on the basis of different happiness score ranges so that the impact of feature value can be analyzed on different happiness score ranges individually. For instance, countries with a happiness score greater than 7 will be grouped together. Similarly, the other ranges will be 6-7, 5-6, 4-5, 3-4 and 2-3. For data visualization, box plots, histograms, scatter plots and bar graphs will be employed for better visibility of feature importance for happiness score prediction. In the end, the correlation of every feature attribute will be calculated with the happiness score for the verification of the hypothesis.

**References**

<https://www.kaggle.com/unsdsn/world-happiness>

<https://www.kaggle.com/unsdsn/world-happiness#2019.csv>