Data Analysis of the World Happiness Report 2021

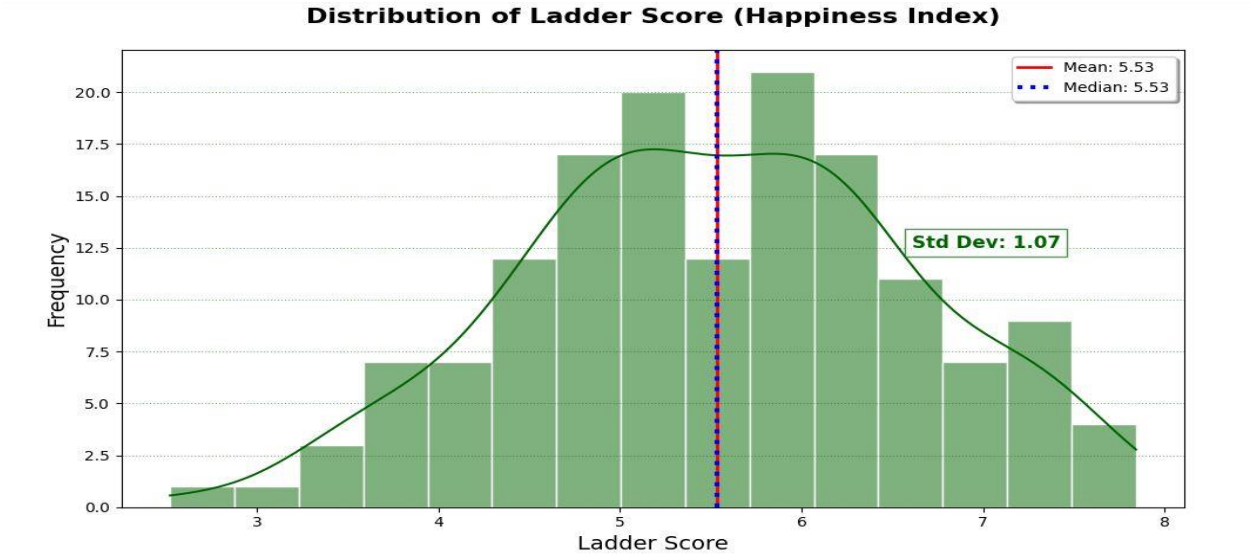
Jay Nakrani, Student ID No. : 18057042

GitHub *Link*

1. Introduction:

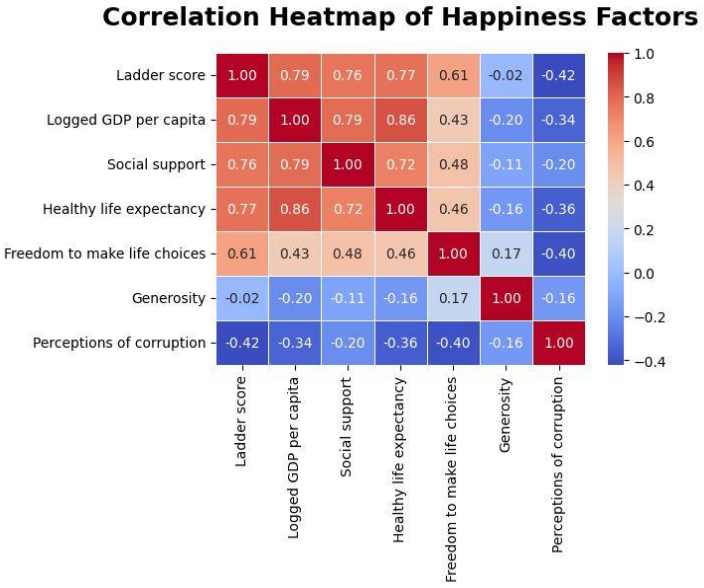
This report gives an overview for the performed analysis on the data of world happiness report for the year of 2021. This data has been created from different countries in the world. The data from year 2021 focuses on many key factors that affects happiness and overall wellbeing of their people, such as GDP per capita, life expectancy, and social support. In this study I have tried to identify the patterns and relationship between happiness score (ladder score) and different aspects provided in data through using different visualization methods.

2. Histogram for the Distribution of ladder score (Happiness score):



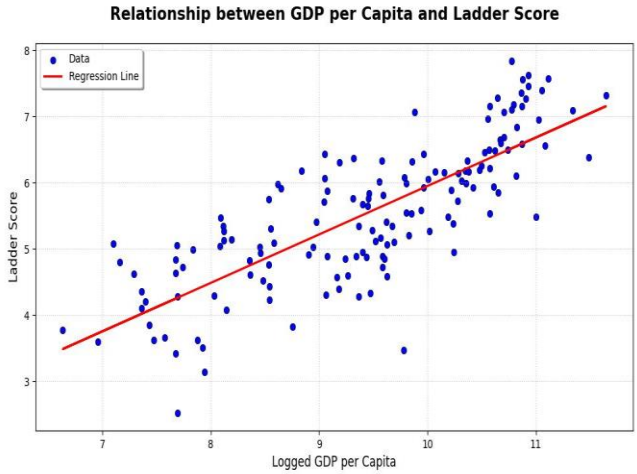
The above histogram highlights the allocated distribution of ladder scores (Happiness Index) from the data. This distribution is covering all the countries with their Happiness Index score (Ladder Score). Graph clearly shows the trend is moderately biased towards right, with most countries scoring Ladder Score between 4.5 and 7. With the mean (red line) at 5.53 and median (Blue line) at the same it can be indicated that there is balanced trend. The standard deviation of this distribution is 1.07, which illustrates a difference in happiness amongst countries. The graphic above shows that most countries cluster around average scores, with very few having extreme scores.

3. Heat map to understand the correlations between Happiness and other factors:



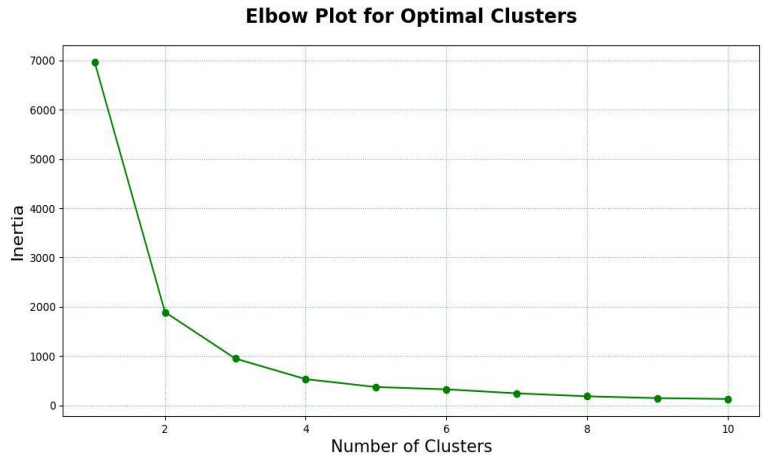
Second visualization performed to find the correlations between happiness and other key factors provided in the data. Heat map allows to understand the correlation of Happiness with other key factors. Dark red color shows a strong positive relation and dark blue color shows a very negative relation with Ladder Score (Happiness Score). For example, there is a strong positive relationship between Ladder Score (Happiness Score) and Social Support (0.76), Healthy Life Expectancy (0.77) and highest with Logged GDP per Capita (0.79). On the other hand, there is a negative correlation between Ladder Score and perceptions of corruption (-0.42), between Ladder Score and Generosity (-0.02), which clearly explains that GDP per capita, Healthy Life Expectancy and Social Support plays a major role in defining Ladder Score (Happiness Score). Further analysis has been done on Logged GDP per capita as it shows a strong relation with Happiness.

4. Scatter plot to understand the relationship between GDP and Happiness:



The third visualization performed to find the relationship between Ladder Score and The Logged GDP per capita. On the scatter plot, each blue point represents a country with displaying country's Ladder Score and Logged GDP per capita on y-axis and x-axis respectively. The regression line, which can be seen as a red line on scatter plot, indicates that there is a strong positive relation between Happiness and GDP per capita. However, scatter plot shows some fluctuation in point distribution around the regression line, which suggests that Happiness is not solely depend on GDP of the country, there are other factors to consider as well. Despite of this, regression line shows a linear trend in relation between Happiness and GDP which explains that economic success greatly boosts happiness.

5. K-means clustering and an Elbow plot:



In the last visualization, an Elbow plot has been created to find the ideal number of cluster for K-means clustering, which can be visualize in the graph on the left. Graph represents the number of clusters (k) and the inertia (a gauge of clustering quality) on x-axis and y-axis respectively. The figure shows the sharp fall in inertia from k=1 to k=3, after that it slows down. This explains that "elbow point" at k=3, three clusters are best to perform k-means clustering for this dataset.

The below graph has been generated by performing K-means clustering using optimal clusters k=3. Red point cluster on the graph highlights the rich countries with the high GDP and high Happiness ratings indicating the correlation between happiness and GDP. Following green square clusters and blue cross clusters shows middle-income countries and low-income countries respectively. This analysis confirms that in 2021, countries economic circumstances played a big roll to determine the well-being of their people.