Concepts and Technologies of AI

5CS037

Title: Analysis of the World Happiness Report: Exploring South Asia and Middle East Perspectives.

Name: Supriya Kunwar
Group: L5CG7
Student ID: 2408484
Module Leader: Siman Giri

Introduction:

The World Happiness Report is a useful publication that seeks to establish which country's citizens are most and least happy. This makes it informative on other variables that determine happiness like social support, income and health hence management can be able to understand what constitutes wellbeing by governments and organizations. It is a good working-working paper for expanding the understanding of how countries and territories feel happiness, and how it is possible to use this knowledge to build policies that will enhance the citizen's well-being.

There are three aims of this report. The first of these is to bring out features about global happiness from the data. The second strategy is to concentrate on South Asia and assess the determinant of happiness in this geography. Finally, it contrasts South Asia and the Middle East to identify how the indices of happiness differ in these two regions.

Problem - 1: Getting Started with Data Exploration

In this section, an analysis of the World Happiness dataset will be made in order to obtain the first insights.

Dataset Overview:

The first step in the analysis was to load the dataset and review the initial 10 rows of data to get a feel for the structure of the data being used. We then analysed the number of rows and columns and listed all the dataset's columns along with their data types.

Basic Statistics:

Mean, median and standard deviation of the score column were all computed to show the spread of the data in terms of happiness. Also, we determined which countries fall into two groups – those with the highest happiness scores and the ones with the lowest.

Missing Values:

We tried to detect if there are any missing values in the dataset and for this purpose to-complete missing values were selected and got to know that there are 3 missing entries in each of the following: Log GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, Perceptions of corruption, and Dystopia + residual.

Filtering and Sorting:

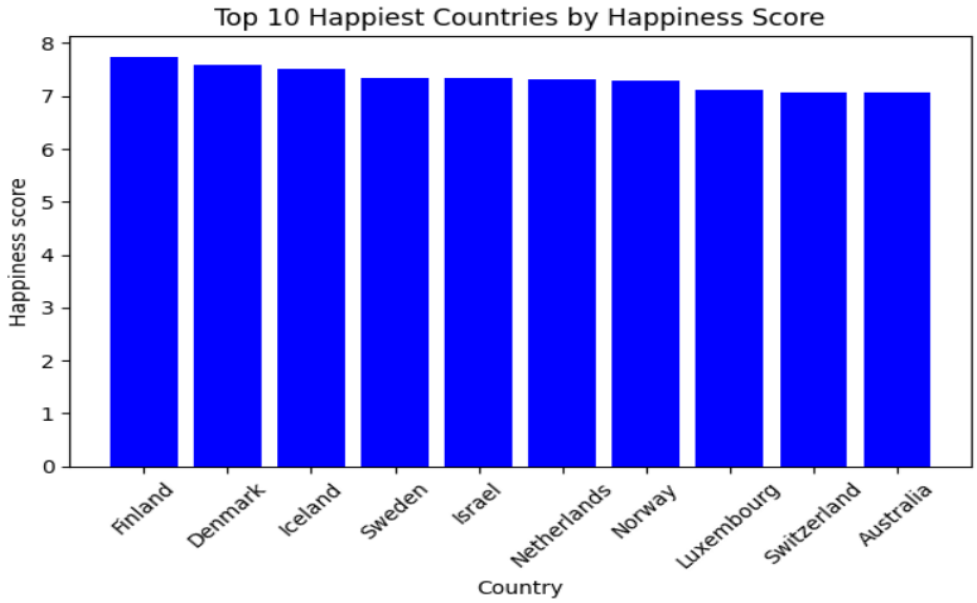
We also used a filter to present only the country data with happiness score more than 7.5, then sort the data base on GDP per Capita in descending order and take the first ten. Since the happiness scores greater than 7.5 were reported only by 3 countries, data of only 3 countries: Finland, Denmark and Iceland were shown.

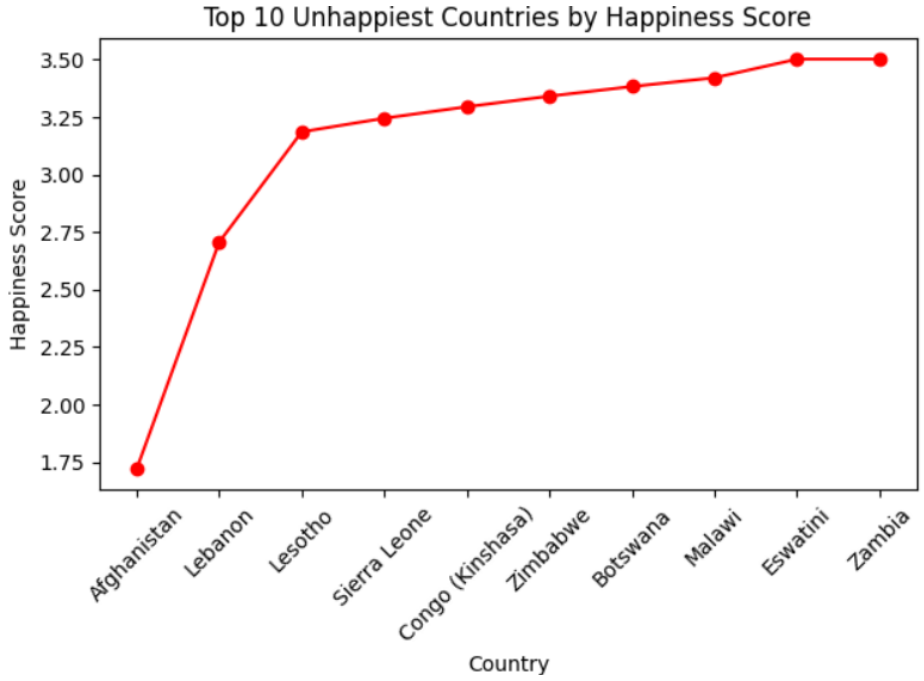
Adding New Columns:

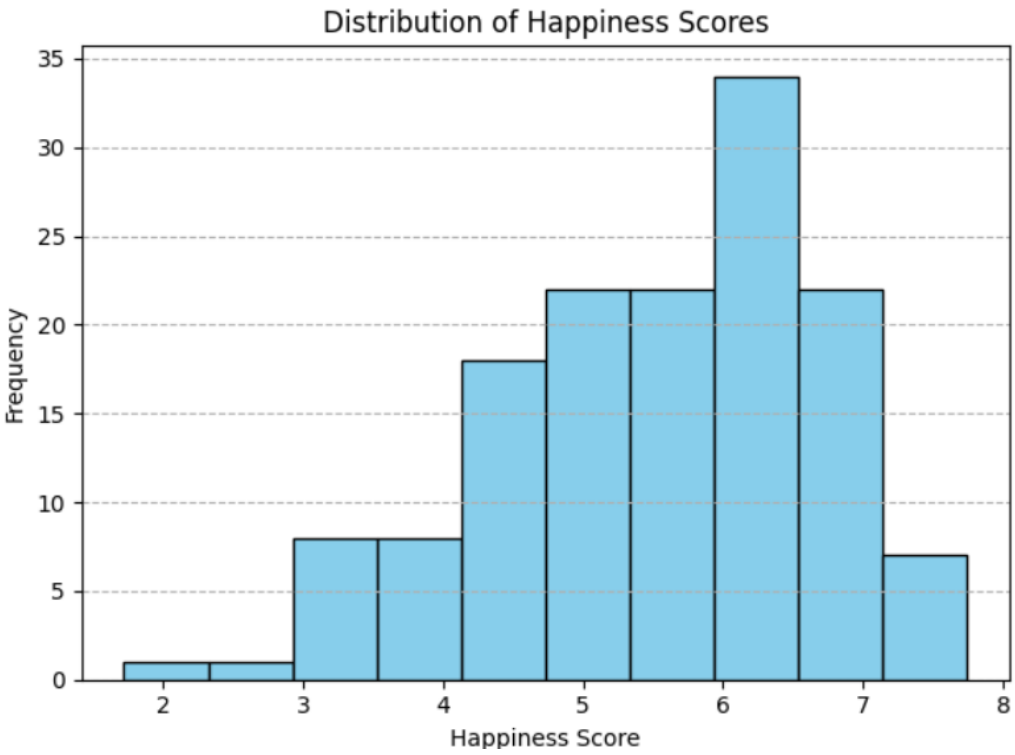
A new column, "Happiness Category," was created. Countries with a score lower than 4 were put in the low, countries with a score between 4 and 6 were put in the medium, and countries with above 6 score were put in the high happiness category.

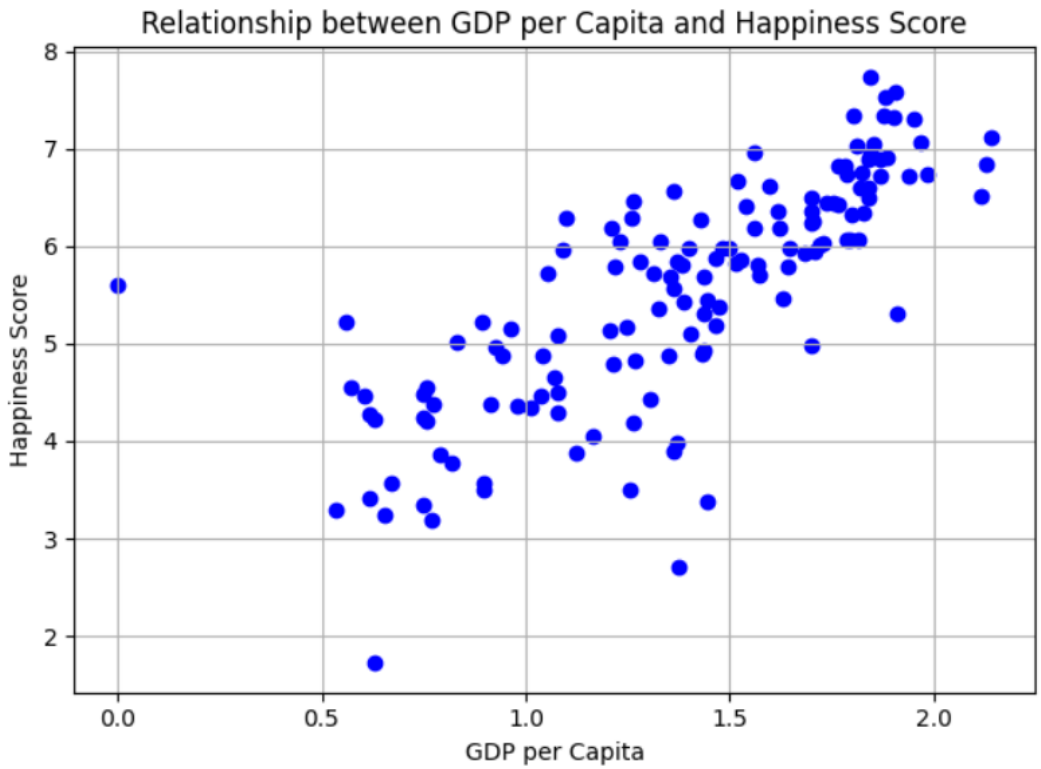
Data Visualizations:

For the visualization, we created bar chart where we show the 10 happiest countries and for the second one – line chart where we compared the scores of 10 most unhappy countries. The histogram of the "Score" column was created to analyze distribution, and that shows certain trends about happiness across the globe. Additionally, a scatter plot was created between GDP per Capita and happiness scores to explore their relationship, revealing potential correlations between wealth and happiness which are shown below:






Distribution of Happiness Scores
Frequency
Happiness Score
0
5
10
15
20
25
30
35
2
3
4
5
6
7
8


Relationship between GDP per Capita and Happiness Score
Happiness Score
GDP per Capita
2
3
4
5
6
7
8
0.0
0.5
1.0
1.5
2.0

Conclusion:

In this problem, we first familiarized ourselves with the structure of the World Happiness dataset, looked for missing values in the data set and gained a first impression of the data. Other findings are the happiest and the unhappiest countries: the former is Finland the latter is Afghanistan. The correlation with happiness was confirmed by a positive association of happiness scores with Log GDP per capita on a scatter plot. Also, a histogram presenting the distribution of the happiness score emphasized that most countries scored between middle and high.

Problem - 2 - Some Advance Data Exploration Task:

In this section, we performed analysis of the 'World Happiness" dataset with an emphasis on the Middle Eastern nations.

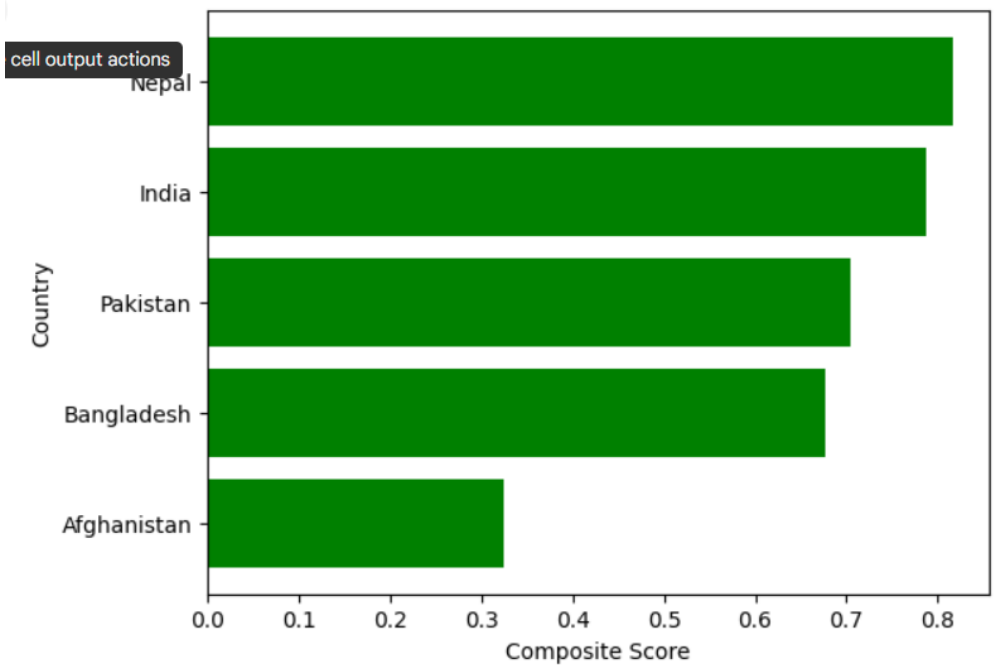
Task - 1 - Setup Task - Preparing the South-Asia Dataset:

We created a subset of the dataset containing only South Asian countries, identified using a predefined list of countries. The dataset was filtered to extract only these countries. The first few rows of the subset were displayed to verify the data and saved the filtered dataset as a separate csv to do upcoming tasks.

Task - 2 - Composite Score Ranking:

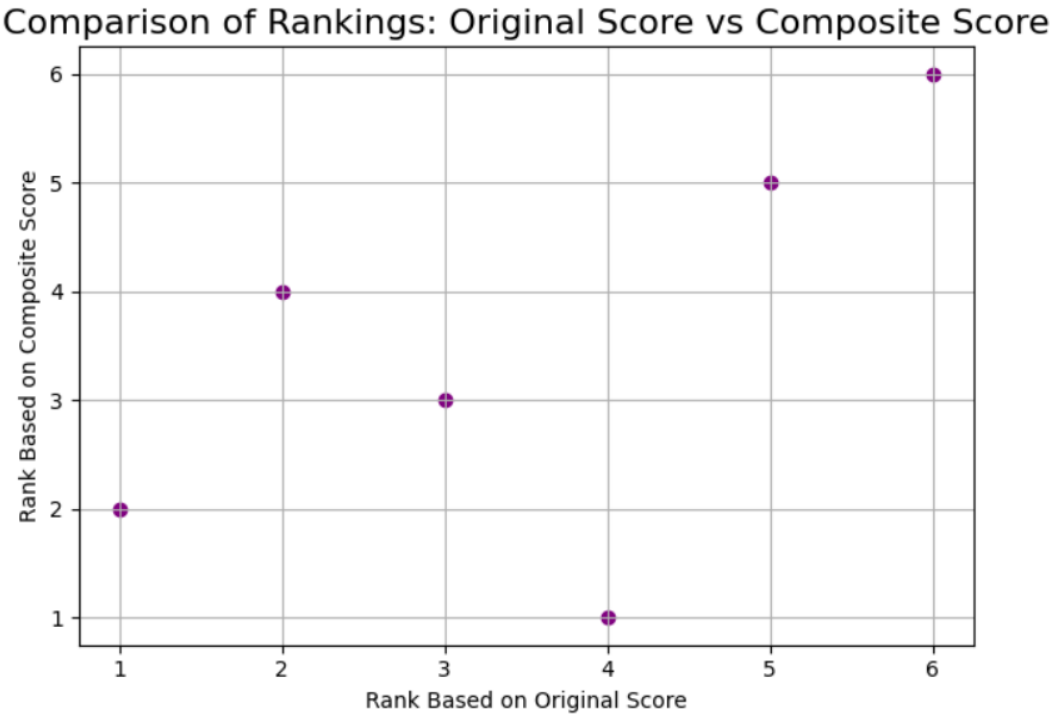
Using the South Asian data frame, we ourselves calculated a Composite Score using the given formula and a new variable column named ‘Composite Score’ was created. The nations were then sorted based on their compound totals; therefore, the higher scores tend to represent higher places. The initially ranked countries had better performance in economic, social and health aspects.

To illustrate the rankings, a horizontal bar chart of the top 5 countries based on composite score was created as shown below:



The bar chart reveals that Nepal has recorded the highest Composite Scores compared to India while Afghanistan poses a major challenge by showing the lowest Composite Scores.

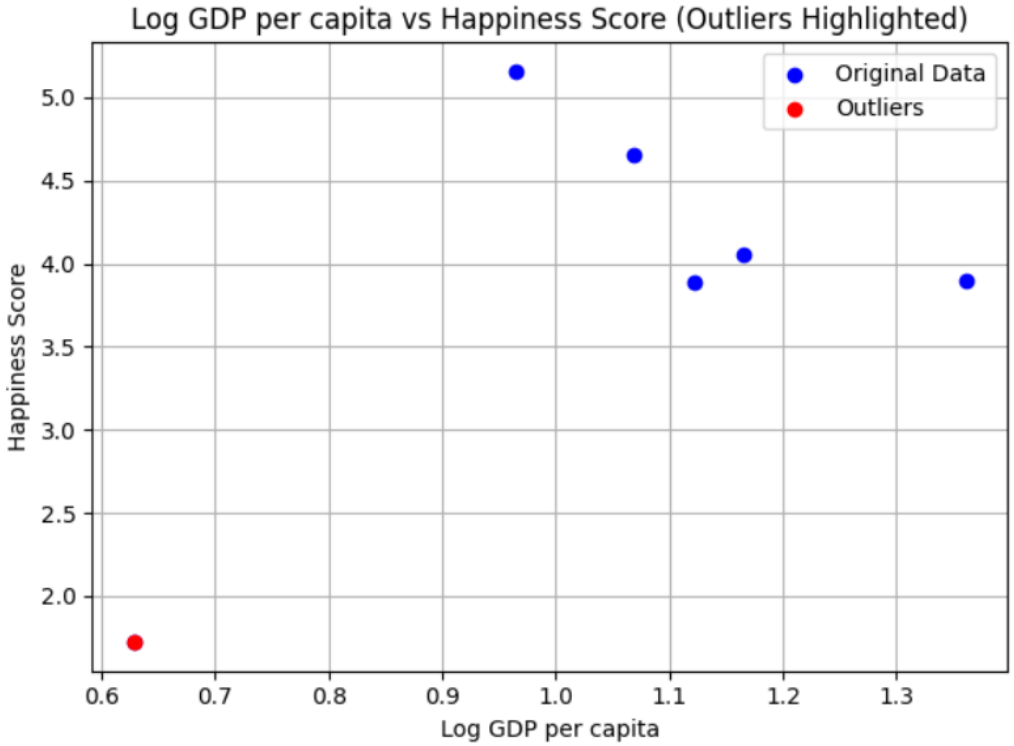
A scatter plot was presented to evaluate how well the Composite Score correlates with the original happiness score. Below is a comparison of the original happiness score with the composite score:



The scatterplot reveals that GDP, social support, and life expectancy (composite score) bring some differences in rankings as compared with the overall happiness score.

Task - 3 - Outlier Detection:

Outlier detection was performed on the South Asian countries using the "Score" and "GDP per Capita" columns. A scatter plot was created to show outliers as below:

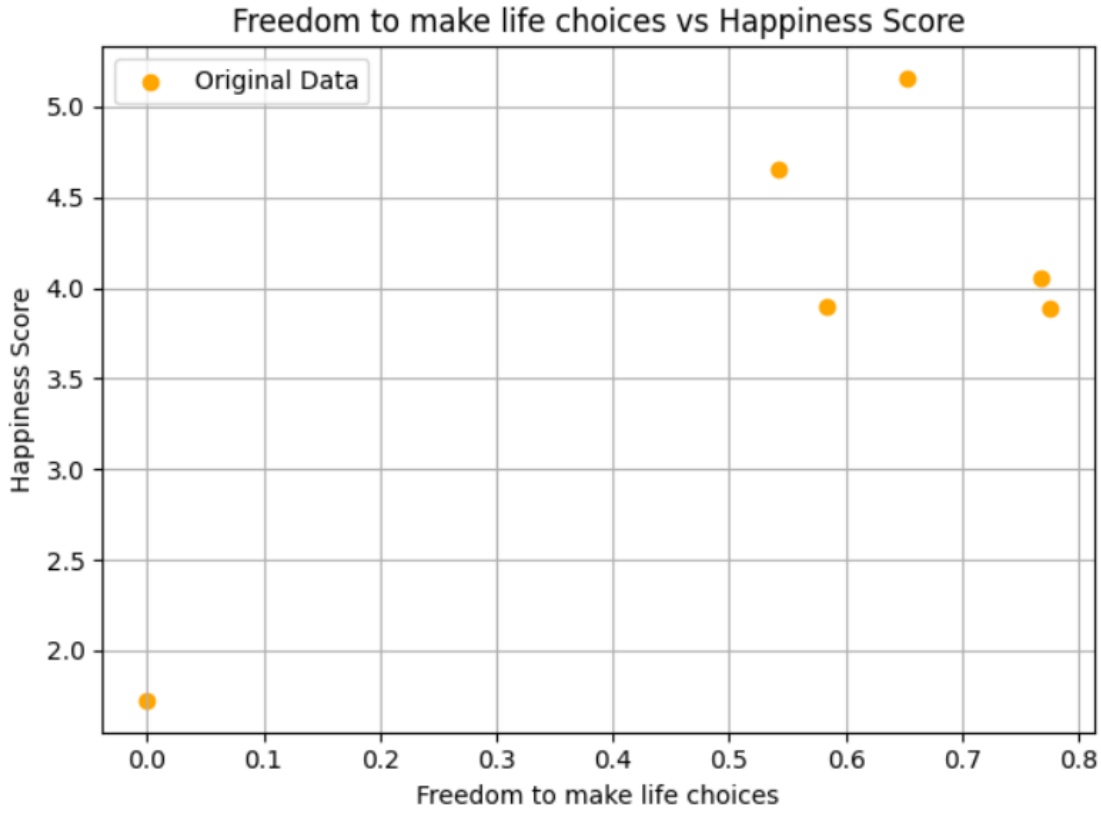


This plot shows that Afghanistan can clearly be classified as an outlier based on low Score and GDP per Capita. Such outliers are highly influential in reducing averages for regions and generate pronounced insights into economic and social disparities within South Asia.

Task - 4 - Exploring Trends Across Metrics:

The correlation of the two metrics (Freedom to make life choices and Generosity) along with the score to South Asian countries was computed. Below is the scatter plot with trendlines for the two metrics against the score:

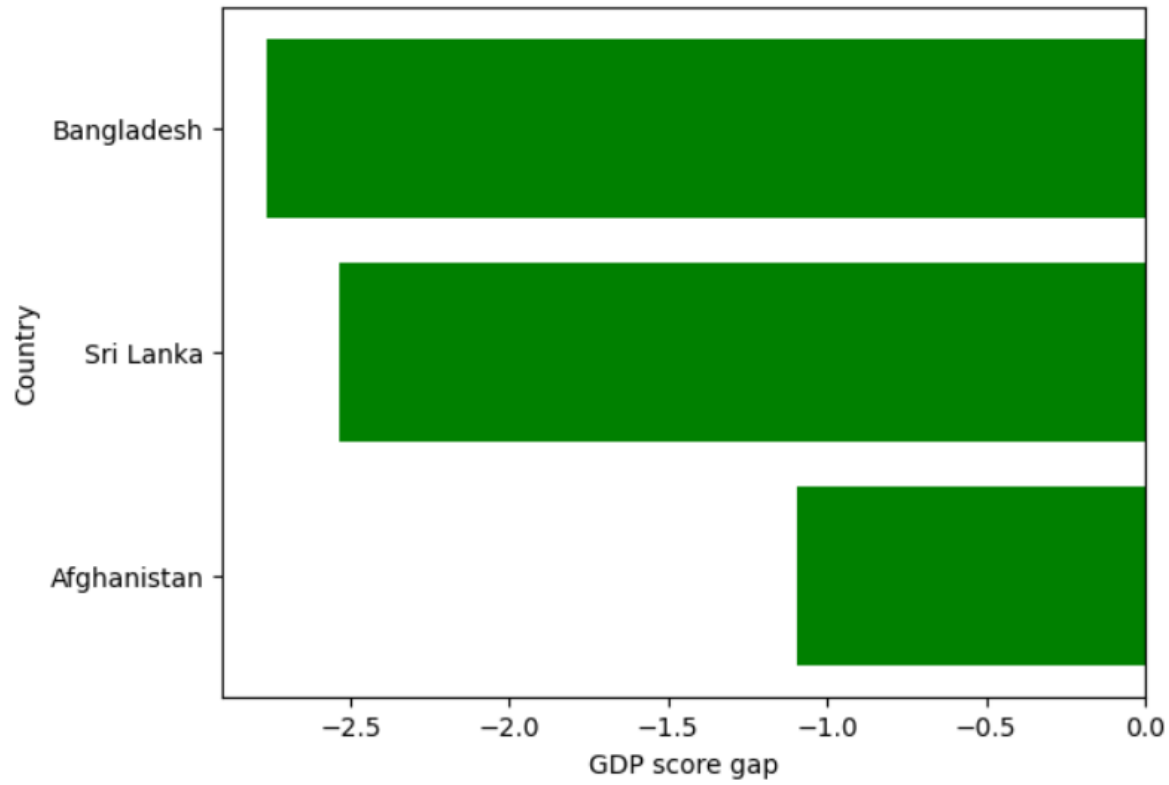
The scatter diagram presents a weak positive correlation score between generosity and happiness score. This means that as generosity increases, happiness score tends to increase slightly, but not significantly.



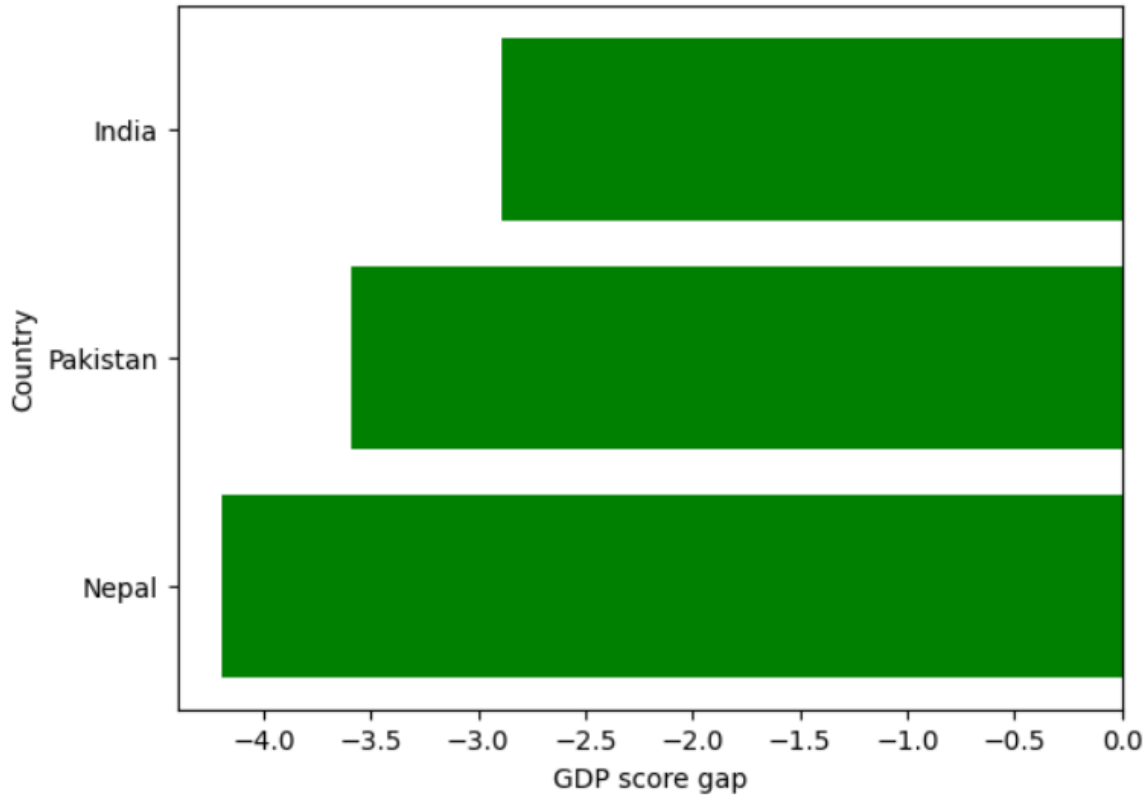
The scatter plot indicates that there is a weak positive correlation among happiness score and freedom of making life choices. This means that as freedom to make life choices increases, happiness score also tends to increase slightly.

Task- 5- Gap Analysis:

A new column GDP score gap was added. Then according to the gap south asian countries were ranked in both ascending and descending order. Similarly, top 3 countries with largest positive and negative gaps using a bar chart were plotted as follows:



Afghanistan has the largest positive gap, followed by Sri Lanka and Bangladesh.



Nepal has the largest negative GDP score gap, followed by Pakistan, and then India.

The GDP score gap likely reflects factors like income inequality, disparities in access to education and healthcare, and uneven economic development across different regions within the countries.

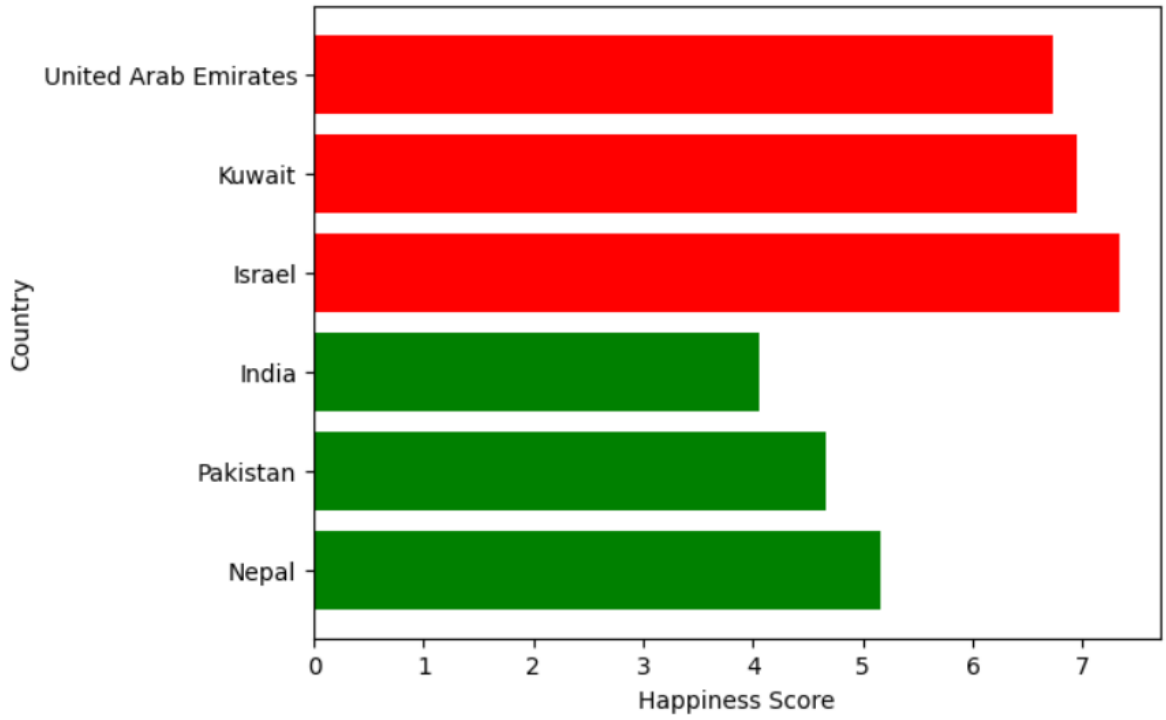
Problem- 3- Comparative Analysis:

1. Descriptive Statistics:

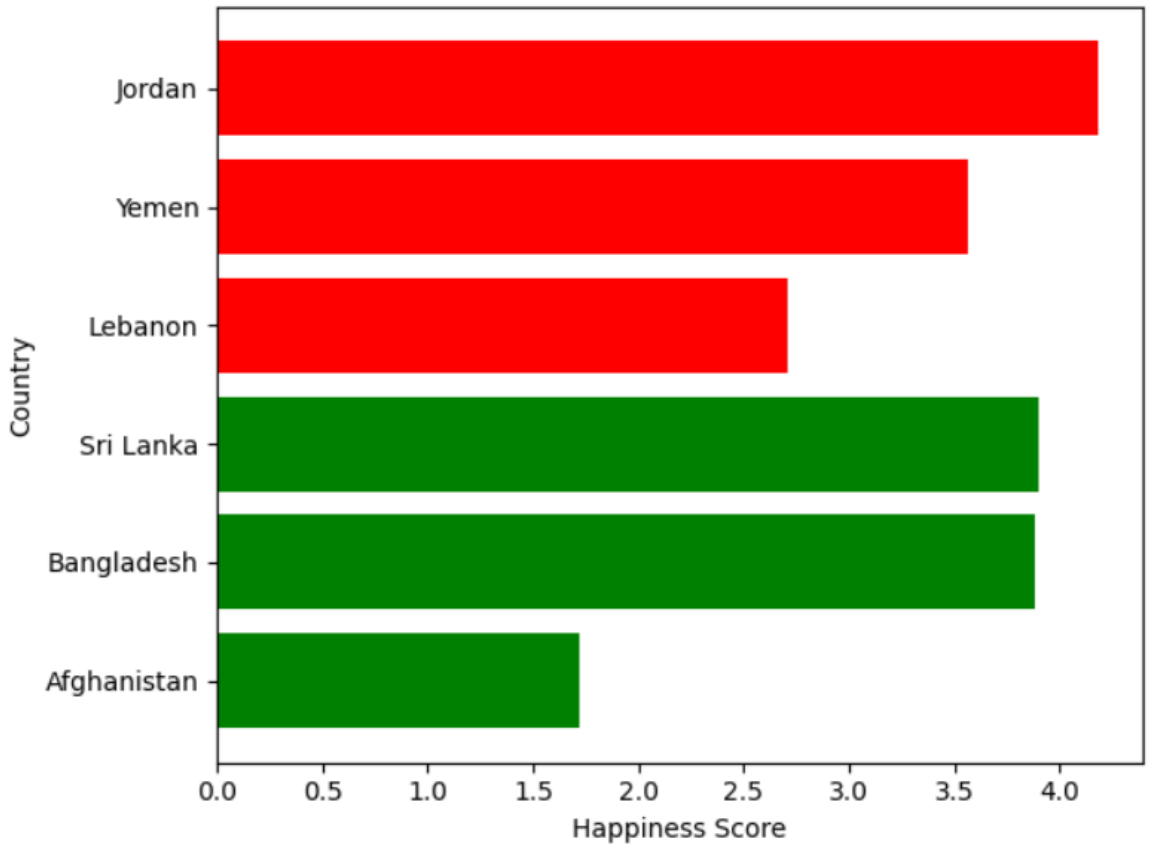
Mean, median and standard deviation of the happiness scores for both south asia and middle east were calculated. According to the data analysis, Middle East has higher happiness scores on average than South Asia.

2. Top and Bottom Performers:

Top 3 and bottom 3 countries in each region based on the score and then comparison was done between them making a bar graph as below:



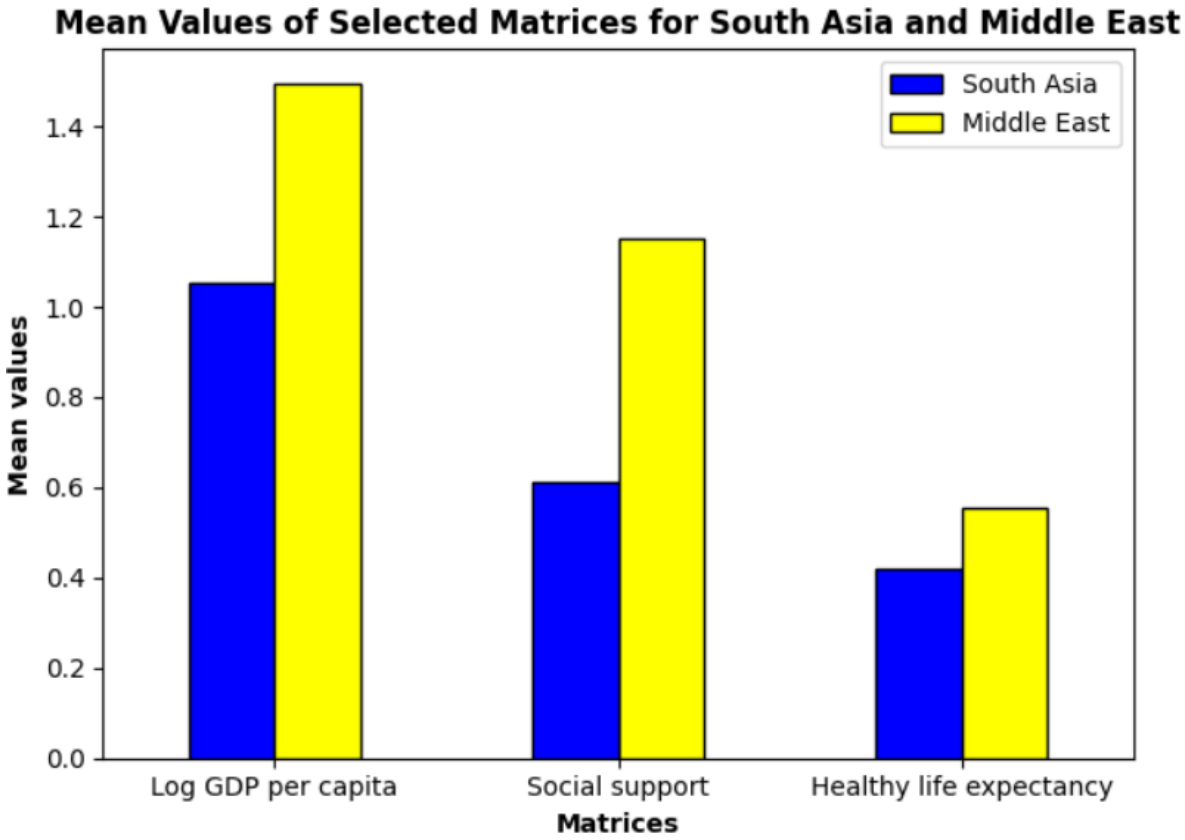
The graph presents happiness scores of the top three countries: three from the Middle East (UAE, Kuwait and Israel) and three from South Asia (India, Pakistan and Nepal). The Middle Eastern countries consistently had higher happiness scores than their South Asian counterparts.



Based on the bottom 3 data presented, South Asian countries tend to have slightly higher happiness scores compared to Middle Eastern countries.

3. Metric Comparisons:

Key metrics like GDP per Capita, Social Support, and Healthy Life Expectancy between the two regions using grouped bar charts as follows:



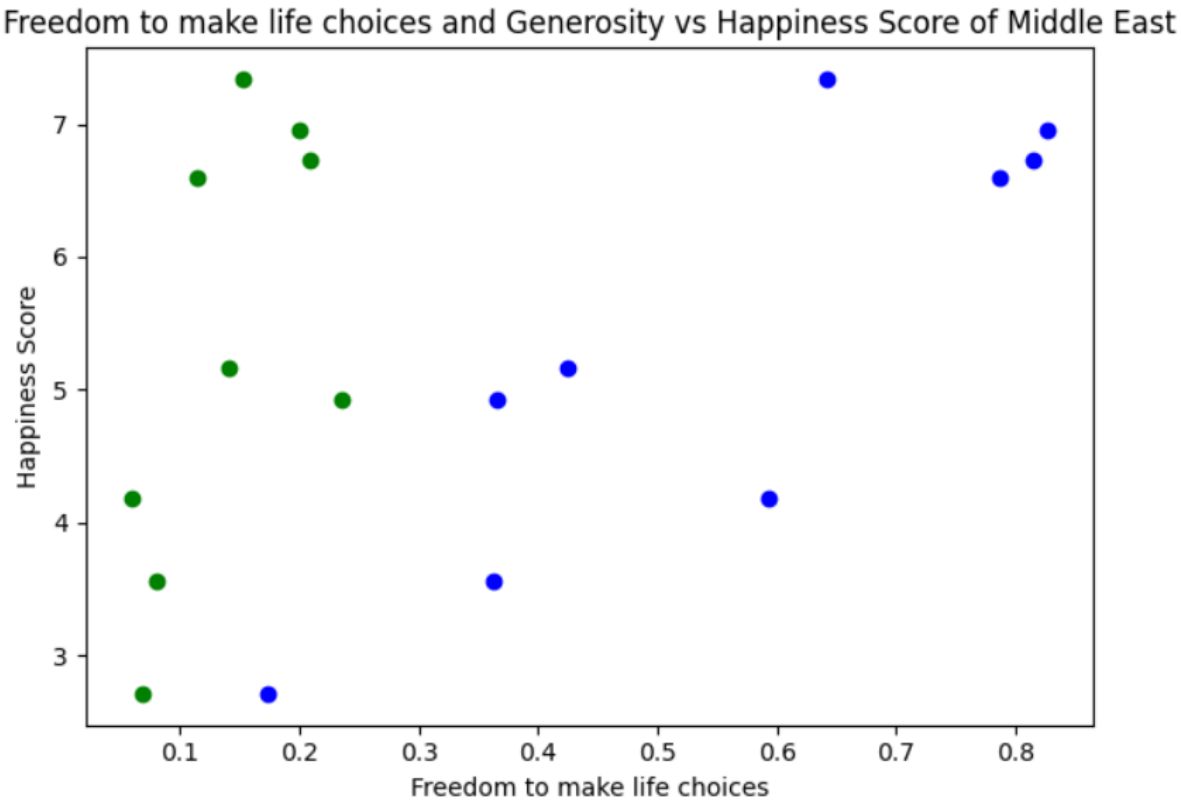
From the analysis of the bar chart, the largest disparity between South Asia and the Middle East is seen in the **Log GDP per capita** metric.

4. Happiness Disparity:

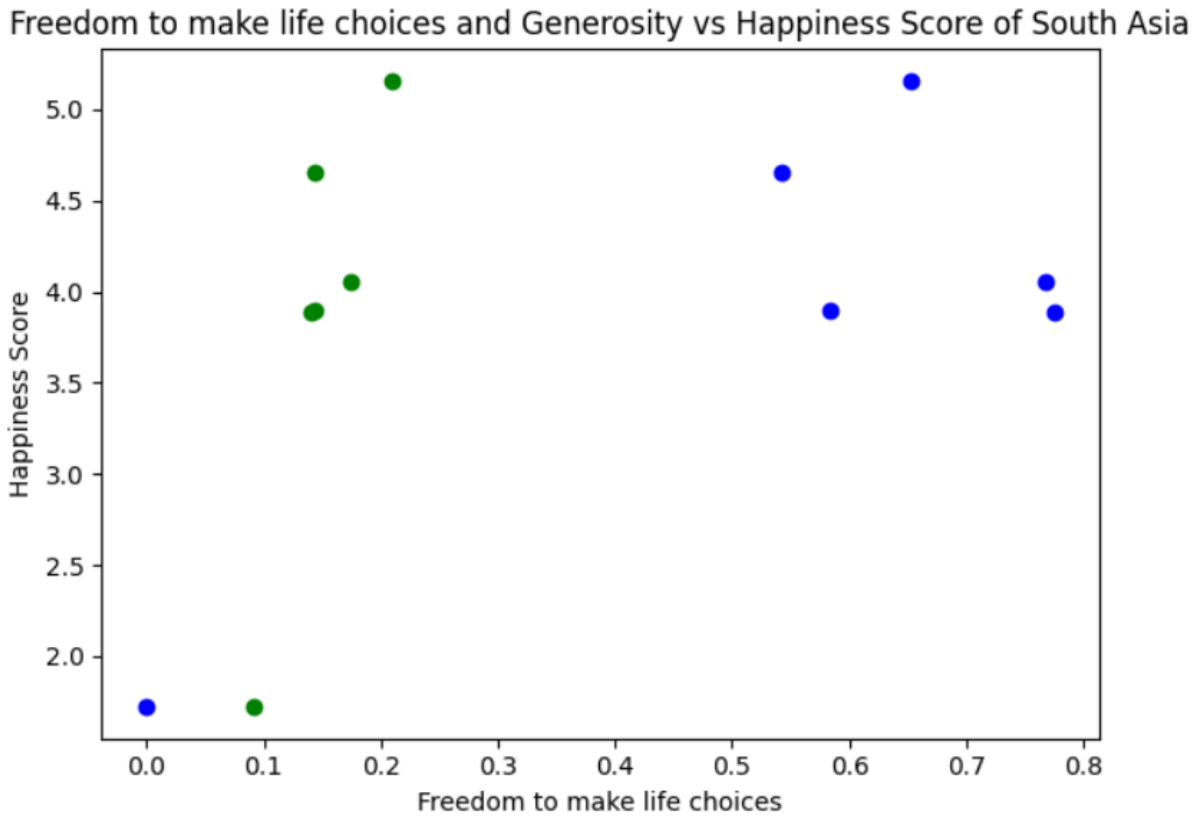
The range (max- min) and coefficient of variation (CV) for Score in both regions was computed. South Asian countries have greater variability than Middle East countries in happiness.

5. Correlation Analysis:

The correlation of Score with other metrics Freedom to Make Life Choices, and Generosity within each region was analysed and scatter plots to visualize and interpret the relationships was plotted as follows:



Countries ranking high on the scale of generosity in the Middle East have been observed to score low on freedom for life choice parameters. While countries with high freedom for life choices in the Middle East usually score low on the generosity scale, a negative relationship is established between the two in the region.



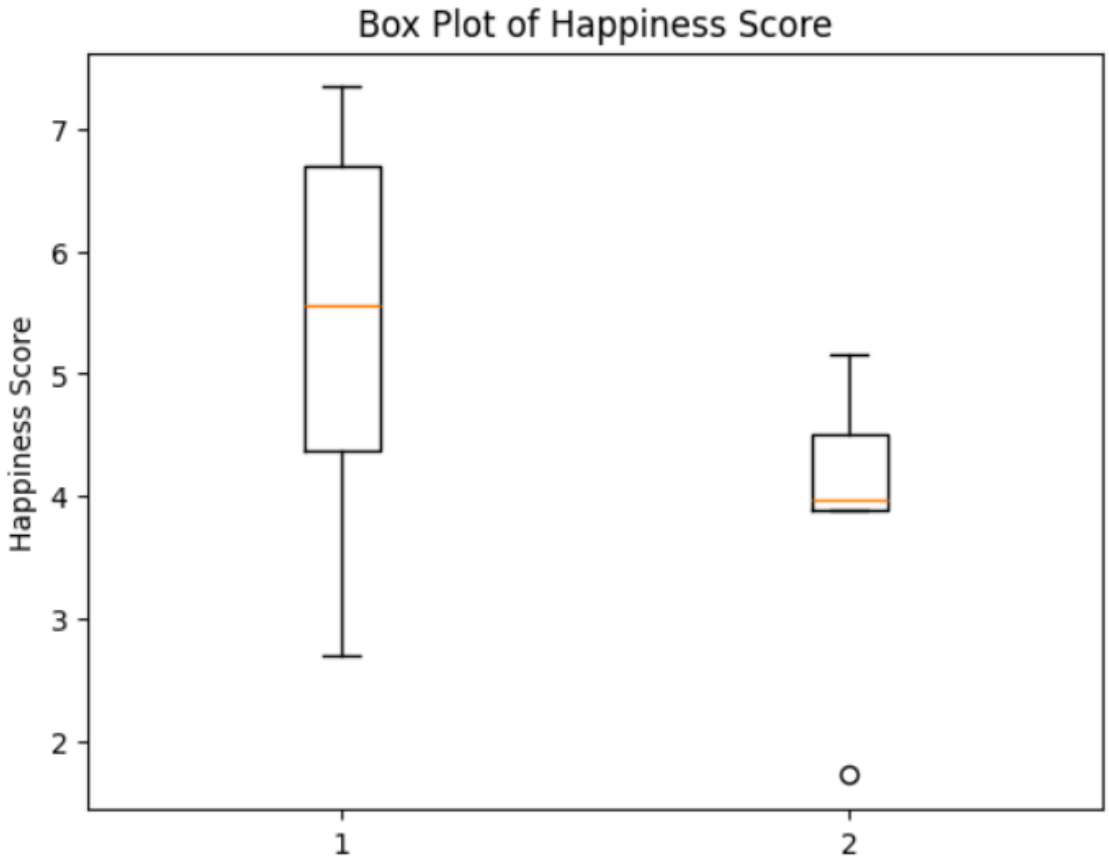
Generosity and Freedom to Make Life Choices are highly correlated positively with Happiness Score in South Asia. Countries scoring highly on any of the two dimensions also have high Happiness Scores.

6. Outlier Detection:

We identified the outliers in both regions based on their Happiness Score and GDP per capita. There is no outliers in middle east but there is one outlier in south asia which is Afghanistan.

7. Visualization:

We created a box plot to compare the distribution of happiness score between South Asian Countries and Middle Eastern Countries:



It's revealed that group one displays a higher median happiness score and a broader range than group two. An outlier in the second group reflects a much lower score on the happiness index.

GitHub Link:

https://github.com/supreeyakunwar/5CSO37-2024-Supriya/blob/main/2408484_Supriya_Kunwar.ipynb