Concepts and Technologies of AI

Assignment 1: Statistical Interpretation and Exploratory Data Analysis

Name: Kritish Dhakal

Group: L5CG20

Student ID: 2408573

Lecturer: Bibek Khanal

Tutor: Bibek Khanal

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

The World Happiness Report is an annual global survey which ranks countries based on their citizens' happiness, using factors like Gross Domestic Product (GDP) per capita, Healthy life expectancy, Social Support, Freedom to make life choices, Generosity of people, Perception of corruption amongst civilians and dystopia + residual. It is a report aimed at highlighting the wellbeing of countries each year for global enthusiasts as well as helping policy makers make better policies and prioritize people and their wellbeing. The report also sets a benchmark for global happiness and quality of life that all countries and policymakers should strive to reach.

This report aims to analyze the World Happiness Report dataset to uncover trends of general happiness of people across different regions of the world. This report will focus on understanding key factors around the world that influence the happiness of the general population. This report will specifically focus on South Asian countries and Middle Eastern Countries and aim to provide meaningful insights.

## Problem 1: Data Exploration, Understanding and Visualization.

In Problem 1, we first loaded the dataset and displayed its rows and columns. We then calculated the mean, median and standard deviation for the score column which represents the happiness score of the countries. Using this data, we found the happiest and the unhappiest countries. Then we filtered the dataset to only have countries that score higher than 7.5, then we sorted this data and displayed it. We then created a new column called Happiness Category and categorized countries based on their happiness score. Countries with a score lower than 4 were put in the low category, countries with a score between 4 and 6 were put in the medium category, and countries with above 6 score were put in the high category. Finally, we created a bar chart to show the 10 happiest countries which turned out as follows:
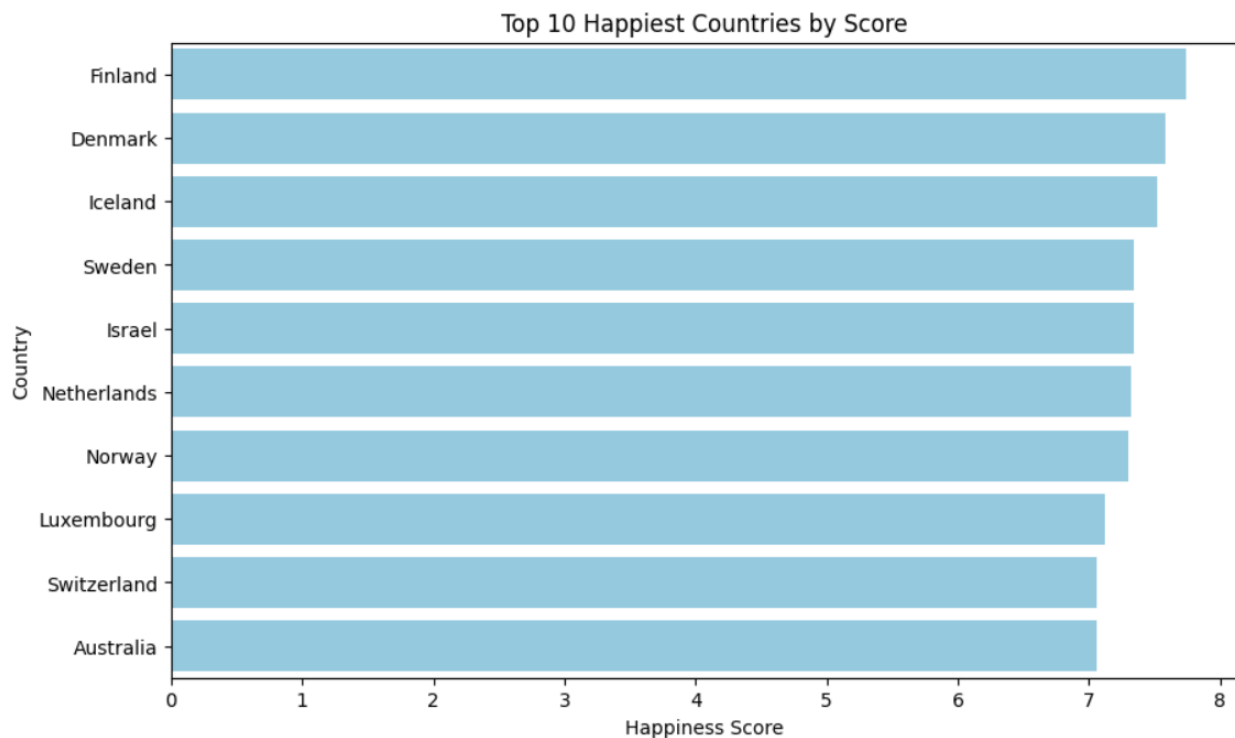
Then we created a line chart to show the 10 unhappiest countries.

Top 10 Unhappiest Countries by Score
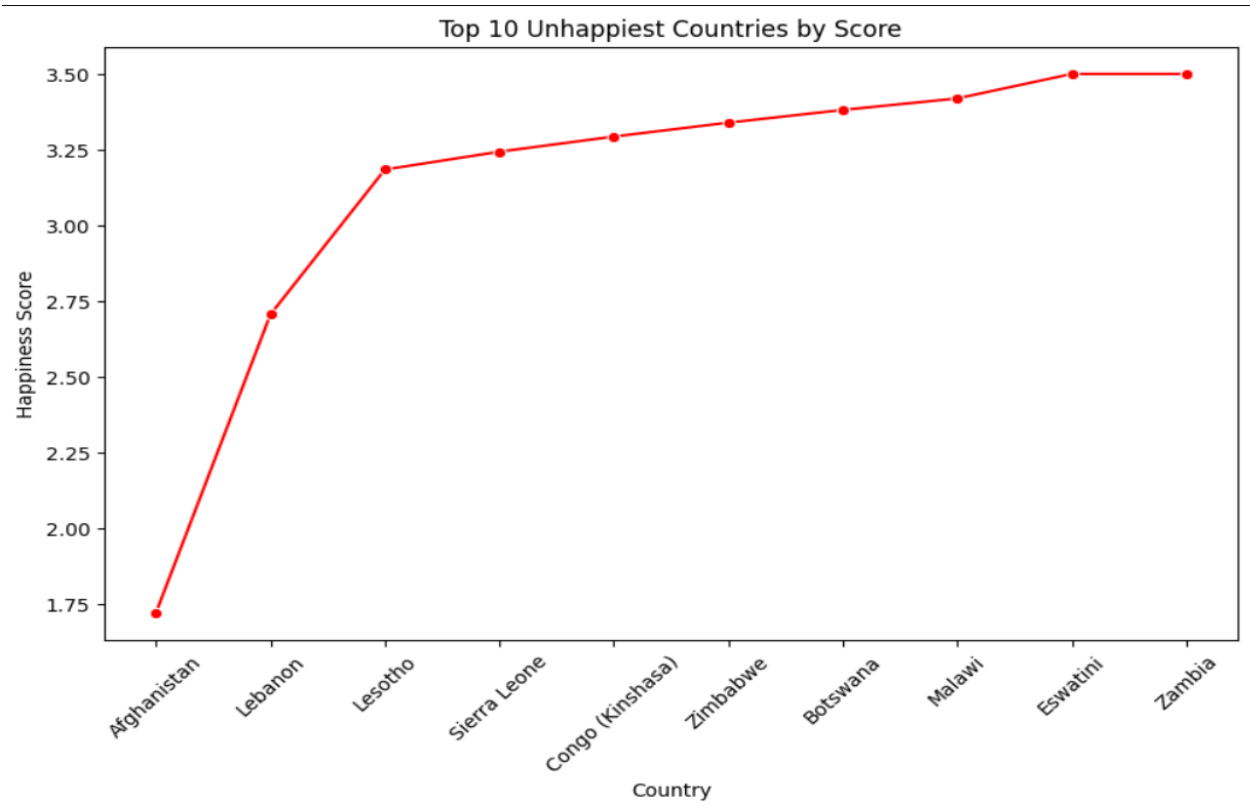
We also created a histogram to show the distribution of happiness scores.

Distribution of Happiness Scores
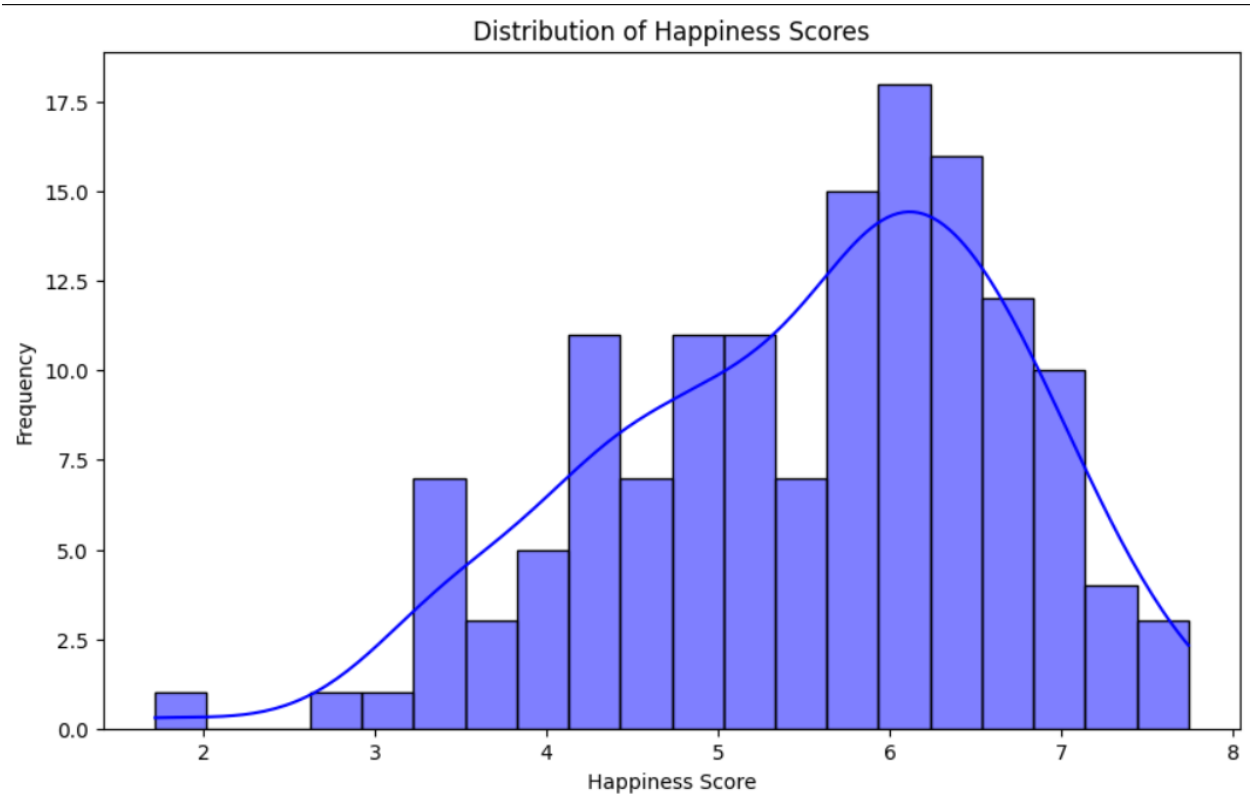
We then created a scatter plot to showcase the relationship between GDP per capita and Happiness scores.

Scatter Plot between GDP per Capita and Happiness Score
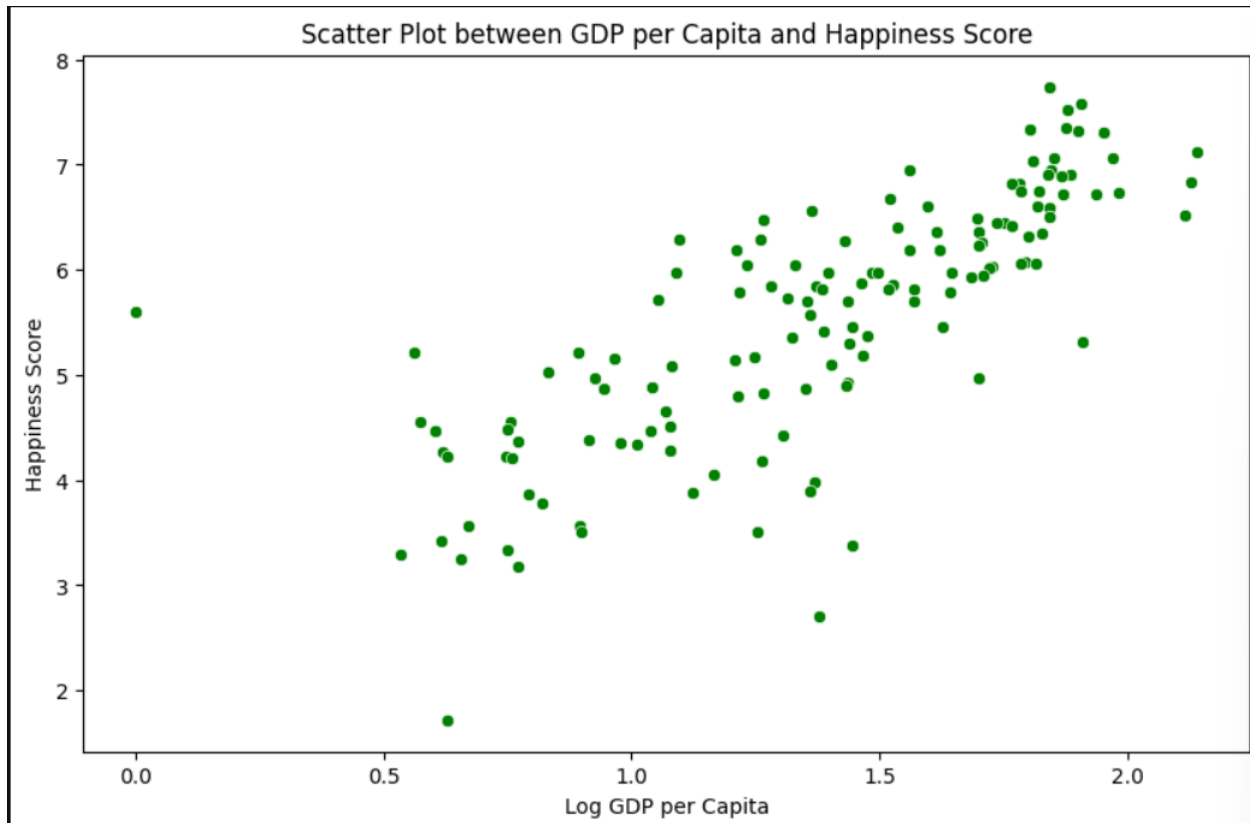
From our analysis, we draw the conclusion that higher GDP per capita correlated with higher happiness scores.

Problem 2: Advance Data Exploration of South Asian Countries

In problem 2, we first filtered South Asian countries from the dataset, then computed a composite score using GDP per capita, social support and healthy life expectancy columns from the dataset. We found that Sri Lanka, Nepal, India, Pakistan and Bangladesh were the Top 5 countries based on the composite scores.

We then ranked South Asian countries based on their composite score and then used the IQR method to detect outliers and highlighted them in a scatter plot:
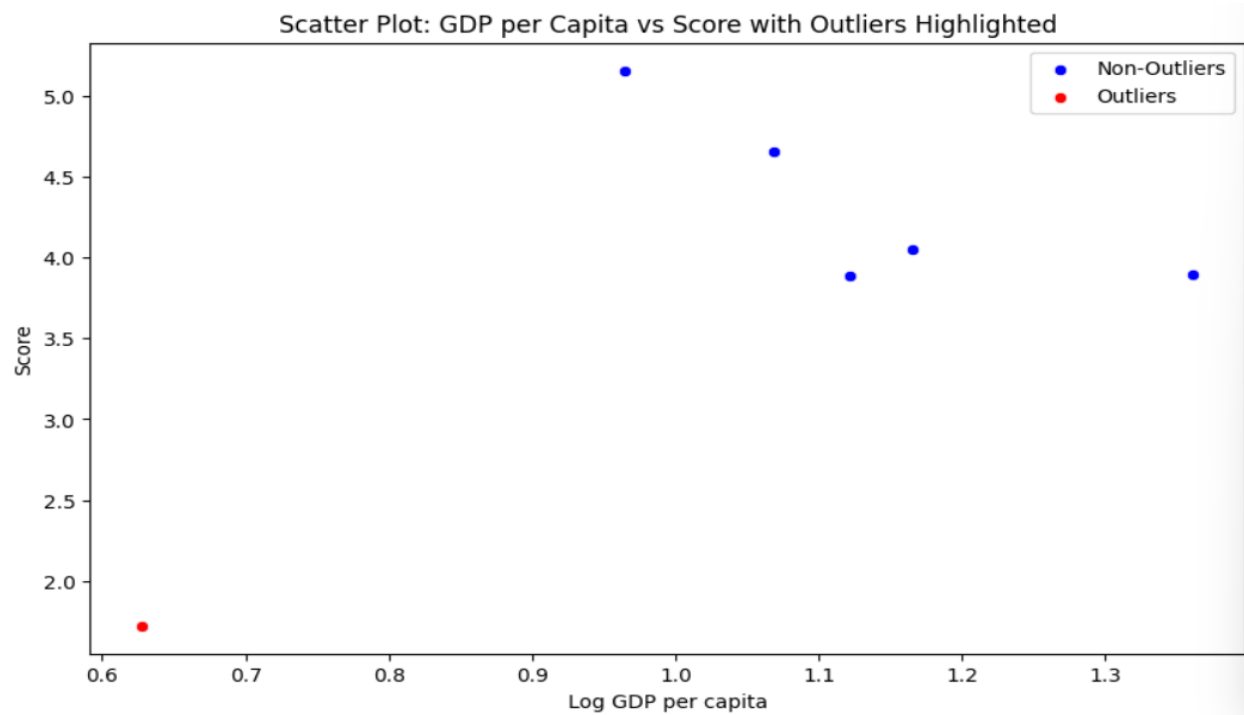
We then chose two metrics, Freedom to make life choices and Generosity and calculated their Pearson correlation with happiness score. This resulted in the following scatter plots:

Scatter Plot: Freedom to Make Life Choices vs Score

Scatter Plot: Generosity vs Score
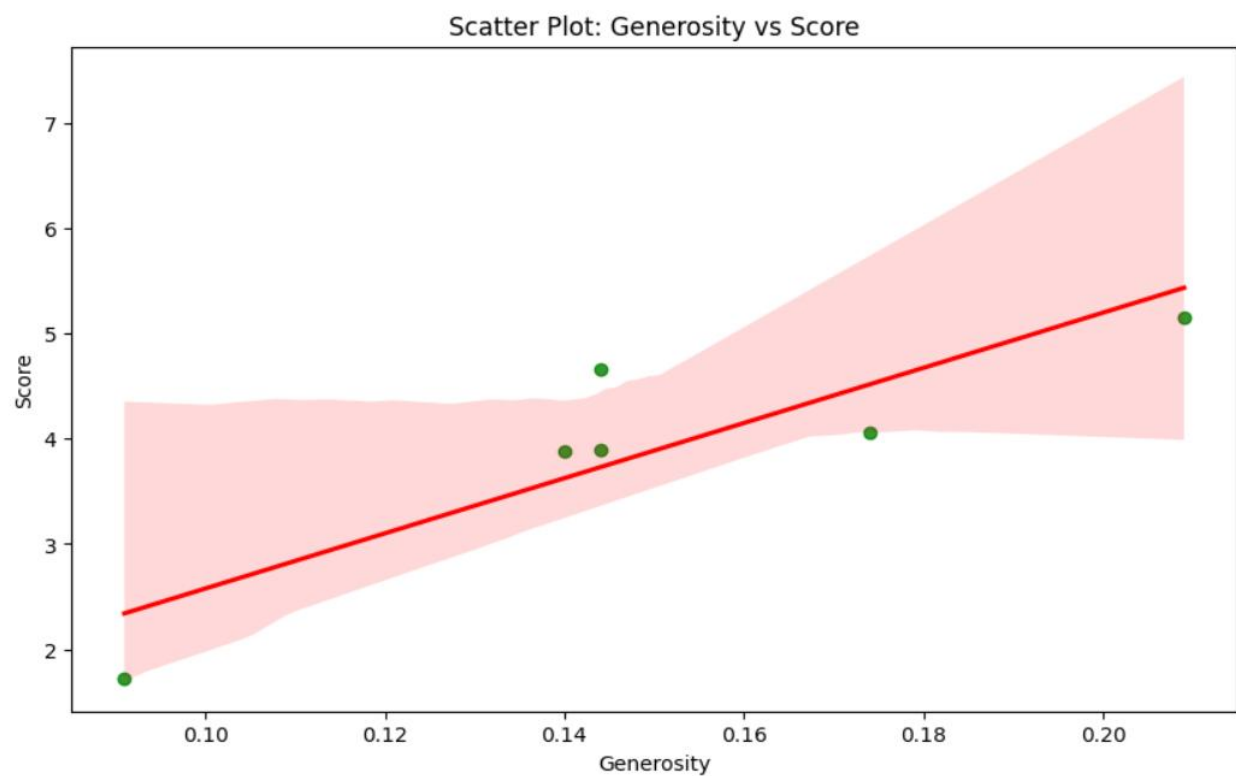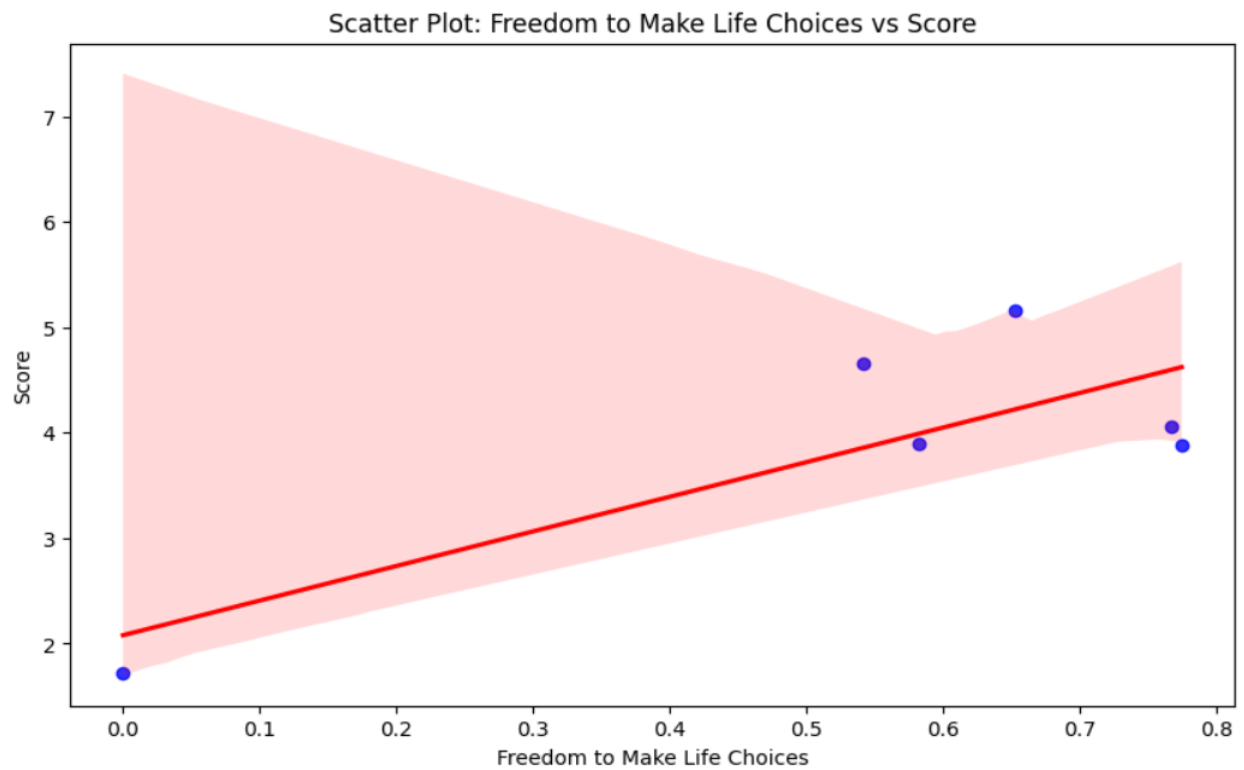
We then analyzed the gap which is difference between GDP per capita and the score for each South Asian country and ranked them based on this in both ascending and descending order. We found that the 3 countries with the largest positive gap were Afghanistan, Sri Lanka and Bangladesh and the 3 countries with the largest negative gap were Nepal, Pakistan and India.

From the analysis done in this problem, we found that rankings based on composite scores are very similar to the happiness scores of these countries. There were some outliers which had high GDP but low happiness which could be an indication of societal discrimination that sadly still exist in these countries.

Problem 3: South Asian Countries vs Middle Eastern Countries

In problem 3, we first filtered Middle Eastern countries and then we compared mean and standard deviation of South Asian countries and Middle Eastern countries and found that Middle Eastern countries had a higher average happiness score. We also identified the top 3 and bottom 3 countries of both regions and these were the result:

Top 3 South Asian Countries were Nepal, Pakistan and India.
Bottom 3 South Asian Countries were Afghanistan, Bangladesh and Sri Lanka.

Top 3 Middle Eastern Countries were Israel, Kuwait and United Arab Emirates.
Bottom 3 Middle Eastern Countries were Lebanon, Yemen and Jordan.

We also compared GDP per capita, Social Support and Healthy Life Expectancy of countries of both regions in a bar chart:
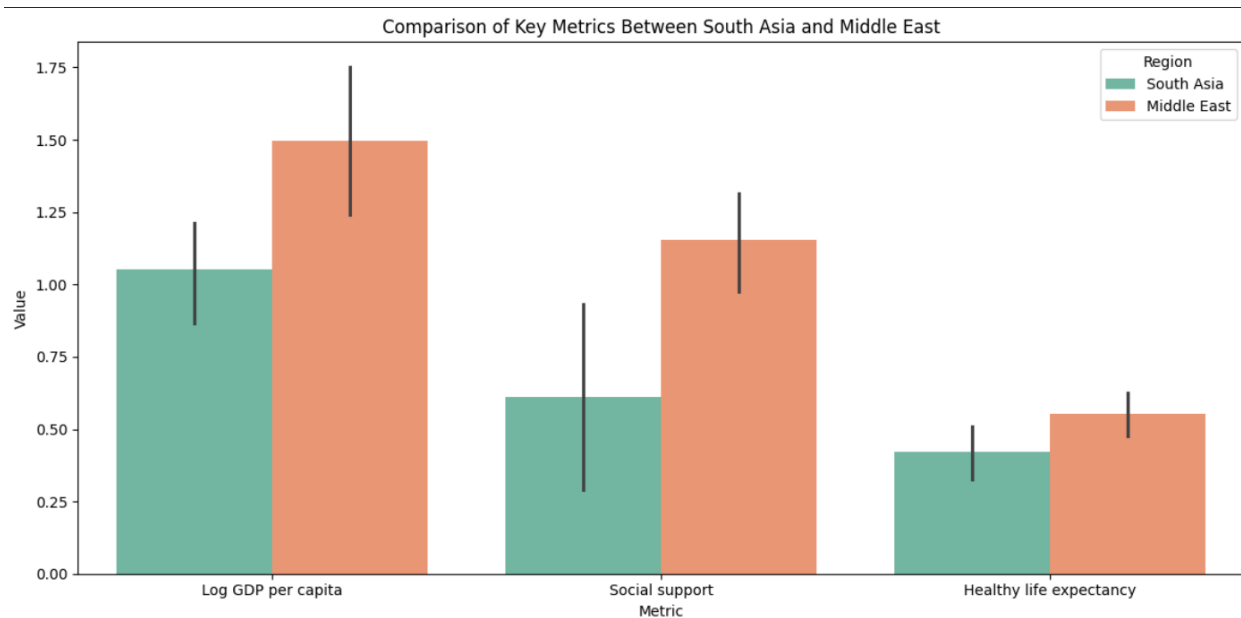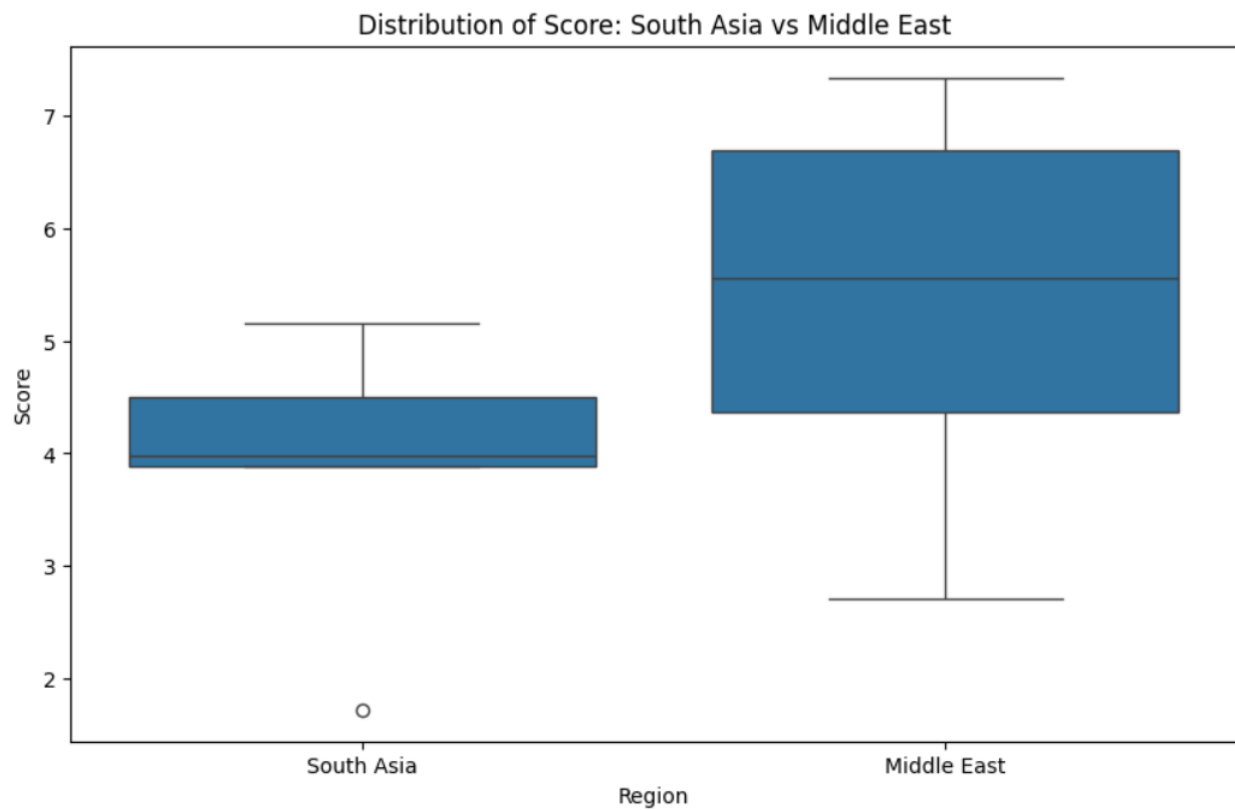

We can see that Middle Eastern Countries are higher than South Asian Countries in every metric.

We then compared the variability of happiness by finding out the range and coefficient of variation for happiness score in both regions and found that South Asia had a higher happiness variation than Middle Eastern.

We also analyzed the correlation of happiness score with Freedom to make life choices and Generosity and plot the results in scatter plots:
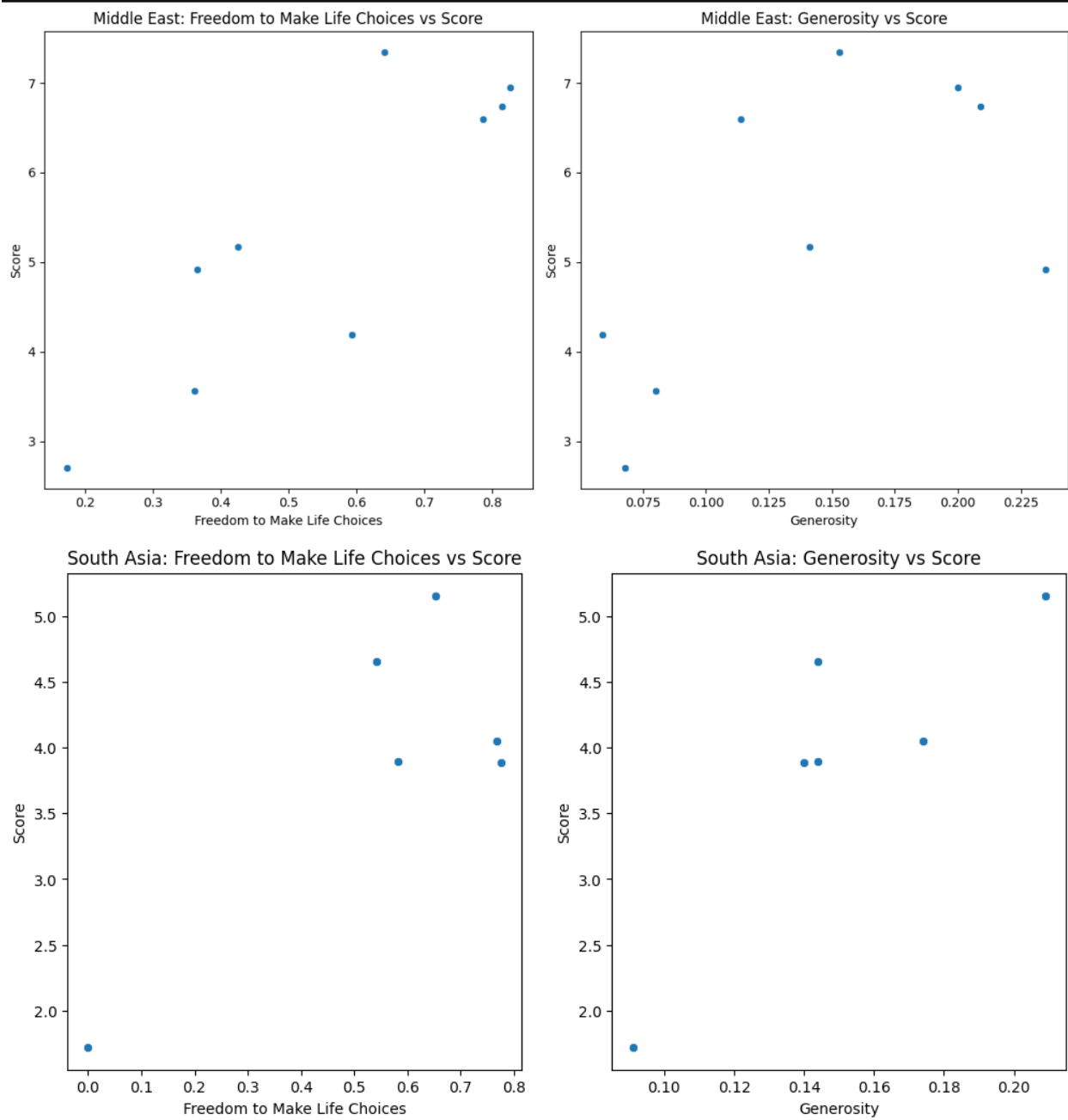
We also identified the outliers in both of these regions based on their Happiness Score and GDP per capita then we created a box plot to compare the distribution of happiness score between South Asian Countries and Middle Eastern Countries:

Distribution of Score: South Asia vs Middle East

From this problem we found out that Middle Eastern Countries are happier than South Asian Countries while South Asian Countries have a greater variance in happiness which means there are both countries with extremely high happiness scores and countries with extremely low happiness scores.

Conclusion:

In our report, we identified patterns in global happiness like the correlation of GDP per capita and happiness scores. When getting a closer look at South Asian region, we found that composite scores were very similar to the happiness scores of most of the countries in this region however the outliers which had high GDP but low happiness indicated that these areas could have societal discrimination. When comparing South Asian data with Middle Eastern data, we found that Middle East scores higher than South Asia in almost all metrics and has lower variance than South Asia in metrics.

Our analysis shows the multi-faceted nature of happiness in different regions of the world which are influenced by health and well-being, social and economic factors.