UNIVERSITY OF
WOLVERHAMPTON

HERALD COLLEGE
KATHMANDU

ing
GROUP

5CS037

Concepts and Technologies of AI

Assignment - I - Statistical Interpretation and Exploratory
Data Analysis.(Exploring South Asia and Middle East Perspectives.)

Name: Bibek Singh Bhandari

Student ID: 2409013

Lecturer:Siman Giri

Tutor:Bibek Khanal

Submitted on:19-12-2024

Acknowledgement

Introduction

An important international study that assesses happiness levels in various nations is the World Happiness Report. It offers thoughtful perspectives on the state of happiness around the world, helping individuals, scholars, and officials in better understanding the elements that affect well-being. Happiness scores and important factors including life expectancy, social support, economic production, independence, lack of corruption, and goodwill are included in the dataset used in this research.

In order to explore the World Happiness paper dataset, this analysis aims to explore the dataset through three primary objectives. It first studies the data to find trends and patterns in happiness around the world. The analysis of South Asian happiness levels and the factors influencing them comes next. In order to provide light on the regional and cultural differences in happiness, it concludes by comparing the happiness levels, measurements, and inequalities between South Asia and the Middle East. This research aims to offer a region-specific and comparative understanding of happiness through these investigations.

Objectives

The following are the primary objectives of this report:

1. To investigate and evaluate the World Happiness Report data in order to find important trends and patterns.

2. To investigate South Asian happiness levels and the variables affecting them.

3. To compare the Middle East and South Asia in terms of unusual values, related measurements, and happiness levels.

4. To understand and explain the geographic differences in happiness between these two areas.

Data Exploration and Understanding:

•Dataset Overview:

	Country name	score	Log GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Dystopia + residual
0	Finland	7.741	1.844	1.572	0.695	0.859	0.142	0.546	2.082
1	Denmark	7.583	1.908	1.520	0.699	0.823	0.204	0.548	1.881
2	Iceland	7.525	1.881	1.617	0.718	0.819	0.258	0.182	2.050
3	Sweden	7.344	1.878	1.501	0.724	0.838	0.221	0.524	1.658
4	Israel	7.341	1.803	1.513	0.740	0.641	0.153	0.193	2.298
5	Netherlands	7.319	1.901	1.462	0.706	0.725	0.247	0.372	1.906
6	Norway	7.302	1.952	1.517	0.704	0.835	0.224	0.484	1.586
7	Luxembourg	7.122	2.141	1.355	0.708	0.801	0.146	0.432	1.540
8	Switzerland	7.060	1.970	1.425	0.747	0.759	0.173	0.498	1.488
9	Australia	7.057	1.854	1.461	0.692	0.756	0.225	0.323	1.745

Basic Statistics:

1. Find the Score column's mean, median, and standard deviation. Mean 5.52758041958042

→ The score's median was 5.785, and its standard deviation was 1.1707165099442993.

2. Determine which nations have the highest and lowest happiness ratings.

→ As per the World Health Organization's(WHO) studies, Finland and Afghanistan have the highest and lowest happiness levels, accordingly.

3. Missing Values: The data sets show that there are no missing values.

4. Filtering and Sorting

→ Finland, Denmark, and Iceland scored more than 7.5
→ The top 10 nations by GDP per capita: These nations were also among the happiest.

According to GDP per capita data, the 10 wealthiest countries were also the happiest.
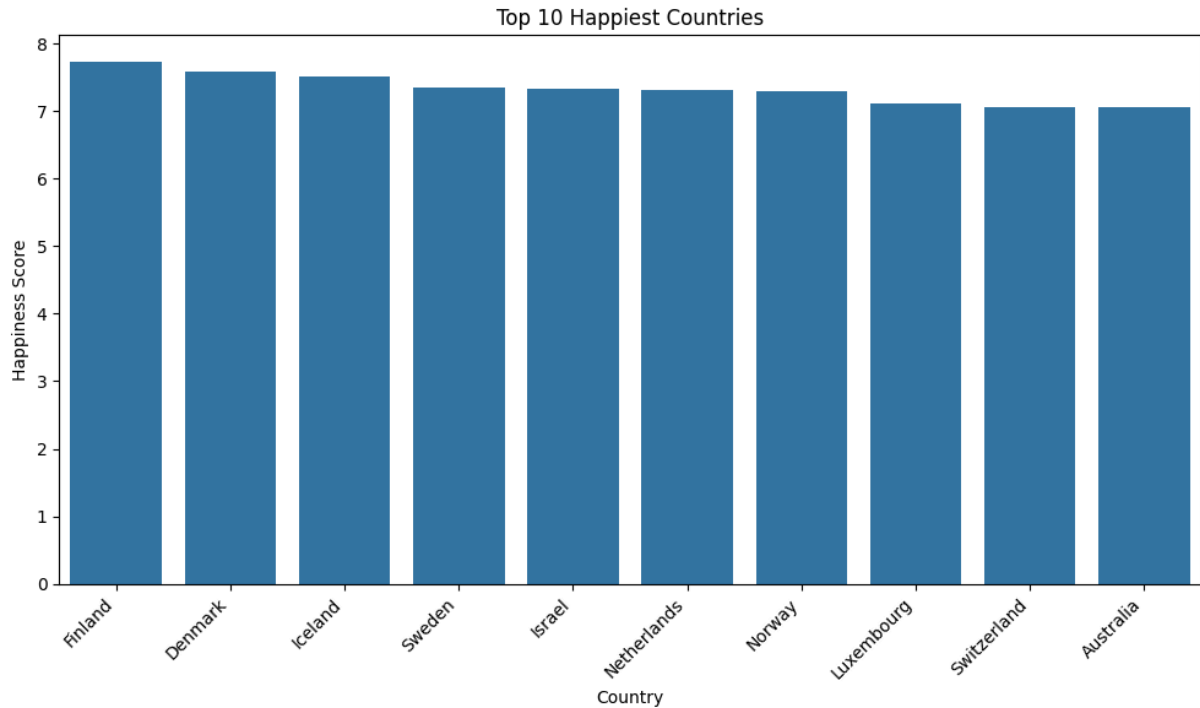
1.5. Including an Extra Column

A new column named Happiness Category was added in order to group countries based on their happiness scores:

→ Low: Under four
→ Score < 6 ≤ Moderate: 4

1.6 Data Visualizations

1. Bar Plot of Top 10 Happiest Countries -

Finland is at the top of this line plot, which shows the data of the top ten happiest nations. Denmark and Iceland are next in line. Here, you can see the image of the following graph as represented.
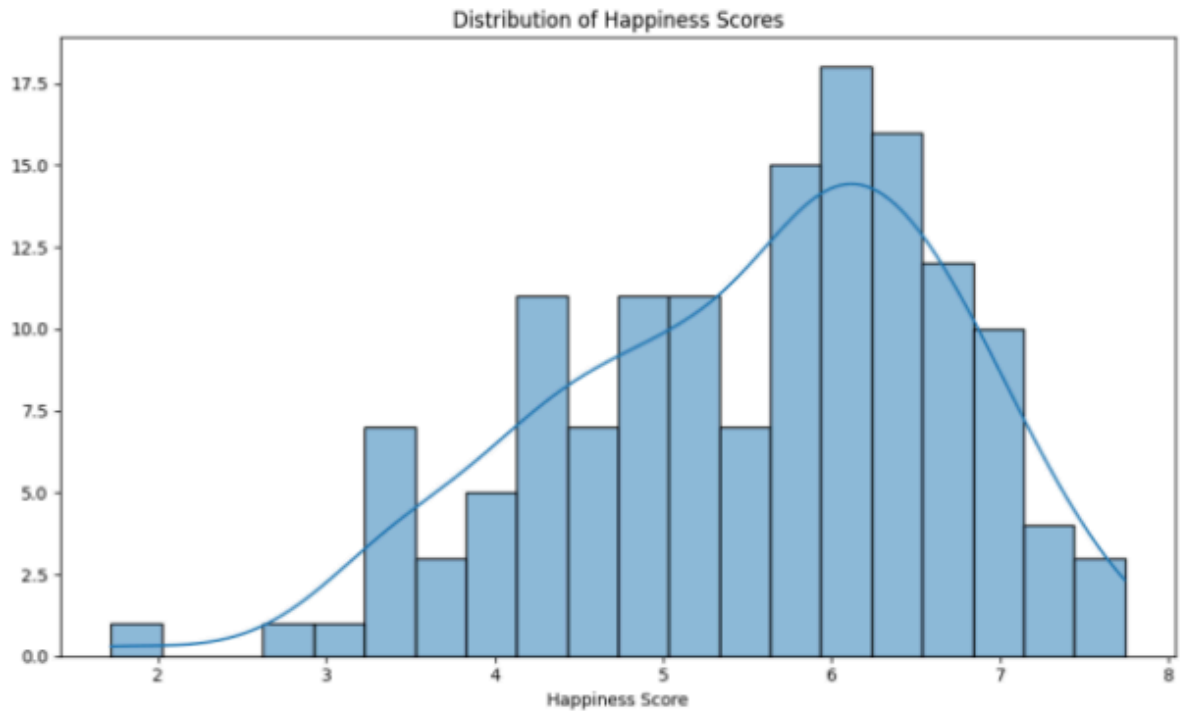
Top 10 Happiest Countries

Line Plot:

Using a line chart, these plots show the ten unhappiest nations by score, with Afghanistan at the bottom and Lebanon and Lesotho above it.Here you can see the graph as follows.
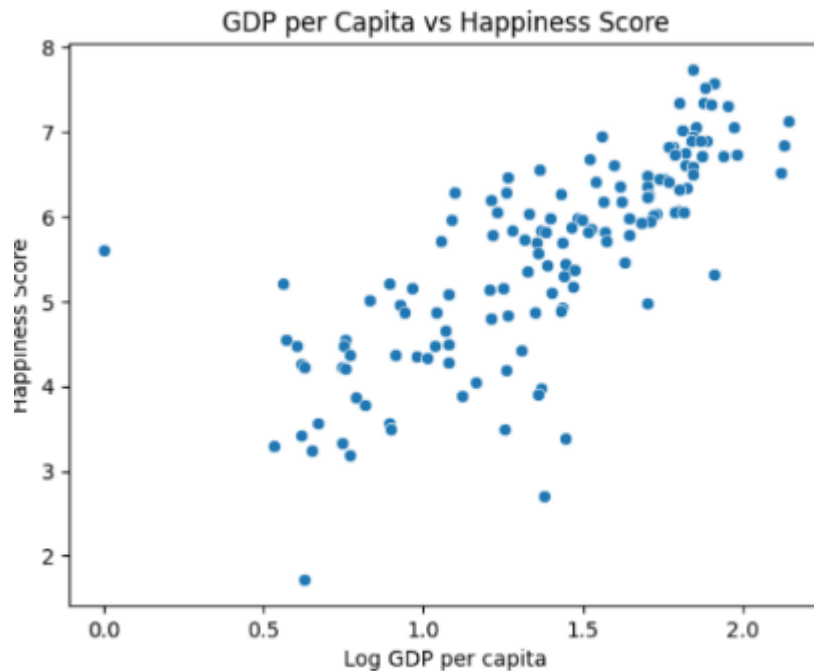
Unhappiest country:

138	Congo (Kinshasa)	3.295	0.534	0.665	0.262	0.473	0.189	0.072	1.102	Low
139	Sierra Leone	3.245	0.654	0.566	0.253	0.469	0.181	0.053	1.068	Low
140	Lesotho	3.186	0.771	0.851	0.000	0.523	0.082	0.085	0.875	Low
141	Lebanon	2.707	1.377	0.577	0.556	0.173	0.068	0.029	-0.073	Low
142	Afghanistan	1.721	0.628	0.000	0.242	0.000	0.091	0.088	0.672	Low

Distribution of Happiness Scores

Scatter Plot:

This Scatter plot shows the Happiness Score and GDP per capita.The graph is plot below:

GDP per Capita vs Happiness Score

Problem 2-Some Advance Data Exploration Assignment:

Task 1: Getting the South Asian Dataset Ready
A new CSV file named south_asia_data.csv containing the following countries was saved: Nepal, India, Sri Lanka, Pakistan, Afghanistan, Bhutan, Bangladesh, and Maldives.
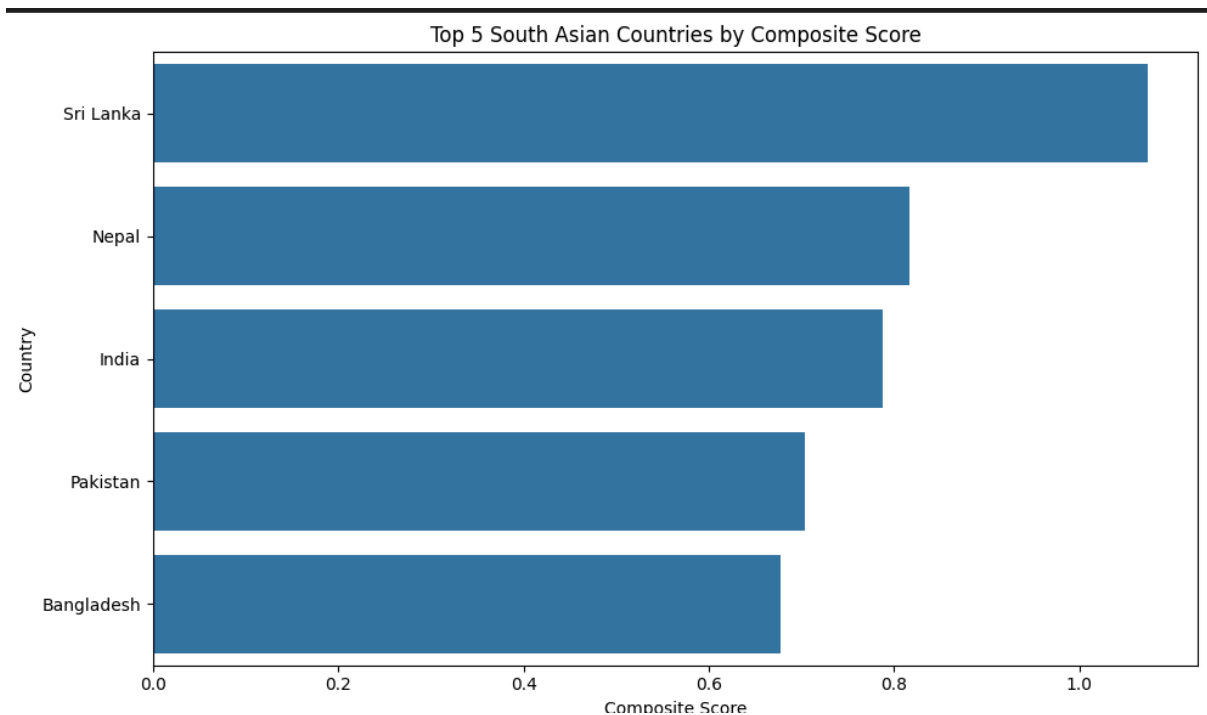
Task 2: Using the formula below, a new column called composite score was formed.
→ **0.40 × GDP per capita plus 0.30 × social support equals the composite score.**
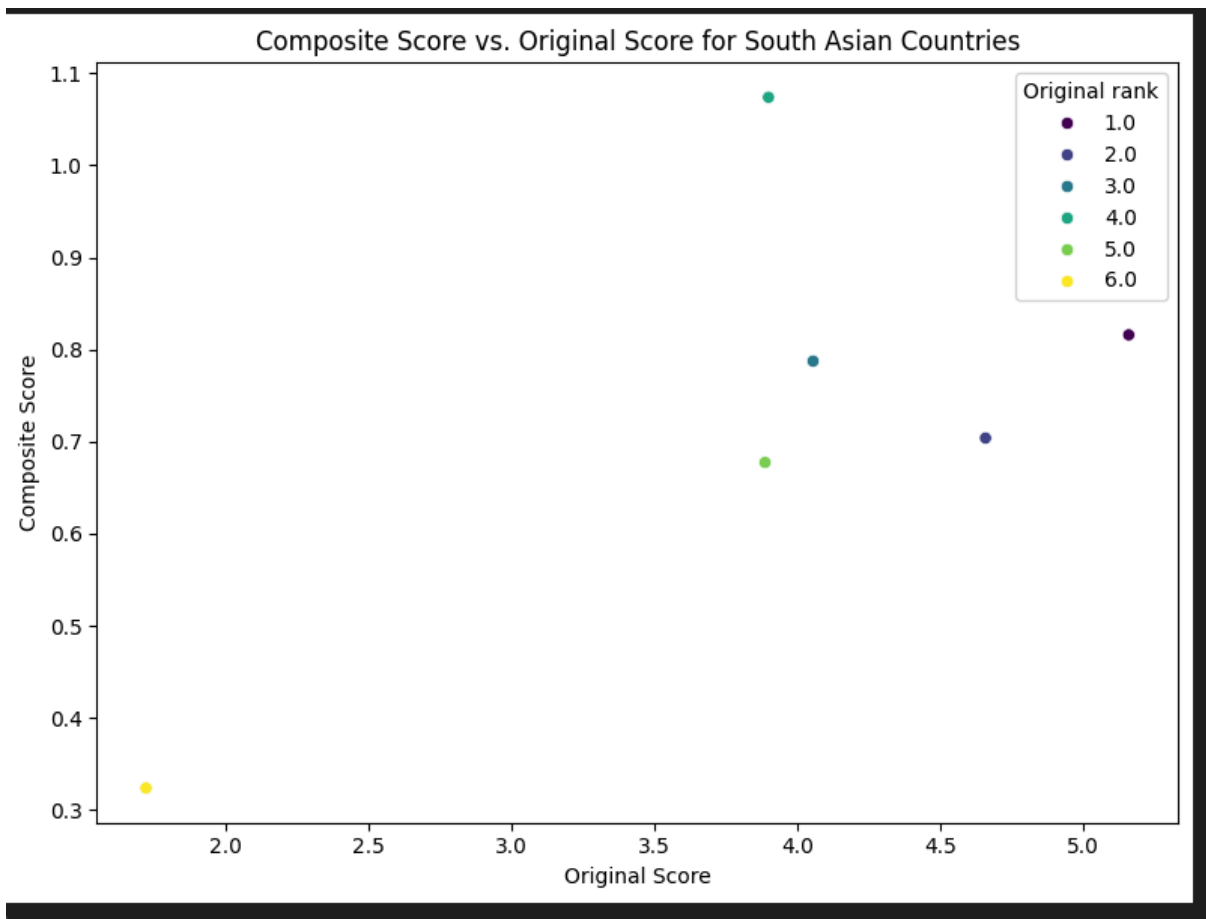**+ 0.30 × Life Expectancy at Good Health**

2. The nations were arranged in descending order based on their composite scores.

	Country name	Composite Score
127	Sri Lanka	1.0739
92	Nepal	0.8159
125	India	0.7874
107	Pakistan	0.7039
128	Bangladesh	0.6774
142	Afghanistan	0.3238

3. It shows five countries are visualized using horizontal graph showing the composite score.

Top 5 South Asian Countries by Composite Score

4.As you can see according to the graph we can visualized that the composite score and original score are not aligning with each other.

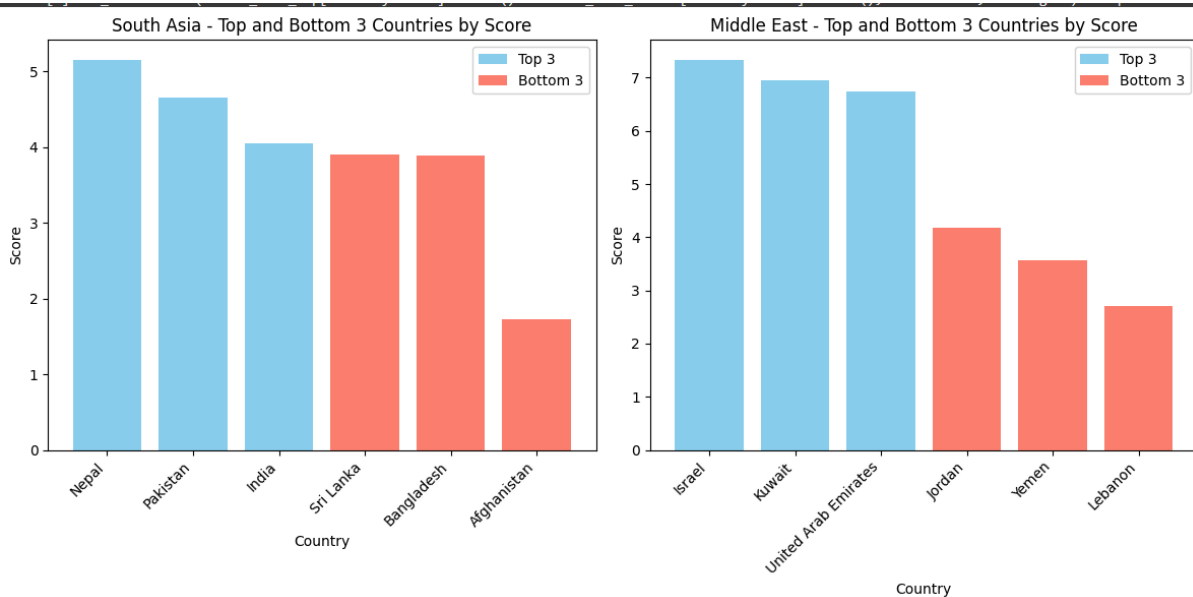Composite Score vs. Original Score for South Asian Countries

Task 3

When it comes to both score and GDP per capita, Afghanistan stands out among South Asian countries.Outliers are data points that fall outside of the range indicated by

[Q1−1.5×IQR, Q3+1.5×IQR], and IQR=Q3−Q1IQR = Q3 - Q1IQR=Q3−Q1. [Q1 minus 1.5 times the IQR, Q3 plus 1.5 times the IQR] Q3 + 1.5×IQR, Q1 − 1.5×IQR. Any number that is less than Q1−1.5×IQRQ1 - 1.5 \times IQRQ1−1.5×IQR or larger than Q3+1.5×IQRQ3 + 1.5 \times IQRQ3+1.5×IQR is considered to be an outlier. Outliers can be caused by measurement errors, rare events, or truly extraordinary performances in specific circumstances. Data points that differ from the majority of the dataset and are either clearly higher or lower than the interquartile range (IQR) are known as outliers.
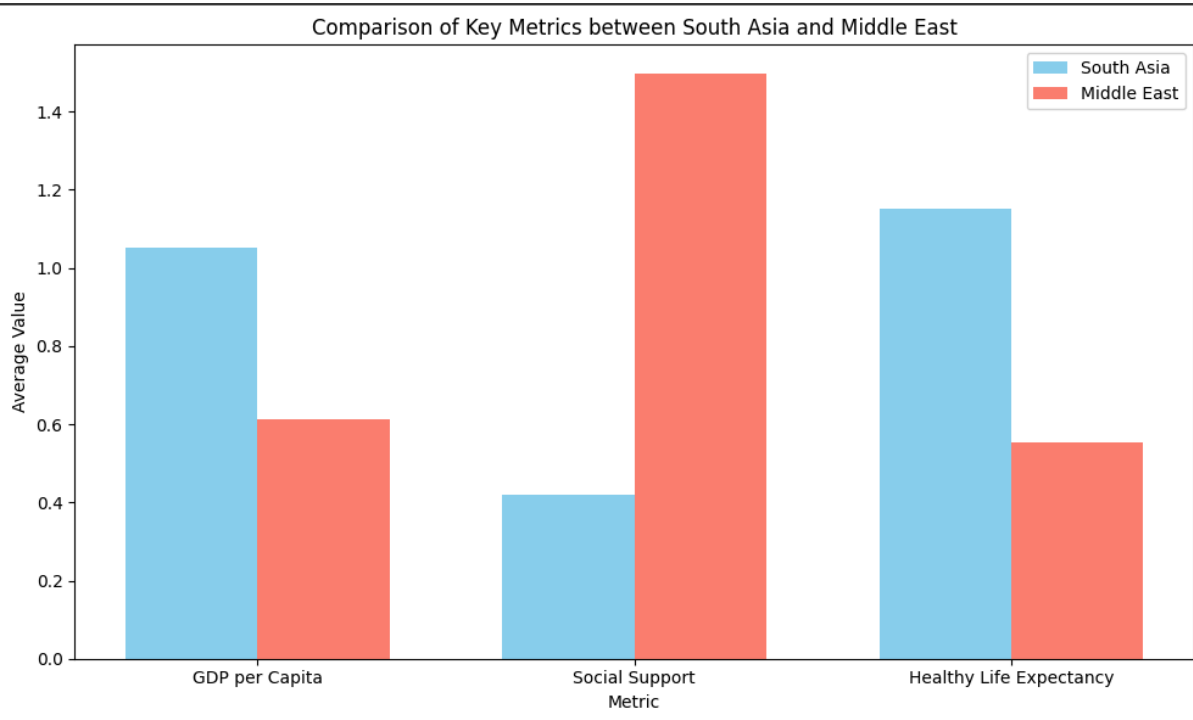
```
South Asia - Mean Score: 3.895666666666667, Standard Deviation: 1.1770690152521504
Middle East - Mean Score: 5.412100000000001, Standard Deviation: 1.5662011684327144

Middle East has higher happiness scores on average.
```

2. The graph below compares the countries in the Middle East and Southeast. In South Asia, the top three nations are Israel, Nepal, Pakistan, and India. The top three lowest-ranked nations are Sri Lanka, Kuwait, and the United Arab Emirates in the Middle East. South Asia's Bangladesh, Afghanistan, and Jordan As you can see the plotted graph shows Yemen and Lebanon in the Middle East.
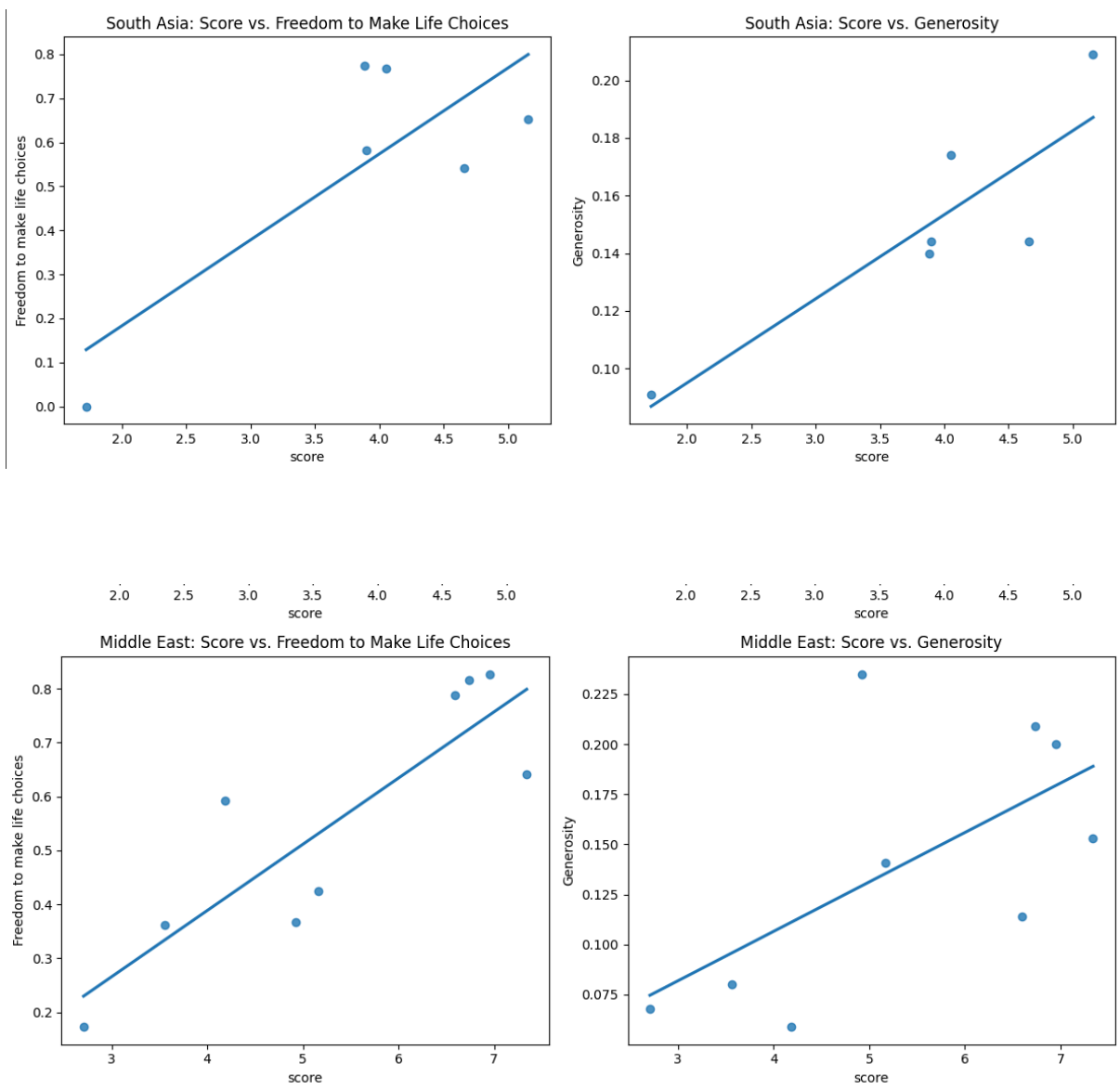
South Asia - Top and Bottom 3 Countries by Score

Middle East - Top and Bottom 3 Countries by Score

3. The graph that follows compares the GDP per capita, social support metric, and health life expectancy of Middle Eastern and South Asian nations.With an average value of 0.4 in South Asia and 1.4 in the Middle East, the peer support statistic shows the stark differences between the two regions.
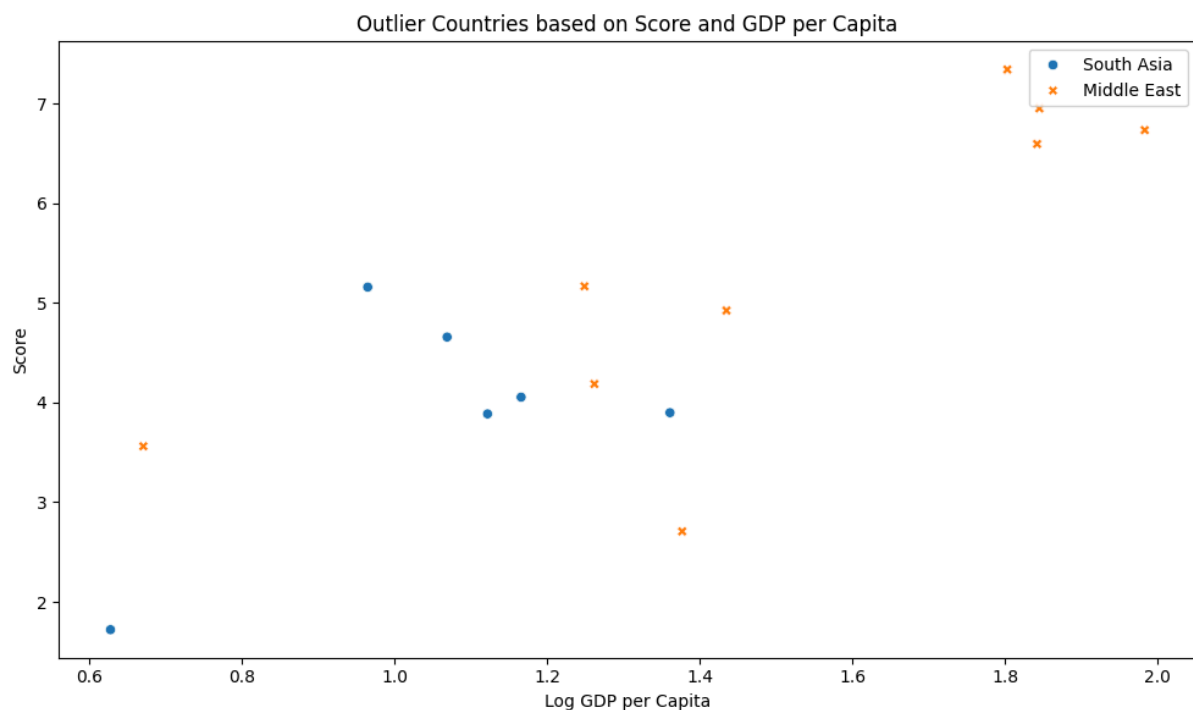

Comparison of Key Metrics between South Asia and Middle East

4. South Asia: Coefficient of Variation: **30.21%**,
  Range: 3.4370000000000003
  Middle East: Coefficient of Variation: 28.94%,
  Range: 4.634 Happiness is more variable in South Asia.

5. A scatter plot is made to show and analyze the associations when the score is correlated with other metrics, such as Generosity and Freedom to Make Life Choices, within each region.

University of Wolverhampton

6. Outlier nations in both areas are identified based on GDP per capita and score. Outlier nations frequently have unique features that affect their GDP per capita and happiness ratings. It is important to carefully analyze regional statistics since these outliers have a chance to distort regional averages and comparisons.
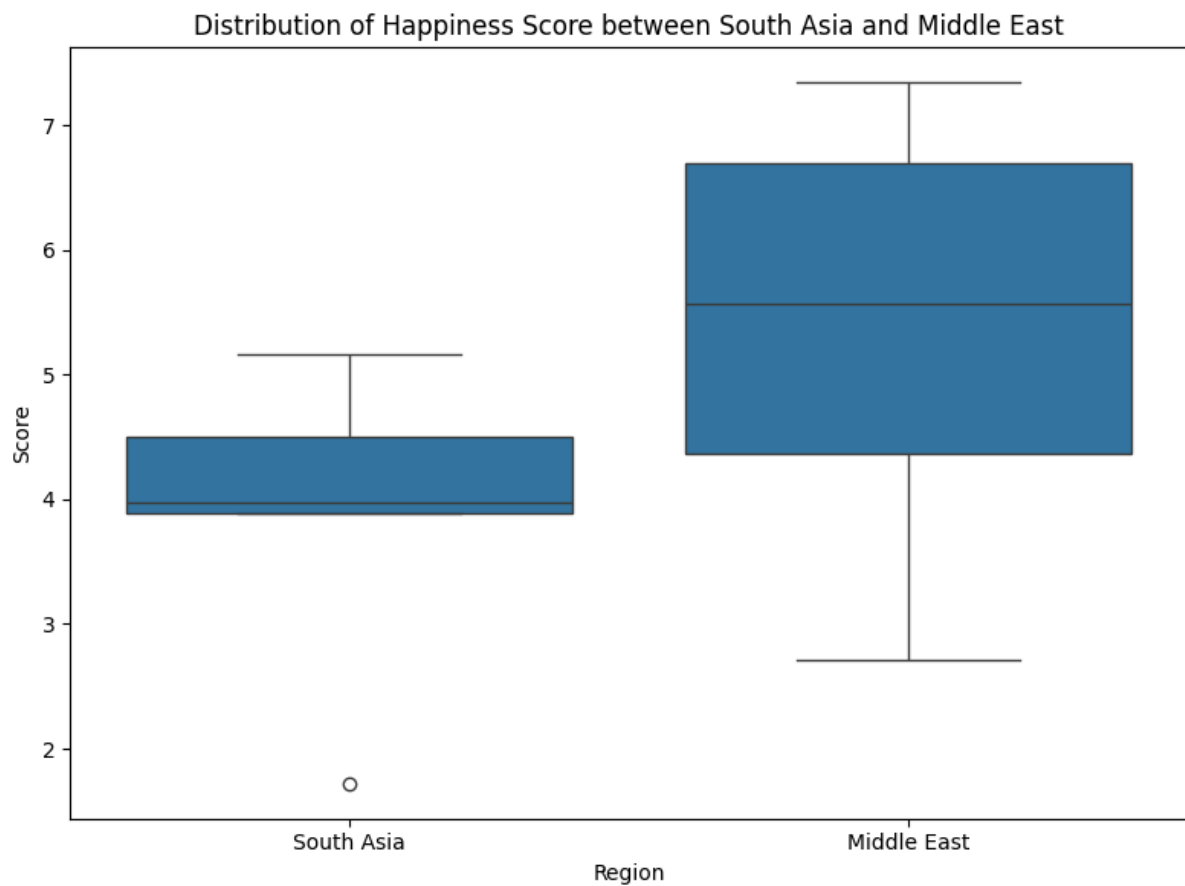

Outlier Countries based on Score and GDP per Capita

7. Boxplots are made to compare how scores are distributed in the Middle East and South Asia.

Distribution Shapes: The general form of each region's happiness score distribution is shown by the boxplots.

Medians: The median happiness score is shown by the line inside each box. A greater median suggests that people in that area are generally happier.

Outliers: Outlier countries with very high or low happiness scores are shown by points outside the boxplots' borders.

Distribution of Happiness Score between South Asia and Middle East

Conclusion:

In conclusion, this study compares South Asia and the Middle East through a study of the World Happiness Report. Compared to the Middle East, South Asia has more variation in happiness and lower averages for GDP per capita and social support. Happiness is most strongly positively correlated with generosity, but freedom to make life decisions is substantially less strongly related. Negative GDP-Happiness gaps show resilience and strong social ties, while positive gaps show inequality and materialism. Regional averages are greatly impacted by outliers such as Lebanon (Middle East) and Afghanistan (South Asia).In both areas, social support and access to healthcare are important factors that influence happiness. Improving happiness levels requires lowering inequality and promoting social harmony.Officials can develop ways to enhance well-being and lessen inequities with the help of an understanding of these trends. Additional research could examine the political or social factors of happiness.