5CS037

Concept And Technology

of AI

Assignment 1

Name: Ritesh Chaudhary

Group: L5CG21

Student ID: 2438464

Table of Content

Page No:

Introduction

The World Happiness Report presents a thorough international assessment that measures nations based on the well-being levels expressed by their citizens. The overall well-being score is influenced by several factors, including economic productivity, social support, life expectancy, individual freedom, governmental integrity, and benevolent behavior.

This report includes three primary aims:

- To examine the overarching patterns in the World Happiness Report information.
- To analyze the well-being measurements of nations in South Asia and Middle East, identifying key components affecting their positions.
- To compare and contrast the well-being trajectories between South Asia and the Middle East to understand regional distinctions.

Problem 1: Data Exploration and Understanding

Overview of the Activity

Here, we do some preliminary exploration of the World Happiness Report dataset: import the dataset, have a glimpse into its structure, perform basic statistics, check for missing values, clean and organize data, and plot the key variables.

1.1 Dataset Overview:

We first imported the dataset and viewed the first 10 rows. There are 158 rows and 12 columns in the dataset, each representing different factors related to happiness. The columns include: 'Country', 'Score', 'GDP per Capita', 'Social Support', 'Healthy Life Expectancy', and so on and so forth.

	Country name	score	Log GDP per capita	Social support	\
0	Finland	7.741	1.844	1.572	
1	Denmark	7.583	1.908	1.520	
2	Iceland	7.525	1.881	1.617	
3	Sweden	7.344	1.878	1.501	
4	Israel	7.341	1.803	1.513	
5	Netherlands	7.319	1.901	1.462	
6	Norway	7.302	1.952	1.517	
7	Luxembourg	7.122	2.141	1.355	
8	Switzerland	7.060	1.970	1.425	
9	Australia	7.057	1.854	1.461	

	Healthy life expectancy	Freedom to make life choices	Generosity	\
0	0.695	0.859	0.142	
1	0.699	0.823	0.204	
2	0.718	0.819	0.258	
3	0.724	0.838	0.221	
4	0.740	0.641	0.153	
5	0.706	0.725	0.247	
6	0.704	0.835	0.224	
7	0.708	0.801	0.146	
8	0.747	0.759	0.173	
9	0.692	0.756	0.225	

	Perceptions of corruption	Dystopia + residual
0	0.546	2.082
1	0.548	1.881
2	0.182	2.050
3	0.524	1.658
4	0.193	2.298
5	0.372	1.906

1.2 Basic Statistics:

The mean, median, and standard deviation values of the happiness score were calculated. We found which countries have the highest and lowest happiness scores, having Finland at the top and Afghanistan at the lowest.

Code cell output actions

```
[ ]  #Calculate the mean, median, and standard deviation for the Score column.
     mean_scores = data['score'].mean()
     median_score = data['score'].median()
     std_dev_score = data['score'].std()

     print(f'Mean score: {mean_scores}')
     print(f'Median score: {median_score}')
     print(f'Standard deviation of score: {std_dev_score}')
```

```
Mean score: 5.52758041958042
Median score: 5.785
Standard deviation of score: 1.1707165099442995
```

1.3 Missing Value

We checked for incomplete values and did not find any significant problem in this data set.

```
# Missing Values:
# 1. Check if there are any missing values in the dataset. If so, display the total count for each column.

missing_values = data.isnull().sum()
print("Missing values in each column:")
print(missing_values)
```

```
Missing values in each column:
Country name                      0
score                             0
Log GDP per capita                3
Social support                    3
Healthy life expectancy           3
Freedom to make life choices      3
Generosity                        3
Perceptions of corruption         3
Dystopia + residual               3
dtype: int64
```

1.4 Filtering & Sorting

The countries having a score greater than 7.5 were filtered out. We then sorted that cleaned data by GDP per Capita in descending order in order to find out what countries were the richest with the highest scores for happiness.

```
# Filtering and Sorting:
# 1. Filter the dataset to show only the countries with a Score greater than 7.5.

filter = data[data['score']>7.5]
print(filter)
```

```
  Country name  score  Log GDP per capita  Social support  \
0      Finland  7.741               1.844           1.572
1      Denmark  7.583               1.908           1.520
2      Iceland  7.525               1.881           1.617

   Healthy life expectancy  Freedom to make life choices  Generosity  \
0                    0.695                         0.859       0.142
1                    0.699                         0.823       0.204
2                    0.718                         0.819       0.258

   Perceptions of corruption  Dystopia + residual
0                      0.546                2.082
1                      0.548                1.881
2                      0.182                2.050
```

1.5 Adding New Column

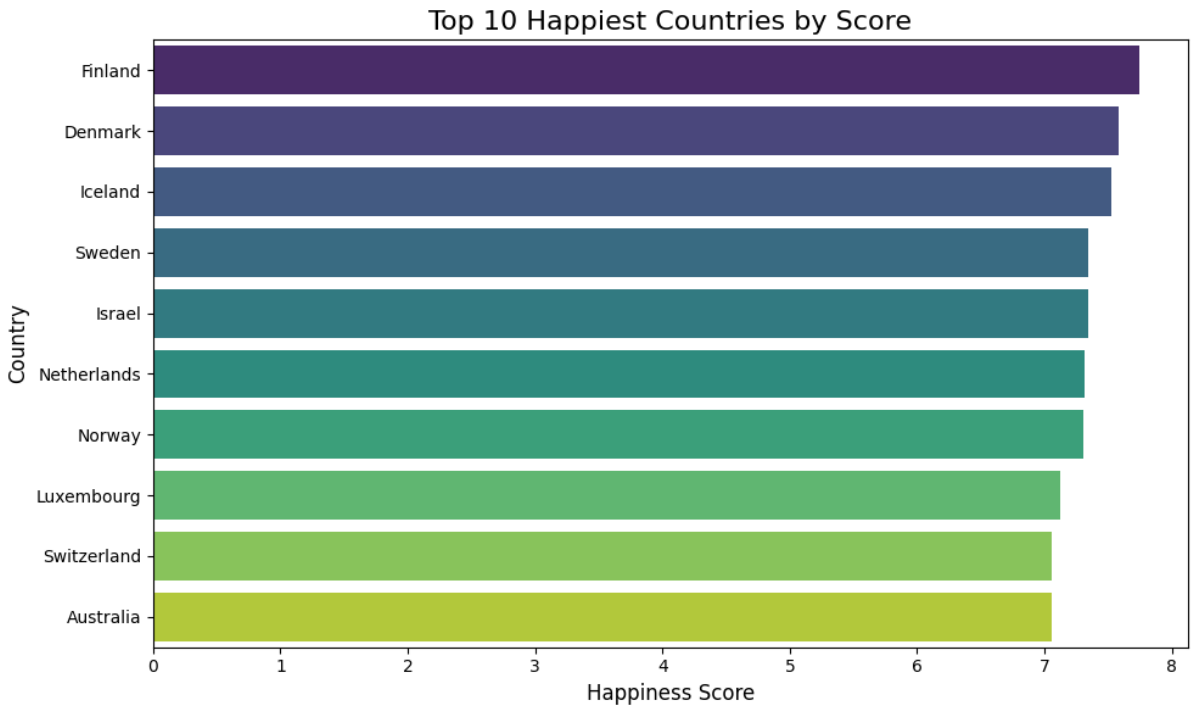
Extra column "Happiness Category" added that categorizes countries as "Low", "Medium", or "High" based on their score.

```python
# Add a new column "Happiness Category"
def categorize_happiness(score):
    if score < 4:
        return 'Low'
    elif 4 <= score <= 6:
        return 'Medium'
    else:
        return 'High'

# Apply the function to create the new column
data['Happiness Category'] = data['score'].apply(categorize_happiness)

# Display the first few rows to confirm the changes
print(data[['score', 'Happiness Category']].head(10))
```

```
   score Happiness Category
0  7.741               High
1  7.583               High
2  7.525               High
3  7.344               High
4  7.341               High
5  7.319               High
6  7.302               High
7  7.122               High
8  7.060               High
9  7.057               High
```

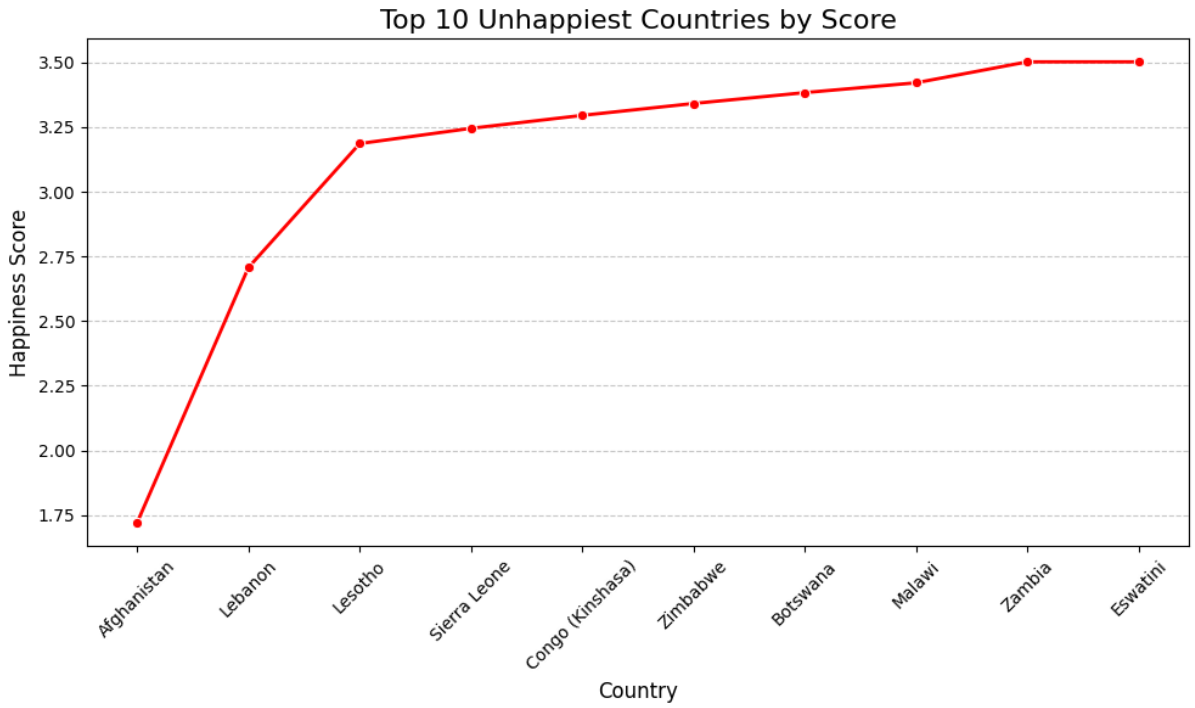
1.6 Data Visualizations:

- Bar Plot: Top 10 Happiest countries by Score.
- Line Plot: Bottom 10 Least happy countries by Score.
- Histogram: Score Column.
- Scatterplot of GDP per Capita vs. Score.



Observations and Insights:

Countries like Finland, Denmark, and Switzerland have the highest scores, whereas countries like Afghanistan and some Sub-Saharan countries rank low. The happy scores are skewed to the right, which indicates most countries attain high levels in the happiness scale.


Top 10 Unhappiest Countries by Score
Happiness Score
3.50
3.25
3.00
2.75
2.50
2.25
2.00
1.75
Afghanistan
Lebanon
Lesotho
Sierra Leone
Congo (Kinshasa)
Zimbabwe
Botswana
Malawi
Zambia
Eswatini
Country

Problem 2: Advanced Data Exploration Tasks

2.1 Preparing the South-Asia Dataset

The goal of this task was to extract data for South Asian countries and prepare it for further analysis. We filtered the data from those countries of South Asia: India, Pakistan, Bangladesh, and so on using a provided list. We save filtered data in a CSV format that may be utilized for other use.

```
#3.2 Problem 2 Some Advanced Data Exploration Task
#Task 1 Setup Task Preparing the South-Asia Dataset
#.1
#  Define the countries in South Asia with a list for example:
south_asian_countries = ["Afghanistan", "Bangladesh", "Bhutan", "India","Maldives", "Nepal", "Pakistan", "Srilanka"]
print(south_asian_countries)
```

```
['Afghanistan', 'Bangladesh', 'Bhutan', 'India', 'Maldives', 'Nepal', 'Pakistan', 'Srilanka']
```

2.2 Composite Score Ranking

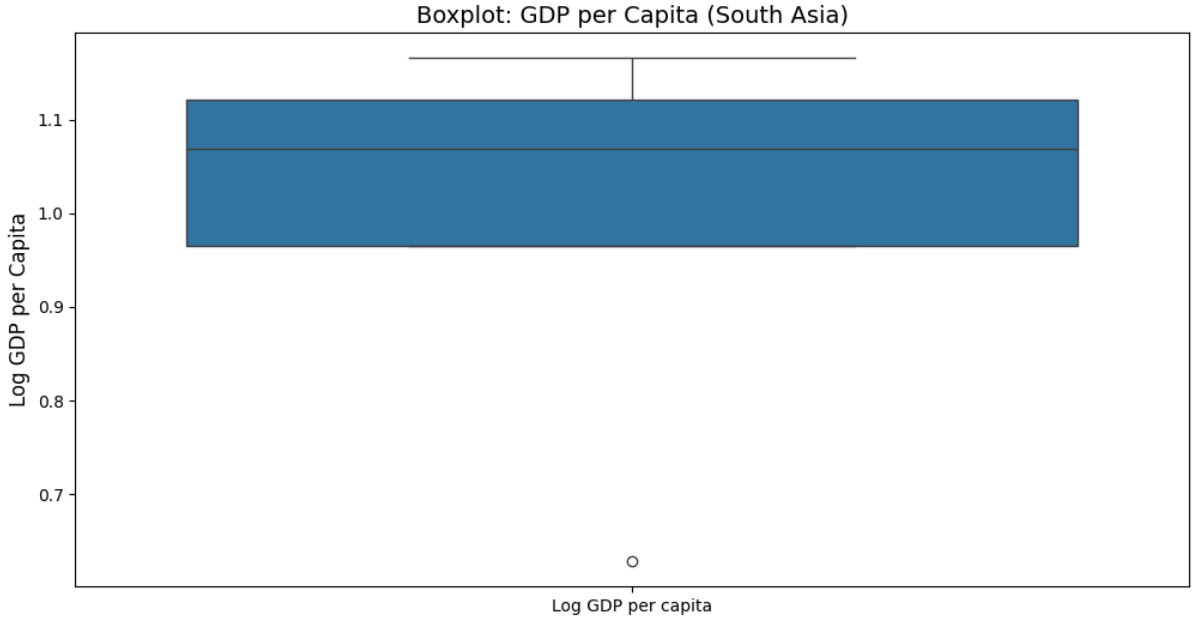
According to the formula combining Per Capita GDP, Social Support, and Life Expectancy Healthy using weighted measures, we computed a new composite score for South Asian countries. We used this formula to create the "Composite Score" new column. The countries were then ranked according to the same composite score. Bar Chart**:** Top 5 South Asian countries by Composite Score. There is a fair relationship between GDP per Capita and the happiness score in South Asia, but other factors such as social support and life expectancy also have major contributions to overall happiness.

```
#Task - 2 - Composite Score Ranking:
#.1


# Create a copy of the data to avoid modifying the original
SouthAsia = filtered_data.copy()

# Calculate Composite_Score
SouthAsia['Composite_Score'] = 0.40 * SouthAsia['Log GDP per capita'] + 0.30 * SouthAsia['Social support'] + 0.30 * SouthAsia['Healthy life expectancy']

# Display the result with Country name and Composite_Score
print(SouthAsia[['Country name', 'Composite_Score']])
```

```
    Country name  Composite_Score
92         Nepal           0.8159
107     Pakistan           0.7039
125        India           0.7874
128   Bangladesh           0.6774
142  Afghanistan           0.3238
```

2.3 Outlier Detection

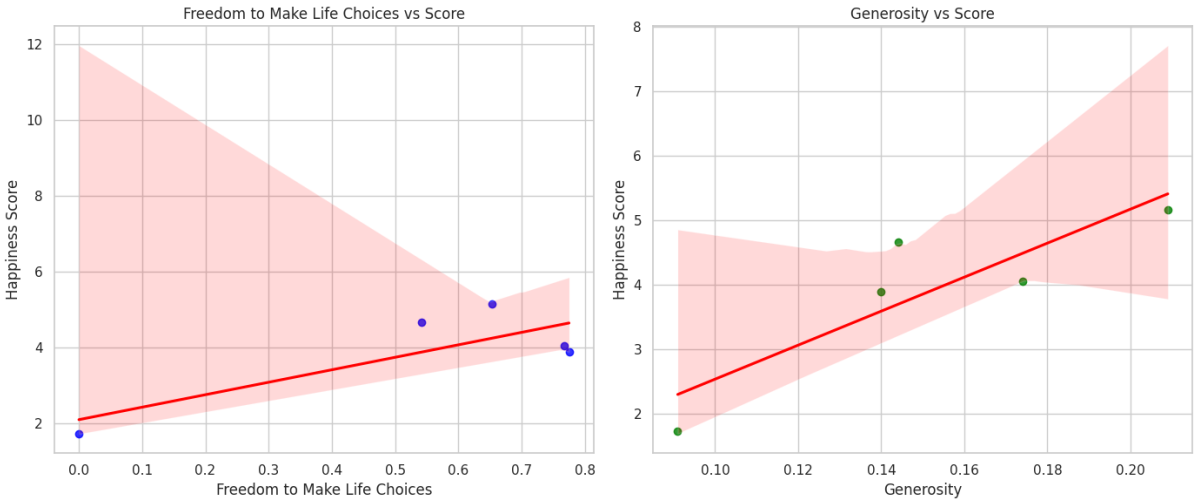
This exercise was based on the detection of outliers from the South Asian dataset using the Interquartile Range Rule in both the Score and GDP per Capita columns. We used the rule of 1.5 * IQR to identify outliers in both happiness score and GDP per Capita. We highlighted the outlier countries in a scatter plot. Some countries, such as India and Sri Lanka, were clearly shown to be outliers because of the mismatch between the two variables of GDP per Capita and happiness scores.




```
Outliers based on GDP per Capita (IQR method):
  Country name  Log GDP per capita
4  Afghanistan                0.628

SouthAsia with GDP_Z (Z-scores for GDP):
  Country name  Log GDP per capita     GDP_Z
0        Nepal                0.965 -0.129511
1     Pakistan                1.069  0.409255
2        India                1.166  0.911758
3   Bangladesh                1.122  0.683819
4  Afghanistan                0.628 -1.875320

Outliers based on Z-score for GDP per Capita:
Empty DataFrame
Columns: [Country name, Log GDP per capita, GDP_Z]
Index: []
```

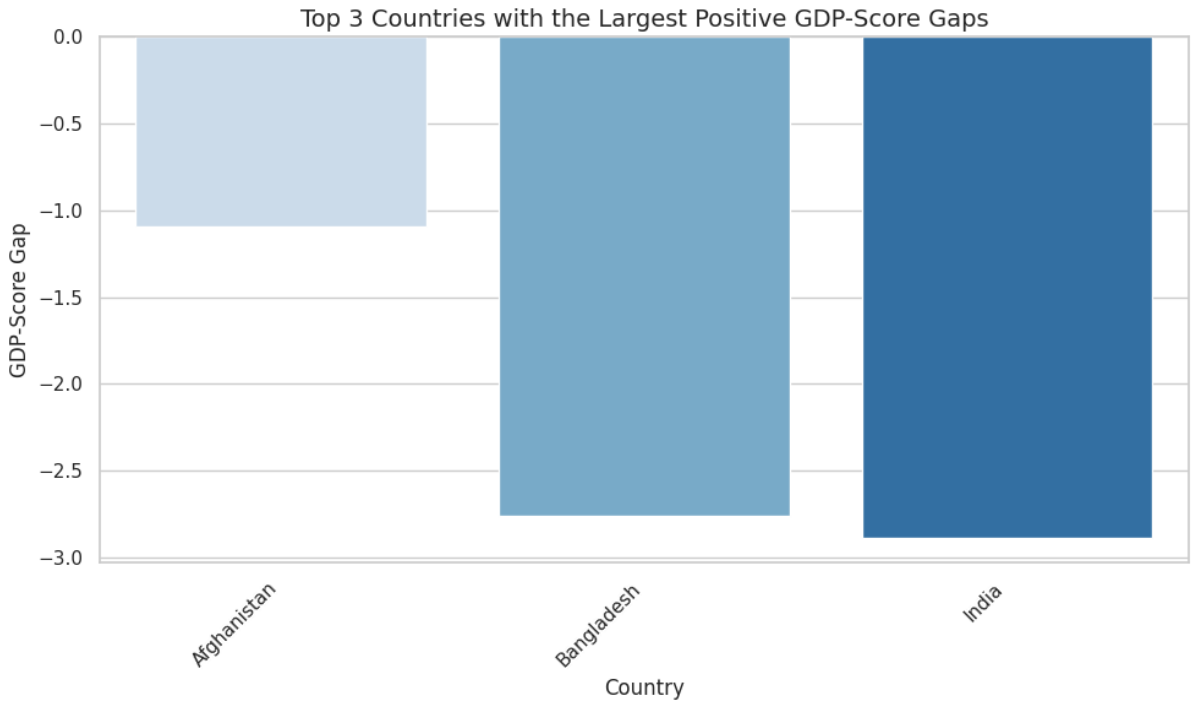
2.4 Exploring Trends Across Metrics

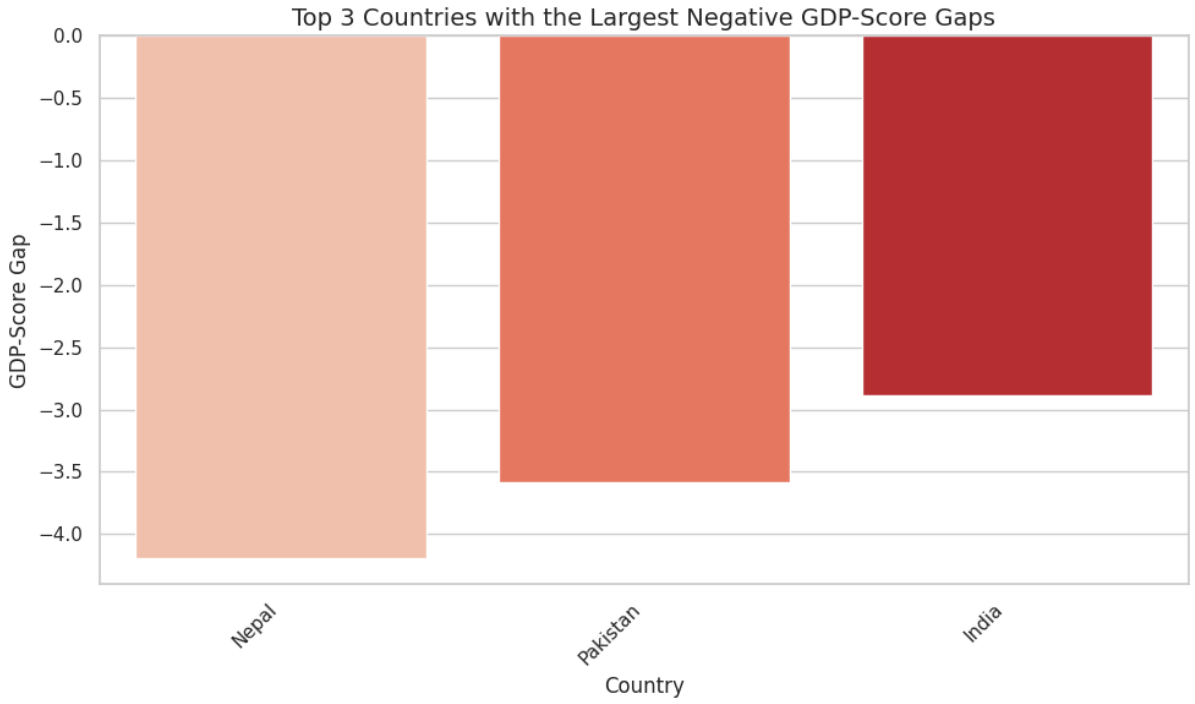
We have analyzed the relationship between the happiness score and other metrics such as Freedom to Make Life Choices and Generosity. For the selected metrics, Pearson's correlation coefficients were calculated. The relations are visualized by means of scatter plots.



2.5 Gap Analysis

We analyzed the gap between GDP per Capita and the happiness score using the "GDP-Score Gap" column. Created the GDP-Score gap column and ranked the countries by the largest positive and negative gaps. Bar Chart Highlighting the countries with the largest positive and negative GDP-Score gaps.





Problem 3: Comparative Analysis between South Asia and the Middle East

3.1 Preparing the Middle Eastern Dataset

Just as for the South Asian analyses, the Middle Eastern countries were excluded and a separate dataset created.

1 to 10 of 10 entries Filter

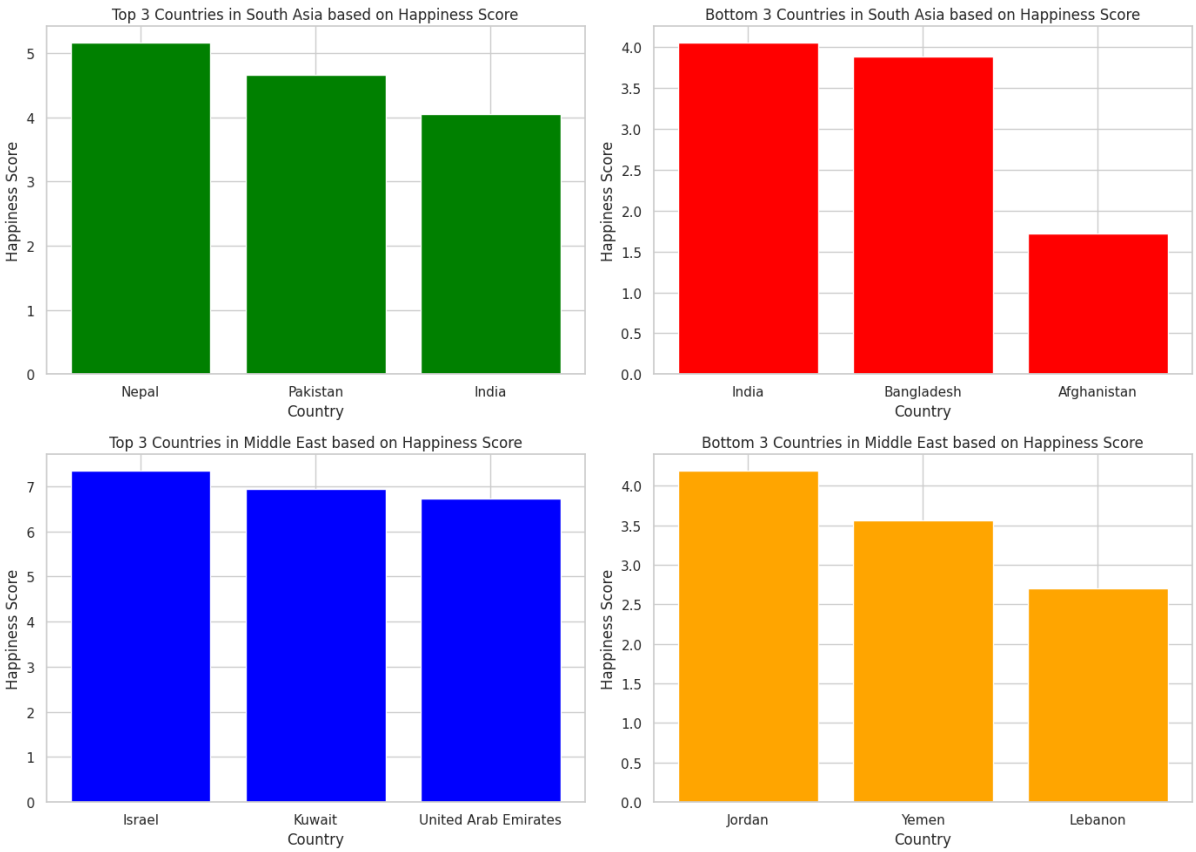
Country name	score	Log GDP per capita	Social support	Healthy life expectancy ▲	Freedom to make life choices	Generosity	Perceptions of corruption	Dystopia + residual	Happiness Category
Bahrain	5.959								Medium
Yemen	3.561	0.671	1.281	0.293	0.362	0.08	0.113	0.76	Low
Iraq	5.166	1.249	0.996	0.498	0.425	0.141	0.048	1.809	Medium
Saudi Arabia	6.594	1.842	1.361	0.511	0.787	0.114	0.188	1.79	High
Lebanon	2.707	1.377	0.577	0.556	0.173	0.068	0.029	-0.073	Low
United Arab Emirates	6.733	1.983	1.164	0.563	0.815	0.209	0.258	1.741	High
Iran	4.923	1.435	1.136	0.571	0.366	0.235	0.123	1.057	Medium
Jordan	4.186	1.262	0.983	0.594	0.593	0.059	0.189	0.504	Medium
Kuwait	6.951	1.845	1.364	0.661	0.827	0.2	0.172	1.884	High
Israel	7.341	1.803	1.513	0.74	0.641	0.153	0.193	2.298	High

3.2 Descriptive Statistics

We made a comparison of the average and standard deviation of the happiness scores for South Asia and the Middle East. Mean and standard deviation have been derived for both regions. Bar Chart: Comparison of top 3 and bottom 3 countries in both regions.

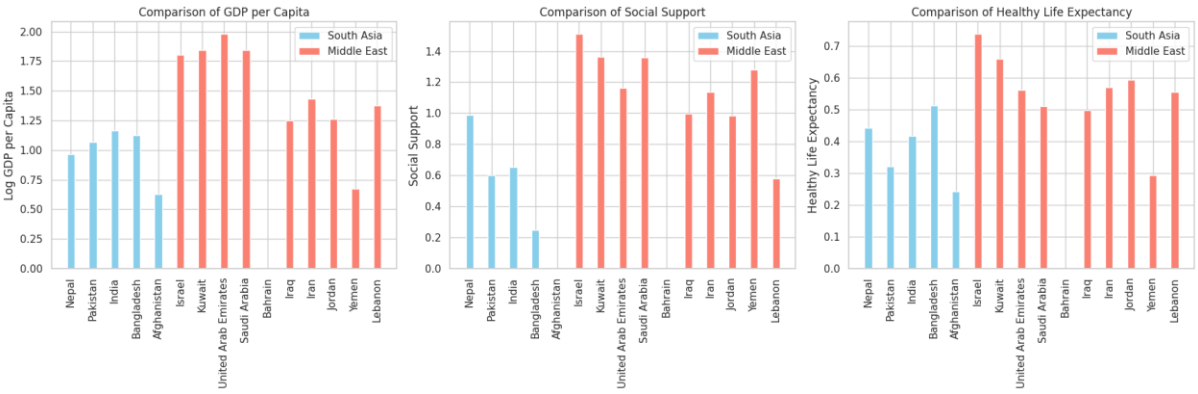
```
South Asia - Mean of Score: 3.8952000000000004
South Asia - Standard Deviation of Score: 1.3160025455902433

Middle East - Mean of Score: 5.412100000000001
Middle East - Standard Deviation of Score: 1.5662011684327144
```

```
Top 3 countries in South Asia based on happiness score:
  Country name  score
0        Nepal  5.158
1     Pakistan  4.657
2        India  4.054

Bottom 3 countries in South Asia based on happiness score:
  Country name  score
2        India  4.054
3   Bangladesh  3.886
4  Afghanistan  1.721

Top 3 countries in Middle East based on happiness score:
           Country name  score
0                Israel  7.341
1                Kuwait  6.951
2  United Arab Emirates  6.733

Bottom 3 countries in Middle East based on happiness score:
  Country name  score
7       Jordan  4.186
8        Yemen  3.561
9      Lebanon  2.707
```



3.3 Metric Comparisons

We have set up comparison with few key metrics like GDP per capita, social support, and healthy life expectancy for both regions. Grouped bar charts comparisons show these in between south Asia and middle east.



3.4 Happiness Disparity

We have compared both value ranges and coefficients of variations (CV) of happiness scores as well as their differences between the two regions.

```
Coefficient of Variation (CV) for South Asia:
CV: 34.23%

Coefficient of Variation (CV) for Middle East:
CV: 35.86%

Middle East has greater variability in happiness.
```

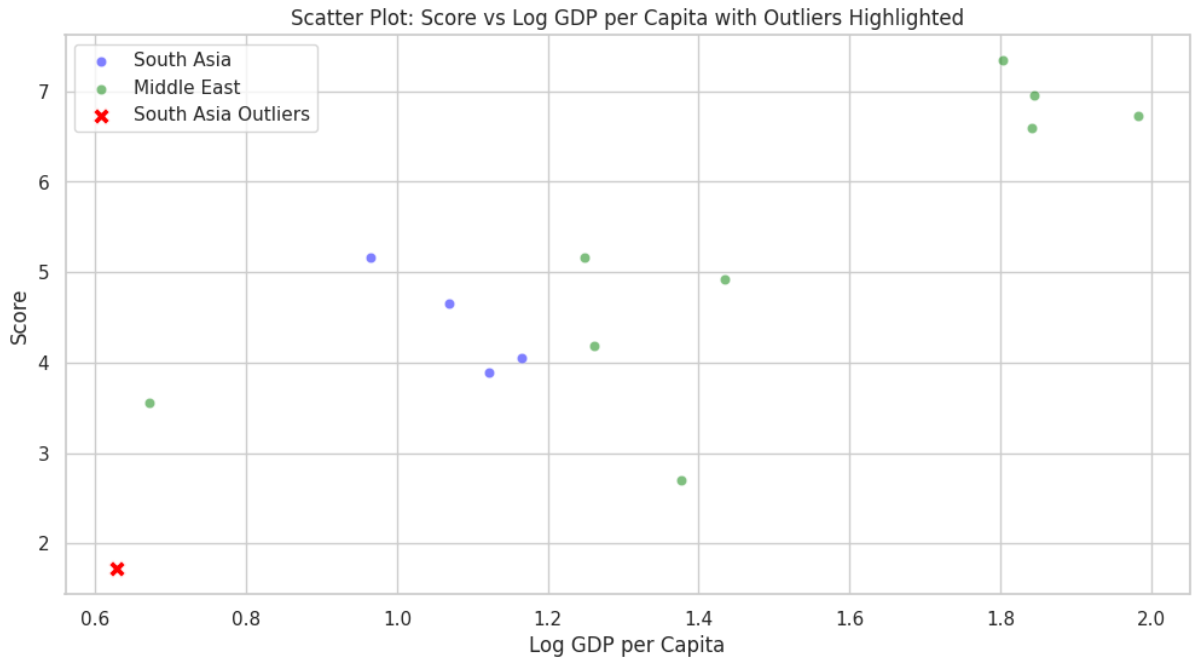
3.5 Correlation Analysis

We could analyze Its mean happiness score along with the correlation of freedom to make life choices, and generosity for the two regions.

```
Correlation Matrix for South Asia:
                                score  Freedom to make life choices  \
score                        1.000000                      0.801498
Freedom to make life choices 0.801498                      1.000000
Generosity                   0.877333                      0.740660

                             Generosity
score                          0.877333
Freedom to make life choices   0.740660
Generosity                     1.000000

Correlation Matrix for Middle East:
                                score  Freedom to make life choices  \
score                        1.000000                      0.863220
Freedom to make life choices 0.863220                      1.000000
Generosity                   0.627524                      0.388854

                             Generosity
score                          0.627524
Freedom to make life choices   0.388854
Generosity                     1.000000

South Asia: Correlation between score and Freedom to make life choices: 0.80
South Asia: Correlation between score and Generosity: 0.88

Middle East: Correlation between score and Freedom to make life choices: 0.86
Middle East: Correlation between score and Generosity: 0.63
```

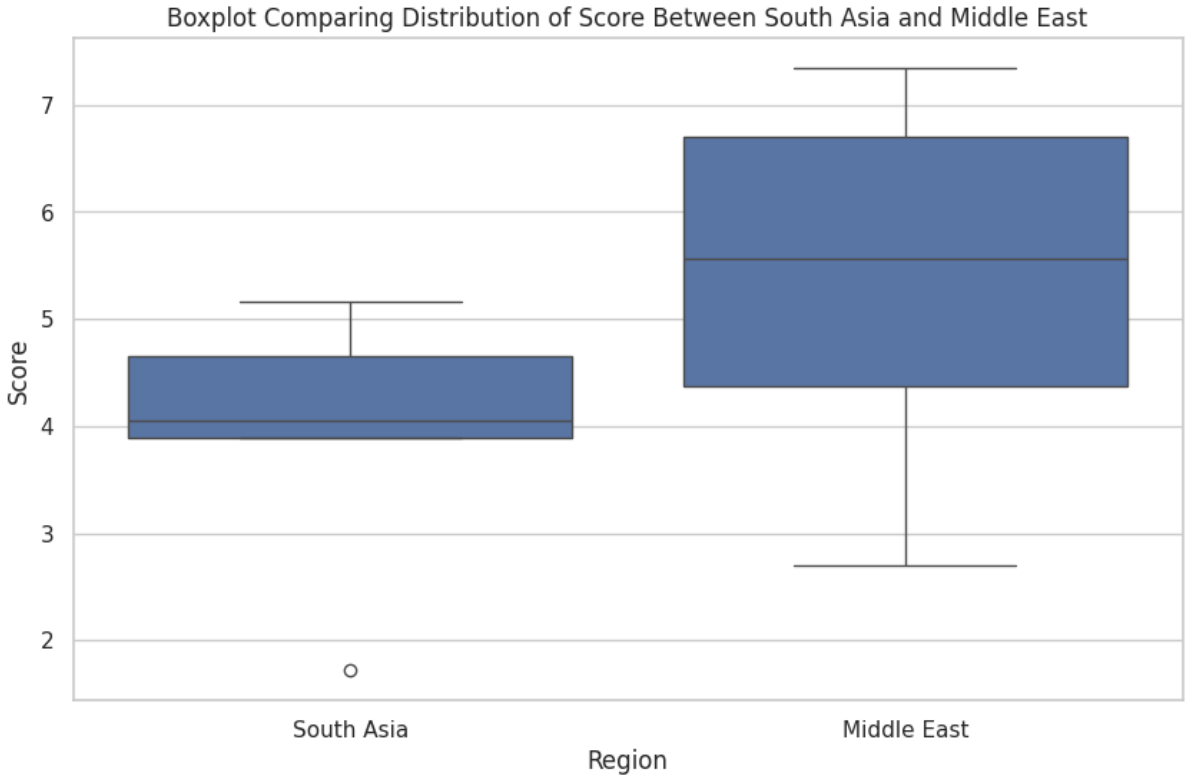
3.6 Outlier Detection

Outliers were identified for both areas using Score and GDP per Capita.



3.7 Visualization

We made boxplots that show distribution of happiness scores for South Asia and the Middle East against one another.



Conclusion

The analysis, therefore, discloses a distinction in happiness scores by countries in South Asia and countries in the Middle East. Happiness differs not only in terms of across countries in both regions but also GDP per Capita, social support, and healthy life expectancy factors that determine happiness scores. Economically speaking, South Asian countries despite having a higher GDP per capita than some others are yet to achieve good social support and life expectancy levels, which are instrumental in scaling down the happiness rank of the region.