Year	Module	Assessment Type
2024	Concepts and Technologies of AI	Academic Report

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

Student Id : 2431328
Student Name : Ugesh KC
Section : L5CG22
Module Leader : Mr. Siman Giri
Tutor : Mr. Ronit Shrestha
Submitted on : 12/20/2024

1.Introduction

The World Happiness Report is a survey which shows global happiness and ranks it according to happiness level. It considers factors like economic prosperity, social support, health and freedom. As it is a report of happiness of people worldwide, the report helps to identify or assume conditions of people living in specific country.

The World Happiness report is an important resource which helps to understand people's livelihoods in different countries. It helps to provide actionable options for improving quality of life by looking factors that contribute to happiness.it helps to address social economic and governance challenges.

The objective of this report is to understand the dataset through the means of Statistical figures like Histograms, Bar charts, Box plots which help in visual representation of numerical data. It analyzes happiness trends specifically within South Asia to identify regional patterns. Likewise, this report includes the comparison of South Asia and Middle East to look at the differences in happiness levels. Overall, it focuses on data exploration, South Asia analysis, and South Asia versus Middle East comparison.

Problem 1: Getting Started with Data Exploration

Steps:

1. Dataset Overview:

Import library of pandas using **import pandas as pd.** Load the dataset using **pd.read_csv('World_Hapiness_Report.csv')** store it in a variable **'df'**. Display the first 10 rows of the dataset using **.head()** method. Identify number of rows and columns using **.shape[]**method. List columns and their data types using **.dtypes** method.

2. Basic Statistics:

Calculate the mean, median and standard deviation for the score column using the **.mean()**, **.median()**, **.std()** methods respectively. Use **.loc[[].idxmax()], .loc[[].idxmin()]** method to identify the country with the highest and lowest happiness scores.

3. Missing Values:

Use the **.isnull().sum()** method to display the total count of missing values for each column.

4. Filtering and Sorting:

Filter the dataset for countries with a Score greater than 7.5 using the comparative operator **'>'** and the condition **['score'] > 7.5** which will return only the rows where the score is greater than 7.5.For the Filtered dataset, arrange the dataset by GDP per Capita in descending order using the **.sort_values()** method with **ascending=False**.After sorting display the top 10 rows using the **.head()** method.

5. Adding New Columns:

Create a new column called Happiness Category using **df['Happiness Category']** and make function **categorize_happiness(score)** which categories countries into three categories based on their score using if elif statement:

-low if the score is less than 4.

-Medium if the score is between 4 and 6.

-High if the score is greater than 6.

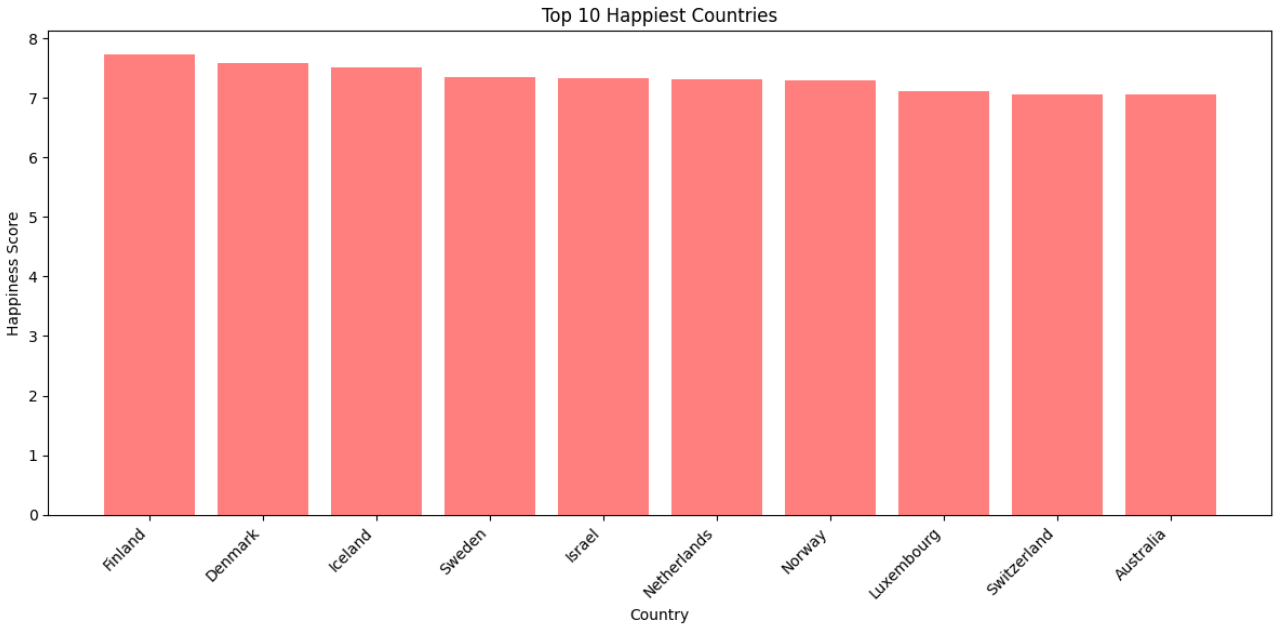
The function is applied to the score column using **.apply()** to create a new column.

6. Data Visualizations:

For data visualizations **import matplotlib.pyplot as plt** and **import seaborn as sns.**

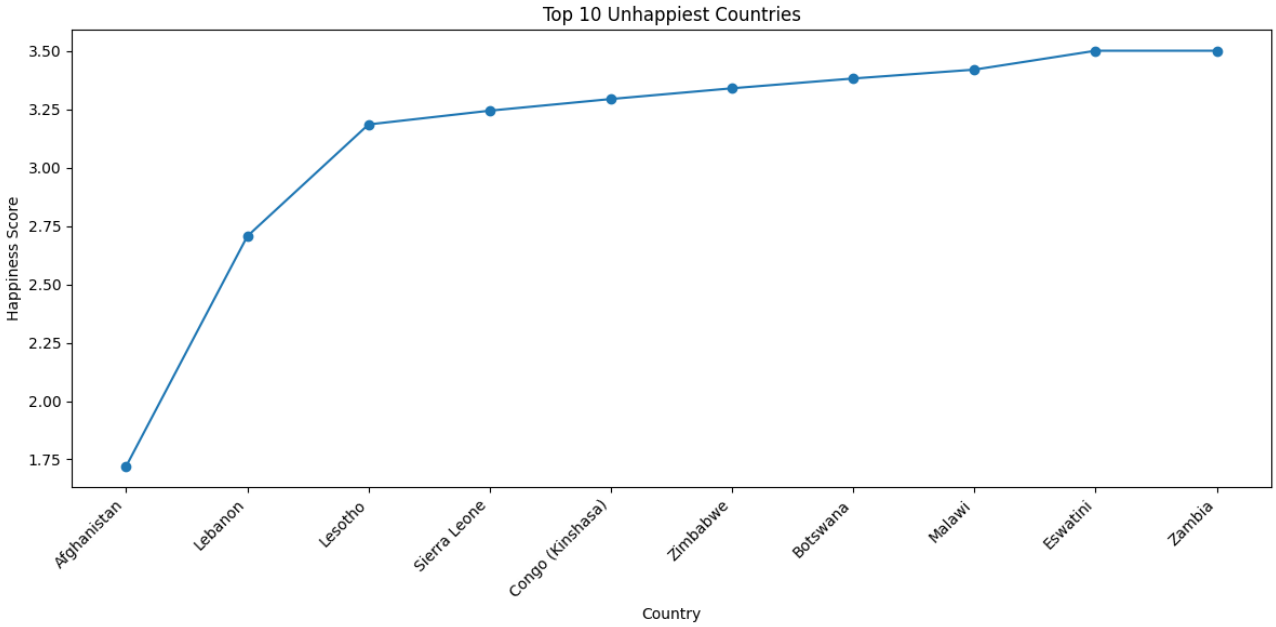
For all visualizations, **plt.xlabel(), plt.ylabel(),** and **plt.title()** are used to label axes and add a title. In bar and line plots**, plt.xticks()** is used to rotate labels for better readability.

a. **Bar Plot**: Use **plt.bar() method** to display data as vertical bars. It compares categories of scores of the top 10 happiest countries.



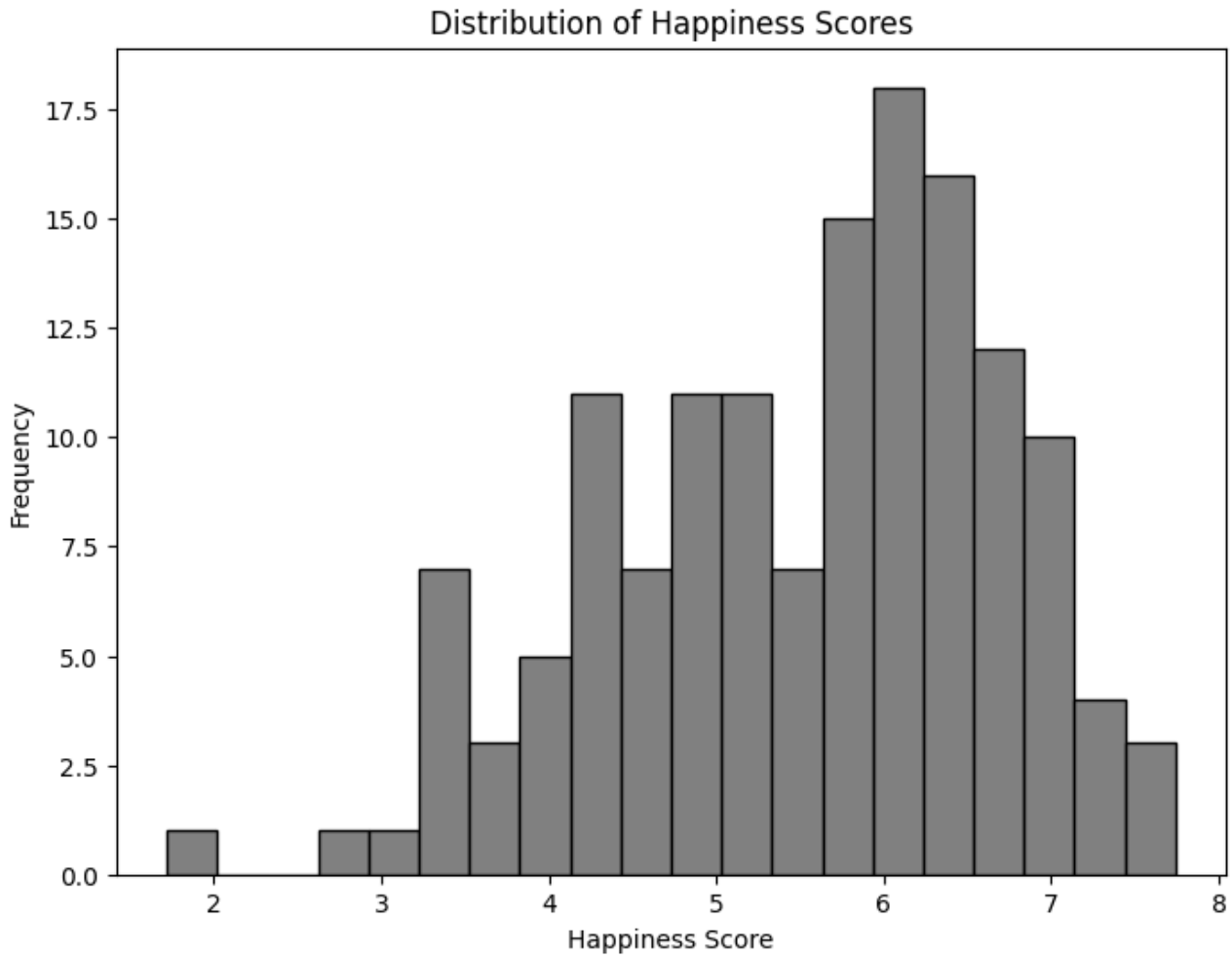
Fig(a)

b. **Line Plot**: Use **plt.plot()** with markers to connect data points which makes it easier to visualize trends of the top 10 unhappiest countries.



Fig(b)

c. **Histogram**: Use **plt.hist()** to display the frequency distribution happiness scores across all countries.

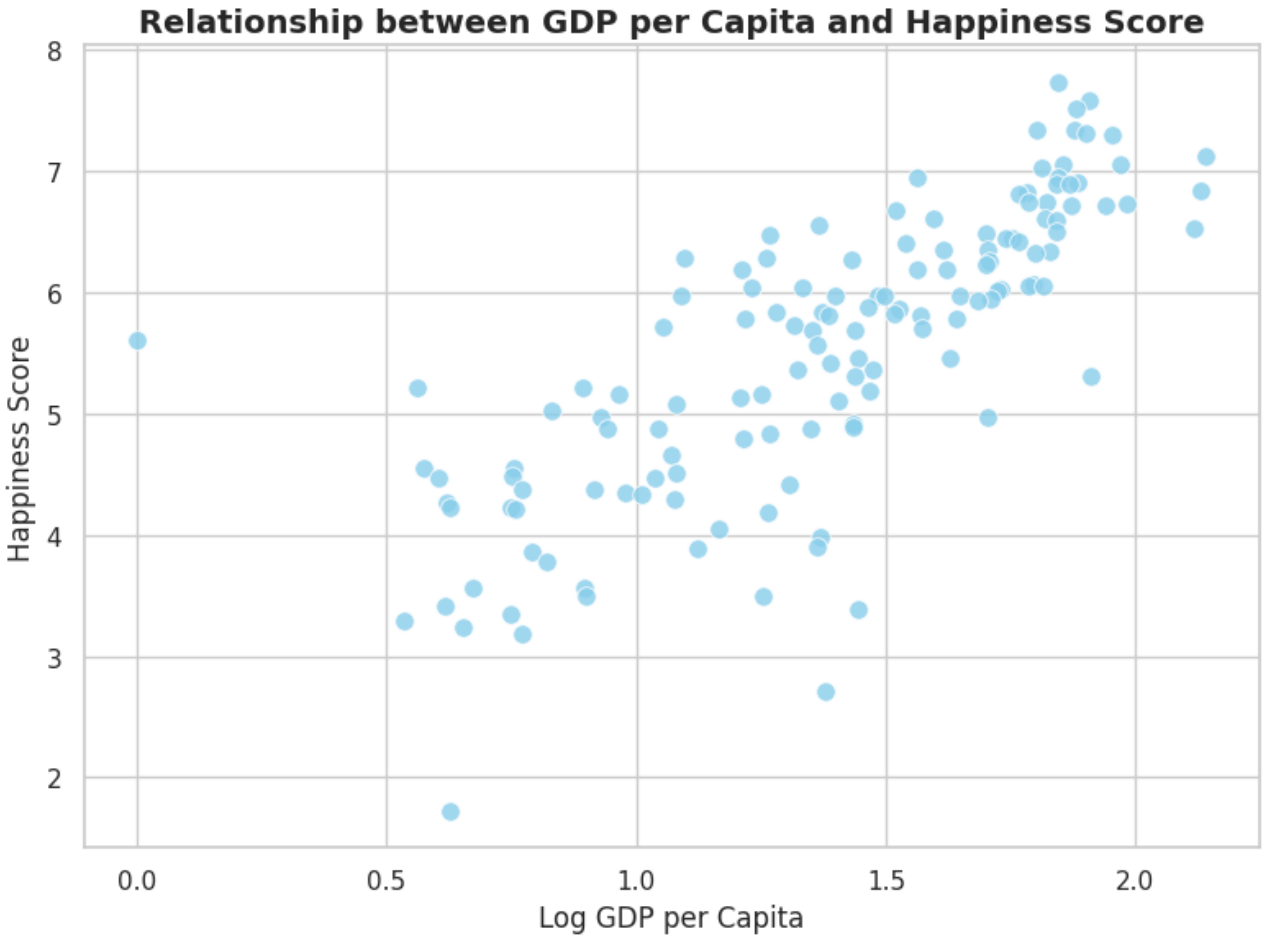


Fig(c)

Interpretation of Histogram

The above histogram provides happiness score distribution across countries. The score range between 2 and 8.The most frequent scores are around 6 and the highest frequency is at 18 which means many countries fall in this range. The distribution is slightly left skewed which means there are fewer countries with very high scores than lower scores. Outliers can be seen near 2 and close to 8.

d. **Scatter Plot**: Use **sns.scatterplot()** to plot data points which helps to visualize relationships between GDP per Capita and Happiness Score.


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

Fig(d)

Problem 2: Analyzing South Asia

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

Steps:

1.Define the countries:

A list of south Asian countries is defined with countries like Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka.

2. Filter the dataset:

To load the dataset use **pd.read_csv()** and to filter the matching dataset from the list use **.isin()** method.

3. Save the filter dataframe:

To save the filtered dataframe as separate CSV files for future use **.to_csv()** method can be used.

Task - 2 - Composite Score Ranking:

Steps:

1.Composite Score Calculations:

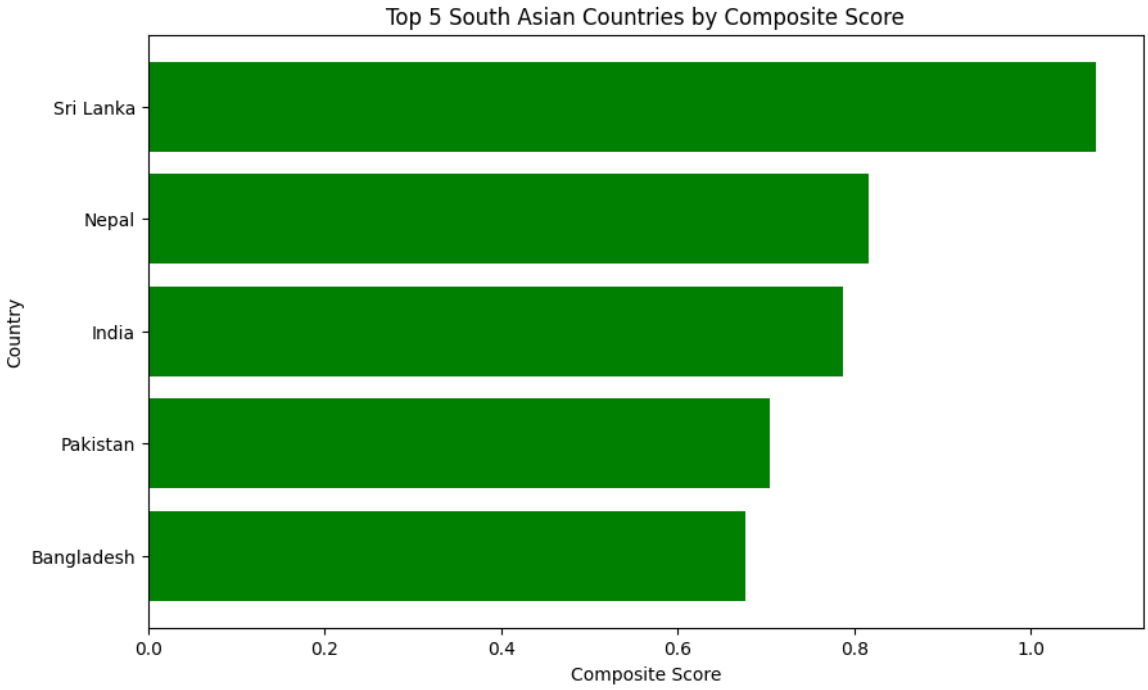
Create a new column **Composite Score** in the **South Asia Dataframe.** The score is calculated as **Composite Score = 0.40 × GDP per Capita + 0.30 × Social Support + 0.30 × Healthy Life Expectancy**

2.Ranking by Composite Score:

Rank the South Asian countries based on the Composite Score in descending order using **.sort_values()** method.

3.Visualization of Top 5 Countries:

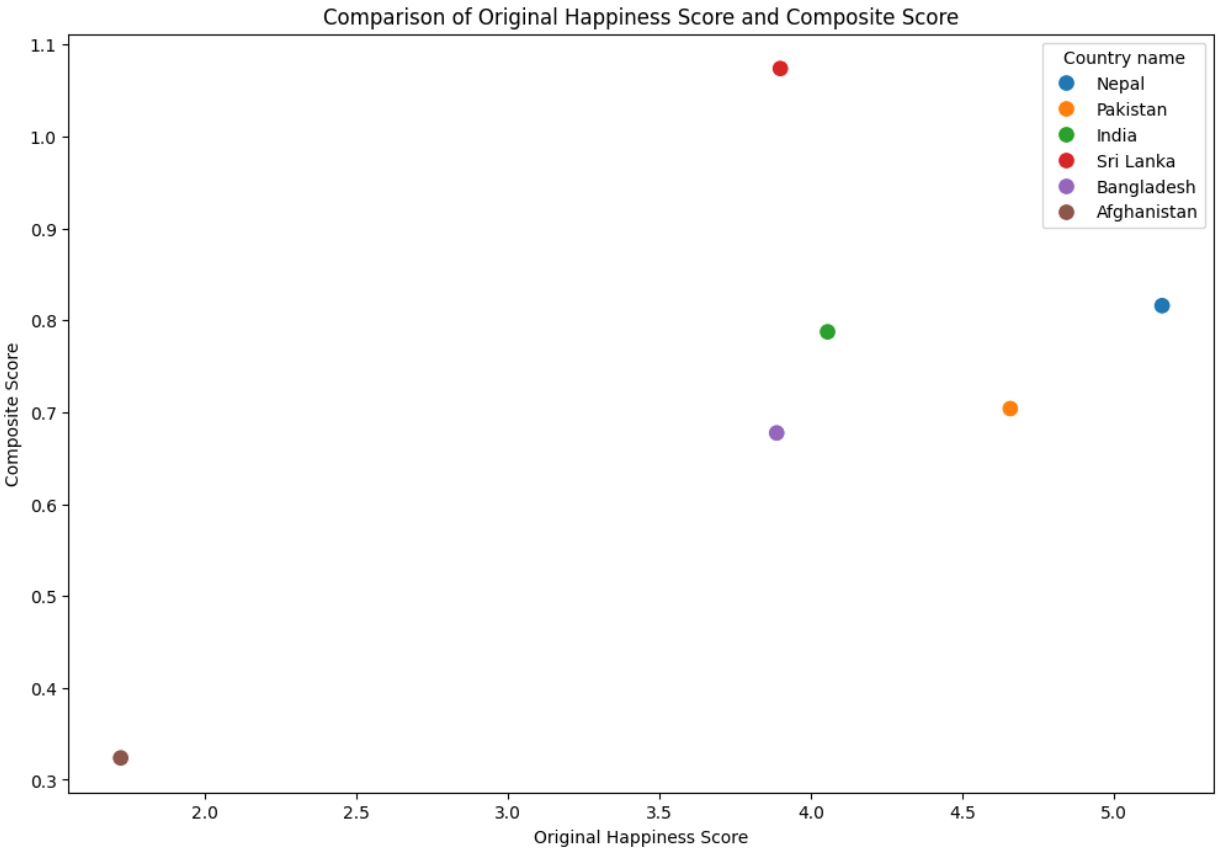
Visualize the top 5 countries using horizontal bar chart showing the Composite Score using **plt.barh()** method with the countries on the y-axis and Composite Scores on the x-axis.



Fig(e)

4.Comparison with Original Happiness Score:

Create a scatter plot to compare 'Score' and 'Composite Score' using the **sns.scatterplot()** method.



Fig(f)

Discussion:

The rankings based on the Composite Score do not completely match with the Original Happiness Score. Sri Lanka performs better in the Composite Score than Nepal, Pakistan and India but in original score it does not. Nepal original score is 5.158 which is top-ranked but it is not on top in terms of Composite Score. Afghanistan score is consistently low on both. This difference shows that the Composite Score includes other factors like social, economic or well being measures. Composite Score gives broader view of happiness than the Original Score.

Task - 3 - Outlier Detection:

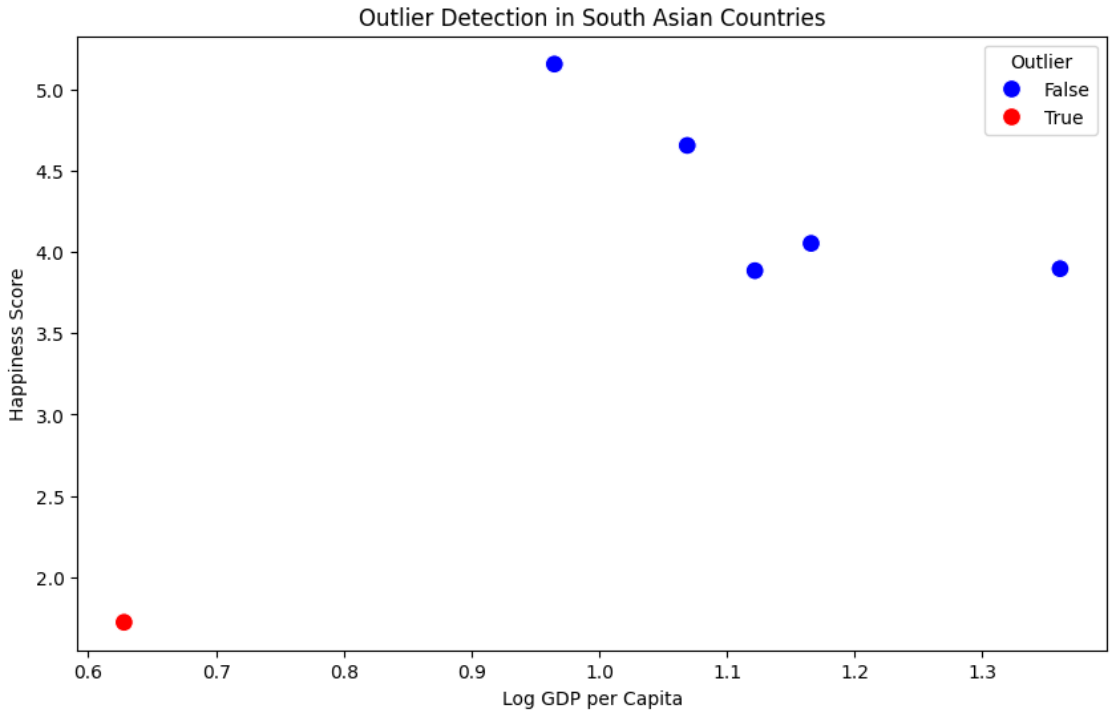
Steps:

1.Identification of outlier countries:

Identify Outliers Using the 1.5 * IQR Rule. Where IQR = Q3 – Q1. Outliers are values below Q1-1.5*IQR and above Q3+1.5*IQR. Create two separate lists of outliers for Happiness Score and GDP per Capita then combined to get unique outliers using **.concat()** method and **.drop_duplicates** to ensure there are no duplicates in the final result**.**

2.Create Scatter plot:

To create a scatter plot use **.scatterplot()** with GDP per Capita on x- axis and Happiness Score on the y-axis.



Fig(g)

Characteristics of the Outlier:

a. Very low GDP per Capita approximately 0.6 and Happiness Scores around 1.7.

b. The value is far from the main group which have higher GDP values ranging between 0.9 and 1.3 and Happiness Scores between 3.5 and 5.1.

c. The value of the Outliers are likely the result of economic challenges or instability.

Impact on Regional Averages:

a. Presence of outliers can lowers average GDP and Happiness Scores.

b. It increases variability which makes averages less reliable.

c. this highlights the need to address inequalities in the region.

Task - 4 - Exploring Trends Across Metrics:

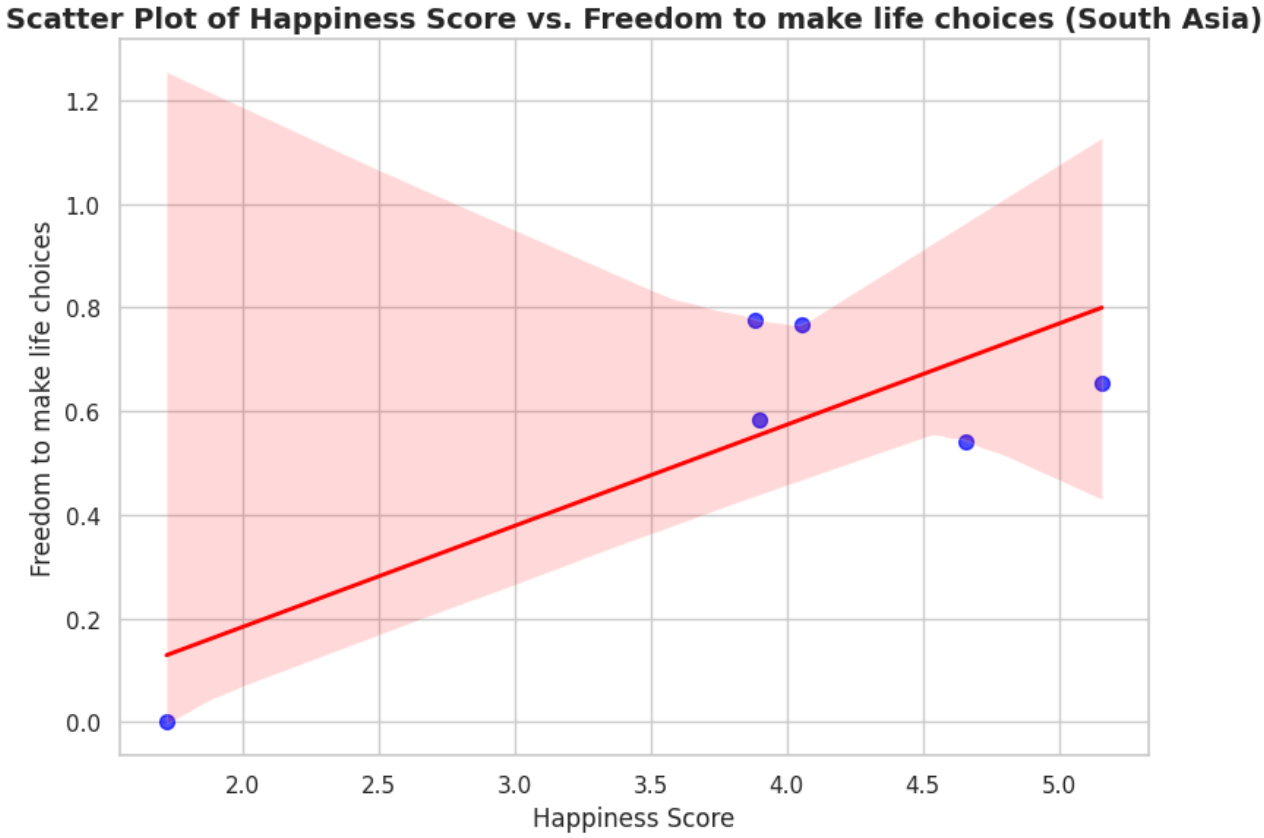
Steps:

a. Correlation Calculation and Scatter plots:

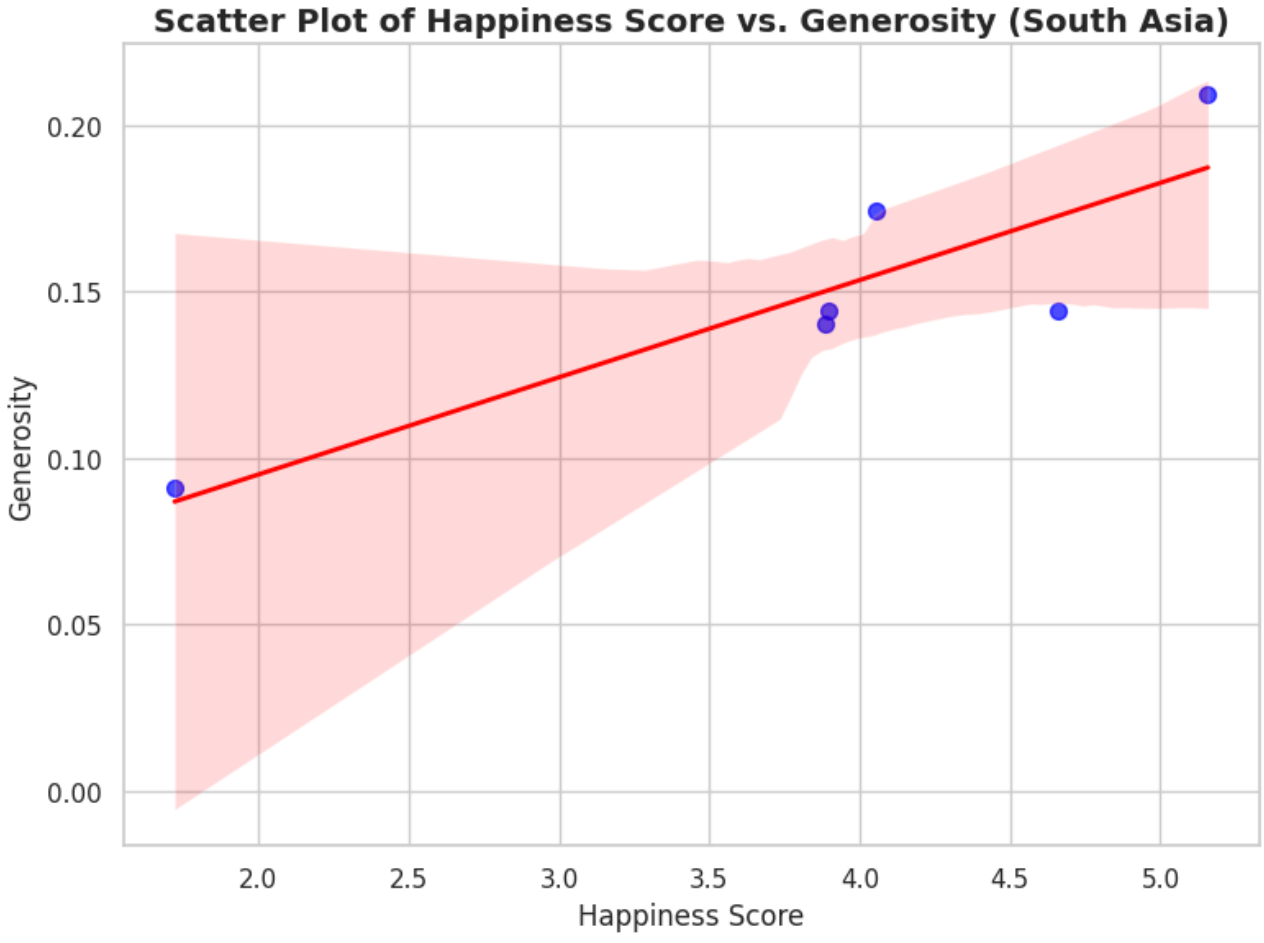
Calculate the correlation between the Happiness Score and selected metrices(Freedom to Make Life Choices and Generosity) using Pearson's correlation.to plot the scatter plots with trendlines to visualize the relationships use **sns.regplot().**

Discussion:

The strongest Relationship is between Happiness Score and Freedom to Make Life Choices which indicates impact of personal freedom on happiness. The weakest Relationship is between Happiness Score and Generosity. Freedom to make life choices has a clear positive trend with happiness. Generosity shows a weaker positive trend with more scattered points.


Scatter Plot of Happiness Score vs. Freedom to make life choices (South Asia)
Freedom to make life choices
Happiness Score
0.0
0.2
0.4
0.6
0.8
1.0
1.2
2.0
2.5
3.0
3.5
4.0
4.5
5.0

Fig(h)


Scatter Plot of Happiness Score vs. Generosity (South Asia)
Generosity
Happiness Score
0.00
0.05
0.10
0.15
0.20
2.0
2.5
3.0
3.5
4.0
4.5
5.0

Fig(i)

Task - 5 - Gap Analysis:

Steps:

1.Calculate GDP-Score Gap:

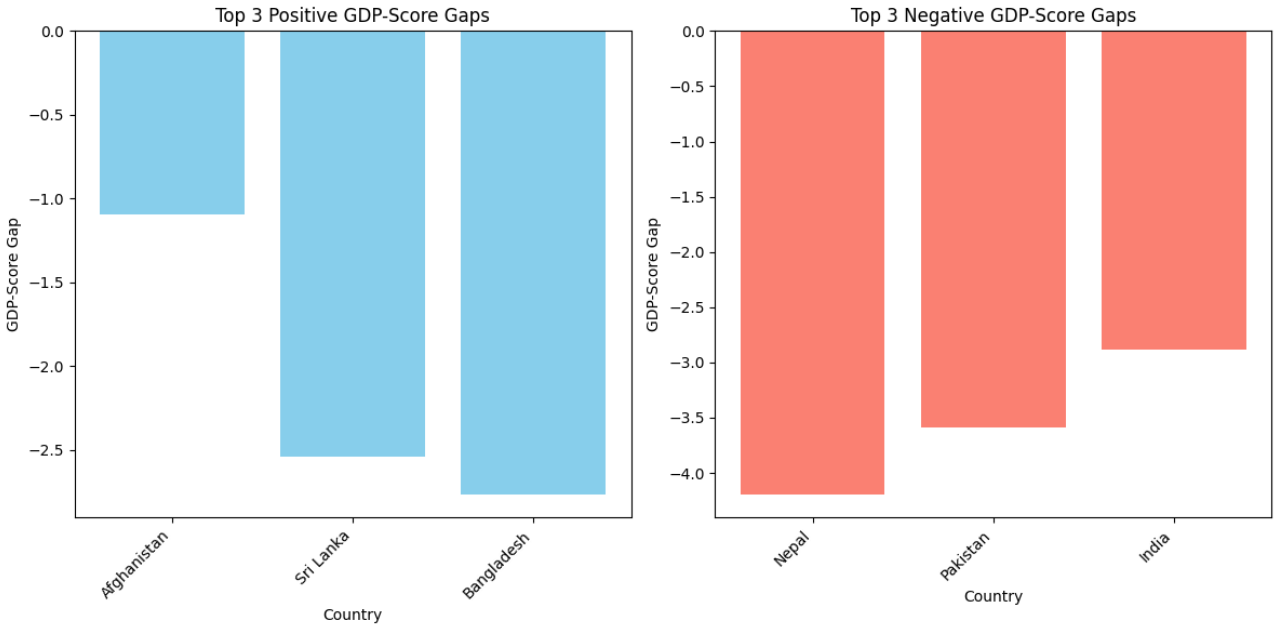
Add a new column **GDP-Score Gap** in dataset which shows the difference between the Log GDP per capita and the Score for south Asian country.

2.Ranking countries by the gap:

Rank the South Asian countries by this gap in both ascending and descending order using .sort_values() method and .**sortvalues(ascending = False)** method respectively.

3.Visualization:

Use **plt.bar()** to create bar charts for the top 3 countries with the largest positive and negative GDP-Score Gaps. In fig(j) sky blue color is used to show Top 3 Positive GDP-Score Gaps whereas for Negative Salmon color is used.



Fig(j)

Reasons behind the gaps

Top 3 Positive GDP-Score Gaps:

Countries like Sri Lanka, Bangladesh, and Afghanistan shows lower happiness scores than expected for their GDP. Possible reasons are:

a. High levels of inequality

b. Economic or political instability

c. Social or environmental challenges that GDP by itself does not consider

Top 3 Negative GDP-Score Gaps :

Countries like Nepal, Pakistan, and India shows happiness scores than expected for their GDP.

Possible reasons are:

a. Strong community ties and social support systems

b. Cultural emphasis on non-material aspects of life such as spirituality, family values)

Problem - 3 - Comparative Analysis:

Steps:

1. Preparing the Middle Eastern Dataset:

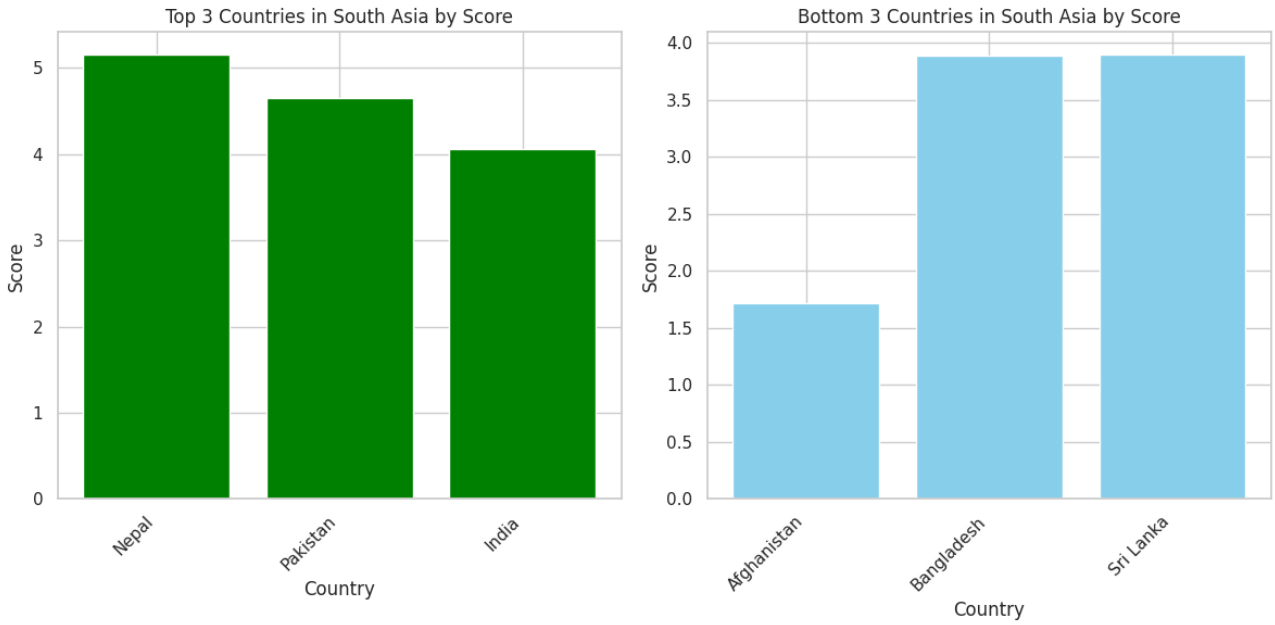
Define a list of Middle Eastern countries which includes Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Syria, United Arab Emirates, and Yemen. Load the dataset using **pd.read_csv().**Filter the countries matching the defined list using **.isin()** method. Store the filtered Dataframe in the variable middle_east_df.

2. Descriptive Statistics:

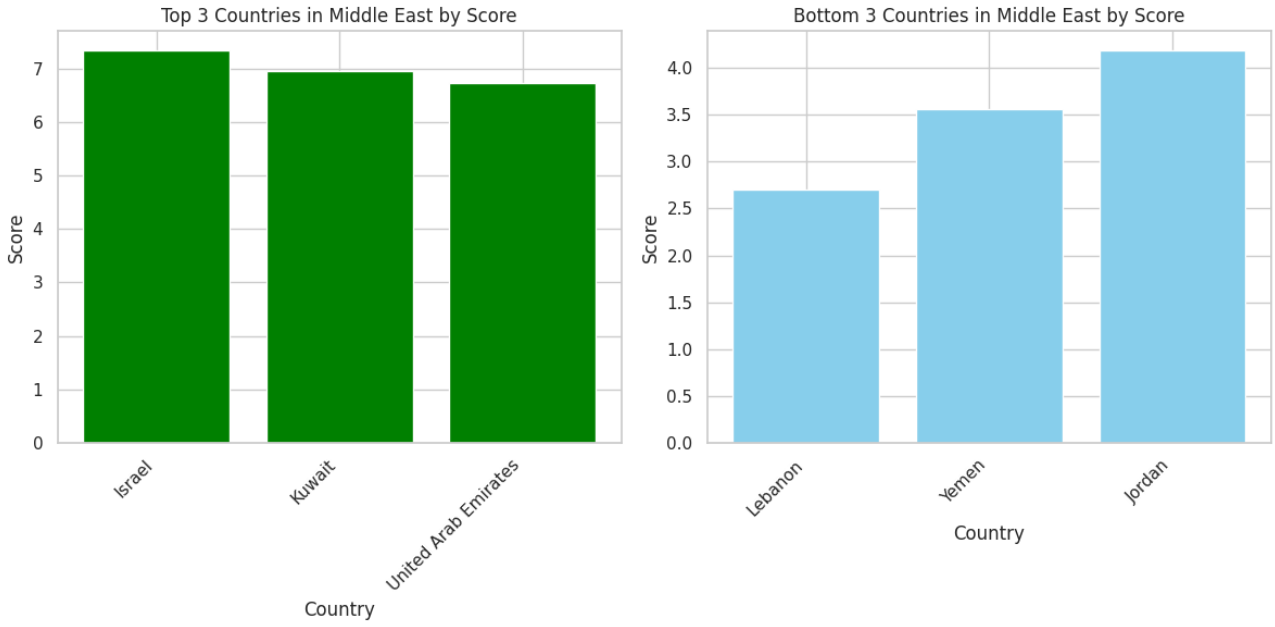
Calculate the mean and standard deviation of the Happiness Score for both South Asia and the Middle East using .mean() and .std() methods. Compare the average Happiness Scores of the two regions to determine which has a higher score using **if elif else** statement. The result should be “The Middle East has a higher average happiness score.”

3.Identification of Top and Bottom Performers:

Use **.nlargest()** to find the top 3 countries and **.nsmallest()** for the bottom 3 based on the Happiness Score in both region. Create bar charts for top and bottom performers using plt.bar().Use subplots to present both top and bottom performers side by side as shown in fig(k) ang fig(i).



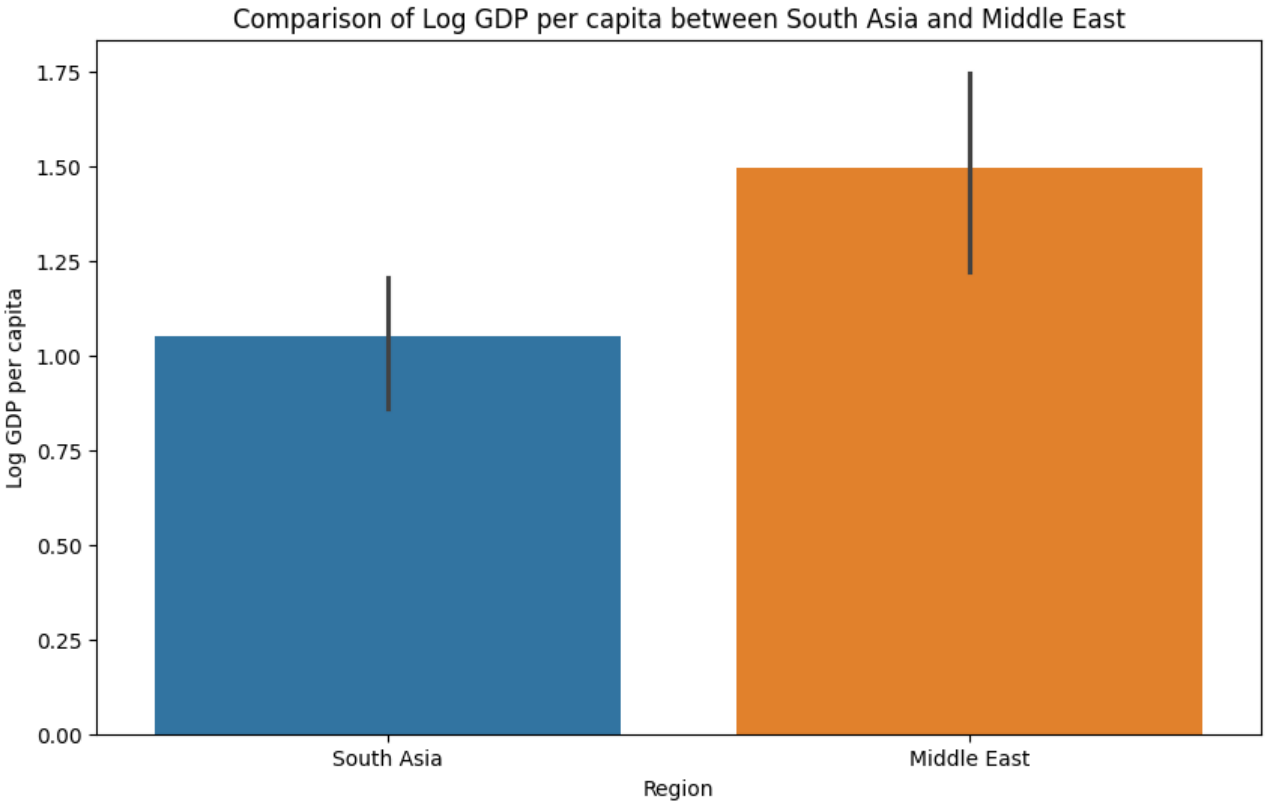
Fig(k)



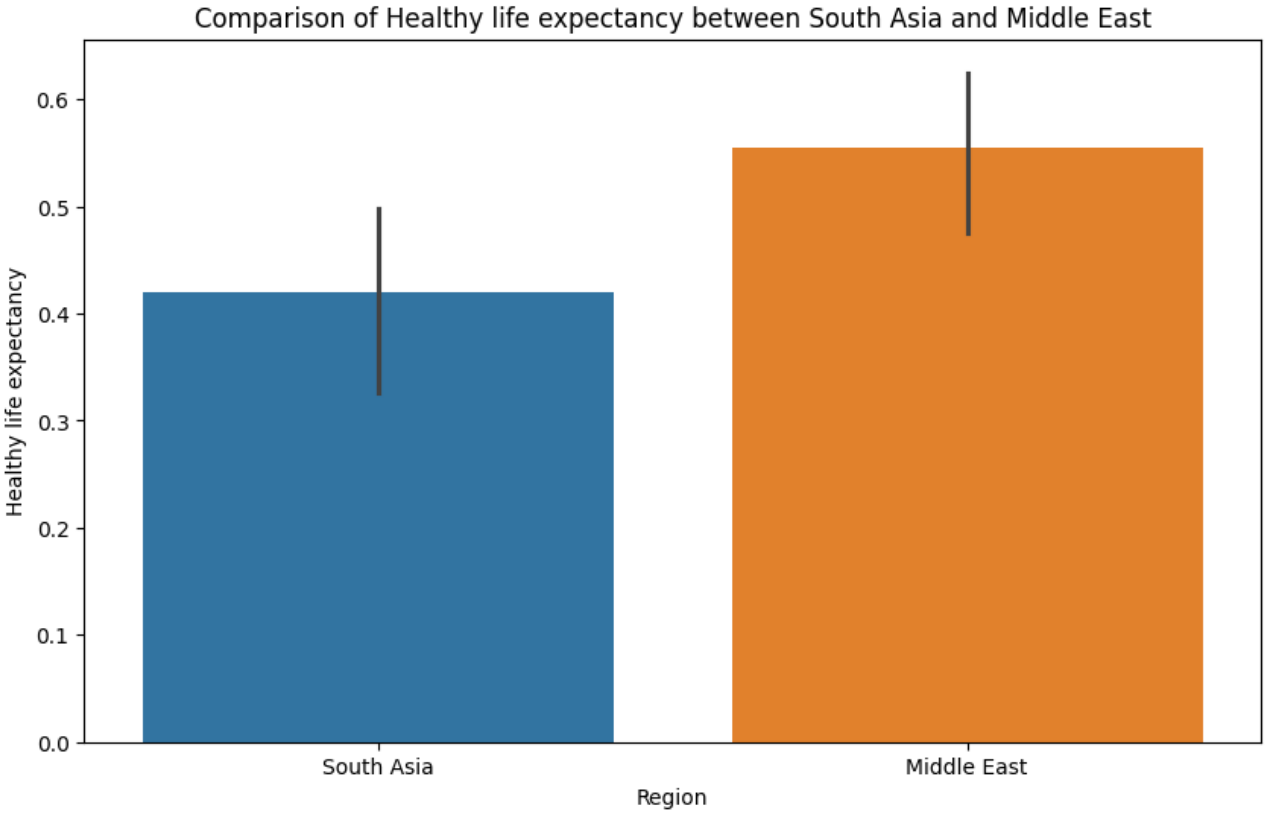
Fig(i)

4. Metric Comparisons:

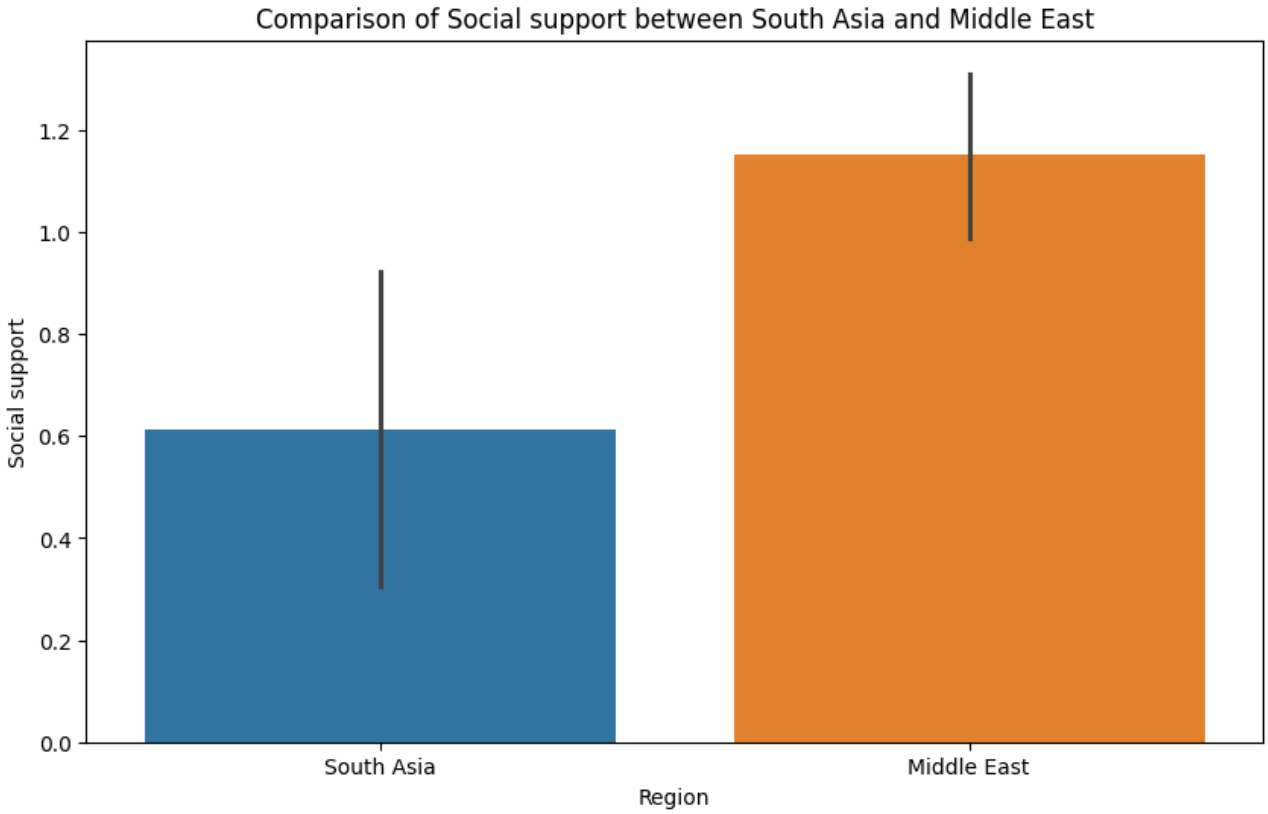
Compare metrics like GDP per Capita, Social Support, and Healthy Life Expectancy between both regions. Create grouped bar charts using **sns.barplot()** to show metric disparities in South Asia and the Middle East. Calculate disparities for each metric by comparing mean values across the two regions.


Comparison of Log GDP per capita between South Asia and Middle East
Log GDP per capita
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
South Asia
Middle East
Region

Fig(m)


Comparison of Healthy life expectancy between South Asia and Middle East
Healthy life expectancy
0.6
0.5
0.4
0.3
0.2
0.1
0.0
South Asia
Middle East
Region

Fig(n)



Fig(o)

Results:

Disparity in Log GDP per capita: 0.44

Disparity in Social support: 0.54

Disparity in Healthy life expectancy: 0.13

5.Happiness Disparity:

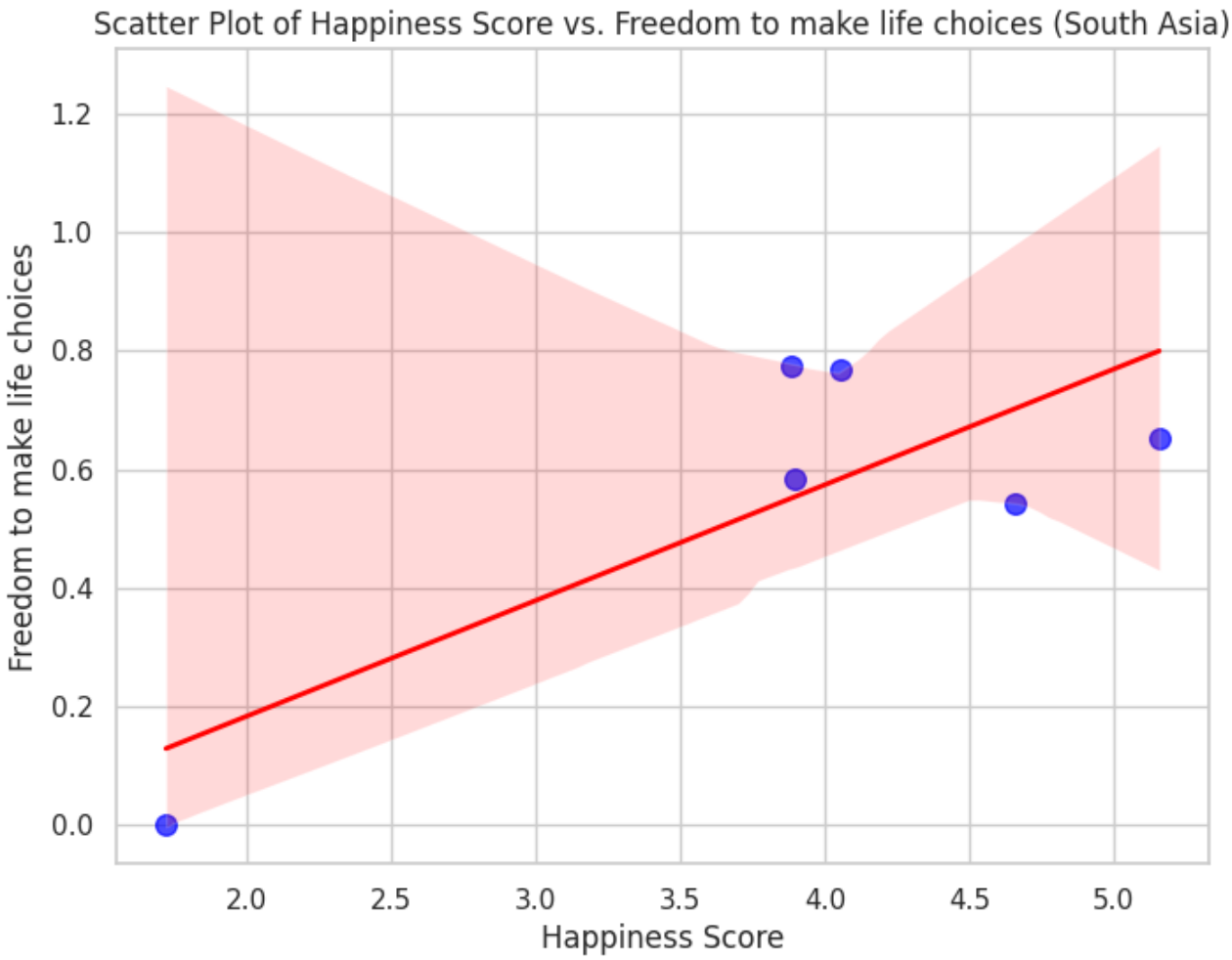
Calculate the range (max - min) of Happiness Scores for both region using **.max()** and **.min().** Calculate the CV as the ratio of standard deviation to the mean. Compare the variability of happiness scores between both regions to know which has greater disparity.

Results:

South Asia has greater variability in happiness.

6.Correlation Analysis

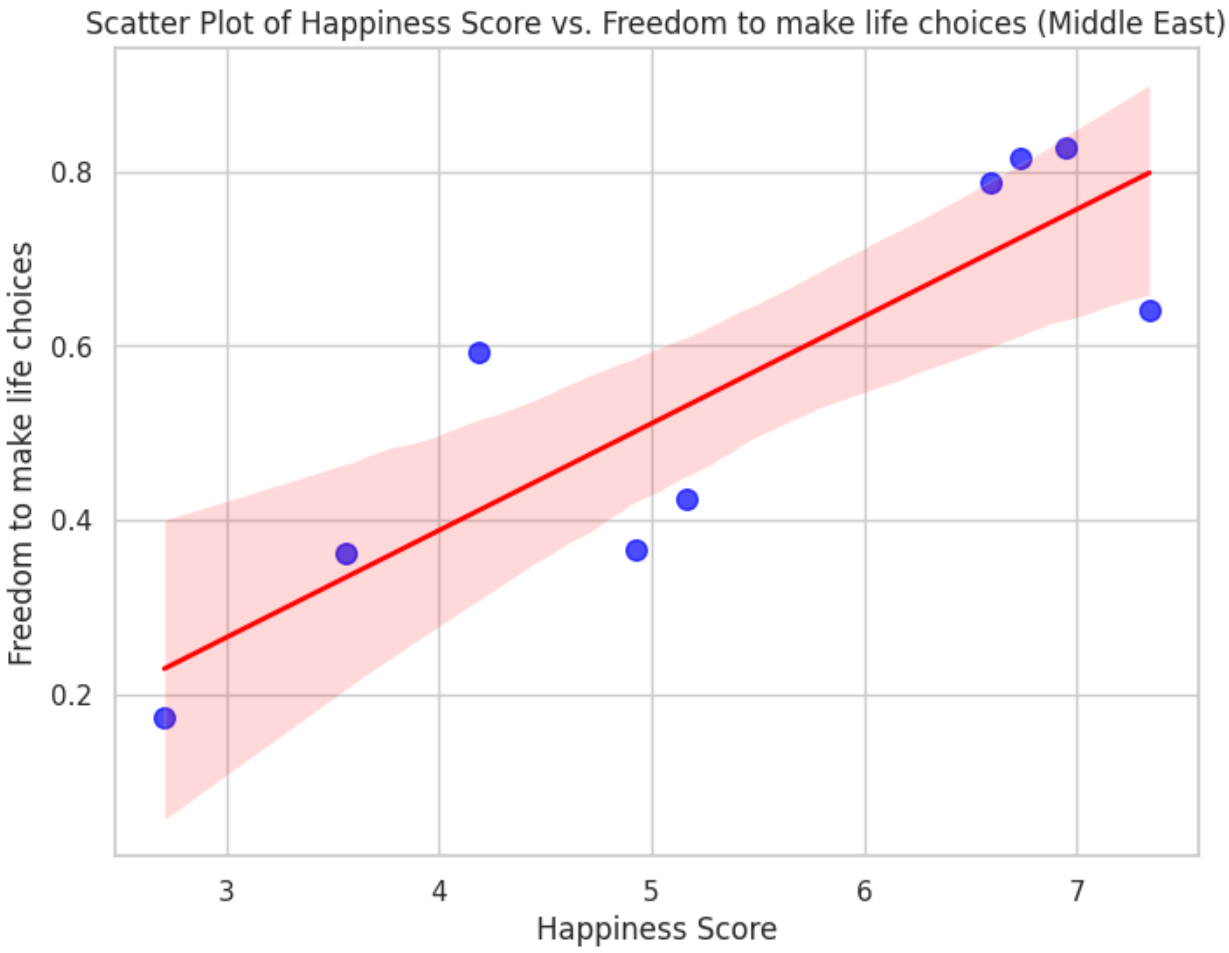
Analyze correlations between the Happiness Score and Freedom to Make Life Choices and Generosity using **.corr(method='pearson').** Create scatter plots using **sns.regplot()** to visualize the relationship, highlighting trends and variability.


Scatter Plot of Happiness Score vs. Freedom to make life choices (South Asia)
Freedom to make life choices
Happiness Score
0.0
0.2
0.4
0.6
0.8
1.0
1.2
2.0
2.5
3.0
3.5
4.0
4.5
5.0

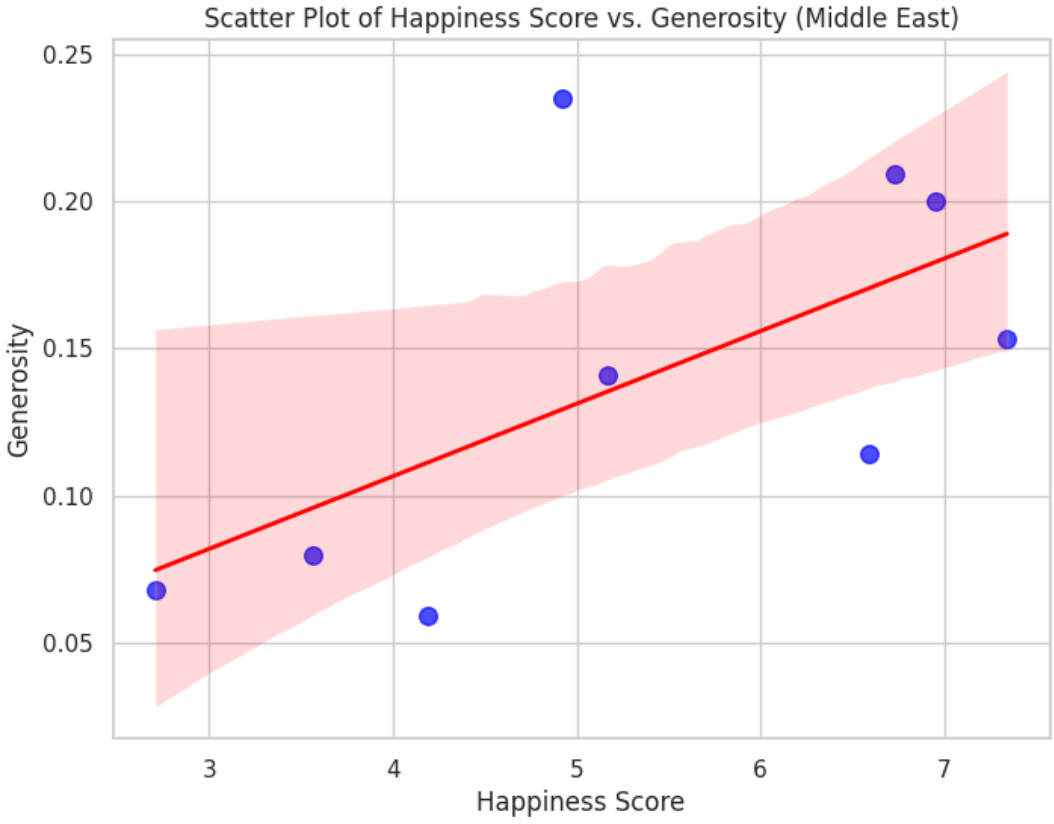
Fig(p)


Scatter Plot of Happiness Score vs. Generosity (South Asia)
Generosity
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Happiness Score

Fig(q)


Scatter Plot of Happiness Score vs. Freedom to make life choices (Middle East)
Freedom to make life choices
0.2
0.4
0.6
0.8
3
4
5
6
7
Happiness Score

Fig(r)


Scatter Plot of Happiness Score vs. Generosity (Middle East)
Generosity
0.05
0.10
0.15
0.20
0.25
3
4
5
6
7
Happiness Score

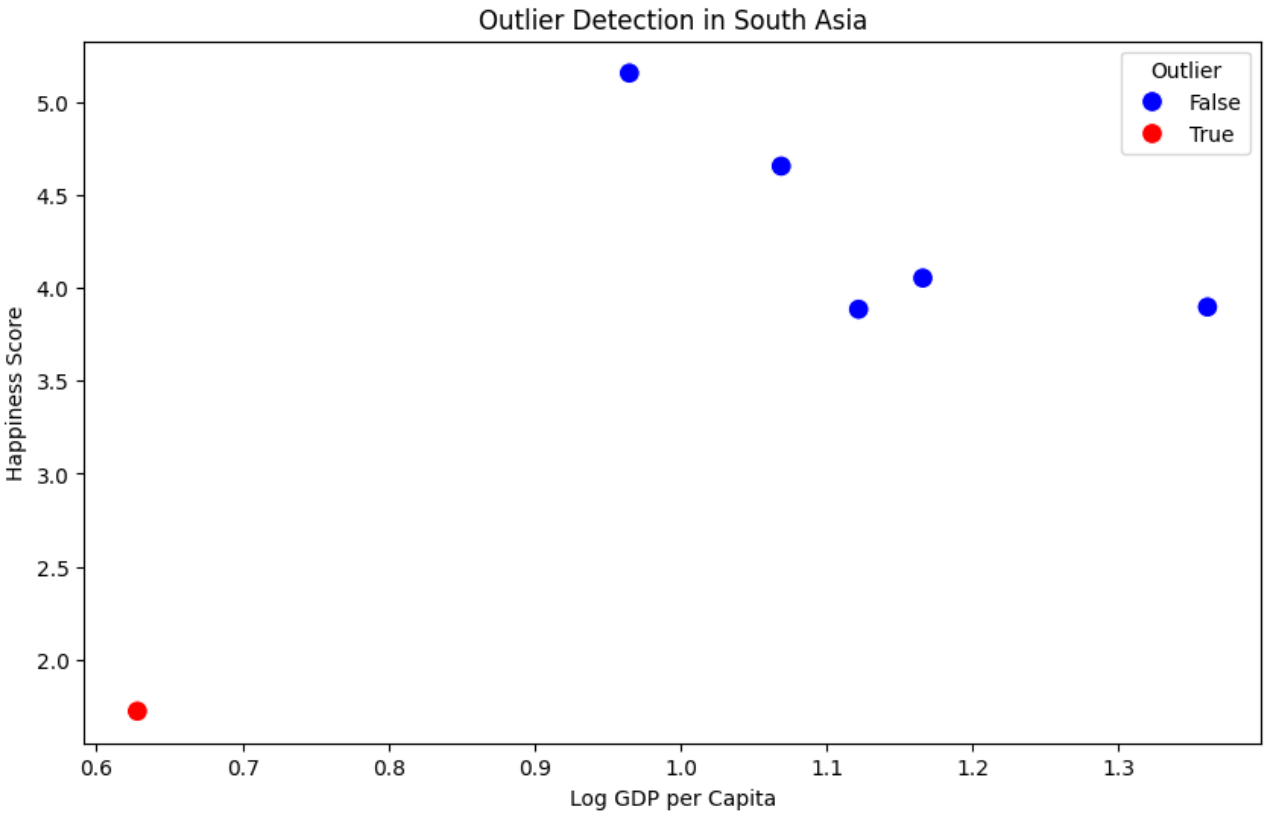
Fig(s)

7.Outlier Detection:

Calculate the IQR for Happiness Score and GDP per Capita for each region. Identify outliers based on the 1.5 * IQR rule. Use **sns.scatterplot()** to create scatter plots.

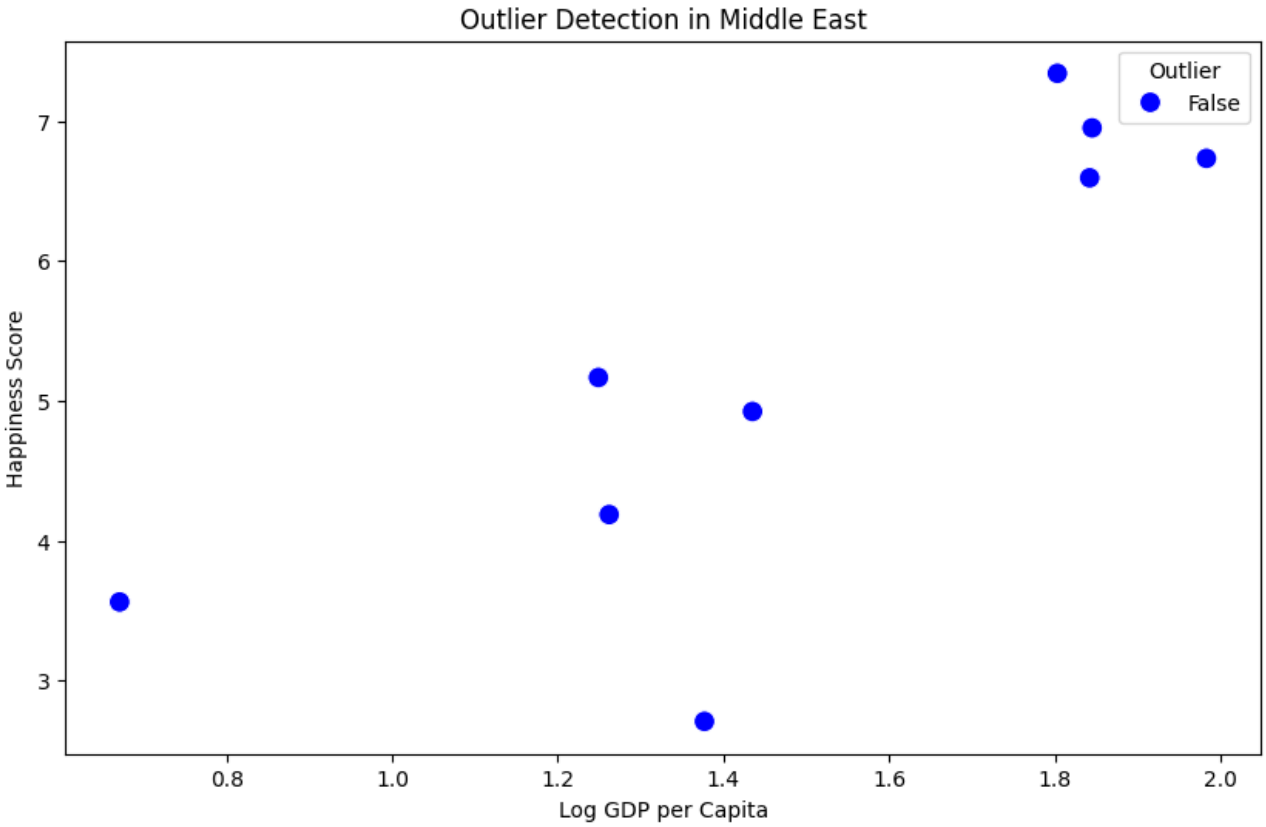
Implications:

Outlier Countries in South Asia is Afghanistan with a Score of 1.721 and Log GPD per Capita of 0.628.It is due to low Happiness Score and GDP per Capita compared to other countries in the region.It indicates that Afghanistan has severe socio economic challenges may be due to conflict instability and lack of infrastructure. It means there is need for economic development and proper governance and social support systems



Fig(t)

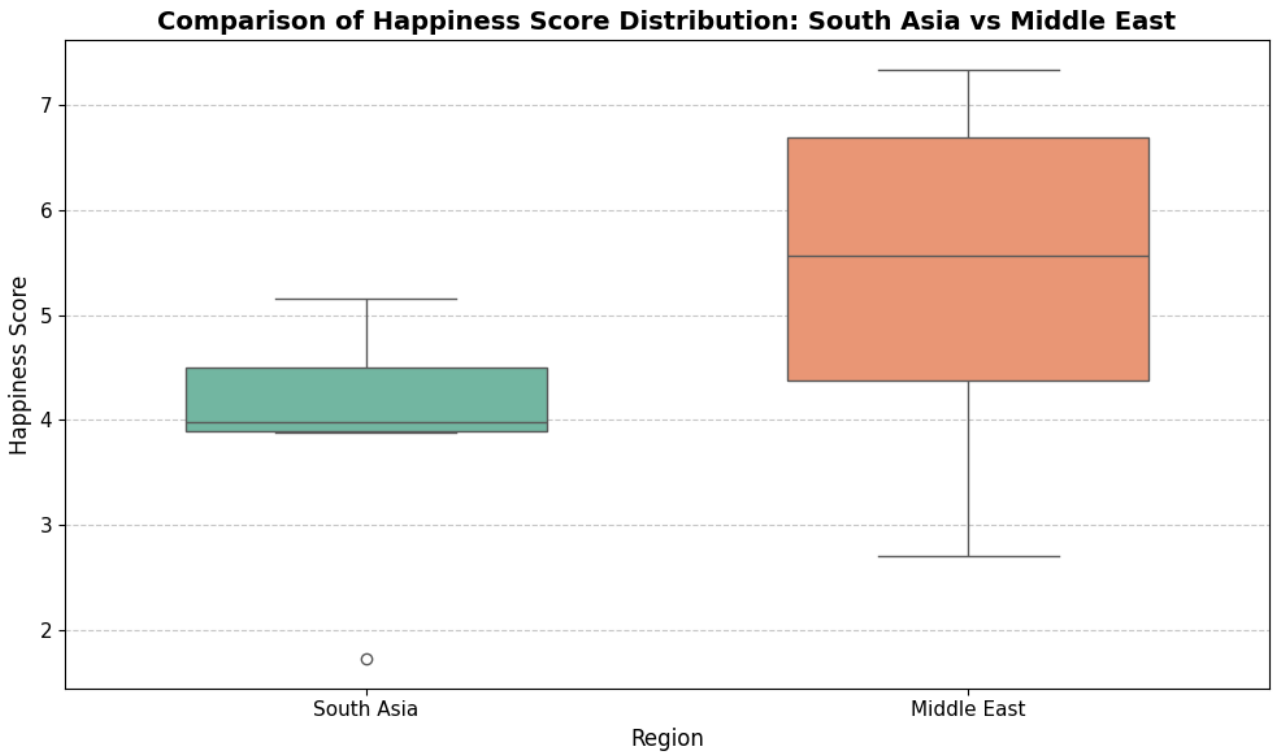
No Outliers was found in the Middle East which means that most countries have a balanced relationship between Happiness Score and GDP per Capita.



Fig(u)

8.Comparison of Score Distribution:

Combine the South Asia and Middle East datasets with a new column, Region, indicating the respective region.Create boxplots using **sns.boxplot()** to compare the distribution of Happiness Scores.



Fig(v)

Interpretation:

Median Comparison:

South Asia has a lower median is approximately 4 compared to the Middle East which is higher and between 5.5 and 6. Middle east report greater happiness compared to South Asia.

Distribution Shape:

South Asia shows low variability with tightly clustered scores, while the Middle East has a wider spread.

Outliers:

South Asia has a lower outlier of 1.7 and the Middle East has no visible outliers.

Conclusion

In problem 1, The Happiness score range between 2 and 8.The most frequent scores are around 6 and the highest frequency is at 18 which means many countries fall in this range. The distribution is slightly left skewed which means there are fewer countries with very high scores than lower scores. Outliers can be seen near 2 and close to 8.

In problem 2, The rankings based on the Composite Score do not completely match with the Original Happiness Score. The strongest Relationship is between Happiness Score and Freedom to Make Life Choices which indicates impact of personal freedom on happiness. The weakest Relationship is between Happiness Score and Generosity. Freedom to make life choices has a clear positive trend with happiness. Generosity shows a weaker positive trend with more scattered points. Countries like Sri Lanka, Bangladesh, and Afghanistan shows lower happiness scores than expected for their GDP.

In problem 3, It shows that South Asia has lower happiness scores than Middle East. Afghanistan is an outlier in South Asia with very low happiness and GDP which is caused to economic, government and social issues. The Middle East doesn’t have any outliers. The data also indicates that personal freedom has a strong link to happiness whereas generosity has less of an impact. In some cases, South Asia is happier compared to expected based on their GDP indicating that community support and culture plays important role in the happiness of the people. Overall, this suggests that social support and equity are important elements for improving well-being.