# Herald College, Kathmandu


## Concepts and Technologies of AI

### 5CS037

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

Name: Sanjiwani Karki
Group: '10'
Student ID: 2418323
Canvas ID: **S.Karki48@wlv.ac.uk**

# Acknowledgement

This report would not have been possible without the guidance and resources provided by Herald College, Kathmandu, and the invaluable dataset shared for the purpose of this analysis. My sincere thanks to my instructors who provided feedback and support throughout this assignment.

# Contents

## Introduction

The World Happiness Report gives important insights into how happy people are in different countries by looking at happiness scores and factors like income, social support, life expectancy, freedom, and generosity. This helps to understand how these factors impact global happiness and aids governments and organizations in improving the quality of life for their citizens.

This report will analyze the World Happiness Report dataset, focusing on three main goals:

**Data Exploration**: This section will examine the dataset, looking at basic statistics and visualizations to understand the global distribution of happiness scores.

**South Asia Analysis**: This part will focus on South Asian countries, analyzing their happiness scores and the factors contributing to them. It will also create composite scores and identify outliers to gather regional insights.

**South Asia vs. Middle East Comparison**: The final section will compare happiness scores and related metrics between South Asia and the Middle East, identifying trends, differences, and correlations like the relationship between GDP per capita and social support.

Using data analysis, visualizations, and comparisons, this report will offer insights into global and regional happiness trends, helping to understand the factors that contribute to happiness in different regions.

# 3.1 Problem 1
## Data Exploration and Understanding:

### Dataset Overview:

First 10 rows of the dataset:
The dataset has 143 rows and 9 columns.
Columns and their data types:

| | 0 |
|---|---|
| **Country name** | object |
| **score** | float64 |
| **Log GDP per capita** | float64 |
| **Social support** | float64 |
| **Healthy life expectancy** | float64 |
| **Freedom to make life choices** | float64 |
| **Generosity** | float64 |
| **Perceptions of corruption** | float64 |
| **Dystopia + residual** | float64 |

**dtype:** object

The dataset contains 143 rows and 9 columns, including factors like GDP, social support, life expectancy, and freedom, which influence the happiness score for each country. It provides insights into the global well-being based on these key metrics.

- Missing Values:

```
Mean score: 5.52758041958042, Median score: 5.785, Standard Deviation: 1.1707165099442995
country with the highest score: Finland
country with the lowest score:Afghanistan
```

The mean happiness score is 5.53, with a median of 5.79 and a standard deviation of 1.17. Finland has the highest happiness score, while Afghanistan has the lowest.

- Missing Values:

```
Missing values:
 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

After handling:
 Country name                      0
score                              0
Log GDP per capita                 0
Social support                     0
Healthy life expectancy            0
Freedom to make life choices       0
Generosity                         0
Perceptions of corruption          0
Dystopia + residual                0
dtype: int64
```

Initially, the dataset had missing values in the columns "Log GDP per capita," "Social support," "Healthy life expectancy," "Freedom to make life choices," "Generosity," "Perceptions of corruption," and "Dystopia + residual," each with 3 missing entries. After handling the missing values, all columns are now complete with no missing data.

Filtering and Sorting:

| | Country name | score | Log GDP per capita | Social support | Healthy life expectancy | Freedom to make life choices | Generosity | Perceptions of corruption | Dystopia + residual |
|---|---|---|---|---|---|---|---|---|---|
| **7** | Luxembourg | 7.122 | 2.141 | 1.355 | 0.708 | 0.801 | 0.146 | 0.432 | 1.540 |
| **16** | Ireland | 6.838 | 2.129 | 1.390 | 0.700 | 0.758 | 0.205 | 0.418 | 1.239 |
| **29** | Singapore | 6.523 | 2.118 | 1.361 | 0.769 | 0.743 | 0.168 | 0.575 | 0.788 |
| **21** | United Arab Emirates | 6.733 | 1.983 | 1.164 | 0.563 | 0.815 | 0.209 | 0.258 | 1.741 |
| **8** | Switzerland | 7.060 | 1.970 | 1.425 | 0.747 | 0.759 | 0.173 | 0.498 | 1.488 |
| **6** | Norway | 7.302 | 1.952 | 1.517 | 0.704 | 0.835 | 0.224 | 0.484 | 1.586 |
| **22** | United States | 6.725 | 1.939 | 1.392 | 0.542 | 0.586 | 0.223 | 0.169 | 1.873 |
| **85** | Hong Kong S.A.R. of China | 5.316 | 1.909 | 1.184 | 0.857 | 0.485 | 0.147 | 0.402 | 0.333 |
| **1** | Denmark | 7.583 | 1.908 | 1.520 | 0.699 | 0.823 | 0.204 | 0.548 | 1.881 |
| **5** | Netherlands | 7.319 | 1.901 | 1.462 | 0.706 | 0.725 | 0.247 | 0.372 | 1.906 |

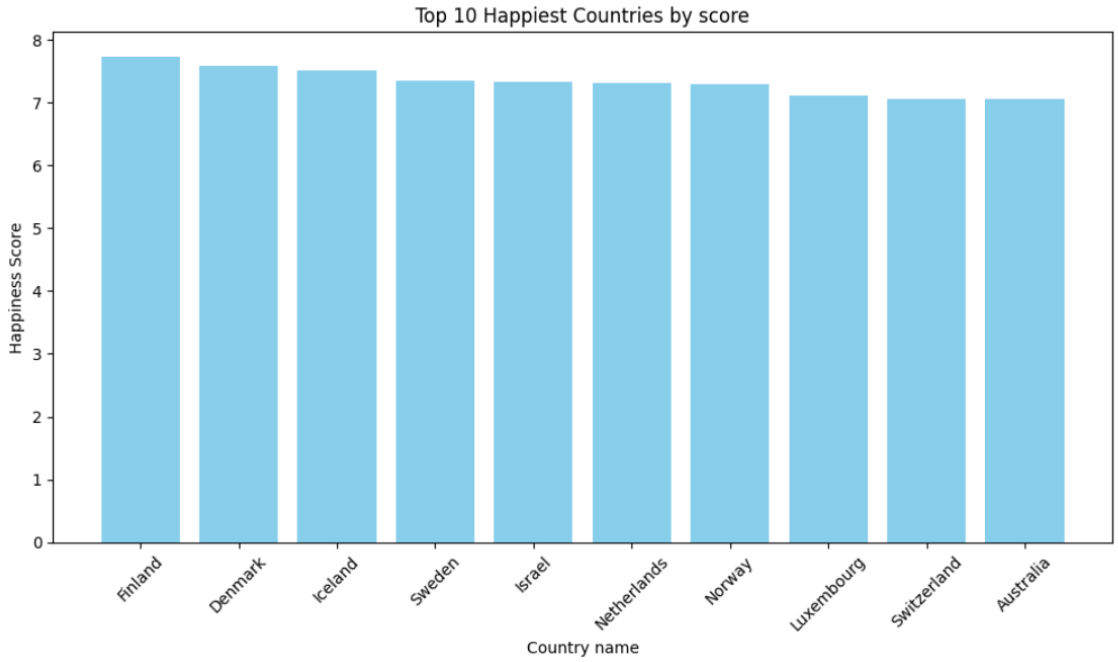
Activate Windows

Here are 10 countries from the dataset with their happiness scores and related metrics, including economic and social factors like GDP per capita, social support, and freedom to make life choices. Countries such as Denmark, Norway, and Switzerland rank high in happiness due to strong social and economic indicators, while others like Hong Kong have lower scores influenced by varying metrics.

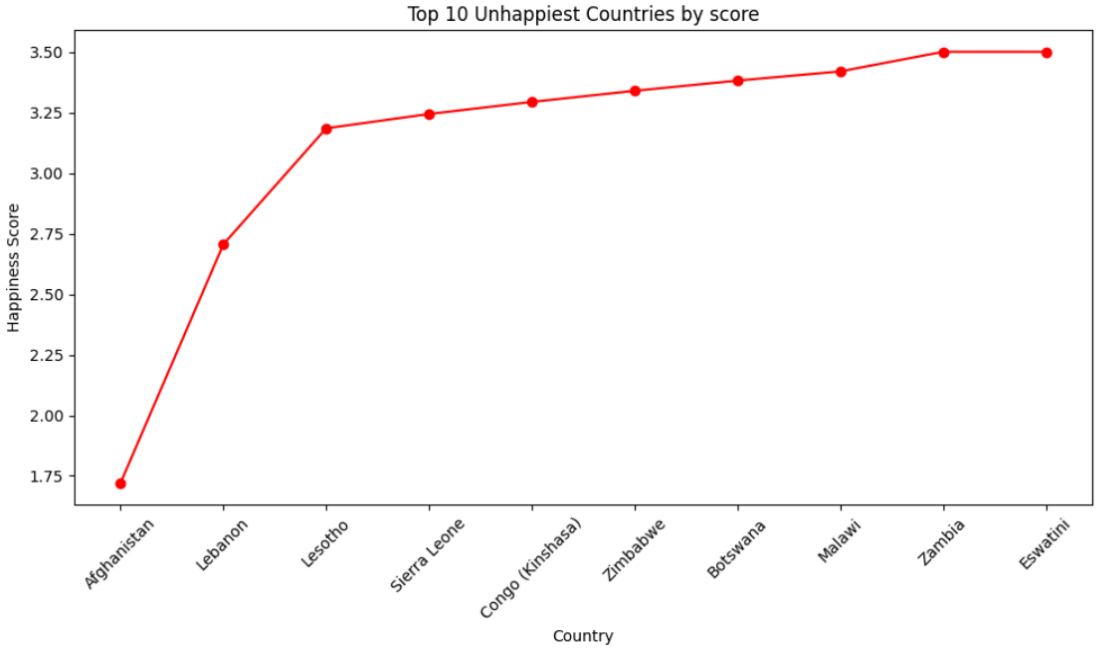
Adding New Columns:

```
          Country name happiness category
0              Finland               High
1              Denmark               High
2              Iceland               High
3               Sweden               High
4               Israel               High
..                 ...                ...
138  Congo (Kinshasa)                Low
139      Sierra Leone                Low
140           Lesotho                Low
141           Lebanon                Low
142       Afghanistan                Low

[143 rows x 2 columns]
```

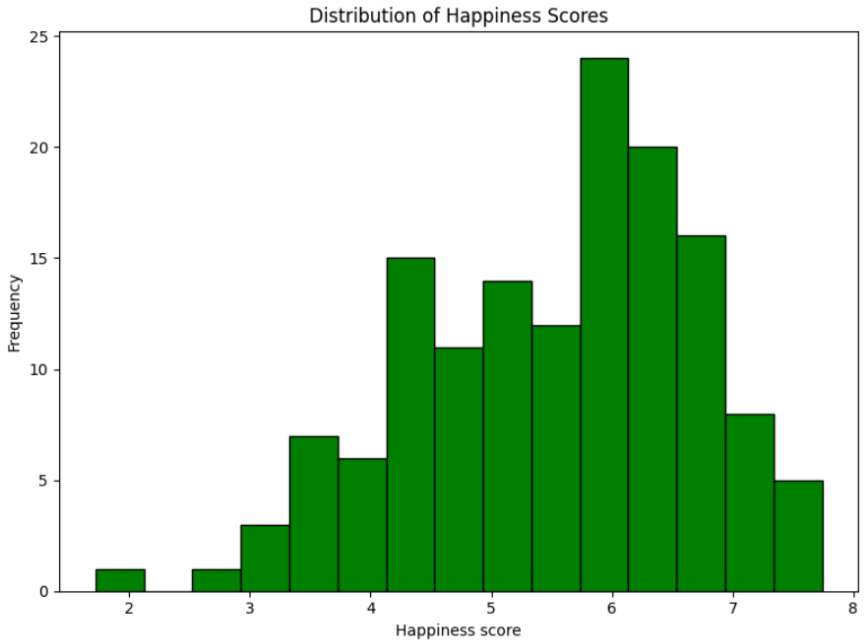
The dataset categorizes countries into three happiness categories based on their scores: **High**, **Medium**, and **Low**. Countries like Finland, Denmark, and Iceland fall into the **High** category, while countries such as Afghanistan, Lebanon, and Sierra Leone are in the **Low** category. This classification helps in understanding global happiness distribution.



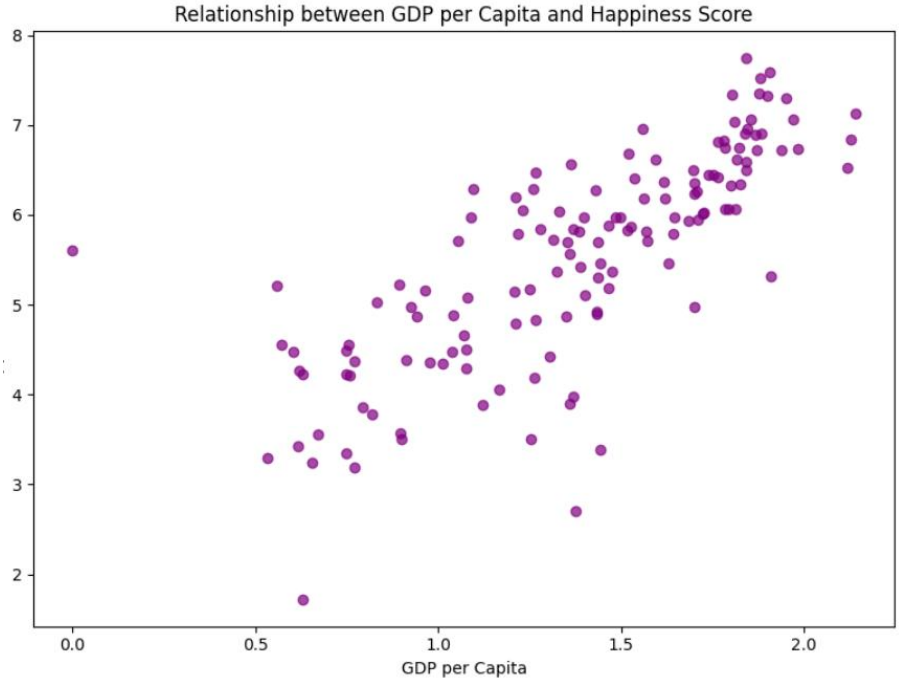

The output is a bar plot showing the top 10 happiest countries ranked by their happiness scores. The x-axis lists the countries, the y-axis shows their scores, and the bar height reflects the score. Taller bars indicate higher happiness, with the highest bar representing the happiest country.




The line plot shows the 10 countries with the lowest happiness scores, with markers highlighting their scores. The red line emphasizes their low happiness levels for easy comparison.




The histogram visualizes the distribution of happiness scores across countries. It groups scores into 15 bins, showing how frequently scores fall within specific ranges. Most scores are concentrated in the middle, reflecting a normal-like distribution.




The scatter plot illustrates the relationship between GDP per capita and happiness scores. It shows a positive trend, indicating that countries with higher GDP per capita generally tend to have higher happiness scores. However, there are exceptions where some countries with moderate GDP still achieve notable happiness levels.

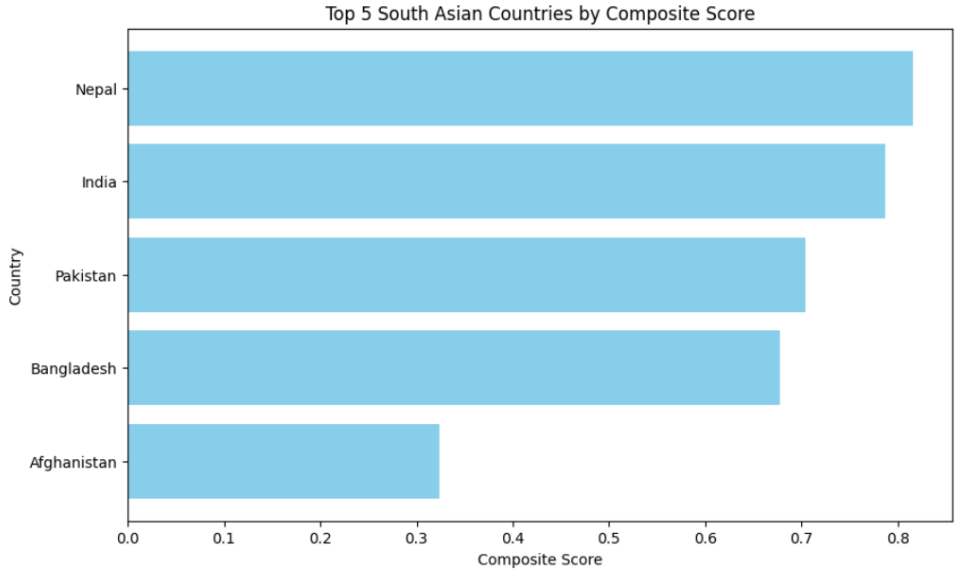
## 3.2 Problem-2

Setup Task - Preparing the South-Asia Dataset:

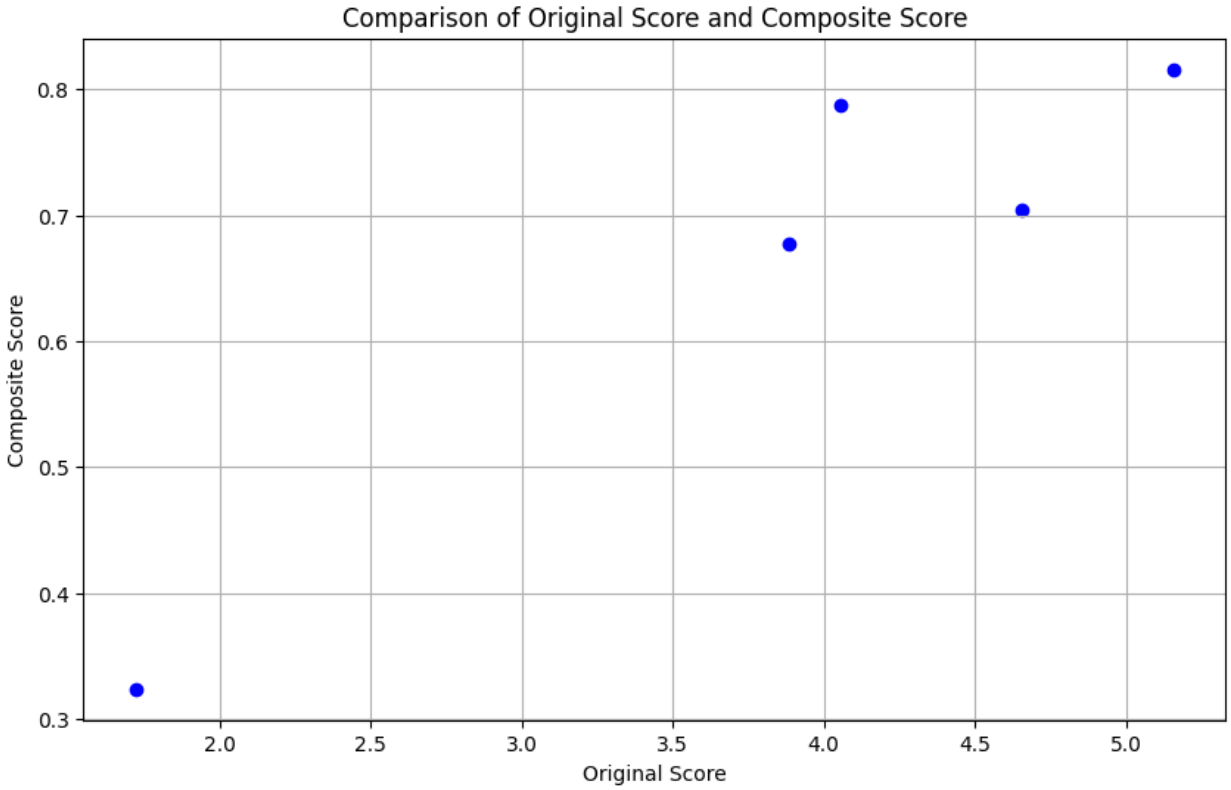
```
South Asia dataset saved as 'South_Asia_Happiness.csv'
```

The code filters the dataset to include only South Asian countries and saves it as a new CSV file, South_Asia_Happiness.csv.

Composite Score Ranking:



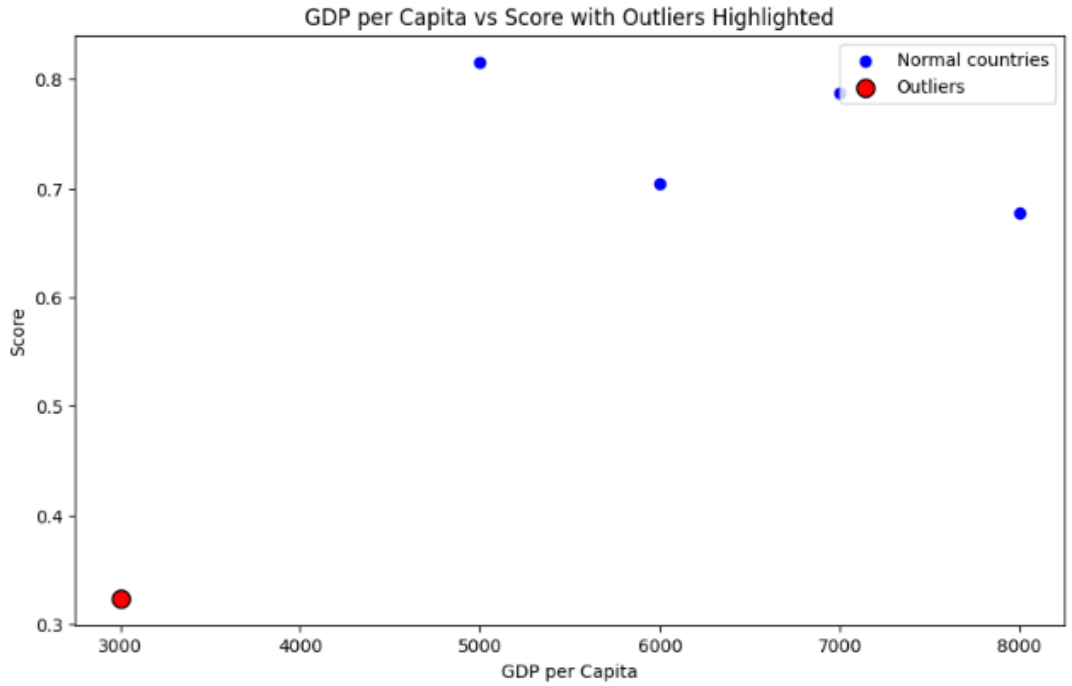

The code calculates a composite happiness score for five South Asian countries using GDP, social support, and life expectancy. It ranks them, visualizes the top 5 in a bar chart, and compares the original happiness score with the composite score using a scatter plot. The results are displayed in a Data Frame.




| | Country name | score | Composite Score |
|---|---|---|---|
| 0 | Nepal | 5.158 | 0.8159 |
| 1 | India | 4.054 | 0.7874 |
| 2 | Pakistan | 4.657 | 0.7039 |
| 3 | Bangladesh | 3.886 | 0.6774 |
| 4 | Afghanistan | 1.721 | 0.3238 |

The code calculates a composite happiness score for five South Asian countries using GDP, social support, and life expectancy. It ranks them, visualizes the top 5 in a bar chart, and compares the original happiness score with the composite score using a scatter plot. The results are displayed in a Data Frame.

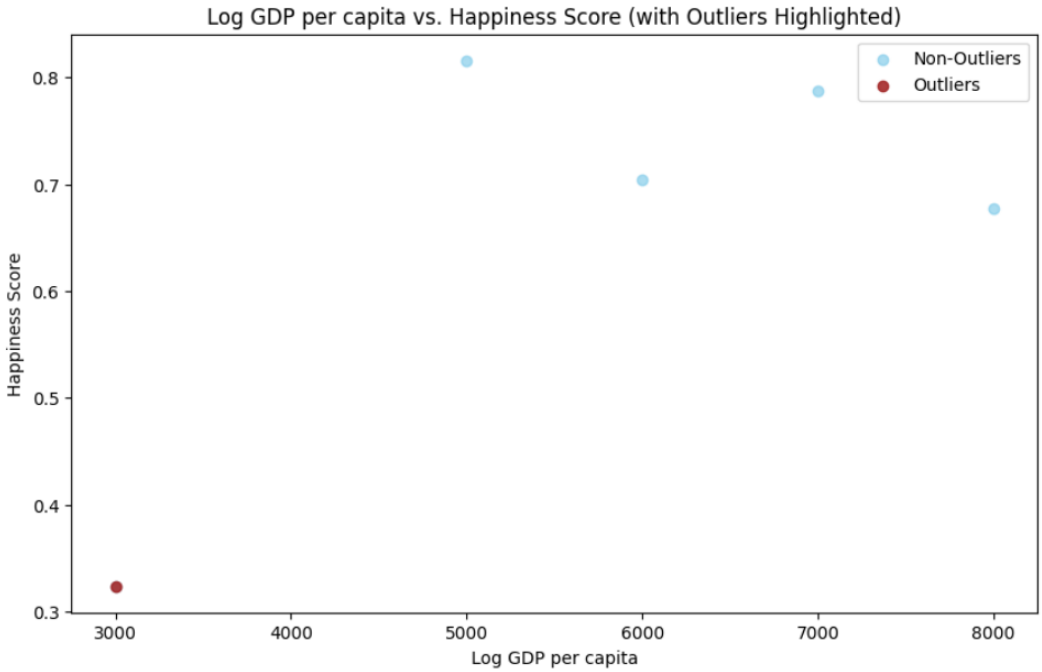
```
Debugging Score:
Q1 (25th percentile): 0.6774
Q3 (75th percentile): 0.7874
IQR (Interquartile Range): 0.10999999999999999
Lower bound: 0.5124
Upper bound: 0.9523999999999999
Debugging GDP per Capita:
Q1 (25th percentile): 5000.0
Q3 (75th percentile): 7000.0
IQR (Interquartile Range): 2000.0
Lower bound: 2000.0
Upper bound: 10000.0

Outliers based on Score:
  Country name   Score
4  Afghanistan  0.3238

Outliers based on GDP per Capita:
Empty DataFrame
Columns: [Country name, GDP per Capita]
Index: []
```




```
Outliers based on Score or GDP per Capita:
  Country name  GDP per Capita   Score
4  Afghanistan            3000  0.3238
```

The code detects outliers in **Score** and **GDP per Capita** using the **1.5 * IQR rule** and visualizes them on a scatter plot. Outliers are highlighted in **red**, and normal data points are in **blue**. It then prints the countries identified as outliers and shows the relationship between GDP and happiness score.

```
Identified outlier countries:
  Country name   score  Log GDP per capita
4  Afghanistan  0.3238                3000
```



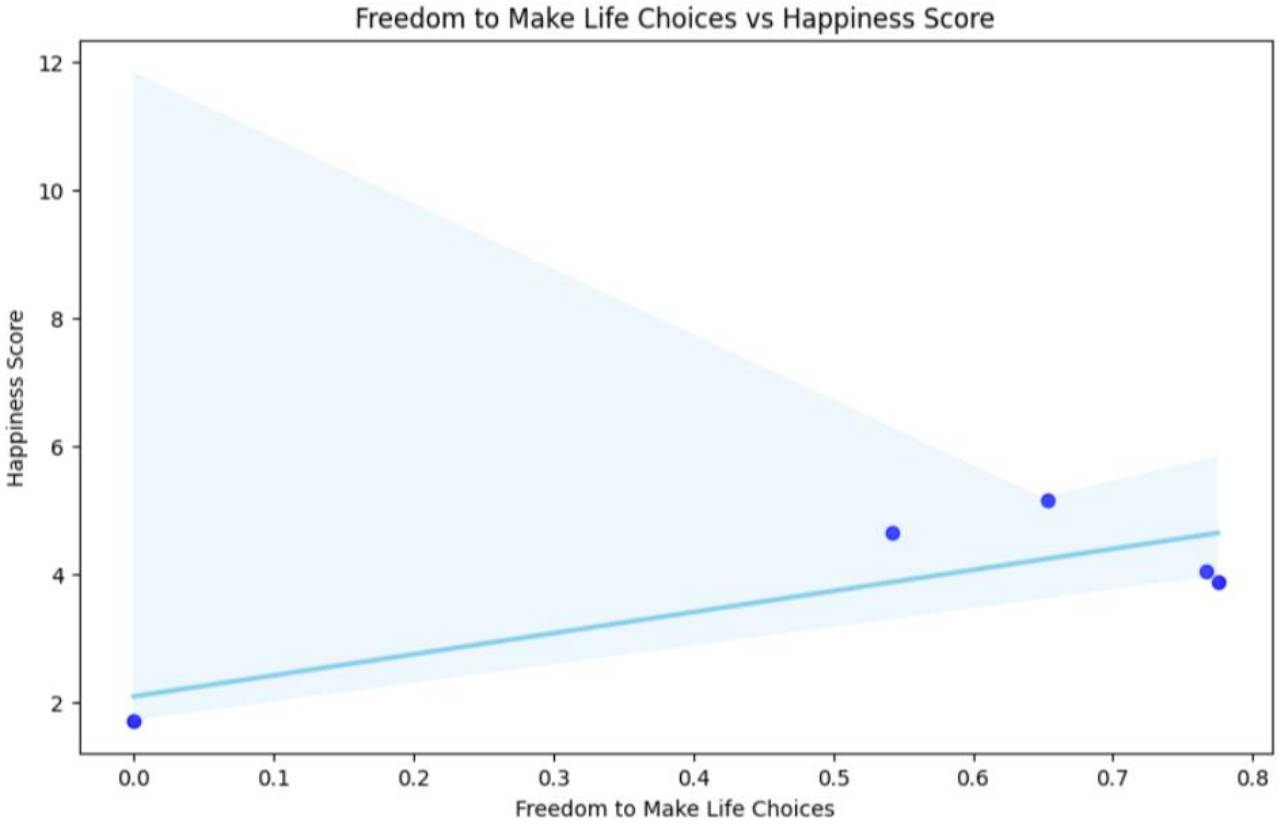

```
Outliers with extreme GDP or Happiness Scores can skew regional averages and impact overall interpretations.
```

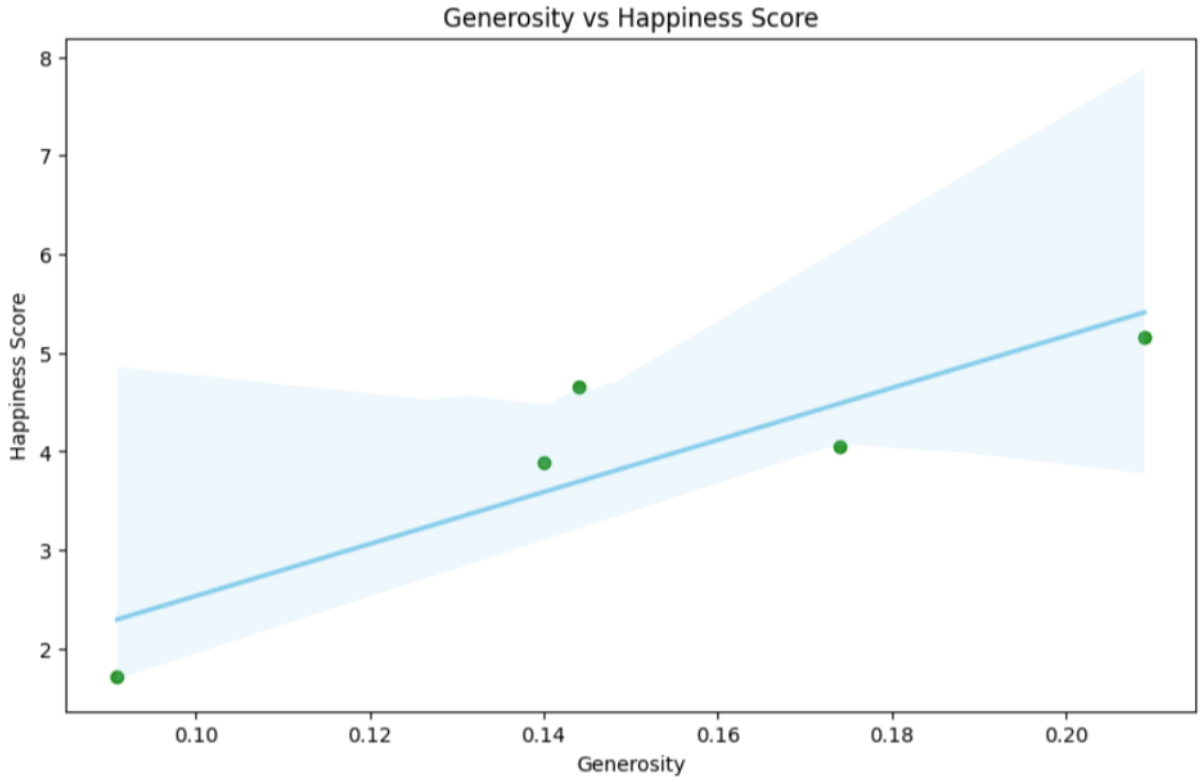
The code detects outliers in happiness scores and GDP per capita using the 1.5 * IQR rule, then highlights them in a scatter plot. Outliers are marked in brown, while normal data points are in sky blue. The identified outliers are printed, and the plot shows how extreme values can affect data interpretation.

Outliers in the dataset, such as countries with extreme happiness scores or GDP per capita, can skew regional averages, making the overall picture misleading. These outliers may represent unique circumstances like economic instability or high wealth. Ignoring them could lead to incorrect conclusions about the region's general well-being, as they distort average values, masking the true regional trends. Identifying and understanding outliers helps in drawing more accurate insights.

**4.** Exploring Trends Across Metrics:

```
Freedom to make life choices    0
Generosity                      0
score                           0
dtype: int64
Pearson correlation between 'Freedom to Make Life Choices' and 'Score': 0.8015
Pearson correlation between 'Generosity' and 'Score': 0.8773
```

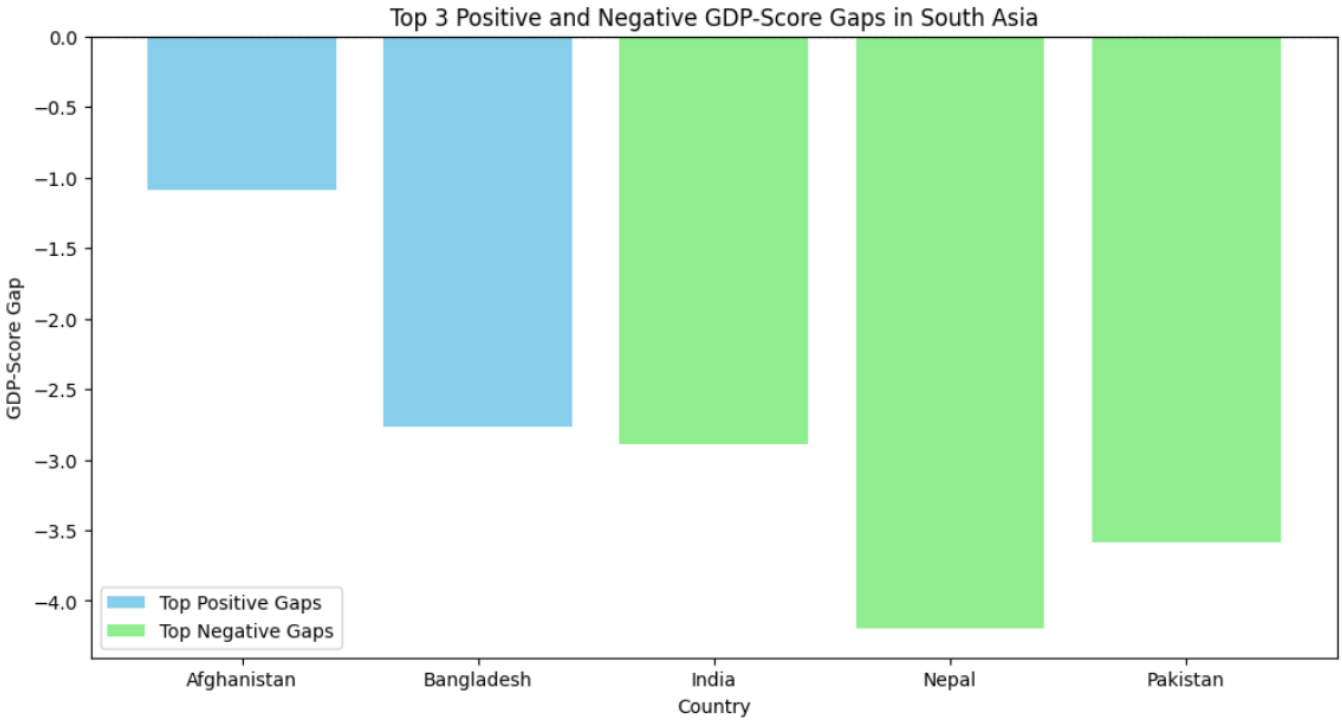





```
The strongest correlation is between 'Generosity' and 'Score'.
The weakest correlation is between 'Freedom to Make Life Choices' and 'Score'.
```

The code analyzes the relationship between **freedom to make life choices**, **generosity**, and **happiness score** in South Asia. It checks for missing values, calculates Pearson correlations, and creates scatter plots with trend lines. It identifies which factor has the strongest and weakest correlation with happiness.

5. Gap Analysis:




The code calculates the **GDP-Score Gap** (difference between GDP per capita and happiness score) for South Asian countries. It then sorts countries by this gap and highlights the top 3 with the largest positive and negative gaps. A bar chart visualizes these gaps, with positive gaps in **sky blue** and negative gaps in **light green**, showing the disparity between GDP and happiness.

## 3.3 Problem - 3
### Comparative Analysis:

| | Country name | score | GDP per Capita | Social support | Healthy life expectancy |
|---|---|---|---|---|---|
| 0 | Bahrain | 6.1 | 1.25 | 0.87 | 0.75 |
| 1 | Iran | 5.7 | 1.05 | 0.75 | 0.71 |
| 2 | Iraq | 4.5 | 0.90 | 0.65 | 0.60 |
| 3 | Israel | 7.2 | 1.30 | 0.89 | 0.80 |
| 4 | Jordan | 5.9 | 1.10 | 0.82 | 0.77 |
| 5 | Kuwait | 6.3 | 1.20 | 0.77 | 0.79 |
| 6 | Lebanon | 5.0 | 1.00 | 0.70 | 0.70 |
| 7 | Oman | 6.1 | 1.00 | 0.75 | 0.74 |
| 8 | Palestine | 5.8 | 1.10 | 0.73 | 0.70 |
| 9 | Qatar | 7.3 | 1.40 | 0.90 | 0.82 |
| 10 | Saudi Arabia | 6.7 | 1.50 | 0.84 | 0.85 |
| 11 | Syria | 4.1 | 0.95 | 0.60 | 0.68 |
| 12 | United Arab Emirates | 7.8 | 1.60 | 0.85 | 0.88 |
| 13 | Yemen | 3.5 | 0.80 | 0.55 | 0.66 |

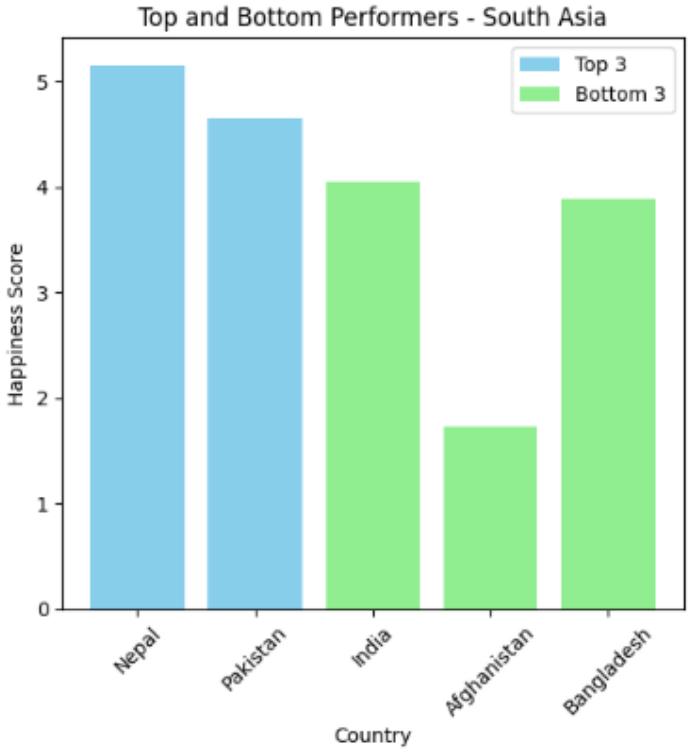
The code creates a **pandas** Data Frame with example data for 14 Middle Eastern countries, including their **score**, **GDP per capita**, **social support**, and **healthy life expectancy**. The data represents various metrics for each country, useful for analysis or visualization.

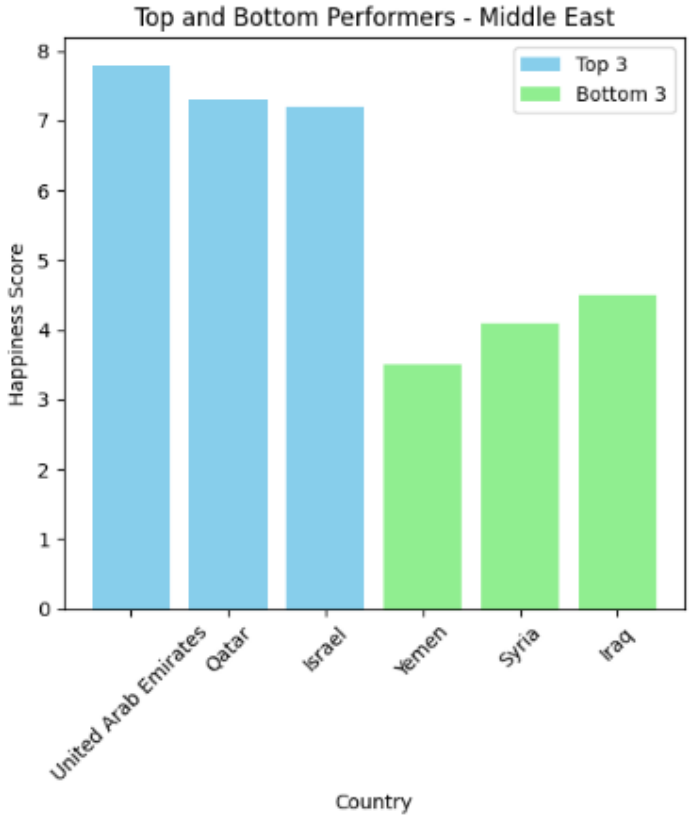
1. Descriptive Statistics:

```
South Asia - Mean Score: 3.90, Standard Deviation: 1.32
Middle East - Mean Score: 5.86, Standard Deviation: 1.24
Middle East has a higher average happiness score.
```

The code compares the average happiness scores of South Asia and the Middle East. It calculates the mean and standard deviation for both regions and then determines which region has the higher average score. The result is printed based on the comparison.

2.Top and Bottom Performers:

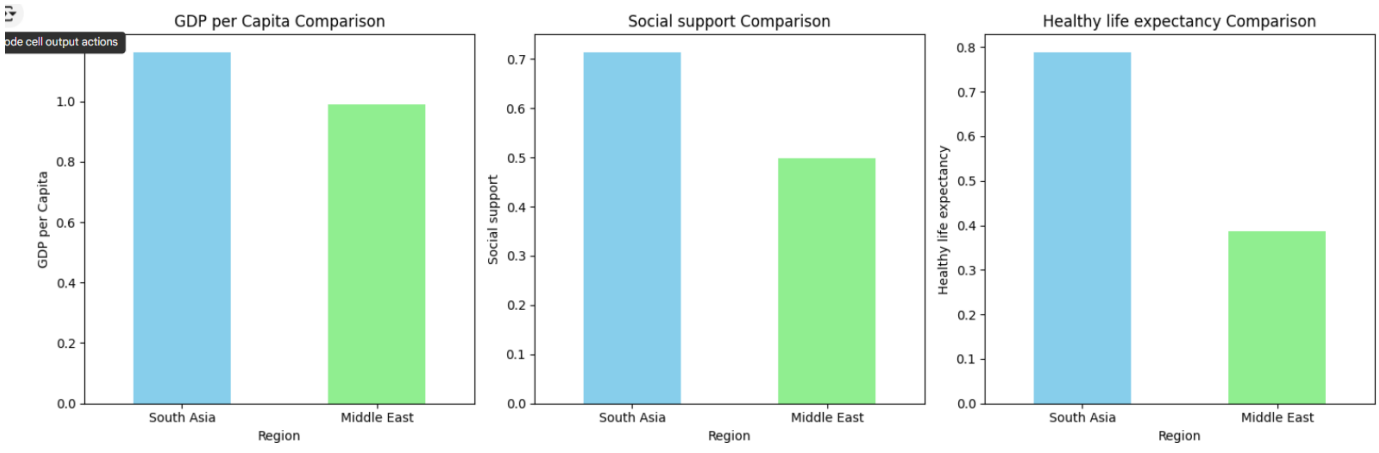



```
Top 3 Performers - South Asia:
  Country name  score
0        Nepal  5.158
1     Pakistan  4.657
2        India  4.054

Bottom 3 Performers - South Asia:
  Country name  score
4  Afghanistan  1.721
3   Bangladesh  3.886
2        India  4.054

Top 3 Performers - Middle East:
            Country name  score
12  United Arab Emirates    7.8
9                  Qatar    7.3
3                 Israel    7.2

Bottom 3 Performers - Middle East:
   Country name  score
13        Yemen    3.5
11        Syria    4.1
2          Iraq    4.5
```

The code compares happiness scores for South Asia and the Middle East by identifying and plotting the top 3 and bottom 3 performers from each region. It creates two bar charts to visualize the performance, then prints the top and bottom 3 countries for both regions.

## 3.Metric Comparisons:




| Region | GDP per Capita | Social support | Healthy life expectancy |
|---|---|---|---|
| Middle East | 1.162929 | 0.714286 | 0.789286 |
| South Asia | 0.990000 | 0.498400 | 0.387200 |

Activate Windows
Go to Settings to activate Windov

This code compares South Asia and the Middle East by plotting grouped bar charts for GDP per capita, social support, and healthy life expectancy. It calculates the average values for each metric in both regions and visualizes the differences. The results show the disparity between the regions based on these key indicators.

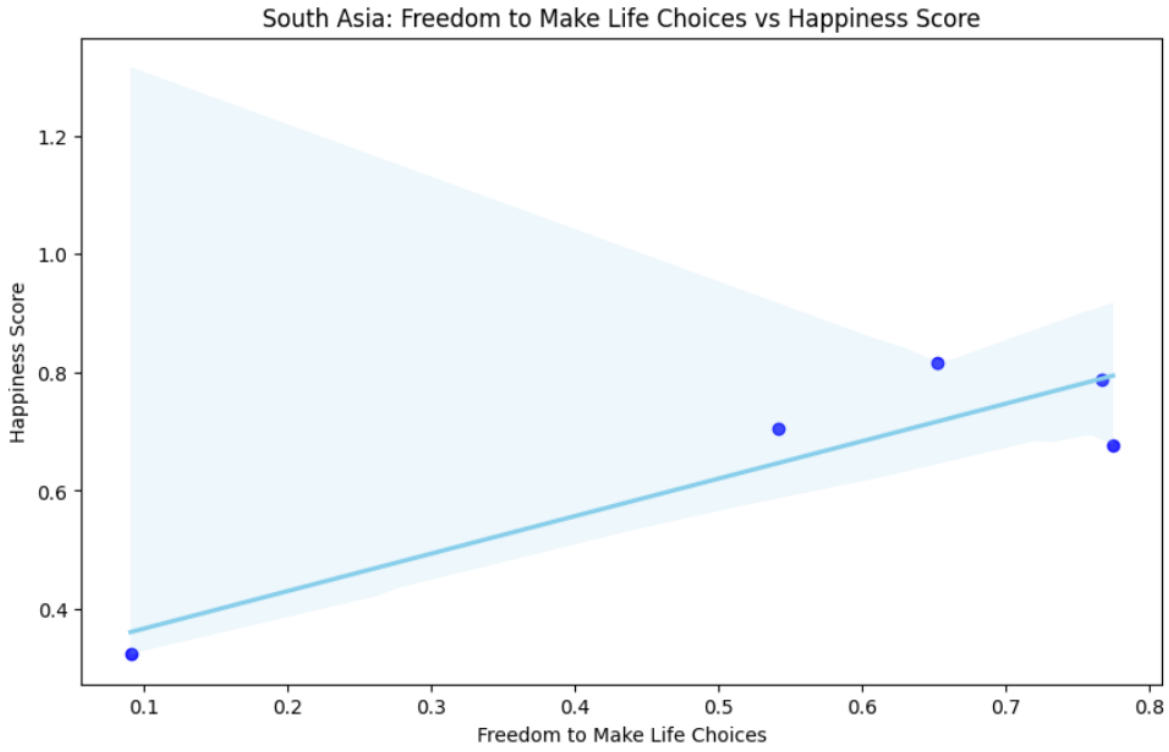
## 4.Happiness Disparity:

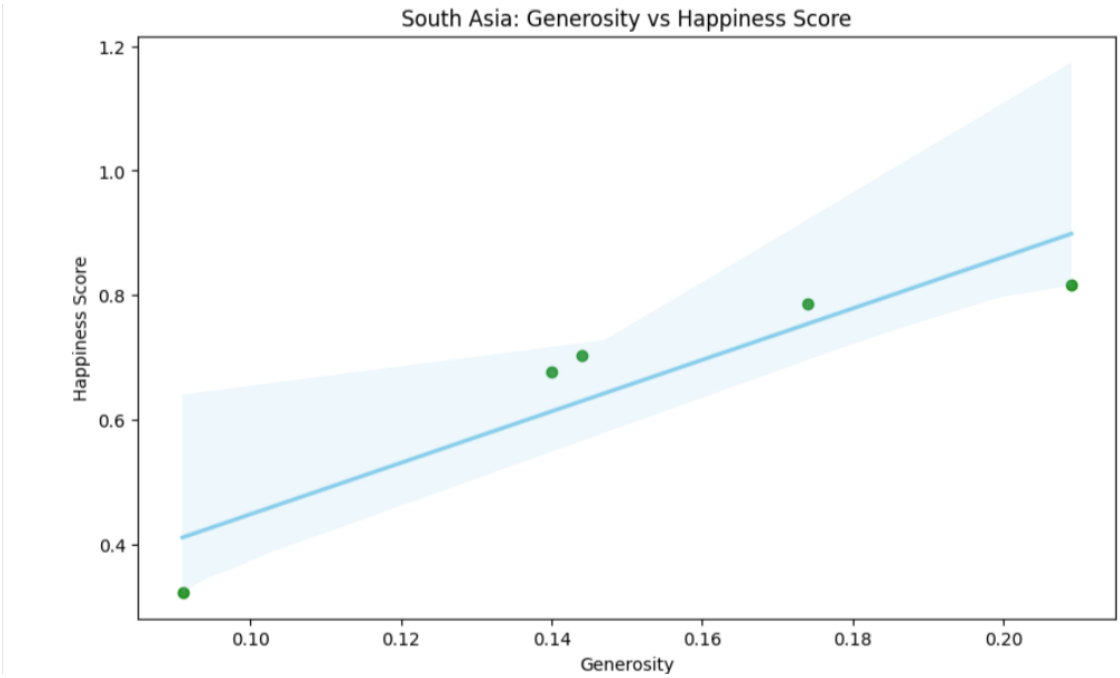
```
Range of happiness scores for South Asia: 0.4921
Coefficient of Variation (CV) for South Asia: 29.82%

Range of happiness scores for the Middle East: 0.565
Coefficient of Variation (CV) for the Middle East: 25.73%

South Asia has greater variability in happiness scores.
```

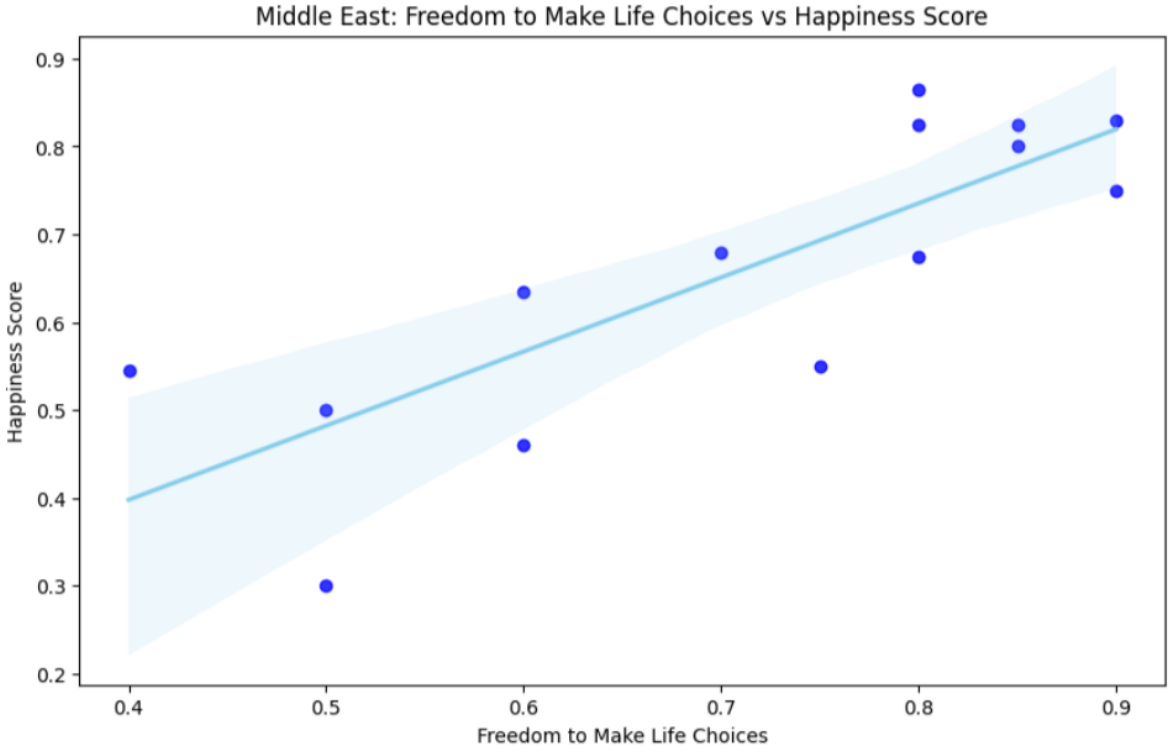
This code calculates the range and coefficient of variation (CV) for happiness scores in South Asia and the Middle East. It compares the variability of happiness scores between the two regions. The region with the higher CV has greater variability in its scores.

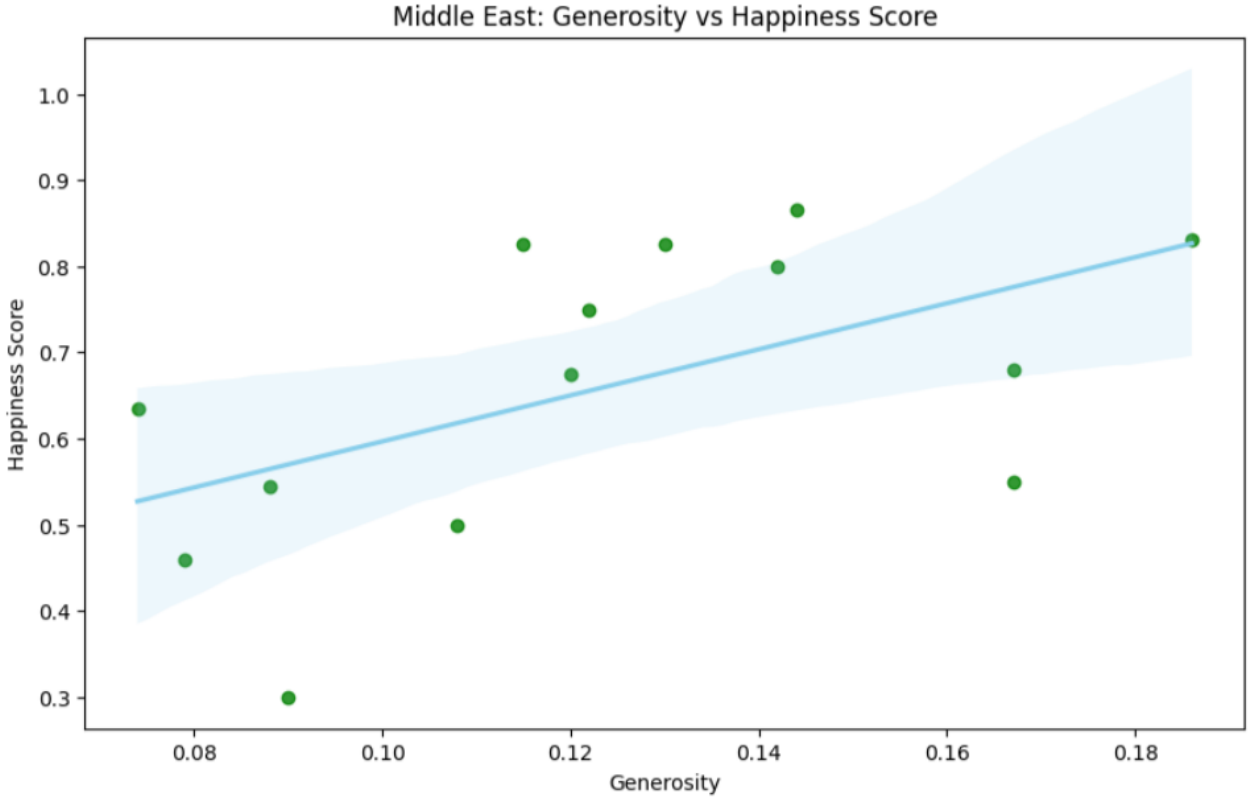
## 5.Correlation Analysis:

```
South Asia - Pearson correlation between 'Score' and 'Freedom to Make Life Choices': 0.9053
South Asia - Pearson correlation between 'Score' and 'Generosity': 0.9159
Middle East - Pearson correlation between 'Score' and 'Freedom to Make Life Choices': 0.8105
Middle East - Pearson correlation between 'Score' and 'Generosity': 0.5459
```



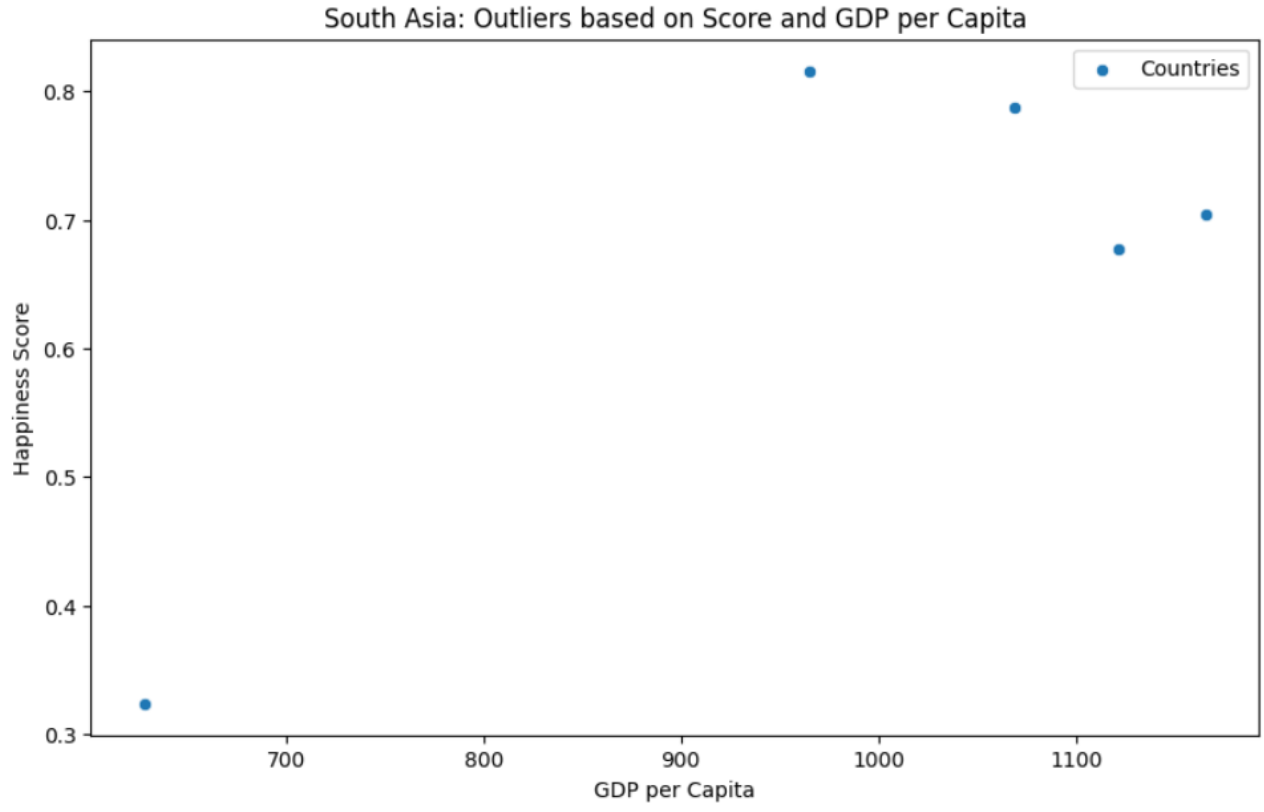

Middle East: Freedom to Make Life Choices vs Happiness Score
Happiness Score
Freedom to Make Life Choices
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.4
0.5
0.6
0.7
0.8
0.9

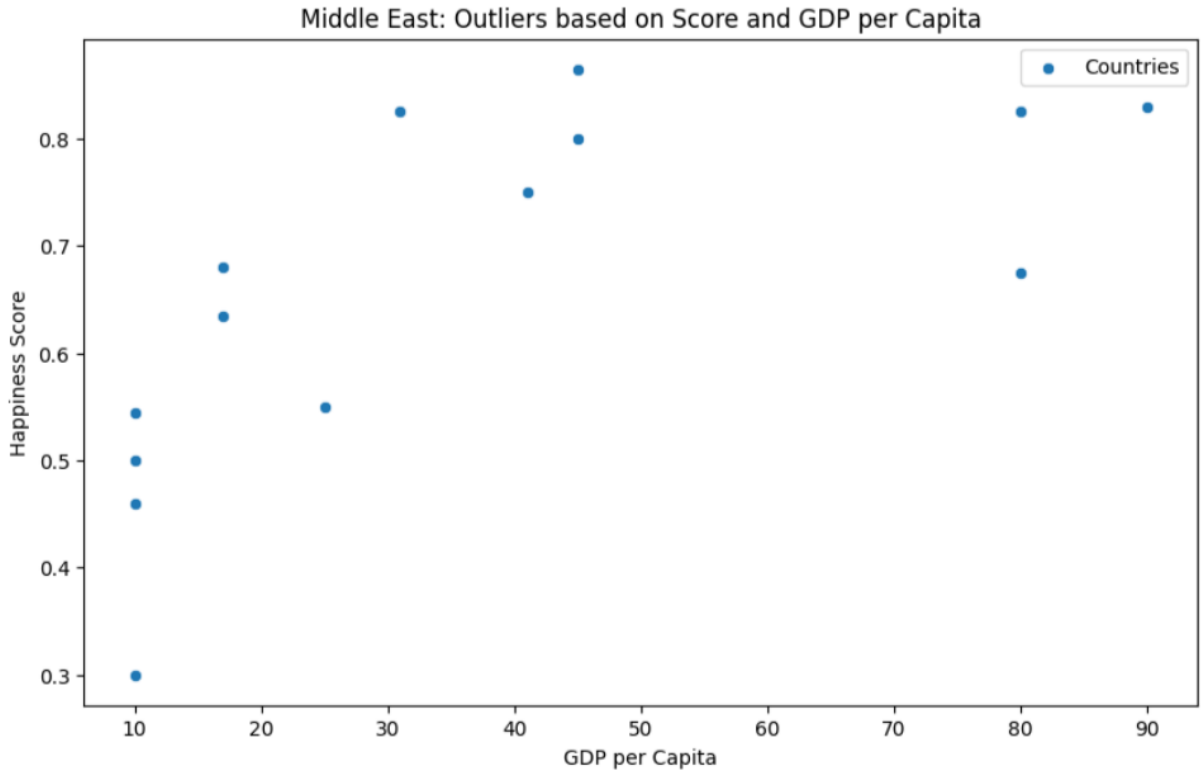

Middle East: Generosity vs Happiness Score
Happiness Score
Generosity
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.08
0.10
0.12
0.14
0.16
0.18

This code calculates the Pearson correlation between happiness scores and two factors ('Freedom to make life choices' and 'Generosity') for both South Asia and the Middle East. It then generates scatter plots for each region, showing the relationship between these factors and happiness scores with regression lines. The results are printed as correlations for each region.

6. Outlier Detection:

```
South Asia Outliers:
Empty DataFrame
Columns: [Country name, score, GDP per capita, score_z, GDP_z]
Index: []

Middle East Outliers:
Empty DataFrame
Columns: [Country name, score, GDP per capita, score_z, GDP_z]
Index: []
```

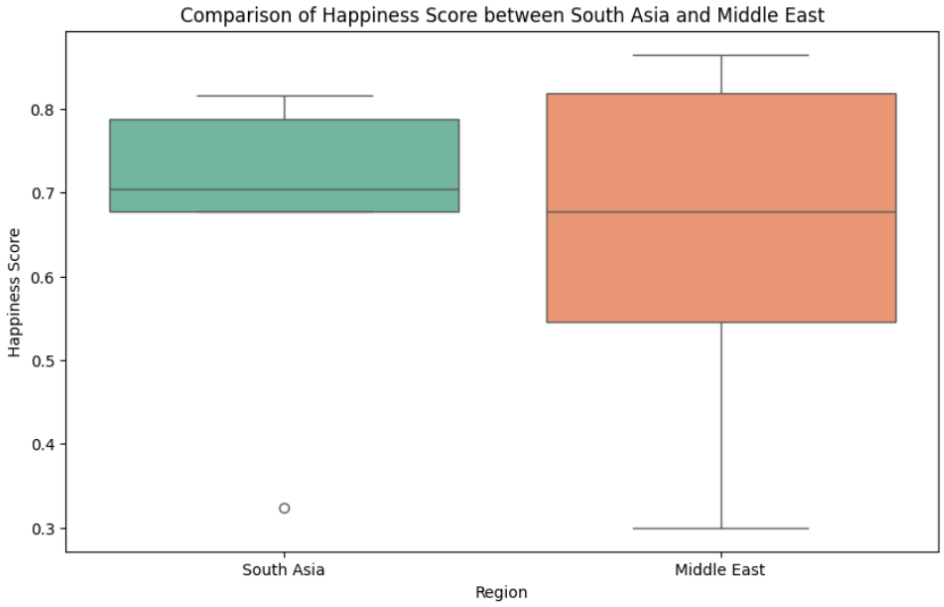





This code identifies outliers in happiness scores and GDP per capita for South Asia and the Middle East using Z-scores (values greater than 3 or less than -3). It then visualizes the outliers in scatter plots for both regions, highlighting the outliers in red.

7. Visualization:

```
<ipython-input-77-b9e7b143a41b>:28: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.boxplot(x='Region', y='score', data=combined_df, palette="Set2")
```




Activate V
Go to Setting

The code creates a boxplot to compare the happiness scores between South Asia and the Middle East, using data from both regions. The boxplot helps visualize the distribution of scores, highlighting medians, quartiles, and potential outliers in both regions.

## Conclusion

In conclusion, this report highlights the differences in happiness scores between South Asia and the Middle East, showing the unique challenges each region faces. South Asia, with generally lower and more similar happiness scores, points to issues like economic development, social support, and freedom, especially in countries like Afghanistan. On the other hand, the Middle East has a wider range of happiness scores, with countries like Qatar and the UAE having higher scores, while nations like Yemen and Syria face lower scores due to political instability and conflict. This suggests that the happiness in the Middle East is largely affected by political situations, while South Asia struggles with improving economic conditions and social services. The report shows that both regions need different approaches to improve happiness—South Asia should focus on better economic opportunities and social support, while the Middle East needs to focus on stability and personal freedoms. With more research and focused policies, both regions can work towards improving well-being and achieving higher levels of happiness.