CONCEPTS OF TECHNOLOGIES AND AI

(5CS037)

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

Student ID: 2408624

Student Name: Prayas Shrestha

Group: L4CG15

Module Name: Concepts of Technologies and AI(5CS037)

Submission Date: 2024-12-20

Contents

Introduction

The World Happiness Report is a significant global survey that evaluates happiness based on various socioeconomic and personal well-being factors. It gives insight into the factors that contribute the most to happiness in the world, such as economic productivity, social support, and life expectancy. It evaluates countries with a happy index, informing the world about inequalities and areas for improvement.

This report is an analysis of the data set with the help of visual and statistical tools, and it is mostly focused on South Asia and the Middle East. The job includes testing the dataset, computing composite scores, and comparing the trends of the regions. This study examines correlations as well as identifying outliers, thus the study identifies major patterns and differences in happiness metrics. The study is designed to provide an overview of the social and economic determinants of happiness, and it aims to propose policies for a higher level of well-being at both regional and global levels.

Problem 1 Getting Started with Data Exploration - Some Warm-up Exercises:

1. Data Exploration and Understanding

First, we loaded the WHR_5CS037.csv and made a data frame from it. We then displayed the first ten rows of the dataset using head () to analyze the rows and columns. After that, we used ‘shape’ to identify how many rows and columns the dataset holds and finally used ‘d types’ to examine the data type of different columns.

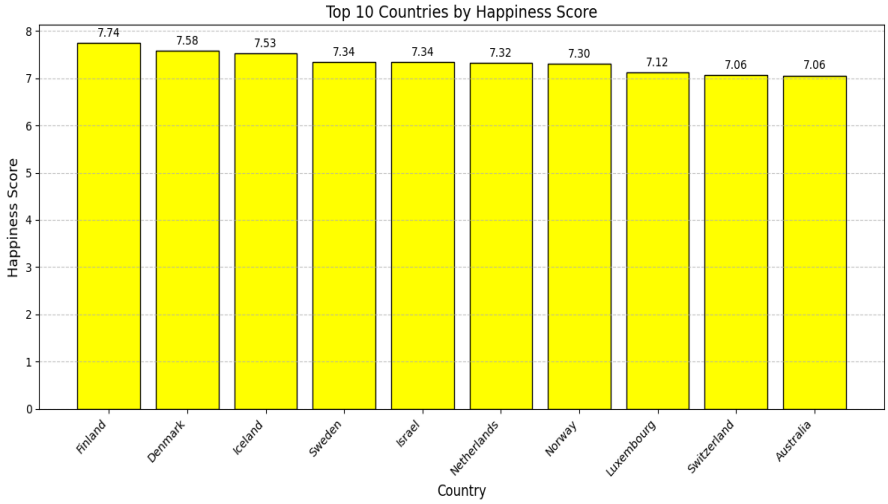
We then performed basic statistics operations on the dataset, calculating the mean, median, and standard deviation of the ‘score’ column. Then, using the max() and min() methods, we calculated the countries with the maximum and minimum happiness scores, with Finland having the highest score and Afghanistan having the lowest.

We then handled missing values with the isnull (). sum () function, which calculated the total number of null values in each row and displayed them.

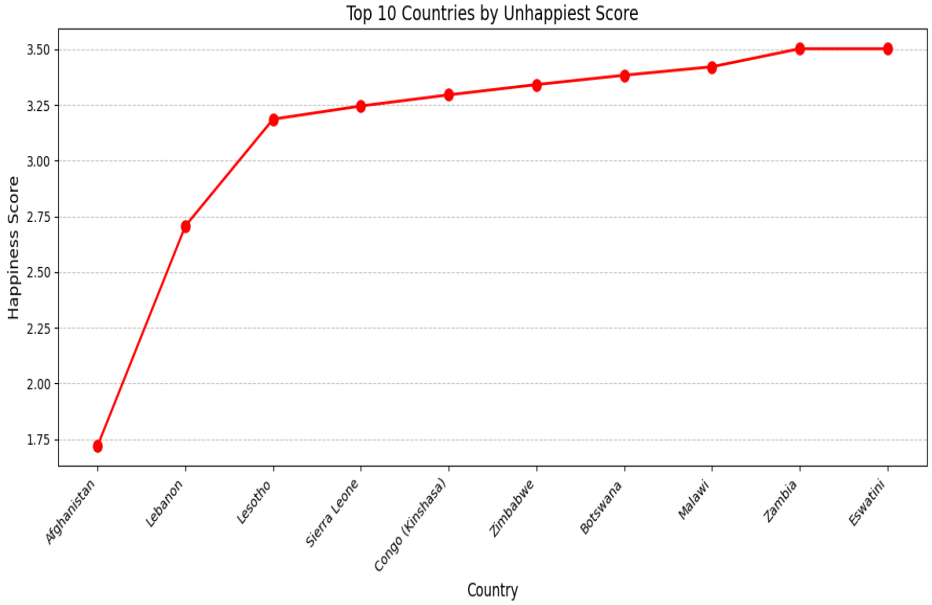
We then filtered the dataset where the happiness score is greater than ‘7’ then displayed them in descending order using ‘sort_values’ where ‘ascending = false’ and displayed the top 10 rows of the filtered dataset

For separating the happiness category, I created a function whose parameter is score and compared then with a condition that returns high, medium, and low when the condition is met. Then, create a new column that applies the function on the ‘score’ column and categorizes the happiness score.

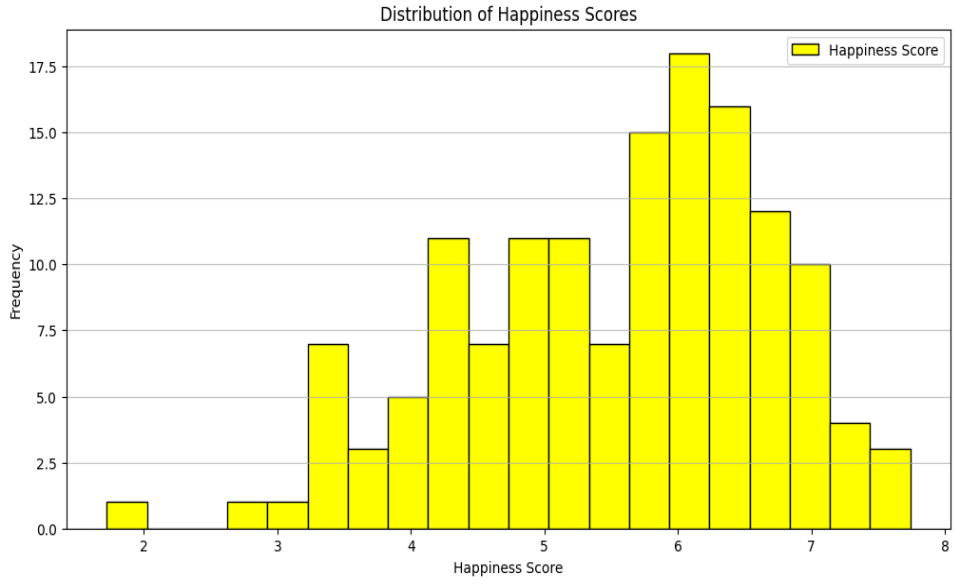
2. Data Visualization:



Here, I plotted a bar plot utilizing the matplotlib. pyplot which displays the top 10 happiest countries by their happiness score. Where x-axis represents the top 10 countries by their happiness score and Finland with 7.74 is the happiest country in the dataset y-axis represents the happiness score on a scale. The height of the bar indicates the happiness score of a country. The happiness score is displayed at the top of the bar for clarity. I made use of n_largest to find the top 10 happiness scores to plot the data.

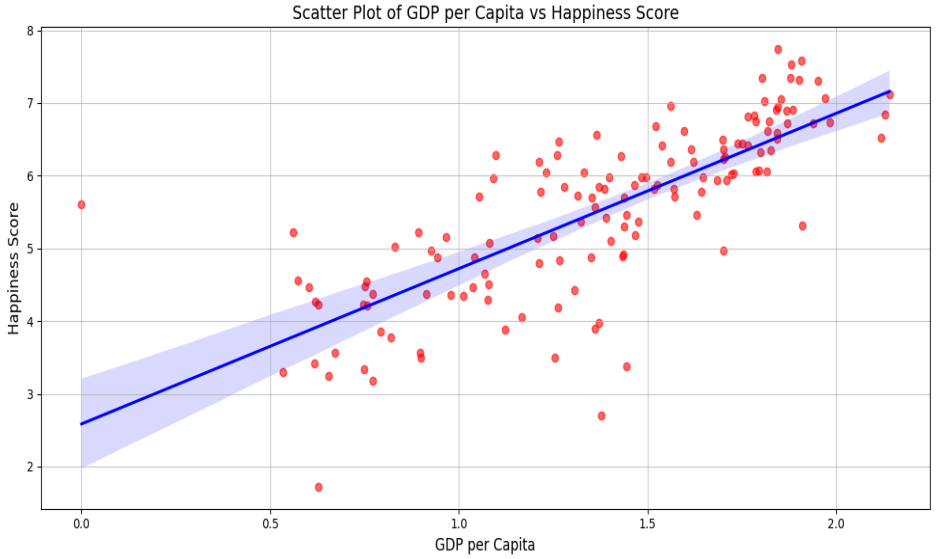


This is a line chart that showcases the top 10 unhappiest countries by score. The X-axis represents the name of the country arranged in ascending order according to the score. The Y-axis represents the happiness score on a numerical scale which ranges from 1.75 to 3.5. The line connects the score for continuity. From the plot, we can observe that Afghanistan is the country which the lowest happiness score of 1.75. The score gradually moves up from then to Eswatini with 3.5. Made use of nsmallest for finding the top 10 unhappiest scores of the data.



The histogram shows the distribution of the happiness score, with the x-axis representing the happiness score and the y-axis representing the frequency. The distribution of happiness is skewed to the right meaning there are more happiness scores in the lower and fewer in the higher ranges. There is a peak in the frequency around the

happiness score of 6. The data suggests that to increase the happiness score we have to focus on lower score range to have more impact.



This scatterplot shows the relationship between the happiness score of the country and its GDP per capita. I used regplot where my x is GDP per capita, and y is happiness score, the scatter plot represents the country and the line represents the linear trend line, showing the direction of the relationship. The plot shows a positive correlation between the two which means the countries with higher GDP per capita tend to have higher happiness scores.

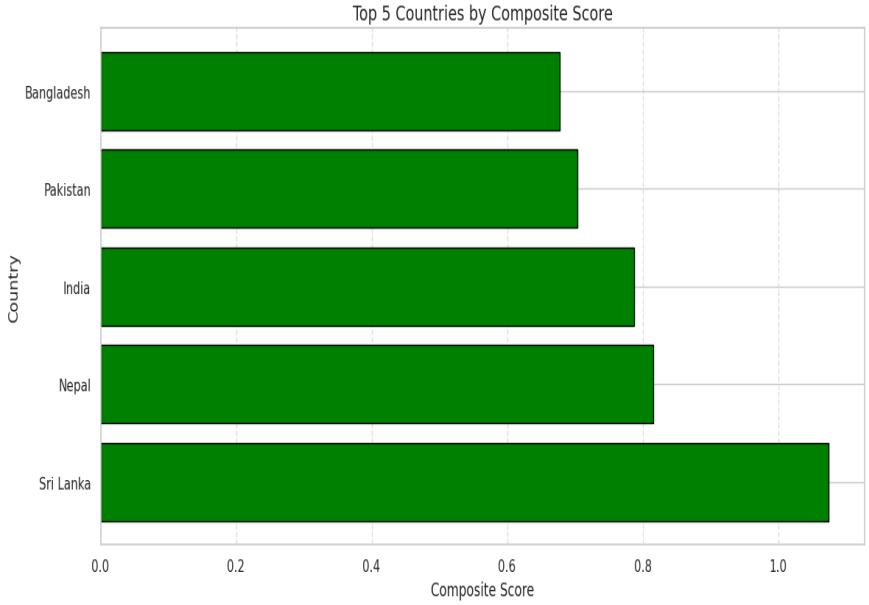
Problem 2 - Some Advance Data Exploration Task:

Task1 - Setup Task - Preparing the South-Asia Dataset:

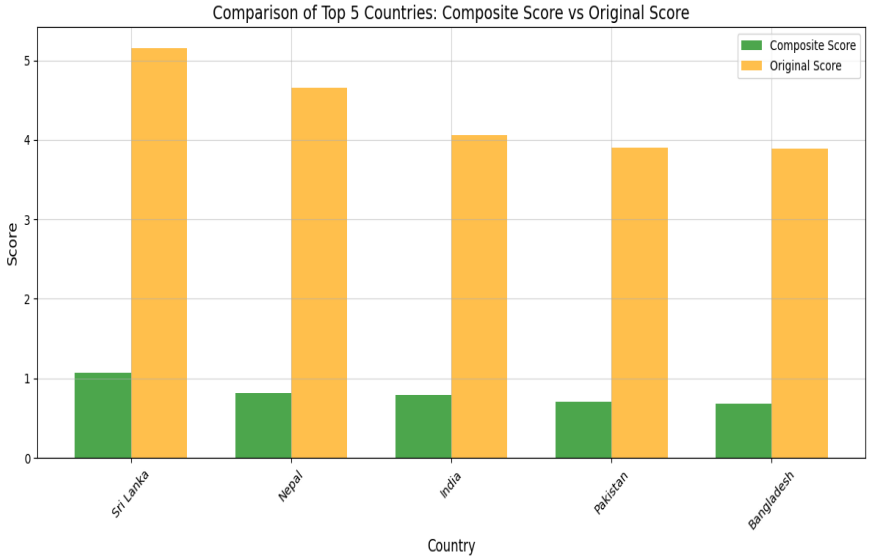
For this task, we created a new data frame consisting of South Asian countries where we extracted the data from the original CSV and saved it as a separate CSV file.

Task2 - Composite Score Ranking:

For this task, we read the previously saved CSV with the data from South Asian countries. We added a new column named ‘Composite Score’ to the data frame where the composite score is calculated using the weighted average of GDP per capita (40%), social support (30%), and healthy life expectancy (30%). The data frame is sorted in descending order according to ‘Composite score’ using ‘sort_values.

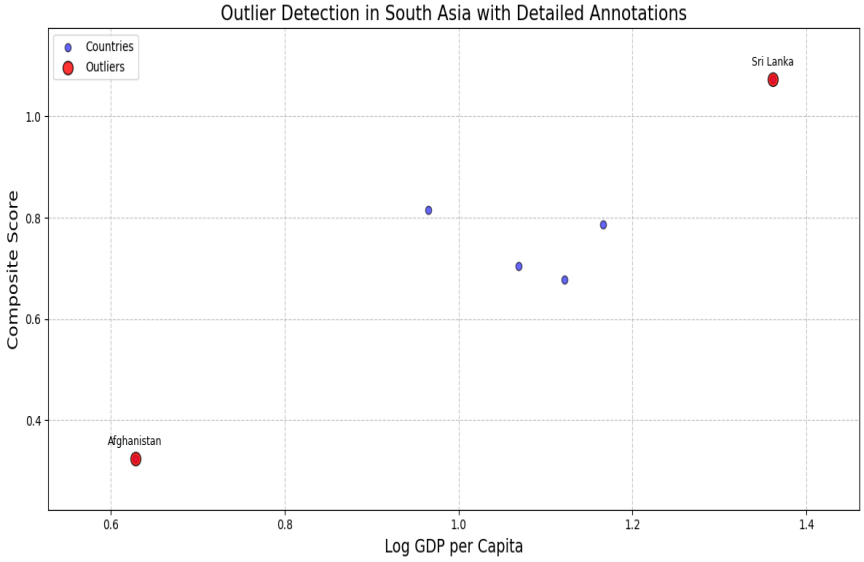


The horizontal bar plot shows the top 5 countries in the South Asian region by their composite score. The x-axis represents the composite score whereas the y-axis represents the top 5 countries in ascending order to their composite score. We can observe that Sri Lanka has the highest score, followed by Nepal, India, Pakistan, and Bangladesh. This is achieved by utilizing plt.barh () where parameters are composite score, country name, color and edge color.



This bar plot represents the composite score vs the score of the South Asian countries. The x-axis represents the countries, and the y-axis represents the scores in scale. The green bar represents the composite score, and the orange bar represents the original score. By observing the plot, we can see that the original score is comparatively higher than the composite score because we used the weighted average to calculate the composite score. The difference between the two scores is uniform.

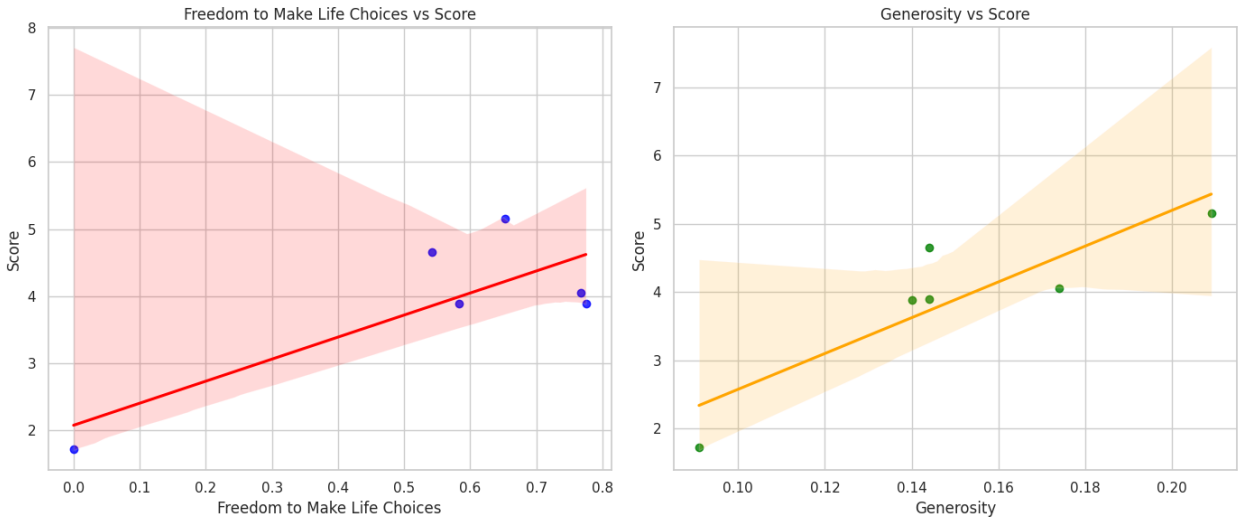
Task3 - Outlier Detection:



For calculating the outliers we first calculate the Q1,Q3 and IQR of the composite score and calculate the lower bound and upper bound score. Then compare them with the composite score and set positive and negative outliers.

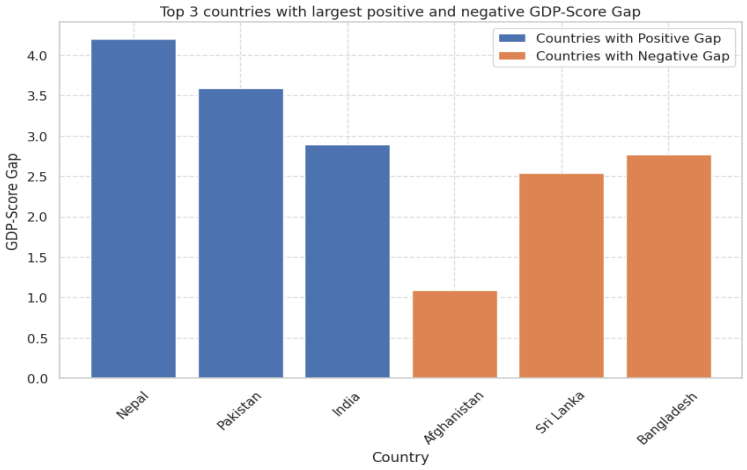
This scatter plot displays the composite score against the GDP per capita of the South Asian countries. The x-axis represents GDP per capita which captures the country's economic productivity, and the y-axis represents the composite score of the country derived from the original score where the blue dot represents the normal countries, and the bigger red dots represent the outliers – countries which deviate from the general trend. Afghanistan is a negative outlier, and Sri Lanka is a positive outlier. The plot showcases a positive correlation between GDPs per capita and composite score but outliers like Afghanistan and Sri Lanka deviate significantly from this trend. This shows that the composite score may vary due to other factors such as social norms, and generosities.

Task4 - Exploring Trends Across Metrics:



This scatter plot displays the relationship between two variables (Freedom to make life choices and Generosity) and their scores. The plot also consists of a regression line to represent the trend in the data. In the plot, we can that freedom to make life choices and scores have a positive correlation as freedom to life choices increases the score also tends to rise. In the second plot, we can see that generosity and score also have a positive correlation as higher generosity is associated with a higher score. The plot emphasizes the importance of social factors influencing the score.

Task5 - Gap Analysis:



This bar plot displays the top 3 countries with positive and negative GDP-Score Gap where the x-axis represents the countries, and the y-axis represents the GDP-Score gap in scale. The blue represents the top positive gaps, and the orange represents the top negative gaps. The plot shows that Nepal has the highest positive gaps, and Afghanistan has the smallest overall. Sri Lanka and Bangladesh show negative gaps, suggesting that these countries

achieve relatively high scores despite having lower GDP levels, likely due to factors such as governance, social policies, or other well-being indicators.

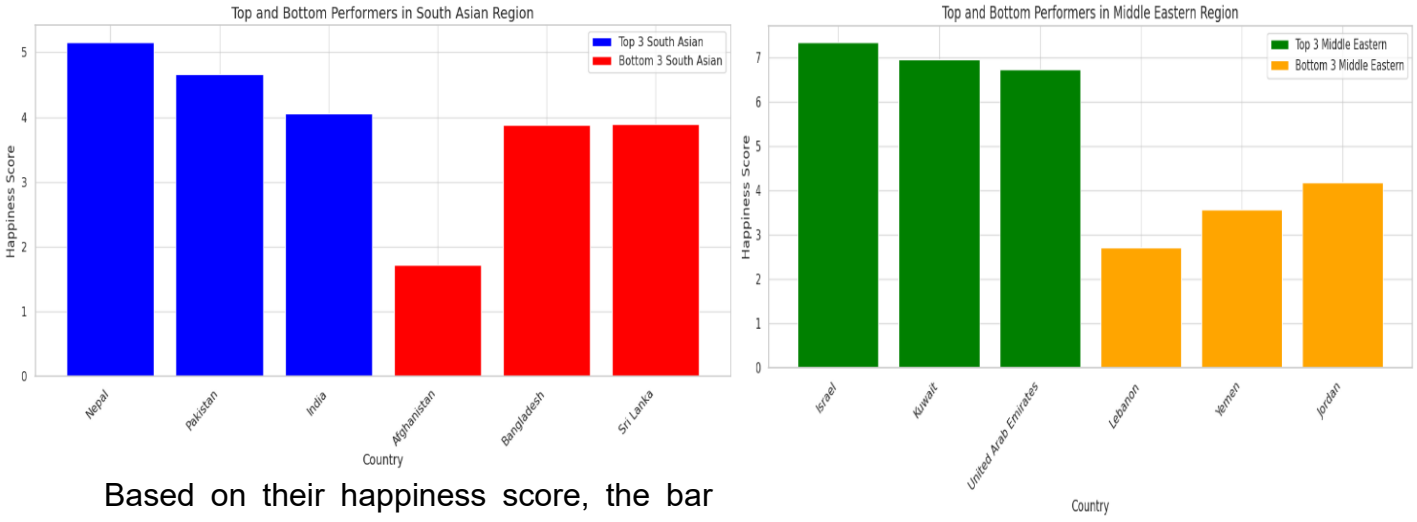
Problem 3

In this task, we created a data frame by extracting all the Middle Eastern countries from the original CSV and saved it as a new CSV file for operation

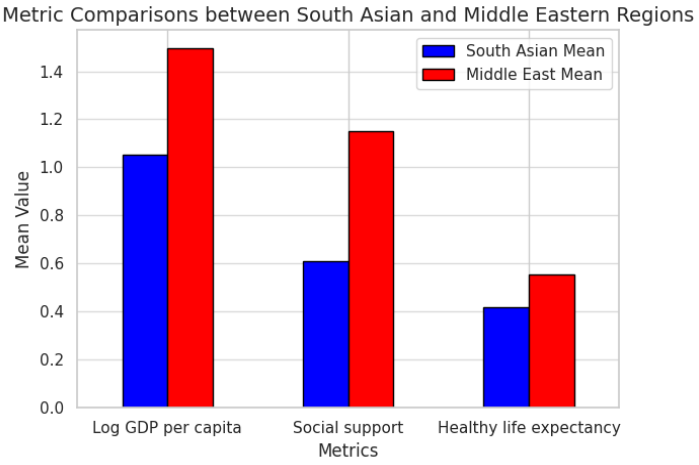
1. Descriptive Statistics

We calculated the mean, median, and standard deviation of the Middle East score and compared it to the mean, median, and standard deviation of the South Asia score where we observed that the Middle East has a higher happiness score compared to South Asia score.

2. Top and Bottom Performers



Based on their happiness score, the bar plot displays the top 3 and bottom 3 performers in both South Asia and Middle East regions. Made using plt.bar, the x-axis represents the countries of the region, and the y-axis represents the happiness score in the scale. The plot shows that the Middle East has a happiness score on a higher scale of 7 compared to South Asia's score of 5. The bottom performers show relatively low scores with Afghanistan in South Asia and Yemen in the Middle East. The plot shows great variation in the score in the Middle East compared to the Southeast.

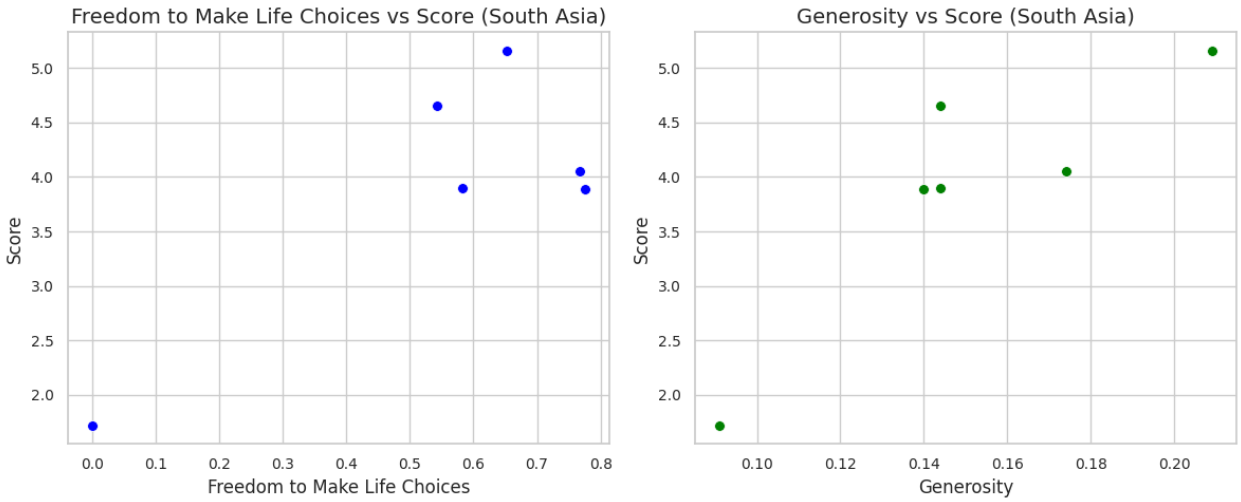


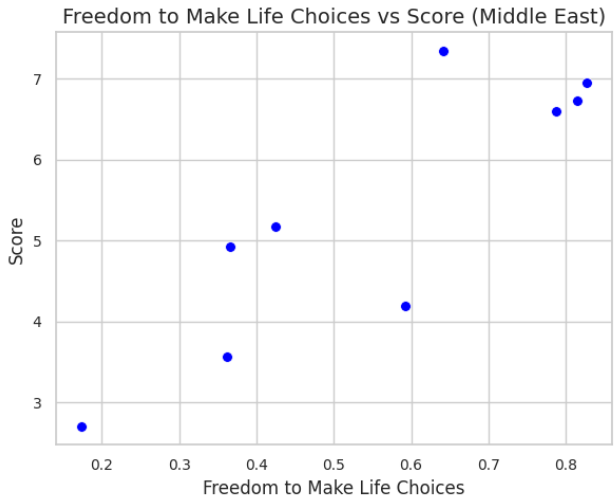
3. Correlation Analysis

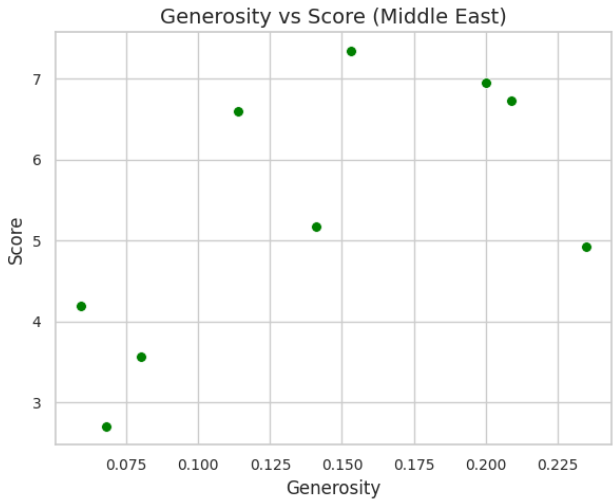
The bar plot compares 3 key metrics (GDP per capita, social support metrics, and healthy life expectancy across both regions. In the plot, we can see that the Middle East performs consistently better on all three metrics which shows that the region has an advantage on all three metrics and South Asia has areas for improvement in all three metrics.

4. Happiness Disparity: Here we calculated the max, min, variance, and standard deviation of each region and compared them to find the greater variability of the happiness in comparison. After the calculation, the Middle East has a greater variability of happiness.

5.Correlation Analysis: We made use of corr. Function to find the correlation of Freedom to make choices and Generosity against score of both region and created scatter plot.

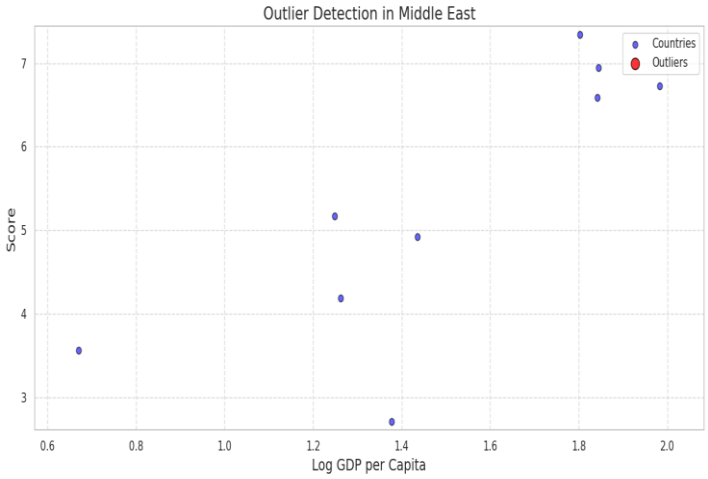


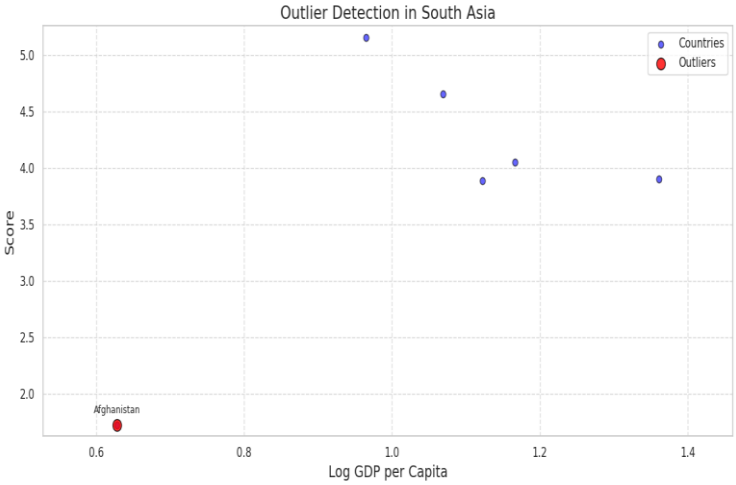




The scatter plots above display the relationship between the score and two metrics (Freedom to make choices and Generosity). From the plot, we can clarify that Freedom to make life choices plays a crucial role in the Middle East compared to South Asia and has a stronger relationship to score than Generosity in the Middle East. South Asia has weaker variability and correlation.

6. Outliers Detection

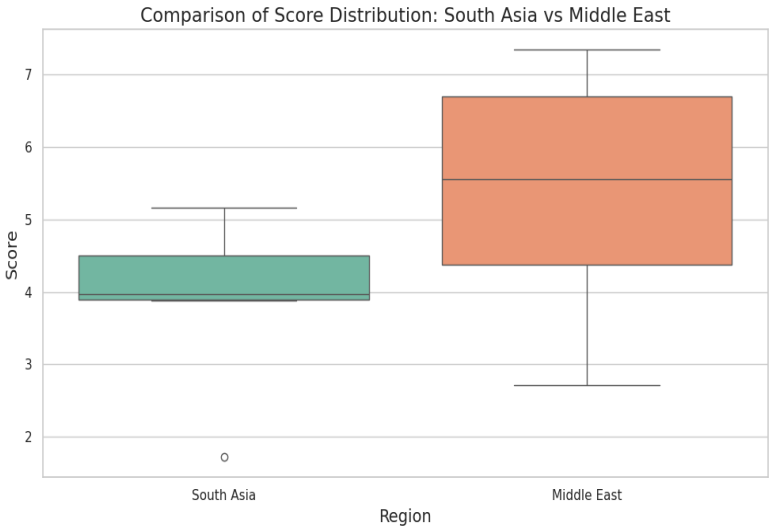




This scatter plot displays the score against the GDP per capita of the South Asian countries and the Middle East. The x-axis represents GDP per capita which captures the country's economic productivity, and the y-axis represents the country's score where the blue dot represents the normal countries. The bigger red dots represent the outliers – countries that deviate from the general trend. Afghanistan is a negative outlier in South Asia whereas there are no outliers in the Middle

East. The plot shows a positive correlation between GDPs per capita and scores in South Asia and the Middle East. Still, Afghanistan deviates significantly from this trend in South Asia and the Middle East doesn't have any outlier that is deviating from the trend.

7. Visualization:



In this task, we create a boxplot and compare the scores of each region. The box plot compares the scores of South Asia and the Middle East. South Asia has a lower median and less range showing consistency in scores. While the middle east has a higher median and wide range which shows great variability. The whisker represents the main range of each region, but an outlier appears below South Asia. Overall, the Middle East has a higher score whereas South Asia has a more compact score.

Conclusion

This report analyzes the World Happiness Report dataset, especially focusing on data exploration, regional analysis of South Asia, and Comparative analysis of the Middle East. With statistical calculations and different visualization methods, we see the general trend in global happiness with different factors such as GDP per capita, generosity, and many more highlighting their importance in the happiness score.

For Problem 1, we explored the dataset and analyzed that Finland is the happiest country and Afghanistan is the least happy country in the world. GDP per capita and other factors influence the happiness score of each country. From Problem 2, we analyzed that Sri Lanka had the highest composite score, which shows greater variability in happiness by

economic and social metrics. Outliers like Afghanistan and Sri Lanka show a significant change in trend where regional factors vary.

From Problem 3, we compared South Asia and the Middle East. It showed that the Middle East has higher happiness scores on average and greater variability because of diverse regional factors. Correlation analysis revealed that ‘Freedom to make life choices’ has a stronger relationship with the score of the Middle East whereas South Asia has a weaker relationship with most metrics. Afghanistan was a negative outlier in South Asia whereas the Middle East has no outlier, indicating uniformity despite variability.

The analysis's findings highlight the significant role of economic and social factors in shaping the regions' happiness scores. The analysis shows that South Asia should focus more on GDP, social support, and life expectancy, while the Middle East should focus on reducing the variability to ensure equitable well-being. Outliers remind us that targeted, context-specific strategies are vital. By addressing the above, regions can create condition for happiness.