Concepts and Technologies of AI

Assignment 1: Statistical Interpretation and Exploratory Data Analysis

Student ID : 2431338
Student Name : Swayam Shrestha
Section :G26
Tutor : Nabin Acharya
Module Leader :Bibek Khanal

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

The World Happiness Report is an annual global survey of how happy people are in which countries. It assesses happiness levels based on GDP per capita, healthy life expectancy, social support, freedom to make life choices, generosity, perception of corruption and a factor called dystopia + residual. The report aims to showcase the well-being of countries, presenting valuable information for people around the world and assisting American decision-makers in creating place-centered, people-centered policies that enhance overall wellbeing, happiness, and quality of life.

The purpose of the study is to analyze the World Happiness Report dataset, looking at patterns of general happiness across continents. This analysis aims to explore globally and regionally significant insights by pinpointing factors associated with happiness.

South Asian and Middle Eastern countries will also be analyzed separately to explore both regional differences and similarities within happiness determinants. This analysis focuses on Africa and Asia to offer insights and actionable findings that can be relevant to policymakers, researchers, and anyone interested in the dynamics of wellbeing and happiness in these regions.

Problem 1: Data Exploration, Understanding and Visualization.

For this, we started with loading the dataset and looking how it looks like, how many rows and columns does it have etc. Initially, we calculated the mean, median, and standard deviation of the "Happiness Score" column representing happiness level of countries across the globe. With these statistics, we figured out the happiest and unhappiest places in the data set.

Subsequently we filtered the dataset to the countries that had a higher happiness score than 7.5. This subset was filtered in personally ordered descending happiness score and was presented for further analysis.

To provide more insight, we added a new column labeled "Happiness Category", which classifies countries according to their happiness scores:

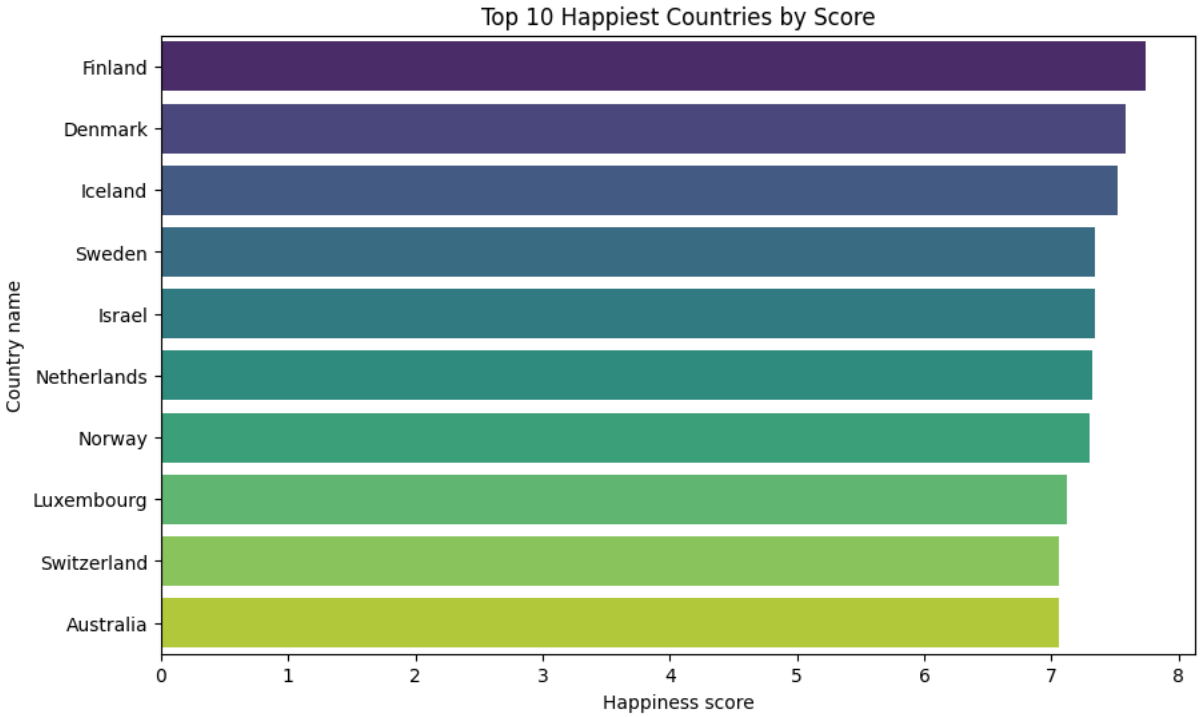
Low: Scores below 4

Medium: 4 to 6 scores

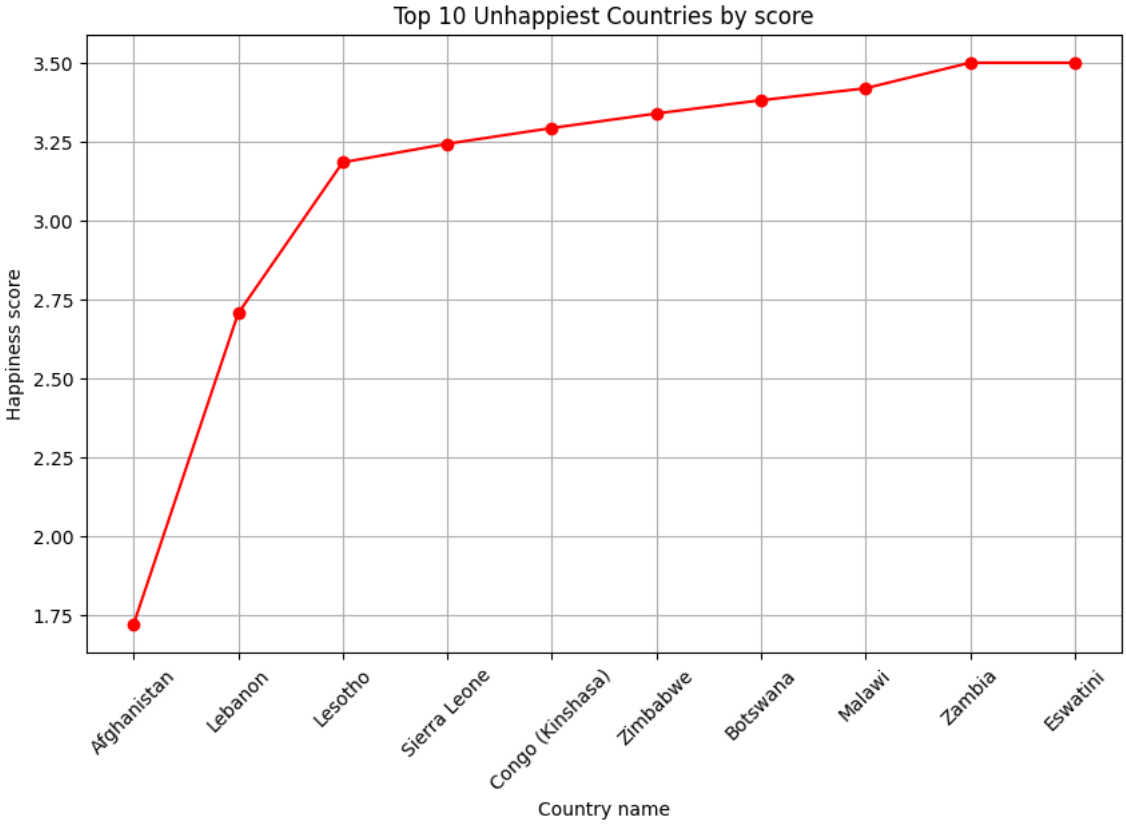
High: Scores above 6

Lastly, we plot the data — we generate a bar-plot of 10 happiest countries.

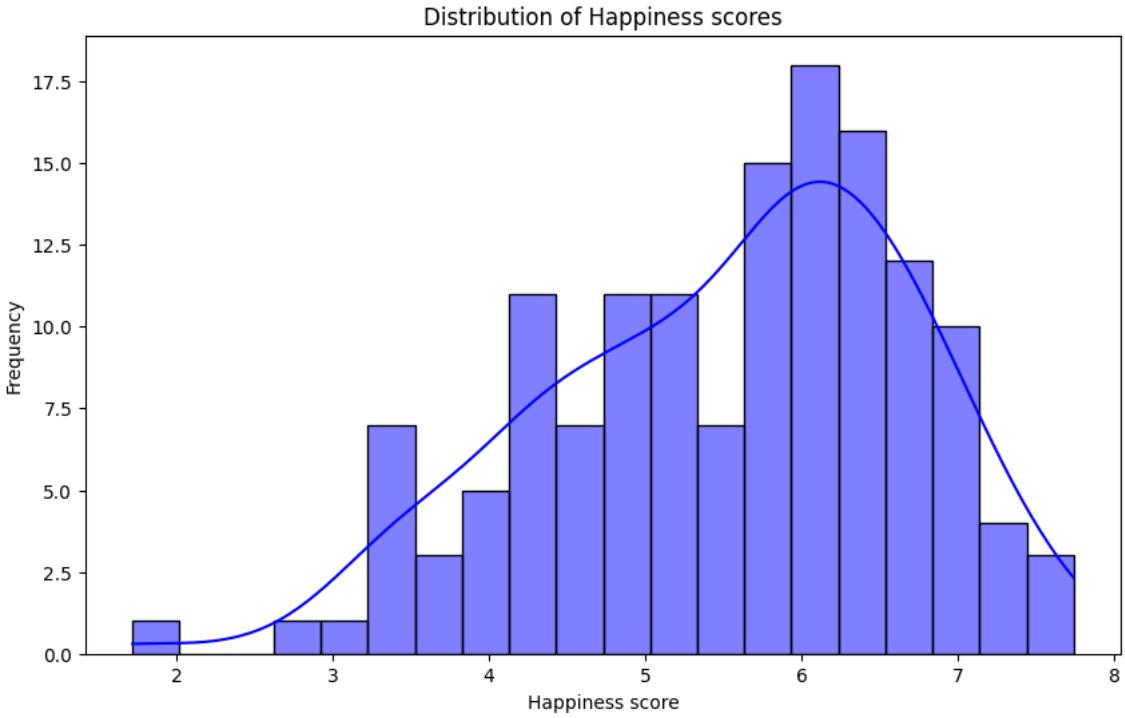
This will provided a clear graphical representation of the top-performing nations in terms of happiness.



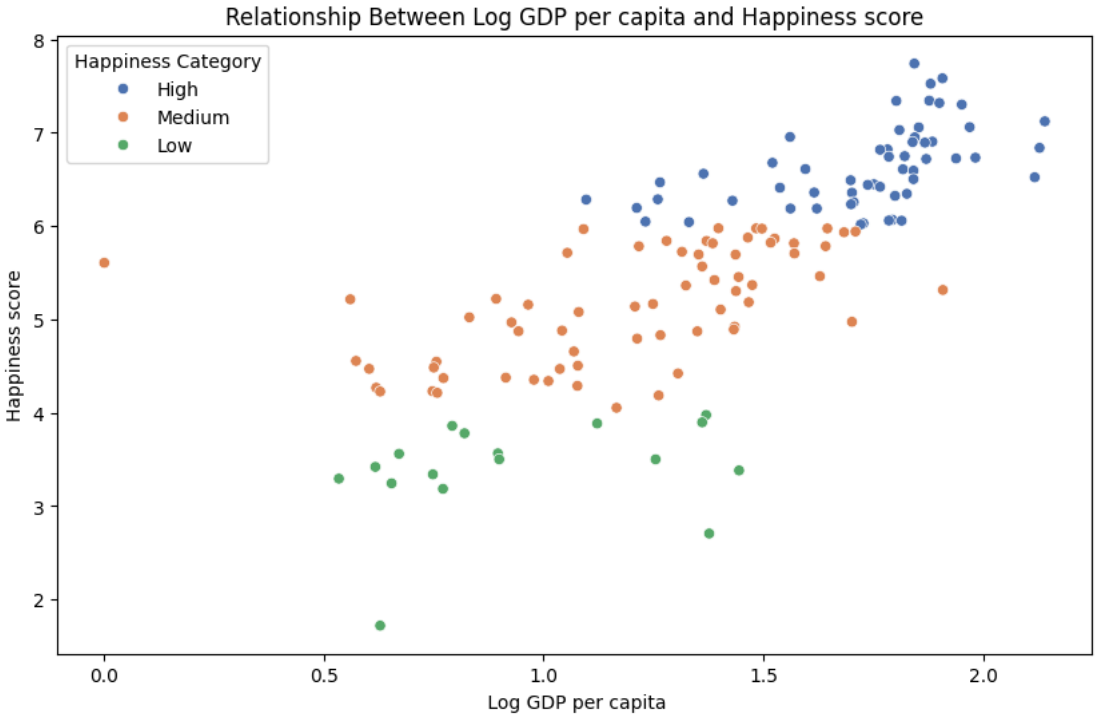
Now we created a line chart to show the 10 unhappiest countries.



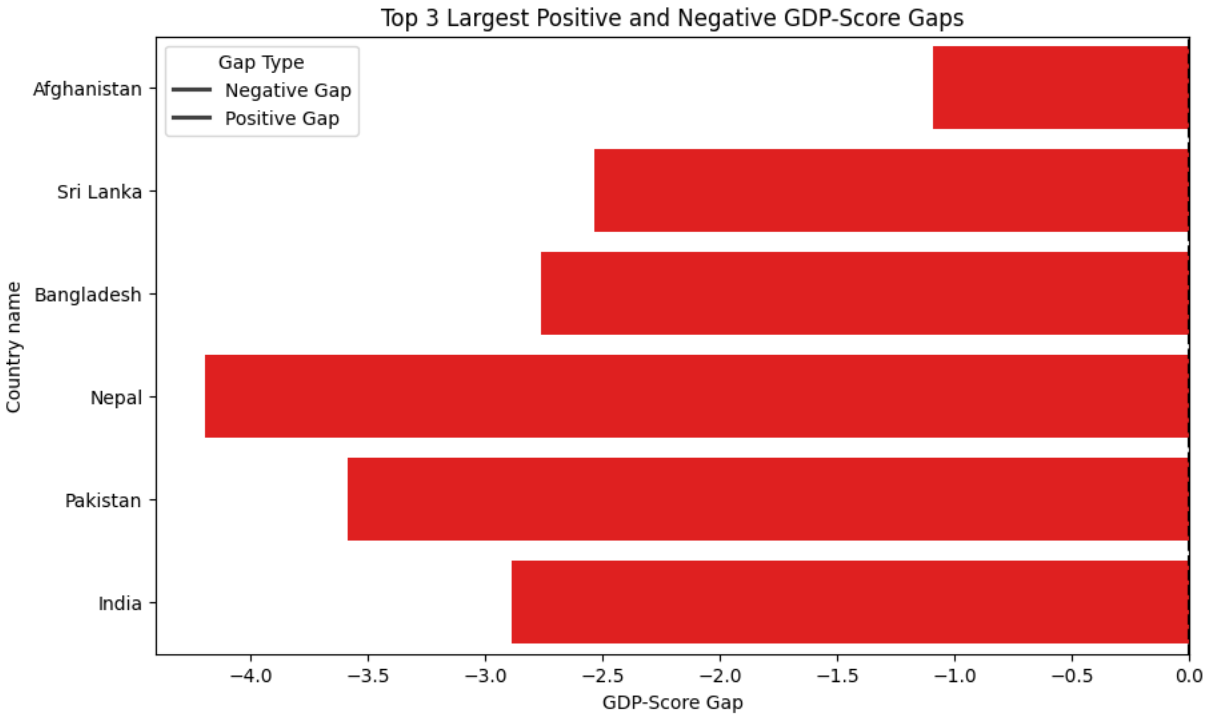
We also created a histogram to show the distribution of happiness scores.



Now we created a scatter plot to show the relationship between GDP per capita andHappiness scores.



From our analysis, we concluded that higher GDP per capita is positively correlated with higher happiness scores.



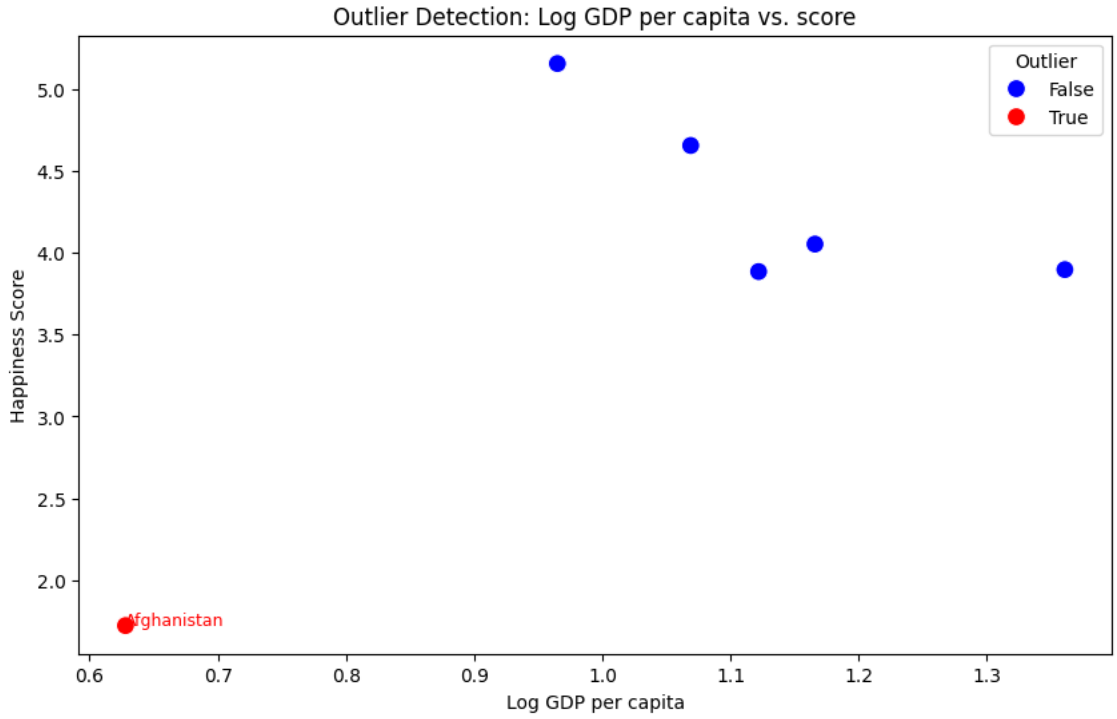
This code will create a bar chart comparing the top 3 largest positive and negative GDP-Score gaps for South Asian countries. Positive gaps are colored in green and negative gaps in red. A vertical dashed line at 0 separates the two types of gaps. The chart includes a title, axis labels, and a legend to distinguish between positive and negative gaps.

Problem 2: Advance Data Exploration of South Asian Countries.

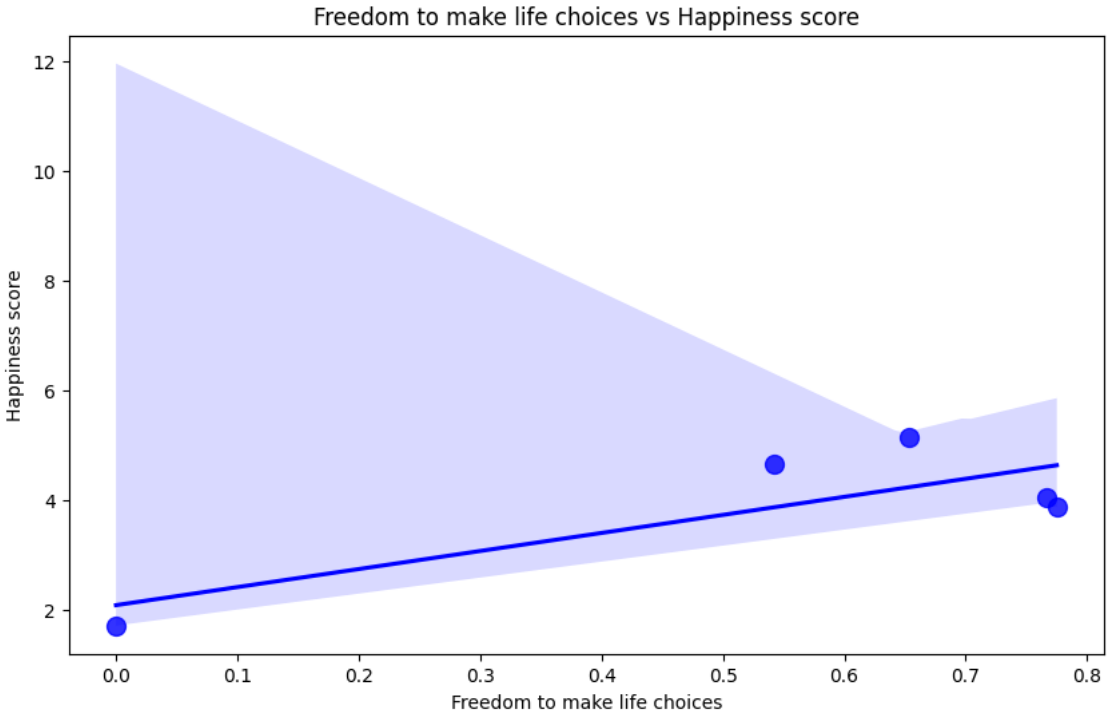
In this analysis we started the process by removing South Asian countries from the sample. Finally, we combined the three columns (GDP per capita, social support, and healthy life expectancy) to create single composite index. Using this index, we evaluated the five most inclusive South Asian countries: Sri Lanka, Nepal, India, Pakistan, and Bangladesh.

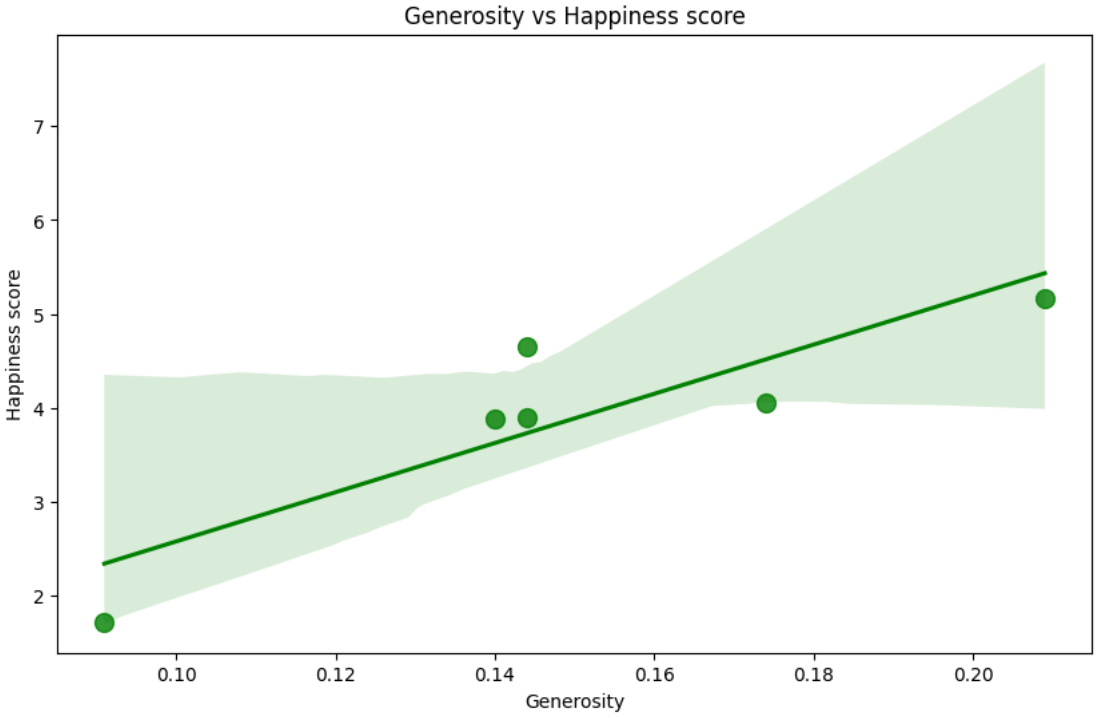
We then ranked all countries in the region according to their composite index values. Hence, we used the Interquartile Range (IQR) method to find outliers,

which were then highlighted in a scatter plot for better visualization of the data distribution.



Next, we selected two metrics—**Freedom to make life choices** and **Generosity**—and calculated their Pearson correlation with the happiness score. This analysis was visualized through the following scatter plots:





We then calculated the gap-the difference between GDP per capita and the happiness score-for each country in South Asia. After ranking the countries based on this gap, both in ascending and

descending order,our analysis showed that the three countries with the largest positive gap (that is, where GDP per capita exceeded the happiness score) were Afghanistan, Sri Lanka, and Bangladesh. On the other hand, the three countries with the largest negative gap (where happiness scores were lower than might have
been expected from GDP per capita) were Nepal, Pakistan, and India.

From this analysis, we have found that
the ranking composite indexdeveloped a very close correspondence to t he happiness scores forthese countries. However, there were some outliers. Some of these countries, even with high GDP, may rank lower on happiness scores,possibly indicating underlying issues, like societal discrimination, stillpersist in those regions.

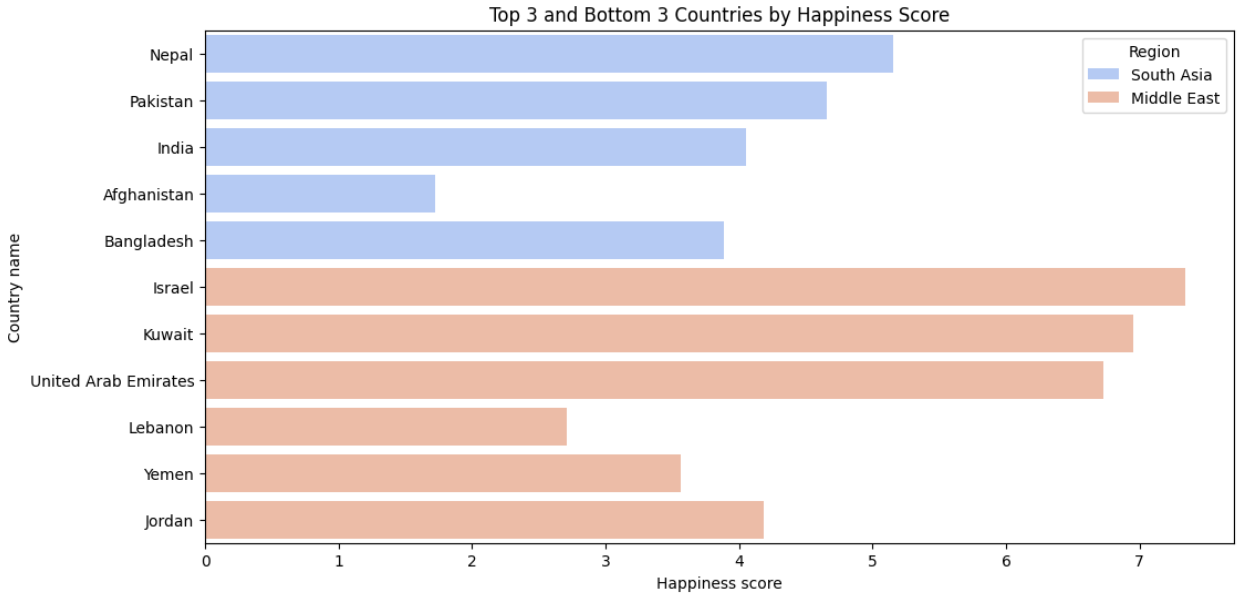
Problem 3: South Asian Countries vs Middle Eastern Countries.

In this analysis, we focus on the Middle Eastern countries
and comparethem with South Asian countries. First, we calculated the mean and standard deviation of happiness scores for both
regions. It can be seen that the Middle Eastern countries have a higher average happiness score as compared to South Asian countries.

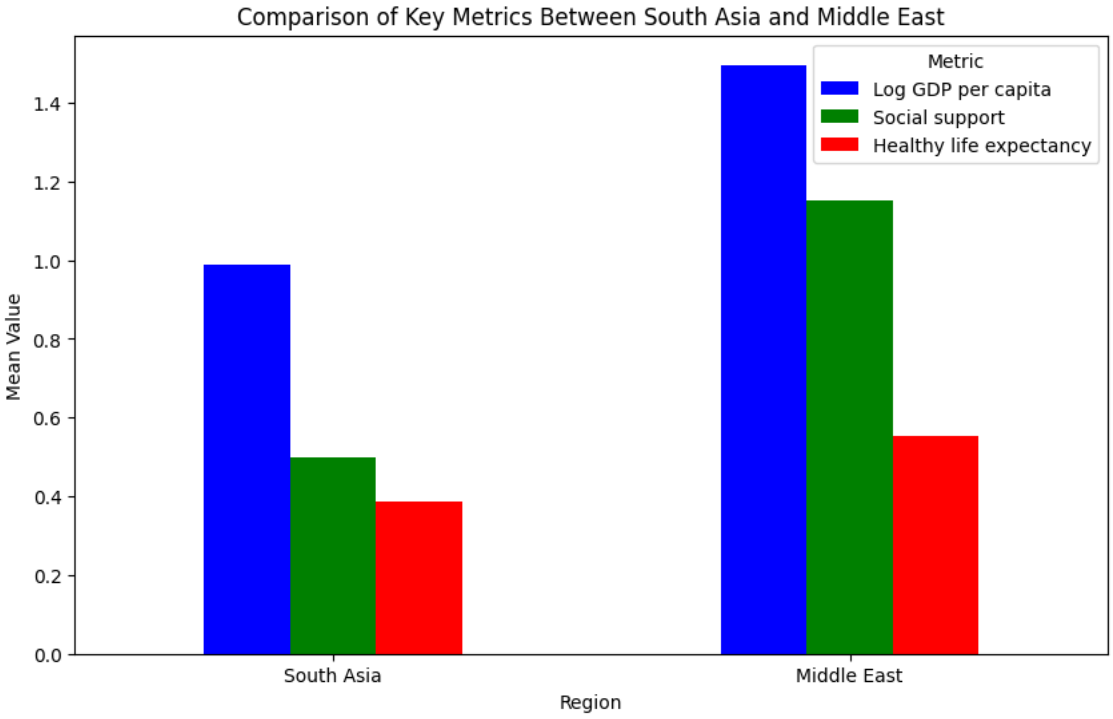
We then identified the top three and bottom three countries based on their happiness scores in both regions:

Top 3 South Asian Countries: Nepal, Pakistan, India
Bottom 3 South Asian Countries: Afghanistan, Bangladesh, Sri Lanka
Top 3 Middle Eastern Countries: Israel, Kuwait, United Arab Emirates
Bottom 3 Middle Eastern Countries: Lebanon, Yemen, Jordan
We also made pairwise comparisons between countries of these two

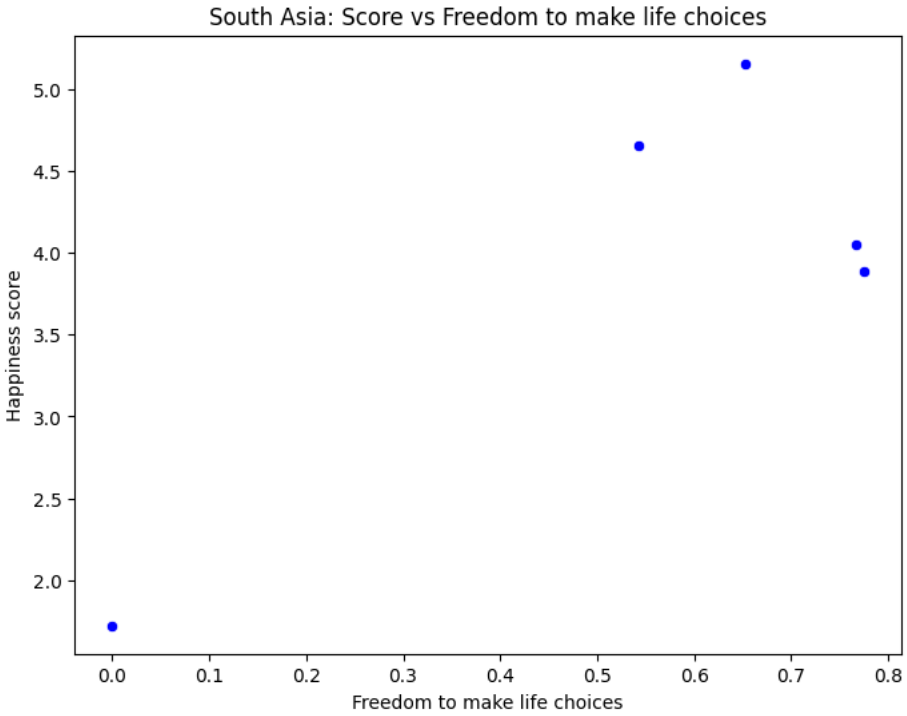
regions in terms of three metrics-GDP per capita, social support, and healthy life expectancy-and then presented them through the bar chartbelow.

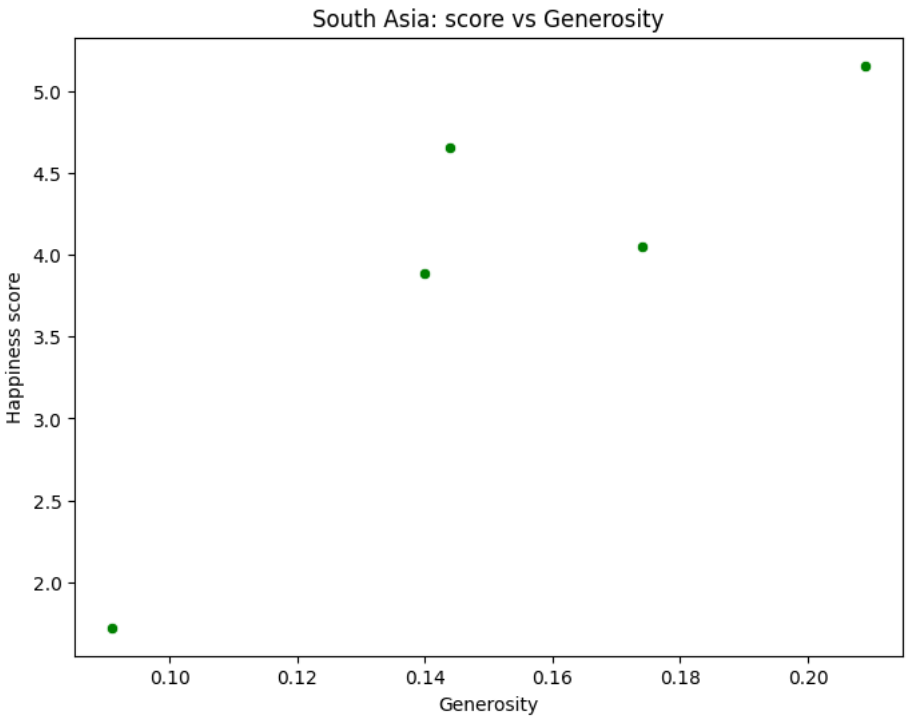


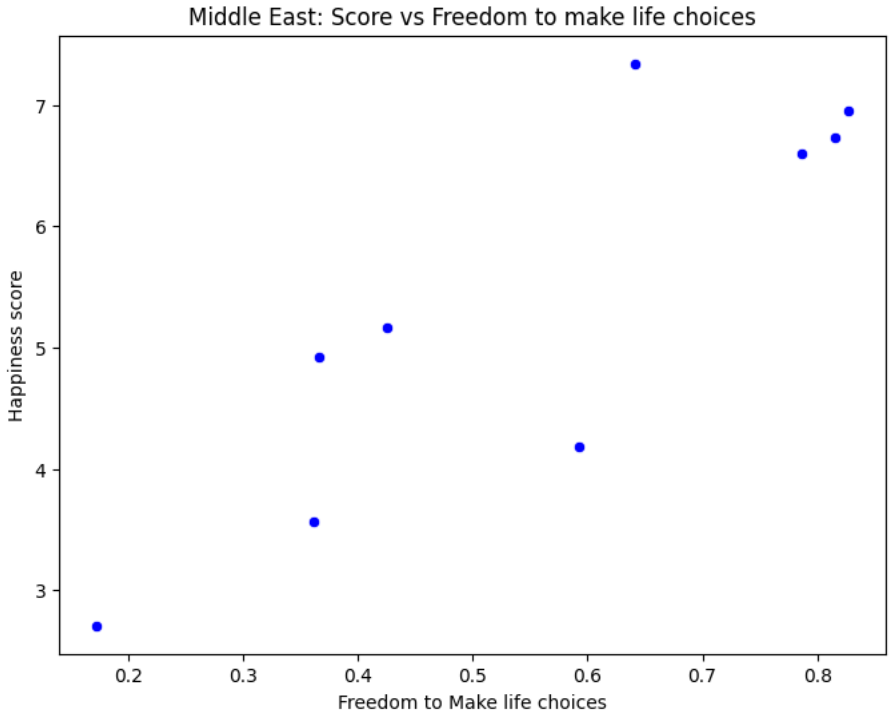
Now, This code will create a grouped bar chart comparing three metrics between South Asia and the Middle East: Log GDP per capita, Social support, and Healthy life expectancy. It creates a DataFrame containingthe mean values for each region, then plots the values with distinct colors for each metric. The chart is customized with a title, axis labels, and a legend, and is displayed using plt.show().

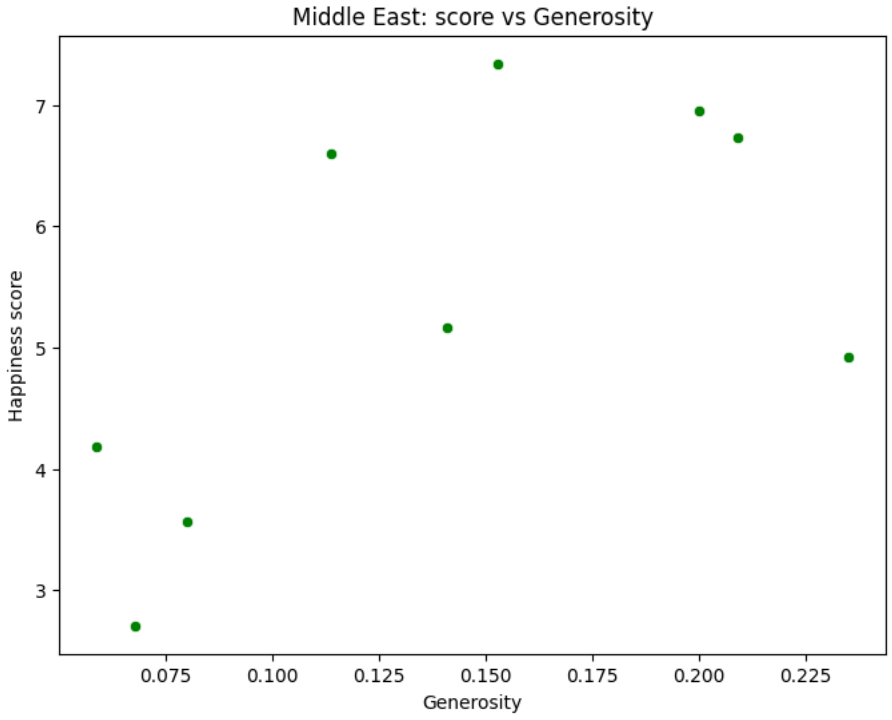


We can observe that Middle Eastern Countries are greater than South Asian Countries in each
of the metrics.
We then compared the variation of happiness by determining the range and coefficient of
variation for happiness score in both regions and determined that South Asia has a higher
happiness variation than Middle Eastern.
We also determined the correlation of happiness score with Freedom to make life choices
and Generosity and plot the results in scatter plots:

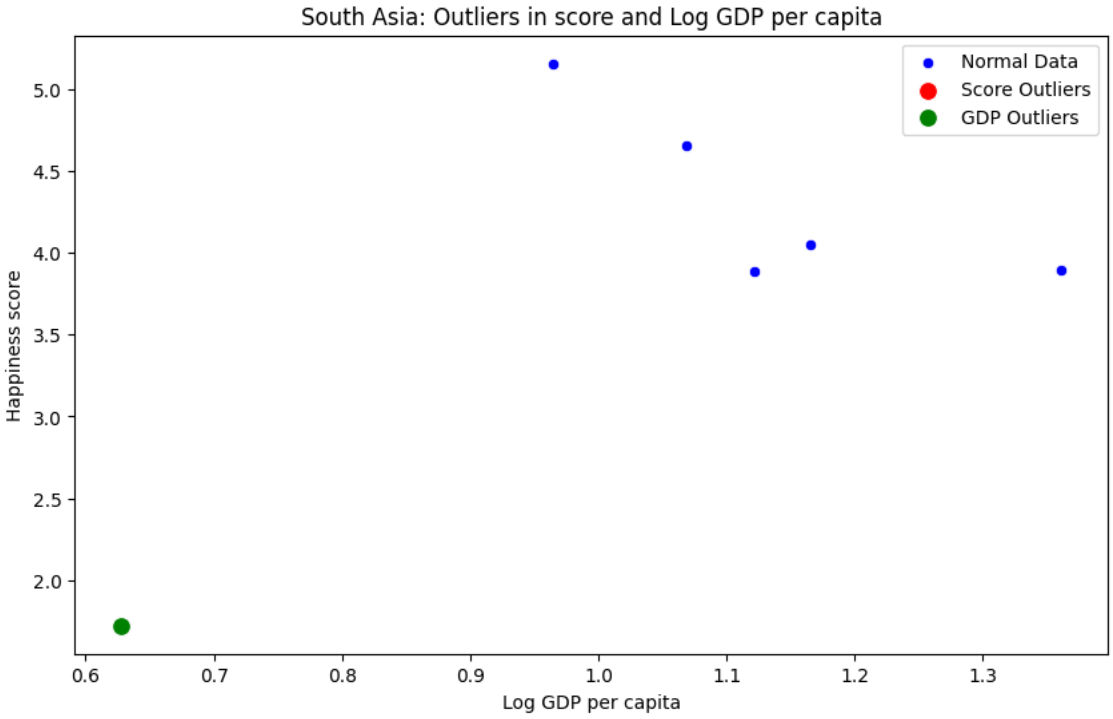

South Asia: Score vs Freedom to make life choices
Happiness score
Freedom to make life choices
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

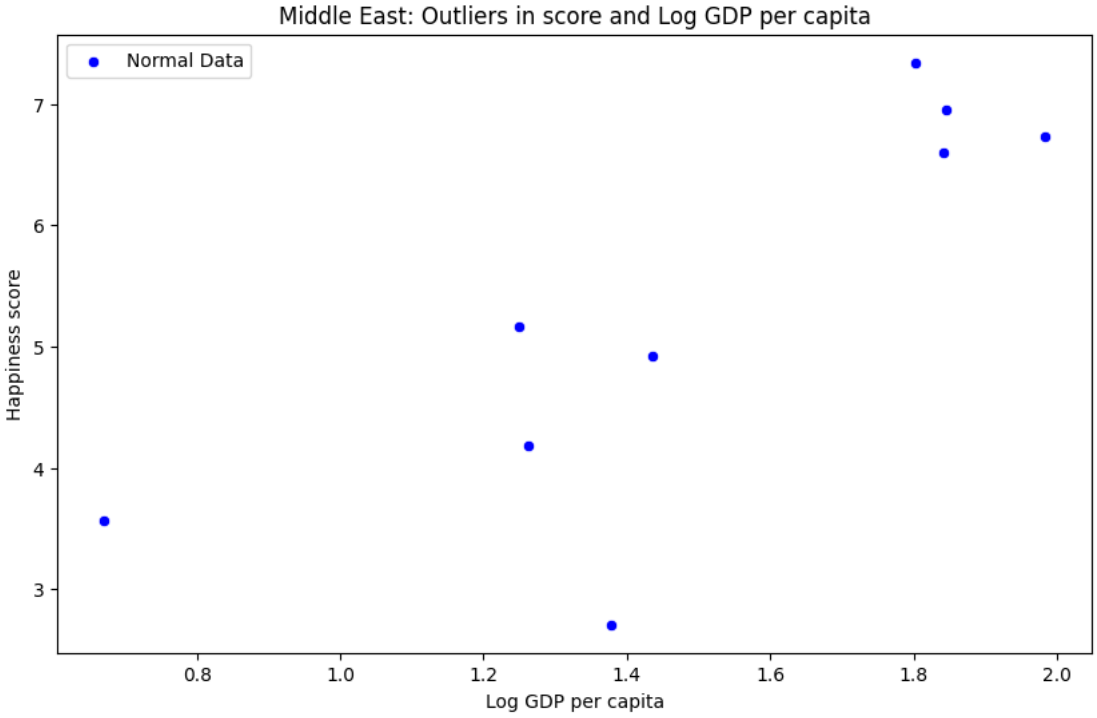

South Asia: score vs Generosity
Happiness score
Generosity
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0.10
0.12
0.14
0.16
0.18
0.20





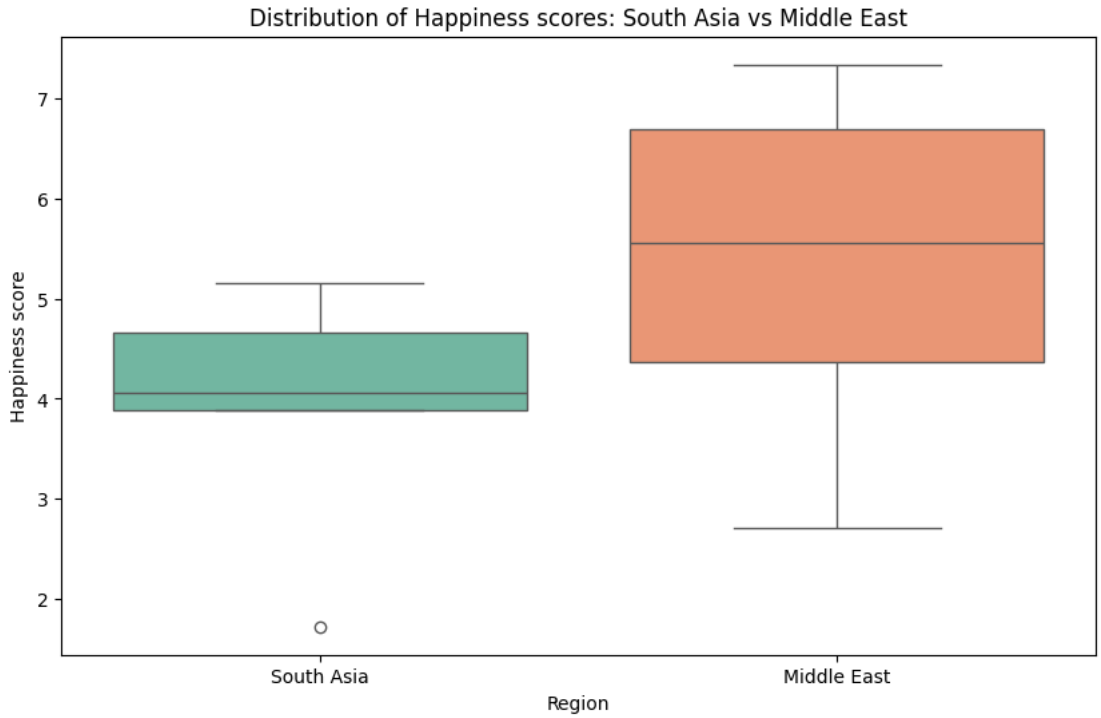
We detected anomalies of both South Asia and the Middle East to checkfor Happiness Scores against GDP per capita. Following that, we createda comparative box plot of Happiness Scores across countries in South Asia and the Middle East.





This code will generate two scatter plots: one for the South Asia region and one for the Middle East. Both visualize the relationship between "Log GDP per capita" and "Happiness Score," drawing normal

data points in blue, score outliers in red, and GDP outliers in green. Each plot includes titles and axis labels, along with a
legend to distinguish these.



From this analysis,
we found that countries from the Middle East have a higher average Happiness Score compared to South Asian countries. However, South Asian countries have more variability in their Happiness
Scores, showing that there are countries that have extremely high happiness scores and countries with very low happiness scores.

Conclusion:

In our report, we identified some global patterns in happiness, such as the tendency of GDP per capita to track the scores in
happiness. In thecase of the South Asian region, composite
scores mirrored happiness scores in most countries except for

those outliers that had a high level of GDP
but scored low on happiness, indicating the presence of societal discrimination in those regions. Comparing the results of the data between South Asia and the Middle East, it is observed that
the latteroutperforms the former on a majority of indicators with lesser variability in many of them.Our analyses have once again pointed toward the very complex and multidimensional aspect of happiness influenced by various factors such as health, social conditions, and economic status.