Academic Year	Module	Assessment Number	Assessment Type
S24	Concepts and Technologies of AI	01	Report Writing

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

Student Id   :2358431
Student Name   : Anupam Poudel
Section   : L5CG10
Module Leader   : Mr. Siman Giri
Tutor   : Ms. Prakriti Regmi
Submitted on.   : 2024/12/20

Table of Contents

Table of Figures

1. Introduction

Each year, the World Happiness Index publishes the results of the World Happiness Survey, which measures the extent to which people in different countries and regions are happy, and gives us the information about overall life satisfaction all over the world. The study, a collaborative effort of several organizations, contains data from the Gallup World Poll and is used to place the countries on a list by the subjective well-being of their respective inhabitants. This happiness index is the product of a scientific method based on the Cantril ladder, which is a life evaluation scale. In this scale, people express their quality of life by rating their life as the best possible (10) or the worst

possible (0). Furthermore, the report doesn't stop at just showing the happiness ladder; it also looks at other circumstances, like social, economic, and environmental, around the world to give a full view of the situation.

Vital to the happiness score variations are six crucial aspects, namely: GDP per capita, social support, life expectancy, freedom to make life choices, corruption perceptions, and generosity. These components provide a solution to the question of why some nations got higher scores of happiness than others, although they do not directly determine the main index. The results for each factor are shown as the difference from the hypothetical conditions of a locale called "Dystopia" which represents the lowest national average for each metric. In the form of a pure scale, this stress-test allows us to order the importance of each individual cause of life dissatisfaction, e.g., so that we may see the valid means to be happier.

The report has the additional value of being a guideline for decision-makers and other stakeholders, as well as for the general public, who might want to have an insight into how happiness is made and what sociopolitical consequences it involves. At the same time, however, the F-sheet also shows the critical impact of good governance, the challenge of inequality, the resilience in environmental management, and the vitality of community networks on human happiness on the national and global levels. Since the document is a work in progress, it underscores that the only way to go is by using a holistic concept in the context of fostering well-being and resilience in a world that is increasingly more interconnected but at the same time more demanding.

2. Objectives

Therefore, this paper will endeavor to expand on the statistical interpretation of happiness with a focus on the sub-concepts of AI. The primary objectives include:

- Conduct a comprehensive data analysis to assess the GDP per capita, social support, and other contributing factors to happiness in various countries.
- Explore statistics and correlations among key indices in South Asia, identifying trends and outliers that impact regional happiness scores.
- Compare and contrast the happiness metrics between the South Asian and Middle Eastern regions, emphasizing differences in average happiness levels and underlying factors.
- Utilize AI techniques to enhance data visualization and interpretation, providing deeper insights into the relationships between happiness and economic or social indicators.
- Investigate the impact of outliers on regional averages and discuss their implications for understanding happiness in the context of AI-driven analysis.
- vi. Reflect on the findings to contribute to the broader discourse on happiness and well-being, highlighting the role of data science in interpreting complex social phenomena.

3. Data Exploration and Understanding

3.1 Data Preparation and Overview

In brief, we began by loading the "World Happiness Report" dataset to conduct the preliminary investigation. We have been able to construct a clear picture of the structure of the dataset and its main statistical characteristics by analyzing the first 10 lines, which are crucial for understanding the factors that contribute to happiness across different countries. This string of numbers includes data from 143 countries with such feature, using a red-

green-blue scale: happiness score, GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption, Dystopia, and residual. This groundwork exploration will then lead us to a more in-depth study, e.g. how these variables influence a country's overall happiness.

3.2 Descriptive Statistics

In this part, we made a set of statistics of the major traits of the data called descriptive statistics that summarize the data for the user. These statistics data, in general, offer a very clear understanding of the distribution and/or mean of the different variables, thus, they make it a simple task of detecting the trends and the anomalies inside the data.

3.3 Filtering and Sorting

Our next step was to filter the dataset only to include the countries with a happiness rating of more than 7.5 and then have these countries-sort in descending order based on GDP per capita. The most joyful countries were the ones mostly detected in the process while also the less poor people were identified. By rearranging the specimen of 10 countries according to their economic return, it will be easier to comprehend the relation between wealth and happiness.

3.4 Missing Values

In this step, missing values were analyzed in the data set. One of the essential tasks of any data set is dealing with partial data, as it can lead to misinterpretation and draw inaccurate conclusions. Detection and management of missing values is the key to the validity of our conclusions.

3.5 Adding Column

We created the "Happiness Category" column to group the countries into three ways based on their happiness level: low, moderate, and high. Countries in the low category group are those with poor overall quality of life while medium category includes those with just moderate well-being. On the other hand, high-category countries indicate that the population is experiencing good health and happiness. The ranking gives a more precise representation of the spread of happiness among countries and is, thus, a tool to decide which region should be targeted or what policies should be implemented.

3.6 Data Visualizations

Data visualization is the main aspect of the transformation of the data into the visual forms. We depict different types of graphs to perform the analysis:

Bar Charts: These graphs are used to show categorical data by using vertical or horizontal bars.
Line Plots: They are focused on demonstrating correlations and trends between the variables over time.

Histograms: They are a form of graph that shows the distribution of a numerical variable, which means they need to be a numerical variable.

Scatter Plots: This kind of graph, use to show the relationship between two variables (quantitative) and help to figure out if there is a linear relationship between the variables (the correlation of GDP per capita and happiness with the score).

The charts and graphs bring clarity to our understanding of the connections in the data and allow us to gain better insights into the factors that impact happiness.
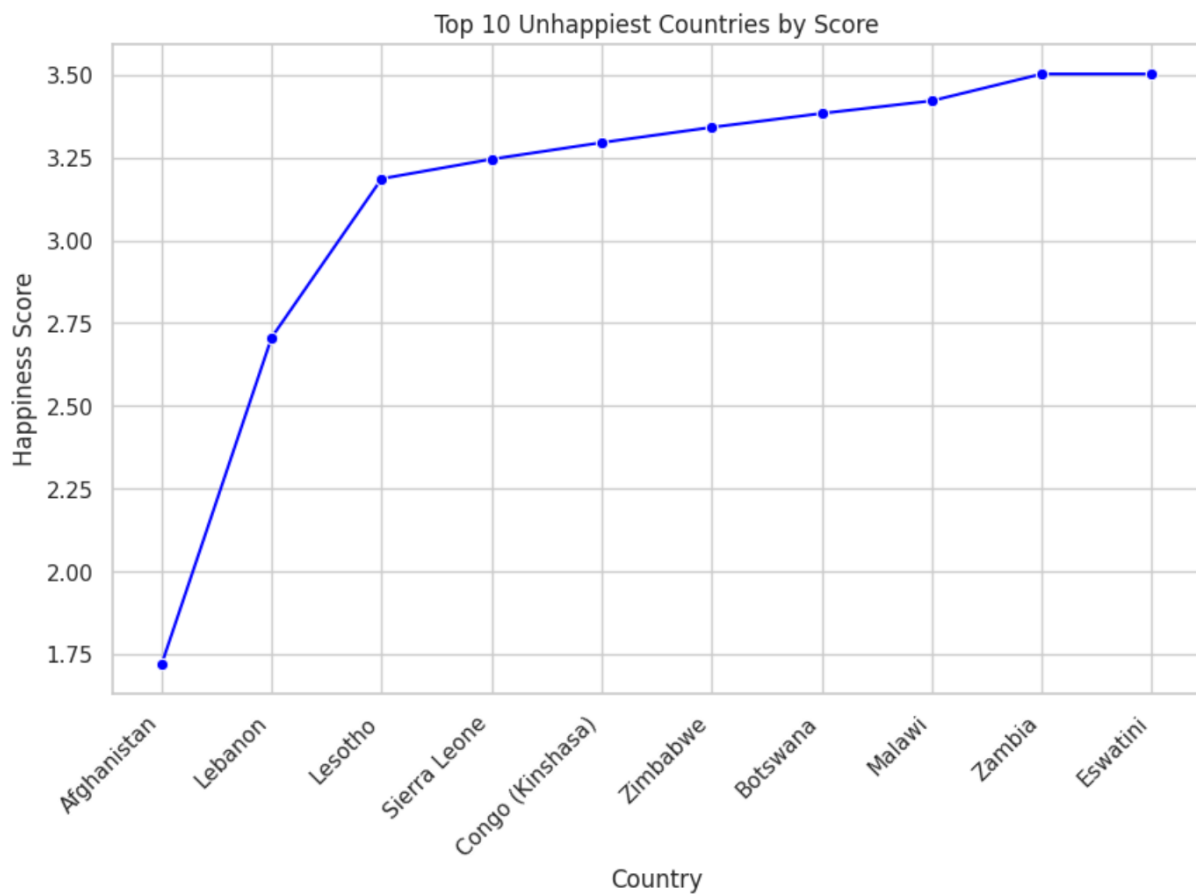
Figure 1 Top 10 unhappiest countries represented in line graph

This is a line graph presenting top 10 Unhappiest countries. Here we can see Eswatini with score of 3.50 representing least unhappiest country in the chart and Afghanistan with score below 2.75 making it most unhappiest country in the chart.
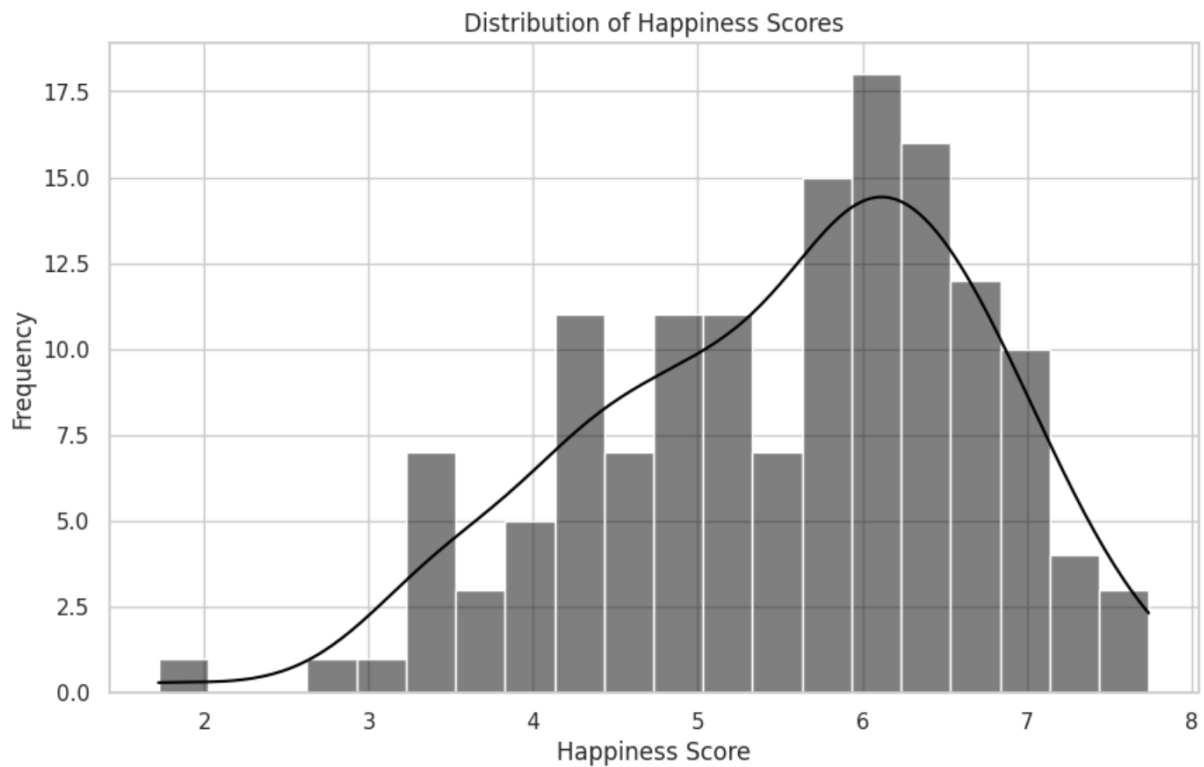
Figure 2 distribution of happiness represented in histogram plot.

The happiness scores of the people from the sample are portrayed in the form of a histogram graph that is made up of bars

According to this chart, the scores that represent the population are mainly spread out. The mode shows that the vast number of people in the population have a score between 6 and 7. Through the graph, it is observed that those that have scores "in the middle," are higher, than other ones close to the margin of the bell curve. Moreover, the indicators "high" and "low" are barely noticed here which infers that the data provides a respectable level of credibility.
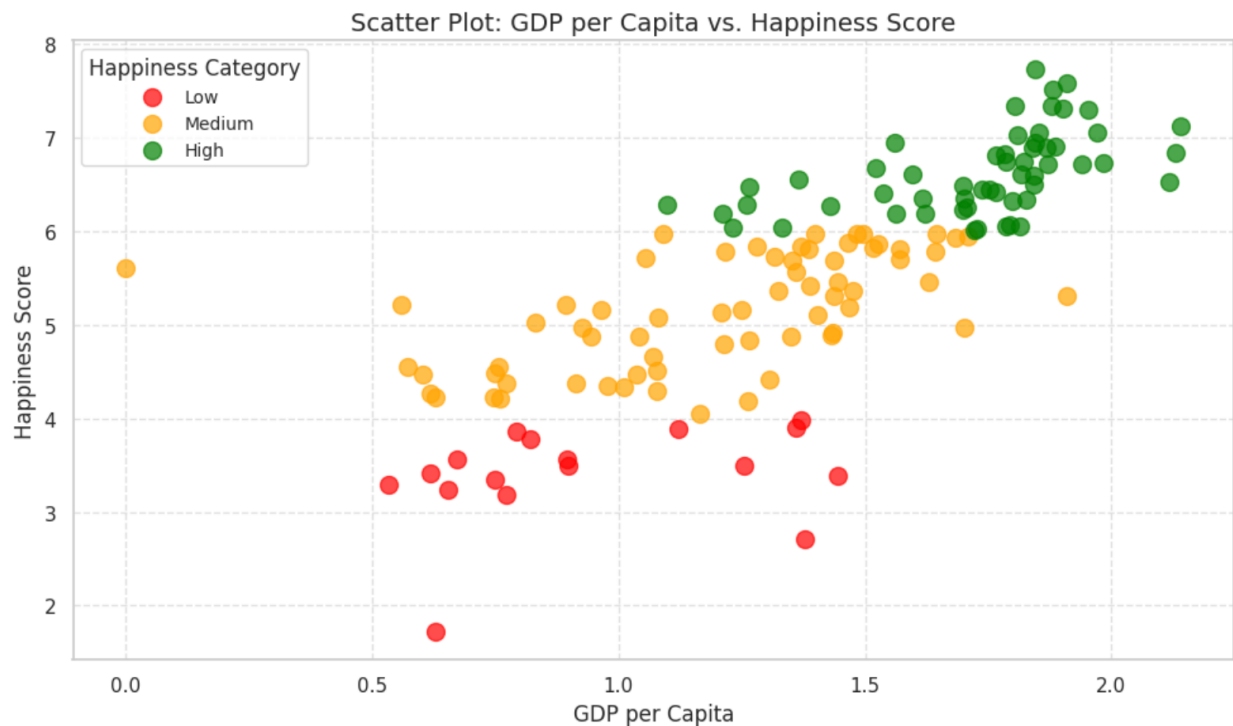
Figure 3 Here's a scatter plot showing how GDP per capita and Happiness Score are related.

This scatter plot illustrates the relationship between a country's GDP and its Happiness Score. The reason is that some nations that have higher happiness scores are not that wealthy in terms of per capita earnings, while some others have lower happiness scores though they can afford more. That is, these two things are not always equal, and by solely looking at the GDP of a country we can show that some people can be misled to measure how happy, or a country is doing.

4. South Asia Analysis

4.1 Data Filtering and Preparation

We started by making a list of South Asian countries. This is crucial because it gives us a reference to filter the data we need. So, we took the whole dataset and only kept the South Asian countries. We saved this in a separate CSV file to make analysis easier. This way, we can easily look at the trends of happiness in different regions of South Asia and quickly compare things.

4.2   Composite Score Analysis

Let's see the data on South Asian countries again from the data filtered before. We decided to introduce a new metric that we give the name 'Composite Score' to help us with the assessment of a country's happiness. This Composite Score comprises three fundamental metrics:
Composite Score: 0.40 x GDP per Capita + 0.30 x Social Support + 0.30 x Healthy Life Expectancy
This score is the right answer to the social, individual, and community-based health outcomes making the situation not too lopsided in any dimension. We will use a graphical method to highlight the best

performers in the region according to these scores. We will use the scatter plot to get the data from the original data and analyze then compare them with the new data.

4.3   Correlation and Outlier Detection

Outliers are the points which seem different from the crowd and do not obey a useful pattern. Usually, they are the reason why our model performance is bad, and our conclusions are not trustful. So, we must realize and exclude them. We are going to take the 1.5 x IQR scale to detect the outliers based on the happiness score versus the method of making money (GDP per Capita). A presentation is going to be made, outlining how the outliers are related clearly with GDP using scatter plots. The presentation marks the case of the outliers with a red color. The anomaly of Afghanistan in the South Asian dataset, which is an extremely challenged economy is accepted as an example, because of the low levels of social assistance, healthcare, and the other states in the region have, this is shown by the graph. It is possible for one abnormal point to be present among them to give a delusion of the real status of the territory bringing the focus of the policymakers to undertake specific interventions.

4.4   Gap Analysis

The new we will include is 'GDP-Score Gap' and the countries will be displayed first in ascending (positive gaps) and then in descending (negative gaps) order. Most of the countries who are on the good side of the scale (positive gaps) are those with a higher income as experienced in the GDP index than in the happiness index. However, when looking at the case of countries with a lower GDP than their happiness scores, a negative family is formed with these countries. The main intention here is to highlight the top 3 countries from each group and then illustrate them via a bar graph graphically.
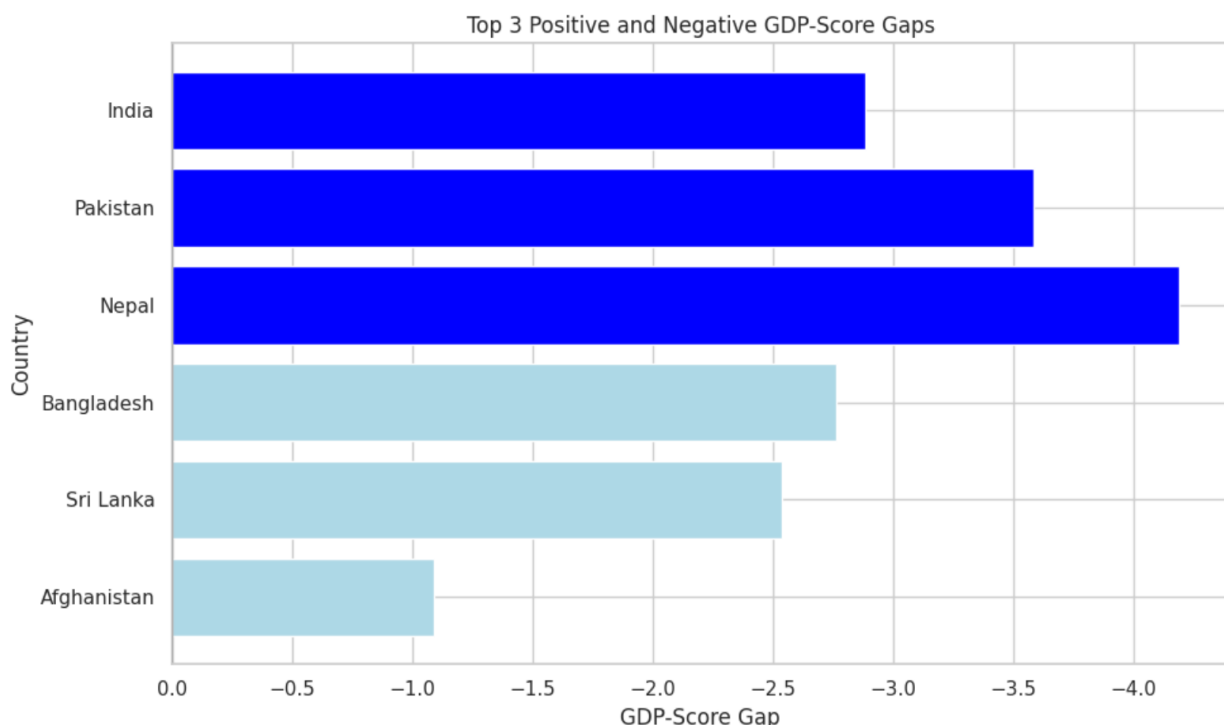

Figure 4 Green bars show the good stuff, while red bars show the not-so-good stuff.

Here are some key points to consider:

Some countries have higher GDPs, which does not translate to better lives. Nepal, Pakistan, and India have such positive gaps. Easy monetary growth does not automatically make people happier. Social safety nets would help bridge these gaps of inequality for economic well-being.

Countries such as Sri Lanka, Bangladesh and Pakistan have lower GDPs and larger Scores. This shows that wellness is not just about making money. Factors like society, culture, and healthcare also contribute to well-being all the same.

5.   South Asia and Middle East Comparison

5.1   Comparative Descriptive Statistics

Let's filter the Middle East data frame and save it in a separate CSV file to make further analysis easier. We'll calculate the measures of central tendency and dispersion for each region. Middle East has higher variability in happiness compared to South Asia.

5.2   Top and Bottom Performers

This section will help us understand the top and bottom performers in South Asia and Middle East.

5.3   Metric Comparisons

We'll create a bar chart to compare three different metrics in each region. Middle East consistently outperforms South Asia in all factors, indicating that South Asia needs to focus on improving all metrics to achieve a better quality of life.

5.4   Happiness Disparity

We'll compare the happiness ranges of Middle East and South Asia. Middle East has a higher happiness range of 4.63, indicating a larger difference between the happiest and least happy countries in the region. South Asia's happiness range is smaller, but the variability in happiness scores is higher compared to the range.

5.5   Correlation Analysis

We'll explore the relationship between happiness scores and two other parameters: 'Freedom to make life choices' and 'Generosity' in both regions. Both parameters have average coefficients of 0.80 and 0.86, respectively. A positive correlation indicates that higher scores in these parameters are associated with higher happiness. The correlation is slightly stronger in the Middle East, suggesting that happiness is more influenced by freedom. In South Asia, generosity shows a stronger correlation, indicating that higher scores are linked to higher well-being.

5.6   Outlier Detection
We'll identify any outliers in the data to ensure the accuracy and reliability of our analysis.

In the previous task, we noticed an outlier in South Asia. But in the Middle East, we don't see any outliers, which means the data are pretty consistent and uniform. This balance helps prevent big differences.

5.7   Visualizations: Score Distributions
When we look at the box plot, we will see that the Middle East region is the one with the highest median score. And the standard deviation (SD) values reveal that the scores are less scattered when compared to the South Asia area. The South Asia region also has a number

of highly unusual cases, with scores that are way too high in contrast to the rest of the distribution. At the same time, the scores for the well-being or quality of life metrics that are measured by the 'score' variable seem to be more in line with each other in the Middle East compared to South Asia.

6. Conclusion

We discovered in the research that various factors are responsible for the well-being of the countries. The Middle East is the area where people are more likely to be happy in general but the countries of South Asia had also done very well in terms of happiness average. Whereas nations in both sub-regions show extreme levels of happiness, those in South Asia exhibit a much wider spread of happiness scores, revealing a more significant disparity between the well-being of citizens. The Middle East, conversely, has a more evenly matched spread of happiness scores. For South Asia to address this issue, the focus must be on economic growth, social support, and health system improvements. The development of economic or social support or health in these regions can be a vital step to narrow the gap between them and consequently raise the overall happiness of people in the region.

Github

https://github.com/anupampoudel02/assignment1