

|  |  |  |  |
| --- | --- | --- | --- |
| **Academic**  **Year** | **Module** | **Assessment**  **Number** | **Assessment Type** |
| 2024 | 5CS037 Concepts and Technologies of AI | 1 | Report writing |

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

Student Name**:** Anish Rajlawat

Student Id: 2408483

Group: L5GC7

Module Leader: Simon Giri

Tutor: Simon Giri

**Plagiarism Screenshot**

**Collab Notebook Link:**

https://colab.research.google.com/drive/1ObEpWEoGNl-2qzsbK-kk-y5fyTJpDs-f#scrollTo=jyzDHy6eCKgO

# **Introduction**

## **1.1Overview**

## World Happiness Report is an annual publication that ranks countries from the world according to their levels of happiness. This report goes into the complex interaction of variables associated with well-being, including economic performance, social support, health, freedom, trust, and generosity.

## This study seeks to fully analyze the World Happiness Report dataset pertaining to the South Asian and Middle Eastern countries. In these regions, it is hoped to find insights about the factors affecting happiness, differences in the two regions, and the policy implications for both.

## **Objectives**

1. **Explore** **and prepare the data**: the World Happiness Report dataset will be loaded, cleaned, and explored with respect to missing values and outliers.
2. **Analyze the South Asia:** what are the factors influencing happiness in the South Asian countries? Check for outliers or anomalies in data.
3. **Compare between South Asia and the Middle East:** happiness comparison across these regions, with an emphasis on convergences and divergences in such happiness and the factors influencing it.

# **2. Report Summary**

This report discusses the complete analysis undertaken on the World Happiness Report dataset. The first problem related to the preliminary exploration of the dataset, which involved statistical analyses and visualization. The second problem examined South Asia, with composite score calculations, outlier detections, and relationships between variables. The third problem compared South Asia and the Middle East on salient indicators, identified best and worst performers, and analyzed distribution of happiness scores.

## **2.1 problem - 1 - Getting Started with Data Exploration:**

In this section, we have introduced the data of World Happiness to acquire some starting insights.

### **2.1.1 Dataset overview :**

The data set on the World Happiness Report was first loaded and used to explore the trends that are taking place among the world's happiness. The data comprise 143 rows and 9 columns, as initially inspected to understand how the data is organized and what it contains. It initially featured columns such as country name, regional indicator, happiness score, GDP per capita, social support, Healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. This was the groundwork for later analysis by looking into the data structure on dimensions and variable types.

### **2.1.2 Basic Statistics :**

By using basic statistical analysis, average and variability were revealed for happiness scores. The average happiness score yielded an overall measure of happiness; while the median could be regarded as a measure of central tendency, it is not significantly affected by outlier values. A score measure in standard deviation was used to express how scattered or clustered scores were from the average score and how much it deviated from the average. The analysis indicates how best the overall distribution of happiness scores would fare and what could be possible patterns and anomalies in the data.

**2.1.3 Missing Values :**

The dataset was thoroughly scanned for missingness as this has drastic consequences on the accuracy and reliability of the analysis. In this case, however, no missing values were found, making things easier for the subsequent data cleaning and analysis.

If there had been missing entries, the mean, median, and mode imputation methods would have been used to fill in those values. Because there aren't any missing values in this dataset, any analyses that could have been done with them are straightforward.

**2.1.4 Filtering and Sorting :**

Through sorting the dataset by GDP per capita, it was narrowed down to highlight countries of extremely high happiness scores, so the idea of finding useful correlations between economic and happy country profiles would remain relevant. In this way, the analysis presents opportunities to investigate aspects associated with the conditions that produce these great levels of welfare.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### **2.1.5 Adding New Columns :**

Happiness Category - A new column entered to categorize countries Major Low Medium and High based on their happiness scores. This makes it possible to conduct a more specific analysis on happiness levels and compare different groups. The exploration of country distribution among these categories with this new column shows potential trends or patterns.

### **A screenshot of a computer Description automatically generated**

**2.1.6 Data Visualizations :**

A graph of blue rectangular bars with white text

Description automatically generatedTo present the findings more effectively, several graphs and charts were prepared. For instance, bar charts were used to exemplify the comparison of high and low countries, and the major differentials in happiness levels. Whereas histogram produced an idea of happiness score distribution for all countries, scatter graphs gave a loose idea of GDP per capita versus happiness score. All this visualization helped see trends, patterns, or outliers that bettered the understanding of factors affecting world happiness.

A graph with a red line

Description automatically generated

A graph of happiness scores

Description automatically generated

A graph with blue dots

Description automatically generated

## **Problem - 2 - Some Advance Data Exploration Task:**

The study hereby investigates some of the things that measure happiness in South Asian countries - developing a composite score, which identifies the outliers and establishing the relationship between variables.

### **2.2.1 Task - 1 - Dataset Overview :**

South Asia's focus geographical filters a full list of South Asian countries that are categorized and applied to the initial set of datasets. Despite all the fuss, it produces a whole new dataset saved for future analysis by allowing only data completely relevant to the Indian South countries. This was done to ensure that the next analysis would entirely relate to the South Asia region only.

A screenshot of a computer

Description automatically generated

### **2.2.2 Task - 2 - Composite Score Ranking**

A composite score was devised for each South Asian country to measure well-being more comprehensively. This composite score was a weighted aggregate of GDP per capita, social support, and healthy life expectancy. South Asian countries were further ranked based on their composite scores. A horizontal bar chart is used to visualize the top 5 countries that have the highest total scores.

A bar graph with numbers and symbols

Description automatically generated with medium confidence

For that reason, a scatter plot was prepared to show the relationship across rankings of both by countries based on their respective composite scores derived from the given indices considered for formulating a composite score in happiness and by that of the original happiness score. These major differences are, therefore, analyzed under this notion to understand their cause or factor.

A graph with points in the center

Description automatically generated

Assessment with a composite score against the original happiness score provides finer understanding of the factors describing well-being in South Asia.

### **2.2.3 Task - 3 - Outlier Detection :**

Outlier detection techniques were applied in recognizing certain countries that boast very high or very low scores on happiness scores or GDP per capita. To defining 'outliers' based on the distribution of data, IQR (Interquartile Range) method was implemented.

A graph with blue dots

Description automatically generated

By graphical means, namely scatter plot, the relationship of GDP per capita and happiness score was investigated. It highlighted the outlier countries on the scatter plot as an attempt to determine potential anomalies. Analysis of these outliers would reveal some useful lessons on factors leading to their outlier positions. For instance, countries have a remarkably high score on happiness while it has a low GDP per capita; it could be argued then that there exist forms of social support or other cultural factors contributing to well-being. In contrast, a country that has a good GDP per capita and yet low score on happiness is expected to have other issues of inequality, corruption, or political instability.

In these ways, one learns about the outlier, which rewards the study with more intensive refinement and direction for potential analysis.

### **2.2.4 Task - 4 -Exploring trends across metrics :**

As part of this research work on understanding the relationship between happiness and different factors, an attempt was made to explore the correlation between happiness scores and two critical factors; Freedom to Make Life Choices, Generosity and so on.

A graph with green dots

Description automatically generated

A graph with orange dots

Description automatically generated

The measurement calculated the Pearson correlation coefficient to assess the direction and strength of the linear association of these two variables with the happiness scores. Thus, scatter plots were developed here to show these relationships and show prevalent trends.

By analyzing their correlations and visualizations, some of the strongest and weakest correlations between these indicators and happiness scores can be identified, thus providing information about variables that strongly influence happiness levels in South Asian countries. Such as a strong positive correlation between happiness and freedom from making life choices as indicating that personal freedoms are thus a very important contributor to overall well-being.

### **2.2.5 Task - 5 - Gap analysis :**

The difference between an economy's level of prosperity and the level of people living happily in it has a newly created column "GDP-Score Gap," i.e., the difference between a country's GDP per capita and its happiness score.

A green rectangular bars with white text

Description automatically generated with medium confidence

South Asian countries have been ranked based on this gap, central as well as peripheral; for example, the countries where the gap turned out to be more significant in the context of positive and negative measures. The three richest and the three poorest are respectively represented through bar graphs.

A graph with red squares

Description automatically generated

A close investigation into the said gaps should reveal the extent of these disagreements between an economy on the one hand and happiness on the other. In other words, countries with large positive gaps likely experience something like inequality, corruption, or political instability that hinders economic growth from being translated into the increased happiness of-income citizens. In contrast, countries exhibiting huge negative gaps possess strong social support systems, cultural factors such as expectations regarding happy life, or effective governance.

Understanding these gaps is crucial since they could highlight the hazards for the policymaker or researcher working in areas where possible interventions may be needed to improve growth-happiness interactions.

## **Problem 3**

This section is the comparison of happiness levels and their influencing factors between the regions of South Asia and the Middle East. The following activities will be among some of those included:

### **2.3.1 Task - 1 - Data overview :**

For comparative studies to be made between South Asia and another part of the world, a piece of that data subset was restricted to Middle Eastern countries. A definite list of Middle Eastern countries was defined: Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Syria, United Arab Emirates, Yemen. Using this filtered dataset also constrains an analysis specific to the Middle Eastern region.

A screenshot of a computer

Description automatically generated

**2.3.2 Descriptive Statistics :**

In fact, both the mean and standard deviation in happiness scores is evaluated for both the regions South Asia and the Middle East to get an overall picture of happiness in the two regions.

A close up of a number

Description automatically generated

Mean: Compared to South Asia, the Middle East's average happiness score (5.41) is significantly higher (3.89). This indicates that on average, people of the Middle East are happier than those of South Asia.

Standard Deviation: The fact that both regions have virtually similar standard deviations implies the same level of non-homogeneousness in the happiness scores independent to any geographical boundary. In other words, while happiness averages higher in the Middle East, variation in happiness levels within each region is similar.

**2.3.3 Top and Bottom Performers :**

In the creation of the top and bottom-performing countries under each region, the top 3 and the bottom 3 identified countries were then compared with their happiness scores, and bar charts were produced as visually comparable graphs between these countries. For the Middle East region, the top three countries were Israel, United Arab Emirates, and Qatar, while Yemen, Syria, and Iraq were at the bottom. In South Asia, the top three countries were Bhutan, Maldives, and Sri Lanka, whereas the bottom countries included Afghanistan, Nepal, and Pakistan. Visualization thus provides a clear compare of happiness levels between the two regions while indicating the nations with the most pronounced disparities in happiness measures.

**A graph with red and green bars

Description automatically generated**A red and green bar graph

Description automatically generated

**2.3.4 Metric Comparisons :**

A graph of values of a group of individuals

Description automatically generated with medium confidenceThese are differentially accounted: for such prime indicators as GDP per capita, social support, and healthy life expectancy. This is compared in grouped bar charts. The Middle East region shows an increased GDP percentage when compared to South Asia, indicating that citizens in the Middle East are, on an average, economically very wealthy. While the social support is equally distributed similar in both regions, slight changes have been exhibited in favor of the Middle East. The average healthy life expectancy in the Middle East is longer than that of South Asia. Such discrepancies in economical as well as health outcomes are indicative factors in measuring the difference in happiness levels for these two areas.

**2.3.5 Happiness Disparity :**

To measure the dispersal of happiness scores within each region, range, and coefficient of variation (CV) values were computed. For a wider range of happiness score differences between people, the Middle East ranked the highest. The Middle East has only insignificantly higher CV than the former regions compared to other areas, meaning that both sorts of regions generally illustrate a similar level of relative variability concerning happiness scores. This indicates that, although the happiness scores from the Middle East are much wider ranging than other areas, the relative dispersion of scores within the region is similar.

### **2.3.6 Correlation Analysis :**

To determine correlations with happiness, we considered happiness scores in relation to two important variables namely Freedom to make life Choices and Generosity. Both resulting correlations indicated an unlikely strong positive correlation between happiness scores and Freedom to make life choices, suggesting that happier countries would likely be freer countries and indeed the strongest correlation of all was found between these two factors. Happiness scores showed similar patterns of correlation with Generosity, suggesting countries highest on the generosity scale would have higher happiness among their residents. Scatter plots were employed to visualize these relationships and grant further evidence of the positive associations between these variables. Hence, through an analysis of these correlations, the region would then be able to examine much deeper into the very issues contributing to happiness within the Middle East region.

A graph with blue and green dots

Description automatically generatedA graph with blue and green dots

Description automatically generated

### **2.3.7 Outlier Detection :**

I used an outlier identification process via the 1.5×IQR rule to identify the outlier countries in Score and GDP per Capita in both regions. From scatter plots, outliers such as Qatar and UAE in the Middle East had GDPs that were significantly higher but Scores that were somewhat lower. Furthermore, countries with much lower values for both metrics were Yemen and Afghanistan. The outliers represent very different socio-economic conditions about happiness.

### **2.3.8 Visualization :**

A graph with a box and a line

Description automatically generated with medium confidenceThe box plots comparing the score distributions between South Asia and the Middle East show strong differences. Very individual differences between the Middle East, which had a skewed distribution with higher mediation happiness and more outliers, and South Asia, with shorter and lower medians and reflected its more homogenous happiness levels but lower in comparison.

From this perspective, therefore, the Middle East has generally performed better than South Asia in happiness and economic measurements, even as the differences have been larger. South Asia remained more similar in terms of self-rated happiness despite the many challenges it faced but had lagged on crucial economic and social parameters. Such findings amplify the nexus between economic, social, and cultural dimensions

# **3.Conclusion**

The comparative analysis focusing on happiness levels and their determinants between South Asia and the Middle Eastern region reveals huge contrasts. For instance, there is generally higher happiness among people in the Middle East, which is shown in the higher GDP per capita as well as greater correlations of happiness to freedom, generosity, and the lower variability of happiness scores. In contrast, South Asia deals with challenges of lower economy indicators coupled with possible social and political problems and thus, variable happiness scores.

Such differences between the regions might help policymakers and researchers pinpoint specific areas that may be improved and where impactful future strategies concerning both regions may be drafted. More research is required to understand in detail, the factors influencing happiness in each region as well as intervention opportunities for the arising issues.