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 : 2438424
Student Name : Bhawana Thapa
Section : L5CG4
Module Leader : Mr. Siman Giri
Tutor : Mr.Bibek Khanal
Submitted on : 12/28/2024

Contents

Table of Figures

1. Introduction

The World Happiness Report is a cross-sectional survey that is conducted cross nationally with the aim of ranking nations depending on the level happiness in a given nation. It is prepared annually with the collaboration of the United Nation Sustainable Development Solutions Network and the data is based on the results of the surveys of people all over the world. Some are, Gross Domestic Product per capita, ability to choose how to live, views on corruption, life expectancy without disabilities, social support and so on. It is most useful to know what makes or reduces human's well-being with calls to action and develop the environment that will enhance quality of life.

It is more important as a kind of document which is envision to serve to the developments of the well-being of society. Thus, it actually points out the direction that needs to be developed by qualifying those countries depending upon the reported nation level of happiness. It also reveals the economic, social and the environmental background of the happiness range and explains to this specific nation where it must get to the next kind of welfare. This paper depicts an understanding of how the growth is regrouped within the region.

2. Objectives

Therefore, this paper will endeavor to expand on regional concepts of happiness with special reference to the South Asian and the Middle East region. The primary objectives include:

i. Conduct a data analysis and try to find out how much does the country have GDP per capita, how much social support does it have and so on.
ii. Explore statistics, correlations and outliers for major indices in South Asia.
iii. Comparison between two regions: Middle East and South Asia based on average happiness numbers.

3. Data Exploration and Understanding

3.1 Data Preparation and Overview

To build the analysis, we first start with loading the dataset that is called ‘World Happiness Report’. By obtaining the first 10 rows, there is a possibility to understand the general statistics of the dataset, which is helpful to analyze the reasons for the happiness in countries. As exploring with rows and columns, we can see 143 different countries with 9 different parameters including score, GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, perceptions of corruption, Dystopia and residual.

3.2 Descriptive Statistics

Next task is to make a descriptive analysis of the score column to obtain a better understanding. These included use of mean, median and standard deviation.

The mean which is also called the average is a measure of the degree of center in the distribution of the scores. The mean happiness rating in all countries stands about 5.53. This implies, the average happiness level across the world is slightly above the midpoint. From here it is possible to examine which countries are below or above the average to a greater extent.

Moreover, the standard deviation measures the extent of spread or spread out of the score.
Small standard deviation means most of the countries have nearly equal happiness level while high standard deviation means few countries have very high and may have very low level of happiness.
The SD of 1.17 means that most of the scores of country happiness are from approximately 1.17 higher or lower than the average which means there is not a significant difference in the happiness level in different countries but also not very low variation in the scores given.

3.3 Filtering and sorting

Here we filtered the countries based on score greater than 7.5 and sort 10 of them in the descending order based on GDP per Capita. We can observe that these filtered countries are the happiest one. Meanwhile, sorting makes it easier to identify the wealthiest countries among the happiest ones. Sort the top 10 countries who has the better performances in terms of economy.

3.4 Missing values

Simply in this step we checked for the missing values in the dataset since incomplete data often leads to distorted conclusion.

3.5 Adding column

New column is added called ‘Happiness Category’. Countries are categorized in three different levels low, medium and high. Those who falls under the low, means they are experiencing severe problems with their general quality of life whereas medium means moderate. However, high belong to high well-being state. Categorization helps to show the distribution of the happiness levels cross-sectionally between countries, which may be useful for further work at the regional or policy level.

3.6 Data visualizations

It is a process of converting data into visual formats necessary in data visualization and exploration. We are working on different plots. A bar chart refers to a rectangular representation of data. While, line plot focus on the trend and correlation. A histogram is a graphical method of presenting data on the distribution of a numerical variable. Scatter plots, two quantitative variables are related with one another, and they show the strength of the associations. It helps here to analyze the relation between the variable GDP Capita and the happiness score.

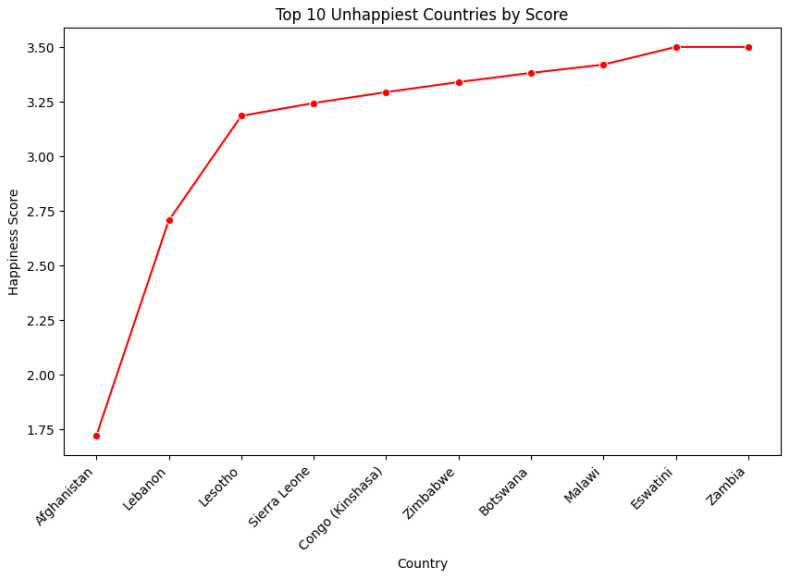

Figure 1: Line plot showing the top 10 unhappiest countries

This diagram presents the Top 10 Unhappiest countries ranked by Happiness Score, with Afghanistan having the lowest score around 1.8 and making it the unhappiest country. The score generally increases from left to right up to 3.4 for Zimbabwe.

Figure 2: Histogram plot showing distribution of happiness scores

The above diagram represents how Happiness Scores are spread out in a population. The curve in above graph shows that the average or mode of the data is around the score 6. Moreover, it is clear that majority of the people in the sample population have relatively "normal" scores, with very few scorings either really high, or really low.

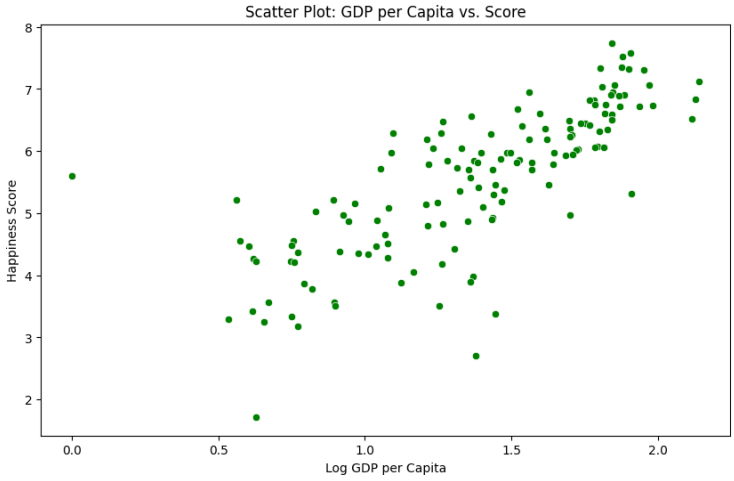

Figure 3: Scatter plot showing interaction between GDP per capita and Happiness Score

The scatter plot draws out the interaction between the GDP and the Happiness Score of a country. Some countries have higher happiness scores though they have lower levels of income per capita, and other simply have lower happiness levels though their level of economic development is way better. Meaning, both metrices are not parallel and this may mislead if one will only use the GDP account to measure the improvement of a nation and the quality of life of its people.

4. South Asia Analysis

4.1 Data Filtering and Preparation

Initially, we created the list of the South Asian Countries. This step is important to ensure that we have a reference we will use to filter the data that we need. So, from the overall dataset we only filtered out or extract the South Asian countries and saved in another CSV file to make analysis easier without having to reapply the filter again. As a result, it enables us to look at the trends of happiness around regions in South Asia as well as quicker access for tasks and comparison.

4.2 Composite Score Analysis

Access the data of South Asian from previous filtered data frame. Next, we introduce 'Composite Score' metric to the data frame which assist to evaluate these countries based on selected happiness. Composite Score consists of three key metrics:

Composite Score: 0.40 x GDP per Capita + 0.30 x Social Support + 0.30 x Healthy Life Expectancy

It makes a certain balance of evaluating the economic possibility, the social needs, and the health outcome. With a visual representation, we determine the top high performers in the region considering these scores. In order to compare it with original score, the scatter plot is made to examine the variations.

4.3 Correlation and Outlier Detection

In a simple word, outliers are points that are far from other data points and does not lie on the curve of data set. Outlier analysis enhance model performance and developing solutions, and ensuring validity of conclusions made on data. Here, we define the outliers using the 1.5 x IQR rule based on the happiness score and economic indicator (GDP per Capita). Through the scatter plot we visualize the outliers and relationship between GDP and score. Red highlighted refers to the outlier in the graph. Afghanistan is the outlier in South Asia data frame with exceptionally low/poor economic condition, social support or health care compared to the regional trend. One outlier can affect the representation of overall region. And also, drag the attention of the policy makers for special intervention.

4.4 Gap Analysis

Adding a new column 'GDP-Score Gap' and manipulate with both ascending (positive gaps) and descending (negative gaps) ranking. Countries with positive gaps seems to have higher GPD compared to the score and vice-versa to the countries with negative gaps. In this task, we extract the top 3 countries from each and visualize through bar graph.

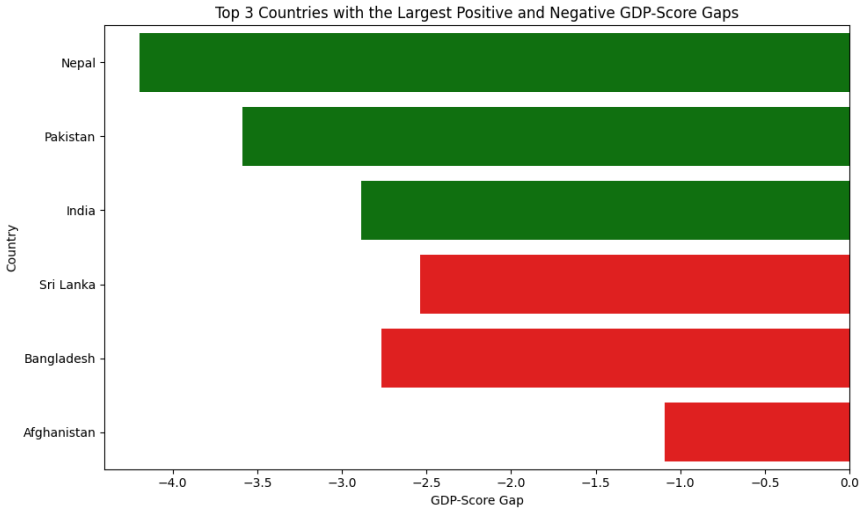

Figure 4: Green bars represent the positive gaps while red represents the negative.

i. Positive gaps: As I already mentioned, the green bar countries Nepal, Pakistan and India have comparatively greater GDP in respect to their actual score. It shows that economic growth does not necessarily equal well-being. To resolve it inequality, social safety must be address.

ii. Negative gaps: Similarly, countries with red bars Sri Lanka, Bangladesh and Pakistan have a lower GDP but a disproportionately higher Score. This suggests that the countries well-being is not only dependent on economic factors. Hence, importance must be given to non-economic factors such a culture, community support and healthcare.

5. South Asia and Middle East Comparison

5.1 Comparative Descriptive Statistics

From the actual dataset filter the Middle East data frame as like South Asia and save it in another CSV file to make further analysis easier. We performed the statistics operation to determine the measures of central tendency and dispersion of data respectively. With comparison Middle East has higher variability in happiness.

5.2 Top and Bottom Performers

With this operation we have a clear understanding about the top and bottom happiness performers in South Asia and Middle East.

5.3 Metric Comparisons

Three different metrics comparisons have been done with the bar chart. In every factor, Middle East is excelling over the South Asia. As a result, these disparities indicates that South Asia really needs to focus on every metrics to achieve quality of life.

5.4 Happiness Disparity

Based on the comparison Middle East has higher range of 4.63 greater than South Asia which indicates that there is higher difference between happiest and the least happy country in Middle East. On the other hand, the range is South Asia is not as large like the scores but the variability comparing to range is more, if compared with the mean value.

5.5 Correlation Analysis

In this analysis, we simply showing the relationship of score with two other parameters called ‘Freedom to make life choices’ and ‘Generosity’ in both regions. Both have average 0.80 and 0.86 coefficients respectively. A positive correlation implies that the freedom of making life choice increases with the scores. This correlation is slightly stronger in the Middle East meaning score is more affected by the freedom. In case of generosity, South Asia show the strong correlation meaning the higher the scores, the higher the well-being.

5.6 Outlier Detection

We had already seen the outlier of South Asia in previous task. However, in Middle East we don’t see any outlier means the data are relatively uniform and consistent. This balance really helps to prevent large disparities.

5.7 Visualizations: Score Distributions

Observing the box plot we get to know that Middle East area has a higher median level, and the SD values indicate that scores distributions are less spread out compared to the South Asia area. Furthermore, South Asia shows a few outlier scores that are considerably higher than the rest of distribution. Overall well-being or quality of life metrices captured by the ‘score’ variable tend to be more consistent in the Middle East compared to the South Asia.

6. Conclusion

Overall, we see that the countries well-being is dependent on different metrices or factors. Throughout this analysis, we observe Middle East has outshone South Asia in average happiness scores. While both regions have disparities between the top and bottom countries, South Asia has a higher variation of happiness scores around its mean, reflecting unequal distribution of well-being. The Middle East, though having a greater range, is more consistent and uniformed. To bridge this gap, South Asia needs to work on economic growth, social support and health system. Addressing these areas can help reduce disparities and improve overall happiness across the region.

Link to GitHub: Click Here