Academic Year	Module	Assessment Number	Assessment Type
5	Concepts and technologies of AI	1	report

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

Student Id : 2408961
Student Name : Samir-Pemi-magar
Section : L5CG13
Module Leader : Mr. Siman Giri
Tutor : Ms. Durga Pokharel
Submitted on : 12/30/2024

Contents

Introduction

World happiness report (WHR) is a file in which the happiness level of many countries are listed. In this report we are provided with the name of Country name, score, Log GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, Perceptions of corruption and Dystopia + residual. In this assignment we dive more into south Asian countries and middle east countries analyzing their happiness level.

Objective

- Explore the data set and analyze the important part
- Analyze the happiness level
- Compare the south Asian countries with middle east countries to identify the key difference between them

Problem 1: Data Exploration and Understanding

Dataset overview

- the WHR is imported through drive and 10 no of rows are displayed through help of .head().
- In WHR total of 142 rows and 9 columns are detected through the help of shape[] where 0 stands for row and 1 stands for columns
- All the columns are listed along with their datatype and concatenated in the print statement to display to user

Basic Statistics

- Mean median and standerd deviation have been calculated and concatenated to the print statement to display to user
- Country with highest and lowest happiness score have been identified with the help of idxmax() and idxmin() along with the use of loc.

Missing values

- Checked for missing values with the help of isnull().sum and displayed to the user

Filtering and sorting

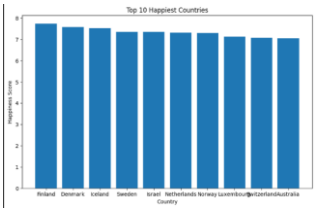
- Dataset were filtered to show only those countries with a score greater then 7.6 with the help of loc including condtition (>7.5).
- The filtered dataset were sorted in descending order by gdp per capita with ascending = false and displayed only top 10.

Adding new columns

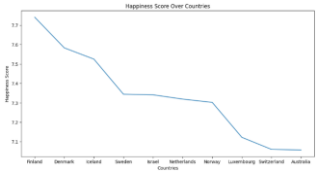
- New column was created with lambda which categorizes the countries into 3 categories: Low − (Score < 4) Medium − (4 ≤ Score ≤ 6) High − (Score > 6).

Data visualization

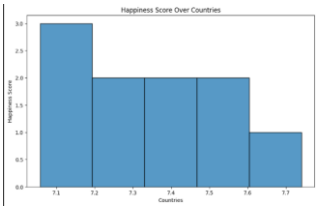
- **Bar plot**: Top 10 happiest countries by score we plotted in bar chart in descending order with the help of plt.



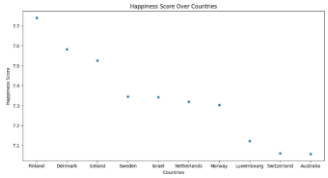
- **Line plot**: top 10 unhappiest countries by score we plotted using a line chart with the help of sns.



- **Histogram:** a histogram was made for the score columns to show its distribution.



- **Scatter plot:** scatter plot was made to plot the relation between gdp per capita and happiness score.



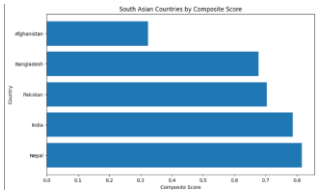
Problem - 2 - Some Advance Data Exploration Task

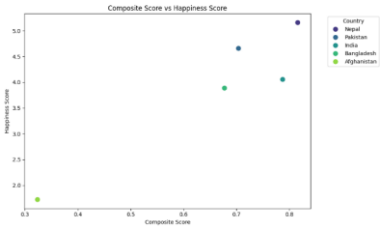
Preparing the South-Asia Dataset

South Asian countries was created with the list of south Asian countries name such as Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri lanka and saved to csv file with the help of .to_csv.

Composite Score Ranking

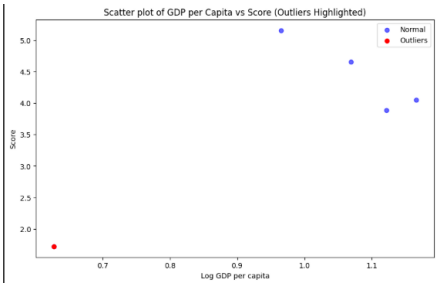
Created a new column called Composite Score and visualized the top 5 using horizontal bar chart. The ranking was also compared with the original and analyzed and discussed whether the ranking align with the original score. Some visual plot were also provided for better understanding i.e. scatter plot were made and the correlation between them were also printed. The composite score were found to be 0.94 which is close to 1 indicating strong relationship between them.





Task - 3 - Outlier Detection

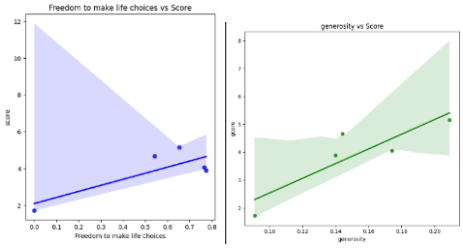
- Outliers countries were identified based on their score and gdp per capita with the use of 1.5 × IQR rule.
- Scatter plot were made with GDP per Capita on the x-axis and Score on the y-axis, highlighting outliers in a different color.



- The characteristics of these outliers and their potential impact on regional averages were discusses i.e how it may either get the value of average of region higher or lower.

Task -4 - Exploring Trends Across Metrics

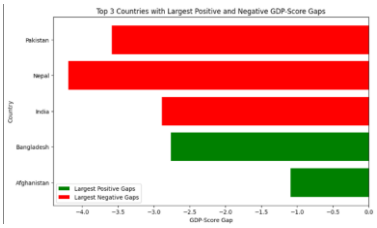
- Two matrices were chosen and their correlation with score were calculated and displayed to the user. The correlation between them were calculated with the help of .corr().
- Scatter plot with trend lines were created with the help of sns.regplot.



- the strongest and weakest relationships between these metrics and the Score for South Asian countries were identified. Metrics having higher correlation have the strongest relationship with each other. correlation between the freedom to make life choices and score is lower than that of generosity and score correlation which are 0.80 and 0.88 respectively indicating generosity and score correlation have higher relationship.

Task - 5 - Gap Analysis

- a new column, GDP-Score Gap was added which is the difference between GDP per Capita and the Score for each South Asian country.
- The countries were ranked according to gap in ascending and descending order with the use of .sort_value and ascending = True for ascending and ascending = False of descending.
- Top 3 positive and negative gaps were visualized in bar chart in horizontal using plt.barh and .head(3) for accessing top 3.



- The reason behind these gaps and their implication on south Asian countries were analyzed and described

Problem - 3 - Comparative Analysis

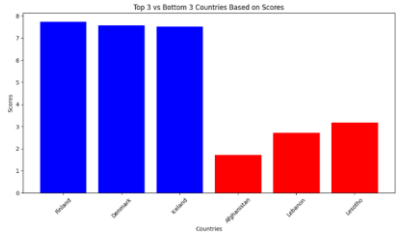
a data frame from middle east countries were made and filtered according to task 2 question 1. A csv file of middle east countries was also made.

Descriptive Statistics

Mean and standard deviation of middle east countries and south Asian countries were calculated and displayed with the help of .mean() and print. After analyzing the mean of both countries the region with higher happiness score is measured.

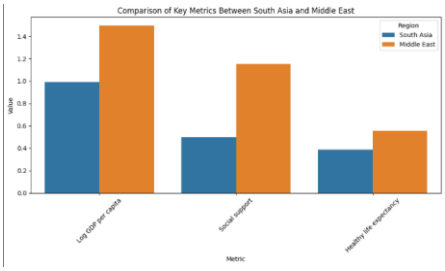
Top and Bottom Performers

Top 3 and bottom 3 in each region were identified and displayed to the user with the help of .head(3) but keeping top in descending score while bottom in ascending score. Both were concatenated and stored and stored in data frame and then the bar chart of the combined data frame is displayed.



Metric Comparisons:

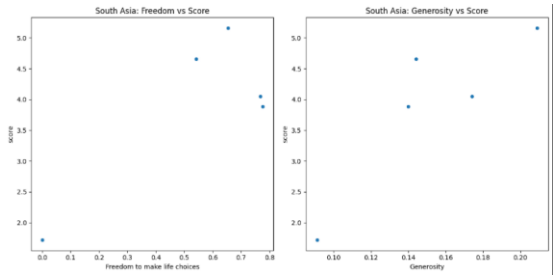
- Compared metrics like GDP per Capita, Social Support, and Healthy Life Expectancy between the regions using grouped bar charts using .mean().reset_index(). Concatenating both and storing them in one data frame, the bar chart of the data frame is created.



- After analyzing country with greater disparity was detected and explained

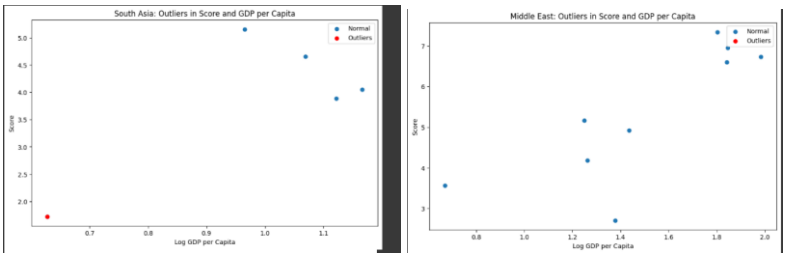
Correlation Analysis

- the correlation of Score with other metrics Freedom to Make Life Choices, and Generosity within each region were alanyzed with the help of .corr()
- a scatter plot was created visualize and interpret the relationships with the help of sns.scatterplot().



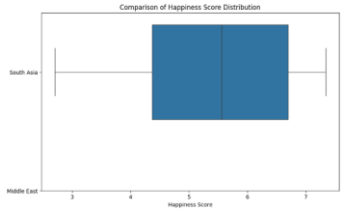
Outlier Detection

- outlier countries in both regions based on Score and GDP per Capita were identified with finding the interquartile range and then finding the lower bound and upper bound and comparing the score with the lower bound and upper bound .
- these outliers were plotted with the help of plt.scatter().



Visualization

- With the help of sns.boxplot, boxplots comparing the distribution of Score between South Asia and the Middle East was created.



- The key difference between the distribution shape median and outliers are identified and discussed about. distrubution of the shape is the shape which shows us how the values are spread out. it shows us wether the value are skewed to left or right, symmetrical. while median is the middle value of the order in an arranged data in order while in any other data is it something that represents the centre of the value. while outliers are data that have way higher or lower data value then that of other values.

Conclusion

Through this we get the insight of world happiness data. We can understand how the data are calculated and the space for more development in happiness level. With the help of this analysis we can see the difference between the south Asian countries and middle east countries and what may be the cause for such difference. Through such understanding we can find more better way to cope to this and create a better country with higher happiness level.