**Herald College, Kathmandu**

****

**Concepts and Technologies of AI**

**5CS037**

Assignment-1 - Statistical Interpretation and Exploratory Data Analysis

Analysis of the World Happiness Report: A Data-Driven

Exploration of Global and Regional Trends.

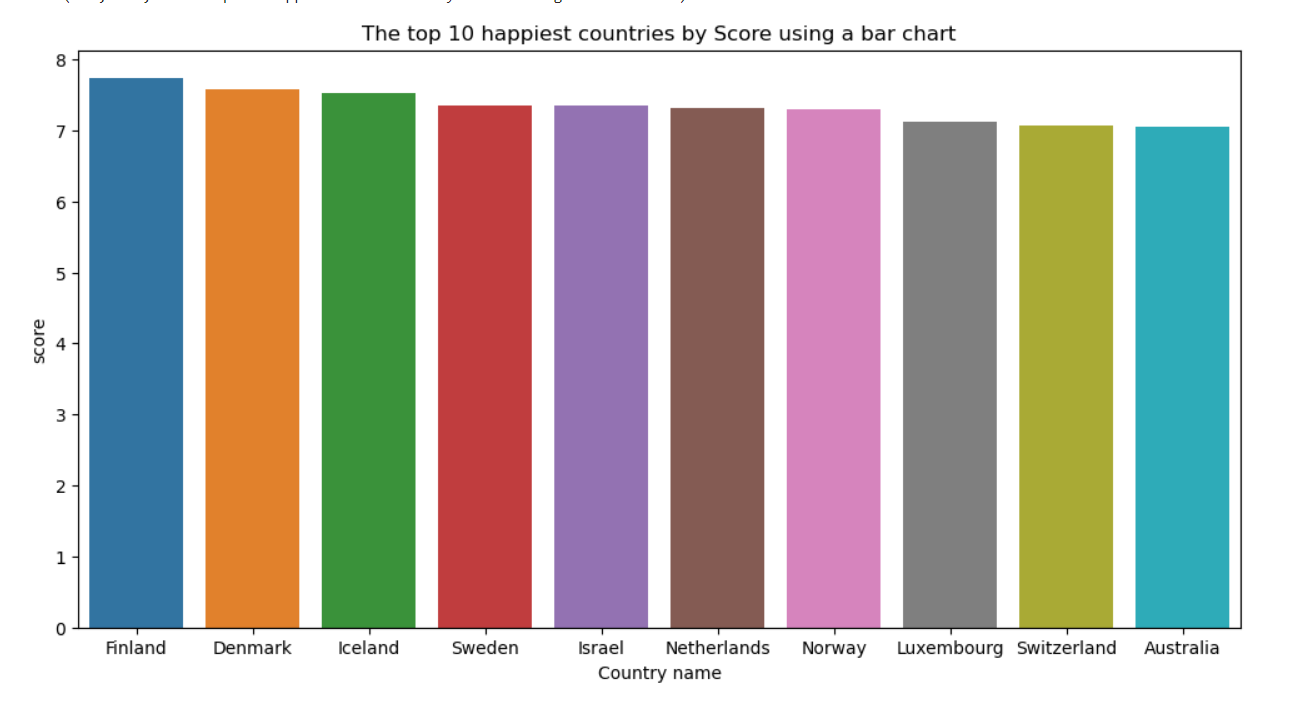
December 02, 2024

1. Interpretation for the Numerical data of original dataset (WHR-2024-5CS037.csv)

* The given data set is about the details of the country and their important indexes i.e. Score , GDP per capita . There are many countries. When the shape of the dataset is calculated the shape was found to be (143 rows & 9 columns).
* The repeated data types of the given original dataset were found to be ‘float64’ and the unique data type was found to be Object that is ‘Country name’.
* While getting the basic statistics data for the dataset, the following the dates were found:
  + Count : 143
  + Mean: 5.52
  + Standard Deviation: 1.17
  + Minimum value: 1.721
  + First Quartile: 4.72
  + Second Quartile: 5.785
  + Third Quartile: 6.416
  + Maximum Value: 7.741
* The highest score on the original dataset is 7.741 which happens to be of Finland.
* The lowest score on the original dataset is 1.721 which happens to be of Afghanistan.
* While checking for any missing values in the dataset , it appears that all the column has 3 missing values except for the column ‘Country name’ & ‘score’.
* While filtering for the score whose value were greater than 7.5 was found to be only of three countries i.e. Finland, Denmark, Iceland.
* We are now adding the new column ‘Happiness\_category’ to the dataset which categorizes on the basis of score. We made a function which categorizes the county happy or not according to the score it has.

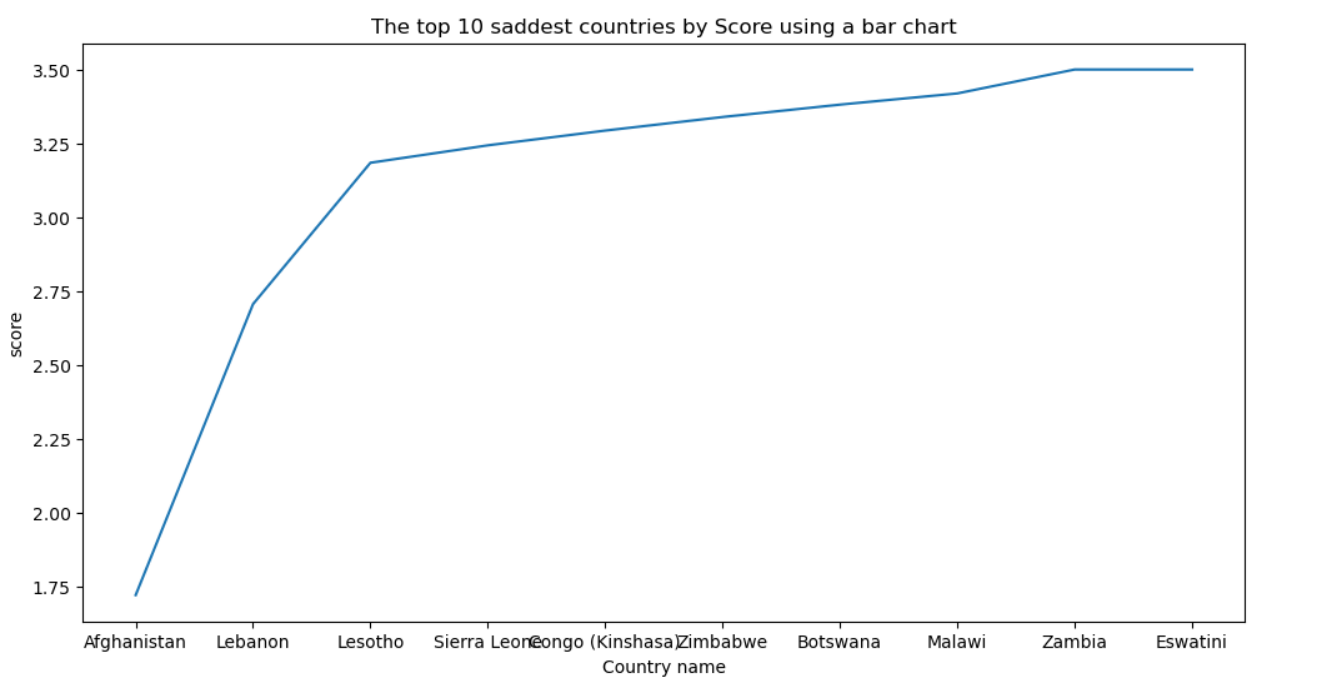
1. Interpreting the visualization of original dataset (WHR-2024-5CS037.csv)

While plotting the most happiest country from the dataset , the followings were my founding from analyzing the data.

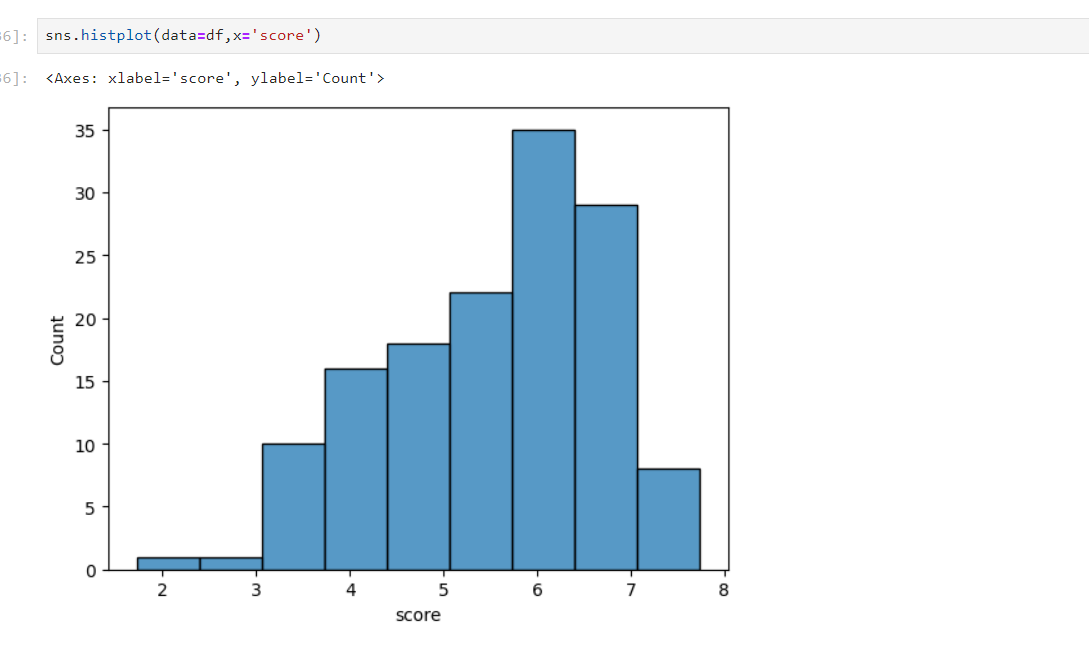


According to the above bar chart, Finland is the most happy country with the score of ‘7 – 7.8’.

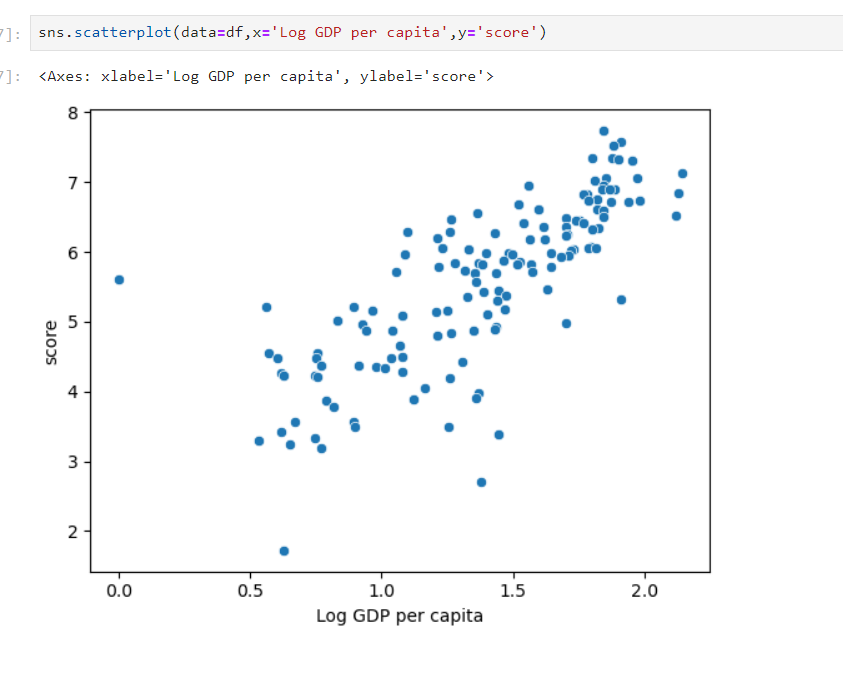
While plotting the line plot for the saddest countries in the dataset, the following seems to be the generalization,



According to the line plot, the saddest the country from the dataset, Afghanistan is the saddest country with score 1.6.



According to the above hist plot, The most common score in the data frame is somewhere between 6 – 7.



From the above scatterplot, it can be seen that the relationship between ‘Log GDP per capita’ and ‘score’ seems to be somewhat linear.

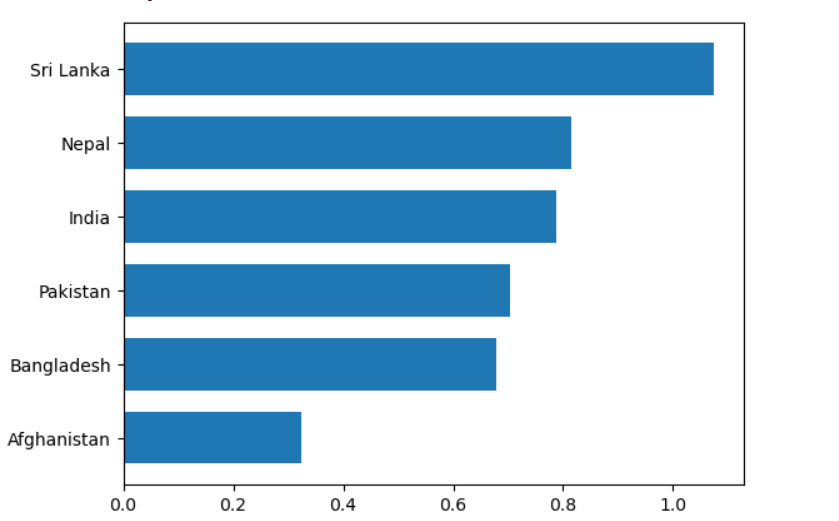
1. Analysis of the filtered dataset which contains south Asian countries.

* At first, I created the function to checks whether the given countries are in the Data frame or not and push into the new csv.
* A new column called Composite score was created by using the formula,

Composite Score = 0.40 × GDP per Capita + 0.30 × Social Support

+ 0.30 × Healthy Life Expectancy.

* The new dataset was sorted in a descending order and was plotted in a horizontal bargraph



According to the above barplot , Sri Lanka is at the top as it’s Composite is high in dataset of the given South Asian Countries.