5CS037 -

Concepts and Technologies

of Artificial Intelligence

Assignment 1:

Statistical Interpretation and

Exploratory Data Analysis

Name: Nayanika Dubey

University Student ID: 2408008

Group: L5CG14

Lecturer: Mr. Siman Giri

Tutor: Ms. Sunita Parajuli

Table Of Contents:

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

Introduction

The World Happiness Report is a report developed taking into account 6 variables, each of which correlates to the overall happiness of people. It discusses the science behind happiness and provides a statistical view on the happiness variance worldwide.

This report provides leading governments to create a base for building new policies and strategies that enhance the overall happiness of the people residing in the nation. It also shows how each variable in the overall study affects people and which sector needs improvement.

In this report, we will discuss the steps taken, libraries used and the concepts applied to analyse and interpret the world happiness data. The report will primarily address the course of action applied to complete this assignment, the challenges faced and the resulting knowledge-base development.

The report will discuss each of the task problems in its respective subsections.

Problem 1

In the first problem of the assignment, we tackled data exploration through a csv file. It was a pretty basic data analysis task, requiring basic panda methods. For the data visualization portion of the task, we implement matplotlib.pyplot functions to generate line graphs, barcharts, scatterplots and so on.

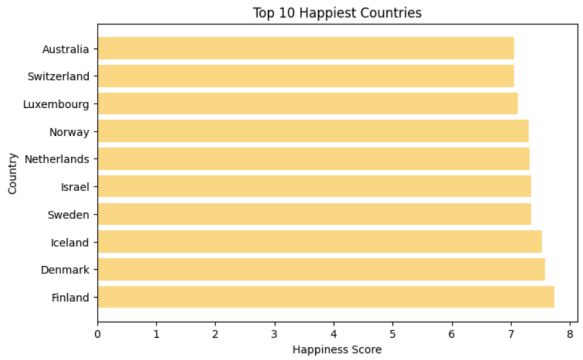
Task Flow

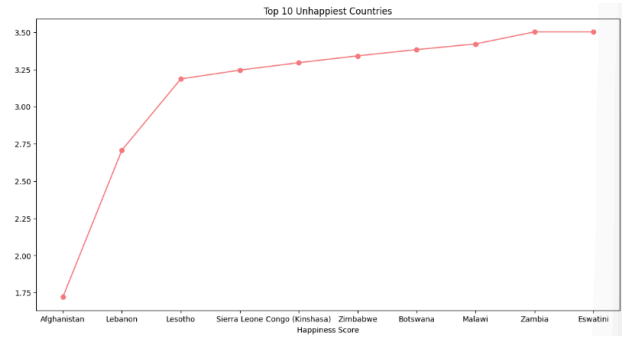
AS the preliminary tasks in the assignment, the tasks were easy and required basic method implementation. Most tasks were practiced in previous tutorials hence further research wasn't as required as future tasks.

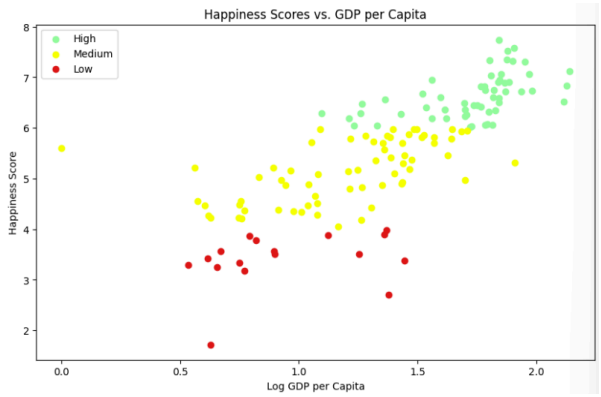
Challenges

There were no particular challenges faced in this section as it was a relatively simple task.

Visual Observation







Problem 2

Moving into the second problem of the task, we read data out of a csv file and filtered required data into a new csv file for South Asian Countries. We worked with the new data set to visualize and compute data as specified by the assignment.

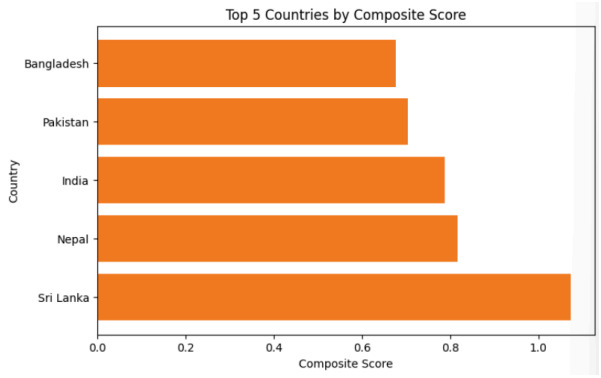
Task Flow

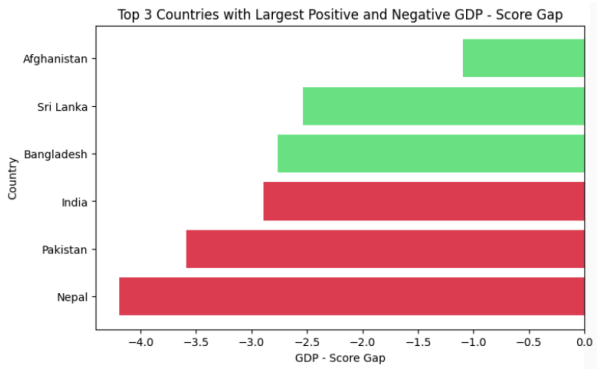
This task required a new csv file to be generated and used to play with the data. The tasks were pretty simple as general panda methods were enough to tackle the problems. Most of the formulae used in this section were taught in the workshops hence it was an easy implementation task.

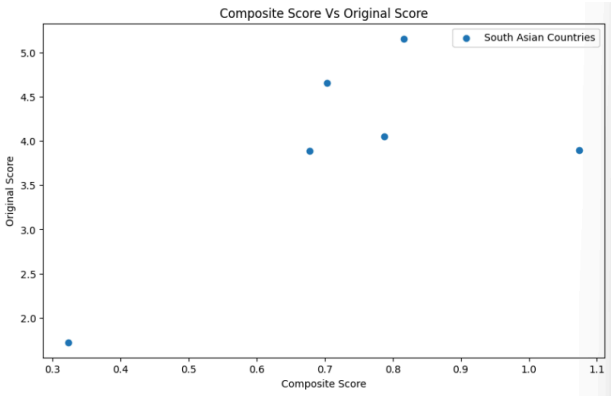
Challenges

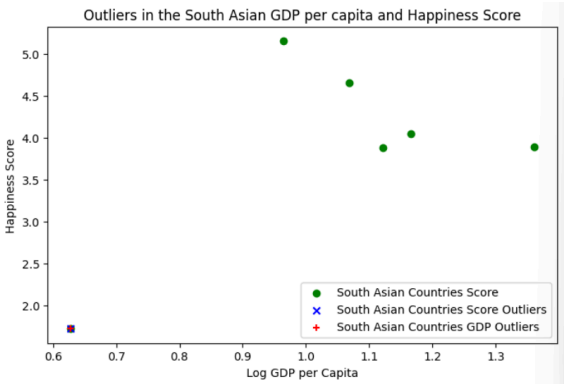
A particular challenge faced in this task was confusion caused by conflicting results to desired results. The task 5 of this portion requested data be highlighted in largest positive and negative gaps between two variables. However, the result of the code was entirely negative. This discrepancy between expected vs. actual results created a delay furthering the task flow.

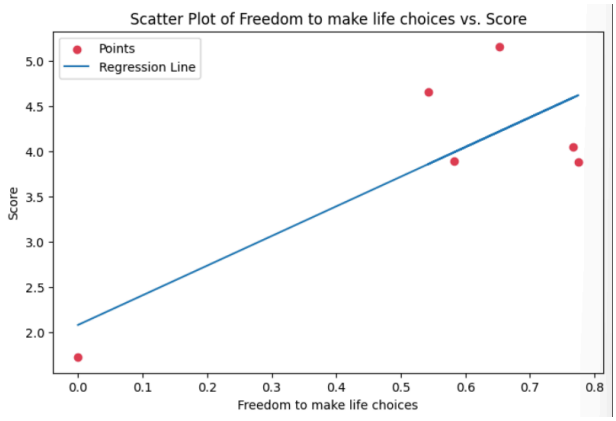
Visual Observation

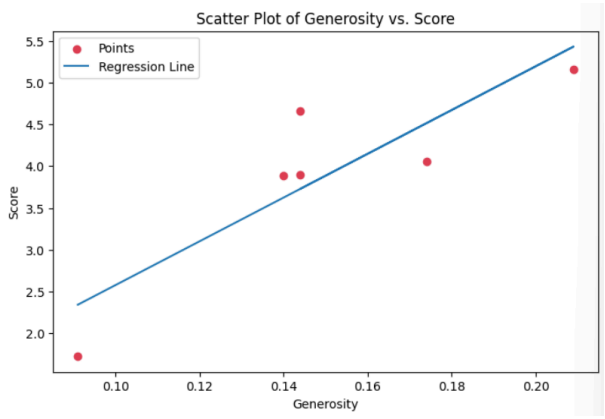





Composite Score Vs Original Score
South Asian Countries
Original Score
Composite Score


Outliers in the South Asian GDP per capita and Happiness Score
Happiness Score
Log GDP per Capita
South Asian Countries Score
South Asian Countries Score Outliers
South Asian Countries GDP Outliers


Scatter Plot of Freedom to make life choices vs. Score
Points
Regression Line
Score
Freedom to make life choices


Scatter Plot of Generosity vs. Score
Points
Regression Line
Score
Generosity

Problem 3

Similar to problem 2, we generated a csv file for Middle Eastern countries and computed the data to receive the desired result.

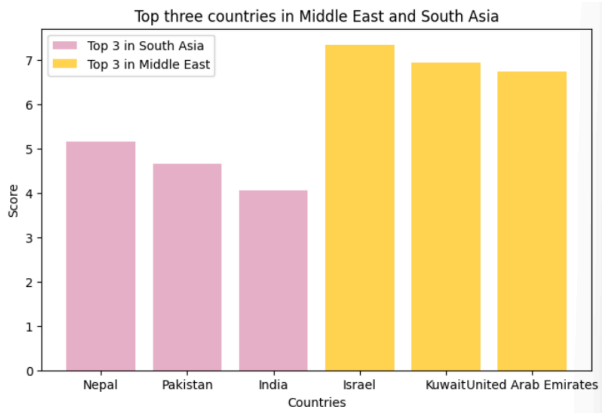
Task Flow

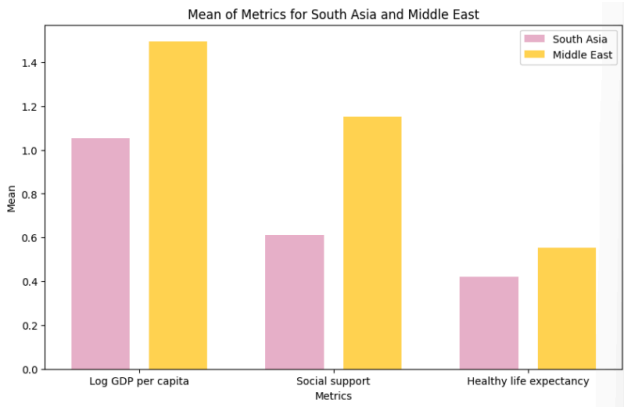
To start this problem, the primary requirements were to build a csv file for the middle eastern dataset. From the newly rendered CSV file, we computed the various questions requirements for the data and presented them through visual representation using bar graphs, scatter plots and box plots. An interesting part of this portion of the task was to generate the slope and intercept using the data to plot it in a graph. This was quite different from the rest of the questions and generated errors when first attempted. However, after further research and debussing, the desired output was achieved and the task was a learning experience.

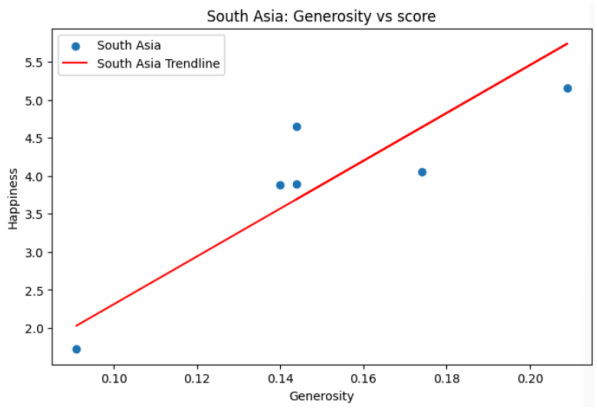
Challenges

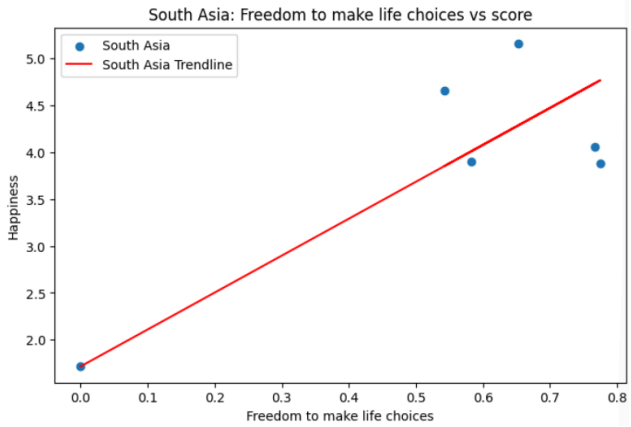
A particular challenge faced in this section was caused by an error in the list of countries of the Middle Eastern region, where we had stated “Palestine” instead of “State of Palestine” as marked in the dataframe. This had resulted in discrepancy in the results from the desired results, however it was easily solved by correcting the human error.

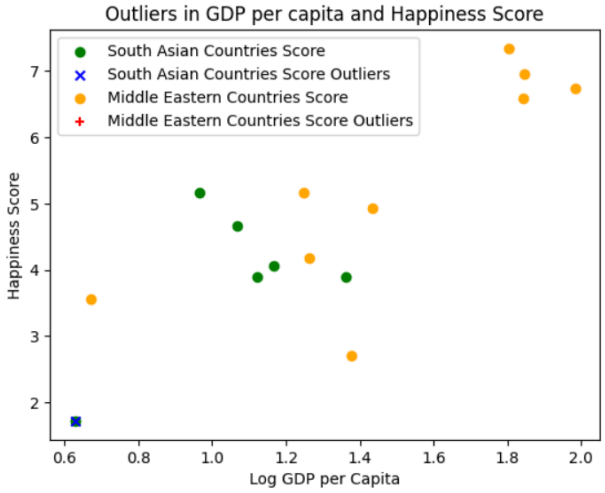
Visual Observation

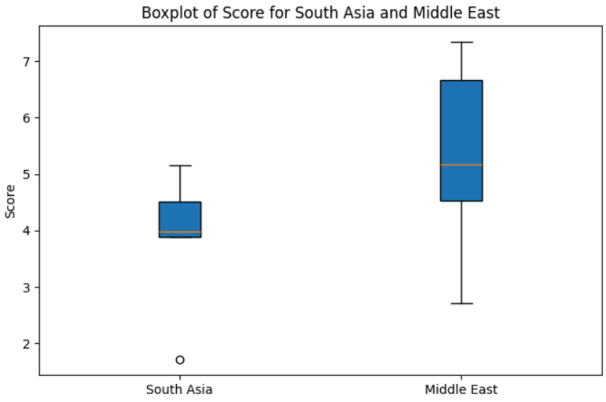












Despite the minor hiccup faced in this section, this was one of the most interesting portions of this assignment as it required us to venture out of lecture and workshop content and do our own research to compute certain results as well as learn new content to tackle the problems.

Conclusion

In this assignment, we practised data analysis and cleaning tactics to help interpret and analyse data in a better and efficient manner. The learnings of this assignment can be implemented in real life scenarios to get a better understanding of statistical data and draw conclusions from a well formatted and clean dataframe.

References:

https://worldhappiness.report/about/#:~:text=The%20World%20Happiness%20Report%20reflects,and%20national%20variations%20in%20happiness.

https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.boxplot.html