Assignment - I - Statistical Interpretation and Exploratory Data Analysis.

Student Name: Saksham Rajkarnikar
Student ID: 2462291
Course Name: 5CS037 Concepts and Technologies of AI
Date of Submission: 9th Jan 2026

Table of Content

Introduction

The Human Development Index (HDI) is a composite indicator developed by the UNDP that assesses a country's average performance in key aspects of human development, including a long and healthy life (life expectancy), access to education (educational attainment), and an adequate standard of living (GNI per capita).

The objective of this analysis is to perform statistical analysis and exploratory data analysis of HDI across different countries and regions for different time frames.

The analysis focuses on:

- Examining HDI values for a specific year (2022)
- Examining HDI trends from 2020 to 2022
- Performing a detailed regional study of South Asia
- Comparing South Asia with the Middle East region

The study's scope is limited to descriptive statistics, data cleaning, visualization, and comparative analysis based only on the dataset's variables, predictive modeling and outside variables are not included.

Problem-wise Analysis

Problem 1A: HDI Analysis for the Year 2022

Methods / Approach

The initial dataset was first filtered to include the data from the year 2022. Exploratory data analysis (EDA) methods were applied, such as examining the data structure, detecting missing values, and confirming the correctness of data types. Rows with missing values were dropped. Descriptive statistics were applied to summarize the overall distribution of the dataset. Finally, countries were filtered and ranked according to their HDI values and Gross National Income (GNI) per capita.

Key Results

- There is a noticeable difference between nations in the 2022 global HDI distribution.
- The highest HDI country falls under the “Very High Human Development” category, while the lowest HDI country belongs to the “Low Human Development” category.
- A number of nations attained HDI scores greater than 0.800, signifying exceptionally high levels of human development.
- High HDI does not always correspond directly to the highest GNI per capita, suggesting the importance of non-income factors.

Visualizations and Tables

Summary statistics of HDI for the year 2022

Statistics	HDI Value
Mean	0.7229
Median	0.7395

Standard Deviation	0.1530

Top 10 Countries with Highest HDI sorted by GNI

country	year	hdi	life_expectancy	pop_millions	hdi_f
Liechtenstein	2022	0.942	84.656	0.039327	0.924163
Qatar	2022	0.875	81.559	2.695122	0.892561
Singapore	2022	0.949	84.133	5.975689	0.945325
Ireland	2022	0.950	82.716	5.023109	0.942838
Luxembourg	2022	0.927	82.591	0.647599	0.920482
United Arab Emirates	2022	0.937	79.196	9.441128	0.922940
Switzerland	2022	0.967	84.255	8.740472	0.949369
Norway	2022	0.966	83.393	5.434319	0.957148
United States	2022	0.927	78.203	338.289857	0.927693
Hong Kong, China (SAR)	2022	0.956	84.315	7.488865	0.942671

HDI classification Table

	country	hdi	HDI Category
32	Afghanistan	0.462	Low
65	Albania	0.789	High
98	Algeria	0.745	High
131	Andorra	0.884	Very High
164	Angola	0.591	Medium

Interpretation and Discussion

The COVID-19 pandemic's worldwide effects on health systems, education, and economic activity are primarily to blame for the 2020 HDI decline. Regional differences are still noticeable, emphasizing structural differences in development and uneven recovery rates. Despite the fact that income and HDI have a strong correlation, diminishing returns at higher income levels imply that more extensive social investments are necessary for long-term human development.

Problem 1B: HDI Visualization and Trend Analysis (2020–2022)

Methods / Approach

The dataset was filtered and a new dataset was created to include data from the years 2020, 2021, and 2022. Cleaning procedures were implemented after variables like country, year, and HDI were examined for missing or inconsistent values. Multiple visualizations were created to analyze trends over time, regional differences, and relationships between HDI and income.

Key Results

- Many countries experienced stagnation or decline in HDI during 2020.
- For a number of nations, a slow recovery in HDI values is seen in 2021 and 2022
- Europe and North America consistently show the highest average HDI, while South Asia and Sub-Saharan Africa remain lower.
- The HDI and GNI per capita are positively correlated.

Visualizations and Tables

Line Chart of selected countries HDI trends

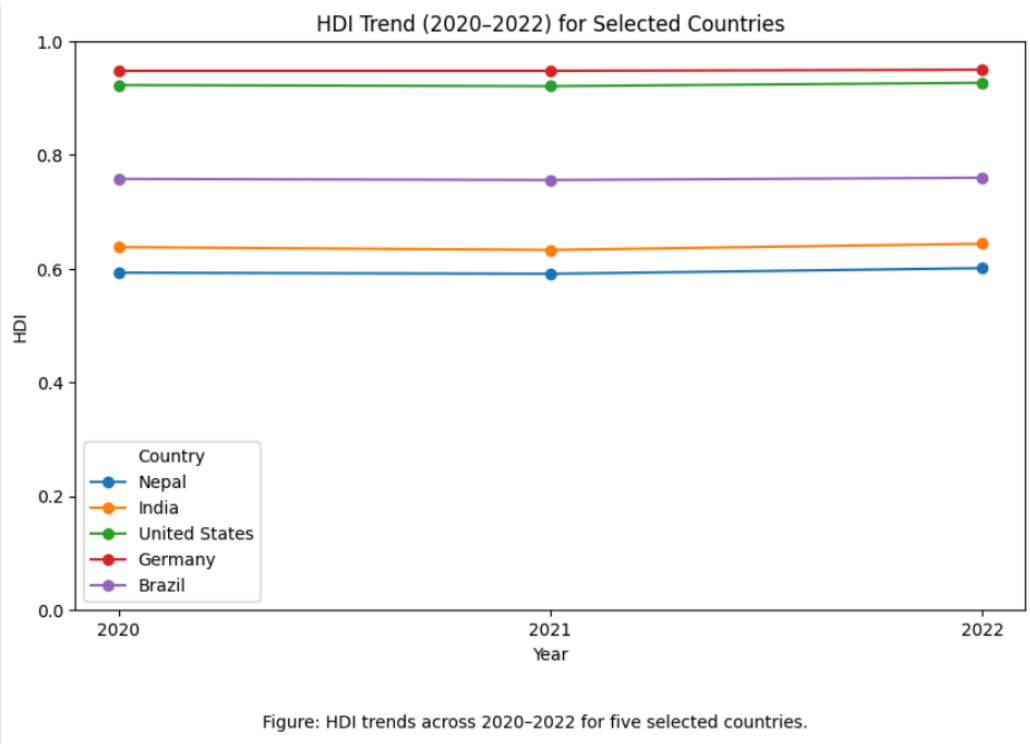
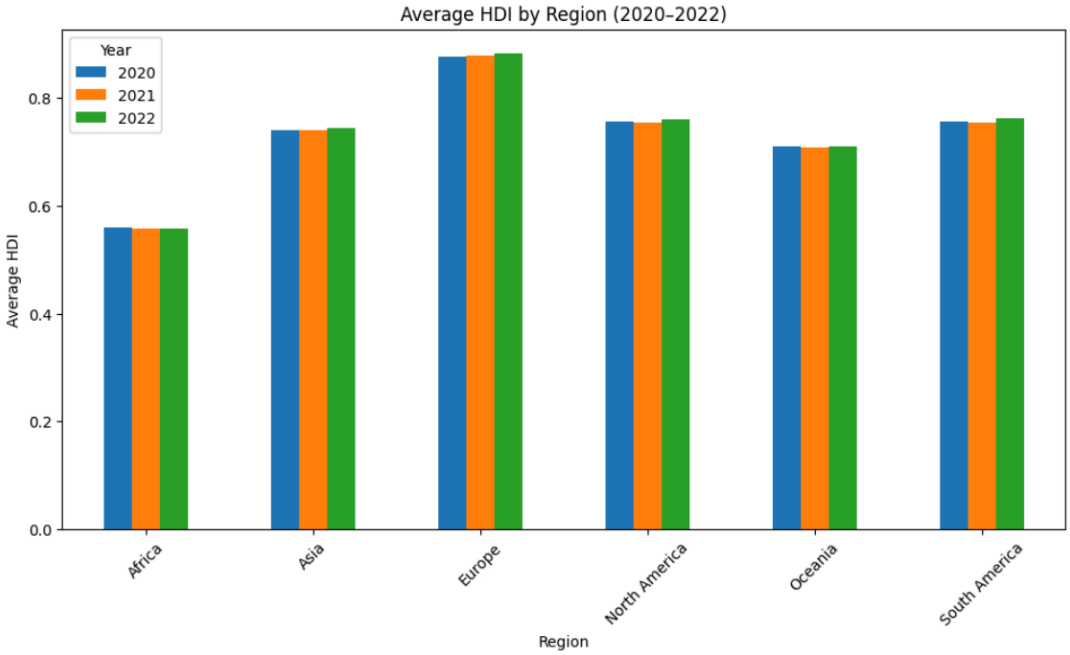
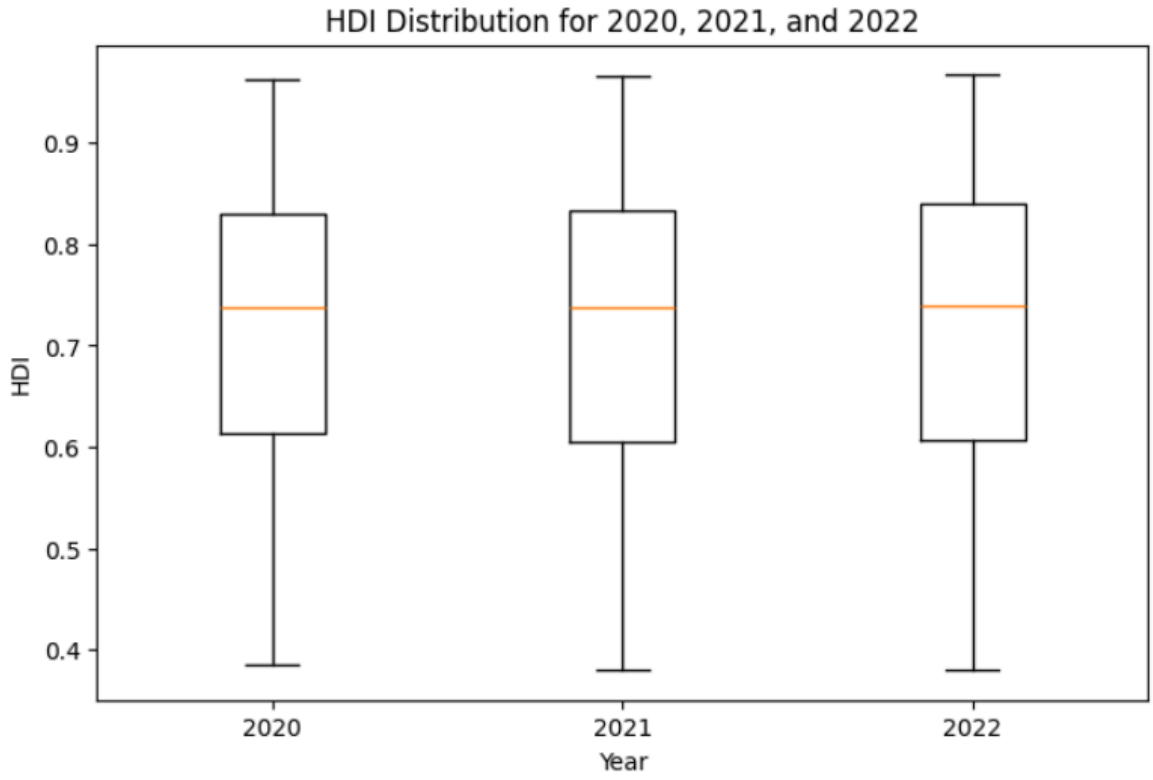


Figure: HDI trends across 2020–2022 for five selected countries.

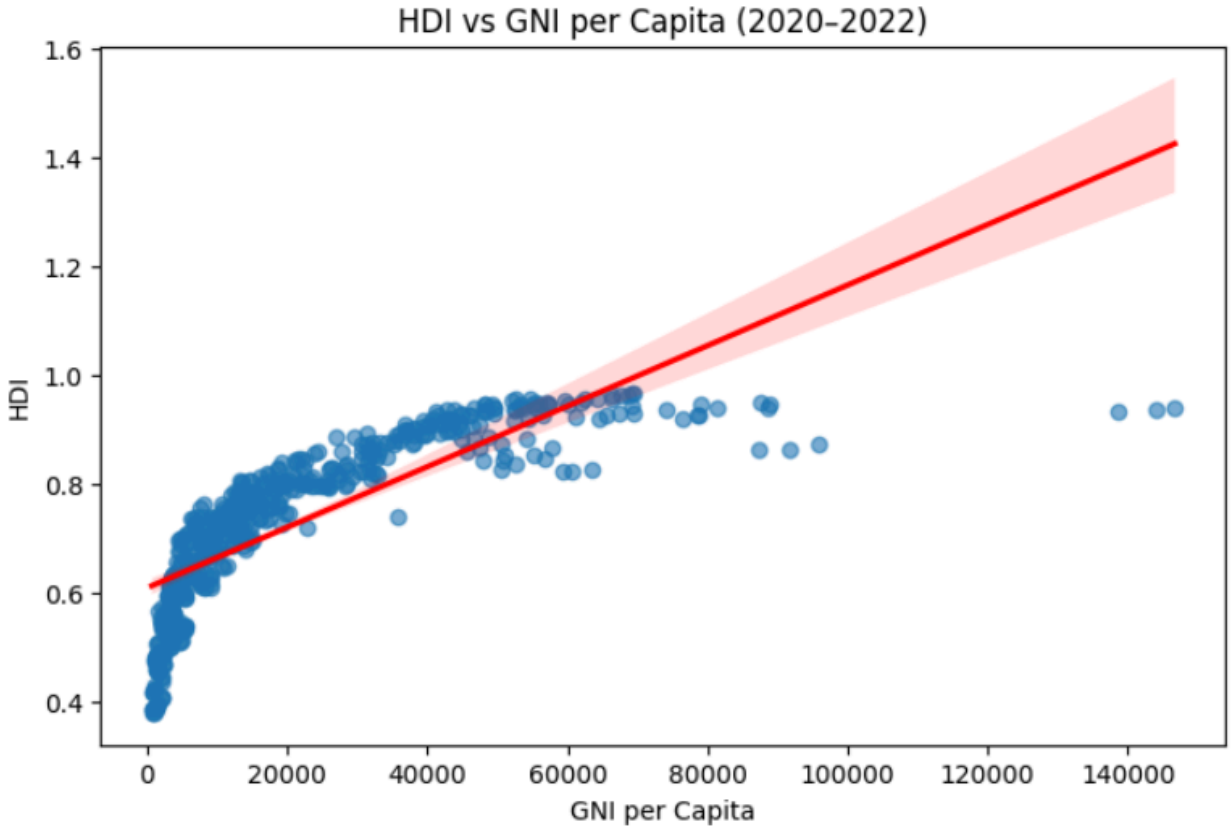
Bar-Chart of Average HDI by Region from 2020-2022



Box plot of HDI distribution for 2020,2021 and 2022



Scatter plot of HDI vs GNI per capita



Interpretation and Discussion

In summary, the decrease in HDI in 2020 can be attributed largely to the impact of the COVID-19 pandemic worldwide, which affected health systems, education, and economic activities. There are still set regional disparities, showing unequal recovery velocities and structural differences in development. While income is strongly correlated with HDI, diminishing returns at higher levels of income suggest that broader social investments are necessary for continued human development.

Problem 2: Advanced HDI Exploration – South Asia

Methods / Approach

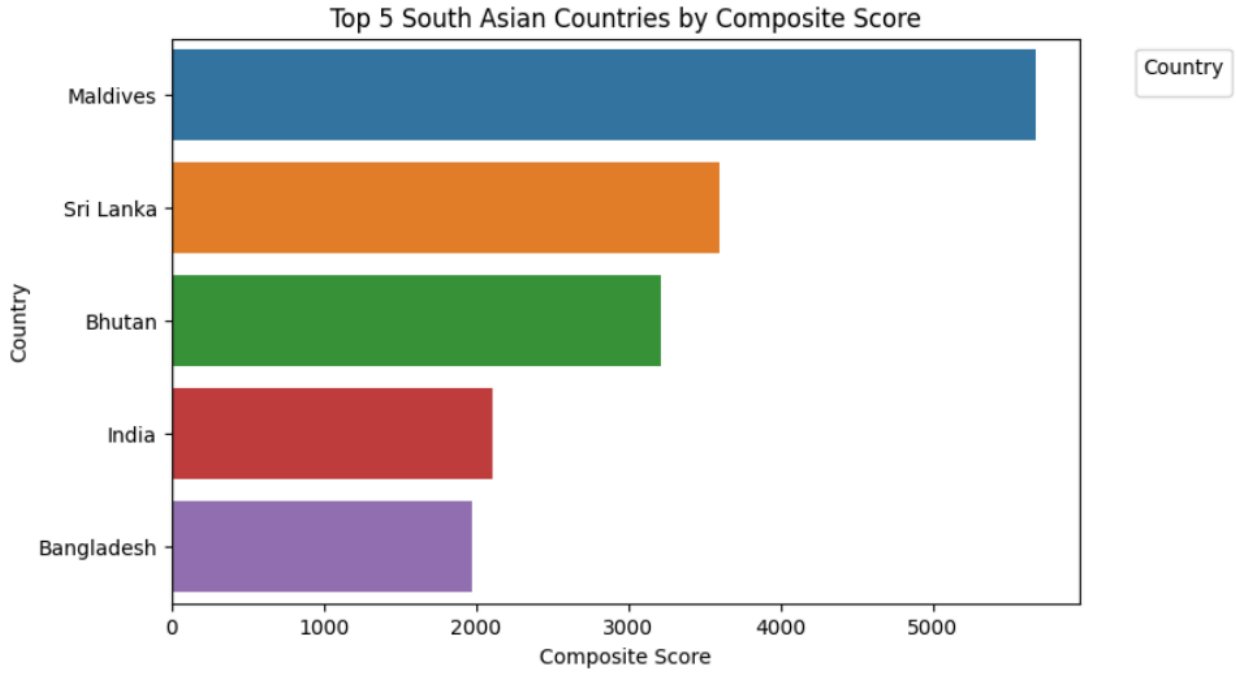
A regional dataset consisting of South Asian countries was created by filtering the original main dataset. A Composite Development Score was calculated using life expectancy and GNI per capita indicators. Rankings based on this composite score were compared with HDI rankings. Outliers were identified using the 1.5×IQR method, and correlation analysis was performed to get relationships between HDI and selected development indicators.

Key Results

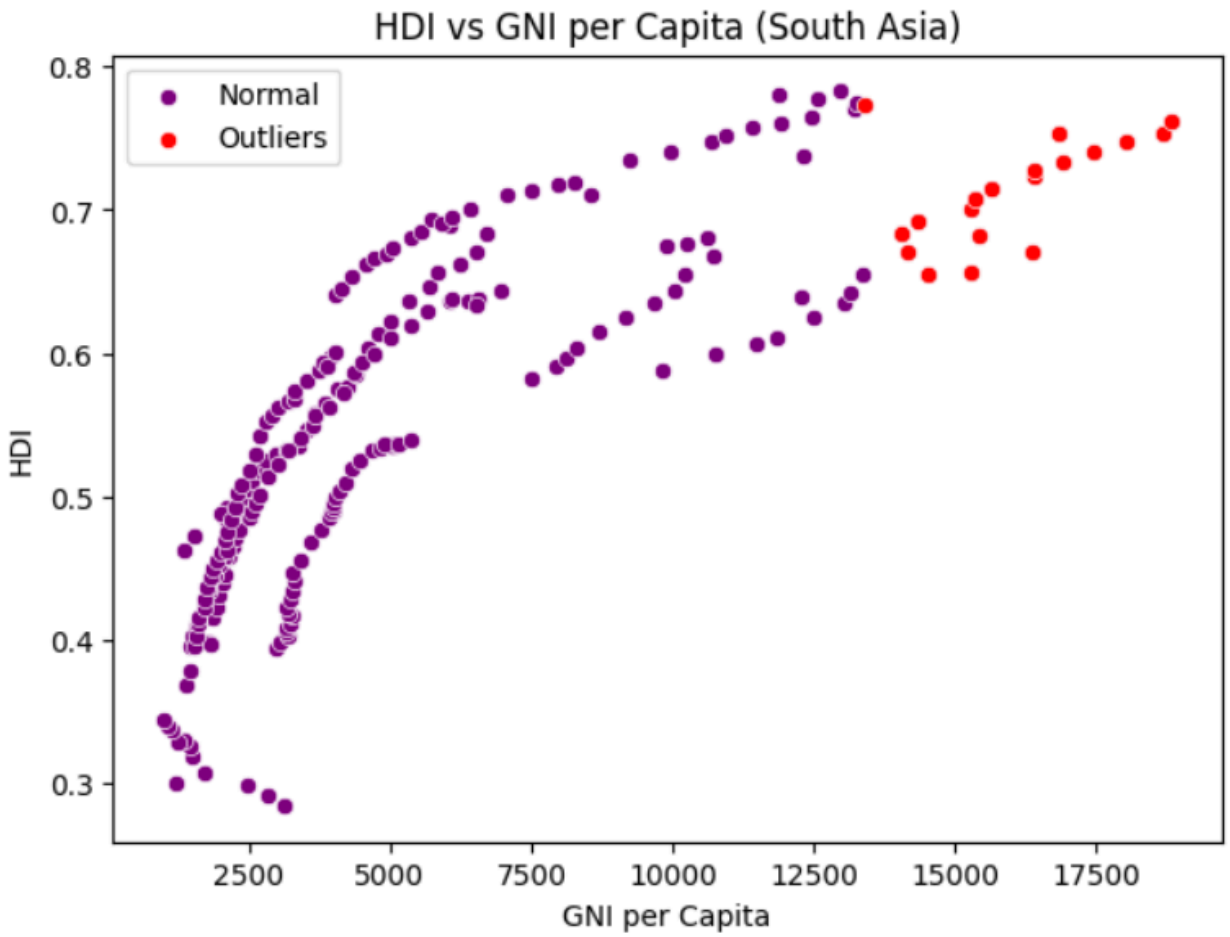
- Sri Lanka and Maldives rank higher in composite scores due to stronger social indicators.
- The comparison between the rankings of HDI and the composite score brings out the influence of indicator weighting on rankings.
- Strong positive correlations were observed between HDI and life expectancy.
- In a number of countries, a divergence can be found between income levels and outcomes for HDI.

Visualizations and Tables

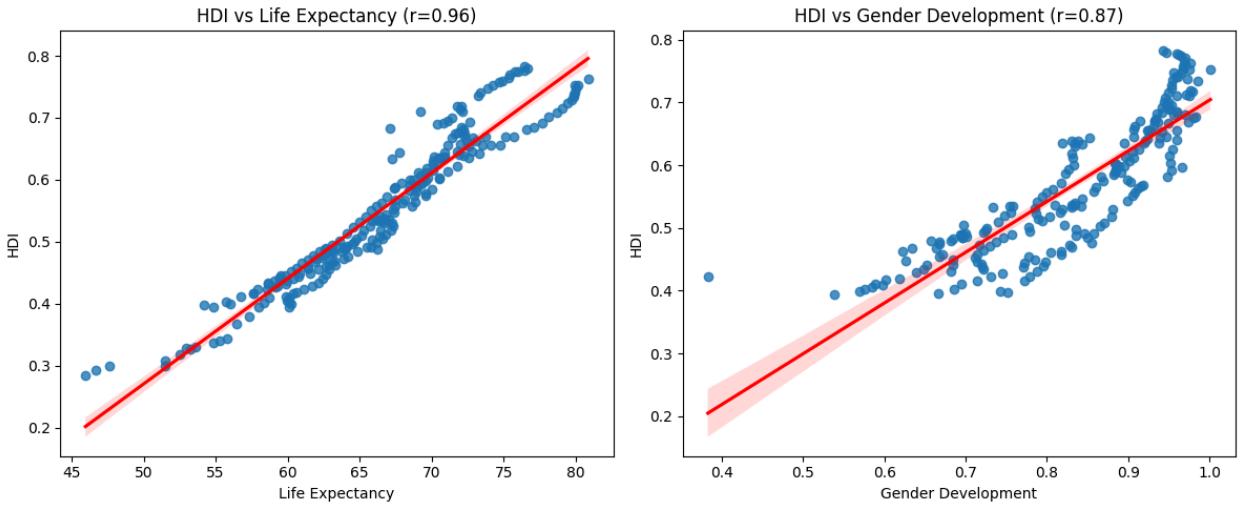
Bar Chart of top 5 South Asian COuntries by Composite Score



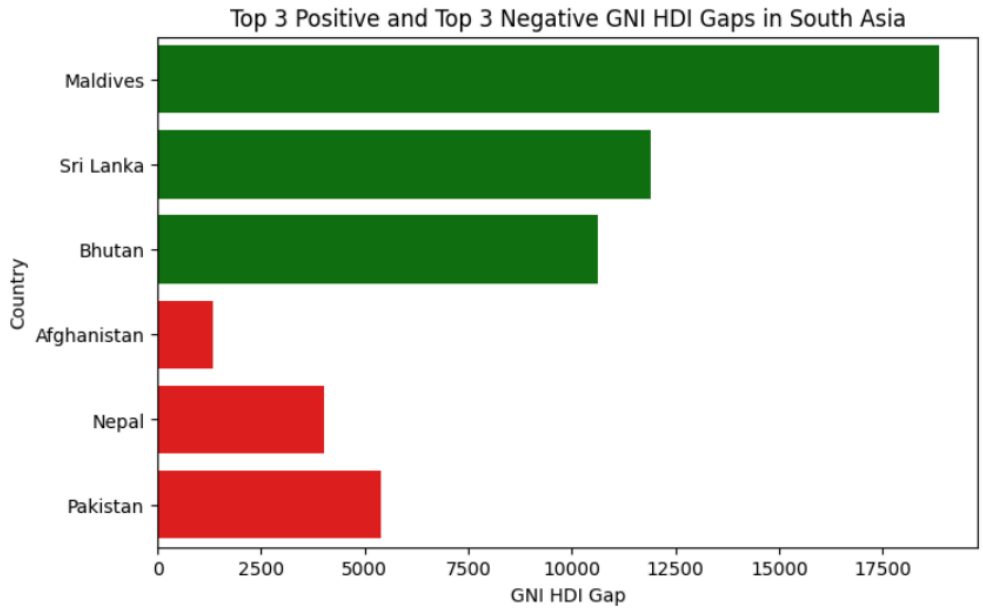
Scatter plot of HDI vs GNI outliers



Scatter plot for metric relationship



Bar Chart of Top 3 positive and top 3 negative GNI Gaps in South Asia



Interpretation and Discussion

According to analysis, the relationship between higher incomes and higher levels of human development is not automatically achieved in the region of South Asia. Nations with effective health and educational structures usually perform better than other nations with lower incomes. Outliers provide evidence of effective policy and investment for better outcomes in development.

Problem 3: Comparative Regional Analysis – South Asia vs Middle East

Methods / Approach

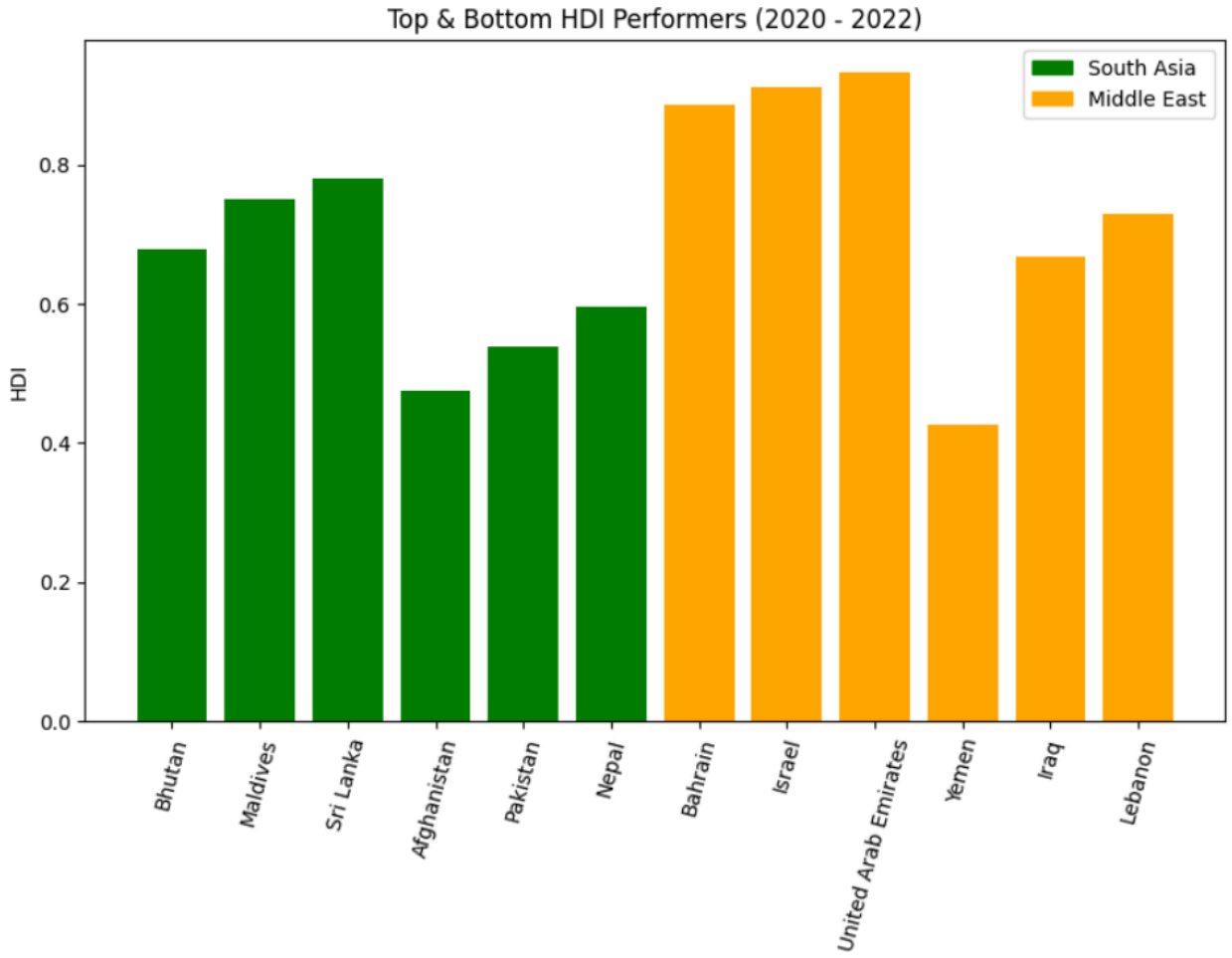
Subsets for South Asia and the Middle East were created using the cleaned 2020–2022 dataset. Descriptive statistics, including mean, standard deviation, range, and coefficient of variation, were computed for HDI and related indicators. Comparative visualizations were used to assess performance, disparities, correlations, and outliers across regions.

Key Results

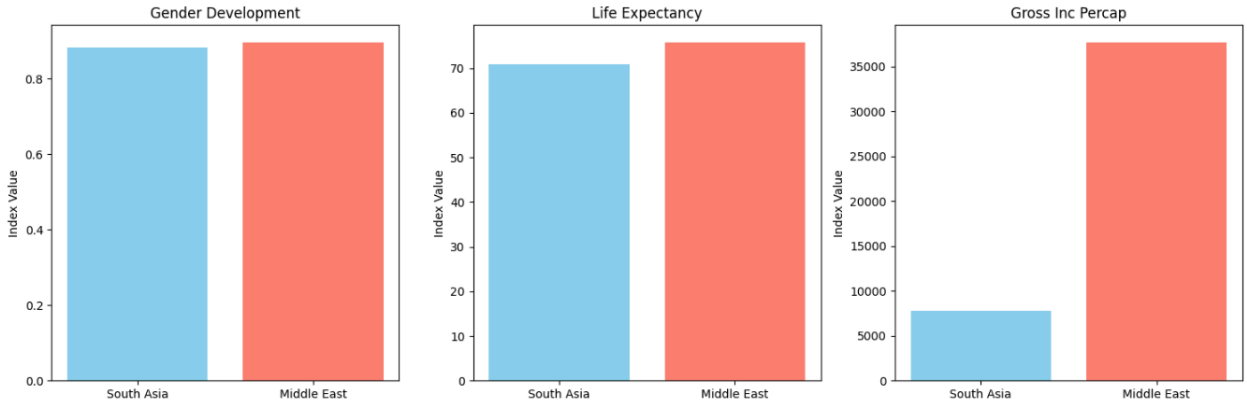
- The Middle East region demonstrates a higher average HDI compared to South Asia.
- South Asia exhibits greater variability and inequality in HDI values.
- Life expectancy and income indicators show the largest disparities between the two regions.
- There are positive correlations between the HDI and the development indicators in both regions.

Visualizations and Tables

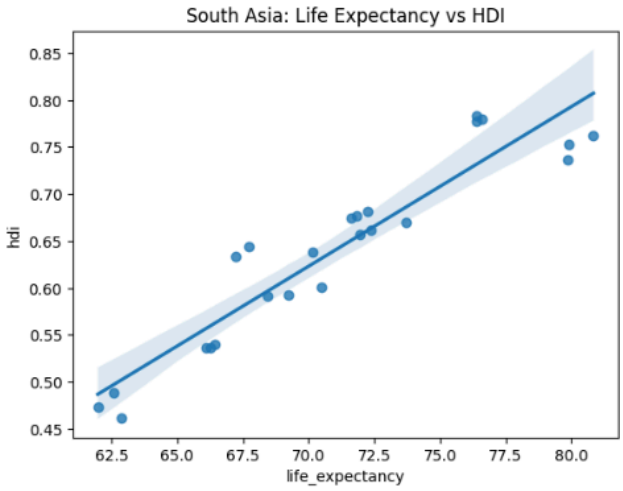
Bar chart for the Top 3 and Bottom 3 HDI performers of South Asia and Middle East

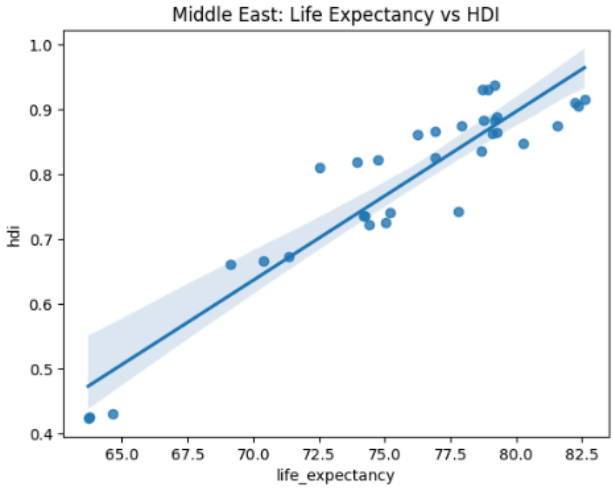


Barchart for metric comparisons (Gender development, Life Expectancy and Gross Inc Percap)

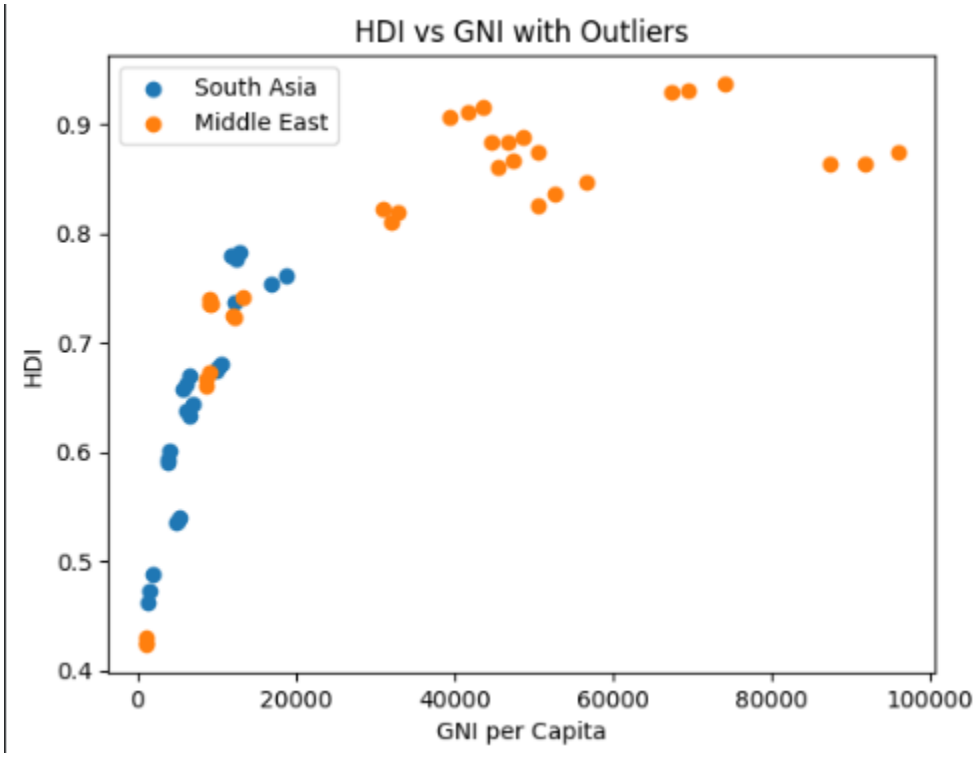


Scatter plot for life expectancy vs HDI for South Asia and Middle East


South Asia: Life Expectancy vs HDI
hdi
life_expectancy
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0


Middle East: Life Expectancy vs HDI
hdi
life_expectancy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
65.0
67.5
70.0
72.5
75.0
77.5
80.0
82.5

Scatter plot for outlier detection for South Asia and Middle East



Interpretation and Discussion

The comparative study reveals structural distinctions between the Middle East and South Asia's developmental trajectories. Higher income levels benefit the Middle East, but South Asia's greater variability points to unequal access to healthcare and education. The existence of outliers in both areas highlights how national policies and institutional efficacy affect the results of human development.

Conclusion

Regional and global differences in human development are highlighted by this exploratory data analysis. Although income is a significant factor in determining HDI, it is insufficient on its own. Development is significantly influenced by health and education outcomes.

The analysis highlights the need for inclusive, balanced development policies, especially in areas like South Asia, and shows how global events like the COVID-19 pandemic affect HDI trends.

Overall, methodical data preparation, statistical analysis, and interpretation based on visualization helped to achieve the assignment's goals.

Github Link

Github Repository Link:
https://github.com/Kaede-7/5CS037-Concepts-and-Technologies-of-AI-Assignment1