UNIVERSITY OF WOLVERHAMPTON

BIRATNAGAR INTERNATIONAL COLLEGE

ing

Analysis of the Human Development Index (HDI)

Statistical Interpretation and Exploratory Data Analysis

Course: Concepts and Technologies of AI (5CS037)

Student Name: Kuldeep Mandal

Student ID: 2505925

Tutor: Ayush Regmi

Table of Contents

1. Introduction

The (Human Development Report) is a composite indicator developed by the United Nations Development Programme (UNDP) to measure a country's overall development. Unlike purely economic indicators, HDI combines health (life expectancy)**,** education**,** and standard of living (GNI per capita) to provide a more holistic picture of human well-being.

This report presents a statistical interpretation and exploratory data analysis (EDA) of the Human Development Index Dataset (1990–2022). The analysis focuses on identifying recent development patterns, temporal trends, regional disparities, correlations between variables, and outlier countries. Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn were used for data cleaning, analysis, and visualization.

2. Problem 1A – Single Year HDI Exploration (2022)

The dataset was filtered to include only observations from the year 2022. Data cleaning steps included handling missing values, converting numeric columns stored as text into numeric format, removing duplicate records, and standardizing country names. Countries were also classified into official UNDP HDI categories (Low, Medium, High, Very High).

2.1 Data Preparation and Cleaning

The dataset was first examined to identify all available years, after which **2022** was selected as the latest year for focused analysis. Rows with missing HDI values were removed because HDI is the key variable of interest. Numeric columns such as HDI, GNI per capita, and life expectancy were converted from text to numeric format to ensure accurate calculations.

This cleaning step ensured that the dataset used for analysis was consistent, reliable, and suitable for statistical interpretation.

2.2 Descriptive Statistics

Basic descriptive statistics (mean, median, and standard deviation) were calculated for HDI in 2022. These measures provide an overview of the central tendency and spread of global development levels.

- The **mean HDI** reflects the average development level across countries.
- The **median HDI** helps identify skewness in the distribution.
- The **standard deviation** indicates the degree of inequality in HDI values.

```
HDI 2022 - Mean:   0.723
HDI 2022 - Median: 0.740
HDI 2022 - Std:    0.153

Country with highest HDI in 2022: Switzerland (0.967)
Country with lowest HDI in 2022: Somalia (0.380)
```

Table 1: Summary statistics (Mean, Median, Standard Deviation) of HDI in 2022

2.3 Highest and Lowest HDI Countries

The country with the highest HDI in 2022 represents very high human development, typically associated with strong healthcare systems, advanced education, and high income.
In contrast, the country with the lowest HDI highlights persistent challenges such as poverty, limited access to healthcare, and weak educational infrastructure. This contrast clearly demonstrates global inequality in human development.

2.4 Filtering and HDI Categorization

Countries with HDI greater than 0.800 were filtered and sorted by GNI per capita. While many high-HDI countries also have high income, this is not universally true, showing that income alone does not fully determine human development.

Countries were then classified into Low, Medium, High, and Very High HDI categories using UNDP thresholds. The updated dataset was saved as “HDI category added.csv”, fulfilling the assignment requirement.

2.5 Interpretation

The 2022 analysis shows that although many countries have achieved high or very high HDI, a substantial number remain in lower categories. This suggests that global development progress is uneven and highlights the need for targeted social and economic policies.

3. Problem 1B – HDI Trend Analysis (2020–2022)

Data for the years 2020, 2021, and 2022 were extracted to analyze short-term HDI trends. Average HDI values were calculated by region and year. Line charts, bar charts, box plots, and scatter plots were used to visualize trends and relationships.

3.1 Data Extraction and Cleaning

Data for **2020, 2021, and 2022** were extracted to study short-term HDI trends. Cleaning steps similar to Problem 1A were applied to ensure consistency across years.

3.2 HDI Trends for Selected Countries

A line chart was created for five South Asian countries (Nepal, India, Bangladesh, Pakistan, and Sri Lanka). Most countries show gradual improvement, although the rate of change differs.

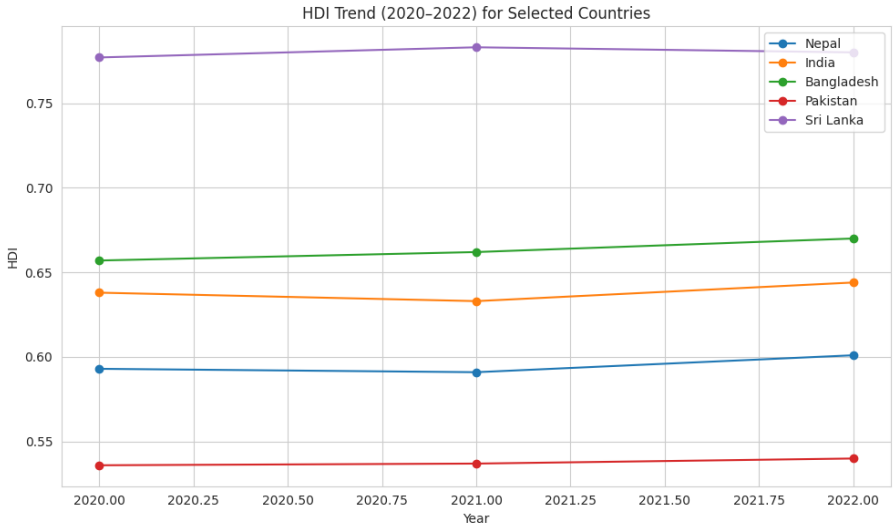

Figure 1: HDI trends (2020–2022) for selected South Asian countries

This line graph illustrates changes in HDI values over time for selected South Asian countries. Most countries show gradual improvement, although the rate of growth differs, reflecting varying development progress.

3.3 Regional Comparisons over Time

Bar charts and box plots were used to compare HDI across regions and years.

- **Bar chart:** Shows average HDI by region for each year.
- **Box plot:** Highlights changes in HDI distribution by year over time.

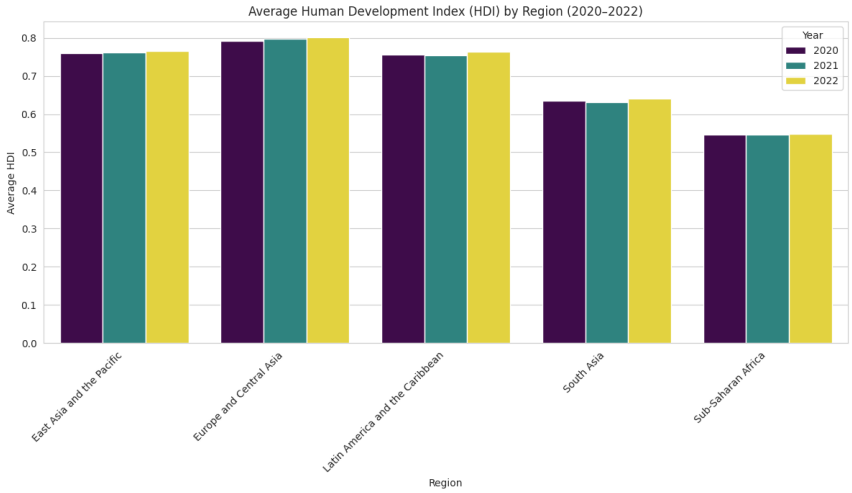

Figure 2: Average HDI by region (2020–2022)

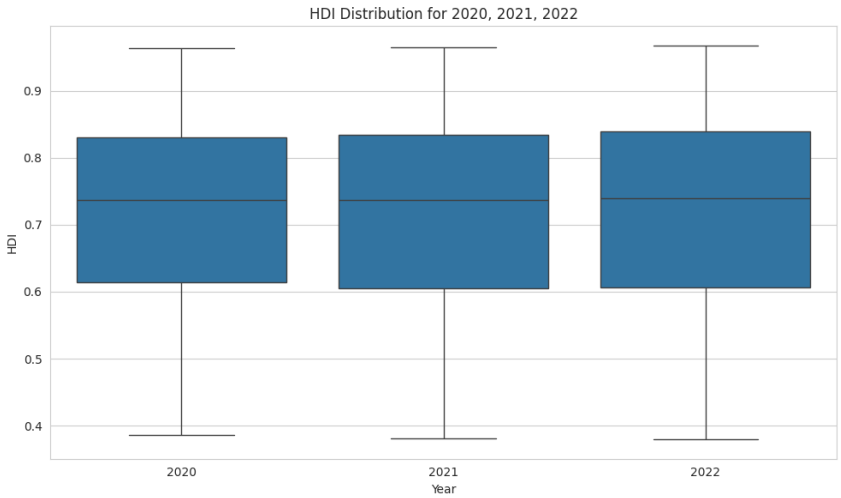

Figure 3: HDI distribution by year using box plots

The box plot displays the median, spread, and outliers of HDI values for each year. A slight upward shift in the median suggests overall improvement, while the spread indicates persistent disparities.

3.4 Relationship between HDI and Income

A scatter plot of **HDI vs GNI per capita** shows a clear positive relationship, although several countries deviate from the general trend.

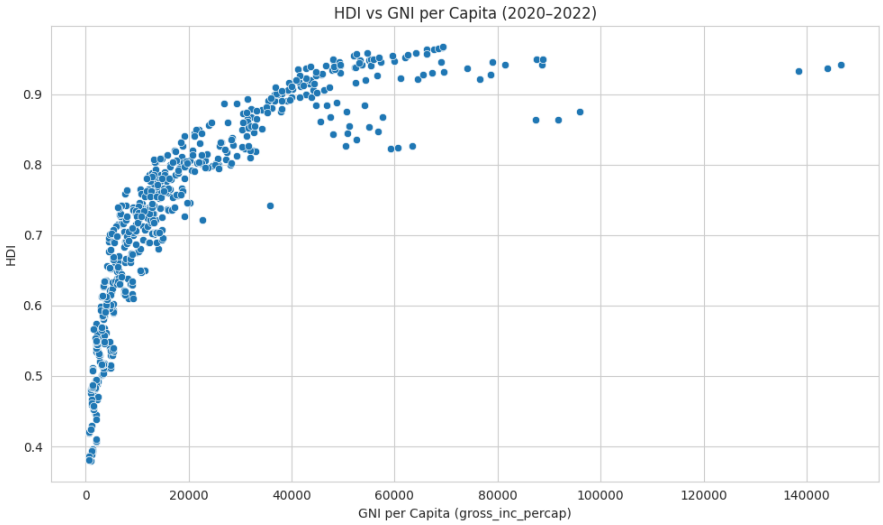

Figure 4: Relationship between HDI and GNI per capita (2020–2022)

This confirms that while income is important, social factors also play a significant role.

3.5 Interpretation

The analysis suggests that global HDI trends were influenced by external shocks such as the COVID-19 pandemic. Regions with stronger healthcare and social protection systems demonstrated faster recovery, while others experienced stagnation.

4. Problem 2 – Advanced HDI Exploration (South Asia)

A subset of South Asian countries was created for detailed analysis. Composite scores, correlation analysis, gap analysis, and outlier detection were performed to better understand development dynamics within the region.

4.1 Composite Development Score

A composite score combining life expectancy and GNI per capita was calculated for South Asian countries. Rankings based on this score were compared with HDI rankings.

South Asian countries ranked by Composite Score (2022):

	country	hdi	Composite Score
3530	Maldives	0.762	5678.289357
5477	Sri Lanka	0.780	3592.832541
659	Bhutan	0.681	3209.130864
2573	India	0.644	2105.481239
461	Bangladesh	0.670	1975.446053
4388	Pakistan	0.540	1632.210427
4091	Nepal	0.601	1228.811605
32	Afghanistan	0.462	419.425420

Table 2: South Asian countries ranked by Composite Score (2022)

Differences between rankings indicate that some countries perform better economically than socially, or vice versa.

4.2 Outlier Detection

Using the **1.5 × IQR rule**, outliers were identified for HDI and income. These outliers represent countries whose development outcomes differ significantly from regional norms.

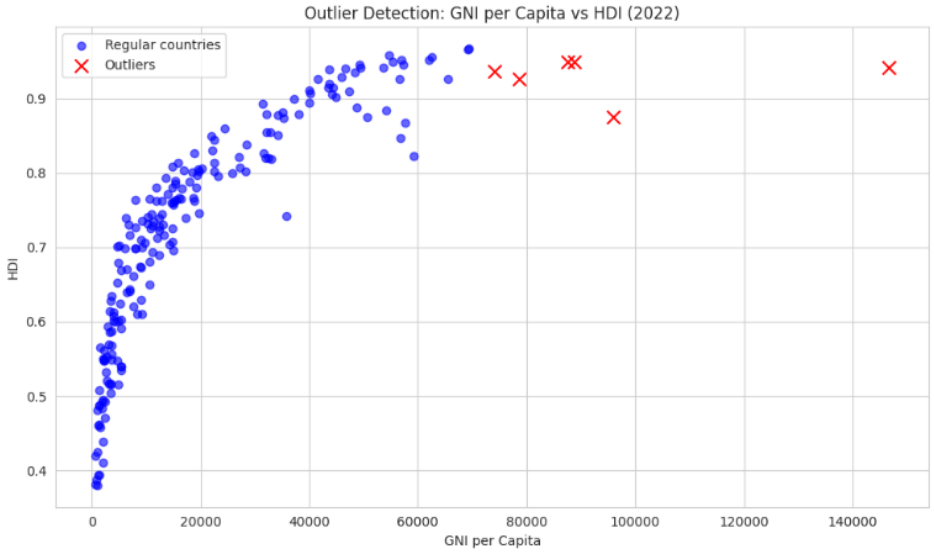

Figure 5: Outliers in South Asia (HDI vs GNI per capita)

4.3 Correlation Analysis

Correlation analysis revealed that life expectancy has a stronger relationship with HDI than gender development in South Asia. This suggests health outcomes are a critical driver of development in the region.

4.4 GNI–HDI Gap Analysis

The difference between GNI per capita and HDI highlights countries where economic growth has not translated effectively into human development.

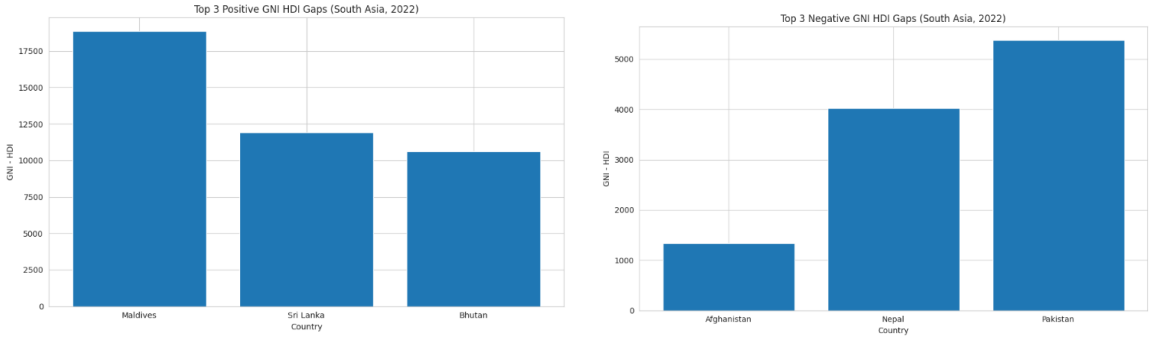

Figure 6: Top 3 positive and negative GNI–HDI gaps in South Asia

This visualization identifies countries where economic performance does not translate proportionally into human development, revealing efficiency gaps in development outcomes.

4.5 Interpretation

South Asia displays significant internal disparities. While some countries perform well despite limited income, others struggle to convert economic growth into broader human development, highlighting the importance of effective social investment.

5. Problem 3 – Comparative Regional Analysis (South Asia vs Middle East)

A subset of South Asian countries was created for detailed analysis. Composite scores, correlation analysis, gap analysis, and outlier detection were performed to better understand development dynamics within the region.

5.1 Descriptive Comparison

The Middle East shows a higher average HDI, while South Asia displays greater variability.

```
Mean and Std of HDI for South Asia vs Middle East (2020–2022):
                mean       std
Region
Middle East  0.788909  0.141295
South Asia   0.639583  0.098273
```

Table 3: Mean and standard deviation of HDI by region (2020–2022)

5.2 Top and Bottom Performers

Top-performing and bottom-performing countries were identified in each region.

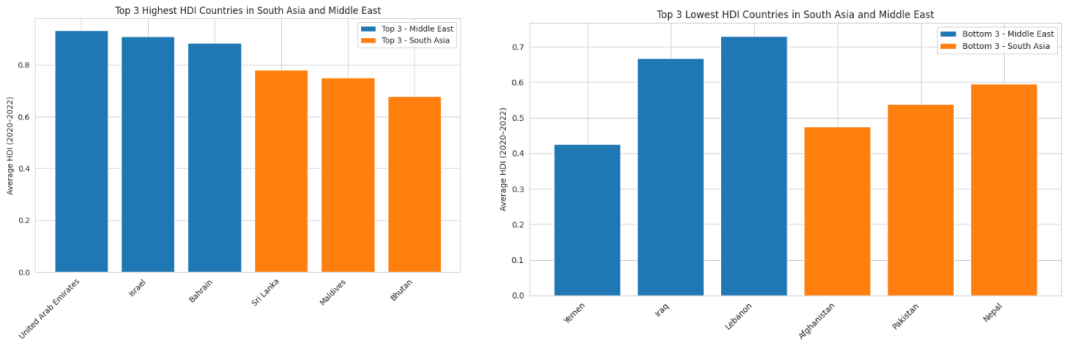

Figure 7: Top and bottom HDI countries in South Asia and Middle East

5.3 Interpretation

The findings suggest that economic resources contribute significantly to regional differences, but health and education remain critical determinants of HDI. Higher income does not automatically guarantee higher human development.

6. Conclusion

This analysis demonstrates that HDI is influenced by multiple factors, not income alone. Health and education play a vital role in determining development outcomes. South Asia continues to make progress but faces persistent inequality, while the Middle East generally performs better but still contains internal disparities.

The findings reinforce the importance of balanced development policies that invest in social infrastructure alongside economic growth.

References

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L., & Wang, S. (2024). World Happiness Report 2024. *World Happiness Report*.

Human Development Report. (n.d.). *Human Development Report*.

L, Y., & Human Development Index Dataset (1990–2022). (n.d.). Human Development Index Dataset [1990-2022]. *Human Development Index Dataset [1990-2022]*.

UNDP. (n.d.). *United Nations Development Programme (UNDP). (n.d.). Human Development Index (HDI) Methodology*.

GitHub Repository Link:

ClickHere

Or

https://github.com/kuldeepmandal/Concepts-TechnologiesAI/tree/main/FinalAssesment