Concepts and Technologies of AI

5CS037

Statistical Interpretation and Exploratory Data Analysis of the Human Development Index (HDI)

Name: Krishna Kumar Mandal

University ID: 2508742

College ID: np02cs4a240074

Group: L5CG2

Table of Contents

1. Introduction

The Human Development Index (HDI) is score that compares the performance of different countries in global scale. Instead of just focusing at money, it focuses on how people are actually living.

It focuses on three simple things:

- **Health:** The average lifespan of people.
- **Education:** Literacy of people.
- **Wealth:** The average income of people.

Rather than looking just at money, HDI focuses on the whole picture of quality-of-life people are living.

The main objective of this paper is to study the numbers and look for patterns in the Human Development Index information from the last few years. This report looks at how HDI scores are distributed, how they change over the years, the differences between regions, and how the different parts of the score affect each other. This report is prepared by different Problems such as (e.g. 1A, 1B, 2, and 3) following guidelines of the assignment.

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

2.1 Method and Data Preparation

We first looked through the data to find every year listed, then we pulled out the recent year (2022) to focus on. The filtered data was saved as hdi_2022_df. The saved data was used to complete all the task of Problem 1A.

2.2 Data Cleaning and Justification

Different data cleaning steps were followed:

- The data was checked thoroughly to see if any columns had missing information. Countries with missing HDI scores were deleted since HDI is the main thing being studied and we can't accurately guess the missing numbers.
- Numbers like average income, were changed into a proper number format that were saved as text.
- duplicate rows were removed.
- special characters (e.g., “–”) representing missing data was replaced with NaN.

Following these steps ensured that the results were correct and the data was organized.

2.3 Basic Statistical Analysis

The mean, median, and SD for 2022 were calculated to see how the scores are distributed. The top HDI score in 2022 shows where development is best, while the lowest score points to areas that are struggling.

2.4 Filtering, Sorting, and HDI Categories

Countries which had an HDI above 0.800 were selected and sorted from highest to lowest by their average income (GNI). This shows the countries with high development and how much they earn.

A new column called HDI Category was added based on the UNDP's levels:

Low (< 0.550)

Medium (0.550–0.699)

High (0.700–0.799)

Very High (≥ 0.800)

We made sure every country was in the right group, then saved the file as HDI_category_added.csv.

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

3.1 Data Cleaning and Justification

After filtering the dataset, it was saved as HDI_problem1B.csv which only included year 2020,2021 and 2023.

The cleaning steps involves:

- Deleting rows that had no country, year, or HDI values.
- Changing number columns that were saved as text.
- Making sure every country name is written the same way.
- Deleting any duplicate entries.

These steps were followed to make sure the yearly comparisons were accurate.

3.2 Visualizations and Interpretations

Figure 1: HDI Trends for Selected Countries (2020–2022)

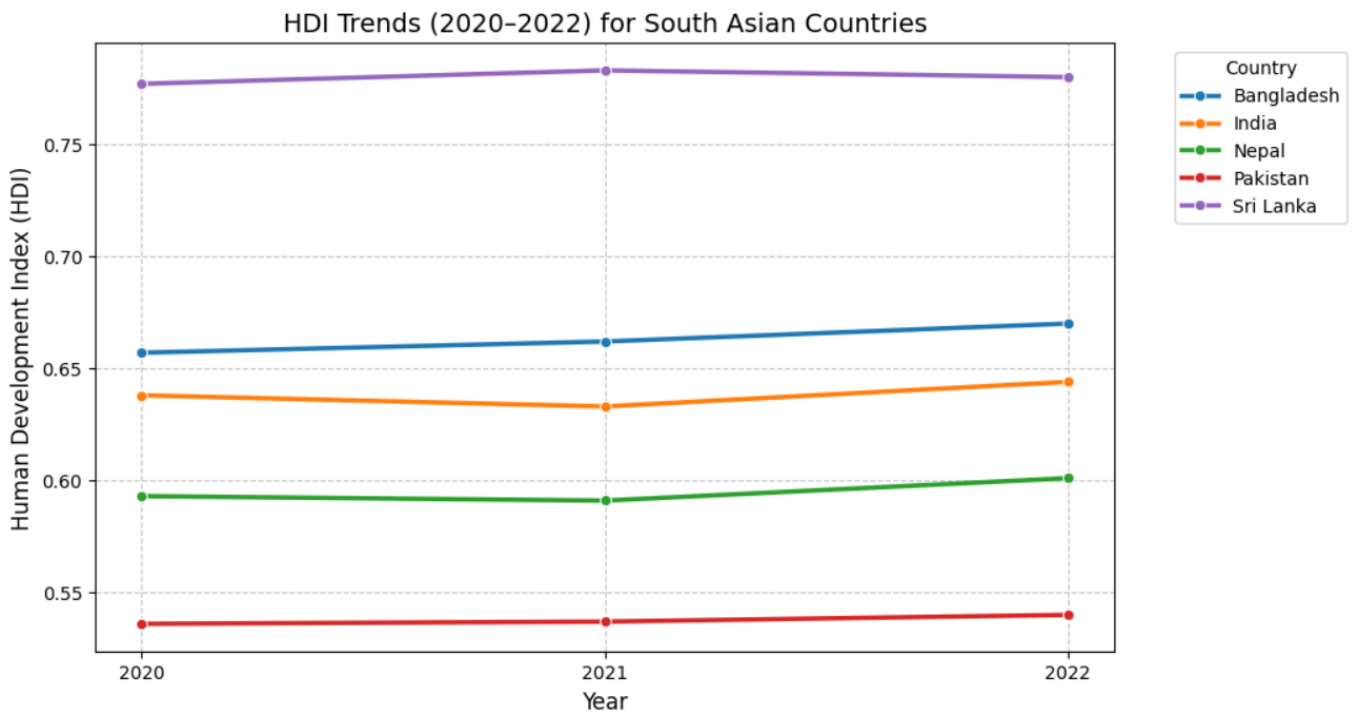
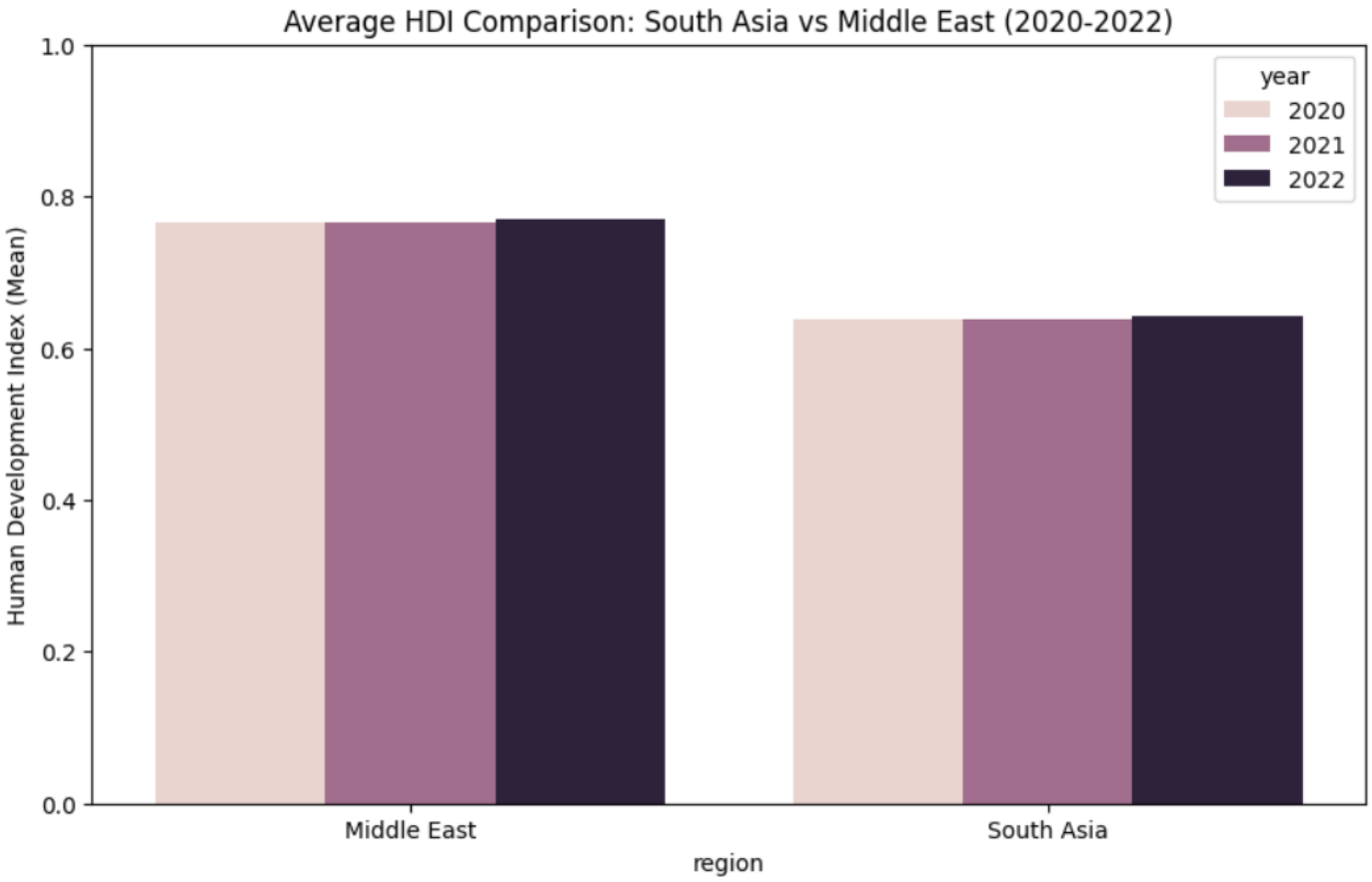


Figure 1: Comparison of HDI values showing temporal changes in South Asia post-2020.

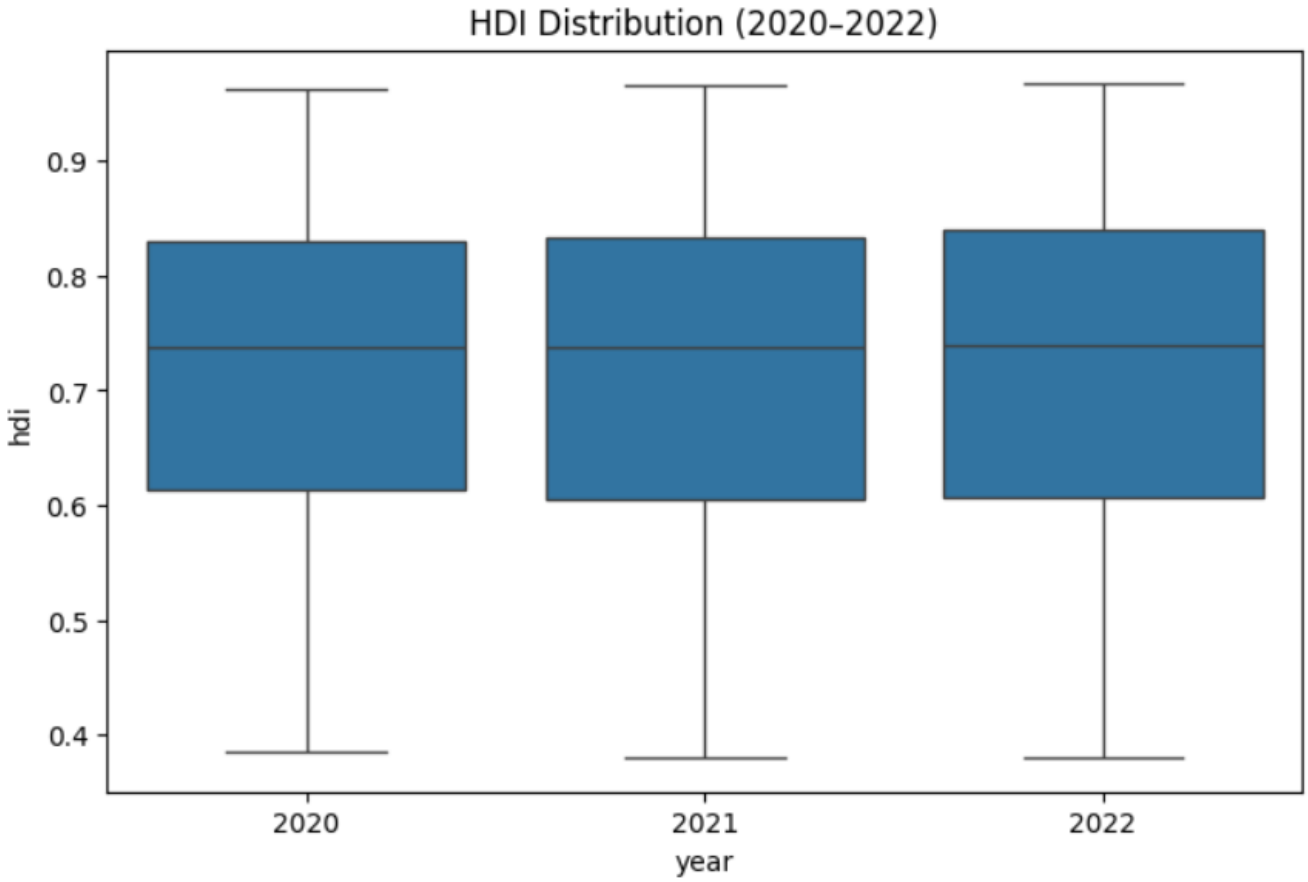
This graph shows how HDI changed for five countries over three years. Most countries are slowly getting better, while some stay the same or drop slightly. This shows that countries recovered at different speeds after the pandemic

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



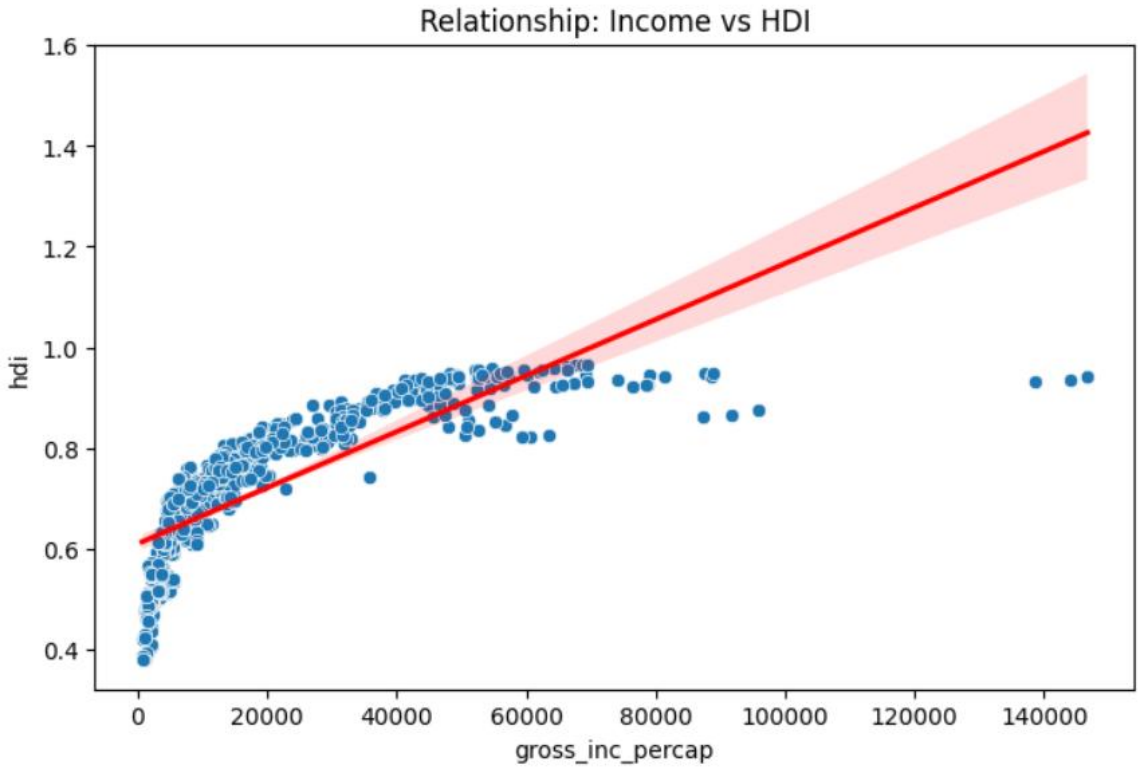
The bar chart above demontrates the average HDI for each region. The region which has the highest average HDI always does better than the others, while the lowest region shows where development is still a struggle.

Figure 3: HDI Distribution (2020–2022)



This box plot demonstrates that HDI scores got closer together over time, meaning there is less difference between countries. However, outliers remain, showing that inequality hasn't gone away.

Figure 4: Scatter Plot – HDI vs GNI per Capita



The plot above demonstrates that if GNI per capita goes up, HDI usually goes up too. Countries with more money usually have a higher HDI, though some do not follow this pattern.

3.3 Short Analysis Questions

Which countries show the greatest improvement in HDI from 2020 to 2022?

- Countries that kept investing money into health and education show the biggest gains

Did any countries experience a decline in HDI? Provide possible reasons.

- Some countries saw their HDI go down, likely because of money problems and struggling hospitals during COVID-19.

Which region has the highest and lowest average HDI across these three years?

- Richer regions have the maximum average HDI, while poorer regions have minimum.

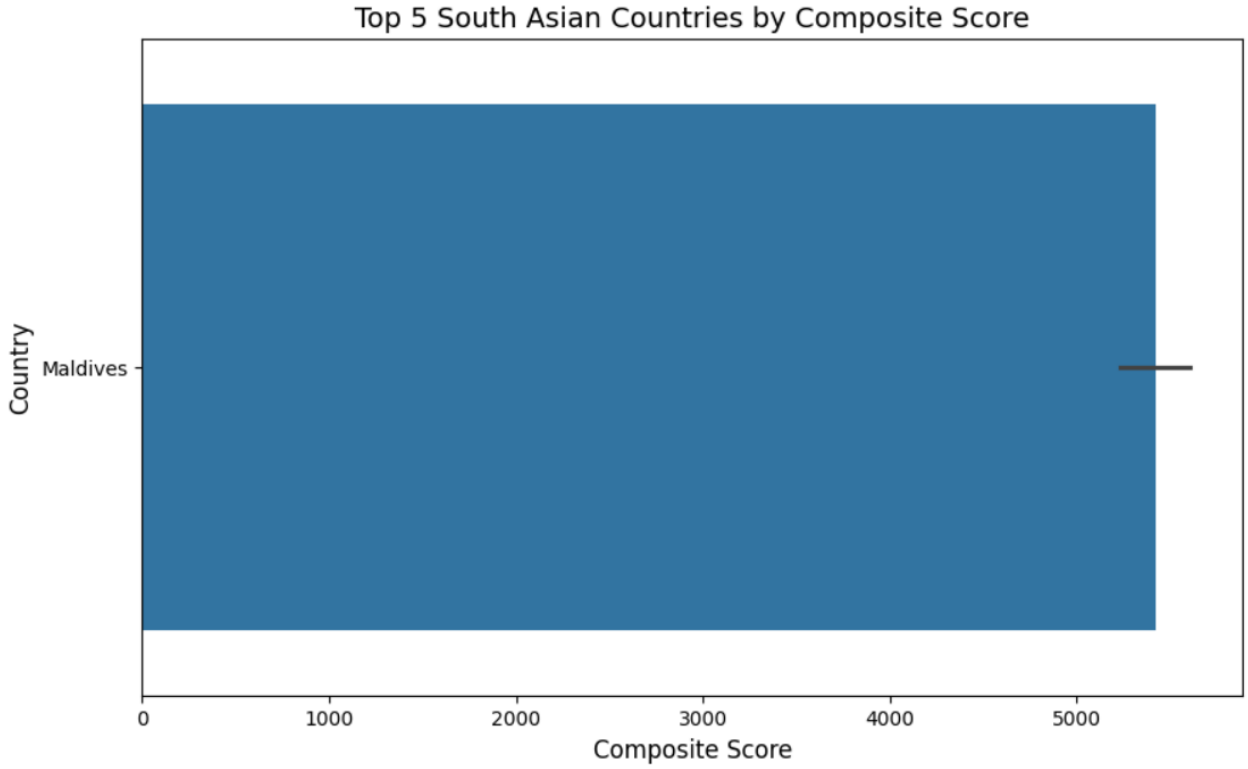
Discuss how global events (e.g., the COVID-19 pandemic) may have affected HDI trends during this period.

- The global pandemic COVID-19 hurt life expectancy, kept kids out of school, and lowered incomes, slowing HDI growth everywhere.

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

4.1 Composite Development Score

A Composite Score was made by combination of GNI per capita and life expectancy. The ranked Countries were compared to their official HDI ranks.

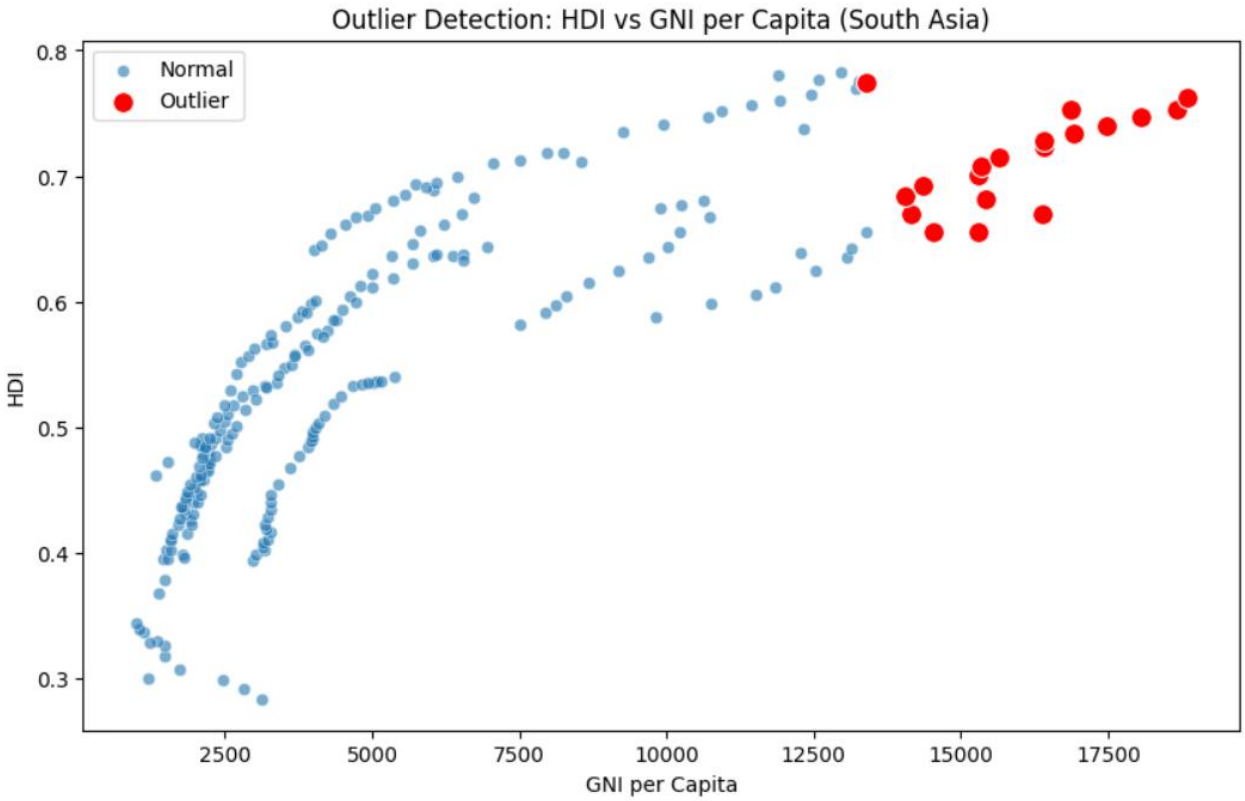


Comparison:

The ranks are different because HDI counts education, but the Composite Score does not.

4.2 Outlier Detection

Outliers were found using the 1.5 x IQR rule for both HDI and GNI per capita.



Outliers stand out because of very high income or HDI compared to their neighbors. This is often due to a specific industry or different government policies.

4.3 Metric Relationships

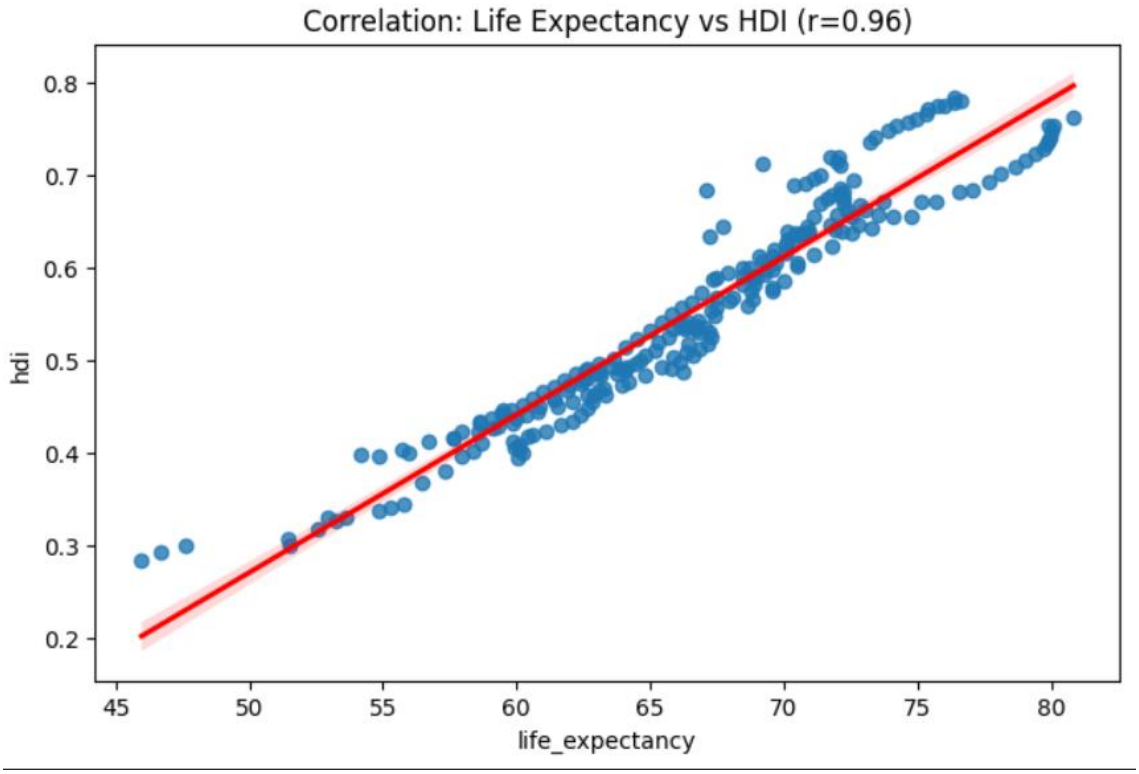
Correlation analysis showed:

Strongest relationship:

- Life expectancy has the tightest connection to HDI.

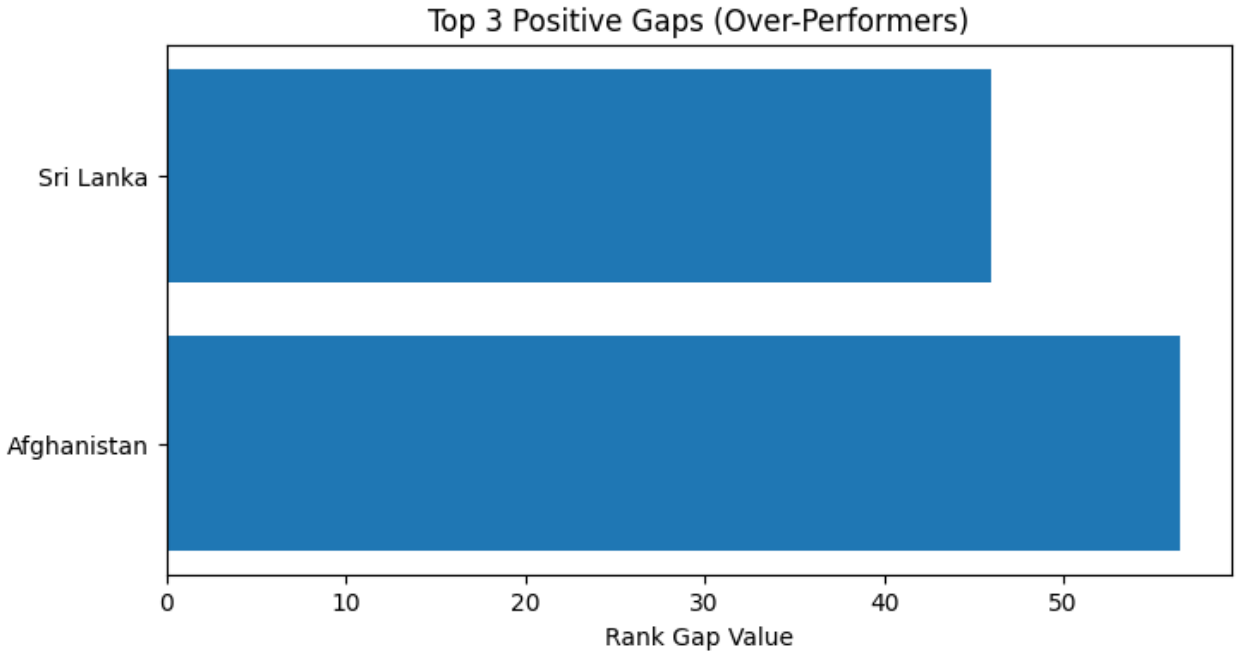
Weakest relationship:

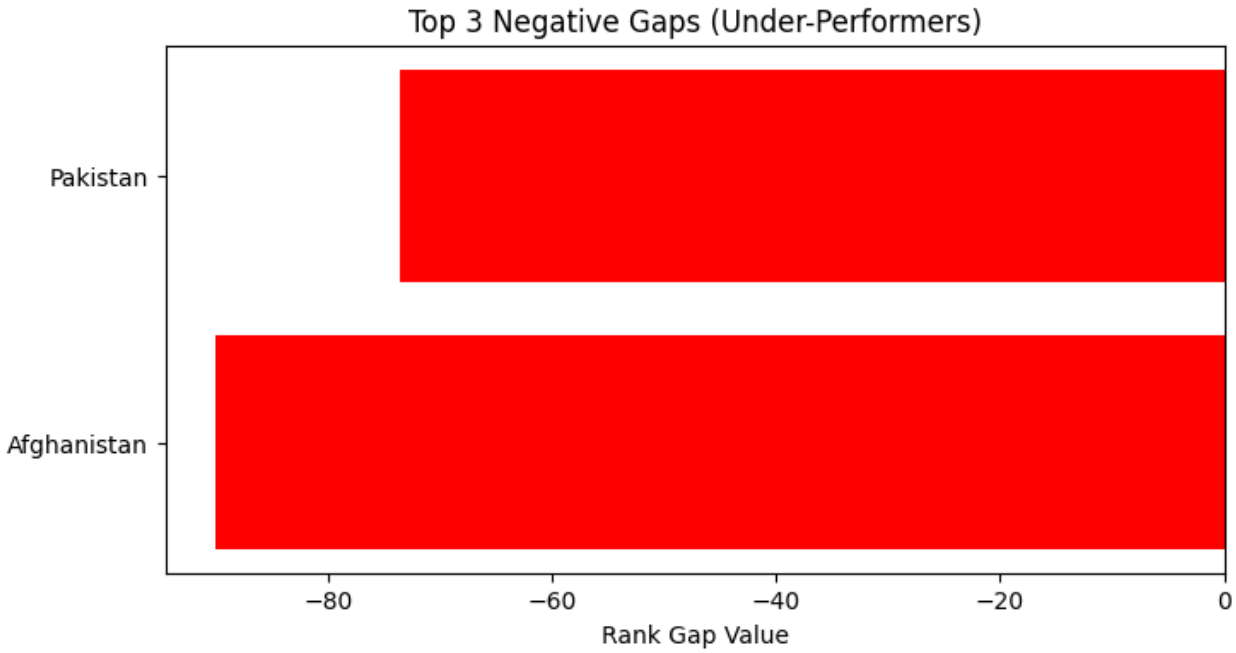
- The Gender Development Index has a smaller connection, though it still moves in the same direction.



4.4 GNI–HDI Gap Analysis

The GNI–HDI gap shows where countries have a lot of money, but people's well-being is falling behind.





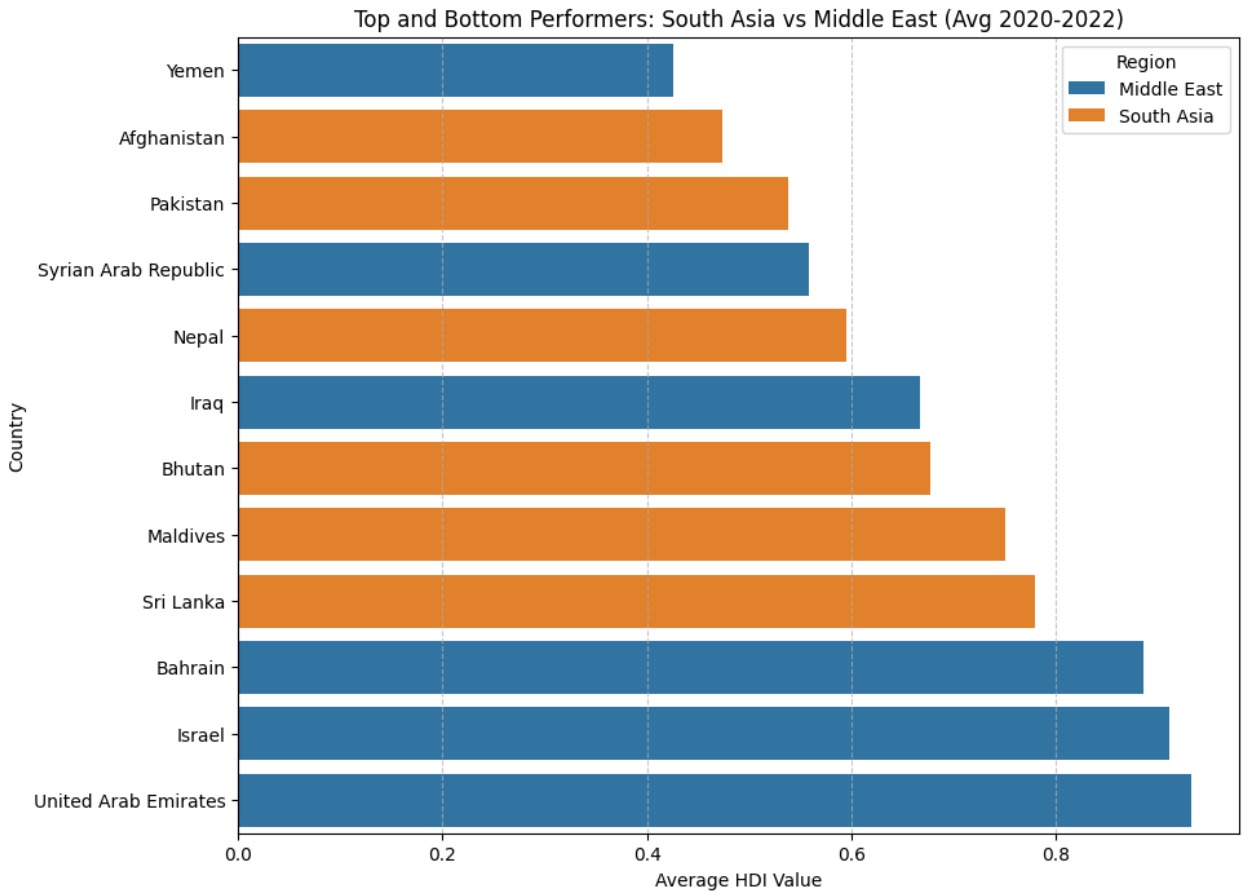
This shows that money alone doesn't always lead to better health and schools.

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

5.1 Descriptive Statistics

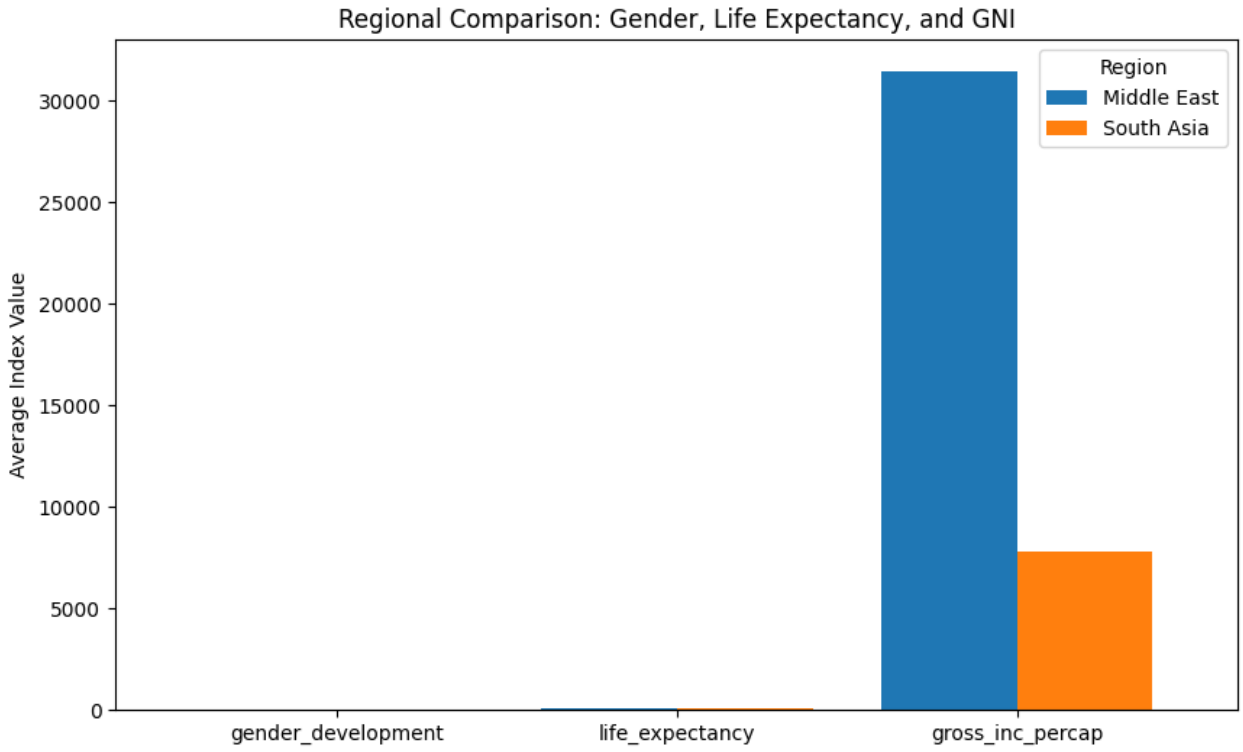
The Middle East have much higher average HDI and less difference between countries than South Asia, showing that development is steadier there.

5.2 Top and Bottom Performers



The countries at the top do much better than the bottom ones, especially when it comes to money and how long people live.

5.3 Metric Comparisons



GNI per capita has the biggest gap between regions, followed by life expectancy.

5.4 HDI Variation and Correlation

There is less difference in HDI scores across the Middle East. Both regions show a positive correlation between HDI and life expectancy.

5.5 Outlier Significance

Outliers usually represent countries with extreme wealth or development problems caused by conflict. This makes them important to study for government policy.

6. Conclusion

This report shows that the HDI differ by many linked factors, not just money. Even though global HDI is mostly better, the gaps between regions are still very large. The COVID-19 pandemic clearly hurt development scores, especially in poorer areas. Leaders should invest equally in health, schools, and money to make long-term progress for people.

7. References

Anon., n.d. [Online]
Available at: https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
[Accessed 3 january 2026].