

5CS037 - Concepts and Technologies of AI.

A Data-Based Analysis of Global and Regional Human Development

Name : Samip Regmi
Group: L5CG4
University Id: 2511198
Identification number: np02cs4a240105

Submitted To: Ayush Regmi Sir

Introduction	4
Human Development Index	4
Objectives of Report	4
Problem 1	5
3.1 Problem -1A- Single Year HDI Exploration (Latest Year 2022)	5
1. Extract Latest Year	5
2. Data Exploration	5
3. Missing Values and Data Cleaning	5
Justification	5
4. Basic Statistics	6
5. Filtering and Sorting	6
6. Adding HDI Category Column	6
3.2 Problem 1B- HDI Visualization and Trend Analysis (2020-2022)	7
1. Data Exploration and Saving	7
2. Data Cleaning	7
3. Visualization Tasks	7
1. A. Line Chart -HDI Trend (Country-Level)	7
Interpretation of Figure 1:	7
2. B. Generate Visualizations	8
Interpretation of Figure 2:	8
- Box Plot: HDI Distribution for 2020, 2021 and 2022	8
Interpretation of Figure 3	8
- Scatter Plot: HDI vs GNI Per Capital	9
Interpretation of Figure 4	9
4. Short Analysis Questions	9
Problem 2	10
Advanced HDI Exploration	10
1. Create South Asia Subset:	10
2. Composite Development Score	10
Interpretation of Figure 5	10
Interpretation of Figure of comparison of HDI and composite score	11
3. Outlier Detection	12
Interpretation of Figure 7	12
4. Exploring Metric Relationships	12
Interpretation of Figure 8	13
5. Gap Analysis	13
Interpretation of Figure 9	13
Problem 3	14
Comparative Regional Analysis: South Asia vs Middle East	14
1. Create Middle East Subset:	14
2. Descriptive statistics	14
3. Top and Bottom performers	14
Interpretation of Figure 10	15
Interpretation of Figure 11	15
4. Metric Comparisons	16
Interpretation of Figure 12	16
Interpretation of Figure 13	16
Interpretation of Figure 14	17
5. HDI Disparity	17
6. Correlation Analysis	18
Interpretation of Figure 15 and 16	18
7. Outlier Detection	19
Interpretation of Figure 17	19
8. Conclusion	20
9. Github Link	20
10. References and Citations	20
11. Appendix	21

Introduction

The Human Development Index (HDI) is a composite indicator used to assess a country's overall level of development. It reflects average achievements across three key dimensions: health, education, and standard of living. Health is measured using life expectancy at birth. Education is evaluated based on the mean years of schooling for adults aged 25 and above, together with the expected length of formal education for children at the start of their schooling. The standard of living is measured through gross national income (GNI) per capita, which is adjusted using a logarithmic scale. These three dimension-specific indices are then combined using a geometric mean to calculate the final HDI score.

The HDI is often used to compare development outcomes between countries and to examine how nations with similar income levels can achieve different results in health and education. Such comparisons can encourage discussion about policy decisions and national development priorities.

While useful, the HDI does not capture all aspects of human development. It does not directly address issues such as inequality, poverty, personal security, or empowerment. To complement the HDI, the Human Development Report Office (HDRO) offers additional composite indices that provide higher insights into inequality, gender gaps, and multidimensional poverty.

Objectives of Report

- To Explore the 2022 HDI dataset and compute data using Exploratory Data Analysis Techniques.
- To analyze multi year patterns from 2020 to 2022 to explore changes, regional differences and trends across countries.
- To perform advanced data analysis, with focus on South Asian Countries, detection of outliers, composite metrics and metrics relationships and gap analysis.
- To perform and compute related metrics between Middle East and South Asia, comparative analysis of HDI

Problem 1

3.1 Problem -1A- Single Year HDI Exploration (Latest Year 2022)

In problem 1A, we explored the dataset of HDI for the latest year and used basic Exploratory Data Analysis (EDA) Techniques.

1. Extract Latest Year

- The latest year in the database was identified using .unique() constraint in year column
- Observations from year 2022 was only included by comparing year column with 2022
- The filtered dataframe was saved as hdi_2022_df using .to_csv() method

2. Data Exploration

- First 10 rows of 2022 dataset was displayed using .head(10) method
- Number of rows and columns were identified using .shape property

rows	columns
206	30

- All columns and their datatype was identified using .info() method

3. Missing Values and Data Cleaning

- Missing values were checked in each column and their count was computed using .isnull().sum() in dataframe
- Columns datatypes were checked using .dtypes property which verified that numeric columns have correct datatype
- Inconsistent Country Names were checked using .strip
- Duplicated rows were checked using .duplicated.sum() method and were handled by dropping the duplicated values
- Data with missing symbols '-' were replaced with NA and null values were handled by removing rows with null values less than 10% and more than 10% were imputed with column mean

Justification

Missing data were handled based on the proportion of missingness. For columns with 10% or less missing values, rows containing missing observations were removed to maintain accuracy. For columns with more than 10% missing values, missing entries were imputed using the column mean to preserve the data distribution

4. Basic Statistics

- The mean, standard deviation and median for HDI in year 2022 were computed using mean(), std() and median() methods

Mean	Median	Standard Deviation
0.7268333333333336	0.7415	0.15219828657299067

- The Highest and Lowest value of HDI was computed using loc() method with the help of idxmax() method

Ranking	Country	HDI
Highest	Switzerland	0.967
Lowest	Somalia	0.38

5. Filtering and Sorting

- Countries with hdi greater than 8 were computing using comparison
- And these countries were sorted in descending order by their GNI values
- Top ten countries ordered by their GNI values were displayed using .head(10) method

Country	GNI Values
Qatar	95944.37754
Singapore	88761.14559
Ireland	87467.51319
Luxembourg	78554.23640
United Arab Emirates	74103.71494
Switzerland	69432.78669
Norway	69189.76165
United States	65564.93798
Denmark	62018.95694
Brunei Darussalam	59245.63485

6. Adding HDI Category Column

- A new column called HDI_Categorgy was created which classified each country into four official Human Development Index groups.

HDI Category	HDI Range
Low	< 0.550
Medium	0.550 - 0.699
High	0.700 - 0.799
Very	>= 0.800

3.2 Problem 1B- HDI Visualization and Trend Analysis (2020-2022)

In Problem 2B section we analyzed multi-year HDI patterns (2020, 2021 and 2022) and explored temporal changes, regional differences and trends across countries.

1. Data Exploration and Saving

- Dataset was filtered to only include years 2020, 2021 and 2022 using .isin() method
- The dataset which was filtered was saved as HDI_problem1B.csv with the help of .to_csv() method

2. Data Cleaning

- Missing values in the hdi, country and year was checked using .isnull().sum() method

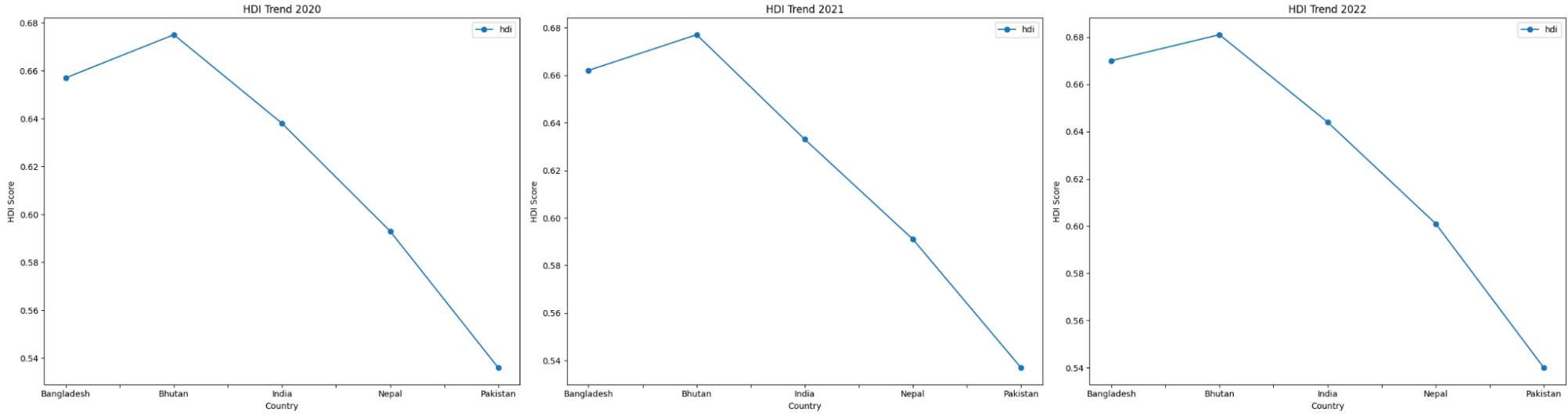
Columns	Null Count
HDI	8
Country	0
Year	0

- Missing data were handled based on the proportion of missingness. For columns with 10% or less missing values, rows containing missing observations were removed to maintain accuracy. For columns with more than 10% missing values, missing entries were imputed using the column mean to preserve the data distribution
- Duplicated data were removed using the .drop_duplicates() method.

3. Visualization Tasks

1. A. Line Chart -HDI Trend (Country-Level)

- I selected five countries which are Nepal, India, Bangladesh, Bhutan and Pakistan which are from South Asia Region using .isin() method
- Hdi Values for each countries were plotted across 2020,2021 and 2022



Interpretation of Figure 1:

- The figure illustrates a consistent regional trend across all three years, where Bhutan maintains the highest Human Development Index (HDI) score and Pakistan consistently records the lowest.

2. B. Generate Visualizations

- Bar Chart: HDI Average By Region (2020 - 2022)

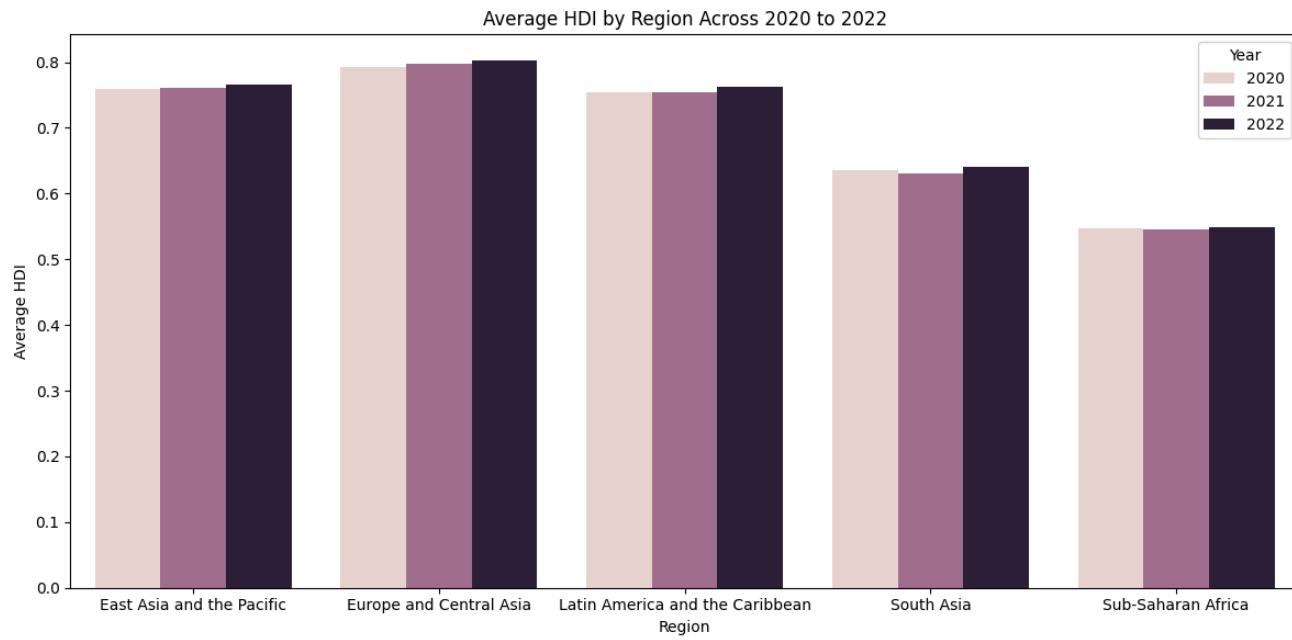


Figure 2: Average Human Development Index (HDI) scores by region for the years 2020, 2021, and 2022.

Interpretation of Figure 2:

- Europe and Central Asia consistently show the highest mean HDI values, whereas Sub-Saharan Africa records the lowest levels throughout all three years. The results point to a clear development gap, with Europe, East Asia, and Latin America all sustaining average HDI scores above 0.75, well ahead of South Asia and Sub-Saharan Africa.
- Box Plot: Distribution of HDI for 2020, 2021 and 2022

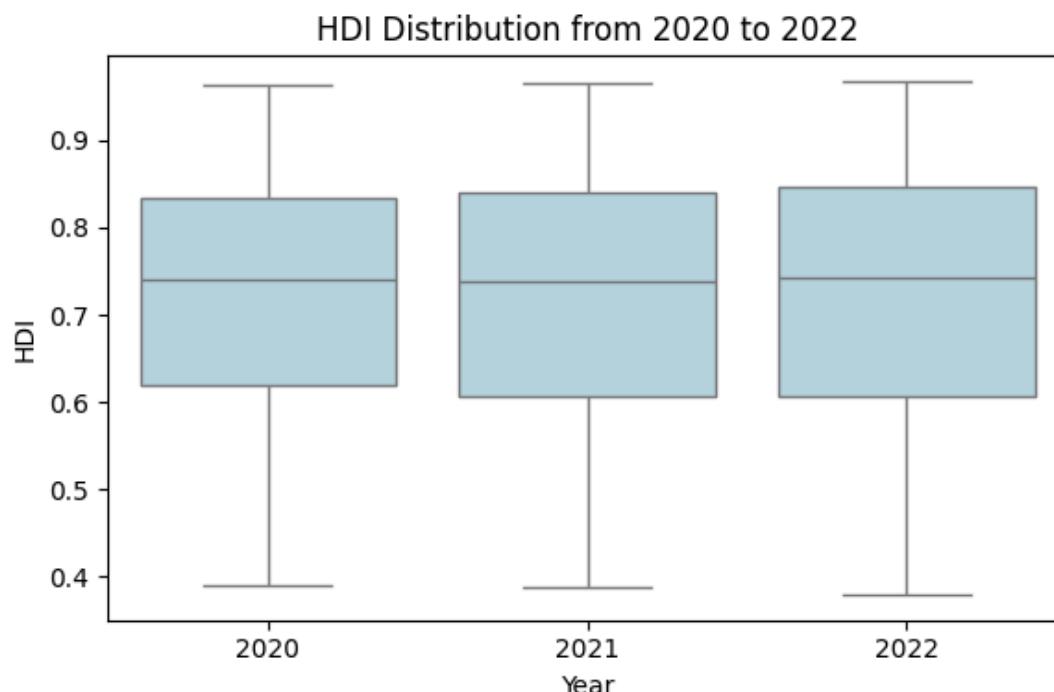


Figure 3: Distribution of Human Development Index (HDI) scores for the years 2020, 2021, and 2022

Interpretation of Figure 3

- The box plots reveal a remarkably stable distribution of HDI scores across 2020, 2021, and 2022, with the median remaining consistent around 0.74 and the interquartile ranges showing almost no variation.

- Scatter Plot: HDI vs GNI Per Capita

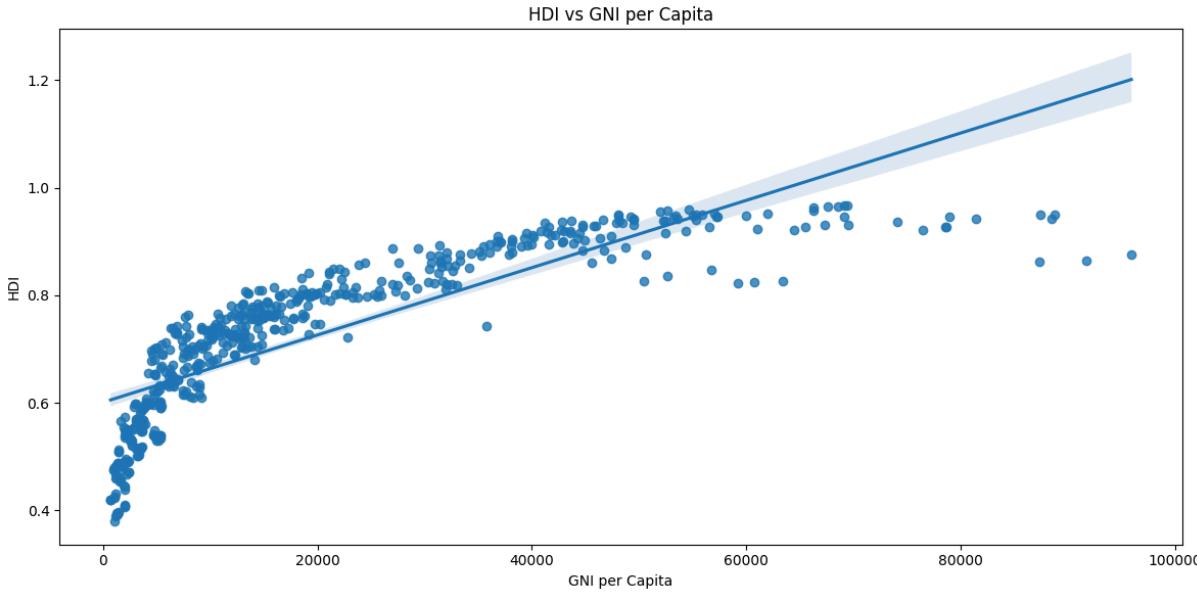


Figure 4: Relationship between Gross National Income (GNI) per capita and Human Development Index (HDI).

Interpretation of Figure 4

- The chart demonstrates a strong positive correlation between GNI per Capita and HDI, confirming that higher national income is closely linked to better human development outcomes. However, the relationship follows a logarithmic curve rather than a straight line, indicating that while wealth initially boosts HDI significantly it also relates to other factors.

4. Short Analysis Questions

- Which Country Shows the greatest improvement in HDI from 2020 to 2022?**
 - We can see Bangladesh showing the greatest improvement in HDI from 2020 to 2022 with a positive value of 0.013. This improvement is mainly due to recovery in life expectancy, education continuity and economic activity after the initial impact of COVID 19 pandemic.
- Did Any Countries experience a decline in HDI? Provide possible reason**
 - No, Based on the calculations above, we can see none of the selected countries experienced a decline in HDI. All countries show positive HDI growth and overall development in human development in this period.
 - Also many countries can also experience a decline in HDI
 - COVID 19 related deaths, reducing life expectancy
 - Economic slowdown, lowering income levels
 - Disruptions in Education, such as prolonged school closures
- Which region has the highest and lowest average HDI across these three years?**
 - The Highest average of HDI region was Europe and Central Asia with hdi value of 0.797 and Lowest average of HDI region was Sub-Saharan Africa with 0.54733
- Discuss how global events (e.g the COVID-19 pandemic) may have affected HDI trends during this period.**
 - The COVID-19 pandemic caused a global setback in HDI, particularly in 2020 and 2021. Major impacts included:
 - Reduced life expectancy
 - Loss of jobs and income
 - Interrupted education systems

Problem 2

Advanced HDI Exploration

Advanced analysis of HDI data was done giving much priority on South Asian countries , outlier detection,composite metrics, metric relationships and gap analysis

1. Create South Asia Subset:

- South Asian Countries were defined.
- HDI dataset were filtered to include only these countries using `.isin()` method
- The filtered dataset was saved as `HDI_SouthAsia.csv` using `to_csv()` method

2. Composite Development Score

- New metric called `Composite_Score` was added using the formula **$0.30 * \text{Life Expectancy Index} + 0.30 * \text{GNI Per Capita Index}$**
- South Asian Countries were ranked based on their composite Score
- Top 5 Countries were plotted in horizontal bar chart

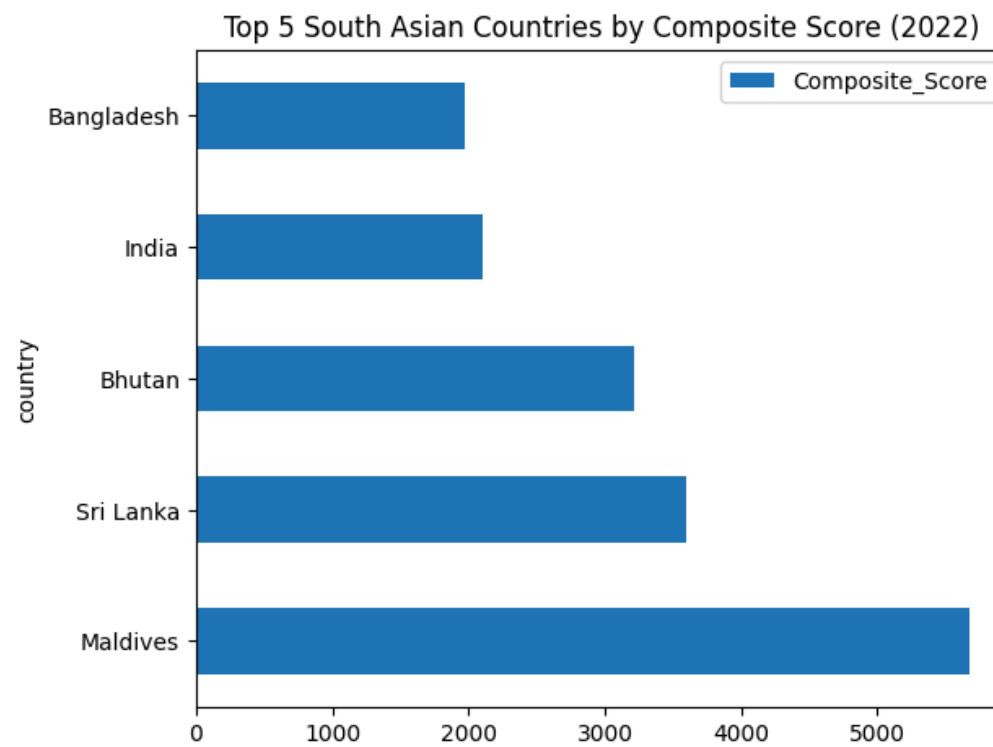


Figure 5: Top five South Asian countries ranked by composite score in 2022.

Interpretation of Figure 5

- The chart identifies the Maldives as the top performing country in South Asia for 2022, achieving a Composite Score significantly higher than the runner up, Sri Lanka. This ranking highlights a regional trend where smaller nations like the Maldives and Bhutan currently outperform larger economies like India and Bangladesh on this specific combined measure of life expectancy and income.

- Comparison between the countries by Composite_Score and HDI

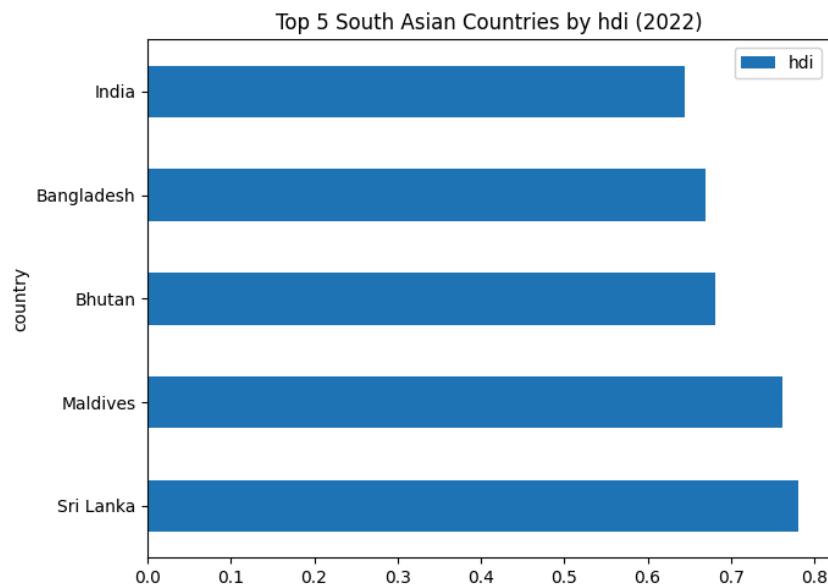


Figure 6: Top five South Asian countries ranked by Human Development Index (HDI) in 2022.

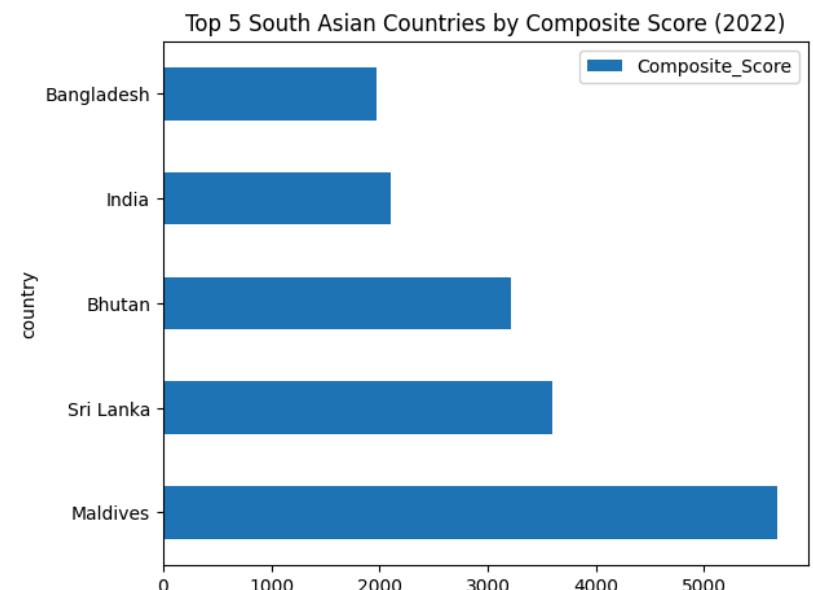


Figure 5: Top five South Asian countries ranked by composite score in 2022.

Interpretation of Figure of comparison of HDI and composite score

- While Composite Score highlights overall national strength and capacity, HDI better captures quality of life. Smaller nations like Maldives and Sri Lanka excel in both, whereas larger South Asian countries still face challenges translating economic scale into human development.

3. Outlier Detection

- Outliers were detected in HDI and GNI using $1.5 * \text{IQR}$ Rule, custom function was made and was applied to HDI and GNI column using `.apply()` method
- Scatter plot of GNI per capita vs HDI

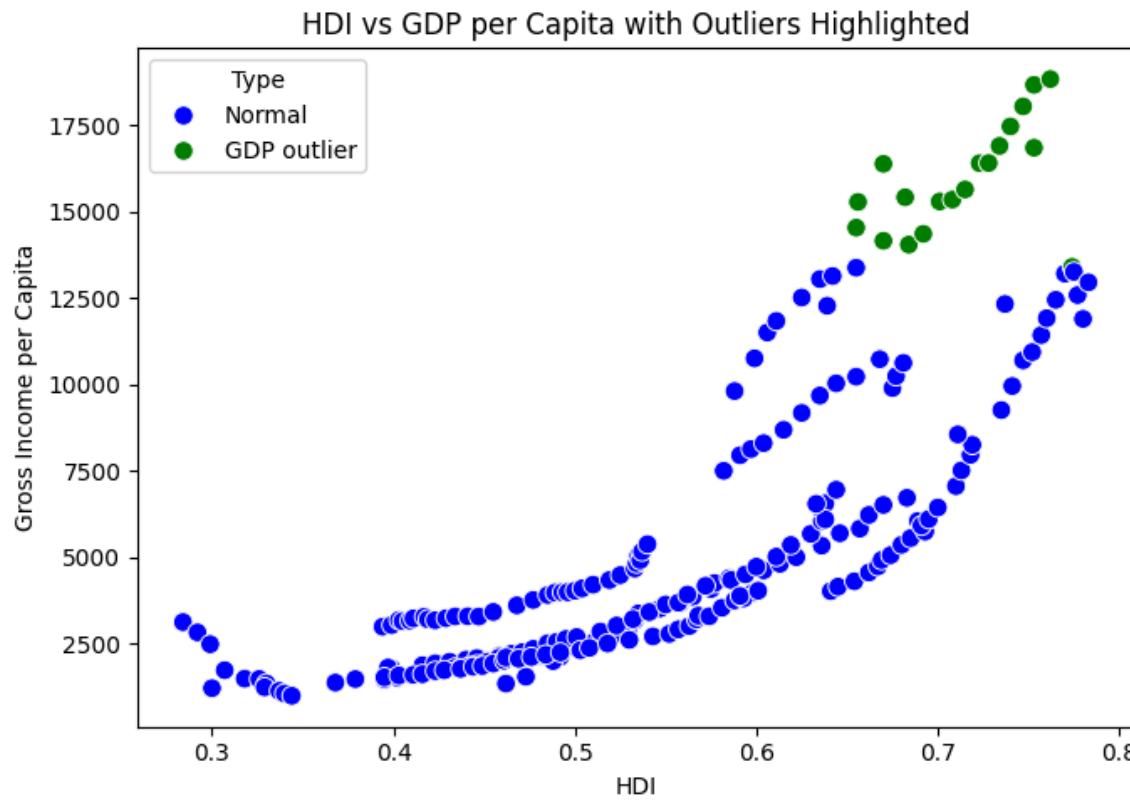


Figure 7: Relationship between Human Development Index (HDI) and Gross Income per Capita for South Asian countries,

Interpretation of Figure 7

- The plot illustrates a positive, nonlinear correlation where Gross Income per Capita rises exponentially as HDI improves, indicating that economic growth accelerates significantly at higher development levels. The green outliers show a distinct economic trajectory that deviates from the standard South Asian pattern.

4. Exploring Metric Relationships

- Two HDI components which are Life Expectancy Index and Gender Development Index were selected
- The `.corr()` method was used to compute the Pearson correlation of each metric with HDI.
- Scatter plot to visualize relationship

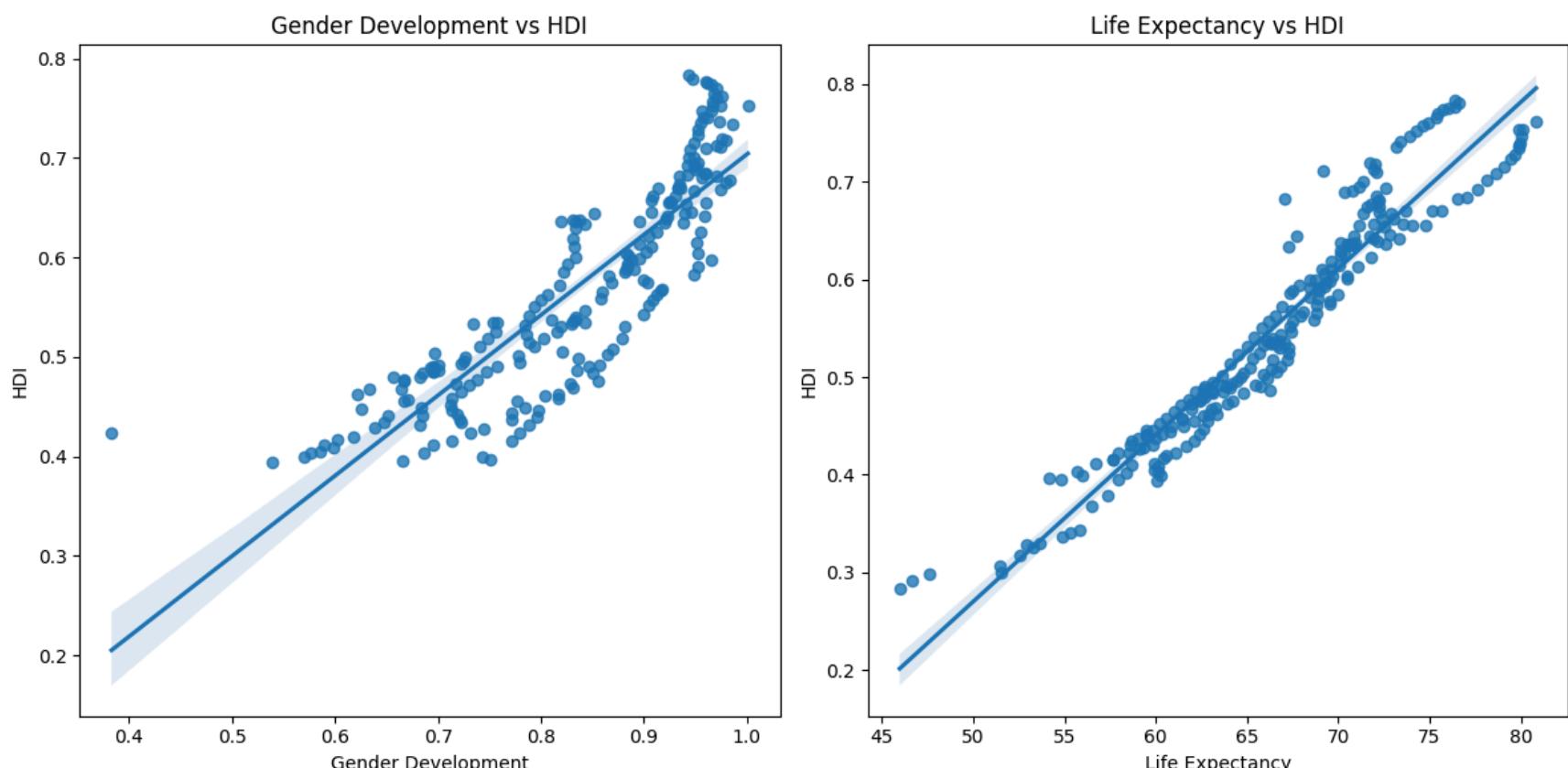


Figure 8: Relationship between Human Development Index (HDI) and two key development indicators in South Asia.

Interpretation of Figure 8

- The visual comparison reveals that Life Expectancy is a significantly stronger and more consistent predictor of HDI in South Asia than Gender Development, as evidenced by the extremely tight linear clustering of data points . While Gender Development also shows a positive correlation, the noticeably wider dispersion of points indicates a weaker relationship with more variability .
- Discussion
 - Which metric is most strongly related to HDI and shows the weakest relation with HDI.
 - The analysis of correlations shows that Life Expectancy has the strongest relationship with the Human Development Index (HDI), with a correlation coefficient of approximately 0.959. This indicates that countries with higher life expectancy tend to have higher HDI scores. In contrast, Gender Development exhibits the weakest relationship with HDI, with a correlation of around 0.866, suggesting that although it is positively associated with HDI, the strength of this relationship is lower compared to life expectancy

5. Gap Analysis

- A new metric called GNI_HDI_Gap was calculated with formula “gross_inc_percap” - “hdi”
- South Asian countries were ranked by GNI_HDI_Gap in descending order and ascending order using .sort_values() using by = “GNI_HDI_Gap” setting ascending = True
- Plot the top three positive gaps and top three negative gaps.

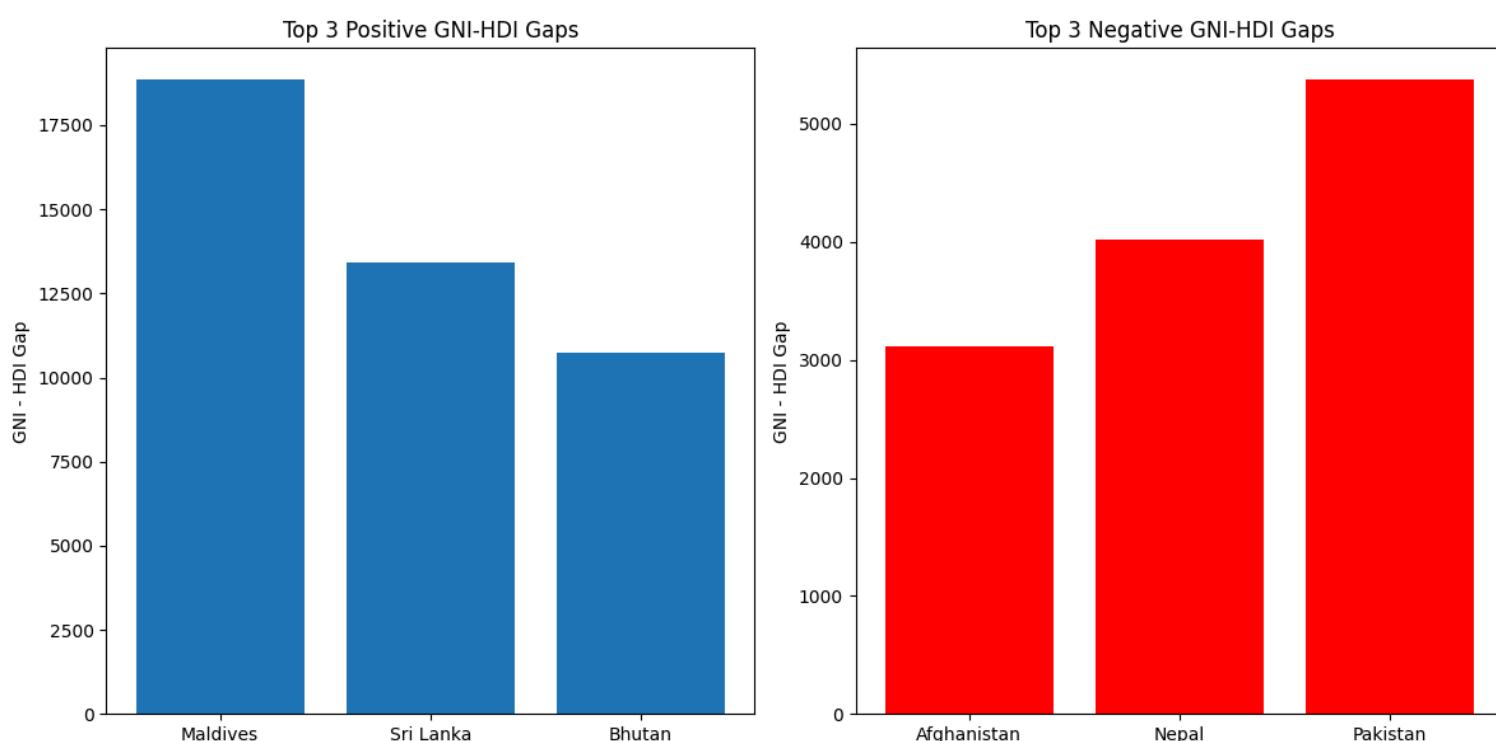


Figure 9: Comparison of the top three positive and negative gaps between Gross National Income (GNI) and Human Development Index (HDI)

- Discuss the implications of the gap, eg cases when GNI is high but HDI is lower than expected

A noticeable gap between high GNI per capita and lower than expected HDI suggests that economic wealth alone does not guarantee human development. In several cases, countries with high income levels do not achieve equally high HDI scores because HDI also depends on health outcomes, education quality, and income distribution.

This gap highlights the importance of inclusive development policies that prioritize health, education, and social equity rather than focusing solely on economic growth.

Interpretation of Figure 9

- The charts highlight an economic disparity within South Asia, identifying the Maldives, Sri Lanka, and Bhutan as having the largest positive gaps driven by their substantially higher national incomes relative to their HDI scores. In contrast, the negative gaps for Afghanistan, Nepal, and Pakistan underscore these nations as the most economically constrained, indicating that low income levels are the primary limiting factor in their overall human development.

Problem 3

Comparative Regional Analysis: South Asia vs Middle East

In Problem 3, we analyzed and compared HDI and its related indicators for South Asia and the Middle East, using the 2020–2022 dataset from Problem 1B.

1. Create Middle East Subset:

- A List of Middle East countries were selected and were defined as
[“Bahrain”, “Iran”, “Iraq”, “Israel”, “Jordan”, “Kuwait”, “Lebanon”, “Oman”, “Palestine”, “Qatar”, “Saudi Arabia”, “Syria”, “United Arab Emirates”, “Yemen”]
- Dataset from problem 1B (HDI_problem1B.csv) was filtered to create subsets for Middle East and South Asia using df_filtered dataframe and countries were checked in it by .isin() method
- These datasets were saved and exported using the .to_csv() method.

2. Descriptive statistics

- Standard Deviation and Mean of HDI was computed for each region using .mean() method in individual region

Region	Mean	Standard Deviation
Middle East	0.776733334	0.142622224
South Asia	0.639583333	0.098273317

- Identify the region that performs better on average

3. Top and Bottom performers

- South Asian and Middle Eastern countries were grouped by their HDI values by their mean across the years.

Regions	South Asia	HDI Value
Top 3		
South Asia	Sri Lanka	0.78000
	Maldives	0.750667
	Bhutan	0.677667
Middle East	United Arab Emirates	0.932667
	Bahrain	0.885333
	Saudi Arabia	0.867667
Bottom 3		
South Asia	Afghanistan	0.474333
	Pakistan	0.53667
	Nepal	0.595000
Middle East	Yemen	0.426333
	Iraq	0.667000
	Lebanon	0.730000

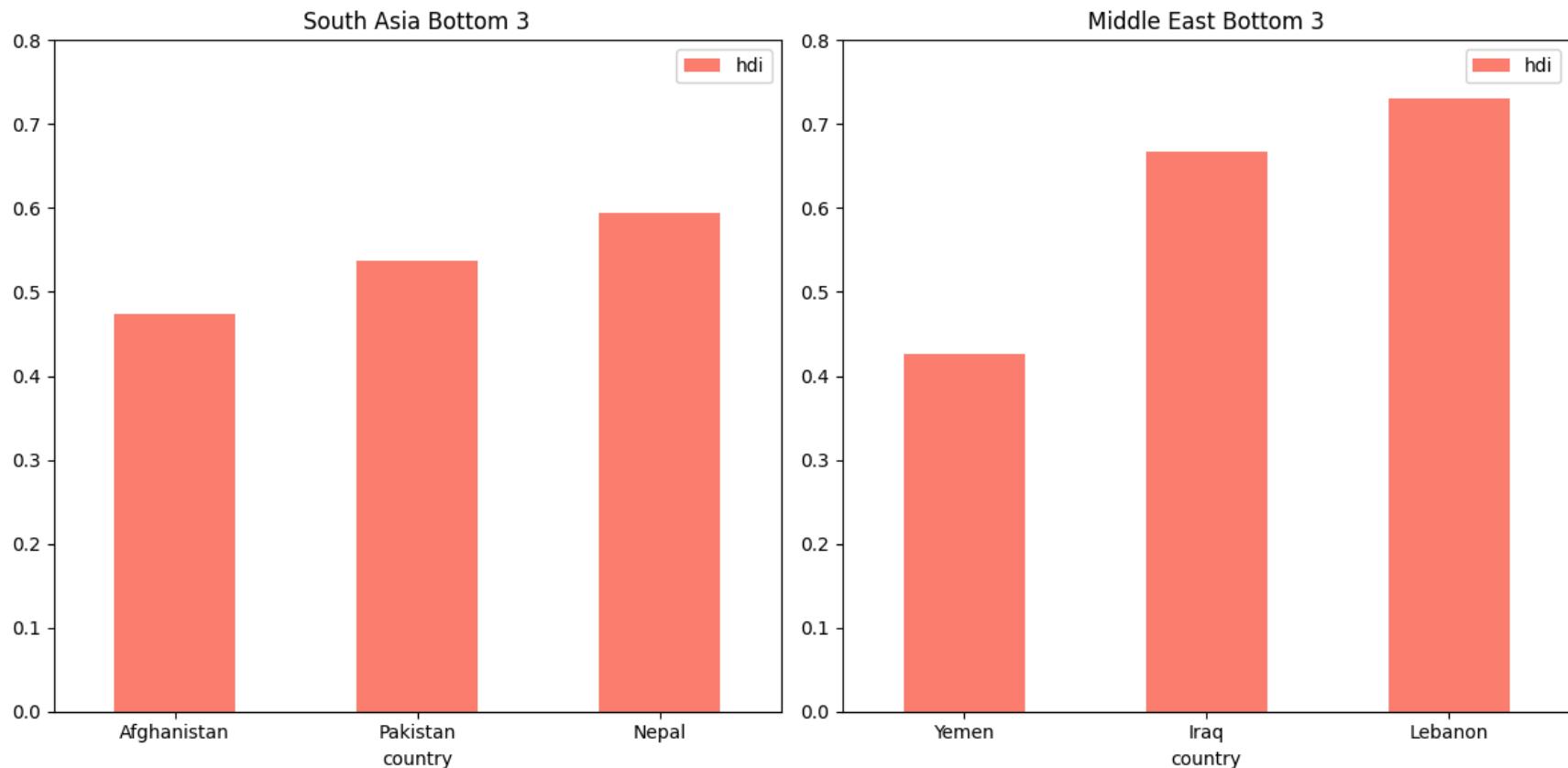


Figure 10: Comparison of the bottom three countries by Human Development Index (HDI) in South Asia and the Middle East.

Interpretation of Figure 10

- The figure highlights a distinct regional disparity where the Middle East's nations significantly outperform South Asia's bottom countries, indicating a much higher development baseline in the Middle East. However, the similarly critical scores of Afghanistan and Yemen suggest that severe instability creates comparable depths of human development crises regardless of the region's broader economic strength.

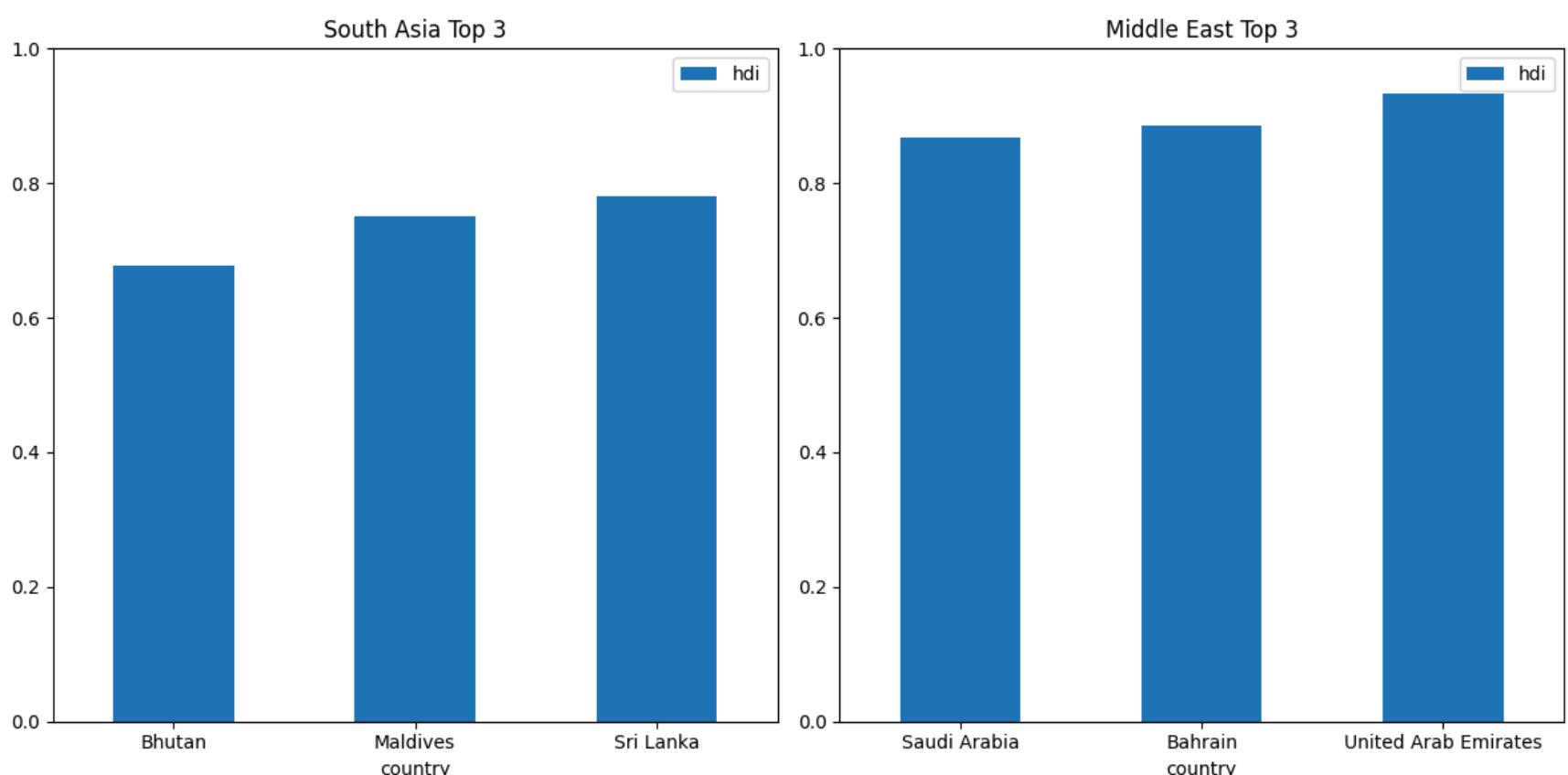


Figure 11: Comparison of the top three countries by Human Development Index (HDI) in South Asia and the Middle East.

Interpretation of Figure 11

- The comparison reveals a significant regional disparity, as the top performing Middle Eastern nations all achieve very high HDI scores that substantially exceed South Asia's leaders. This contrast highlights that the development ceiling in the Middle East is currently much higher

4. Metric Comparisons

- Metrics like gender_development, life_expectancy, gross_inc_percap was calculated by their mean() and compared

Metrics	South Asia	Middle East
Gender Development	0.881458333	0.881458333
Life Expectancy	70.970083333	70.970083333
Gross Income Per Capita	7766.5099314	7766.509931

- Grouped Barcharts
 - Gender development Index across South Asia and Middle East

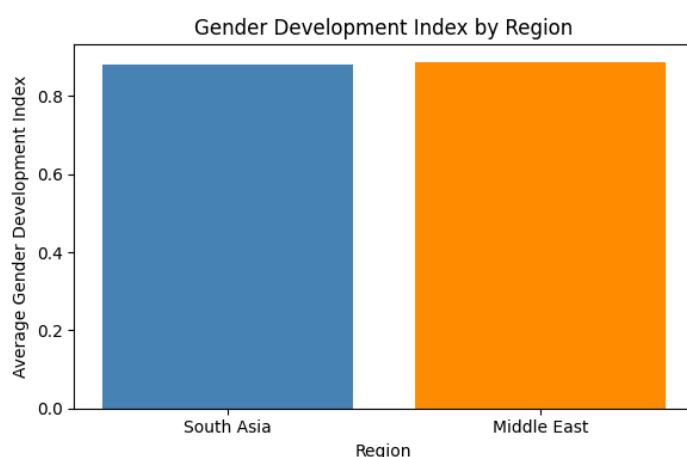


Figure 12: Comparison of the average Gender Development Index between South Asia and the Middle East.

Interpretation of Figure 12

- The chart illustrates that the average Gender Development Index in South Asia is nearly identical to that of the Middle East. This suggests that despite the Middle East's substantial economic advantage, both regions face comparable challenges regarding gender based development disparities, with neither holding a significant lead in gender equity.
- Comparing Life Expectancy in South Asia and the Middle East

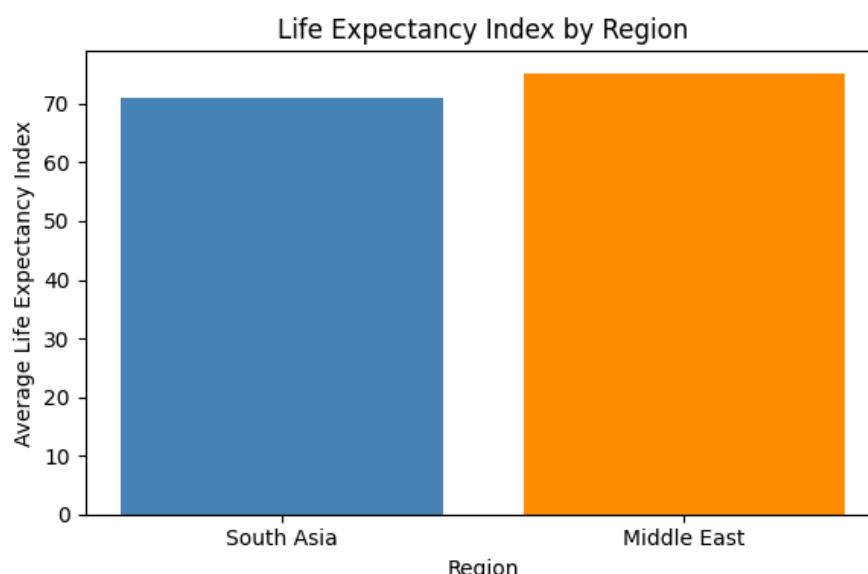


Figure 13: Comparison of the average Life Expectancy Index between South Asia and the Middle East.

Interpretation of Figure 13

- The chart demonstrates that the Middle East maintains a higher average life expectancy than South Asia, reflecting better overall health infrastructure and living standards in the region. However, the relatively narrow gap between the two bars indicates that South Asia performs surprisingly well in public health outcomes, trailing the wealthier Middle East by a much smaller margin than seen in economic metrics like GNI.

- GNI per capita index across South Asia and Middle East

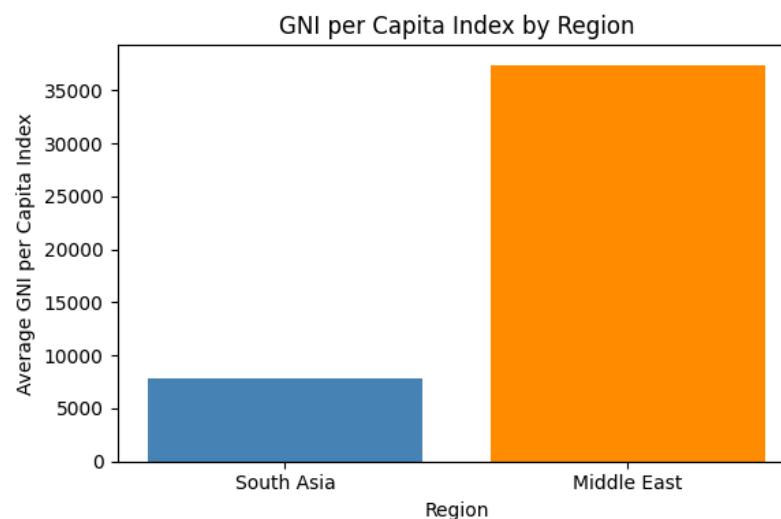


Figure 14: Comparison of the average Gross National Income (GNI) per capita index between South Asia and the Middle East.

Interpretation of Figure 14

- The chart reveals a massive economic divide, with the Middle East having an average GNI per capita higher than South Asia. This overwhelming disparity identifies national income as the primary driver of inequality between the two regions, significantly outweighing the smaller gaps seen in health and gender metrics.

5. HDI Disparity

- The range of HDI of each region was computed using max - min formula

Regions	HDI Value
South Asia	0.321
Middle East	0.513

- Coefficient of Variation for HDI was computed using ($CV = \text{std}/\text{mean}$) formula

Region	Coefficient of Variation
South Asia	0.1536520
Middle East	0.18361800392921748

- Region which exhibits more variation in HDI
 - Middle east is the region with more variation in HDI

6. Correlation Analysis

- For each region , correlation of HDI was computed with Life Expectancy Index and Gender Development Index using .corr() method

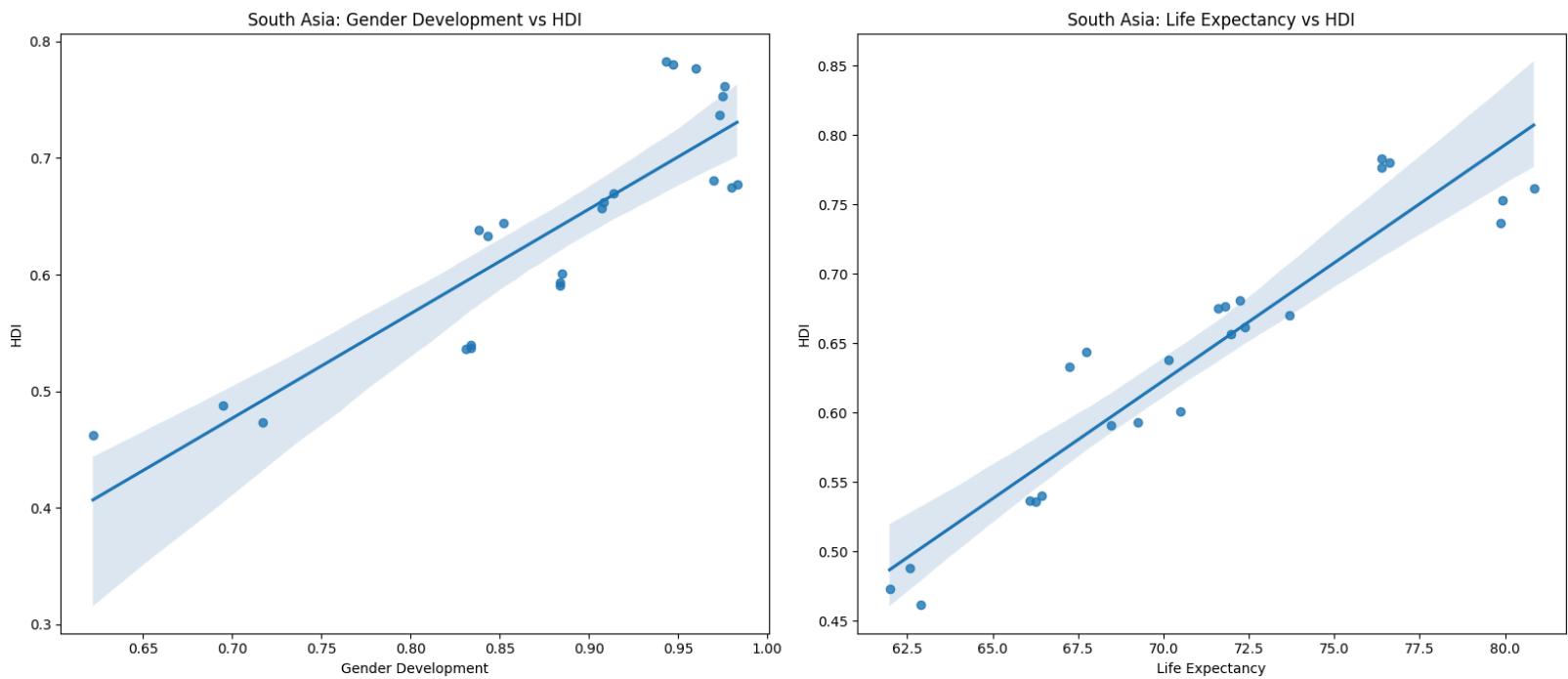


Figure 15: Relationships between Human Development Index (HDI) and two key development indicators in South Asia.

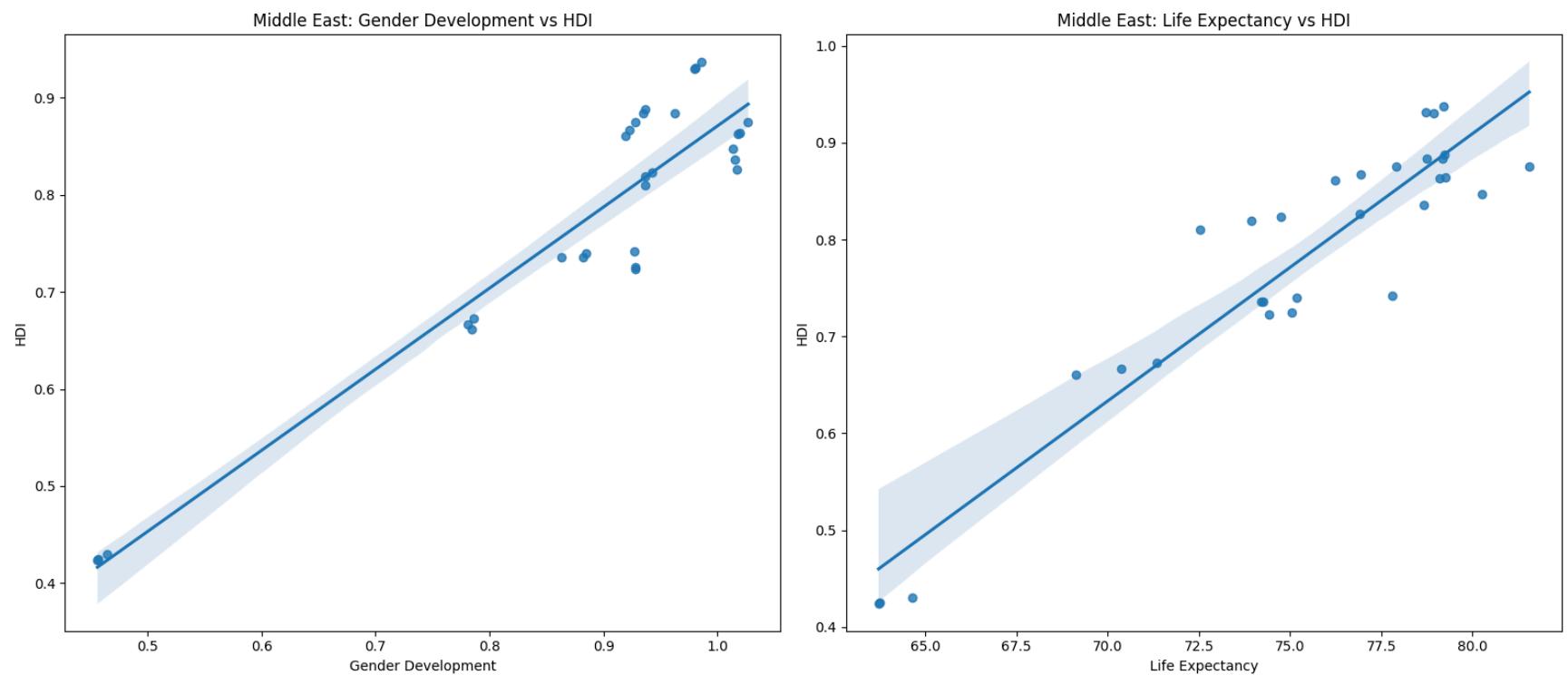


Figure 16: Relationships between Human Development Index (HDI) and key development indicators in the Middle East.

Interpretation of Figure 15 and 16

- We can see from the above graph and data that the Middle East is largely related to HDI and gender development with correlation of 0.9360954290299537 and is also largely related to HDI and life expectancy 0.9360212027493279.
- But in the case of South Asia, South Asia is comparatively more correlated to HDI and Life expectancy with correlation of 0.9387641385416623 compared to that of 0.8742177004831327 with HDI and gender development.

7. Outlier Detection

- Outliers were detected using the outliers() method where series data is sent as an argument and the column series data was sent as an argument
- Scatter plot for outliers

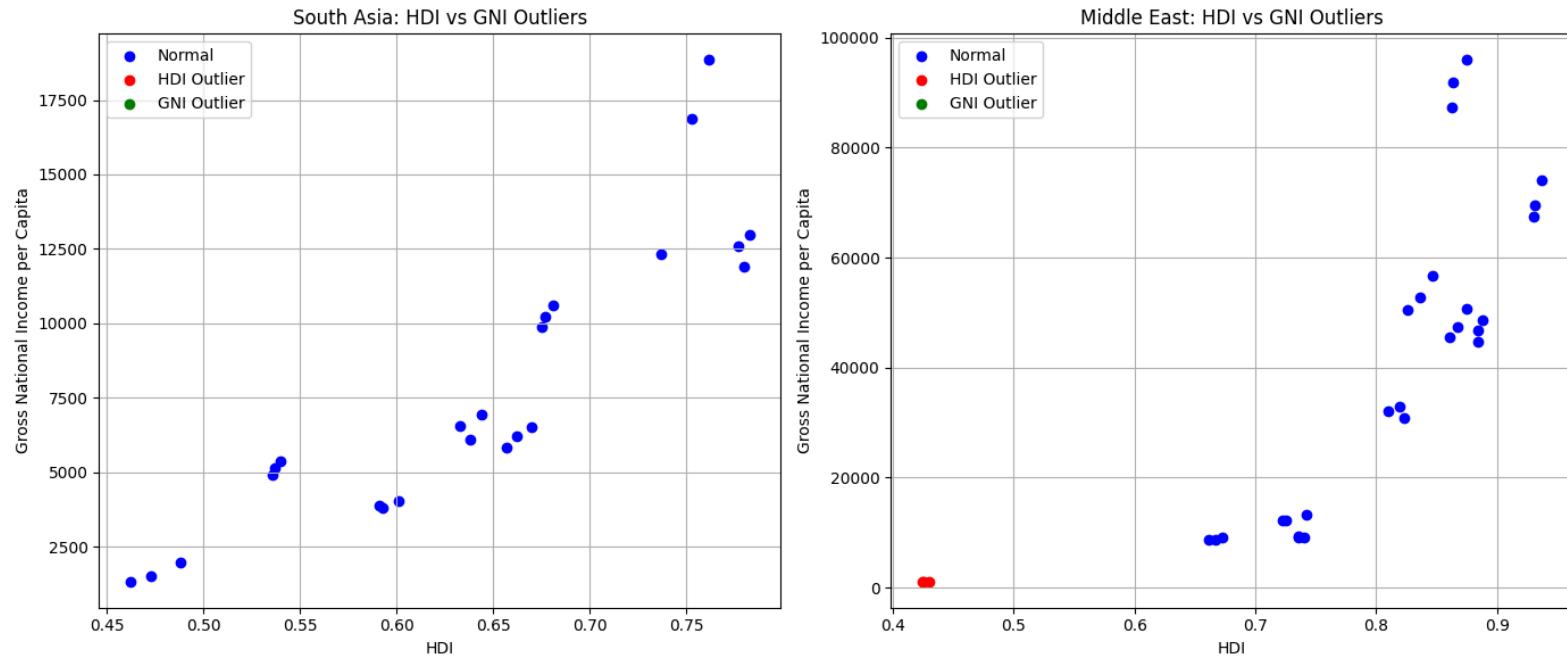


Figure 17: Scatter plots showing the relationship between Human Development Index (HDI) and Gross National Income (GNI) per capita for South Asia (left) and the Middle East (right)

Interpretation of Figure 17

- Outliers are data points that lie far from the main cluster and can significantly affect how a group appears overall. When included without consideration, they may distort comparisons and lead to inaccurate interpretations of a country's performance. For this reason, outliers are often better analyzed separately. In this case, several Middle Eastern countries exhibit noticeable outliers, whereas the South Asian countries show a more consistent pattern with no significant outliers.

8. Conclusion

This report provided a comprehensive exploration of Human Development Index (HDI) to analyze the global and regional development trends, especially on South Asia and Middle East. Using systematic exploratory data analysis, visualization and comparative methods, it analyzed the HDI trend, components and regional differences of development from 2020 to 2022.

The single year analysis in 2022 showed that there was a large inequality among countries. The human development level ranged from the top of Switzerland to the bottom of Somalia. The multi year trend in HDI revealed a continuous increase of HDI for most countries after the COVID 19. However, the pandemic caused an evident decline in health, education and income in some years.

The indepth exploration in South Asian countries indicated that HDI rises with economic growth, but not all countries can reach a high level of human development. The composite development score and gap analysis indicate that countries with a relatively large GNI per capita do not reach a proportionally high HDI. This implies that in addition to economic growth, more attention should be paid to the balanced development of health, education and social inclusion. The correlation analysis indicated that life expectancy and gender development are strongly correlated with HDI, especially in South Asia.

The comparative regional analysis showed that the MiddleEast outperforms South Asia on average HDI but also contains greater inequality and variability within the region. The presence of outliers in Middle Eastern countries suggests uneven development pattern caused by differences in income distribution and social policies.

Overall, this study reinforces the idea that human development is multidimensional and cannot be measured by economic indicators alone. Policymakers should focus on inclusive and sustainable development strategies that prioritize health, education, and equality alongside economic growth to achieve long term improvements in human well being.

9. Github Link

10. References and Citations

Title of web page	Web link + date accessed	Short reflection
Human Development Index (HDI)	https://hdr.undp.org/data-center/human-development-index#/indices/HDI I January 4, 2026	In the HDI framework, health is computed using life expectancy at birth. Education is assessed based on the average years of schooling completed by adults aged 25 and above, along with the expected years of schooling for children entering school. Standard of living is represented by gross national income (GNI) per capita, which is adjusted using a logarithmic scale. These three dimensions are combined using a geometric mean to calculate the final HDI value. The following data was used from HDI website and the above content was referenced from it
Human Development Index	https://www.who.int/data/nutrition/nlis/info/human-development-index January 7, 2026	The Human Development Index (HDI) evaluates a nation's development by integrating measures of health, education, and quality of life. The following data was used from World Health Organization (WHO) website and the above content was referenced from it

11. Appendix

Extra Datas and Graphs

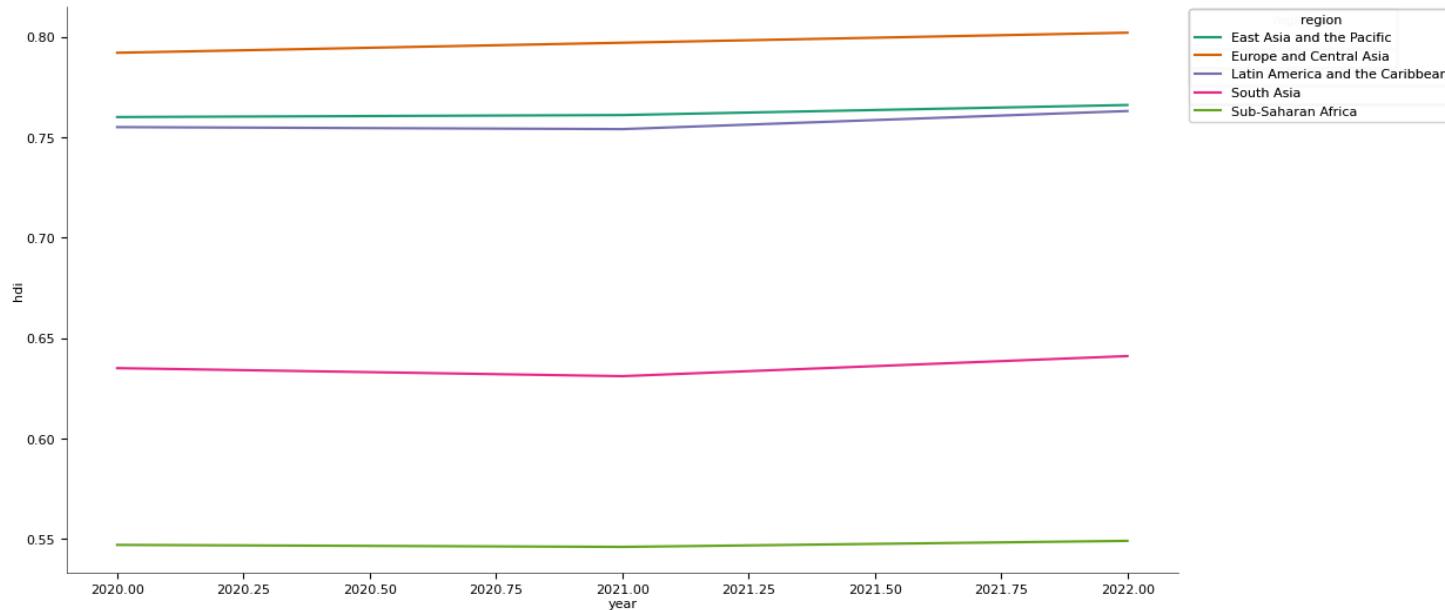
- Problem 1B 3. Visualisation Task

Selected Countries	Nepal
	India
	Bangladesh
	Bhutan
	Pakistan

- Composite Score Formula

$$\text{Composite Score} = 0.30 \times \text{Life Expectancy Index} + 0.30 \times \text{GNI per Capita Index}$$

- Time Series graph of regions of hdi and year



- Plot of South Asian Countries across time and hdi

