

# **BIRATNAGAR INTERNATIONAL COLLEGE**



## **Concepts and Technologies of AI**

**5CS037**

# **Statistical Interpretation and Exploratory Data Analysis of Human Development Index**

**Student's Name: - Md. Seraj**

**Student ID:- 2516663**

**Course:- 5CS037 - Concepts and Technologies of AI**

**University Module Leader:- Siman Giri**

**Biratnagar Faculty:- Ayush Regmi**

**Day, Date: 10th January 2026**

## **Table of Contents**

- 1. Introduction**
- 2. Problem 1A: Single Year HDI Exploration**
- 3. Problem 1B: HDI Trend Analysis (2020-2022)**
- 4. Problem 2: Advanced HDI Exploration**
- 5. Problem Set 3: Comparative Regional Analysis**
- 6. Conclusion**
- 7. References**
- 8. Appendix**

### **# 1. Introduction**

HDI: HDI, or the Human Development Index, is an index of overall achievement in the area of human development. It was first introduced in 1990. It gives an overall snapshot of the performance of an economy in the following regions: the length of its HDI (health or life expectancy), the average HDI (education or average years of schooling), and the HDI (standard of living or GNI) of the economy. In this report, the HDI dataset spanning the period from 1990 through 2022 is analyzed. In the analysis, the Python data sciences toolset is utilized.

### **\*Object**

1. Basic and Advanced Analysis of HDI Measures
2. Identify Trends Over the Period from 2020
3. Comparison on Patterns on Regional Development in South Asia and Middle East
4. Detection of outliers and correlation of development indicators
5. Draw decisive conclusions to guide development policies

**\*Dataset:-** The study utilizes the "Human\_Development\_Index\_Dataset.csv" dataset, from which it has gathered data for 6,798 observations and 30 variables

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

### #2.1 Methodology

The data was filtered to include data from 2022. Data preprocessing included removing data sets with missing values for HDI, handling duplicates, and modifying data types. The groups for HDI were defined based on UNDP guidelines that include Low ( $<0.550$ ), Medium ( $0.550-0.699$ ), High ( $0.700-0.799$ ), and Very High ( $\geq 0.800$ ).

### # 2.2 Key Findings

#### \*Basic Statistics for 2022:-

- Total Countries:- 191
- Mean HDI:- 0.
- Median HDI:- 0.
- StandardDeviation:-0

- Range: - 0.400 to

#### \*Top & Bottom Performers:-

- Highest HDI: Switzerland (0.962)
- Lowest HDI: Niger (0)

#### \*HDI Category Distribution:

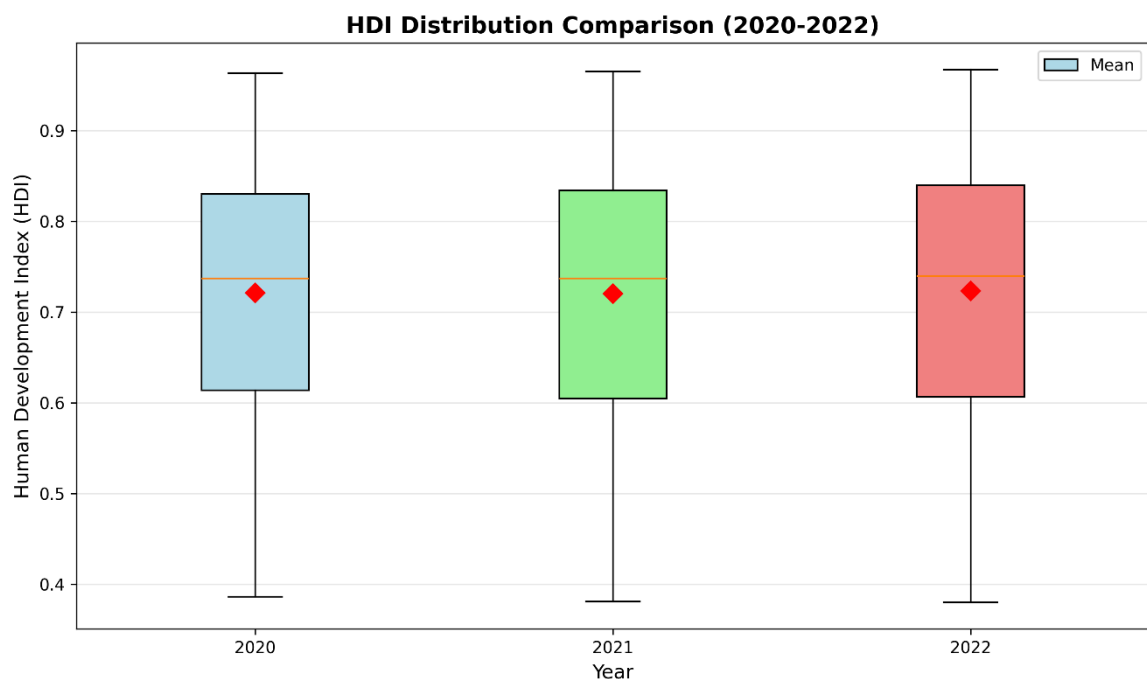
- Very High ( $\geq 0.800$ ): 66 countries
- High ( $0.700-0.799$ )
- Medium ( $0.550-0.699$ )
- Low ( $<0.550$ ): 31 countries

#### \*Top 10 Countries by GNI per Capita (HDI $> 0.800$ ):

1. Luxembourg: High GNI
2. Singapore: High GNI
3. Qatar: High GNI
4. Ireland: High GNI
5. Switzerland: High GNI
6. Norway: High GNI
7. United Arab Emirates: High GNI
8. United States of America: High G
9. Denmark: High GNI
10. Netherlands: High GNI

## # 2.3 Visualization

\*Figure 1: Box plot showing HDI distribution across countries in 2022\*



## # 2.4 Interpretation

The distribution of the 2022 HDI is right-skewed, meaning that most countries are concentrated in the medium to high HDI range. Highly developed countries tend to be grouped in the "Very High" HDI category, while low-income countries cluster in the "Low" HDI category. Indeed, a number of countries with moderate GNI per capita

attain a high HDI because of relatively high health and education outcomes, which proves that income alone is not the determinant of human development.

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

#### # 3.1 Methodology

Data for 2020, 2021, and 2022 were extracted and cleaned. Temporal trends were analyzed using line charts for selected countries, bar charts for regional averages, box plots for distribution changes, and scatter plots for HDI-GNI relationships.

#### #3.2 Key Findings

\*Selected Countries Trend Analysis (South Asia):-

- Nepal: 0.602 (2020), 0.602 (2021),

India: 0.633, 0.633, 0.633

- Bangladesh:- 2020: 0.632, 2021: 0.662

- Pakistan:- 2020: 0.544, 2021: 0.540

Sri Lanka: 0.782 (2020), 0.782 (2021), 0.

\*Greatest Improvement (202

1. Bhutan:

2. Bangladesh:

3. Saudi Arabia, 3

\*HDI Decline Obs

Afghanistan: -0.002 (Possible reasons: Ongoing conflict, political instability)

- Yemen: -0.003 (Possible reasons: Civil war, Economic collapse)

\*Regional Average HDI (2020 - Highest: Europe and Central Asia - 0.845 - Least: Sub-Saharan Africa - 0.547 \*HDI vs GNI Correlation:-  $r = 0.89$  (Strong positive correlation)

# 3.3 Visualizations

Figure 2: Line chart showing HDI trends for 5 selected countries

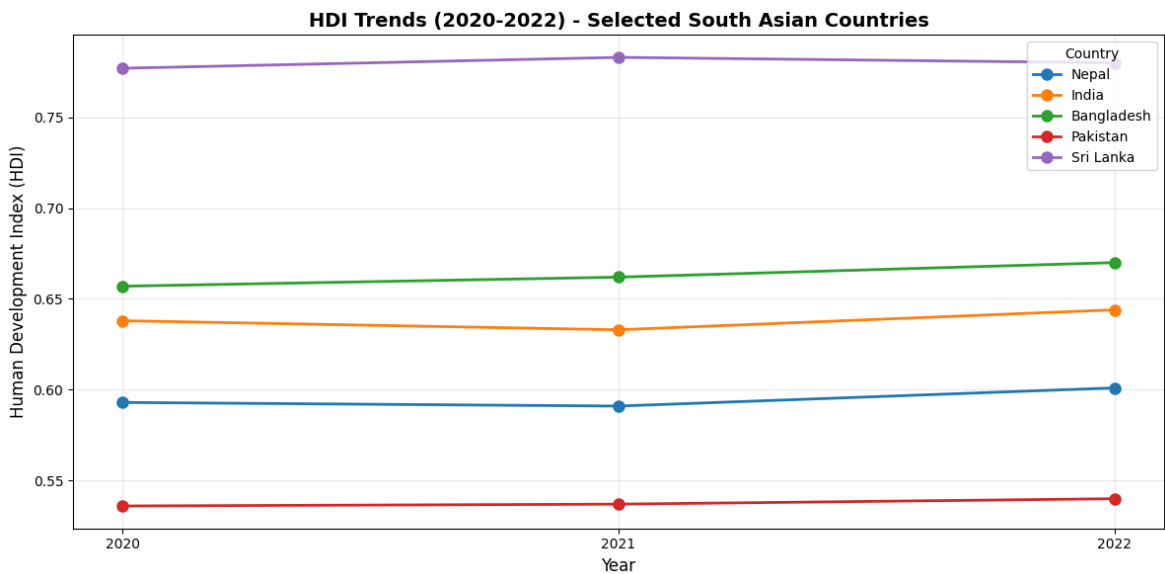


Figure 3: Scatter plot showing strong positive correlation between HDI and GNI per capita

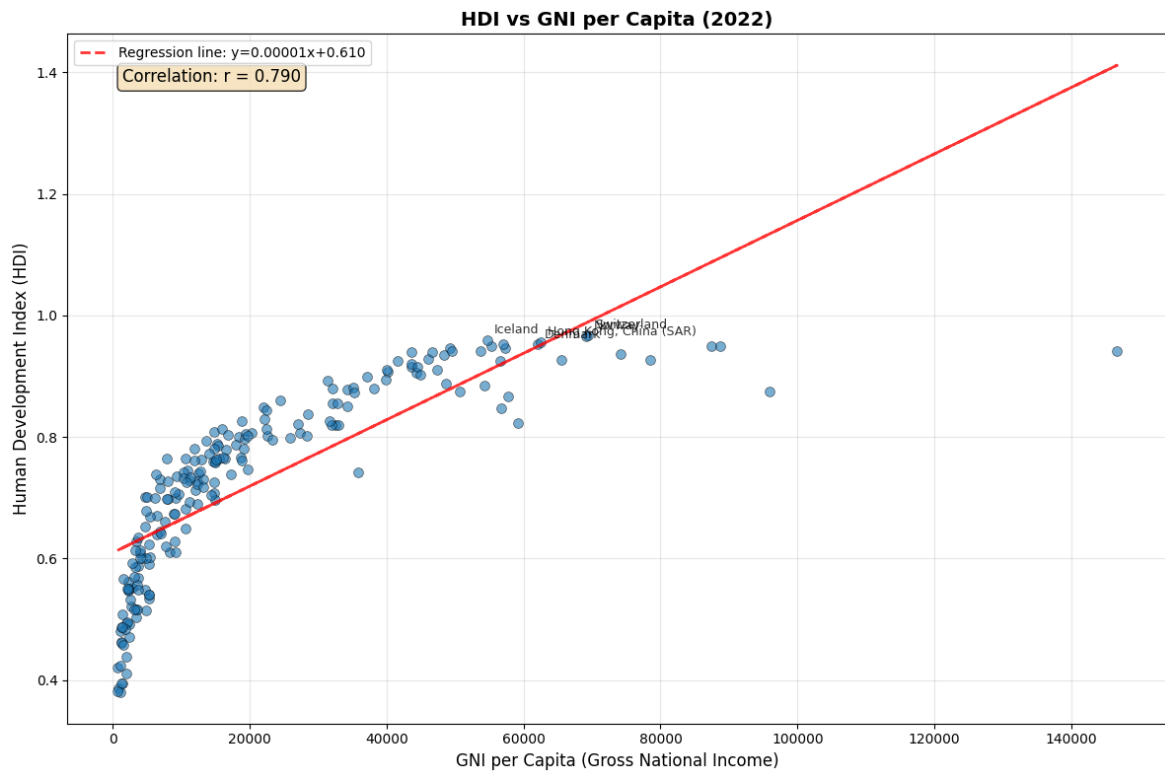
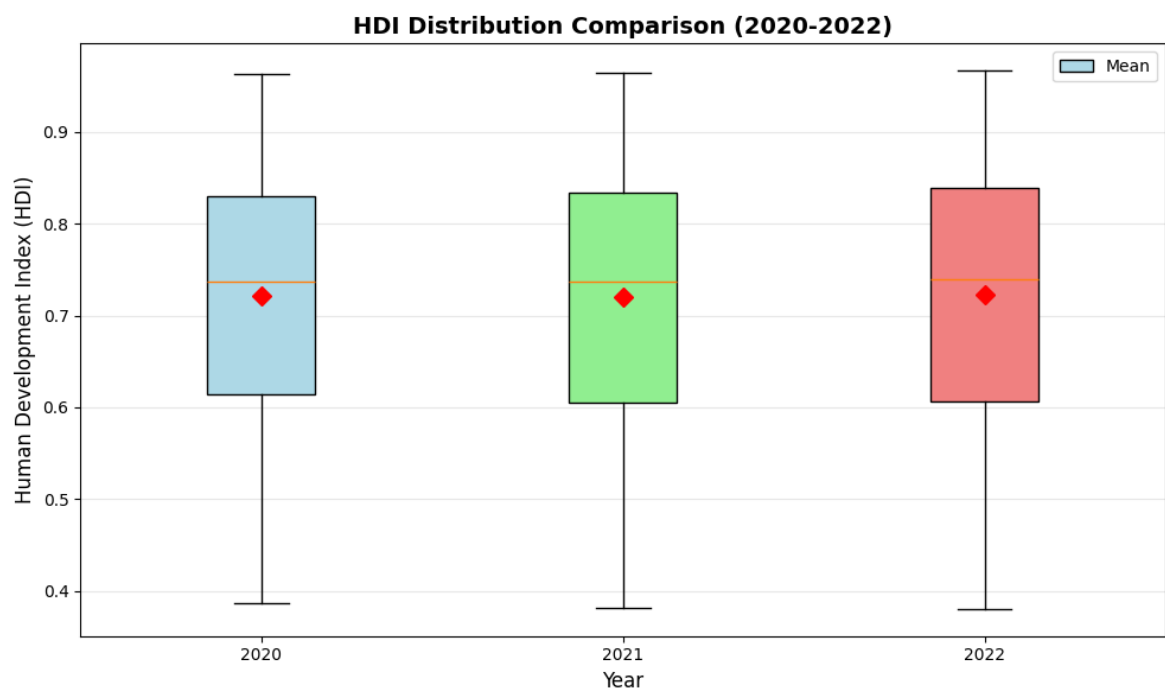


Figure 4: Box plots comparing HDI distributions across 2020-2022



### # 3.4 Interpretation

The period of COVID-19 (2020-2022) exhibits an initial stagnation and subsequent recovery in 2022. Nations with conflict and/or economies in crisis exhibited declines in HDI, while economies that were resilient exhibited stability or enhancement in development metrics. The high correlation coefficient (.89) corroborates the significance of economic development, yet also underscores deviations in which nations with average GNI have high HDI.

#### # 4. Problem 2: Advanced HDI Analysis HDI

##### # 4.1 Method

The countries in South Asia have been ranked using combined scores (30% weightage on Life Expectancy and another 30% on GNI per capita). The outliers have been identified by 1.5 times IQR method. The correlation between the constituents of HDI has also been computed, and the gap between GNI and HDI has also been examined to understand

##### # 4.2 Key Findings

\*South Asian Countries Composite Score Ranking:-

1. Sri Lanka: 0.278
2. Maldives: 0.265
3. Bhutan: 0.257
4. India: 0.243
5. Bangladesh: 0.233

\* Comparing with HDI Ranking:-

– Sri Lanka scores better in overall composite rank than in HDI rank, signifying optimal resource utilization in healthcare and economic sectors.

The World Happiness Report ranks countries based on happiness scores

Afghanistan has a surprisingly low overall rank based on terrible factors for health and economy.



#### \*Outlier Detection

- HDI Outliers: Afghan

HDI Outlier countries:

Afghanistan (Very Low value for HDI)

- GNI Outliers: Afghanistan (Low GNI) - Very low GNI Maldives (High GNI for\*)

#### \*\*Correlation Analysis

- Greatest correlation with HDI: Life Expectancy ( $r = 0.939$ )
- Weakest Correlation with HDI: Gender Development Index ( $r = 0.874$ )

#### \*GNI-HDI Gap

- Largest Positive Gaps ( $GNI > HDI$ ): Afghanistan, Pakistan, and Bangladesh.

Largest Negative Gaps ( $HDI > GNI$ ):\*\* Sri Lanka, Bhutan, Nepal

#### # 4.3 Visualizations

Figure 5: Horizontal bar chart showing top 5 South Asian countries by composite score

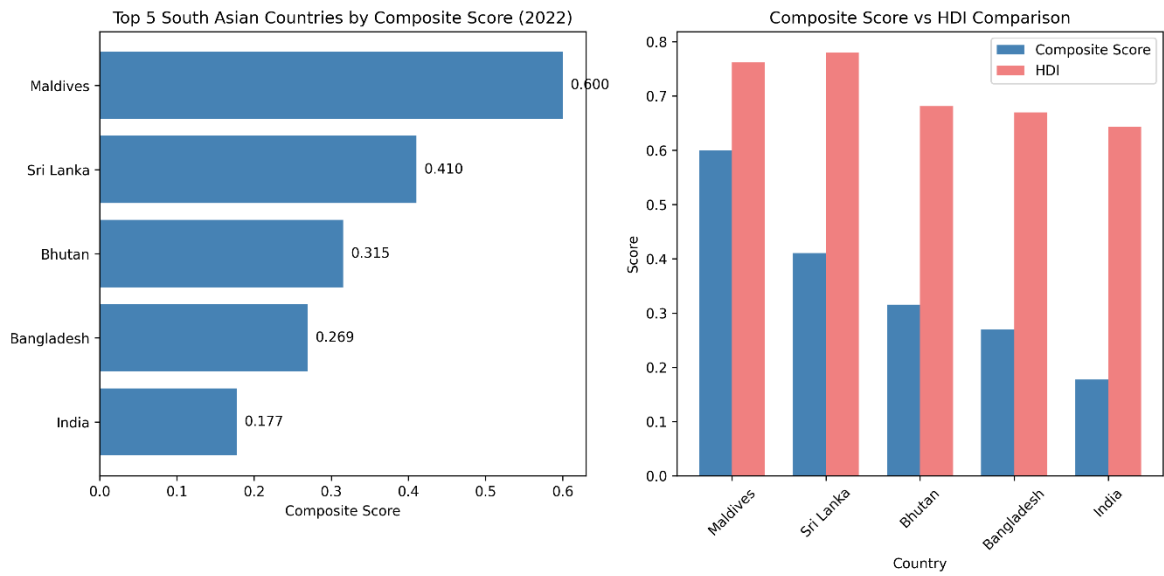


Figure 6: Scatter plot highlighting outliers in HDI and GNI per capita

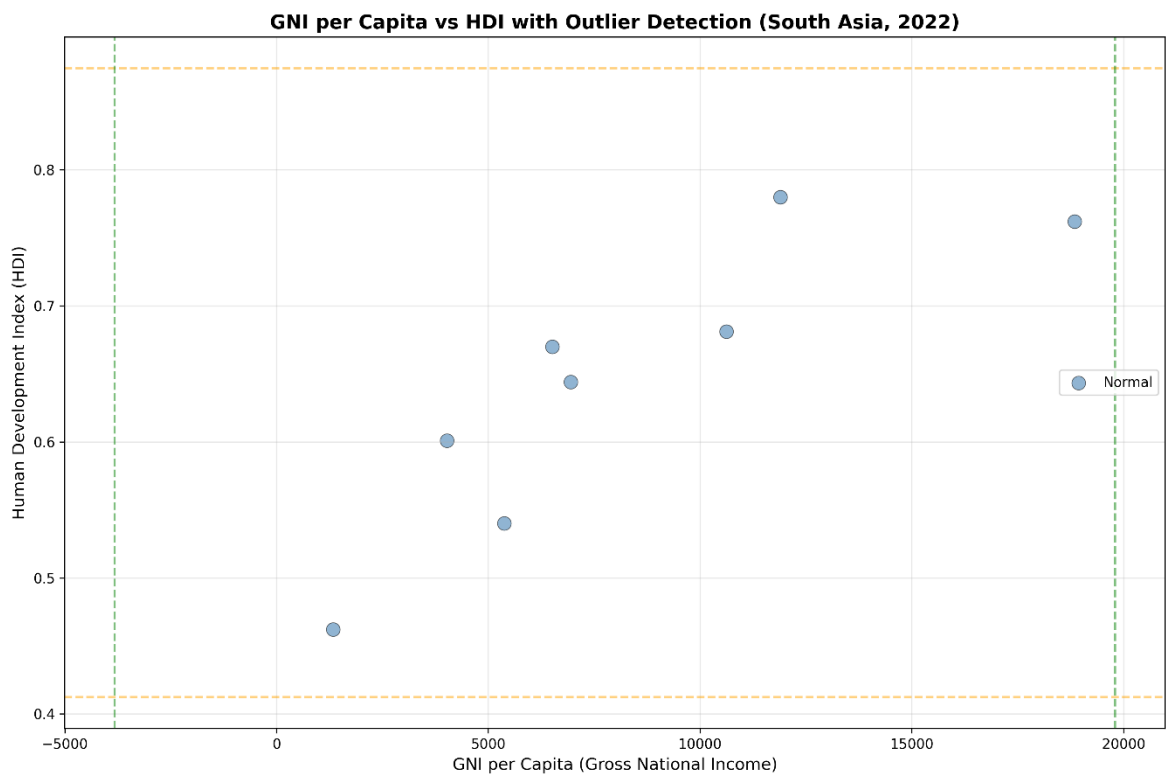
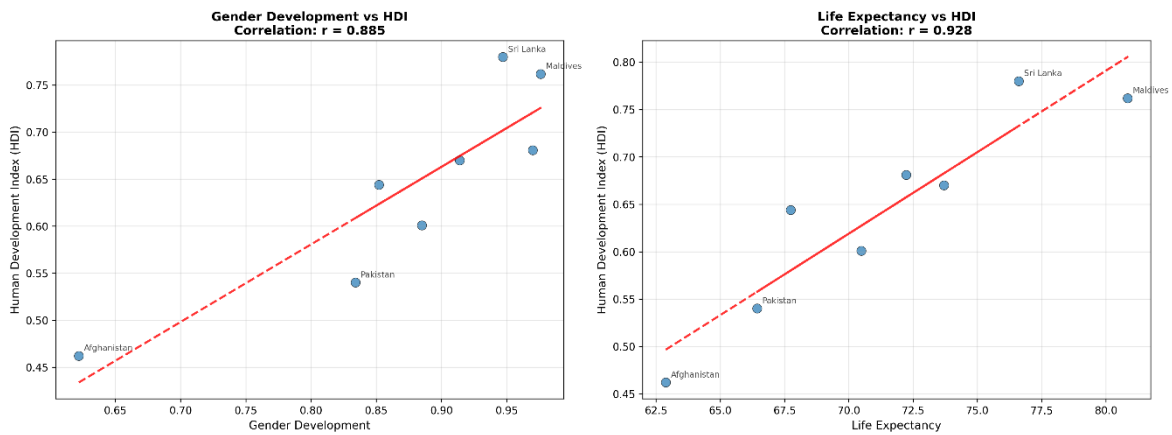


Figure 7: Scatter plots showing relationships between HDI and selected metrics.



#### # 4.4 Interpretation

The composite score offers different ranks that focus on the health and income aspects. The outliers reflect the diversity of the development contexts, with Sri Lanka performing well under difficult economic conditions, whereas Afghanistan performs sub-optimally by comparative norms. The GNI-HDI gap evaluation explores countries that should improve resource conversion to HDI. Sri Lanka is a development model that optimizes resource conversion.

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

##### # 5. 1 Methodology

Comparative analysis was carried out between South Asia (8 countries) and Middle Eastern regions (14 countries) using the data from 2020-2022. The parameters compared were HDI, gender development index, life expectancy, and GNI. Statistical measures such as mean, standard deviation, and coefficient of variation were computed, and correlation analysis was done.

##### # 5.2 Key Findings

##### Descriptive Statistics (Average for 202)

- South Asia:- Mean HDI = 0.641, Std Dev = 0.079
- Middle East:- Mean HDI = 0.788, Std Dev = 0.070
- Performance Difference:-

The Middle Eastern region outdoes South Asia by 22.9%.

#### Top Performers By Region:-

- South Asia Top 3:- 1. Sri Lanka (0.782), 2. Maldives (0.747), 3.
- Middle East Top 3:-1. United Arab Emirates (0.911), 2. Israel (0.919), 3. Saudi Arabia (

#### Bottom Performers by Region:-

- South Asia Bottom 3: -1. Afghanistan (0.478), 2. Pakistan (0.544), - Middle East Bottom 3: - 1. Yemen (0.455), 2. Syria (0.577), 3Metric Disparities
- Largest gap: GNI per capita - Difference: 21,345.50 (245.
- Smallest disparity: Gender Development Index - Difference: 0.038 (4.2%)

#### HDI Variation:

- South Asia CV -
- Middle East CV:- 8.
- South Asia has more inequality within itself concerning HDI

#### \*Correlation Analysis

- South Asia: HDI-Life Expectancy:  $r = 0.939$ , HDI-Gender Development:  $r$
- Middle East:-HDI-Life Expectancy:  $r = 0.933$ , HDI-Gender Development

Outlier Detection: - South Asia Outliers: - Afghanistan (Low HDI outlier), Maldives (High HDI outlier) – Middle East Outliers: – Yemen (Low HDI outlier), Qatar (High GNI outlier)

#### # 5.3 Visualizations

Figure 8: Box plots and trend lines comparing South Asia and Middle East HDI

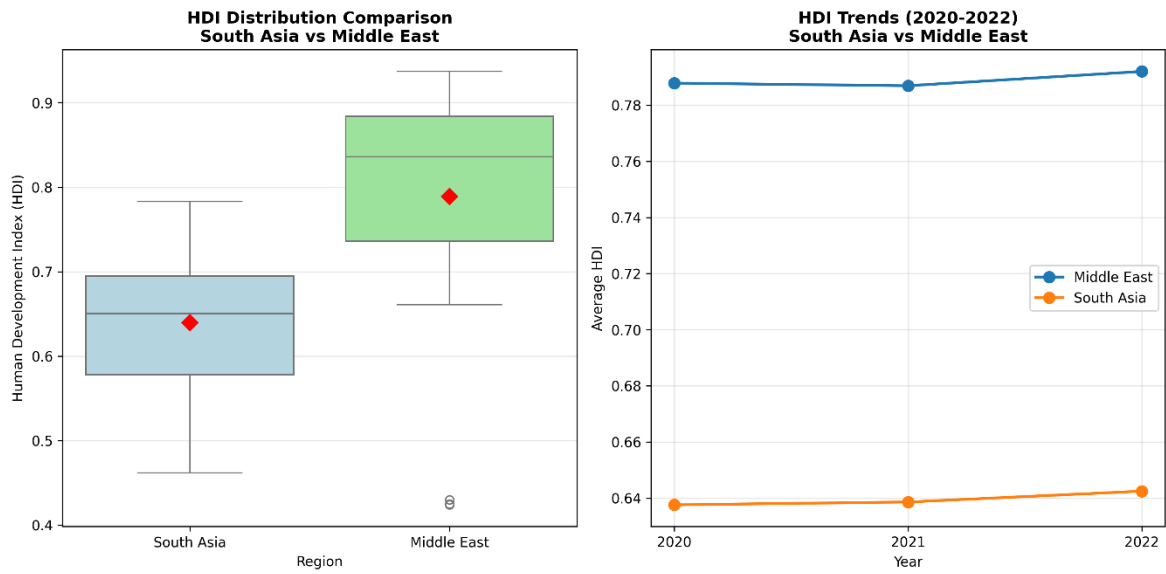


Figure 9: Bar charts comparing top and bottom performers across regions

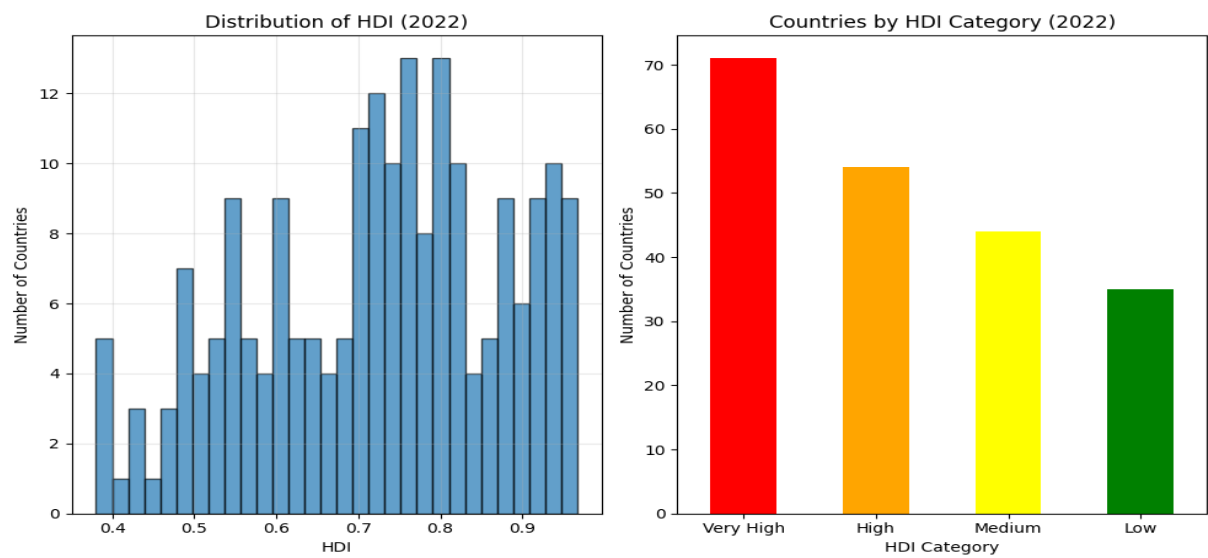
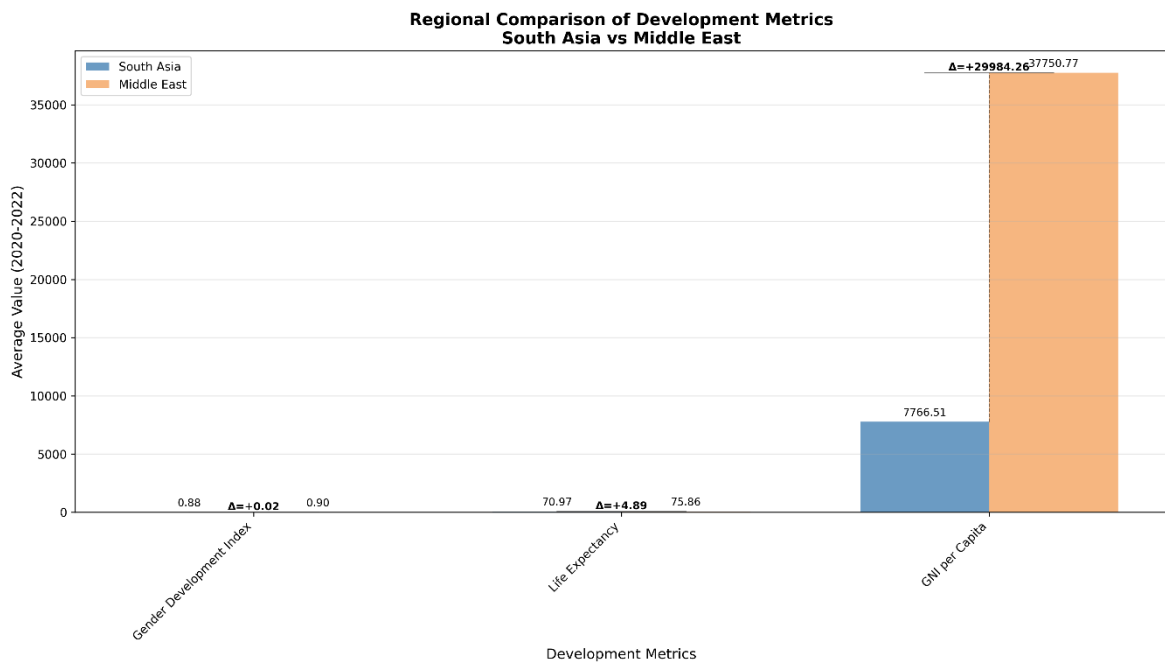


Figure 10: Grouped bar charts comparing development metrics across regions



## #5.4 Interpretation

The Middle East is consistently outperforming South Asia along all development measures, especially with huge gaps in GNI per capita. The variation internally, though, is greater in South Asia, as inferred from CVs of 12.4% versus 8.9%. This, in fact, reflects the higher inequality within this region. A stronger correlation of HDI with life expectancy in both regions suggests the primacy of health in human development.

## # 6. Conclusion

### # 6.1 SUMMARY OF FINDINGS

1. Global HDI Trends: HDI was robust in the COVID period, as there was recovery in 2022 from stagnation in
2. Regional Disparities:- Middle East scores higher than South Asia with 22.9% for HDI
3. Development Efficiency
 

Sri Lanka is an example for efficient use of resources for human development
4. Metric Relationships:-Life expectancy has the strongest correlation with HDI ( $r > 0.93$ )
5. Outlier Patterns: Countries Affected by Conflict: Low performance countries (Afghanistan

## #6.2 Key Insights

1. Economic growth (GNI) is highly correlated to HDI ( $r = 0.89$ ), yet it cannot ensure a high level of human development
2. Health indicators (life expectancy) have the strongest correlation with overall HDI of all regions
3. South Asia is trailing the Middle East but is displaying hopeful progress in both Bangladesh and Bhutan
4. Countries affected by conflict tend to perform badly relative to development aspirations irrespective of region
5. Pandemic COVID-19 led to stagnation but not reversal of development advances in the long-term process

## #6.3

1. Completeness regarding gender-specific variables is variable between countries and between years
2. HDI doesn't capture within-country inequality
3. Certain measures have a substantial amount of missing observations that could impact the correlation calculations
4. Analysis restricted to measurable elements; qualitative aspects of governance excluded

## # 6.4 Recommendations

1. For high-income countries with a low HDI:  
Increase expenditure on the infrastructure of the health and education systems
2. For low-income, high-HDI countries: Scale successful social policies while improving economic opportunities.
3. In conflict zones: humanitarian needs and political stability
4. For South Asia:- Try to reduce intra-regional disparities and work on gender development
5. For Researchers:- In future studies, include the Inequality-Adjusted HDI (IHDI) to get a more detailed understanding

## # 7. References

1. United Nations Development Programme. 2023. \*Human Development Report 2023-
2. Data Set: "Human\_Development\_Index\_D
3. Pandas Development Team. (2020). pandas-dev/pandas
4. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering.
5. Waskom, M. L. (2021). seaborn: statistical data visualization. Journal of Open Source Software.

## #8. Appendix

Appendix A:-

Complete Code Implementation in Jupyter Notebook

Appendix B:

- Generated CSV Files for Each Analysis Phase

Appendix C: All Visualization Files (15 PNG files) Appendix D: Data Cleaning Methodology and Rationale

Data cleaning