

PAPER NAME

2548200_ROJITDAHAL-2.docx

AUTHOR

-

WORD COUNT

1407 Words

CHARACTER COUNT

7674 Characters

PAGE COUNT

21 Pages

FILE SIZE

715.9KB

SUBMISSION DATE

Jan 9, 2026 9:01 PM GMT+5:45

REPORT DATE

Jan 9, 2026 9:02 PM GMT+5:45

● 11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 2% Internet database
- 1% Publications database
- Crossref database
- Crossref Posted Content database
- 11% Submitted Works database



Exploratory Data Analysis and Statistical Interpretation⁵ of the Human Development Index

Individual Assessment

Course: 5CS037 – Concepts and Technologies of AI

Name: Rojit Dahal

ID: 2548200

Date: Jan 10th, 2026

Table Of Topics

Table Of Topics

Introduction.....	1
Understanding HDI	1
Aims of the report	1
Scope of the report	1
Problem Analysis (One by One).....	1
Problem I(a)– Single Year HDI Exploration	2
Method Used	2
Results	2
Interpretation.....	2
Problem I(b) – HDI Trend Analysis (2020-2022)	3
Methods.....	3
A sample of the data between the years 2020, 2021, and 2022 was taken out. This dataset was cleaned and validated and then used to plot the temporal trends and distributions.....	3
Results	3
Visualizations	3
Interpretation.....	5
Problem II - Developed HDI Exploration (South Asia)	6
Methods.....	6
Results	6
Visualizations	7
.....	9
.....	9
Interpretation.....	10
Problem III - ¹ South Asia vs Middle East Comparative Regional Analysis.....	11
Methods.....	11

Results/Comparisons	11
Visualizations	11
.....	16
Interpretation.....	16
Conclusion.....	17
References	18

Introduction

Understanding HDI

HDI consists of ² health (life expectancy), education (mean years of schooling and expected years of schooling) and standard of living on a scale between 0 and 1. It is the index used to classify nations into four levels.

The goal ¹⁰ of this analysis is to complete a statistical interpretation/EDA assignment on HDI data. Pandas, NumPy, Matplotlib and Seaborn are among the libraries used for this analysis. This involves examining this dataset by annual trends, longer-term changes, regional variation and other complex analysis, including analysis of overall scores, outliers and gaps.

Aims of the report

To investigate global HDI data as of the year 2022, to estimate trends of 2020-2022, outliers, and conduct a region-wise study.

Scope of the report

Data exploration, trends analysis, and regional comparisons of ⁸ South Asia and the Middle East in terms of Human_Development_Index_Dataset.csv have been discussed in this report.

Problem Analysis (One by One)

Problem I(a)– Single Year HDI Exploration

Method Used

The initial step involved analyzing the dataset in order to extract all the available years. The most recent year was 2022, which was obtained and explored as a single year.

The steps involved in data cleaning were:

- Changing numeric columns to suitable data types
- Dropping duplicate records
- Dealing with missing values Deleting rows that contain a missing critical field like country and HDI.

Results

The descriptive statistics (mean, median and standard deviation) have been computed to get values of HDI in 2022. The highest¹¹ and the lowest nations in terms of HDI were identified to show the ends of the development count in the world. The discovery of economically strong countries with high human development outcomes was due to the use of a filter and sorting by¹² Gross National Income (GNI) per capita of countries whose HDI scores were above 0.800.

Interpretation

All the countries were divided into four official categories of the UNDP HDI: Low, Medium, High, and Very High. Such classifications gave a more qualitative description of development levels, and this was stored as a processed CSV file to be used in the future.

Problem I(b) – ¹HDI Trend Analysis (2020-2022)

Methods

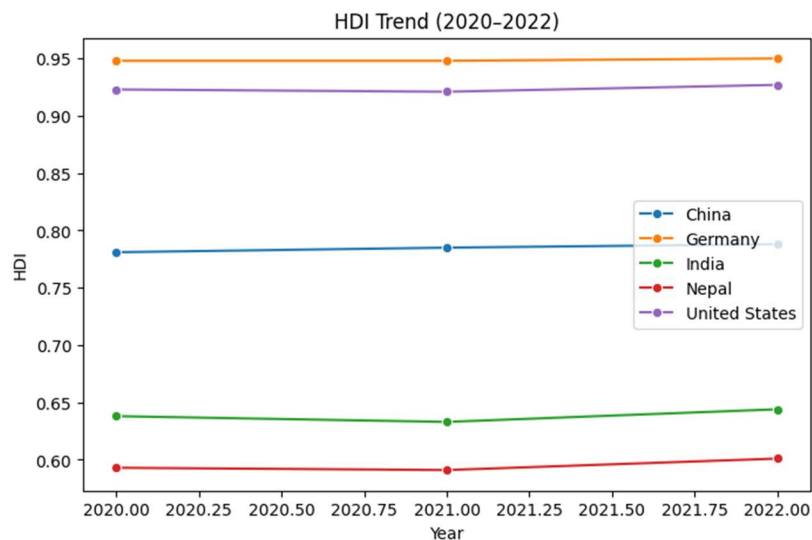
A sample of the data between the years 2020, 2021, and 2022 was taken out. This dataset was cleaned and validated and then used to plot the temporal trends and distributions.

Results

- Compared to 2020, HDI values do not change significantly, but their difference is minimal between 2020 and 2022.
- It is noticeable that there is a slight slowdown or a dip in the period around 2020-2021, which is probably caused by the global disruptions.
- The median HDI has shown little variation which indicates ³that there is a general stability.
- Affirmative correlation between HDI and GNI per capita.
- The HDI difference is not entirely explained by income alone and this leads to the contribution of the health and education factors.

Visualizations

Line Chart: Trends of HDI of five countries selected revealed slow alteration in time indicating recovery and adaptation to the world shocks.



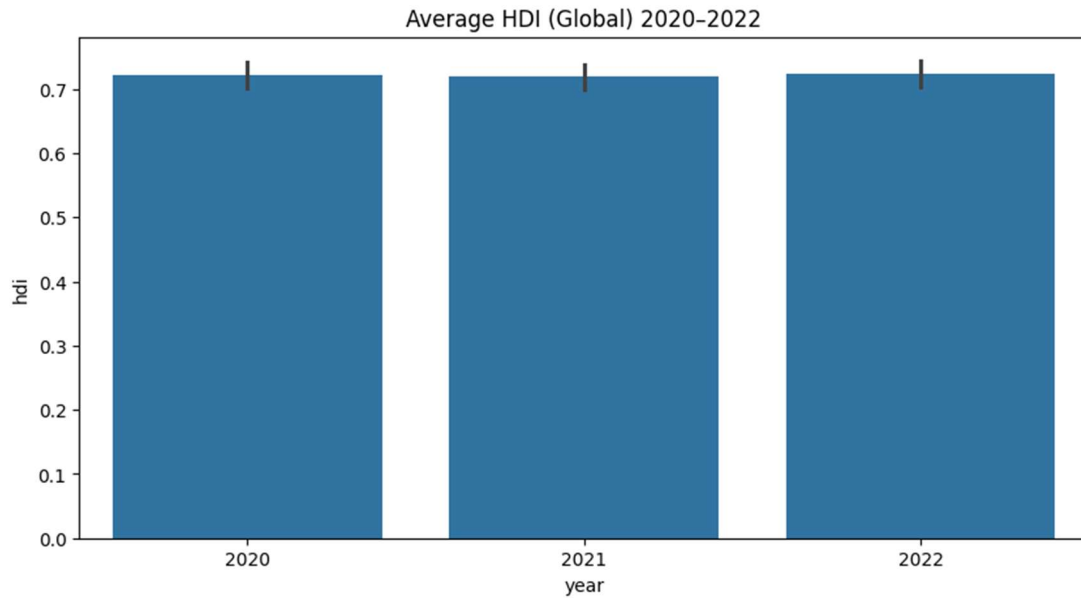
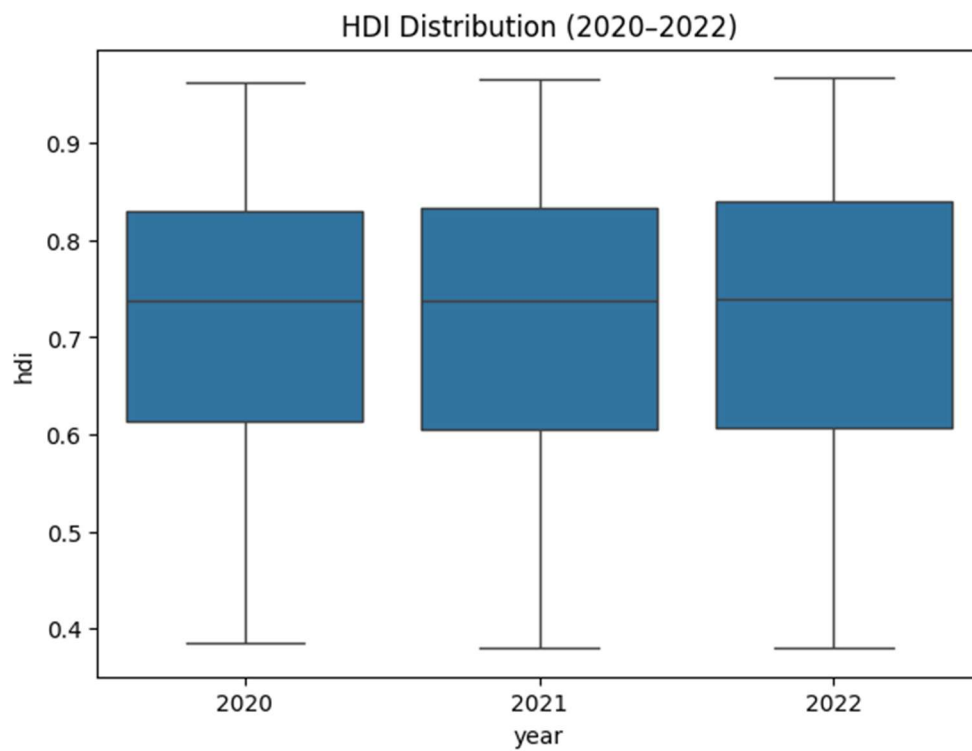
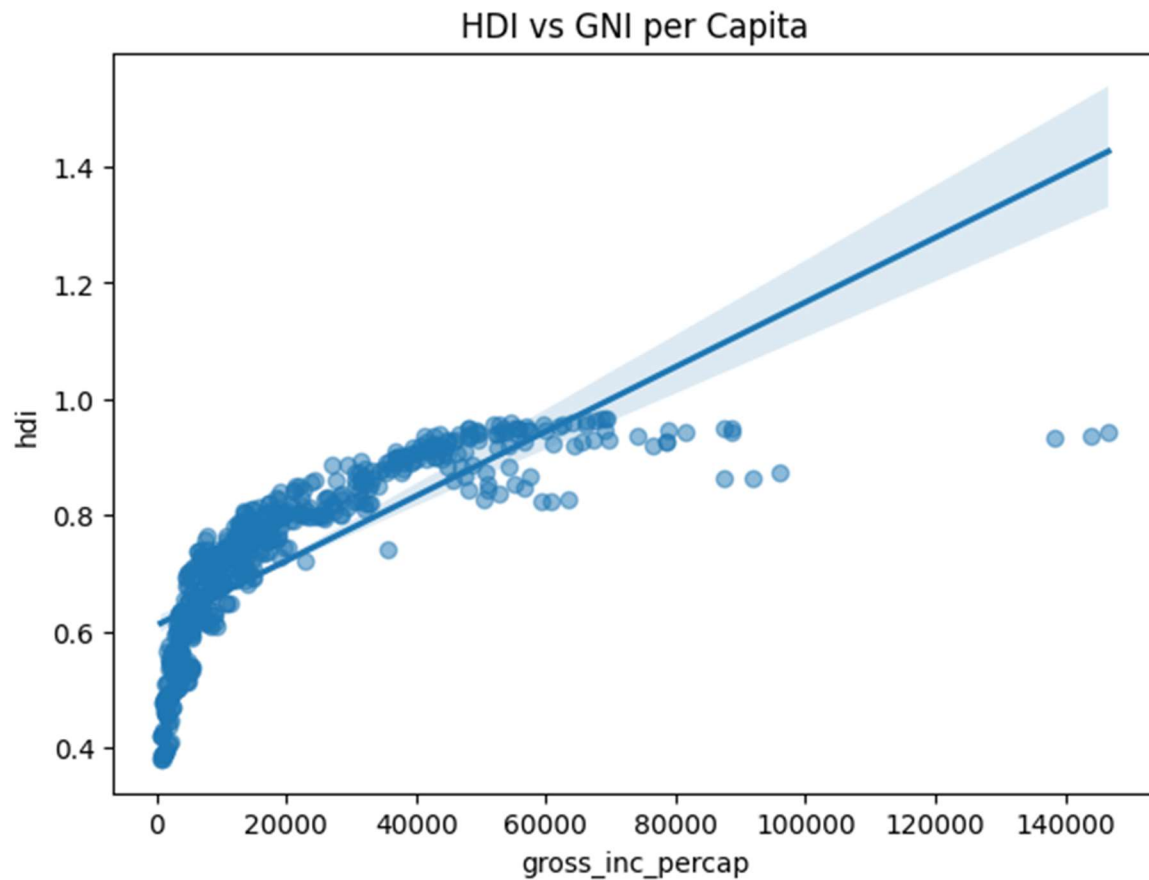


Fig: Bar Diagram (Average HDI)

Box Plot: Medians of the HDI distribution in the three years were relatively stable with slight fluctuations which demonstrated that there is not much volatility in the short term.



Scatter Plot (HDI vs GNI): The relationship between income and HDI was found to be positive, however not linear, indicating the importance of the non-income determinant, i.e. health and education.



Interpretation

Based on the analysis, it is indicated that income has a very strong impact on HDI, yet it does not entirely define the outcomes of human development. The COVID-19 pandemic is also a global event that might have slowed progress or provoked temporary negative results in certain countries.

Problem II - Developed HDI Exploration (South Asia)

Methods

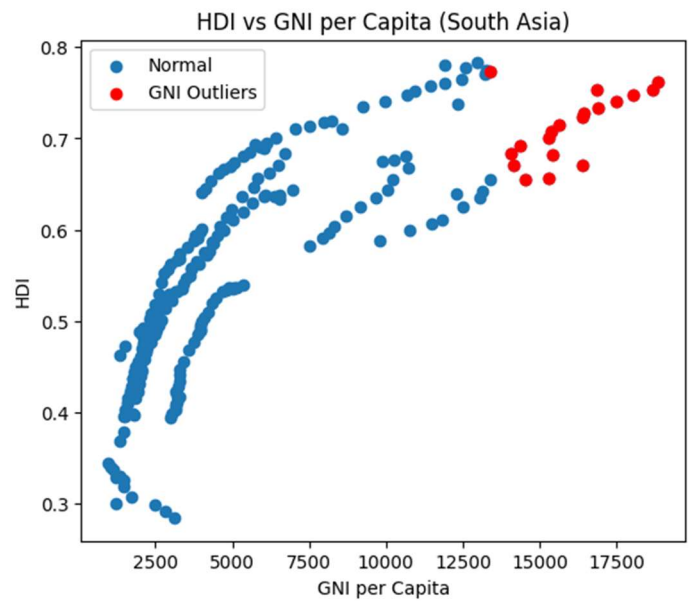
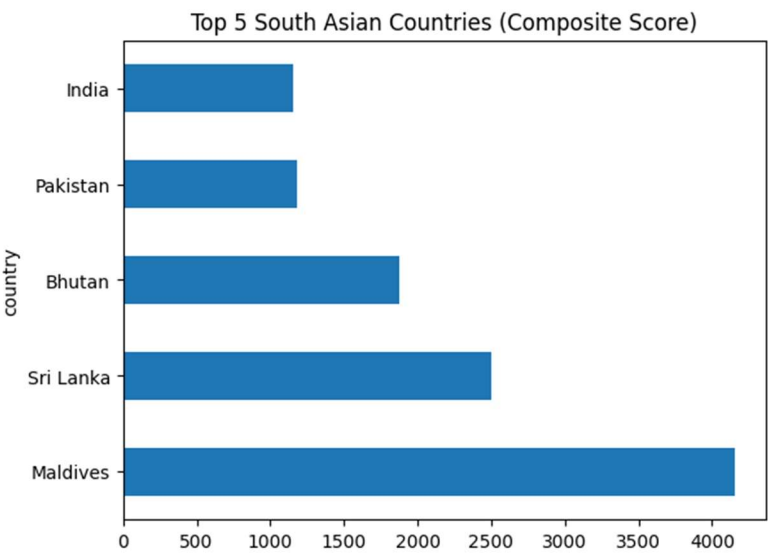
There was a formation of a regional sub-group of countries in South Asia. Other measures were also provided to allow further analysis, such as:

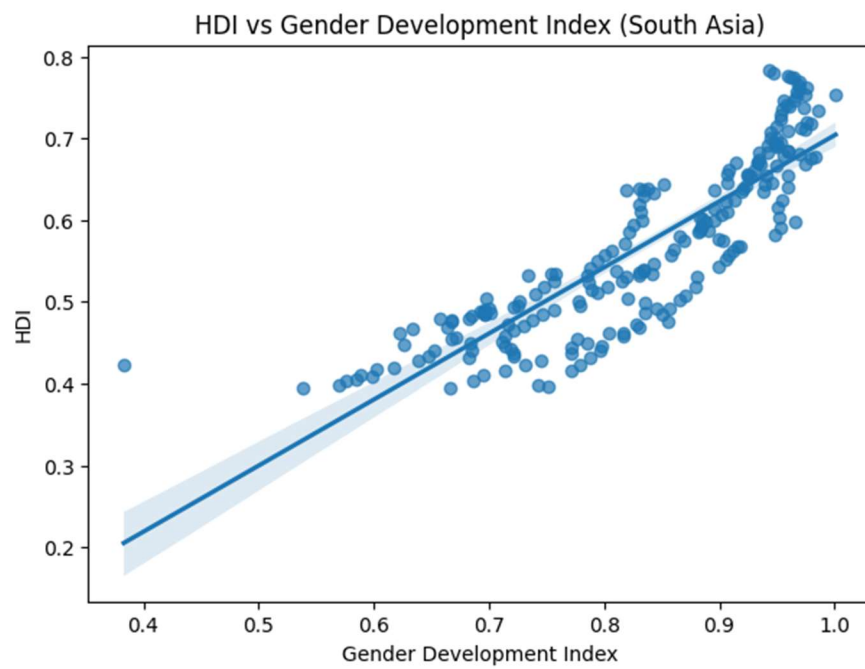
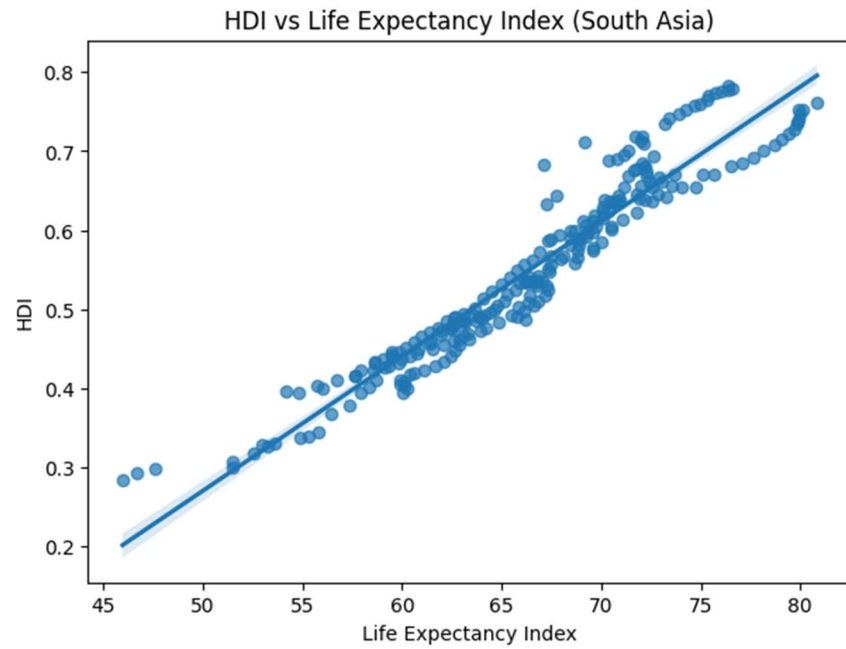
- Composite Development Score
- GNI-HDI Gap
- The 1.5xIQR rule of outlier detection.
- Composite Score Analysis
- Outlier Detection

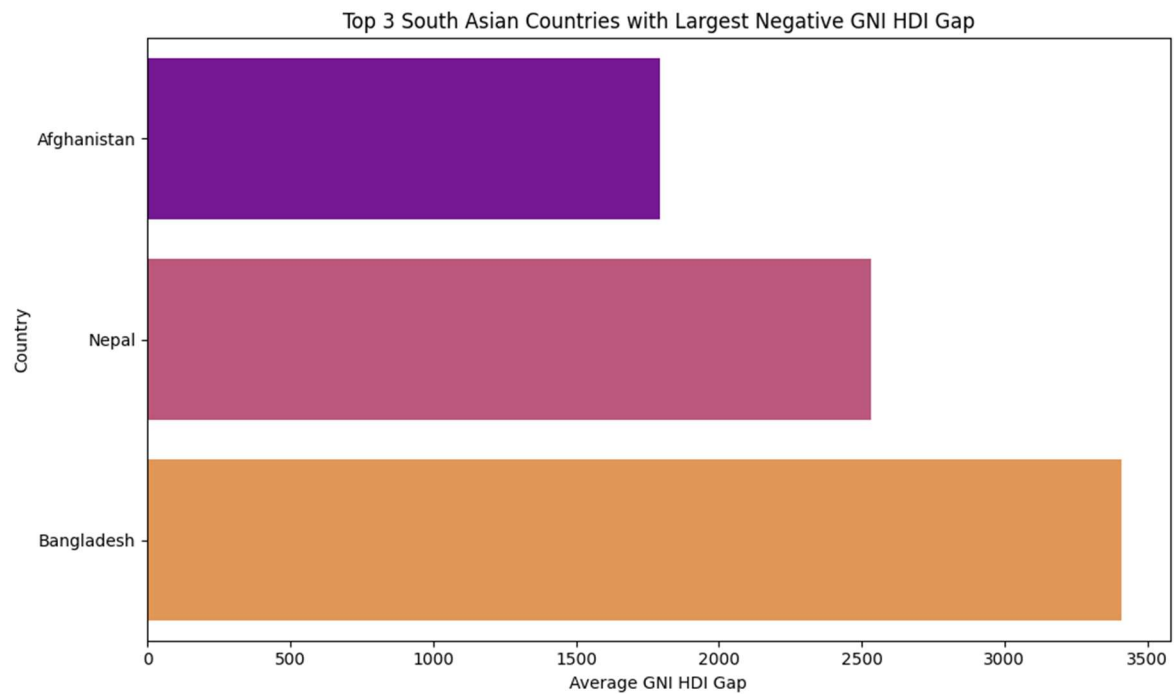
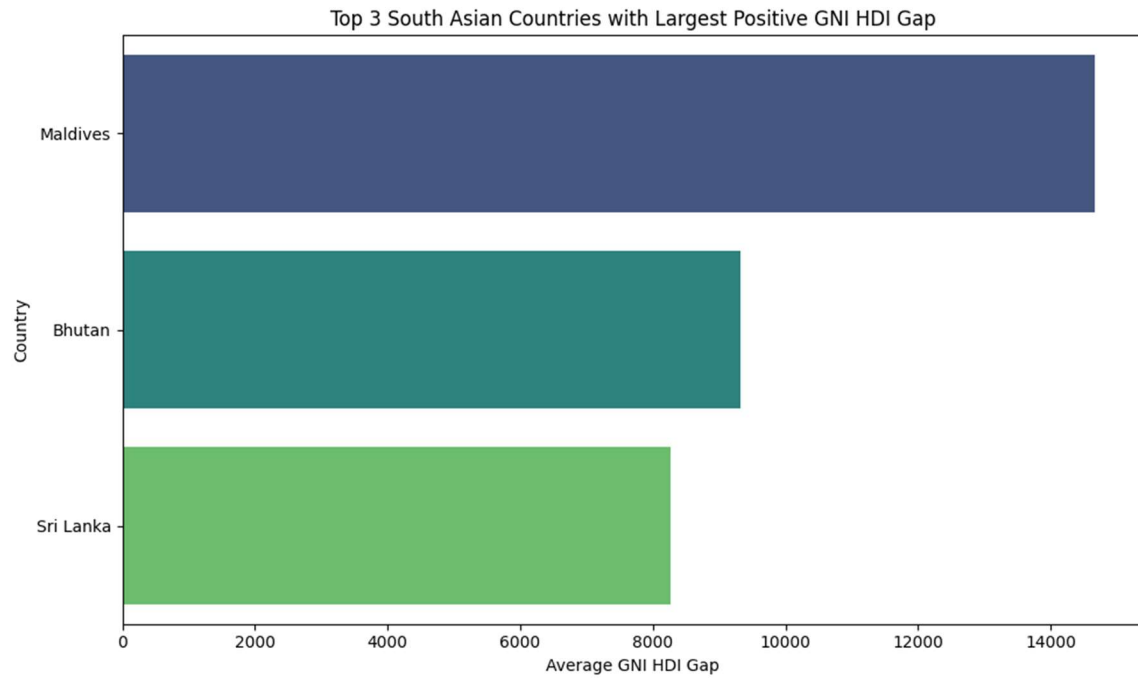
Results

- The differences in HDI, life expectancy and income levels are evident in South Asian countries.
- The Composite Score ranking is contrasted with HDI ranking indicating that income and health do not entirely define HDI.
- There are no outliers in HDI calculated with the IQR method, but GNI per capita has obvious outliers, meaning that there is income inequality.
- There is a GNI-HDI Gap in different countries, and it appears that the higher the income, the comparatively higher the human development.
- The HDI is more correlated with life expectancy than the gender development.

Visualizations







Interpretation

These findings imply that economic growth is not the sole factor that is relevant but health and social factors are important in enhancing human development in South Asia. There are also income inequalities in the region and transferring economic benefits to overall human development is a major issue.

Composite Score Analysis: A Composite Score, a composite measure of life expectancy and GNI per capita to rank the South Asian countries. This ranking was slightly different in the traditional HDI rankings, which means that income and health are not the only indicators of successful development of a country.

Outlier Detection: The IQR technique values showed that there were no significant outliers in the values of HDI since they were clustered. However, there are clear outliers in GNI per capita which means that income distribution in the region was unequal.

Gap Analysis: The GNI-HDI Gap illuminated the instances when the comparatively big incomes did not reflect similarly high HDI scores. This implies ineffectiveness of conversion of economic resources into the general human development.

Problem III - ⁶South Asia vs Middle East Comparative Regional Analysis.

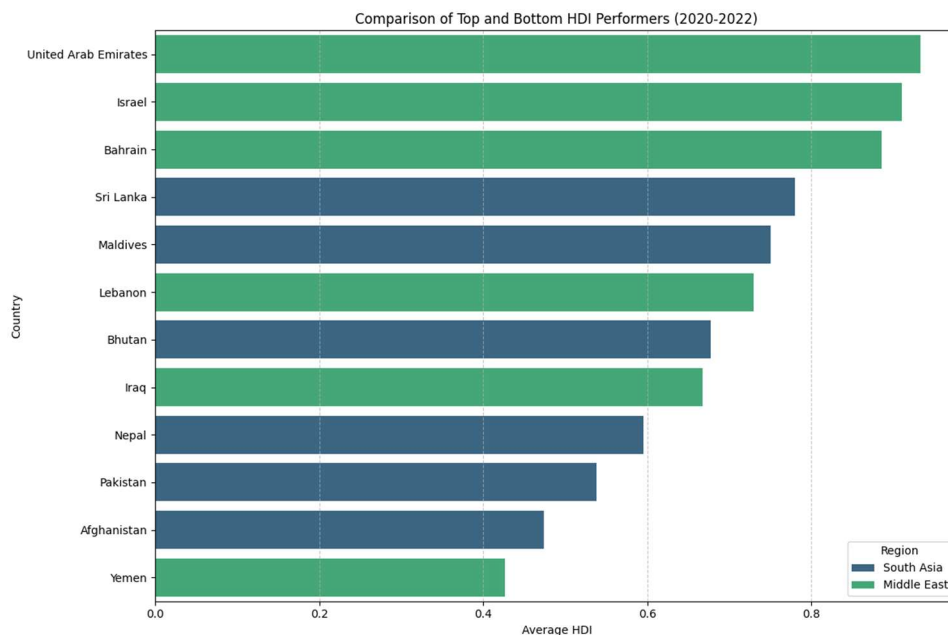
Methods

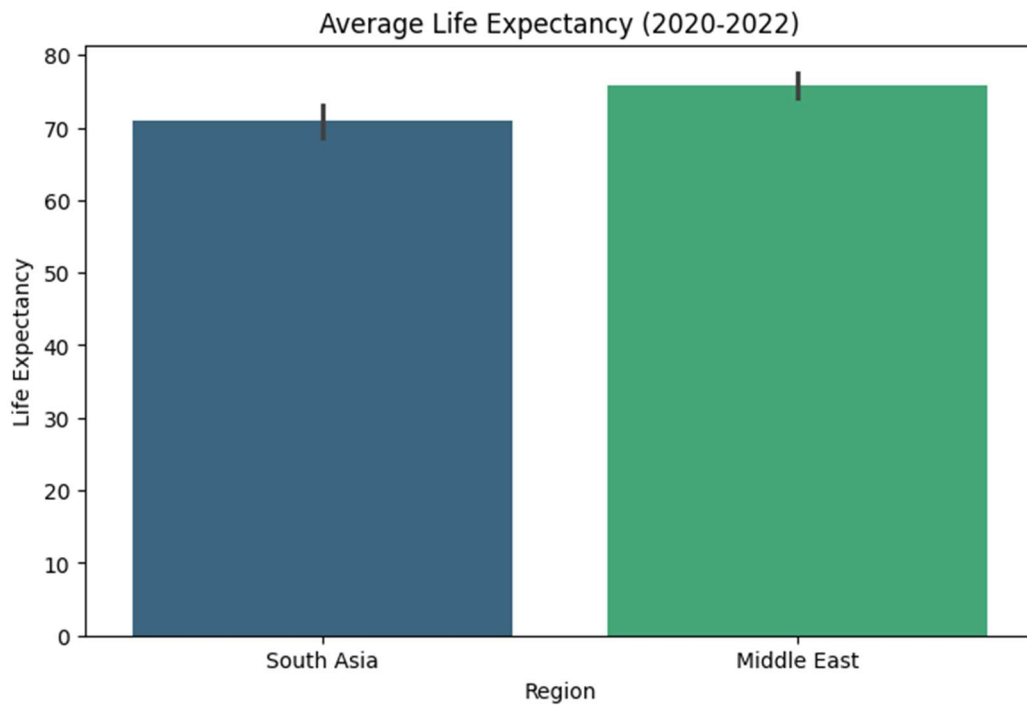
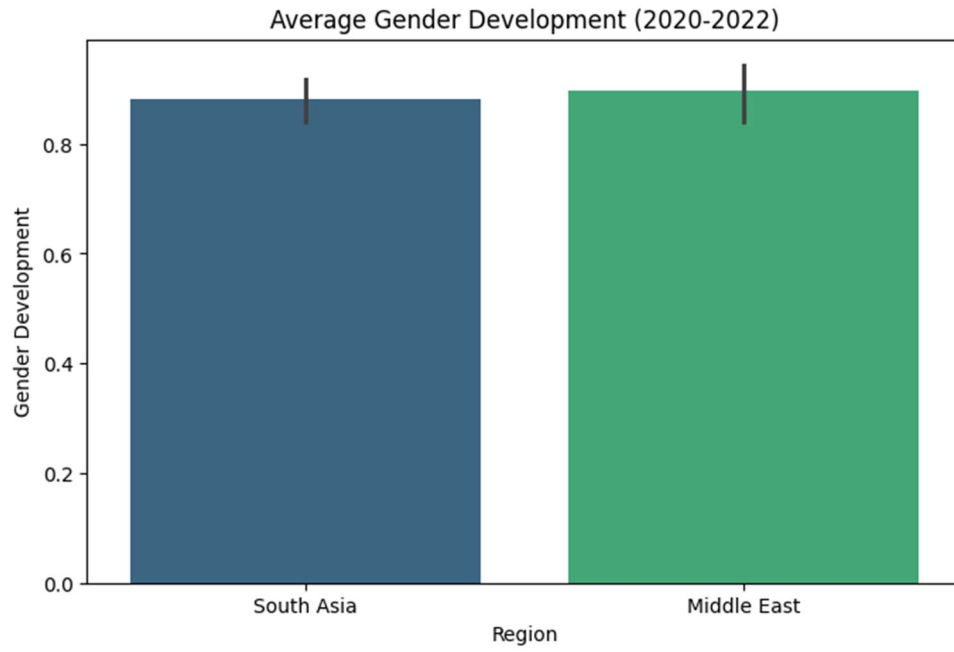
The comparison of South Asia and Middle East subsets of 2020-2022 was done using descriptive statistics, measures of variation, and correlation analysis.

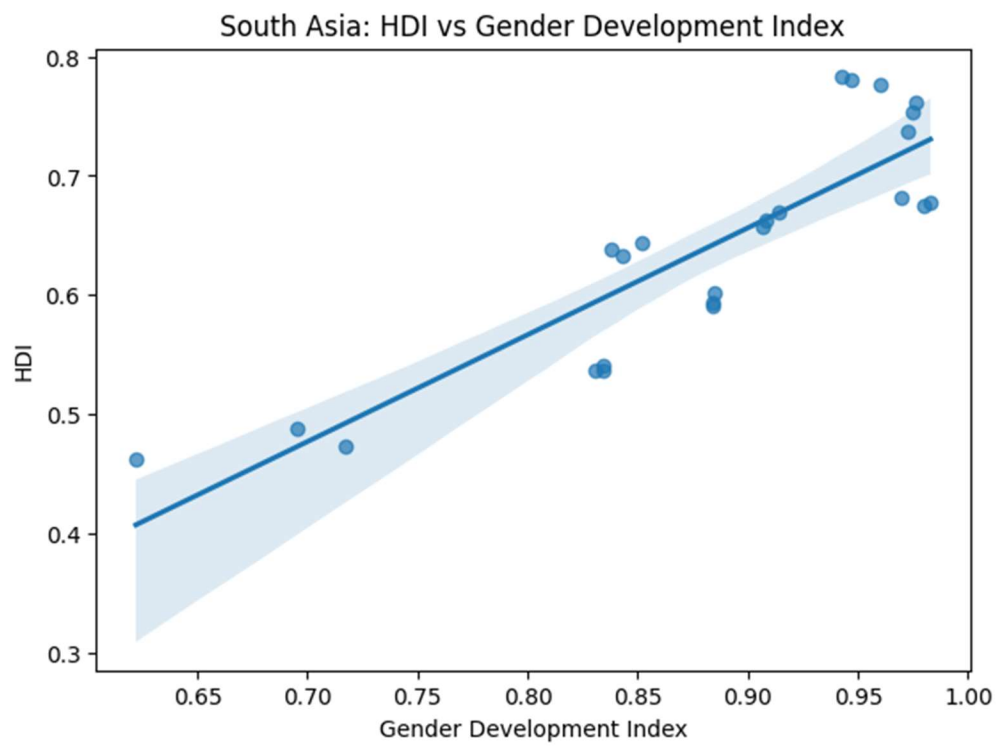
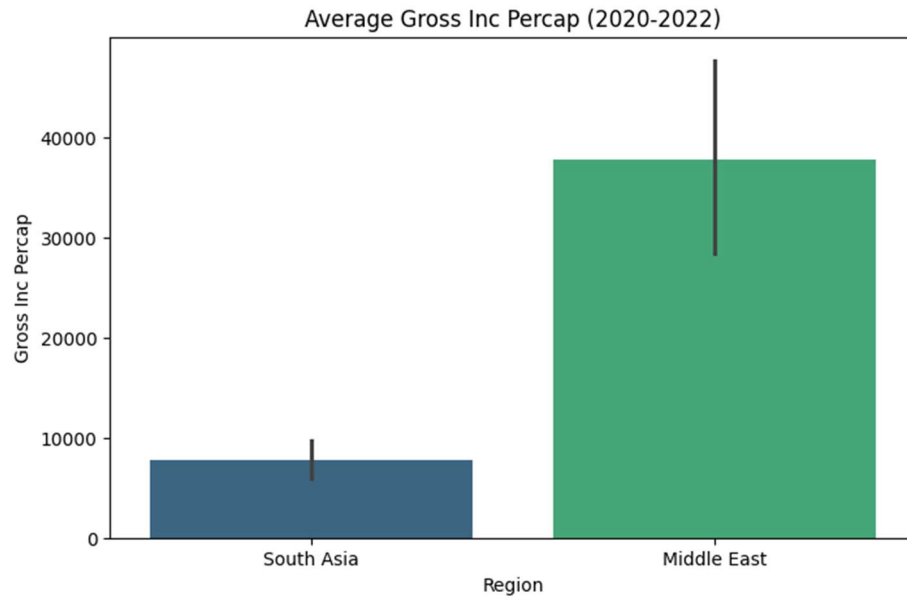
Results/Comparisons

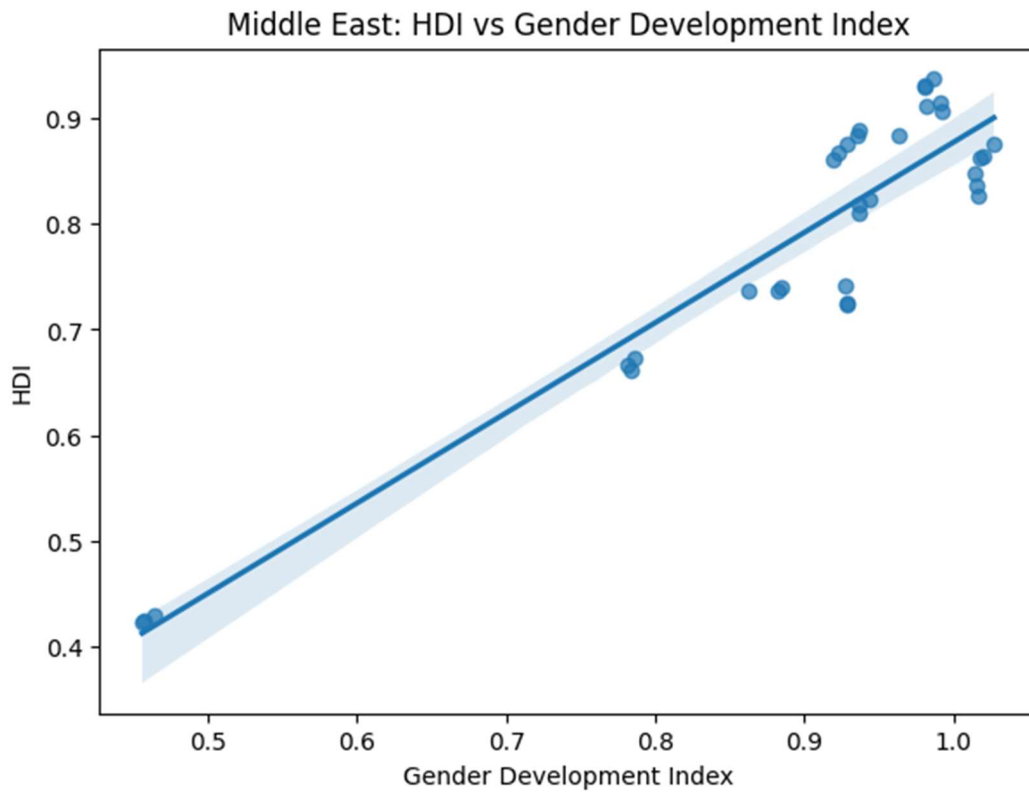
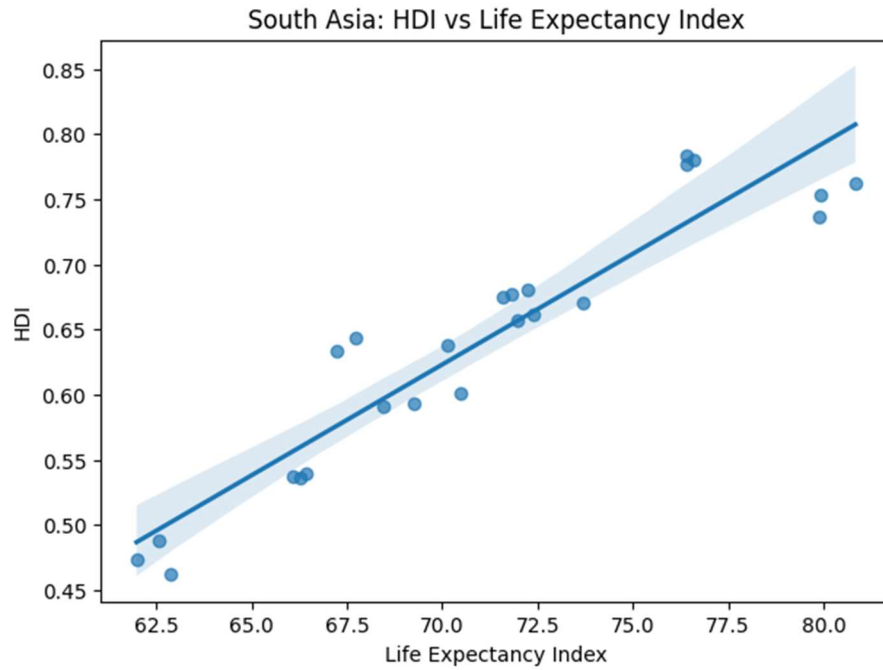
- Mean HDI level in most of Middle Eastern countries were higher than that in ⁷South Asia.
- The South Asia region was more heterogeneous in HDI, ³range and coefficient of variation.
- Correlation analysis indicated ⁴that there has a strong positive correlation between life expectancy and HDI in the two areas.

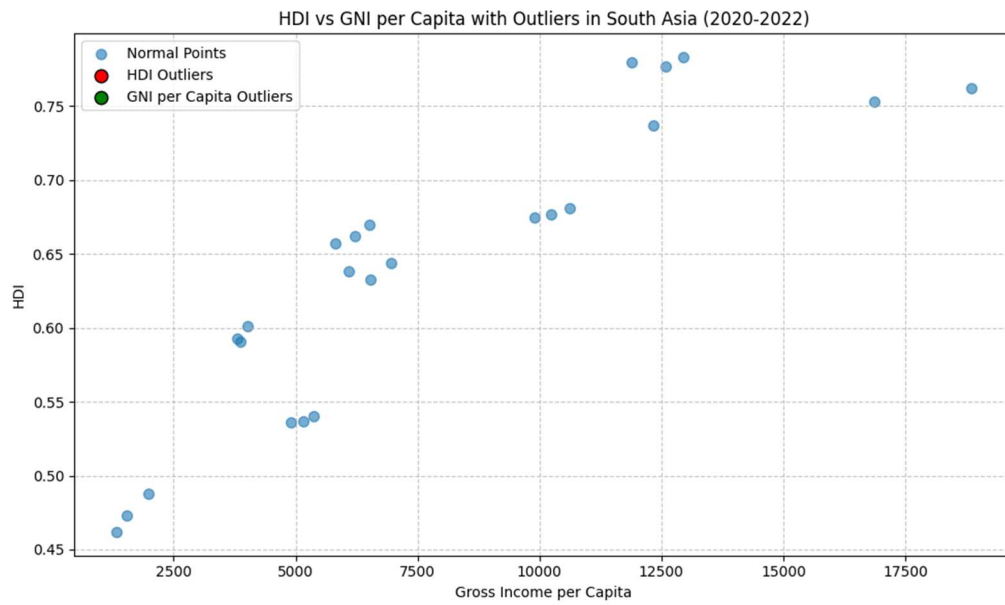
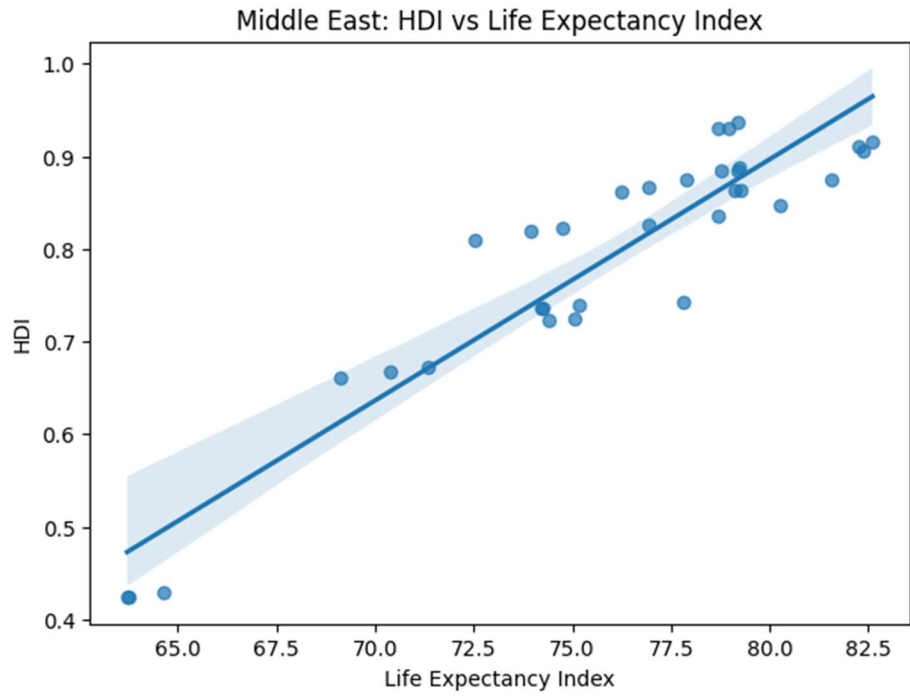
Visualizations

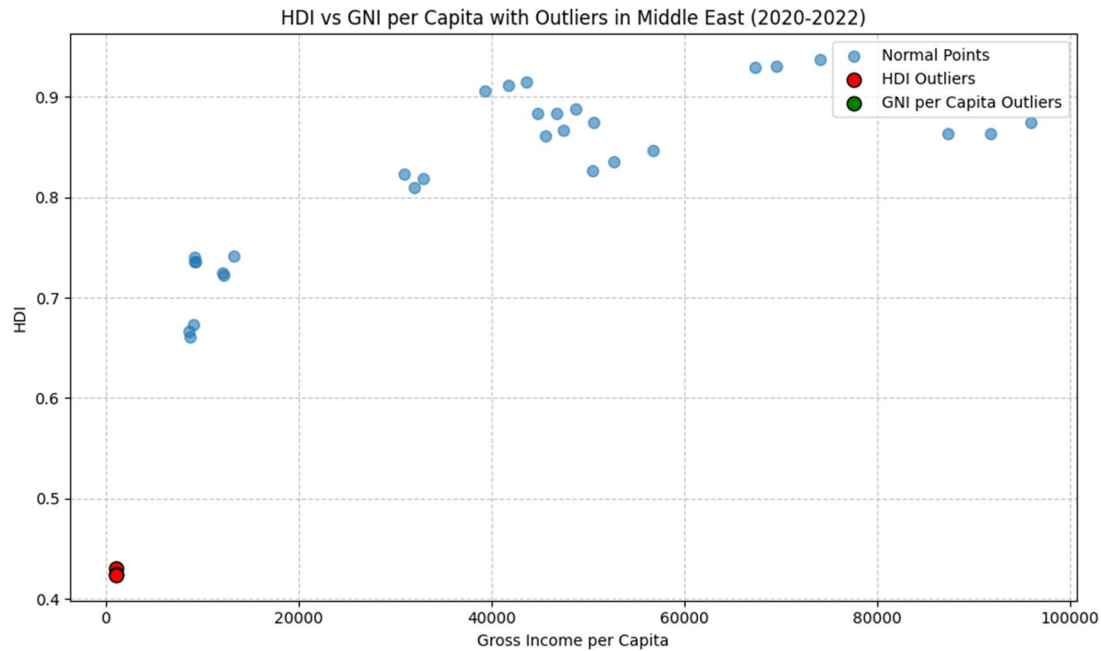












Interpretation

The GNI per capita outliers were more notable in the Middle East, which is representative of the unreasonable income level in some countries. Such outliers had an important impact on regional averages.

South Asia:

HDI outliers point to those countries whose human development is excessively low or high. GNI per capita outliers indicate either very good economic performance or extreme poverty. Combined outliers (high GNI with moderate HDI) imply that there is an unequal distribution of wealth or that there is inefficient investment in human development.

Middle East:

HDI outliers can represent humanitarian crisis at the lower end, and good welfare, health, and educational systems at the top. The oil rich countries are usually the outliers in terms of GNI per capita; in cases where exceptionally high GNI does not equate to high HDI, inequality or poor public expenditure, or lack of diversification of the economy are likely causes.

Conclusion

This discussion illustrates the applicability of HDI as a multidimensional scale of development as well as showing the weaknesses of HDI. The relationship between income, health and education is complex and the high economic output does not necessarily lead to human development.

Comparisons on a regional basis show that there are unchanging disparities, indicating the necessity of equal policy actions. Although the analysis is constrained by the extent of available indicators, it offers useful information about the trend in global and regional development.

¹References

United Nations Development Programme Human Development Reports.

(Yukiolmafuko, 2023)

● 11% Overall Similarity

Top sources found in the following databases:

- 2% Internet database
- 1% Publications database
- Crossref database
- Crossref Posted Content database
- 11% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	University of Wolverhampton on 2026-01-09 Submitted works	2%
2	University of Galway Canvas on 2025-11-18 Submitted works	1%
3	University of Wolverhampton on 2026-01-06 Submitted works	1%
4	University of Wolverhampton on 2026-01-07 Submitted works	1%
5	University of Wolverhampton on 2026-01-08 Submitted works	1%
6	University of Wolverhampton on 2026-01-09 Submitted works	1%
7	University of Wolverhampton on 2026-01-08 Submitted works	<1%
8	University of Wolverhampton on 2026-01-07 Submitted works	<1%

9	University of Wolverhampton on 2026-01-08 Submitted works	<1%
10	University of Wolverhampton on 2026-01-08 Submitted works	<1%
11	University of Wolverhampton on 2026-01-09 Submitted works	<1%
12	University of Wolverhampton on 2026-01-09 Submitted works	<1%