STATISTICAL INTERPRETATION AND EXPLORATORY DATA ANALYSIS

NAME: BIPLOV MAHARJAN

STUDENT ID: 2462258

CONCEPTS AND TECHNOLOGIES OF AI

TUTOR: RAJU KARKI

DATE: 9TH JAN, 2026

Contents

INTRODUCTION

HDI (Human Development Index):

A composite metric used to evaluate a nation's social and economic progress is the Human progress Index (HDI). In 1990, the United Nations Development Program (UNDP) introduced it. In contrast to focusing solely on GDP (income), HDI takes into consideration standard of living, health, and education to present a more comprehensive picture of people's well-being.

Components of HDI (Human Development Index):

There are three major components are:

1. Health (life expectancy)
2. Education (level of literacy)
3. Standard of living (life style)

Why HDI is important

- Focus on the welfare of individuals rather than only economic progress (GDP).
- Aids in comparing living standards between nations.
- Help legislators create improvements to income, education, and health.

Objectives of the Analysis

1. To know human Development Levels.
2. Determine The Strengths and Weaknesses
3. Direct Policy and Decision-Making
4. Compare Countries or Regions
5. Track Development Over Time

Scope of the report

This report analyzes Human Development Index (HDI) data using Python to understand global development patterns while building practical data analysis skills. By examining relationships between HDI components including health, education, and income, the project develops essential

skills in data handling, analysis, and visualization relevant to data science, AI, and advanced analytics.

Problem-wise Analysis:

Single Year HDI Exploration

1. Methods and approach

The Human Development Index (HDI) for 2022 is examined in this analysis. Data cleaning was done to fix missing values and guarantee reliability after the dataset was filtered to only contain records from the chosen year. To illustrate the differences in global development, descriptive analysis was used to determine which nations had the highest and lowest HDI values. To compare income and development levels, additional research concentrated on nations with HDI values greater than 0.800, which were arranged by GNI per capita. To comprehend the structure and facilitate precise analysis, basic dataset exploration was also carried out.

2. Visualizations and tables

	iso3	country	year	hdi	life_expectancy	pop_millions	hdi_f	hdi_m	life_expec_f	life_expec_m	...	gender_development	gender_inequality
3333	LIE	Liechtenstein	2022	0.942	84.656	0.039327	0.924163	0.974295	86.129	83.019	...	0.949	NaN
4719	QAT	Qatar	2022	0.875	81.559	2.695122	0.892561	0.869100	83.101	80.626	...	1.027	0.212
5214	SGP	Singapore	2022	0.949	84.133	5.975689	0.945325	0.954078	86.295	81.981	...	0.991	0.036
2706	IRL	Ireland	2022	0.950	82.716	5.023109	0.942838	0.951392	84.412	81.017	...	0.991	0.072
3399	LUX	Luxembourg	2022	0.927	82.591	0.647599	0.920482	0.927311	84.770	80.445	...	0.993	0.043
6105	ARE	United Arab Emirates	2022	0.937	79.196	9.441128	0.922940	0.936027	81.412	77.720	...	0.986	0.035
5610	CHE	Switzerland	2022	0.967	84.255	8.740472	0.949369	0.977383	85.932	82.484	...	0.971	0.018
4323	NOR	Norway	2022	0.966	83.393	5.434319	0.957148	0.970415	85.055	81.718	...	0.986	0.012
6171	USA	United States	2022	0.927	78.203	338.289857	0.927693	0.923405	81.015	75.479	...	1.005	0.180
2475	HKG	Hong Kong, China (SAR)	2022	0.956	84.315	7.488865	0.942671	0.969515	86.893	81.754	...	0.972	NaN

10 rows × 29 columns

secondary_education_f_%	secondary_education_m_%	seats_in_parliament_f_%	seats_in_parliament_m_%	labour_participation_f_%	labour_participation_m_%	co2_emissio
NaN	NaN	28.000000	72.000000	52.79	67.26	3.
81.832741	71.417976	4.444444	95.555556	61.73	95.33	39.
80.464020	85.932037	29.126214	70.873786	63.37	77.00	9.
88.585639	86.417415	27.397260	72.602740	59.40	70.51	7.
96.579350	89.328283	33.333333	66.666667	57.99	65.09	13.
82.029137	86.145172	50.000000	50.000000	55.32	89.45	25.
96.939726	97.517437	39.024390	60.975610	61.49	71.94	4.
99.094031	99.274980	44.970414	55.029586	62.53	69.59	7.
95.424278	95.067093	28.060264	71.939736	56.79	67.97	14.
77.866913	84.067802	NaN	NaN	52.91	64.71	4.

co2_emission_tons	mat_footprint_percap_tons
3.736776	NaN
39.884274	63.5912
9.397481	43.1508
7.529751	26.3467
13.185395	49.1903
25.333276	28.1632
4.117550	33.5913
7.573541	32.0515
14.932487	29.3498
4.373242	NaN

the following images are the final representation of our first problem

3. **Key results**

   after basic statisics we can see

```
Mean is  0.7228872549019609
median is  0.7395
Standard Deviation is  0.15302880386427828
```

4. **Interpretation and discussion**

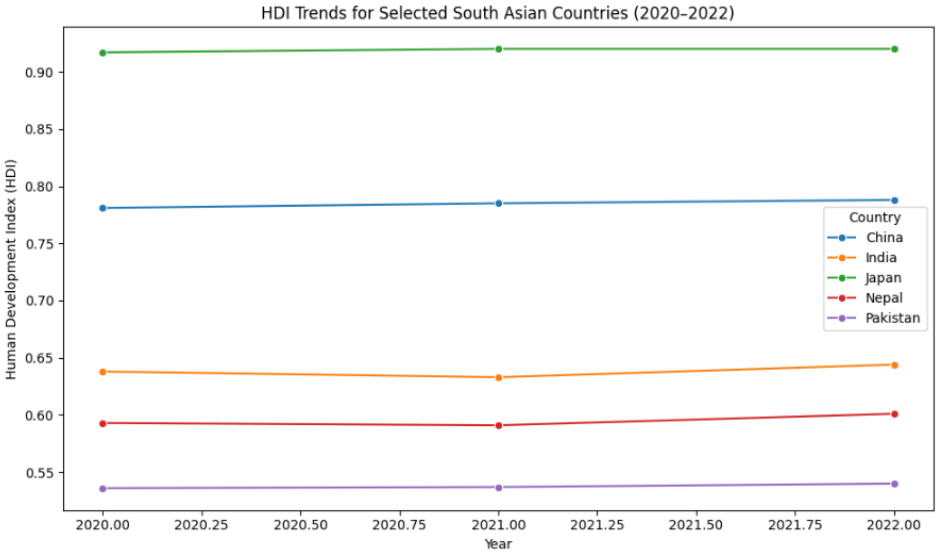
   The output show significant gaps in human development around the world in 2022. There is a strong correlation between income and overall development, as countries with higher HDI values typically have higher GNI per capita. These findings highlight the necessity of concentrated efforts to advance human development in nations with lower HDI scores.

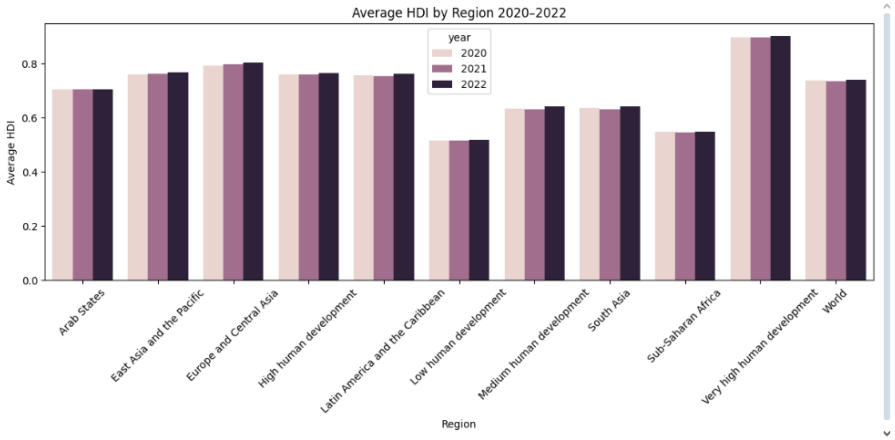
HDI Trend Analysis (2020–2022)

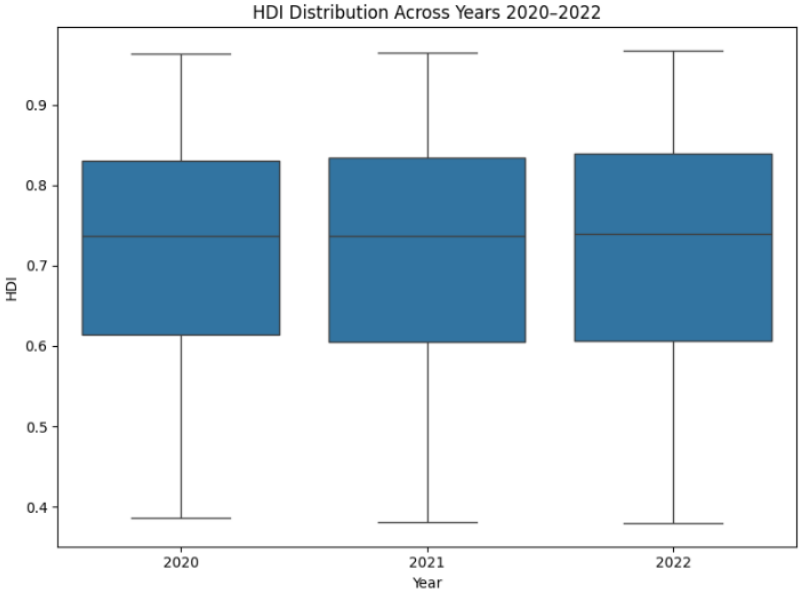
1. **Methods and approach**

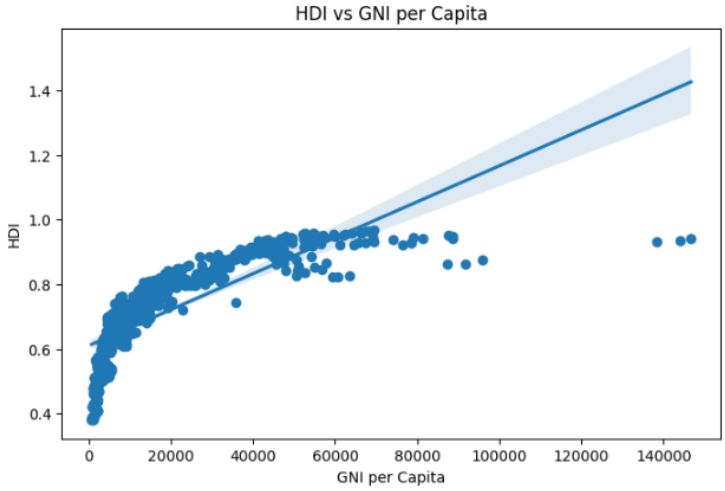
   This analysis examines Human Development Index (HDI) data for the years 2020, 2021, and 2022. The dataset was filtered to include only these years and cleaned to address absent values, inconsistencies, and duplicate records to ensure data reliability. Exploratory analysis and visualizations were then used to study HDI trends across countries and regions, including comparisons over time and relationships between HDI and economic indicators such as GNI per capita. These methods support a perspicuous understanding of development patterns and changes during the selected period.

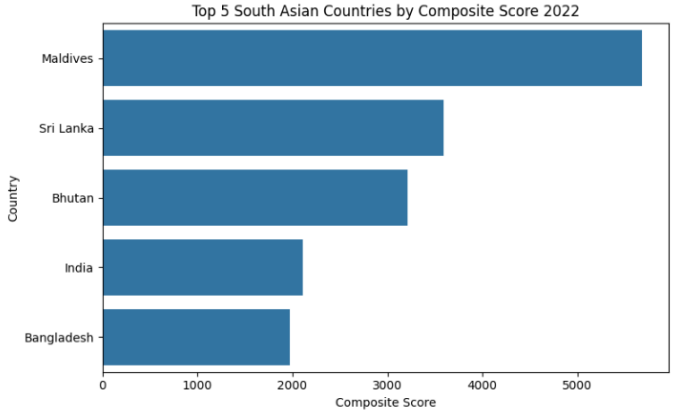
2. Visualizations and tables








HDI vs GNI per Capita
HDI
1.4
1.2
1.0
0.8
0.6
0.4
0
20000
40000
60000
80000
100000
120000
140000
GNI per Capita


Top 5 South Asian Countries by Composite Score 2022
Country
Maldives
Sri Lanka
Bhutan
India
Bangladesh
0
1000
2000
3000
4000
5000
Composite Score

3. **Key results**

   The COVID 19 pandemic significantly contributed to a decline in overall HDI during 2020–2022

4. **Interpretation and discussion**

   With high income countries stable and low to middle income countries showing more variation. HDI correlates powerfully with GNI per capita, though some countries achieve high HDI despite infernal income. territorial disparities persist, highlighting the need for targeted development policies.

Advanced HDI Exploration

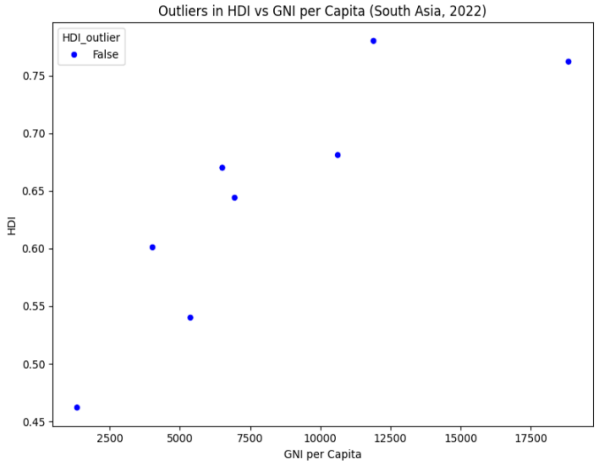
1. **Methods and approach**

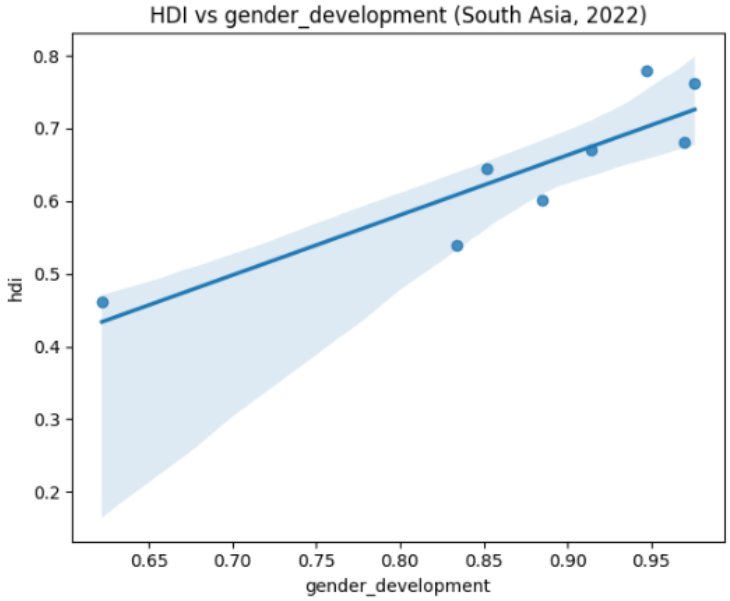
   The HDI data for South Asian countries was filtered and cleaned for analysis. With top performers visualized in bar charts. Outliers in HDI and GNI were identified using the 1 .5 × IQR rule and highlighted in scatter plots. Relationships between HDI components and gross HDI were examined using Pearson correlation, and a GNI HDI Gap metric was used to assess discrepancies between income and development outcomes.

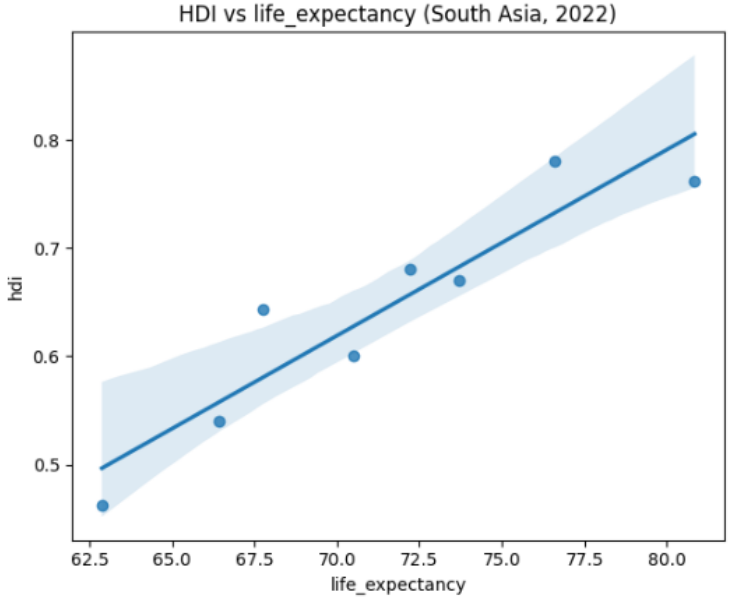
2. **Visualizations and tables**

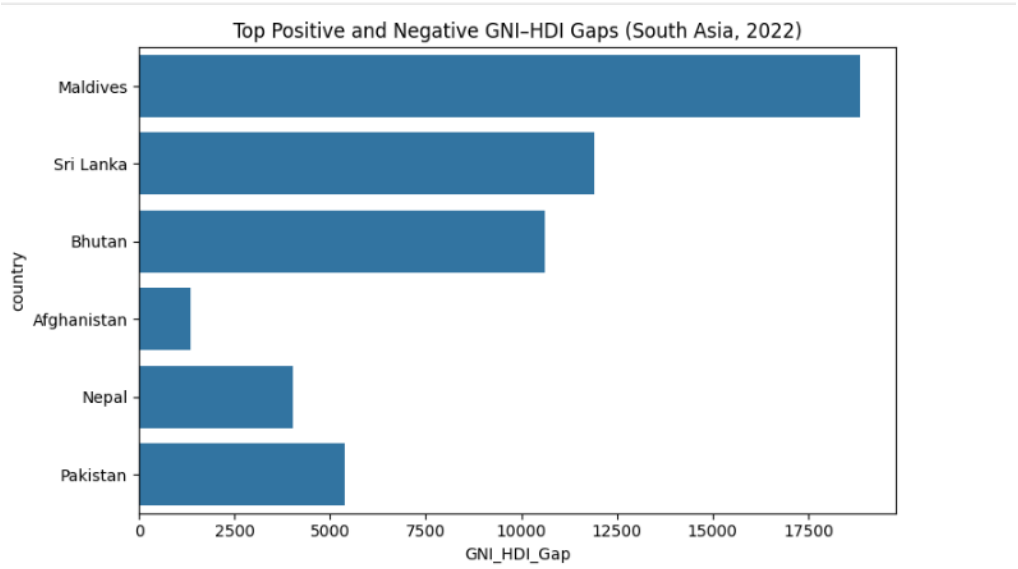
   Ranking South_Asian_countries based on given Score

	country	Composite Score	hdi
3531	Maldives	5678.289357	0.762
5478	Sri Lanka	3592.832541	0.780
660	Bhutan	3209.130864	0.681
2574	India	2105.481239	0.644
462	Bangladesh	1975.446053	0.670
4389	Pakistan	1632.210427	0.540
4092	Nepal	1228.811605	0.601
33	Afghanistan	419.425420	0.462


Outliers in HDI vs GNI per Capita (South Asia, 2022)
HDI_outlier
False
HDI
GNI per Capita
0.75
0.70
0.65
0.60
0.55
0.50
0.45
2500
5000
7500
10000
12500
15000
17500


HDI vs gender_development (South Asia, 2022)
hdi
gender_development
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.65
0.70
0.75
0.80
0.85
0.90
0.95


HDI vs life_expectancy (South Asia, 2022)
hdi
life_expectancy
0.8
0.7
0.6
0.5
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0



3. Key results

Between 2020 and 2022, Middle Eastern countries scored higher on the HDI than South_ Asian_countries, and there was a important difference in GNI per capita between the two countries. Optimistic trends show that long life expectancy and greater gender related development are linked to higher HDI levels.

4. Interpretation and discussion

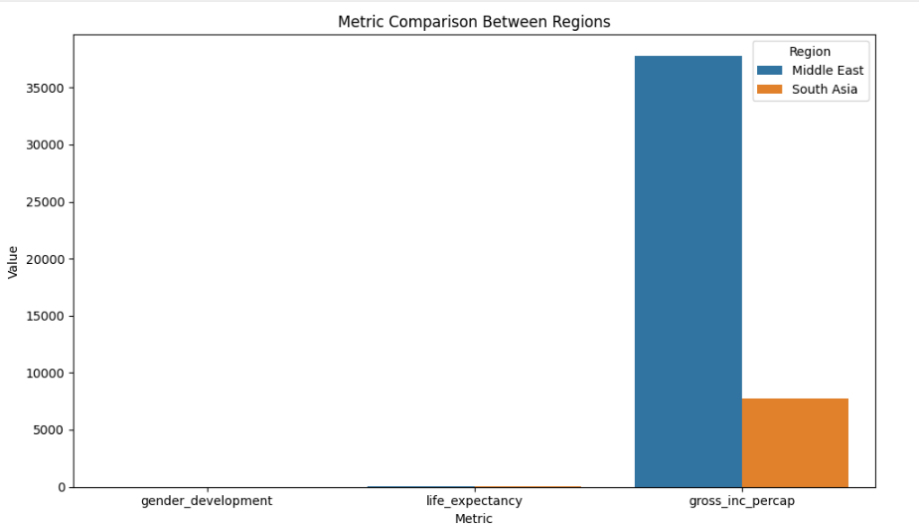
The analysis shows variation in development across South Asian countries. The Composite Score highlighted top performers, while outliers revealed disparities between income and HDI. Correlation analysis identified which HDI components near influence development and the GNI HDI Gap highlighted cases where income does not fully translate into better weak development outcomes.

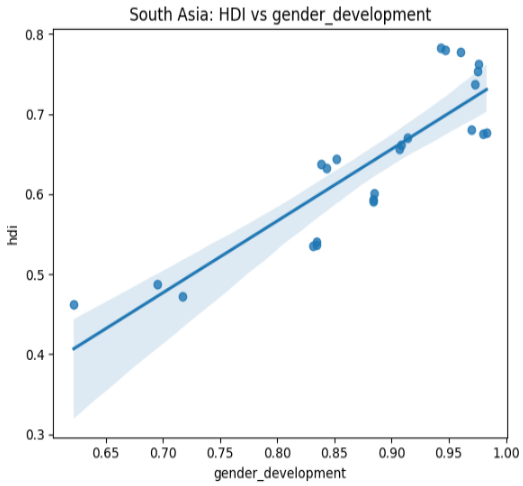
Regional Analysis between South Asia and Middle East

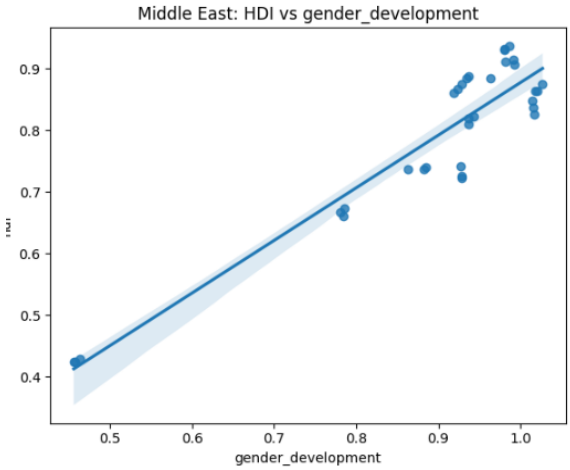
1. Methods and approach

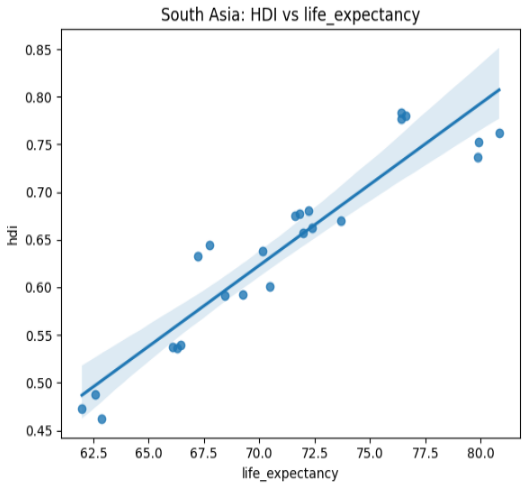
The Middle East and South Asia HDI data from 2020 to 2022 was cleaned and filtered. Bar charts and synchronic data were used to compare top and bottom performers as well as territorial performance. The 1.5 × IQR criterion was used to identify outliers that deviated from territorial patterns, and correlation analysis looked at the correlations between the HDI and its components.

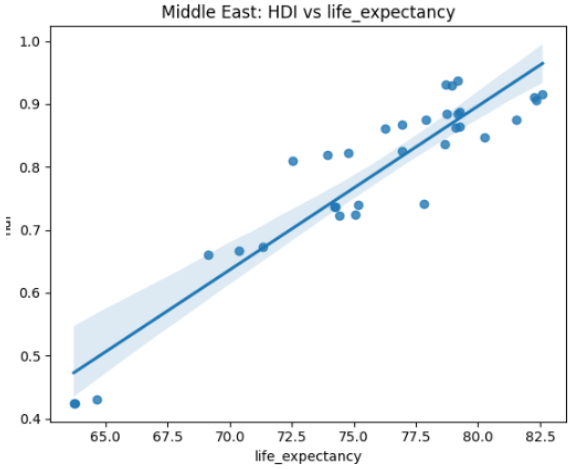
2. Visualizations and tables

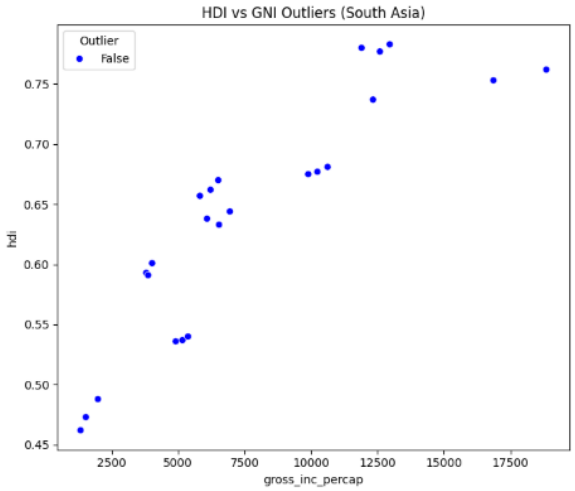


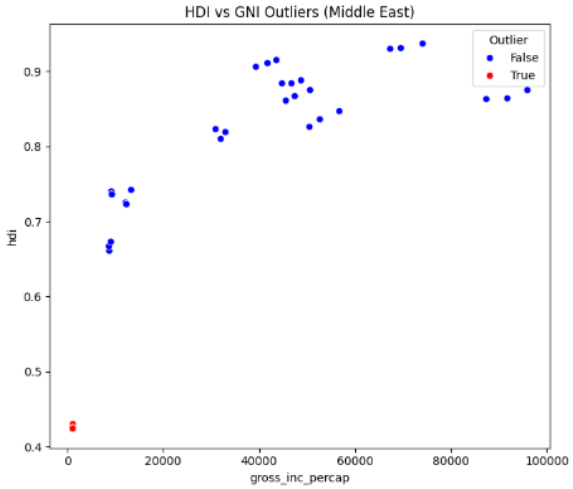

South Asia: HDI vs gender_development
hdi
gender_development
0.8
0.7
0.6
0.5
0.4
0.3
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00


Middle East: HDI vs gender_development
gender_development
0.9
0.8
0.7
0.6
0.5
0.4
0.5
0.6
0.7
0.8
0.9
1.0


South Asia: HDI vs life_expectancy
hdi
life_expectancy
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0


Middle East: HDI vs life_expectancy
life_expectancy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
65.0
67.5
70.0
72.5
75.0
77.5
80.0
82.5


HDI vs GNI Outliers (South Asia)
Outlier
False
hdi
gross_inc_percap
0.45
0.50
0.55
0.60
0.65
0.70
0.75
2500
5000
7500
10000
12500
15000
17500


HDI vs GNI Outliers (Middle East)
Outlier
False
True
hdi
gross_inc_percap
0.4
0.5
0.6
0.7
0.8
0.9
0
20000
40000
60000
80000
100000

3. **Key results**

   In general, between 2020 and 2022, Middle Eastern countries scored higher on the HDI than South Asian countries. The Middle East and South Asia had a big difference in GNI per capita. Higher gender development and long-life expectancy are linked to higher HDI levels, according to optimistic trends seen in all the interactions examined.

4. **Interpretation and discussion**

   According to the analysis, between 2020 and 2022, Middle Eastern countries scored higher on the HDI than South Asian countries. Comparisons of top and bottom performers revealed luminary differences both within and between areas. Countries with abnormally high or low HDI in comparison to their peers were found through outlier detection, informative anomalies in development patterns.

Conclusion

Although some nations differ from this pattern, HDI is positively connected with GNI per capita, life expectancy, and gender development. Although there are differences between countries, the Middle East broadly scores better than South Asia. The impacts of the broad have momently delayed development. Trends and Disparities: Although the HDI represents more broad development indicators, anomalies demonstrate that resources and wealth do not needfully correspond to human development. Compared to South Asia, Middle Eastern nations have higher averages but more variety. Limitations: Long term patterns may not be captured because the analysis is based on current data (2020–2022). Implications: Countries may improve HDI results and lessen inequities by strengthening governance, implementing policies, and making data driven decisions.

References

Python Libraries / Documentation

- Matplotlib (2024). Matplotlib: Python Plotting — Matplotlib 3.1.1 Documentation. [Online] Available at: https://matplotlib.org/
- NumPy (2024). NumPy. [Online] Available at: https://numpy.org/
- Pandas (2018). Python Data Analysis Library. [Online] Available at: https://pandas.pydata.org/
- Seaborn (2012). Seaborn: Statistical Data Visualization — Seaborn 0.9.0 Documentation. [Online] Available at: https://seaborn.pydata.org/

Appendix

- Github link: https://github.com/biplovmaharjan-oss/AI-stuffs/tree/main/Assignment%20-%20I
-