Assignment 1: Statistical Interpretation and Exploratory Data Analysis

by:

Sasika Bista

Level 5

Group 5

Submitted to:

Faculty of Concept And Technologies

Of AI

Bachelor Of Computer Science

Herald College Kathmandu

January 9, 2026

Table of Contents

Table of Figures

Introduction

Brief overview of HDI

The Human Development Index (HDI), introduced by Pakistani economist *Mahbub ul-Haq,* HDI measures a country's achievement in key dimensions of human development, *health* (life expectancy), *education*, and *economic prosperity* (Gross National Income per capital). While HDI is also influenced by income, the HDI, calculated as the geometric mean of these three indices, provides a more comprehensive understanding of development than just economic growth. This dimension is then used to rank countries into four tiers of human development, "low", "medium", "high", and "very high". As such calculation of HDI helps to highlight the areas where countries excel or need improvement.

Objective of the analysis

The main objective of this analysis is for me to develop practical skills in various parts of data science process such as data manipulation and visualization using Pandas, NumPy, and python, visualize key insight using charts and plots for easy interpretation of data using Matplotlib and Seaborn. Besides the main objective of this analysis, other objectives that are to be achieved during this process are to explore the Human_Development_Indec.csv file, from which I am to gain useful insights into global human development trends throughout the decades, as well as identify and understand the patterns and relationship among various factors affecting HDI

Scope of the Report

The scope of this report includes the methodologies applied for basic Exploratory Data Analysis (EDA) performed on a small HDI dataset and discussion on their significant findings. It further covers the interpretation of data visualization and the analysis of key insights. Additionally, the report highlights observed HDI trends and disparities in HDI values.

The scope of this report includes to the completion and documentation of basic data science tasks performed on a small dataset on Human Development Index (HDI). It details the methodology used for data cleaning and processing tasks like handling missing values, determining maximum and minimum HDI, computing metric across regions, years, or country and many more. By using line plots, bar charts, scatter plots, and box plots, the report further explores patterns and relationships within the data. Finally, the report discusses meaningful insights from both statistical summaries and graphical representations to better understand global human development trends. As such the scope of this report is limited to exploration analysis's methodology and significant findings, but doesn't extend to intermediate outputs, step-by-step computation, or code used.

Problem wise Analysis

1. Problem 1A – Single Year HDI Exploration

1.1. Method and Approach

This analysis focused on the latest available year, 2022, to apply basic Exploratory Data Analysis (EDA) techniques on the HDI dataset. Data corresponding to the latest year was extracted after identifying unique years in the dataset and stored separately for analysis. Initial exploration involved examining dataset dimensions, column names, and data types to get familiarized with the data structure.

Then data cleaning was performed to address missing values, duplicate entries, inconsistent country names, incorrect data types, and special characters. Columns with minimal missing values were removed due to their insignificance to the analysis, while remaining missing numerical values were handled using aggregation methods and categorical values were imputed using the mode. Object columns were checked to determine if there exist any object columns having numeric values by attempting conversion to numeric type, to ensure no issues would arise during subsequent computations.

Country name inconsistencies and misspellings were resolved using the pycountry and rapidfuzz libraries. Country names were first looked up using pycountry. Those names that could not be looked up were compared against the full list of countries provided by pycountry, and rapidfuzz's fuzz ratio was used to identify the closest match along with its similarity score. Country names with similarity scores below 80 were dropped from the data set, while those with scores of 80 or higher were left untouched.

For basic statistics, NumPy and Python's built-in functions were used to perform operations including calculation mean, median, standard deviation, and identifying the highest and lowest HDI values for 2022. Countries with an HDI greater than 0.800 were filtered. The filtered countries were then, sorted by Gross National Income (GNI) in descending order to analyze the top 10 countries with the highest HDI based on GNI.

At last, a new column, HDI Category, was added, which classified countries into the four official Human Development Index tiers low, medium, high, and very high based on their 2022 HDI values.

1.2. Key insights

After completing the analysis, the dataset was found to contain 30 columns, indicating 29 factors (excluding the HDI column) were present in the dataset that potentially influenced the Human Development Index. Before the cleaning process, the dataset consisted of 206 countries, which was later reduced to 190 meaningful entries after data cleaning, to ensure higher data quality to perform our further analysis.

Basic statistical analysis of HDI values for the year 2022 showed that the mean HDI was 0.722, while the median HDI was 0.7395, suggesting that the data is slightly left-skewed. The standard deviation of 0.155 indicates uneven level of human development across the countries for this year.

Further statistics revealed that Switzerland ranked the highest HDI in 2022 with a value of 0.967, while Somalia ranked the lowest HDI at just 0.38. This difference in HDI highlights significant global disparities in human development.

Further analysis in this problem set involved filtering countries with high HDI values and then sorting them by Gross National Income (GNI) per capita. The results showed that countries with higher GNI per capita generally tend to have higher HDI values. However, it was observed that higher GNI did not necessarily mean the highest HDI. For example, Liechtenstein, despite having the highest GNI per capita, did not have the highest HDI. This suggests that human development is influenced by multiple factors beyond income alone.

index	country	hdi	gross_inc_percap
3332	Liechtenstein	0.942	146673.2415
4718	Qatar	0.875	95944.37754
5213	Singapore	0.949	88761.14559
2705	Ireland	0.95	87467.51391
3398	Luxembourg	0.927	78554.2364
6104	United Arab Emirates	0.937	74103.71494
5609	Switzerland	0.967	69432.78669
4322	Norway	0.966	69189.76165
6170	United States	0.927	65564.93798
1550	Denmark	0.952	62018.95694

Figure 1 Top 10 countries with high HDI values ranked by Gross National Income (GNI) per capital 2022

Finally, the introduction of the HDI Category column, grouped countries into four development tiers of HDI. This provided a clear overview of global human development levels. The results showed that 67 countries fall into the '*Very High*' category, 47 countries in the '*High*' group, 41 countries as '*Medium*', and 35 countries in the '*Low*' category. This distribution highlights the uneven spread of human development across countries.

1.3.Interpretation and discussion

This analysis summarized key statistical characteristics of the 2022 HDI data and provided a clearer understanding of how countries are distributed across different tiers of human development index. It highlighted moderate variation in HDI values emphasizing global development disparities among the countries. Further supported by a noticeable gap between the highest and lowest performing countries.

The introduction of the HDI Category column further revealed that while a larger portion of countries fell into the *Very High* and *High* HDI groups, a significant portion remains in the *Medium* and *Low* categories. Overall, highlighting the uneven spread of human development across countries.

2. Problem 1B - HDI Visualization and Trend Analysis (2020 – 2022)

2.1. Method and Approach

This analysis focused on exploration of HDI patterns for the years 2020, 2021, and 2022 to uncover regional differences, and trends across countries. Data corresponding to these years were extracted and stored separately for analysis.

The data cleaning was performed to address missing values, duplicate entries, inconsistent country names, incorrect data types, and special characters for essential columns for this problem set mainly speaking for hdi, country, and year columns. Columns with minimal missing values were removed due to their insignificance to the analysis, while remaining missing numerical values were handled using aggregation methods. Object columns were checked to determine if there exist any object columns having numeric values by attempting conversion to numeric type, to ensure no issues would arise during future analysis.

Country name inconsistencies and misspellings were resolved using the pycountry and rapidfuzz libraries. Country names were first looked up using pycountry. Those names that could not be looked up were compared against the full list of countries provided by pycountry, and rapidfuzz's fuzz ratio was used to identify the closest match along with its similarity score. Country names with similarity scores below 80 were dropped from the data set, while those with scores of 80 or higher were left untouched.

To explore patterns and trends in the Human Development Index (HDI) across countries and regions, several visualizations plots were created using the Python libraries Seaborn and Matplotlib.

For the **line chart**, Ghana, Canada, Belgium, Albania, and Somalia were selected at random to demonstrate HDI trends over the years 2020 to 2022. The dataset was filtered to include only these countries. The line chart plotted years on the x-axis and HDI values on the y-axis, using different colors to represent each country.

For the **bar chart**, each country was mapped to a region using pycountry and pycountry_convert. The pycountry provided the country names with their respective two-letter country code while pycountry_convert used the country code to get its respective continent name. For any country whose respective region value was not determined were dropped. The mean HDI for each region was calculated and a grouped bar chart was created to compare average HDI values across regions and years.

The **box plot** was used to visualize the distribution of HDI values for each year (2020, 2021, and 2022). The box plot displayed years on the x-axis and HDI values on the y-axis.

Finally, a **scatter plot** was created to examine the relationship between gross national income prosperity and human development. However, the dataset was first checked to ensure the presence of the gross_inc_percap (GNI per capita) column, if not a suitable message was relayed. The scatter plot displayed GNI per capita on the x-axis and HDI on the y-axis, with a regression line added to indicate the general trend.

2.2.Key insights

The **line chart** illustrates country-level HDI trends. Since the countries were selected at random, the trends for the most countries were same except for one anomaly. High-HDI countries such as Canada and Belgium showed small but steady improvements over the three-year period, reflecting ongoing progress in human development. Middle-HDI countries revealed mixed trends. While Albania, displayed gradual improvement, Somalia, however, experienced a sharp decline in 2022, likely due to unusual disruptions such as economic instability, or health crises/ pandemic. While low-HDI countries such as Ghana remained largely stable, indicating slow or steady progress, Overall, the trends suggest that higher-HDI countries tend to improve steadily, middle-HDI countries improve gradually, and low-HDI countries may experience inconsistent growth.
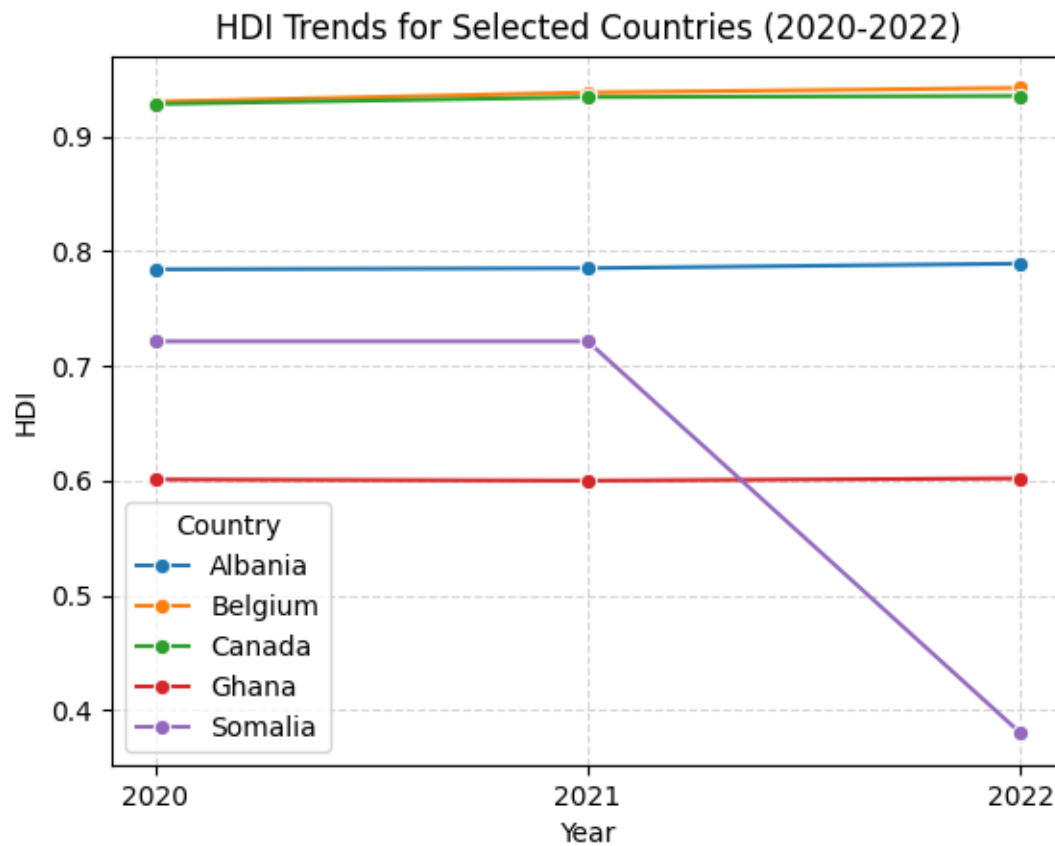

Figure 2 Line chart showing the HDI trends of five selected countries

region	2020	2021	2022
Africa	0.562560686095932	0.56037550091074688	0.557037037037037
Asia	0.7465057718579234	0.7466932718579234	0.7519641051912568
Europe	0.8728669081204727	0.8749134197483798	0.8793552802134961
North America	0.7562608695652174	0.7537391304347826	0.7604347826086956
Oceania	0.7095714285714285	0.709	0.7098571428571429

Figure 3 Mean HDI for each region-year pair

The **bar chart** comparing average HDI by region highlighted regional gaps. Europe consistently had the highest average HDI from 2020 to 2022, whereas Africa had the lowest. Most regions, apart from Oceania, showed slight upward trends, suggesting gradual improvement in human development. Oceania's HDI remained relatively constant, whereas Africa showed a slight decline by 2022. All regions follow expected patterns throughout these years, with no major anomalies observed.
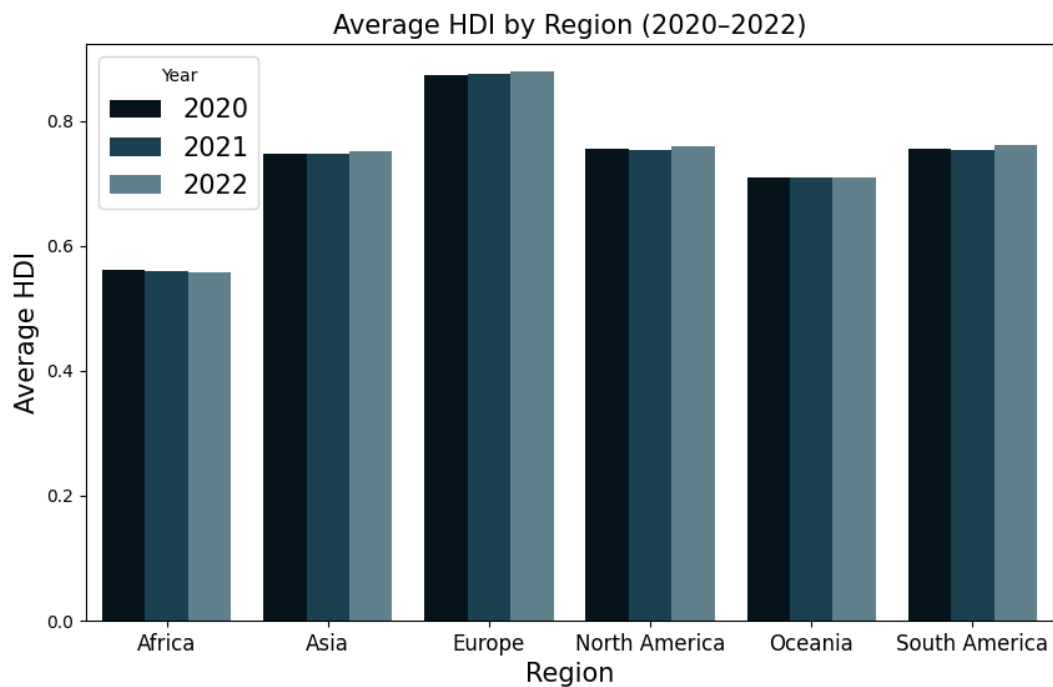
Figure 4 Bar chart showing the average HDI of different regions
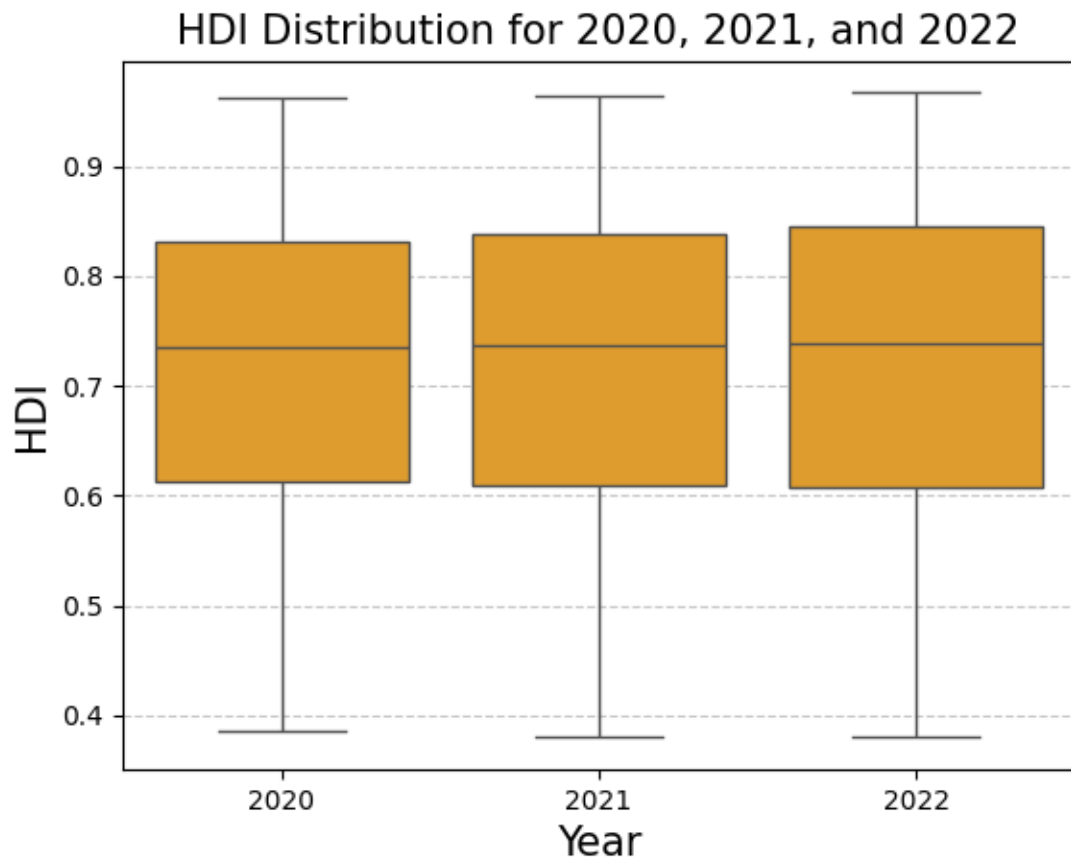
Figure 5 Box plot showing the HDI spread from 2020 to 2022

The **box plot** examining HDI distribution across years indicated stability in central tendencies and variability. The median value ("middle country") remains the same throughout the year, indicating a stable central tendency without major fluctuations. In short, the HDI hasn't changed much from 2022 to 2022 for the median value country. The interquartile range can be observed to be similar across the years, which also suggests that the variation in HDI among countries hasn't changed significantly over the years. While the median is slightly above the center of the box, especially in 2020 and 2022.This suggests a slight negative (left) skew, meaning more countries are above the median, but the skew is not very pronounced. And, finally, since there are no points outside the whiskers, it suggests that there are no outliers in the HDI for year 2020,2021, and 2022.
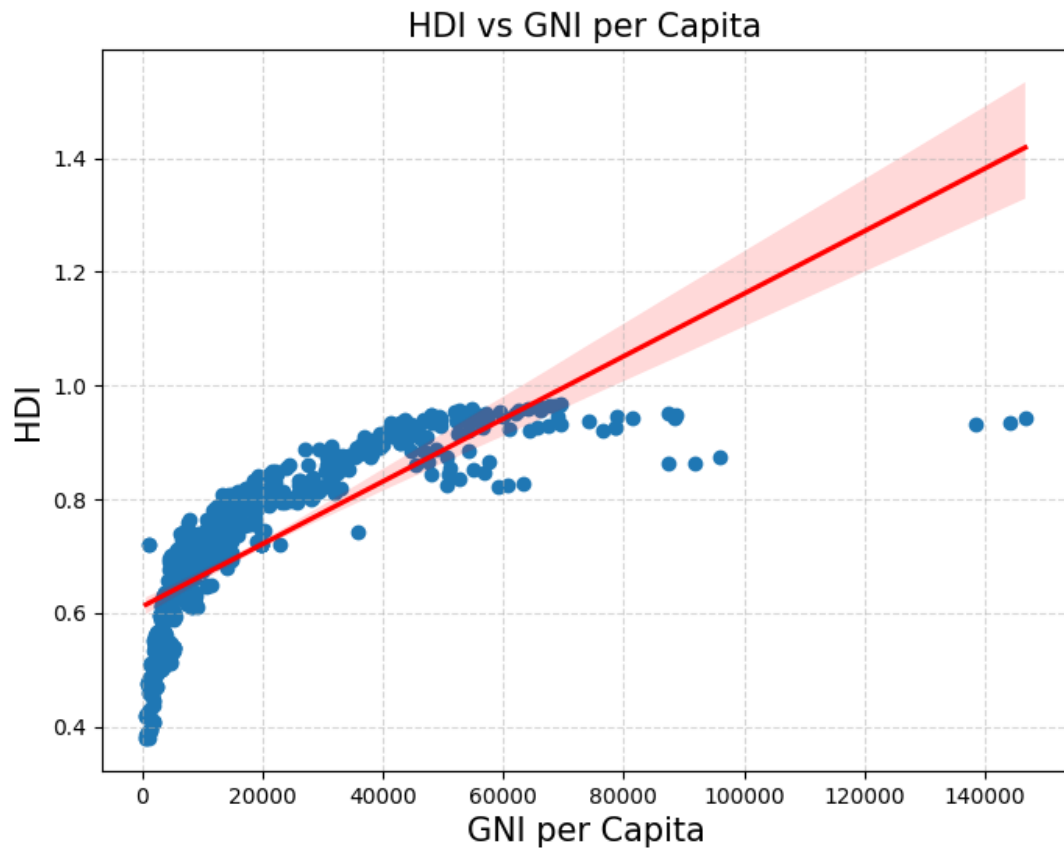
Figure 6 Scatter plot showing relation between HDI and GNI per Capital.

The **scatter plot** of HDI versus GNI per capita illustrated the relationship between income and human development. There was a clear positive correlation between HDI and GNI, countries with higher GNI per capita generally had higher HDI values. The relationship was non-linear, with HDI rising rapidly at lower income levels but leveling off at higher incomes, showing influence of gross national income on human development. While the regression line captured the general trend, it overestimated HDI at very high-income levels. Beyond middle-to-high income levels, HDI values clustered around 0.9–0.95, indicating that additional income contributes less to further HDI gains, this suggests that after certain point in income growth only is not substantial to human development. A few anomalies were also observed: some very high-income countries had HDI values similar to other high-income nations, and some lower-income countries displayed relatively higher HDI than their peers, suggesting that factors beyond income play a role in human development.

2.3.Interpretation and Discussion

The analysis provides a comprehensive view of global human development patterns from 2020 to 2022. Country-level trends show that high-HDI nations tend to improve steadily, medium-HDI countries make gradual progress, and low-HDI countries may experience little or no growth or decline. Regional comparisons highlight persistent gaps, with Europe consistently achieving the highest average HDI and Africa the lowest. Distribution analysis confirms that most countries cluster around similar HDI ranges, with limited variation and no major outliers. The scatter plot of HDI versus GNI per capita reveals a positive correlation, though income alone does not fully explain human development, as when a country's GNI per capita is low, even a small increase in income can lead to a noticeable improvement in HDI. But for already wealthy countries, adding more income results in only small or minimal increases in HDI, because other factors (like education, health) start to matter more. Overall, these insights emphasize that human development is influenced by multiple factors beyond economic prosperity and that gap remain both across countries and regions.

3. Problem 2: Advanced HDI Exploration

3.1.Method and Approach

The advanced HDI exploration for South Asia analysis focuses specifically on south Asia's regional exploration, examination, and related metrics for patterns and trends in human development. By creating a subset of the dataset including only South Asian countries: Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka, the subset was saved separately for further analysis and served as the basis for all subsequent calculations and visualizations for this problem set.

A Composite Development Score was created by combining life expectancy and gross national income per capital with equal weights of 30% each. Initially, when countries were ranked by this score, top positions were dominated by the same country appearing multiple years, making cross-country comparison less meaningful. To address this issue, the analysis was restricted to the most recent years (2020–2022). The dataset was then grouped by country and year, and the average composite score was calculated to handle duplicate entries. Top-performing countries across these years were then visualized using a grouped horizontal bar chart to compare development trends effectively.

Next, outlier detection was performed for both HDI and GNI per capita using the Interquartile Range (IQR) method, with values outside 1.5×IQR considered outliers. A scatter plot was used to visually highlight normal countries versus outliers. GNI on x-axis and HDI on y-axis.

The relationship between HDI and other metrics, Gender Development Index and Life Expectancy Index were then explored using Pearson correlation and scatter plots with regression lines to visualize their relationship.

Finally, a GNI–HDI Gap metric was calculated as the difference between GNI per capita and HDI for each country. Countries were then ranked based on this gap and the largest positive and negative countries were identified. To ensure meaningful comparisons, the analysis was restricted to the most recent year (2022), and the top three positive and negative countries in each gap category were visualized using a color-coded bar chart.

3.2.Key Insights

The composite score analysis for South Asian countries from 2020 to 2022 revealed clear differences in development performance across the countries. Initially, Maldives was overtaken by Sri Lanka, with Maldives recording a composite score of 3,799.86, approximately 76.433439 points lower than Sri Lanka. But then Maldives ranked among the top performers in 2021 and 2022, achieving composite scores above 5,000. Throughout the entire period, Bhutan, India, and Bangladesh consistently ranked third, fourth, and fifth, respectively, indicating stable but comparatively moderate development performance. Overall, these findings highlight the importance of using multiple indicators rather than relying on a single metric when assessing development levels across countries.
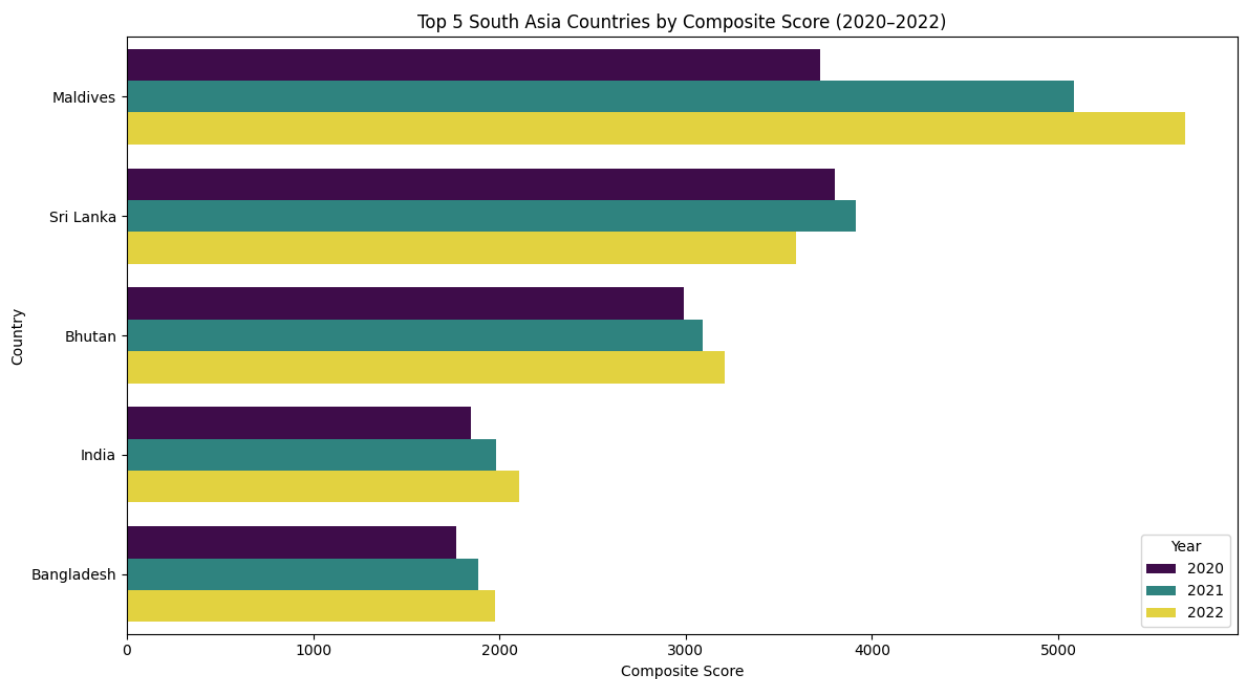

Figure 5: Horizontal plot showing Top 5 South Asia Countries
by Composite Score (2020–2022).

Figure 7 horizontal plot showing Top 5 South Asia Countries by Composite Score (2020–2022).

index	country	hdi	gross_inc_percap	year
3514	Maldives	0.67	16391.40757	2006
3521	Maldives	0.715	15648.36655	2013
3522	Maldives	0.723	16410.16419	2014
3523	Maldives	0.728	16414.93039	2015
3524	Maldives	0.734	16911.45422	2016
3525	Maldives	0.74	17474.64201	2017
3526	Maldives	0.747	18055.72465	2018
3527	Maldives	0.753	18679.92536	2019
3529	Maldives	0.753	16857.60546	2021
3530	Maldives	0.762	18846.79219	2022

Figure 8 Maldives identified as a persistent outlier in HDI–GNI per capita analysis across selected years

Outlier detection revealed that all identified outliers belonged to Malaysia across different years, rather than multiple countries. This indicates that Malaysia consistently deviated from the general South Asian development pattern in terms of HDI and GNI per capita. The repeated appearance of the same country as an outlier suggests, Malaysia despite having comparatively higher income levels had lower HDI comparative to those with lesser GNI than Malaysia and emphasizes that human development is influenced by multiple factors beyond economic prosperity.
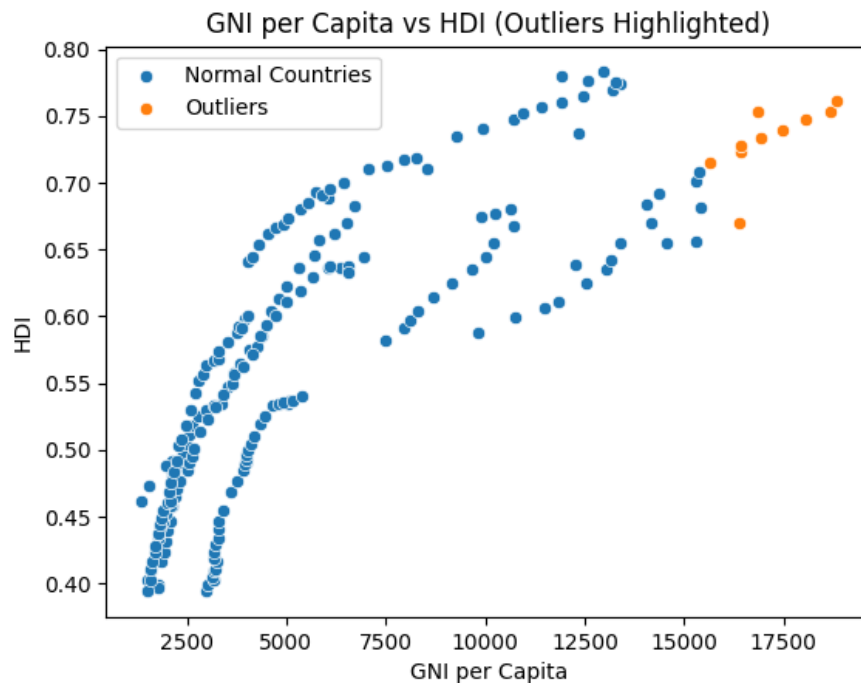

Figure 9 Scatter plot showing normal points and outliers among HDI VS Gross income per capital

The metric relationship analysis showed that the Life Expectancy Index has a strong positive correlation with HDI, since the points are tightly clustered around the trendline. This shows a strong, clear linear relationship with HDI. As life expectancy increases, HDI increases consistently with little variation, indicating that improvements in health outcomes play a crucial and consistent role in enhancing human development. In contrast, the Gender Development Index displayed a weaker correlation with HDI. Since the points are more spread out around the trendline, it indicates greater variability and a weaker correlation with HDI compared to life expectancy. While HDI still increases as gender development improves, the relationship is less direct.
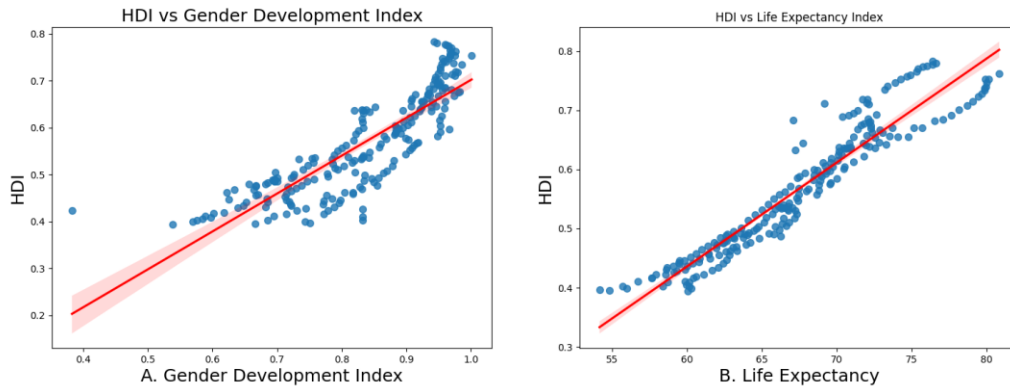

Figure 10 A. Scatter plot between Gender Development Index,

B. Scatter plot between Life Expectancy Index,

The GNI–HDI gap analysis across selected South Asian countries identified top positive and negative countries based on their difference between Gross National Income and Human Development Index. Taller bars indicate countries with higher income to have higher HDI, while shorter or negative bars indicate the opposite.
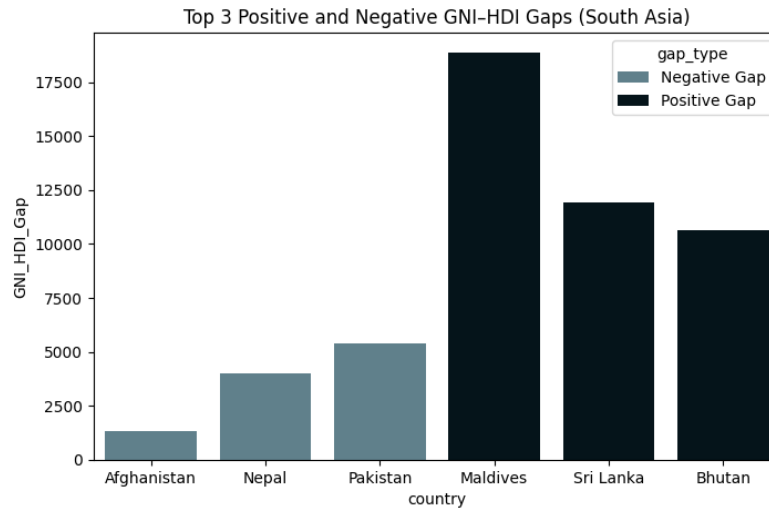

Figure 11 GNI-HDI Gap analysis (Bra graph of Top 3 Positive and Negative Countries)

Maldives, Sri Lanka, and Bhutan ranked as the top three countries with the largest **positive** GNI–HDI gaps, recording gaps of 18846.79219, 11899.49847, and 10624.87388 respectively.

index	country	GNI_HDI_Gap	hdi	gross_inc_percap
3530	Maldives	18846.03019	0.762	18846.79219
5477	Sri Lanka	11898.71847	0.78	11899.49847
659	Bhutan	10624.192879999999	0.681	10624.87388

Figure 12 Top positive South Asian Countries Based on GNI and HDI Gap

In contrast, the countries with the largest negative GNI–HDI gaps recorded 1334.74, 4024.95, and 5373.73 by Afghanistan, Nepal, and Pakistan respectfully.

index	country	GNI_HDI_Gap	hdi	gross_inc_percap
32	Afghanistan	1334.743733	0.462	1335.205733
4091	Nepal	4024.953685	0.601	4025.554685
4388	Pakistan	5373.730423	0.54	5374.270423

Figure 13 Top negative South Asian Countries Based on GNI and HDI Gap

However, because HDI and GNI are on very different scales, HDI ranging between 0 to 1 while GNI ranging from 1000 to 10000+, taking their difference doesn't yield any significant results. It can be observed from the figure 12 and 13, that the values of GNI-HDI gap and GNI itself are very close to one another. Hence, the countries recorded as top three positive and negative based on GNI-HDI gaps are also top three positive and negative countries based on GNI index alone.

3.3.Interpretation and Discussion

The advanced HDI analysis of South Asian countries shows clear differences in development outcomes across the region and confirms that human development is influenced by multiple factors, not income alone. The composite score analysis highlights that countries with higher life expectancy and GNI per capita generally perform better, but changes in rankings across years demonstrate that development performance is not static. Visualizing the outliers helped highlight that higher income does not always translate into proportionally higher HDI. Metric relationship analysis further showed that life expectancy has a strong and consistent positive relationship with HDI, while gender development exhibits a weaker and influence. Some countries underperform relative to their income levels, while others achieve stronger HDI outcomes despite lower GNI, emphasizing the importance of effective investment in other metrics such as health, gender equality, and education.

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

4.1.Method and Analysis

The Comparative Regional Analysis focuses on examination and comparison of human development index across South Asias and Middle East regions. Previously created dataset from problem 1B (HDI_problem1B.csv) was used to filter datasets to include data of these two regions for the years 2020 to 2022.

Basic statistical measures such as maximum and minimum HDI for each region were calculated using NumPy and Python. Additionally, top three and bottom three countries were also determined for each region and by using a grouped bar graph three top and bottom countries of both regions were visualized and compared based on HDI for comparative regional analysis.

Further metric comparison were analyzed of selected columns: gender development index, life expectancy index, and GNI index. Since the scale of these selected columns vastly differed, min-max scaling was applied on life expectancy index and gross national income index to make them the same range as gender development index. The metrics were then visualized using a grouped bar graph, for analysis and comparison of disparity between these selected columns between each region. Similarly, disparity between HDI index of South Asia region and Middle East region were analyzed by computing range and coefficient of variation for HDI to observe which region had higher variance.

Furthermore, correlation of gender development , and life expectancy with HDI for both regions was also calculated. Scatter plots with a regression line were used to visualize the strength and direction of these relationships to determine the strongest HDI influencing factor.

Finally, outliers in HDI and GNI were identified by using the 1.5 rule, for each region were calculated. These outliers were then visualized using a scatter by differentiating normal data points and outlier data points for each region.

4.2. Key insights

Basic statistical operation for mean and standard deviation revealed that the mean HDI for Middle East countries (0.7889090909090909) were more than for South Asia Countries (0.6395833333333333), as well as in terms of standard deviation Middle east region has higher standard deviation (0.14) than South Asia (0.098).

index	hdi
count	24.0
mean	0.6395833333333333
std	0.09827331747496995

Figure 14 Mean and standard deviation of South Asia region based on HDI

This suggests Middle East countries generally achieve higher HDI values, while South Asia shows lower average HDI, highlighting development gaps. Although South Asia has comparatively low standard deviation than Middle East region suggesting more consistent human development across the countries, South Asia overall development is lower comparative to Middle East.

index	hdi
count	33.0
mean	0.7889090909090909
std	0.14129476362297624

Figure 15 Mean and standard deviation of Middle East region based on HDI

Hence, it is concluded that Middle East countries perform better on average than South Asia countries.

Analysis of range and coefficient variation for HDI support this observation. The Middle East region with range 0.513 and coefficient variation 0.179 exhibits more variation in HDI comparative to Soth Asia with range of 0.321 and coefficient variation of 0.152.
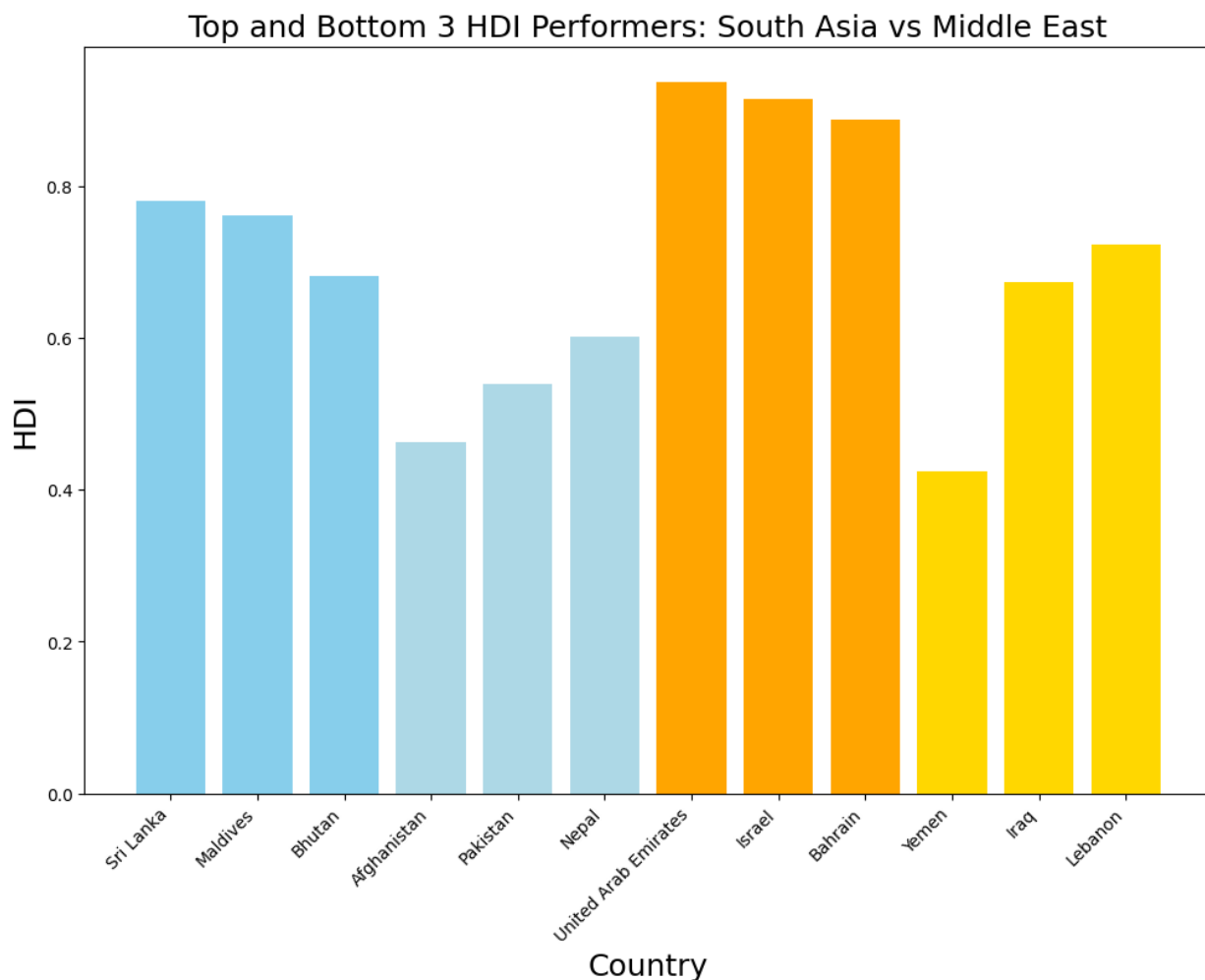
Figure 16 Top and Bottom 3 HDI Performers of South Asia and Middle East Regions

This conclusion is further reinforced by the analysis of top three and bottom HDI performers from each region.

index	region	country	hdi	group
23	South Asia	Sri Lanka	0.78	Top 3
14	South Asia	Maldives	0.762	Top 3
8	South Asia	Bhutan	0.681	Top 3
2	South Asia	Afghanistan	0.462	Bottom 3
20	South Asia	Pakistan	0.54	Bottom 3
17	South Asia	Nepal	0.601	Bottom 3

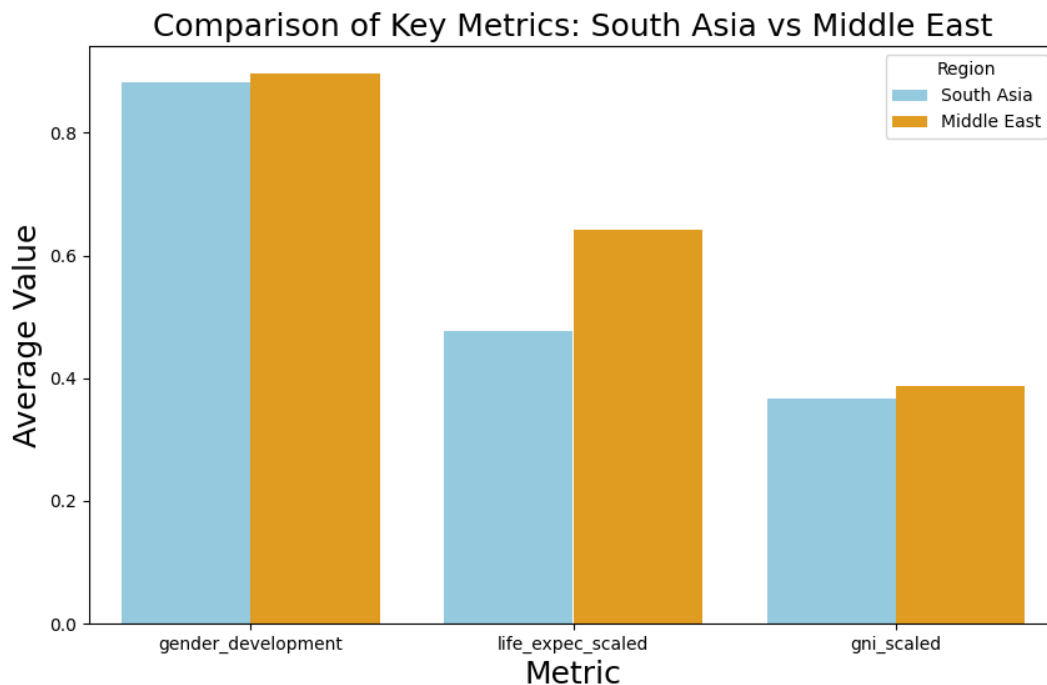*Figure 17 Top 3 and Bottom 3 HDI Performers of South Asia*

While the highest HDI performing countries in South Asia records HDI index of 0.68-0.78, countries in Middle East Countries achieve substantially higher, in range of 0.88-0.93.

index	region	country	hdi	group
29	Middle East	United Arab Emirates	0.937	Top 3
8	Middle East	Israel	0.915	Top 3
2	Middle East	Bahrain	0.888	Top 3
32	Middle East	Yemen	0.424	Bottom 3
5	Middle East	Iraq	0.673	Bottom 3
17	Middle East	Lebanon	0.723	Bottom 3

*Figure 18 Top 3 and Bottom 3 HDI Performers of Middle East Region*

Additionally, even the lowest HDI performing countries of Middle East countries (0.42-0.72) are generally exhibit higher HDI than the lowest performers from South Asia (0.46-0.60).

Further comparison across other key metric like gender development, life ,expectancy, and Groess National Income Index, showed that Middle East regions continue to outperform Soth Asia countries indicating balanced human development in the middle east countries.

Following correlation analysis of gender development and life expectancy with HDI highlights how these metrics influence the overall HDI performance for both these regions.
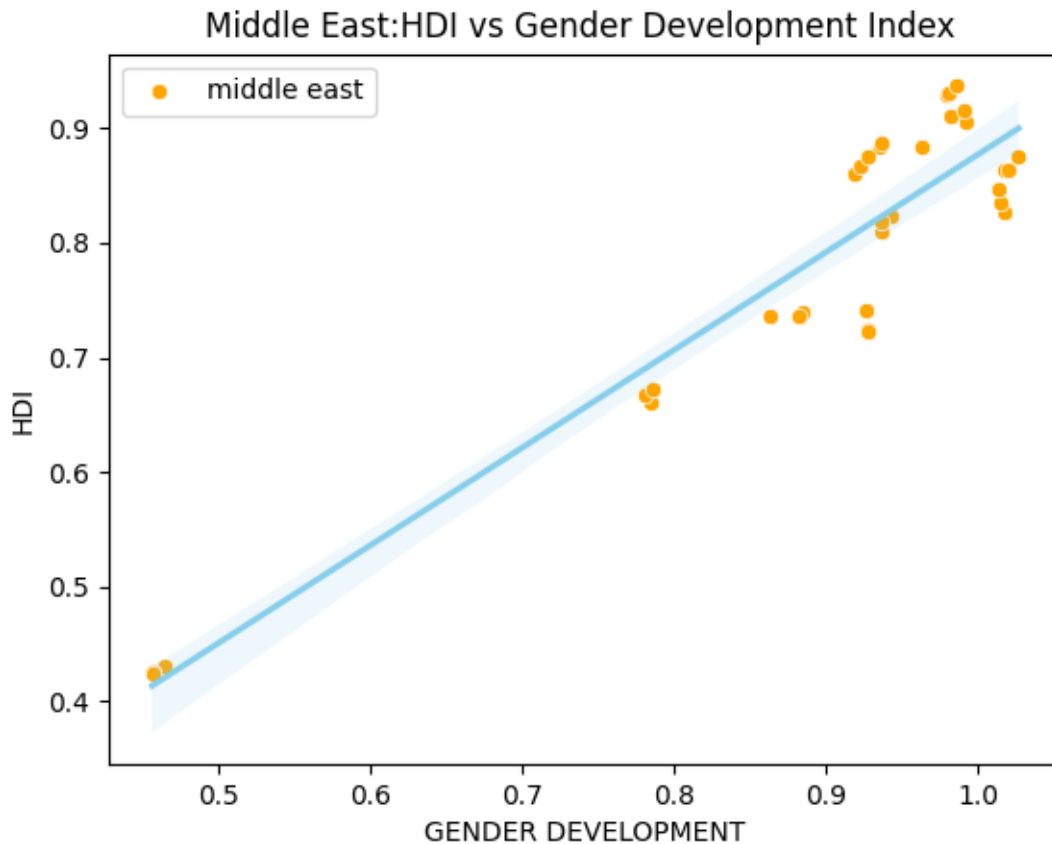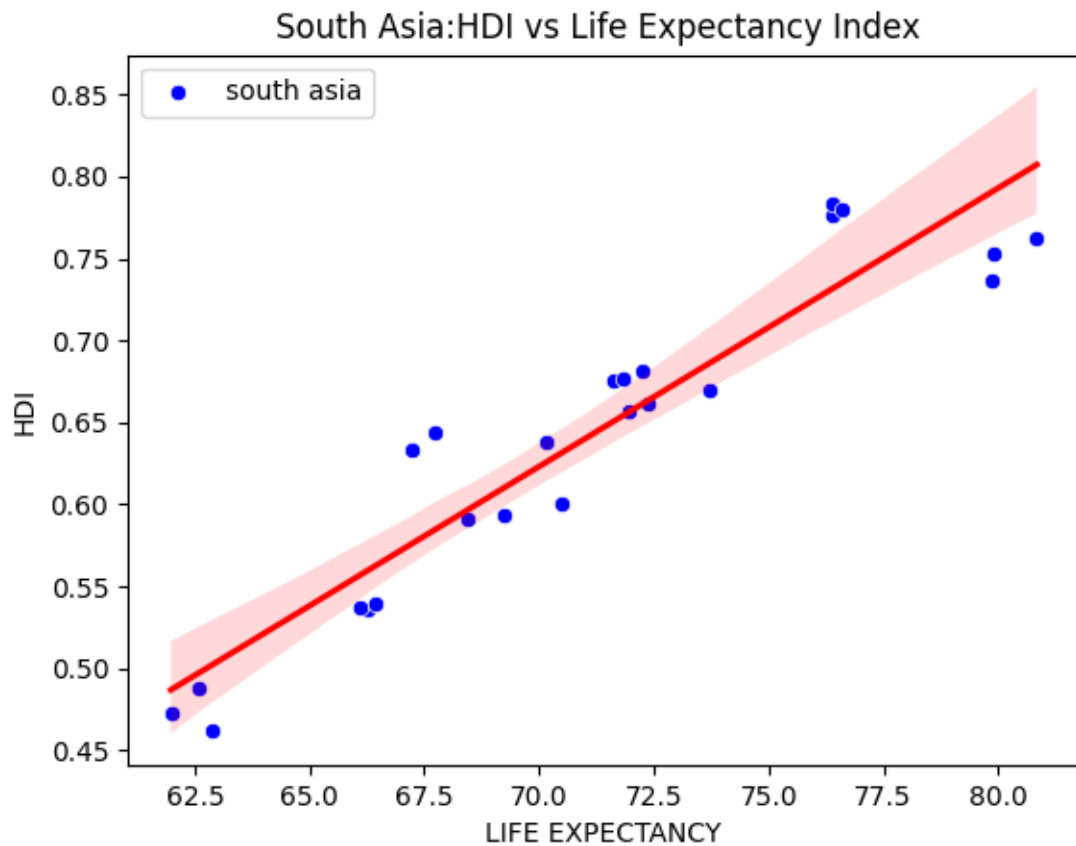

Figure 20 Correlation of Gender Development with HDI for Middle East Region

Figure:11 shows the upward trendline between Gender Development and HDI indicates a positive relationship, as gender development improves/increases, HDI also increases. The data points are closely clustered around the trendline, also suggests a moderate to strong positive correlation. As such, Gender development appears to have a moderately strong positive influence on overall human development in Middle East, as measured by HDI.
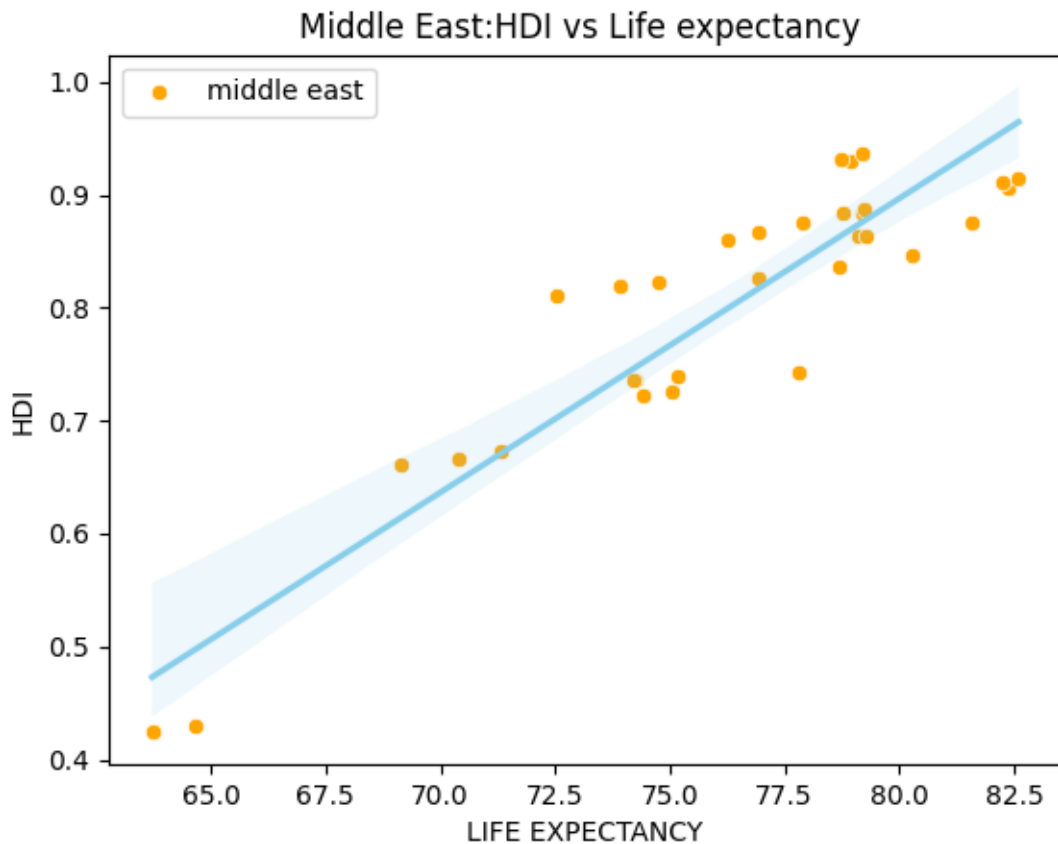
Figure 21 Correlation of Life Expectancy with HDI for Middle East Region

Figure:12 also shows the upward trendline indicates a positive relationship, as gender development improves/increases, HDI also increases. The data points are closely clustered around the trendline, suggesting a moderately strong positive correlation between HDI and Life Expectancy. Although it shows few outliers, this overall analysis suggests that Life expectancy has a moderately strong positive influence on overall human development in Middle East, as measured by HDI.

Figure 22 Correlation of Gender Development with HDI for South Asia Region

Figure :13 shows the upward trendline indicates a positive relationship meaning as life expectancy improves/increases, HDI also increases. Additionally, the data points are loosely clustered around the trendline, suggesting Life expectancy to have a moderate positive influence on overall human development in South Asia, as measured by HDI.
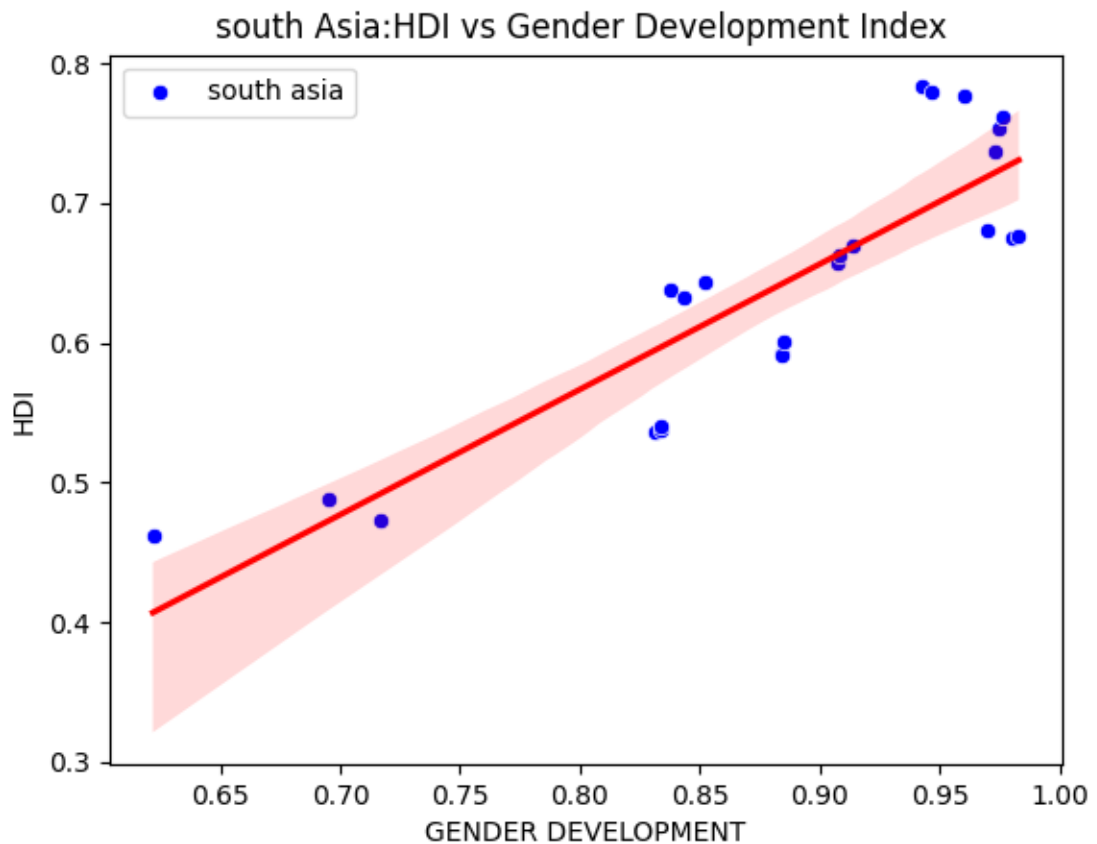
Figure 23 Correlation of Life Expectancy with HDI for South Asia Region

Figure 14: shows the upward trendline indicates a positive relationship, as gender development improves/increases, HDI also increases. The data points are closely clustered around the trendline, suggesting a moderate to strong positive correlation. Although there is relatively larger deviation among certain points, since they still follow the overall upward trend and are also not isolated, we can conclude that there are no outliers suggesting that gender development can positively affect overall human development in south Asia region.

Hence, Middle East region's superior performance across other metrics further solidifies the observation that it outperforms South Asia in terms of HDI.
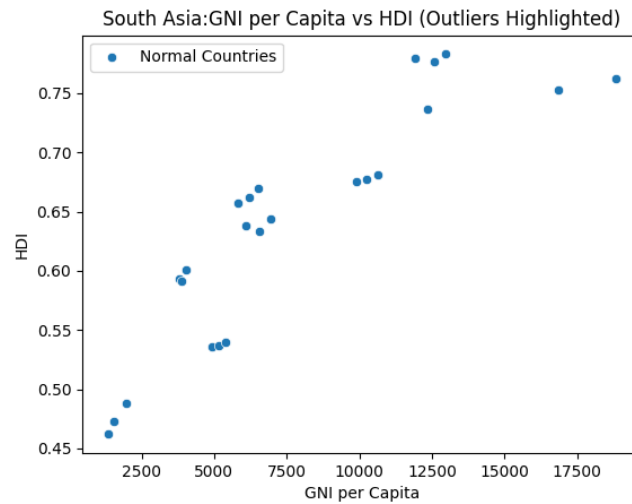
Figure 24 Figure 15 Outlier Analysis for South Asia Region (2020-2022)

Finally, outlier analysis for the period 2020-2022 on South Asia and Middle East region revealed that there were no outliers in South Asia region, suggesting that no country deviated from their general positive trend. In contrast, Yemen emerged as an outlier for all three years, deviating from the regional trend of Middle East region likely due to, its low GNI Index.
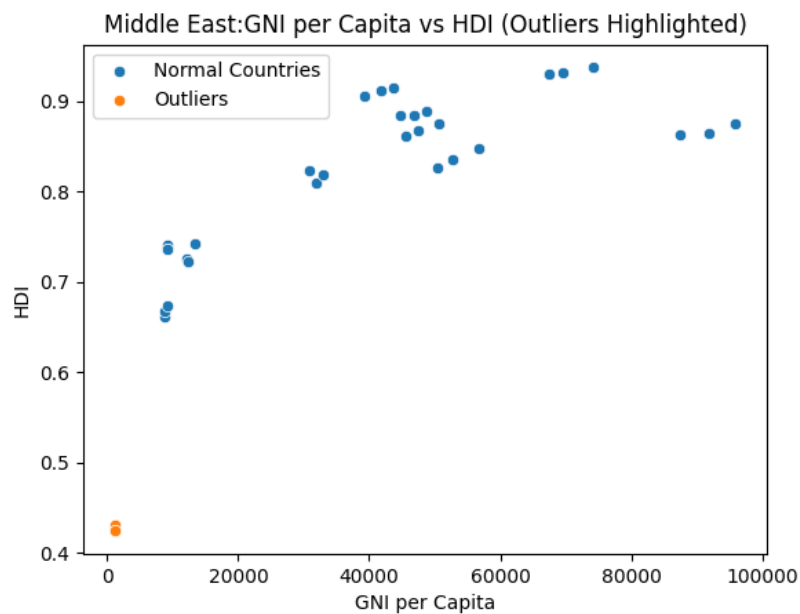

Figure 25 Outlier Analysis for Middle East Region (2020-2022)

4.3.Interpretation and Discussion

The comparative Regional Analysis of South Asia region and Middle East region for the year 2020 to 2022 indicates that the Middle East region consistently outperforms South Asia in terms of the Human Development Index (HDI). Although, South Asia has comparatively low standard deviation than Middle East region suggesting more consistent human development across the countries, South Asia's overall development is lower comparative to Middle East. Also, supported by our outlier analysis which indicated outliers in Middle East region while none in South Asia Region. Similarly, Middle East region also exhibits more variation in HDI comparative to Soth Asia region. Moreover, even in terms of factors like gender development, life ¸expectancy, and Groess National Income Index Middle East regions still outperforms Soth Asia countries indicating balanced human development in the Middle East Countries. Although, Middle East countries

Conclusion:

The analysis of HDI trends for single year, multi-year, region wise, and country wise all indicates that Gross National Income (GNI) strongly influences the overall Human Development Index (HDI). However, the result also shows that the other factors, such as Gender Development and Life Expectancy, can moderate or divert the positive trend in HDI set by Gross National Income, highlighting that human development is driven by multiple interrelation dimensions beyond income only.

Although the provided HDI dataset lacks additional factors such as social, political, happiness index and many more which could further explain variation in human development, this analysis demonstrates that economic prosperity, gender equality, and health remain the major factors driving HDI performance. Additionally, correctness of the data cleaning and processing can significantly affect the result obtained from the analysis. For instant, because the library used rapidfuzz's, fuzz ratio could not give an accurate country match to "Hong Kong, China (SAR)" it was dropped entirely. Since China is one of the most developed countries in the world, it's exclusion could have altered the outcome.
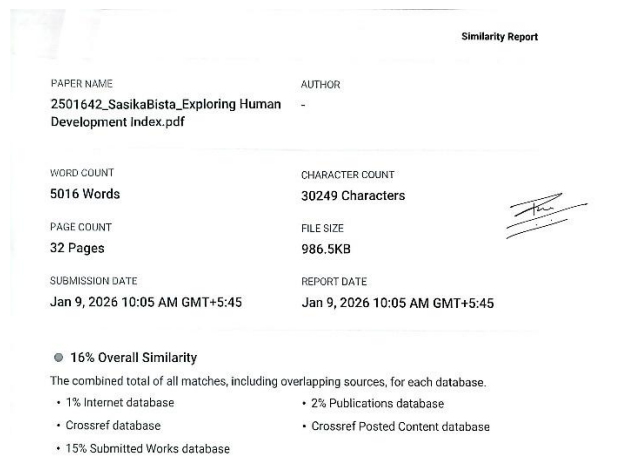
Appendix

Plagiarism Test :

Figure 26 Plagiarism Test For This Report with Plagiarism Percentage and Tutor's Signature

Git Hub Link to Notebook File:

https://github.com/sasika-bista/Concept_and_Technologies_of_Ai.git