REPORT WRITING

Concepts and Technologies of AI (5CS037)

-Statistical Interpretation and Exploratory

Data Analysis of HDI

Submitted To:

Ayush Regmi Sir

Submitted By:

Sandhya karki

BIRATNAGAR
INTERNATIONAL COLLEGE

ing

UNIVERSITY OF
WOLVERHAMPTON

Table of Contents:

BIRATNAGAR
INTERNATIONAL COLLEGE

UNIVERSITY OF
WOLVERHAMPTON

Introduction

The Human Development Index (HDI) is a composite statistic that was launched by the United Nations Development Programme (UNDP) to assess the success of a nation in three primary areas of human development: health, education, and standard of living. It is a combination of indicators of life, expected years of schooling, mean years of schooling, gross national income (GNI) per capita into a single index with a range of 0-1. Nations that score high in terms of HDI are said to be experiencing better general human development outcomes. This report shows a statistical interpretation and exploratory data analysis (EDA) of HDI data between the years 1990 and 2022 with specific emphasis made on 2020-2022 period. The analysis will be organized into four primary issues, each of which will cover various points of the dataset. Problem 1A deals with the last year of HDI (2022), with some base statistics and treatment of missing cases, and management of countries into official HDI classes. Problem 1B explores the trends in HDI in several countries and regions and takes the form of visualizations in the form of line charts, bar charts, box plots, and scatter plots. Problem 2 is devoted to South Asia and it presents some sophisticated methods, including composite scoring, outlier detection, correlations, and gap analysis. Lastly, Problem 3 compares South Asia to the Middle East which shows regional differences and similarities in HDI and other measures. The aim of this report is to offer some insights on global and regional development trends, existing inequalities, and examples of using Python-based data analysis methods. The purpose of the report is to illustrate both positive and negative aspects of human development in the world by using statistical interpretation and visualization of the data.

BIRATNAGAR
INTERNATIONAL COLLEGE

UNIVERSITY OF
WOLVERHAMPTON

Problem 1

Basic Data Exploration & Trend Visualization

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

The first step in the analysis involved filtering the dataset to encompass that of the past year 2022 only and this was named HDI 2022df. The subset consists of 206 rows and 30 columns, i.e. the countries where the HDI data are available in the past year. The initial 10 rows were used to take a look at the format and contents of the dataset info () and.dtypes were used to revise column names and types, and verify that the majority of variables could be quantitative, whereas a few were categorical (country name and ISO code).

An analysis of missing values has shown that some of the columns had incomplete data. Particularly, hdi_f, hdi_m, gross_inc percap f and gross income percap m had 13 values each, whereas hdi had 2 missing values. This was done by and dropping rows that had critical missing values in HDI and GNI per Capita, which guaranteed the integrity of the further statistical analysis. There were also duplicate rows that were eliminated.

Key Results
Mean HDI (2022): 0.7229

Median HDI (2022): 0.7395

Standard Deviation: 0.1530

The country with the highest HDI is Switzerland (HDI = 0.967).

Lowest HDI Country: Niger (HDI = 0.400)

Those countries that had a HDI that exceeded 0.800 were filtered and ranked in terms of

Gross National Income (GNI) per Capita. The top 10 were Liechtenstein, Qatar, Singapore,

Ireland, Luxembourg, UAE, Switzerland, Norway, and the United States which are

considered to be among the nations with the highest HDI scores, and it can be said that they are also very high.

The new category HDI Category was created based on UNDP thresholds. This was distributed as follows:

Very High: 71 countries

High: 54 countries

Medium: 44 countries

Low: 35 countries

The resulting cleaned data was stored as HDI category added.csv which would be used in subsequent activities

B. Generate Visualizations:

One of the subsets was called hdi_trend_df which filtered the dataset to only contain the years 2020, 2021 and 2022 in order to look at the short-term trends in HDI. The rows that had a null value were removed and key columns such as hdi, nation, and year were checked to shed light on missing data. The grossincpercap field was purged by replacing non-numeric characters and converting it to a float and non-duplicated records were removed. A region mapping was adopted to make regional comparisons and country names were formatted using strings to make them standardized.

Figure 1: Line Chart – HDI Trend (2020–2022) for Selected Countries

The chart presents trends of Human Development Index (HDI) of five countries namely Norway, Switzerland, the United States, India and Nepal between the years 2020 and 2022. The developed countries, that is, Norway, Switzerland and the USA, constantly ensure that HDI is above 0.9 representing their high achievement in terms of health, education and income indicators. All these countries present slight variations in the three years,

which emphasizes the steadiness and stability of their system of development even under such global pressure as the COVID-19 pandemic. Compared to this, the developing countries, India and Nepal, show slow growth in the values of HDI with slight declines in 2021 which, probably, were caused by the pandemic-related impacts on healthcare and education. Nevertheless, the 2022 year reports about the recovery of both countries, which shows the good dynamic in the developmental trend of these countries. Notably, HDI does not decrease significantly in any of the countries of this choice and this indicates that despite the external shocks, human development is resilient. The general trend highlights the existing disparity between developed and developing countries, and also demonstrates positive growth among the emerging markets.

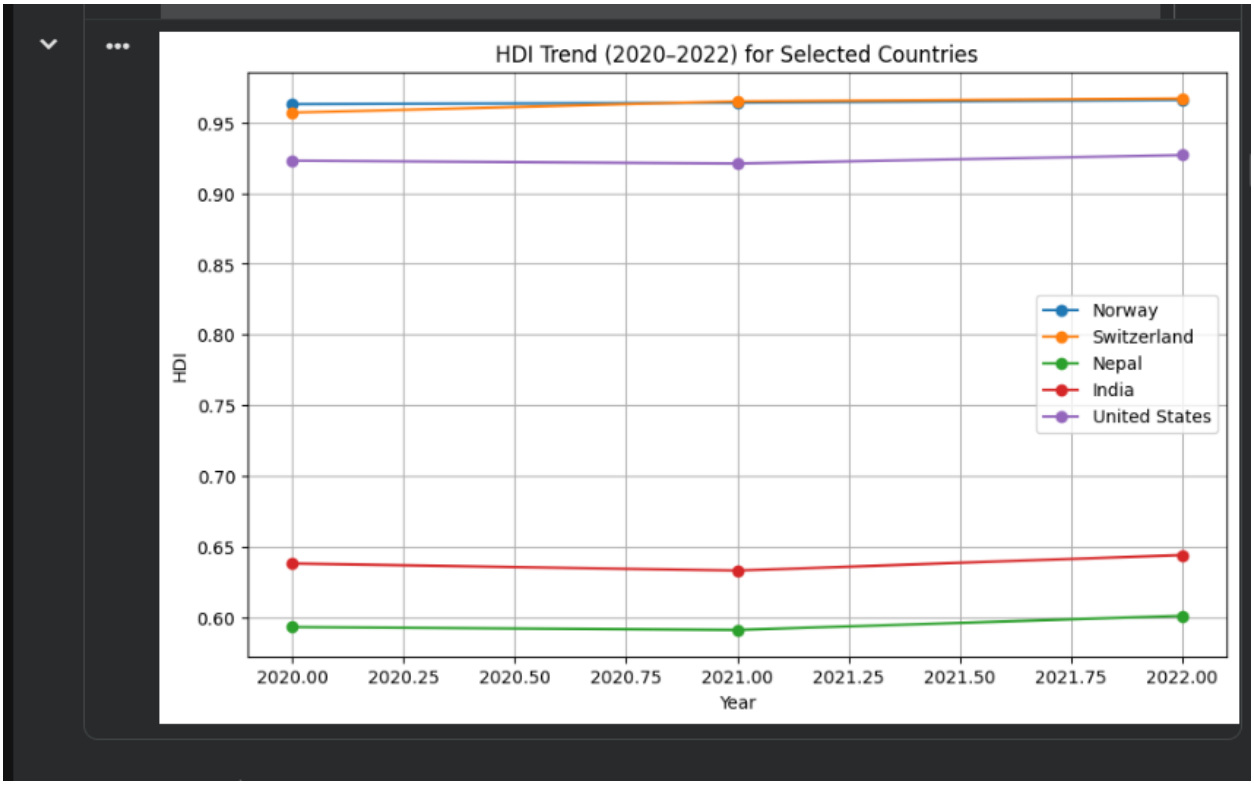


Figure 2: Bar Chart – Average HDI by Region (2020–2022)

The bar chart compares the average Human Development Index (HDI) per different regions during the years 2020, 2021 and 2022. It is clearly presented in the visualization that Europe and North America have always attained the highest HDI averages which are above 0.9 and as a result they have been very active in the health, education and income measures. Oceania also does not lag behind these top regions and is second only to them, whereas Asia and Africa obtain quite low averages, which reveal the existing global inequality in human development. Africa, especially, is at the very bottom as HDI values are concentrated around 0.6, which highlights the many difficulties encountered by most nations in the form of poverty, lack of access to healthcare, and educational disparities. Asia is showing moderate performance, with the average HDI being far lower than Europe and North America, indicating the variation in the level of development in Asia as a continent. Nevertheless, the chart also shows that the value of HDI in all regions slightly improves during 2020-2022, which indicates a slow gradual growth in the world even despite such issues like the COVID-19 pandemic. Such gradual increase points to stability and gradual improvement, yet the margin between high-performing and low-performing areas is still quite high, and it is necessary to have specific policies and international collaboration to minimize inequalities and achieve sustainable human development across the globe.

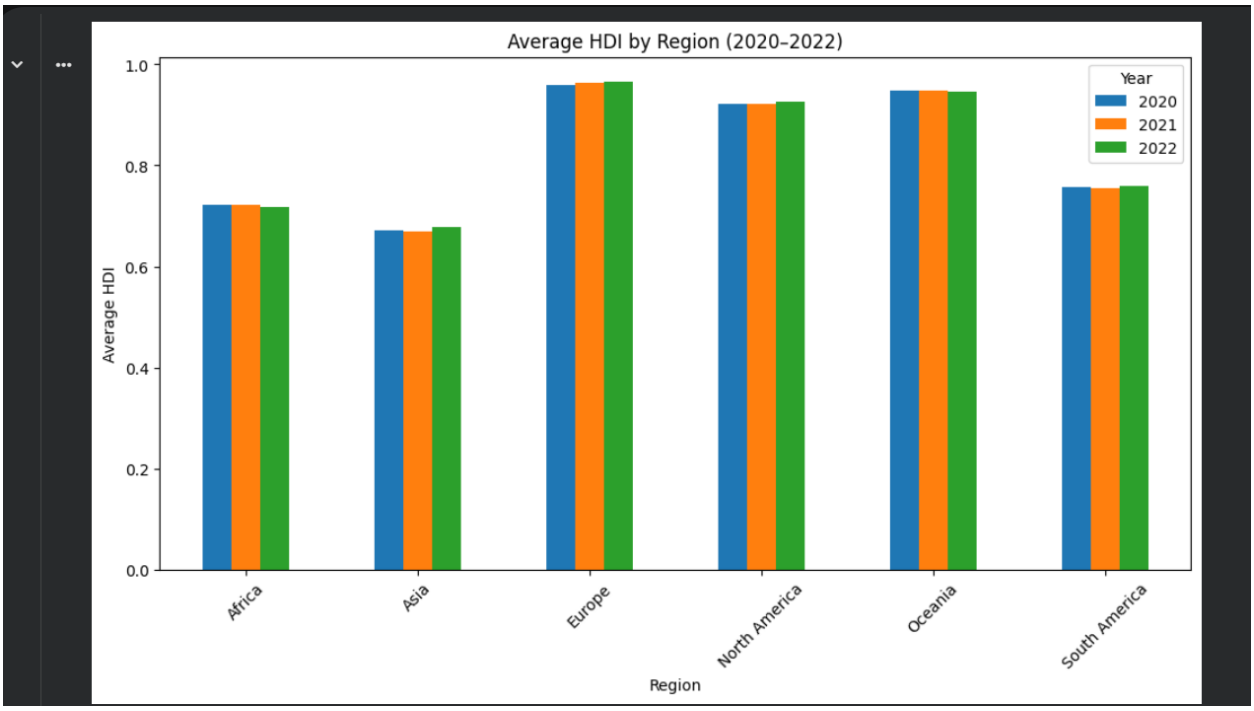


Figure 3: Box Plot – HDI Distribution (2020–2022)

Figure 3 is a box plot that enables the visualization of the values of the Human Development Index (HDI) in 2020, 2021, and 2022, and it allows us to see clearly both central tendencies and the variability of the development of the world. Each box symbolizes the interquartile range (IQR) and the position indicates where the 50 percent of the countries lie and the horizontal line within the box represents the median HDI of that year. The plot shows that the median HDI slightly rises between 2020 and 2022, which indicates the gradual human development improvement at the global scale despite the challenges in the world, including the COVID-19 pandemic. The majority of countries lie between the HDI of 0.6 and 0.8 or the medium and high levels of development based on the classification of the United Nations Development Programme. That clustering implies that the percentage of countries that are steadily developing but have not yet attained the very high level of development is high. The box plot whiskers stretch to both high and low values of the HDI indicating the level of development of the different countries. There are some

countries which show outliers under 0.5, which are countries which have consistently been low in HDI because of various reasons like conflict, poverty, or access to healthcare and education. Even at the top, there is always a group of nations that would always have HDI scores higher than 0.9, which further supports the divide between the well-developed countries and those ones that are still lagging behind. Notably, the general distribution pattern does not change much over the three years, which supports the pattern of gradual yet consistent changes as opposed to radical changes. This stability implies that there is resilience in the development of humankind, and gradual improvements are made on an international scale, and also, the differences between regions and income groups remain evident.

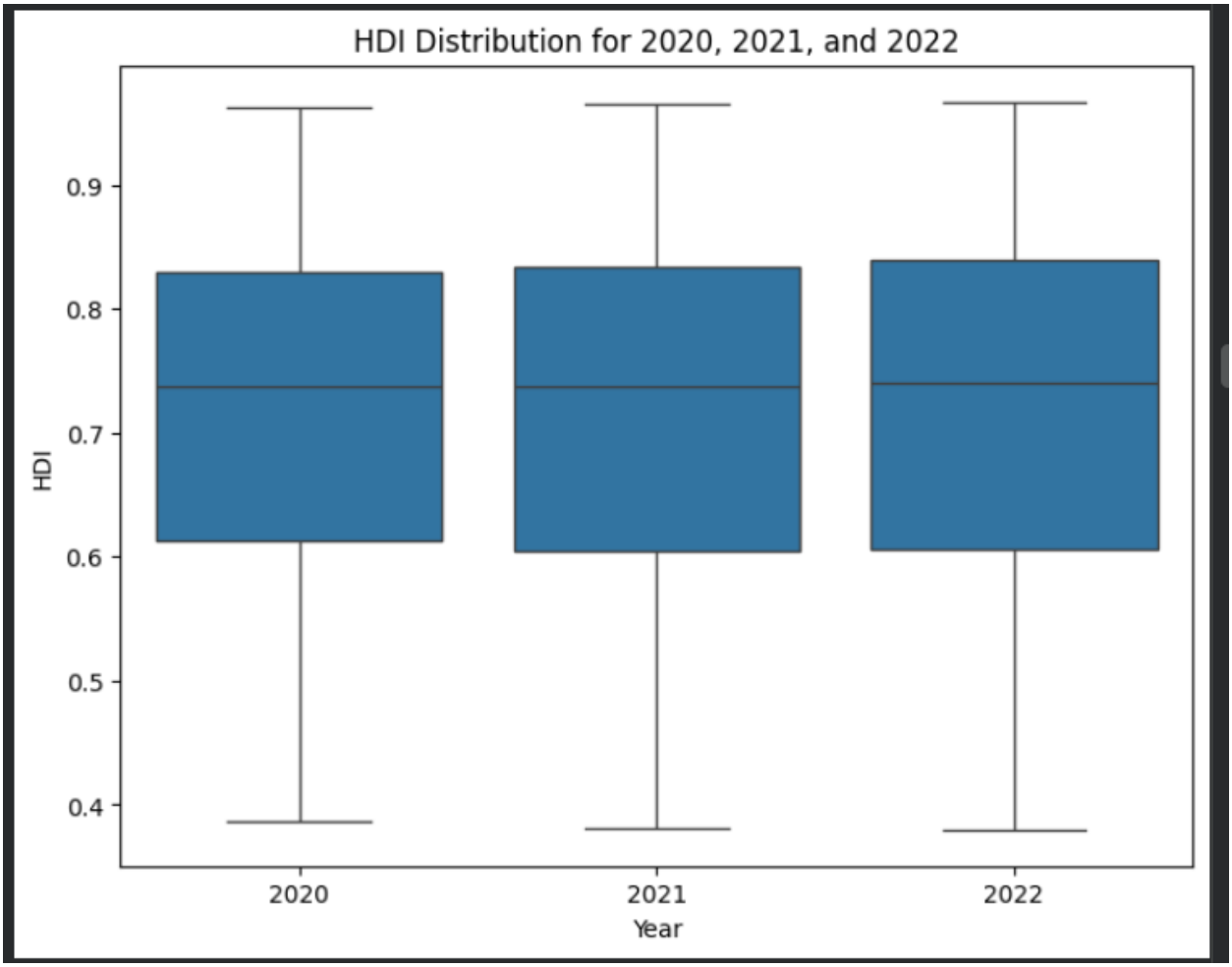
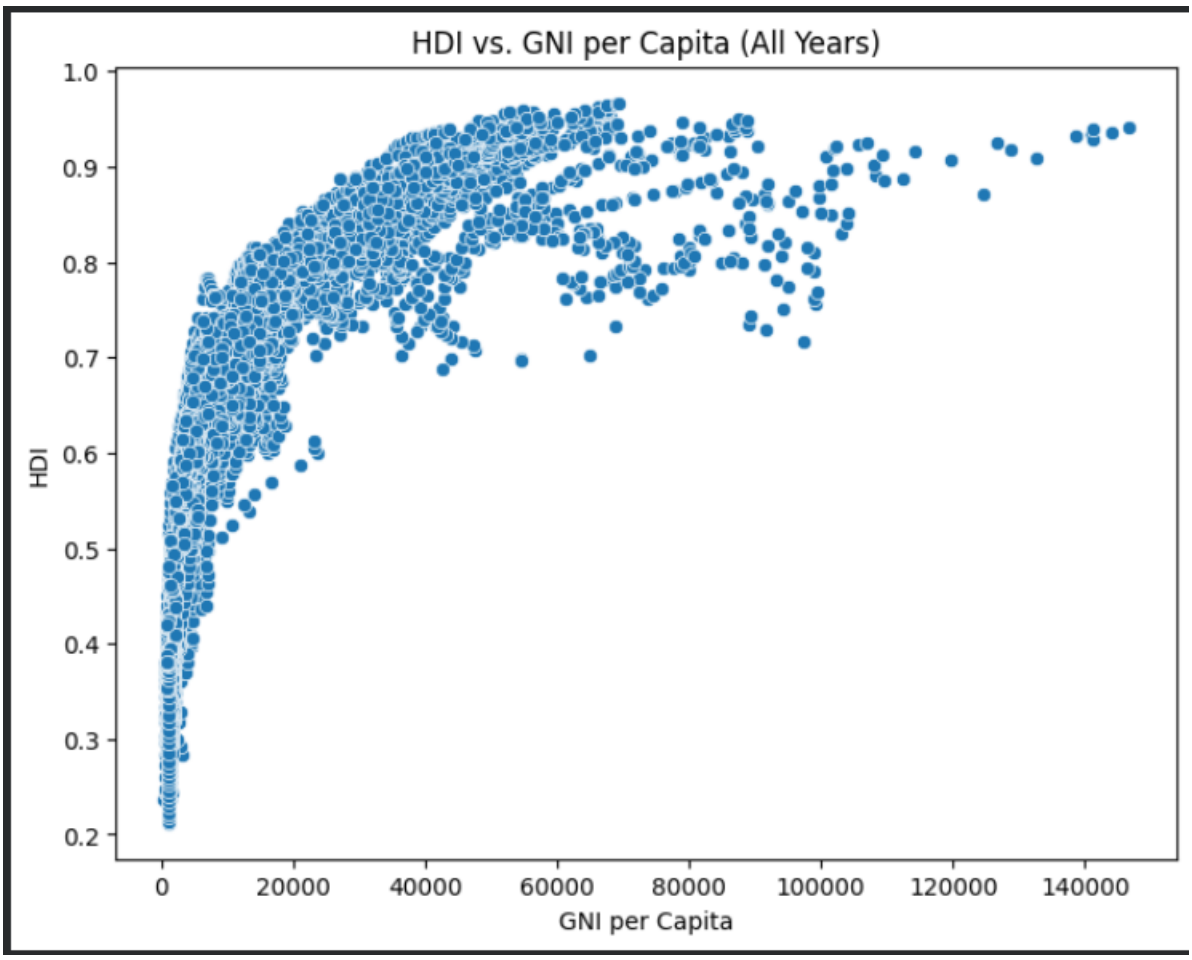


Figure 4: Scatter Plot – HDI vs GNI per Capita

Figure 4 below depicts the scatter plot of GNI per capita and Human Development Index (HDI), which is very informative on the impacts of economic prosperity on overall human development. There is a definite positive correlation in the plot, where the greater the GNI per capita in a country, the higher the value of HDI. This tendency proves that the income is an important contributor to the rise in the health, education and standard of living, which are three parameters that HDI measures. Nevertheless, there are also significant details that the visualization brings into focus. Some moderate-income countries are located quite high on the HDI scale, which proves that income is not the sole determinant of the development outcomes. The cases indicate that through good governance, good social policies and investments in education and healthcare it is possible to see high human development in countries without extremely high levels of income. At the high end of the income level, the association seems to level off, that is, when nations hit extremely high levels of GNI per capita, additional rises in income causes only a slight growth in HDI. This trend shows a decreasing marginal, in which more wealth does not mean similar growth in human development. The scatter plot thus supports the notion that economic growth is a key driver of development, but that it has to be supplemented by policies that can guarantee even distribution of resources, quality education, and healthcare systems. On the whole, the graph shows the significance of income in determining HDI and the constraints of using only economic indicators to measure human advancement, which may indicate the necessity to find a compromise between sustainable development and economic benefits.



Problem 2 – Advanced HDI Exploration (South Asia Focus)

Task 1: Creating the South Asia Subset

The initial analysis in this exploration was to take the South Asian region out of the world HDI data. As a requirement, country names were also normalized by removing any surplus spaces and changing them to title case. It was then narrowed down to a subgroup that includes only the eight South Asian nations; Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka. The number of entries per country was 33, which proved that the dataset consisted of annual records of 1990-

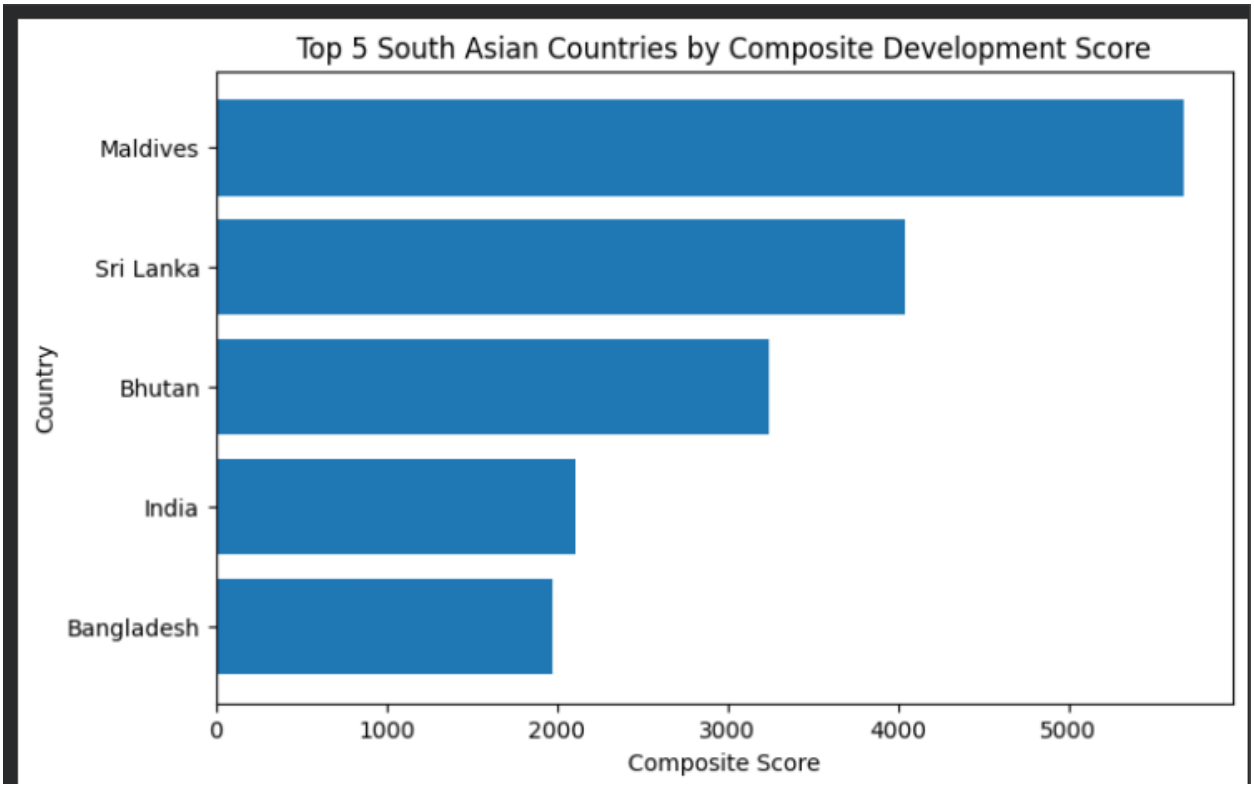
2022. This was a significant step since it enabled the analysis to narrow down on South Asia and avoid noise created by the global data and thereby gain a better understanding of the patterns of development within the region. To save this subset, it was necessary to save it as HDI_SouthAsia.csv to be able to do further work and guarantee the reproducibility of the results. The limited scope of the analysis would help identify differences, advances, and exceptions peculiar to South Asia that would otherwise not be visible in a global dataset.

Task 2: Composite Development Score

In order to acquire a more refined perspective of development, another index named Composite Development Score had been implemented. This score was the combination of two important dimensions of human development namely life expectancy and gross national income (GNI) per capita based on the following formula: Composite Score = 0.30 × Life Expectancy + 0.30 × GNI per Capita This weighting placed a strong focus on health as well as income in influencing the development outcomes. Upon ranking the countries on this score, Maldives was ranked as the obvious winner, followed by Sri Lanka, Bhutan, India and Bangladesh. A horizontal bar chart was drawn to illustrate the top five countries and the findings indicated great inequalities. Maldives was much higher than its neighbours, as it has high investments in the health care and comparatively high income levels. Sri Lanka and Bhutan have also done well with the indications which have a balanced development of the two nations. From this, India and Bangladesh were lower on the scale, but both still ranked in the top five showing a consistent improvement. This exercise has shown that although HDI can give an overall measure, composite scoring could offer more information on how health and income interact to influence development. It was also evident that Afghanistan, Pakistan and Nepal were underdeveloped compared to the rest of the region as seen in the visualization.

To illustrate these results, a horizontal bar chart was created, which shows the five leading countries out of South Asians in terms of the Composite Development Score. The difference is evidently depicted in the chart as Maldives is at the top since the scores significantly differ and then the scores decrease gradually in the other four countries. The x-axis of the chart is

Composite Score, and the y-axis is in a descending order of the countries. This approach of using an inverted y-axis ensures that the country that scored the highest is at the top and therefore, this makes the comparison easy to visualize and more effective. This chart supports the numerical data and gives a good picture of the performance of the regional development. It brings out how Maldives has been able to use its resources well to attain high development results whereas other countries such as India and Bangladesh, though developing, still have more to improve.

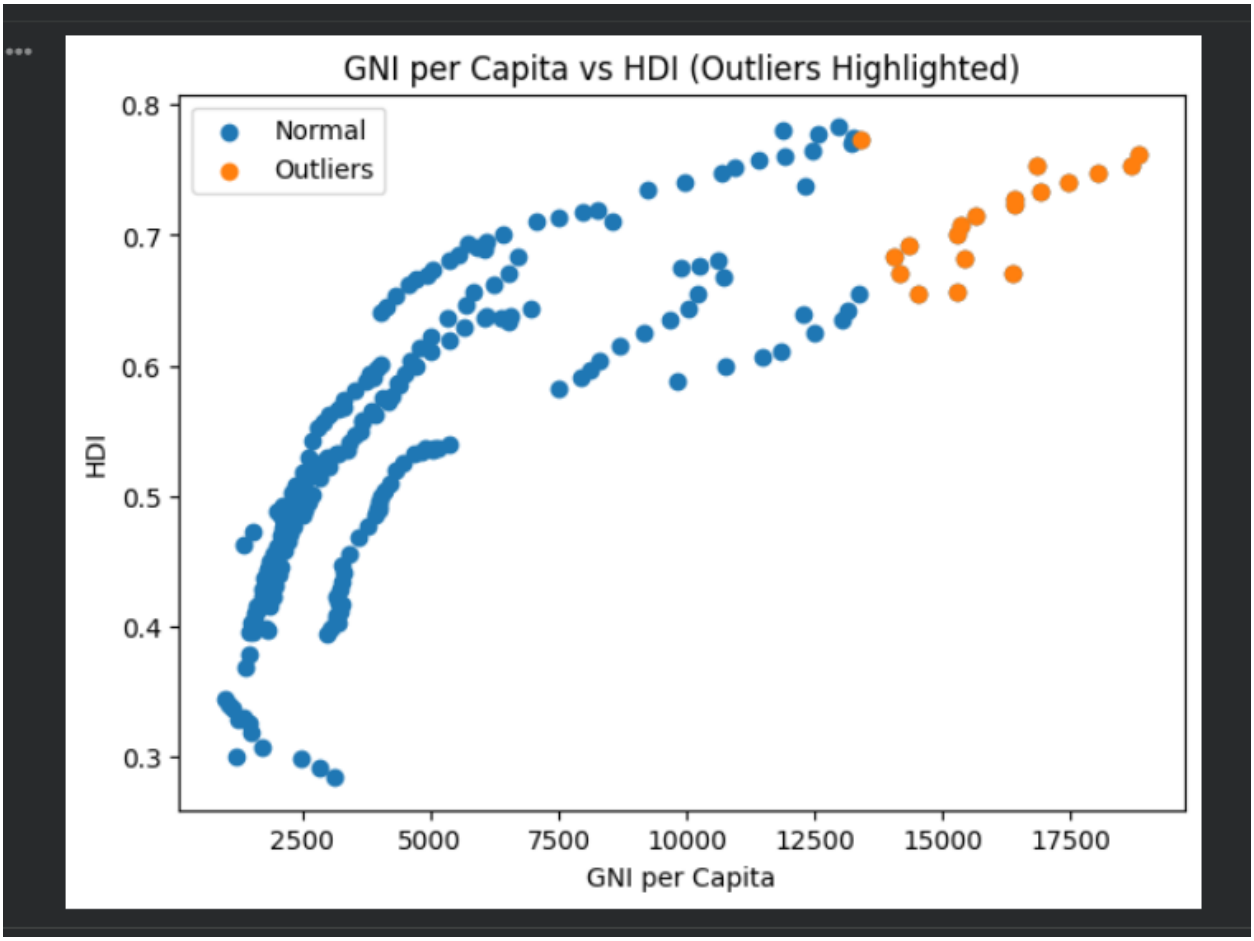


Task 3: Outlier Detection and Visualization

Outlier detection was performed using the Interquartile Range (IQR) method, a robust statistical technique for identifying values that fall significantly outside the expected range. This method was applied to two key indicators: Human Development Index (HDI) and Gross National Income (GNI) per capita. The results revealed that there were no HDI outliers, indicating that all South Asian countries fell within a reasonable range of development outcomes. However, the analysis identified 19 GNI outliers, most of which were associated with Sri Lanka. These anomalies suggested that Sri Lanka's income distribution was notably different from its regional peers, potentially due to economic fluctuations, reporting inconsistencies, or structural imbalances in its economy.

To visualize these findings, a scatter plot was generated with GNI per Capita on the x-axis and HDI on the y-axis. The plot included two categories of data points: blue dots representing “Normal” entries and orange dots representing “Outliers”, as indicated in the legend. The majority of countries followed a clear upward trend, confirming the expected positive correlation between income and human development—as GNI per capita increased, HDI generally rose as well. However, the orange outlier points, particularly those clustered around Sri Lanka, deviated from this trend. These outliers were concentrated in the higher GNI range but did not correspond to proportionally high HDI values, suggesting that income gains were not translating effectively into broader development outcomes.The scatter plot served as a powerful visual tool for identifying and interpreting anomalies. It highlighted how most South Asian countries align with global development patterns, while a few—like Sri Lanka—exhibit unusual behavior that warrants further investigation. This visualization reinforced

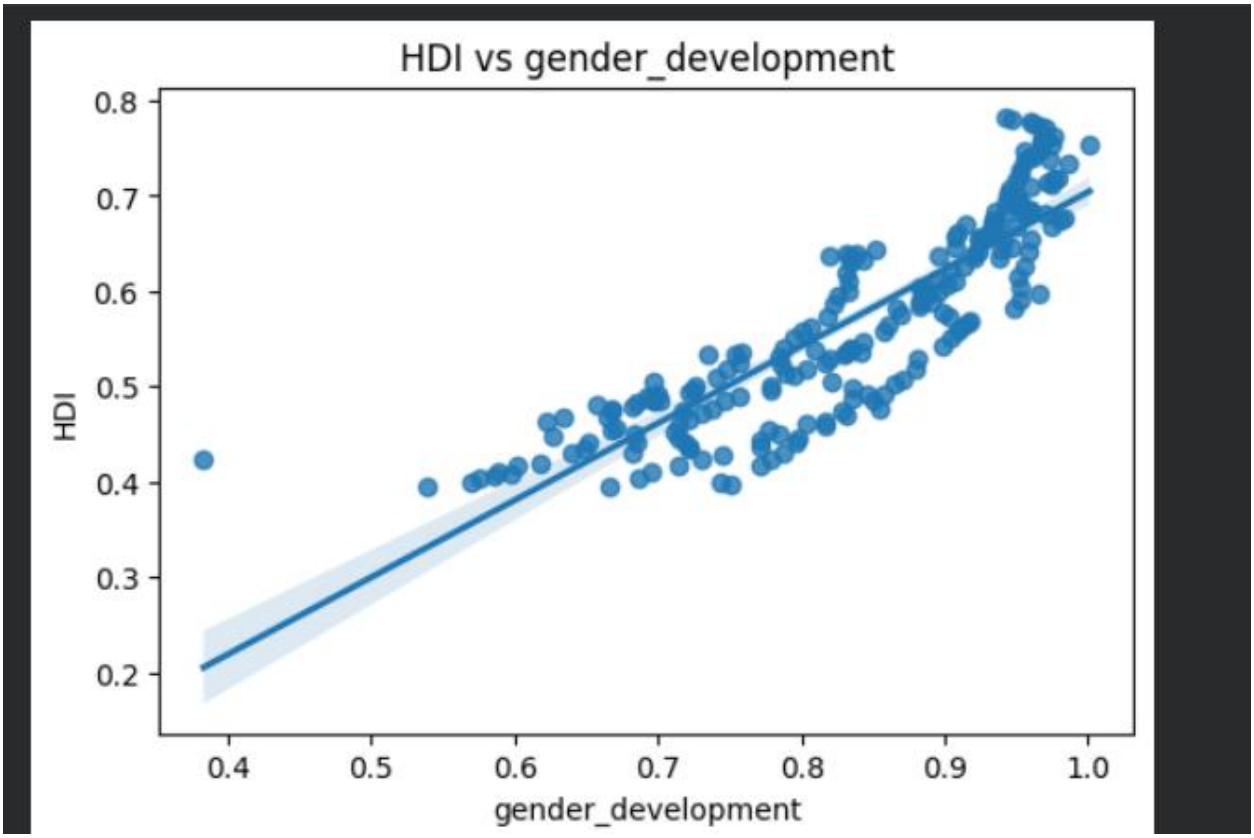
the importance of outlier analysis in regional assessments, as it helps uncover hidden disparities and challenges that may not be visible through averages alone. Moreover, it emphasized that economic indicators must be interpreted in context, and that high income does not automatically guarantee high human development. Overall, Task 3 demonstrated how combining statistical techniques with visual analytics can enhance the depth and clarity of socioeconomic analysis.

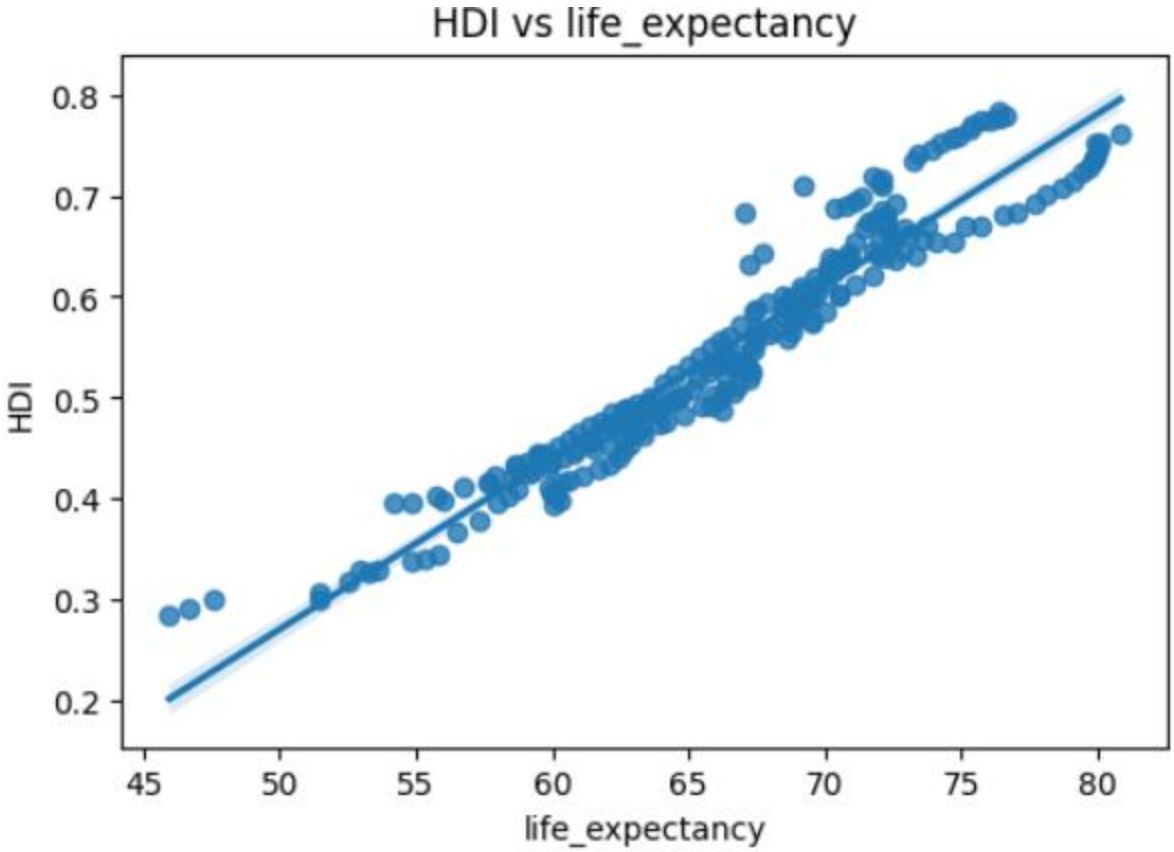


Task 4: Metric Relationships and Regression Visualization

orrelation was used in this exercise to determine the relationship between the Human Development Index (HDI) and two key ones: Gender Development Index and Life Expectancy. These measures were chosen as they capture the major aspects of social equity of human development and health and are the key points on which the differences and the improvement are observed throughout South Asia. It was found that HDI and Life Expectancy have a very strong positive association (r = 0.959), which proves that better health outcomes are directly related to the general human development. The

correlation between the level of life expectancy and the level of HDI was significant, and it was always observed that countries with good life expectancy had greater HDI values, which further supports the role of healthcare access, quality, and longevity in development outcomes. Relationship with Gender Development Index was not as strong and yet statistically significant ($r = 0.866$), indicating that gender equality also plays a role in human development but the development in this regard is not evenly spread across the region. The relationship between these metrics using scatter plots with regression lines were visualized to ensure their confirmation. The plot with the first title being HDI vs gender development represented the data points as country year entries and the scores of the gender development in the x-axis and the HDI in the y-axis. The data were fitted into a regression line with a shaded confidence interval. The trend of the plot was clearly an upward trend that indicated that gender development was also becoming better thus HDI was inclined to go up. Nevertheless, the dispersion of the data points showed that the nations having lower gender development index tended to fall on the lower part of HDI, which signifies that there are still gender inequalities. This presentation underscored the necessity of specific policies that can be used to fix gender inequality as a hindrance to sustainable development. The second plot, HDI vs life expectancy, demonstrated that there was even a closer clustering of the data along the regression line and this demonstrates the strength of the correlation. The values of life expectancy were indicated in the x-axis with values between 45 to 80 years and the HDI values indicated in the y-axis. The positive gradient of the regression corroborated the fact that the higher the life expectancy of the country, the higher the HDI score. This story supported the fact that human development in South Asia can best be predicted through health outcomes. The low

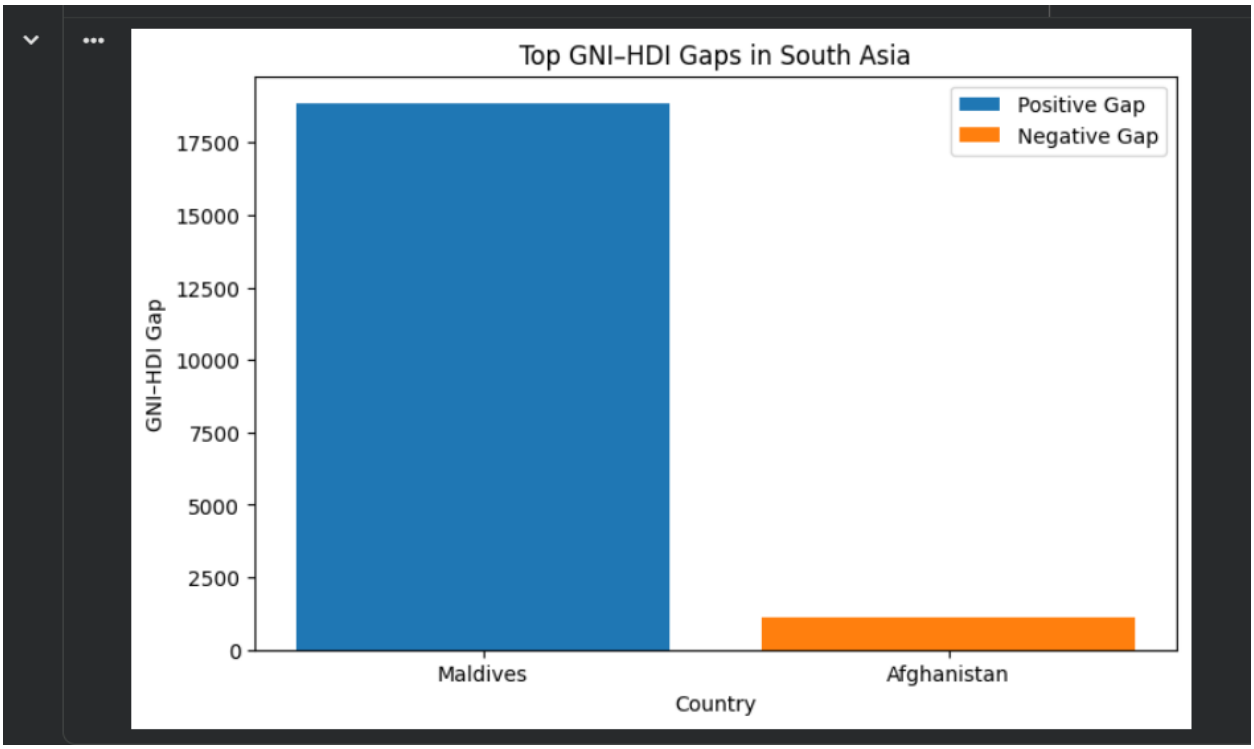

HDI vs gender_development
HDI
gender_development
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.4
0.5
0.6
0.7
0.8
0.9
1.0



Task 5: Gap Analysis and Visualization

The last activity in this exploration was centered on establishing inequities between the economic performance and the human development outcomes using a Gap Analysis. It was done by the computation of the difference between the Gross National Income (GNI) per capita and Human Development Index (HDI) of each of the country-year observations in the South Asia subset. The resulting measure, entered as GNI_HDI_Gap was used as a measure of the extent to which the income levels were turned into greater development successes of a country. A large value of the positive gap implied that a country was much higher in income than what its HDI may indicate, whereas a negative gap indicated that income was not high enough

to give expected development results. Analysis found that Maldives had the largest positive gap i.e. its GNI per capita was higher than HDI disproportionately. This implies that Maldives has a good economic performance; however, its HDI is not entirely accurate due to its income advantage- perhaps because it has weak education, gender equality, or other social balance indicators that cannot be measured by income alone. At the other extreme, Afghanistan had the worst gap depicting that its income levels were way below the threshold that would sustain higher human development. This is an indication of the current poverty, war and inadequate availability of basic services like health and education in the country. To display these findings in form of a bar chart, a bar chart was created with the title Top GNI-HDI Gaps in South Asia. The chart used two countries on the x-axis; that is, Maldives and Afghanistan and the corresponding GNI-HDI gaps on the y-axis. A tall blue bar was the Maldives which showed a large positive gap whereas short orange bar was a representation of a negative gap as was the case with Afghanistan. The visualization was easy to grasp and effective because the color-coded legend was clearly defining between positive and negative gaps. This chart was effective to bring out the differences between the two countries as well as bring out the larger theme of imbalance in terms of income and development. In addition to these extremes, the gap analysis also showed that the moderate income countries like Pakistan and Bangladesh had relatively lower level of HDI score. This implies that there are inefficiencies in transforming economic resources into real gains in education, healthcare and social equity. These results suggest a role of governance, effectiveness of policies, and equal distribution of resources in the development outcomes. Economic growth of a country should be coupled with wise investments in human capital to make sure that the growth in income can be translated to significant increase in quality of life. Altogether, Task 5 helped to underline one important lesson: economic prosperity in itself is not a prerequisite to human development. The discrepancy between the GNI and HDI can be used as a diagnostic measure of what areas development activity might be failing, and why a policy response is most necessary. The synthesis of both quantitative analysis and visual narration made this task an interesting story on the structural challenges and opportunities of countries in South Asia in striving to achieve inclusive and sustainable development.



Overall Interpretation

The task based analysis of South Asia shows an intricate and unbalanced development image. Maldives has always been on the forefront of the region and the indicators of high health and income propel its high Composite Score. Sri Lanka and Bhutan are doing well too, which is the indicator of equal development in several directions. India and Bangladesh are improving steadily but with low scores in comparison with regional leaders. Afghanistan always is ranked least, which supports the effect of conflict, poverty, and inaccessibility to the necessary services on human development. The outliers analysis reveals the necessity to analyze abnormalities because the GNI patterns of Sri Lanka indicate structural imbalances. The close relationship between HDI and life expectancy proves that the health outcomes represent a highly important determinant of the human development in the South Asian region, and the less significant relationship with gender development indicates

persistence of inequalities. The gap analysis also highlights that economic growth is inadequate and that the income gains should be converted into financial gains in terms of education, health, and gender equality in the countries. Generally, South Asia is a region of resilience, inequality, and opportunity where certain countries become progressively developed, whereas others have no possibility of returning to their historical root issues. The results reveal the necessity of specific policies, regional collaboration, and long-term investment in the social sector to help overcome the gap between economic success and human development.

Problem 3

Comparative Regional Analysis: South Asia vs Middle East

1. Create Middle East Subset:

Preparation of the dataset to be used in the comparative study of South Asia and the Middle East was the first step of Problem 3. To make the analysis concentrate on the trends in development in the last years, the global HDI dataset was narrowed down to the years between 2020 and 2022. This period was selected as it includes the most recent data, as it is possible to see the effect of the recent world events (including the COVID-19 pandemic) yet to make the year-to-year comparisons. After the dataset had been limited to this period, two regional subsets were generated. In the case of South Asia, this was Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka. In the case of the Middle East the subset consisted of 14 countries which were Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Syria, United Arab Emirates and Yemen. All the subsets were clearly designated with a new column named region, where the values of South Asia and Middle East were respectively written. The significance of this labelling was that it enabled the two datasets to be combined in the future and also could maintain the regional differences. This was a necessary task as it put the base of

all further comparisons. The analysis was also made clear and reproducible because the dataset was filtered carefully and explicitly defined regions were considered. It also ensured that it was possible to directly compare the HDI values, distributions and trends between South Asia and the Middle East without confusion. Concisely, Task 1 converted a massive international dataset in two clean, well-organized regional subsets, which preconditioned further comparative analysis of the next tasks.

2. Descriptive Statistics:

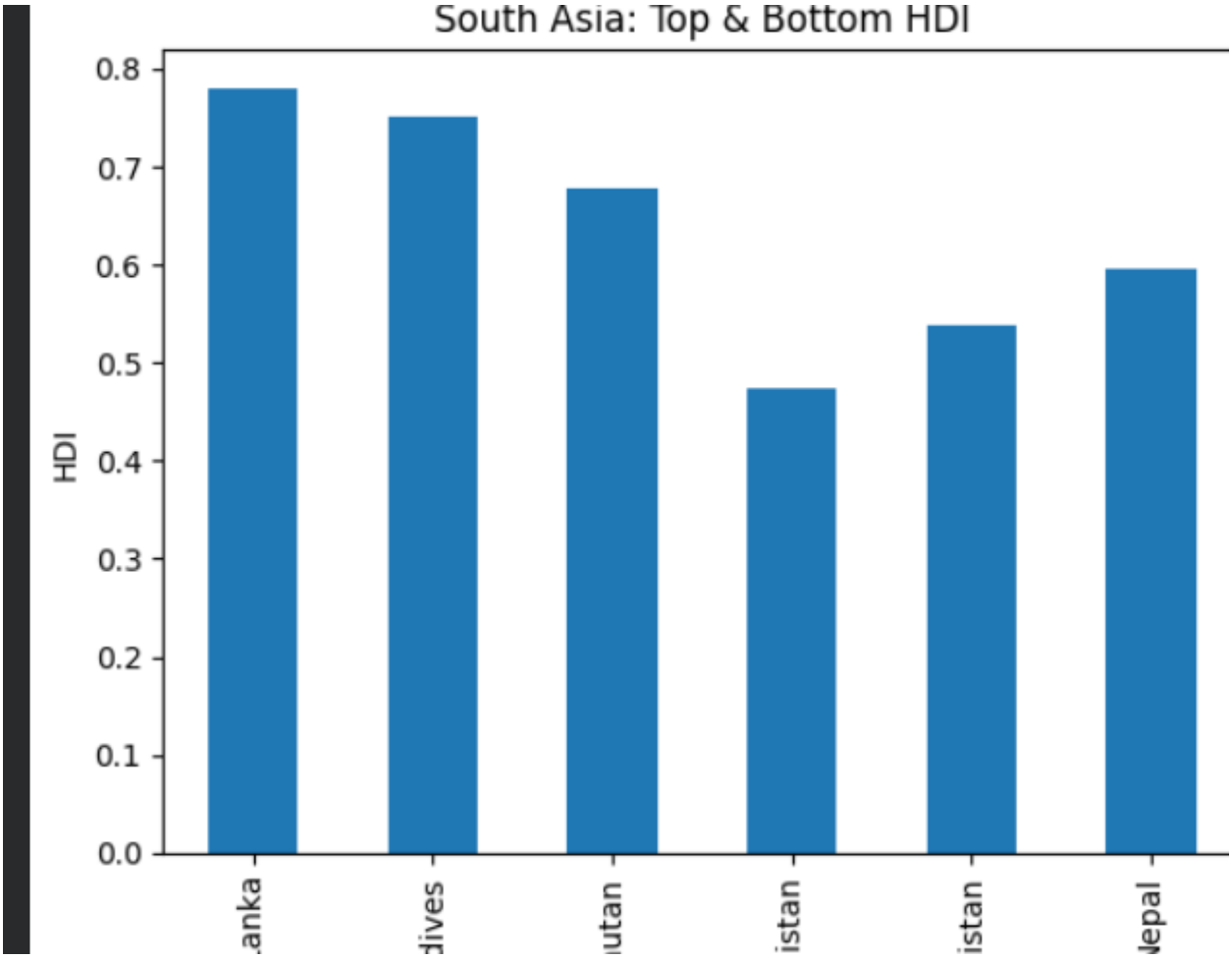
The comparison of the general Human Development Index (HDI) levels of South Asia and the Middle East in 202022 became the subject matter of the analysis in this task, which shifted the preparation of regional subsets to the overall level. The mean and standard deviation of HDI values were used to determine the mean and standard deviation of HDI in every region using the pandas agg () function. The central tendency was represented by the mean HDI; they gave an average level of human development of the countries in the region, whereas the standard deviation was a measure of variability; it was used to indicate the similarity of levels of development of countries within the region or the level of development was highly unequal. Such statistical method was significant as it enabled the analysis to shift away from the analysis of individual countries, but instead show regional trends. On an example a mean HDI of the Middle East being higher than the South Asian would imply

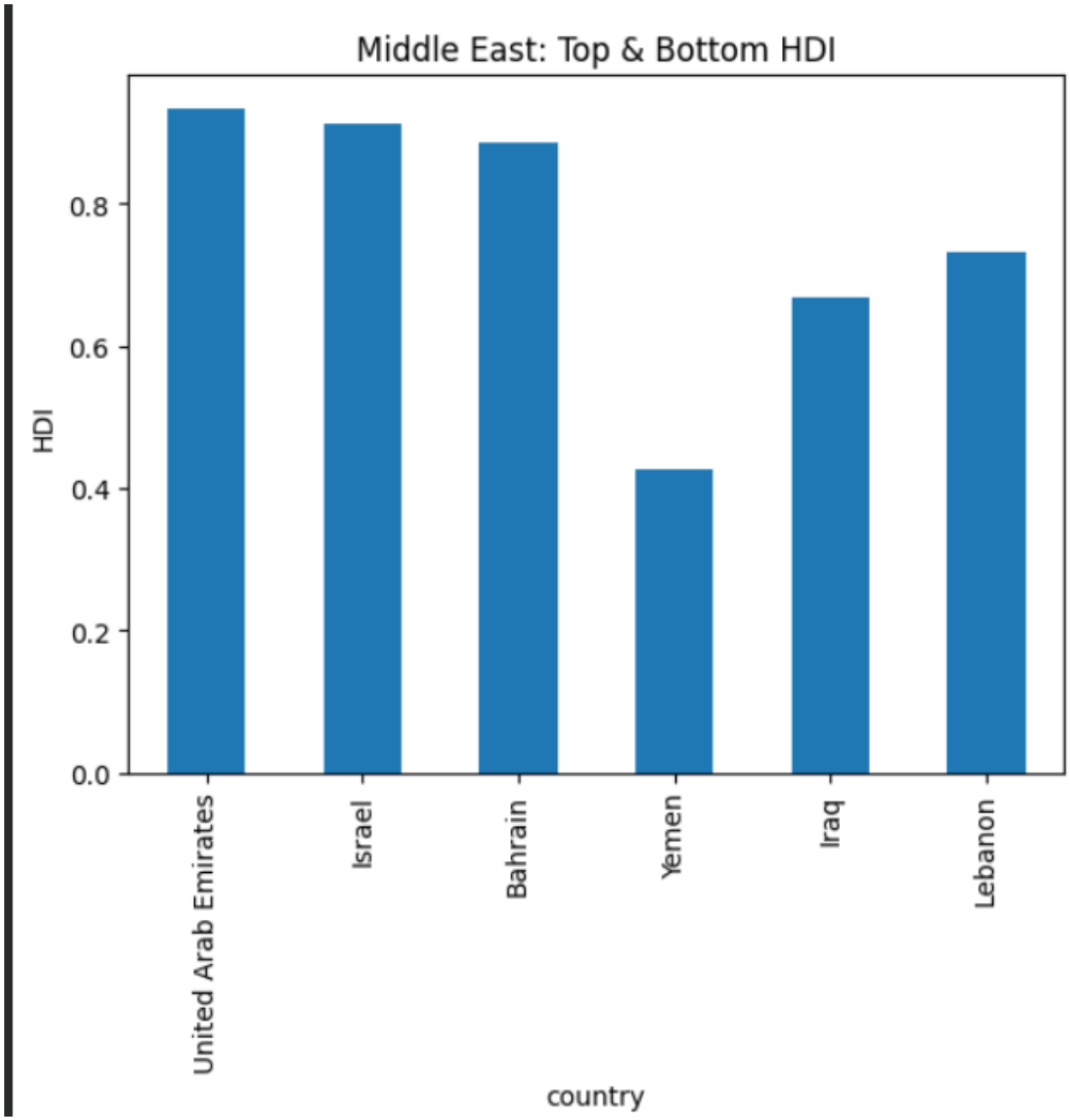
better overall performance by the Middle East whereas having higher standard deviation of the South Asian would imply more disparities in the countries of the South American region with some countries being fairly better than others. The calculations and the resultant comparison of these values resulted in Task 2 giving a clear quantitative background on which these regions could be compared and thus the eventual understanding on whether the presence of differences between the South Asia and the Middle East was owing to general levels of development or internal differences within the regions. This move provided a basis of further analysis in subsequent assignments, including the analysis of distributions, outliers, and correlation to other indicators.

3. Top and Bottom Performers:

This task was aimed at analyzing the internal distribution of HDI in each of the regions and determining the best and worst performers. In the first section of the code, the average HDI per nation during the years 20202022 was computed with the use of groupby() and mean() functions. In particular, the average HDI of each country in South Asia was calculated as sa_avg = south_asia_df.groupby('country')['hdi'].mean() whereas the Middle Eastern countries were computed in a similar way as me_avg = middle_east_df.groupby('country)['hdi].mean() The averages were then sorted to get the first three and bottom three countries in every region using sort values and head 3. This ranking gave a clear picture of the countries that were either doing well or doing poorly in the development of humans in their respective regions. In order to have a visual representation of these rankings, bar charts were created with matplotlib. The title of this first chart was South Asia: Top and Bottom HDI, which showed the HDI of the three highest and three lowest performing countries of South Asia. The countries were placed on the x-axis and the y-axis took the values of HDI. The chart revealed that the HDI scores were the highest in Sri Lanka, Maldives and Bhutan and Afghanistan, Pakistan and Nepal occupied the lowest end. The second chart, the Middle East: Top and Bottom HDI was structured in the same

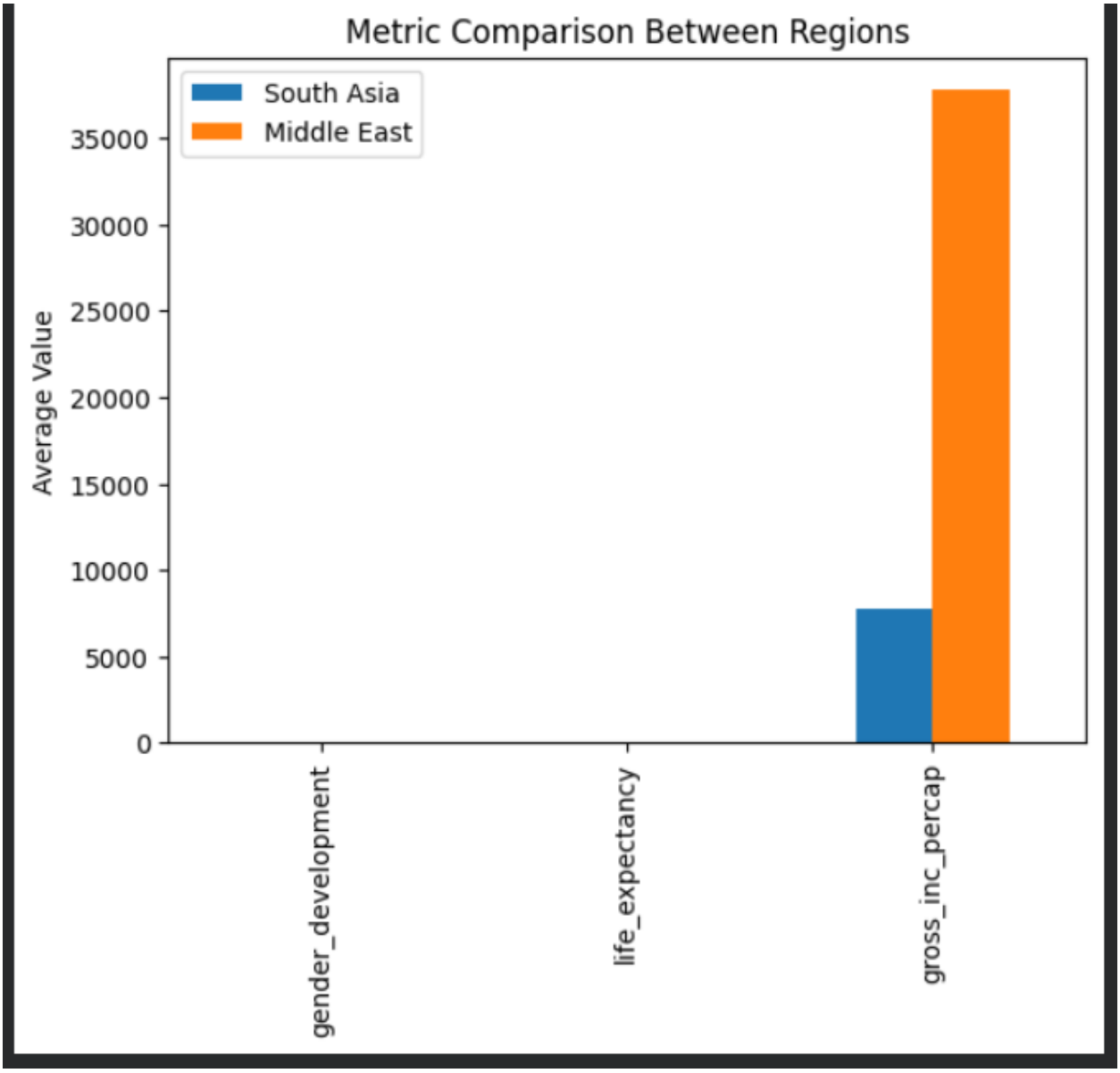
way and showed that United Arab Emirates, Israel and Bahrain, were the most performing countries with the lowest HDI values in the area with Yemen, Iraq and Lebanon on the lowest position. These visualizations played a significant role in bringing out the differences in each region and are an easy way of comparing the development outcomes. On the whole, Task 3 provided me with statistical and visual information about the distribution of HDI in the countries of South Asia and the Middle East. The analysis that was based on identifying the extremes and placing them against each other did not only determine the countries that are doing well or doing bad, but also the level of inequality in each region. This preconditions further reading in further tasks, e.g. outlier detection and correlation analysis.





4: Metric Comparison Between Regions

This exercise was to analyze the major development indicators of the south Asia and the middle east with an aim of understanding the greater regional differences of the region in question not only in HDI. The three measures that have been chosen are the gender development, life expectancy, and gross income per capita, all of which indicate the social equity, health outcomes, and economic strength. The code was used to compute the average of each measure in both regions using pandas. In particular, it is computed that South Asia is the mean of the values of the metrics in sa/mean south/asia/df[metrics].mean() and Middle East is the mean of the values of the metrics in me/means = middle/east/df[metrics].mean. Such findings were then assembled into a new DataFrame known as comparison_df to arrange the metrics side by side allowing comparison directly. A bar chart was created in order to visualize these differences and this chart was named as Metric Comparison Between Regions. The chart showed all the metrics in the x-axis, average values in the y-axis, and color-coded bars to differentiate the ones in South Asia (blue) and Middle East (orange). It was seen in the visualization that the Middle East had gross income per capita on average that was much higher which means that the Middle East shows good economical performance. Conversely, the gender development and life expectancy bars seemed equal in terms of the height indicating that both regions have equal results in the former or that the difference was insignificant to be seen in the existing scale. This chart was especially instrumental in pointing out the fact that economic advantage does not necessarily translate into the best performance in all the development dimensions. In sum, Task 4 offered a simple and intuitive way of comparing strengths and weaknesses of the regions, which supports the significance of considering several indicators to examine the human development.



4. HDI Disparity:

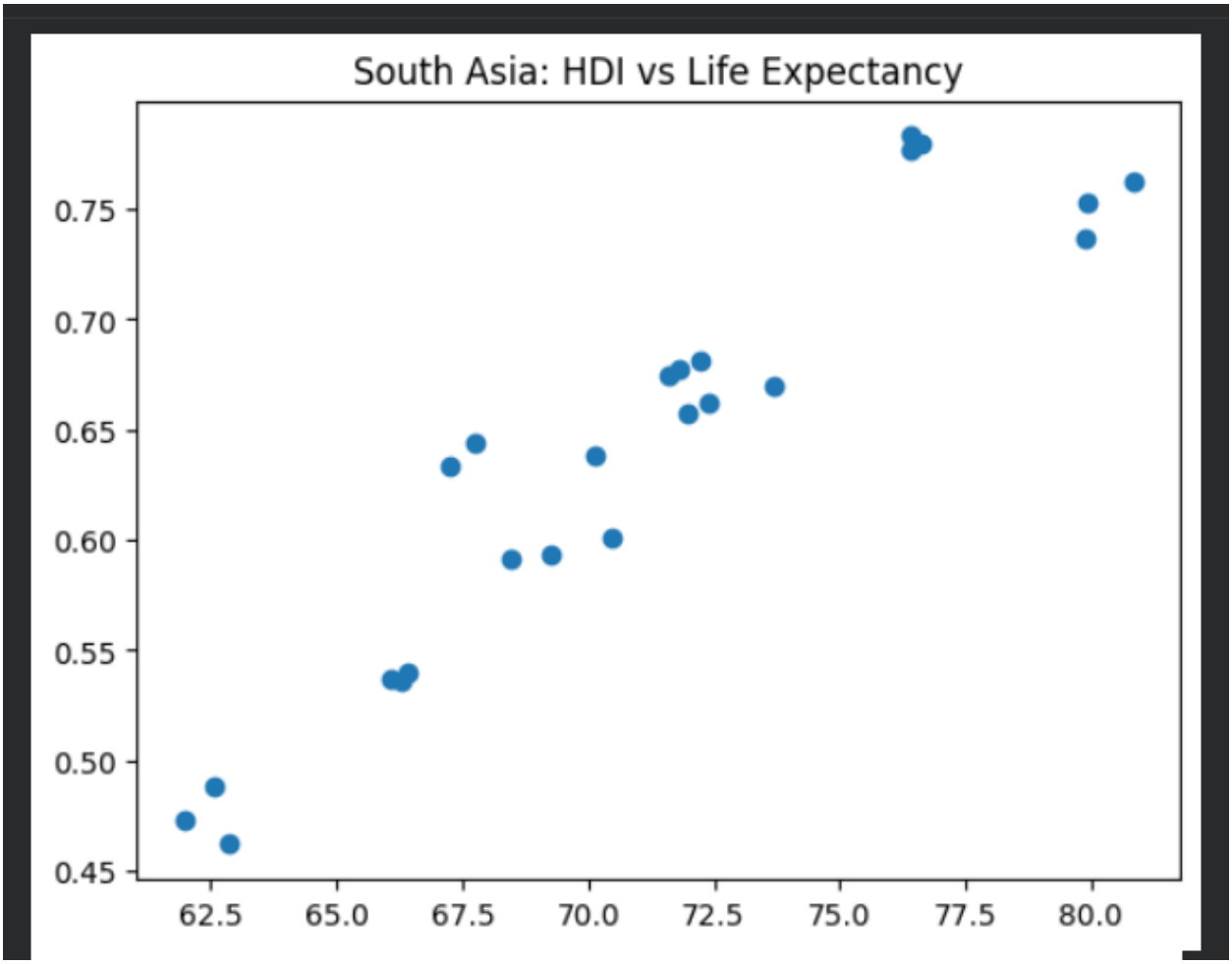
The code employs statistical estimates for the analysis and comparison of the level of variation of HDI value between South Asia and Middle East regions. It begins with calculating the range between HDI value for each region. It calculates it by determining the difference between the maximum HDI value for each dataset (south_asia_df dataframes and middle_east_df dataframes). It calculates it by subtracting the minimum HDI value from the maximum value for each dataset. The value obtained gives an estimate of diversity between maximum levels of human development in each region. After that, it calculates the Coefficient of variation for each region based on the division of standard deviation for HDI value for each region by its mean HDI value. Coefficient of variation gives an estimate of relative diversity for an element or group of elements being measured in comparison to its average value. Its calculation gives an estimate of diversity similar to that obtained from calculating standard deviation but is more significant for

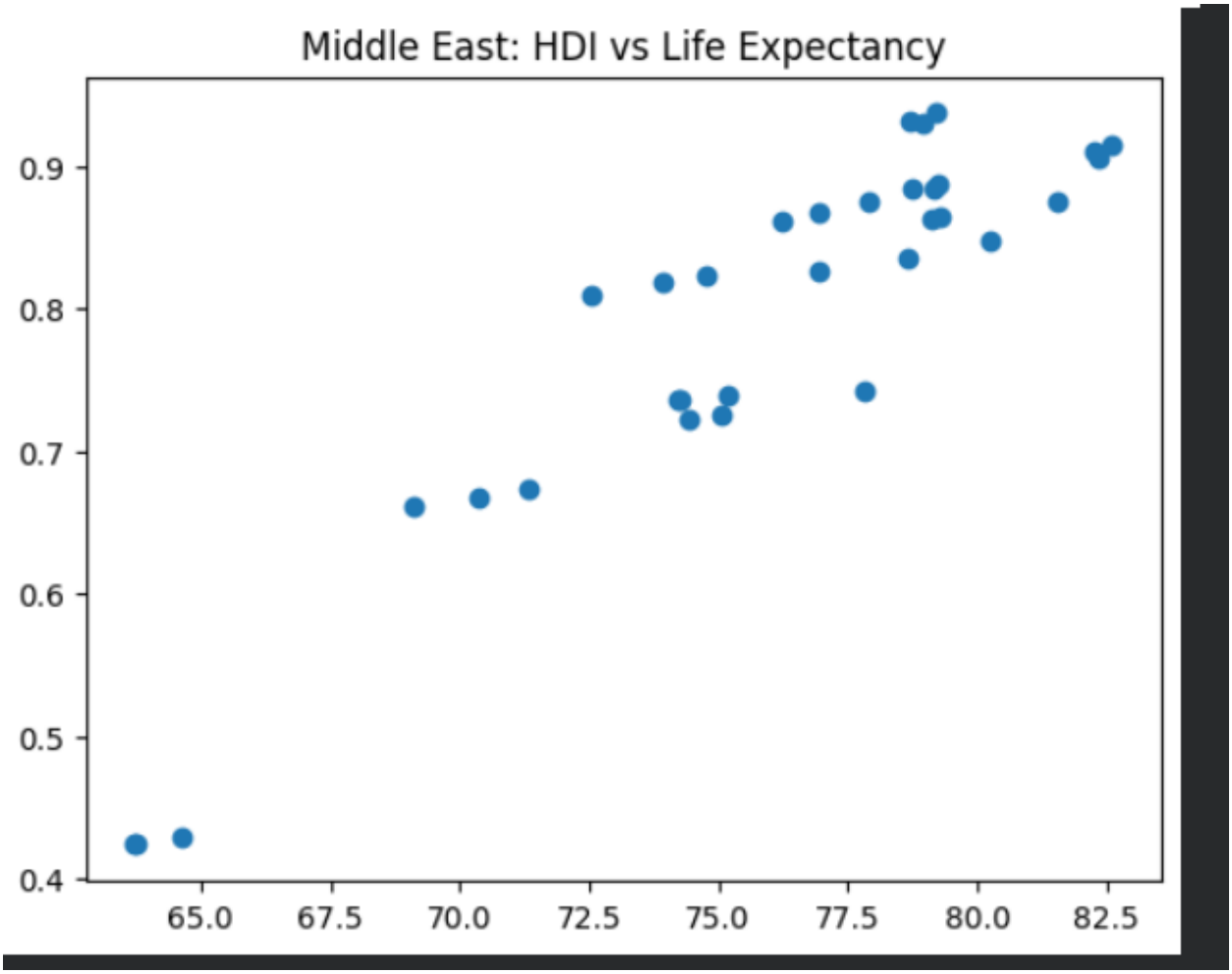
comparative assessment between two regions which are most likely not equal in every detail such as human development level for regions under assessment. After completing these calculations, it applies conditional logic based on 'if/else logic statements' for the algorithm helps identify areas with higher levels of inequality or dispersion of HDI values using a step-by-step comparison process, which provides an indication of how human development is spread between the two regions.

Correlation Analysis:

In Task 6, we examine the correlation of the Human Development Index (HDI) and gender development with life expectancy for the regions of South Asia and the Middle East. This is achieved through the use of correlation analysis and scatter plots. The code for this task calculates correlation matrices for the variables of choice using the `.corr()` function. The correlation matrices show the strength and direction of linear correlation for the variables of HDI, gender development,

and life expectancy. The closer the correlation coefficients are to +1, the more strongly they are positively related. The code for this task uses `plt.scatter()` for the creation of scatter plots for a graphical illustration of the relationship between life expectancy and HDI for each of the regions of South Asia and the Middle East. Each data point in the scatter plot represents a country. Roles of life expectancy and HDI are reversed for illustration of their relationship. The upward trend of data points represents a positive relationship between life expectancy and HDI. The combination of the above statistical methods provides a rich interpretation of how life expectancy positively or negatively affects the development of humans in the two regions.


South Asia: HDI vs Life Expectancy
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



Outlier Detection:

The emphasis of this section of the Comprehensive Report is to identify and compare outliers in the Human Development Index (HDI) and Gross National Income (GNI) per capita for South Asia and the Middle East. This is achieved by defining a function named detect_outliers, which employs the Interquartile Range (IQR) method, a widely used statistical tool for identifying outliers. The detect_outliers function accepts any variable as input and produces the first quartile (Q1) and

the third quartile (Q3) of that variable, along with the IQR, which represents the distance between Q1 and Q3. Subsequently, it computes the lower and upper bounds for any outlier by calculating Q1 - 1.5 times the IQR and Q3 + 1.5 times the IQR, respectively.

New columns were established for both the South Asia Dataset and the Middle East Dataset to indicate whether each country's HDI and GNI per capita for their respective regions are classified as outliers based on the detect_outliers function. This facilitates a structured comparative analysis of the developmental extremes or outlier characteristics of the regions. Furthermore, the total count of HDI and GNI per capita outliers for each region was computed, and each individual country exhibiting unusually high or low GNI per capita was identified. The initial lines of both the South Asian Dataset and the Middle Eastern Dataset, along with the outlier identifiers, illustrate the distribution of extreme values across the two regions.

References

- United Nations Development Programme (UNDP). *Human Development Report 2023/2024: Reimagining Cooperation in a Polarized World.* New York: UNDP.
-
- UNDP Human Development Index Data Center. *Human Development Index (HDI) Dataset and Methodology.*
-
- Security Risks Monitor. *South Asia Human Development Index 2023–24 – A Review.*
- World Population Review. *Human Development Index (HDI) by Country 2023.*
-
- UNDP. *Human Development Report 2021/2022: Uncertain Times, Unsettled Lives.*

CONCLUTION:

rt examined the evolution of human development in South Asia and the Middle East from 2020 to 2022, emphasizing both general advancements and disparities among countries. The analysis reveals a distinct difference between the two regions by integrating statistics, visual comparisons, and regional benchmarks.

Generally, Middle Eastern countries usually achieve higher scores on the Human Development Index (HDI) and possess greater income per capita compared to nations in South Asia. Nevertheless, South Asia exhibits significantly greater inequality. The region consists of relatively strong performers like Sri Lanka and the Maldives, alongside nations such as Afghanistan, where development results are still very low. This significant disparity underscores the inconsistent advancement throughout South Asia.

The study additionally discovered that life expectancy is the key factor influencing HDI scores in both areas. This emphasizes the crucial significance of access to high-quality healthcare and public health services. Gender development was another crucial element affecting human development; however, advancements in this field have been inconsistent—particularly in South Asia—hampering lasting and sustainable progress.

Additional comparisons showed that economic growth alone does not guarantee improved human development. For instance, the Maldives has comparatively high income levels, yet this does not entirely result in similarly high human development outcomes. Conversely, Afghanistan's extremely low income levels are strongly associated with its inadequate HDI performance, illustrating the profound connection between economic and social issues.

In summary, these results highlight that human development involves more than just income. Successful policies should adopt a balanced strategy—enhancing healthcare, education, gender

equality, and governance while promoting economic growth. In South Asia, addressing development disparities and enhancing equitable access to resources are critical priorities. For the Middle East, the difficulty is in sustaining economic advancement while making sure that growth is advantageous for all segments of society.

https://github.com/karkisandhya2061-boop/2547175_Sandhyakarki.git