Concepts and Technologies of AI (5CS037)

Report

Statistical Interpretation and Exploratory Data Analysis.

Ritesh Khatri
2548382

Table of Contents:

INTRODUCTION :

The Human Development Index (HDI) is a way of comparing the development levels of various nations. The HDI mainly depends on three important features of human life. These features are health, education, and income.In the calculation of HDI, health is estimated through the life expectancy of citizens. Education is estimated through the average years of instruction received by the population. On the other hand, income is estimated through Gross National Income (GNI) per capita.

Dataset used in this review covers the years 1990 to 2022. It includes data for many countries where HDI is available. The HDI value is calculated using all three components together, so if one value is low, it affects the final HDI score.

This dataset helps us study how countries have developed over time. It also helps compare regions and find countries that perform better or worse than expected. Even if two countries have similar income, their HDI can be different because of differences in health and education.

Single Year HDI Exploration:

– Methods / Approach:

1. Made new data frame containing data of year 2022.
2. Checked data types, missing values, duplicates, and inconsistencies.
3. Filled missing data using media.
4. Calculated basics statistics
5. Filtered countries have HDI greater than 0.800.
6. low, medium, high, very high category is added to the country depending upon their HDI value.

– Key results

1. Switzerland had highest HDI in 2022 (0.967)
2. Somalia had lowest HDI in 2022 (0.38)
3. Found countries have high, low, medium HDI range.
4. Traced countries associated with greater HDI than 0.800.

– Visualizations and tables

countries associated with greater HDI than 0.800.

	country	HDI_Category	hdi
3332	Liechtenstein	Very High	0.942
4718	Qatar	Very High	0.875
5213	Singapore	Very High	0.949
2705	Ireland	Very High	0.950
3398	Luxembourg	Very High	0.927
6104	United Arab Emirates	Very High	0.937
5609	Switzerland	Very High	0.967
4322	Norway	Very High	0.966
6170	United States	Very High	0.927
2474	Hong Kong, China (SAR)	Very High	0.956

Fig: HDI>0.800

– Interpretation and discussion

Countries consisting higher HDI in general had higher earning , which was not the only case; some performed well and some performed bad even with high income, indicating there is a crucial role of education and health as well.

Problem 1B :

HDI Trend Analysis (2020–2022)

Methods / Approach:

1. Extracted the required data to include only the year 2020 to 2022.
2. Missing values are handled with media because it is robust for outliers.
3. Various visualization techniques, including line charts, bar charts, box plots, and scatter plots, were used to analyze the data.

– Key results of the analysis

For countries India, Nepal, Bangladesh, Pakistan, Sri Lanka: Sri Lanka had more HDI than other countries, while Pakistan had lowest. Nepal however lied in middle of the countries given.

Region-wise analysis shows that Europe and Central Asia record the highest HDI, whereas Sub-Saharan Africa exhibits the lowest HDI. GNI per capita and HDI has positive correlation.

– **Graphs and Output Tables**

A. Line Chart:

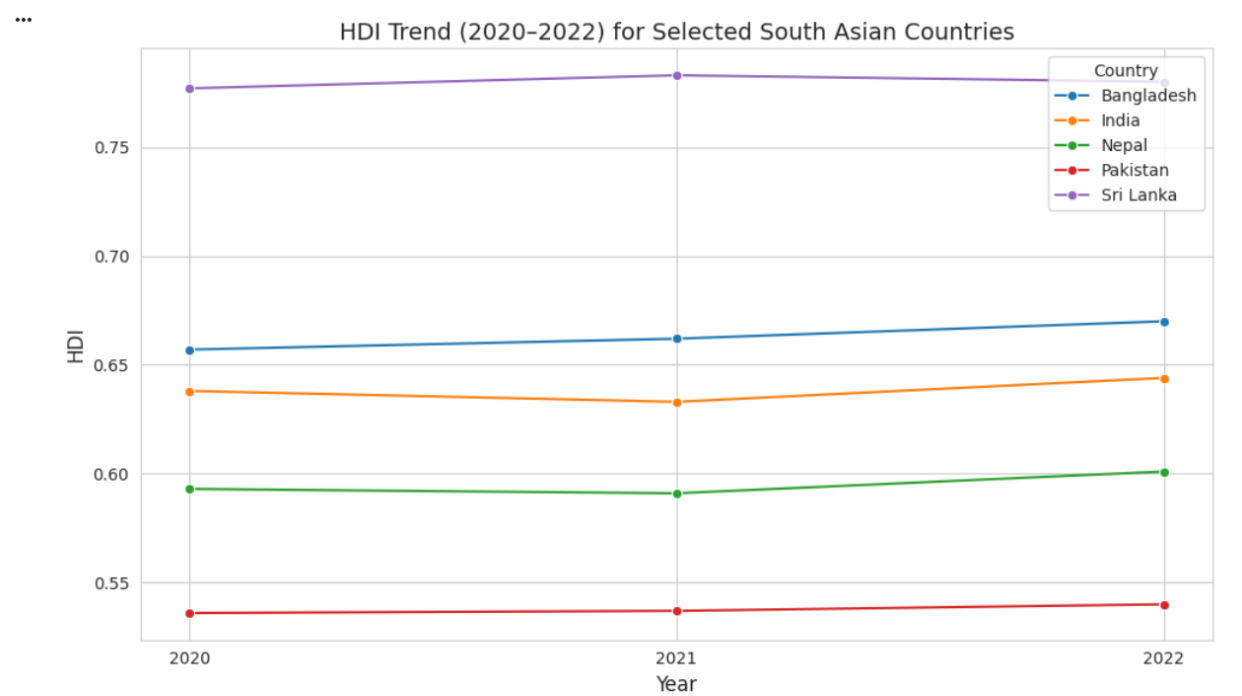


Figure 1: HDI trends from 2020 to 2022 for India, Nepal, Bangladesh, Pakistan, and Sri Lanka.
The chart shows human development progression differences across these countries over three years.

Given line chart shows HDI data of countries India, Nepal, Bangladesh, Pakistan, Sri Lanka, from year 2020 to 2022.

From the graph, Sri Lanka had more HDI than other approximately hovered HDI value of 0.77., while Pakistan had lowest around 0.35. Nepal however lied in middle of the countries given.

– Bar Graph

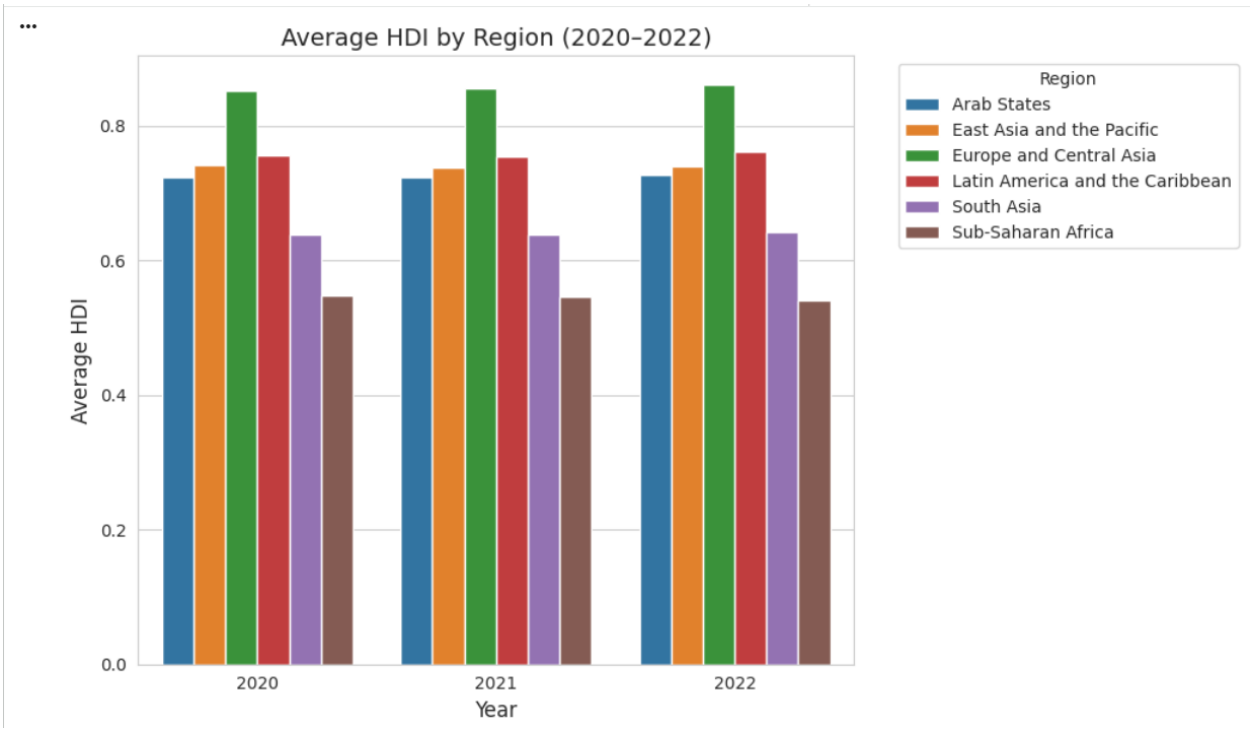


FIG: Bar_Diagram

Given bar chartt points the mean HDI values of six given regions(2020-2022)

Overall, Europe and Central Asia had the highest HDI in all three years, while Sub-Saharan Africa had the lowest. Most regions show a small increase in HDI over time.

Europe and Central Asia stayed well ahead of other regions during the period. Latin America and the Caribbean, along with East Asia and the Pacific, also showed fairly high and stable HDI values. The Arab States remained at a moderate level with only small changes.

South Asia and Sub-Saharan Africa had the lowest HDI scores, showing a clear gap in development compared to other regions. Overall, although there was slight improvement in HDI between 2020 and 2022, regional inequality in human development is still clearly visible

-Box Plot Diagram:

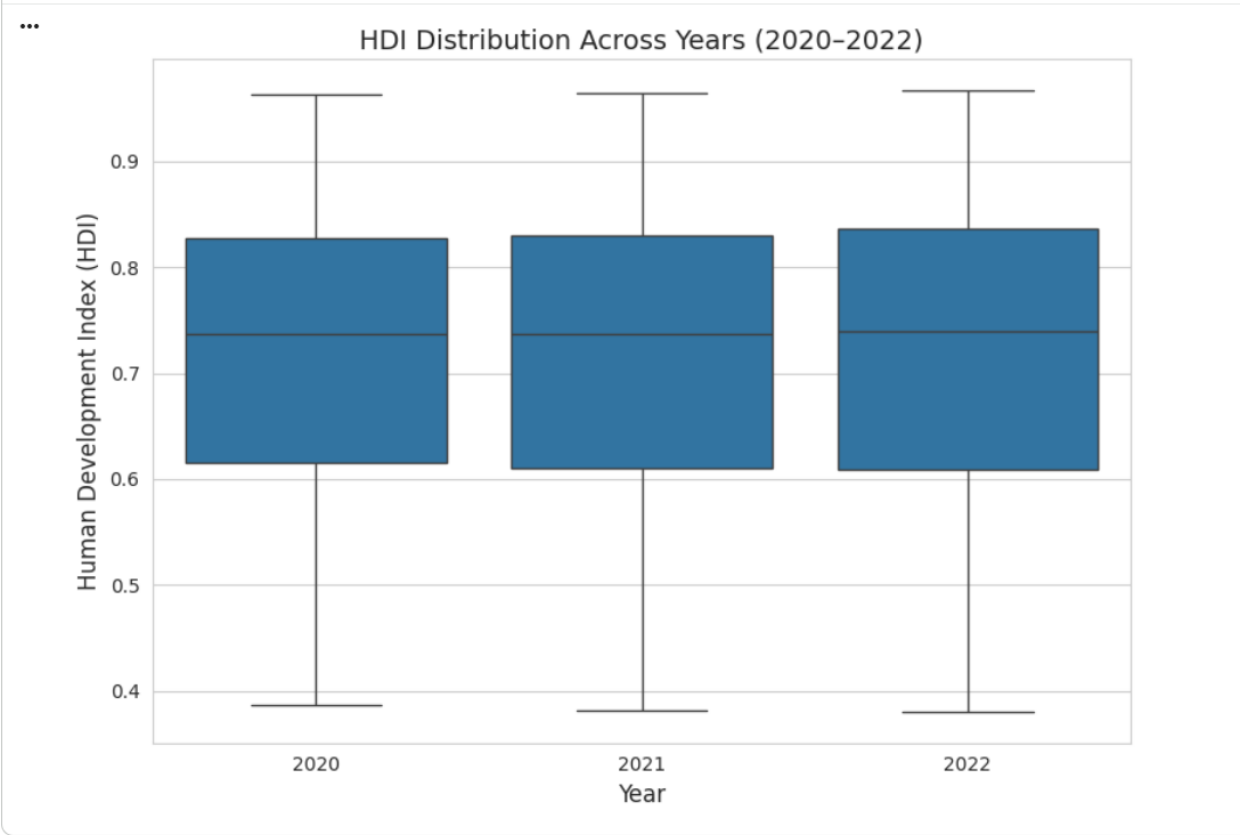


Fig: Box Plot (2020-2022), of six regions.

The box plot shows HDI did not change much from 2020 to 2022. The middle value (median) is about 0.74 for all years. The boxes are same size, so the middle 50% of countries have similar HDI. The lowest and highest values also stay same. The median is a little lower in box, so few countries consist high HDI. There were no outliers found. This tells us global HDI is very stable in these years.

– Scatter Plot: (HDI vs GNI_per_Capita)

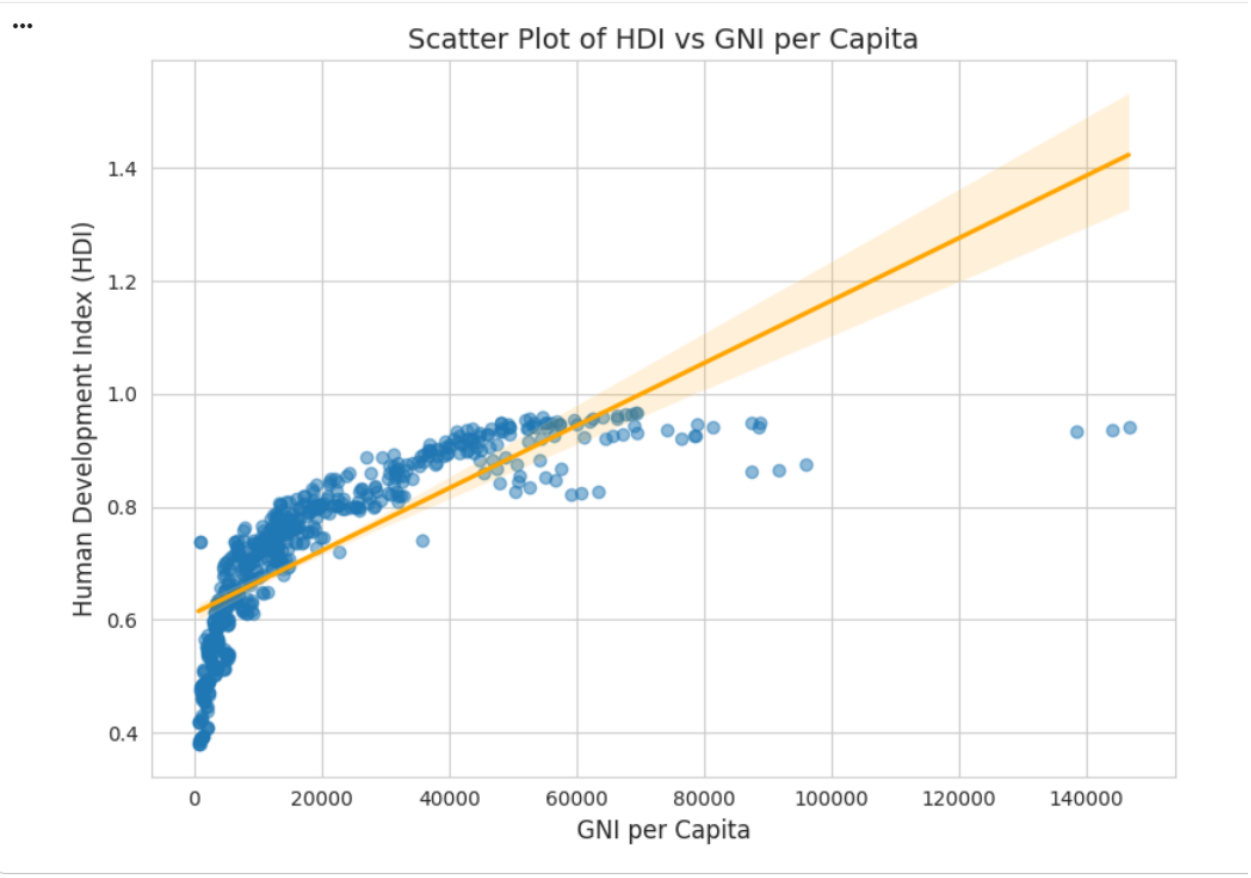


Fig: Scatter(GNI_per_capita vs HDI)

Given scatter plot shows relation of GNI per capita and HDI for countries. most country follow a upward trend, so more income often mean more HDI. but some points far from main group, like maldives or sri lanka, they having very high GNI per capita compare to rest. these point is outlier, it show few country is much richer than most, even if their HDI not so much higher.

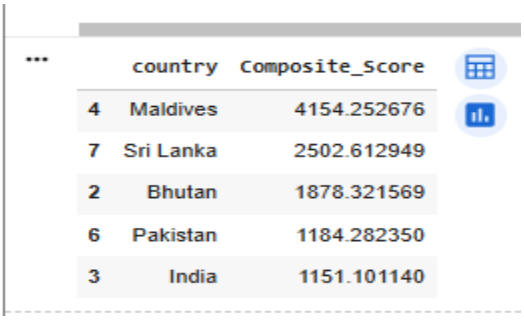
 # Problem 2 – Advanced HDI Exploration

– Methods / Approach

1. Calculated Composite score using the formula:
   Composite Score = 0.30 × Life Expectancy Index + 0.30 × GNI per Capita Index}
2. Ranked South Asian countries based on Composite Score.
3. Made a comparison between composite score and HDI.
4. Detected outliers
5. Pearson correlation of each column with HDI Calculated.

– Key Findings:

1. The countries are ranked from highest to lowest HDI as follows: 1st Maldives, 2nd Bhutan, 3rd India, 4th Pakistan, 5th Sri Lanka



	country	Composite_Score
4	Maldives	4154.252676
7	Sri Lanka	2502.612949
2	Bhutan	1878.321569
6	Pakistan	1184.282350
3	India	1151.101140

2. No outliers were detected in HDI

```
HDI Outliers:
Empty DataFrame
Columns: [country, year, hdi]
Index: []

GNI per Capita Outliers:
       country  year  gross_inc_percap
145   Maldives  2003       14543.14634
146   Maldives  2004       15287.20714
148   Maldives  2006       16391.40757
149   Maldives  2007       14162.42816
150   Maldives  2008       15423.81910
151   Maldives  2009       14050.92438
152   Maldives  2010       14363.23615
153   Maldives  2011       15300.31364
154   Maldives  2012       15355.44089
155   Maldives  2013       15648.36655
156   Maldives  2014       16410.16419
157   Maldives  2015       16414.93039
158   Maldives  2016       16911.45422
159   Maldives  2017       17474.64201
160   Maldives  2018       18055.72465
161   Maldives  2019       18679.92536
163   Maldives  2021       16857.60546
164   Maldives  2022       18846.79219
259  Sri Lanka  2018       13405.80369
```

3. Maldives/Sri Lanka stand out as outliers in GNI per capita. Because their income in that years was high as compared to other countries.

-Visualizations with graph and tables:

Scatter-plot of the GNI_Per_capita vs HDI

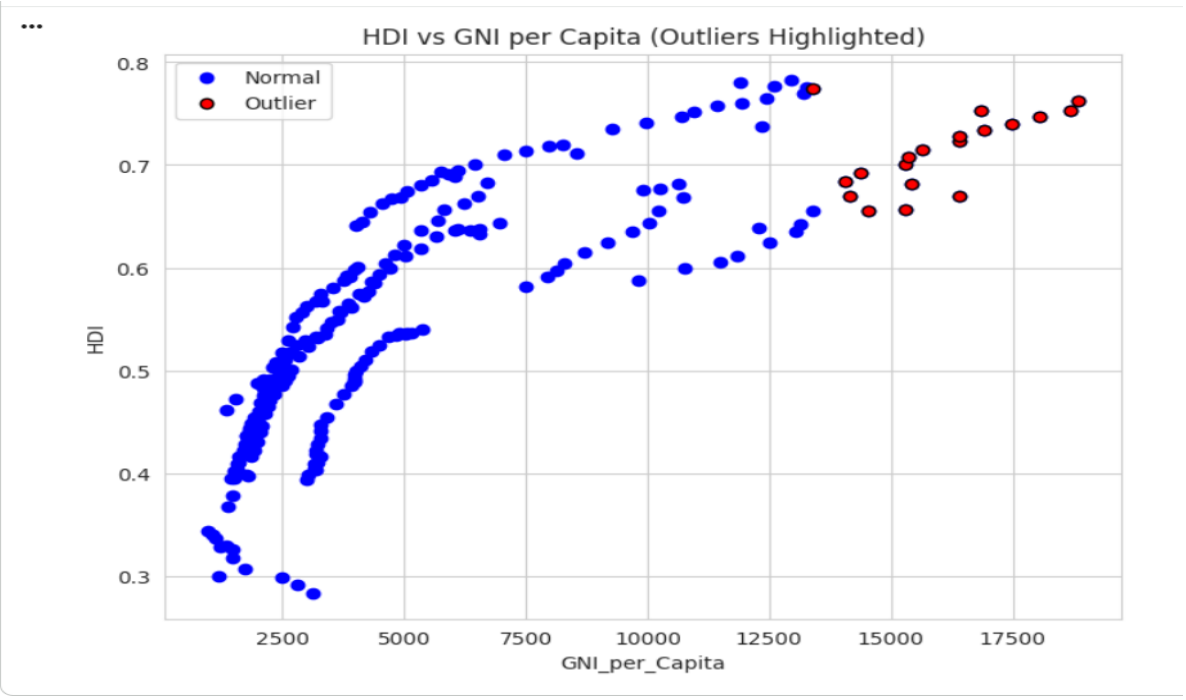


Fig: scatter plot of GNI per Capita vs HDI (South Asia)

Based on the data, we can see a positive linkage of Gender Index with HDI. When a country's HDI score is going up, the gender equality also gets better. This shows that rich nations usually gives better chances for both

men and women. But, some countries still have big gap where they have much money but low gender equality. Overall, the graph prove that we cannot get high development without supporting all gender the same way."

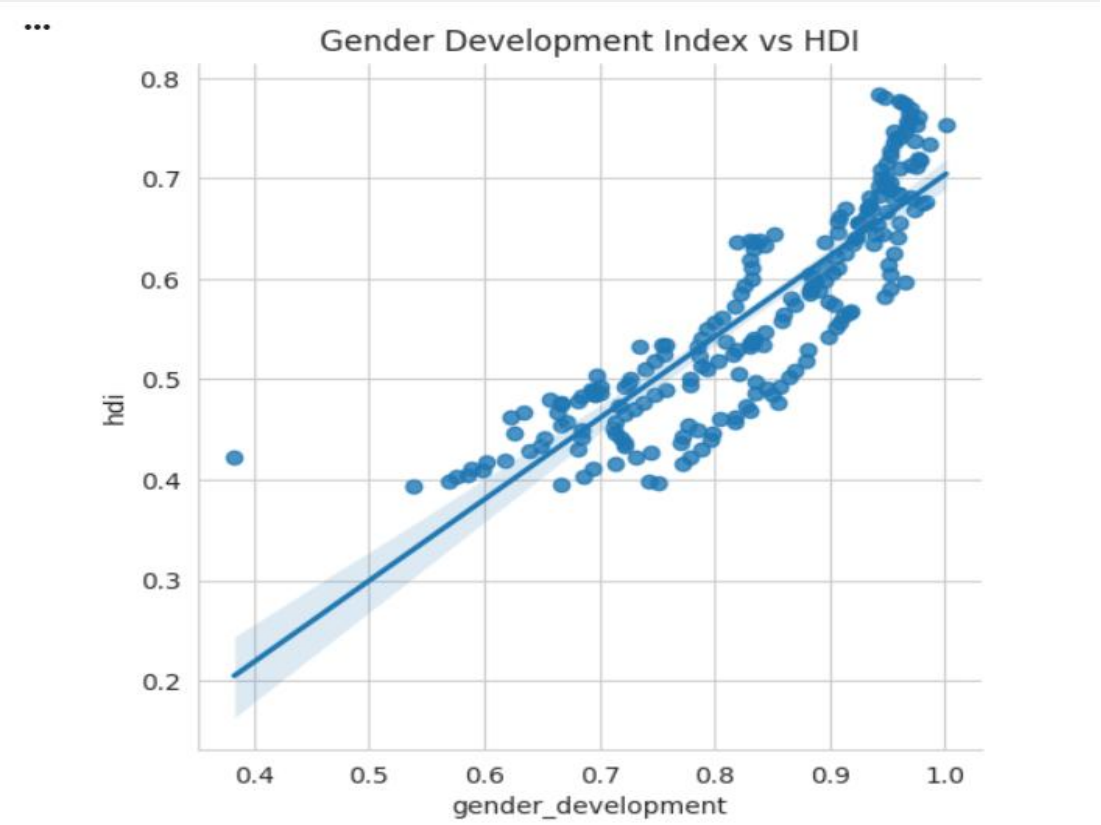


Fig: gender development vs HDI (South Asia)

the first plot show gender development index and HDI. trend mostly goes up, mean when country do better on gender, its HDI also gets better. but some dot not follow close, they have lower gender index but still okay HDI maybe. this say equality not always same as overall development.

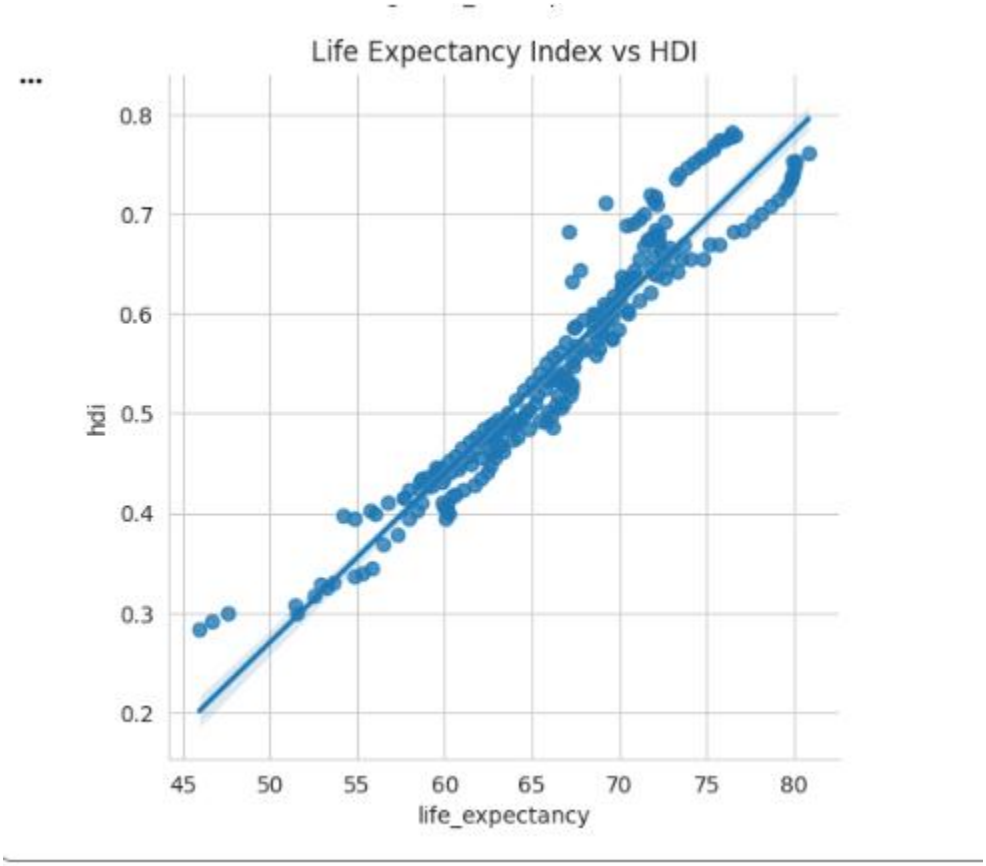


Fig: Life Expectancy Index vs HDI (South Asia)

second plot is life expectancy index versus HDI. dots go from bottom left to top right in strong pattern. it mean people living longer very connect to higher HDI scores. no dots far away from line, so life expectancy is a good predictor for development.

Regional wise Analysis: :

– Methods / Approach

A comparative analysis was conducted between South Asia and the Middle East. First, I make list of countries in each region like instruction. Then I filter main dataset from Problem 1B - the 2020-2022 data - to make two new CSV files: HDI SouthAsia 2020 2022.csv for south Asia region and another is HDI_MiddleEast 2020 2022.csv representing middle east data.
After data ready, I calculate some descriptive statistics in Python with Pandas. I calculated Average(mean) and standard deviation(SD) for metric 'HDI' for two regions. Also I see top 3 and bottom 3 countries in each group to know difference of development. Then I calculate Range and Coefficient of Variation (CV) to see inequality in region.

After that I make grouped bar charts for compare metrics and scatter plots with trendlines using Matplotlib and Seaborn. This to check correlation between HDI, GDI and income.

Key results

HDI increase with gender development in two regions.

HDI and life expectancy has positive correlation to each other.

Average Score: Middle East has better HDI average (0.766) than South Asia (0.639).

```
South Asia - Mean HDI: 0.6395833333333333
South Asia - Std HDI: 0.09827331747496995
Middle East - Mean HDI: 0.766452380952381
Middle East - Std HDI: 0.13910010474666892
```

Best Countries: Top countries in Middle East are UAE, Israel and Bahrain,

but in South Asia top are Sri Lanka, Maldives and Bhutan.
Lowest Countries: Yemen have lowest score in Middle East, and Afghanistan is lowest in region: South Asia.

```
South Asia - Top 3 Countries by HDI
      country       hdi
7   Sri Lanka  0.780000
4    Maldives  0.750667
2      Bhutan  0.677667

South Asia - Bottom 3 Countries by HDI
       country       hdi
0  Afghanistan  0.474333
6     Pakistan  0.537667
5        Nepal  0.595000
Middle East - Top 3 Countries by HDI
                  country       hdi
12  United Arab Emirates  0.932667
3                Israel  0.910667
0               Bahrain  0.885333

Middle East - Bottom 3 Countries by HDI
                  country       hdi
13                  Yemen  0.426333
11  Syrian Arab Republic  0.558667
2                   Iraq  0.667000
```

The Gap: Middle East have bigger gap between rich and poor countries because CV (0.181) more than South Asia (0.153).
Main Difference: Biggest difference between two region is money (GNI per capita), but health (Life Expectancy) is more same.

– Visualizations and tables

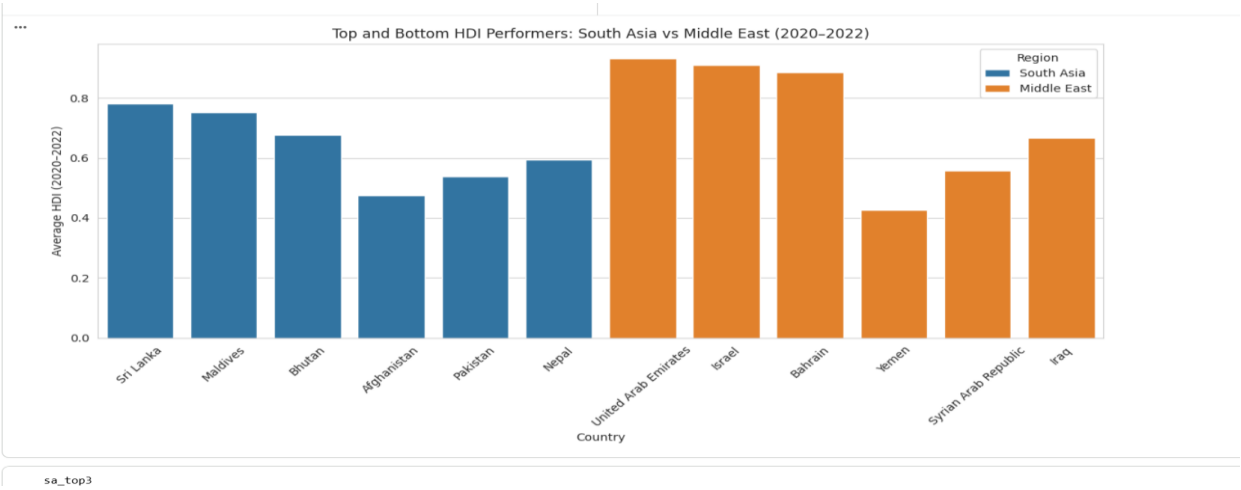


Fig: Top and Bottom HDI Performers: South Asia vs Middle East (2020–2022)

This chart show top and bottom HDI countries from South Asia and Middle East based on average HDI (2020–2022).
In South Asia, Sri Lanka is highest HDI country (around 0.78), then Maldives and Bhutan. Nepal and Pakistan are middle, and Afghanistan is lowest with HDI below 0.5. South Asia mostly have medium to low human development.
In Middle East, UAE, Israel and Bahrain have very high HDI (above 0.88), show strong development. On other side, Yemen is weakest with very low HDI. Syria and Iraq also lower than regional leaders.
Overall, Middle East perform better than South Asia, but both region have big gap between top and bottom countries**.**

– Interpretation and discussion

My results show that the Middle East is the better-performing region on average. However, the higher standard deviation and Coefficient of

Variation (CV) in the Middle East show there is a lot of inequality. For example, there is a very big gap between a rich country like the United Arab Emirates and a country like Yemen.

In South Asia, the average is lower, but the countries are closer to each other because the spread is smaller. When looking at the metrics, GNI per Capita shows a very big difference between the regions. The scatter plots show that having more money usually helps a country have better human development, but it is not the only thing that matters. Overall, the Middle East has higher scores, but South Asia is more "balanced" even though it is at a lower level.

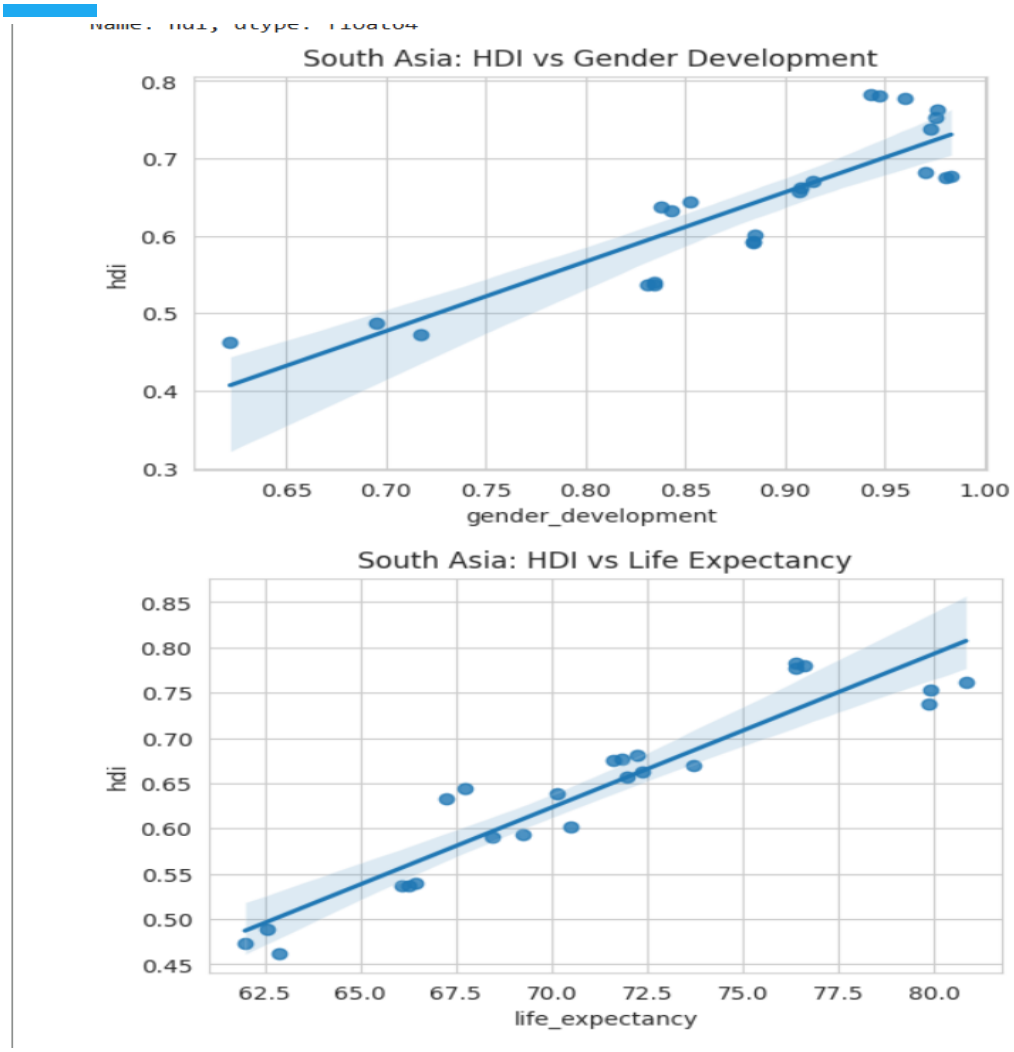


Fig: life expectancy and gender development vs HDI(south_Asia)

This figure show the relationship between HDI and Gender Development, and HDI and Life Expectancy in South Asia.

In the first chart, there is a positive relationship between gender development and HDI. As gender development index increase, the HDI value also increase. Countries with higher gender equality tend to have better human development. Few countries with low gender development clearly show lower HDI, which indicate gender gap affect overall development.

In the second chart, HDI increase with life expectancy. Countries where people live longer usually have higher HDI. The trend line go upward, showing strong connection between health condition and human development. Lower life expectancy countries remain at lower HDI level. Overall, both charts clearly show that gender equality and life expectancy play important role in improving HDI in South Asia.

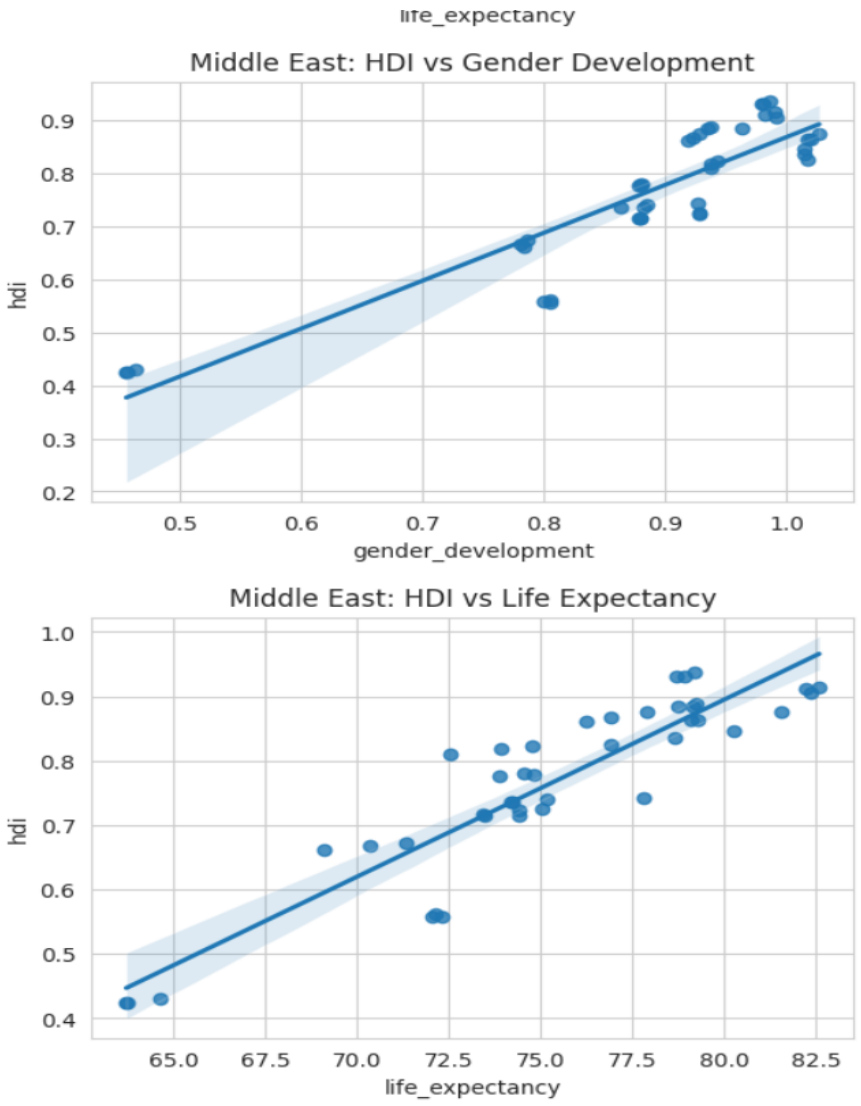


Fig: life expectancy and gender development vs HDI(Middle East)

The graph is about the relationship between Human Development Index (HDI) and life expectancy in the Middle East. It shows how life expectancy is linked with overall human development in Middle Eastern countries. The horizontal axis represent life expectancy in years, while the vertical axis show HDI values. Each point on the graph stand for a country, and the trend line summarize the overall pattern.

Looking at the values, countries with higher life expectancy (around 75–80 years) have very high HDI (about 0.85–0.92). Countries with lower life expectancy (around 65–70 years) fall into lower HDI range (around 0.65–0.75). The points are closely grouped around the trend line, showing a strong and consistent relationship.

Both regions show that gender equality and life expectancy are key drivers of HDI, but the Middle East outperform South Asia in translating these factors into higher human development. South Asia need stronger social, economic, and institutional support to turn improvements in gender and health into higher overall HDI.

Conclusion:

This study show that Human Development Index (HDI) is not only about income but also strongly depend on health and education. From the analysis, Middle East countries perform better in average then South Asia, mainly because of higher income level, but they also have more bigger inequality between rich and poor countries. South Asia have lower HDI overall, but the countries are more close with each other, which show a more balanced pattern but at lower level. The result clearly show that life expectancy and gender development have strong positive relation with HDI in both the regions. Even some countries with good income still can have lower HDI if health and gender equality is weak. Overall, to improve human development, countries should not only focus on economic growth but also need to improve health, education and give equal opportunity for all people.