



Analysis on Human Development

Tejamanikanta, Gudla

Instructor: Ms. Gahangir Hossain

INFO 5709 - Data Visualization

Term Project 24/07/2023

Introduction:

The Human Development Index (HDI) is a pivotal tool for assessing the overall development and well-being of countries across the globe. Comprising indicators related to health, education, and income, the HDI provides valuable insights into the quality of life and standard of living experienced by individuals within a nation. In this project, we delve into a comprehensive dataset available on Kaggle, featuring essential information on the HDI and its key components for various countries.

The dataset encompasses crucial metrics such as HDI rank, life expectancy at birth, expected years of schooling, mean years of schooling, gross national income (GNI) per capita, and GNI per capita rank minus HDI rank. Our primary objective is to conduct a thorough analysis of the dataset and uncover significant trends and patterns related to human development on a global scale.

Understanding the factors contributing to a higher HDI is of paramount importance for policymakers, governments, and international organizations. It equips them with essential knowledge to design targeted and effective strategies for fostering sustainable development and improving the well-being of their citizens. By identifying countries that have made remarkable strides in achieving higher HDI scores, we can draw valuable lessons from their successes. Simultaneously, analyzing regions or countries with lower HDI scores can reveal the challenges they face and enable us to extend support and aid where it is most needed.

Moreover, the correlation analysis between different HDI components can unravel the intricate relationship between health, education, and income levels within societies. This deeper understanding can further inform policymakers and stakeholders about the priority areas for interventions to uplift human development levels.

Dataset Overview:

The dataset provided contains a comprehensive set of indicators related to the Human Development Index (HDI) and its components for various countries. HDI is a composite measure used to evaluate a country's overall development status, considering key factors such as life expectancy, education, and income. This dataset offers valuable insights into the quality of life and standard of living for individuals within each nation, facilitating cross-country comparisons and in-depth analysis of global development trends.

It Contains 11 Columns and 195 Columns after Cleaned.

Dataset Link: <https://www.kaggle.com/datasets/rajkumarpandey02/human-development-index-and-components>

Description of Attributes:

1. **HDI Rank:** This attribute represents the ranking of countries based on their Human Development Index (HDI) scores. Countries are ordered from highest to lowest HDI scores, providing a clear hierarchy of human development levels on a global scale.
2. **Country:** This attribute indicates the name of each country included in the dataset. It serves as a unique identifier for each country, enabling efficient data organization and facilitating cross-referencing with other datasets.
3. **HUMAN DEVELOPMENT:** This attribute might provide additional information or classification related to the level of human development in each country. It could be used to group countries based on their development status, such as "Very High", "High", "Medium", "Low" or "Others".
4. **Human Development Index (HDI):** The HDI is a composite index that measures a country's overall development. It considers life expectancy at birth, expected years of schooling, and mean years of schooling, providing a single value that represents the country's level of human development.
5. **Life Expectancy at Birth:** The number of years a newborn infant could expect to live if the prevailing patterns of age-specific mortality rates at the time of birth remain constant throughout the infant's life. It is a key indicator of a country's overall health and well-being.
6. **Expected Years of Schooling:** This attribute denotes the number of years of education a child of school-entry age can expect to receive, assuming the education system's current enrollment rates remain constant.
7. **Mean Years of Schooling:** This attribute indicates the average number of years of education received by individuals aged 25 years and older in each country. It provides insights into the country's education system and the level of knowledge and skills possessed by its population.
8. **Gross National Income (GNI) per Capita:** GNI per capita measures the average income earned by a country's citizens and is an essential indicator of economic prosperity and standard of living.
9. **GNI per Capita Rank minus HDI Rank:** This attribute represents the difference in ranking a country based on its GNI per capita and its HDI value. A negative value indicates that the country is better ranked by GNI per capita than by HDI value, meaning its income ranking is higher compared to its HDI ranking. Conversely, a positive value suggests that the country's income ranking is lower than its HDI ranking.

Tools Used: Tableau and Jupyter Notebook

Exploratory Data Analysis:

- Removing Duplicate Column named as HDI_rank1.
- Missing values have been replaced with 0 as there are only a few missing values which won't affect the data in large operations.

```
In [1]: 1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 # Load the data
5 data = pd.read_csv("HDI.csv")
6 print(data.columns)

Index(['HDI_rank', 'Country', 'HUMAN_DEVELOPMENT', 'Human_Development_Index',
      'Life_expectancy_at_birth', 'Expected_years_of_schooling',
      'Mean_years_of_schooling', 'Gross_national_income_per_capita',
      'GNI_per_capita_rank_minus_HDI_rank', 'HDI_rank1'],
      dtype='object')
```

```
In [2]: 1 # Clean the data as we can see duplicate column
2 data.drop(columns=['HDI_rank1'], inplace=True)
3 print(data.columns)

Index(['HDI_rank', 'Country', 'HUMAN_DEVELOPMENT', 'Human_Development_Index',
      'Life_expectancy_at_birth', 'Expected_years_of_schooling',
      'Mean_years_of_schooling', 'Gross_national_income_per_capita',
      'GNI_per_capita_rank_minus_HDI_rank'],
      dtype='object')
```

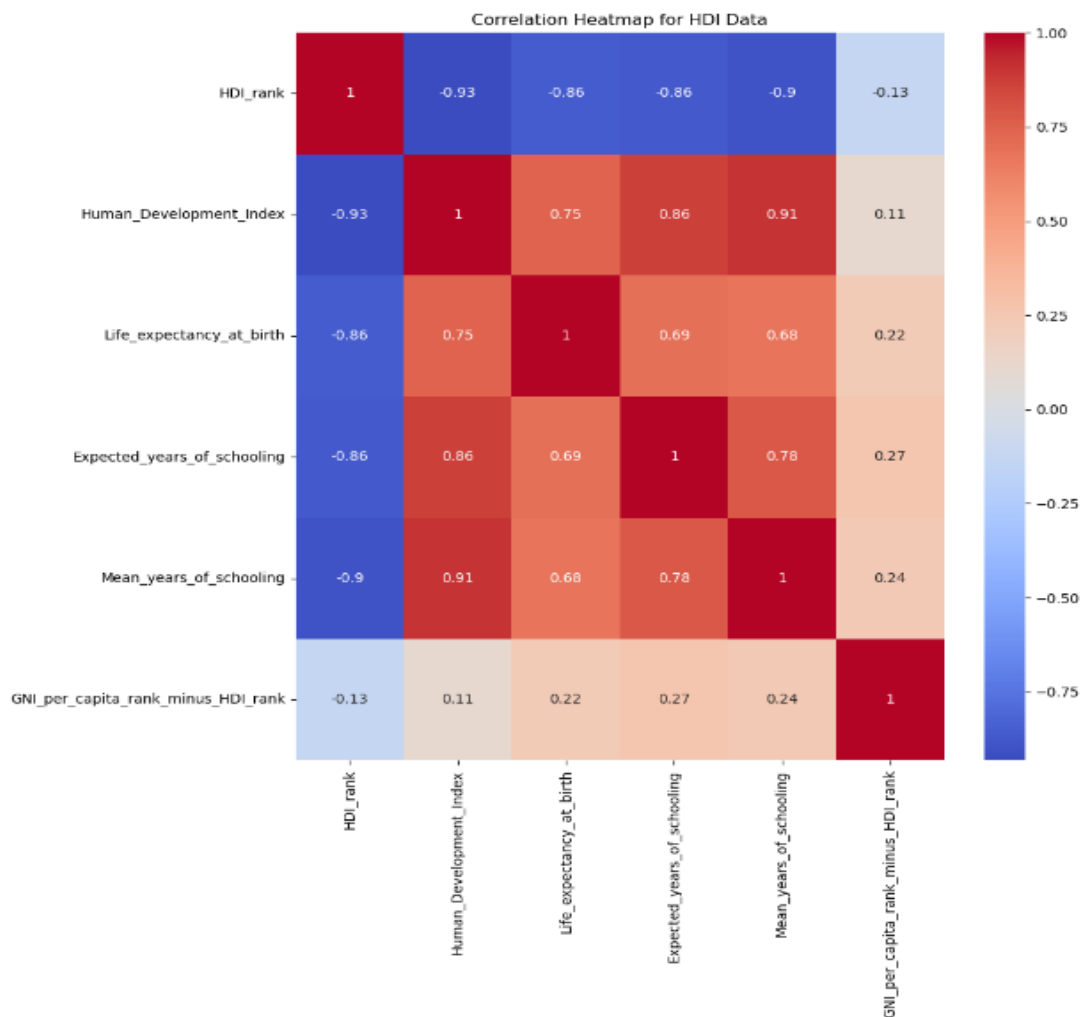
- After Filling Missing Values again verified by checking null values count in all the columns.

```
In [3]: 1 # Check for missing values
2 missing_values = data.isnull().sum()
3 print(missing_values)
4 # Filling missing values with zero
5 data_clean = data.fillna(0)
6 print('****Updated Columns after Cleaning****')
7 missing_values_1 = data_clean.isnull().sum()
8 print(missing_values_1)
```

```
HDI_rank      0
Country       0
HUMAN_DEVELOPMENT  0
Human_Development_Index  4
Life_expectancy_at_birth  0
Expected_years_of_schooling  2
Mean_years_of_schooling  4
Gross_national_income_per_capita  2
GNI_per_capita_rank_minus_HDI_rank  4
dtype: int64
****Updated Columns after Cleaning****
HDI_rank      0
Country       0
HUMAN_DEVELOPMENT  0
Human_Development_Index  0
Life_expectancy_at_birth  0
Expected_years_of_schooling  0
Mean_years_of_schooling  0
Gross_national_income_per_capita  0
GNI_per_capita_rank_minus_HDI_rank  0
dtype: int64
```

Correlation Matrix:

```
In [17]: 1 numeric_columns = []
2 for column in data_clean.columns:
3     if pd.api.types.is_numeric_dtype(data_clean[column]):
4         numeric_columns.append(column)
5
6 data_clean = data_clean[numeric_columns]
7
8 # Create a correlation matrix
9 correlation_matrix = data_clean.corr()
10
11 # Create a heatmap of the correlation matrix
12 plt.figure(figsize=(10, 10))
13 sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
14 plt.title("Correlation Heatmap for HDI Data")
15 plt.show()
```

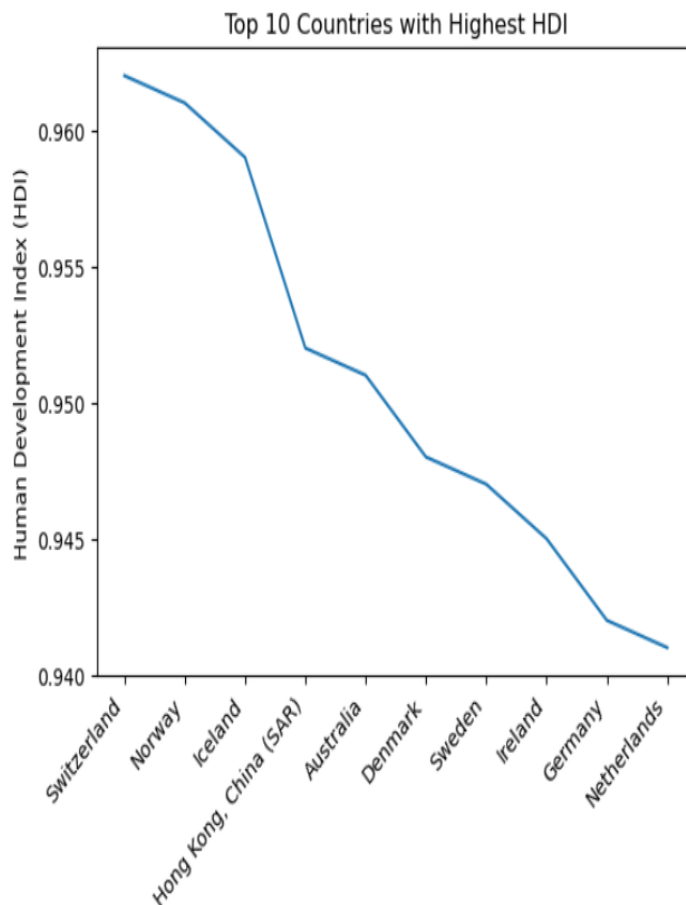


Outcome:

The heatmap shows the correlation between the Human Development Index (HDI) and three other variables: life expectancy at birth, expected years of schooling, and gross national income per capita. The darker the color, the stronger the correlation. The heatmap shows that there is a strong positive correlation between HDI and all three variables. This means that countries with a higher HDI tend to have higher life expectancy, more years of schooling, and higher income.

1. Line plot on Top 10 countries with the highest 'Human_Development_Index' (HDI).

```
In [4]: 1 # Plot 1: Top 10 countries with the highest 'Human_Development_Index' (HDI)
2 top_10_hdi_countries = data_clean.nlargest(10, 'Human_Development_Index')
3 plt.plot(top_10_hdi_countries['Country'], top_10_hdi_countries['Human_Development_Index'])
4 plt.xticks(rotation=45, ha='right')
5 plt.xlabel('Country')
6 plt.ylabel('Human Development Index (HDI)')
7 plt.title('Top 10 Countries with Highest HDI')
8 plt.show()
```

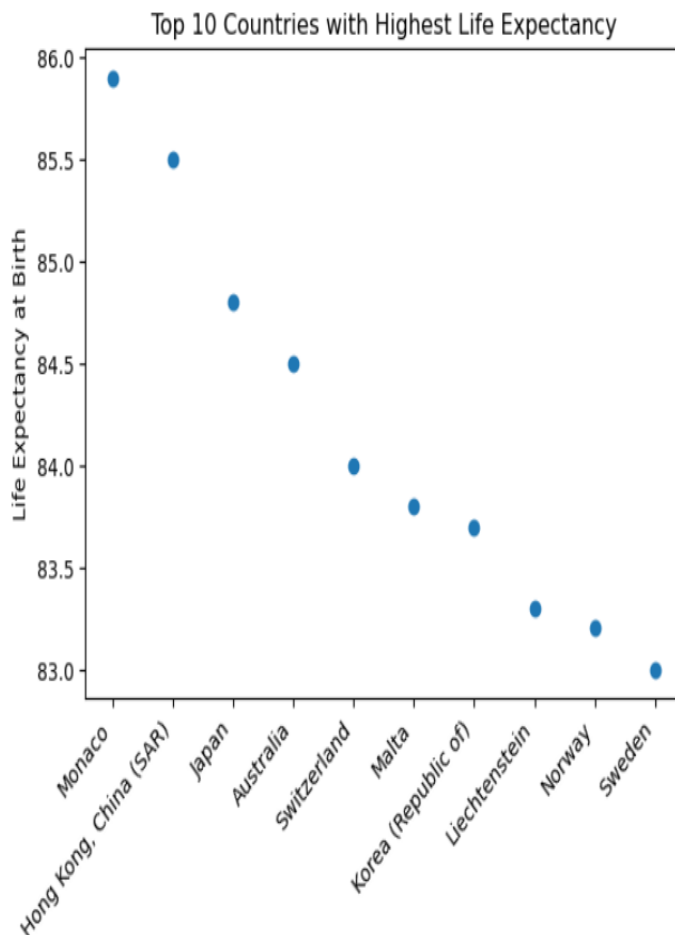


Outcome:

The result shows the top 10 countries with the highest HDI. The HDI is a composite index that measures a country's achievements in three dimensions: health, education, and income. The countries in the image are ranked according to their HDI score, which is a number on a scale from 0 to 1. Switzerland has the highest HDI value with 0.962.

2. Scatter plot on Top 10 countries with the highest 'Life_expectancy_at_birth'.

```
In [5]: 1 # Plot 2: Top 10 countries with the highest 'Life_expectancy_at_birth'
2 top_10_life_expectancy_countries = data_clean.nlargest(10, 'Life_expectancy_at_birth')
3 plt.scatter(top_10_life_expectancy_countries['Country'], top_10_life_expectancy_countries['Life_expectancy_at_birth'])
4 plt.xticks(rotation=45, ha='right')
5 plt.xlabel('Country')
6 plt.ylabel('Life Expectancy at Birth')
7 plt.title('Top 10 Countries with Highest Life Expectancy')
8 plt.show()
```

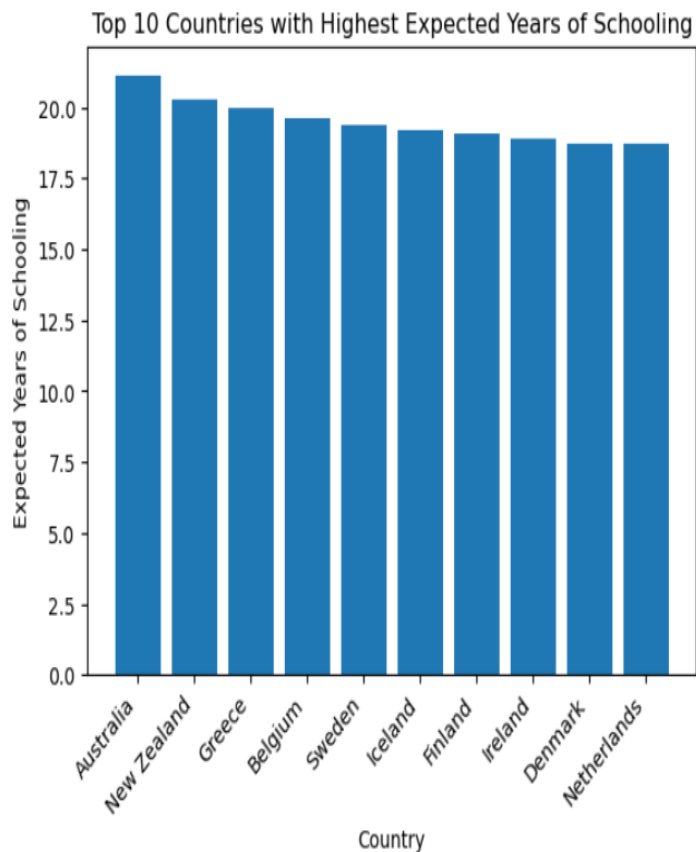


Outcome:

The result shows the top 10 Countries with highest Life expectancy. It is measured in Number of Years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life. Monaco has the highest value with 85.9 years followed with other countries.

3. Bar Plot Top 10 countries with the highest 'Expected_years_of_schooling'.

```
In [6]: 1 # Plot 3: Top 10 countries with the highest 'Expected_years_of_schooling'
2 top_10_expected_schooling_countries = data_clean.nlargest(10, 'Expected_years_of_schooling')
3 plt.bar(top_10_expected_schooling_countries['Country'], top_10_expected_schooling_countries['Expected_years_of_schooling'])
4 plt.xticks(rotation=45, ha='right')
5 plt.xlabel('Country')
6 plt.ylabel('Expected Years of Schooling')
7 plt.title('Top 10 Countries with Highest Expected Years of Schooling')
8 plt.show()
```

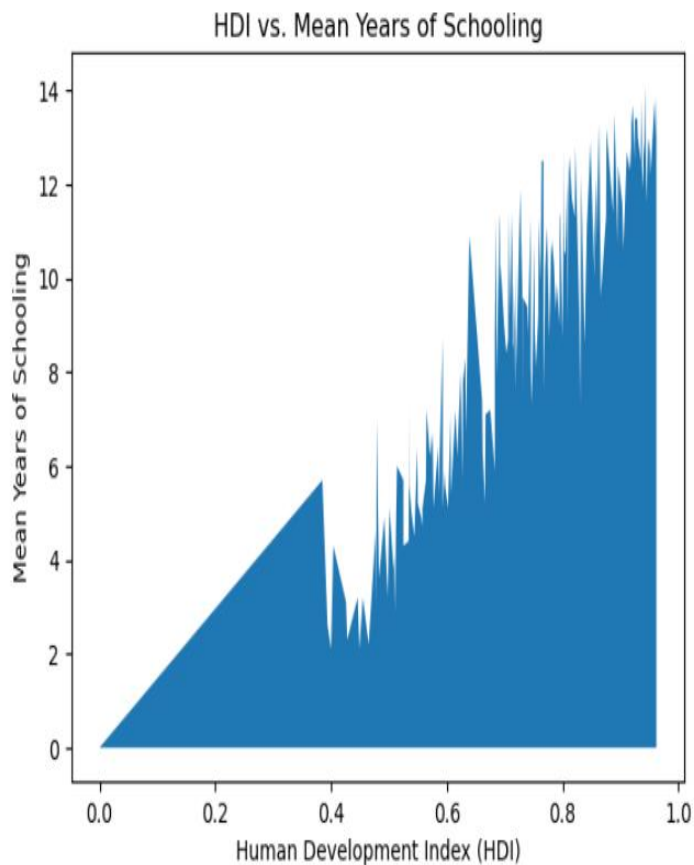


Outcome:

The above graph displays the Top 10 Countries with Expected Years of Schooling. It is measured in Number of years that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout the child's life. Australia has the highest value of 21.1 Years.

4. Area plot between 'Human_Development_Index' and 'Mean_years_of_schooling'.

```
In [7]: 1 # Plot 4: Area plot between 'Human_Development_Index' and 'Mean_years_of_schooling'
2 plt.fill_between(data_clean['Human_Development_Index'], data_clean['Mean_years_of_schooling'])
3 plt.xlabel('Human Development Index (HDI)')
4 plt.ylabel('Mean Years of Schooling')
5 plt.title('HDI vs. Mean Years of Schooling')
6 plt.show()
```

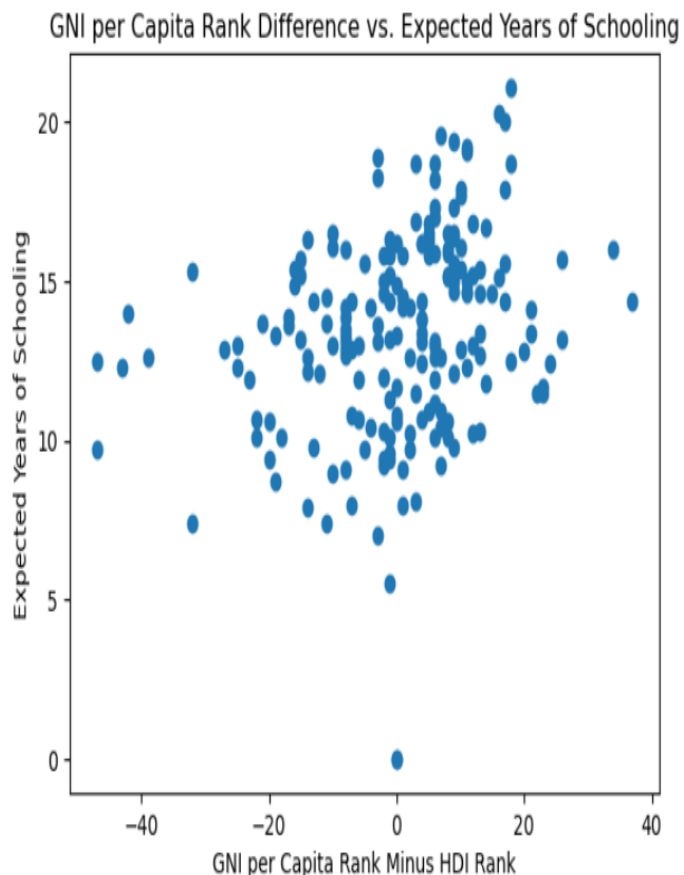


Outcome:

The result of the area plot displays the Mean Years of Schooling related to HDI. It is measured in Number of Years. The meantime of the Schooling will be represented in it. We can observe there are a few countries which still has zero in count. By this we can see that it weakly correlated to the Human Development Index for most of the countries.

5. Scatter plot between 'GNI_per_capita_rank_minus_HDI_rank' and 'Expected_years_of_schooling'.

```
In [8]: 1 # Plot 5: Scatter plot between 'GNI_per_capita_rank_minus_HDI_rank' and 'Expected_years_of_schooling'
2 plt.scatter(data_clean['GNI_per_capita_rank_minus_HDI_rank'], data_clean['Expected_years_of_schooling'])
3 plt.xlabel('GNI per Capita Rank Minus HDI Rank')
4 plt.ylabel('Expected Years of Schooling')
5 plt.title('GNI per Capita Rank Difference vs. Expected Years of Schooling')
6 plt.show()
```

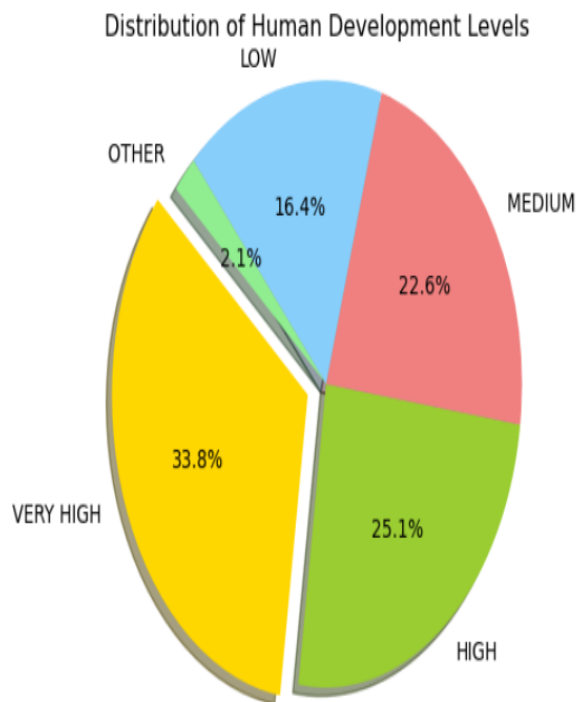


Outcome:

The above Scatter plot is arranged between expected years of schooling and GNI per capita rank minus HDI rank. The GNI per capita rank difference is calculated by subtracting the GNI per capita rank from the HDI rank. A positive rank difference means that the country has a higher GNI per capita than HDI rank, while a negative rank difference means that the country has a lower GNI per capita than HDI rank.

6. A pie chart to show the distribution of human development levels.

```
In [9]: 1 # plot 6:A pie chart to show the distribution of human development levels
2 hd_counts = data_clean['HUMAN_DEVELOPMENT'].value_counts()
3 labels = hd_counts.index
4 sizes = hd_counts.values
5 colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue', 'lightgreen']
6 explode = (0.1, 0, 0, 0, 0) # Explode the first slice (very high) to highlight it
7
8 plt.pie(sizes, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%', shadow=True, startangle=140)
9 plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
10
11 plt.title('Distribution of Human Development Levels')
12 plt.show()
```



Outcome:

The pie chart shows the distribution of human development levels in the world. The very high level is 33.8%, which means that a third of the world's countries have a very high level of human development. This is followed by the High level, which is 25.1%. The other levels are medium, low, and other. That shows most of the countries are being developed in terms of Human Development.

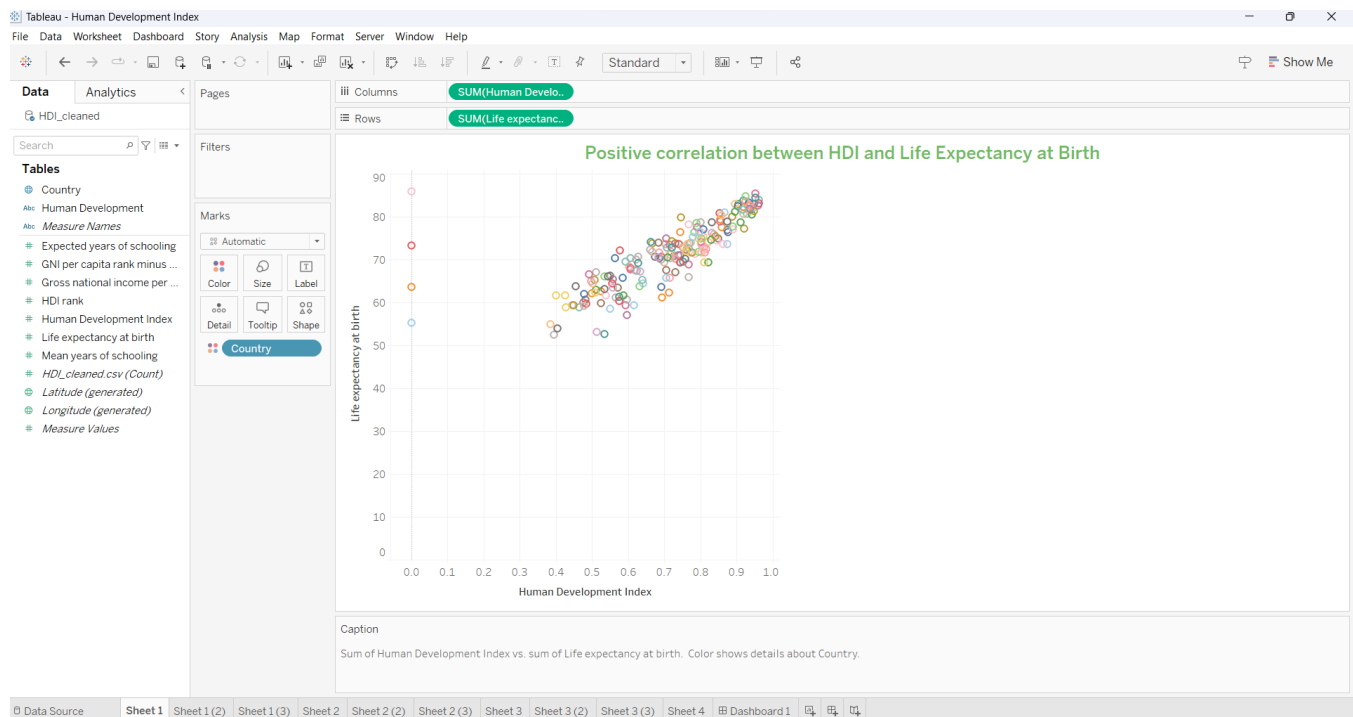
Hypothesis:

1. Positive correlation between HDI and Life Expectancy at Birth.
2. Positive correlation between HDI and Expected year of schooling.
3. Positive correlation between HDI and Mean year of Schooling.
4. Distribution of Human Development levels.

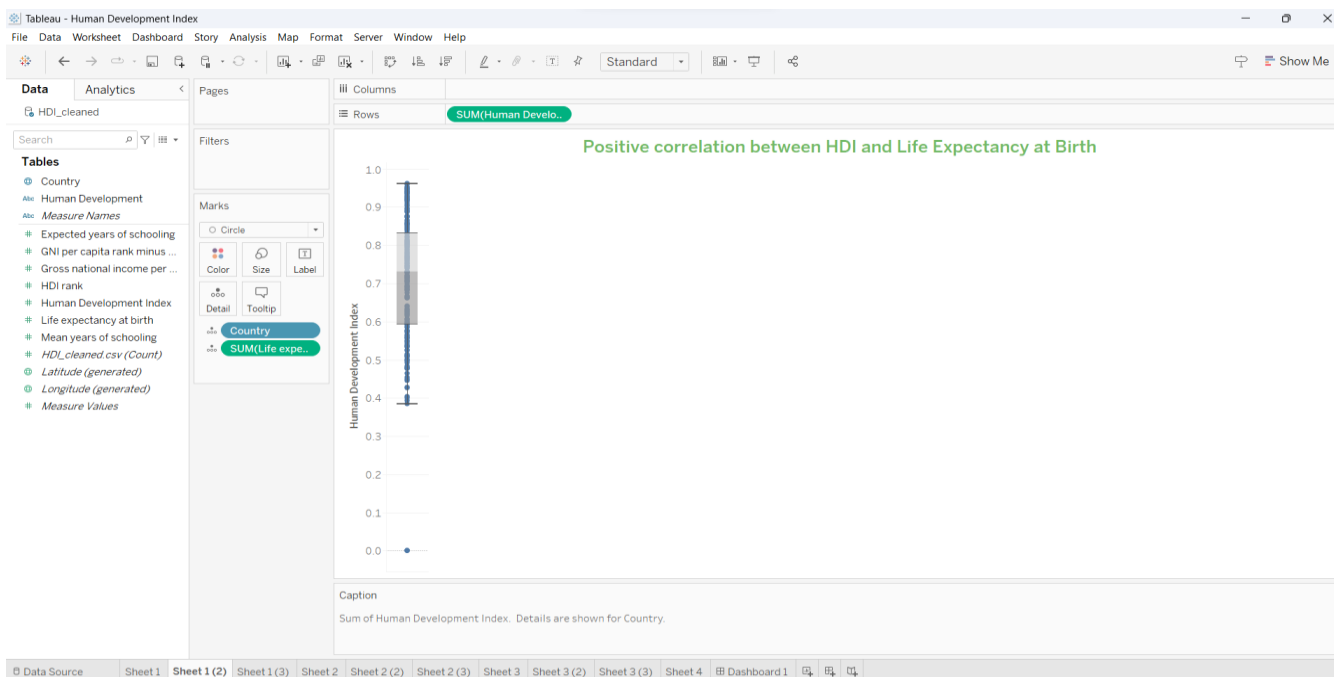
1. Positive correlation between HDI and Life Expectancy at Birth.

The Positive correlation between HDI (Human Development Index) and Life Expectancy at Birth means that as the HDI of a country increase, the life expectancy of its citizens also tends to increase. In other words, there is a strong association between higher levels of human development and longer life expectancies.

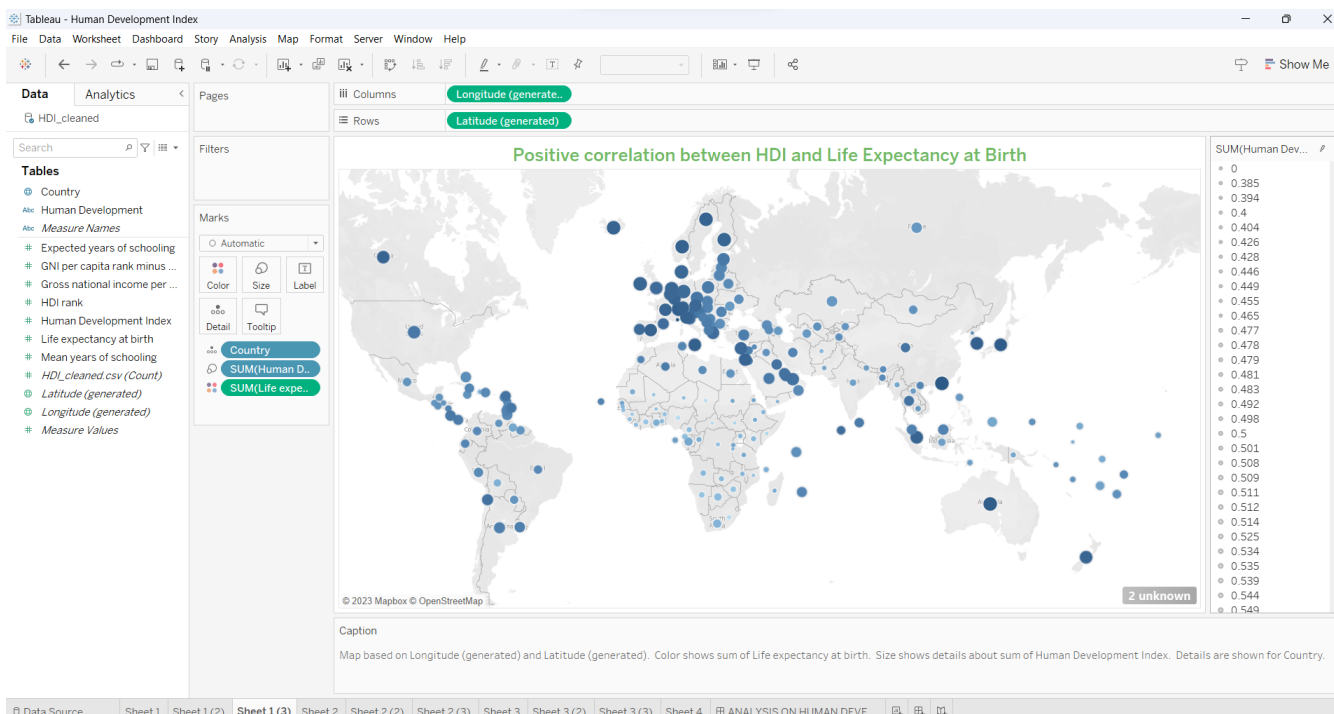
For this Hypothesis, I have chosen 3 different plots Scatter, box plot and Geographical map with size of value.



In this scatter plot shows the relationship between the Human Development Index (HDI) and life expectancy at birth for all countries. Monaco has the highest Life expectancy with value of 85.90.



In This box plot shows the distribution of life expectancy at birth for countries with different HDI levels. The median life expectancy for countries with a high HDI is 75 years, and the interquartile range (IQR) is 5 years. This means that 50% of the countries with a high HDI have a life expectancy between 70 and 80 years, and 25% of the countries have a life expectancy below 70 or above 80 years. There are two outliers, one at 65 and one at 85. This suggests that there are two countries with a life expectancy that is significantly lower or higher than the rest of the countries with a high HDI.

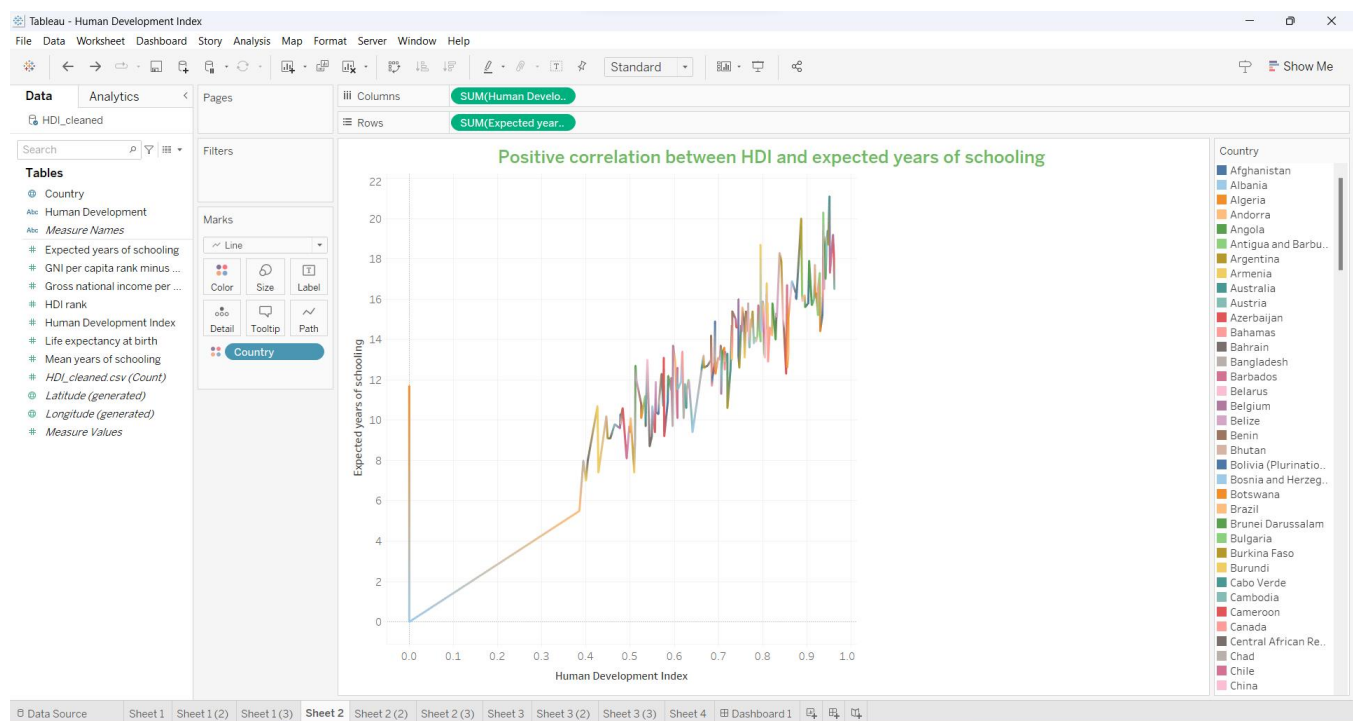


In this map we can observe countries in Europe are shaded with dark blue mostly it says it has high Life Expectancy at Birth and where as in Africa it is light Blue which is it has low values. Size of the Circle represents HDI value.

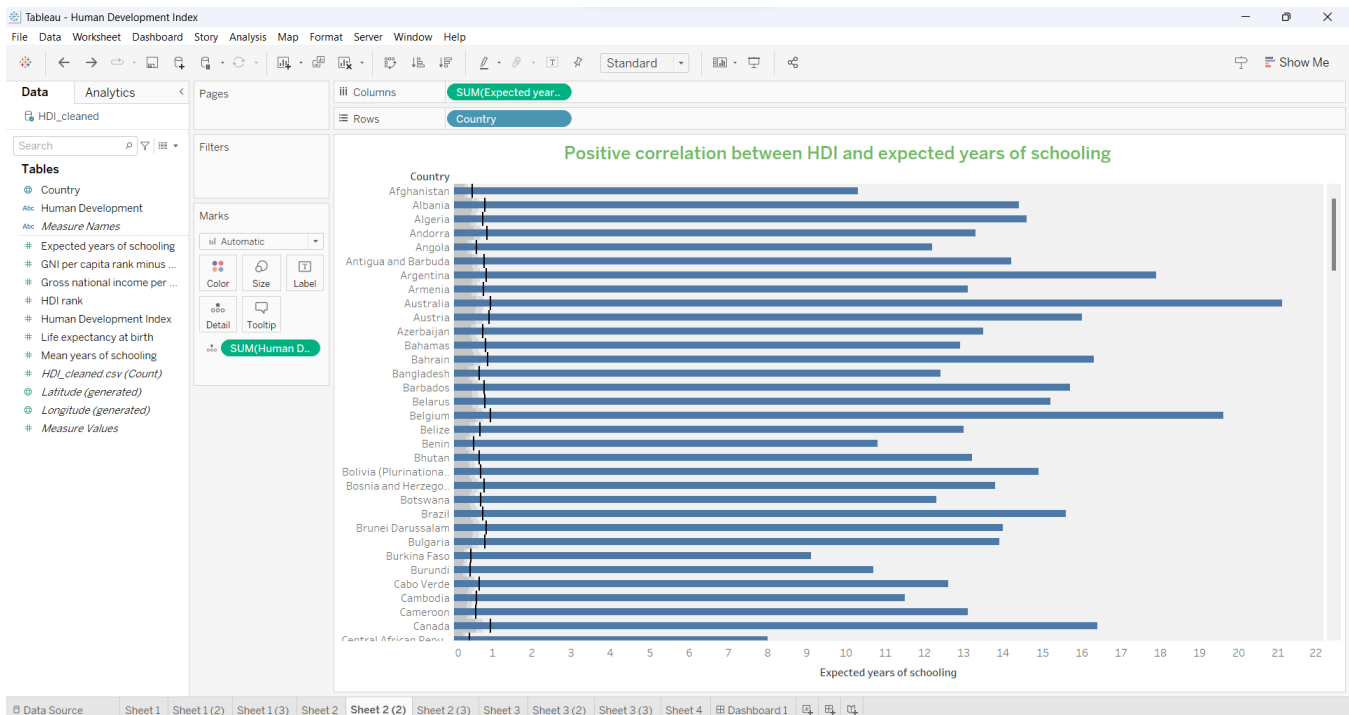
2. Positive correlation between HDI and Expected year of schooling.

The positive correlation between HDI (Human Development Index) and Expected Years of Schooling means that as the HDI of a country increase, the expected years of schooling for its children also tend to increase. In other words, there is a tendency for countries with higher human development levels to provide more years of schooling to their children.

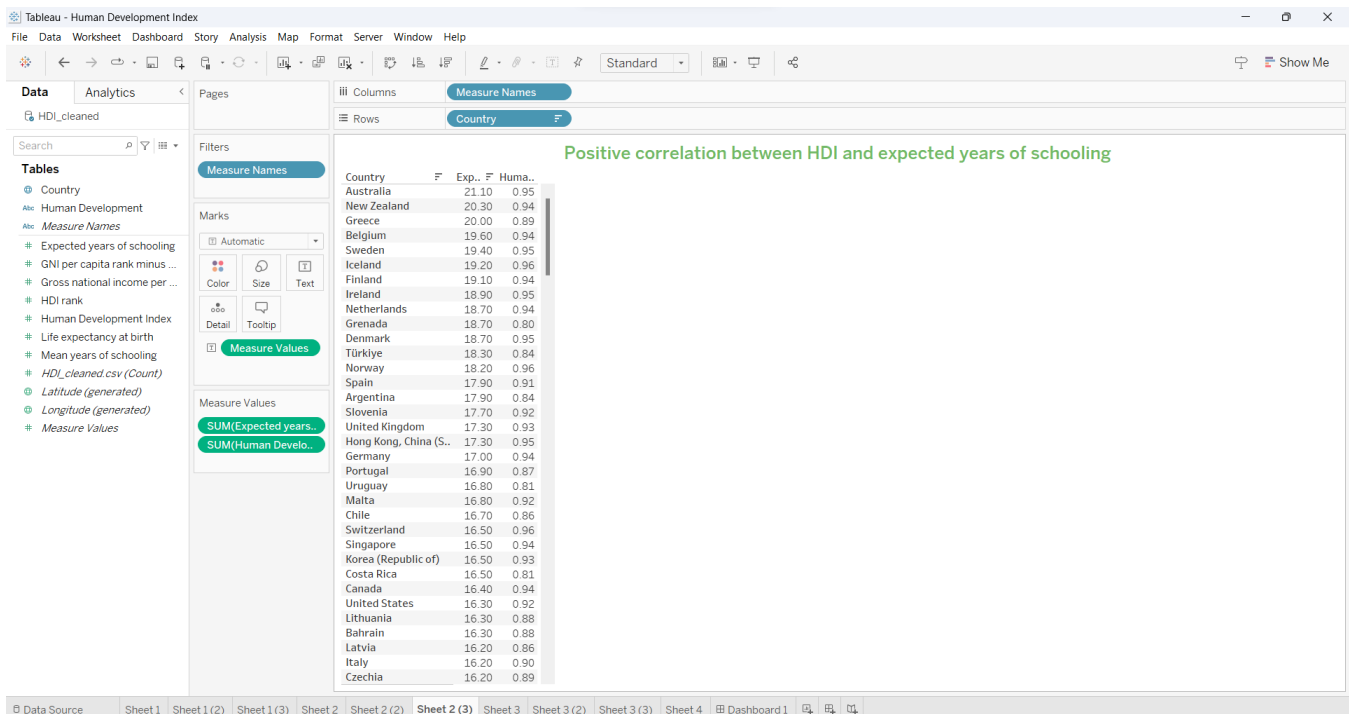
For this Hypothesis, I have chosen 3 different plots line, bullet plot and Table view.



Line plot shows Countries in Europe and North America tend to have a higher HDI and expected years of schooling than countries in Africa and Asia.



In this Bullet Bar plot, we can see Norway, which has an HDI of 0.96, has an expected years of schooling of 18.2 years. On the other hand, Niger, which has an HDI of 0.535, has an expected years of schooling of 10.1 years.



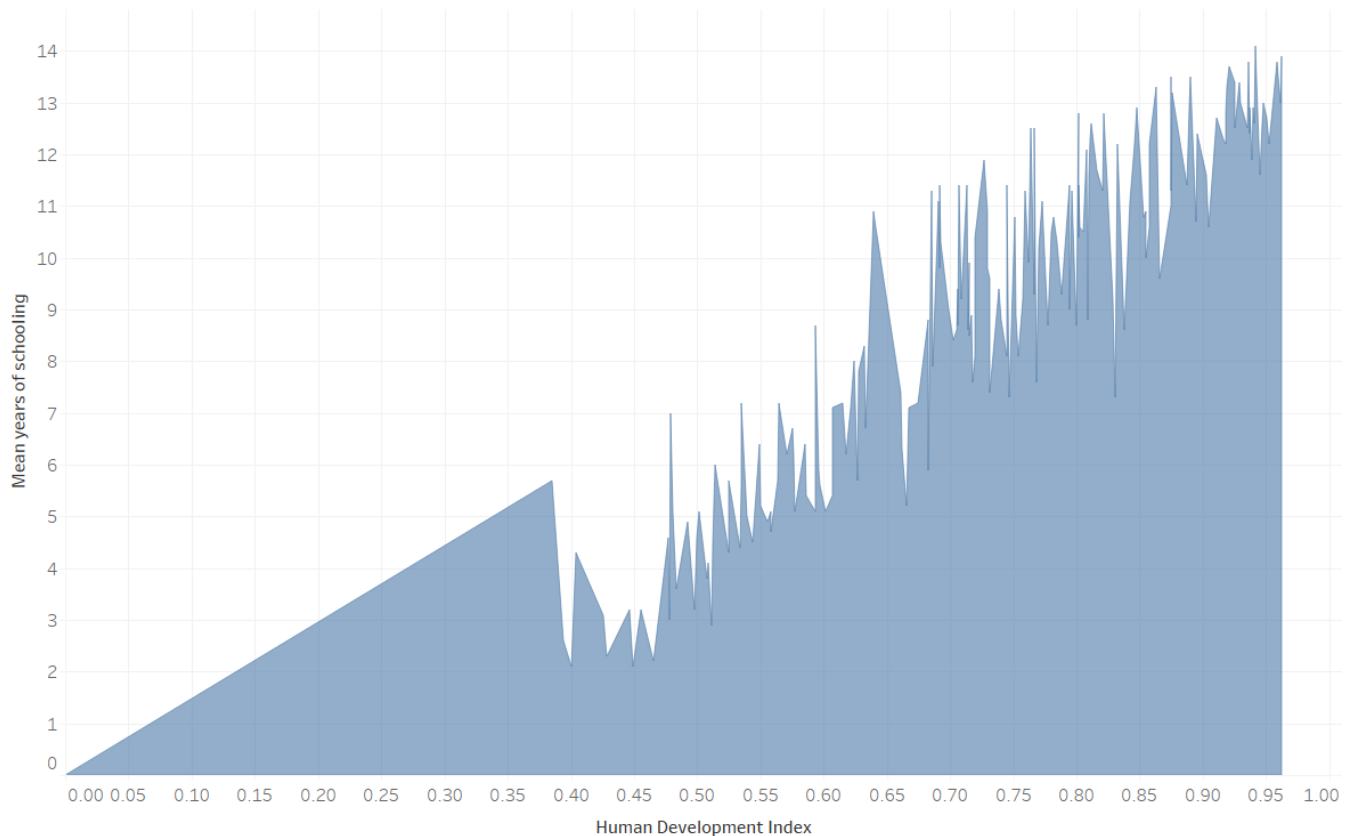
In this Text View we can easily see the highest Expected years of schooling is in Australia and it has HDI of 0.95.

3. Positive correlation between HDI and Mean year of Schooling.

The positive correlation between HDI (Human Development Index) and Mean Years of Schooling signifies that as a country's HDI increases, there is a tendency for the average number of years of education received by its population to also rise.

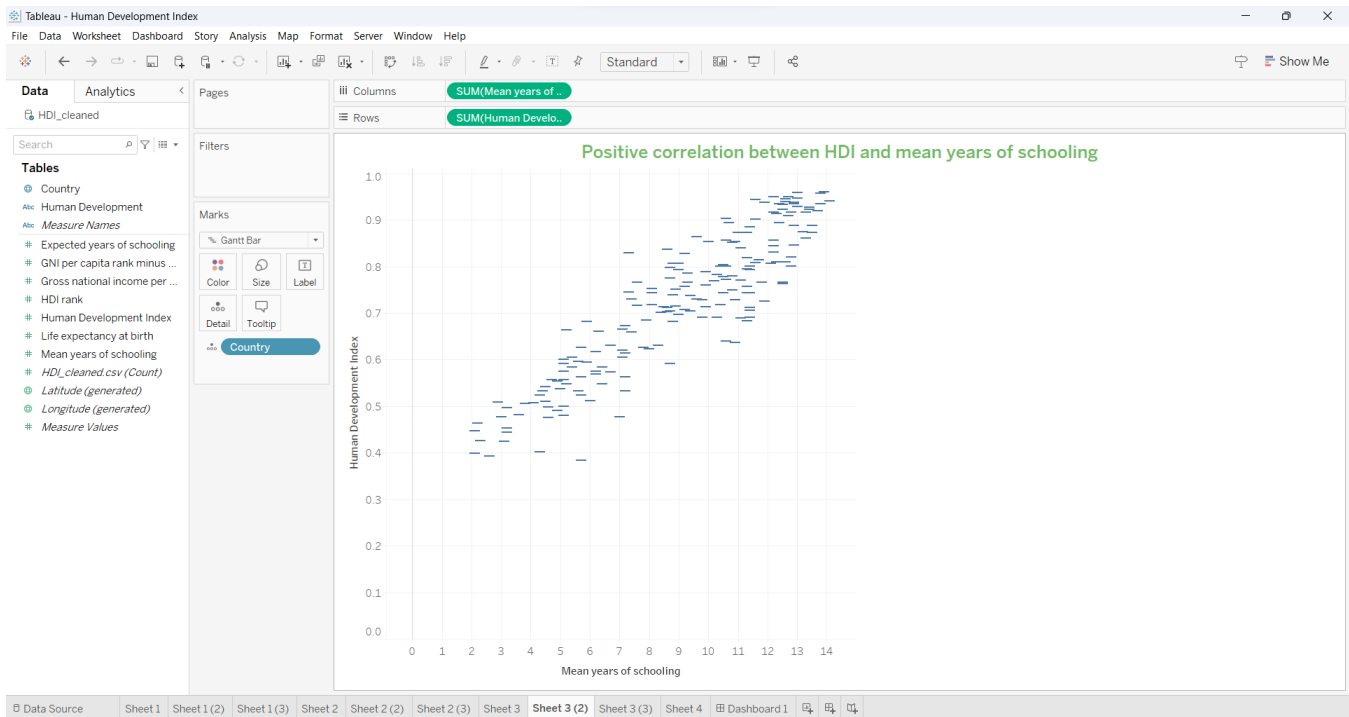
For this Hypothesis, I have chosen 3 different plots Area, Gantt plot and Map.

Positive correlation between HDI and mean years of schooling

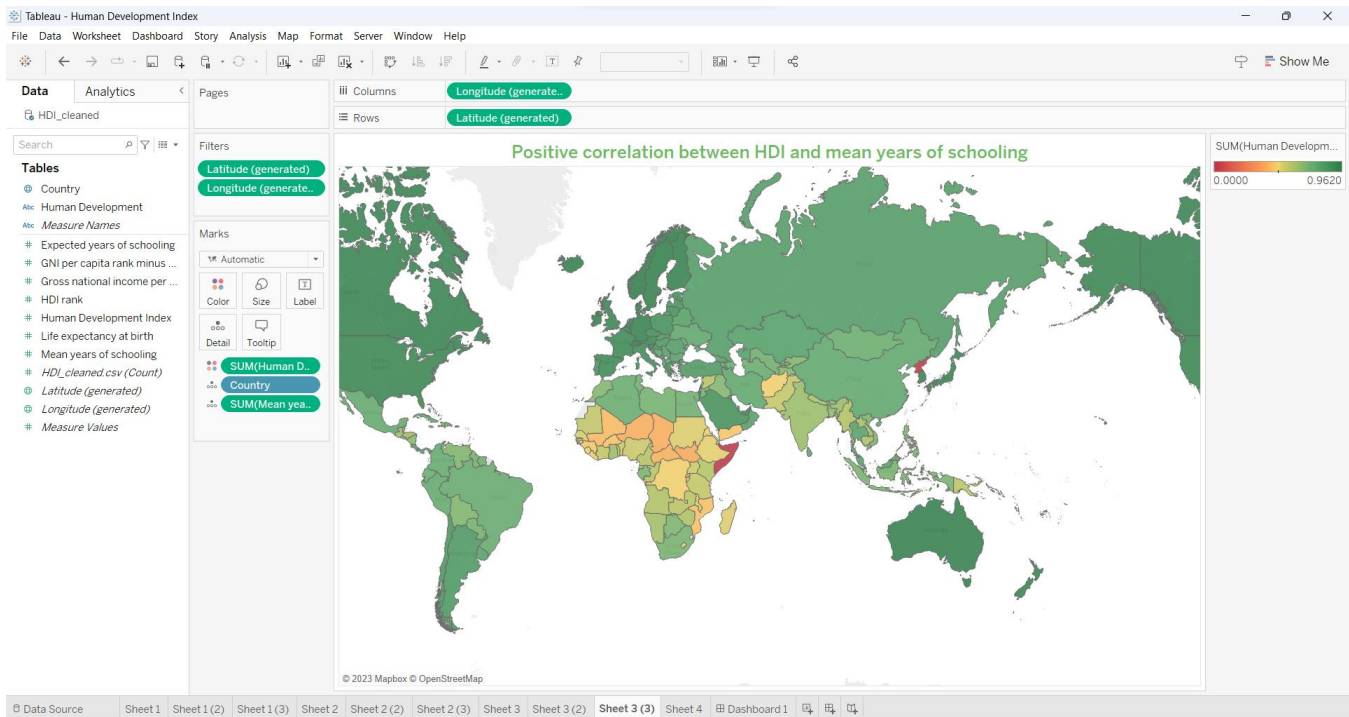


Sum of Human Development Index vs. sum of Mean years of schooling. Details are shown for Country.

Area plot shows Countries in Europe and North America tend to have a higher HDI and expected years of schooling than countries in Africa and Asia.



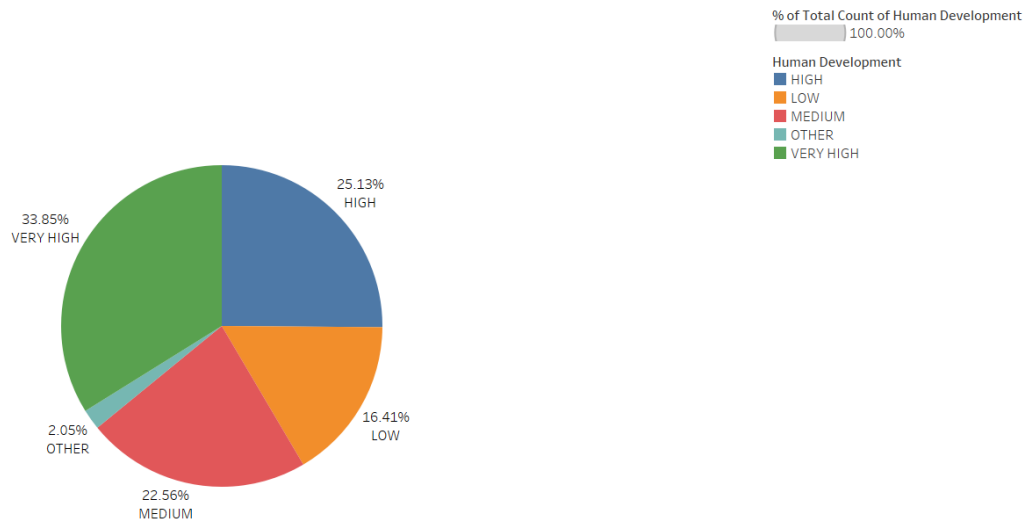
In this Gantt chart we can see most of the values are in higher region that is above 0.7 that says it has a positive relation between them.



From map we can see that African countries has the lowest HDI and those are need to develop in terms of schooling and living aspects but coming to Europe or US has high HDI which is a symbol of developed countries as HDI is one of the key aspects.

4. Distribution of Human Development levels.

Distribution of human development levels



% of Total Count of Human Development and Human Development. Color shows details about Human Development. Size shows % of Total Count of Human Development. The marks are labeled by % of Total Count of Human Development and Human Development. The view is filtered on Human Development, which keeps HIGH, LOW, MEDIUM, OTHER and VERY HIGH.

The above pie chart says about Human Development factor, and it is separated into group based on HDI value or The 3 factors which are Life expectancy at birth, Expected years of schooling and Mean years of Schooling based on the columns present in the dataset.

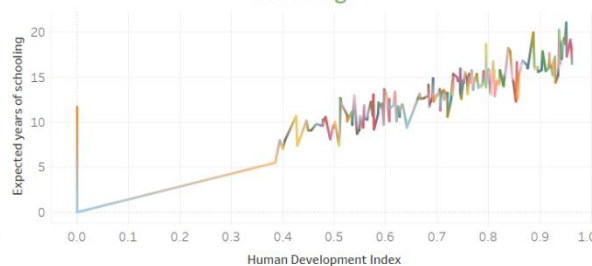
Dashboard:

ANALYSIS ON HUMAN DEVELOPMENT INDEX AND ITS COMPONENTS

Positive correlation between HDI and Life Expectancy at Birth



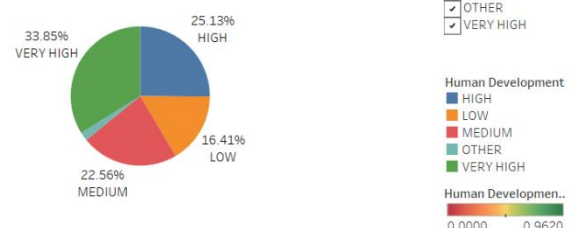
Positive correlation between HDI and expected years of schooling



Positive correlation between HDI and mean years of schooling

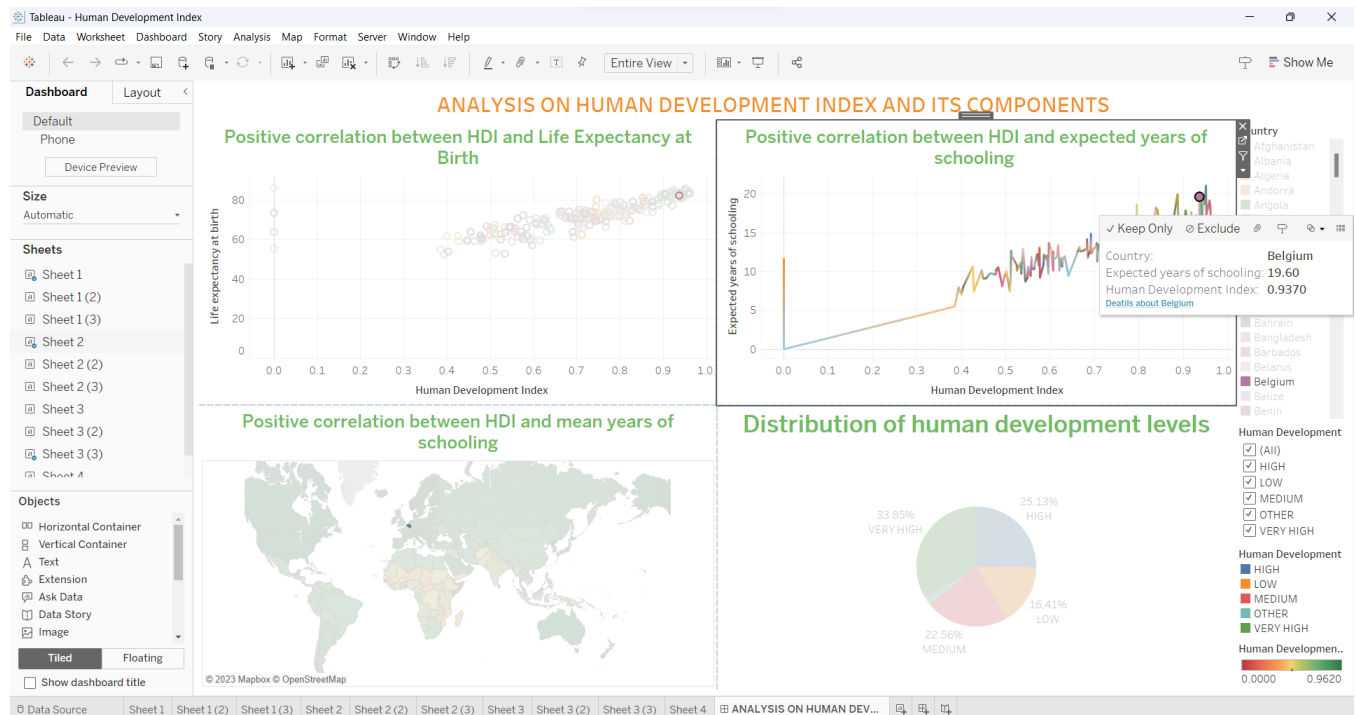


Distribution of human development levels



The above image is a Dashboard where we can add our worksheets into it make our own changes and link sheets to perform filters and tag a website to the variables.

I have chosen a few filters which display the right side of the Dashboard and attached a web link to the country field as we can go to undp official page, and it says information related to respective country.



The screenshot shows the UNDP Reports search results for 'Belgium'. The search results are displayed on a website with the URL hdr.undp.org/search?search_api_fulltext=Belgium. The search results are filtered to show 2 results for 'Belgium'.

2 results for "Belgium"

REFINE SEARCH RESULTS (News (1))

Search Results:

- Belgium**
 - ... Belgium ... Belgium ... country ...
- NEWS**
 - Recasting Human Development measures – Expert Group Meeting**
 - Saisana, Joint Research Center, European Commission, Belgium. Michaela Saisana is a senior

Conclusion:

In conclusion, this analysis has uncovered important details about the intricate connection between HDI and its constituent parts, illuminating the variables that influence human development in various nations. Decision-makers may create educated policies and interventions that support sustainable development, raise living standards, and promote greater well-being for populations all over the world by having a greater awareness of these processes.

References:

[1]. Human Development Index Data Set is collected from Kaggle.

<https://www.kaggle.com/datasets/rajkumarpandey02/human-development-index-and-components>

[2]. Human Development Index by UNDP.

<https://hdr.undp.org/>

[3]. Human Development Index information using Wikipedia.

https://en.wikipedia.org/wiki/Human_Development_Index

[4]. HDI by World Population Review

<https://worldpopulationreview.com/country-rankings/hdi-by-country>