

# DATA SCIENCE REPORT

## ON BILLIONAIRES

### STATISTICS DATASET\*

MERCY OKEBIORUN

### Dataset Overview:

- The dataset comprises information on **2,640** individuals with **35** features.

Key features include 'rank','finalWorth','age','birthYear','birthMonth','birthDay','cpi country','cpi change country','gross tertiary education enrollment','gross primary education enrollment country','life expectancy country','tax revenue country country','total tax rate country','population country','latitude country,' and 'longitude country.'

### Initial Data Inspection:

- The dataset was loaded successfully, containing billionaires' details such as rank, final worth, and various demographic indicators.

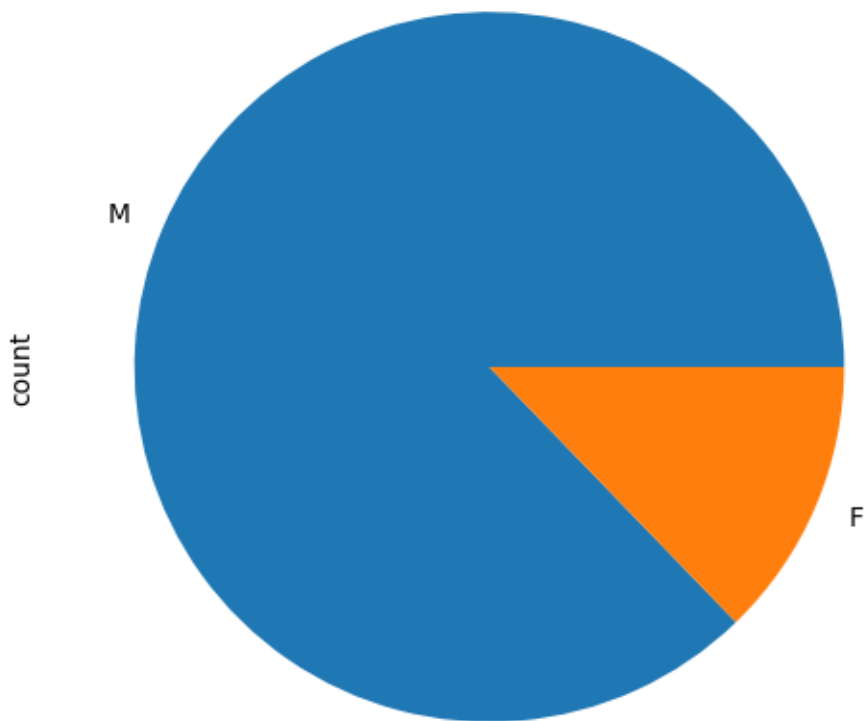
### Data Cleaning:

- Initial shape of the dataset: **(2640, 35)**.
- Columns with non-essential information, including 'date,' 'city,' 'category,' etc., were dropped.
- Initially, **4,228** missing values were identified and addressed using forward filling and mode imputation.
- After data cleaning, there are no missing values or duplicate entries in the dataset.

Final dataset: **(2640, 26)**

### Demographic Analysis:

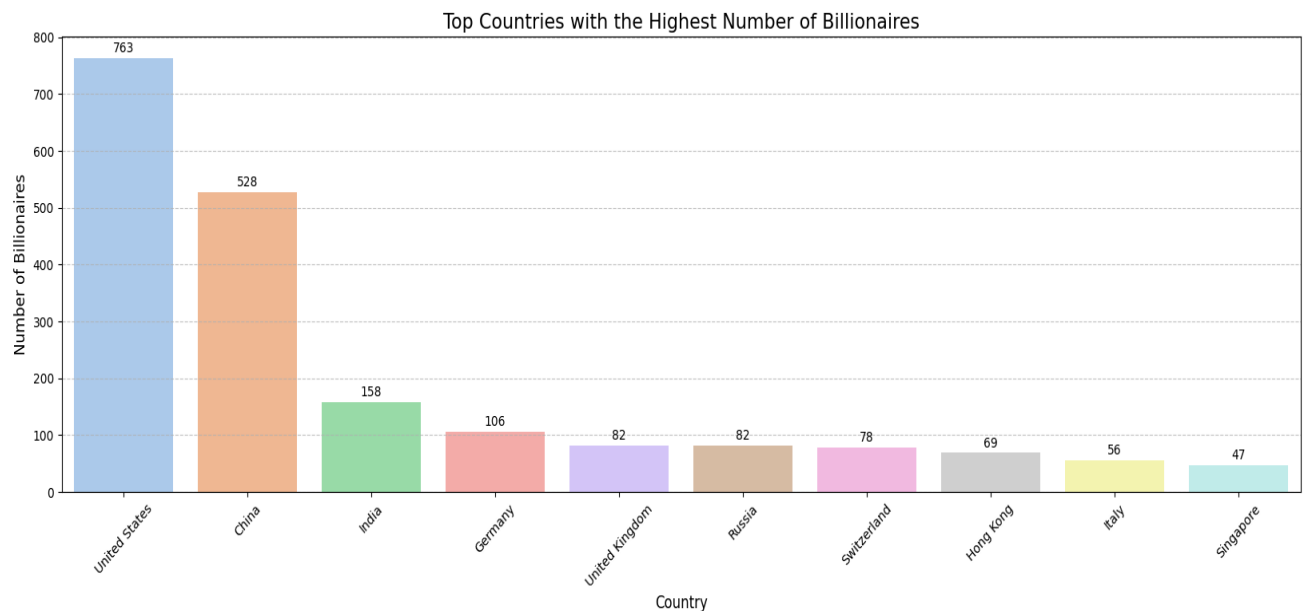
- The dataset predominantly consists of male billionaires **(2,303)** compared to female billionaires **(337)**.



- The average age of billionaires is approximately 65 years.

### Geographical Analysis:

- The top countries with the highest number of billionaires are visualized, showcasing the United States as the leader.



#### Conclusion:

- The dataset is now clean, free of missing values and duplicates, providing a solid foundation for further analysis.
- Demographic insights reveal a notable gender imbalance among billionaires.
- Geographical analysis emphasizes the concentration of billionaires in specific countries, with the United States leading the pack.

#### **\*Analysis and Insights Report\***

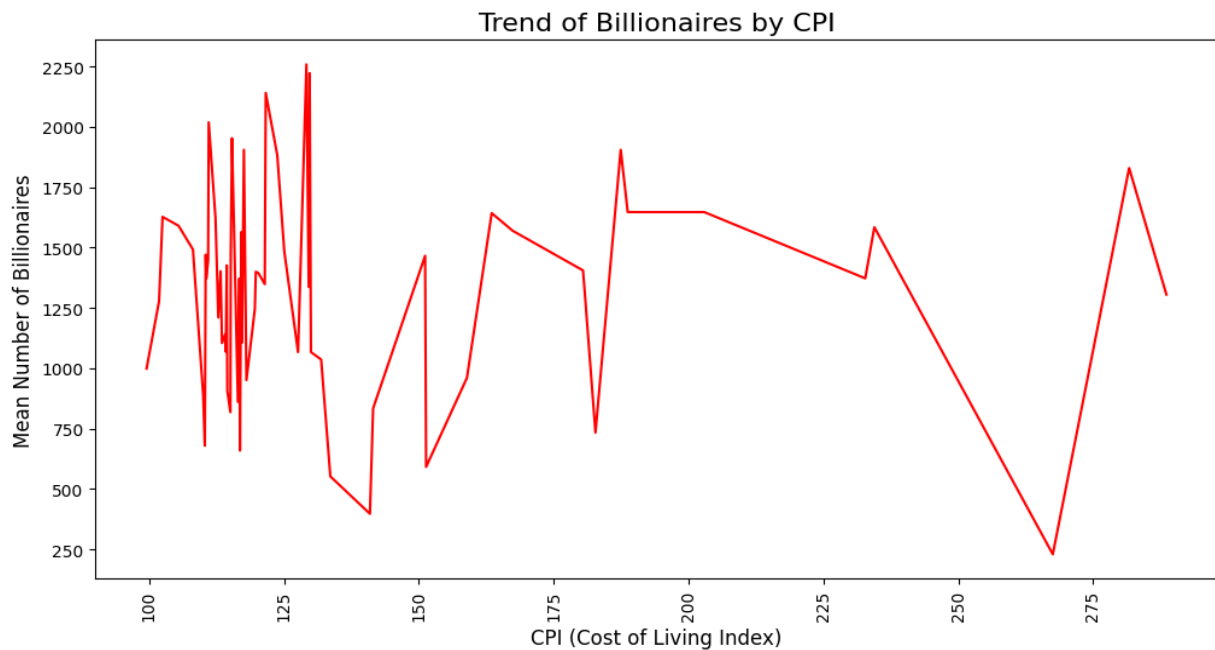
##### *Top Sources of Wealth:*

- The bar chart visualizes the top sources of wealth among billionaires, with 'Finance & Investments' being the most common source, followed by 'Technology' and 'Fashion & Retail.'



### CPI and Number of Billionaires:

- The line plot illustrates the trend of billionaires concerning the Cost of Living Index (CPI).



### Population and Number of Billionaires:

- A correlation coefficient of **0.12** suggests a weak positive correlation between a country's population and the number of billionaires.

### Average Total Tax Rate:

- In countries with the most billionaires, the average total tax rate is approximately **44.76%**, indicating variations in tax policies among wealthy nations.

### Wealth Distribution:

- The dataset introduces a new column, 'finalWorth\_round,' and calculates the total wealth and its percentage distribution by country.

	country	finalWorth	Percentage
0	Algeria	4600	0.04

1	Andorra	1500	0.01
2	Argentina	11000	0.09
3	Armenia	1200	0.01
4	Australia	173500	1.42
..	...	...	...
73	United Kingdom	370700	3.04
74	United States	4592100	37.62
75	Uruguay	1800	0.01
76	Uzbekistan	14400	0.12
77	Vietnam	12600	0.10

### Country-wise Wealth Distribution:

- Notable countries like the United States, China, and India exhibit a significant share of total wealth, highlighting the concentration of billionaires in certain regions.

### Conclusion:

- The analysis provides valuable insights into the distribution and factors influencing billionaire wealth globally, emphasizing the need for comprehensive and multi-faceted investigations in the field of wealth economics.

### \*Billionaires Insights Report\*

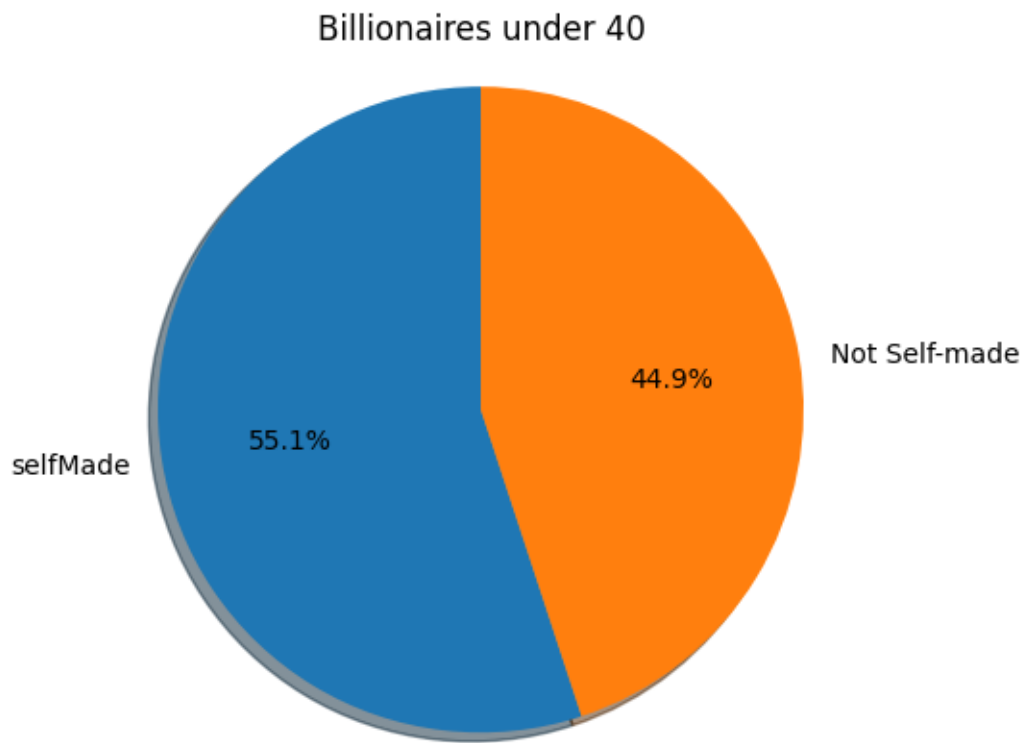
### Geographical Distribution:

- The map displays the geographical distribution of billionaires globally, emphasizing clusters in specific regions. The clustering analysis using DBSCAN further illustrates the spatial patterns of billionaires which cannot be displayed here.

### Demographic Analysis:

- There are 89 billionaires under the age of 40, indicating a considerable presence of young billionaires in the dataset.

- Among these, 49 are self-made, highlighting the entrepreneurial spirit among young billionaires.

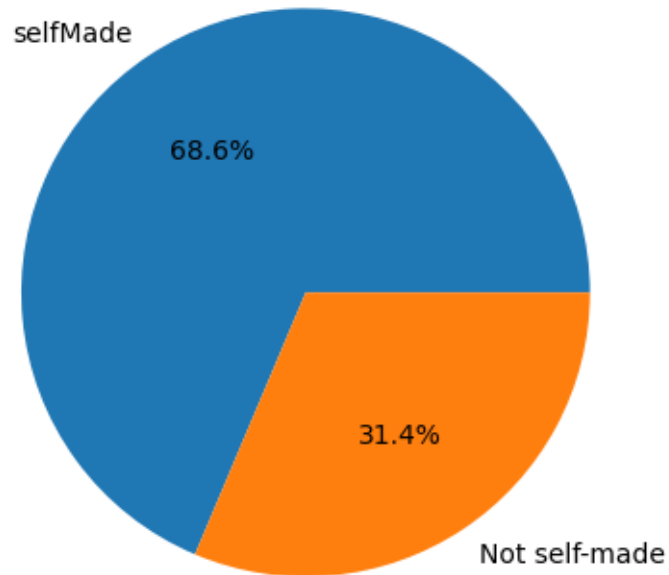


**Wealth and Self-Made Status:**



- A pie chart depicts the proportion of self-made billionaires among all billionaires. Approximately 68.6% are self-made, showcasing the significance of individual efforts in

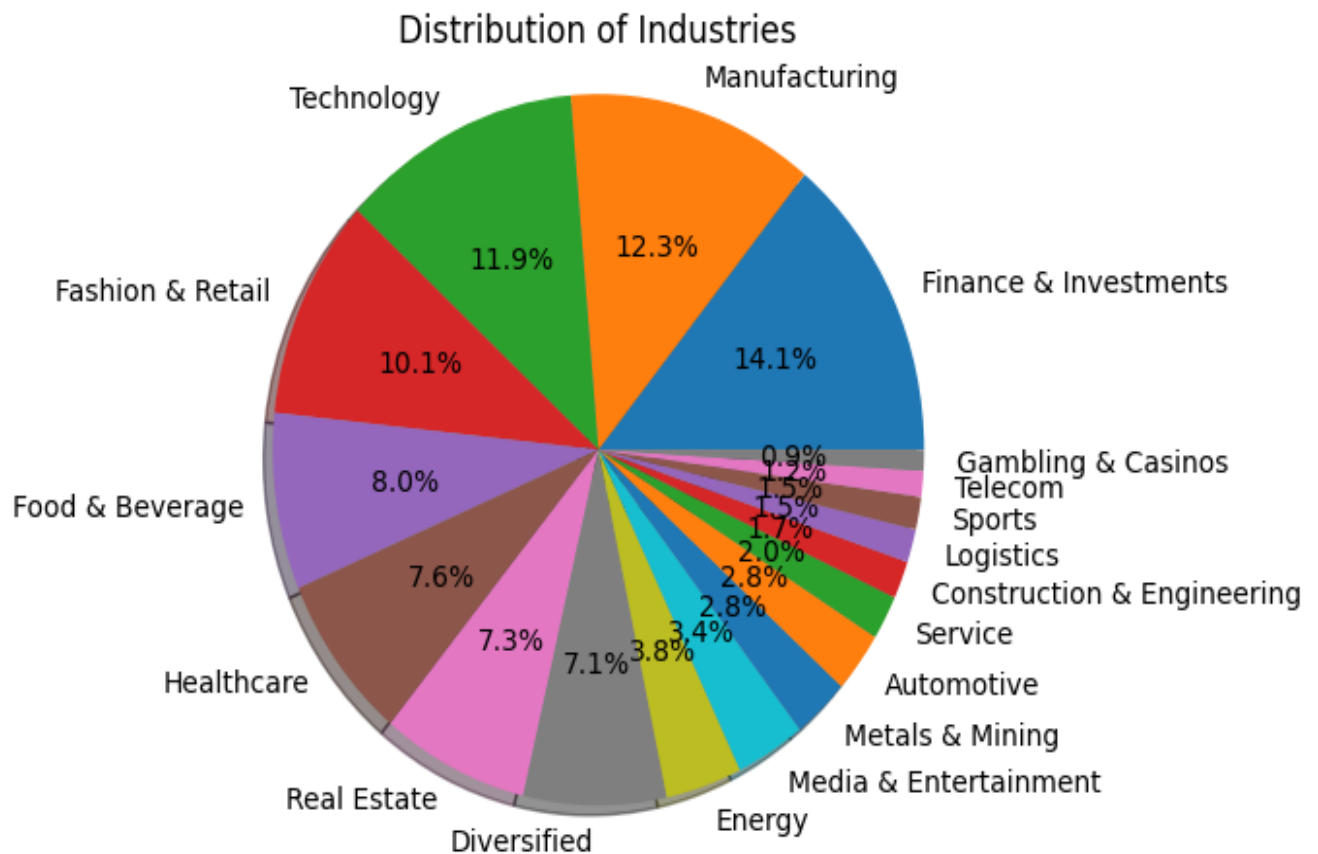
### Self-made vs All Billionaires



accumulating wealth.

### Industries Overview:

- The dataset is dominated by billionaires in the 'Finance & Investments,' 'Manufacturing,' and 'Technology' sectors.
- The pie chart visually represents the distribution of billionaires across various industries, providing insights into the diversification of wealth.



#### Conclusion:

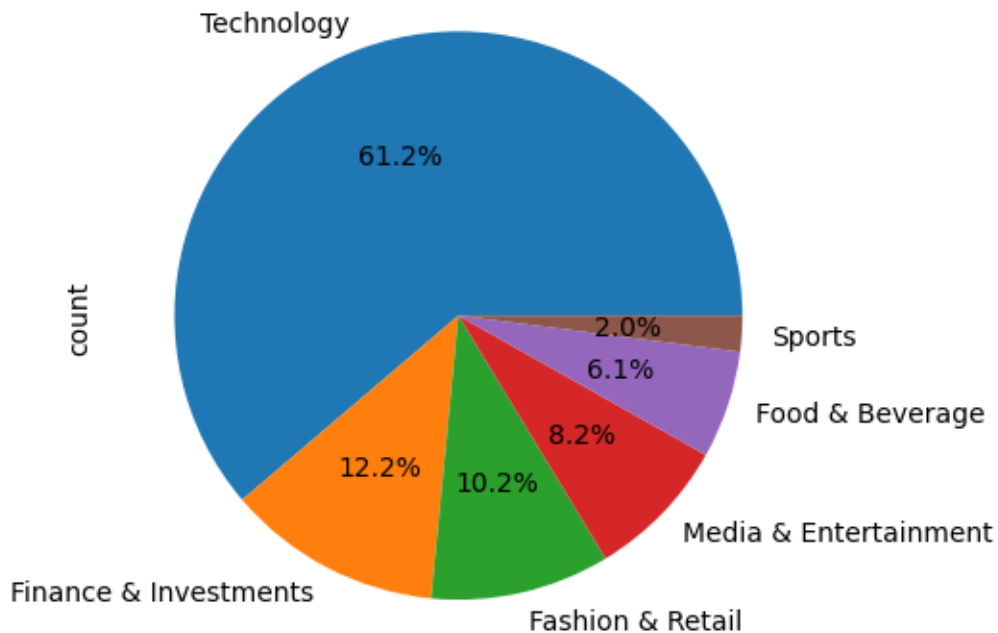
- The analysis offers valuable insights into the demographic and wealth characteristics of billionaires, providing a foundation for more detailed explorations. Understanding the geographical, demographic, and industrial dimensions of billionaire wealth is crucial for comprehending global economic dynamics.

#### **\*Self-Made Billionaires Under 40: Industries and Wealth Correlations\***

##### *Industries of Self-Made Billionaires Under 40:*

- The pie chart illustrates the distribution of industries among self-made billionaires under the age of 40. The majority are involved in 'Technology,' 'Finance & Investments,' and 'Manufacturing,' showcasing the dominance of these sectors in fostering young entrepreneurs.

## Industries of self-made billionaires under 40



### Pearson correlation coefficients between wealth and various economic indicators.

1. \*\*Pearson Correlation between Wealth and GDP: 0.02\*\*

- Interpretation: A correlation coefficient of 0.02 indicates a very weak positive correlation between wealth and Gross Domestic Product (GDP). The correlation is close to zero, suggesting that there is almost no linear relationship between the two variables.

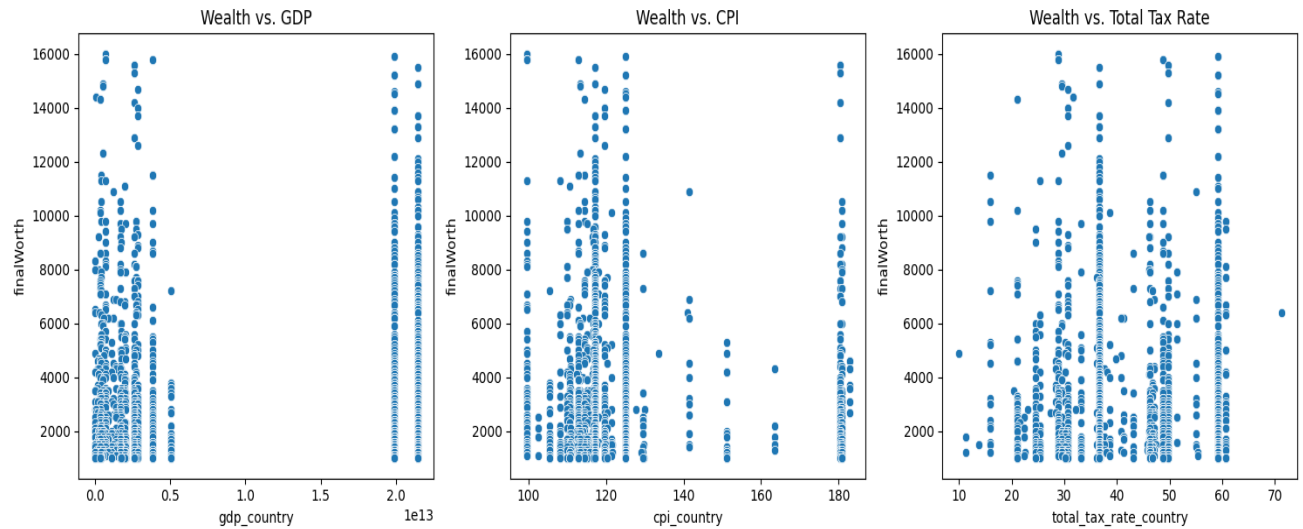
2. \*\*Pearson Correlation between Wealth and CPI: -0.07\*\*

- Interpretation: A correlation coefficient of -0.07 suggests a weak negative correlation between wealth and the Consumer Price Index (CPI). Although the correlation is still relatively close to zero, the negative sign implies a slight tendency for wealth to decrease as CPI increases, and vice versa.

3. \*\*Pearson Correlation between Wealth and Total Tax Rate: -0.10\*\*

- Interpretation: A correlation coefficient of -0.10 indicates a weak negative correlation between wealth and the total tax rate. Similar to the previous case, the negative

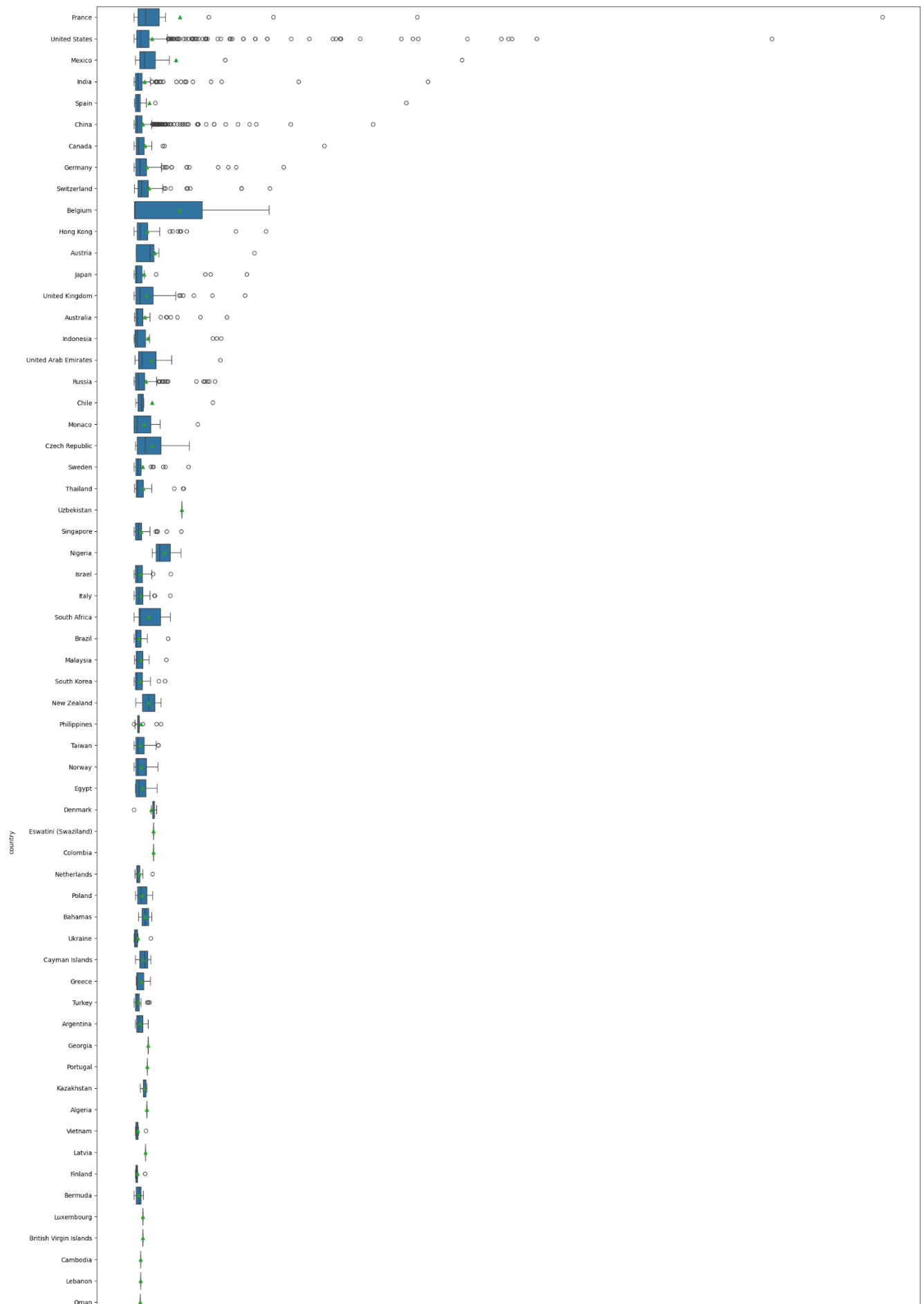
correlation suggests that there might be a slight tendency for wealth to decrease as the total tax rate increases.



### Country-wise Wealth Distribution:

- A box plot provides insights into the distribution of final worth among different countries. This visualization helps identify variations in wealth accumulation across

nations, with mean values indicated for each country.



### ### Correlation Analysis: Education and Wealth

#### #### Pearson Correlation Coefficients

The dataset reveals interesting insights into the correlation between wealth and education enrollment at the tertiary and primary levels:

#### #### Wealth and Tertiary Enrollment:

- \*\*Pearson Correlation Coefficient: 0.09\*\*
- A positive correlation coefficient of 0.09 suggests a modest, positive relationship between an individual's wealth and the tertiary education enrollment rate of their respective country.
- While the correlation is not exceptionally strong, it indicates that, on average, countries with higher tertiary education enrollment rates may have individuals with higher wealth. However, causation cannot be inferred solely from correlation.

#### #### Wealth and Primary Enrollment:

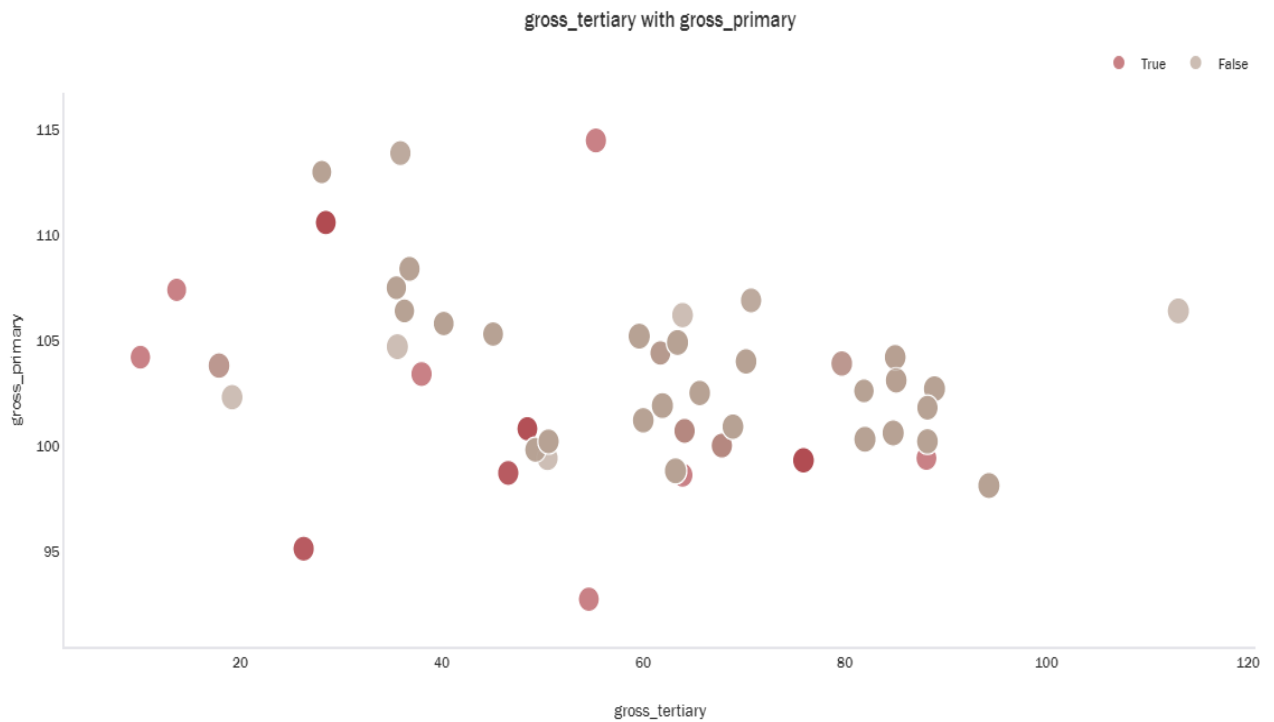
- \*\*Pearson Correlation Coefficient: 0.01\*\*
- The correlation coefficient of 0.01 implies a very weak positive correlation between wealth and the primary education enrollment rate.
- In this case, the correlation is close to zero, suggesting that there is almost no linear relationship between an individual's wealth and the primary education enrollment rate in their country.

### ### Interpretation:

- The positive correlations, though relatively small, hint at potential associations between wealth and education. Higher wealth might be associated with higher tertiary education enrollment, indicating a possible trend where wealthier individuals are more likely to pursue higher education.

### ### Conclusion:

Understanding the correlation between wealth and education enrollment is crucial for policymakers and researchers. While these correlations provide valuable insights, further investigation and consideration of multiple factors are essential for a holistic understanding of the dynamics at play in wealth and education relationships.

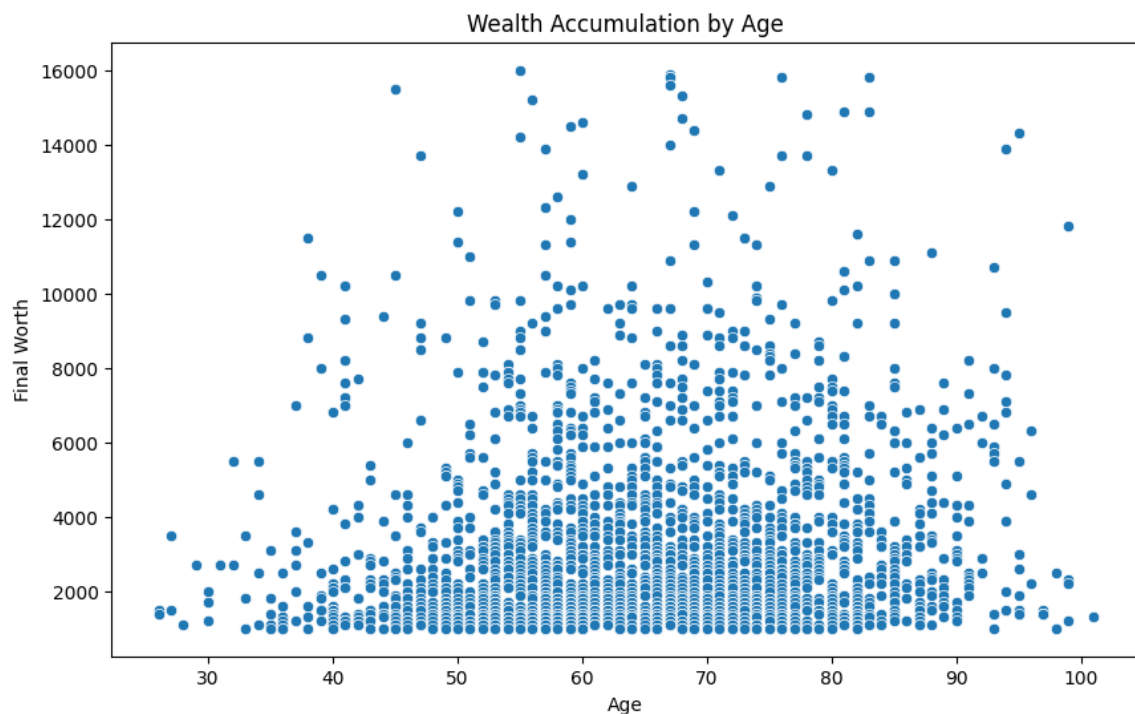


## PEARSON CORRELATION BETWEEN AGE AND WEALTH



The pearson correlation between age and wealth is 0.11. This value indicates a weak positive linear relationship between the two variables. The correlation coefficient ranges from -1 to 1, where 1 signifies a perfect positive correlation, 0 indicates no correlation, and -1 represents a perfect negative correlation. In this case, the positive correlation suggests that, on average, as age increases, wealth tends to increase slightly. However, the strength of this relationship is considered weak.

### A SCATTER PLOT TO VISUALLY EXPLORE THE RELATIONSHIP BETWEEN AGE AND WEALTH IN THE DATASET



**\*\*Conclusion: Data Science Report on Billionaires Statistics Dataset\*\***

In this comprehensive data science report, we undertook a thorough examination of a dataset focused on billionaires' statistics. Our analysis covered various dimensions, providing valuable insights into the demographics, wealth distribution, geographic trends, and correlations within the dataset.

**\*\*Key Findings:\*\***

1. **Wealth Distribution:**

- The dataset encompasses a broad spectrum of wealth categories, with individuals representing diverse industries and sectors.
- Finance & Investments, Manufacturing, and Technology emerge as dominant sectors in terms of billionaire representation.

2. **Demographic Insights:**

- Demographic features, including age and gender, were explored to understand the composition of the billionaire dataset.
- The average age of billionaires is approximately 65 years, with a notable gender imbalance, where males significantly outnumber females.

3. **Geographic and Economic Trends:**

- Geographic data, as represented by latitude and longitude, provides a global perspective on the distribution of billionaires.
- Economic indicators, such as GDP, CPI, and tax rates, exhibit substantial variations across the represented countries.

4. **Correlation Analysis:**

- Correlation analysis revealed a weak positive correlation between wealth and tertiary education enrollment (0.09) and a minimal correlation with primary education enrollment (0.01).

5. **Outlier Handling:**

- Outlier removal was employed to refine the dataset, resulting in 2,334 observations for more focused analyses.

6. **Visual Insights:**

- Visualizations, including the scatter plot examining the relationship between age and wealth, provided initial insights into distribution patterns.

**\*\*Implications and Future Directions:\*\***

- This dataset serves as a valuable resource for gaining a deeper understanding of wealth accumulation, demographic trends, and economic interplays among billionaires.
- Future analyses could explore specific industry trends, regional wealth dynamics, and the impact of external factors on billionaires' financial standing.

In conclusion, our data science report offers a comprehensive exploration of the billionaire statistics dataset, shedding light on critical aspects that contribute to the broader understanding of wealth distribution and demographic characteristics among the world's wealthiest individuals. Further research avenues may uncover more nuanced patterns and contribute to a richer comprehension of this dynamic landscape.