**Exploratory Data Analysis on the Billionaire Dataset**

Ajla Korman1

1 International Burch University, Faculty of Engineering, Natural and Medical Sciences, Sarajevo, Bosnia and Herzegovina  
ajla.korman@stu.ibu.edu.ba

**Abstract.** This research paper analyzes billionaire wealth distribution through several different segments. Examined dataset uncovered important information about the most profitable industries, geographic wealth distribution, and gender difference. Results showed that the largest amount of money is held in the hands of a dozen people. The study also confirmed two major discoveries from earlier - most billionaires are still self - made and male billionaires outnumber female ones.

**Keywords:** Billionaires, Wealth Distribution, Age, Gender Demographics, Geographical Distribution

# Introduction

Wealth distribution whether that be in one country, on the continent, or on a global scale is a topic discussed in many social sciences. The count of billionaires in one country is highly significant since it shows the country's current economic state and points out investments that could be highly feasible. Billionaire is a term used to describe a special minority of rich individuals whose net worth surpasses an upper limit of the average financial means of one individual based on certain criteria. Most of billionaires' money is earned through manufacturing industry sectors together with product placement on the worldwide market. Since billionaires place their products on the worldwide market, their investment decisions are of high importance not only for their companies' further development but could also affect the economic growth and transformation of that region.

This research will use a dataset that will provide details about billionaires' net worth together with information about their businesses, as well as gender and age demographics and economic data inside of the country they reside. Analysis of this dataset will be beneficial to understanding critical factors that enable billionaires to accumulate large sums of capital.

An exploratory data analysis (EDA) will be used on the billionaire dataset to discover essential distribution patterns of wealth between industries and regions as well as explain gender and age characteristics. Dataset inspection will help to answer questions, such as which business sectors generate the highest numbers of billionaires, is billionaires' wealth inherited or self-made, does the billionaire demographic differs across different sectors based on their age.

This research will use step-by-step data preprocessing and analyzing methods to examine the billionaire population. The aim will be to reveal all elements that enable people to create financial abundance for themselves.

# Literature Review

Authors of the paper “The Origins of the Superrich” [1] Freund and Oliver created a billionaire database from where it was possible to conclude that technological and financial institutions are more profitable and therefore enable the creation of a larger number of billionaires in comparison to the real estate and manufacturing industries.

Additionally, Piketty analyzed wealth accumulation and inheritance of wealth in families in his famous book “Capital in the Twenty-First Century” [2]. It was concluded that it is easier for people who already have wealth to continue accumulating it or find it easier to establish a strong economic position in the market.

The impact of billionaires on economic development has been a part of a heated discussion. Bagchi and Svejnar [3] explore in their work how billionaire concentration in one location influences the economic growth of that region. It was discussed that even though billionaires' investments can help one region financially flourish, that will also negatively affect the broader region since all the investments would be focused on one small part.

A major gender gap between billionaires was identified in the work “Billionaires – Reflections on the Upper Crust” [4] by West. Due to historical, socioeconomic, and political factors, male billionaires were dominating all industry sectors.

The percentage rate between self-made and inherited wealth was discussed in Zucman’s “The Hidden Wealth of Nations” [5]. The wealth that passes through generations raises the social standards to a point where less wealthy people find it hard to reach these social levels. On the contrary, the rapid development of the technological and financial sectors has enabled many people to earn massive amounts of money through different startups.

# Methods and Material

Billionaire Statistics Dataset [6] was used for the analysis. This dataset includes various attributes related to this selected population, such as their total net worth, industry sectors in which they operate, countries of residency, and countries' economic data such as GDP, taxes, and information on whether the billionaire's wealth was inherited or self-made.

To perform this analysis, various Python libraries were used to analyze the dataset and provide graphs. For data preprocessing and analysis pandas library was used. Numpy was used for mathematical operations on a large number of records. For the creation of graphs, mathplotlib was used.

## 3.1 **Dataset Attributes**

The dataset originally contained a wide range of columns (35 columns), as detailed in the table below.

### Table 1. Original dataset columns

| **Column Name** | **Column Meaning** |
| --- | --- |
| rank | Billionaire position based on wealth |
| finalWorth | The billionaire’s total net worth (in billions of USD) |
| category | The broad classification of the billionaire’s industry |
| personName | The full name of the billionaire |
| age | The age of the billionaire at the time of data collection |
| country | The country where the billionaire primarily resides |
| city | The city where the billionaire primarily resides |
| source | The main company or asset responsible for the billionaire’s wealth (e.g., Amazon, Tesla) |
| industry | The specific industry in which the billionaire has accumulated wealth |
| countryOfCitizenship | The country where the billionaire holds official citizenship |
| organization | The main organization or business associated with the billionaire. |
| selfMade | Whether the billionaire is self-made (True) or inherited (False) |
| status | The state of billionaire |
| gender | The gender of the billionaire (Male/Female) |
| birthDate | The full birthdate of the billionaire |
| lastName | The billionaire’s last name (if applicable) |
| firstName | The billionaire’s first name (if applicable) |
| title | Official role inside of the company |
| date | The date when the billionaire’s wealth was last updated |
| state | The state where the billionaire resides (if available) |
| residenceStateRegion | The region or province where the billionaire lives |
| birthYear | The year of birth of the billionaire |
| birthMonth | The month of birth of the billionaire |
| birthDay | The day of birth of the billionaire |
| cpi\_country | The Consumer Price Index (CPI) of the billionaire’s country (indicates inflation level) |
| cpi\_change\_country | The inflation trends |
| gdp\_country | The Gross Domestic Product (GDP) of the billionaire’s country. |
| gross\_tertiary\_education\_enrollment | The percentage of the population enrolled in higher education in the billionaire’s country. |
| gross\_primary\_education\_enrollment\_country | The population percentage with primary education. |
| life\_expectancy\_country | The average life expectancy in the billionaire’s country. |
| tax\_revenue\_country\_country | The GDP percentage |
| total\_tax\_rate\_country | The total tax rate in the billionaire’s country |
| population\_country | The total population of the billionaire’s country |
| latitude\_country | The latitude coordinate of the country |
| longitude\_country | The longitude coordinate of the country |

3.2 **Dataset Selection**

Even though the dataset originally had 35 columns, not all were relevant for analysis. Therefore, 20 columns were removed, and the analysis was focused on the 15 other columns that provided the most significant information based on the research goal. These 15 selected attributes are:

* rank
* personName
* finalWorth
* age
* gender
* industries
* countryOfCitizenship
* selfMade
* gdp\_country
* tax\_revenue\_country
* total\_tax\_rate\_country
* population\_country
* source
* organization
* city

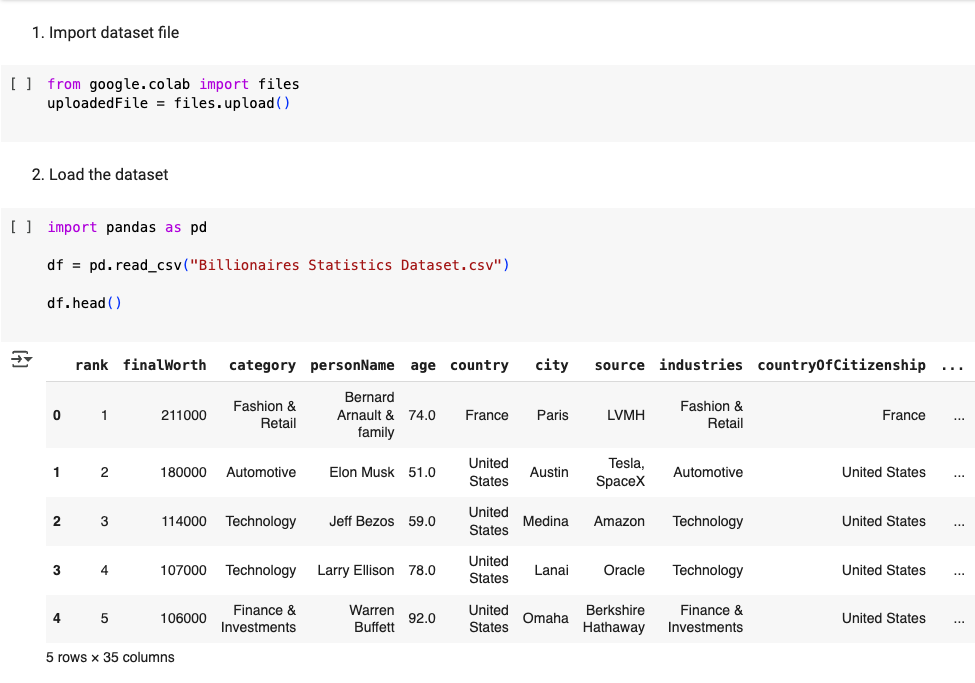
# 4 Dataset Preprocessing and Cleaning

To get precise and correct results, firstly preprocessing steps were taken, such as:

* Handling missing values
* Dropping unnecessary columns
* Transforming data types
* Fetching missing values using APIs
* Removing duplicates and final dataset adjustments

Alongside this, the dataset originally contained 2640 rows. where empty values existed, those values got removed. Afterward, the dataset had 2457 rows.

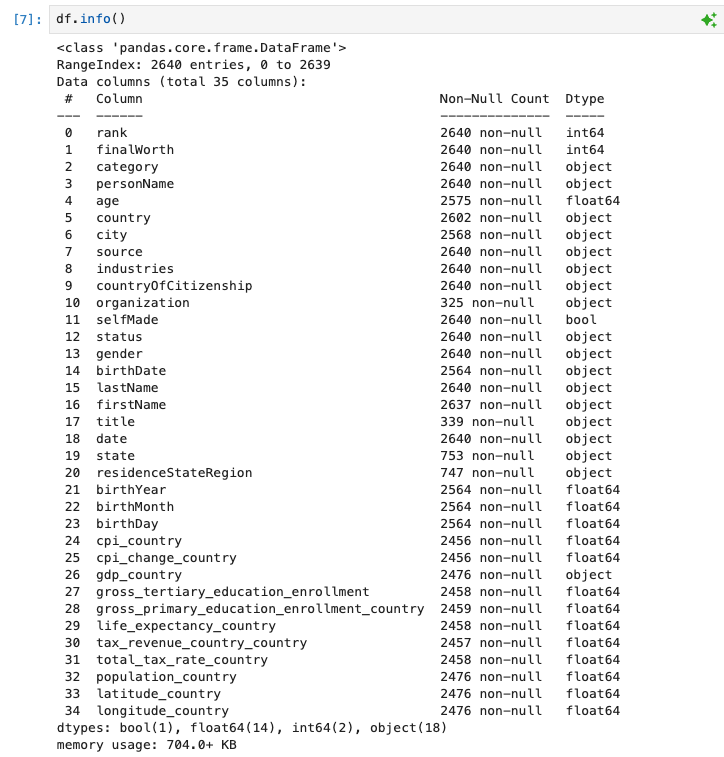
### Figure 1. Original dataset presentation

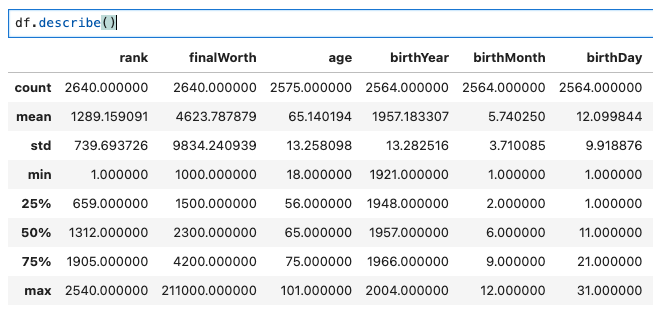


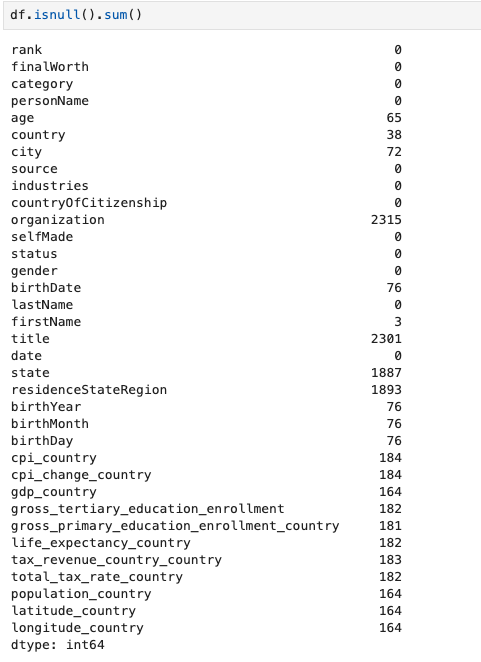
## 4.1 Initial Data Inspection

To check column names, missing values, and data types, commands as shown in the figures below. This helped to identify unnecessary columns and missing values.

### Figure 2. Initial Dataset Inspection



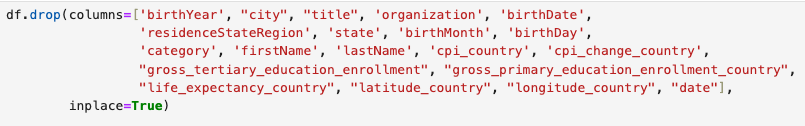




## 4.2 Dropping Unnecessary Columns

As stated earlier, 20 columns were removed. They were either removed since they had large amounts of records missing data or the information inside of them was not necessary for the analysis at all.

### Figure 3. Removing Unnecessary Columns



## 4.3 Handling Missing Values

A minor number of records in certain columns had missing values, therefore these rows were removed so as not to affect the final analysis results.

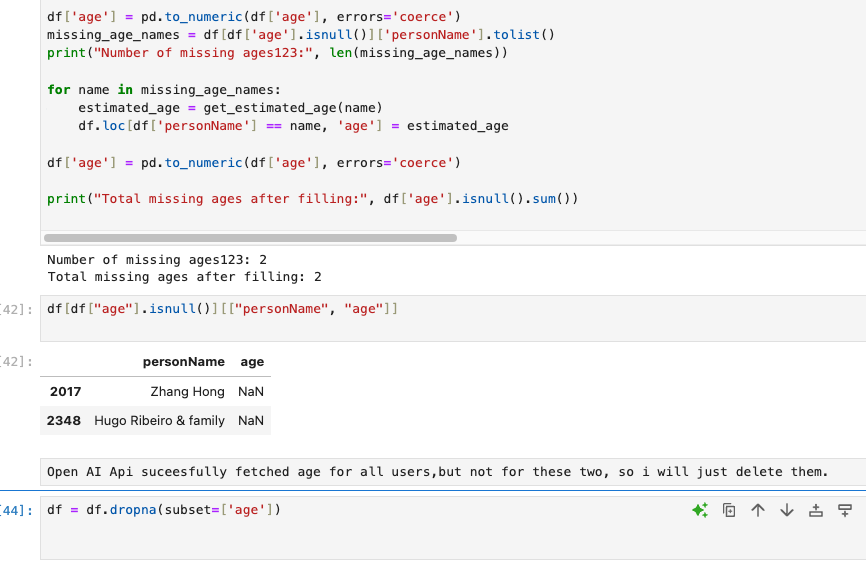
### Figure 4. Handling Missing Values



## 4.4 Filling Missing Age and Economic Indicator Values Using APIs

For 53 billionaires, their ages were missing. Age information was a crucial part of the analysis. Therefore, to make sure that the results are as precise, as possible, values from columns first, last, and company name were used in combination with OpenAI API to get ages for those missing records. The API successfully retrieved ages for most billionaires, but it failed for a minor number of records ( 2 in total), therefore records with those two individuals were completely removed. Missing economic values were filled by connecting with World Bank API.

### Figure 5. Filling Missing Values Using APIs

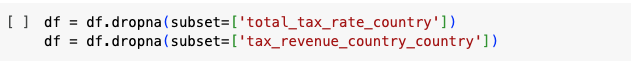


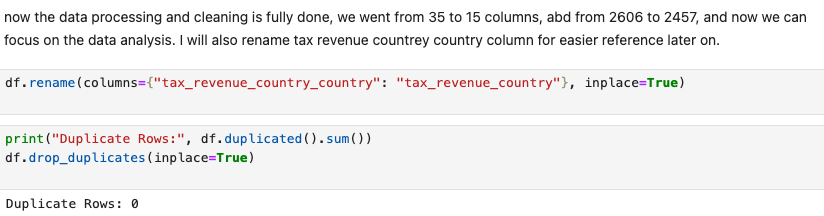
#### 

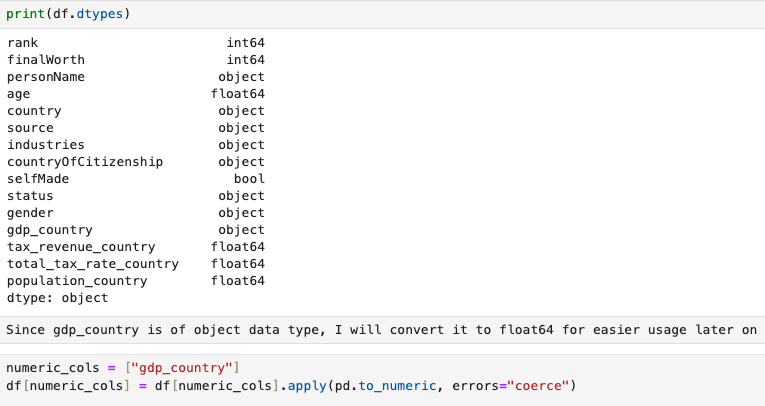
## 4.5 Final Cleaning and Data Type Fixes

In the end of data preprocessing, confusing column names were renamed for easier understanding throughout the whole data analysis process. Additionally, duplicate rows were removed, and all numeric columns were converted to the most suitable number data type for easier usage later on.

### Figure 6. Final Clean Up







# 5 **Exploratory Data Analysis (EDA)**

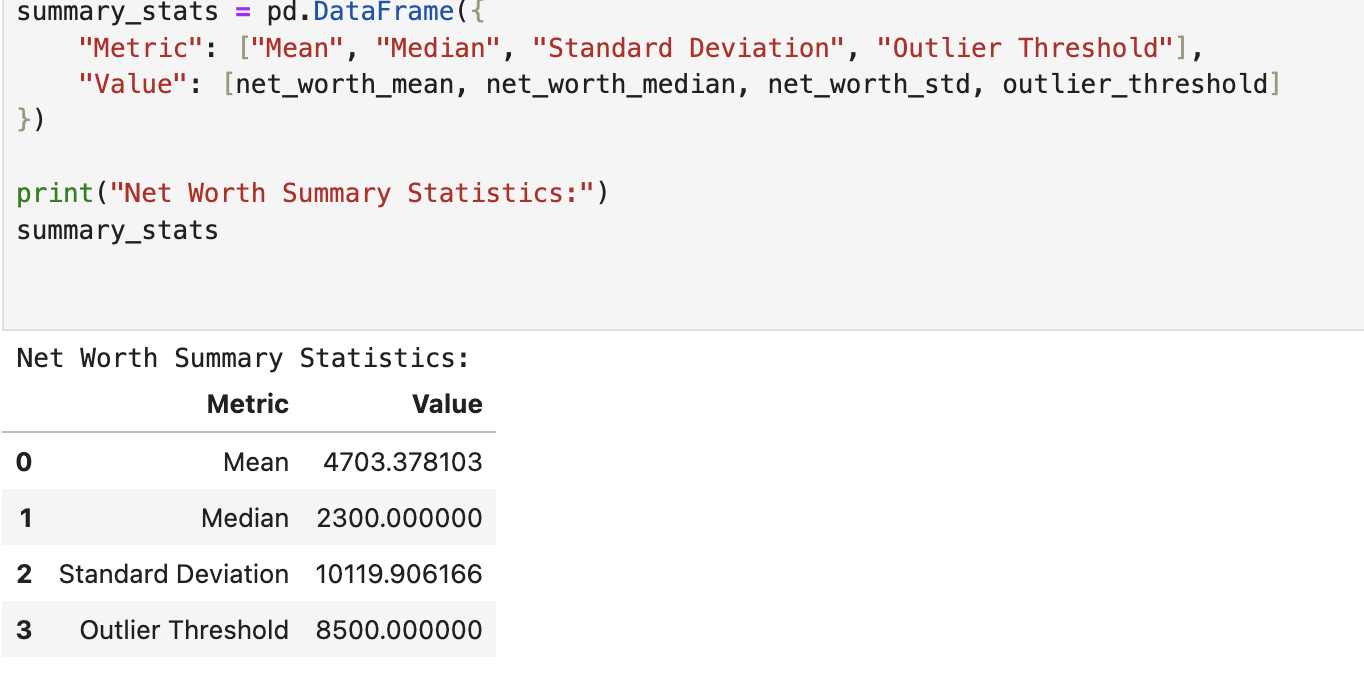
This section will provide the results of the data analysis that will be necessary to determine key factors such as how billionaires are made, the most profitable industries, age and gender demographics, and similar.

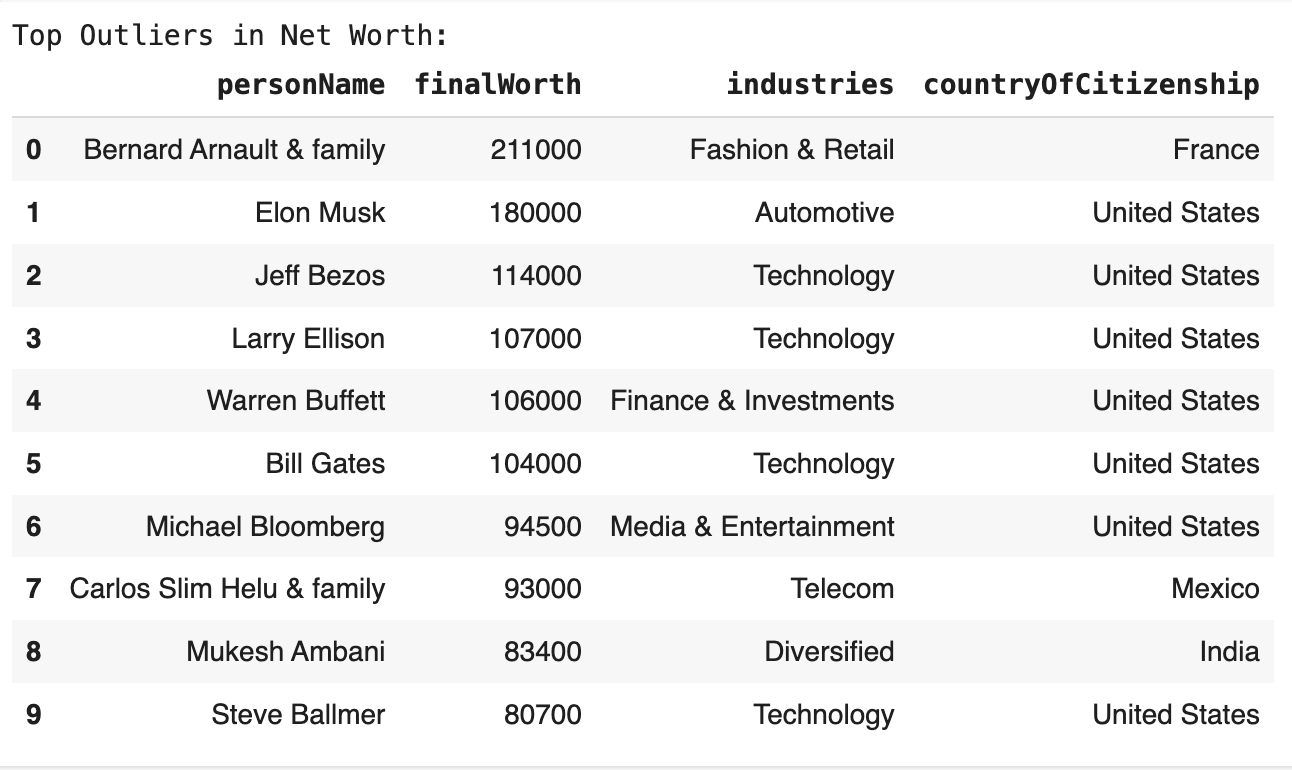
## 5.1 **Net Worth Analysis**

Here, billionaires' net worth distribution was analyzed. It was researched what countries have the highest number of billionaires. Additionally, the top ten billionaires who had extremely more than others were identified as outliers.

**Summary Statistics.** To get an overview of a billionaire's net worth, the average billionaire's wealth was calculated using the mean. Additionally, to identify the top ten billionaires standard deviation and interquartile range methods were used.

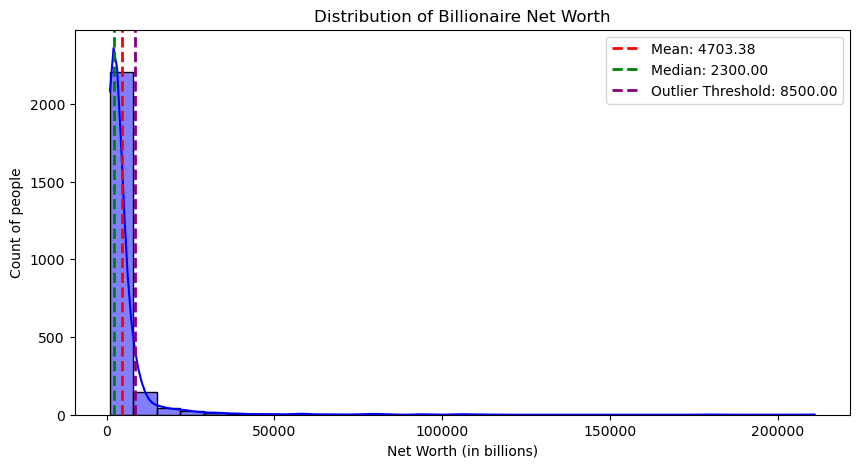
### Figure 7. Billionaire Statistics and Outliers





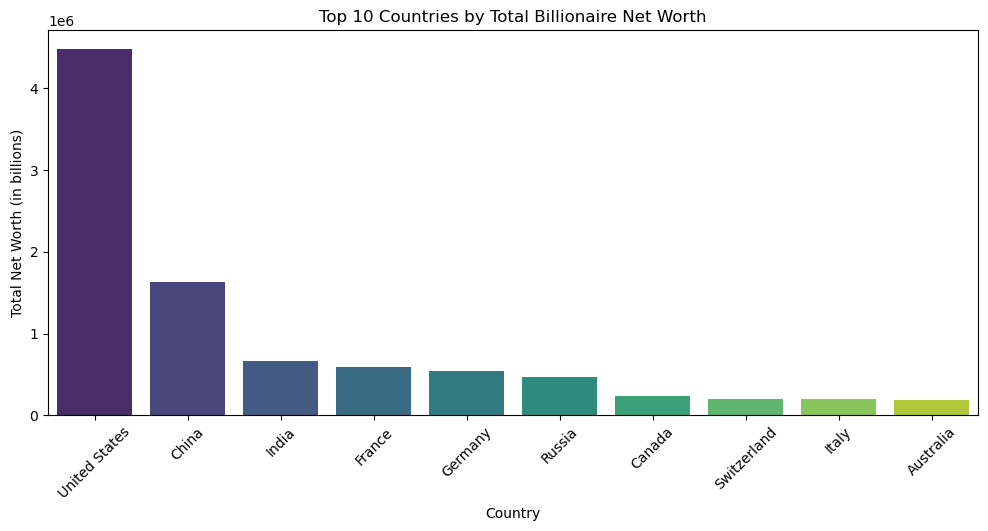
**Visualize Net Worth Distribution.** The billionaire wealth distribution is shown through a histogram. The wealth distribution shows strong skewness because massive sums of money belong to only a small number of billionaires while the rest possess much more modest fortunes. Vertical lines demonstrate the positions of the mean value and median value together with the outlier threshold.

### Figure 8. Distribution of Billionaire Net Worth



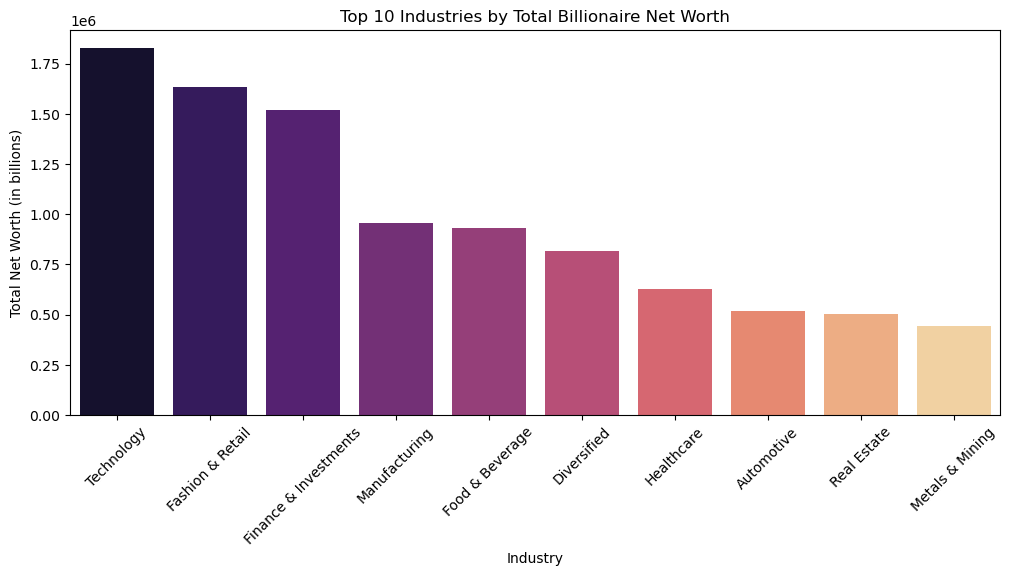
**Countries with the Highest Total Net Worth.** A global overview of wealth distribution focused on identifying which 10 countries possess the biggest amount of billionaires. The United States and China lead in total billionaire net worth.

##### Figure 9. Distribution of Billionaires per Country



**Industries with the Most Billionaire Wealth.** Sectors in which billionaires work were examined to determine the top ten leading money-making industries. Technology, Finance, and Retail are declared as the most lucrative industries.

### Figure 11. Top 10 Industries by Total Billionaire Net Worth

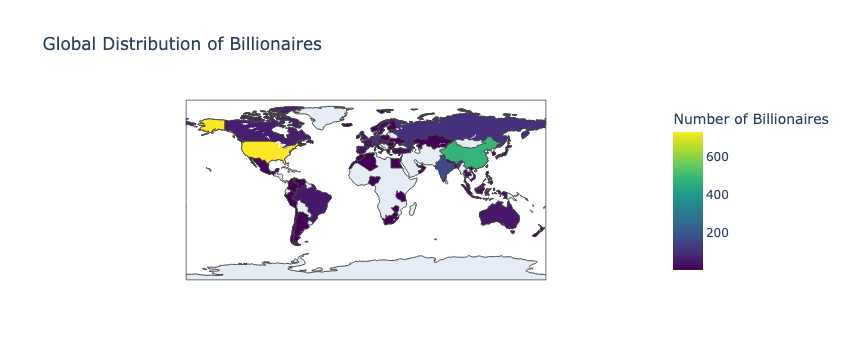
****

**5.2 Geographical Distribution**

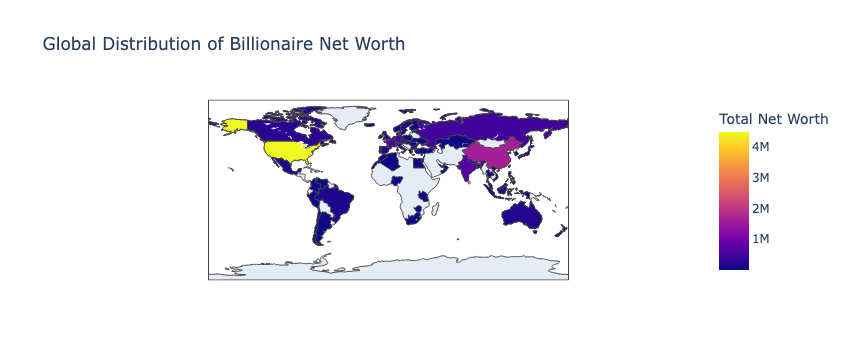
Complete geographical analysis shows how the billionaire's wealth is distributed worldwide. The first choropleth map shows number of billionaires per country, while the second map shows billionaire wealth totals per country.

The analysis has shown that the United States together with China are residencies for the highest number of billionaires. The United States stands apart from other nations because it holds the greatest total billionaire wealth due to its dominant economic position on the market and worldwide. India alongside Germany and Russia joins the United States and China in their population of billionaires which indicates that emerging markets are also creating billionaires and that the Western market does not hold a monopoly on billionaires creation. A lower number of billionaires in poorer countries tends to correspond to diminished total national wealth because economic power is controlled by dominant nations.

### Figure 12. Global Distribution of Billionaires



### Figure 13. Global Distribution of Billionaire Net Worth

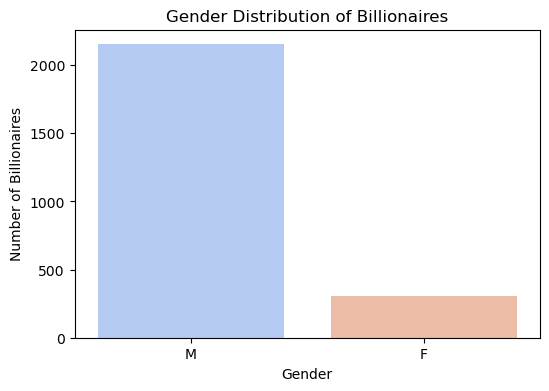


## 5.3 Gender Analysis

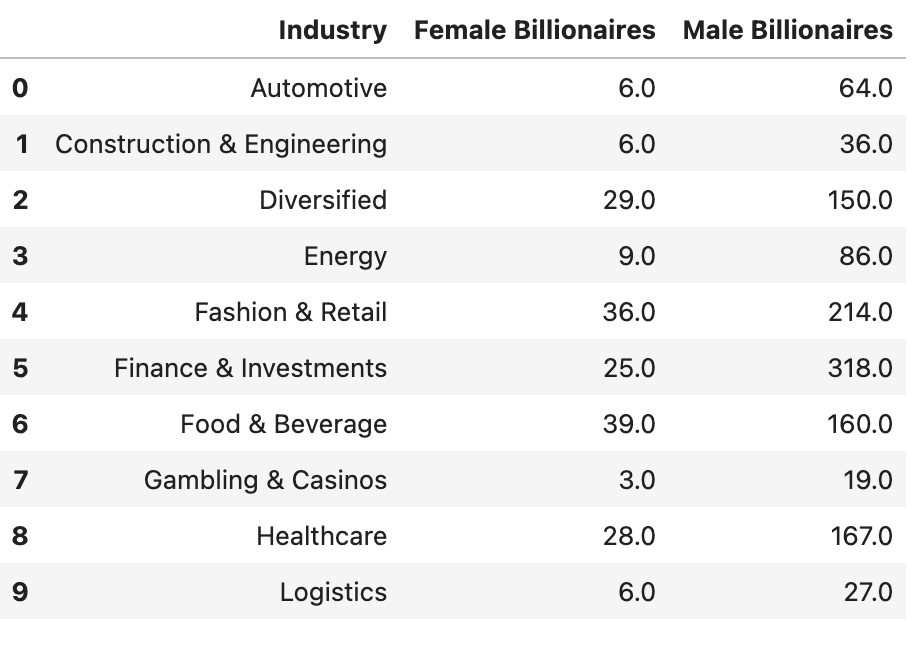
The gender distribution analysis has revealed a significant difference between the number of male and female billionaires. The majority of billionaires are males. The reasons for this inequality could be traced to the combination of historical and socio-economic circumstances, and industry patterns.

Furthermore, almost every business sector demonstrates male billionaire dominance. fashion, food, and retail industries maintain higher portions of billionaire women than the automotive, energy, finance and investment industries which favor male entrepreneurs. Female entrepreneurs find better chances to succeed in particular business sectors.

### Figure 14. Gender Distribution of Billionaires



### Figure 15. Rate of Billionaires per Gendre in Top 10 Industries



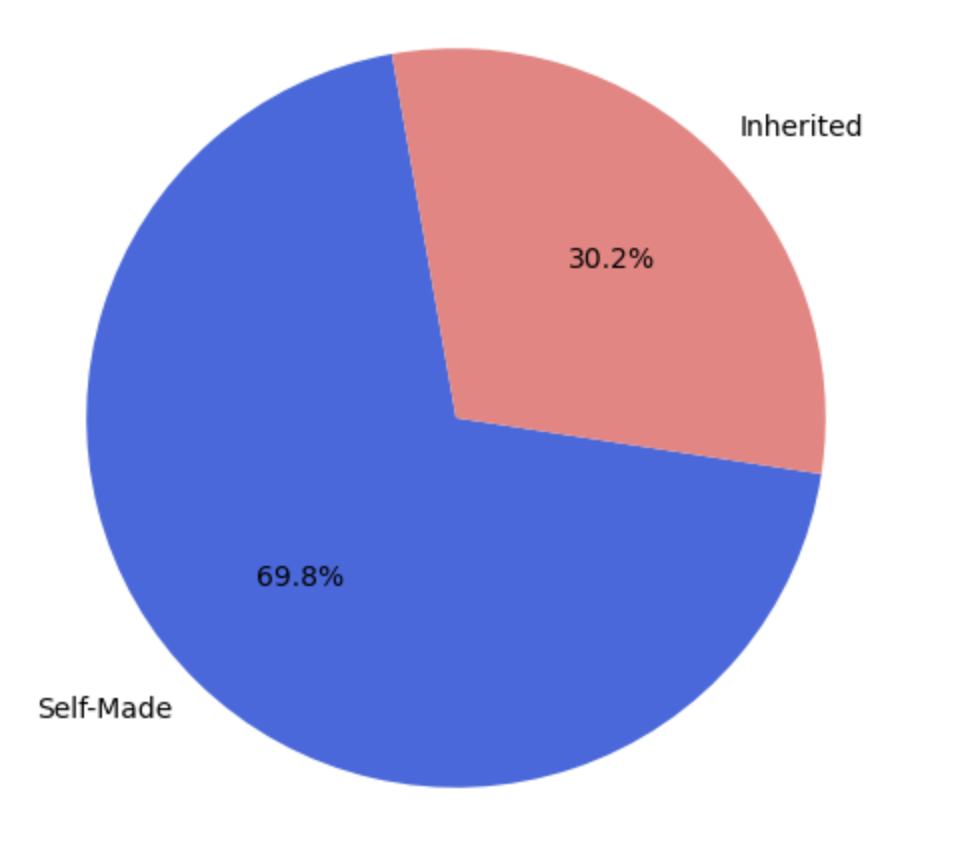
## 5.4 Source of Wealth

Self-made billionaires represent the largest group of billionaires according to the data analysis. According to the pie chart, self-made billionaires account for 69.8 percent of the total whereas inheritors of wealth represent 30.2 percent. The statistics demonstrate that a major portion of billionaires achieved their wealth without receiving inherited money.

Certain business sectors contain a higher number of self-made billionaires. Technology, finance, and investment together with manufacturing prove to be industries where most billionaires obtain their wealth through self-generated resources. Fashion and retail together with real estate contain a bigger proportion of inherited money compared to other industries.

The existence of the entrepreneurial culture in China and the United States is confirmed by the highest amount of self-made billionaires that reside in both of these countries. An equal number of self-made and inherited billionaires is visible in countries such as Germany and Russia. Additionally, India has an increasing number of self-made billionaires which shows that startup culture is constantly being developed there.

### Figure 16. Self-made vs inherited billionaire wealth chart



### 

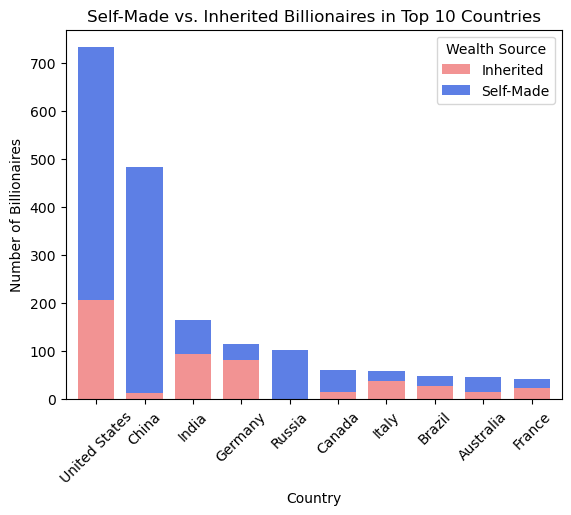
### 

### 

### Figure 17. Self-made vs inherited billionaire per industry



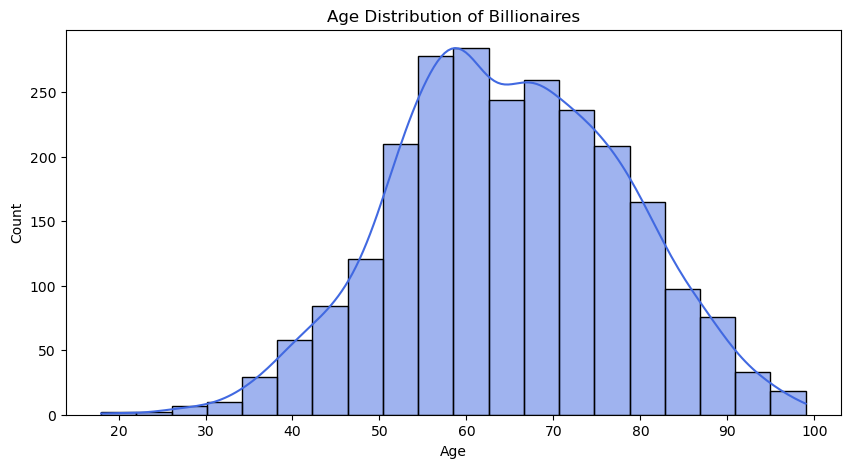
### Figure 18. Self-made vs inherited billionaire per country



## 5.5 Age Distribution

Most of the billionaires are around the age of 60. When this is plotted on a histogram, it shows a perfect bell curve distribution pattern because billionaire success does not happen overnight, rather it takes decades of work to achieve it.

### Figure 19. Age Distribution Between Billionaires



The analysis also examined the link between the industry sector, its average net worth, and the billionaires' average age who work in this industry. This has brought up interesting conclusions, such as that the automotive, telecom and logistics sectors contain the largest number of senior billionaires indicating that these sectors require extensive experience and substantial financial resources if you want to achieve success in these fields. The technology industry contains the youngest group of billionaires who average is 56 years old. This aligns with the fact, that today tech entrepreneurs with startups can easily get rich following current trends. Older billionaires exist primarily in metals and mining along with energy and finance and investments sectors since both industries require significant market experience and long-term capital investment.

### Figure 20. Correlation between Industry, Average Net Worth, and Average Age

#### 

# 6 Results and Discussion

The analysis of billionaire net worth has shown that the biggest amount of money is held by a handful of people making the whole distribution uneven. Leaders in the field who have the biggest number of billionaire residents are the USA and China which confirms that their markets are ever-evolving both on domestic and international markets making space for high-profit businesses. This aligns with the Bagchi and Sevjnar’s work mentioned in the literature review, affirming once again, that high concentration of billionaires in one place makes other regions poor of investment opportunities. The research has also confirmed that the most popular industry sectors that allow people to financially succeed are IT, finance, and manufacturing.

Even though these three sectors have a lot of potential for innovation, still it takes a decent amount of time for one individual to achieve billionaire status in these sectors. Average age for a billionaire in tech industry appears to be 56 years old, which is not young, however it does implicate that is faster to become billionaire in this industry than in energy and automotive sectors where average age is 60 years old. Billionaire status in traditional business sectors depends on experience together with long-term investment approaches, but new technology along with startup culture make it possible for younger people to achieve billionaire status at a faster rate.

Additionally, the study has shown that 70% of billionaires are self-made with the majority of them owning businesses in one of the three sectors mentioned above. 30% of the inherited billionaires rely more on family businesses in the field of retail and fashion, meaning that traditional businesses have not died out, rather they have kept a balanced rate of development.

Furthermore, this research results confirmed gender difference results from previous researches as well. The gender gap stayed substantial, with a smaller increase of female billionaires in consumer - driven industries such as fashion, retail, food and beverage. Once again, factors for this seem to be of historical, and socioeconomic background making them hard for women to overcome them.

# 7 Conclusion

This research document explores the complete patterns which affect billionaire wealth distribution. It takes into account age and gender demographics, as well as billionaire geographical distribution and the both positive and negative economic impact it makes on local and broader region. Billionaires country of residence is also put into focus since its international placement has impact on potential opportunities for creating high-revenue businesses.

It was concluded that substantial profits come from innovations in industries who are of high importance to the current needs of people, such as development in technology, finance and manufacturing sectors, as well as combining innovations in two or more of these fields together.

Even though this research provides extensive information into current billionaire distribution pattern and their correlated attributes, it could be still improved. One area for improvement would be analysising wealth accumulation of billionaires over a period of couple of decades, and comparing high and lows in the wealth amount with thevpolitical and economical situation that was trending at the moment. Another potential improvement would be implementing machine learning techniques to predict future potential investments or industries who would have high turnover.

**References**

1. Kaplan, S., Rauh, J.: Origins of the Superrich: Billionaire Characteristics Database. Peterson Institute for International Economics.<https://www.piie.com/publications/working-papers/origins-superrich-billionaire-characteristics-database>, last accessed 2025/02/10.
2. Freeland, C.: Plutocrats: The Rise of the New Global Super-Rich and the Fall of Everyone Else. Harvard University Press, Cambridge (2012).<https://www.hup.harvard.edu/books/9780674430006>.
3. Bagchi, S., Svejnar, J.: Does Wealth Inequality Matter for Growth? The Effect of Billionaire Wealth, Income Distribution, and Poverty. Peterson Institute Working Papers (2013).
4. West, D.: Billionaires – Reflections on the Upper Crust. Brookings Institution Press, Washington, D.C. (2014).
5. Zucman, G.: The Hidden Wealth of Nations. University of Chicago Press, Chicago (2015).
6. Nelgiriya, W.: Billionaire Statistics Dataset. Kaggle,<https://www.kaggle.com/datasets/nelgiriyewithana/billionaires-statistics-dataset/>, last accessed 2025/02/10.
7. Korman, A.: Billionaire Statistics Dataset Analysis Code. Google Drive,<https://drive.google.com/file/d/1o09uO-ZSZpkg7Inwi1TfAo3mjC94LLCd/view?usp=sharing>, last accessed 2025/02/10.