Exploring Wealth Dynamics: A Comprehensive Big Data Analysis of Wealth Accumulation Patterns

Karim Mohammed Rezaul[1], Mifta Uddin Khan[2], Nnamdi Williams David[3], Kazy Noor e Alam Siddiquee[4], Tajnuva Jannat[5], Md Shabiul Islam[6]
Wrexham University, UK[1]
Northumbria University London, UK[2, 3]
Multimedia University, Malaysia[4]
Centre for Applied Research in Software & IT (CARSIT), UK[5]
Multimedia University, Malaysia[6]

***Abstract*—The study offers a thorough examination of the accumulation and distribution of wealth among billionaires through the application of big data analytics methodologies. This research centres on an extensive dataset known as "Billionaires.csv," [19] which encompasses a range of information about billionaires from diverse nations, including their demographic characteristics, company particulars, sources of wealth, and more details. The study aims to get a deeper understanding of the determinants that change the net worth of billionaires and detect trends in the worldwide financial system that can guide entrepreneurial ventures and investment possibilities. The dataset is subjected to analysis and visualisation through the utilisation of Python tools and libraries, including but not limited to Pandas, NumPy, Matplotlib, and Seaborn. The results of this study offer valuable insights into the distribution of wealth among billionaires, the factors that contribute to industry success, gender disparities, age demographics, and other factors that influence the accumulation of billionaire wealth.**

Keywords*—Big data; python; billionaires; net worth; wealth accumulation; wealth inheritance; geographic location; statistical analysis

I. Introduction

The rapid growth of wealth among billionaires has garnered significant attention in recent years, with the number of billionaires increasing from 1,001 in 2010 to 2,153 in 2019 [1]. This surge in wealth accumulation has led to a growing interest in understanding the factors that contribute to the success of these individuals. Big Data Analytics, with its ability to process large volumes of data, offers a promising approach to uncovering patterns and insights related to wealth distribution, wealth creation, and accumulation among billionaires [2].

This paper explores the dynamics of wealth accumulation and its distribution among billionaires worldwide using a comprehensive dataset. The study uses advanced big data analytics techniques to show trends, correlations, and patterns that underlie wealth creation and proliferation among the world's richest individuals. The research's significance lies in its contribution to academic literature and its potential to inform policy decisions and economic strategies. By elucidating factors contributing to wealth accumulation, it looks to provide insights into economic and social policies that could promote a fairer distribution of wealth and opportunities. The methodological approach combines quantitative analysis with sophisticated data visualization techniques to offer a panoramic view of wealth dynamics. The findings are expected to shed light on wealth accumulation trajectories and contribute to a deeper understanding of the economic forces shaping our world.

The objective of this investigation is to conduct an analysis of the "Billionaires.csv" dataset utilising Python tools for Big Data Analysis, with a particular emphasis on regions and cities. This paper aims to investigate several research questions about the distribution of billionaires across the globe. The study's objectives include identifying the ten nations with the highest concentration of billionaires, the industries with the greatest growth, the nations with the largest percentage of female billionaires, the age groups where there are the most and least billionaires belong, as well as other elements that may affect the development of wealth, for example, inheritance.

The paper is structured into four tasks: Problem Domain, Data Description, and Research Question; Solution Exploration; Solution Development; and Evaluation and Future Development. Each task delves into different aspects of the research process, from formulating hypotheses and evaluating methodologies to analyzing the dataset and discussing the potential impact of the results. This research endeavors to enhance comprehension of wealth accumulation trends among billionaires and offer valuable insights for policymakers, investors, and entrepreneurs by utilizing Big Data Analytics and the Python programming language.

II. Literature Review

Between 1999 and 2019, the Hurun Rich List was diligently compiled, and research was done to confirm billionaires and their families. They compare the Hurun Rich List to each company's annual report to determine which billionaires own it [3]. The systematic billionaire wealth factors were not considered in the univariate tests. Age was disregarded, and economic activity might potentially be important [4]. A study constructed a comprehensive metric of wealth inequality on a global scale, utilizing Forbes magazine's roster of billionaires, and subsequently conducted a comparative analysis of its impact on economic growth to the effects of income inequality and poverty [5]. An additional research study demonstrated that industries that concentrate wealth do so with purpose, and this

*Corresponding Author

may occur due to a reliance on the state, the presence of market failures, or the inheritance of extreme wealth [6]. A recent study conducted during the pandemic in Canada has demonstrated that low-wage workers have been disproportionately affected. Additionally, data from Statistics Canada has indicated that women and racialized workers are overrepresented within the low-wage worker demographic [7]. An article aims to compare asset ownership in three common marital regimes in France and other European countries: Property regimes, marriage, non-marital cohabitation, and registered partnership (PACS) [8]. The results show that luck plays a big part at the very ends of the range of economic outcomes, and the fact that empirical regularities tend to disappear in the far tails can be used to look at any group of successful or failed people [9]. A study finding indicates that there is a detrimental association between wealth inequality and economic growth., however, when considering the influence of political connections on the acquisition of wealth by billionaires, it is observed that politically connected wealth inequality has a negative impact on economic growth. On the other hand, politically unconnected wealth inequality, income inequality, and initial poverty do not exhibit any significant relationship with economic growth [5]. An empirical approach to studying the potential influence of education and cognitive ability on the accumulation of extreme wealth involves analyzing high-wealth groups and retrospectively evaluating the level of education and cognitive ability of these individuals in the past [10]. When comparing billionaires, it is evident that wealth is not correlated with attractiveness. Neither does it pertain to the level of schooling achieved [14]. The concentration of wealth and its distribution has been a longstanding concern in economic literature and societal discourse. Studies on wealth inequality and its consequences have highlighted the need for a deeper understanding of the factors contributing to the accumulation of wealth among the elite class. In this section, we provide a brief overview of relevant literature that explores the determinants of billionaire net worth and sheds light on the dynamics of wealth accumulation.

A study looked at a continuous-time DSGE model with several types of households and a financial sector, and they were able to make the case of studying how financial frictions affect families by looking at how wealth is distributed [15]. An agent-based model of microscopic wealth sharing in a dynamic network is used in another study to investigate the topological aspects of economic inequality [16]. The model changes in two steps that happen back and forth: the connected agents conservatively trade wealth, and the links are rewired based on the agents' wealth. Moreover, an analysis finds financialization, rentiers, and labour exploitation as the primary drivers of billionaire wealth in the U.S. sectors with the highest number of billionaires [17]. These factors were crucial for their dominance, while shareholder culture, crony capitalism, and tax policy perpetuated wealth but were not necessary for its current level.

Gender disparities in wealth distribution have also garnered significant attention. Numerous studies have documented the persistent gender wealth gap, demonstrating that women tend to have lower levels of wealth compared to men. However, the examination of gender disparities among billionaires remains relatively limited. To clarify the gender dynamics within this wealthy class and put light on the larger issue of disparities between genders in the accumulation of wealth, it is imperative to know whether there is an important variation in assets between male and female billionaires. An article looks at how the trends in inequality found by multiple well-known inequality indices in the Forbes 400 richest families and compared with each other [18]. Other researchers have looked at other parts of the data set. A paper makes the first Distributional Wealth Accounts (DWAs) for Europe from 1970 to 2018 by putting together state accounts, tax records, and surveys. According to our new database, the amount of wealth compared to national income has changed about the same in both the US and Europe. However, since the mid-1980s, wealth inequality has grown much faster in the US than in Europe.

Finally, the impact of wealth inheritance on billionaire net worth has attracted scholarly attention. Inherited wealth often carries advantages in terms of family legacies, established networks, and access to resources. However, the extent to which inherited wealth plays a role in billionaire net worth compared to self-made wealth is a topic of ongoing debate. Investigating the difference in net worth between billionaires who inherited their wealth and those who built their fortunes from scratch adds to the understanding of intergenerational wealth transfer and its implications for wealth inequality. By reviewing the existing literature on age and wealth, gender disparities, industry sectors, and wealth inheritance, we situate our study within the broader research landscape. This background provides a foundation for our analysis of the billionaire dataset and highlights the gaps and opportunities for further exploration. The findings from this study contribute to the existing knowledge base, inform policy discussions, and offer insights into the complex dynamics of wealth accumulation among billionaires.

III. Problem Domain

The open-source dataset utilized in this paper is the billionaire's dataset from Kaggle. Understanding wealth distribution and wealth trends may be done extremely well by using this dataset. This report looks at the "Billionaires.csv" dataset analysis using Python tools designed for large-scale data analysis. The database includes data about billionaires from different nations, such as their ages, genders, industrial sectors, net worths, and places of wealth. Understanding the variables that affect the creation and distribution of billionaires' wealth is the aim of the investigation. To better comprehend these patterns and discover potential prospects for company starts and investments, it may be helpful to examine current trends in the global economy.

IV. Investigation

The paper aims to investigate the following measures:

1) To examine the relationship between demographic factors (age, gender) and net worth among billionaires.

2) To explore the variations in net worth across different industry sectors among billionaires.

3) To investigate the differences in net worth between self-made billionaires and those who inherited their wealth.

4) To analyze the correlation between age and net worth among billionaires and determine the significance of the relationship.

5) To compare the net worth of billionaires across different industry sectors and identify sectors with the highest net worth individuals.

6) To evaluate the differences in net worth between self-made billionaires and those who inherited their wealth using statistical tests.

7) To provide insights into the patterns and trends in the distribution of wealth among billionaires based on demographic factors and industry sectors.

V. Research Methodology

- Data Collection: The open-source dataset obtained from Kaggle was used in this study.
- Data Preprocessing: To deal with any missing numbers, outliers, and other mistakes, the data will be pre-processed.
- Data visualization: Data visualization is the process of using visual elements like charts, graphs, or maps to represent data.
- Hypothesis Testing: Formulate specific hypotheses based on research questions and objectives. Selection of appropriate statistical tests to test the hypotheses.
- Interpretation of Results: The results will be interpreted to identify key findings.

VI. Data Description

The dataset helps research variables related to wealth accumulation and distribution among billionaires. The net worth may be impacted by age and industrial sectors. A study of inheritance and the gender difference in wealth accumulation might potentially be done using the dataset. There are 22 variables and 2614 rows in the dataset. Table I provides a brief description of each variable.

TABLE I. BRIEF DATA INTERPRETATION

Feature	Description
name	Billionaires name.
rank	Annual list of billionaires ranked by net worth.
year	Year of data collection.
company. founded	The year that the business was formed
company.name	The billionaire's company's name
company. relationship	The connection between the business and the billionaire.
company. sector	Area of billionaires' businesses.
company. type	What kind of business the billionaire owns—public, state-owned, public etc.
demographics.age	Individual’s age.
demographics. relationship	Billionaire's gender.
location. citizenship	Nation in which the billionaire was born.
location. country code	The Country's ISO code where the billionaire lives.
location.gdp	Country's GDP where the billionaire lives.
location. region	Location of the billionaire in the globe.
wealth. type	Wealth's source (Inherited. Self-made, etc.)
wealth. worth in billions	The billionaire's total wealth, expressed in billions of dollars.
wealth.how. category	Method by which each billionaire acquired their wealth
wealth.how. from emerging	If the billionaire's prosperity was derived from a developing market.
wealth.how. industry	The branch of industry where the billionaire made their money.
wealth.how. inherited	The billionaire's wealth was acquired through inheritance.
wealth.was. founder	If the millionaire was a company founder.
wealth.how.was political	Billionaire has political experience or not.

VII. Research Questions

1) Which ten international countries have the maximum number of billionaires?

2) Which industries and sectors are the most successful?

3) Which sectors have the highest awareness of female billionaires?

4) What age institution do most people and a minority of billionaires belong to?

VIII. Hypothesis

A crucial component of inferential statistics is hypothesis testing, which enables us to conclude unobserved data, frequently the population, using data from a sample of observed data. When analyzing experimental data in economics, testing multiple null hypotheses simultaneously is a common practice [11]. To research this hypothesis, numerous statistical strategies will be used, inclusive of t-tests, ANOVA, and F-assessments. The outcomes of those experiments could shed mild on the variables concerned with wealth accumulation and guide the improvement of policies geared toward reducing wealth inequality. In our research, we will look at the principle of different factors that can be intently related to a billionaire's wealth. We REJECT the null hypothesis in favor of the alternative if the P-value is less than the threshold for significance (= 0.05). The correlation is statistically significant, we conclude. Or, to put it another way, we come to the simple conclusion that x and y in the population at the -level are linearly related.

If $p > 0.05$, then not correlated.

If $p < 0.05$, correlated.

The following hypotheses have been formulated for our study.

1) Does a billionaire's age affect his or her wealth?

Ho: A billionaire's net worth is not significantly related to age.

HA: A billionaire's net worth is highly related to age in a significant way.

2) Men with billion-dollar wealth are wealthier than women.

Ho: A billionaire's net wealth is unaffected by gender in a significant way.

H_A: Gender significantly affects a billionaire's financial worth.

3) The net worth of billionaires varies considerably between various industry sectors.

Ho: A billionaire's net worth is significantly influenced by the industry sector.

HA: A billionaire's net worth is not much impacted by their industry sector.

IX. Exploration of Solutions

Big data is the term used to describe the vast and diversified informational resources that are accumulating quickly. Software that can be used to handle data effectively and perform pre-processing and cleaning etc., is known as big data technology. The five main qualities of big data are truth, diversity, value, and velocity. In seconds, minutes, hours, or days, data volume is measured [12]. Velocity describes the speed at which new data is produced and transferred, whereas variety describes the wide range of data forms that are produced often. It's vital to note that velocity also shows how rapidly a company reacts to big data-derived business insight. The chance that the data's quality and consistency are inadequate, that is, less uniform, consistent, and controllable, is referred to as veracity. The fifth V, "value," stands for the ability of a person or an organization to translate enormous amounts of data into beneficial outcomes, which includes the capability to acquire and then use data [13].

To address the issues of huge information programs, numerous methodologies, and technologies have evolved these days. Those are Apache Hadoop, MongoDB, Apache Kafka, Elastic Search, Python, Seaborn, and Plotly.

X. The Tools or Technologies used in this Research

The Python programming language and modern libraries such as Pandas, NumPy, Matplotlib, and Seaborn have been selected as the technologies that will be used for this project. This decision was made about the technology that would be implemented. It is beneficial to use open-source software solutions since they are free of charge, have an interface that is simple to use, and can give comprehensive analysis and visualization capabilities. Other technologies, on the other hand, can need payment or prior expertise. When it comes to the Python Seaborn visualization toolkit, the matplotlib module is an absolute must. A graph interface that is both high-level and insightful of data will be utilized. Python packages such as Pandas, NumPy, and SciPy can be utilized to edit and analyze data. The Pandas data frame makes the process of processing data simpler. NumPy can deal with low-level data, whereas SciPy is responsible for statistical analysis.

XI. Solution Development

A key part of current data-driven decision-making is coming up with practical solutions to problems based on data analysis. Using the "Billionaires" dataset as an example, this paper talks about the main steps and things to think about when coming up with good solutions for the research questions we have. We are going to initially preprocess our data to check whether there is any noise, further, we will analyze our dataset to answer those research questions we formulate in the previous part. In the meantime, we will test the hypothesis for each question. Finally, we will show our findings and future scopes.

A. Data Preprocessing

It is necessary to preprocess the dataset to prepare it to provide true insights into the data to answer the research questions. To preprocess data, numerous processes are taken. We are going to take advantage of the procedures that are necessary in order to finish our assignment.

B. Visual Exploratory Data Analysis

Fig. 1 determines statistics at the variables' names, counts, and their respective information kinds. There are a complete of thirteen objects, along with 4 integers, 2 floats, and 3 Booleans.

Fig. 2 presents comprehensive statistical information for each numerical column, encompassing the count of values in a row, the mean, standard deviation, minimum and maximum values, and the quartiles at 25%, 50%, and 75% of the dataset. In terms of the dataset, the period under examination spans from 1996 to 2014. When compared to the oldest company, which was established in 1610, the most recent company was established in 2012. The individual who is the youngest among them is only 12 years old, while the individual who is the oldest is 98 years old.Beyond $3.5 billion, the median net worth of a billionaire is greater than that amount.

Fig. 3 illustrates the number of unique values in each variable. Here, we are going to present a visual exploratory information analysis. Commonly, specific variables incorporate less than twenty precise values and there may be a repetition of values, this means that the records can be grouped by manner of these precise values. At this level of the study process, we're investigating the effects of visible evaluation on a spread of specific variables derived from the dataset.

Fig. 4 demonstrates the dataset comprising individuals belonging to three distinct genders: males, females, and married couples. Moreover, there exist five discrete categories of billionaires' wealth.

The four graphs in Fig. 5 and Fig. 6 depict a visual analysis of six distinct variables within the dataset. Fig. 5 illustrates the distribution of wealth and the corresponding industries associated with wealth. Fig. 6 illustrates six distinct instances of form inheritance, along with a classification indicating whether each billionaire is a founder or not.

```
[1097] billionaire_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2614 entries, 0 to 2613
Data columns (total 22 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   name                     2614 non-null   object
 1   rank                     2614 non-null   int64
 2   year                     2614 non-null   int64
 3   company.founded          2614 non-null   int64
 4   company.name             2576 non-null   object
 5   company.relationship     2568 non-null   object
 6   company.sector           2591 non-null   object
 7   company.type             2578 non-null   object
 8   demographics.age         2614 non-null   int64
 9   demographics.gender      2580 non-null   object
 10  location.citizenship     2614 non-null   object
 11  location.country code    2614 non-null   object
 12  location.gdp             2614 non-null   float64
 13  location.region          2614 non-null   object
 14  wealth.type              2592 non-null   object
 15  wealth.worth in billions 2614 non-null   float64
 16  wealth.how.category      2613 non-null   object
 17  wealth.how.from emerging 2614 non-null   bool
 18  wealth.how.industry      2613 non-null   object
 19  wealth.how.inherited     2614 non-null   object
 20  wealth.how.was founder   2614 non-null   bool
 21  wealth.how.was political 2614 non-null   bool
dtypes: bool(3), float64(2), int64(4), object(13)
memory usage: 395.8+ KB
```

Fig. 1. Dataset information.

```
[1098] billionaire_df.describe()
```

	rank	year	company.founded	demographics.age	location.gdp	wealth.worth in billions
count	2614.000000	2614.000000	2614.000000	2614.000000	2.614000e+03	2614.000000
mean	599.672533	2008.411630	1924.711936	53.341239	1.769103e+12	3.531943
std	467.885695	7.483598	243.776546	25.333320	3.547083e+12	5.088813
min	1.000000	1996.000000	0.000000	-42.000000	0.000000e+00	1.000000
25%	215.000000	2001.000000	1936.000000	47.000000	0.000000e+00	1.400000
50%	430.000000	2014.000000	1963.000000	59.000000	0.000000e+00	2.000000
75%	988.000000	2014.000000	1985.000000	70.000000	7.250000e+11	3.500000
max	1565.000000	2014.000000	2012.000000	98.000000	1.060000e+13	76.000000

Fig. 2. Statistical view of numerical data.

```
# Finding unique values for each column
# TO understand which column is categorical and which one is Continuous
# Typically if the numer of unique values are < 20 then the variable is likely to be a category otherwise continuous
billionaire_df.nunique()

name                        2077
rank                         468
year                           3
company.founded              178
company.name                1577
company.relationship          74
company.sector               520
company.type                  18
demographics.age              76
demographics.gender            3
location.citizenship          73
location.country code         74
location.gdp                  81
location.region                8
wealth.type                    5
wealth.worth in billions     183
wealth.how.category            9
wealth.how.from emerging       1
wealth.how.industry           19
wealth.how.inherited           6
wealth.how.was founder         1
wealth.how.was political       1
dtype: int64
```

Fig. 3. Checking unique values.

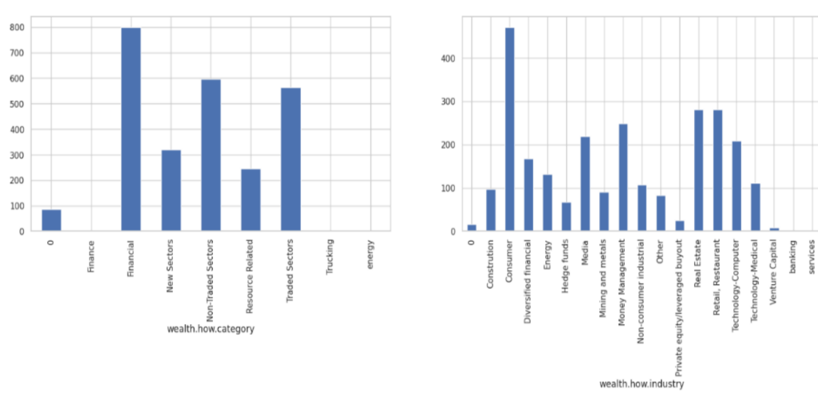


Fig. 4. Gender and wealth distribution visually.

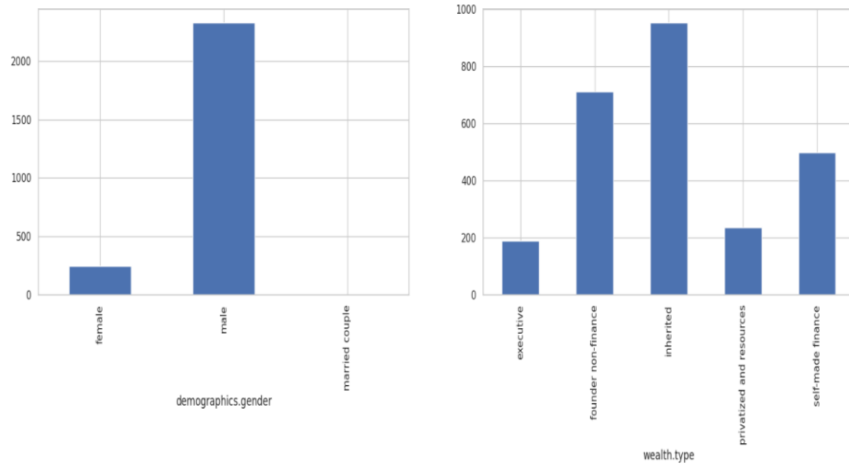


Fig. 5. Distribution of ‘wealth category and ‘wealth industry’ by bar graph.

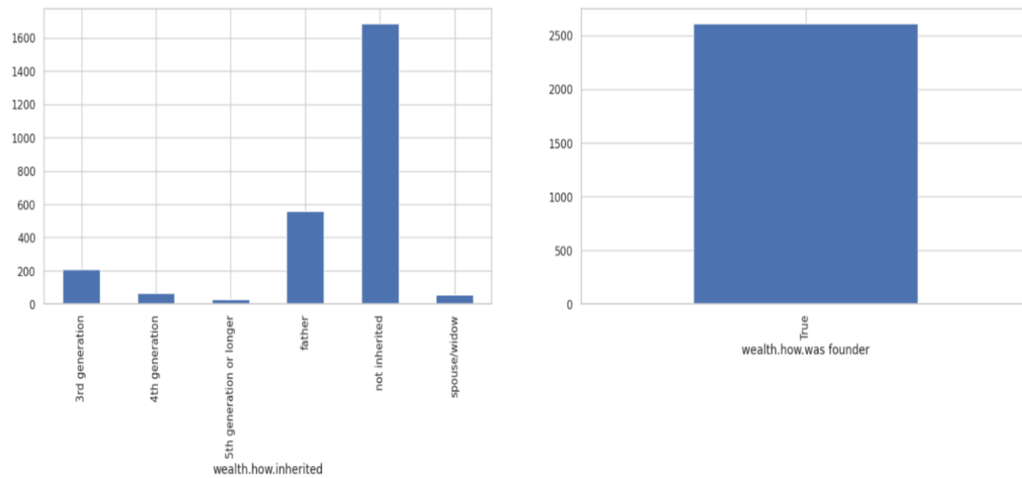


Fig. 6. Distribution of ‘wealth inherited’ and ‘wealth.how.was founder.

The dataset contains 22 variables, but only a few are essential for answering research questions. To maintain the desired number of variables, only a few are required, resulting in the loss of unnecessary ones. The remaining variables will be maintained for research purposes.

Fig. 7 displays a dataset after irrelevant columns have been removed, presenting a streamlined view of wealthy individuals' demographics and financial information.

	name	demographics.gender	location.citizenship	company.sector	demographics.age	wealth.worth in billions	wealth.type	wealth.how.category	wealth.how.inherited	wealth.how.was founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2614 rows × 10 columns

Fig. 7. After removing irrelevant columns - new dataset.

By doing so, we will rename the column name as follows in order to make the analysis simpler:

Fig. 8 represents the dataset with renamed columns to provide a clearer understanding of the attributes of each individual listed.

C. Handling Missing Values

At this point, we determine which values in the dataset are absent. One of the most critical components of records evaluation is the management of lacking values. The precise study topic, the number of missing data, and the influence that the missing values can have on the evaluation are all elements that must be taken into consideration whilst determining whether to put off missing values or impute them.

As can be seen in Fig. 9, 34 cells are missing from the gender column, 23 cells in the company_sector column, 22 cells in the wealth_type column, and one cell in the wealth_source column.

In Fig. 10, We check the Total missing values in our dataset and the proportion of missing on different columns. The percentage of values that are missing from the dataset is quite low (most of them are much lower than 0%). Therefore, we are going to ascribe those values to the requirements that we have.

	name	gender	citizenship	company_sector	age	net_worth_billion	wealth_type	wealth_source	wealth_inherited	was_founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2614 rows × 10 columns

Fig. 8. Renaming columns.

```
#1.Checking missing values dataset
print(billionaire_df.isnull().sum())
```

```
name                 0
gender              34
citizenship          0
company_sector      23
age                  0
net_worth_billion    0
wealth_type         22
wealth_source        1
wealth_inherited     0
was_founder          0
dtype: int64
```

Fig. 9. Number of missing values on each column.

```
# get the number of missing data points per column
missing_values_count = billionaire_df.isnull().sum()
# how many total missing values do we have?
total_cells = np.product(billionaire_df.shape)
total_missing = missing_values_count.sum()

# percent of data that is missing
(total_missing/total_cells) * 100
```

```
0.306044376434583
```

```
# Calculate the percentage of missing values in each column for updated dataframe
missing_values = billionaire_df.isnull().sum() / len(billionaire_df) * 100

# Print the percentage of missing values in each column
print(missing_values)
```

```
name               0.000000
gender             1.300689
citizenship        0.000000
company_sector     0.879878
age                0.000000
net_worth_billion  0.000000
wealth_type        0.841622
wealth_source      0.038256
wealth_inherited   0.000000
was_founder        0.000000
dtype: float64
```

Fig. 10. Missing values in percentage.

As a first step, we strive to impute the cells that might be lacking from the gender column. Inside the gender column, there are a complete of 34 values that are lacking, and most people of these billionaires have covered terms related to their families alongside their names. Due to the absence of gender values, data analysis is impeded. The presence of the term "family" in the name gives the impression that the individual in question is not a billionaire but rather a member of a family that is a billionaire, which has the potential to skew the results of the study. Consequently, removing these rows guarantees that the analysis is founded on data that is both reliable and pertinent.

Fig. 11 demonstrates the word "family" is present in most of the cells that are absent from the gender column. We are going to get rid of those rows. Since Oeri Hoffman and Sacher appear to be a "married couple" right before our eyes, we shall impute one missing cell as belonging to a "married couple.". Rest of the cells we can delete the rows as there are a smaller number of missing cells.

Fig. 11 demonstrates the word "family" is present in most of the cells that are absent from the gender column. We are going to get rid of those rows. Since Oeri Hoffman and Sacher appear to be a "married couple" right before our eyes, we shall impute one missing cell as belonging to a "married couple.". Rest of the cells we can delete the rows as there are a smaller number of missing cells.

In Fig. 12, following the elimination of cells that were absent from the gender subgroup, we are left with 2581 individual observations. The next step is for us to analyze our research using this dataset. We are still trying to impute a few cells that are missing from our database.

Fig. 13 shows above are the remainder of the values that are absent from our dataset. We are going to try to impute those using statistical methods such as the mean, the mode, and the median. As a method for dealing with missing values, we use the median for continuous variables and mode for categorical measures.

```
# here we can see the the column gender has missing gender because most of the billionaire
# name is included family .In data set only 1.3% of rows missing gender.
#and most of those are family business
# and missing data percentage is very less so we can delete those rows
missing_gender_df = billionaire_df[billionaire_df['gender'].isnull()]

names = missing_gender_df.loc[:, 'name']

for name in names:
    print(name)
```

```
Oeri Hoffman and Sacher
Haniel family
Wonowidjojo family
Merck family
Henkel family
Boehringer family
Seydoux/Schlumberger families
Brenninkmeijer family
Shin Kyuk-Ho
Lemos family
Von Siemens family
Porsche family
Funke family
Verspieren family
Moores family
Goulandris family
Rochling family
Peugeot family
Simon family
Freudenberg family
Juffali family
Leibbrand family
Reimann family
Conle famle
Bemberg family
Isono family
Ryusuke Kimura
Kim Suk-won
Larragoiti family
Strwher family
Werhahn family
Otani Family
Junichi Murata
Autrey family
```

Fig. 11. Checking missing values in gender with their names.

```
[340] #It is not possible to determine the gender of individuals based solely on their name or family name.
      billionaire_df = billionaire_df.dropna(subset=['gender'])
      billionaire_df
```

	name	gender	citizenship	company_sector	age	net_worth_billion	wealth_type	wealth_source	wealth_inherited	was_founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2581 rows × 10 columns

Fig. 12. Dropping irrelevant cells from gender subset.

```
[341] #1.2.Checking Missing values for updated Dataframe
      print(billionaire_df.isnull().sum())

name                  0
gender                0
citizenship           0
company_sector       11
age                   0
net_worth_billion     0
wealth_type          10
wealth_source         1
wealth_inherited      0
was_founder           0
dtype: int64
```

Fig. 13. Checking total missing values remained in the updated data frame.

Fig. 14 illustrates the code snippet of a Python script using pandas to impute missing values in the dataframe.

After imputing the remainder variables, we got the updated dataset with 'zero' missing values (Fig. 15).

D. Checking Duplicate

As part of the data cleansing process, we conducted a check to identify any duplicated rows (as in Fig. 16). The duplicated () method in pandas allows us to identify and eliminate any duplicated rows in the dataset. This measure was implemented to ensure the distinctiveness of each discovery and to prevent any potential interference with subsequent analyses. Therefore, we did not discover any duplicate rows as a result.

```
[342] #I am treating missing values with Median for Continuous values, and Mode for categorical values.
      # Treating missing values of categorical variable with MODE value
      billionaire_df['company_sector'].fillna(value=billionaire_df['company_sector'].mode()[0], inplace=True)
      billionaire_df['wealth_type'].fillna(value=billionaire_df['wealth_type'].mode()[0], inplace=True)
      billionaire_df['wealth_source'].fillna(value=billionaire_df['wealth_source'].mode()[0], inplace=True)
      billionaire_df

<ipython-input-342-d863120c5587>:3: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

<ipython-input-342-d863120c5587>:4: SettingWithCopyWarning:
```

Fig. 14. Missing values imputation by statistical techniques.

```
[343] #1.3.Checking Missing values for updated Dataframe
      print(billionaire_df.isnull().sum())

name                 0
gender               0
citizenship          0
company_sector       0
age                  0
net_worth_billion    0
wealth_type          0
wealth_source        0
wealth_inherited     0
was_founder          0
dtype: int64
```

Fig. 15. Data frame after imputation.

```
[344] #4.Checking for duplicated rows in a dataset after dropping the rows with missing values
      # Check for duplicated rows
      duplicated_rows = billionaire_df[billionaire_df.duplicated()]

      # Print the duplicated rows
      print(duplicated_rows)

Empty DataFrame
Columns: [name, gender, citizenship, company_sector, age, net_worth_billion, wealth_type, wealth_source, wealth_inherited, was_founder]
Index: []
```

Fig. 16. Checking duplicate.

E. Outlier Analysis

To guarantee the precision and dependability of the studies, we did an outlier treatment of the dataset by using the interquartile range (IQR) approach in Python. This allowed us to recognize and exclude any intense values that would have a sizeable impact on the effects.

Fig. 17 and Fig. 18 show that the 'age' column contains significant values that need to be addressed. The dataset includes both horrible and zero age graphs, and outliers are identified at significantly lower than 20 ages. It is important to note that billionaires' age cannot be zero or impoverished.

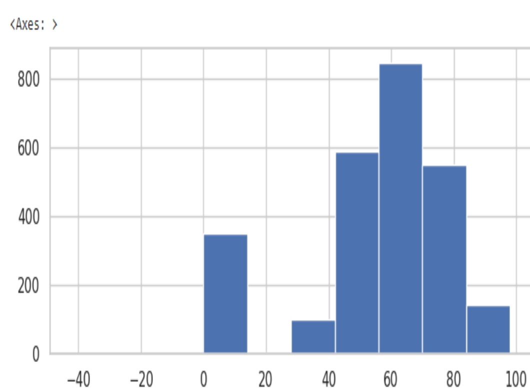


Fig. 17. Visual distribution of 'Age' by histogram.

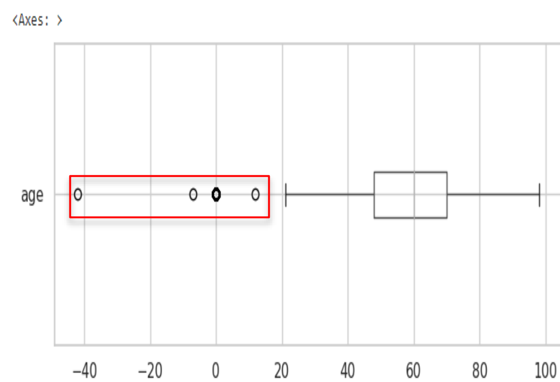
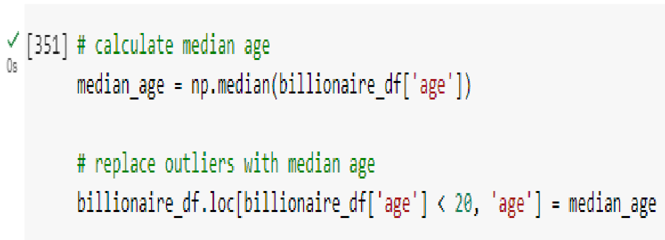


Fig. 18. 'Age' column where outliers detected.

Among the options available for dealing with the outliers, one option is to substitute the median age charge of the relaxation of the dataset for the values of bad and 0 years old. The outliers are going to be dealt with by imputing them using the median statistical technique. This will be done to address the issue.

It is through this method that the median age of the information is determined, which consequently alters the age of billionaires who are significantly younger than twenty years old (Fig. 19).



```
[351] # calculate median age
      median_age = np.median(billionaire_df['age'])

      # replace outliers with median age
      billionaire_df.loc[billionaire_df['age'] < 20, 'age'] = median_age
```

Fig. 19. Treating outliers in 'Age' column.

The net worth column that is displayed in Fig. 20 does not demonstrate any discernible increasing or declining trend. In addition to this, it demonstrates that the 'net_worth' column does not contain any outliers. Further evidence that there is a substantial association between 'age' and 'net worth' is shown by the graph.

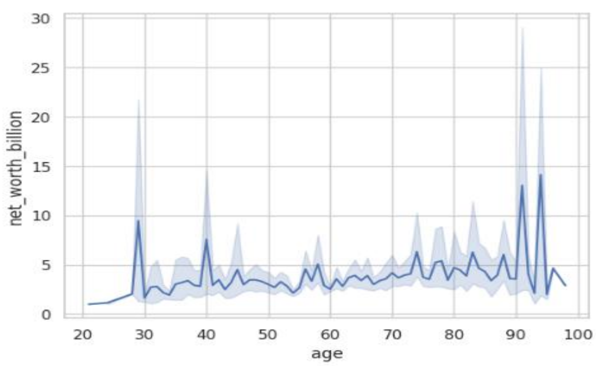


Fig. 20. Line plot of age vs. net worth billion.

XII. Dataset Exploration

The top ten countries with the highest number of billionaires are depicted in Fig. 21.

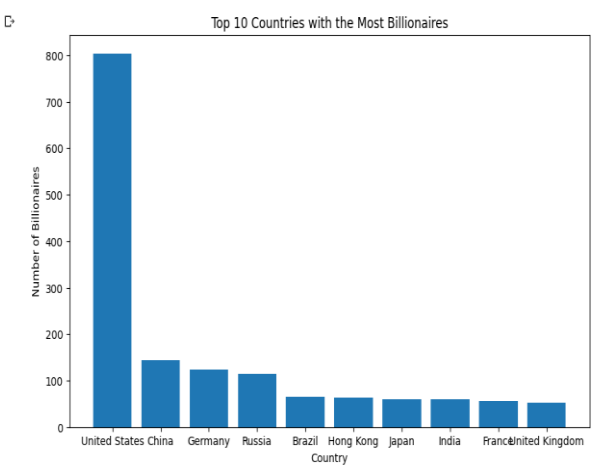


Fig. 21. Top ten billionaires in a bar graph.

Perspectives from Fig. 21:

- Fig. 21 above displays a bar chart of the top 10 richest nations and their billionaires.
- The most billionaires reside in the United States.
- China ranks second among countries with 100 billionaires.
- The number of billionaires in the United Kingdom, which is ranked 10th, is just under fifty.

Fig. 22 depicts the top five industries with the largest number of billionaires.

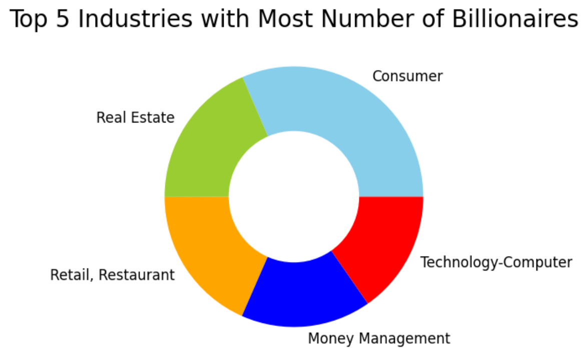


Fig. 22. Top five industries with the most billionaires are shown in a pie chart.

Perspectives from Fig. 22:

- The consumer goods industry has the most billionaires, as shown in the pie chart above.
- A significant number of billionaires are also involved in real estate.

Fig. 23 depicts the number of women billionaires by sector.

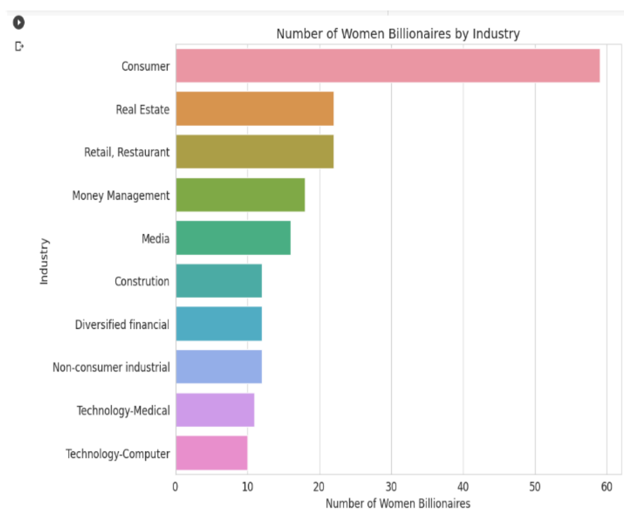


Fig. 23. Graph depicting the number of female billionaires by sectors.

Perspectives from Fig. 23:

- With over 55 female billionaires, the consumer enterprise has the most female billionaires.
- The retail and real estate industries are 2nd, with little more than 20 lady billionaires.

- With only 10 women billionaires working in the IT area, it has the lowest number of female billionaires.

Fig. 24 represents the billionaires by age range.

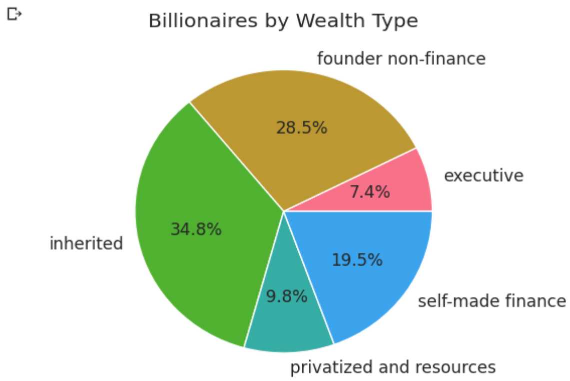


Fig. 24. Proportion of billionaires in each age group shown in the bar chart.

Perspectives from Fig. 24:

- Many of the billionaires in the dataset are between the ages of 60 and 70.
- There are very few billionaires aged 30 and under.
- Approximately two-thirds of billionaires are between the ages of 40 and 80.

Fig. 25 illustrates the proportion of billionaires by wealth category.

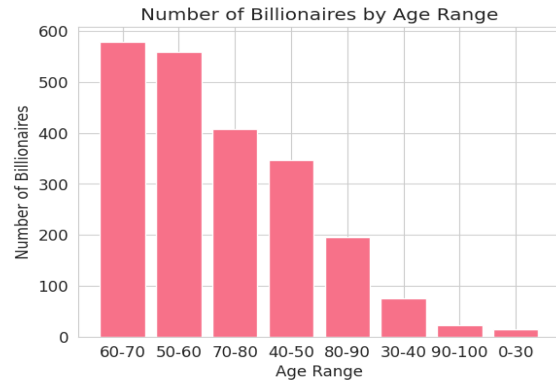


Fig. 25. A pie chart displaying the proportion of billionaires in each wealth category.

Perspectives from Fig. 25:

- The vast majority of the billionaires in the sample inherited their fortunes. They account for 34.8% of all billionaires.
- Self-made billionaires comprise 19.5% of the total.
- Executives account for the smallest proportion of billionaires, accounting for only 7.4% of all billionaires.

Fig. 26 depicts the 2014 net worth of the top 10 billionaires.

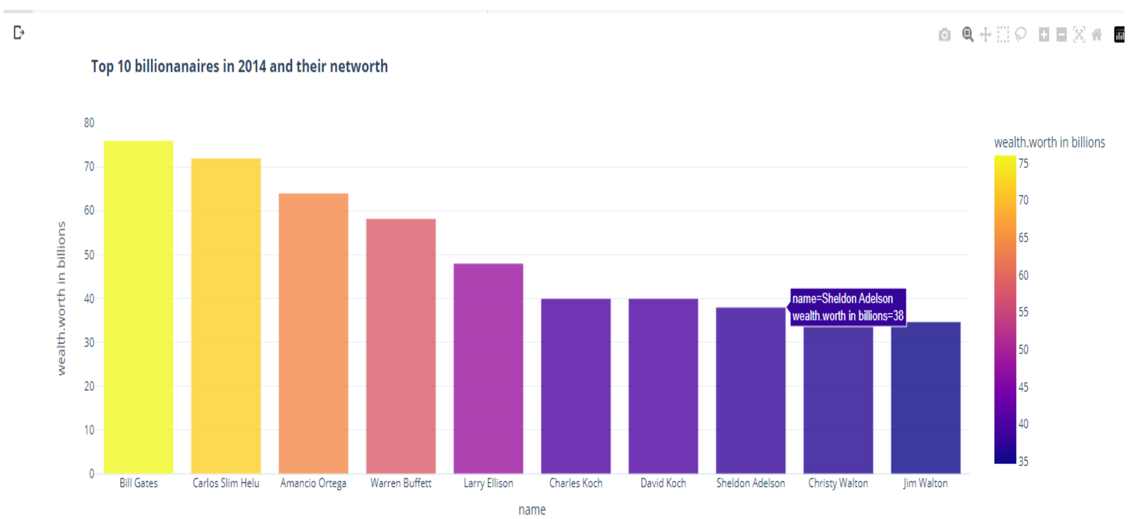


Fig. 26. A bar chart depicting the top ten billionaires and their net wealth in 2014.

Perspectives from Fig. 26:

- The bar graph above shows the most current top 10 billionaires alongside their net worth since 2014 is the latest year that is included in the dataset.
- With an overall net worth of $76 billion, Bill Gates is the first.
- After that, Carlos Slim comes in second place, boasting a net worth of seventy-two billion dollars.
- Christy Walton is the simplest female among the Top Ten rich individuals, with a net worth of thirty-eight billion greenbacks.

The top 10 billionaires' sources of income are shown in Fig. 27.

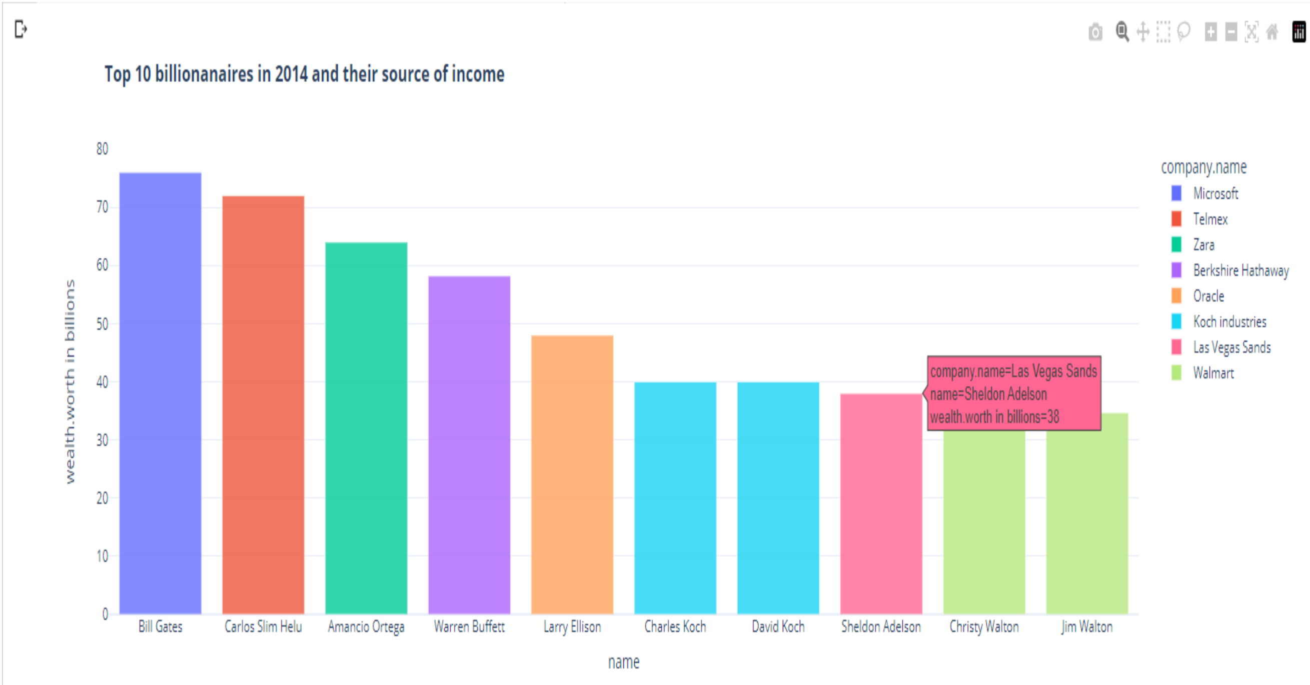


Fig. 27. The top 10 billionaires and their sources of income are shown in a bar graph.

Perspectives from Fig. 27:

- Bill Gates, the richest billionaire, receives funding from Microsoft.
- We can see from the picture above that some colors are repeated twice; these are family businesses or businesses that have produced more than one billionaire, like Walmart and Koch Industries.

The ages of the top 10 billionaires are shown in Fig. 28.

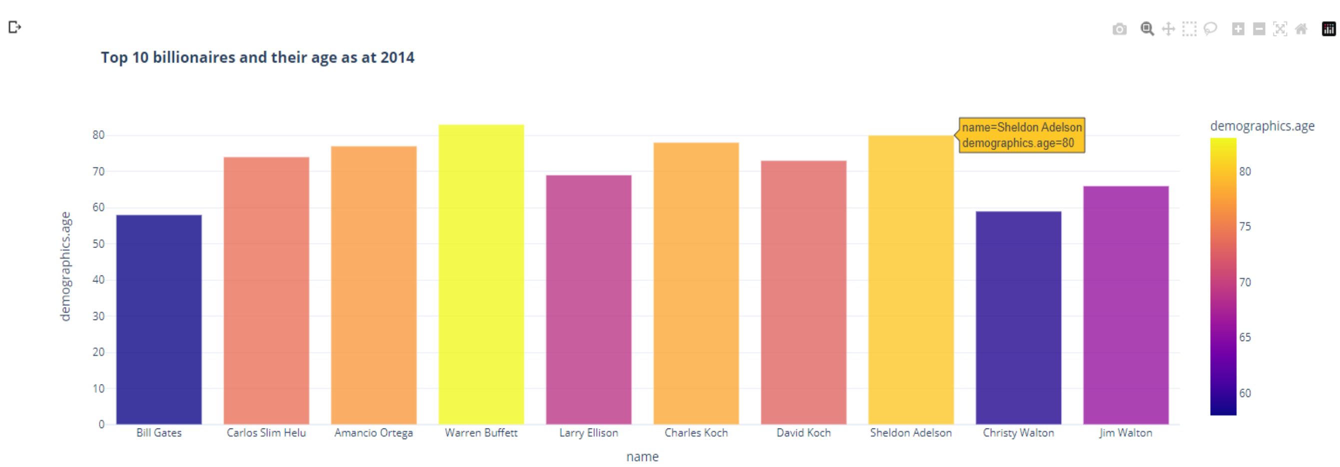


Fig. 28. A bar graph showing the ages of the top 10 billionaires.

Perspectives from Fig. 28:

- The world's number one billionaire is 58 years old.
- The top ten richest people are all over 50 years old.
- Warren Buffett, who is over 80 years old, is the oldest person among the top ten billionaires.

Fig. 29 presents the top ten youngest billionaires.

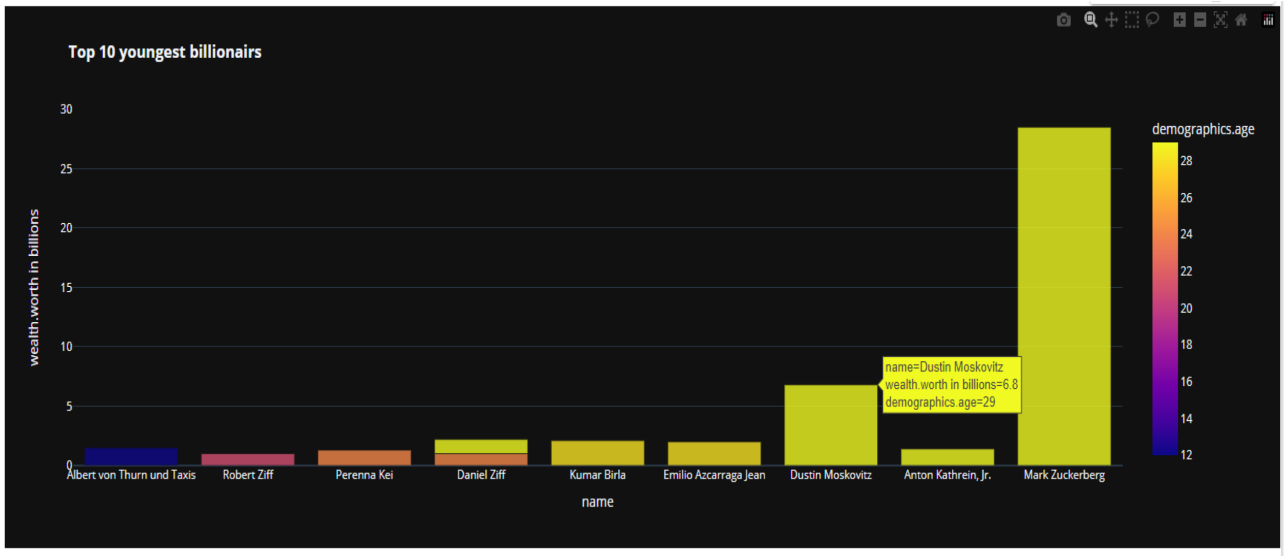


Fig. 29. The top ten youngest billionaires are represented by a bar chart.

Perspectives from Fig. 29

- Albert Thurn, who is only 12 years old and worth a billion dollars, is the youngest billionaire.
- Mark Zuckerberg is the wealthiest young person, with a net worth of more than $25 billion.

Fig. 30 shows the industries of the top ten youngest billionaires.

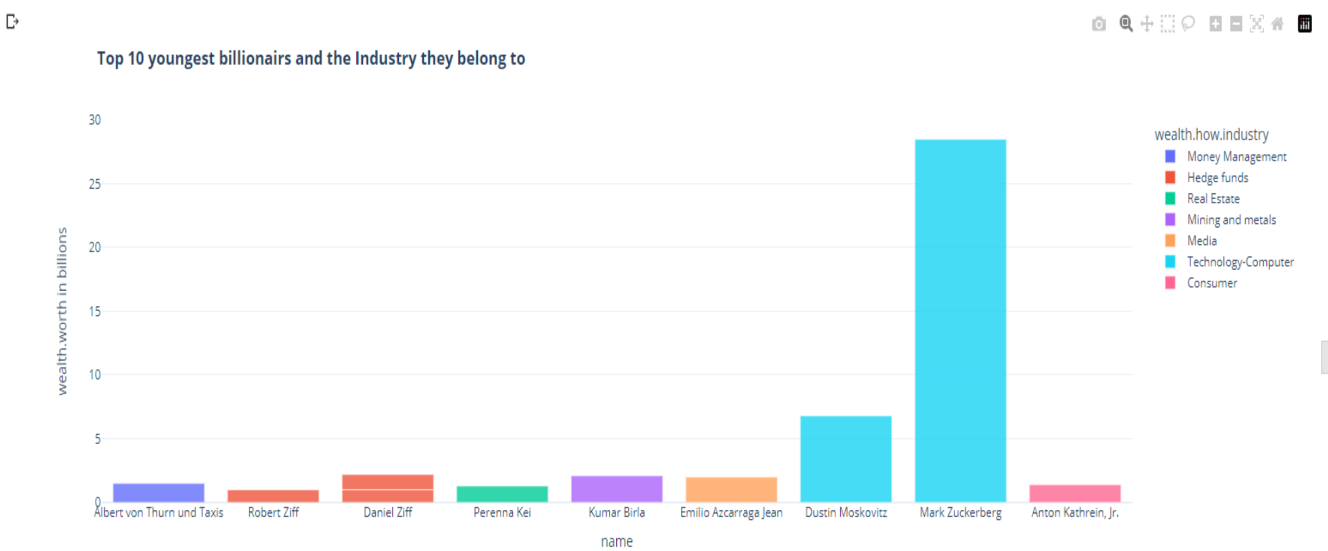


Fig. 30. A bar graph depicting the industries of the world's youngest billionaires.

Perspectives from Fig. 30:

- The youngest billionaire works in the financial services industry.
- The richest young billionaire is from the technology-computer industry.

Fig. 31 exhibits the top ten female billionaires and their ages.

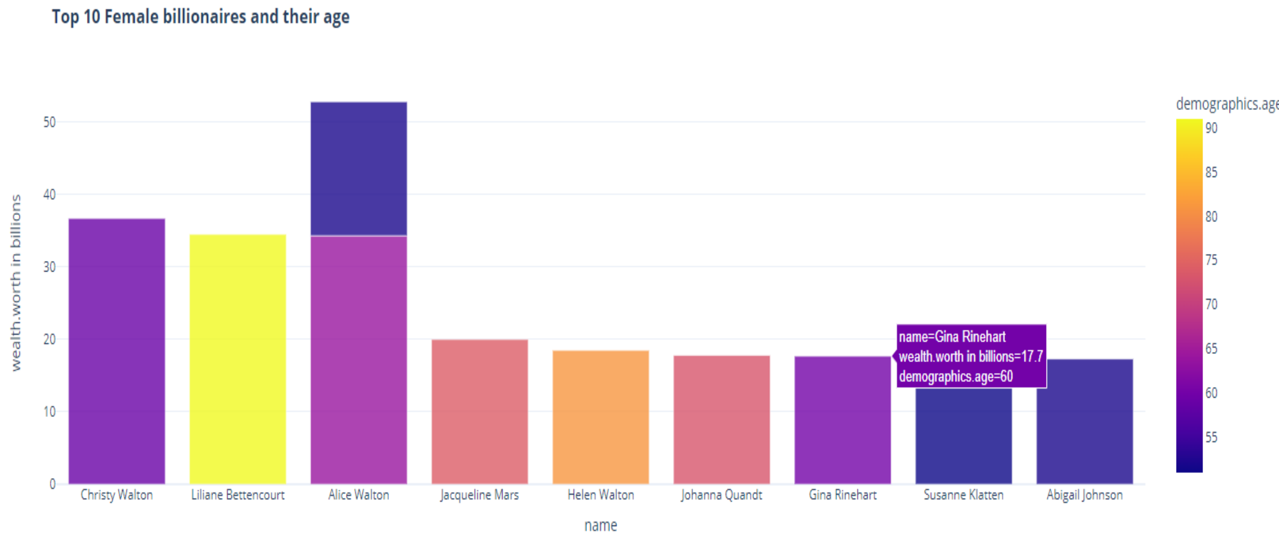


Fig. 31. The ages of the top ten female billionaires.

Perspectives from Fig. 31:

- Christy Walton, the wealthiest female billionaire, is 59 years old and worth more than 36 billion dollars.
- Lillian Bettercourt, who ranks second among female billionaires, is over 80 years old.

Fig. 32 highlights the top ten billionaires and the industries in which they work.

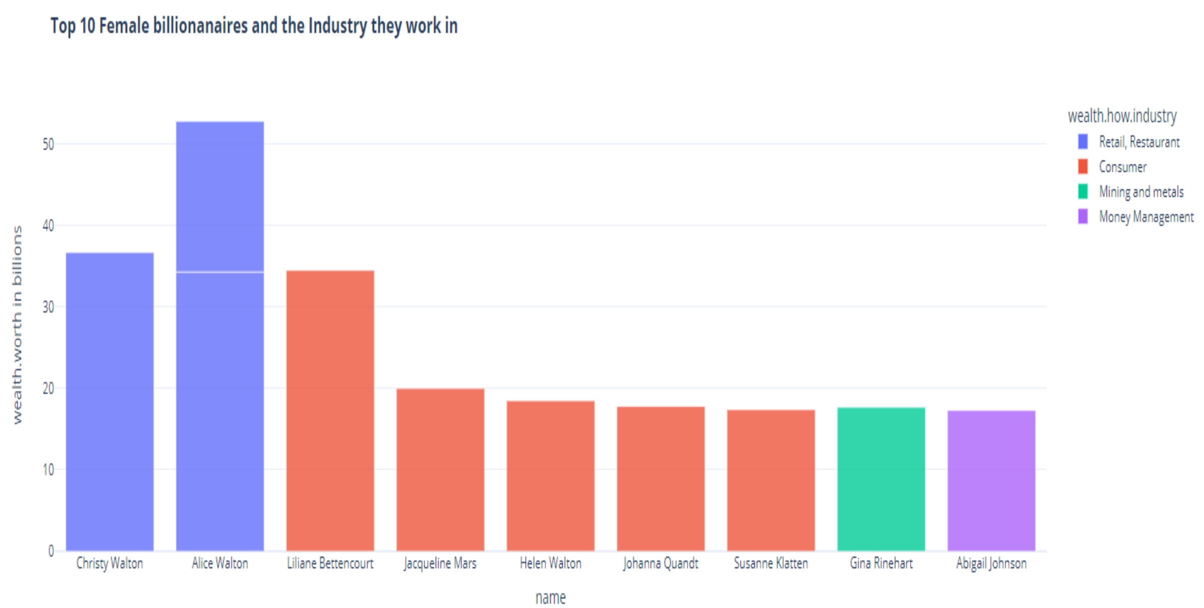


Fig. 32. A bar chart showcasing the top 10 female billionaires and the industries in which they work.

Perspectives from Fig. 32:

- The richest female billionaire works in retail and restaurants.
- Most female billionaires work in the consumer business.

The top ten industries' billionaires combined net worth is shown in Fig. 33 for 2014.

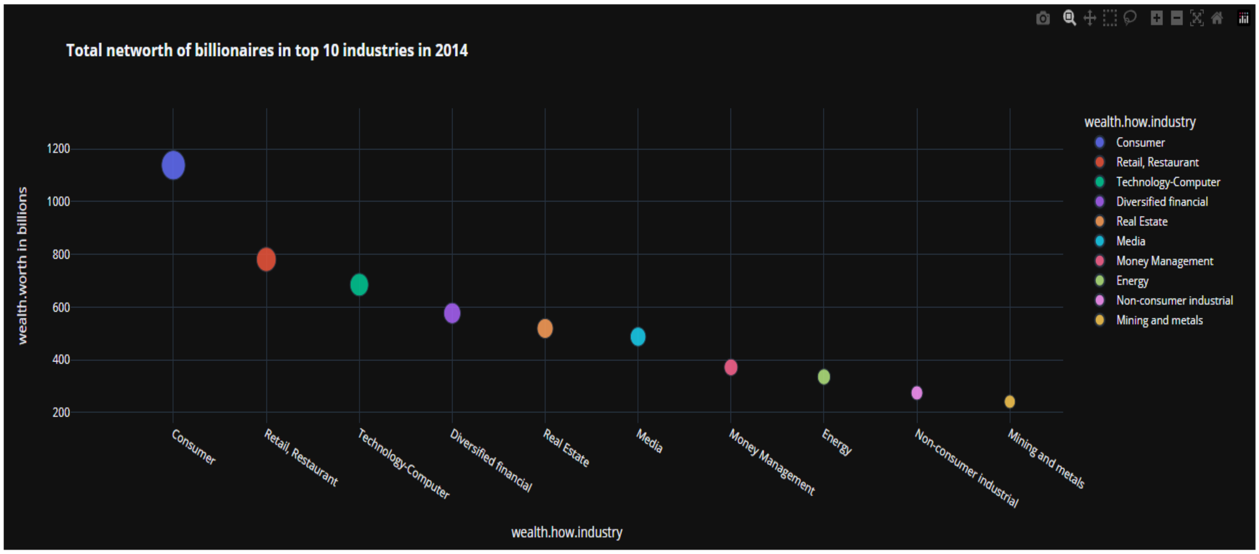


Fig. 33. Billionaires combined net worth in the leading ten sectors in 2014.

Perspectives from Fig. 33:

- The largest sector, valued at over $1 trillion, is the consumer goods sector.
- The retail, restaurant industry comes in second, with a market value of $800 billion.

XIII. Testing the Hypothesis

1) Hypothesis 1: Does a billionaire's age affect his or her net worth?

Ho: A billionaire's net worth is not significantly related to age.

H_A: A billionaire's net worth is related to age in a significant way.

The code for evaluating the first hypothesis is shown in Fig. 34.

```
import pandas as pd
from scipy.stats import pearsonr

# Compute Pearson correlation coefficient and p-value

# THIS TESTS THE HYPOTHESIS OF THE RELATIONSHIP BETWEEN AGE AND NETWORTH

corr, p_value = pearsonr(billionaires['demographics.age'], billionaires['wealth.worth in billions'])

# Print the results
print("Pearson correlation coefficient:", corr)
print("p-value:", p_value)
```

```
Pearson correlation coefficient: 0.08750438991654515
p-value: 3.990971769605956e-05
```

Fig. 34. Test code for Hypothesis 1.

Perspectives from Fig. 34:

- The association between age and net worth was examined using the Pearson correlation coefficient.
- Considering a p-value of 0.00004, it is evident that the age of a billionaire and their wealth are strongly correlated. We agree with an alternate concept as a result.
- The more elderly billionaire is most likely to be richer.

2) Hypothesis 2: Men with billion-dollar wealth are wealthier than women?

Ho: A billionaire's net wealth is unaffected by gender in a significant way.

H_A: Gender significantly affects a billionaire's financial worth.

The code to test the second hypothesis is shown in Fig. 35.

```
import scipy.stats as stats

# Filter the data for male and female billionaires
male_df = billionaires[billionaires['demographics.gender'] == 'male']
female_df = billionaires[billionaires['demographics.gender'] == 'female']

# Conduct the two-sample t-test assuming equal variance
t_stat, p_val = stats.ttest_ind(male_df['wealth.worth in billions'], female_df['wealth.worth in billions'], equal_var=True)

# Print the t-statistic and p-value
print("t-statistic:", t_stat)
print("p-value:", p_val)

t-statistic: -0.9940802457505286
p-value: 0.32029332217806183
```

Fig. 35. Test code for Hypothesis 2.

Perspectives from Fig. 35:

- This hypothesis was tested using a two-sample t-test.
- Gender and net worth are unrelated, as indicated by the p-value of 0.32, which is larger than 0.05.
- We can rule out the alternative hypothesis and adopt the null hypothesis which is billionaires' net worth is unaffected by gender.

3) Hypothesis 3: The net worth of billionaires varies considerably depending on the industry area.

Ho: A billionaire's net worth is significantly influenced by the industry sector.

H_A: A billionaire's net worth is not much impacted by their industry sector.

Fig. 36 illustrates the code for testing the third hypothesis.

```
import scipy.stats as stats

# Group the data by industry and select the net worth column
grouped_data = billionaires.groupby('wealth.how.industry')['wealth.worth in billions']

# Perform one-way ANOVA test
f_statistic, p_value = stats.f_oneway(*[grouped_data.get_group(x) for x in grouped_data.groups])

print("F-statistic:", f_statistic)
print("p-value:", p_value)

F-statistic: 2.745769472721782
p-value: 0.00022408174979504465
```

Fig. 36. Test code for Hypothesis 3.

Perspectives from Fig. 36:

- This hypothesis was investigated using the one-way ANOVA test where the p-value is less than 0.05.
- We can accept the null hypothesis and concur that the industry sector significantly influences billionaires' net wealth.

XIV. Evaluation

The study revealed how gender, age, and industry affect billionaire wealth accumulation. The retail industry has the most female billionaires, and most billionaires are men. According to our research, billionaires are usually 50–60 years old. The most senior billionaire was 98 years old, while the youngest was 21. Furthermore, real estate, media, and construction have more millionaires than other industries. The insights on billionaire wealth distribution can inform investment decisions and corporate strategies across sectors.

Our investigation has shown the key variables that enable billionaires to become rich. The above insights could help decision-makers, investors, and businesspeople understand wealth distribution and develop successful financial strategies in the global economy.

XV. Findings and Results

The following is the overall finding from the billionaire dataset analysis:

- The United States is the country that has the highest number of billionaires, which has 804 in total.
- The sector with the most billionaires is the consumer goods sector.
- The majority of billionaires are older than 50.
- 34.8% of billionaires received their wealth through inheritance.

- With a cumulative net worth of 76 billion greenbacks, invoice Gate is the richest billionaire.
- Christy Walton has a net worth of $38 billion, making her the richest female billionaire.
- The youngest billionaire is Albert Thurn, who's just 12 years old and possesses a $1 billion fortune.

XVI. Conclusions

In summary, our examination of the "Billionaires.csv" dataset utilizing Python tools for big data analytics yielded significant findings regarding the determinants that impact the accrual and allocation of riches among billionaires. Our observation revealed a noteworthy accumulation of wealth within a minority of individuals, indicating the existence of income disparity within the billionaire demographic.

The study found that entrepreneurship and technology helped a significant number of billionaires in the sample become wealthy. Technology and finance were the most profitable industries for billionaires. The analyses also showed that the US and China have the most billionaires. Billionaires also showed a trend toward gender diversity. This study sheds light on billionaires, but it has limitations. The dataset covers billionaires until 2021 and may not reflect current trends. The analysis only considered quantitative variables, ignoring qualitative factors that may contribute to wealth accumulation.

Our examination also provides a valuable contribution to the comprehension of the fluctuations in affluence within the billionaire population, exposing trends linked to entrepreneurial pursuits, sectors of operation, geographical location, and gender. Subsequent investigations ought to adopt a more all-encompassing methodology by combining qualitative and longitudinal data to acquire a more intricate comprehension of the accumulation of billionaire wealth and its societal ramifications.

XVII. Limitations and Future Work

There are a few limitations to our dataset and methodology, even though our analysis offers insightful information about the elements that lead to self-made billionaires' wealth accumulation.

- The dataset excludes billionaires who received their riches through inheritance and is only comprised of self-made billionaires. The generalizability of our findings to the total billionaire population may be impacted by this exclusion.
- The dataset may not be typical of people who have less net worth because only billionaires with total assets of more than one billion dollars are included in it. Additionally, because the data was taken over a limited period, probably, it doesn't accurately reflect patterns and shifts in wealth distribution.
- Descriptive statistics and exploratory data analysis methods played a significant role in our investigation. Although these techniques help understand the data, they do not prove causation or take into consideration confounding variables.
- The dataset may contain incomplete or erroneous data, which could produce biased or insufficient results. These restrictions must be understood and considered when interpreting the findings of our investigation.

Notwithstanding these constraints, the dataset can yield valuable insights into the determinants that contribute to the ascent of self-made billionaires. The determinants encompass the sector, age, gender, and regional attributes of the billionaires. The findings derived from this dataset can offer precious insights for companies and traders in their choice-making procedures, as well as for an academic look at the societal implications of wealth inequality. This information has the potential to be utilized for several objectives. To augment our understanding of the intricacies surrounding the wealth of billionaires and its far-reaching consequences, future research undertakings should integrate both quantitative and qualitative methodologies, incorporate longitudinal data, and investigate more extensive array of variables. This will facilitate a more holistic comprehension of the complexities associated with the amassing and dispersal of wealth among individuals in the billionaire class.

References

[1] Njangang, H., Beleck, A., Tadadjeu, S., & Kamguia, B. (2022), ‘Do icts drive wealth inequality? evidence from a dynamic panel analysis’, *Telecommunications Policy*, 46(2), 102246. Available at: https://doi.org/10.1016/j.telpol.2021.102246.

[2] Torgler, B. and Piatti, M. (2013) ‘Extraordinary wealth, globalization, and corruption’, Review of Income and Wealth, 59(2), pp. 341–359. Available at: doi:10.1111/roiw.12027.

[3] Kong, G., Xu, L. and Zhang, W. (2022) ‘The benevolence of the billionaires: Evidence from China’s Hurun Rich List’, Finance Research Letters, 48, p. 103030. Available at: doi: 10.1016/j.frl.2022.103030.

[4] Ghosh, S. (2016) ‘Billionaire wealth, firm performance and Financial Crisis: An empirical analysis for India’, South Asian Journal of Macroeconomics and Public Finance, 5(2), pp. 133–156. Available at: doi:10.1177/2277978716671050.

[5] Bagchi, S. and Svejnar, J. (2013) ‘Does wealth inequality matter for growth? the effect of billionaire wealth, income distribution, and poverty’, SSRN Electronic Journal [Preprint]. Available at: doi:10.2139/ssrn.2351448.

[6] Jacobs, D., 2015 ‘Extreme wealth is not merited’, Oxfam Discussion Papers.

[7] Hemingway, A. and Rozworski, M. (2020) ‘Canadian billionaires’ wealth skyrocketing amid the pandemic’, Ottawa, ON, CA: Canadian Centre for Policy Alternatives.

[8] Frémeaux, N. and Leturcq, M. (2022) ‘Wealth accumulation and the gender wealth gap across couples’ legal statuses and matrimonial property regimes in France’, European Journal of Population, 38(4), pp. 643–679. Available at: doi:10.1007/s10680-022-09632-5.

[9] Hamermesh, D. and Leigh, A. (2022) ‘“Beauty too rich for use”: Billionaires’ assets and attractiveness’, SSRN Electronic Journal [Preprint]. Available at: doi:10.2139/ssrn.4114299.

[10] Wai, J. and Lincoln, D. (2016) ‘Investigating the right tail of wealth: Education, cognitive ability, giving, network power, gender, ethnicity, leadership, and other characteristics’, Intelligence, 54, pp. 1–32. Available at: doi: 10.1016/j.intell.2015.11.002.

[11] List, J., Shaikh, A., & Xu, Y. (2016) ‘Multiple hypothesis testing in experimental economics’, Experimental Economics. Available at: https://doi.org/10.3386/w21875

[12] Gillis, A.S. (2021) ‘The 5 Vs of big data, Data Management.’ Available at:

https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data (Accessed: 07 May 2023).

[13] Nguyen, T.L. (2018) ‘A framework for five big Vs of Big Data and organizational culture in firms’, 2018 IEEE International Conference on Big Data (Big Data) [Preprint]. Available at: doi:10.1109/bigdata.2018.8622377.

[14] Hamermesh, D. and Leigh, A. (2021) “Beauty too rich for use”: Billionaires’ assets and attractiveness [Preprint]. Available at: doi:10.3386/w29361.

[15] Jesús Fernández-Villaverde, Hurtado, S., and Galo Nuño (2023), ‘Financial Frictions and the Wealth Distribution’, Econometrica, 91(3), pp.869–901. doi: https://doi.org/10.3982/ecta18180.

[16] Kohlrausch, G. and Goncalves, S. (2023) ‘Wealth distribution on a dynamic complex network’ [Preprint]. doi:10.2139/ssrn.4673263.

[17] Piper, R. (2023). The Institutional Drivers Contributing to Billionaire Wealth at the Sector Level. Class, Race Corporate Power, [online] 11(1). doi: https://doi.org/10.25148/crcp.11.1.010593.

[18] Gastwirth, J.L., Luo, R. and Pan, Q. (2024). A statistical examination of wealth inequality within the Forbes 400 richest families in the United States from 2000 to 2023. METRON. doi: https://doi.org/10.1007/s40300-024-00267-6.

[19] Dataset - Billionaires CSV File, https://corgis-edu.github.io/corgis/csv/billionaires/ [Accessed on 20 March 2024].

XVIII. Key Terms and Definitions

Big Data Analytics: The procedure of using advanced computer and statistical tools to analyze and clean meaningful data from huge and complex datasets.

Data Visualisation: The graphical presentation of data to highlight patterns, trends, and insights to make challenging material more approachable and understandable.

Inheritance: The practice of transferring wealth, assets, or property from one generation to the next, usually using close family ties.

Economic Growth: The amount of goods and services produced by an economy increases with time. GDP growth, which measures the total value of a nation's finished goods and services, is typically used to measure it. Economic growth reveals productivity, living standards, and prospects for both businesses and people.

Hypothesis Testing: A method that uses statistical analysis to determine whether a hypothesis or claim made about a population is supported by the data from a sample of that community.

Data Pre-processing: The manipulation of raw data using a variety of methods to prepare it for subsequent processing is what is referred to as "data pre-processing," and it is an essential part of the data preparation process.

Python: A high-level programming language that is well-known for its ease of use and readability and that finds widespread use in the fields of scientific computing, machine learning, and data analysis.