**Exploring the Evolution of Billionaires**

**A Data-Driven Analysis for the Last Two Decades**

**Omar Emara 228248**

**Amir Mohamed 224030**

Table of Contents

[1. Introduction 3](#_Toc165737880)

[2. Datasets 3](#_Toc165737881)

[2.1. Billionaries Dataset 3](#_Toc165737882)

[2.2. Genders Dataset 4](#_Toc165737883)

[2.3. Countries Dataset 4](#_Toc165737884)

[3. The Journey Through Data 5](#_Toc165737885)

[3.1. Preparation and Initial Setup 5](#_Toc165737886)

[3.2. Cleaning, Organizing, and Extending Dataset 5](#_Toc165737887)

[3.3. Research Design 6](#_Toc165737888)

[7](#_Toc165737889)

[3.4. Hypothesis Testing 8](#_Toc165737890)

[3.5. Concluding Insights 8](#_Toc165737891)

[4. Exploratory Questions 9](#_Toc165737892)

[4.1. Visualizations for Internal Data Analysis Questions 9](#_Toc165737893)

[14](#_Toc165737894)

[4.2. Visualizations for External Data Analysis Questions 15](#_Toc165737895)

[5. Key Insights 18](#_Toc165737896)

[5.1. Analysis of Age Across Industries and Nationalities 18](#_Toc165737897)

[5.2. Impact of Gender and Citizenship 18](#_Toc165737898)

[5.3. Guaranteed Investment Countries 18](#_Toc165737899)

[5.4. Hypothesis Testing on Age and Net Worth 18](#_Toc165737900)

[6. Conclusion 19](#_Toc165737901)

## Introduction

The "Exploring the Evolution of Billionaires: over the last two decades" project aims to analyze two decades of billionaire wealth data. Through detailed data analysis, we seek to uncover trends, shifts, and determinants influencing billionaire wealth accumulation. Leveraging a comprehensive dataset covering net worth, industry affiliations, and demographic attributes, our goal is to provide insights into the factors driving the emergence, growth, and distribution of billionaire fortunes. This project offers a nuanced understanding of global wealth dynamics, benefiting researchers, policymakers, and those interested in the economics of success.

## Datasets

### 2.1. Billionaries Dataset

The dataset has 31,732 records and 19 features and can be accessed through [from here](https://www.kaggle.com/datasets/guillemservera/forbes-billionaires-1997-2023?select=all_billionaires_1997_2023.csv). An excel sheet of the dataset is also included in the same directory as this proposal. The dataset encompasses the following features:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature Name |  | Feature Type | Description |
| Year | Integer | Interval | The year in which the billionaire data was recorded. |
| Month | Integer | Interval | The month in which the billionaire data was recorded. |
| Rank | Integer | Ordinal | The rank of the billionaire based on their net worth in a particular year. |
| Net Worth | Float | Ratio | The net worth of the billionaire in USD. |
| Last Name | String | Nominal | The last name of the billionaire. |
| First Name | String | Nominal | The first name of the billionaire. |
| Full name | String | Nominal | The full name of the billionaire. |
| Birth Date | Date | Interval | The birth date of the billionaire. |
| Age | Integer | Ratio | The age of the billionaire. |
| Gender | String | Nominal | The gender of the billionaire. |
| Country of Citizenship | String | Nominal | The country of citizenship of the billionaire. |
| Country of Residence | String | Nominal | The country of residence of the billionaire. |
| City of Residence | String | Nominal | The city of residence of the billionaire. |
| ****Business Category**** | String | Nominal | The category of the billionaire's business (e.g., technology, finance). |
| ****Business Industries**** | String | Nominal | The industries associated with the billionaire's businesses. |
| ****Organization Name**** | String | Nominal | The name of the organization associated with the billionaire. |
| Position in Organization | String | Nominal | The position held by the billionaire in the organization. |
| Self-Made | Boolean | Nominal | Indicates whether the billionaire is self-made or not. |
| ****Wealth Status**** | String | Nominal | Indicates if the billionaire's net worth has increased, decreased, remained stable, or returned to the list compared to a previous data point. |

### 2.2. Genders Dataset

This dataset attributes first names to genders, giving counts and probabilities. It combines open-source government data from the US, UK, Canada, and Australia. For more details about the dataset visit [this link](https://archive.ics.uci.edu/dataset/591/gender+by+name). This dataset was used to deduce the gneders of Billionaries’ records that were missing this feature.

### 2.3. Countries Dataset

This dataset provides detailed information on countries, including their capitals, populations, and geographical areas. It covers various demographic and territorial attributes, offering a comprehensive overview of nations worldwide. This dataset was used to fill in the city\_of\_residence feature for the records missing it. For more information you can access the dataset following [this link](https://www.kaggle.com/datasets/sirishasingla1906/countries-of-the-world?resource=download).

## The Journey Through Data

### 3.1. Preparation and Initial Setup

* Dependencies Installation: Before diving into the data, necessary libraries are installed.
* Data Loading: The journey begins with loading the "billionaires.csv" dataset, setting the stage for the analyses that follow.

### 3.2. Cleaning, Organizing, and Extending Dataset

* Initial Exploration: Examination of the dataset's structure, identifying data types, and visualizing missing values.
* Data Cleaning: Cleaning includes filling missing values, standardizing the format of numerical entries, and removing unnecessary columns. This step was crucial for ensuring the accuracy of our analyses.
* Feature Engineering: New features are computed, such as deducing if billionaires are self-made, based on yearly data.
* A screen shot of a computer program

  Description automatically generatedA computer screen shot of a program code

  Description automatically generatedExternal Data Integration: Additional data sources are integrated, like startup-friendly and GNI indexes, providing broader economic context and enhancing the robustness of our study.

The followings are code snippets of filling missing birth\_year and birth\_date features and cleaning the business\_industries feature.

### 3.3. Research Design

* Question Formulation: We formulated specific questions that could be answered directly from the dataset and identified gaps that required additional data from external sources.
* Data Extenstion through Web Scraping: For questions beyond the scope of the initial dataset, we augmented our analysis with data scraped from various online resources, enriching our dataset with broader socio-economic and demographic insights.

A computer screen shot of a program

Description automatically generatedThe following snippet is for the code used to scrap some of the data needed for the analysis.

### A computer code on a black background Description automatically generatedA computer screen shot of text Description automatically generated

### 3.4. Hypothesis Testing

* Formulating Hypothesis: A hypothesis is articulated, proposing that there is no significant correlation between age and net worth of billionaires.
* Statistical Testing: Through detailed calculations of mean, variance, and t-scores, the hypothesis was tested.

### 3.5. Concluding Insights

* Visual Insights: We used visualizations to clearly present the answers derived from our data, making complex trends easily understandable.
* Hypothesis Outcome: The hypothesis testing revealed a significant correlation between age and net worth, leading to the rejection of the null hypothesis and providing new insights into wealth accumulation factors.

The following snippet of code is for the hypothesis testing and applying T-test to the hypothesis.

A screenshot of a computer program

Description automatically generated

## **Exploratory Questions**

This section highlights the key questions we explored in the analysis of billionaire wealth evolution over the last two decades.

### 4.1. Visualizations for Internal Data Analysis Questions

#### 4.1.1. How does the average age at which billionaires achieve their status differ across industries, countries of origin, and gender over the years of study?

A graph of different colored lines

Description automatically generated

A graph of blue and orange bars

Description automatically generated

#### 4.1.2. Do specific countries or regions have a higher concentration of billionaires who specialize in particular business categories?

#### A computer screen with many squares Description automatically generated

#### 4.1.3. How does the rate of wealth growth or decline vary between self-made and non-self-made billionaires over time?

#### **A graph of a line graph Description automatically generated with medium confidence**

#### 4.1.4. What are the top 3 industries with the highest concentration of billionaires, and how does this distribution change over the years?

#### **A graph with different colored lines Description automatically generated**

#### 4.1.5. What are the industries that are more guaranteed to generate profit based on the number of billionaires that work in the industry and the amount of net worth they have from their industries?

**A graph with many colored dots

Description automatically generated**

#### 4.1.6. What are the safest industries to invest in that grew (or) took the least damage during the global economic crisis in 2008 and Covid?

#### A chart of a number of companies Description automatically generated with medium confidence

### A chart of a number of companies Description automatically generated with medium confidence

### 4.2. Visualizations for External Data Analysis Questions

#### 4.2.1. How do tax rates and GNI index jointly influence the average net worth of billionaires in different countries?

A graph with different colored dots

Description automatically generated

#### 4.2.2. What are the most start-up friendly countries for the top industries?

A graph with different colored lines

Description automatically generated

#### 4.2.3. What are the most guaranteed countries to invest in?

A graph of different colored bars

Description automatically generated

## Key Insights

### 5.1. Analysis of Age Across Industries and Nationalities

* There is a noticeable difference in the average age of billionaires across industries. Billionaires in industries like "Gambling & Casinos" and "Hotels & Resorts" tend to be older, while those in tech-related fields achieve billionaire status at a younger age. This pattern has remained consistent over the study period.
* Billionaires from China and Russia tend to be younger compared to their counterparts from other nations, suggesting these countries may offer faster paths to significant wealth, particularly in technology and manufacturing sectors.

### 5.2. Impact of Gender and Citizenship

* The analysis explored how gender and citizenship affect if individuals become billionaires. Results indicate that these factors significantly influence the distribution of wealth and status achievement across different industries and countries.

### 5.3. Guaranteed Investment Countries

* The analysis identified countries like Lithuania, Canada, France, Portugal, and Poland as "guaranteed" investment destinations based on a combination of high Startup Friendly Scores and strong economic indicators like GNI.
* These countries offer a balance of economic stability and favorable conditions for startups, making them recommended locations for investment.

### 5.4. Hypothesis Testing on Age and Net Worth

* The hypothesis testing confirmed a significant correlation between age and net worth among billionaires. This suggests that as billionaires age, their net worth tends to increase, possibly reflecting the accumulation of wealth over time and the expansion of their business interests.

## Conclusion

Our journey through "Exploring the Evolution of Billionaires: A Data-Driven Analysis over the last two decades" has been both enlightening and impactful. This comprehensive analysis has not only deepened our understanding of how billionaires accumulate wealth but also provided valuable insights into the dynamics that influence this accumulation over time.

We discovered significant variations in the age of billionaires across industries and nationalities, noting younger billionaires in tech and emerging markets, particularly in China and Russia. Our study also highlighted the role of gender and citizenship in shaping the paths to wealth, unveiling a complex interplay of economic, social, and geographic factors.

Moreover, our analysis of investment environments in various countries identified promising opportunities in nations like Lithuania, Canada, France, Portugal, and Poland, underpinned by robust economic indicators and startup-friendly policies.

The project supported in a powerful affirmation of our hypothesis, revealing a significant correlation between age and net worth among billionaires, which underscores the importance of experience and maturity in wealth accumulation. In essence, this project has provided insights that not only enhance our comprehension of billionaire wealth but also offer practical guidance for policymakers, investors, and entrepreneurs globally.