The Global Billionaire Landscape: An Analysis of Wealth, Industry, and Demographics

Name: Sean Dowdall
Student Number: S00210945

Table of Contents

1. Introduction

1.1 Project Overview

This project aims to explore the global distribution and characteristics of billionaires and wealth accumulation. This research also aims to explore the patterns and dynamics of extreme wealth in the world today. Throughout this research, various aspects of billionaires profiles, including demographics, industry affiliation and geographic distribution have been analysed, outlining growing economic differences and providing insight into the accumulation and distribution of wealth at the highest level of society today.

1.2 Dataset Description

The dataset chosen for this project was the Billionaires Statistic Dataset 2023. This is a dataset analysing the world's wealthiest people, and was created by Nidula Elgiriyewithana. This dataset was posted and made available for download on Kaggle in 2023. The dataset contained 2540 rows and 35 columns of billionaire's data. The billionaires in the dataset were ranked from highest net worth to lowest net worth and some key variables of this dataset include the age, gender, net worth, industry and country of origin of the billionaires.

2. Exploratory Data Analysis (EDA)

2.1 Data Cleaning and Preparation

Understanding that the quality and integrity of my data analysis hinges on the quality of the underlying data, the data was cleaned thoroughly to ensure valid results could be achieved. This involved first identifying and removing duplicate entries in the dataset. When handling missing data, the following approach was taken: Median values were used to impute the missing numerical data to ensure minimal skewness. Modal values were used for categorical data, maintaining the integrity of the non numerical information. Following the completion of these steps, there were still some rows in the dataframe with a high percentage of missing data. As the rows were not going to be used in my analysis, they were dropped. At this stage it was confirmed that the entire dataset had been cleaned effectively, and the dataset was ready for exploratory data analysis.

2.2 EDA Interpretations

The exploratory analysis provided a fundamental understanding of the billionaires dataset. Starting out with some simple graphical representations of the top 10 countries with the most billionaires, in the form of a bar chart and moving onto a geospatial view of all the countries with the most billionaires, calculated by their country of origin and mean latitude and longitude. This analysis highlighted the United States as the predominant home of the most billionaires, which raised questions about the economic and policy factors in America contributing to this concentration. Analysis of the biggest billionaire producing industries revealed sectors such as technology and finance as significant wealth generators. Another notable finding was the gender imbalance of billionaires in 2023. Out of all billionaires in our population, 88.1% of them were male. This may prompt considerations of social and economic barriers to wealth accumulation that could be considered moving forward. Self-made vs inherited billionaires were also analysed, concluding that 70% of all the billionaires in our population were self-made and did not inherit their wealth.

3. Questions

3.1 What questions were explored

Several questions were raised following the initial exploratory data analysis. The following questions asked in this section proved helpful in understanding the global billionaire landscape in 2023. The following questions were raised:

1. How is wealth distributed across different industries and are certain industries more likely to produce billionaires?

2. Which countries or regions have the highest concentration of billionaires and how does this relate to the economic indicators for those countries.

3. Is there a correlation between the age of billionaires and their net worth? Do younger billionaires tend to be wealthier in certain industries compared to older ones?

4. What is the gender distribution among billionaires and are there significant differences in the average net worth of male and female billionaires?

5. Is there a relationship between the level of education (e.g., gross tertiary education enrollment) in a billionaire's country of citizenship and their net worth?

6. What proportion of billionaires are self-made versus those who inherited their wealth? Is there a significant difference in the average net worth between these two groups?

These questions proved pivotal in moving forward with the analysis and encouraged future hypothesis tests, correlations and predictions in the later sections of the analysis.

4. Analysis Results

4.1 Hypothesis Testing

Several statistical tests were used in the study to further explore the relationships found in the data. The average net worth of self-made versus inherited billionaires was compared using t-tests. The tests did not show a statistically significant difference in the net worth of these two groups. Chi-squared tests were applied to investigate the industry distribution between the different genders. The findings revealed a statistically significant difference, suggesting that there are gender differences in the representation of billionaires in the industry.

Anova tests were also carried out to discover significant differences between the average net worth of billionaires across different countries of citizenship. The Anova test results revealed that the country of citizenship does not have a significant impact on the net worth of billionaires.

To ensure integrity from the hypothesis tests, for each population test taken, two samples were also taken and tested for each hypothesis. This allowed us to compare the results and ensured we were making correct assumptions based on our data analysis.

These tests revealed that the country of citizenship does not have a significant impact on the net worth of billionaires, and interestingly, in the distribution of genders across different industries hypothesis test, discrepancies were found between the sample tests and population test.

There was significant evidence to reject the null hypothesis that the distribution of industries is the same across different genders following the population test, but the opposite determination

was found during the sample tests. This was likely due to the underrepresentation of women in the small samples, leading to inaccurate results.

4.2 Correlations

A key component of this study was correlation analysis, which highlighted complex relationships in the data. One important finding showed us that, although education contributes to wealth accumulation, it is not the primary factor. This weak positive correlation was found between tertiary education enrollment rates and billionaires net worth. Covariance between gross tertiary education enrollment and net worth was calculated at this point, alongside the p-value and Pearson Correlation Coefficient. This data was visualised on a scatter point with a regression line to conclude the result of a weak positive correlation.

Furthermore, the research revealed a noteworthy association between the age of a billionaire and their industry of work, emphasising that the younger generations gravitate towards the technology sector. Another chi-squared hypothesis test was utilised in this section of the analysis. The chi-squared test result suggests that age significantly influences the choice of industry for wealth accumulation among billionaires. The data was viewed on a contingency table before moving forward with a bar chart visualisation of the wealth distribution of age groups across different industries, and a heatmap visualisation of the same results.

4.3 Predictions

The study used logistic and linear regression analyses to try to predict the relationships outlined in the previous section. A weak but statistically significant correlation was confirmed by the linear regression analysis, which looked at the relationship between net worth and tertiary education enrollment rates. Based on an individual's age, the probability of them being connected to the technology sector was estimated using logistic regression. The findings showed a strong trend of young billionaires being in the technology industry. Multiple visualisations accompany this data to provide a thorough understanding of the impact.

5. Conclusion

In conclusion, this study provided many fascinating insights into the global billionaire landscape in 2023. The discovery was made that the US and China, are home to a significant number of the billionaire’s in the world today. That the billionaire landscape is largely male dominated and that the majority of billionaires had self-made rather than inherited wealth. The industry of choice for these billionaires included finance and investing, technology and manufacturing. Interestingly we could determine through regression analysis that younger billionaires predominantly operated and generated wealth predominantly in the technology sector, and that a country's third level (tertiary) education enrollment rate had a minor positive correlation with their billionaires net worth. Overall this analysis provided interesting and insightful information regarding wealth creation, accumulation and distribution and the many factors that influence the global billionaires landscape in 2023.

6. References

www.kaggle.com. (Elgiriyewithana, D. 2023.). *Billionaires Statistics Dataset (2023)*. [online] Available at: https://www.kaggle.com/datasets/nelgiriyewithana/billionaires-statistics-dataset.