Lesley-w19 / **Phase3-Project** Public

Billionaire Dataset exploration in this Project 3

0 stars 0 forks Activity

Star Notifications

Code Issues Pull requests Actions Projects Security Insights

main Go to file

Lesley-w19 Update on the readme file ... now 22

View code

README.md

PHASE 3 PROJECT

Stepping into Billionaire Territory!!

Project Overview

This project requires the use Classification Model to generate insights for a given Agency. This agency provides all sorts of social support in the different industries, mostly volunteer work. In a bid to increase their profits for the next financial year, their have been looking through different datasets that will enable them make a choice of industriy investment.

Business Understanding

Stakeholder Audience

Our Stakeholder are Soar High Research Agency, a non-governmental organization that deals with Social Development. They are looking for an industry to invest in that will yield profit that will be able to fund their various Activities (eg. school sponsorships). Based on their source question, we are required to provide insights on the wealth distribution status and profitable industries they are to venture in. Also, as the kind of industries they are able to encourage their students to take.

Business Objectives

The aim of this project is to analyze the provided dataset on King Country and eventually answer to the Stakeholders questions. With this in mind, the objectives of this project include: The objectives of this project based on the dataset chosen is to find out:

1. The Wealth Status of Billionaires (whether self-made or not)
2. What industry/sources is more inclined to produce billionaires?
3. Demographic analysis of billionaires (age, gender, country)
4. Provide classification to the wealth status of billionaires based on the features.
5. Provide insights into which industry are likely to produce billionaires in future (logistic regression)g

Data Understanding

For this project, we will use the Billionaires Statistics dataset under the file path `data\...` This folder contains historical data of billionaire information captured in a northwestern country, which we will modify for the purpose of our analysis.

The data folder includes the following file:

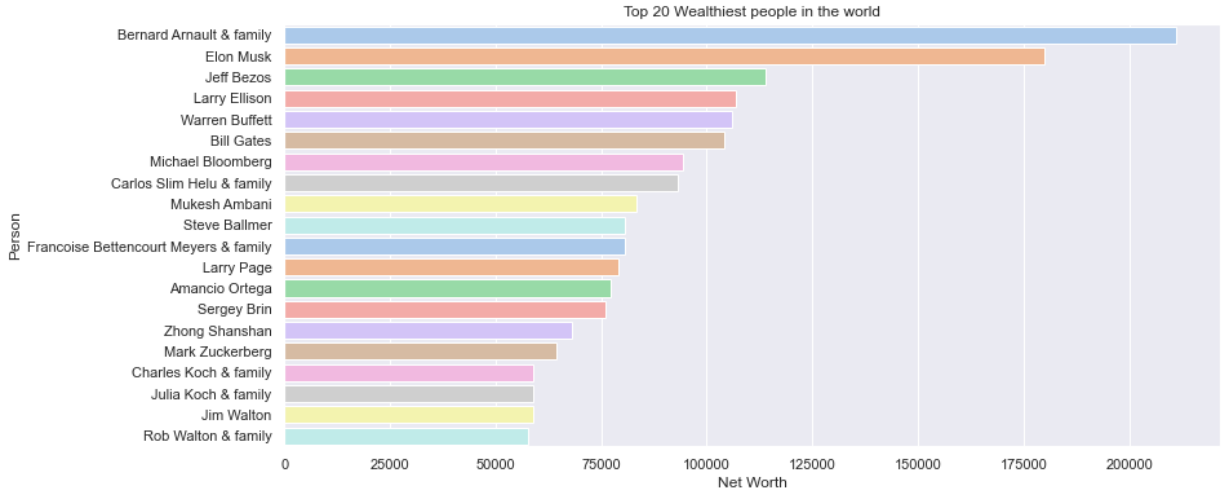
1. `Billionaires Statistics Dataset.csv` - provides the billionaire dataset which will be used for our analysis

The dataset represented has the following data columns as some of the header categories: `rank` : The ranking of the billionaire in terms of wealth. `finalWorth` : The final net worth of the billionaire in U.S. dollars. `category` : The category or industry in which the billionaire's business operates. `personName` : The full name of the billionaire. `age` : The age of the billionaire. `country` : The country in which the billionaire resides. `city` : The city in which the billionaire resides. `source` : The source of the billionaire's wealth. `industries` : The industries associated with the billionaire's business interests. `countryOfCitizenship` : The country of citizenship of the billionaire.

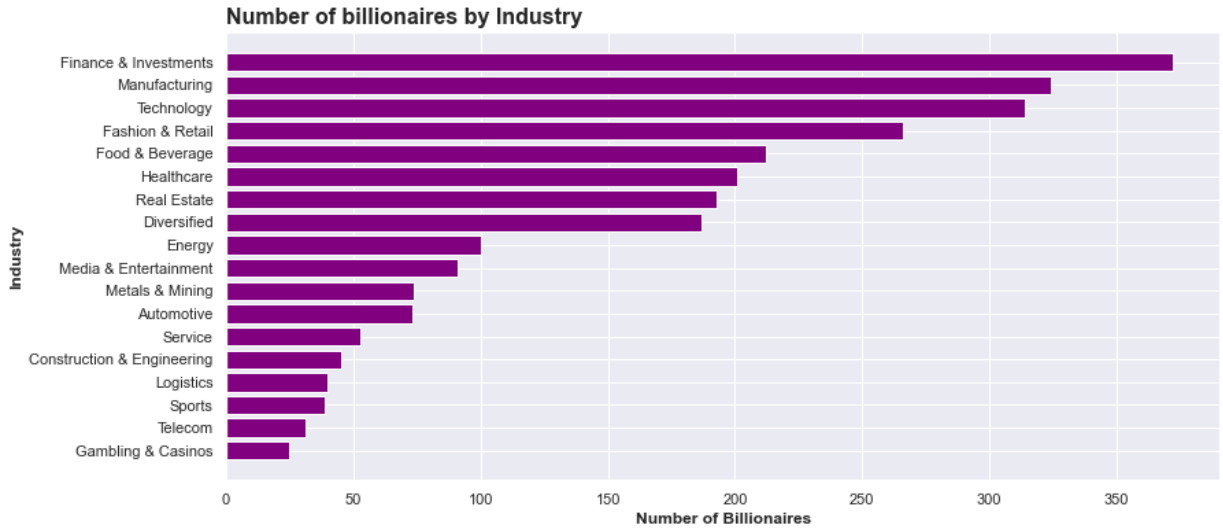
Visualization

This section shows the visual representation of billionaire information analysed. Some of the information represented include:

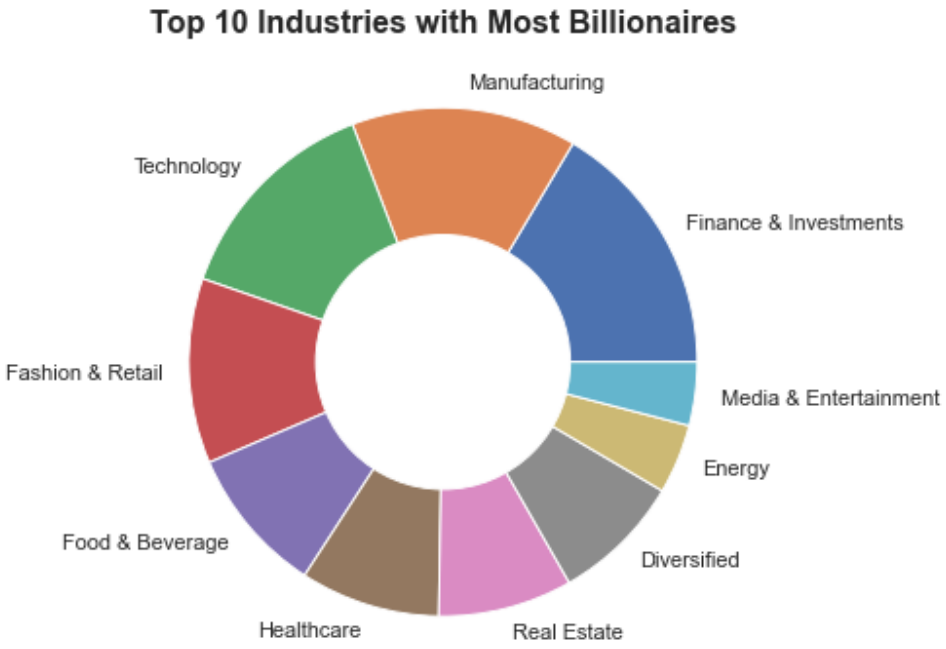
1. Top Billionaire information

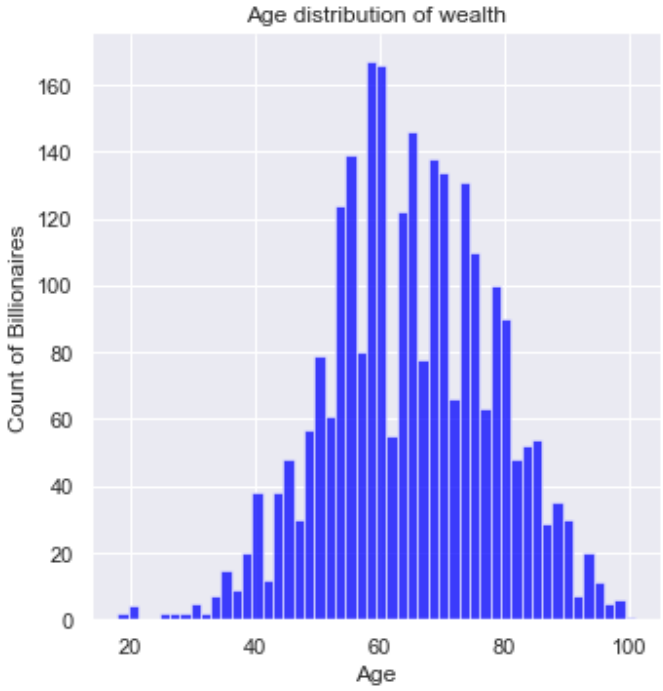


2. Top Industries



3. Industry Information with wealth





4. Billionaire age distribution

Modelling

1. Perform a Train-Test Split
2. Preprocess Data
3. Model Fitting
4. Model Evaluation

Recommendation

Findings

1. From The age distribution analysis, Most Billionaires in the world lie in the age group bracket of 50 -60 years
2. The top five countries of citizenship for most billionaires include : United States, China, India, Germany, Russia
3. The top industry with most number of billionaires in the world are: Finance and Investments, Manufacturing, Technology, Fashion & Retail, Food & Beverage
4. The major sources of wealth include: Investments, Real Estate, Software, Pharmaceuticals
5. On the question on whether Most billionaire acquired their wealth themselves or Inherited, we have seen that most of them have inherited

We are able to use the prediction model to provide insights into what variables are likely to affect the industry wealth distributuin and whether or not the choice of industry affect a persons ability to become a billionaire

Recommendations

To the Research Agency, I would recommend investing in the Finance & Investments, Technology industries, as it shows potential of growing, not only for the younger generation but also the aging. Also,

Conclusions

From the project above, we are able to see that the industry with a high number of wealth distributed is the Finance & Investments industry. Hence we are able to conclude that, based on the wealth age distribution, wealth is accumulated over the years, having best invested in the right industry.

The various logistic modelling, present an almost perfect accuracy score based on the variables indicated. The technologies used for this project include: Data Preparation, Exploratory Data Analysis methods, Data Modelling

Releases

No releases published

Packages

No packages published

Languages

- **Jupyter Notebook** 100.0%