# Project Report GitHub URL

<https://github.com/Rachelm475/UCDPA_rachelmoloney2>

# Abstract

This project is an analysis of Public Company data from the Forbes Global 2000 list from 2018 to 2021, to identify market trends across various regions with a deeper dive into the US market in 2021. I used Jupiter Notebook for my analysis and publicly available datasets in Kaggle.

# Introduction

I chose this project use case as I thought it would be interesting to analyse how the world’s biggest companies and the largest regional markets endured during the Covid19 pandemic. Given the huge disruptions we have seen across the economy in the past couple of years as a result of the pandemic, I was interested to analyse where investors might be hedging their bets and what sectors may have not just survived the pandemic but may be prospering.

Since 2003, Forbes’ Global 2000 list has measured the world’s largest public companies in terms of four equally weighted metrics: assets, market value, sales and profits. I wanted to use a subset of the Top 500 Public Companies from this data to draw some insights on global markets and which sectors investors may favour in the next couple of years.

# Dataset

All dataset csv files have been added to the git hub repository as a folder ‘Datasets’.

1. The Forbes 2018, 2019, 2020 and 2021 datasets were sub-setted from the following source:

<https://www.kaggle.com/arjunprasadsarkhel/forbes-top-200020172021>

I chose this source as it was publicly available and provided Year over Year data for my analysis, that allowed me to get a picture of the data post pre and post pandemic.

1. The Fortune 1000 dataset was sourced from Kaggle also:

<https://www.kaggle.com/winston56/fortune-500-data-2021>

I chose this source as it was publicly available and furnished my Forbes dataset with further variables to do a deeper analysis of high performing sectors in 2021.

# Implementation Process

## Phase One: Forbes Global 2000 Analysis (2018-2021)

I began by editing the I started by editing 3 of the Forbes Global 2000 csv datasets (2018-2020), to take just the first 500 companies from each year. I then imported each of the csv datasets as DataFrames (dfs) and cleaned individually by checking for missing values. For any of the Global 2000 companies listed across each year, I checked for missing values across the variables I might run analysis on (country, sales, profits, assets & market value). I removed any companies whereby there was missing data in any of these variables, as to not affect sums or means calculations further on in the project.

I noticed in the Forbes 2021 dataset there was 2100 rows whereas there was 2000 in other datasets. I imported this as was to identify where there might be an error. Upon checking which columns had null values and cross-checking against the official online Forbes 2021 list (link in sources), I noticed the rankings in the dataset were mislabelled and there was duplicate rows in the last 100 lines.

I created a “New Rank” column to correct the labelling and deleted the duplicates from the last 100 lines. I then checked the Forbes online list again to ensure the first and last company matched my data, by printing the new F21 DataFrame (df). I exported the new F21 df as a CSV, removed the last 1500 rows to take the top 500 and re-uploaded to by notebook (renaming it as F21 and checking again for any missing data). I renamed to the ‘Rank’ column in F18, F19, F20 to ensure this matched against F21.

I then combined all 4 dfs into a new df, FComp, using concat given that there wasn’t a common column to join or merge them (as the companies listed and their rankings vary each year). I checked the shape to identify the number of companies that were removed in total over all years due to missing data (20 in total).

I wanted to perform some analysis to understand which Countries were most represented in the Forbes Global 2000 list over each year, by the count of companies grouped by each region. I opted to run the analysis across the DataFrame created for each year (F18, F19, F20 and F21).

I grouped each by Country, and sorted largest to smallest to get the top 5 countries by number of companies who made the list each year. I then manually created a new DataFrame (df\_1) to show this. I then created a grouped bar chart to show Country representation of the Top 5 Year over Year. **Please see Results #1 for more info.**

I then wanted to understand which Countries had the highest average Market Value over the given time period. Given that the Forbes list is chosen by measuring the world’s largest public companies in terms of four equally weighted metrics: assets, market value, sales and profits, I felt it might make an interesting sample to draw some high-level conclusions on the size of the respective economies of those countries most represented.

I used the merged DataFrame (FComp) to analyse the Market Value data for each Country, grouping by both Country & Year and adding the variables from each Country accordingly. This created a new DataFrame called grouped\_multiple. I then got the Top 5 Countries with the highest average market value over the time period of 2018-2021.

I then created a list ‘search\_values’, to run a search on the grouped\_multiple DataFrame and create a new DataFrame to show the variable totals from each region, segmented by year. I then create another grouped bar chart to show the Market Value trends by those countries YoY. **Please see below in Results #2.**

To verify if my sample was representative of the size of economies, I created an dictionary called GDP\_nom\_21 which shows the top 10 countries ranked by nominal GDP (in $bn) in 2021 (My source for this data is listed as #2 in References). All countries in the Top 5 by Market Value feature in the Top 10 by Nominal GDP list. I also created a segmented df called g\_m\_21 to pull out the total market values by country in 2021. I merged both dfs and created a scatterplot which seems to show correlation:

Chart, scatter chart

Description automatically generated

## Phase Two: United States; Further Analysis of Market Growth in 2021

Given the apparent dominance of the US market across all metrics tested so far, and the sharp increase in combined market value of US Public Companies from the Forbes Global 500 sample, I was interested to do some further analysis on the US market in 2021.

Using the Forbes-2021 dataset, I segmented out the US companies and created a new dataset: Forbes\_21\_US. I then found another dataset Fortune\_1000\_data with additional information on US Public Companies from 2021, including sector labelling and employee headcount data per company.

I uploaded both datasets and merged them on the ‘Company’ column. These datasets are not a like for like match, with Fortune having a different process to select and rank companies than Forbes do for their list. To create a new sample dataset (US\_merge), I ran an isnull() call to identify any rows from the newly created DataFrame that did not have all variable data points I wanted to analyse [Sales, Profits, Assets, Market Value, Num. of Employees and Sector]. I removed these rows (companies), leaving me with a new sample dataset of 374 US Public Companies and their 2021 data to analyse further.

Firstly, I grouped the US companies by sector and counted them to understand which sectors were most represented in Forbes 2021 Global 2000 lists. **Please refer to Results#3 for the resulting pie-chart, showing the sectors where the highest number of companies made the list.**

I then wanted to assess the top sectors by across a number of variables; average Sales, Market Value, Assets and Profitability. I decided to average the variables, rather than sum it in order to get a more balanced picture of the data given that some sectors, such as Financials, are more represented in the sample dataset. Only 4 sectors made the ‘Top 10’ list across all variables – **please see Results #4.**

I wanted to better understand then the drivers for Market Value, and see which variable had the biggest impact on growing Market Value. I ran 3 scatter plots, to see if Sales performance, Profitability or Assets showed the most correlation to Market Value. **Please see Results #5 for outputs.** My findings outlined in Results #5 prompted me to look more closely at Profitability, to understand if any US companies that made the Forbes Global 500 list actually ran losses for 2021.

I used a ‘for loop’ to create a new column in my US\_merge df, whereby any companies with a figure less than 0 in the Profits column would be listed as ‘Not Profitable’. I then sliced the df to create a new df ‘losses’, that showed there were 46 US companies that made the Global 500 list in 2021 even while running losses. I created a seaborn scatterplot to verify my findings in results#5 (**detailed in Results #6**). Out of curiosity, I then grouped them by sector and pulled the 5 with the highest number of companies reporting losses – **please see Results #7 for the output**.

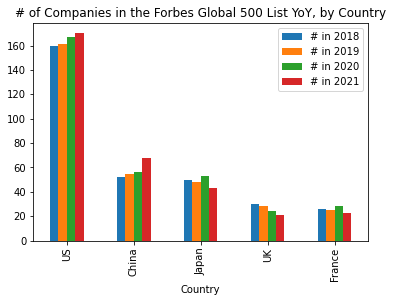
Given that I had employee count data from the Fortune 1000 dataset, I wanted to also look at the productivity levels across different sectors by analysing employee headcount and total sales. I added a column to US\_merge called ‘Productivity’ that divided Sales by Headcount. As the Fortune 1000 data listed employees by actual number (1000’s) while the Forbes data was abbreviated to 100’s to represent $billions, I had to multiply the output by 100 to get the relative productivity by sector when we use the groupby and nlargest functions. The Financial Sector seems to outpace other industries in it’s ability to reach high volume sales with lower headcounts, and Technology sector again makes the Top 10 List across this variable.

# Results

### # of Companies in the Forbes Global 2000 List YoY, by Country

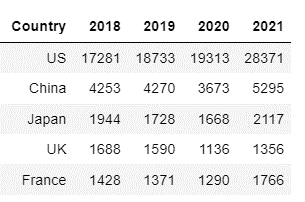
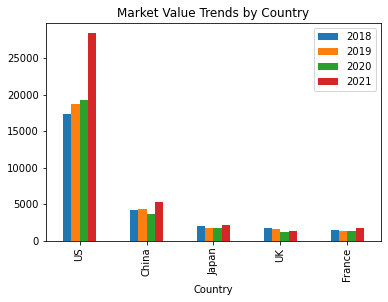
Table

Description automatically generated



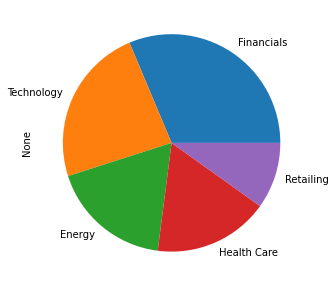
* The United States dominates these lists YoY, with US companies represented in the Top 500 growing from 32% in 2018 to 34% in 2021.
* By comparison, in second place to the US, Chinese companies made up 10% in 2018 and grew to 14% in 2021.
* While US companies are more heavily represented, Chinese representation has grown more quickly over the past 4 years, with an increase of US companies of 6% from 2018-2021 vs an increase of nearly 31% of Chinese firms in the same time.
* Interestingly , the UK has seen a steady decline in representation of the Top 500 in the past 4 years, which could be a result of either / both Brexit and the Covid19 pandemic.

### Market Value Trends by Country; Forbes Global 500 List



* Unsurprisingly, the US shows the largest combined market value given it’s high representation in the Forbes Global 500 sample.
* China interestingly saw a drop in combined market value from 2019 to 2020, despite the total number of companies represented in the sample increasing by 1 in 2021. This is likely due to the market uncertainty caused by the Covid19 pandemic. In fact, apart from the US, all countries in the Top 5 by Market Value saws drops in combined market value across this sample from 2019 to 2020.
* All countries saw uplifts again in combined market value in 2021 which may more reflective of the recovery of the wider economy as more of society re-opened with easing Covid19 restrictions in each region.
* The US has seen the biggest increase in market value from 2020 to 2021; a 47% uplift in combined market value despite only seeing less than 2% uplift in the number of companies represented in this time period.

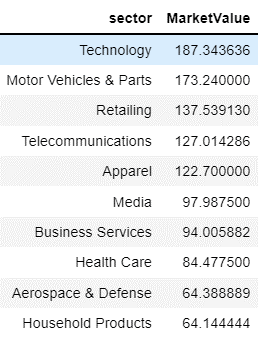
### US Most Dominant Sectors; Based on Forbes Global 2000 US 2021 Sample

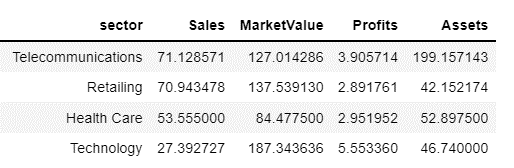
 Table

Description automatically generated

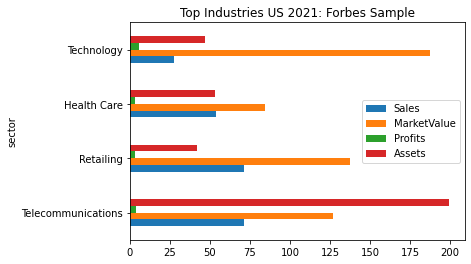
### Average Sector Performance in US 2021: Sales, Profits, Assets & Market Value



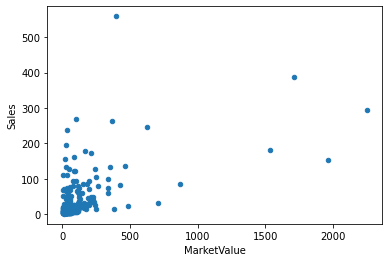
* While Food & Drug Stores see the highest average Sales, and Financials is the most Asset rich industry, neither of these feature in the Top 10 by Average Market Value.
* I then merged all 4 DataFrames to identify if any sectors perform in the top 10 across all metrics;

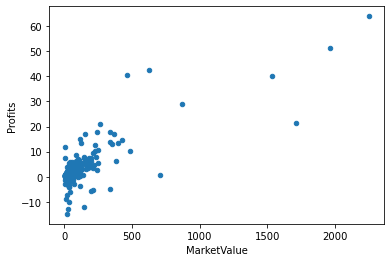


* I used a scatterplot to visualise the sector with the highest market value, that also sees strong sales, profits and assets management and the Technology sector does seem to come out on top from both a profitability and market value point of view:



### Performance across multiple variables and it’s relationship with Market Value





# 

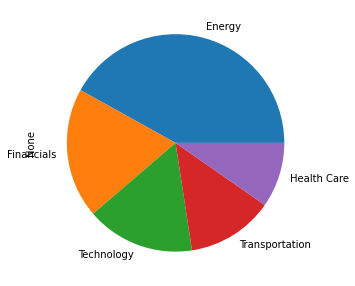
* My interpretation of the above data is that there seems to be the highest degree of correlation out of the 3 variables between Profits and Market Value, and the least between Assets and Market Value.

### US Sample: Profitability and Market Value 2021

# 

* Although there are some companies running losses, they do mostly appear to be on the lower end of the Market Value axis and there does appear to still be a correlation between the 2 variables.

### US Sectors with the highest number of companies who saw losses in 2021; Forbes Sample

 Text

Description automatically generated with medium confidence

* Interestingly, Energy saw the highest amount of companies reporting losses in this sample, which may be unsurprising given how increasingly subsidised this sector is.

# Insights

1. The initial shockwaves of the Covid19 pandemic on Markets from 2019 to 2020 seems to have eased in 2021, with the Top 5 performing Countries all seeing growth in combined market value from 2018 to 2021 despite most seeing drops from 2019 to 2020.
2. The exception to the above is the UK, which may also be also be impacted by on-going market unsteadiness arising from Brexit negotiations.
3. While the US has largely dominated the Top 500 Companies picked by Forbes in the past 4 years, in both the number of companies represented and their combined market value, China is a fast growing market that investors should be considering.
4. The Forbes list does seem to be representative of a region’s overall economy, with overlap / correlation shown between nominal GDP in 2021 and the representation from a company in the Forbes sample.
5. Deeper analysis into the US market shows Technology to be a reliable sector for investment, with Top 10 representation across all variables: Average Sales, Average Profits, Average Assets and Average Productivity.
6. Profitability should be a key consideration when analysing companies to invest in, as it seems to correlate most strongly with Market Value of a company.

# Further Analysis with Machine Learning

If I wanted to build my analysis further, I could use regression analysis to predict the Top 5 performing Countries that will appear in the Forbes Global 2000 list in 2022. Using time series forecasting, I could analyse the past performance of all regions across Sales, Profits, Assets and Market Value from 2018 to 2021 to predict how each region will perform in 2022 across these variables.

References

* <https://www.forbes.com/lists/global2000/#5f256b585ac0>
* <https://statisticstimes.com/economy/projected-world-gdp-ranking.php>