Forbes

Fortune 500 Companies

What Factors Contribute to Ranking the Fortune 500 Companies?

Team Members: Anjuman Alam, Avani Patel, Hiril Raval

Course: MBC 638 – Data Analysis and Decision Making

Descriptive Analysis

We have taken 16 variables, out of which 8 variables are quantitative and the rest are qualitative.

Objective

Every year Forbes creates a list of top 500, 1000 largest American companies and rank them based on various factors. We have taken this year's top 500 list for our data analysis. Together, the 500 corporations on this year's list generated $13.8 trillion in revenue, or some two-thirds of the U.S. economy.

The objective here to analyze what were the factors that contributed to the ranking of these Fortune 500 Companies.

The data was collected online from Kaggle website. https://www.kaggle.com/winston56/fortune-500-data-2021

Population of Interest

The dataset is not a random sample. It is a population dataset and does not represent population.

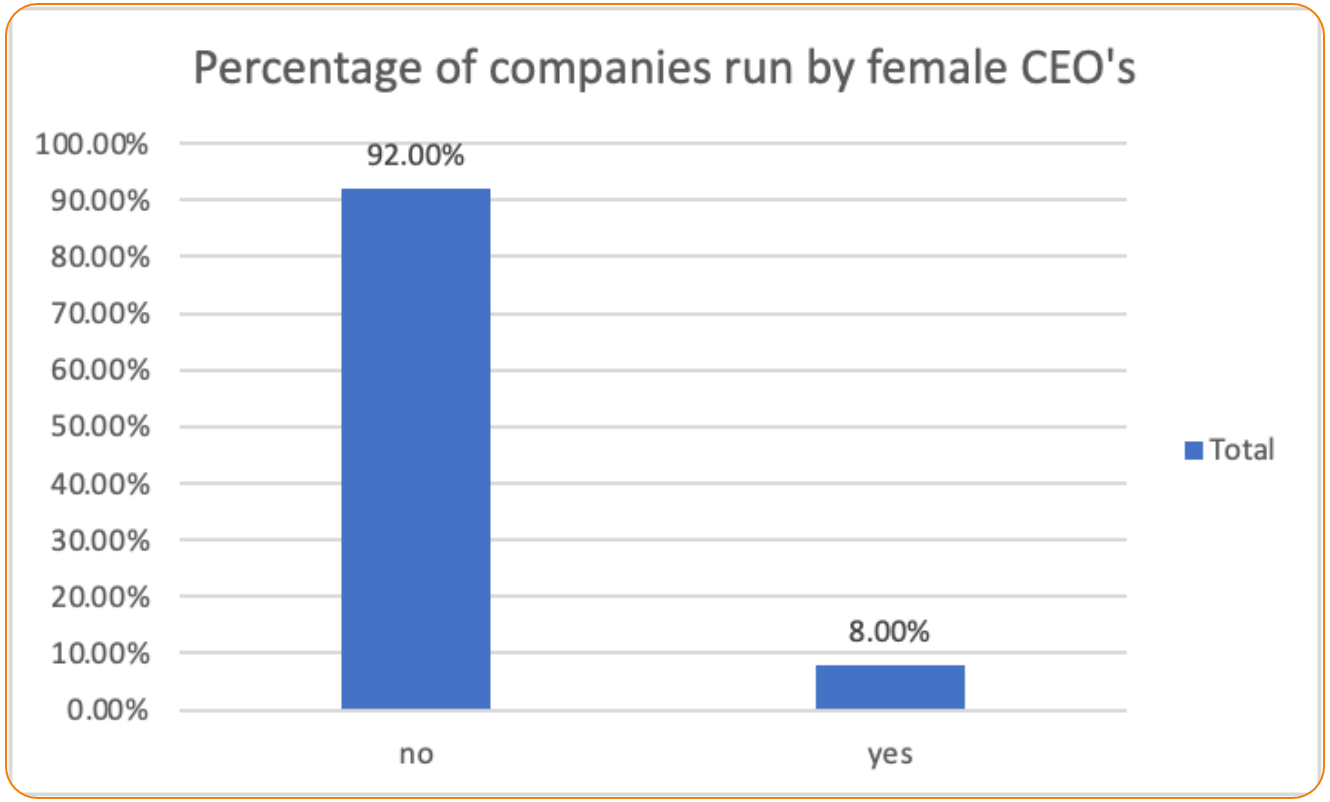
Data Collection from Kaggle

Variable Name	Description	Units of Measurement
Company	The name of the company	
Rank Change	The change in the rank from 2020 to 2021. There is only a rank change listed if the company is currently in the top 500 and was previously in the top 500.	
Revenue	Revenue of each company which was used to rank each company.	Millions
Profit	Profit of each company	Millions
Num. of Employees	The number of employees each company employs.	
Sector	The sector of the market the company is operating in.	
City	The city in which it has its headquarters	
Assets	Total worth of assets owned by the company	Millions

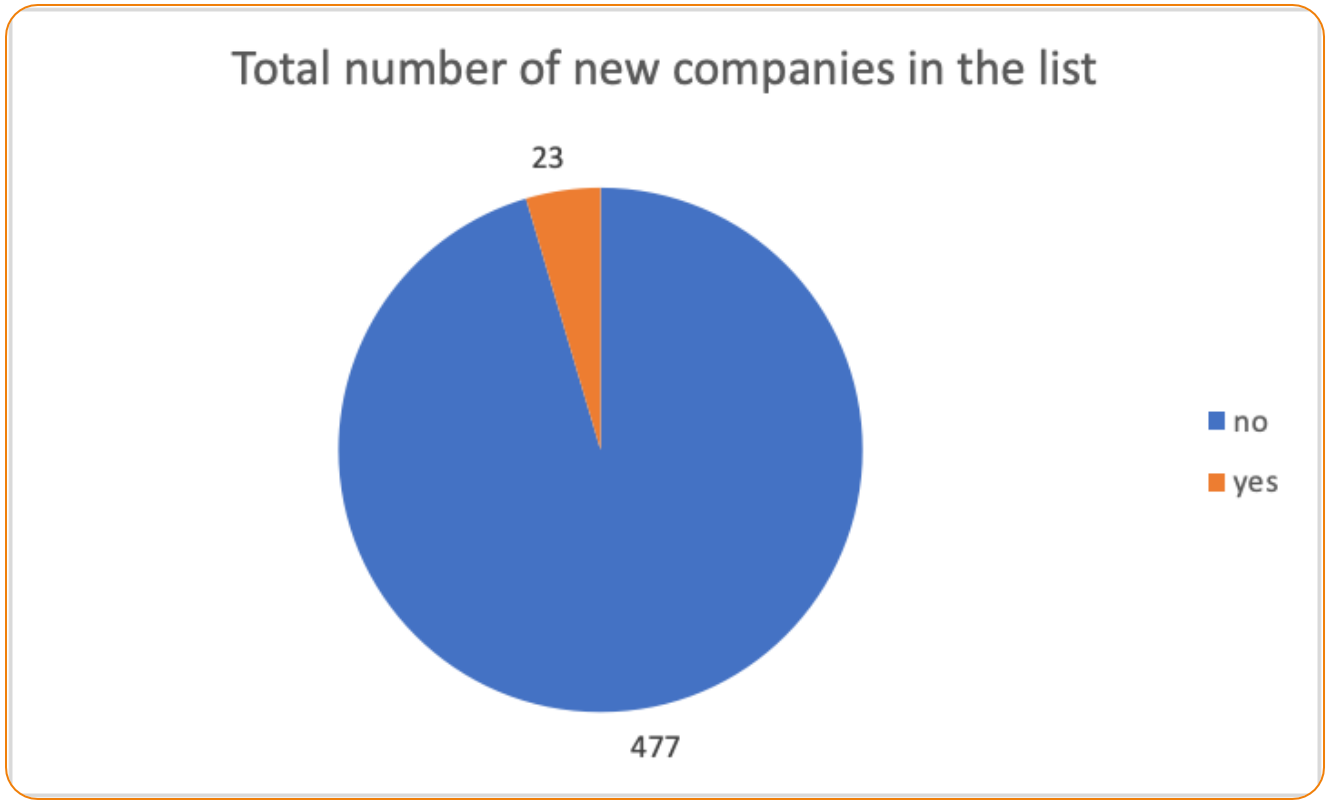
Data Collection from Kaggle

Variable Name	Description	Units of Measurement
CEO Woman	Indicates whether the CEO of the company is a woman ("yes" or "no").	
Profitable	Indicates whether the company is profitable or not ("yes" or "no").	
Prev. Rank	The 2020 rank of the company, as established by Fortune if it is not a newcomer.	
Market Cap	The market cap (or value) as per January 20, 2021.	Millions
Rank	The 2021 rank established	
CEO Founder	Indicates whether the CEO of the company is also the founder ("yes" or "no").	
State	The state where the company's headquarters is located	
Newcomer	Indicates if the company is new to the top Fortune 500 ("yes" or "no"). No value will be listed for companies outside of the top 500.	

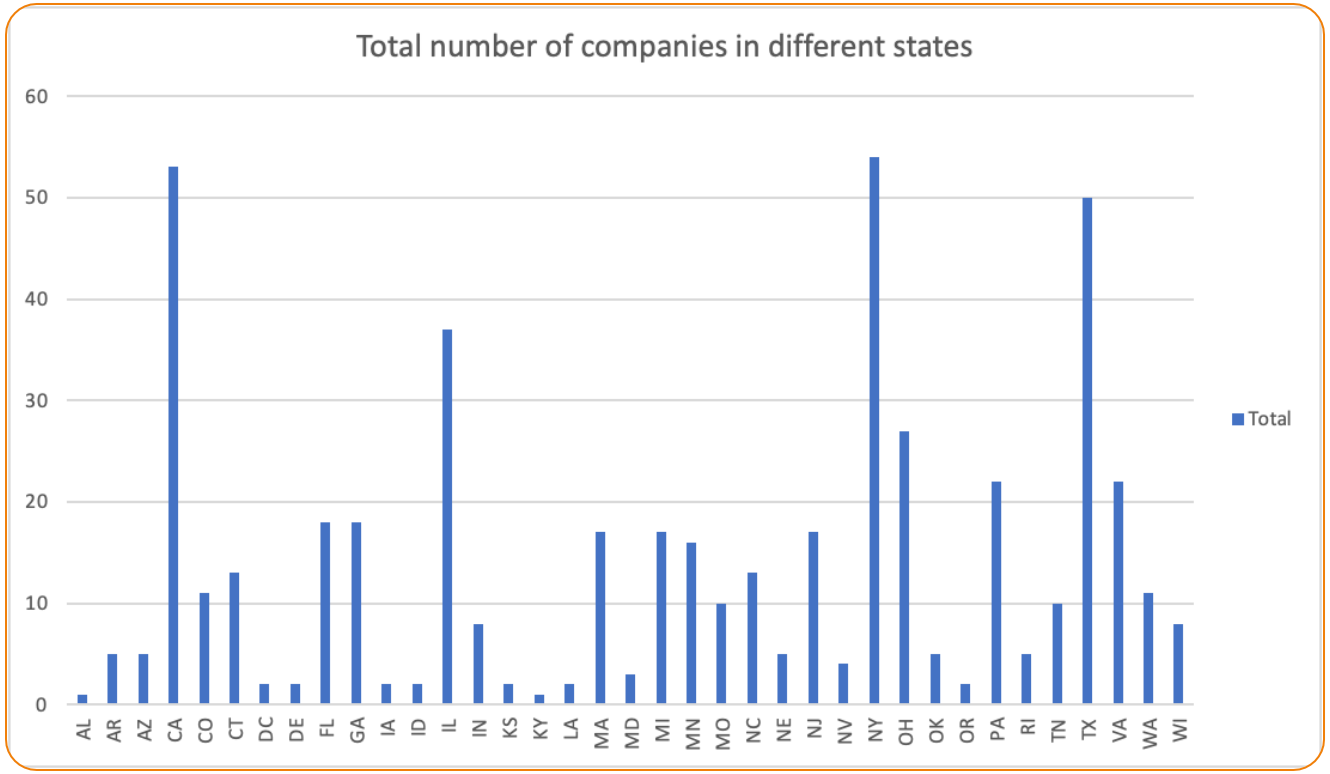
Percentage of companies run by Females CEO's



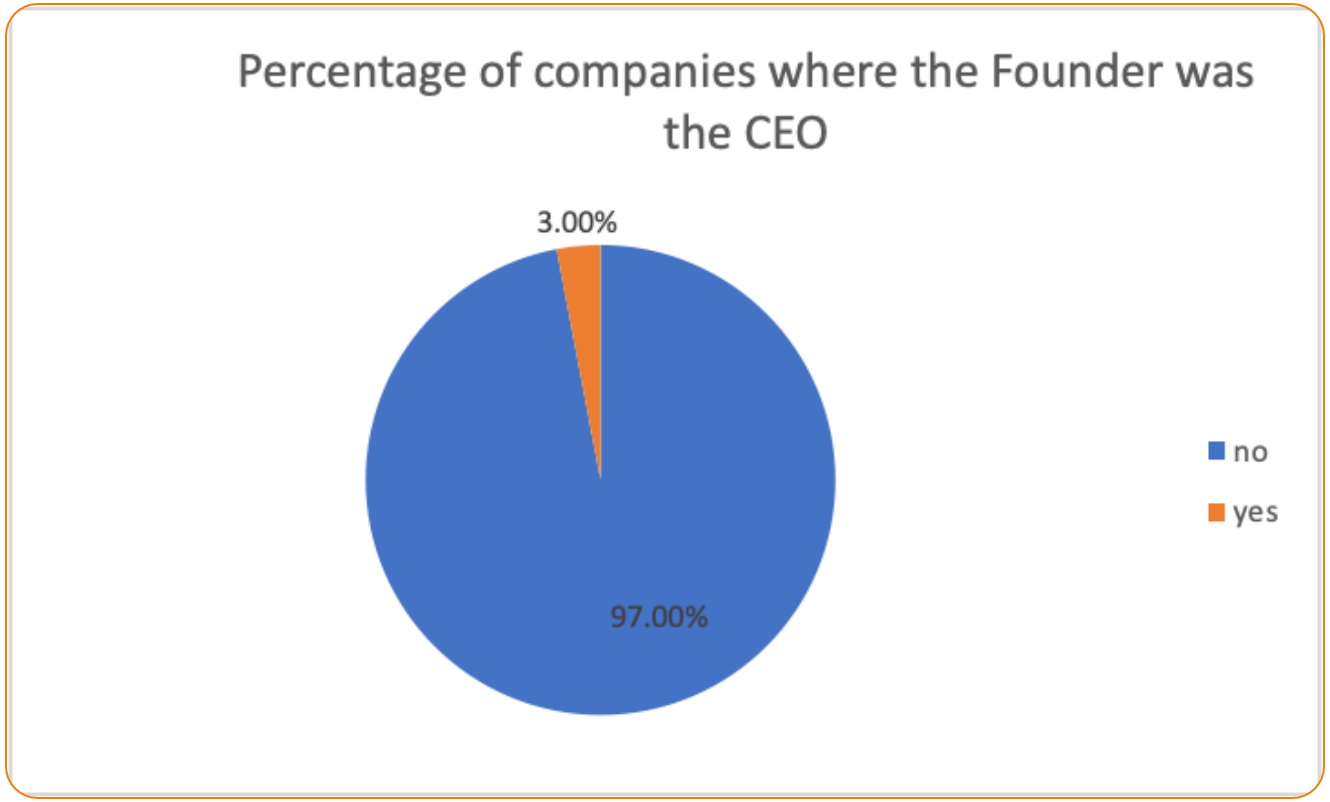
Total number of new companies in the list



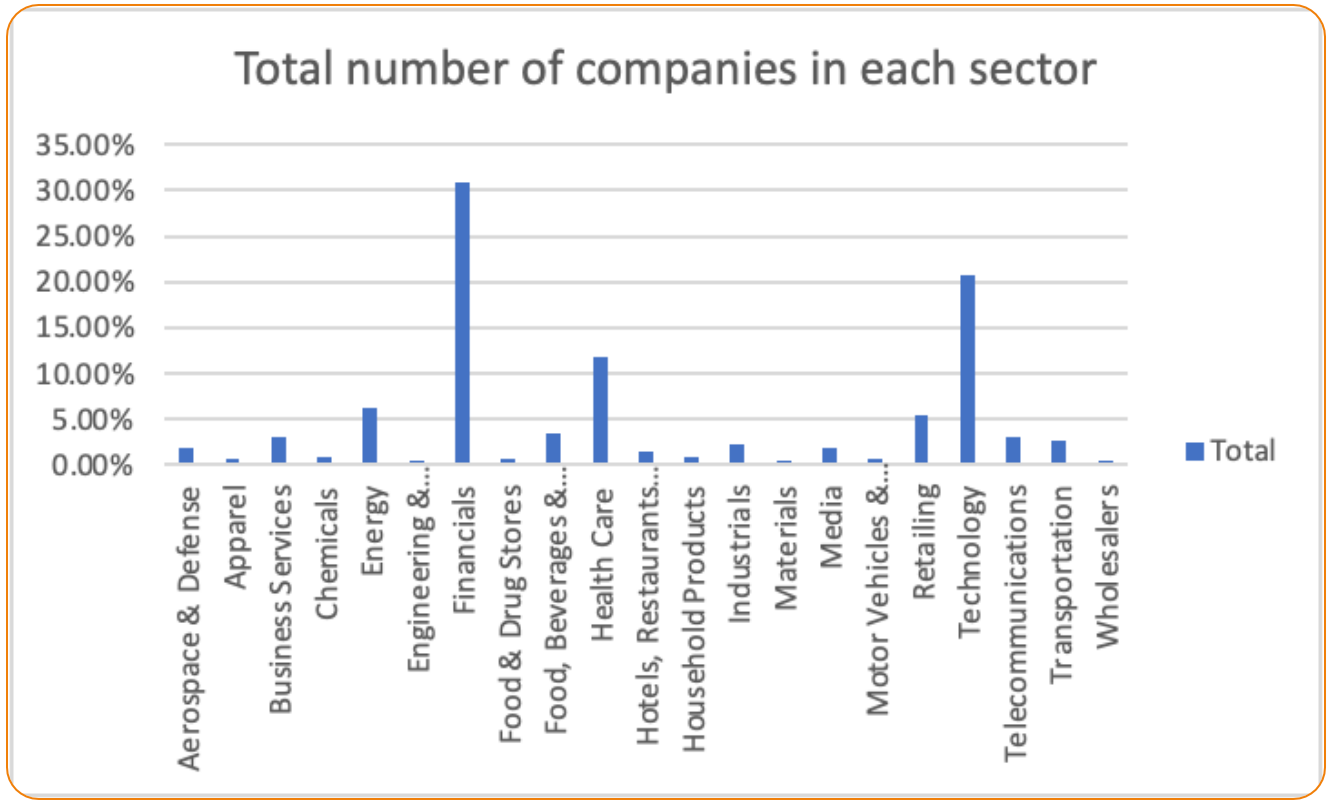
Total number of companies in different states



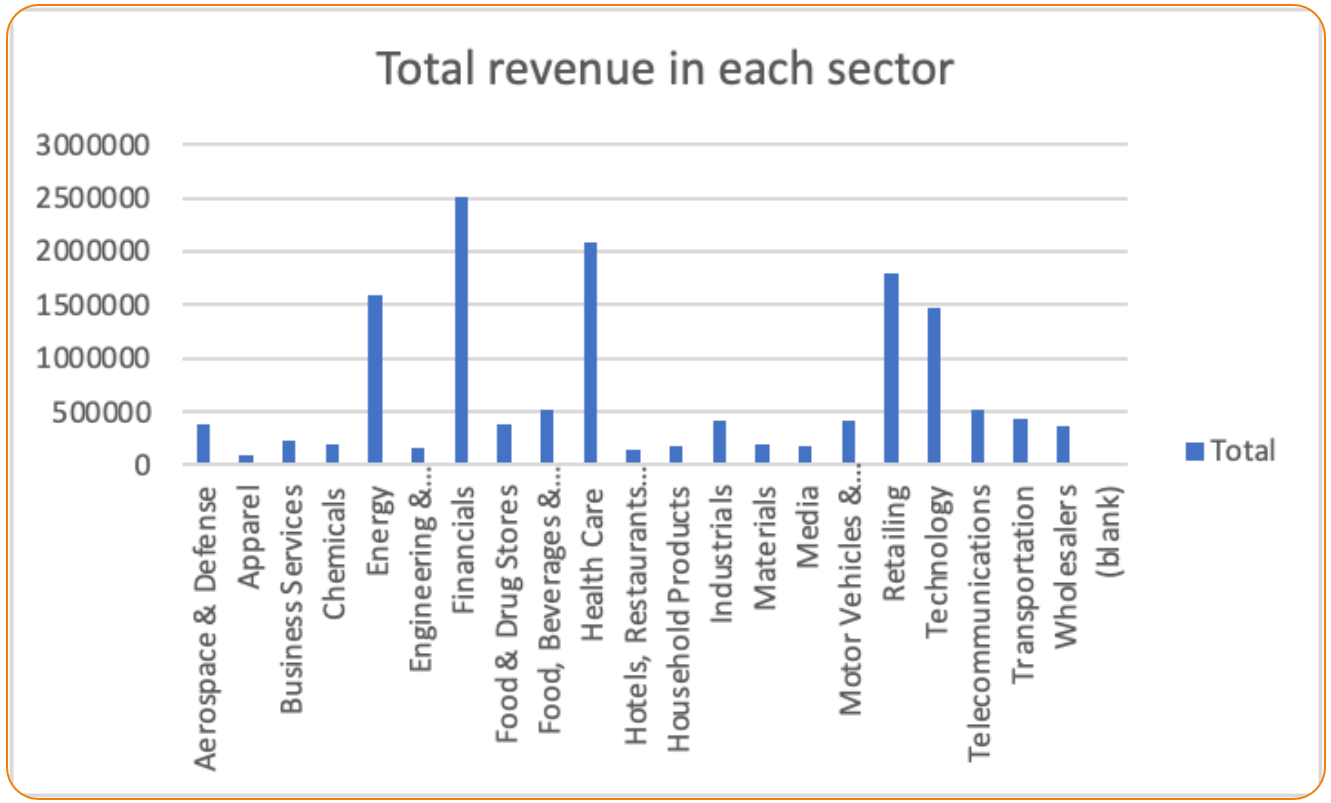
Percentage of companies where the Founder was the CEO



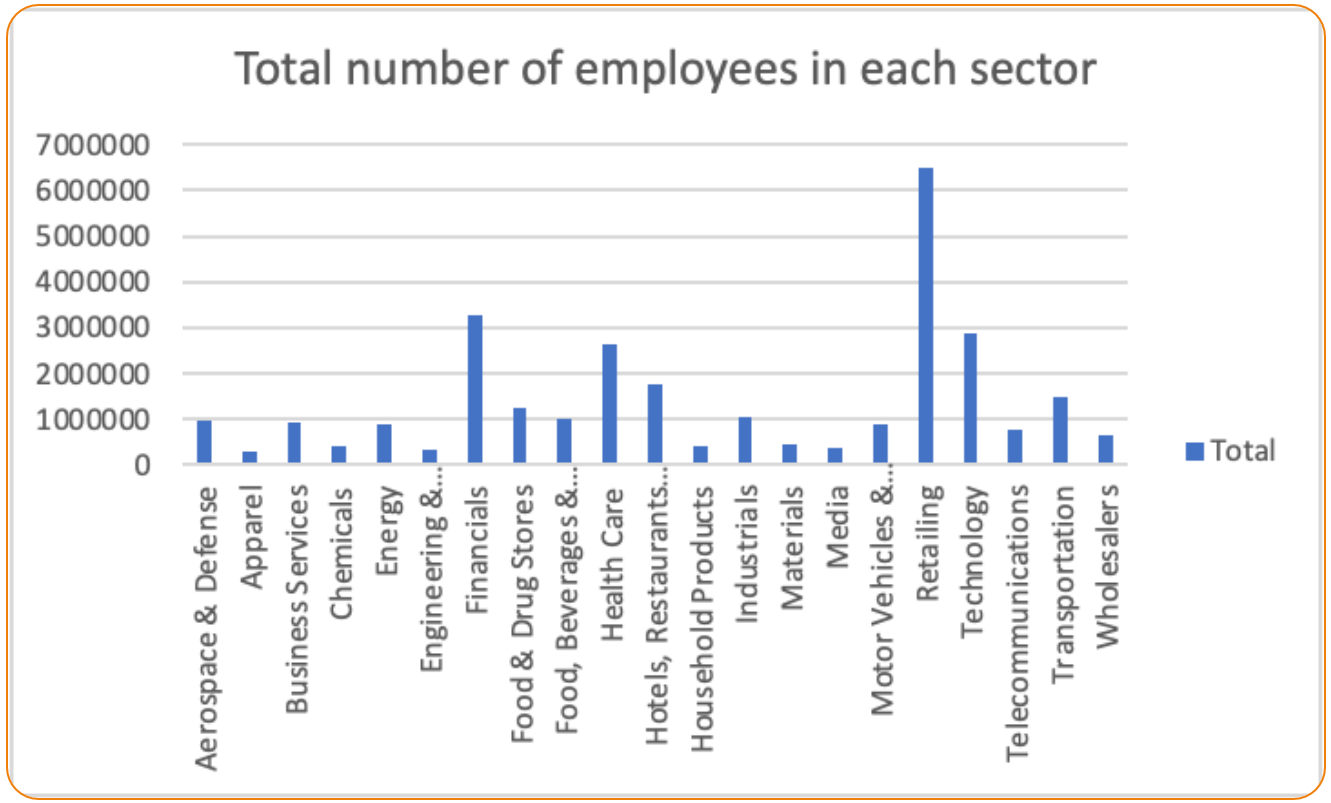
Total profit in percentage in different sectors



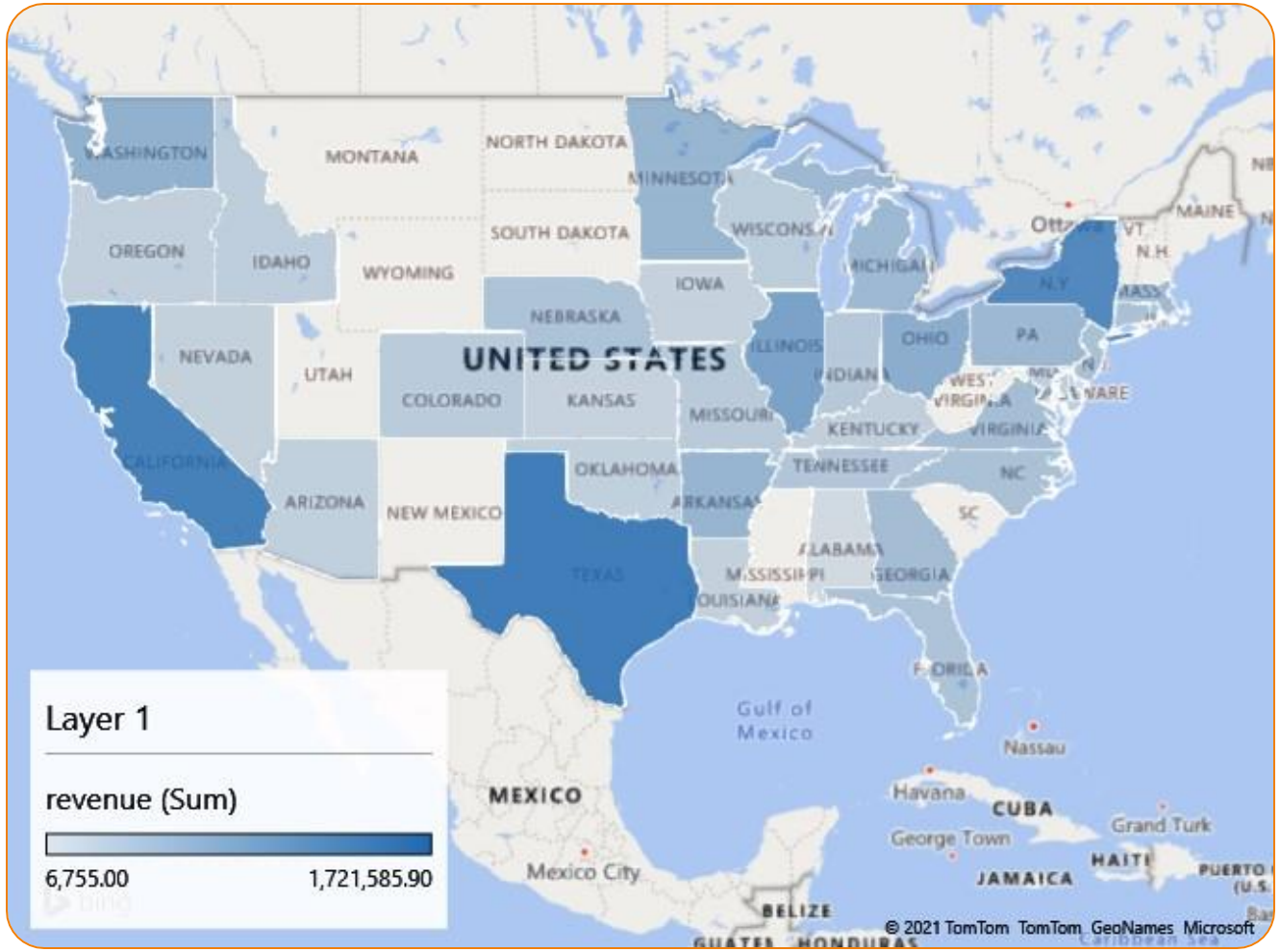
Total revenue in different sectors



Number of employees in different sectors

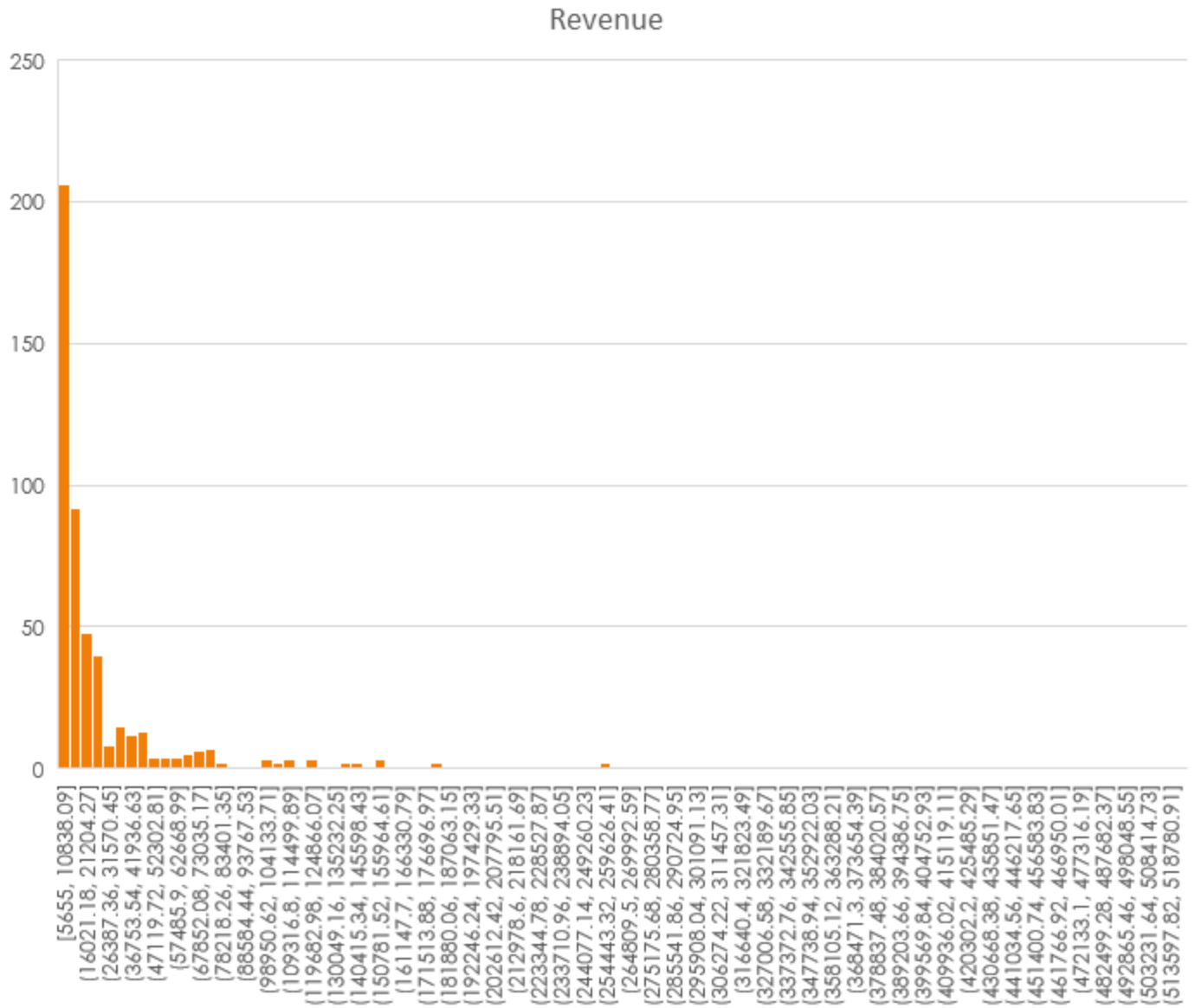


Revenue by state



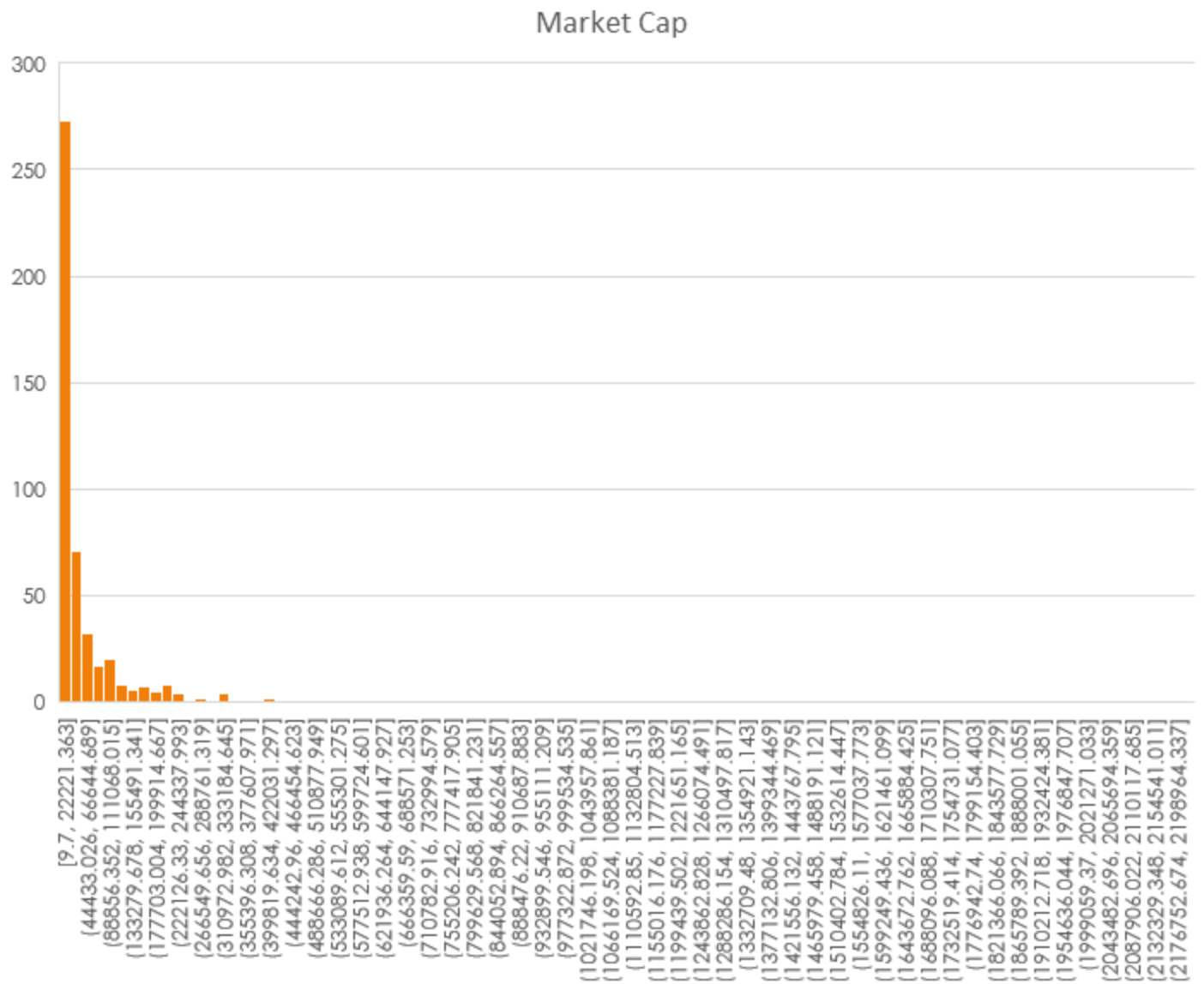
Distribution of revenue data

- Since mean>median, the chart for revenue is right skewed.
- There are outliers found, which is expected.
- The minimum amount is 5655 and the maximum is 523964. The mean was found to be 28420.2092 and the median is 12831.85.



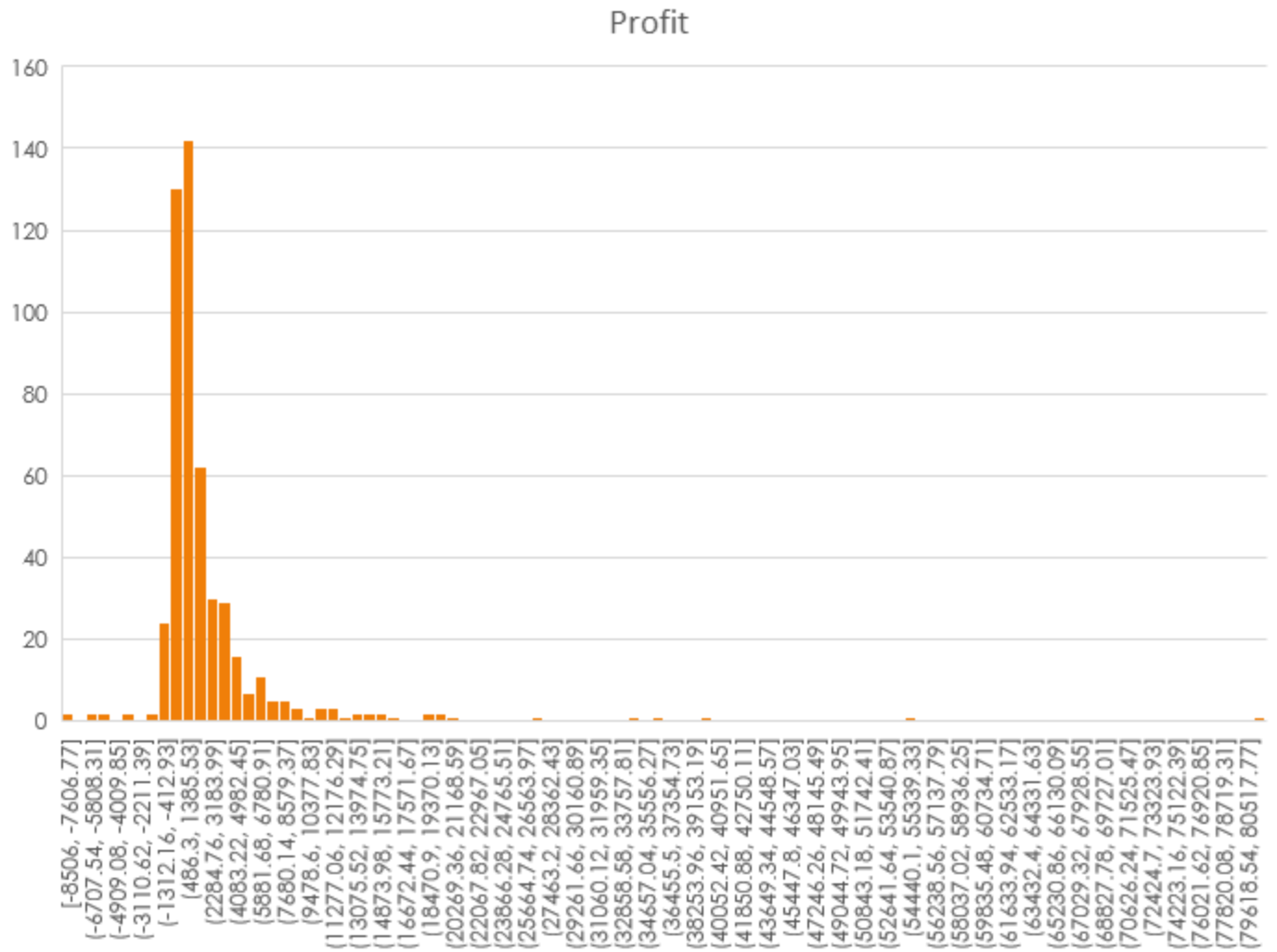
Distribution of Market Cap data

- Since mean>median, the chart for revenue is right skewed.
- There are outliers.
- The minimum amount is 9.7 and the maximum is 2221176. The mean was found to be 62225.59 and the median is 15920.



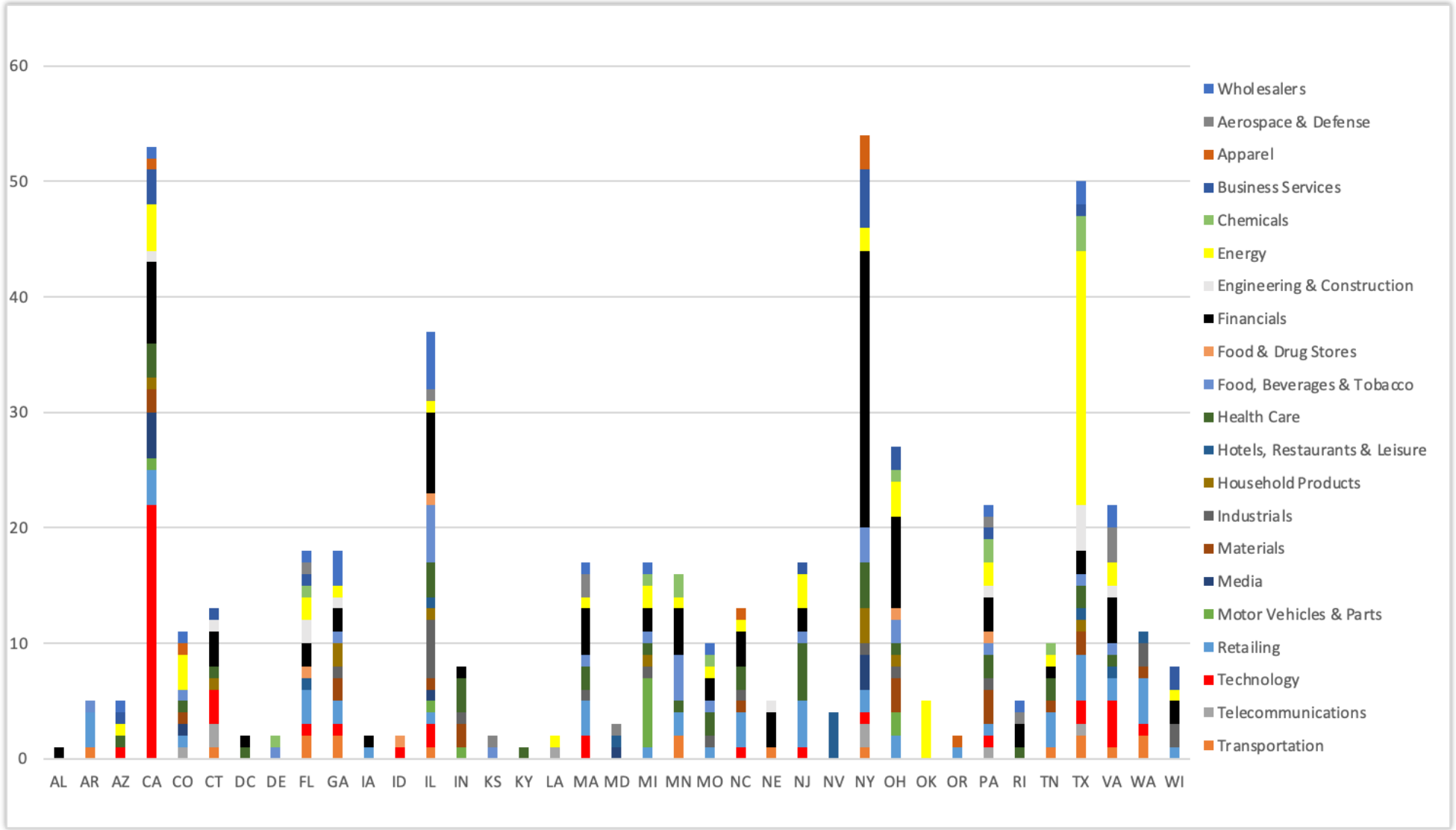
Distribution of Profit data

- Since mean>median, the chart for revenue is right skewed.
- There are outliers.
- The values ranged from negative -8506 to 81417. The mean was found to be 2445.367 and the median is 889.5.



Overall Findings of Descriptive Analysis

- Overall, we found that since the highest number of Finance companies were in New York and the highest amount of revenue generated was by finance sector, New York showed the highest amount of revenue generation.
- Moreover, the second highest number of companies were in the technology sector which were mostly situated in California, and since technology sector was in the top 5 for the revenue generation; California also had high revenue generation.
- Interestingly, Texas had high revenue income as it had the highest number of energy sector companies. And since the Energy sector ranked 4th in terms of revenue generation.


60
50
40
30
20
10
0
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY LA MA MD MI MN MO NC NE NJ NV NY OH OK OR PA RI TN TX VA WA WI
Wholesalers
Aerospace & Defense
Apparel
Business Services
Chemicals
Energy
Engineering & Construction
Financials
Food & Drug Stores
Food, Beverages & Tobacco
Health Care
Hotels, Restaurants & Leisure
Household Products
Industrials
Materials
Media
Motor Vehicles & Parts
Retailing
Technology
Telecommunications
Transportation

Analysis of Associations

Part B: Correlation and Regression Analysis

Objective

- Understand what factors affect the ranking of the Fortune 500 list the most.
- Conduct correlation analysis to understand the correlation between the dependent variables and the explanatory variables.
- Figure out which explanatory variables contribute to rank the most.
- Forecast the revenue of companies based on explanatory variables.

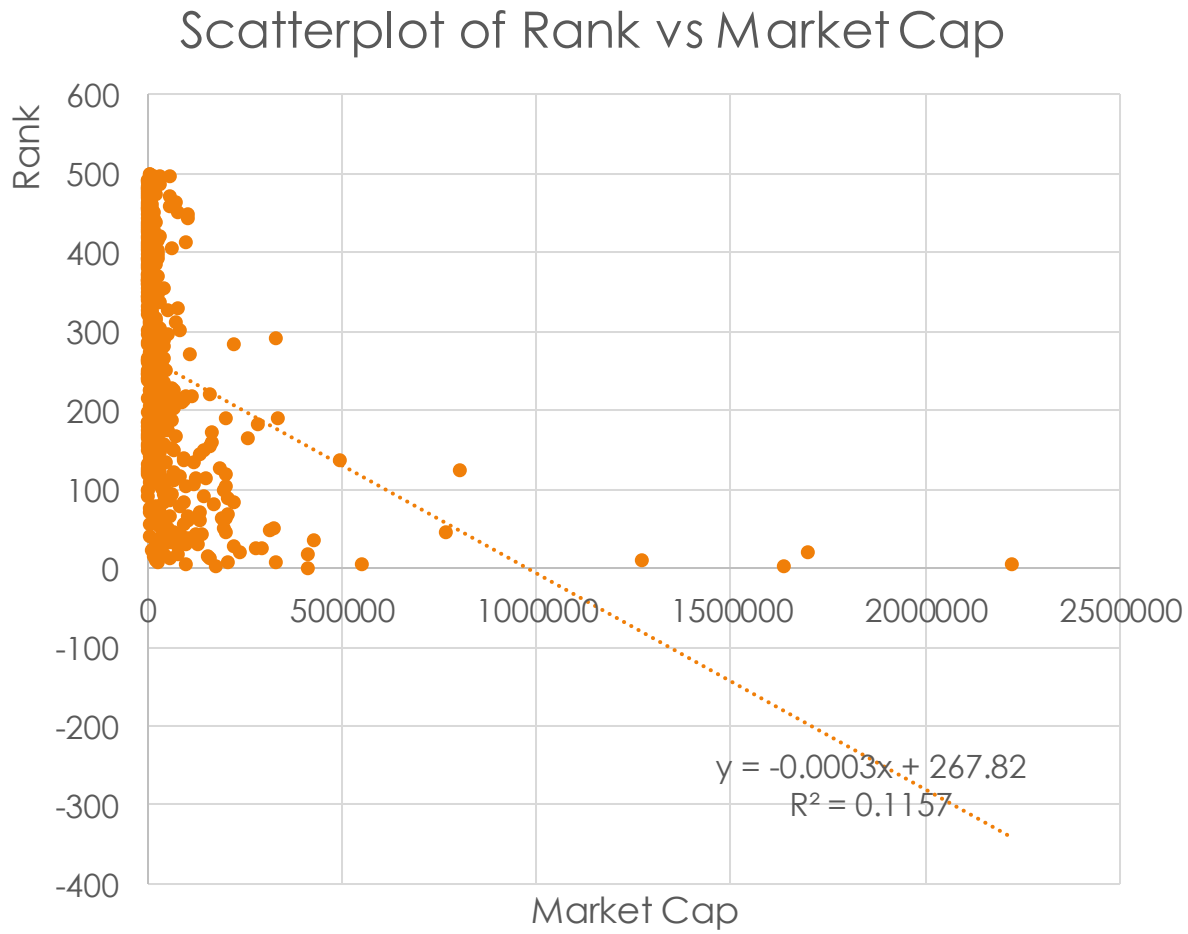
Variables

Y (Dependent Variables)	X (Explanatory Variables)
Rank	• Market Cap • Profit • Revenue • Number of Employees • Sector
Revenue	• Market Cap • Profit • Number of employees • Sector • Assets

Expectations and Predictions

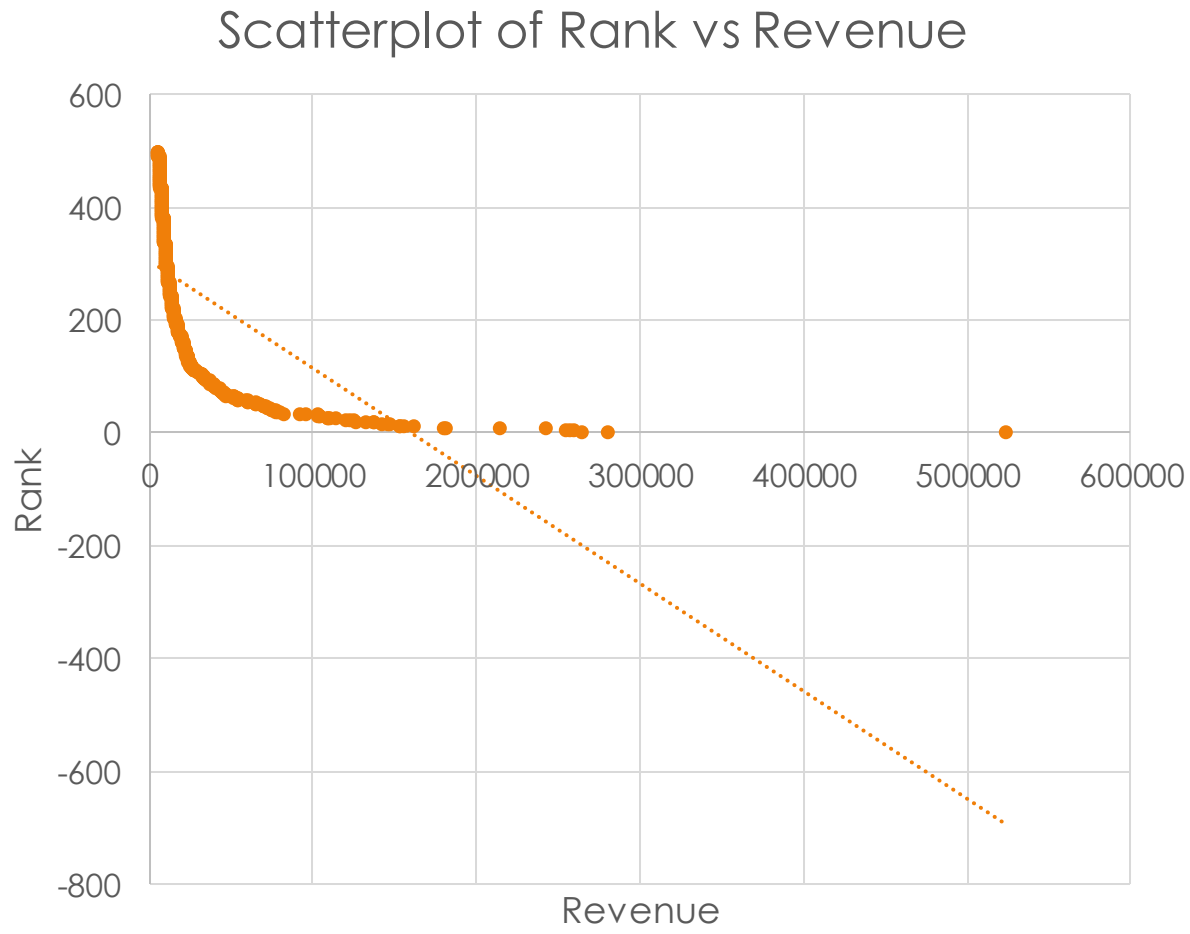
- We expect ranking to be positively correlated with profit, market cap, and revenue.
- We expect revenue to be positively correlated with profit, rank, and market cap, but have no correlation with number of employees.
- We expected for regression that rank would be depended on profits, revenue, market cap, assets, CEO founder.
- But we found that rank has no relation with profits because negative profits (loss making companies) also had high ranks.

Rank vs Market Cap



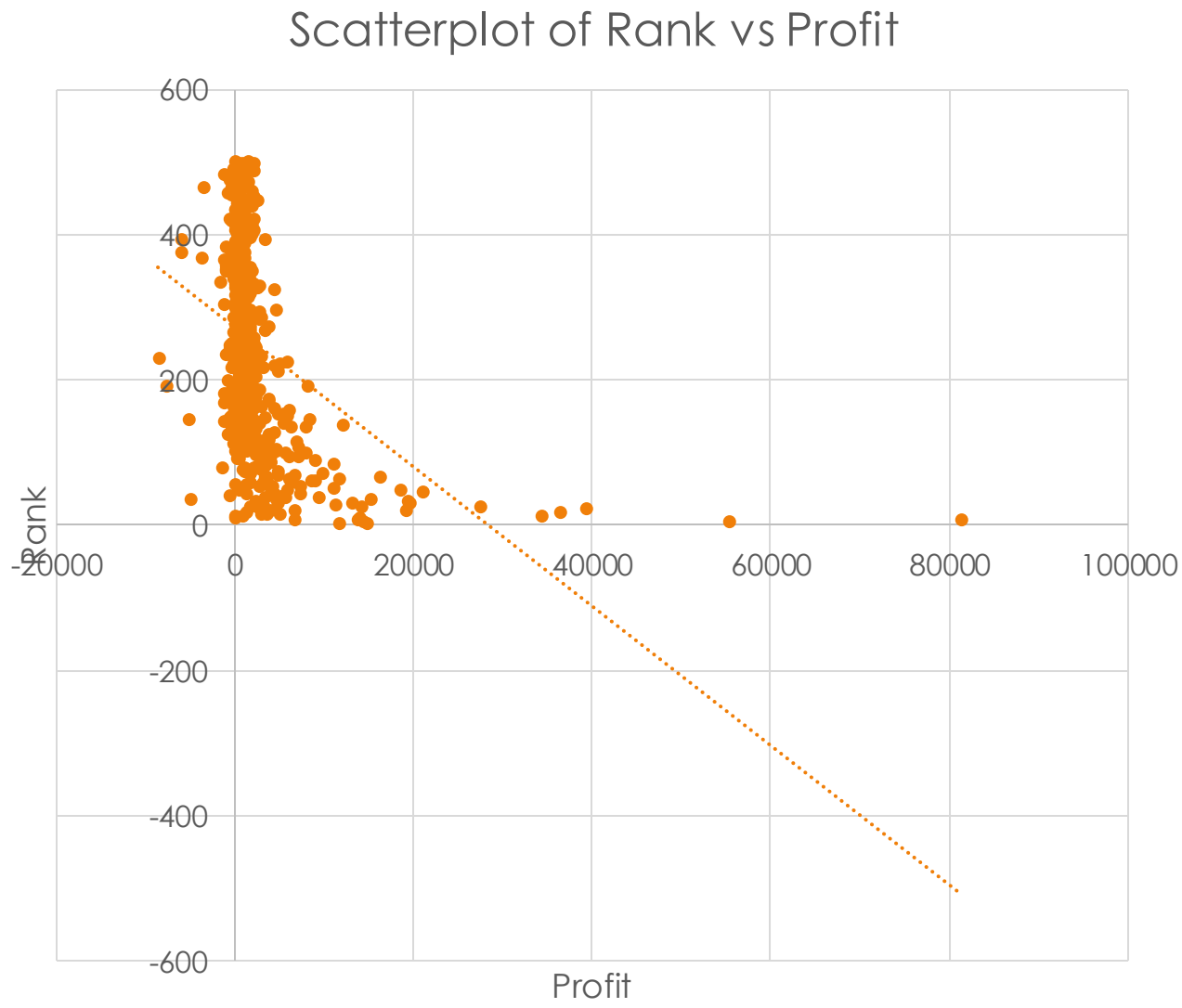
Correlation -0.340

Rank vs Revenue



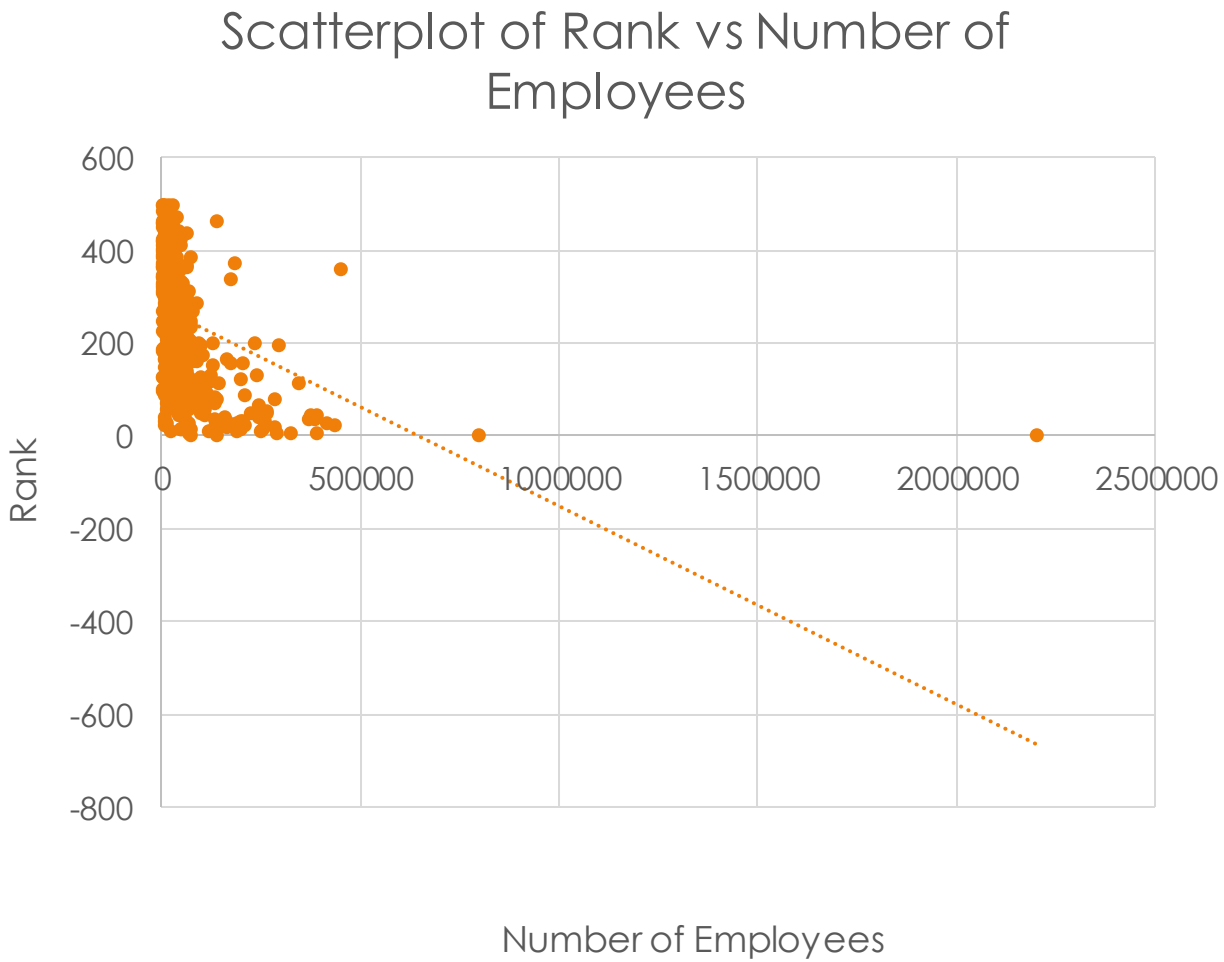
Correlation -0.607

Rank vs Profit



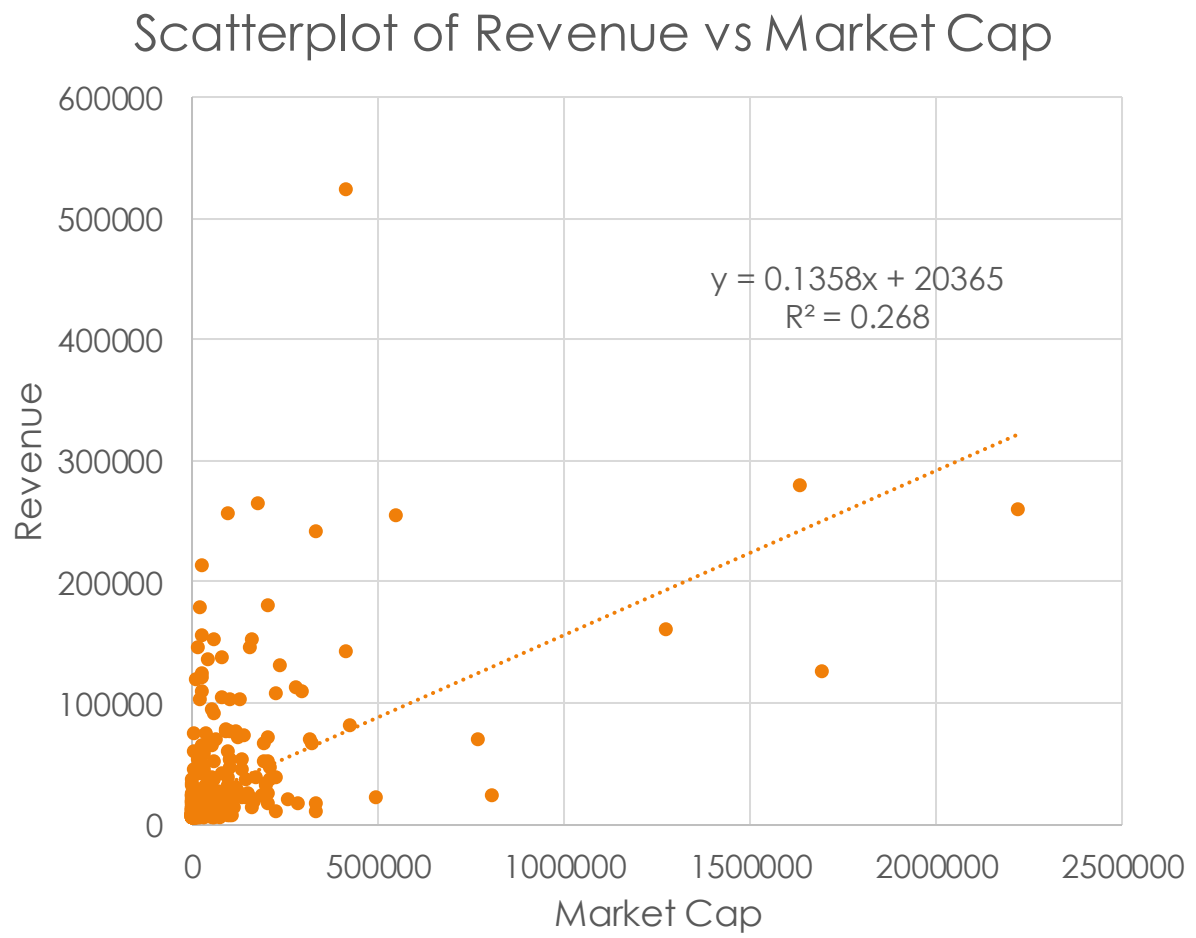
Correlation -0.411

Rank vs Number of Employees



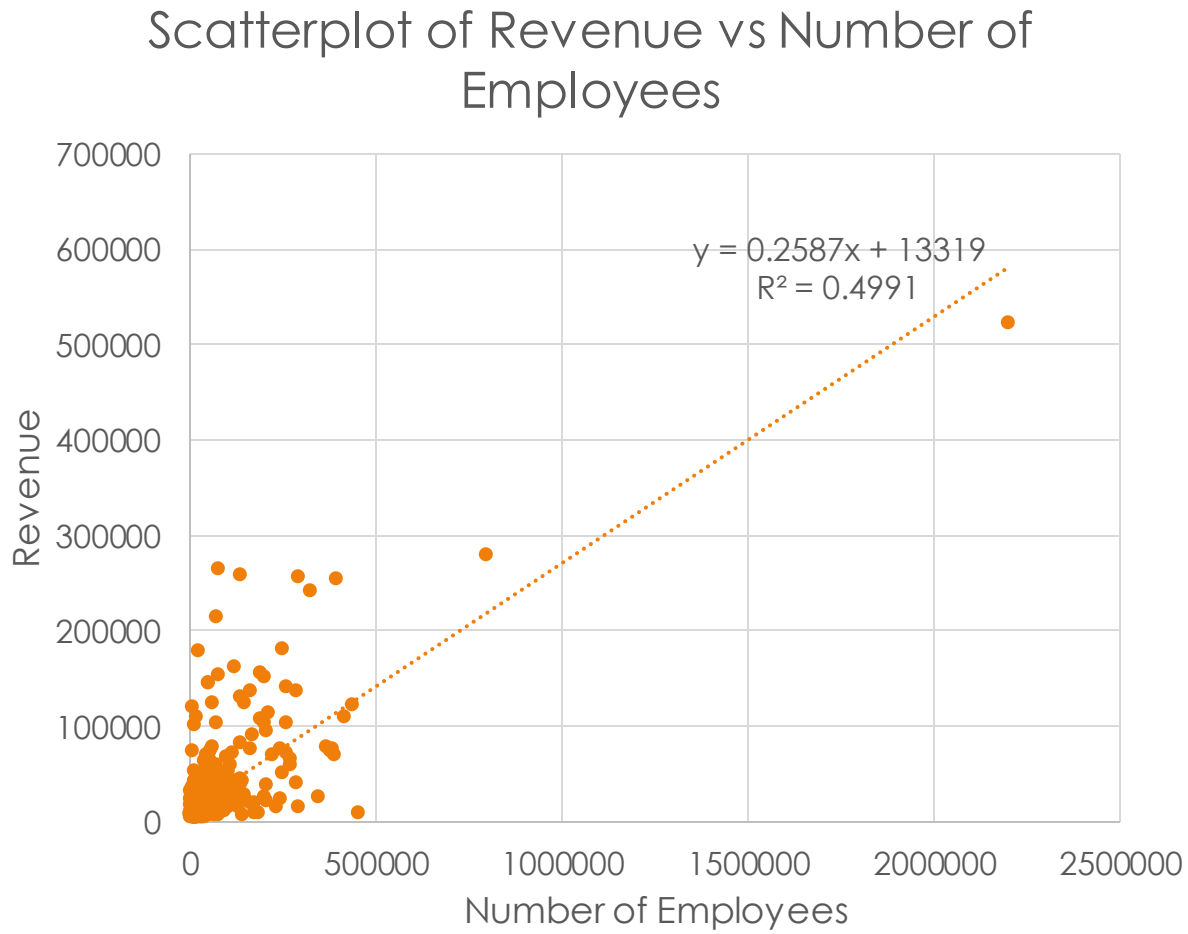
Correlation -0.370

Revenue vs Market Cap



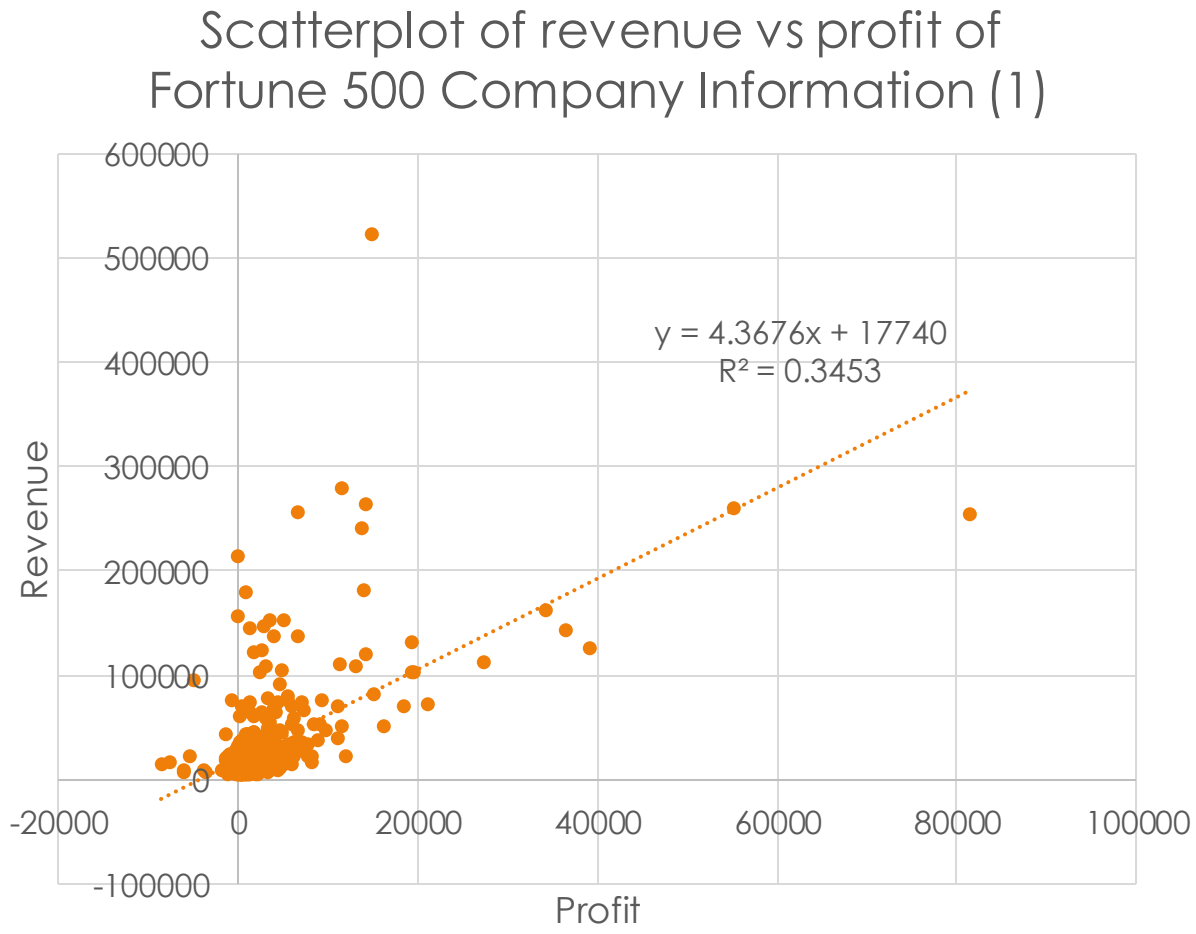
Correlation 0.518

Revenue vs Number of Employees



Correlation 0.706

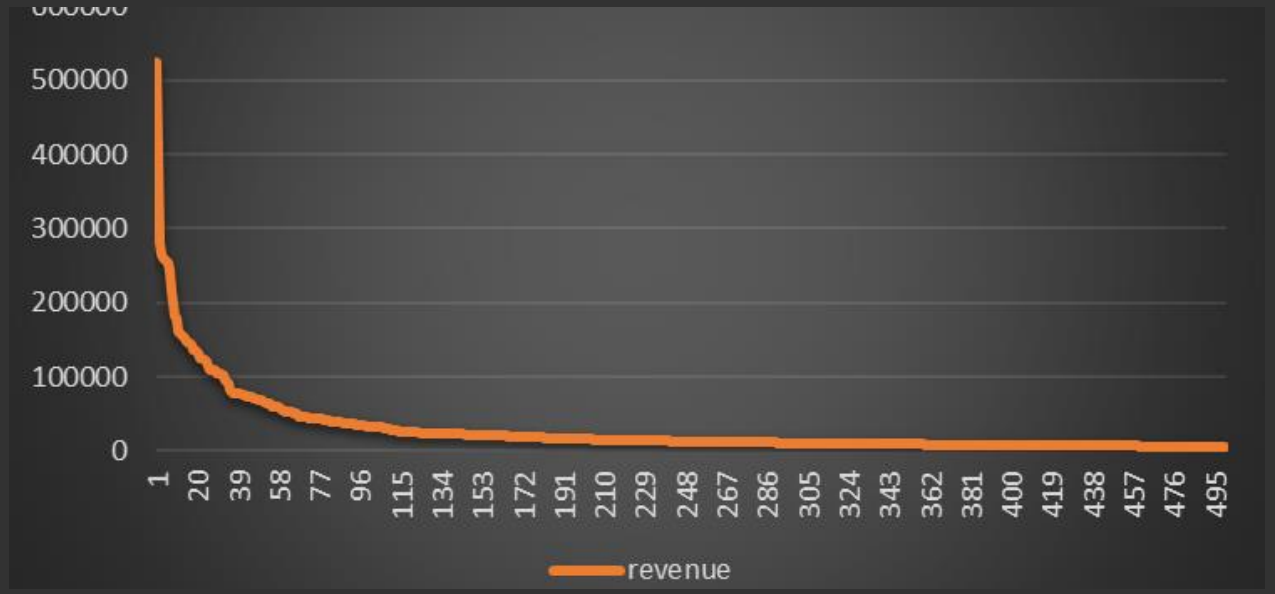
Revenue vs Profit

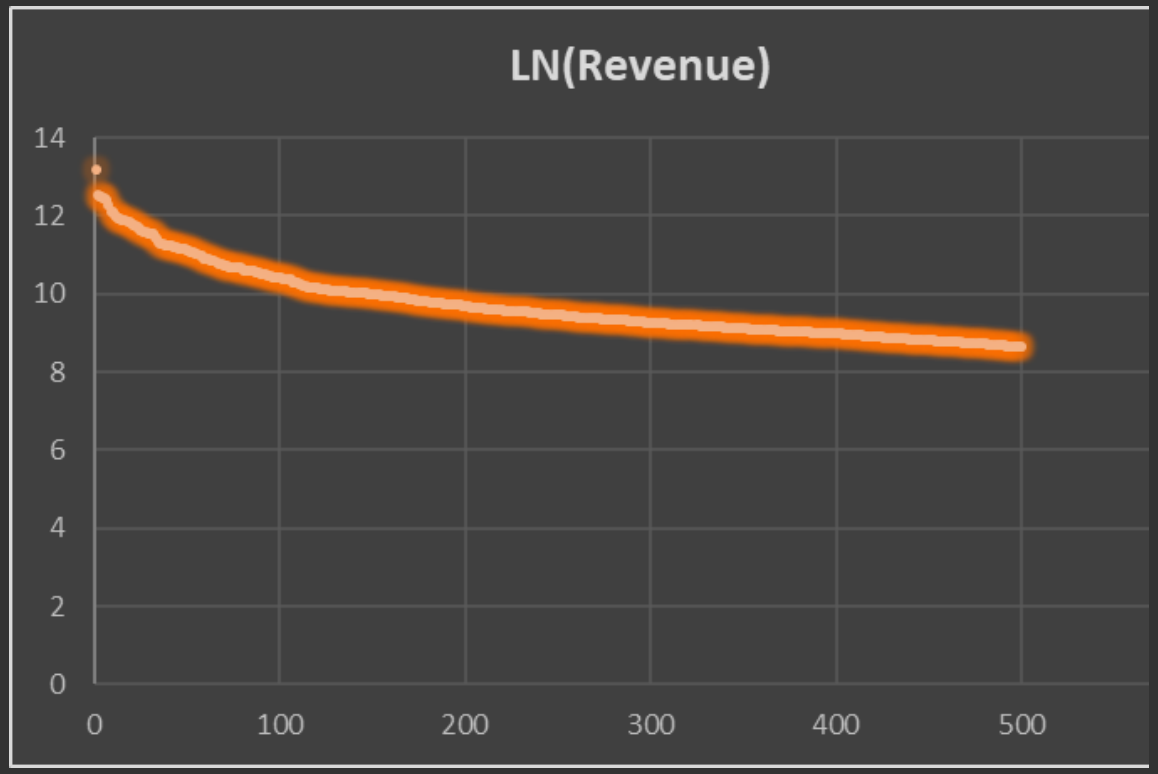


Correlation 0.588

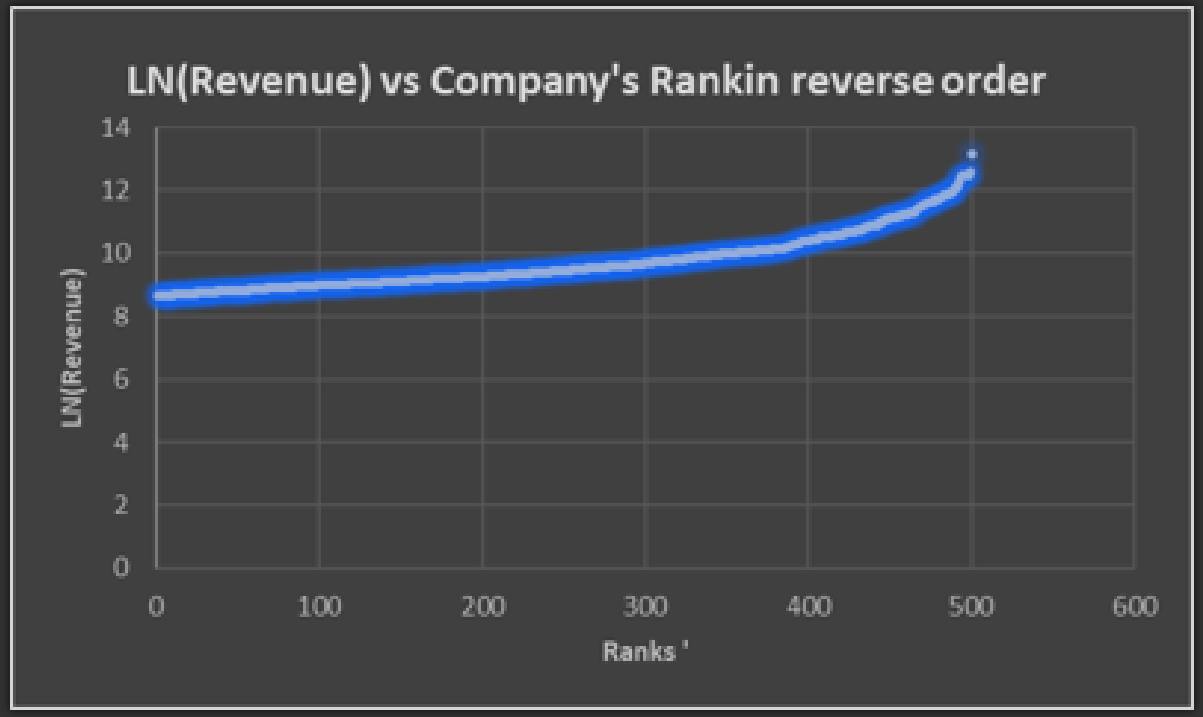
Regression models and Forecasts

Rank vs Revenue plots



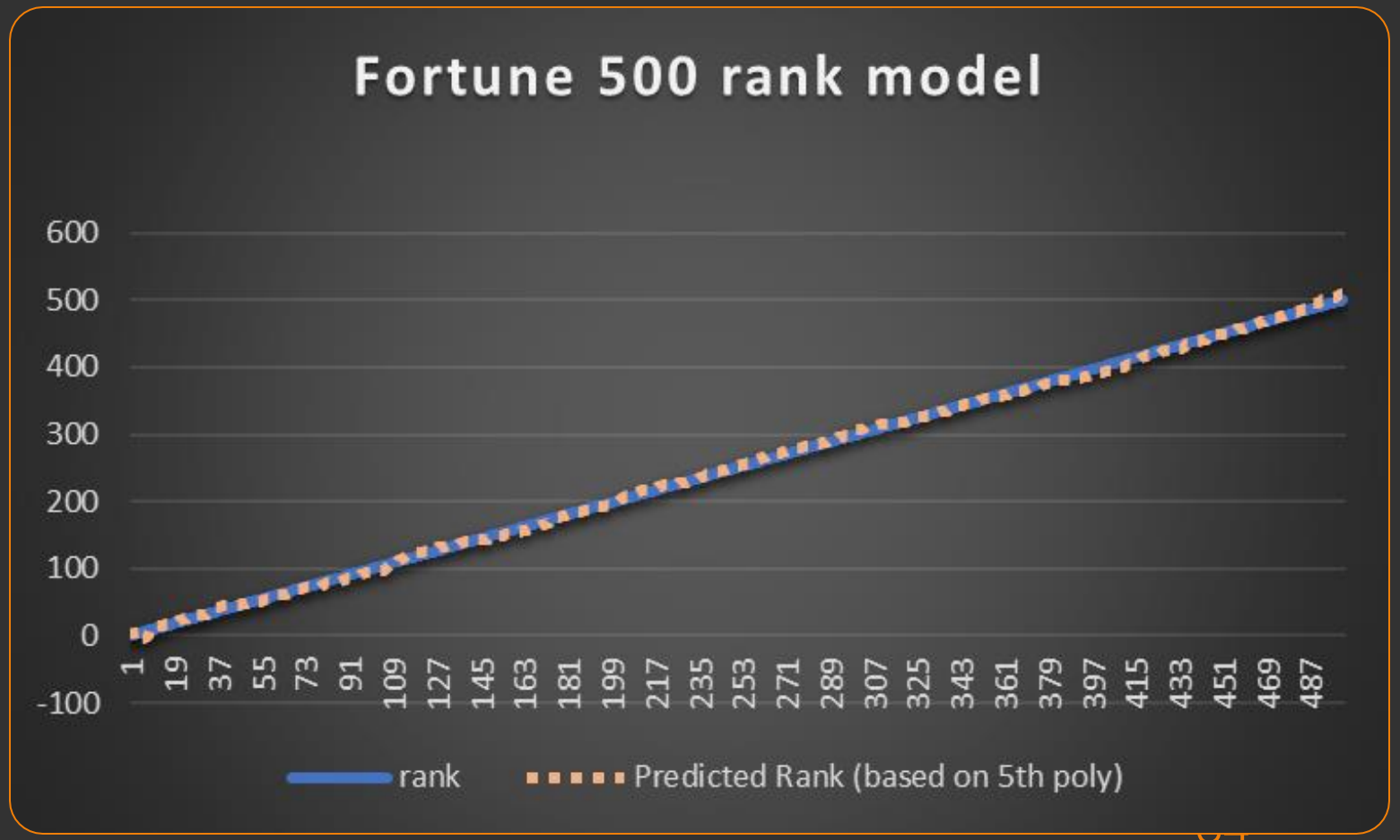


Reversing the graph for simplicity

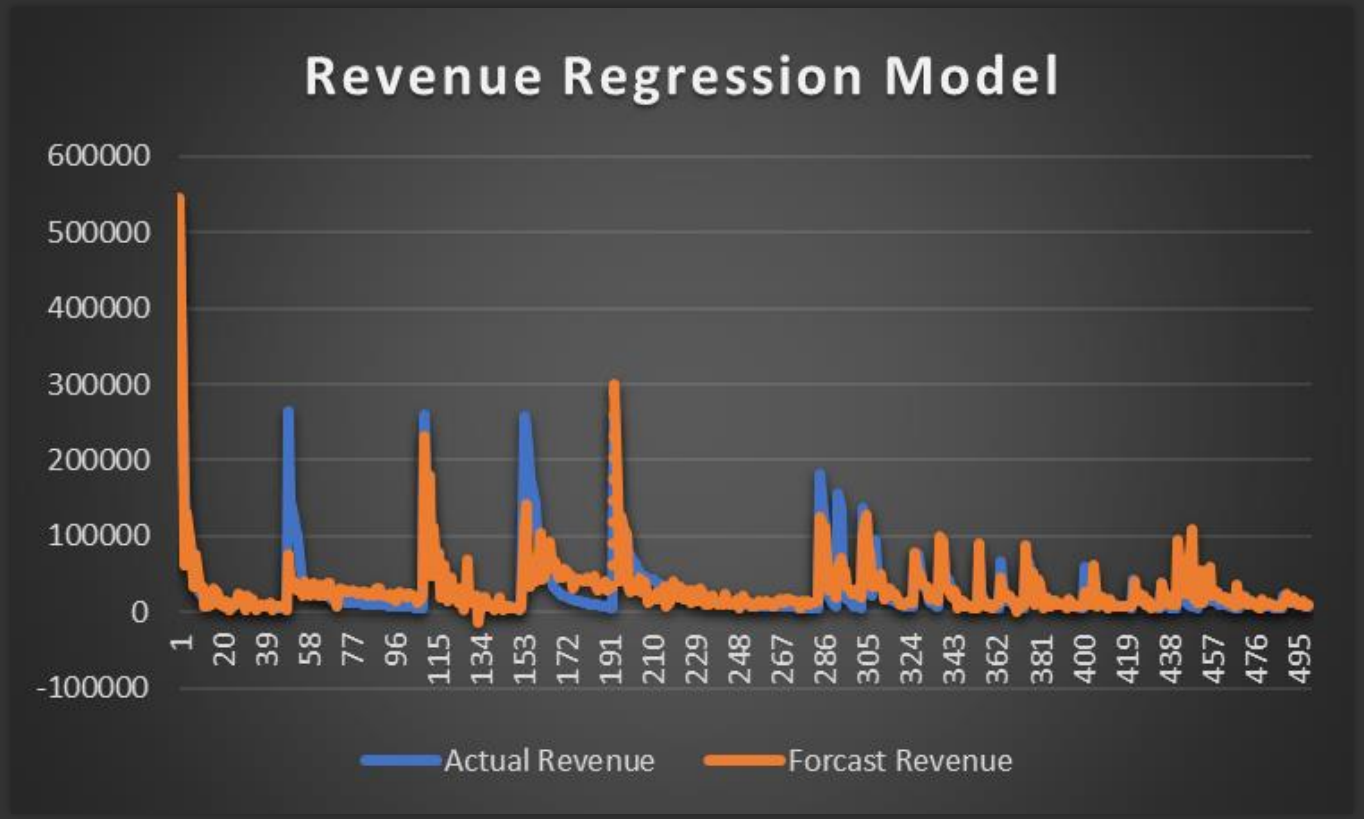


Rank Regression Model
Adj R square = 0.9992

- Ranks are most likely decided on the Total Revenue of company theatrically.
- We found that The P values of other factors, categorical or numeric were statistically unsignificant.
- Practically revenue and profit were two factors with most statistical significance found but profit couldn't be added because of other limitations.



Revenue Regression model (Adj R square = 0.6984)



Variables with most statistical significance in this model

-Profit
-Market Cap
-Number of Employees
-Sector (as categorical)
-Assets

Overall Findings of Association Analysis

- Using logarithmic scale on revenue and polynomial regression of 5th degree, we could get a forecast model which could predict ranks of fortune 500 company based on its revenue.
- In real world, rank of fortune 500 companies are decided based on total revenues. Which we could prove in our model as rest all the variables were statistically insignificant in terms of p-values.
- To understand what factors revenue is dependent on, we used a backward regression and concluded the number of employees, profits, assets, market cap and sectors where the deciding factors of revenue.
- We found that for regression model p-value for revenues where low that means high relationship, but when we checked correlation value it shows no relation. This was maybe because the relation was exponential and not linear.