
California State University, Los Angeles

CIS 5250 Visual Analytics

Instructor: Dr. Shilpa Balan

SAS Project

# Forbes Global 2000 – 2019

Prepared by: Vishav Patel

# INDEX

## 1. Dataset

Data Set link: https://data.world/aroissues/forbes-global-2000-2008-2019/workspace/file?filename=Forbes+Global+2000+-+2019.csv

# 2. Introduction

## 2.1 Objective

Specifically, a dataset on Forbes Global from 2000 to 2019 is the main subject to examine and show. It attempts to research the top 2000 worldwide corporations' profits, revenues, industry types, and related areas. My ability to determine which industries are profitable and which are not will be aided by this data. which business generates the most and least money. Furthermore, based on this information, it is possible to identify which companies were performing well and poorly at any given period. In addition, we can identify which industry is controlling the marketplace.

## 2.2 Aim of Analysis

This project aims to find multiple analysis with this dataset.

- Market value of the company
- Which sectors are making huge profits
- Which sectors are making less profits
- Types of Industry of the top companies
- Companies with highest asset holdings
- Difference in revenue of top companies

After the data cleaning in the excel this all analysis can be done in the SAS studio to generate the desired results and visualizations. Further, Tableau also can be used to generate additional visualizations.[1]

[1] Krasnove, B. (2019, June 30). *Global 500 2000*. Fortune. Retrieved November 2, 2022, from https://fortune.com/global500/2000/

Wikimedia Foundation. (2022, October 30). *Forbes Global 2000*. Wikipedia. Retrieved November 2, 2022, from https://en.wikipedia.org/wiki/Forbes_Global_2000

## 3. Data Description

| | |
|---|---|
| Company | Name of the company |
| Market Value | Market value of the company |
| Revenue | Revenue of the company |
| Profits | Profit made by the company |
| Assets | Assets of the company |
| Rank | Global rank of the company |
| Sector | Sector of the company |
| Industry | Focus of the Industry |
| Continent | Continent name |
| Country | Country located |
| Headquarters | Location of headquarters |
| State | Name of the state |
| CEO | Name of the CEO |
| Forbes Webpage | Source (URL) |
| Profits as % of Assets | Profit in terms of Assets |
| Profits as % of Revenue | Profit in terms of company revenue |

## 4. Data Cleaning

Data science requires the clean-up of existing data. Working with contaminated data may be quite challenging. And today we'll talk about the same thing. Business might suffer from bad or inaccurate data since it can seriously impair choices that depend on it.

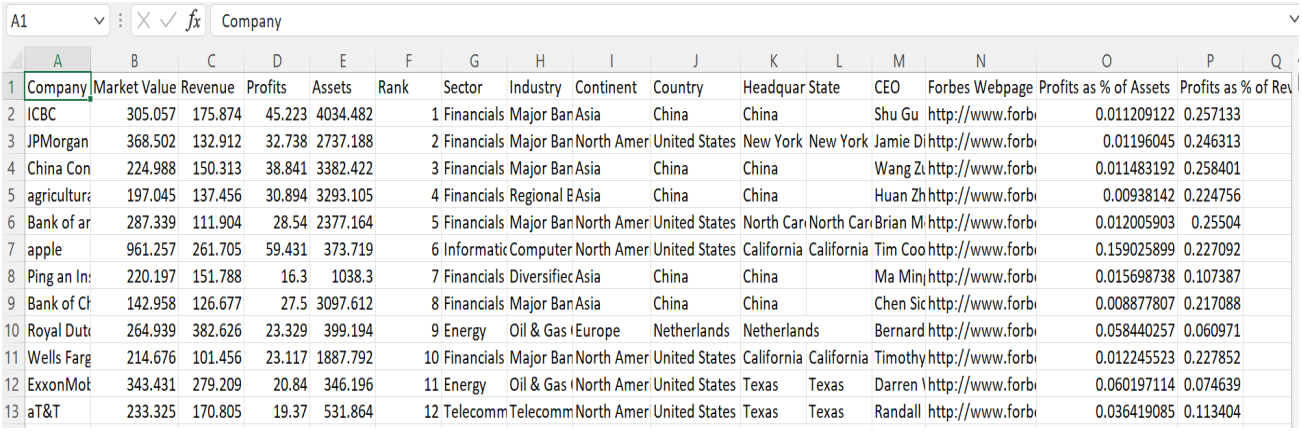
Although it may appear uninteresting and tedious, data cleansing is one of the most crucial activities a data science expert must perform. Having inaccurate or poor-quality data might hurt your procedures and analyses. In my case the initial data was in a good quality. However, it seemed little uninteresting and inconsistent. So, I did make sure that data is interesting and not distracting and for this I used two tools one of them is Microsoft Excel and the other is Open Refine which is a open source data cleaning tool

Further this project will describe the data cleaning with details in three categories

- Standardization
- Uniformity
- Consistency

## 4.1 Order and Standardization

The columns in dataset were reordered from its initial order to ensure that the data flow is seamless and easy to read and follow the data flow. To do this I used a tool Open Refine it has a built-in feature which allows user to reorder the columns and the columns with currency didn't have any indication so indeed renamed the columns and ensured that everything is in billion US dollars. Further there was a column with many null values, and it was unnecessary so eliminated the entire column. Everything was done in Open Refine.



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Company | Market Value | Revenue | Profits | Assets | Rank | Sector | Industry | Continent | Country | Headquar | State | CEO | Forbes Webpage | Profits as % of Assets | Profits as % of Rev |
| 2 | ICBC | 305.057 | 175.874 | 45.223 | 4034.482 | 1 | Financials | Major Ban | Asia | China | China | | Shu Gu | http://www.forb | 0.011209122 | 0.257133 |
| 3 | JPMorgan | 368.502 | 132.912 | 32.738 | 2737.188 | 2 | Financials | Major Ban | North Amer | United States | New York | New York | Jamie Di | http://www.forb | 0.01196045 | 0.246313 |
| 4 | China Con | 224.988 | 150.313 | 38.841 | 3382.422 | 3 | Financials | Major Ban | Asia | China | China | | Wang Zu | http://www.forb | 0.011483192 | 0.258401 |
| 5 | agricultura | 197.045 | 137.456 | 30.894 | 3293.105 | 4 | Financials | Regional B | Asia | China | China | | Huan Zh | http://www.forb | 0.00938142 | 0.224756 |
| 6 | Bank of ar | 287.339 | 111.904 | 28.54 | 2377.164 | 5 | Financials | Major Ban | North Amer | United States | North Car | North Car | Brian M | http://www.forb | 0.012005903 | 0.25504 |
| 7 | apple | 961.257 | 261.705 | 59.431 | 373.719 | 6 | Informatic | Computer | North Amer | United States | California | California | Tim Coo | http://www.forb | 0.159025899 | 0.227092 |
| 8 | Ping an Ins | 220.197 | 151.788 | 16.3 | 1038.3 | 7 | Financials | Diversifiec | Asia | China | China | | Ma Ming | http://www.forb | 0.015698738 | 0.107387 |
| 9 | Bank of Ch | 142.958 | 126.677 | 27.5 | 3097.612 | 8 | Financials | Major Ban | Asia | China | China | | Chen Sic | http://www.forb | 0.008877807 | 0.217088 |
| 10 | Royal Dutc | 264.939 | 382.626 | 23.329 | 399.194 | 9 | Energy | Oil & Gas ( | Europe | Netherlands | Netherlands | | Bernard | http://www.forb | 0.058440257 | 0.060971 |
| 11 | Wells Farg | 214.676 | 101.456 | 23.117 | 1887.792 | 10 | Financials | Major Ban | North Amer | United States | California | California | Timothy | http://www.forb | 0.012245523 | 0.227852 |
| 12 | ExxonMob | 343.431 | 279.209 | 20.84 | 346.196 | 11 | Energy | Oil & Gas ( | North Amer | United States | Texas | Texas | Darren \ | http://www.forb | 0.060197114 | 0.074639 |
| 13 | aT&T | 233.325 | 170.805 | 19.37 | 531.864 | 12 | Telecomm | Telecomm | North Amer | United States | Texas | Texas | Randall | http://www.forb | 0.036419085 | 0.113404 |

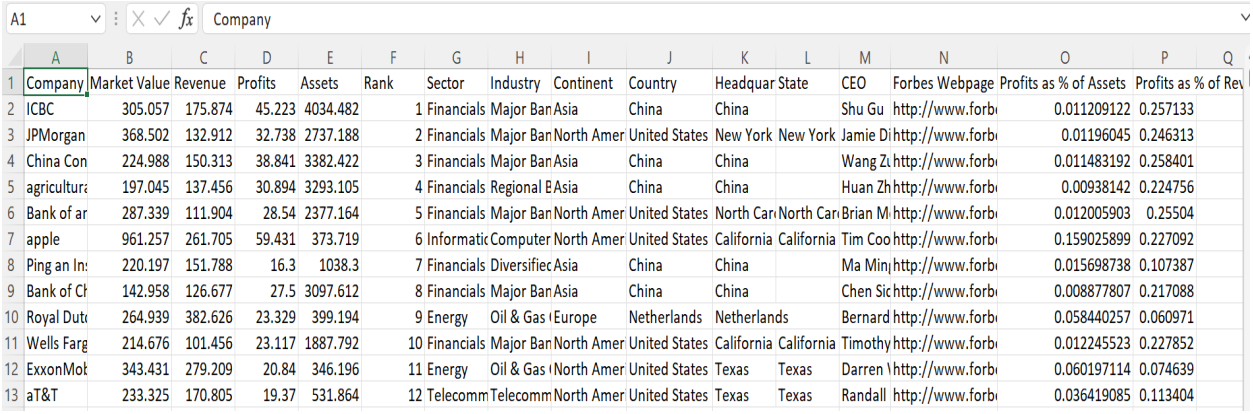
Figure 1.1 - Pre-cleaning

| All | | Rank | Company | Headquarters | Industry | Market Value $B | Assets $B | Revenue $B | Profits $B | Profits as % of Revenue | Profits as % of Assets | CEO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 1 | ICBC | China, China | Major Banks - Financials | 305.057 | 4034.482 | 175.874 | 45.223 | 0.257132947 | 0.011209122 | Shu Gu |
| 2. | 2 | JPMorgan Chase | New York, United States | Major Banks - Financials | 368.502 | 2737.188 | 132.912 | 32.738 | 0.24631335 | 0.01196045 | Jamie Dimon |
| 3. | 3 | China Construction Bank | China, China | Major Banks - Financials | 224.988 | 3382.422 | 150.313 | 38.841 | 0.258400804 | 0.011483192 | Wang Zuji |
| 4. | 4 | Agricultural Bank of China | China, China | Regional Banks - Financials | 197.045 | 3293.105 | 137.456 | 30.894 | 0.224755558 | 0.00938142 | Huan Zhao |
| 5. | 5 | Bank of America | North Carolina, United States | Major Banks - Financials | 287.339 | 2377.164 | 111.904 | 28.54 | 0.255040034 | 0.012005903 | Brian Moynihan |
| 6. | 6 | Apple | California, United States | Computer Hardware - Information Technology | 961.257 | 373.719 | 261.705 | 59.431 | 0.227091573 | 0.159025899 | Tim Cook |
| 7. | 7 | Ping An Insurance Group | China, China | Diversified Insurance - Financials | 220.197 | 1038.3 | 151.788 | 16.3 | 0.107386618 | 0.015698738 | Ma Mingzhe |
| 8. | 8 | Bank of China | China, China | Major Banks - Financials | 142.958 | 3097.612 | 126.677 | 27.5 | 0.217087553 | 0.008877807 | Chen Siqing |
| 9. | 9 | Royal Dutch Shell | Netherlands, Netherlands | Oil & Gas Operations - Energy | 264.939 | 399.194 | 382.626 | 23.329 | 0.060970765 | 0.058440257 | Bernardus Margriet van Beurden |
| 10. | 10 | Wells Fargo | California, United States | Major Banks - Financials | 214.676 | 1887.792 | 101.456 | 23.117 | 0.227852468 | 0.012245523 | Timothy Sloan |

Figure 1.2 - Post-cleaning

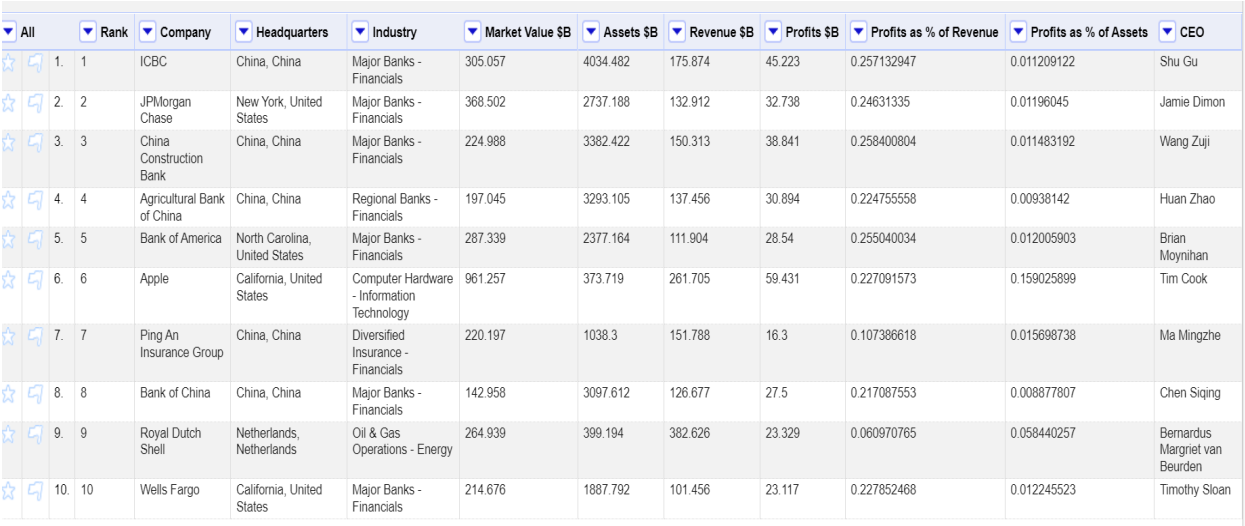
## 4.2 Uniformity

The data must be normalized to a consistent form as the next stage in enhancing the database's quality. The main purpose of this technique is to make it easier to search the database for details about a certain organization. So, to ensure the uniformity of the data flow some columns were split and then merged into one. In this dataset country and headquarters were merged to remove distraction as there were multiple columns with location like continent, country, headquarters, and state. As you can see in figure 2.2 the continent and state columns were merged and then removed further country was merged into headquarters column. This was done in Open Refine.



| | Company | Market Value | Revenue | Profits | Assets | Rank | Sector | Industry | Continent | Country | Headquar | State | CEO | Forbes Webpage | Profits as % of Assets | Profits as % of Rev |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | ICBC | 305.057 | 175.874 | 45.223 | 4034.482 | 1 | Financials | Major Ban | Asia | China | China | | Shu Gu | http://www.forbe | 0.011209122 | 0.257133 |
| 3 | JPMorgan | 368.502 | 132.912 | 32.738 | 2737.188 | 2 | Financials | Major Ban | North Amer | United States | New York | New York | Jamie Di | http://www.forbe | 0.01196045 | 0.246313 |
| 4 | China Con | 224.988 | 150.313 | 38.841 | 3382.422 | 3 | Financials | Major Ban | Asia | China | China | | Wang Zu | http://www.forbe | 0.011483192 | 0.258401 |
| 5 | agricultura | 197.045 | 137.456 | 30.894 | 3293.105 | 4 | Financials | Regional B | Asia | China | China | | Huan Zh | http://www.forbe | 0.00938142 | 0.224756 |
| 6 | Bank of ar | 287.339 | 111.904 | 28.54 | 2377.164 | 5 | Financials | Major Ban | North Amer | United States | North Car | North Car | Brian M | http://www.forbe | 0.012005903 | 0.25504 |
| 7 | apple | 961.257 | 261.705 | 59.431 | 373.719 | 6 | Informatic | Computer | North Amer | United States | California | California | Tim Coo | http://www.forbe | 0.159025899 | 0.227092 |
| 8 | Ping an In | 220.197 | 151.788 | 16.3 | 1038.3 | 7 | Financials | Diversifiec | Asia | China | China | | Ma Min | http://www.forbe | 0.015698738 | 0.107387 |
| 9 | Bank of Ch | 142.958 | 126.677 | 27.5 | 3097.612 | 8 | Financials | Major Ban | Asia | China | China | | Chen Sic | http://www.forbe | 0.008877807 | 0.217088 |
| 10 | Royal Dute | 264.939 | 382.626 | 23.329 | 399.194 | 9 | Energy | Oil & Gas | Europe | Netherlands | Netherlands | | Bernard | http://www.forbe | 0.058440257 | 0.060971 |
| 11 | Wells Farg | 214.676 | 101.456 | 23.117 | 1887.792 | 10 | Financials | Major Ban | North Amer | United States | California | California | Timothy | http://www.forbe | 0.012245523 | 0.227852 |
| 12 | ExxonMob | 343.431 | 279.209 | 20.84 | 346.196 | 11 | Energy | Oil & Gas | North Amer | United States | Texas | Texas | Darren \ | http://www.forbe | 0.060197114 | 0.074639 |
| 13 | aT&T | 233.325 | 170.805 | 19.37 | 531.864 | 12 | Telecomm | Telecomm | North Amer | United States | Texas | Texas | Randall | http://www.forbe | 0.036419085 | 0.113404 |

Figure 2.1 – Pre cleaning

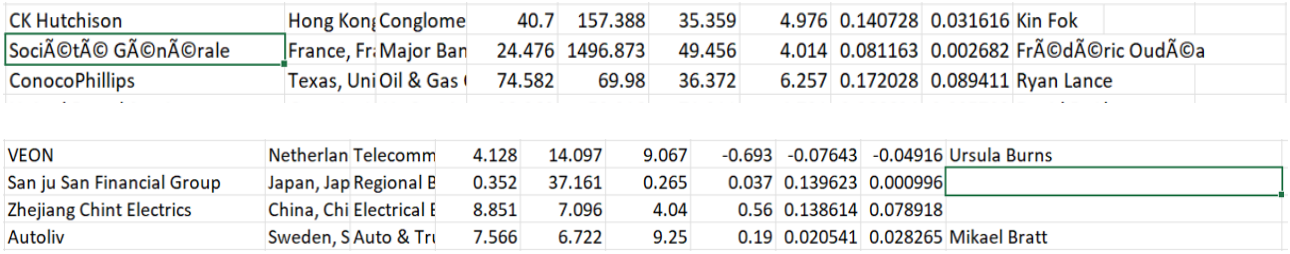


| All | Rank | Company | Headquarters | Industry | Market Value $B | Assets $B | Revenue $B | Profits $B | Profits as % of Revenue | Profits as % of Assets | CEO |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 1 | ICBC | China, China | Major Banks - Financials | 305.057 | 4034.482 | 175.874 | 45.223 | 0.257132947 | 0.011209122 | Shu Gu |
| 2. | 2 | JPMorgan Chase | New York, United States | Major Banks - Financials | 368.502 | 2737.188 | 132.912 | 32.738 | 0.24631335 | 0.01196045 | Jamie Dimon |
| 3. | 3 | China Construction Bank | China, China | Major Banks - Financials | 224.988 | 3382.422 | 150.313 | 38.841 | 0.258400804 | 0.011483192 | Wang Zuji |
| 4. | 4 | Agricultural Bank of China | China, China | Regional Banks - Financials | 197.045 | 3293.105 | 137.456 | 30.894 | 0.224755558 | 0.00938142 | Huan Zhao |
| 5. | 5 | Bank of America | North Carolina, United States | Major Banks - Financials | 287.339 | 2377.164 | 111.904 | 28.54 | 0.255040034 | 0.012005903 | Brian Moynihan |
| 6. | 6 | Apple | California, United States | Computer Hardware - Information Technology | 961.257 | 373.719 | 261.705 | 59.431 | 0.227091573 | 0.159025899 | Tim Cook |
| 7. | 7 | Ping An Insurance Group | China, China | Diversified Insurance - Financials | 220.197 | 1038.3 | 151.788 | 16.3 | 0.107386618 | 0.015698738 | Ma Mingzhe |
| 8. | 8 | Bank of China | China, China | Major Banks - Financials | 142.958 | 3097.612 | 126.677 | 27.5 | 0.217087553 | 0.008877807 | Chen Siqing |
| 9. | 9 | Royal Dutch Shell | Netherlands, Netherlands | Oil & Gas Operations - Energy | 264.939 | 399.194 | 382.626 | 23.329 | 0.060970765 | 0.058440257 | Bernardus Margriet van Beurden |
| 10. | 10 | Wells Fargo | California, United States | Major Banks - Financials | 214.676 | 1887.792 | 101.456 | 23.117 | 0.227852468 | 0.012245523 | Timothy Sloan |

Figure 2.2 – Post cleaning

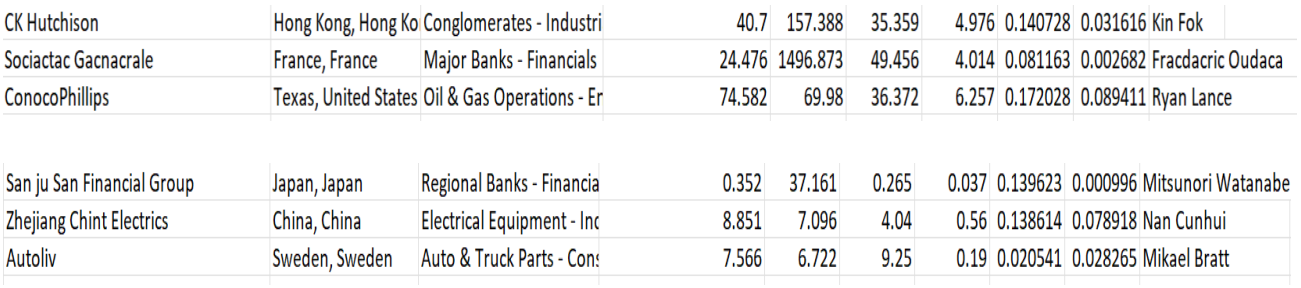
## 4.3 Consistency

In this dataset some columns contained special characters like @ ©Ã which were replaced with letters to ensure the consistency of the information and making it easy to read. Moreover, there was missing information in the CEO column removing the missing value completely might remove useful insights from your data. So, the missing data was filled manually. Everything was done in Excel by using find and replace feature.



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CK Hutchison | Hong Kon | Conglome | 40.7 | 157.388 | 35.359 | 4.976 | 0.140728 | 0.031616 | Kin Fok |
| SociÃ©tÃ© GÃ©nÃ©rale | France, Fr | Major Ban | 24.476 | 1496.873 | 49.456 | 4.014 | 0.081163 | 0.002682 | FrÃ©dÃ©ric OudÃ©a |
| ConocoPhillips | Texas, Uni | Oil & Gas | 74.582 | 69.98 | 36.372 | 6.257 | 0.172028 | 0.089411 | Ryan Lance |
| VEON | Netherlan | Telecomm | 4.128 | 14.097 | 9.067 | -0.693 | -0.07643 | -0.04916 | Ursula Burns |
| San ju San Financial Group | Japan, Jap | Regional B | 0.352 | 37.161 | 0.265 | 0.037 | 0.139623 | 0.000996 | |
| Zhejiang Chint Electrics | China, Chi | Electrical E | 8.851 | 7.096 | 4.04 | 0.56 | 0.138614 | 0.078918 | |
| Autoliv | Sweden, S | Auto & Tru | 7.566 | 6.722 | 9.25 | 0.19 | 0.020541 | 0.028265 | Mikael Bratt |

Figure 3.1 – Pre cleaning



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CK Hutchison | Hong Kong, Hong Ko | Conglomerates - Industri | 40.7 | 157.388 | 35.359 | 4.976 | 0.140728 | 0.031616 | Kin Fok |
| Sociactac Gacnacrale | France, France | Major Banks - Financials | 24.476 | 1496.873 | 49.456 | 4.014 | 0.081163 | 0.002682 | Fracdacric Oudaca |
| ConocoPhillips | Texas, United States | Oil & Gas Operations - Er | 74.582 | 69.98 | 36.372 | 6.257 | 0.172028 | 0.089411 | Ryan Lance |
| San ju San Financial Group | Japan, Japan | Regional Banks - Financia | 0.352 | 37.161 | 0.265 | 0.037 | 0.139623 | 0.000996 | Mitsunori Watanabe |
| Zhejiang Chint Electrics | China, China | Electrical Equipment - Inc | 8.851 | 7.096 | 4.04 | 0.56 | 0.138614 | 0.078918 | Nan Cunhui |
| Autoliv | Sweden, Sweden | Auto & Truck Parts - Cons | 7.566 | 6.722 | 9.25 | 0.19 | 0.020541 | 0.028265 | Mikael Bratt |

Figure 3.2 – Post cleaning

## 5. Data Visualizations

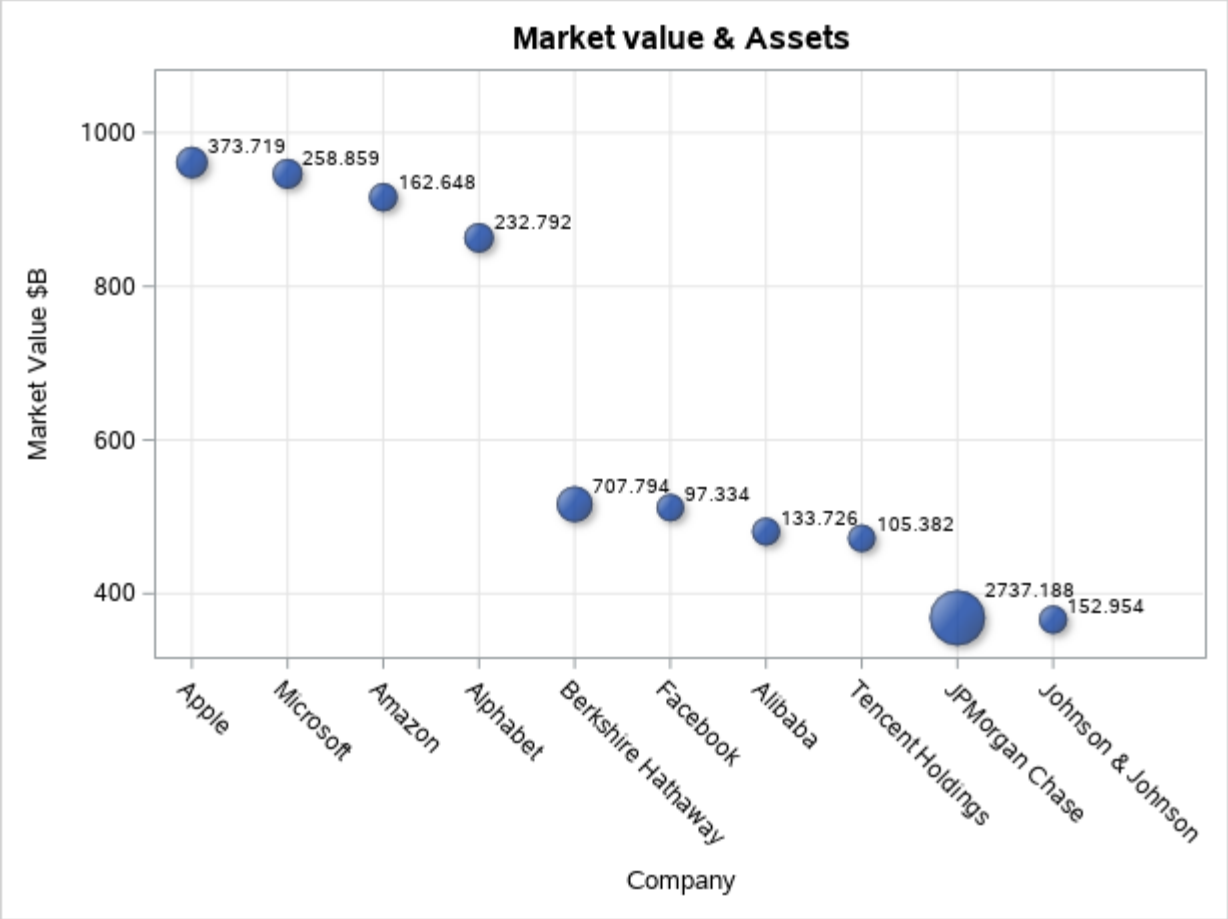
The graphic display of information and data is known as data visualization. Data visualization tools offer an easy approach to observe and analyze trends, outliers, and patterns in data by utilizing visual components like charts, graphs, and maps. Additionally, it offers a great tool for staff members or business owners to clearly deliver data to non-technical audiences.

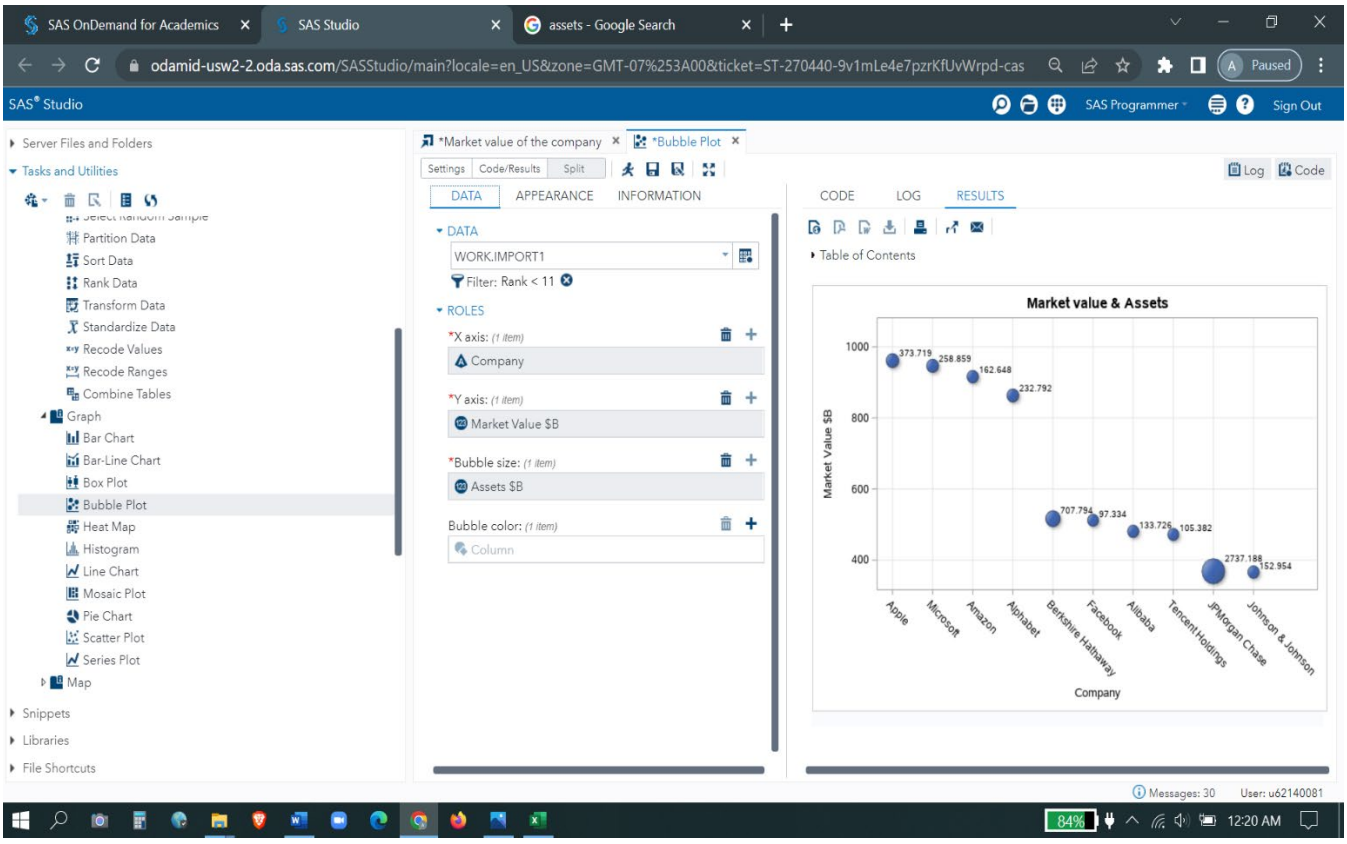
Further we will see the following data visualizations using various graphs such as

1. Companies with the highest market value – Bubble plot
2. Which sectors are making huge profits – Pie Chart
3. Which sectors are making lowest profits – Bar chart
4. Types of Industry of the top companies – Bubble plot
5. Companies with highest assets holdings – Box Plot
6. Difference in revenue of top companies – Bar chart

## 5.1 Companies with highest market value

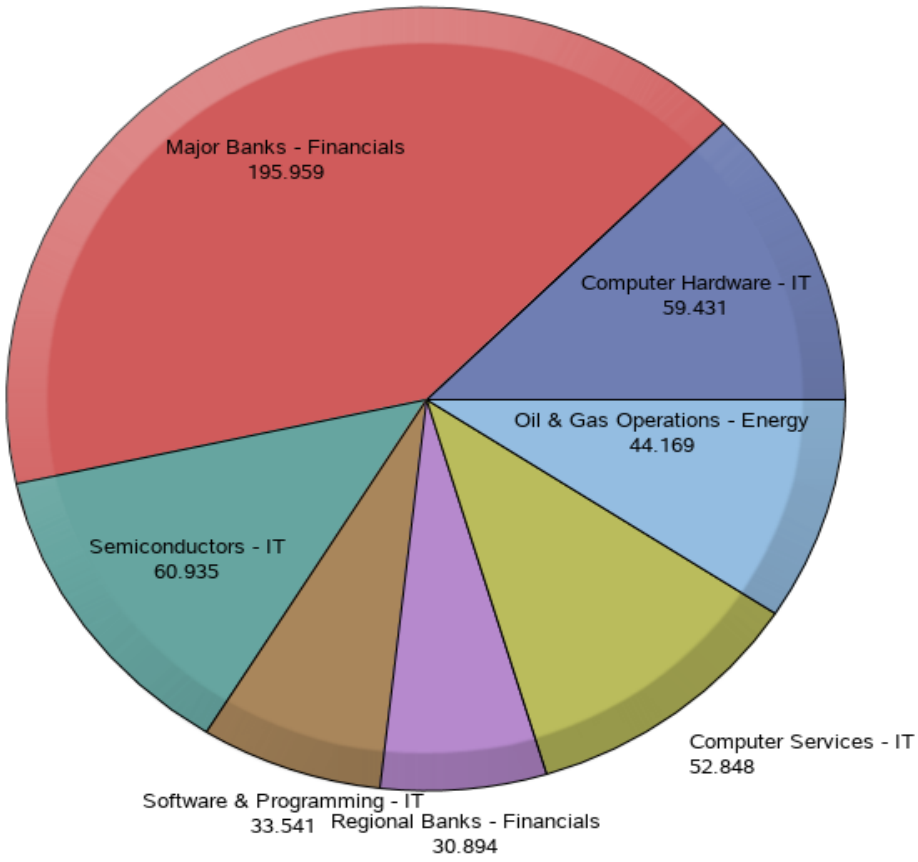
In this visualization a bubble plot is used to show the companies with the highest market value and their respective assets. X-axis is used to show companies and Y-axis shows the figure of the market value in billion USD and the blue bubble represents the third information being the assets of their respective companies. As can be seen in the graph the apple has highest market value with more than $900 billion compared to Johnson & Johnson being the lowest with around $ 300 billion. Furthermore, JPMorgan Chase has the highest asset holding and lowest being the Facebook.

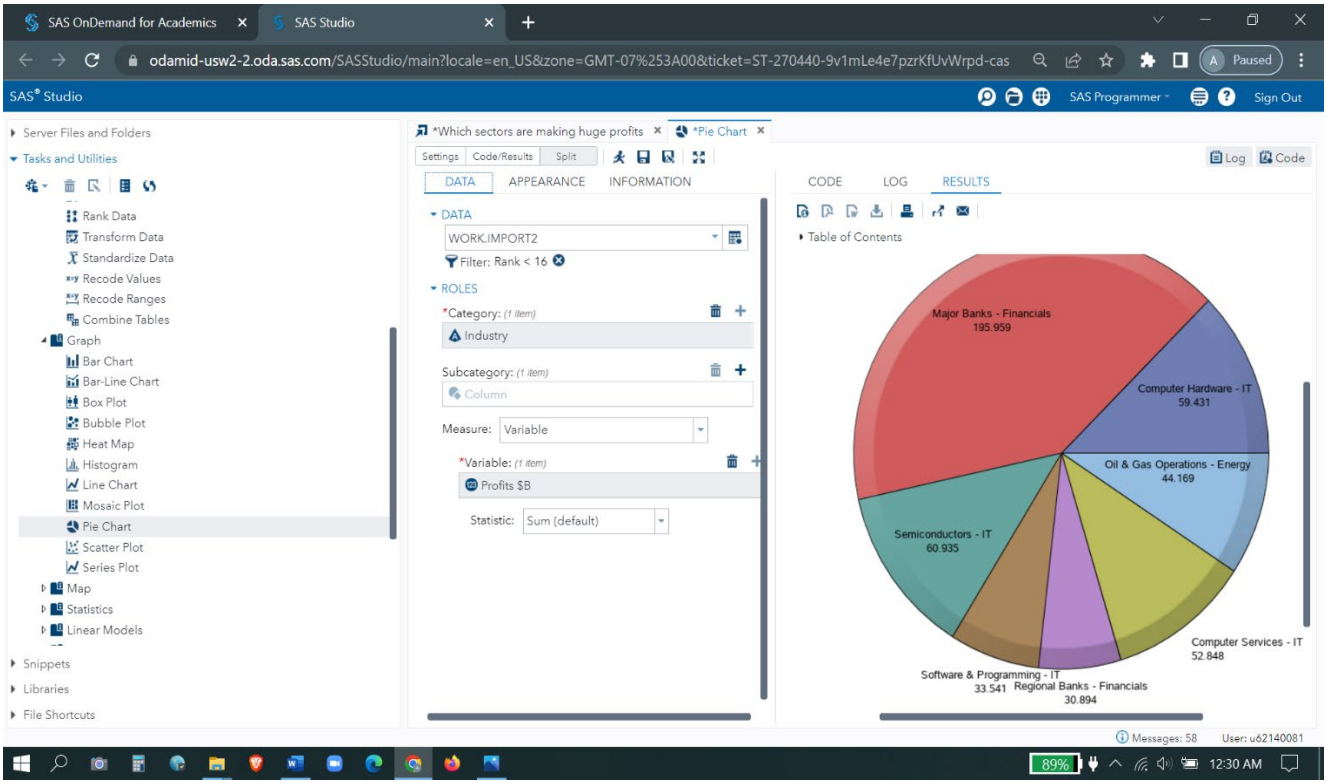



SAS OnDemand for Academics
SAS Studio
assets - Google Search
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
Paused
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Partition Data
Sort Data
Rank Data
Transform Data
Standardize Data
Recode Values
Recode Ranges
Combine Tables
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Pie Chart
Scatter Plot
Series Plot
Map
Snippets
Libraries
File Shortcuts
*Market value of the company
*Bubble Plot
Settings
Code/Results
Split
Log
Code
DATA
APPEARANCE
INFORMATION
DATA
WORK.IMPORT1
Filter: Rank < 11
ROLES
*X axis: (1 item)
Company
*Y axis: (1 item)
Market Value $B
*Bubble size: (1 item)
Assets $B
Bubble color: (1 item)
Column
CODE
LOG
RESULTS
Table of Contents
Market value & Assets
Market Value $B
1000
800
600
400
373.719
258.859
162.648
232.792
707.794
97.334
133.726
105.382
2737.188
152.954
Apple
Microsoft
Amazon
Alphabet
Berkshire Hathaway
Facebook
Alibaba
Tencent Holdings
JPMorgan Chase
Johnson & Johnson
Company
Messages: 30
User: u62140081
84%
12:20 AM

## 5.2 Sectors with the huge profits

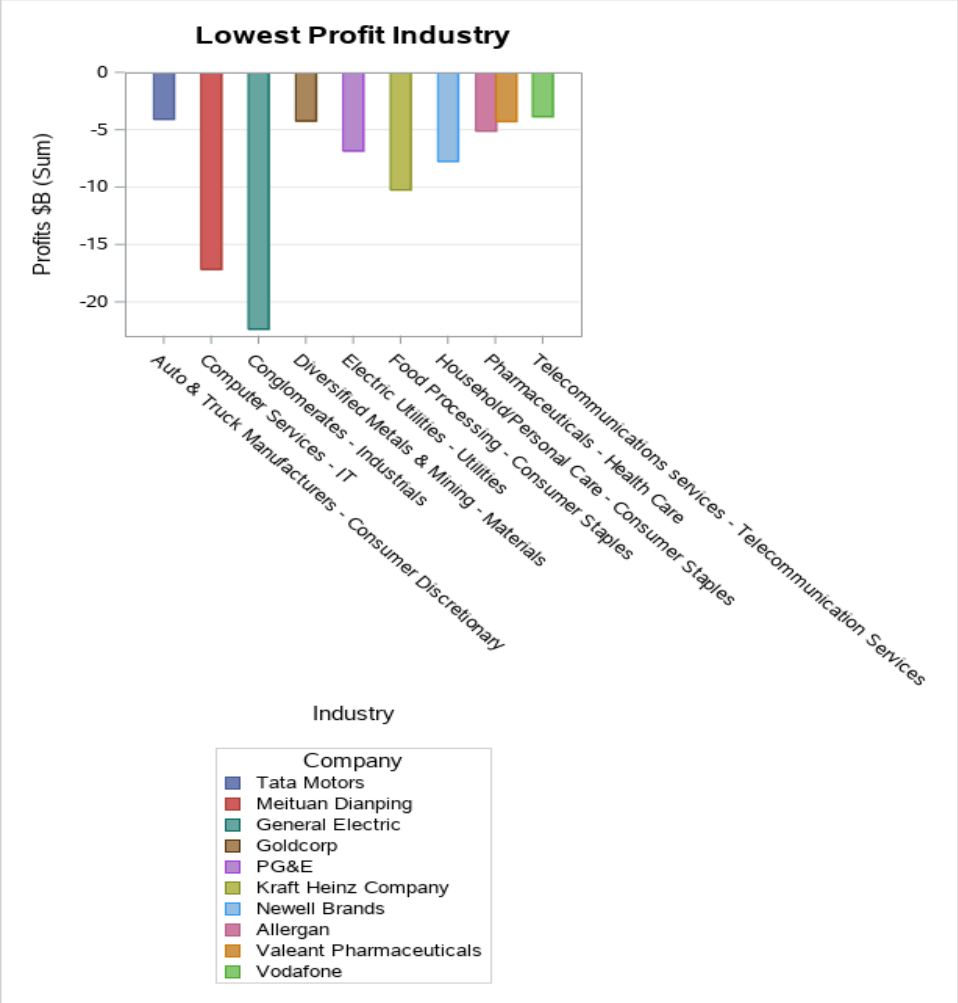
Sectors define the entire organization and their future so it will be very insightful to see that which sectors are making the highest profits. A pie chart is a good example to show the proportion of sectors that are dominating the industry with there profits. As can be seen in this visualization here the proportion is shown in billion USD and the enormous amount of profit is generated by banks of course they will dominate as they circulate the money. The second highest profit generating sector is semiconductors IT sector and with nearly the same amount of profit is generated by Computer hardware industry. Further, surprisingly the regional banks have the lowest profit in the among every other sectors.

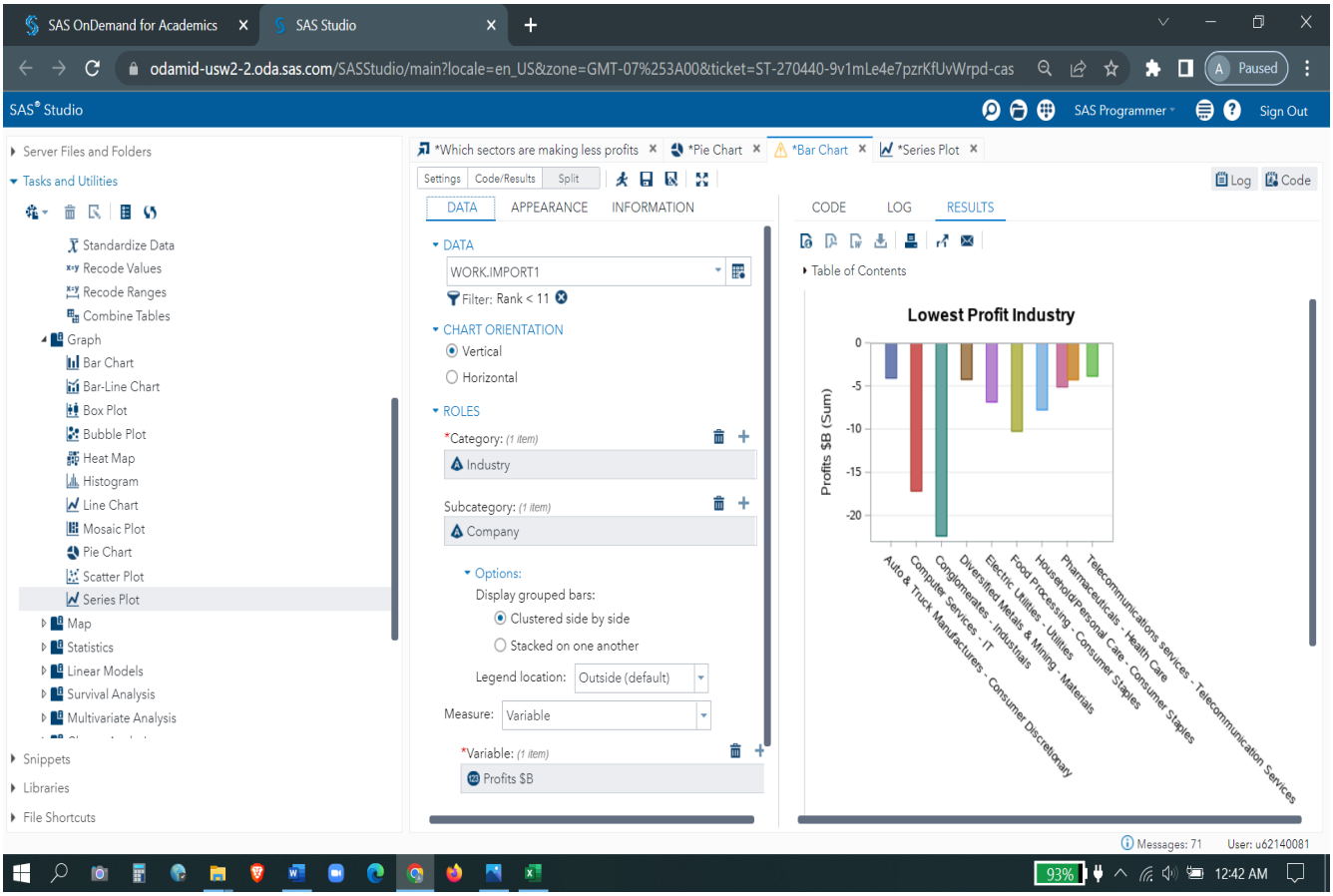



SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Rank Data
Transform Data
Standardize Data
Recode Values
Recode Ranges
Combine Tables
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Pie Chart
Scatter Plot
Series Plot
Map
Statistics
Linear Models
Snippets
Libraries
File Shortcuts
*Which sectors are making huge profits
*Pie Chart
Settings
Code/Results
Split
DATA
APPEARANCE
INFORMATION
DATA
WORK.IMPORT2
Filter: Rank < 16
ROLES
*Category: (1 item)
Industry
Subcategory: (1 item)
Column
Measure: Variable
*Variable: (1 item)
Profits $B
Statistic: Sum (default)
Log
Code
CODE
LOG
RESULTS
Table of Contents
Major Banks - Financials
195.959
Computer Hardware - IT
59.431
Oil & Gas Operations - Energy
44.169
Computer Services - IT
52.848
Regional Banks - Financials
30.894
Software & Programming - IT
33.541
Semiconductors - IT
60.935
Messages: 58
User: u62140081
89%
12:30 AM

## 5.3 Sectors with the loss and lowest profit

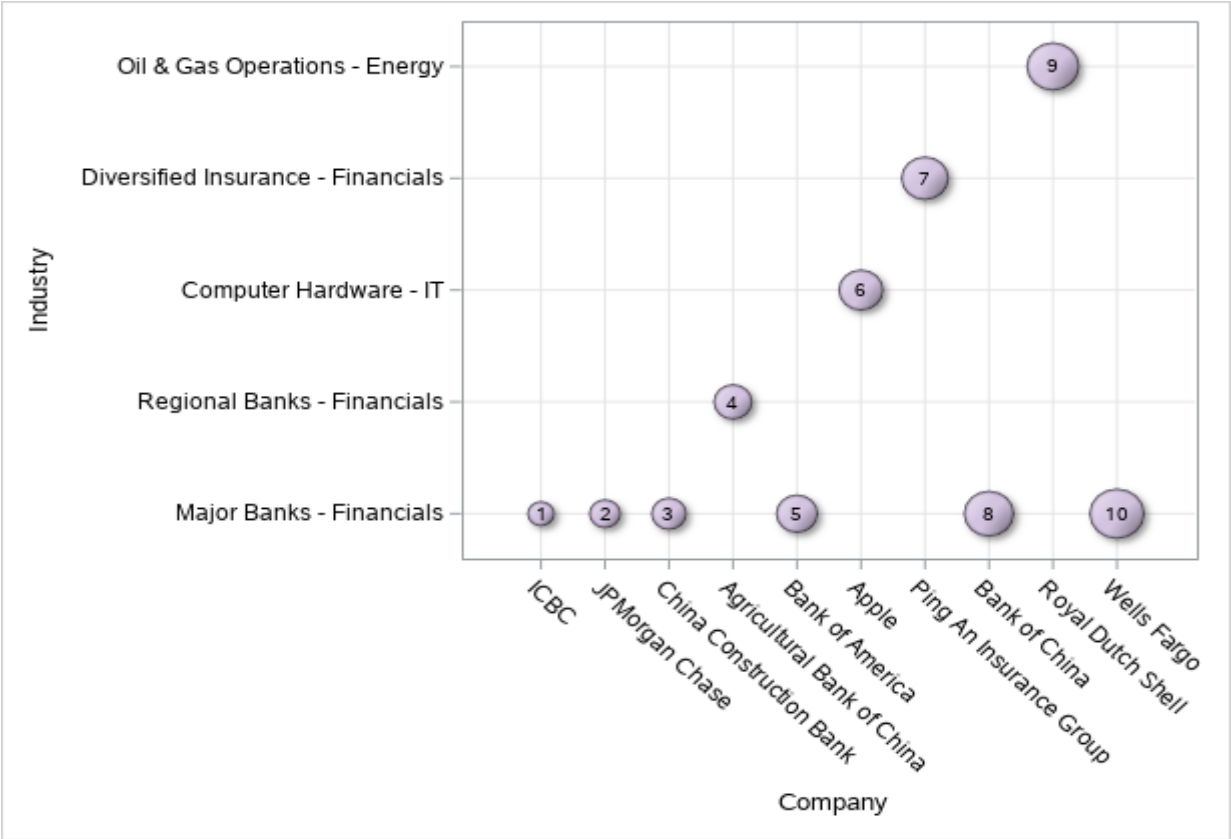
As the organizations make huge profits there are many sectors which makes lowest or nearly no profit, and this visualization exactly shows which sectors are in loss and making nearly no profit. As can be seen the conglomerates suffers the highest loss with more than $20 billion as they are multiple business entities. The IT sectors is as bad as good as you can see it is the second highest suffering sector. Further, the lowest loss is suffered in telecommunication sector with less than $5 billion. Moreover, the company names are given in their respective sectors as a third information which turns out to be very insightful.

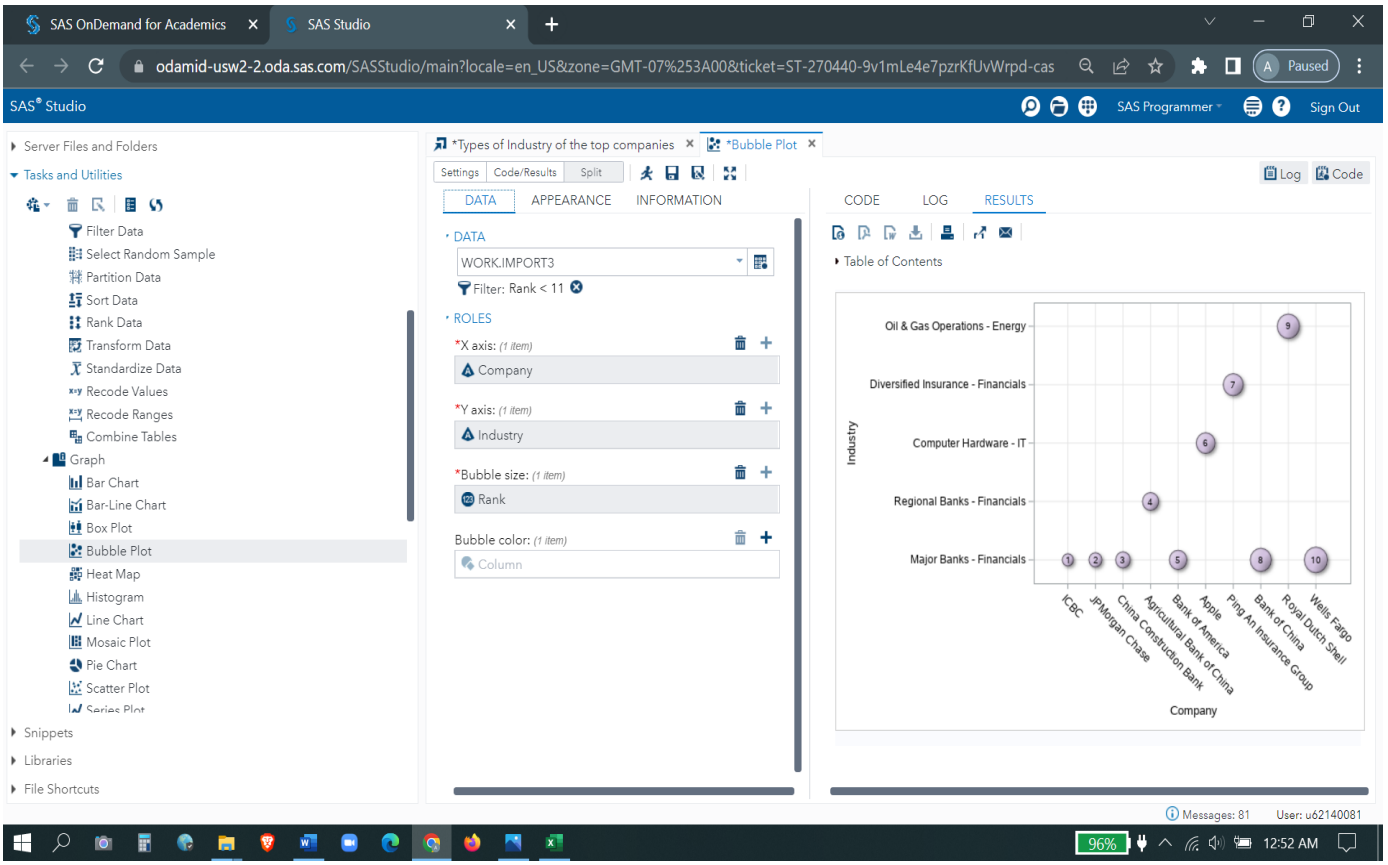



SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
Paused
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Standardize Data
Recode Values
Recode Ranges
Combine Tables
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Pie Chart
Scatter Plot
Series Plot
Map
Statistics
Linear Models
Survival Analysis
Multivariate Analysis
Snippets
Libraries
File Shortcuts
*Which sectors are making less profits
*Pie Chart
*Bar Chart
*Series Plot
Settings
Code/Results
Split
Log
Code
DATA
APPEARANCE
INFORMATION
CODE
LOG
RESULTS
DATA
WORK.IMPORT1
Filter: Rank < 11
CHART ORIENTATION
Vertical
Horizontal
ROLES
*Category: (1 item)
Industry
Subcategory: (1 item)
Company
Options:
Display grouped bars:
Clustered side by side
Stacked on one another
Legend location:
Outside (default)
Measure:
Variable
*Variable: (1 item)
Profits $B
Table of Contents
Lowest Profit Industry
Profits $B (Sum)
0
-5
-10
-15
-20
Auto & Truck Manufacturers - Consumer Discretionary
Computer Services - IT
Conglomerates - Industrials
Diversified Metals & Mining - Materials
Electric Utilities - Utilities
Food Processing - Consumer Staples
Household/Personal Care - Consumer Staples
Pharmaceuticals - Health Care
Telecommunications services - Telecommunication Services
Messages: 71
User: u62140081
93%
12:42 AM

## 5.4 Top companies and their industry

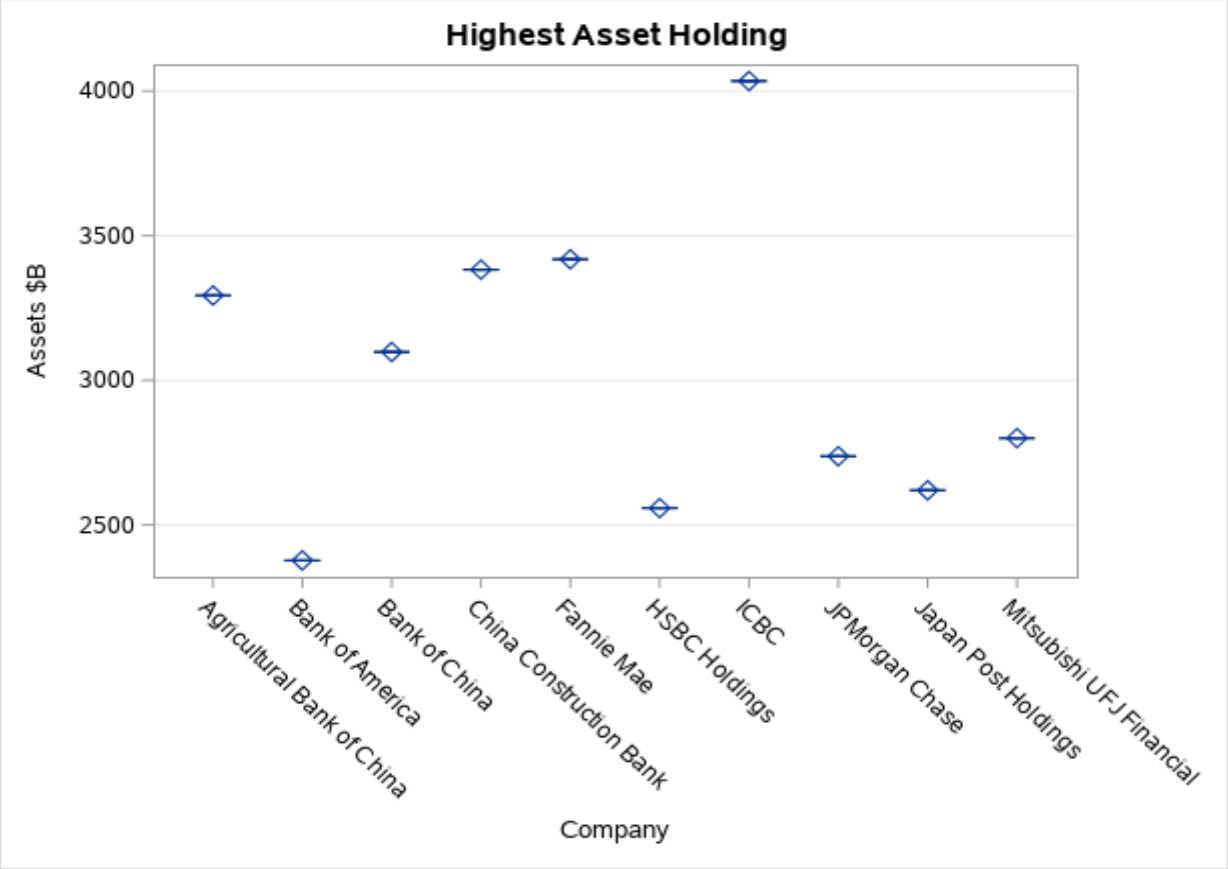
It is always fascinating to know the industry of the worlds top companies and this visualization gives a good information on the same by using bubble plot with the industry on y-axis and companies being on x-axis and third information being the bubble itself showing the global ranking of the companies. The major banking like ICBC and JP Morgan are ranked on the top with highest domination overall. Further Oil & Gas industry has only one company in the top ranking and same goes for the  IT industry and as always the highest number of top ranked companies are in financial sector with 7 being out of 10 companies.

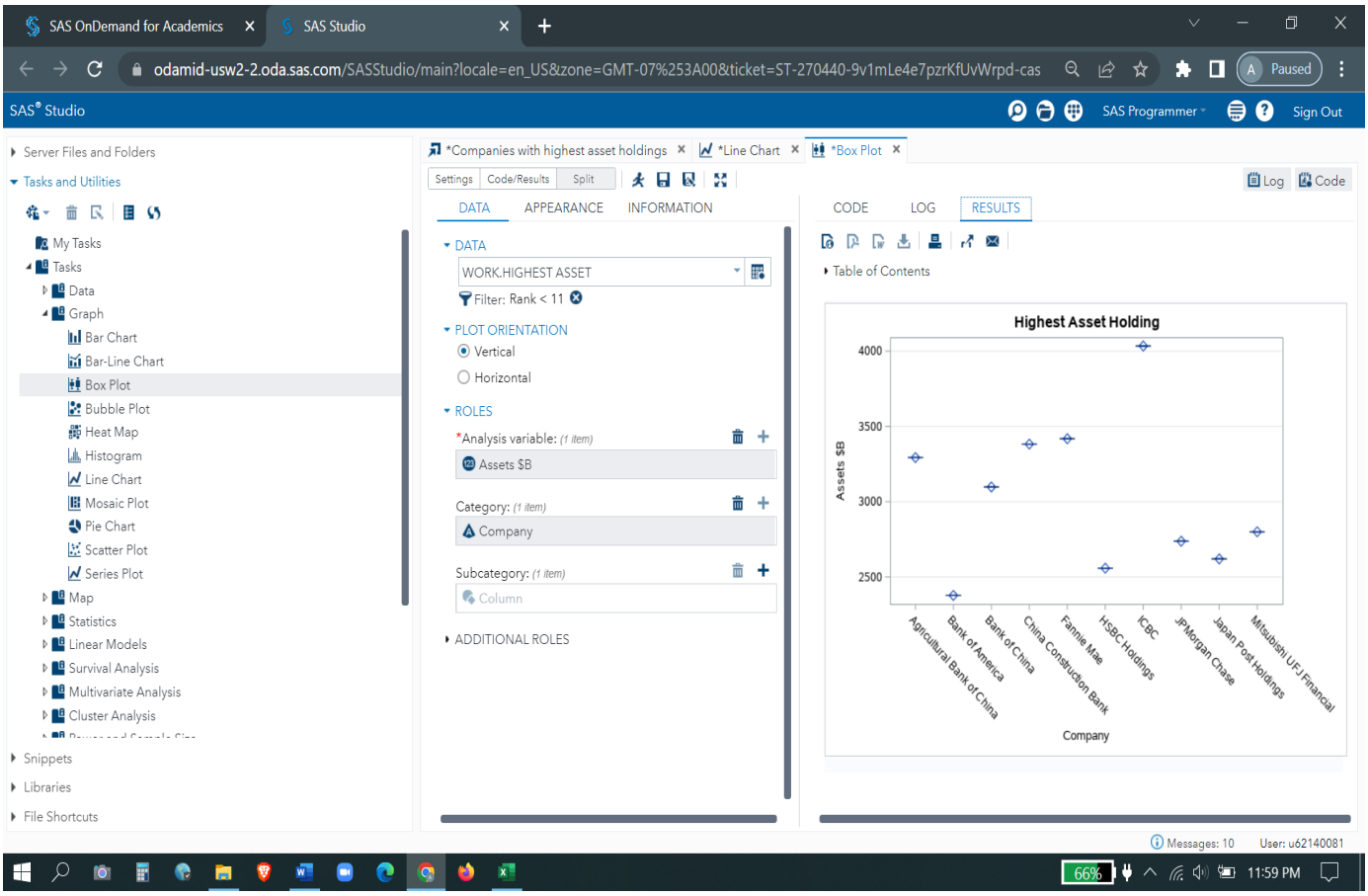



SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
Paused
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Filter Data
Select Random Sample
Partition Data
Sort Data
Rank Data
Transform Data
Standardize Data
Recode Values
Recode Ranges
Combine Tables
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Pie Chart
Scatter Plot
Snippets
Libraries
File Shortcuts
*Types of Industry of the top companies
*Bubble Plot
Settings
Code/Results
Split
Log
Code
DATA
APPEARANCE
INFORMATION
DATA
WORK.IMPORT3
Filter: Rank < 11
ROLES
*X axis: (1 item)
Company
*Y axis: (1 item)
Industry
*Bubble size: (1 item)
Rank
Bubble color: (1 item)
Column
CODE
LOG
RESULTS
Table of Contents
Oil & Gas Operations - Energy
Diversified Insurance - Financials
Computer Hardware - IT
Regional Banks - Financials
Major Banks - Financials
Industry
ICBC
JPMorgan Chase
China Construction Bank
Agricultural Bank of China
Bank of America
Apple
Ping An Insurance Group
Bank of China
Royal Dutch Shell
Wells Fargo
Company
Messages: 81
User: u62140081
96%
12:52 AM

## 5.5 Companies with highest asset holdings

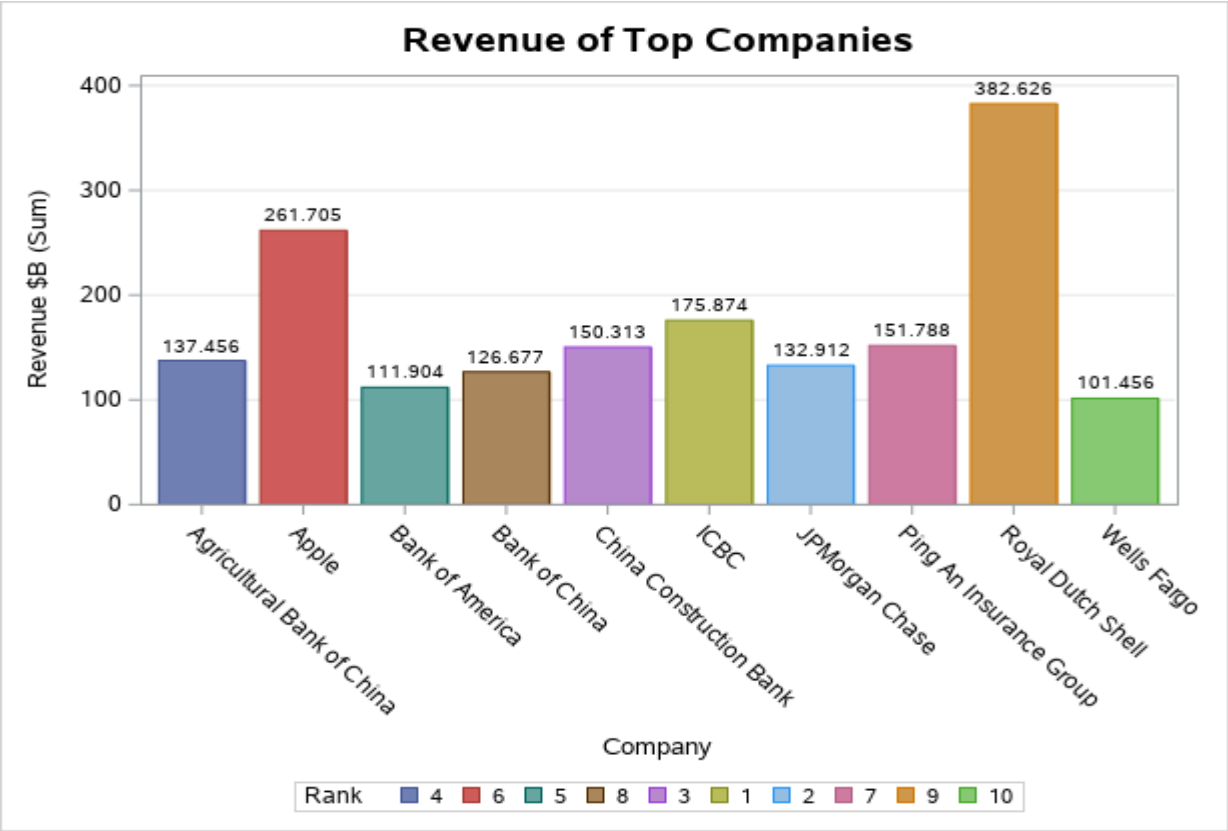
This box plot chart is not generally utilized but it is useful as this box plot shows the highest asset holding companies in the market. The ICBC hold the highest amount of assets than any other company out there with assets around $4000 billion. There is a bit difference with the second highest with nearly $3500 billion owned by Fannie Mae. Furthermore, the lowest being the bank of America. No matter how much the companies make the profits but the one with the highest number of assets last longer in the market.

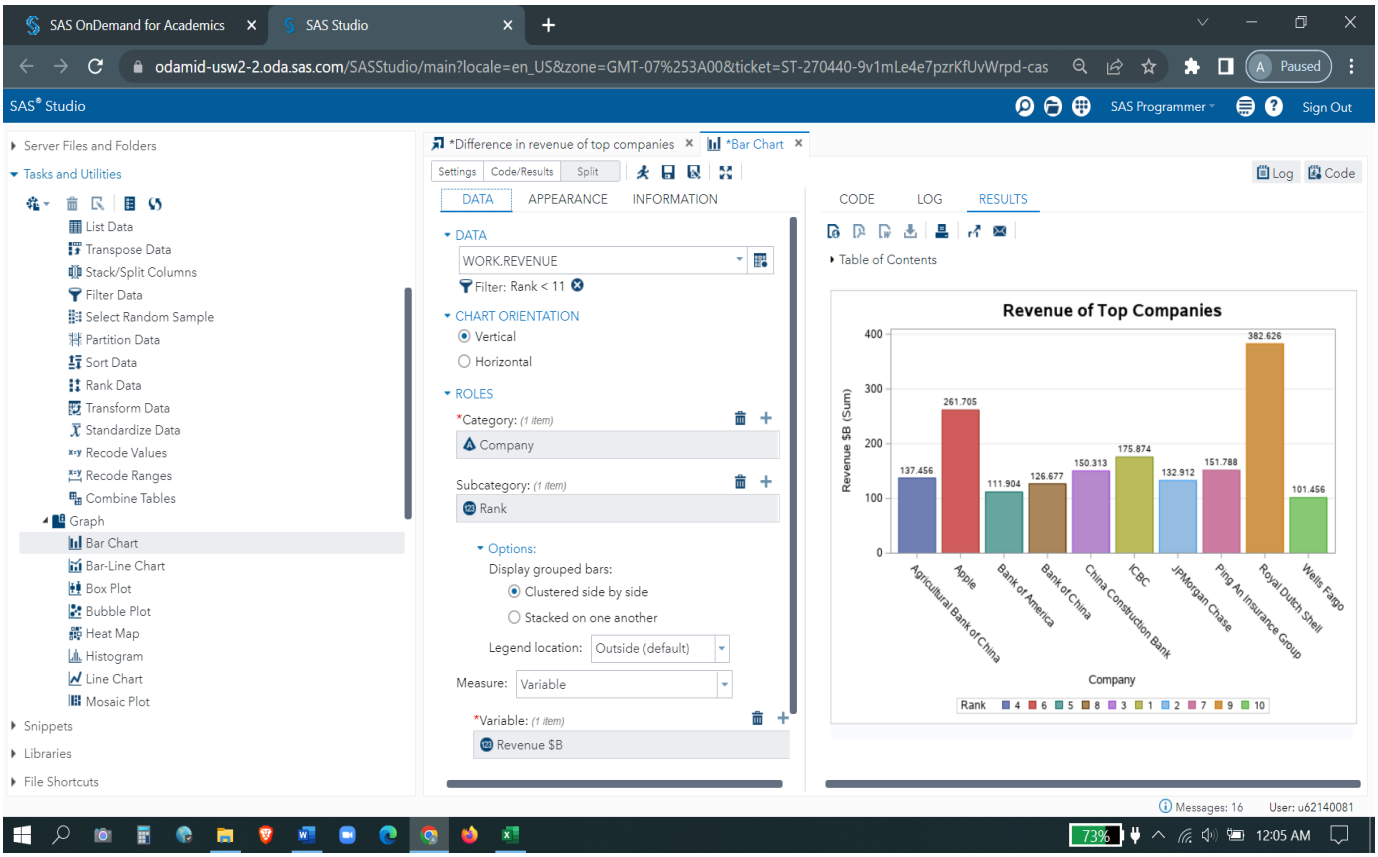



SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
Paused
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
My Tasks
Tasks
Data
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Pie Chart
Scatter Plot
Series Plot
Map
Statistics
Linear Models
Survival Analysis
Multivariate Analysis
Cluster Analysis
Snippets
Libraries
File Shortcuts
*Companies with highest asset holdings
*Line Chart
*Box Plot
Settings
Code/Results
Split
Log
Code
DATA
APPEARANCE
INFORMATION
CODE
LOG
RESULTS
DATA
WORK.HIGHEST ASSET
Filter: Rank < 11
PLOT ORIENTATION
Vertical
Horizontal
ROLES
*Analysis variable: (1 item)
Assets $B
Category: (1 item)
Company
Subcategory: (1 item)
Column
ADDITIONAL ROLES
Table of Contents
Highest Asset Holding
Assets $B
4000
3500
3000
2500
Agricultural Bank of China
Bank of America
Bank of China
China Construction Bank
Fannie Mae
HSBC Holdings
ICBC
JPMorgan Chase
Japan Post Holdings
Mitsubishi UFJ Financial
Company
Messages: 10
User: u62140081
66%
11:59 PM

## 5.6 Difference in revenue of top companies

The representation of the information in a bar graph is very easy to follow which can be seen in this visualization it shows the revenue of the top companies with their respective global ranks. Royal Dutch shell has the highest revenue with $382 billion with the apple being the second with $261 billion and Wells Fargo is the lowest in the list with $101 billion. Lastly, in terms of ranking the ICBC is ranked one globally and wells Fargo being the ranked 10th in the list.

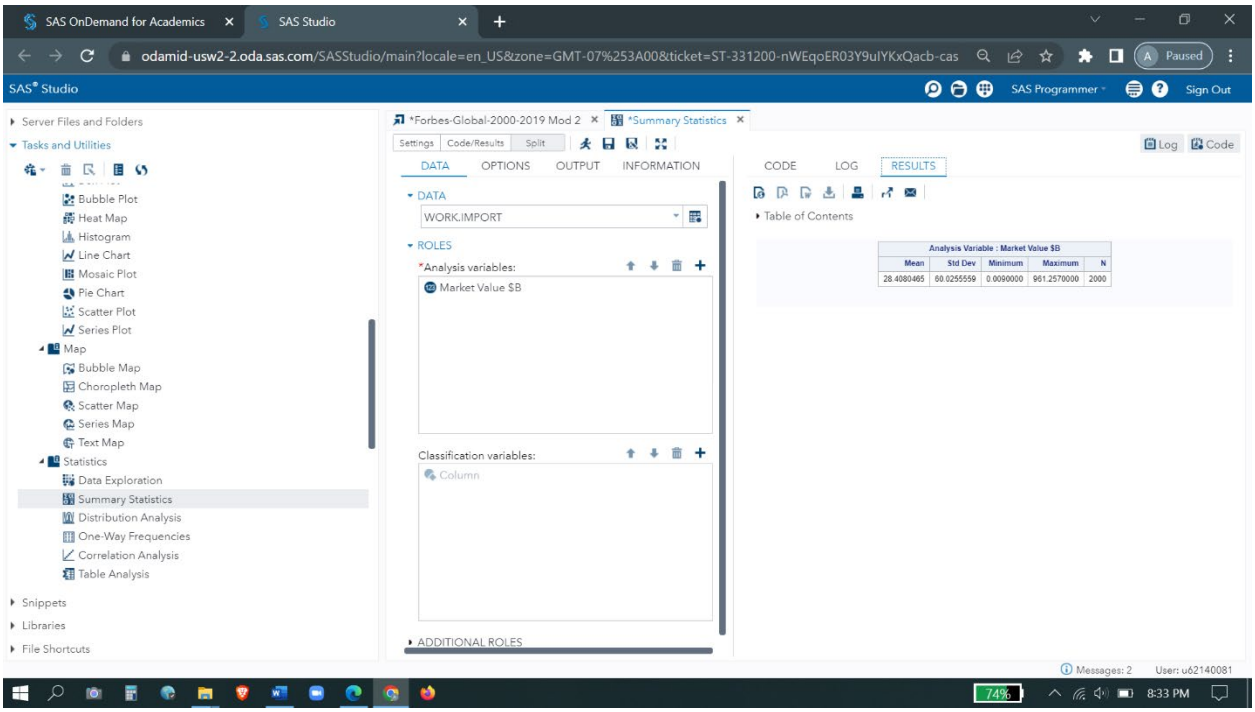



SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-270440-9v1mLe4e7pzrKfUvWrpd-cas
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
List Data
Transpose Data
Stack/Split Columns
Filter Data
Select Random Sample
Partition Data
Sort Data
Rank Data
Transform Data
Standardize Data
Recode Values
Recode Ranges
Combine Tables
Graph
Bar Chart
Bar-Line Chart
Box Plot
Bubble Plot
Heat Map
Histogram
Line Chart
Mosaic Plot
Snippets
Libraries
File Shortcuts
*Difference in revenue of top companies
*Bar Chart
Settings
Code/Results
Split
Log
Code
DATA
APPEARANCE
INFORMATION
DATA
WORK.REVENUE
Filter: Rank < 11
CHART ORIENTATION
Vertical
Horizontal
ROLES
*Category: (1 item)
Company
Subcategory: (1 item)
Rank
Options:
Display grouped bars:
Clustered side by side
Stacked on one another
Legend location:
Outside (default)
Measure:
Variable
*Variable: (1 item)
Revenue $B
CODE
LOG
RESULTS
Table of Contents
Revenue of Top Companies
Revenue $B (Sum)
400
300
200
100
0
137.456
261.705
111.904
126.677
150.313
175.874
132.912
151.788
382.626
101.456
Agricultural Bank of China
Apple
Bank of America
Bank of China
China Construction Bank
ICBC
JPMorgan Chase
Ping An Insurance Group
Royal Dutch Shell
Wells Fargo
Company
Rank
4
6
5
8
3
1
2
7
9
10
Messages: 16
User: u62140081
73%
12:05 AM

## 6. Statistical Summary

| Statical Summary | Explanation |
|---|---|
| Mean | Average of market value |
| Std Deviation | High value of std dev is 60.03 |
| Minimum | The least valuation of a company is $0.009 billion for the minimum value. |
| Maximum | The highest valued company has a maximum market value of $961 billion. |
| N | 2000 rows |

| Analysis Variable : Market Value $B | | | | |
|---|---|---|---|---|
| Mean | Std Dev | Minimum | Maximum | N |
| 28.4080465 | 60.0255559 | 0.0090000 | 961.2570000 | 2000 |

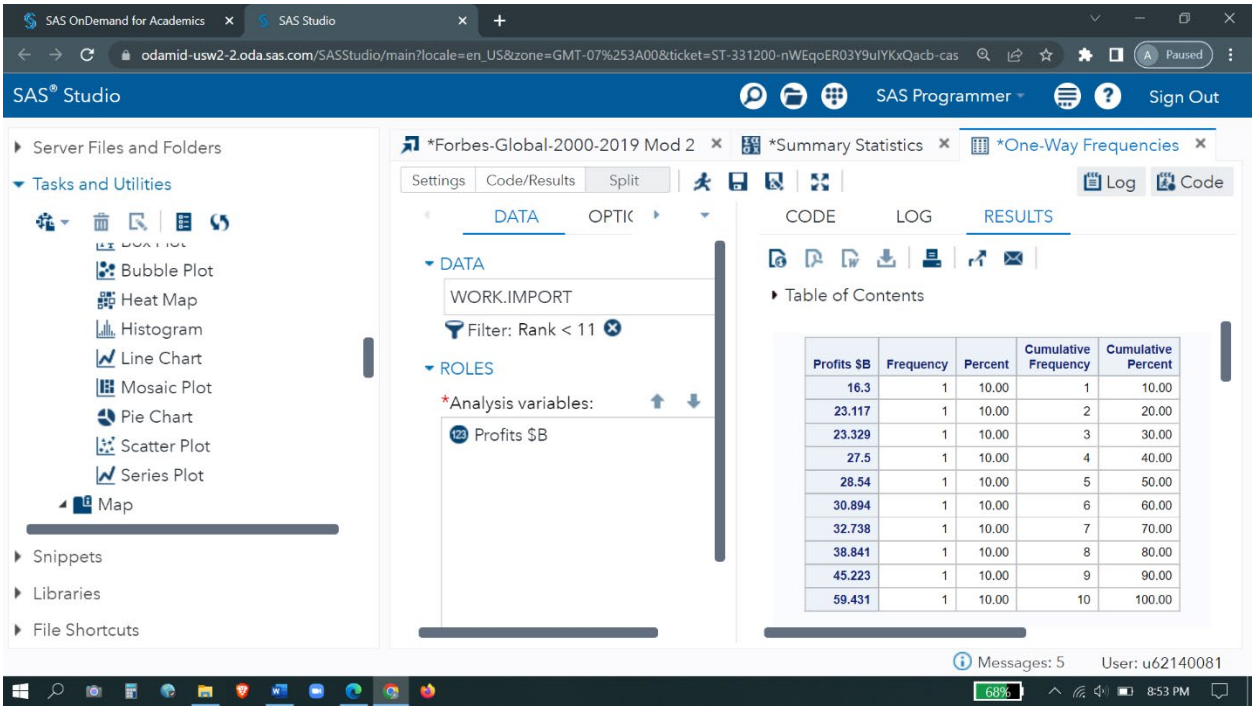


## 7. Statistical Tests

### 7.1 One-way frequency

A "One-Way Frequency" analysis provides insight into one variable under study. This table shows the analysis for profits in $ billion. The graph only displays a set of the top 10 ranked companies, as it will not be possible to display all the rows. The number of value repeated in the profit $B are displayed in the column "Frequency." The full table also reveals that many of the quantities are distinct.

| Profits $B | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| 16.3 | 1 | 10.00 | 1 | 10.00 |
| 23.117 | 1 | 10.00 | 2 | 20.00 |
| 23.329 | 1 | 10.00 | 3 | 30.00 |
| 27.5 | 1 | 10.00 | 4 | 40.00 |
| 28.54 | 1 | 10.00 | 5 | 50.00 |
| 30.894 | 1 | 10.00 | 6 | 60.00 |
| 32.738 | 1 | 10.00 | 7 | 70.00 |
| 38.841 | 1 | 10.00 | 8 | 80.00 |
| 45.223 | 1 | 10.00 | 9 | 90.00 |
| 59.431 | 1 | 10.00 | 10 | 100.00 |



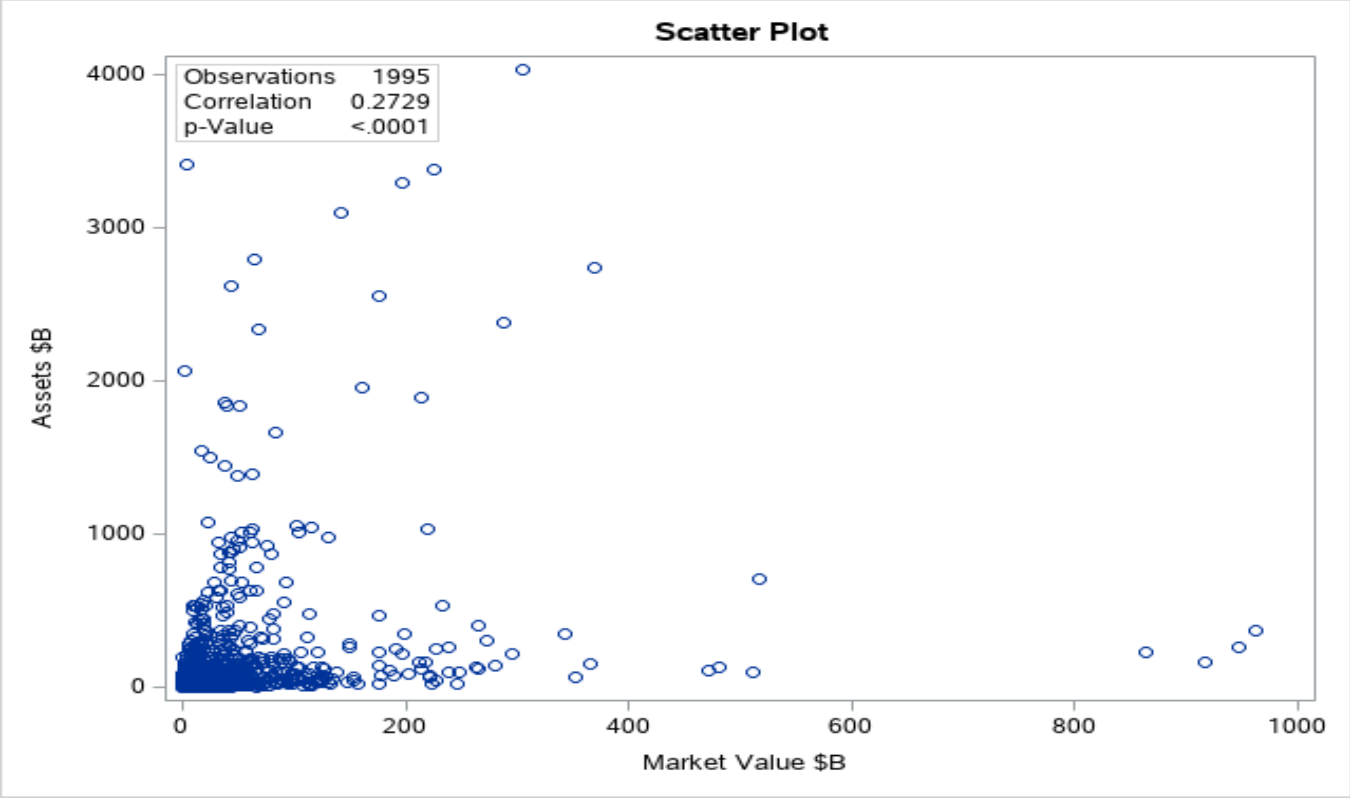

| Profits $B | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| 16.3 | 1 | 10.00 | 1 | 10.00 |
| 23.117 | 1 | 10.00 | 2 | 20.00 |
| 23.329 | 1 | 10.00 | 3 | 30.00 |
| 27.5 | 1 | 10.00 | 4 | 40.00 |
| 28.54 | 1 | 10.00 | 5 | 50.00 |
| 30.894 | 1 | 10.00 | 6 | 60.00 |
| 32.738 | 1 | 10.00 | 7 | 70.00 |
| 38.841 | 1 | 10.00 | 8 | 80.00 |
| 45.223 | 1 | 10.00 | 9 | 90.00 |
| 59.431 | 1 | 10.00 | 10 | 100.00 |

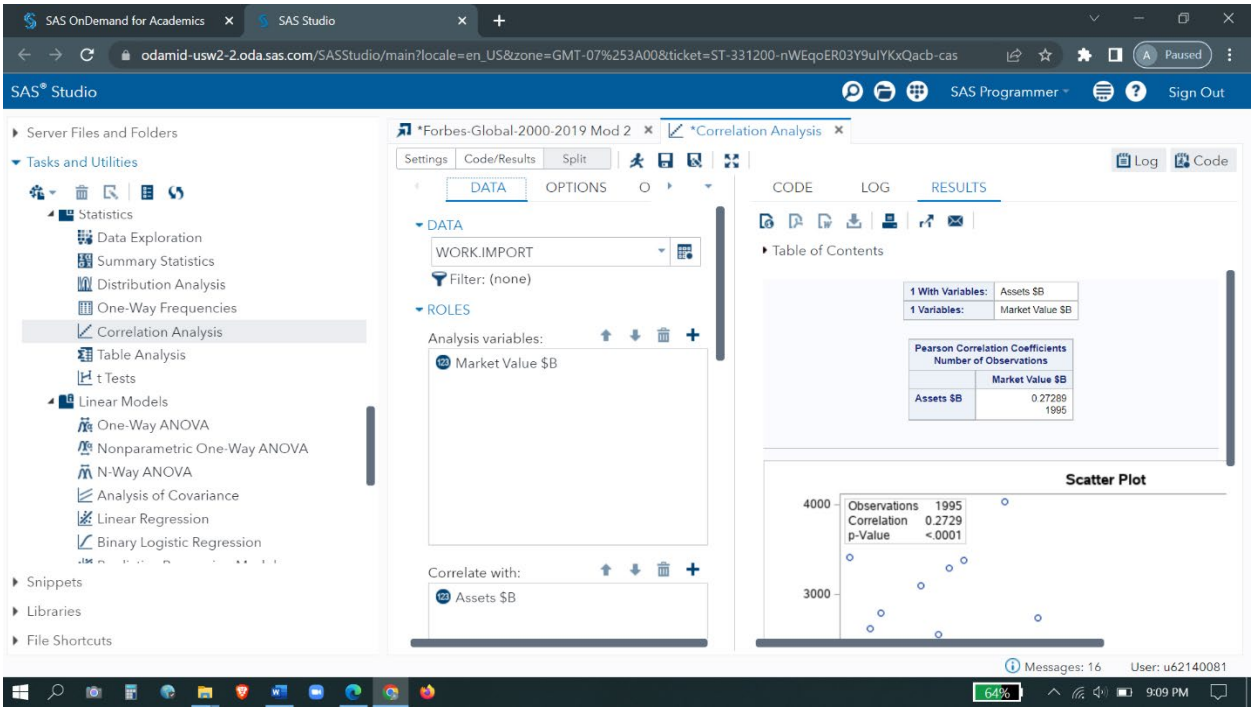
## 7.2 Correlation analysis (Scatter plot chart)

In this correlation analysis, our goal is to determine the Pearson Correlation Coefficient. The sample correlation coefficient typically lies between -1 and 1. This number measures the correlation between the two variables. Based on the Coefficient value of 0.27289, the sample demonstrates a moderate association between Assets and Market value.

| | |
|---|---|
| **1 With Variables:** | Assets $B |
| **1 Variables:** | Market Value $B |

| **Pearson Correlation Coefficients Number of Observations** | |
|---|---|
| | **Market Value $B** |
| **Assets $B** | 0.27289<br>1995 |




SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-331200-nWEqoER03Y9uIYKxQacb-cas
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Statistics
Data Exploration
Summary Statistics
Distribution Analysis
One-Way Frequencies
Correlation Analysis
Table Analysis
t Tests
Linear Models
One-Way ANOVA
Nonparametric One-Way ANOVA
N-Way ANOVA
Analysis of Covariance
Linear Regression
Binary Logistic Regression
Snippets
Libraries
File Shortcuts
*Forbes-Global-2000-2019 Mod 2
*Correlation Analysis
Settings
Code/Results
Split
Log
Code
DATA
OPTIONS
CODE
LOG
RESULTS
DATA
WORK.IMPORT
Filter: (none)
ROLES
Analysis variables:
Market Value $B
Correlate with:
Assets $B
Table of Contents
1 With Variables:
Assets $B
1 Variables:
Market Value $B
Pearson Correlation Coefficients
Number of Observations
Market Value $B
Assets $B
0.27289
1995
Scatter Plot
4000
3000
Observations 1995
Correlation 0.2729
p-Value <.0001
Messages: 16
User: u62140081
64%
9:09 PM

## 7.3 T-test

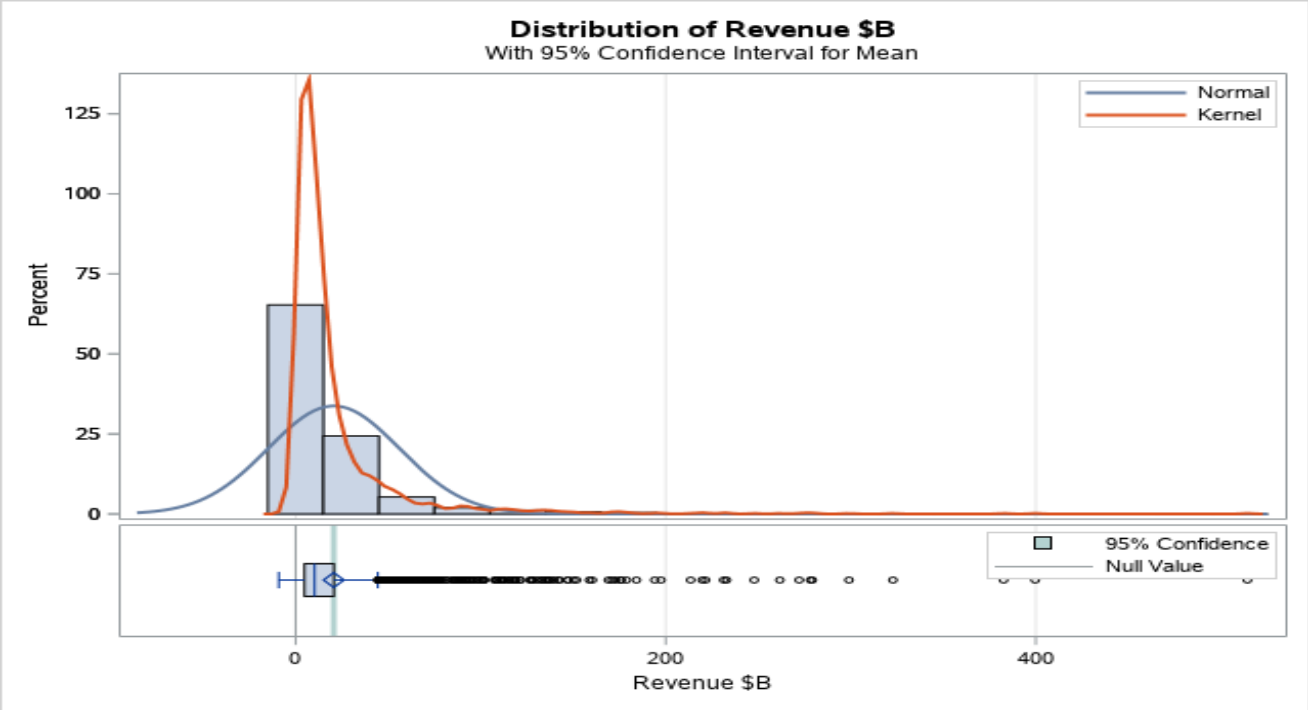
| Description | Explanation |
|---|---|
| Null Hypothesis | Ho= There is no significance of the Investors Count from the mean |
| Alternative Hypothesis | Ha= There is a significance of the Investors Count from the mean |
| Significant Value | Alpha (α) = 0.05 |
| Comparison | 0.01 < 0.05 |
| | We reject the Ho. This indicates that the Revenue is significantly different from the mean. |

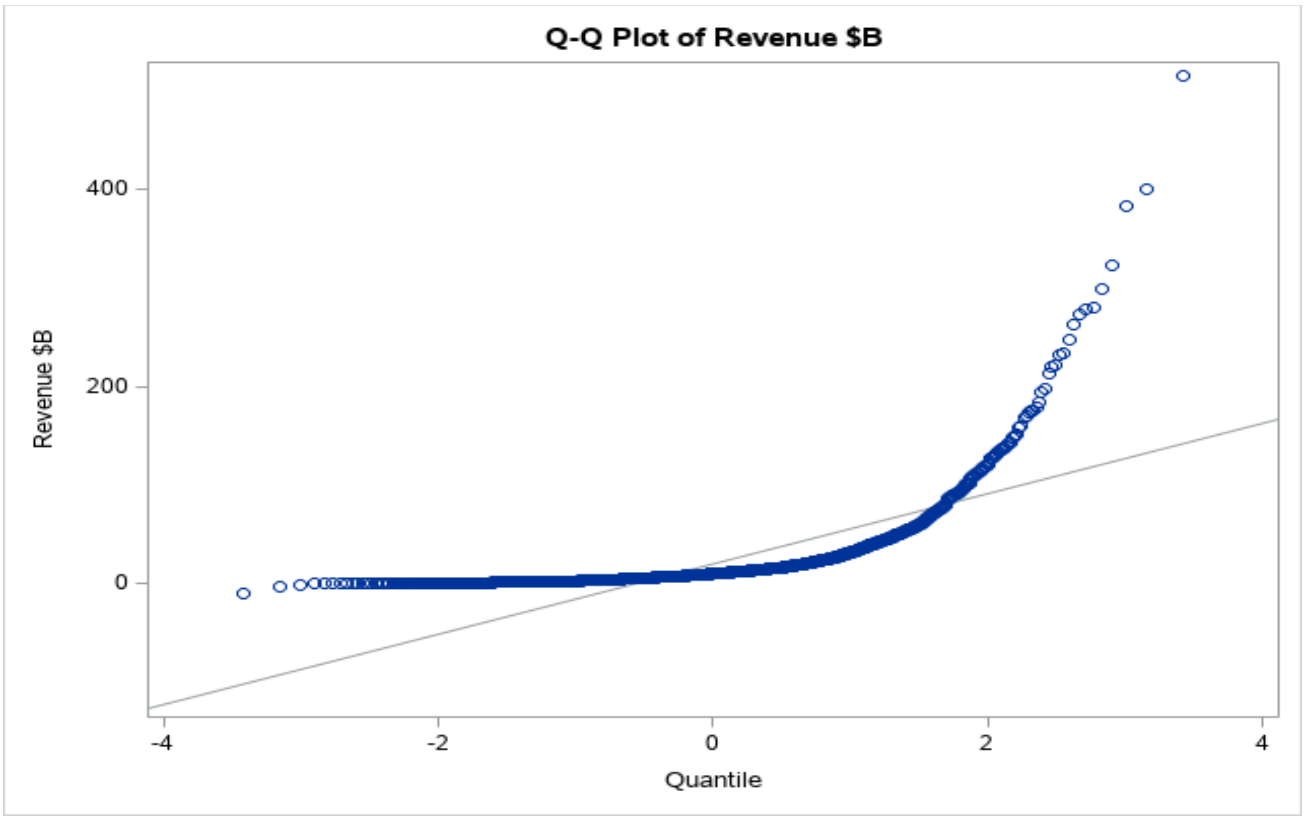
Variable: Revenue $B

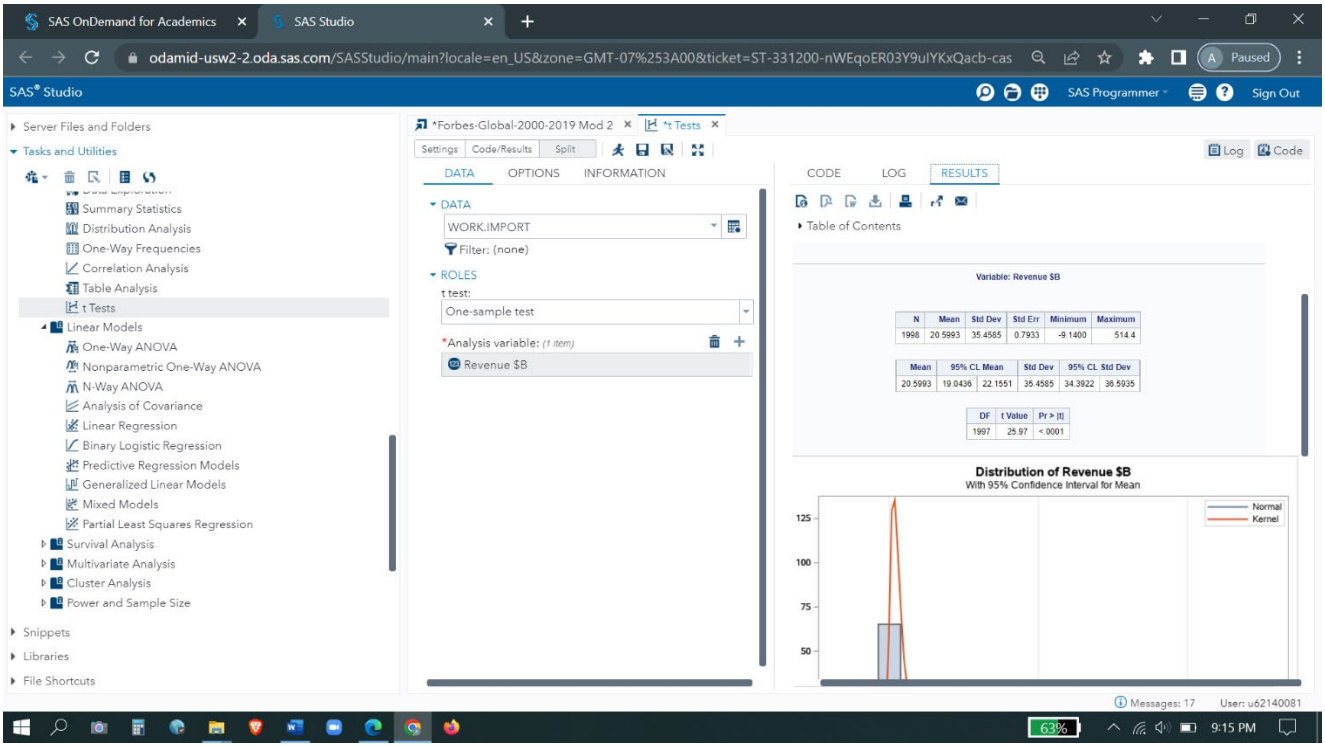
| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|---|---|---|---|---|---|
| 1998 | 20.5993 | 35.4585 | 0.7933 | -9.1400 | 514.4 |

| Mean | 95% CL Mean | | Std Dev | 95% CL Std Dev | |
|---|---|---|---|---|---|
| 20.5993 | 19.0436 | 22.1551 | 35.4585 | 34.3922 | 36.5935 |

| DF | t Value | Pr > \|t\| |
|---|---|---|
| 1997 | 25.97 | <.0001 |




Q-Q Plot of Revenue $B
Revenue $B
400
200
0
-4
-2
0
2
4
Quantile


SAS OnDemand for Academics
SAS Studio
odamid-usw2-2.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-07%253A00&ticket=ST-331200-nWEqoER03Y9ulYKxQacb-cas
SAS® Studio
SAS Programmer
Sign Out
Server Files and Folders
Tasks and Utilities
Summary Statistics
Distribution Analysis
One-Way Frequencies
Correlation Analysis
Table Analysis
t Tests
Linear Models
One-Way ANOVA
Nonparametric One-Way ANOVA
N-Way ANOVA
Analysis of Covariance
Linear Regression
Binary Logistic Regression
Predictive Regression Models
Generalized Linear Models
Mixed Models
Partial Least Squares Regression
Survival Analysis
Multivariate Analysis
Cluster Analysis
Power and Sample Size
Snippets
Libraries
File Shortcuts
*Forbes-Global-2000-2019 Mod 2
*t Tests
DATA
OPTIONS
INFORMATION
WORK.IMPORT
Filter: (none)
ROLES
t test:
One-sample test
*Analysis variable:
Revenue $B
CODE
LOG
RESULTS
Table of Contents
Variable: Revenue $B
Distribution of Revenue $B
With 95% Confidence Interval for Mean
Normal
Kernel
Messages: 17
User: u62140081
63%
9:15 PM

## 8. References


i. "andrea Murphy", " I. C. (n.d.). *The global 2000 2022*. Forbes. Retrieved November 2, 2022, from https://www.forbes.com/lists/global2000/?sh=11aaea0e5ac0

ii. Ponciano, J. (2019, May 19). The world's Largest Public Companies 2019: Global 2000 by the numbers. Forbes. Retrieved November 2, 2022, from https://www.forbes.com/sites/jonathanponciano/2019/05/15/worlds-largest-companies-2019-global-2000/?sh=7cdd7ed64ada

iii. M, S. (2022, October 21). Top 8 excel data cleaning techniques to know in 2022: Simplilearn. Simplilearn.com. Retrieved November 2, 2022, from https://www.simplilearn.com/tutorials/excel-tutorial/excel-data-cleaning

iv. Johnson, D. (2022, September 17). SAS tutorial for beginners: What is & programming example. Guru99. Retrieved November 2, 2022, from https://www.guru99.com/sas-tutorial.html