11/11/2018

New York STOCK EXCHANGE

Pankti Parikh
306613767

A. **Dataset URL :**

Main dataset URL

https://www.kaggle.com/dgawlik/nyse

Related Website Link

https://www.kaggle.com/dgawlik/nyse#securities.csv

B. **Data Insights:**

- In present situation Where Stock market is going up and down, it is necessary to invest consciously in the market whatever it is, the findings and trends of the New-York Stock markets and shares which enables the investors in taking decision regarding investments.
- A stock exchange is the platform where financial instruments like stocks and derivatives are traded. Market participants must be registered with the stock exchange and NYSE to conduct trades. This includes companies issuing shares, brokers conducting the trades, as well as traders and investors. All of this is regulated by the New-York Stock Exchange, which makes the rules of conduct.
- The New York Stock Exchange is still by far the most important equity market in the world. With a market cap of about $21 trillion, the NYSE about three times larger than the NASDAQ, and the two US exchanges together have a larger market cap than the next ten exchanges combined.
- The New York Stock Exchange provides a means for buyers and sellers to trade shares of stock in companies registered for public trading.

- NYSE stocks can be traded via its electronic hybrid market (except for a small group of very high-priced stocks). Customers can now send orders for immediate electronic execution, or route orders to the floor for trade in the auction market.
- NYSE works with US regulators like the SEC and CFTC to coordinate risk management measures in the electronic trading environment through the implementation of mechanisms like circuit breakers and liquidity replenishment points.

C. **Data description:**

- New-York Stock exchange dataset from Kaggle includes four different CSV files as following:

1. **Fundamentals.csv:** most of popular fundamental indicators are derive from annual SEC 10 K filling (2012-2016), Prices are fetched from Yahoo Finance. Whereas fundamentals are fetched from Nasdaq Financials. There are 78 columns in this CSV file describes different properties of company's share and earnings.
2. **Prices-Split-Adjusted.csv:** There have been added adjustments for splits which was same as price.
3. **Prices.csv:** raw, as-is daily prices. Most of data spans from 2010 to the end 2016, for Companies new on stock market date range are shorter. There have been approx. 140 stock Splits in that time, this set doesn't account for that.
4. **Securities.csv:** Each company with division on sectors in general description. Includes

D. **Data Cleaning:**

Here are several key benefits that come out of the data cleaning process:

- It removes major errors and inconsistencies that are inevitable when multiple sources of data are getting pulled into one dataset.
- Using tools to cleanup data will make everyone more efficient since they'll be able to quickly get what they need from the data.
- Fewer errors means happier customers and fewer frustrated employees.
- The ability to map the different functions and what your data is intended to do and where it is coming from your data.

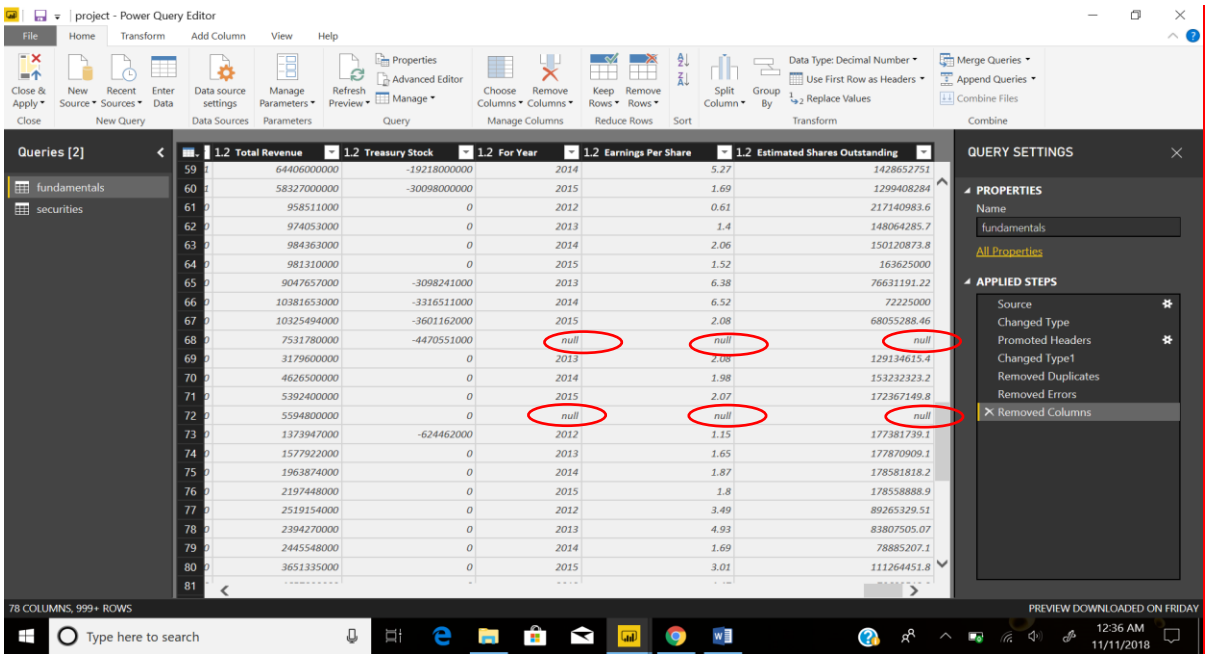
Categories:

1. **Missing values:**

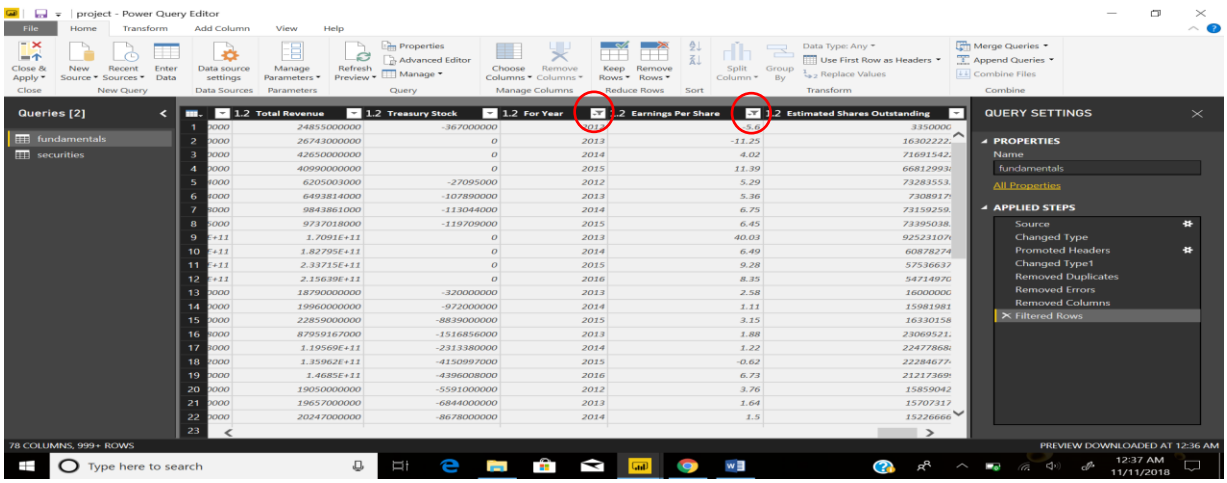
- The goal of such cleaning operations is to prevent problems caused by missing data that can arise when creating a Visuals. This module supports multiple type of operations for "cleaning" missing values, including:
    - Replacing missing values with a placeholder, mean, or other value.
    - Completely removing rows and columns that have missing values.
    - Inferring values based on statistical methods

- As, you can see in following figure. That, I used 2nd type of cleaning by completely removing rows and columns that have missing values .

Before

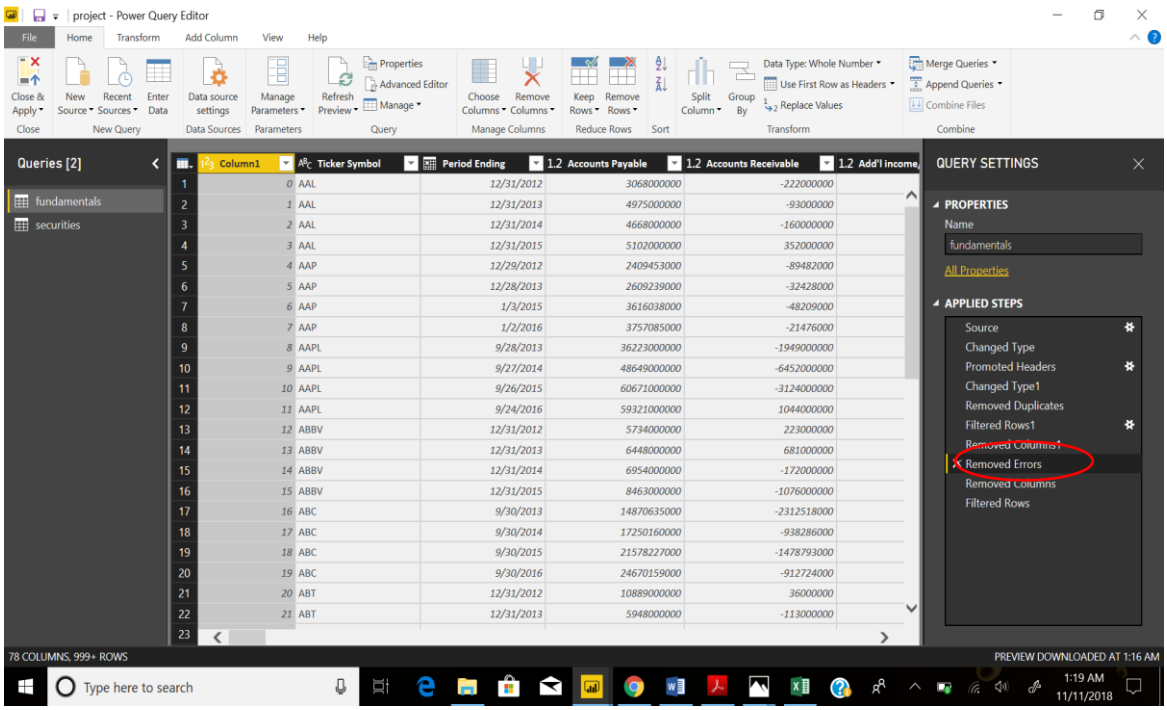


After



2. Erroneous values :

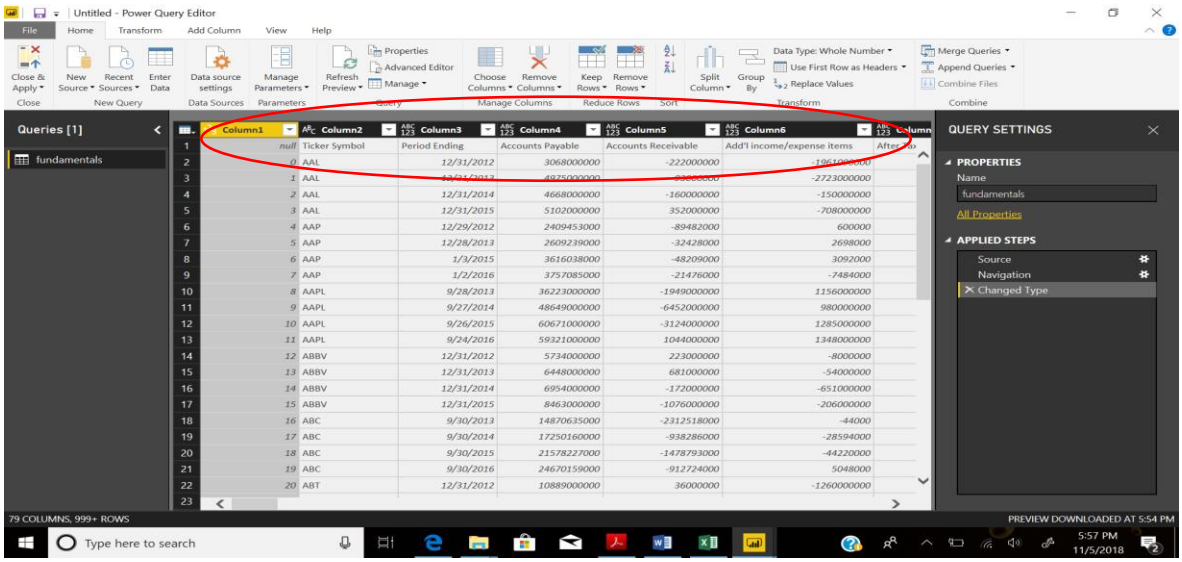
- removing typographical errors or validating and correcting values against a known list of entities. The validation may be strict or fuzzy (such as correcting records that partially match existing, known records). Some data cleansing solutions will clean data by cross checking with a validated data set.
- You can see In the following screenshot error is removed by selecting remove error option from the table properties option.



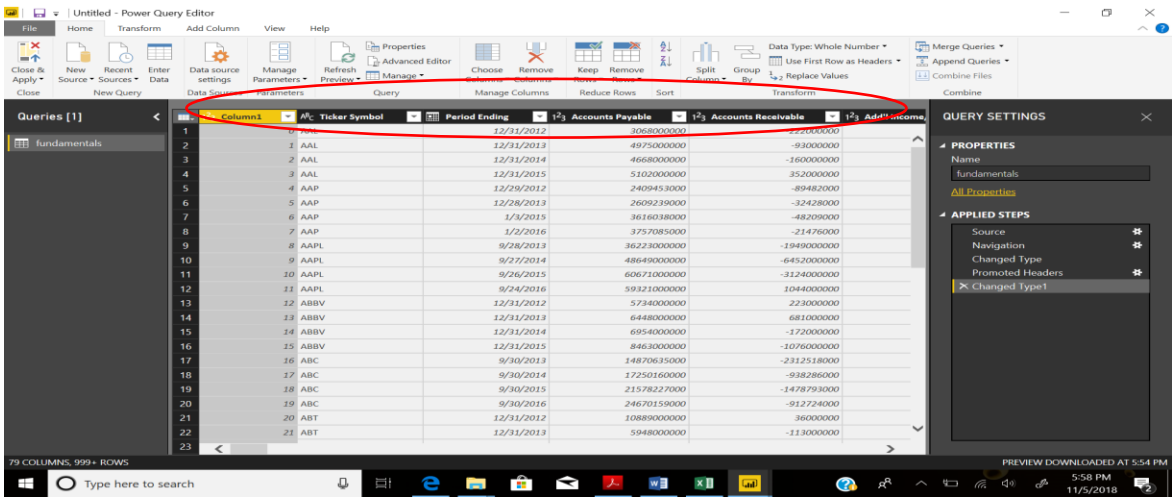
3. Inconsistencies:

- Sometimes data contains some inconsistent values like here, when I inserted data from the csv file it shows that column headers are the elements considered as the first row of the tables. So, I chose the table properties option and selected option called choose first row as the header.

Before

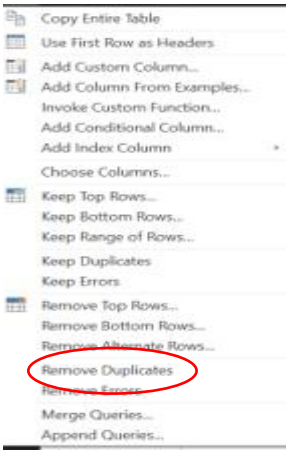


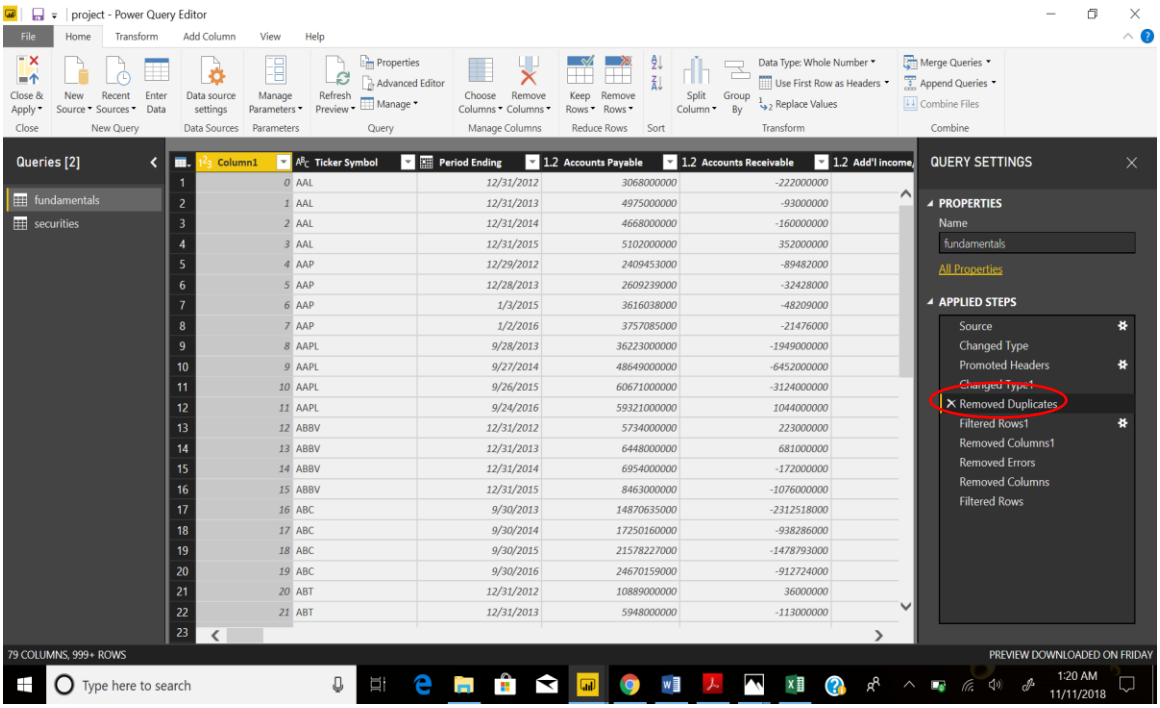
After



4. Duplicate records:

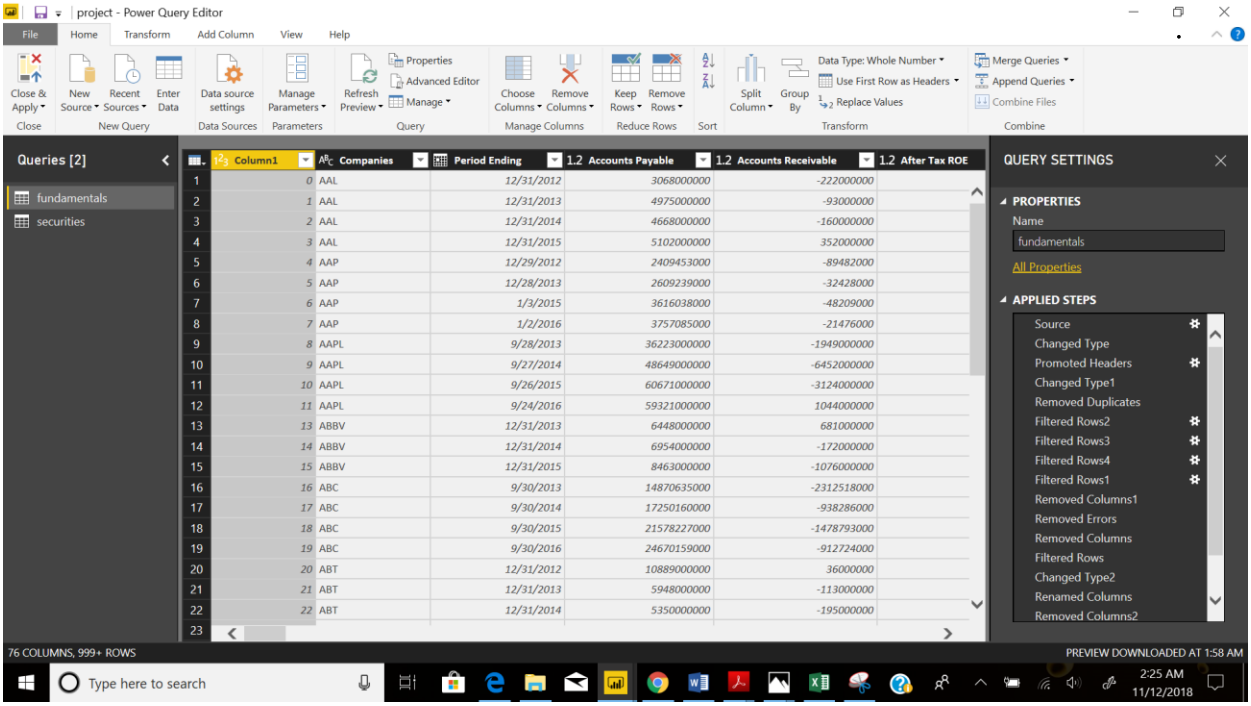
- Duplicate rows are a common problem when you import data. It is a good idea to filter for unique values first to confirm that the results are what you want before you remove duplicate values. By choosing following step you can remove duplicate data from the table.





5. Out of date :

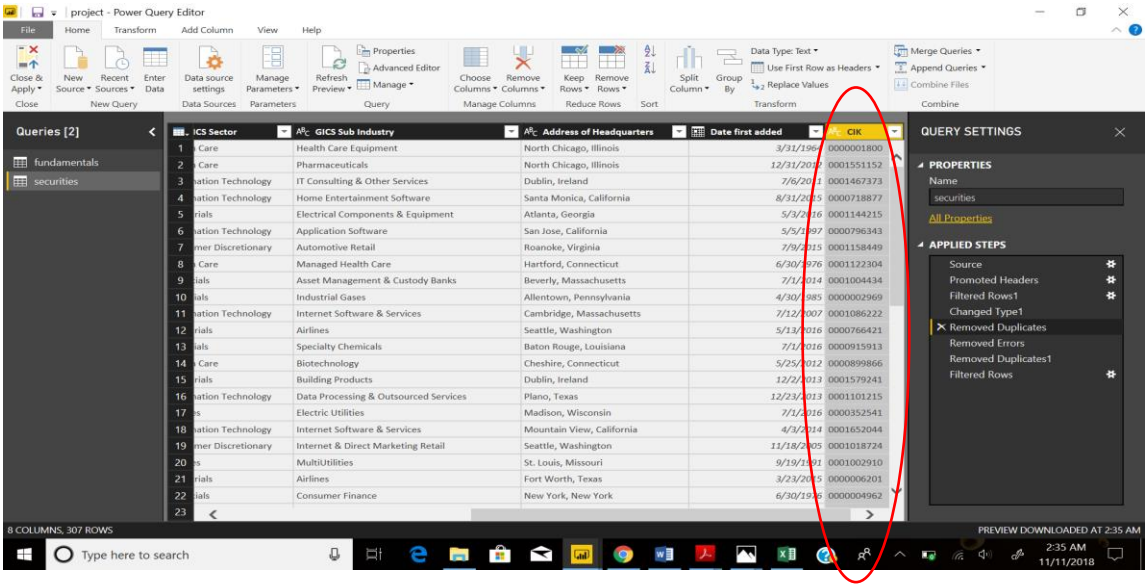
- Means Data having old format or Data Type. No one wants to use the data which is out of the date or old format. But luckily in my dataset not a single column have this kind of problem.



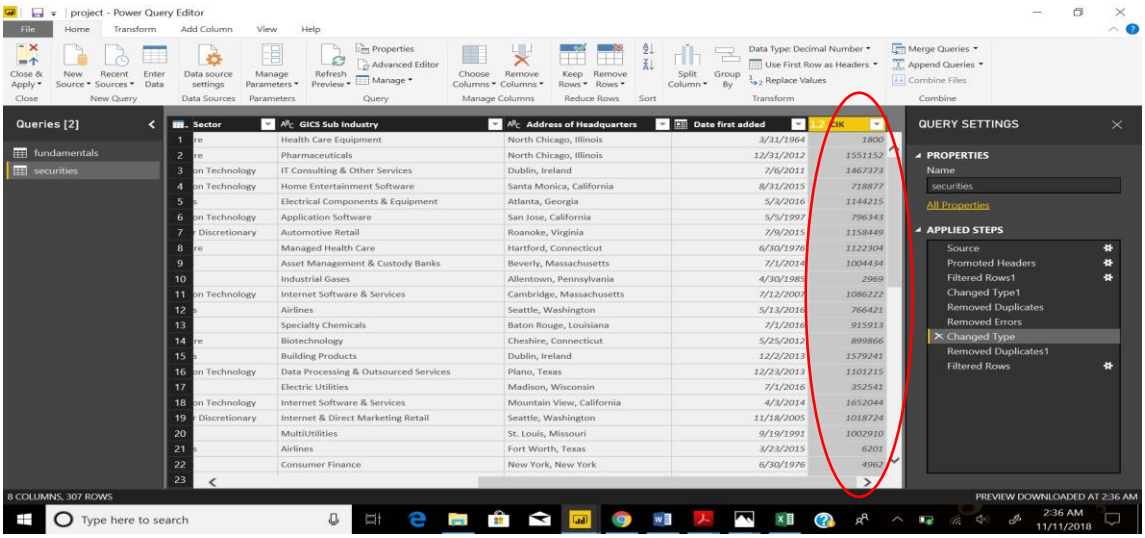
6. Leading or trailing issues :

- Having something before or after the original value of the data called leading or trailing problem in my dataset one column called CIK have certain numeric values and that values have zeroes before the actual value. And this data set have this column as a text by default. But by changing datatype from text to whole numbers of that column I removed this problem.

Before



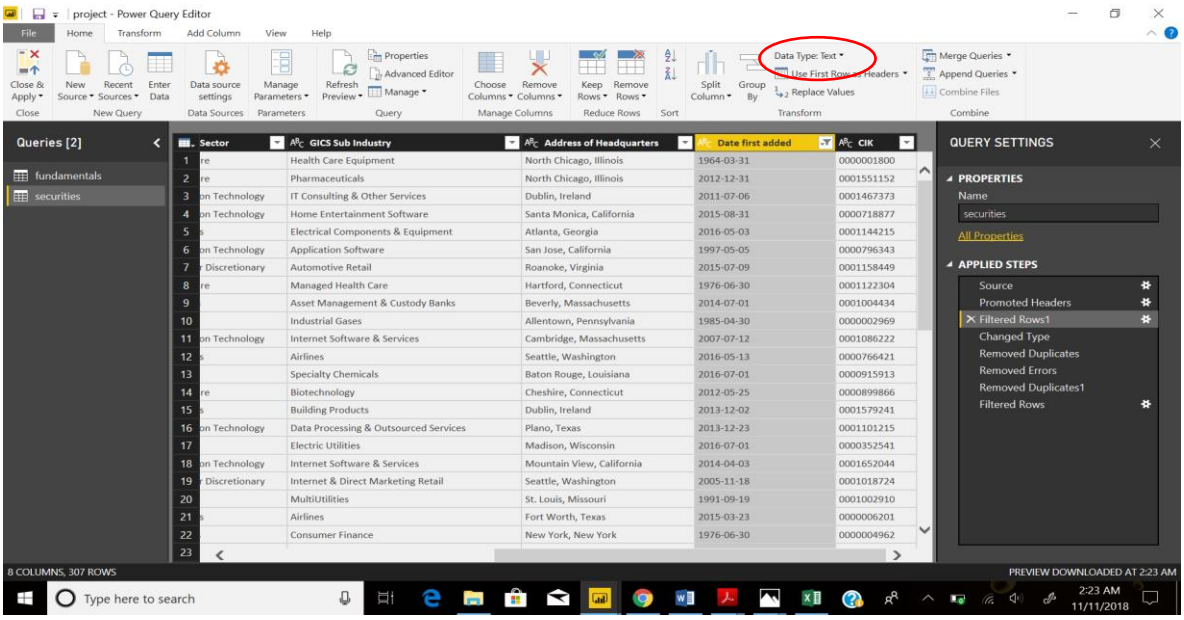
After



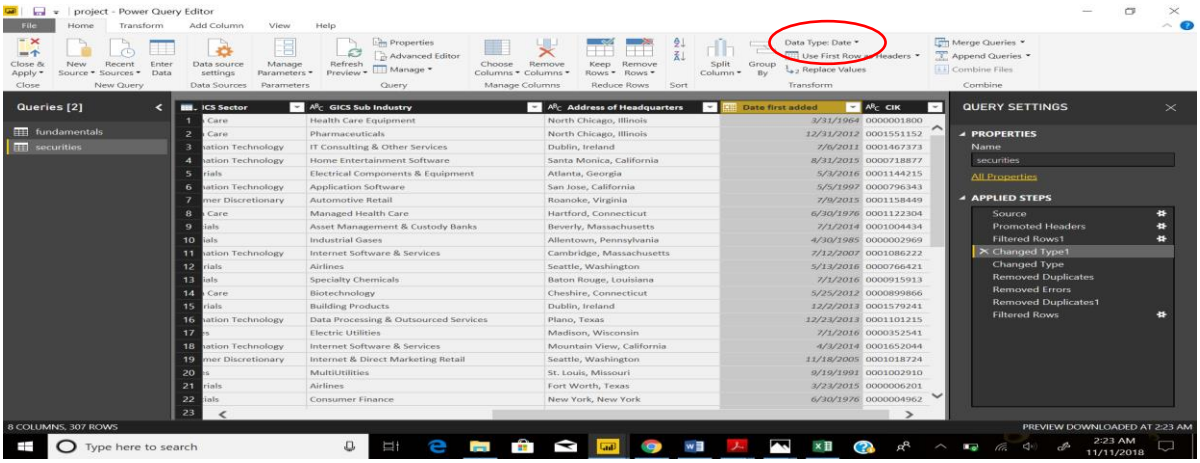
7. **Date format**:

- Dates are always in mm-dd-yyyy or date or date/time format by sometimes raw data contain other data types as a by default like in this dataset one column named Date first added have text date type by default so by correcting that I changed that data type to mm/dd/yyyy format.you can see before and after snapshot as a following.

Before

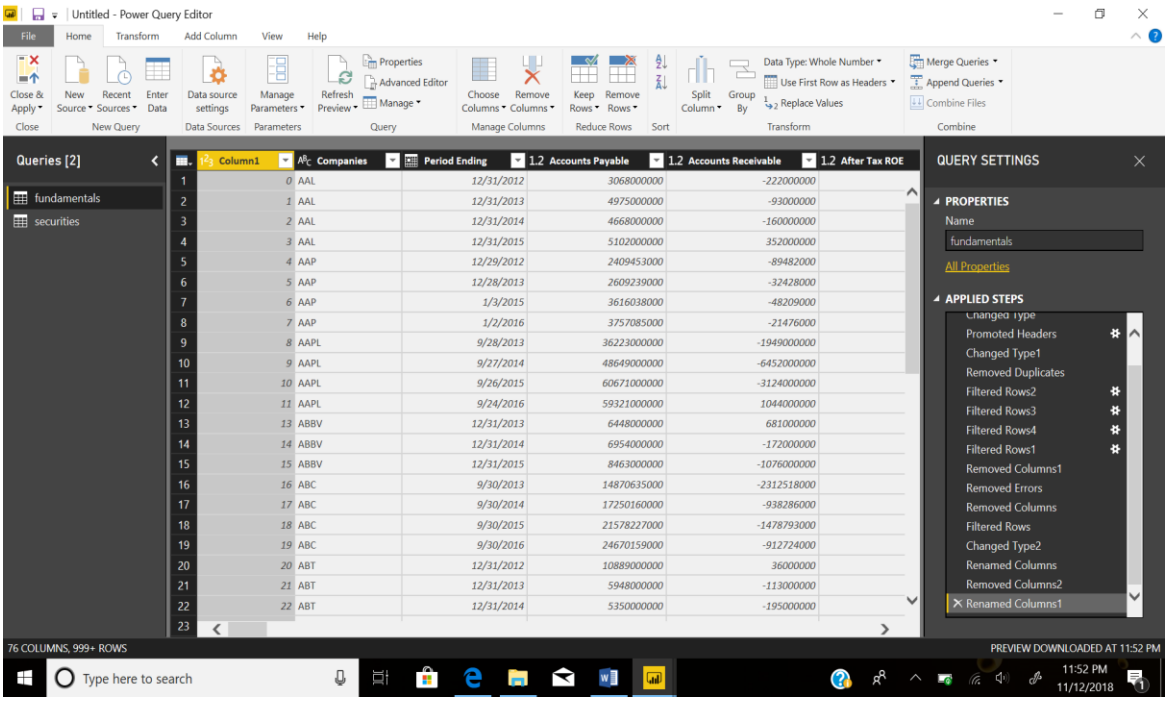


After



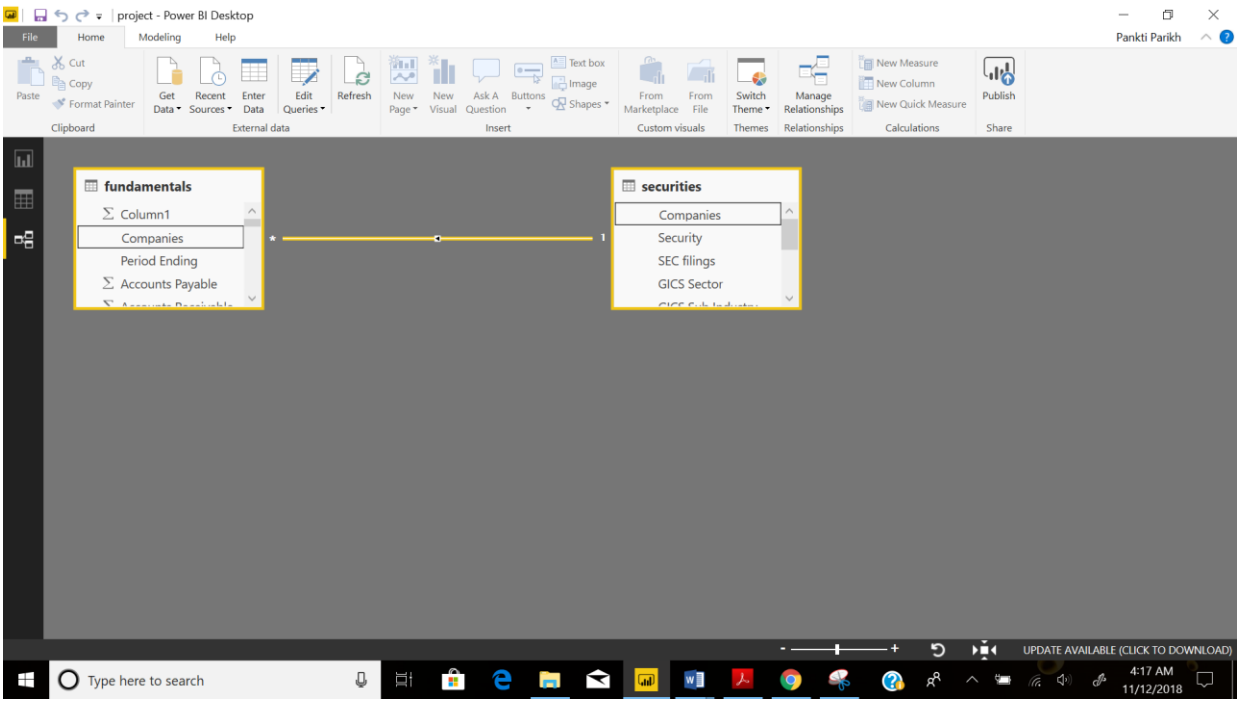
8. Other Data Type:

- This represents the uncommon system characters or line breaks. Any uncertain entry in the columns or rows which is not suitable for the selected datatype. This can be overcome by replacing the error values or removing the complete row or column.



E. **Relationships between Tables:**

- If we query two or more tables at the same time, when the data is loaded, Power BI Desktop will attempt to find and create relationships for us. Cardinality, Cross filter direction, and Active properties are automatically set.
- Power BI Desktop looks at column names in the tables we are querying to determine if there are any potential relationships. If there are, those relationships are created automatically. If Power BI Desktop cannot determine with a high-level of confidence there is a match, it will not automatically create the relationship. we can still use the Manage Relationships dialog to create or edit relationships.



- Here, you can see in the snapshot that two tables are connected through the common column called Companies.

- In this Relationship Companies column of Fundamentals became Primary key while Companies column of securities called Foreign key.

F. Measure:

Here, I created one measure called Net_Income.

Following , one formula I used in visual to calculate Net Income :

EBIT = Net_income + Interest expense + Tax expense

```
Net_Income = sum(fundamentals[Earnings Before Interest and Tax])-(sum(fundamentals[Interest Expense])+sum
(fundamentals[Income Tax]))
```

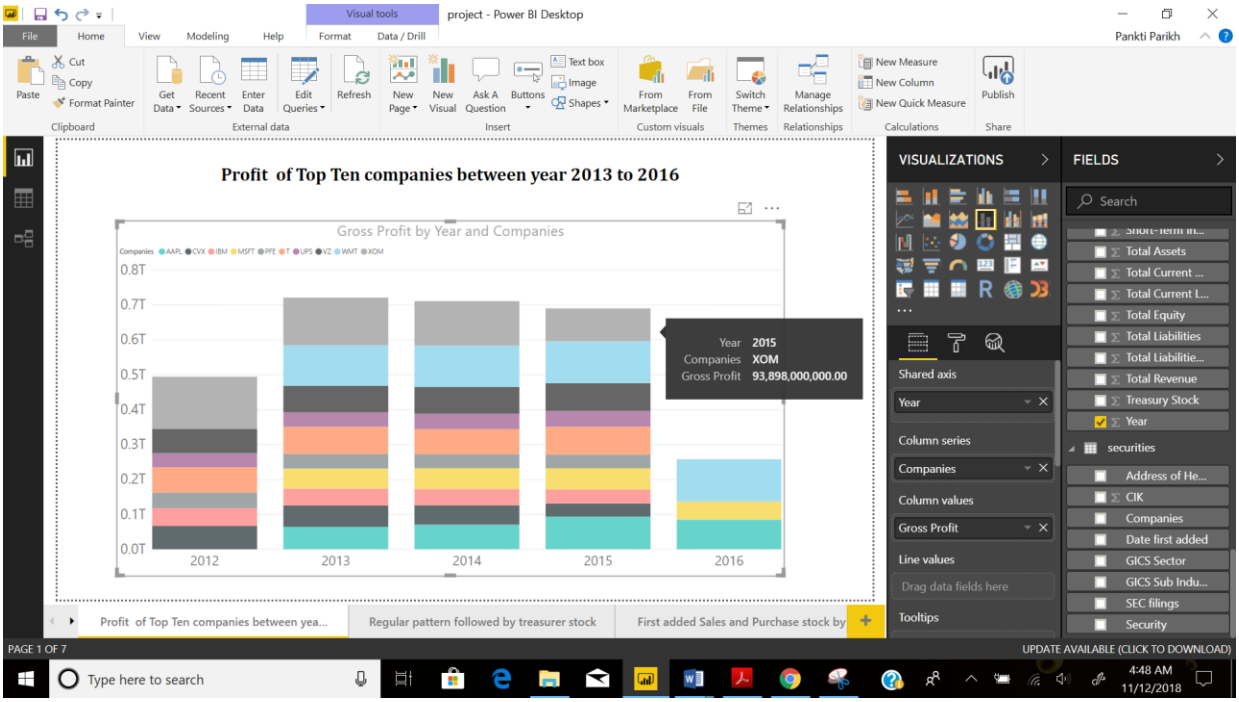
G. Visuals and Dashboard :

1. Stake Column Chart:

- Stake charts are best for comparing data that is grouped by discrete categories. Stake bar charts are best when you don't have too many groups (less than 10 is usually good). Each bar is separated by blank space which indicates that there is no inherent order to your groups.

Question 1 : How much profit made by top ten companies between year 2012 to 2016?

- Here, we can see Profit of Top Ten companies Between year 2012 to 2016. I took Gross Profit in Y-axis, and Years in X-axis. Also took Companies in Column series having filter for top ten values.
- Top Companies made less profit during year 2016 and made more profit in 2013.



2. Line Chart :

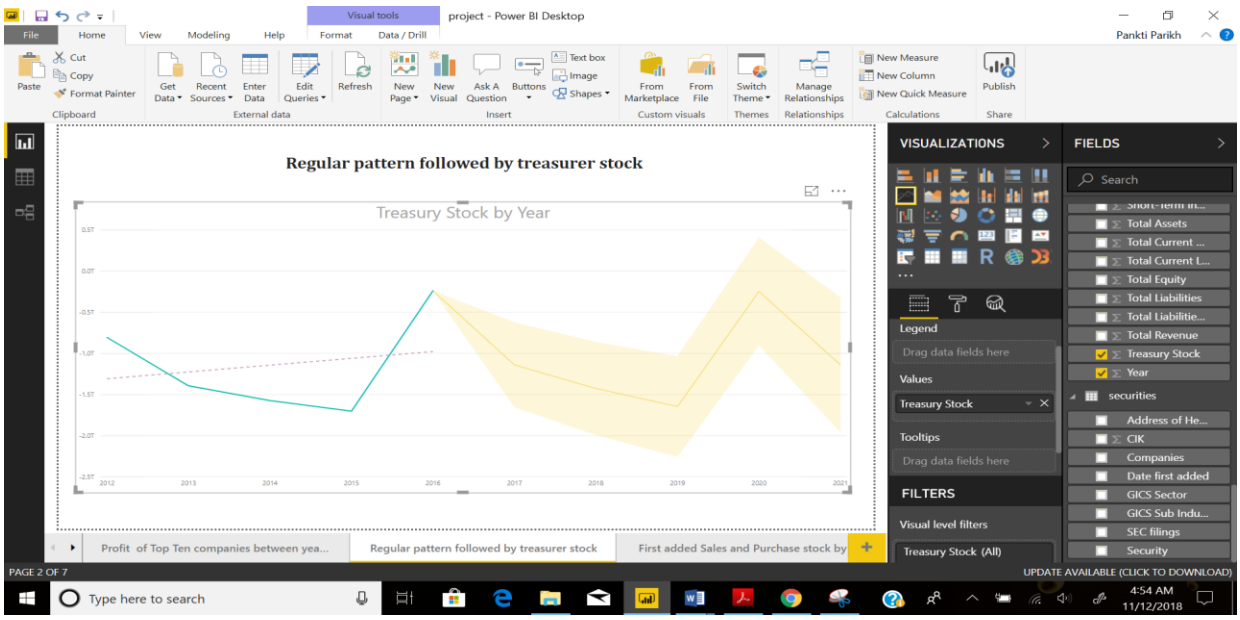
- Line charts are used to show resulting data relative to a continuous variable - most commonly time or money. They are great for projections of performance beyond your data. If you plot your sales vs. month on a line chart over the past two years, it is easy for the reader to identify any trends that may be useful as you plan for the upcoming year.
- A dual axis chart allows you to plot data using two y-axes and a shared x-axis. It's used with three data sets, one of which is based on a continuous set of data and another which is better suited to being grouped by category. This should be used to visualize a correlation or the lack thereof between these three data sets.

Question 2: Is there any regular pattern followed by treasurer stock?

- Y-axis represent treasurer Stock. X-axis represents year given in the dataset (2012-2016) and also have future trend line for years 2016 to 2021 having upper bound and lower bound

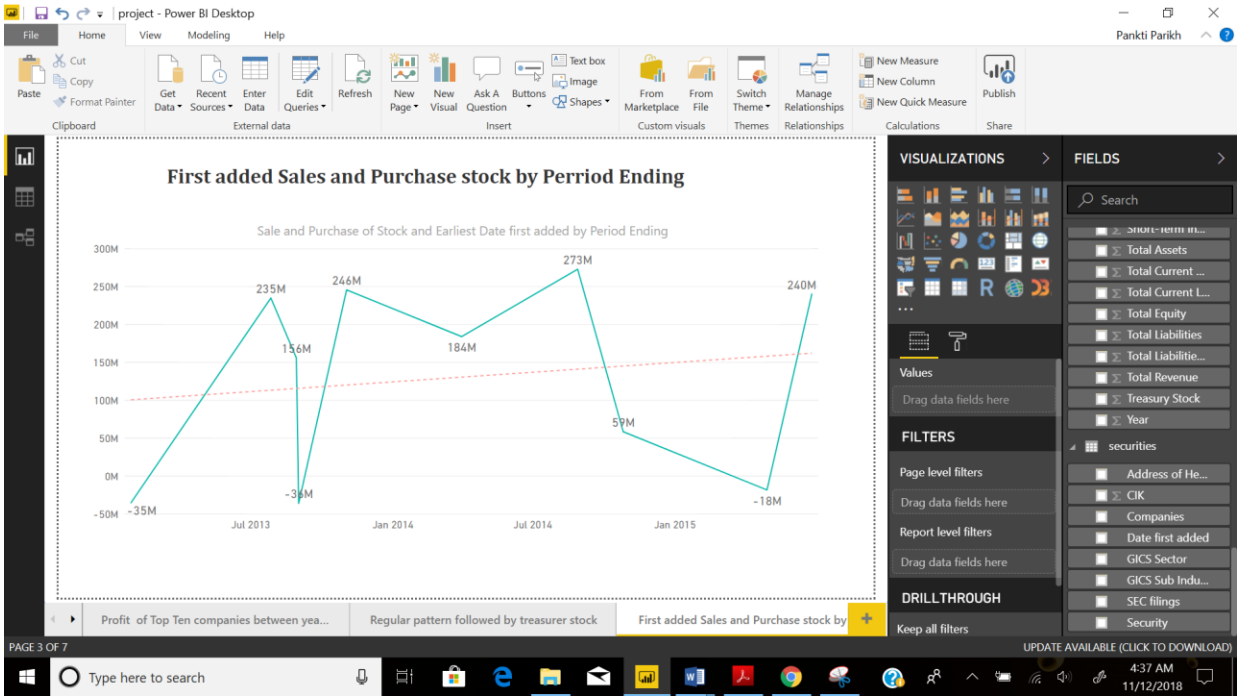
Values. I created forecast trend line using tooltip and also have trend line for year 2012 to 2016 so that we can see specific pattern followed in this year And predict the future trend line using different points.

- In future, Stock price of the treasurer stock will decrease during year 2016 to 2019 and then in between year 2019 and 2020 prices will be increase almost nearest to the stock price of year 2016.



3. Line Chart having Trend Line:

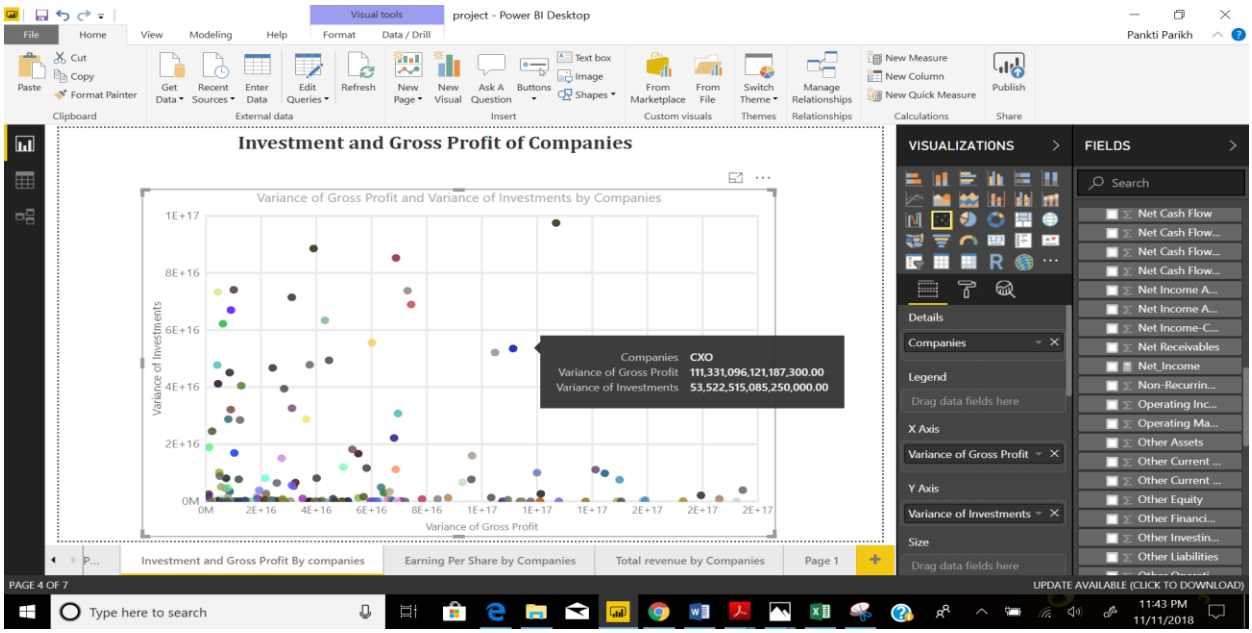
- Following line chart shows the sales and product stock added earliest before the ending period. You can see the sales and product stock prices fluctuation during first end period Jan to July and second end period July to Jan of next year.
- Here, you can see more fluctuation in the stock price during July 2013 to Jan 2014. And lowest fluctuation between year Jan 2014 to July 2014.



4. Scatter Chart :

- A scatter plot chart will show the relationship between two different variables. Scatter plots are useful for quickly understanding if there is a relationship between to variables. If the data forms a band extending from lower left to upper right, there most likely a positive correlation between the two variables. If the band runs from upper left to lower right, a negative correlation is probable. If it is hard to see a pattern, there is probably no correlation.

- This scatter plot portrayed Correlation between Variance of Gross Profit and variance of Investment made by Companies.



5. Pie Chart :

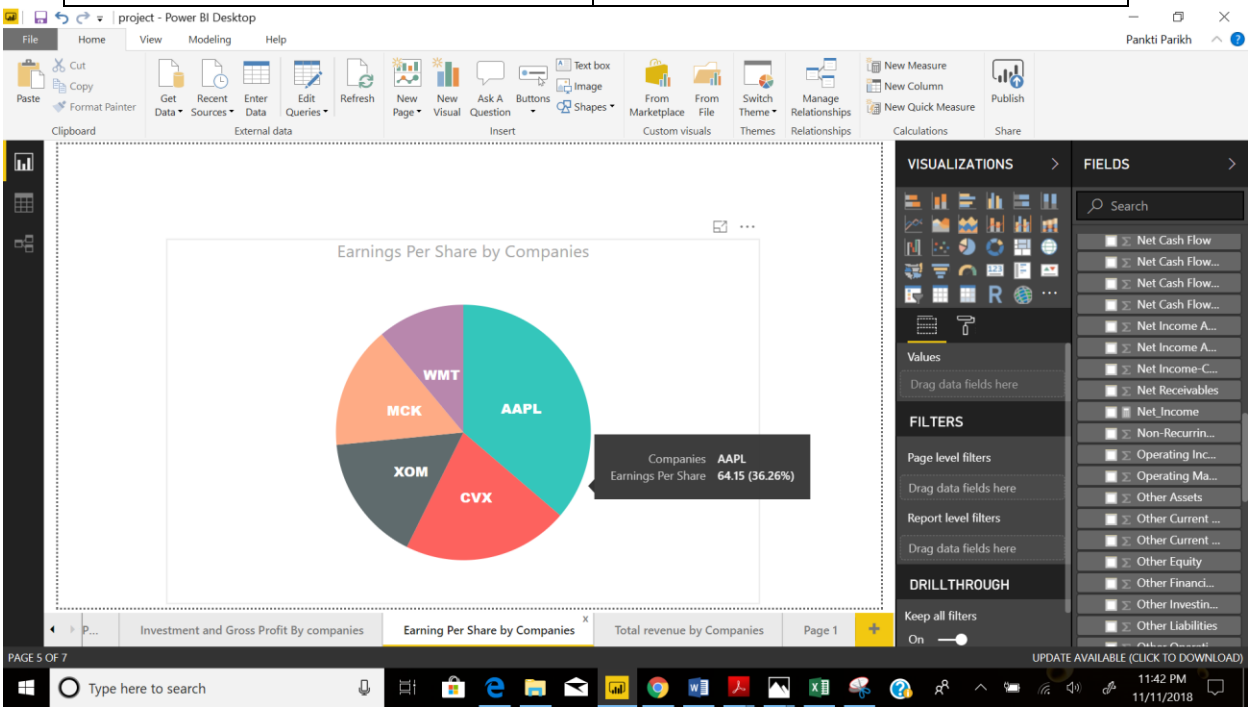
- Pie charts are easy to read and fun to look at making them a great choice if you want to understand the parts of a whole. It's a good practice to order the pieces of your pie according to size an always ensure the total of all the pieces add up to 100%.

Question 3 : Display the top 5 company's Earning per share.

- As you see Earning per share of top 5 companies. Highest Earning made per share by AAPL which is 36.26% having 64.15 shares .

Company Name	Earning per share (in percentage)
AAPL	36.26%
CVX	21.06%

XOM	16.09%
MCK	15.27%
WMT	11.07%

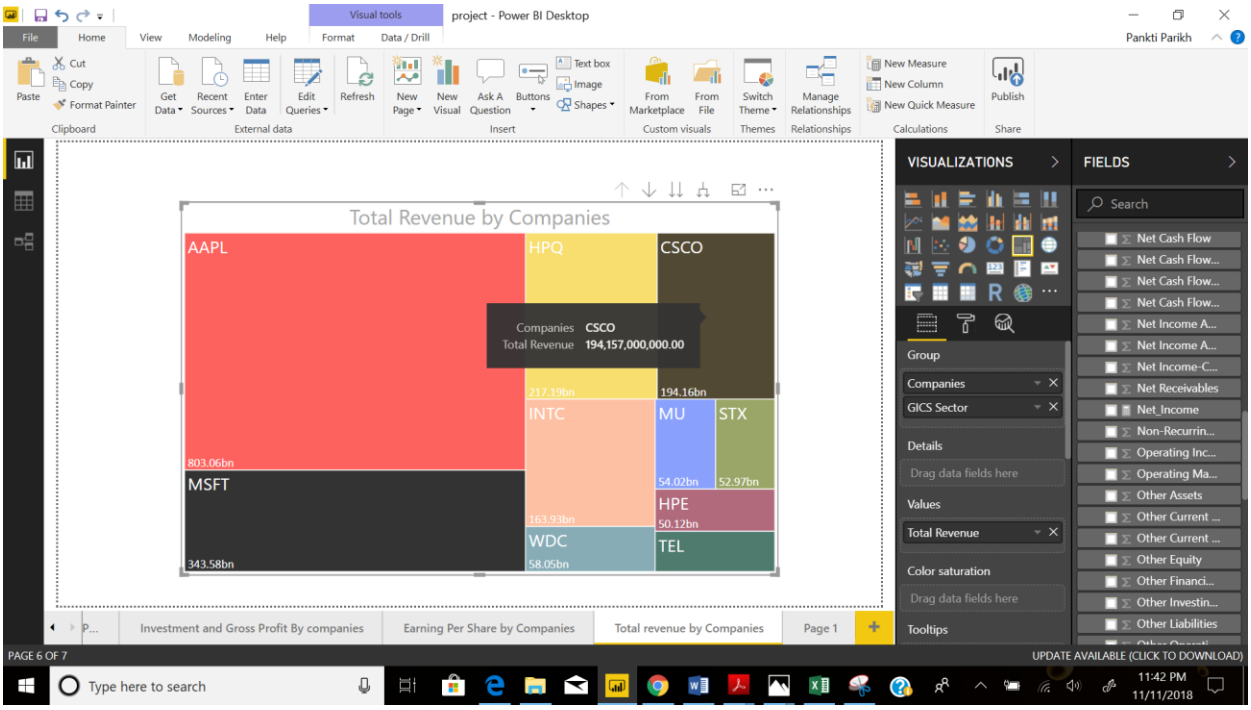


6. Tree Map :

- Treemapping is a data visualization technique that is used to display hierarchical data using nested rectangles; the treemap chart is created based on this technique of data visualization.

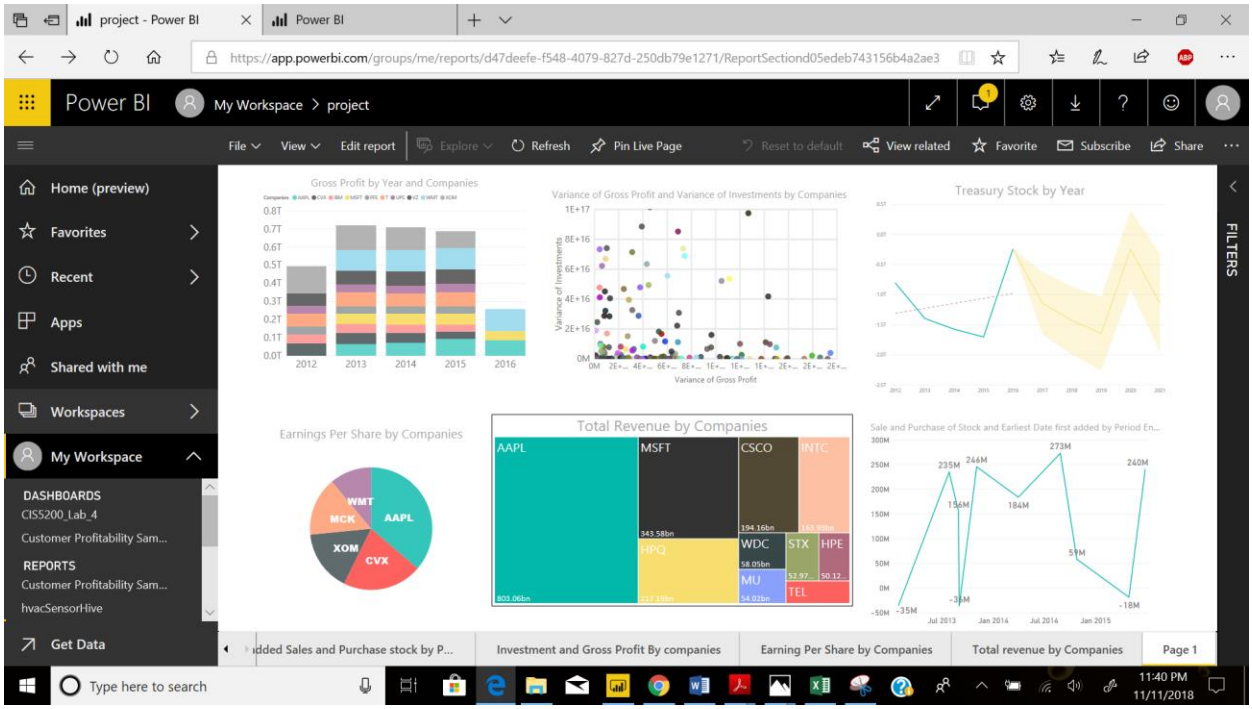
Question 4 : Total revenue generated by Top 10 IT companies.

- There are top 10 IT compnies filtered in following Tree map in which AAPL company made highest revemue because in previous visual you can see they have more Earning per share so they made more revenue 806.03 Billoions**.**



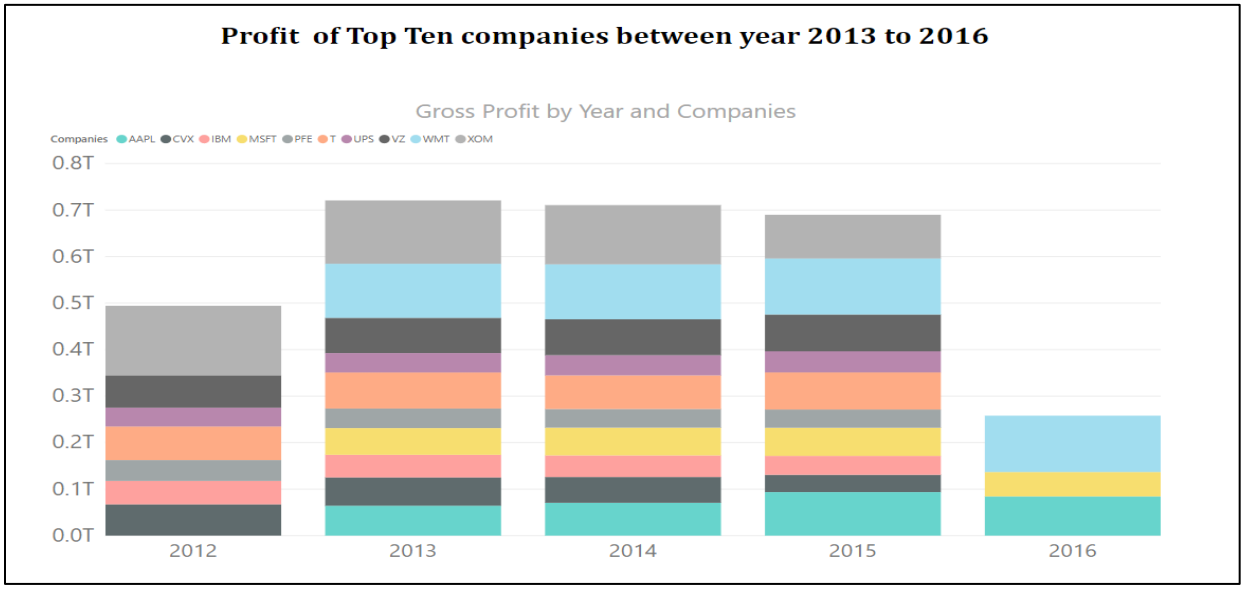
H. Dash Board and Story Telling:

- A Power BI dashboard is a single page, often called a canvas, that uses visualizations to tell a story. Because it is limited to one page, a well-designed dashboard contains only the most-important elements of that story.



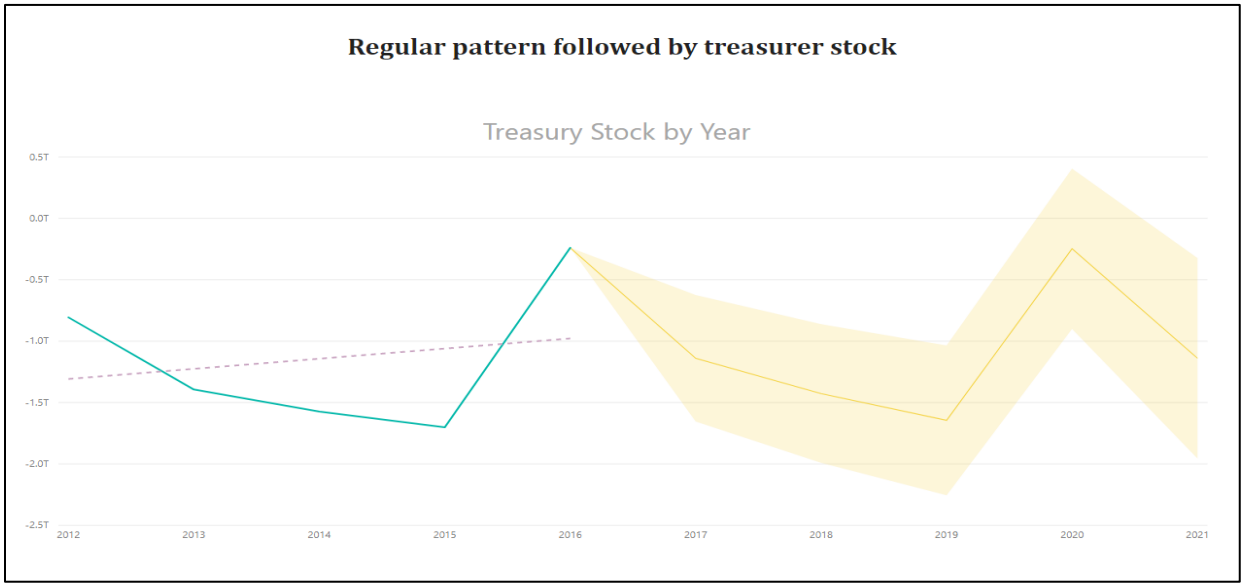
Story Telling:

Chart 1:



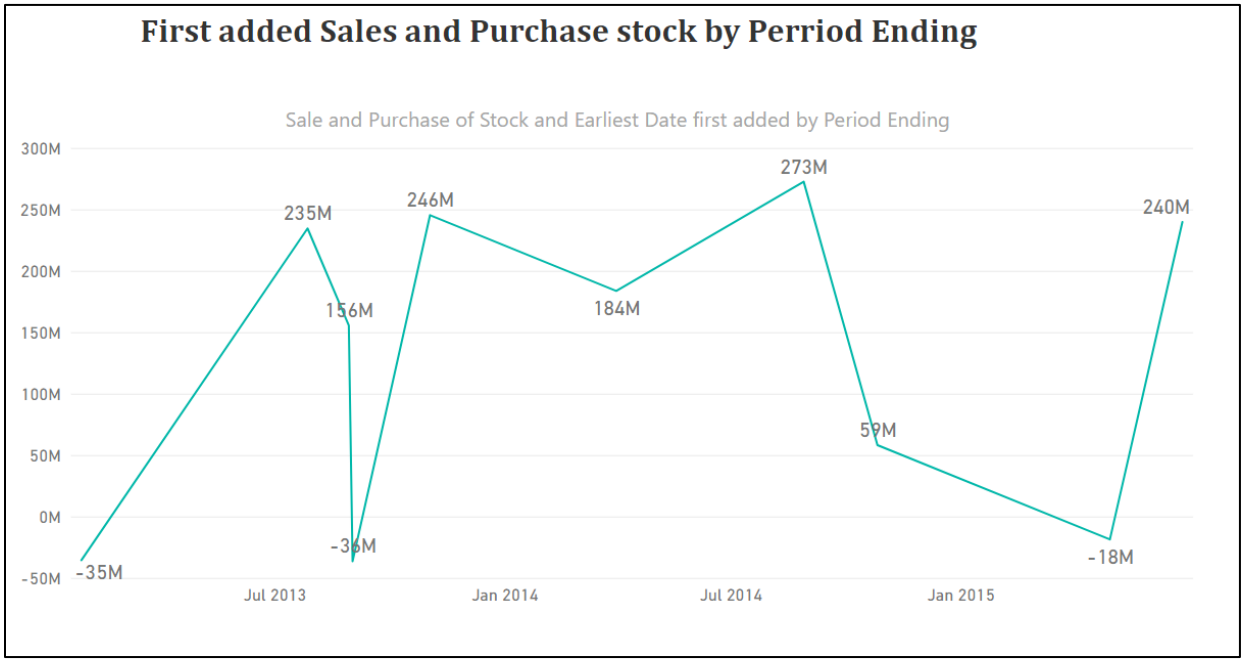
- From above visual we can conclude that only 3 companies out of top 10 companies are able to stay on top [profit wise] in year 2016 compare to last three years.
- WMT and XOM Companies are on top continuously by making more profit during 2012 to 2016.

Chart 2:



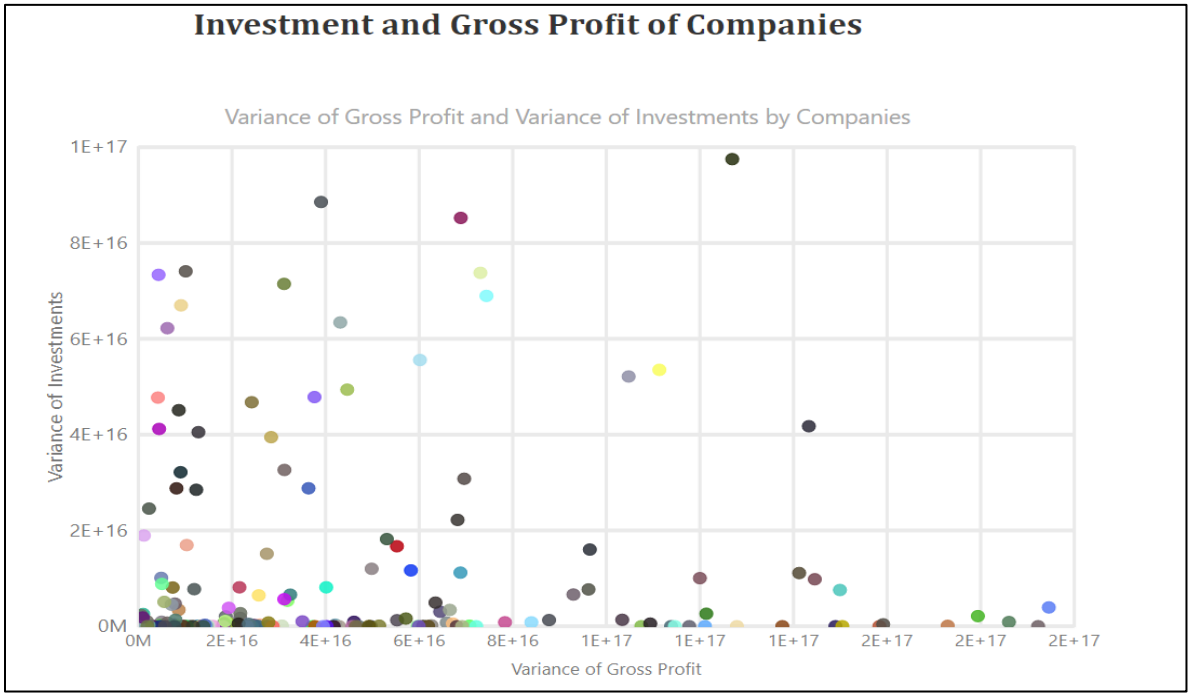
- In future, Stock price of the treasurer stock will decrease during year 2016 to 2019 and then in between year 2019 and 2020 prices will be increase almost nearest to the stock price of year 2016.

Chart 3:



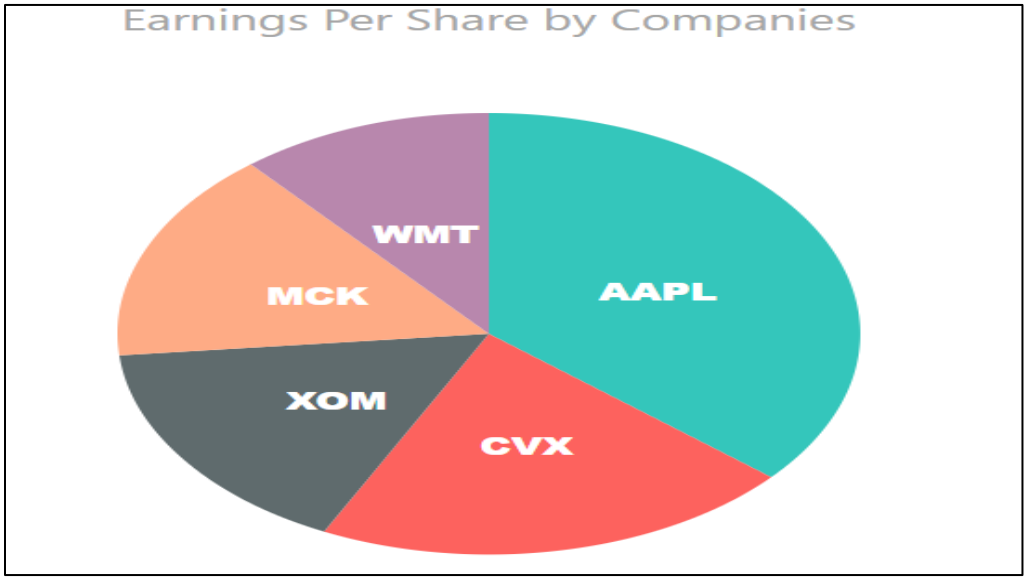
- Prices of stock decreases before January of each year and stock prices are on their lowest level during January. After January Stock prices are suddenly increases.
- Here, you can see more fluctuation in the stock price during July 2013 to Jan 2014. And lowest fluctuation between year Jan 2014 to July 2014.

Chart 4:



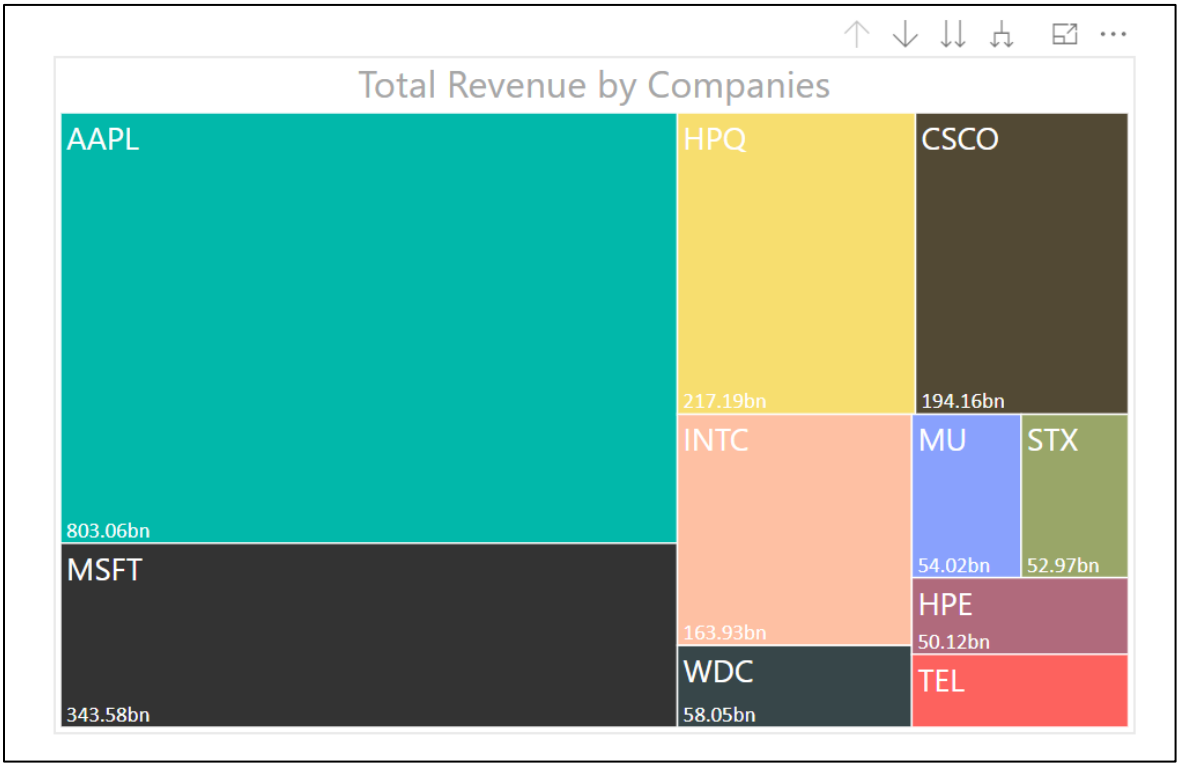
- Investors makes more profit by investing less in companies. While those who invest more their chances becomes less to get more profit.
- We can also say that people used to make less investment and get less profit.

Chart 5:



- AAPL company has almost double earning per share compare to WMT,MCK,XOM , and CWX .
- As you see Earning per share of top 5 companies. Highest Earning made per share by AAPL which is 36.26% having 64.15 shares .

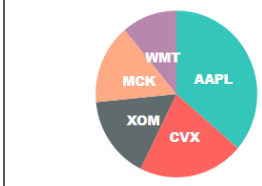
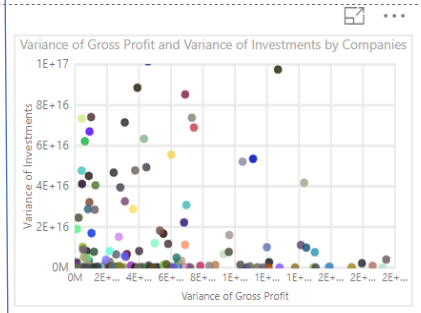
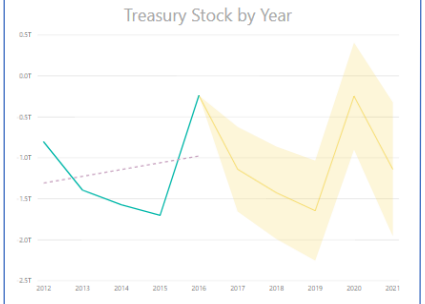
Chart 6:



- Revenue generated by company AAPL is almost 40% in top 10 IT companies.
- There are top 10 IT compnies filtered in following Tree map in which AAPL company made highest revemue because in previous visual you can see they have more Earning per share so they made more revenue 806.03 Billoions**.**

I. Purpose Map:

- Special purpose maps are designed or created for the special purpose. They are not standard map, specialized map displaying particular information.
- There are 3 main elements of purpose Map:
  1. **Explanatory:** Viewer is assisted to interpreting stage beyond perceiving stage in this stage user is scanning document and reading the data.
  2. **Exhibitory:** The viewer has to work to interpret the meaning, relying on their own capacity.
  3. **Exploratory:** Viewer is assisted to the extent of perceiving and then left with finding their own insights.

	Explanatory	Exhibitory	Exploratory
	Annotate/Describe	Display	Manipulate/Introgate
READING			
FEELING