American Stock Market: Tech Giants Analysis

(Public Offering Date -2017)

Jake Horner   
School of Computing, National College of Ireland,Mayor Street, IFSC,Dublin 1, Ireland.  
E-mail: x16318711@student.ncirl.ie

*Abstract*—Some of the highest priced Shares in the current American Stock market, are purely technology-based companies. These industries have changed a lot over the years, since their first conception, or since the company allowed public offerings of shares into the American market. I will be analyzing how they did in the market since their first recorded public offering until 2017.

Index Terms—Stock Market, Shares, Industry, Open Price, Volume.

# Introduction

Since the first computer created in 1946, the technology industry as be flourishing into a very stable and lucrative sector of business. Technology allows for highly competitive or highly innovative companies to soar to new heights within the industry. Constantly changing technology, does not guarantee the top leaders in the current industry to have such a strong presence in the future, as new innovative products and services emerge. The new advancements in tech allow companies to implement business plans and methods to rise to the top of their competitors with providing the best product for their consumers. This intern, by becoming more well known and lucrative will see a direct effect on the company’s share price and volume of shares traded.

I will be analysing two main Stock Market factors of three Top technology-based companies that have emerged since the inception of computers. Firstly, the average open price per year of a particular company, this will give us an overview of how popular the company was and the public or investors trust in the company technology or service. Secondly, volume of shares traded, amount of assets that changes ownership over the course of the year. This would be useful to gauge how many people we are buying and selling with this particular company that given year. i would also like to prove some support of correlation between the Volume sold and Open Prices.

The Three companies I have chosen for analysis are ‘IBM’ (International Business Machines), ‘Intel’, and ‘Microsoft’. I chose these companies as they provide a wide amount of years, to have more complex data analysis from the beginning of data based and technology companies. The companies also have a wide rage of products and specialisations covering Business Computing, Component Manufacturing, Operation Systems, Portable Technology, Software, and Hardware devices.

# Company Review

## IBM (International Business Machines) Overview

‘IBM’ has been in the machine and Information market for over 100 years. ‘IBM; has been a public company since 1911. ‘Thomas Watson’ took charge of the company in 1924 and renamed it to ‘Information Business Systems’, from the previously titled ‘Computing, Tabulating & Recording Company’. The made a household mane for themselves in the1920s by creating the public address system used in schools, and the ‘Social Security Administration’ adopting the company’s punch card machine for a network of Social Security number of all citizens of America.

‘IBM’ was at the forefront of computing machine in 1943, with the ‘Mark 1’, a sequence-controlled calculator. “This was the first device that we would recognize as a modern computer”[1]. Although Computing was not a part of their business plan until the 1950s with ‘Thomas Watson Jr.’s’ takeover of ‘IBM’. In the 1950s and 1960s, the company invented the ‘Hard drive’ that allowed computer to store data. They also created ‘FORTRAN’ a precursor to modern coding languages. The first vacuum tube computer was created by ‘IBM’ which is a precursor to modern microchips. “The modern era of IBM arguably began in 1981 with the Personal Computer 5150, or the ‘PC’. This was one of the first computers intended for consumer use rather than dedicated to business or the government”[1]. In this same era IBM has created architecture for Local Area Networks (LAN). This ear in the 1980 although they were pioneers for personal and business computing is when they started to decline as new and innovative companies joined the Market.

## Intel Overview

Intel is an American company created in 1968 based in California and created by ‘Robert Noyce’ and ‘Gordon Moore’. They provided chipsets and computer hardware for public consumption, business enterprise usage and for device manufacturer hardware components. Intel became a public company in 1971.

Intel early products were memory chip including the first metal oxide semiconductor, the ‘1011’. Future versions of this product, the ‘1013’ in 1970 was very successful as it was a One kilobit DRAM chip, which was a lot of information storage for the time. The paved the way for DRAMs to become the new standard in computer Processor technology. “That same year Intel introduced the erasable programmable read-only memory (EPROM) chip, which was the company’s most successful product line until 1985.” [2].IBM used Intel’s 16-BIT 8088 CPU to be used In the first mass produced ‘Personal Computer’. After this partnership intel expanded to providing their CPUs and hardware to many device manufacturers. Including gearing their processor to work alongside ‘Microsoft’s’ Window Operating System.

Since then, Intel has been providing their innovative processor technology to many manufactures of technology devices, for business and personal use.

## Microsoft Overview

‘Microsoft’ was founded in 1975 by ‘Bill Gates’ and ‘Paul G. Allen’. ‘Microsoft’ today has become a multi-faceted business, with stake in all sectors of computing technologies. From business system to gaming hardware, Microsoft really have a sector for nearly every form of internet application, software tools and hardware devices. ‘Microsoft’ became a publicly traded company in 1986.

‘Microsoft’ stared out as an Operating Systems software provider for the IBM ‘PC’ called ‘MS-DOS’. Since then, Microsoft has had their updated system ‘Windows’ in 1985 which was GUI to interact with their MS-DOS system. Their first manufactured PC being created in 1987, the ‘Altair 8800’.

“Microsoft dramatically expanded its electronic publishing division, created in 1985 and already notable for the success of its multimedia encyclopedia, ‘Encarta’. It also entered the information services and entertainment industries with a wide range of products and services”[3]. This proceeded to branch of too many sectors of computing which Microsoft has entered.

# Data Used

I downloaded ‘Huge Stock Market Dataset’ from ‘Kaggle’[4], which provided me with massive amount of data on publicly traded companies in the American market. This dataset would provide me with data for each trading day such as Date, Open, High, Low, Close, Volume, OpenInt.

I have used this data type of data before, but never to the extent of this project processing thousands of lines of Stock market daily logged data to find considerable and interesting correlations. I also have chosen different companies to my previous project.

I found this data quite in depth storing every trading day of each American traded company since its public stock offering. This allows me analyze thousands of records making the output and accurate and as in depth as possible on a particular company. This the stock price and volume traded can tell you a lot about how a company is doing as that period. Stock price are based off many factors and one of them is how popular it is and how much the public trust in the company’s business plan to create a higher performing selling price.

# Cost Analysis

## AWS EMR

Amazon’s EMR seem to be the most reasonably priced option. I have configured 1 EMR Core node with an EC2 instance of c4.large. I also configured 2 EMR task nodes with an EC2 instance of c4.large. On a monthly basis this should total the monthly price to 56.94 USD if the clusters are used constantly.

For a higher work load you can have 2 core node instances of c5d.24xlarge paired with three task nodes of c5n.2xlarge, this used constantly through the month will total 630.72 USD.

## Oracle Analytics Cloud

The Oracle Analytics cloud – Standard option come to a monthly price of $800 USD, for 1 Standard B89630 cluster that is being used constantly. a cheaper option would definitely be the 1 instance of the Standard – BYOL B89636 cluster, used constantly its monthly price is 240 USD. For a heavier workload, the Enterprise b89631 will come to a monthly cost of 1600 USD.

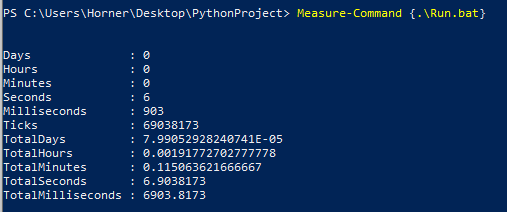
## Microsoft Azure HDInsight

Microsoft Offering of HDInsight with Hadoop Component can scale high depending on your workload. The most affordable option is the head node of 2 instances of E2 v3: 2 cores with 16 GB RAM, and 100 GB Temporary Storage, this is paired with 1 worker node instance of the same instance. This configuration with set you back 363.16 USD per month of consistent use.

On the more expensive end you can have 2 instances of head node E64 v3: 64 cores with 432 GB RAM, 3200 GB Temporary storage. This is paired with E32 v3: 32 cores with 256 GB RAM, 1600 GB Temporary storage, this is totaling at 12,820.26 USD.

# Local Testing

Testing Hadoop Map Reduce on my machine work perfect. Exporting all reducer output file to a directory withing the file called output. This process took on average of 6/7 seconds.



# Amazon Web Services EMR

I followed the week 12 tutorial up to a certain point. I had the cluster set up and the permission accessible, but my ‘MobaXterm’ could not seem to send the files using my AWS key. I tried with multiple different cluster set ups.

# Methodology & Implementation

## Data Cleaning

For my data cleaning I imported each of the relevant datasets and removed the index line of the dataset that stores all the column name for each of the files. I then export new document to be used by the mapper. I used panda to import the data sets to Python.

## Mapper Average

In my mapper for average Open price, I gather all the data needed for each company and refine it down to two columns. The data I store in mapper\_output.txt is the Year within date and the Open Price. I grab the year by only exporting the first four characters in the in the Date.

## Reducer Average

In my reducer for average, I grab the values separated by tab. I then calculate and sum up all the values that have the same date value (year). That total number is then multiplied by the number of instances of that date. I then export the year as an Index, and the Average Open price to \*companyName\*Average\_ReducerOutput.txt in an output folder.

## Mapper Total Volume

In my mapper for total Volume, I gather all the data needed for each company and refine it down to two columns. The data I store in mapper\_output.txt is the Year within date and the Volume value. I grab the year by only exporting the first four characters in the in the Date

## Reducer Total Volume

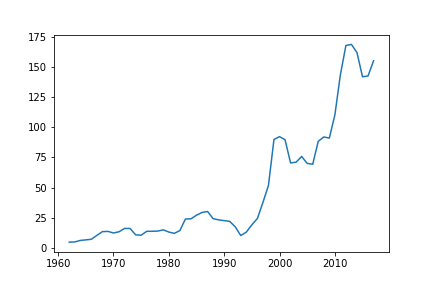
In my reducer for Total Volume, I grab the values separated by tab. I then calculate by adding together all the values that have the same date value (year). I then export the year as an Index, and the Total Volume to \*companyName\* TotalVolume\_ReducerOutput.txt in an output folder.

## Charting

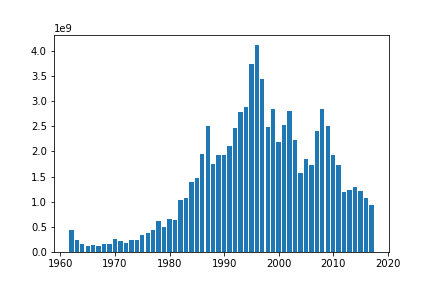
I used panda to import the data sets I made using the reducer scripts. Then I used matplotlib to create and export chart for the data.

# Results and Findings

## IBM



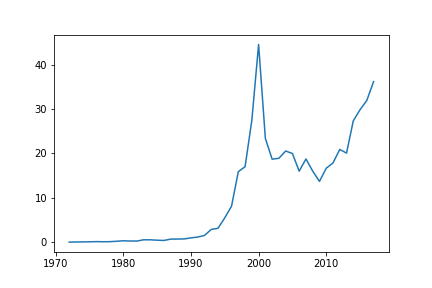
## This is the line plot made in Python using ‘matplotlib’. This line plot repesents the staock market open price average of every year since 1964. One problem with the data set is, due to inflation some of these increases may be exagerated as well as the decreases nullified.



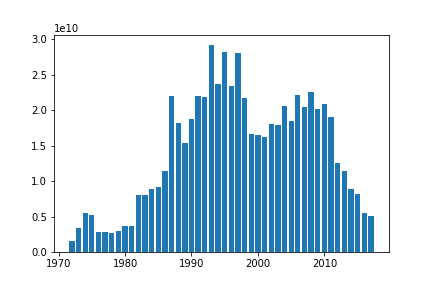
This is the bar chart that represents the total Volume traded each year. Due to the nature of volume those bars are not inflated as a single unit only represent a share which remain consistent until the company release more shares.

It is interesting to note that there are some sight correlations with the extreme peaks of both charts

## Intel



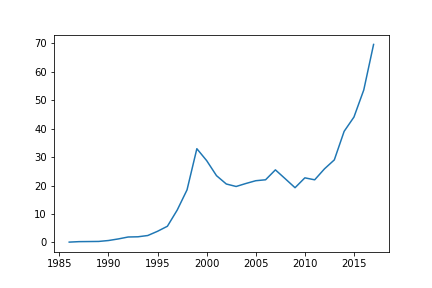
This is a line plot that presents the average open price for each year since Intel’s public offering in 1971. Once gain a problem with this dataset is inflation, but this data set should be accurate as the data set start for the 1970s.



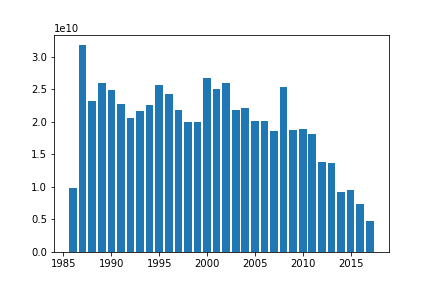
This chart represents the bar plot of the total Volume of assets traded for each year. Inflation does affect these statistics.

There are some correlations between the line plot and the bar chart. As they follow a similar curve.

## Microsoft



This is a line plot representing the average open price for each year since Microsoft first public offering in 1986. This graph clearly represents the tech boom of 1995 to 2001.

This is a bar chart of all the volume of share old in each year. Microsoft has clearly been a high traded and trusted in company. Share movement tends to slow toward more recent year as their share price has skyrocketed.

##### Conclusions and Future Work

It is widely known the more stocks are traded usually mean that the price is going up, which is what share sales Volume represents. I feel these charts and data clearly represent this data. I also found the tech boom of 1995 until 2001was.clearly represented by all the companies spiking in average stock price around these years.

I came into this project expected to do more research on the correlations that effect stock market. I was ultimately let down by my lack of knowledge in Python, Hadoop MapReduce. I ended up settling to prove that more stock is sold usually cause a stock price increase.

##### References

[1] E. Reed, ‘History of IBM: Timeline and Facts’, *TheStreet*. https://www.thestreet.com/personal-finance/history-of-ibm (accessed Dec. 30, 2020).

[2] ‘Intel | History, Products, & Facts’, *Encyclopedia Britannica*. https://www.britannica.com/topic/Intel (accessed Dec. 30, 2020).

[3] ‘Microsoft Corporation | History, Products, & Facts’, *Encyclopedia Britannica*. https://www.britannica.com/topic/Microsoft-Corporation (accessed Dec. 30, 2020).

[4] ‘Huge Stock Market Dataset’. https://kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs (accessed Nov. 22, 2020).

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published**