# INM430-Naveen_Zunjarwad_ProjectProgressReport

November 5, 2017

## INM430 - Analysis Of Stock Market And Exploring Potential Correlation For Future Trading

---

**Student Name :** *Naveen Zunjarwad*
**Project Title:** *Analysis Of Stock Market And Exploring Potential Correlation For Future Trading*

---

## 1 Part-1: Data source and domain description :

This work is an academic excercise to understand and analyse stock market with historic data. Also looking at FANG stocks price movement over the period of time and their effect on stock market. In the end, try to get some insights using correlation between larger set of stocks.

**Data Source**: Yahoo Finance -Using pandas_datareader and then copy to .csv.

Following are the main topics which will be covered in this work: * Analysis of a single stock and exploring different attributes * Brief look at FANG (Facebook,Amazon,Netfilx and Google) stocks and how they relate to S&P500 index. * Finding a correlation between top 50 S&P500 companies. * Conclusion and future scope ***

## 2 Part-2: Analysis Strategy and Plans :

- Analysis of a single stock and exploring different attributes

    - Take example APPLE stock from last 1 year data.
    - Data wrangling: Calculate 20days moving average and additional column.
    - Drop any empty cells.
    - Plot graphs for different attributes e.g: Adj Close, Volumes
    - Explore sub-plotting

- Brief look at FANG stocks and how they relate to S&P500 index.

    - Generate CSV files for FANG stocks
    - Plot a single graph for all FANG stocks for comparision.

- Finding a correlation between top 50 S&P500 companies.

    - Get data for top 50 companies and store it in CV.
    - Plot heatmap graph for the S&P Top 50 companies.

- Conclusion and future scope

    - List observations made along the way.
    - Future improvements

**Require your guidance: Stock Market is a very broad subject. Could you please advise me if I need to do anything more or incorporate any other ideas ?**

# 3 Part-3: Initial investigations on the data sources:

Initial investigation suggests that I could get the data programmatically using 'Yahoo Finance' using pandas_datareader. Also, I made following observations:

- Fetching data from 'Yahoo'remotely may take some time.So will consider storing the data in .csv files and the process them locally.
- I am thinking of having following data set:

    - Apple.csv
    - FANG.csv (or individual files Facebook.csv,Amazon.csv,Netflix.csv and Google.csv)
    - S&P_50.csv
    - Adding additional columns like moving average to the .csv files for analysis.
    - Simple text file S&P 500 company 'ticker' list or hardcoded python list.

# 4 Part-4: Python code for initial investigations

```
In [8]: # Author: Naveen Zunjarwad MSc Data Science City University London.

        # This cell is where you can copy + paste your Python code which loads your data and p
        # When you press CTRL+Enter, this cell will execute and produce some output
        # You can develop your code in Spyder (or another IDE) and copy + paste over here

        # Step-1: Load your data
        # Step-2: Get an overview of the data features, some suggestions to look for:
        #         number of features, data types, any missing values?,
        #         any transformations needed?

        # Start with your import (s) here.

        # The following is essential to get your matplotlib plots inline, so do not miss this

        %matplotlib inline

        import datetime as dt
        import matplotlib.pyplot as plt
        from matplotlib import style
        import pandas as pd
        import pandas_datareader.data as web
```

```python
style.use('ggplot')

# We will look at stock prices over the past year, starting at January 1, 2016
start = dt.datetime(2016,1,1)
end = dt.date.today()

# Let's get Apple stock data; Apple's ticker symbol is AAPL
# First argument is the series we want, second is the source ("yahoo" for Yahoo! Finan
df = web.DataReader("AAPL", "yahoo", start, end)

# Making sure df is a dataframe
type(df)
# Displays first 5 lines.
print(df.head())

#Creates csv file in project directory.
df.to_csv('AAPL.csv')

# We can adjust graph size here.
%pylab inline
pylab.rcParams['figure.figsize'] = (16, 10)

fig = plt.figure()

print("\nThese are some example graphs,I will do more filtering later. And also will ad

plt.subplot(2,2,1)
df['Adj Close'].plot()

plt.subplot(222)
df['Volume'].plot()

plt.show()
```

```
                Open        High         Low       Close    Adj Close  \
Date
2016-01-04  102.610001  105.370003  102.000000  105.349998  101.790649
2016-01-05  105.750000  105.849998  102.410004  102.709999   99.239845
2016-01-06  100.559998  102.370003   99.870003  100.699997   97.297760
2016-01-07   98.680000  100.129997   96.430000   96.449997   93.191338
2016-01-08   98.550003   99.110001   96.760002   96.959999   93.684120

              Volume
Date
2016-01-04  67649400
2016-01-05  55791000
2016-01-06  68457400
2016-01-07  81094400
```

2016-01-08   70798000
Populating the interactive namespace from numpy and matplotlib

These are some example graphs,I will do more filtering later. And also will add more descriptio
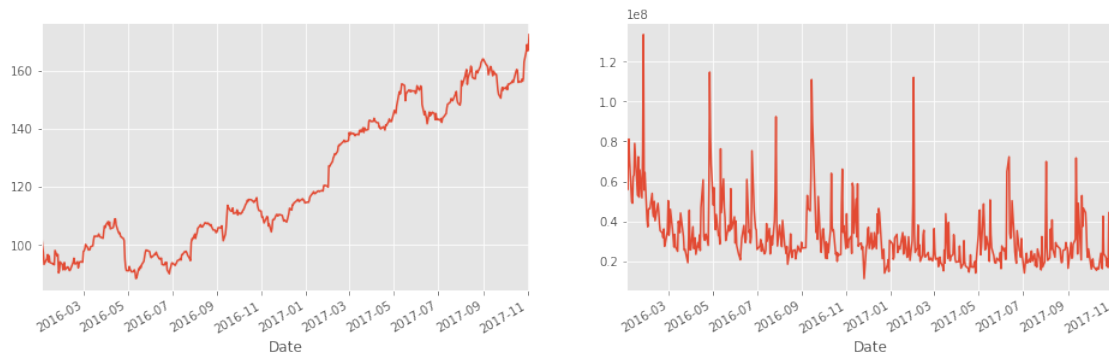


```
In [9]: print("High and Low values of Apple stock")
        df[['High','Low']].plot(label='Apple')

        plt.show()
```

High and Low values of Apple stock