Predicting Stock Prices

**Data Collection:**

🡪Pandas web data reader is an extension of pandas library to communicate with most updated financial data. This will include sources as: Yahoo Finance, Google Finance, Enigma, etc.

🡪We Will Extract Apple Stock Price Using Python. You can View In the Jupyternotebook file.

🡪For Rest of the Analysis We will use the closing Price which remarks the final Price in which the stocks are traded by the end of the day.

Note: Please View the Code To Understand

**Data Analysis:**

**Exploring Rolling Mean and Return Rate of Stocks:**

🡪In this analysis, We analyse stocks using two key Measurements : Rolling Mean and Return rate.

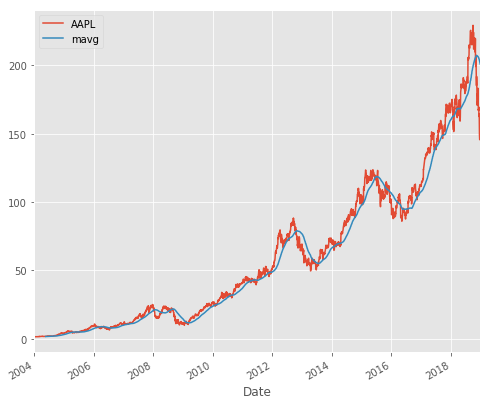
Rolling Mean (Moving Average): to determine trend

🡪Rolling Mean/ MA smooths out price data by creating a constantly average price.

This is useful to cutdown ‘Noise’ in our Price Chart.

🡪FurthurMore, This Moving Average Could act as ‘Resistance’ from the downtrend and uptrend of stocks you could expect it will follow the trend and less likely to deviate outside its resistance point.

🡪For better understanding, lets plot it our with Matplotlib. We will overlay the moving average with our stocks price chart.



In this chart Moving Average Showcases incrases trend the upturn or downturn of the stocks prices.

🡪Logically You should Buy when the stocks are experiencing downturn and sell when the stocks are experiencing upturn.

**Retrun Deviation – to determine risk and return:**

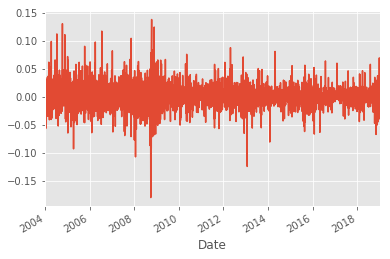
🡪Expected Return measures the mean, or expected value, of the probability distribution of investment returns.

🡪The expected return of a portfolio is calculated by multiplying the weight of each asset by its expected return and adding the values for each investment.

The Following is the formula you could refer to:



Based on the formula we can plot our returns as following



🡪Logically, our ideal stocks should return as high and stable as possible. If you are risk averse(like me), you might want to avoid this stocks as you saw the 10% drop in 2013 and 15 % drop in Near to 2009.

🡪This decision is heavily subjected to your general sentiment of the stocks and competitor analysis.

**Analysing Your Competitors Stocks:**

In this segment, we are going to analyse on how one company performs in relative with its competitor. Let’s assume we are interested in technology companies and want to compare the big guns: Apple, GE, Google, IBM, and Microsoft.

This will return you a slick table of closing prices among the stocks prices from Yahoo Finance. Neat!!

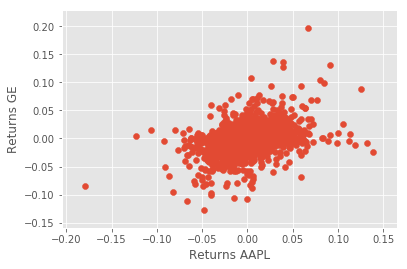
Correlation Analysis – Does one competitor affects others?:

🡪We can analyse the competition by running the percentage change and correlation function in pandas.

🡪Percentage change will find how much the price changes compared to the previous day which defines returns.

🡪Knowing the correlation will help us see whether the returns are affected by other stocks’ returns.

🡪Let’s plot Apple and GE with ScatterPlot to view their return distributions.



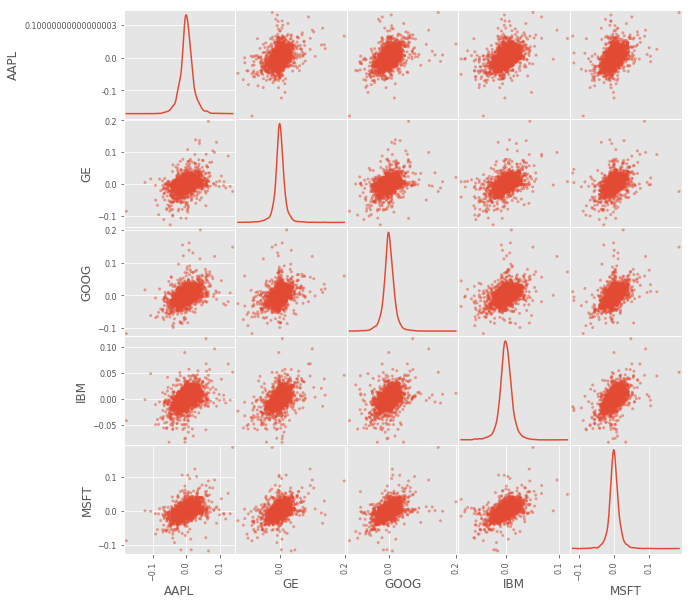
🡪We can see here that there are slight positive correlations among GE returns and Apple returns.

🡪It seems like that the higher the Apple returns, the higher GE returns as well for most cases.

🡪Let us further improve our analysis by plotting the scatter\_matrix to visualize possible correlations among competing stocks.

🡪At the diagonal point, we will run Kernel Density Estimate (KDE). KDE is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

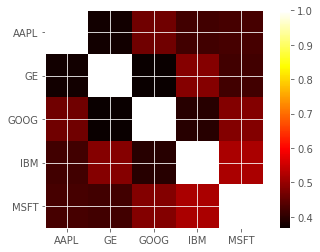
🡪It helps generate estimations of the overall distributions.



🡪From above chart we could see most of the distributions among stocks which approximately positive correlations.

🡪To prove the positive correlations, we will use heat maps to visualize the correlation ranges among the competing stocks. Notice that the lighter the color, the more correlated the two stocks are.

HEAT MAP

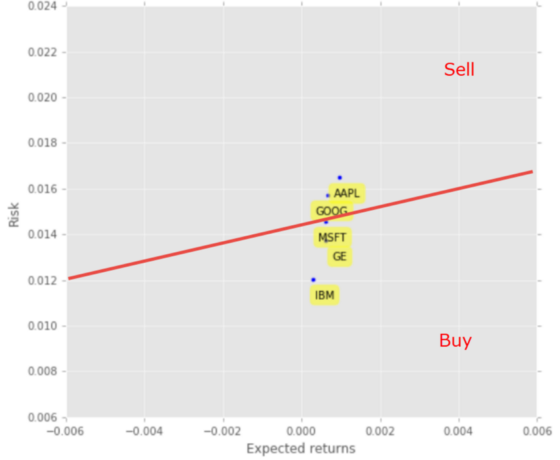


🡪From the Scatter Matrix and Heatmap, we can find great correlations among the competing stocks.

🡪However, this might not show causality, and could just show the trend in the technology industry rather than show how competing stocks affect each other.

**Stock Returns Rate and Risk:**

Apart from correlation, we also analyse each stock’s risks and returns. In this case we are extracting the average of returns (Return Rate) and the standard deviation of returns (Risk).



From above Chart We can Say:

🡪Now you could view this neat chart of risk and return comparisons for competing stocks. Logically, you would like to minimize the risk and maximize returns.

🡪Therefore, you would want to draw the line for your risk-return tolerance (The red line). You would then create the rules to buy those stocks under the red line (MSFT, GE, and IBM) and sell those stocks above the red line (AAPL and GOOG).

🡪This red line showcases your expected value threshold and your baseline for buy/sell decision.

**Predicting Stock Prices:**

**Feature Engineering:**

We will use these three machine learning models to predict our stocks: Simple Linear Analysis, Quadratic Discriminant Analysis (QDA), and K Nearest Neighbor (KNN).

But first, let us engineer some features: High Low Percentage and Percentage Change.

**Pre-processing & Cross Validation**

🡪We will clean up and process the data using the following steps before putting them into the prediction models:

🡪Drop missing value

🡪Separating the label here, we want to predict the AdjClose

🡪Scale the X so that everyone can have the same distribution for linear regression

🡪Finally We want to find Data Series of late X and early X (train) for model generation and evaluation

🡪Separate label and identify it as y

🡪Separation of training and testing of model by cross validation train test split.

View the code To understand

Model Generation :

Simple Linear Analysis shows a linear relationship between two or more variables.

When we draw this relationship within two variables, we get a straight line. Quadratic Discriminant Analysis would be similar to Simple Linear Analysis, except that the model allowed polynomial (e.g: x squared) and would produce curves.

We will plug and play the existing Scikit-Learn library and train the model by selecting our X and y train sets.

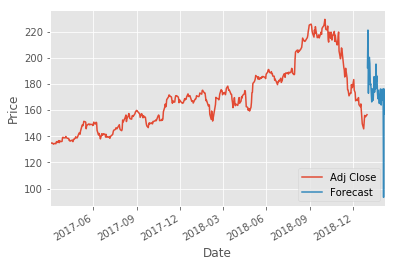
**Evaluation:**

A simple quick and dirty way to evaluate is to use the score method in each trained model. The score method finds the mean accuracy of self.predict(X) with y of the test data set.

This shows an enormous accuracy score (>0.95) for most of the models. However this does not mean we can blindly place our stocks. There are still many issues to consider, especially with different companies that have different price trajectories over time.

**Plotting the Prediction:**

Based on the forecast, we will visualize the plot with our existing historical data. This will help us visualize how the model fares to predict future stocks pricing.



Research Ideas:

Analyse economic qualitative factors such as news (news sourcing and sentimental analysis).

**My Projects On Data Analytics/Science:**

🡪Extraction and Text Analytics Of Twitter Data.

🡪Sentiment Analysis of Text.

🡪Web Data Extraction.

🡪Produces Time Series Forecasting ML algorithm for Twitter Data.

This Time Series Algorithm also helps for Financial Data.