Title Change Request---~~Predictive~~ Performance Analysis of 20 Large Cap US Assets with Pandas

2- Objective and Scope

* Add the following:
* At the end of this presentation, our intention to use data analytics tools to deliver a crash course on “Stock Market 101”
* All analysis was done in Jupyter notebook with Pandas

3- Assets included:

* These stocks are in the large capitalization or large cap category, each with market capitalization of >$10 billion ….
* generally large cap stocks are less volatile to systemic risk/effect
* *attractive to risk averse investors*
* all selected stocks are part of the S&P500 index
* the S&P500 covers over 80% of America equity market by capitalization

4- Extract from jupyter notebook:

* EDIT- (We need to include screenshot of where we imported the 23 csv files to show the effort put into merging and cleaning up of the separate csv files)

5- Plot of monthly returns:

* To compute each stock monthly returns we use pandas “pct\_change” function …note that stock returns is not the same as price. Return is the % change of today’s closing price from yesterday’s closing price
* The line plot here is the monthly returns of all stocks against time…returns is the dependent variable while time is the independent variable ..the time period covered is 120 months from Jan 2010 to Dec 2019
* The returns as seen here are volatile and they all swing on average +/- 2% on a monthly basis except for Netflix which we’ll talk about later in the presentation
* In the next slide we will magnify on the returns of 2 assets in the same industry

6- Monthly returns:

* Same plot as previous slide, but with only 2 stocks
* This is the monthly returns of the 2 US oil & gas giants …as shown in the line plot and the histogram distribution ---their returns have similar spread of +ve and -ve returns except for a few instances when CVX posted >10% return which appeared to be outliers.

7- Cumulative Return of all assets:

* This is a line plot of cumulative returns of all assets against time…for this plot we use the pandas cumprod function to calculate individual stock cumulative returns
* As a non-sophisticated investor your interest should be the growth of your investment …what will be the worth of my investment after x-years.
* Cumulative returns is the measure of the growth of investment after x years
* Again, as you can see in the plot, NFLX completely breakout from the lot … Netflix have seen unprecedented growth since 2009..it has pushed some notable companies out of the industry (Blockbuster) and continue to be a threat to established companies ..Fox, CBS
* A $1000 investment in NfLX in Jan 2010 will be worth $36,300 by the end of 2019, but a note of caution because it’s not all rosy. Netflix actually lost almost 50% of its value in Oct 2011 (shown in the plot) when it announced the split of DVD and streaming (Qwisker) business
* For now, let’s consider the Netflix phenomenon as an outlier and we’ll decouple it from the group in the next slide

8 - Cumulative Returns without NFLIX and S&P500…(NEW Slide)

* As you can see here all 20 companies maintained a healthy growth over the ten years period.
* An investment of $1000 in the least performing asset (XOM) in Jan 2010 will be worth $1,573.30 at the end of Dec 2019
* While an investment of $1000 in the best performing asset (SBUX) in Jan 2010 will be worth $9,873.90 at the end of Dec 2019
* In essence, with large cap companies, your strategy to grow your investment should be long term holding or “Buy and Hold”

9-STD/ Volatility:

* There is always some risk in every undertaking and even greater when you put your money in a company you have no clue about its operations
* Here we use pandas standard deviation function/method to calculate the std of the stocks monthly returns. In essence this is a measure of the volatility of the return.
* Remember in this context, STD is the dispersion of each stock returns relative to its mean value(variance). In essence a higher STD indicates more datapoints are further away from the mean within the data set and a large return range
* All stocks except for few are within 3.5 to 5.5% and of course the standout is NFLX. The range as seen here are typical of the large-cap S&P 500 stocks
* Note that a low STD isn’t necessarily the holy grail of investment….young and aggressive investors typically go for above average std assets
* Also note that std are very susceptible to outliers…

10 - Correlation Matrix…need to round matrix up to 1 decimal place

* Here we created a correlation matrix of the stock returns with pandas CORR function
* The objective of this aspect of our analysis is to understand the degree of linear relationships between pairs of stocks and to use analytics tools to shed more light on the concept of a diversified portfolio. (Do not put all your eggs in one basket)
* The range is -1 to +1.
* Correlation rule of thumb:
* -1 or 1 (perfect correlation)
* >= +/- 0.7 (strong correlation)
* >= +/- 0.5 (moderate)
* <= +/- 0.3 (weak)
* 0 (no linear relation sip)
* Take a look at assets within same industry BAC/WFC; AAL/DAL; CVX/XOM and note that they have strong correlation
* Also take a look at individual assets correlation with S&P 500….very hard to find assets within the S&P500 index with <0.3 correlation with the market.
* Key take away here is to have a mix of lowly correlated assets in your portfolio.
* Next we will focus on 2 assets and make an attempt to predict their future stock price with regression model ….the assets are CVX with the strongest correlation with the market and NFLX with the lowest correlation with the market …also NFLX returns have the highest volatility while CVX have a much lower volatility relative to NFLX

11- CVX Regression—

* Enough talking points in the slide
* Add equation to show how we arrived at forecasted price like I did with newly added multiple regression slide (Clay note- Discuss with group to add or not, may overcrowd slides)

12- NFLX Regression

* What could be responsible for the high spread within the NFLX price distribution and non-conformity with the model?
* All the price breakout above $200 happened in 2019
* 90% of NFLX prices in 2019 (month 100 and above) are above the line equation and could be outliers if regression was limited to Q1 of 2019.
* NFLX members-base more than doubled at the end of Q3 of 2019 with massive international expansion resulting in more than 60% of its subscribers being international.

13- Box plots

* How good is the data? To answer the question around the credibility of the 2 assets data we created the box plots with their price to give us a quick visual understanding of the distribution and skewness/dispersion/variability.
* CVX appears to have normal distribution ..shorter box indicates a less dispersed distribution which aligns with our earlier calculation of CVX std
* However, NFLX has a positive skew and the longer box show greater degree of spread ..this also aligns with our previous observations

14- Added New Slide: Multiple Regression

* We also consider the possibility of a more conservative predictor for price with multiple regression by introducing another dependent variable. We added the US unemployment rate for 10 years coinciding with the years included in our analysis
* For this purpose we imported the LinearRegression module from sklearn.linear\_model
* For CVX: The introduction of an additional variable resulted in a more conservative or better fit predictor for CVX with Jan and Feb prices within the std.
* For NFLX: The model didn’t provide a better predictor further alluding to its non-conformity with others in the pack.

15 - Conclusions---slide need to be updated with additional key take-aways below:

* Our analysis has shown Python/Panda as an effective tool in financial data analytics
* Stock prices are time series with predictable trends especially during period of economic stability
* Day trading is speculative and highly risky. You should stay away from it If you are risk-averse and your strategy is to grow your investment
* Daily/Monthly volatility only appeals to experts, day traders, and those with insider information which is illegal.
* Monthly or daily returns should not be a factor when you are making investment decisions, though they are a good indicator of periodic volatility of an asset
* Cumulative returns should be your primary focus when thinking about growing your capital with stocks
* Make your portfolio selection with a mix of lowly correlated assets . Also include assets that have low correlation with the market to provide some balance against systemic risk
* To use multiple regressions for price prediction there is no one variable fit all to use alongside Time. Each stock unique situation and industry knowledge will provide an insight into variables that impact their price movement.