**Assignment**

Using the given excel file, analyze the data as you see fit. You can choose to focus on all the exchanges and their respective datapoints or narrow your focus to one or a few exchanges. Using what has been learned in class, see if you can find any anomalies, correlations, or even potential trading strategies for the given exchanges. As a note, the exchanges data is random data.

**My Environment/Setup**

* Python 3.7
* Packages:
* Pandas
* Numpy
* Matplotlib
* Statsmodels
* importstatsmodels.formula.api
* Pylab
* Xlsxwriter

**As a Note**

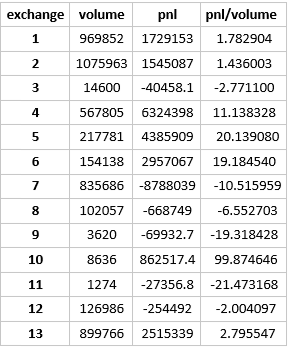
* I left a lot of code in my script that I would typically delete or store separately after its use (mainly output functions that I needed for my write up). I chose to comment it out instead of deleting in case you had questions about how I completed some tasks.
* With more time I would have liked to compare each exchange to numerous benchmarks both small/large cap, domestic and international to see if there is any correlation. I chose to devote my time to studying the actual data provided and comparing the exchanges amongst one another. I wanted to manipulate the data as little as possible, so the outcome was altered least.

**Initial Impressions of Data Set (Follow-up Items, Concerns, Potential Improvements)**

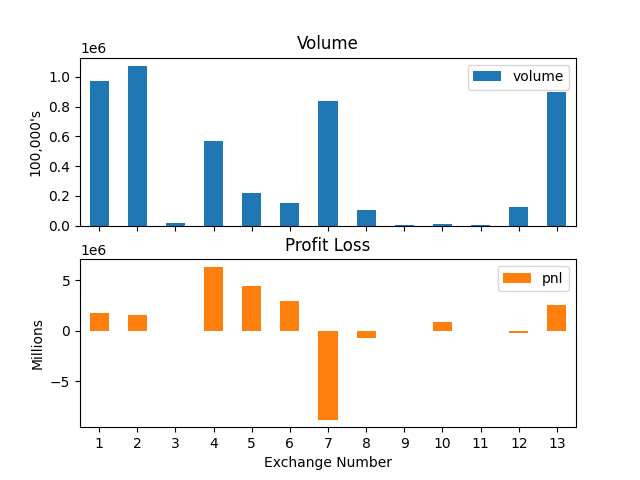
* Initial examination of the supplied csv file displays some column data displaying 0. Noticing the dates fall on 9/3/2018 (Labor Day), 11/22/2018 (Thanksgiving), 12/5/2018 (President Bush Funeral), 12/25/18 (Christmas), 1/1/19 (New Years), 1/21/19 (MLK Day), etc. **it is simple to conclude these are market holidays** and the exchange (exchange\_id 1) is a US exchange as it has US specific holidays. **I chose to remove these dates from my DataFrame for all exchanges if they displayed zero values for a more accurate data set**. I left the lines of code in my script to be commented out or deleted in case these dates should not have been removed. **This would be a follow-up item.**

* *Potential issue found*: 11/22/2018 (Thanksgiving Day) still shows P/L value with zero volume but perhaps this is an international exchange that was open or exposed to currency markets and our fund decided not to trade that day but allowed our exposure to continue that day (unlikely but thought it should be mentioned). The other more likely possibility would be poor data that would need to be removed from the dataset. The following Thanksgiving (2019) had 0 volume and 0 P/L so I would decide to change the P/L to 0 and mark it as an error unless contacts at my company could answer whether or not this was correct or incorrect. For my analysis, I chose to remove these dates that had zero volume but some P/L. **This would also be a follow-up item.**
* The exchanges data all begin on different days making it difficult to do a return comparison by the day. Additionally, the amount of trade days recorded also vary by exchange. Ideally, we would have the same number of trade days across the same trade dates to compare the exchange performance against one another for best analysis. There may be slight variances due to a Leap Year, unannounced closed market day for a tragedy such as a Presidential Funeral or terrorist attack, etc. but keep it as close as possible. Some exchange data has over 500 entries while others have around 300. This makes it difficult to even compare total volume or pnl\* since there was not as much opportunity as other exchanges for gains/losses. If this was my work, I would do what is necessary to either find missing values to complete the data set across all exchanges or clean the data set and evaluate the dates that we do have corresponding data for across all exchanges. **A potential improvement would be to analyze only exchanges with comparable trade dates or back-fill any missing data for the exchanges.** \*For this analysis please use pnl and P/L interchangeably as pnl is the variable name in the tables and graphs.
* I chose to first make a table to display the volume and pnl per exchange. Keeping the thought of the large variance in trading days amongst the exchanges I chose to create a pnl/volume ratio. This helped me see the return for how much volume they traded, removing the time component. The mean of the PV ratio (this is what I will call the pnl/volume ratio) was about 7.2 with a standard deviation of 30.56. **I interpreted it as the exchanges with a higher PV ratio made higher returns per unit traded.**

**Total Profit/Loss Per Exchange**

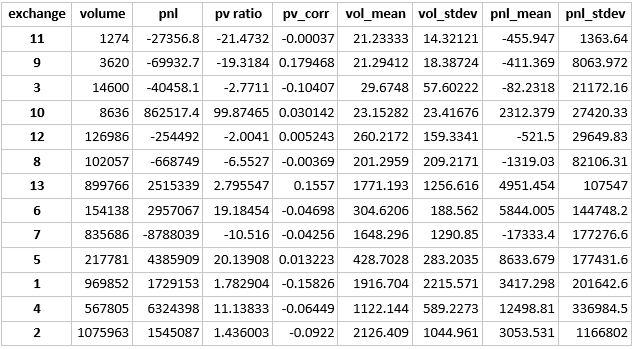


**Visual Comparison of Volume to Profit/Loss**

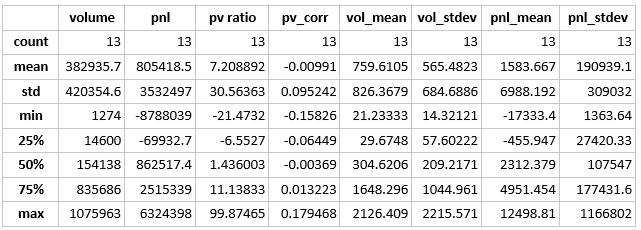


**Evaluating the Profit/Loss Standard Deviations**

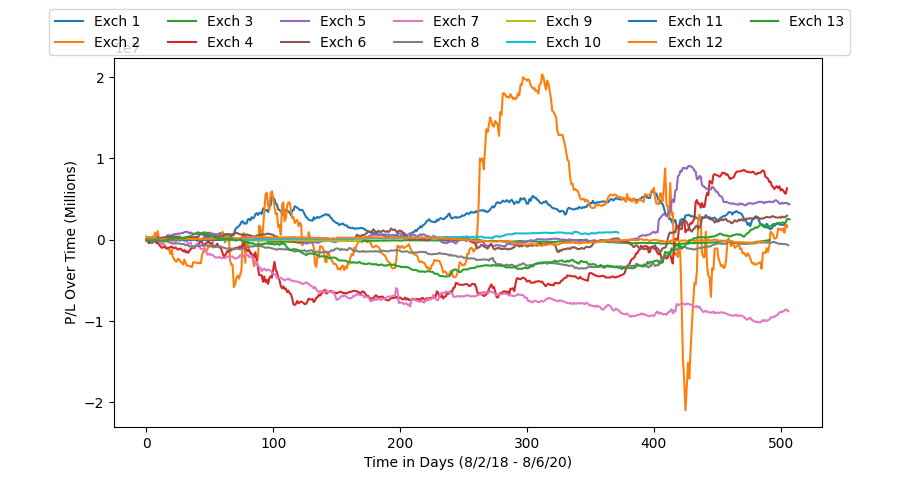
* Data below sorted by P&L standard deviation, ascending. As we can see, exchange 11 had a significantly smaller standard deviation from the rest of the exchanges. This exchange did also have a loss of $27,356 along with the smallest volatility mean and standard deviation. In total it had the smallest volume as well. It had the lowest ratio of pnl/volume (PV ratio). While it may seem the 'least risky' based on the data provided, in comparison **it does not pay off to take on the least risk trading on exchange 11 in comparison to the others**.
* In contrast, exchange 2 had the highest P&L standard deviation of 1,166,802. Exchange 2 had a profit of $1,545,087, a pnl/volume ratio of 1.436 that was a bit below average (although I would potentially eliminate exchange 10's PV ratio from the mean calculation as it is skewing the data as a potential outlier in comparison to the rest of the data. **I would investigate this time depending**).



**Some Additional Statistical Figures**

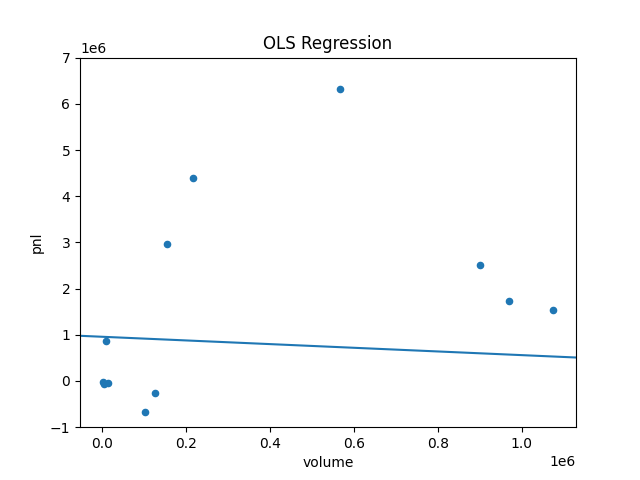


**Visual Output of P/L Over Time for Each Exchange**

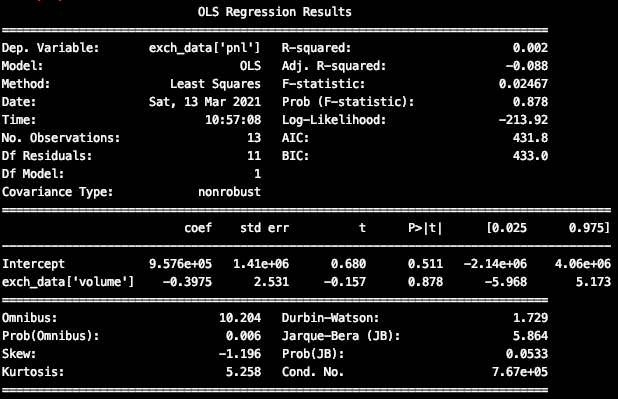


**OLS Regression of PNL on Volume (Exchange Sums)**

* **We can see the OLS regression of pnl on volume may visibly show a very weak negative correlation, likely no correlation**. Based on the p value results in the OLS regression, we would not be able to reject the null of the effect of pnl on volume. This model does not seem to fit the data well and there is an extremely small R-squared value. There is little to no signs of correlation based on these results. **I did a sanity check correlation test of all pnl/volume data across all exchanges provided and I got a very small figure of -0.03661. I would conclude based on the results that there is not much purpose in studying the affect of the sum of pnl on volume amongst the exchanges.**

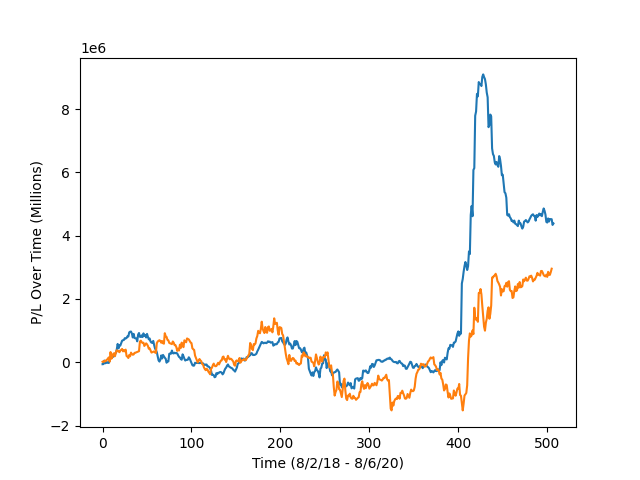






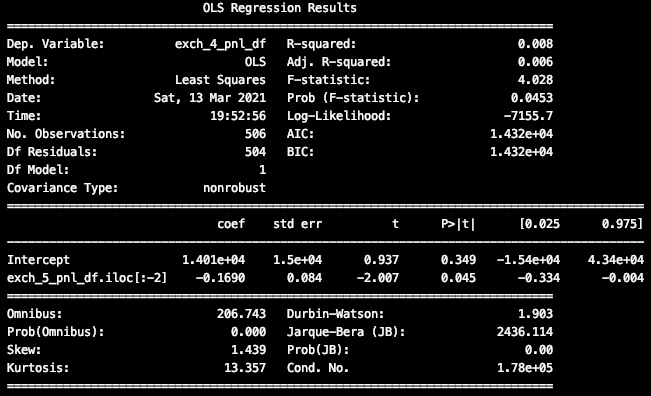
**Further Look at Exchanges 5 and 6**

* I wanted to take a further look at exchanges 5 and 6 as they graphically looked similar before even looking at the actual data. They visually appear to have the most correlation out of all the exchanges.



* The regression below of exchange 4 pnl on exchange 5 pnl has a P value of 0.045. If we assume a 95% confidence interval (α = 0.05) then we would have to reject the null of no correlation between the two. This means **there is some presence of correlation between the returns of these two exchanges**. One caveat to point out is that two trade dates were removed from exchange 5 (the last two trade dates) as OLS required the columns to be of same length. This is not ideal, and it would be preferred to have identical time frame over same trade dates.
* If we were looking to fit a different data model, perhaps one of information criteria, the Akaike and Bayesian have equal values from the regression below so there would be no advantage to one or the other. This makes sense due to the singular variable involved and relatively small sample size. Multivariable regression would likely prefer the AIC as BIC is penalized for a higher amount of observations involved. I wanted to note this as the values were identical.
* I would run this test for all exchanges to be certain there is no correlation, but I wanted to focus on this one for my analysis.





**Autocorrelations**

* Lastly, I wanted to take a look at the autocorrelation of the profit/loss columns for each exchange. **The exchange I might take into consideration is exchange 11. Autocorrelation at lag 1 is 0.47, lag 2 is 0.41, and at lag 3 is 0.30**. While not extremely high, it is much higher than the other exchanges. This could help some technical traders indicate some momentum/trend in exchange 11. Looking at the progressive, large upswings and downswings on the chart of 11's pnl this is no surprise. **If this is maintained, it may be considered an anomaly.**

