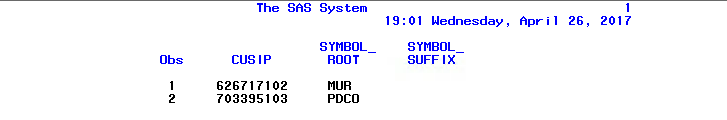
**Name:** Nguyen Thi Cam Hoang **Date:** April 27th, 2017

**Student ID:** 20165327

# Assignment 2

1. **Preparing the TAQ data**
2. Get information

* Find the prior day’s closing day by Yahoo Finance: **June 5, 2013**



* I have information as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cusip number (8 digit)** | **Cusip number (9 digit)** | **Stock Market** | **Symbol** | **Company name** |
| 62671710 | 626717102 | N | MUR | MURPHY OIL CORP |
| 70339510 | 703395103 | Q | PDCO | PATTERSON COMPANIES INC |

1. Export data.

All the data are retrieved by following these rules:

* + Only quotes/trades during normal market hours are considered
* Quotes : 9:00 am ~ 4:00 pm
* Trades : 9:30 am ~ 4:00 pm
  + Negative values of size or prices
  + Normal quote conditions
* Qu\_cond : other than (A, B, H, O, R, and W) delete
  + Canceled quote
* Qu\_cancel : B(delete)
  + Normal Trade conditions
* Tr\_corr : other than “00” delete

After cleaning and screening data, 3 different data; trade, quote, nbbo; are exported into one excel file by using:

proc export data=ct

outfile='C:\Users\Hoang\Google Drive\UNIST\3rd\_Courses\finance\assignment\AS2\data\data.xlsx'

dbms=xlsx;

sheet='trade';

proc export data=cq

outfile='C:\Users\Hoang\Google Drive\UNIST\3rd\_Courses\finance\assignment\AS2\data\data.xlsx'

dbms=xlsx replace;

sheet='quote';

proc export data=nbbo

outfile='C:\Users\Hoang\Google Drive\UNIST\3rd\_Courses\finance\assignment\AS2\data\data.xlsx'

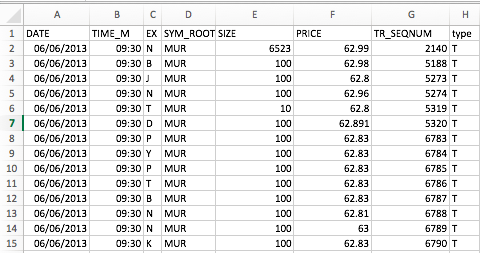
dbms=xlsx replace;

sheet='nbbo';

run;

1. **Trade data analysis**

After preprocessing, the trade data will be like:

****

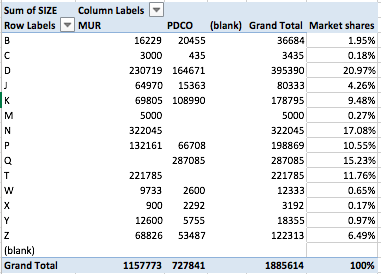
1. What is the largest single trade of the day? When was it executed? at what price?

It can be done by using **max** function in excel.

|  |  |  |  |
| --- | --- | --- | --- |
| **Symbol** | **Time** | **Price** | **Size** |
| MUR | 09:30:31 | 62.99 | 6,523 |
| PDCO | 09:50:57 | 38.42 | 10,000 |

1. How much volume (in shares) was done on each exchange? Which exchange has the largest volume? What is the market share of each reporting exchange?

It can be done by using **Pivot** table



* The exchange with the largest volume is:
* MUR: N
* PDCO: Q
* Total: D

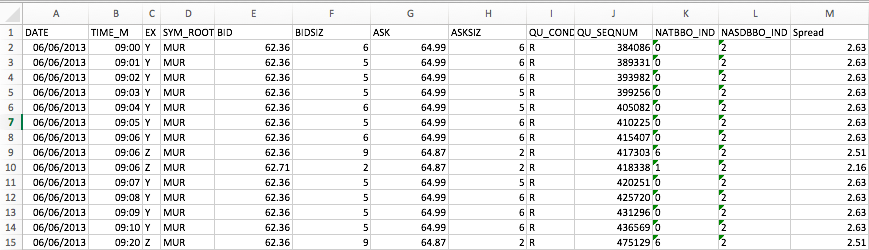
1. Between 9:00 and 16:00, plot the trades. That is, plot price versus time. Price vs time plotting
2. Construct 30-minutes time buckets, i.e. 9:30-10:00, 10:01-10:30, … In each interval, compute the total number of trades. Again, compute the market share of each reporting exchange within time interval.

* To construct 30-min time bucket, the trading time are rounded to the lower limit of the interval. For example, trading time 9:37 will be rounded to 9:30. By taking advantage of Pivot, the below result is archieved.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sum of SIZE** | **Column Labels** | | | | | | | | | | | | | | **Grand Total** |
| **B** | **C** | **D** | **J** | **K** | **M** | **N** | **P** | **Q** | **T** | **W** | **X** | **Y** | **Z** |
| **MUR** | **16229** | **3000** | **230719** | **64970** | **69805** | **5000** | **322045** | **132161** |  | **221785** | **9733** | **900** | **12600** | **68826** | **1157773** |
| 09:30 | 2800 | 1100 | 16003 | 2800 | 10955 |  | 34991 | 10039 |  | 13173 | 400 | 100 | 2800 | 6800 | 101961 |
| 10:00 | 1900 | 400 | 11187 | 2100 | 3000 |  | 14718 | 8808 |  | 10941 | 700 |  | 400 | 3200 | 57354 |
| 10:30 | 2100 | 200 | 29540 | 4900 | 1300 |  | 18388 | 3601 |  | 18786 | 1100 |  | 700 | 5018 | 85633 |
| 11:00 | 400 |  | 10567 | 3764 | 2000 |  | 11186 | 10336 |  | 11408 | 300 | 100 | 800 | 2700 | 53561 |
| 11:30 | 4580 |  | 5953 | 3100 | 3400 |  | 8900 | 15648 |  | 15218 | 500 |  | 500 | 1700 | 59499 |
| 12:00 | 400 |  | 16200 | 8300 | 10100 |  | 15171 | 6545 |  | 17875 | 500 |  | 500 | 4800 | 80391 |
| 12:30 | 300 |  | 13013 | 4827 | 6160 |  | 18798 | 5600 |  | 7743 | 800 | 100 | 300 | 3200 | 60841 |
| 13:00 | 300 |  | 6071 | 3468 | 2800 |  | 16691 | 9323 |  | 12694 | 200 |  | 600 | 2603 | 54750 |
| 13:30 | 700 |  | 15035 | 5150 | 4289 |  | 23147 | 8574 |  | 13850 | 200 |  | 700 | 5200 | 76845 |
| 14:00 | 400 |  | 6366 | 4200 | 3750 |  | 27673 | 9050 |  | 15824 | 400 |  | 700 | 5200 | 73563 |
| 14:30 | 200 |  | 14949 | 4900 | 6600 |  | 22489 | 12504 |  | 15592 | 600 |  | 1500 | 6700 | 86034 |
| 15:00 | 400 |  | 21332 | 5261 | 7651 |  | 28694 | 13786 |  | 22980 | 833 |  | 600 | 5000 | 106537 |
| 15:30 | 1749 | 1300 | 64503 | 12200 | 7800 | 5000 | 81199 | 18347 |  | 45701 | 3200 | 600 | 2500 | 16705 | 260804 |
| **PDCO** | **20455** | **435** | **164671** | **15363** | **108990** |  |  | **66708** | **287085** |  | **2600** | **2292** | **5755** | **53487** | **727841** |
| 09:30 | 2100 |  | 27000 | 1100 | 7197 |  |  | 5407 | 29957 |  | 100 |  | 300 | 4800 | 77961 |
| 10:00 | 1100 |  | 20193 | 1500 | 5800 |  |  | 3838 | 22334 |  |  |  | 100 | 3500 | 58365 |
| 10:30 | 800 |  | 6600 | 800 | 7506 |  |  | 5700 | 13709 |  | 400 |  | 200 | 3700 | 39415 |
| 11:00 | 500 |  | 6010 | 1000 | 6498 |  |  | 3834 | 11387 |  | 300 |  |  | 2300 | 31829 |
| 11:30 | 300 |  | 6078 | 400 | 13116 |  |  | 3205 | 12928 |  | 200 | 500 | 100 | 2800 | 39627 |
| 12:00 | 400 |  | 8375 | 100 | 5300 |  |  | 5466 | 11370 |  |  |  |  | 2400 | 33411 |
| 12:30 | 500 |  | 1700 | 500 | 6699 |  |  | 2600 | 8900 |  | 300 |  |  | 1575 | 22774 |
| 13:00 | 200 |  | 5298 | 1100 | 4100 |  |  | 2300 | 13522 |  | 300 |  |  | 3300 | 30120 |
| 13:30 | 200 |  | 3449 | 500 | 3000 |  |  | 2200 | 12715 |  | 100 | 100 |  | 3100 | 25364 |
| 14:00 | 400 |  | 15777 | 700 | 6200 |  |  | 6300 | 27447 |  | 200 |  | 200 | 3000 | 60224 |
| 14:30 | 1200 |  | 7352 | 1500 | 5900 |  |  | 2700 | 14847 |  | 100 |  | 700 | 2700 | 36999 |
| 15:00 | 2098 |  | 15579 | 500 | 9958 |  |  | 7093 | 20314 |  |  |  | 900 | 4000 | 60442 |
| 15:30 | 10657 | 435 | 41260 | 5663 | 27716 |  |  | 16065 | 87655 |  | 600 | 1692 | 3255 | 16312 | 211310 |
| **Grand Total** | **36684** | **3435** | **395390** | **80333** | **178795** | **5000** | **322045** | **198869** | **287085** | **221785** | **12333** | **3192** | **18355** | **122313** | **1885614** |

1. **Quote data analysis**

After preprocessing, the Quote data will be:

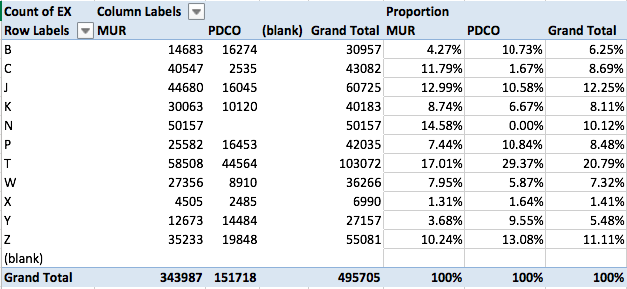


As can be seen, there is no information about best bid in the data. This information cannot be derived easily. To be more specified, in order to observe the liquidity as well as the direction of buy/sell, we can merge trade, quote and nbbo data for more detail analysis. This kind of data will look like:



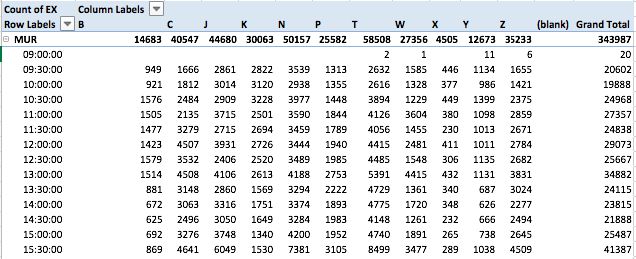
In this exercise, for simplicity, I assume that the bid is represented as best bid so that we can see the difference between problem 3 and 4.

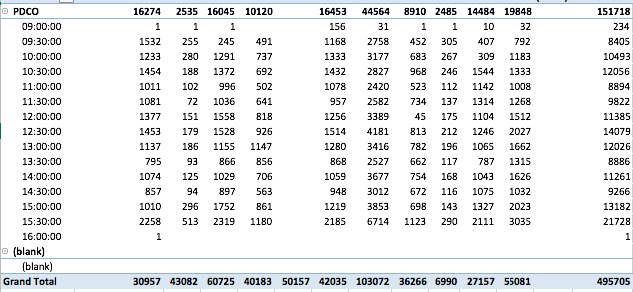
1. During the day, how much the number of quote updates was done one each exchange? Which exchange has frequently updated the quotes? For what proportion of the quotes is the primary listing exchange at the best bid? (Note that the best bid or the best ask do not have to be alone. For example, if the primary exchange is the NYSE, the NYSE is at the best bid if the exchange of the best bid is marked as "N", "XN", or "BNZ".)



From the above table, it can be seen that T and J exchange were high-frequent quoted in the market. Therefore, these stocks seem to be more liquid compared to the others.

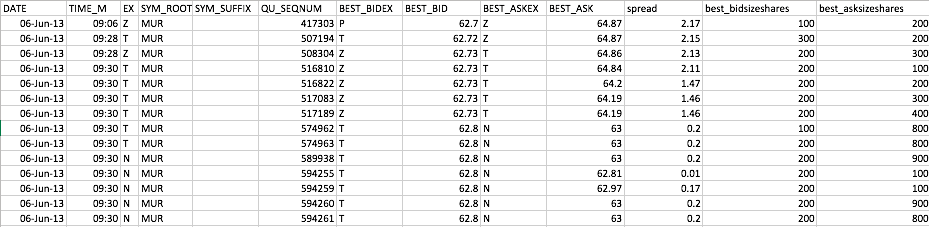
1. In each of the 30-minute time buckets, do the same thing for the above question.

****

****

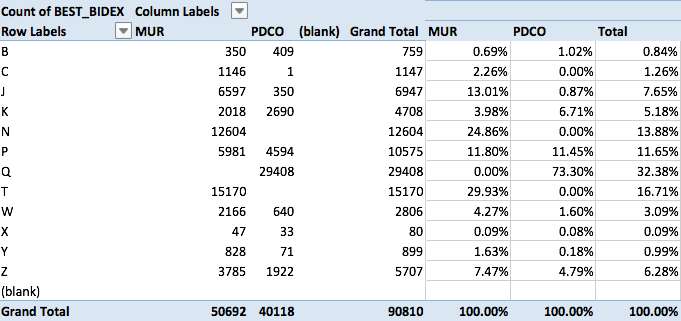
1. **NBBO data analysis**

After preprocessing, the NBBO data looks like:

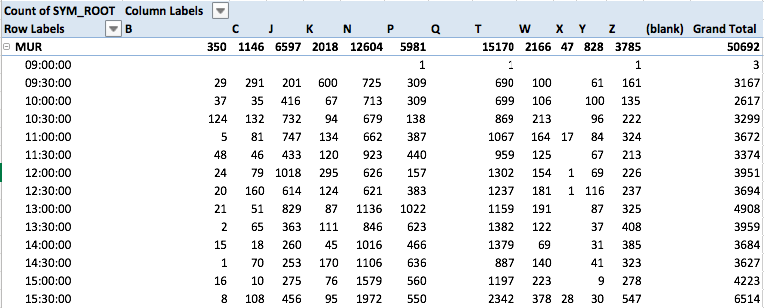
****

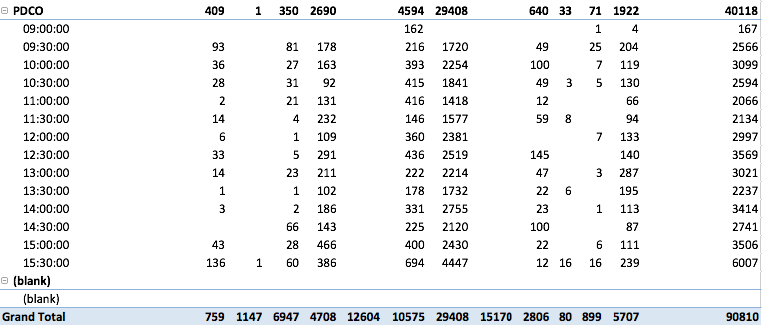
NBBO provides the information on the highest bid as well as the lowest ask price, which means the best bid price in the data can be used for this problem.

1. Do the same analysis in question 3. Are there any differences from results in question 3.

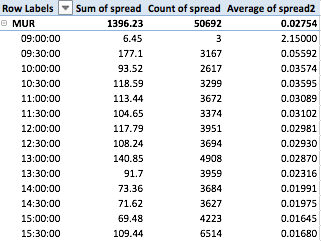


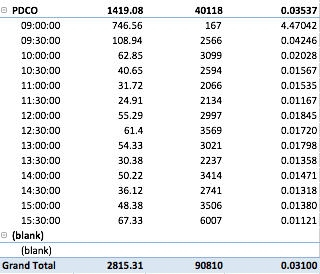
Here the proportion of exchange Q and T are higher compared to the others, which yields different result from problem 3.





1. In each interval, compute the average of the quoted bid-ask spread and plot it.





Because the average of Spread within the first interval is outlier from the others, which was caused by the the amount of stock of the previous trading day. Therefore, it is not plotted in the figures.

|  |  |
| --- | --- |
|  |  |