# THE MESSAGE FILE
## SOLUTION

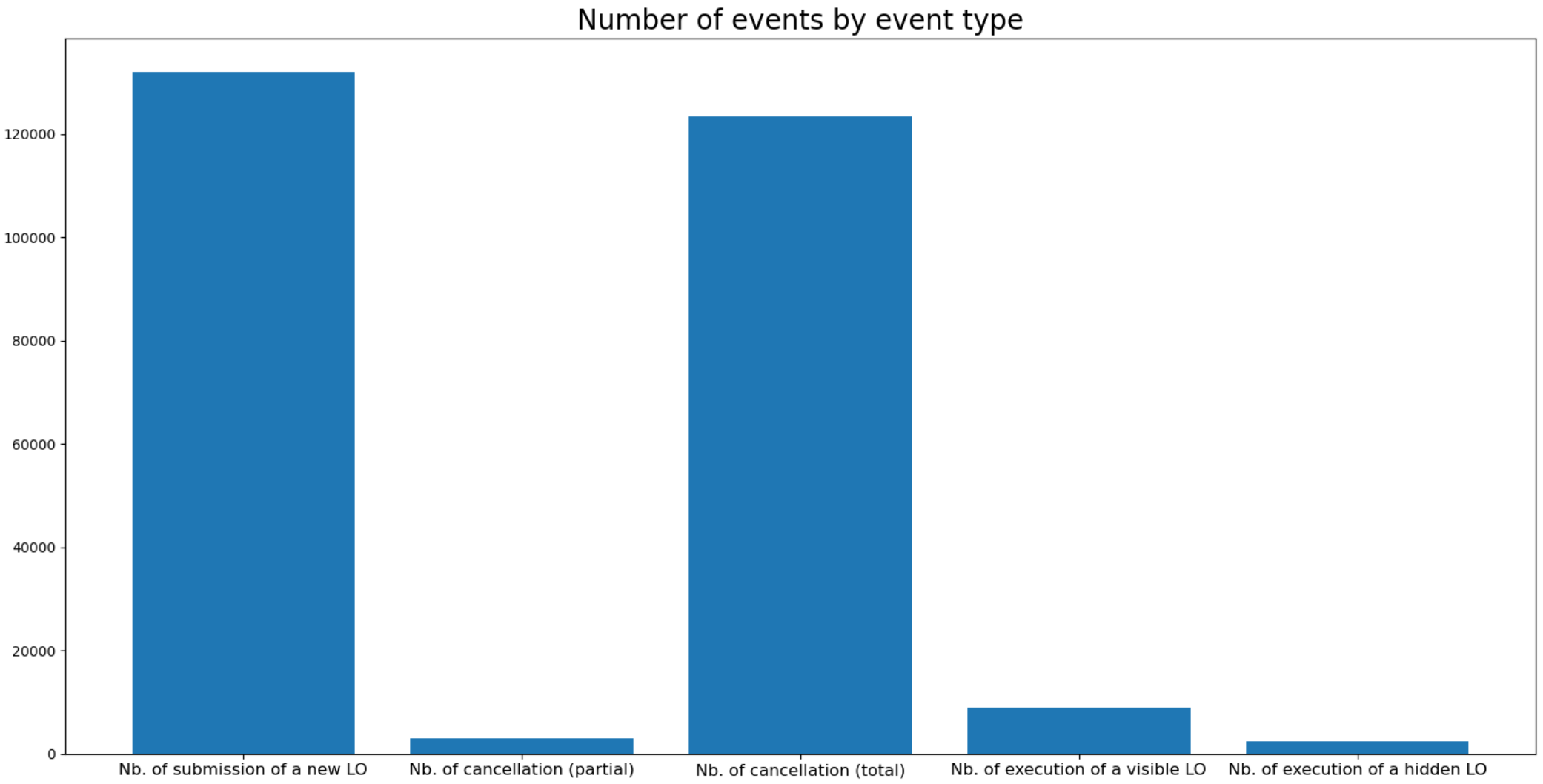
What I have found:

- Total number of event during the trading day 21/06/2012: **269 748**
- Most active trading hour: **3-4 pm**
- Second most active trading hour: **10-11 am**
- Average nb. of event in one minute of trading: **683**
- Min. nb. of order book event in one minute of trading: **0**
- Max. nb. of order book event in one minute of trading: **2 127**

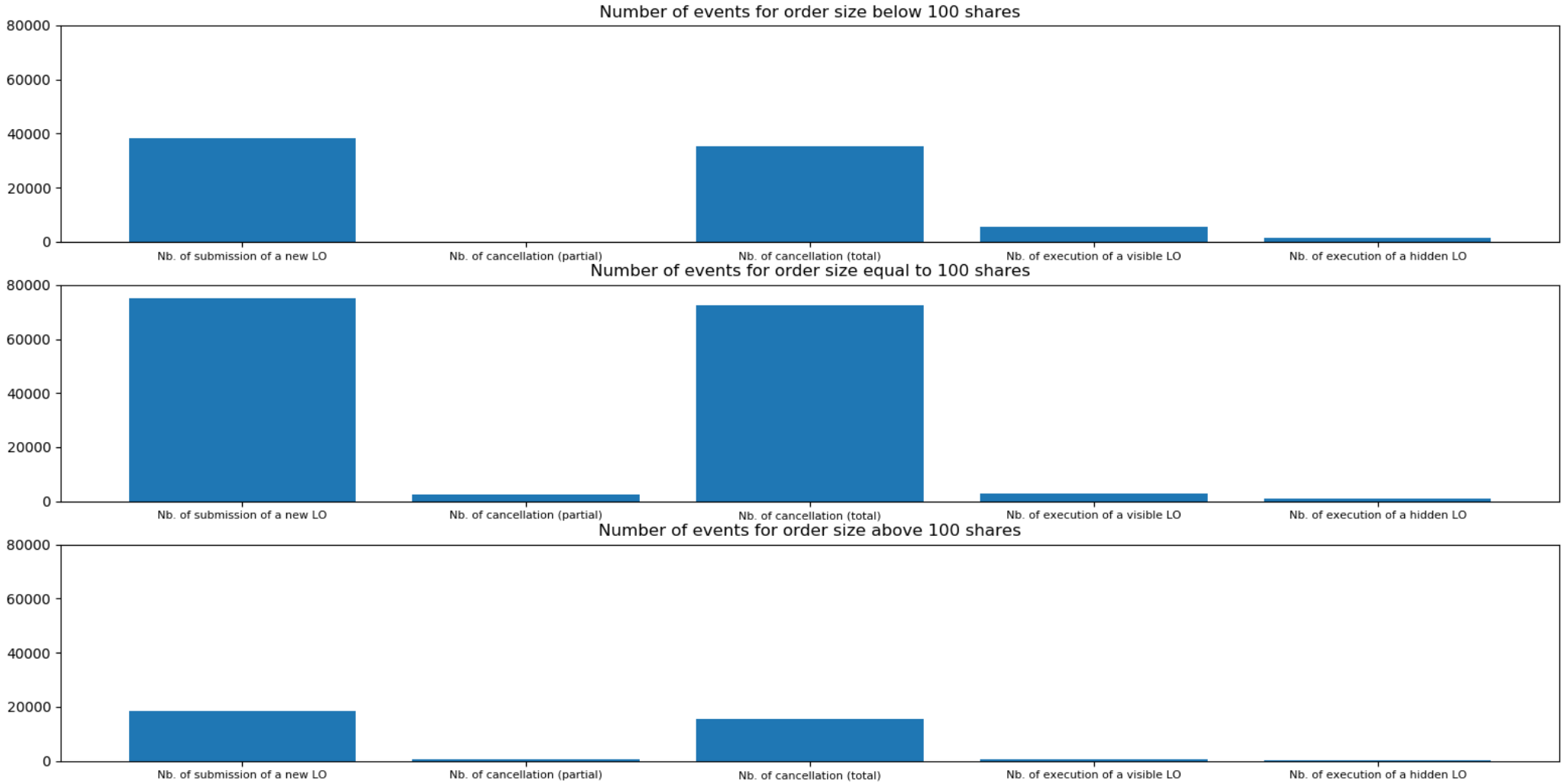
- Imbalance 1: Visible limit order: **-7%**
- Imbalance 2: Hidden limit order: **10%**
- Imbalance 3: Cancelled limit order: **-3%**

- Average time duration of cancelled trade: **5s**
- Median time duration of cancelled trade: **0.7s**

# THE MESSAGE FILE



# THE MESSAGE FILE
## SOLUTION



NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

# LOBSTER DATA
## *THE ORDER BOOK FILE*

| Ask Price 1 | Ask Size 1 | Bid Price 1 | Bid Size 1 | Ask Price 2 | Ask Size 2 | Bid Price 2 | Bid Size 2 | ... |
|---|---|---|---|---|---|---|---|---|
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1186600 | 9484 | 118500 | 8800 | 118700 | 22700 | 118400 | 14930 | ... |
| 1186600 | 9384 | 118500 | 8800 | 118700 | 22700 | 118400 | 14930 | ... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

**variable explanation.**

- Ask Price 1: Level 1 ask price (best ask price)
- Ask Size 1: Level 1 ask volume (best ask volume)
- Bid Price 1: Level 1 bid price (best bid price)
- Bid Size 1: Level 1 bid volume (best bid volume)
- Ask Price 2: Level 2 ask price (second best ask price)
- Ask Size 2: Level 2 ask volume (second best ask volume)
- ...

https://lobsterdata.com/info/DataStructure.php

What is the link between the message file and the order book file?

« *The k-th row in the 'message' file describes the limit order event causing the change in the limit order book from line k-1 to line k in the 'orderbook' file.* »

# PROGRAMMING EXERCISE 2

## *THE ORDER BOOK FILE*

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

✓ Download the order book file (Courses). What is the total number of levels?

**Step 1**

- Show bid and ask level along the order book at the beginning of the trading day
- What is the cumulative market depth at the beginning of the trading day?

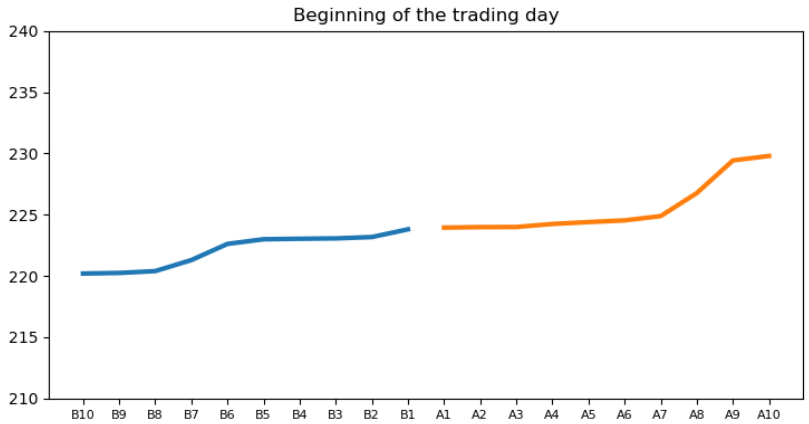
**Step 2**

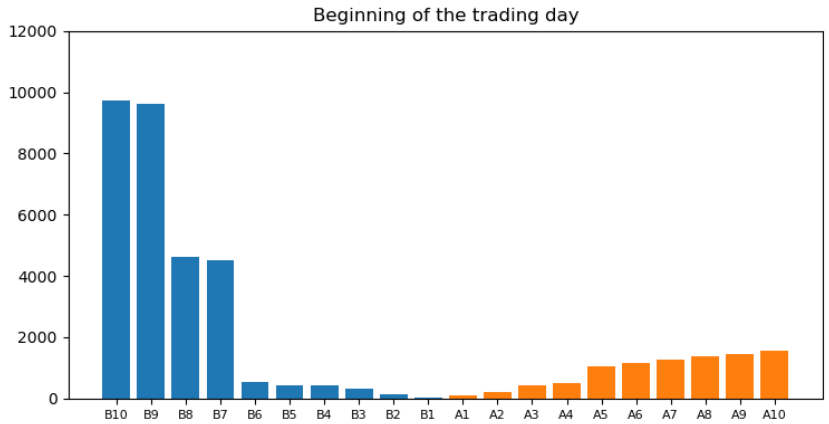
- Show bid and ask level along the order book at the middle of the trading day
- What is the cumulative market depth at the middle of the trading day?

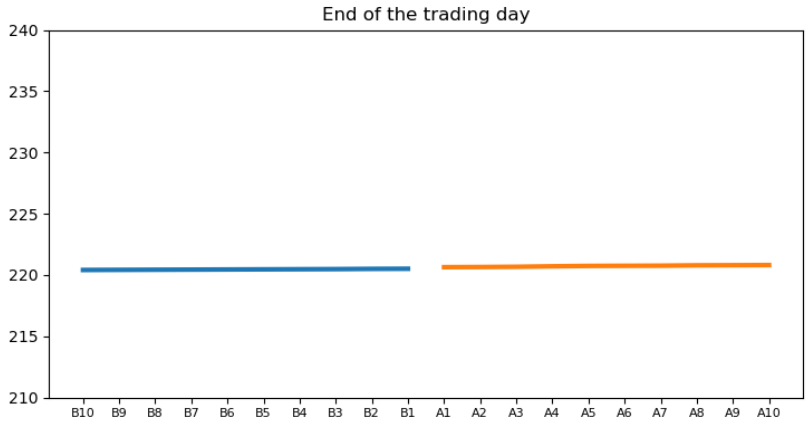
**Step 3**

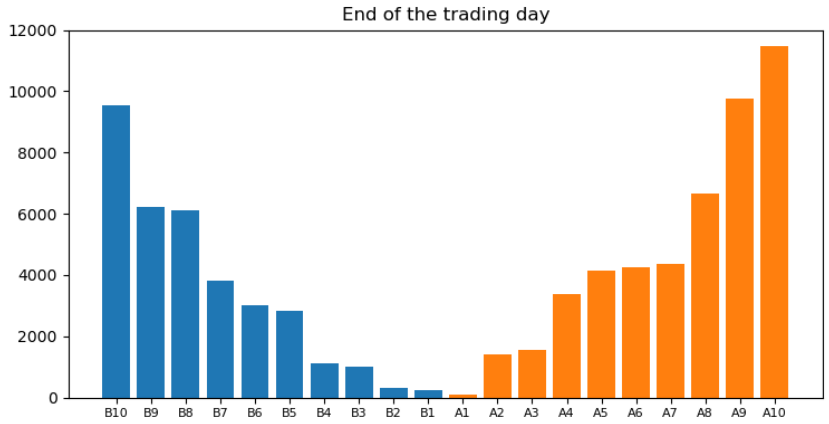
- Show bid and ask level along the order book at the ending of the trading day
- What is the cumulative market depth at the ending of the trading day?

# PROGRAMMING EXERCISE 2
## SOLUTION



# PROGRAMMING EXERCISE 3

## *MESSAGE FILE + LOB FILE*

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

### Option 1

***According to you, which event type impact the most the order book?***

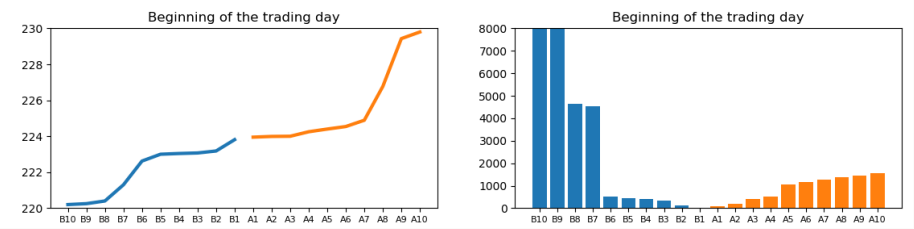
- Submission?
- Cancellation?
- Execution?

I do not provide you a straight methodology, find our own appraoch to tackle this question!

**Complexity: Advanced**

### Option 2

***Create a function that produces:***




**For any hour-minute-second combinaison**

**Ex.:**
**Generate_LOB('14:32:56')**

**Complexity: Moderate**

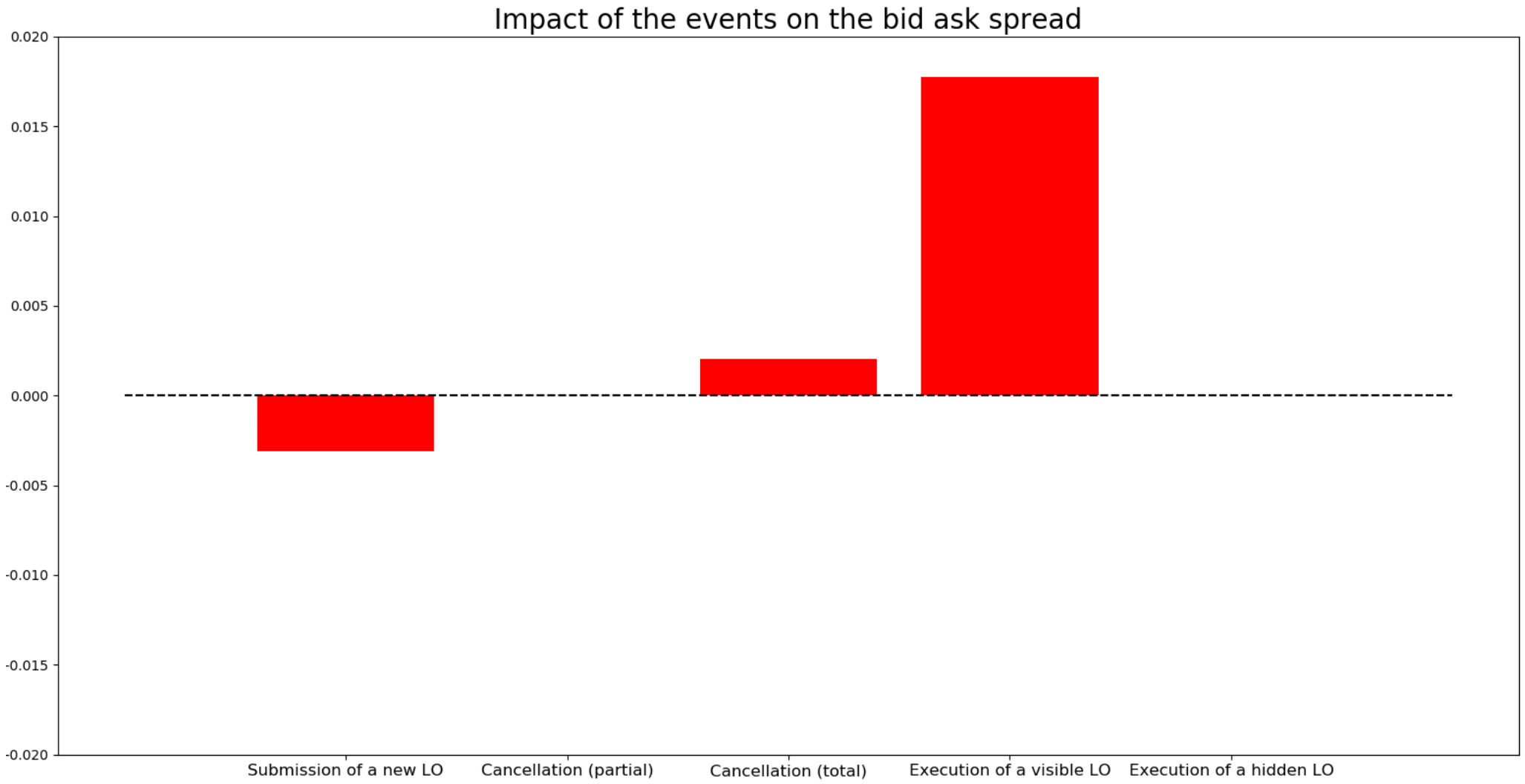
# PROGRAMMING EXERCISE 3

## *MESSAGE FILE + LOB FILE*

***According to you, which event type impact the most the order book?***

- Submission?
- Cancellation?
- Execution?

I do not provide you a straight methodology, find our own appraoch to tackle this question!




*How do you explain the impact "Execution of a visible LO" on the B/A spread?*

# PROGRAMMING EXERCISE 4

## THE TRADE FILE (1)

- Softwares
  - Suggested: Python
  - Accepted: Matlab or R….
- TAQ data from WRDS
- File: **UN_trades_full.xlsx**

1. Date
2. Time
3. Price
4. Volume
5. Exchange Code
6. Sales Condition
7. Correction Indicator
8. Sequence Number
9. Trade Stop Indicator
10. Source of Trade
11. MDS 127 / TRF *(Trade Reporting Facility)* (*)
12. Exclude Record

- Take a look to the TAQ file pdf for column details.

# PROGRAMMING EXERCISE 4
## THE TRADE FILE (2)

1. Filtering

2. How much time separates two transactions?
   i. Few seconds?
   ii. Few minutes?

3. Statistics by trading hour

4. Statistics by stock exchange

5. Generate a python summary file

### EXAMPLE OF PYTHON OUTPUT

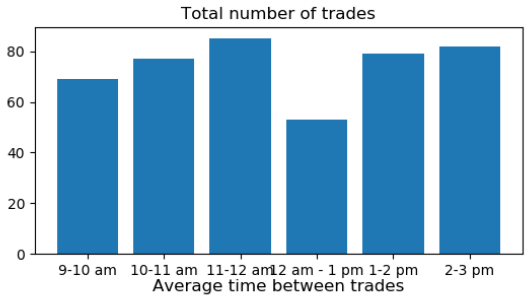
Ticker: AAA
Period: XX/XX/XXXX 09:30:00 – 16:00:00
Nb. of trades: 100
Total volume: 5,222.0
First price: 46.2
Last price: 47.7
Average trade size: 50
Nb. of seconds between trades: 5
Clusters of trades: Morning/Afternoon
.....

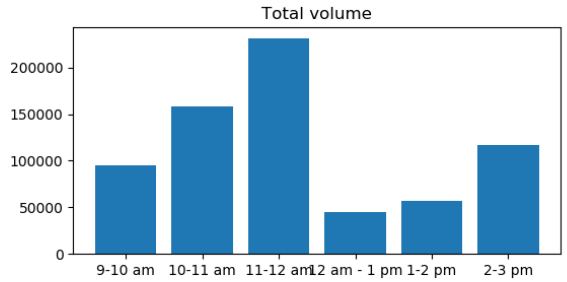
# SCREENSHOT FOR 02/05/1999

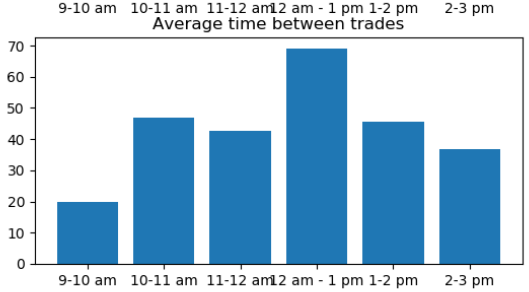
NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

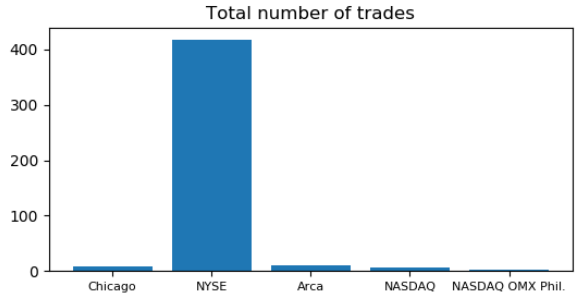
```
Ticker: ['UN']
Trading day: 1999-02-05 00:00:00
Time range: From 09:37:16 to
14:51:24
Nb. of trades: 445
```

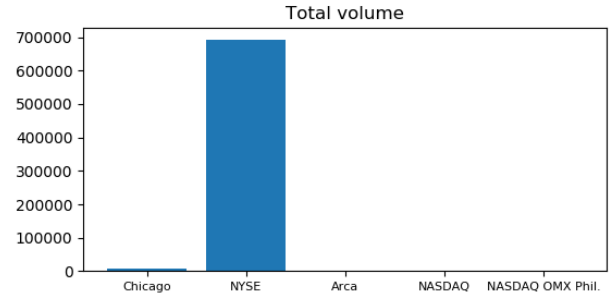
```
Total volume: 702600
average trade size 1579
First price: 72.5
Last price: 71.75
Nb. of seconds between trades: 42.0
```

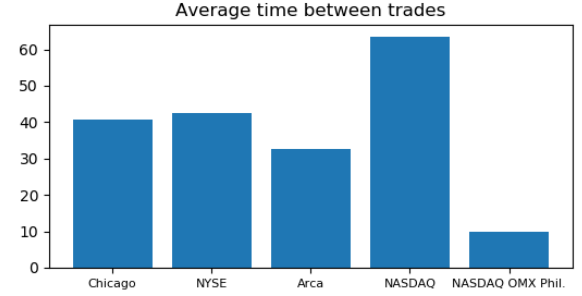


NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

# PROGRAMMING EXERCISE 5
## THE QUOTE FILE

File: **UN_quotes_full.xlsx**

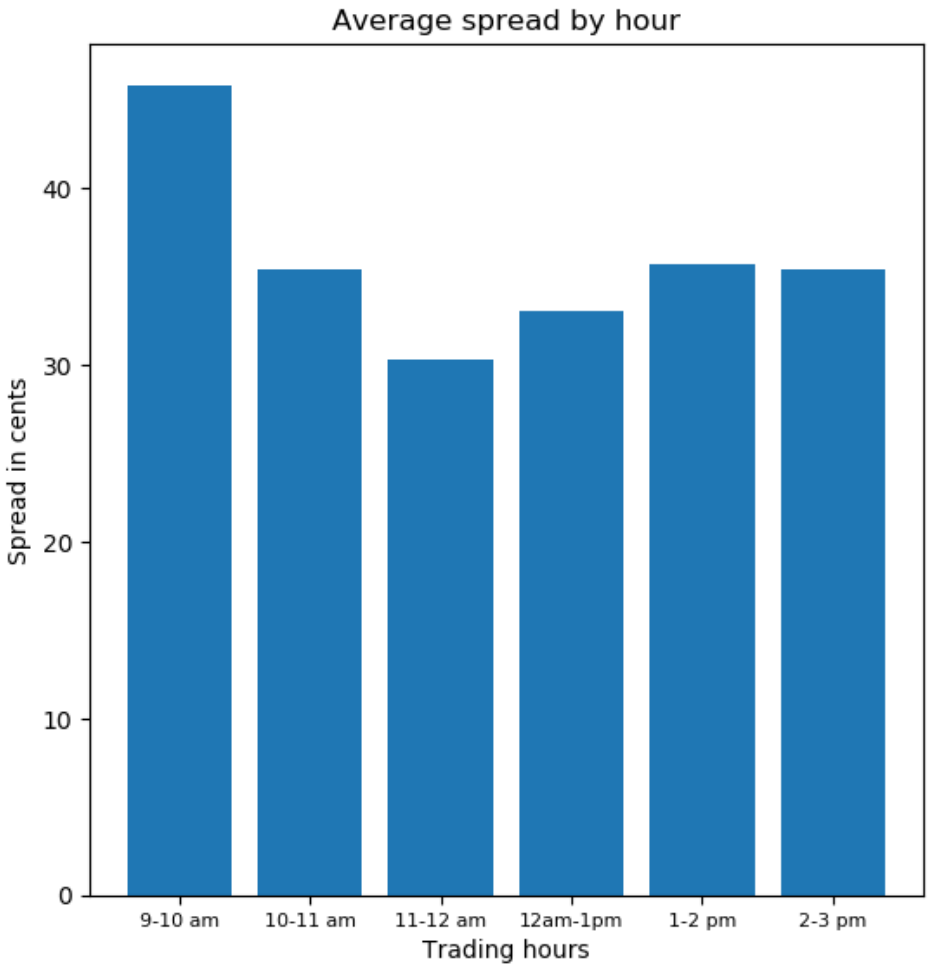
Compute for each market-maker:

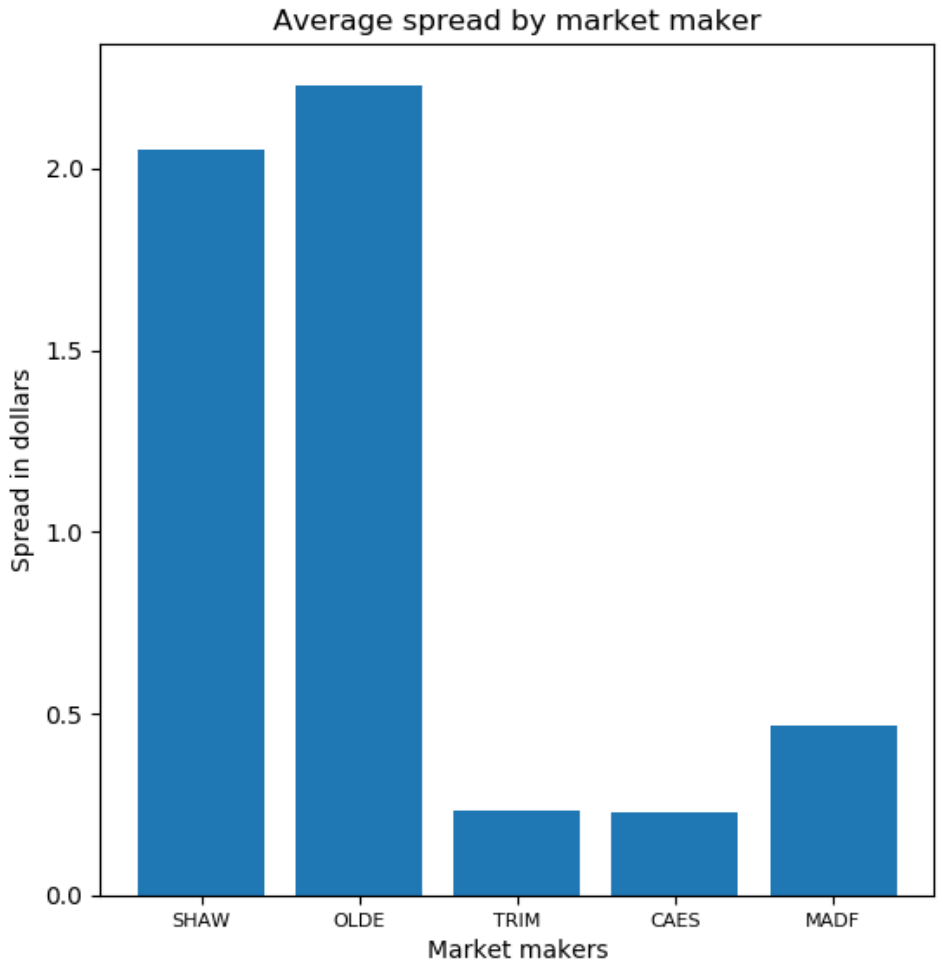
- The best bid of the day
- The best ask of the day
- The average bid ask spread
- More complex: average spread by hour

**EXAMPLE OF PYTHON OUTPUT**

Ticker: AAA
Period: XX/XX/XXXX 09:00:00 – 16:00:00
Best bid of the day: ???
Best ask of the day: ???
Average bid ask spread ($): ???
Average bid-ask spread (%): ???
.....

# SCREENSHOT FOR 08/03/1998
## THE QUOTE FILE



NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

# Next step: Merging the trade file with the quote file.

What can we learn from that?

- ✓ Trade initiation
- ✓ Trading cost
- ✓ Order-flow imbalance

# BEST BID AND ASK
## EXAMPLE 1

| Market maker | Timestamp | Bid | ASk |
|---|---|---|---|
| AA | 04/09/2020 10:12:03 | 60.3 | 62.3 |
| BB | 04/09/2020 10:12:04 | 60.5 | 62.5 |
| CC | 04/09/2020 10:12:05 | 60.0 | 63 |
| | | | |
| Trade | 04/09/2020 10:12:08 | 61 | |

- ✓ Best bid: 60.5
- ✓ Best ask: 62.3

# BEST BID AND ASK
## EXAMPLE 2

| Makert maker | Timestamp | Bid | ASk |
|---|---|---|---|
| AA | 04/09/2020 10:12:03 | 60.5 | 62.6 |
| BB | 04/09/2020 10:12:04 | 60.7 | 62.4 |
| CC | 04/09/2020 10:12:05 | 60.0 | 63 |
| AA | 04/09/2020 10:12:10 | 60.8 | 62.3 |
| | | | |
| Trade | 04/09/2020 10:12:10 | 61 | |
| | | | |

- ✓ Best bid: 60.7
- ✓ Best ask: 62.4

# BEST BID AND ASK

## EXAMPLE 3

| Makert maker | Timestamp | Bid | ASk |
|---|---|---|---|
| AA | 04/09/2020 10:12:03 | 60.9 | 62.0 |
| BB | 04/09/2020 10:12:04 | 60.8 | 62.1 |
| CC | 04/09/2020 10:12:05 | 60.0 | 63 |
| AA | 04/09/2020 10:12:10 | 60.7 | 62.1 |
| BB | 04/09/2020 10:12:11 | 61 | 61.9 |
| CC | 04/09/2020 10:12:11 | 61.1 | 61.8 |
| | | | |
| Trade | 04/09/2020 10:12:12 | 61 | |
| | | | |

- ✓ Best bid: 60.8
- ✓ Best ask: 62.1

# TRADE INITIATION

- I observe a transaction. Is there a way to know who has engaged the trade? Who were the most willing to trade? The buyer or the seller?

- Why is it important?

  - Computing new measures of trading costs
  - Computing order-flow imbalance
  - Quantifying informed-based trading

Some popular variables extracted from the microstructure of financial markets requires the detection of the counterparty that has engaged the trade, that is the the most impatient trader.

# TRADES CLASSIFICATION ALGORITHM

## THE TICK TEST & QUOTE TEST

### Tick rule

- A trade is classified as a buy (sell) if it is executed at a price higher (lower) than that of the previous trade.
- If this is the same price: Use the closest previous price as benchmark.

### Quote rule

- Based on location of the transaction price relative to the quote midpoints.
- If the transaction price is higher than the midpoint, a trade will be classified as a buy.
- If a transaction price is lower than the midpoint, a trade willbe classified as a sell.
- Trade at midpoint ==> It cannot be classified.

# BUYER-INITIATED AND SELLER INITATED TRADES

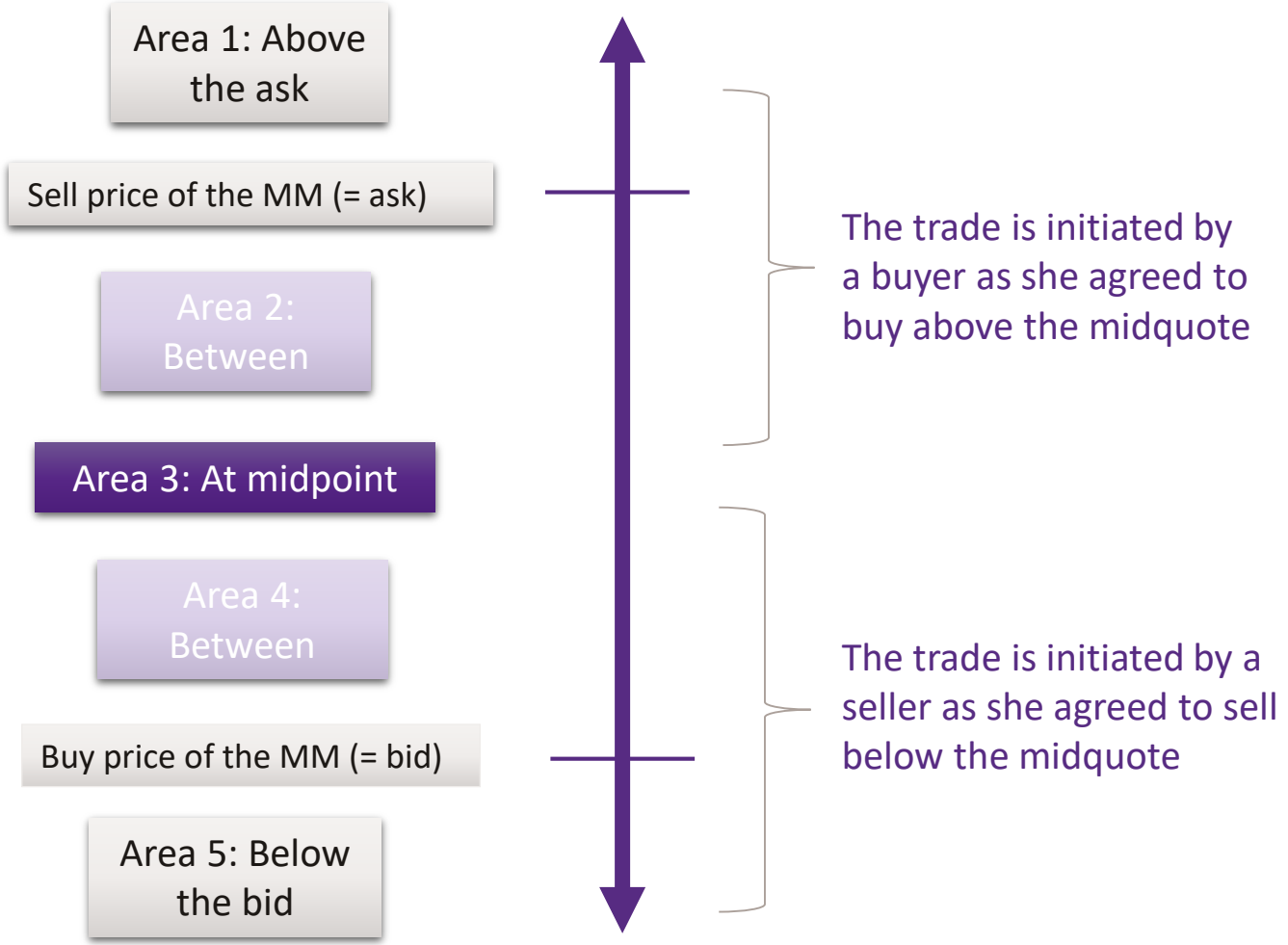
## THE QUOTE RULE

**Buyer-initiated trade**

- A trade is classified as buyer-initiated if it occurs around the ask price.

**Seller-initiated trade**

- A trade is classified as seller-initiated if it occurs around the bid price.



# MORE ADVANCED CLASSIFICATION
## ALGORITHM "CLNV"



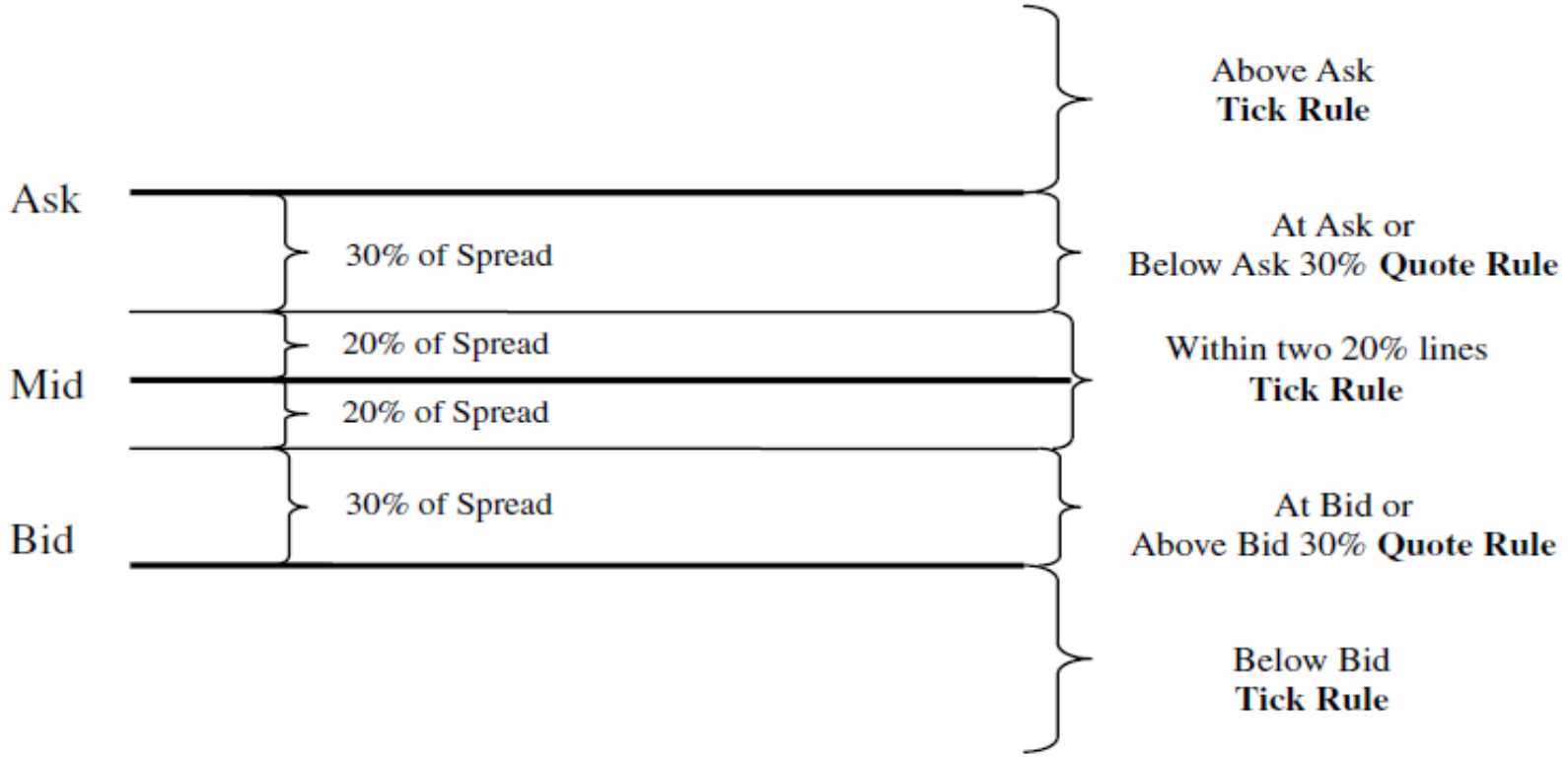



Fig. 1. Our alternative algorithm. (In order to obtain a simpler algorithm, we divided the spread into four quartiles. We use the quote rule for the two quartiles close to ask and bid and tick rule for the two quartiles close to quote midpoints. The overall accuracy rate of our alternative algorithm is 76.45% (not tabulated). One may sacrifice the improvement of accuracy of 0.07% by using this algorithm instead of the one proposed in this figure.

# ORDER-FLOW IMBALANCE

Once you have classified all the trades for a trading day, you can compute the order-flow imbalance:

$$OFI = \frac{|BIT - SIT|}{\frac{BIT + SIT}{2}}$$

- BIT: daily number of buyer-initiated trades
- SIT: daily number of seller-initiated trades

- It increases around corporate events (M&A announcements, earnings announcements …)
- It brings predictive power to explain future returns over very short horizon.

# SPREADs

With the ability to classify trades, new spread can be calculated:

1. Effective spread
2. Realized spread
3. Total price impact

➢ There is no one single formula

References:


- "*Do liquidity measures measure liquidity?'* by Goyenko, Holden and Trzcinka, JFE 2009.
- "*Latency, Liquidity and price discovery*" by Riordian and Storkenmaier, JFM 2012
- "*Do prices reveal the presence of informed trading*?" by Collin-Dufresne and Fos, JF 2015

# QUOTED SPREAD

$$SPR_t = A_t - B_t$$

- Consider a round-trip trade: The bid ask spread is what you loose for two trades.
- The half-spread is the transaction cost per trade.
- Empirical evidence: The wider the spread, the less liquid is the stock.

| | |
|---|---|
| √ *Advantage* | Very easy to calculate |
| ✕ *Drawback* | It ignores the effect of execution inside or outside the quote |

# EFFECTIVE HALF SPREAD

$$EHS = 2 \times D_j \times \left[\ln(P_j) - \ln\left(MQ_j^{before}\right)\right]$$

✓ This is a measure of the execution cost actually paid by the initiator of a transaction.

| **Variable definitions** | |
|---|---|
| $D_j$ | Binary variable that equals 1 for buyer-initiated orders and -1 for seller-initiated orders. |
| $P_j$ | Value-weighted average price ($P_j = \sum_i \frac{q_i}{Q} \times P_i$) |
| $MQ_j^{before}$ | A « benchmark » to select |

- ➢ Calendar-time based benchmark
  - ✓ Midquote point, few seconds/minutes before the transaction
- ➢ Transaction-time based benchmark
  - ✓ Midquote N quotations before the transaction
  - ✓ Alleviates differences in liquidity

# DECOMPOSITION OF THE EFFECTIVE HALF SPREAD

$$
\begin{aligned}
EHS &= 100 \times D_j \times \ln\left(\frac{P_j}{MQ_j^{before}}\right) \\
&= 100 \times D_j \times \ln\left(\frac{P_j}{MQ_j^{before}} \times \frac{MQ_j^{after}}{MQ_j^{after}}\right) \\
&= 100 \times D_j \times \ln\left(\frac{P_j}{MQ_j^{after}} \times \frac{MQ_j^{after}}{MQ_j^{before}}\right) \\
&= \underbrace{100 \times D_j \times \ln\left(\frac{P_j}{MQ_j^{after}}\right)}_{\text{Realized half spread}} + \underbrace{100 \times D_j \times \ln\left(\frac{MQ_j^{after}}{MQ_j^{before}}\right)}_{\text{Permanent price impact}}
\end{aligned}
$$

- ✓ **Component which is subsequently reversed**
- ✓ **Component not subsequently reversed**

# REALIZED HALF SPREAD

$$RHS = 2 \times D_j \times \left[\ln(P_j) - \ln\left(MQ_j^{after}\right)\right]$$

- ✓ Temporary component of the effective spread
- ✓ It measures the revenue of the liquidity suppliers net of the adverse selection costs imposed by the informed on the uninformed traders.

| **Variable definitions** | |
|---|---|
| $D_j$ | Binary variable that equals 1 for buyer-initiated orders and -1 for seller-initiated orders. |
| $P_j$ | Value-weighted average price ($P_j = \sum_i \frac{q_i}{Q} \times P_i$) |
| $MQ_j^{after}$ | Midquotes after the trade |

- ➢ Calendar-time based benchmark
  - ✓ Midquote few seconds/minutes after the transaction
- ➢ Transaction-time based benchmark
  - ✓ Midquote N quotations after the transaction
  - ✓ Alleviates differences in liquidity

# PERMANENT PRICE IMPACT

$$PPI = 2 \times D_j \times \left[\ln\left(MQ_j^{after}\right) - \ln\left(MQ_j^{before}\right)\right]$$

- ✓ Permanent component of the effective spread.
- ✓ It measures the information content of a trade.
- ✓ New information is conveyed by the trade and it impacts the midquote permanently.
  - ✓ unlike inventory costs and other trading costs which are temporary

| Variable definitions | |
|---|---|
| $D_j$ | Binary variable that equals 1 for buyer-initiated orders and -1 for seller-initiated orders. |
| $MQ_j^{before}$ | Midquote before the trade<br>➢ few seconds before |
| $MQ_j^{after}$ | Midquote after the trade<br>➢ several minutes after |

# BID/ASK SPREAD WITH LIMIT-ORDER BOOK

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

| | | |
|---|---|---|
| Quoted spread | $(Ask_t - Bid_t)/(0.5 \times Mid_t)$ | $Ask_t - Bid_t$ |
| Effective spread | $D_t \times (Price_t - Mid_t)/Mid_t$ | $2 \times \lvert Price_t - Mid_{t-s} \rvert$ |
| Realized spread | $D_t \times (Price_t - Mid_{t+x})/Mid_t$ | $2 \times D_t \times (Price_t - Mid_{t+x})$ |
| Price impact | $D_t \times (Mid_{t+x} - Mid_t)/Mid_t$ | $2 \times D_t \times (Mid_{t+x} - Mid_t)$ |

$+x$: few minutes after

"*Does Algorithmic trading improve liquidity?",* by Hendershoott, Jones and Mnekveld, JF 2011
"*Latency, liquidity and price discovery*" by Riordan and Storkenmaier, JFM 2012
"*Trading your neighbor's ETFs: Competition or fragmentation*" by Boehmer and Boehmer, JBF 2003

# FROM TRADE LEVEL TO TIME INTERVAL

- Spreads are computed at trade level (for each transaction).
- How do I aggregate for a given time scale?
  - Morning/afternoon/day
- Dollar-volume-weighted average.

$$\overline{SPREAD}_i = \sum_{k=1}^{K} \frac{q_{i,k}^{\$}}{Q_i^{\$}} \times SPREAD_{i,k}$$

with

$$Q_i^{\$} = \sum_{k}^{K} q_{i,k}^{\$}$$

and

$$q_k^{\$} = P_k \times V_k$$

- i: interval (seconds, minutes, hours)
- k: index
- K: total number of trades in the interval i
- P: transaction price of trade k
- V: volume of the trade k in shares
- Q: volume in dollars

NEOMA
BUSINESS SCHOOL
REIMS · ROUEN · PARIS

# 05

# Final Project

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

# Almost done!

Please allocate the last two hours to your final project.

# FINAL PROJECT
## Instructions

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

- Select one of the corporation below
- Send an email to antoine.noel@neoma-bs.fr to book.
- First come first serve.
- Softwares: Each final project must be done on Python (compulsory).
- Groups of 3, 4 or 5 students
- If a one or several students in your group do not work enough, please inform the course coordinator at and a bonus/malus system will be created.

- Deadline: **06/15/2022 11.59 pm**

# DATASET ALLOCATION

| Corporation name | Ticker | Starting date | Ending date | Group? |
|---|---|---|---|---|
| *Walt Disney* | DIS | 08/03/1998 | 04/30/1999 | |
| *General Electrics* | GE | 08/03/1998 | 04/30/1999 | |
| *Nike* | NKE | 08/03/1998 | 04/30/1999 | |
| *Morgan Stanley* | MWD | 08/03/1998 | 04/30/1999 | |
| *IBM* | IBM | 08/03/1998 | 04/30/1999 | |
| *Pfizer* | PFE | 08/03/1998 | 04/30/1999 | |
| *Unilever* | UN | 08/03/1998 | 04/30/1999 | |
| *Chevron* | CHV | 08/03/1998 | 04/30/1999 | |
| *ETF* | QQQ | 08/03/1998 | 04/30/1999 | |

# FINAL PROJECT

**(DEADLINE: 06/15/2022 11.59 pm)**

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

## *Step 1*

- Create your group (3-4-5 students) and select a corporation.

## *Step 2*

- Merge the trade dataset ("Trade_File_XX.xlsx") with the quote dataset ("Quote_File_XX.xlsx") for one trading day.

  - ✓ Identify the most active MM*. Keep only its bids and asks in the quote file.
  - ✓ Choose a delay of one second to merge the trades with their prevailing bid ask pair.
  - ✓ Python function for merging: **pandas.merge_asof**
  - ✓ https://pandas.pydata.org/pandas-docs/version/0.25.0/reference/api/pandas.merge_asof.html

**: largest amount of bid and ask updates.*

# FINAL PROJECT

**(DEADLINE: 06/15/2022 11.59 pm)**

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

## *Step 3*

- Now, for each transaction, you have now a prevailing bid and a prevailing ask. You can now classify trades. To do so, please implement the CLNV algorithm (slide 120).

## *Step 4*

- Compute the proportion of trades inside quotes, outside quotes and at quotes.

## *Step 5*

- Compute the daily weighted average effective spread.

# FINAL PROJECT

**(DEADLINE: 06/15/2022 11.59 pm)**

NEOMA BUSINESS SCHOOL
REIMS · ROUEN · PARIS

## Step 6

- Repeat step 2 to step 5 for each trading day of your dataset.

## Step 7

- Generate the effective spread and order-flow imbalance time series. Show these two time series on a single plot.

## *Step 8 (BONUS)*

- For each transaction, all prevailing bids and asks are not the ones of the most active MM.
- Rather, keep all market makers and please select the highest bid and the lowest ask among the most contemporaneous bid / ask pair of each market maker (slide 116).