HUAWEI-NUS Innovation Challenge 2023

Enablement Session 1

16 February 2023

In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

Agenda

1. Data Description
2. Demo1: A simple time average strategy
3. Demo2: A simple prediction strategy
4. Frequently Asked Questions

In partnership with:

Ranking- Demo Show

1. Data Description

Table 1 Tickdata File Data Field Description

Serial Number	Field Name	Field Description	Serial Number	Field Name	Field Description
1	COLUMN01	Index	**29**	COLUMN29	Buying Price 2
2	COLUMN02	Stock code	**30**	COLUMN30	Buying Price 3
3	COLUMN03	Tick time (accurate to milliseconds)	**31**	COLUMN31	Buying Price 4
4	COLUMN04	Opening price of the day	**32**	COLUMN32	Buying Price 5
5	COLUMN05	Highest price of the day as of the tick	**33**	COLUMN33	Buying Price 6
6	COLUMN06	Lowest price of the day as of the tick	**34**	COLUMN34	Buying Price 7
7	COLUMN07	Latest transaction price of the tick	**35**	COLUMN35	Buying Price 8
8	COLUMN08	Selling Price 1	**36**	COLUMN36	Buying Price 9
9	COLUMN09	Selling Price 2	**37**	COLUMN37	Buying Price 10
10	COLUMN10	Selling Price 3	**38**	COLUMN38	Buying Volume 1
11	COLUMN11	Selling Price 4	**39**	COLUMN39	Buying Volume 2
12	COLUMN12	Selling Price 5	**40**	COLUMN40	Buying Volume 3
13	COLUMN13	Selling Price 6	**41**	COLUMN41	Buying Volume 4
14	COLUMN14	Selling Price 7	**42**	COLUMN42	Buying Volume 5
15	COLUMN15	Selling Price 8	**43**	COLUMN43	Buying Volume 6
16	COLUMN16	Selling Price 9	**44**	COLUMN44	Buying Volume 7
17	COLUMN17	Selling Price 10	**45**	COLUMN45	Buying Volume 8
18	COLUMN18	Selling Volume 1	**46**	COLUMN46	Buying Volume 9
19	COLUMN19	Selling Volume 2	**47**	COLUMN47	Buying Volume 10
20	COLUMN20	Selling Volume 3	**48**	COLUMN48	Accumulated transaction volume to the tick on the day
21	COLUMN21	Selling Volume 4	**49**	COLUMN49	Accumulated transaction amount to the tick on the day
22	COLUMN22	Selling Volume 5	**50**	COLUMN50	Total commissioned buying volume of the tick
23	COLUMN23	Selling Volume 6	**51**	COLUMN51	Total commissioned selling volume of the tick
24	COLUMN24	Selling Volume 7	**52**	COLUMN52	Limit-up price of the day
25	COLUMN25	Selling Volume 8	**53**	COLUMN53	Limit-down price of the day
26	COLUMN26	Selling Volume 9	**54**	COLUMN54	Weighted average commissioned buying price of the tick
27	COLUMN27	Selling Volume 10	**55**	COLUMN55	Weighted average commissioned selling price of the tick
28	COLUMN28	Buying Price 1			

1. Data Description

COLUMN01: Index. Start from 0.

COLUMN02: Stock code. (e.g. 601888.SH, 002056.SZ)
Each stock has a unique code for its own.
There are two types of suffixes in the codes. “.SH” indicates that this stock is issued by Shanghai Exchange. “.SZ” means that this stock is issued by Shenzhen Exchange.

COLUMN03: Tick time. (e.g. 93000000).
Integer format. Accurate to milliseconds.

COLUMN04: Opening price of the day. (e.g. 1950800)
For each single stock, all values on this column are the same.
Notice that all the price values in the tickdata file has been multiplied by 10000.
For example, 1950800 is actually refers to ¥195.08.

1. Data Description

COLUMN05: Highest price of the day as of the tick.
For each single stock, values on this column are monotonically increasing.

COLUMN06: Lowest price of the day as of the tick.
For each single stock, values on this column are monotonically decreasing.

COLUMN07: Latest transaction price of the tick.
The most current price that this stock is trading at.

COLUMN08 – 47: Selling prices, selling volumes, buying prices, buying volumes.
The selling prices and buying prices are also known as ask prices and bid prices.

1. Data Description

Table 2 An example of selling (buying) prices and volumes

Selling Price10	2.77	Selling Volume10	509100
Selling Price9	2.76	Selling Volume9	369900
Selling Price8	2.75	Selling Volume8	437700
Selling Price7	2.74	Selling Volume7	189200
Selling Price6	2.73	Selling Volume6	280800
Selling Price5	2.72	Selling Volume5	320600
Selling Price4	2.71	Selling Volume4	689200
Selling Price3	2.7	Selling Volume3	300600
Selling Price2	2.69	Selling Volume2	280700
Selling Price1	2.68	Selling Volume1	23804
Latest Price	**2.67**		
Buying Price1	2.67	Buying Volume1	55873
Buying Price2	2.66	Buying Volume2	205400
Buying Price3	2.65	Buying Volume3	611500
Buying Price4	2.64	Buying Volume4	401000
Buying Price5	2.63	Buying Volume5	367600
Buying Price6	2.62	Buying Volume6	358300
Buying Price7	2.61	Buying Volume7	499200
Buying Price8	2.6	Buying Volume8	355000
Buying Price9	2.59	Buying Volume9	27400
Buying Price10	2.58	Buying Volume10	160200

10 lowest prices that sellers are willing to accept for the stock.

corresponding volumes

- An example from the tick of 2022/8/5, 9:30:09:000, 600166.SH.
- The prices and volumes are only valid for this tick, and may change rapidly on the next tick.

10 highest prices that buyers are willing to pay for the stock.

corresponding volumes

In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

1. Data Description

COLUMN48: Accumulated transaction volume to the tick on the day.
For each single stock, values on this column are monotonically increasing.
The exact transaction volume on the current tick, can be calculated by the difference between two adjacent rows.

For example, in tickdata_20220805.csv:
Value at row 365 and column 48 (600166.SH, 9:30:00:000) is: 242500
Value at row 947 and column 48 (600166.SH, 9:30:03:000) is: 455000
Then we can say that the transaction volume on tick 947 is: 455000 – 242500 = 212500 (shares)
That means from 9:30:00:000 to 9:30:03:000, there are 212500 shares of 600166.SH have been traded (but we don't know the exact trading prices for all transactions within this 3 seconds).

COLUMN49: Accumulated transaction amount to the tick on the day.
Similar to COLUMN48.
For example, in tickdata_20220805.csv:
The transaction amount on tick 947 is: 1215301 – 647492 = 567809 (¥)

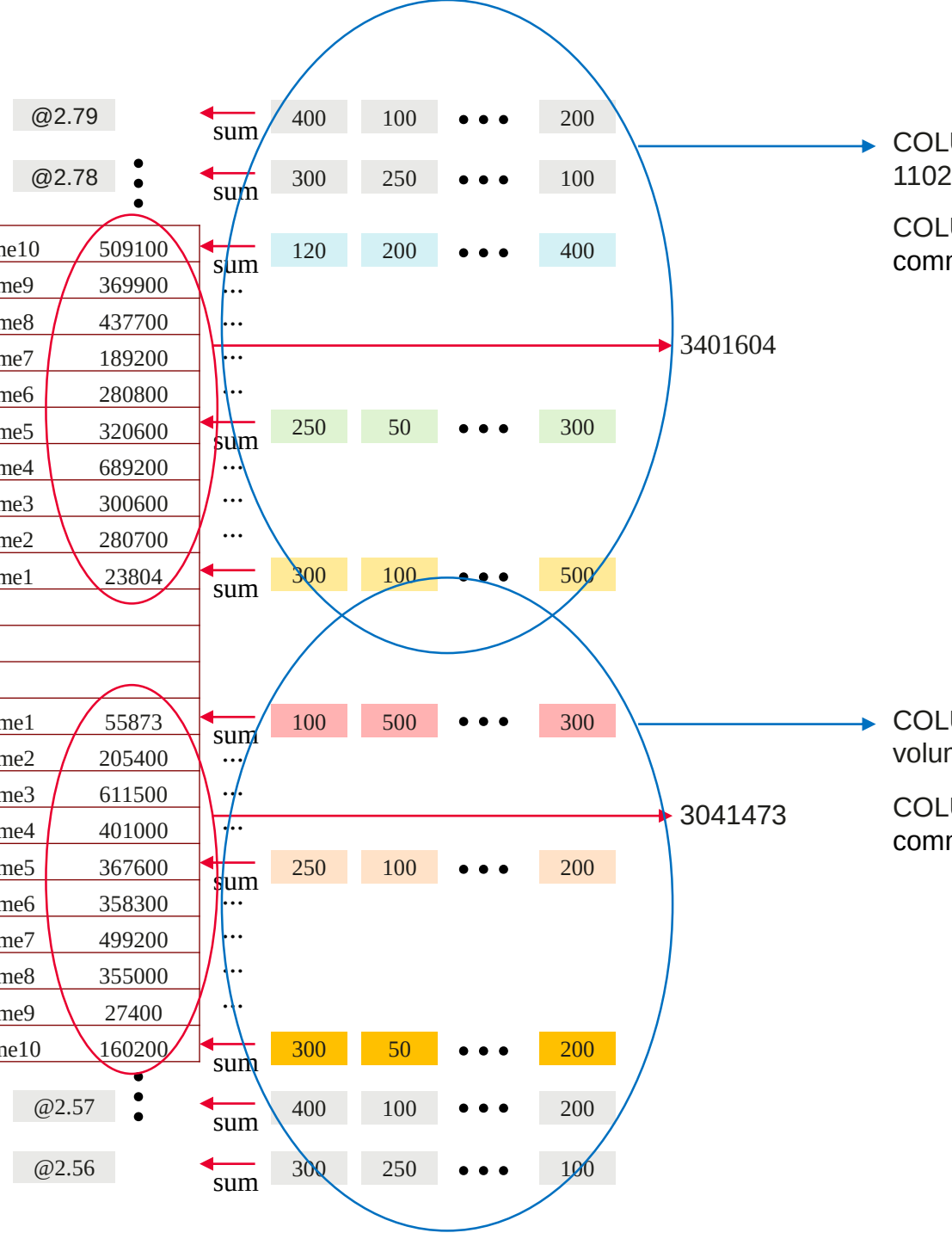
1. Data Description

COLUMN50: Total commissioned buying volume of the tick.
For example, 2022/8/5, 9:30:09:000, 600166.SH:
Value at column50: 3883973
Buying Volume1 + Buying Volume2 + … + Buying Volume10 = 3041473 < 3883973

COLUMN51: Total commissioned selling volume of the tick.
For example, 2022/8/5, 9:30:09:000, 600166.SH:
Value at column51: 11020904
Selling Volume1 + Selling Volume2 + … + Selling Volume10 = 3401604 < 11020904

1. Data Description

Selling Price10	2.77	Selling Volume10	509100
Selling Price9	2.76	Selling Volume9	369900
Selling Price8	2.75	Selling Volume8	437700
Selling Price7	2.74	Selling Volume7	189200
Selling Price6	2.73	Selling Volume6	280800
Selling Price5	2.72	Selling Volume5	320600
Selling Price4	2.71	Selling Volume4	689200
Selling Price3	2.7	Selling Volume3	300600
Selling Price2	2.69	Selling Volume2	280700
Selling Price1	2.68	Selling Volume1	23804
Latest Price	**2.67**		
Buying Price1	2.67	Buying Volume1	55873
Buying Price2	2.66	Buying Volume2	205400
Buying Price3	2.65	Buying Volume3	611500
Buying Price4	2.64	Buying Volume4	401000
Buying Price5	2.63	Buying Volume5	367600
Buying Price6	2.62	Buying Volume6	358300
Buying Price7	2.61	Buying Volume7	499200
Buying Price8	2.6	Buying Volume8	355000
Buying Price9	2.59	Buying Volume9	27400
Buying Price10	2.58	Buying Volume10	160200



COLUMN51: sum of all commissioned volumes, 11020904

COLUMN55: weighted average price of all the commissioned orders, 2.8430

COLUMN50: sum of all commissioned volumes, 3883973

COLUMN54: weighted average price of all the commissioned orders, 2.5970

In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

1. Data Description

COLUMN52: Limit-up price of the day.
A fixed value. Trading price in the current day cannot be higher than the Limit-up price.

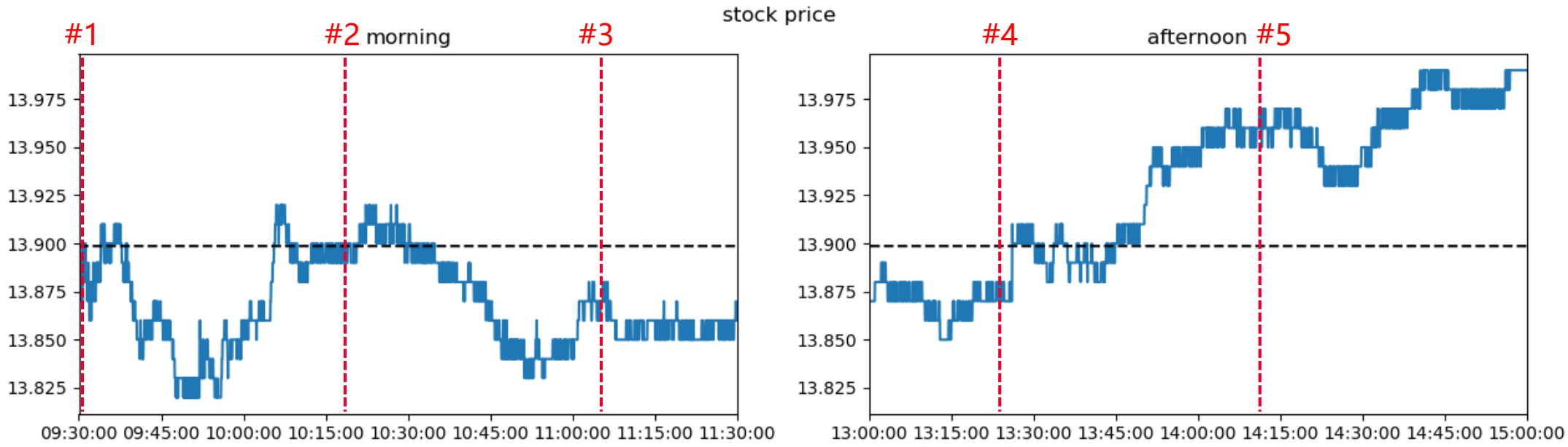
COLUMN53: Limit-down price of the day.
A fixed value. Trading price in the current day cannot be lower than the Limit-down price.

COLUMN54: Weighted average commissioned buying price of the tick.

COLUMN55: Weighted average commissioned selling price of the tick.

2. Demo1: A simple time average strategy

- Break the target volume into equal parts and execute the order equally across a day.
- *od_nCount:* the number of parts



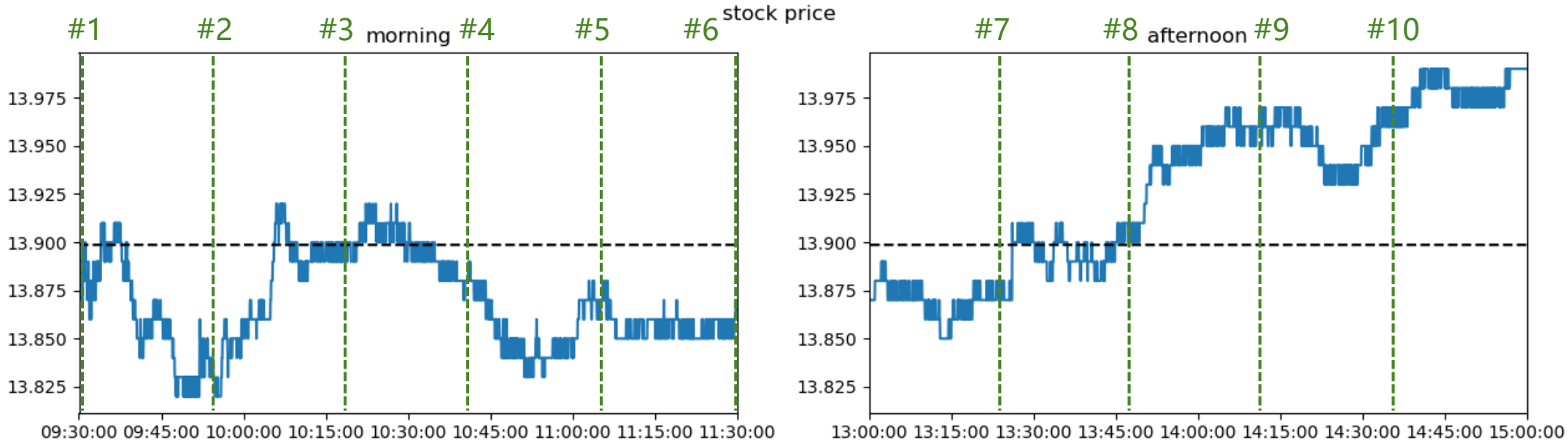
Buy: od_nCount = 5
(Buying every: 240 / 5 = 48 minutes)

In partnership with: Audaque 华傲数据

HUAWEI CLOUD
EVERYTHING AS A SERVICE

2. Demo1: A simple time average strategy

- The buying method and selling method are almost the same here, except the number of parts *od_nCount*.



Sell: od_nCount = 10
(Selling every: 240 / 10 = 24 minutes)

In partnership with: Audaque 华傲数据

HUAWEI CLOUD
EVERYTHING AS A SERVICE

3. Demo2: A simple prediction strategy

- Build a machine learning model that can predict the trends after 5 minutes.
- Then execute a order every 5 minutes according to the prediction result.
- If the result y>=0, that means the price will probably rise.
- Then buy in at the start of this round and meanwhile delay the sell order to the next round.
- If the result y<0, that means the price will probably fall.
- Then sell out at the start of this round and meanwhile delay the buy order to the next round.

In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

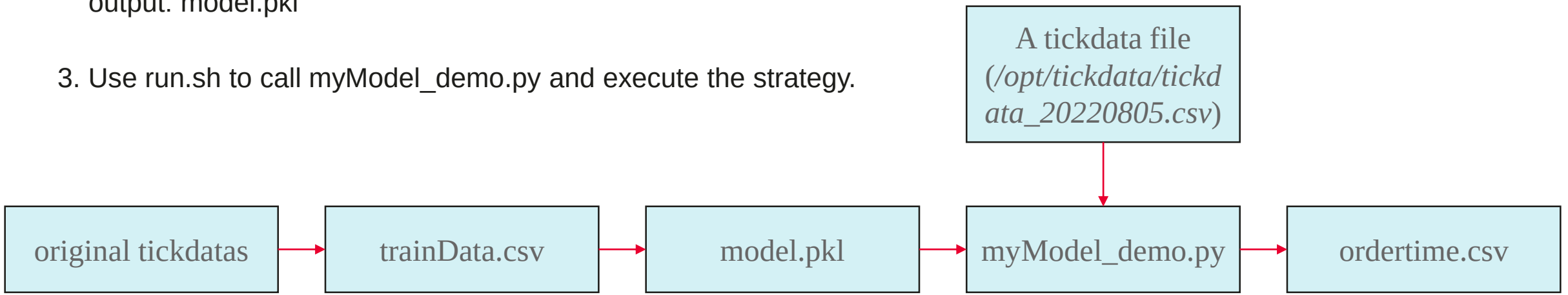
3. Demo2: A simple prediction strategy

Work flow:

1. Use tick2trainData.py to read original tickdata files and produce a single trainData file for training.
   The trainData file includes 11 columns.
   Column y is the price change rate of the next 5 minutes.
   Column x1 to x10 are 10 factors used to predict y.
   output: trainData.csv

2. Use trainModel.py to train a prediction model.
   output: model.pkl

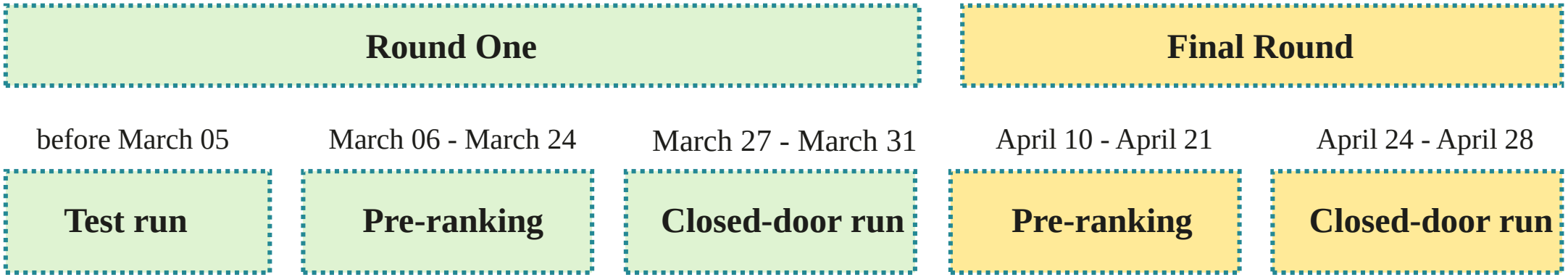
3. Use run.sh to call myModel_demo.py and execute the strategy.



4. Frequently Asked Questions

Q1. Are the pre-ranking phase daily results counted as part of the selection criteria?

A1: No, the daily results in the pre-ranking phase (both in Round One and Final Round) will not be counted as part of the selection criteria. They are only used for reference, which may help you to figure out how well your program performs. Only the results in the closed-door phase will be counted for selection.

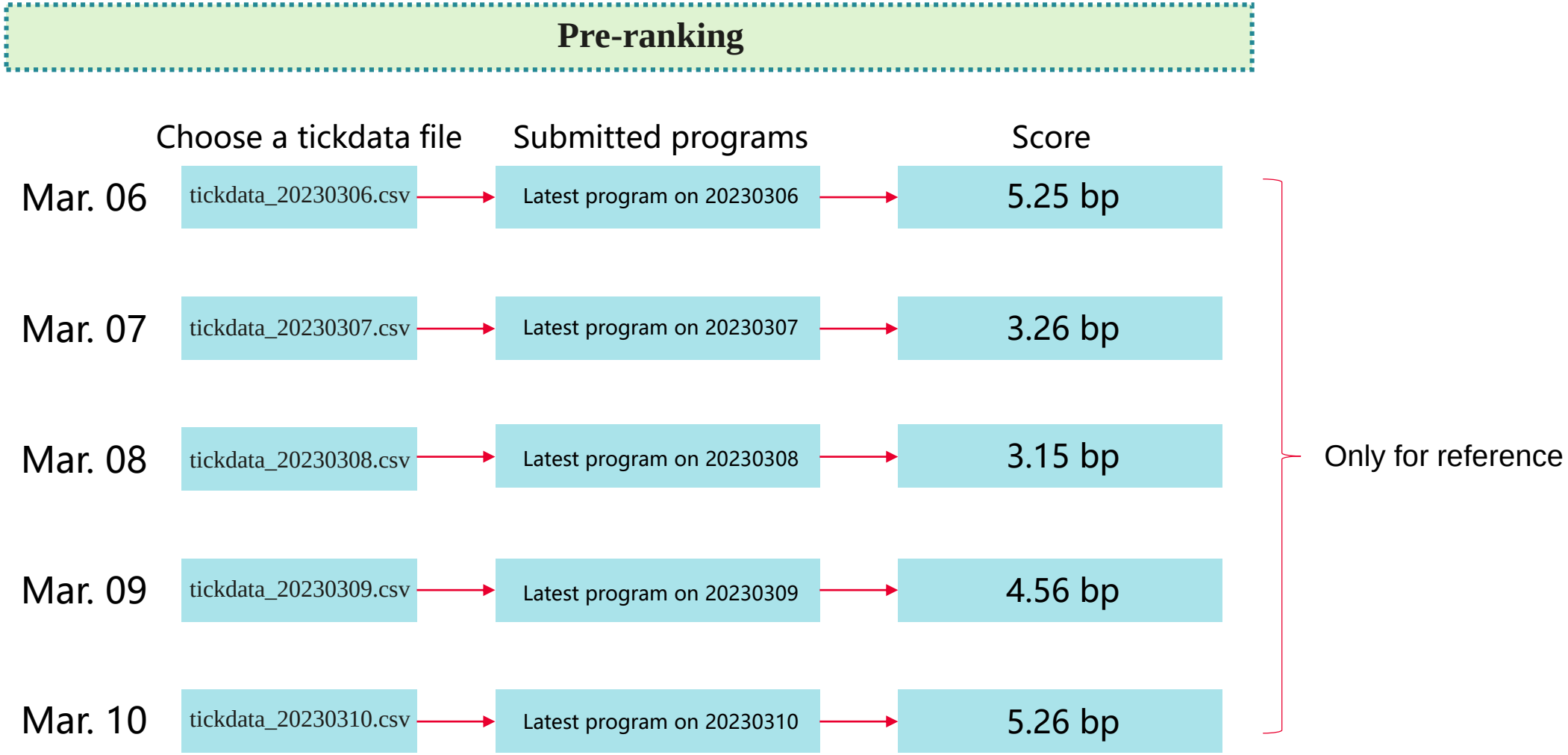


4. Frequently Asked Questions

Q2: What are the released dates for the trading data and are they real time data?

A2: In pre-ranking phase, the latest trading data will be collected every day. Then we will run the programs of all teams. The ranking is given based on the latest data every day. In close-door phase, the ranking is given based on the average score of 5 trading days, and it will be used as the evaluation criterion for the final.

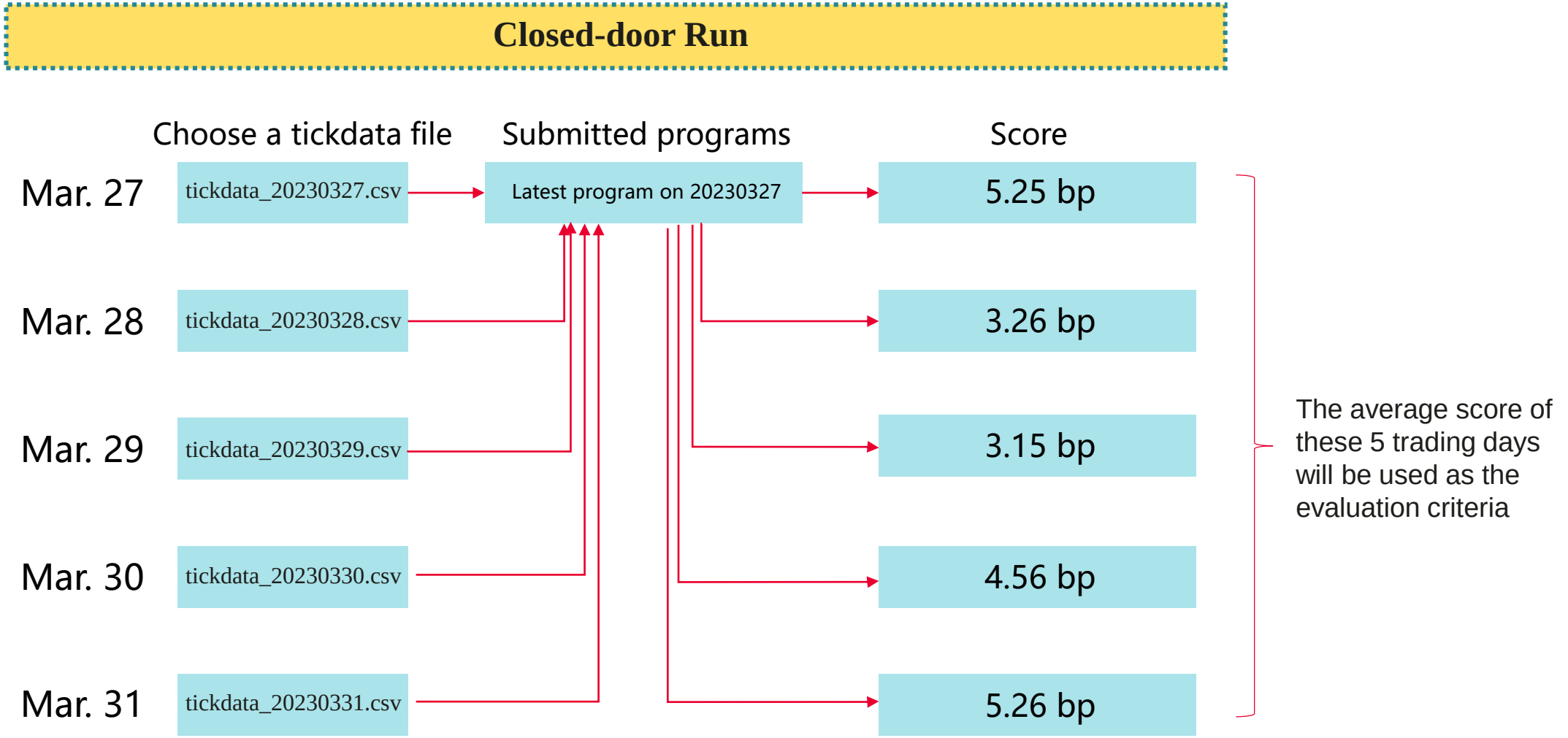
4. Frequently Asked Questions



In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

4. Frequently Asked Questions



In partnership with: Audaque 华傲数据

HUAWEI CLOUD EVERYTHING AS A SERVICE

4. Frequently Asked Questions

Q3: How is the 100 days trading data relevant to the solution model?

A3: The given dataset contains real original trading data of 500 stocks in 100 days with rich trading details. You can build your solution model by data statistics or machine learning base on the dataset. The competition does not limit how to use the dataset. It is not necessary to use all fields in the dataset, as some fields could be redundant and has little effect on model building. It all depends on your idea. The given two demos may give you some inspiration.

Q4: Can participants use external data sources to improve their solution?

A4: Yes. We encourage participants to use external data sources, as long as the data is useful for the solution.

100day tickdata files √

External dataset √

Cannot be too large (maximum 10G)
Must proved to be useful for the solution

4. Frequently Asked Questions

Q5: Can participants get a pre-trading data to improve their solution?

A5: Yes, you can use external data sources, including pre-trading data to improve your model. However, notice that the data should be a closed dataset. That means the data cannot be changed once the solution is submitted.

Q6: Are the participants supposed to analyze the 100days data and daily historical data to decide when to buy and sell the shares?

A6: It is not mandatory to use the 100days data. You can decide to use the data or not. You can also use external data. However, the purpose of this competition is to discover how to make full use of historical data. So if the 100days data is used, the innovation score (Innovation of solution, 30%) in the final round will be higher.

4. Frequently Asked Questions

Q7: Are we required to buy and sell exactly 100 shares for all the 500 stocks every day, no more and no less?

A7: Yes, contestants should buy and sell exactly 100 shares for each stock in the tickdata file, no more and no less. However, the number of stocks in the tickdata file can be different, not always equals to 500. For example, the organizer may input a tickdata file that only contains 100 stocks. The model should be adaptable to the tickdata file and buy and sell exactly 100 shares for those 100 stocks.

Q8: Can we short sell a stock, and does evaluation.py allow us to short sell?

A8: Yes, short sell is allowed. More precisely, the buying and selling tasks are independent. You can imagine that they are in different accounts, even if they are trading the same stock. The evaluation.py will process the buying and selling tasks separately.

4. Frequently Asked Questions

Q9: Are we allowed to use the entire dateset (all 100 files) to train our model?

A9: Yes, contestants can use the entire dateset for both model training and strategy running.

Q10: We installed xgboost (ver 1.7.3) for line 77 in the demo model 2. But after that we encountered an error message when we try to execute the file. We would like to ask if we are using the wrong library or setting? If not, how may we resolve this issue?

A10: The error is caused by version inconsistency. The given model.pkl is trained by python xgboost package, version 1.5.1, and it is incompatible with version 1.7.3. You can either train another model.pkl with version 1.7.3, or just downgrade xgboost to version 1.5.1.

4. Frequently Asked Questions

If you have more questions pertaining to the participation of the innovation challenge, email msba@nus.edu.sg with the subject title “Huawei-NUS Innovation Challenge 2023”.

The reply will be released on:
https://www.sg-innovationchallenge.org/faqs

Questions and Answers

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and organization for a fully connected, intelligent world.

HUAWEI | NUS National University of Singapore

HUAWEI CLOUD
EVERYTHING AS A SERVICE