

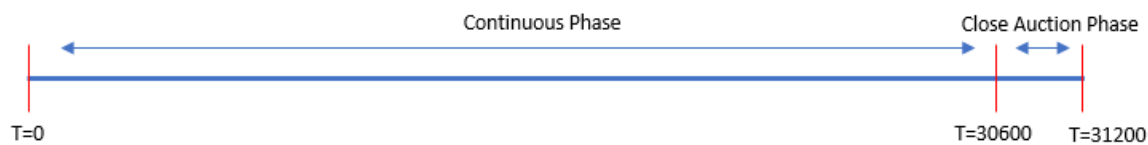
[QUANT] Predict Closing Auction Volume

Description

Close Auction Volume prediction throughout the trading period is essential for various trading strategies to make critical decisions.

Stock trades on exchange in two phases:

1. **Continuous phase** – stock trades in the phase from the start of day till Close Auction takes place. Trades continuously happen in this phase as buyers and sellers agrees on a price. Volume of shares traded in this phase is called Order Book volume.
2. **Close Auction phase** – stock trades in this phase at the end of the trading. In this phase large volume is traded at a single price through an auction, we call this volume as close auction volume.



Objective

You have to develop a model which can predict close auction volume based on historic data provided.

Note: You should use training data to come up with a model that predict the Close volume based on historic data.

Use the model that you have developed (based on training data), and with the input of test data predict the Close volume for respective symbol_ids.

Training Dataset

You can download the training dataset from [here](#)

In the dataset you are given cumulative Order Book volume for various time buckets in a day and close auction volume for that day.

Training dataset is essentially timeseries data for various symbol_ids.

Data Schema

symbol_id: Unique stock identifier.

date_id: Unique date identifier. date_id is sequential.

seconds_bucket: Numbers of seconds elapsed since the start of continuous phase.

cummulative_continuous_volume: total volume traded in continuous phase since the start of the continuous phase till the seconds_bucket.

close_volume: volume traded in Close Auction of the stock on that date. This will tick at the end of the day at seconds_bucket = 31200

Evaluation Criteria

Each prediction (x_i) will be assessed against the realised Close Auction Volume (y_i) for that day using

$$Score = 1 - \left(\frac{1}{n} \sum_{i=1}^n \frac{|x_i - y_i|}{y_i} \right)$$

n -> number of symbol_ids

Score for each test case will be added to produce a final score.

Final score on leaderboard and documentation in the code will be used in evaluation.

Note: This problem statement has additional test cases that'll run post your submission. Your final score is subject to change, your current score is not the final score

Custom Models

You can upload your models to Github/Gitlab **public repository** and download the same from your code.

Here is a sample code you can use.

```
import requests
import os

def download_file(url, filename):
    response = requests.get(url)
    if response.status_code == 200:
        with open(filename, 'wb') as file:
            file.write(response.content)

# Example usage
url = "https://github.com/sp2hari/hackerrank-network-sample/raw/refs/heads/main/userdata1.parquet"
filename = os.path.basename(url)
download_file(url, filename)
```

Please make sure you use this to download only model files. Trying to misuse the network access will result in disqualification.

Network access is enabled only for this question.

Input Format

Input will be a filename. This file will be a CSV which contains a timeseries data for various sybmol_ids. Schema is same as training data.

Timeseries starts with date_id = 1.

CSV will contain full data for first 50 days, and for 51st day you will have data till a certain seconds_bucket.

You will need to predict the close volume for 51st day.

How to read input

```
#!/bin/python3

import pandas as pd

filename = input()
df = pd.read_csv(filename)
```

Constraints

Predicted Close Volume should be a non-negative number.

Output Format

Output needs to be printed in the format shown in below samples.

Sample Input 0

```
sample1.csv
```

Sample Output 0

```
symbol_id,close_volume
1029677210,1473541.0
1077784782,471017.0
1084320043,3454475.0
1106122882,212836.0
```

Explanation 0

You can download the [Sample1.csv](#)

Sample1.csv has historic data for 4 symbol_ids. For 51st date_id, data is present till seconds_bucket 30600.

Output has predicted Close Volume for those symbol_ids.

Sample Input 1

```
sample2.csv
```

Sample Output 1

```
symbol_id,close_volume
1029677210,3088265.0
1077784782,187838.0
```

```
1084320043,2990272.0  
1106122882,78281.0
```

Explanation 1

You can download the [Sample2.csv](#)

Sample2.csv has historic data for 4 symbol_ids. For 51st date_id, data is present till seconds_bucket 14400.

Output has predicted Close Volume for those symbol_ids.

Sample Input 2

```
sample3.csv
```

Sample Output 2

```
symbol_id,close_volume  
1029677210,2038657.0  
1077784782,210855.0  
1084320043,4136057.0  
1106122882,106879.0
```

Explanation 2

You can download the [Sample3.csv](#)

Sample3.csv has historic data for 4 symbol_ids. For 51st date_id, data is present till seconds_bucket 28500.

Output has predicted Close Volume for those symbol_ids.