Order Brushing Detection

Background

Abnormal user behaviours on e-commerce platforms can be detected in various ways. For example, if an item in a shop is a best seller but many of its orders come from the same buyer, we have reason to suspect that the seller is conducting order brushing. Order Brushing is a technique that may be employed by sellers to create fake orders in order to inflate the seller's or a particular item's rating, which may likely push that seller's items up the search results on Shopee. To maintain the integrity of our online shopping experience, we consistently monitor and flag any possible cases of order brushing on Shopee.

Task

1. Identify all shops that are deemed to have conducted order brushing.
2. For each shop that is identified to have conducted order brushing, identify the buyers suspected to have conducted order brushing for that shop.

For the purpose of this question, shops are deemed to have conducted order brushing if their concentrate rate is greater than or equal to 3 at any instance

```
Concentrate rate =
Number of Orders within 1 hour / Number of Unique Buyers within 1 hour
```

For the purpose of this question, suspicious buyers are deemed as the buyer that contributed the highest proportion of orders to a shop that is deemed to have conducted order brushing.

For calculation of the highest proportion of orders to a shop, only include the orders that occurred in instances when order brushing has been deemed to have taken place.

In the case where multiple users share the same highest proportion of orders for a specific shop, all those users are deemed to be suspicious buyers.

Data Description

order_brush_order.csv: It contains orders information.
Columns: [orderid, shopid, userid, event_time]
Each orderid represents a distinct transaction on Shopee.
Each unique shopid is a distinct seller on Shopee.
Each unique userid is a distinct buyer on Shopee.
Event Time refers to the exact time that an order was placed on Shopee.

Examples

A. Determining if a shop has conducted order brushing

Case 1:

Shop 8996761 had 6 orders placed within 1 hour from 11:48:49. These 6 particular orders were placed by two users, user 215382704 and user 2136861.

Concentrate rate for shop 8996761

From event_time [2019-12-31 11:48:49] ~ [2019-12-31 12:48:49]:

= (Orders 3146..35, 3146..31, 3146..96, 3146..20, 3146..04, 3146..61) / (User 215382704 , 2136861)
= 6 / 2 = 3

	orderid	shopid	userid	event_time
5013	31420669820669	8996761	15182097	2019-12-30 23:57:50
5014	31420848484212	8996761	2136861	2019-12-31 00:00:48
5015	31463329902935	8996761	215382704	2019-12-31 11:48:49
5016	31463516755431	8996761	215382704	2019-12-31 11:51:56
5017	31463618079296	8996761	215382704	2019-12-31 11:53:38
5018	31463701425020	8996761	215382704	2019-12-31 11:55:01
5019	31463906062704	8996761	2136861	2019-12-31 11:58:26
5020	31463960795761	8996761	2136861	2019-12-31 11:59:20

Shop 8996761 is deemed to have conducted order brushing when we consider the first event time as [2019-12-31 11:48:49].

Case 2a:

Shop 27121667 had 4 orders placed within 1 hour from [2019-12-31 07:51:59]. These 4 particular orders were placed by two users, user 50769054 and user 183926374.

Concentrate rate for shop 27121667

From event_time [2019-12-31 07:51:59] ~ [2019-12-31 08:51:59]:

= (Orders 3144..84, 3144..27, 3144..88, 3144..92,) / (User 50769054, 183926374)
= 4 / 2= 2

	orderid	shopid	userid	event_time
14032	31449119091684	27121667	50769054	2019-12-31 07:51:59
14033	31449214862927	27121667	183926374	2019-12-31 07:53:35
14034	31449320907988	27121667	183926374	2019-12-31 07:55:20
14035	31449708525992	27121667	183926374	2019-12-31 08:01:48
14036	31460681694470	27121667	38982221	2019-12-31 11:04:41

Shop 27121667 did not conduct order brushing, when we consider the first event time as [2019-12-31 07:51:59].

Case 2b:

Shop 27121667 had 3 orders placed within 1 hour from [2019-12-31 07:53:35]. These 3 particular orders were placed by one user, user 183926374.

Concentrate rate for shop 27121667

From event_time [2019-12-31 07:53:35] ~ [2019-12-31 08:53:35]:

= (Orders 3144..27, 3144..88, 3144..92) / (User 183926374)
= 3 / 1 = 3

	orderid	shopid	userid	event_time
14032	31449119091684	27121667	50769054	2019-12-31 07:51:59
14033	31449214862927	27121667	183926374	2019-12-31 07:53:35
14034	31449320907988	27121667	183926374	2019-12-31 07:55:20
14035	31449708525992	27121667	183926374	2019-12-31 08:01:48
14036	31460681694470	27121667	38982221	2019-12-31 11:04:41

Shop 27121667 is deemed to have conducted order brushing, when we consider the first event time as [2019-12-31 07:53:35].
Note: case 2a and 2b are the same shop with different time intervals. 2a has no order brushing but 2b does. Please consider all possible "1 hour" time interval for each shop, and the shop is deemed to have order brushing if it has at least one time interval meet concentrate rate >= 3 condition.

B. Determining suspicious buyers for shops that are deemed to have conducted order brushing

Case 3:

Shop 145777302 had 9 orders placed within 1 hour from [2019-12-28 12:07:49]. These 9 particular orders were placed by three users, user 201343856, user 107406 and user 101582282.

Calculating concentrate rate for shop

Concentrate rate for shop 145777302
= 9 orders / 3 users
= 3

Shop 145777302 is deemed to have conducted order brushing, when we consider the concentrate rate between event time [2019-12-28 12:07:49] to [2019-12-28 13:07:49].

orderid	shopid	userid	event_time
31205269078694	145777302	201343856	2019-12-28 12:07:49
31205397240383	145777302	201343856	2019-12-28 12:09:57
31205469197102	145777302	201343856	2019-12-28 12:11:09
31205961085182	145777302	201343856	2019-12-28 12:19:21
31206039633750	145777302	201343856	2019-12-28 12:20:40
31206219245988	145777302	201343856	2019-12-28 12:23:40
31207117377024	145777302	107406	2019-12-28 12:38:37
31207546860021	145777302	101582282	2019-12-28 12:45:47
31207600191875	145777302	201343856	2019-12-28 12:46:40
31209716869145	145777302	24336198	2019-12-28 13:21:57
31211630868210	145777302	90067860	2019-12-28 13:53:51
31219427310904	145777302	14448203	2019-12-28 16:03:47
31220878799186	145777302	199891702	2019-12-28 16:27:58
31223463550457	145777302	201343856	2019-12-28 17:11:03

Calculating proportion of orders for each buyer during order brushing period(s)
Order Proportion for User 201343856 = 7 / 9
Order Proportion for User 107406 = 1 / 9
Order Proportion for User 101582282 = 1 / 9

Assume Shop 145777302 only had one instance of order brushing in the entire dataset.

User 201343856* had the highest proportion of orders during the order brushing period, thus
User 201343856 is deemed to be a suspicious buyer for Shop 145777302.
*Note: orderid 31223463550457 is also from user 201343856, but it was not made during an order brushing period, so we excluded orderid 31223463550457 while calculating the order proportion.

Case 4:

Shop 181009364 had 3 orders placed within 1 hour from [2019-12-27 00:20:33]. These 3 particular orders were placed by one user, user 214208720. Shop 181009364 also had another 3 orders placed within 1 hour from [2019-12-29 13:54:31]

Concentrate rate for shop 181009364

Concentrate rate (2019-12-27 00:20:33 ~ 2019-12-27 01:20:33)
= 3 orders / 1 user
= 3

Concentrate rate (2019-12-29 13:54:31 ~ 2019-12-29 14:54:31)
= 3 orders / 1 user
= 3

Shop 181009364 is deemed to have conducted order brushing, when we consider the concentrate rate between event times [2019-12-27 00:20:33] to [2019-12-27 01:20:33] and [2019-12-29 13:54:31] to [2019-12-29 14:54:31].

orderid	shopid	userid	event_time	
31076433523765	181009364	214208720	2019-12-27 00:20:33	
31077003494175	181009364	214208720	2019-12-27 00:30:03	Order Brushing Period 1
31078405144719	181009364	214208720	2019-12-27 00:53:26	
31167197128282	181009364	27847950	2019-12-28 01:33:18	
31200690834058	181009364	214208720	2019-12-28 10:51:30	No Order Brushing
31243796889716	181009364	160784036	2019-12-28 22:49:57	
31248330354574	181009364	18028327	2019-12-29 00:05:30	
31298070356347	181009364	101832161	2019-12-29 13:54:31	
31298092385907	181009364	101832161	2019-12-29 13:54:52	Order Brushing Period 2
31298150262970	181009364	101832161	2019-12-29 13:55:51	

Calculating proportion of orders for each buyer during order brushing period(s)
Order Proportion for User 214208720= 3 / 6^
Order Proportion for User 101832161= 3 / 6

Assume Shop 181009364 only has two instances of order brushing in the entire dataset.

Both User 214208720 and User 101832161 had the same highest proportion of orders during the order brushing period(s), thus both User 214208720 and User 101832161 are deemed to be suspicious buyers for Shop 181009364.
^Note: A total of 6 orders were made to Shop 181009364 during 2 different order brushing periods. Even though orderid 31200690834058 was also made by User 214208720, it was excluded from the calculation for the highest proportion orders because it did not occur within an order brushing period.

Approach

The basic idea is calculating Concentrate rate for all possible 1 hour windows. Here illustrate the strategy:

a. Select a 1 hour window and take a slice from order data.

Note when you set the time condition, both ends are included in the 1 hour window. (Hint: For Python, datetime module and timedelta function can be applied)

b. Find shops which satisfy CR>=3.

For the slice of order data you select, you can use an aggregation function like groupby to calculate order number and unique user number for each shop in this 1 hour window. And then filter shops which satisfy the CR>=3 condition. They are brushing shops.

c. Find the userid who conducts order brushing in brushing shops.

When brushing shops are identified, aggregation function can be applied again to calculate which userid contribute the most orders for each shop. Note if two userid both have the highest order number, then they are both the answer.

d. Repeat the above procedure for all possible 1 hour window.

For example:

2019-12-27 00:00:00 ~ 2019-12-27 01:00:00,

2019-12-27 00:00:01 ~ 2019-12-27 01:00:01, ...

Note:

When you follow the above steps, you may find d step takes a long time since the time period in our data crosses 5 days and the time window is detailed to second. To reduce the time consumption, some advanced strategy might be applied. For example:

a. Multiprocessing to brutally calculate them all.

b. Avoid redundant calculation by detecting windows which cover exactly the same orders.

Submission Format

Check each shop and determine whether it is deemed to have conducted order brushing. If a shop conducted order brushing, list the userid(s) that are identified as suspicious for the corresponding shopid.

Two columns required:

- shopid
- userid

- If a shop is not deemed to have conducted order brushing, assign the value 0
- Else, list the userid(s) that are identified as suspicious for the corresponding shopid
- If there is more than 1 userid identified as suspicious, list all the userids separated by “&”, with the smaller numerical userid first.

shopid	userid
162014252	183926374
321014322	19233237&23421231
22754767	0

Your submission should have 18770 rows (excluding the headers), each with 2 columns.