Promotion Abuse Report

Contents

By : Nguyen Hy Khang

5/6/2023

1. Overview

We now have a sale promotion 30% per order with maximum price as 20K. We know that shops collaborate with their customers to take advantage of this promotion.

To make it clear, we have an order. We can divide it into many smaller orders to exploit this promotion thoroughly. In this case, 20K per order.

Let’s get through an example. Of course, with an order of 600K, we will pay 20K for this order. But if the shop and customer is guileful, they will divide that order into 9 smaller orders (8 orders are worth 67K and 1 order is worth 64K). Instead of paying 20K for an order of 600K, we must pay 179K. In essence, the revenue is stationary, but the cost will rise. That’s why promotion can be counterproductive.

The purpose of this report has answered 3 questions:

1) Which Gmv range has the most transactions ?
2) Which Shop Owner UIDS are most likely fraud ?
3) Which User IDs are most likely fraud ?

I will answer those questions in Detail Analysis part.

2. Data Understanding

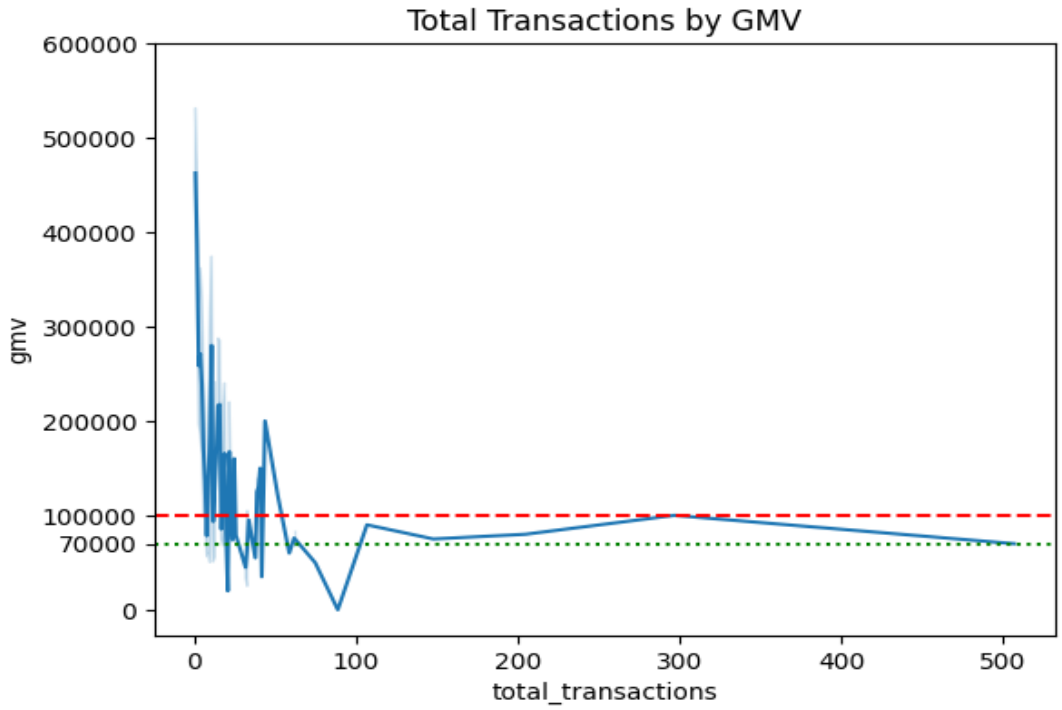
The dataset is the information about the transaction record throughout the sale promotion.

Column Name	Definition	Note
txn_time	The time that the transaction occurs.	
txn_date	A day that the transaction occurs.	Record in 4 days of 2019 (30/10, 31/10, 1/11, 2/11).
order_id	The id of an order.	
uid	User Id who made the transaction.	
shop_id	Merchant Who Receive The Payment.	A shop owner can own many shops.
shop_owner_uid	Id is the Owner of Shop.	
gmv	Transaction amount.	
rebate	Cashback Amount.	Some orders have 0 rebate because they don't apply promotion. So I will remove them to facilitate analysis.

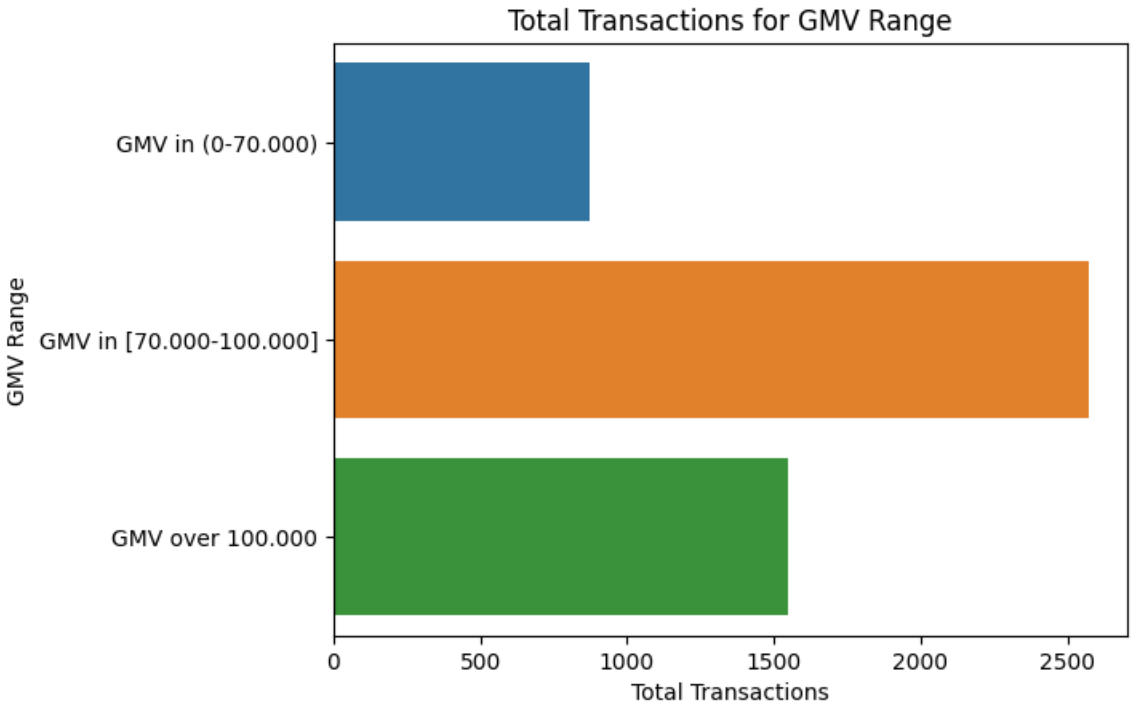
3. Detail Analysis

1. Gmv range has the most transaction

First and foremost, we need to identify the range of Gmv that shop will be most likely fraud. Like I've said before, the suspects will try to divide an order into many smaller orders. Hence, my approach is to check the range that most orders concentrate in.



Through above chart, I figure out that Gmv range between 70.000 – 100.000 has the most transactions. To clarify it, let's see the following chart:



Indeed, GMV in [70.000 – 100.000] accounts for more than 50 % the total transactions. Combining with 2 charts, we can conclude that the Gmv **70.000** has the most transactions.

2. Shop Owner UIDs who are most likely fraud

To find Shop Owner UIDS who are most likely fraud, I have two approaches:

1) Find the target who are possible to fraud in a day

	shop_owner_uid	uid	txn_date	total_transactions	total_rebate
1913	102368516	1026737	10/30/2019	2	20000
1246	100577736	1499917	10/31/2019	2	18000
2558	103117675	100368648	10/30/2019	2	18000
2761	103284688	103868446	10/31/2019	1	20000
2752	103282856	2057216	11/1/2019	1	20000
...	...	...	...	...	...
1404	100899318	103904413	11/2/2019	1	20000
1405	100906499	1418031	10/30/2019	1	20000
1406	100906499	1418031	10/31/2019	1	20000
1407	100906499	1418031	11/1/2019	1	20000
2090	102817559	102661412	11/2/2019	1	20000

4123 rows × 5 columns

It seems like we can’t detect suspects because there is no abnormal thing in this result.

2) Find the target who are possible to fraud in 4 days

We can detect suspects through some criteria: **A number of transactions** (The more transactions they can make, the more suspicious they become), **Gmv per order** (The number is nearer 70.000, the more suspicious they become), **Rebate per order** (The number is nearer 2.000, the more suspicious they become).

	shop_owner_uid	uid	total_transactions	total_gmv	total_rebate	gmv_per_order	rebate_per_order
2331	103280547	1225554	4	280000	80000	70000.0	20000.0
2049	103023297	104029244	4	280000	80000	70000.0	20000.0
1249	100868853	103286330	4	280000	80000	70000.0	20000.0
620	1873720	100451463	4	280000	80000	70000.0	20000.0
2566	103435279	100702363	4	280000	80000	70000.0	20000.0
...	...	...	...	...	...	...	...
1383	101211912	100324327	2	194000	32000	97000.0	16000.0
3546	104030290	1835362	2	157600	31760	78800.0	15880.0
3244	103835619	100531415	2	154000	31100	77000.0	15550.0
1084	100526353	100173458	2	145000	28700	72500.0	14350.0
2483	103366603	101345963	2	168000	27500	84000.0	13750.0

260 rows × 7 columns

There are many people have the same conditions to fraud. It's very tough to identify who are most likely to fraud. But, I have another way to solve it. Let's check how many unique users interact with each shop owner in that table.

	shop_owner_uid	Total users	total_transactions	gmv_per_order	rebate_per_order
62	102817559	53	130	74933.99	20000.0
85	103333474	8	16	77750.00	20000.0
95	103481792	7	14	80142.86	20000.0
96	103483282	7	14	81214.29	20000.0
83	103280547	5	12	72000.00	20000.0

To conclude, top 3 shop owner uids who are most likely fraud:

- **102817559** that interacts with 53 users.
- **103333474** that interacts with 8 users.
- **103483282** and **103481792** interact with 7 users. But I prefer **103481792** because its gmv_per_order is nearer 70.000 than **103483282**'s gmv_per_order.

3. User IDs who are most likely fraud

We just need to count the total Shop Owner UIDs for each suspicious customer. Here is a list:

	uid	Total Shop Owner	gmv_per_order	rebate_per_order
143	103178855	3	70000.00	20000.0
130	103002018	3	70000.00	20000.0
59	100702363	3	70000.00	20000.0
102	101793553	3	70000.00	20000.0
146	103266260	3	75000.00	20000.0
207	104041098	3	75000.00	20000.0
184	103855768	3	75833.33	20000.0
153	103333474	3	76666.67	20000.0
169	103607002	3	79166.67	20000.0

In brief, top 4 User IDs who are most likely fraud:

- **103178855**, **103002018**, **100702363**, **101793553** have the same gmv_per_over (**70.000**) and interact with 3 Shop Owner.

4. Summary

To summarize, we have found out some conclusions:

- Most orders concentrate in Gmv range [70.000 – 100.000] with most transactions is Gmv 70.000.
- There is no way to find suspects because there is no unusual transaction in a day (There is no more than 3 total transactions between Shop Owner and User).
- If we check in 4 days, there are 3 Shop Owner UIDs who are most likely fraud: **102817559**, **103333474**, **103481792**. They have interacted with corresponding number users: **53**, **8**, **7**.
- Top 4 User IDs who are most likely fraud: **103178855**, **103002018**, **100702363**, **101793553**. They have interacted with **3** Shop Owner.