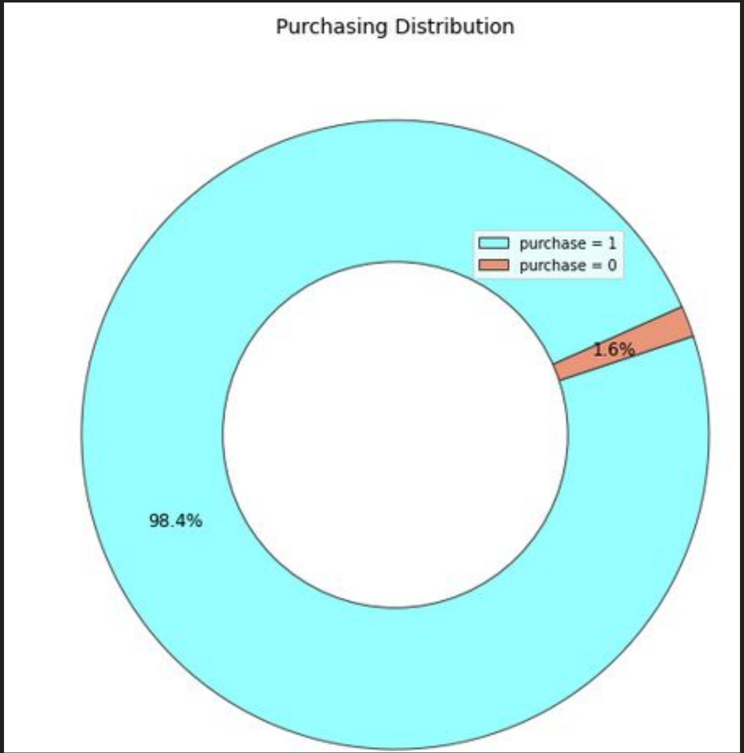
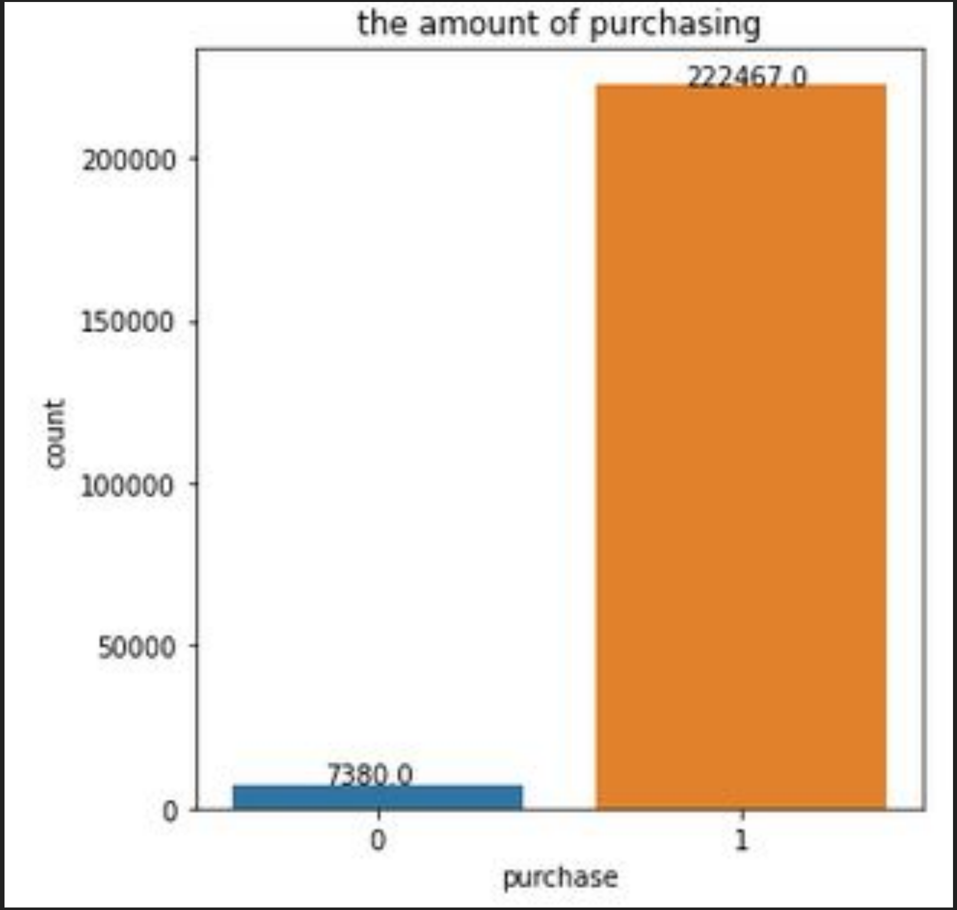
Purchasing Insights

By : Aminah Nurrahmawati



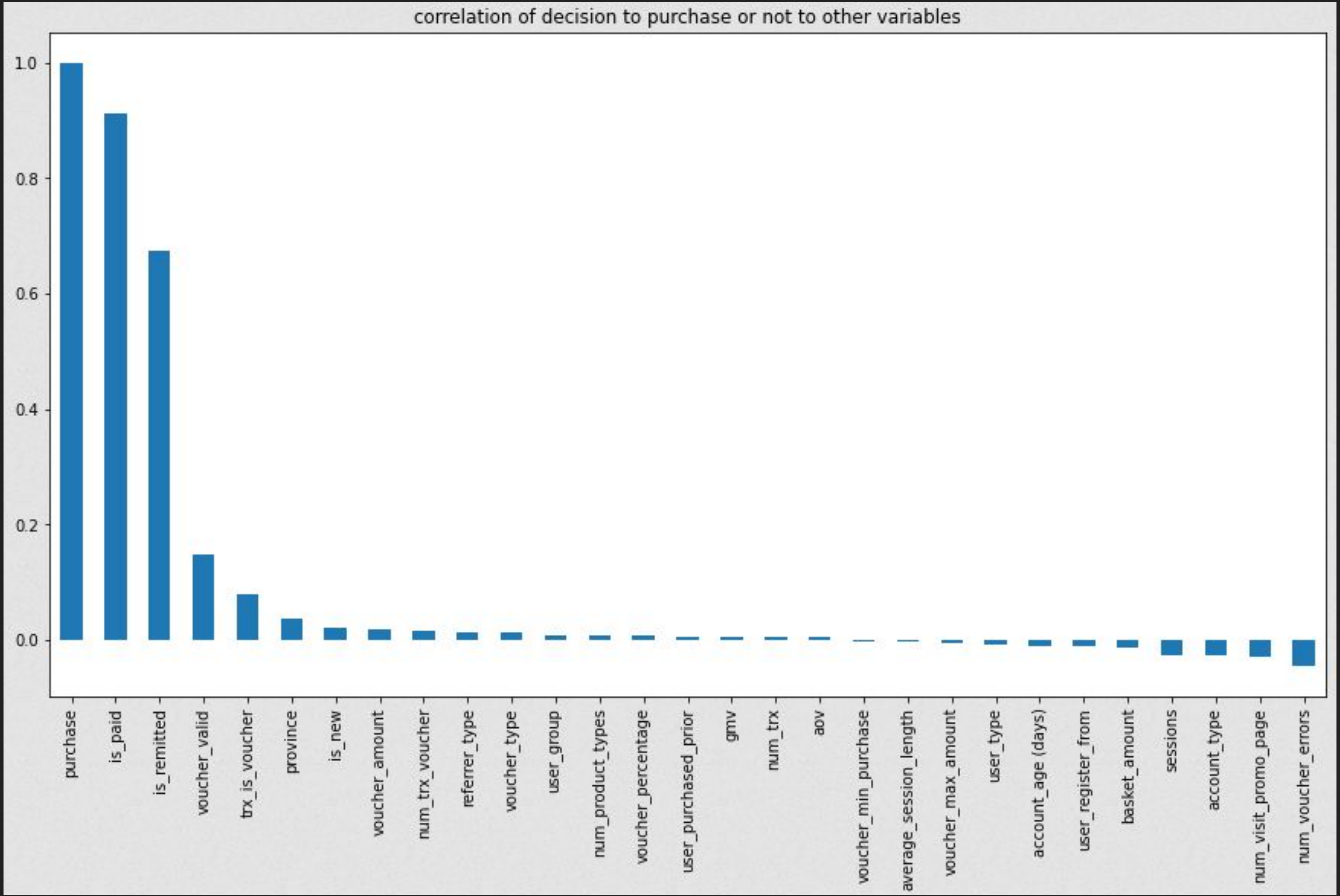
Purchase Distribution



Here's the purchasing distribution percentage. There are 98.4% purchasing's done by the customers, and the rest 1.6%. The total purchasing is 222467 transactions. And the transaction total that is not processed further is 7380.

Factors Influencing the Purchasing

- In general, of course shop owner/seller want their customers who visit their store to make a transaction or purchase their product, after checking out the catalogue. But sometimes there are some customers whom only check out without purchasing.
- There are some variables that have tendency to effect a customer to purchase products on e-commerce platform. So it is really important to carefully analyze what variables that might have strong correlation with decision of a customer purchasing products.
- the correlation value between decision to purchase or not will be shown in the next slide.


correlation of decision to purchase or not to other variables
1.0
0.8
0.6
0.4
0.2
0.0
purchase
is_paid
is_remitted
voucher_valid
trx_is_voucher
province
is_new
voucher_amount
num_trx_voucher
referrer_type
voucher_type
user_group
num_product_types
voucher_percentage
user_purchased_prior
gmv
num_trx
aov
voucher_min_purchase
average_session_length
voucher_max_amount
user_type
account_age (days)
user_register_from
basket_amount
sessions
account_type
num_visit_promo_page
num_voucher_errors

Findings

From the correlation plot shown in the previous slide, I will sort top 5 variables that have highest correlation with decision of a customer to purchase products : (to be noted it does not mean that I ignore the other variable excluding these 5, but for this time let's focus on these 5 first.

1. **is_paid** variable : this is flag to indicate whether the user paid transaction
2. **is_remitted** variable : this is flag whether the transaction is completed
3. **voucher_valid** variable : identifier whether voucher applied is valid
4. **trx_is_voucher** variable : indicate whether the user using voucher in buyer
5. **province** : identifier to user province

Let’s start with the first variable that we assume (based on correlation coefficient) to have contribution to a user to purchase products.

the picture in the right is the plot picturing how many user paid the transaction and how dominant its parameter.

as we can see, the most user who paid will purchasing, just view of them (7380 out of 229847) who didn’t pay the purchase, meanwhile only 1434 user who didn’t pay the transaction but still purchase. we can assume the users who did not pay but still purchase using the paylater method.

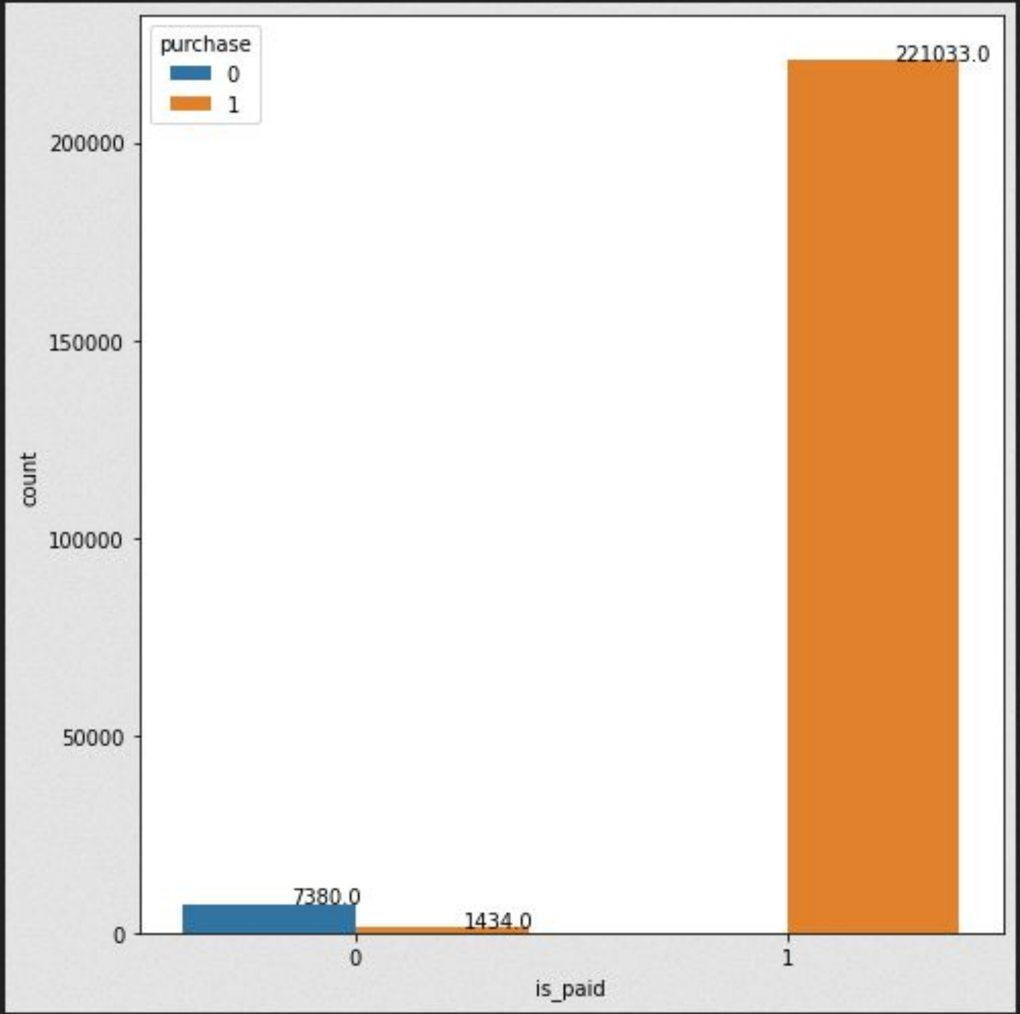
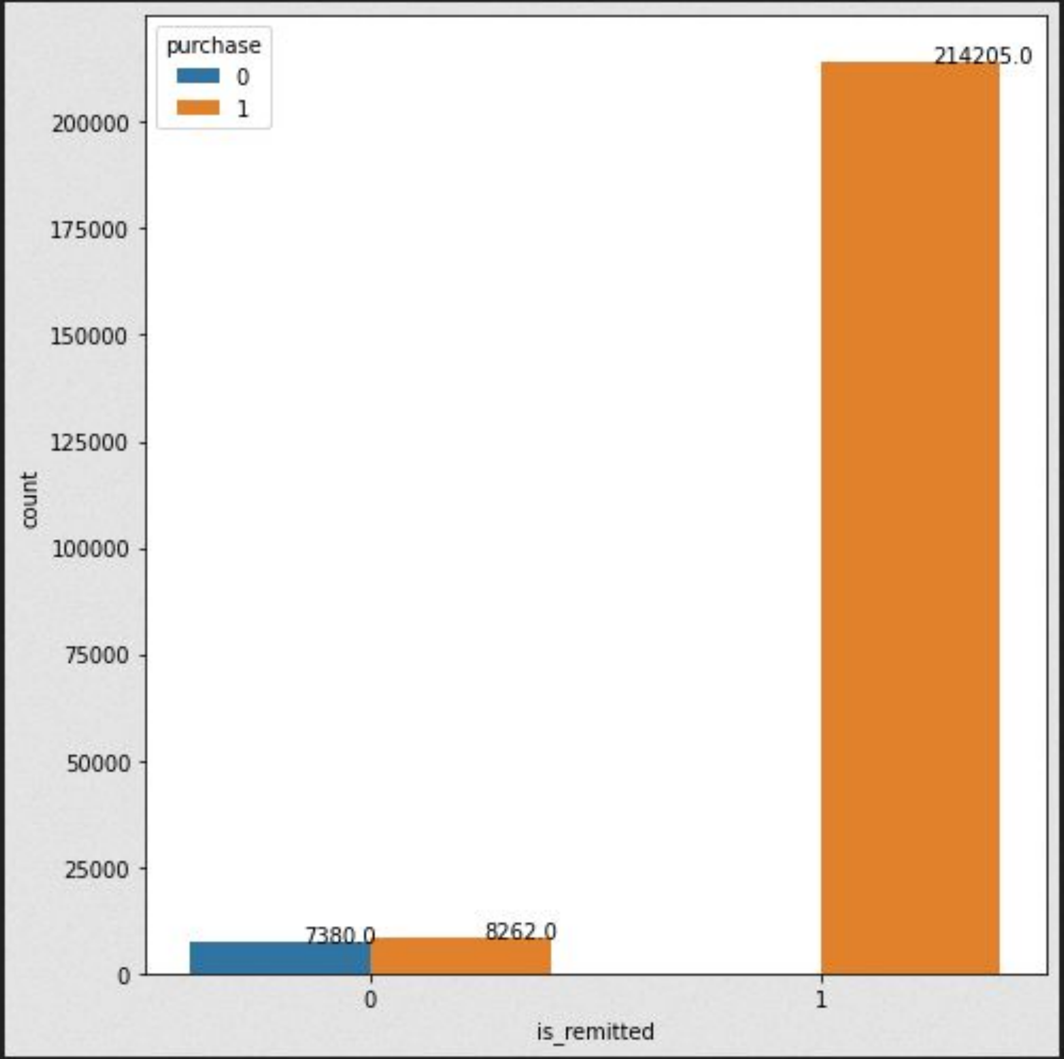


Figure explanation :

1. there are 214205 user id who purchased after completing the transaction
2. there are 7380 user is who did not purchase and completing transaction
3. there are 8262 user id who did not purchase but completing transaction

Most of users who definitely purchased have completed their transaction. and the rest view users who have not completed or completed their transaction didn’t purchase.

So, my recommendation : we can make kind of “reminder” for user who haven’t completed the transaction to completed it, so it can impulse them to purchase.

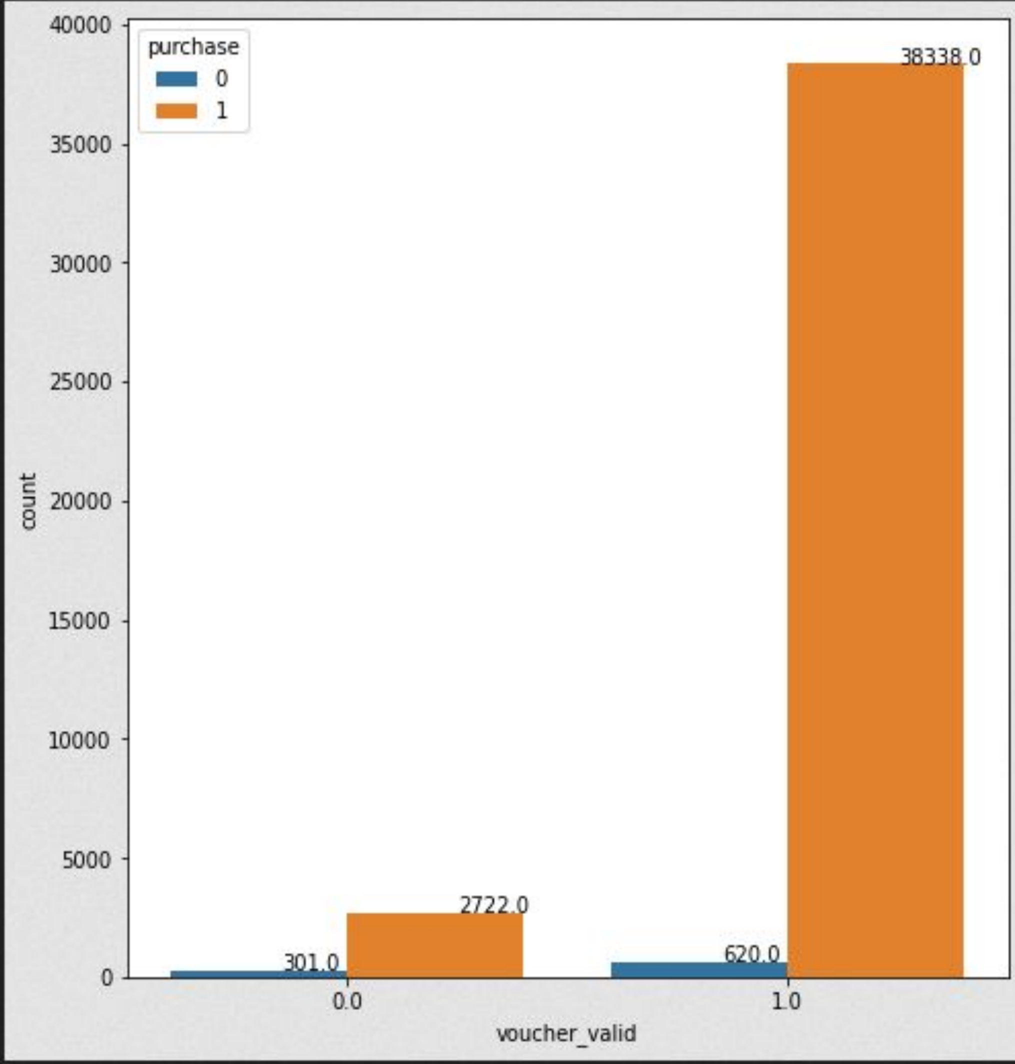


the next one is, we analyze through how the valid vouchers that users had contribute to users in making purchase.

the explanation of barplot :

1. there are 38338 user id who had valid voucher and made purchase and there are only 620 user id who had valid voucher and did not make purchase
2. there are 2722 and 301 user id respectively made purchase and did not make purchase but did not have valid voucher

So, it can bee seen that the valid voucher had strong contribution to users to make purchase. So let's assume that having valid voucher can impulse users to purchase

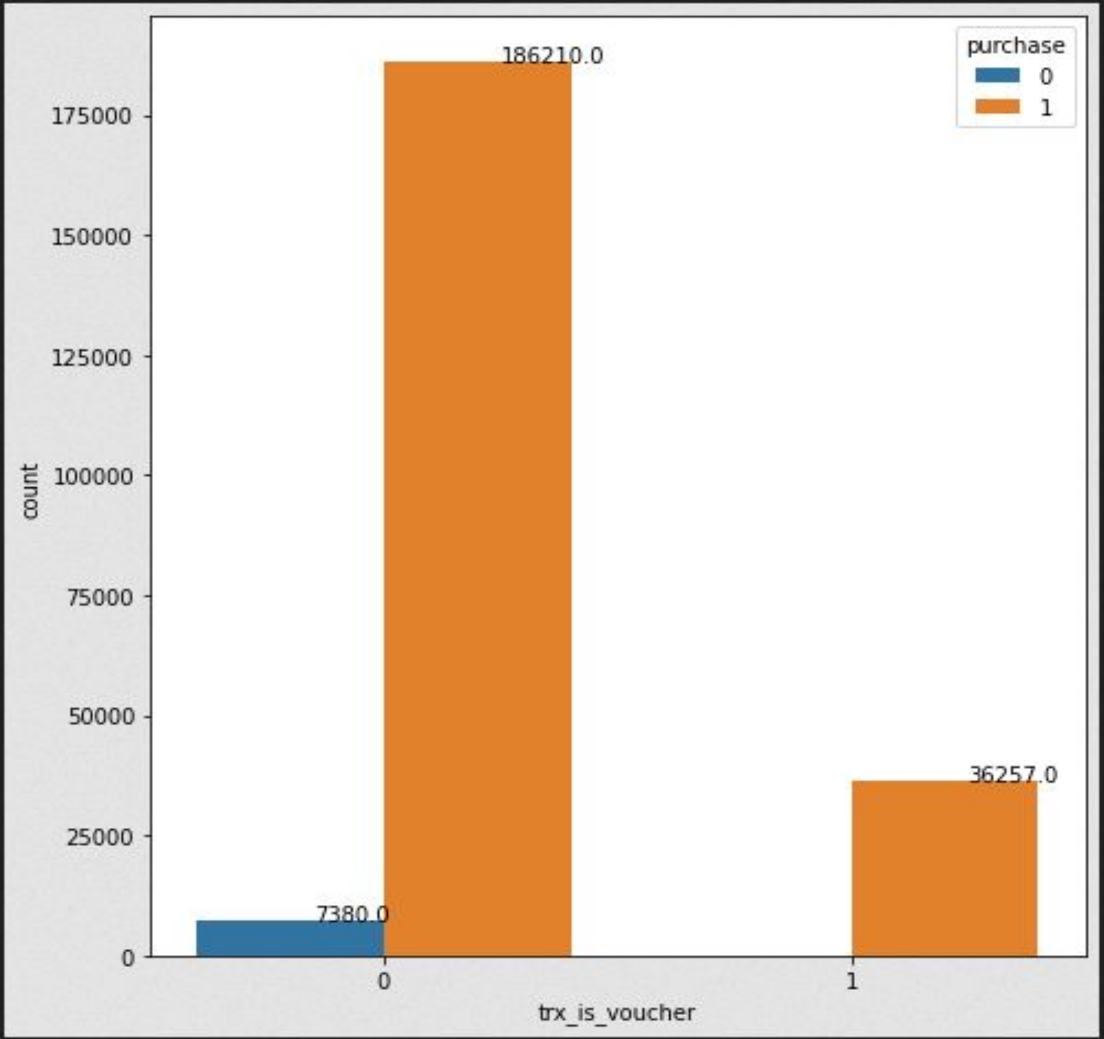


Now, the next parameter is how many users making purchase using voucher.

the explanation :

1. there are 36257 user id using voucher in transaction.
2. there are 186210 user id did not use voucher in their transaction. literally this is the most case out of everything.
3. there are only 7380 user id making purchase without using voucher.

So, from the findings we can say that the contribution of voucher when user make transaction is small. User id mostly don’t use voucher for transaction. We can compare this finding to previous finding about using voucher too.



More about Voucher

from the 2 previous slides, if we conclude the findings of the amount of user_id using voucher transaction and valid voucher :

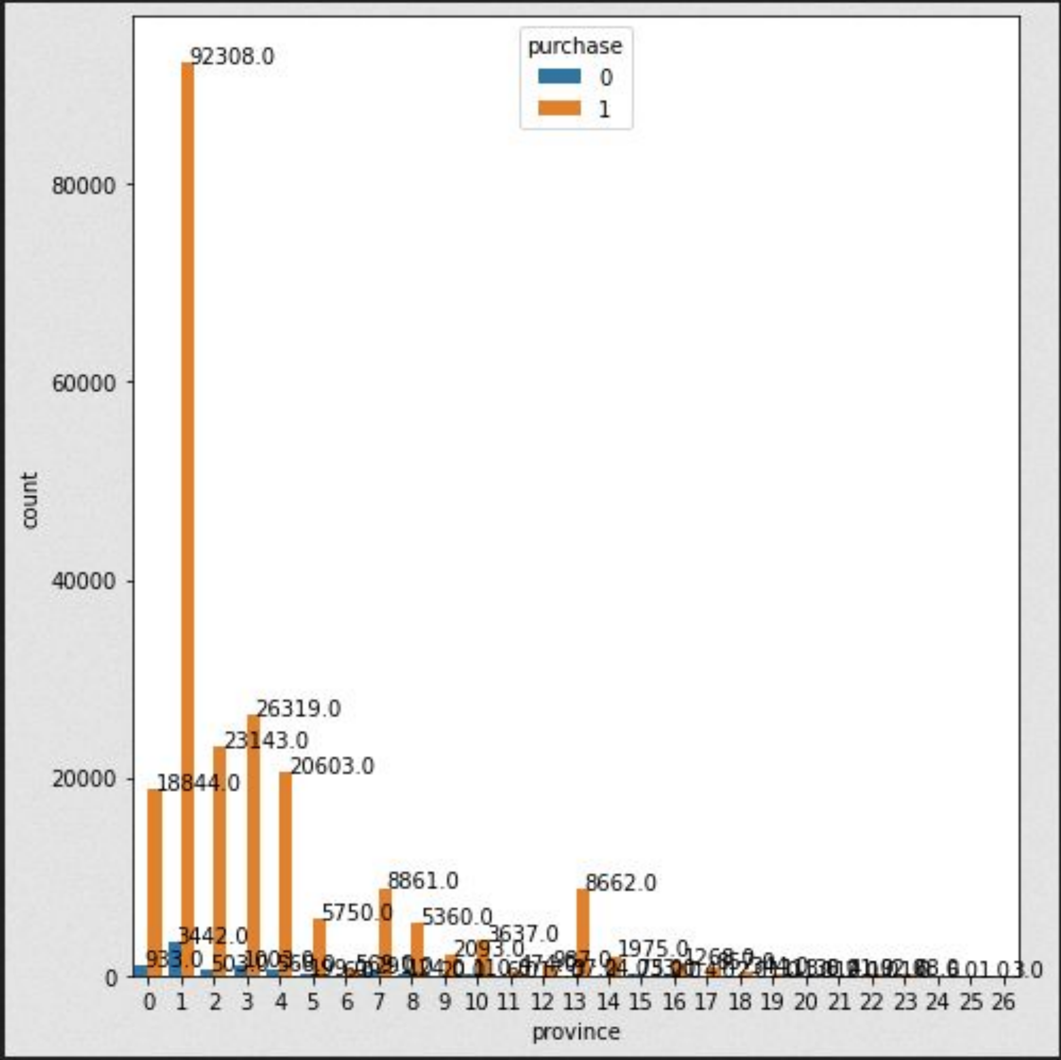
1. mostly they don’t use voucher on their transaction. This can be caused by several probability : maybe the voucher is not available anymore, even though they have voucher. It might be the voucher they have can not be used to the purchasing they made.

recommendation : to analyze more about the use of voucher, we can see from the rating of users who make the purchase with/without voucher. How satisfied they are with the available of voucher. This is can be a feedback to gain the loyal users and to optimize the income of e-commerce

The next parameter is the province that the users come from. the purchase is mostly made in province 1, 3, 2, 4, 0, 13, 7, 8 (those are the top 8 purchase).

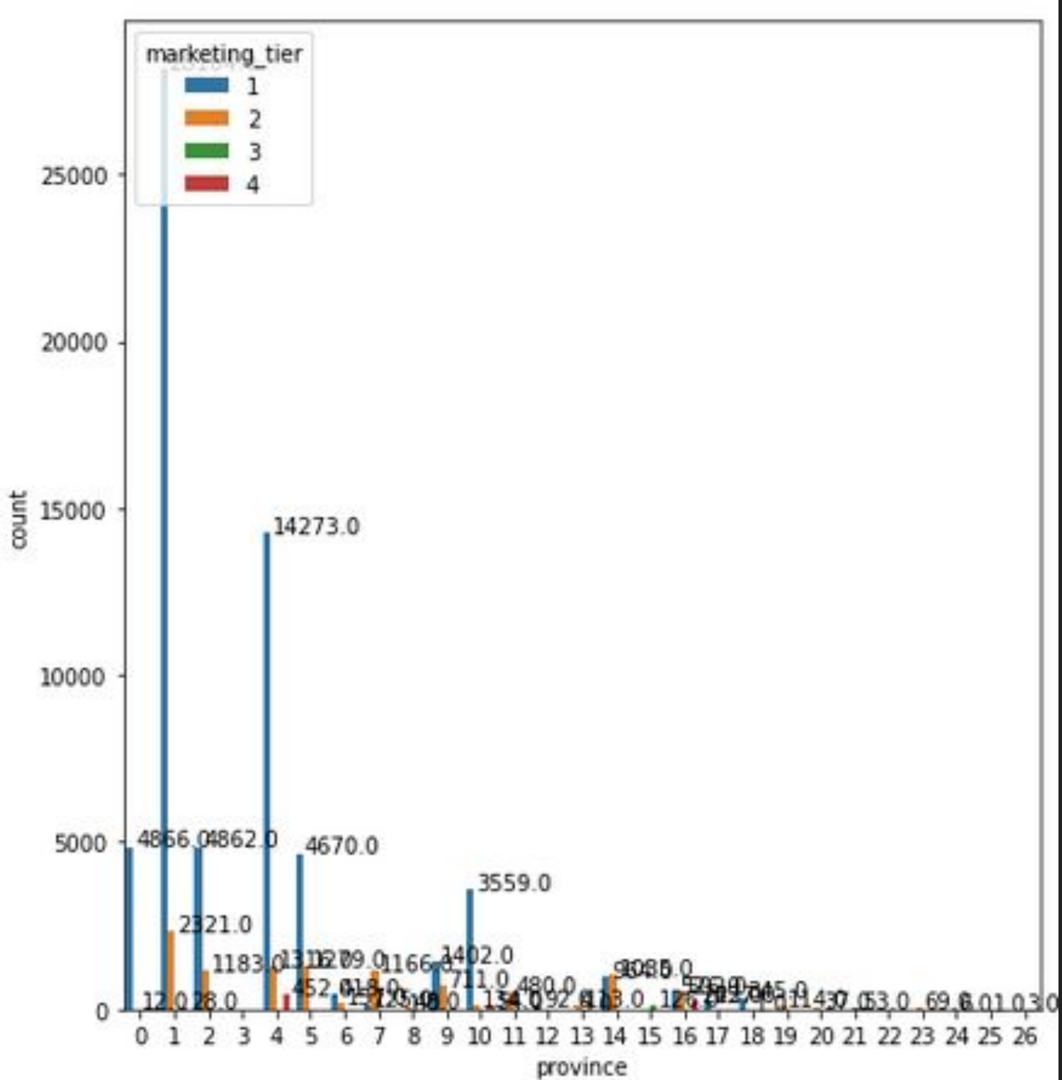
So, from this findings I make the first assumption that people from the rest province excluding those 8 did not have the optimum exposure of the e-commerce.

Recommendation : we can optimize the marketing strategy in the province who have minimum transaction. But



Still relate to the previous parameter of contribution to cause users to make purchase. The province have the contribution to make purchase. So, to make purchase optimum in one province we have to set up the right marketing strategy.

It can be seen from the figure in the right, mostly in some provinces only applied marketing tier 1 and 2 for a few. So if we check once again only province 1, 3, 2, 4, 0, 13, 7, 8 (those are the top 8 purchase) who make the maximum purchase. So, it can be consideration to apply the marketing tier 2 or 3 to the rest of province, to impulse more purchase in those province.

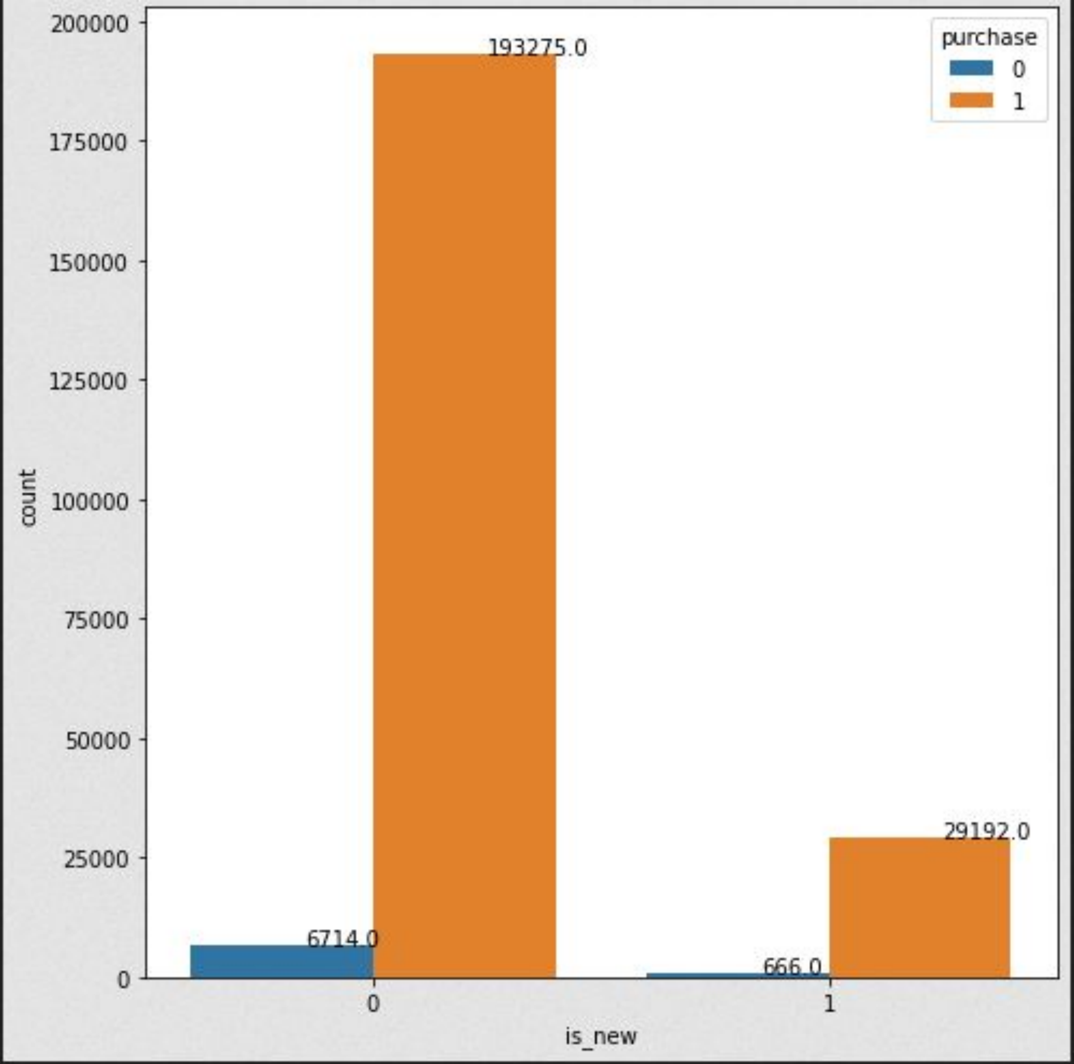


Now, the next parameter is how many new users making prio purchase.

the explanation :

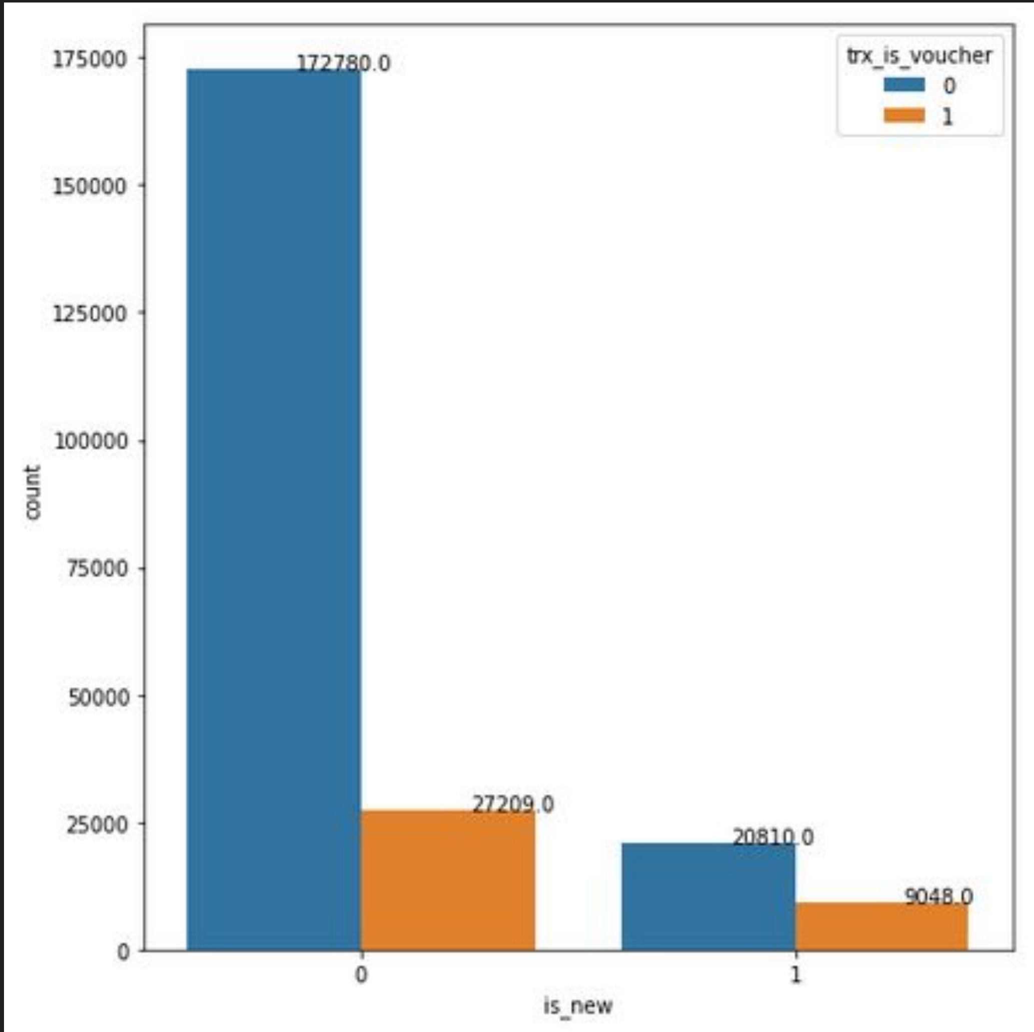
1. there are 193275 user id made purchase and have made purchase before and only 6714 who have prior purchase but don't make purchase anymore
2. there are 29292 user id who never make prior purchase but just start the purchase and only 666 user id who never make purchase at all

so, from these findings, the experience of users give the strong contribution to users to make purchase. So, we can optimize to gain new users by promoting voucher, make marketing tier, and so on.



More about users with prior purchase

this figure shows how many users without prior purchase who make purchase with voucher. We can see that the majority of user is user with prior purchase but don’t use voucher transaction. This can be good as mark of loyal users. So we can optimize marketing strategy to obtain some new users to make purchase.



Customer Segmentation

In the next section, I will segmentate the customer based on some parameters :

1. whether they use transaction voucher
2. whether they purchase and the number of their purchase
3. the basket amount

this aims to find the effectiveness of giving voucher to users based on their characteristics in purchase.

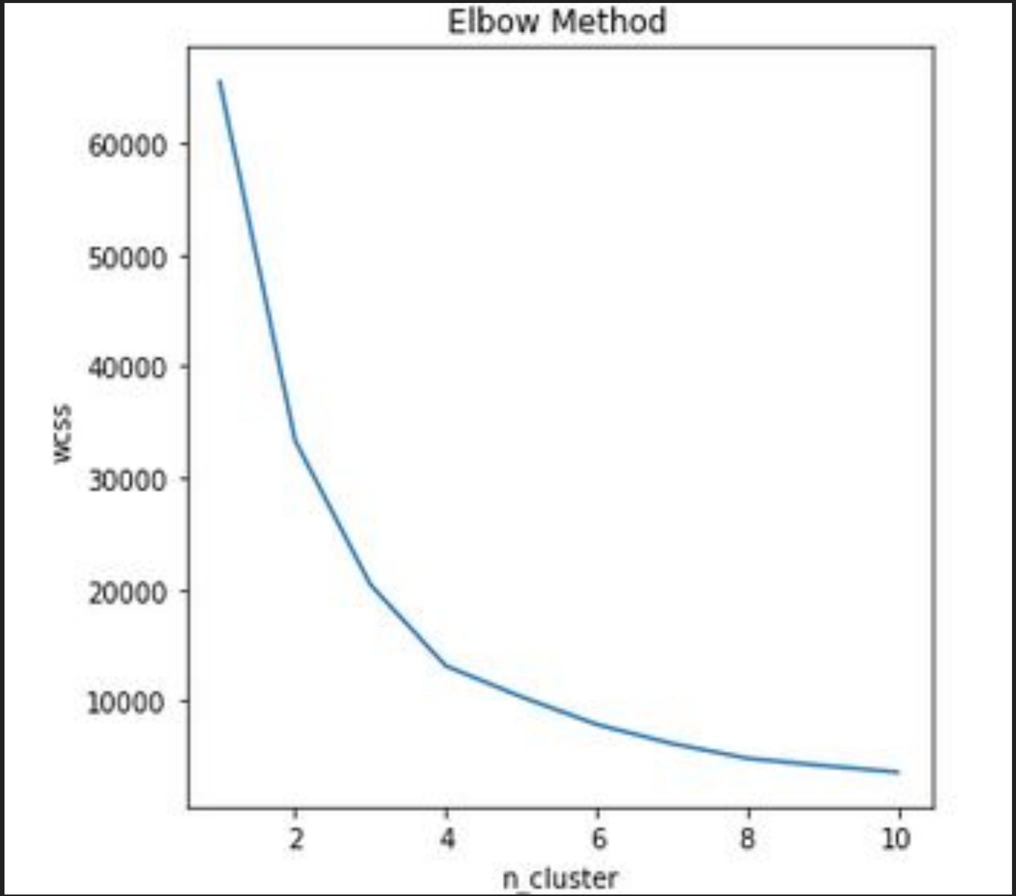
I will use k-means clustering to make the cluster every user based on the 3 parameter above they have

- The picture in the right is the snippet of 3 group parameters that we want to calculate to be clusters.
- I use Elbow Method to calculate how many cluster can be made from our data, before we define every customer's cluster

	purchase	Trx_is_voucher	basket_amount
user_id			
626423711	0	0	0.000009
156651417	0	0	0.000004
237757451	0	0	0.000025
131186871	0	0	0.000010
159001405	0	0	0.000053
...	...	...	...
372071043	32	32	0.005433
2359810	33	32	0.003504
607379603	34	34	0.002723
132287846	39	38	0.005115

We got 4 cluster :

- cluster 1 (0) = barely/don’t use voucher transaction, least transaction of every unique id
- cluster 2 (1) = rarely use voucher transaction, less transaction of every unique id
- cluster 3 (2) = often use voucher transaction, quite many transaction of every unique id
- cluster 4 (3) = most often use voucher transaction, most transaction of every unique id



The snippet of data composition of cluster

user_id	purchase	Trx_is_voucher	basket_amount	cluster
372071043	32	32	0.005433	3
2359810	33	32	0.003504	3
607379603	34	34	0.002723	3
132287846	39	38	0.005115	3
350553855	39	38	0.005302	3

user_id	purchase	Trx_is_voucher	basket_amount	cluster
4148	1	1	0.000032	0
607500078	1	1	0.000010	0
607496591	1	1	0.000023	0
607487747	1	1	0.000023	0
607485610	1	1	0.000023	0

Results

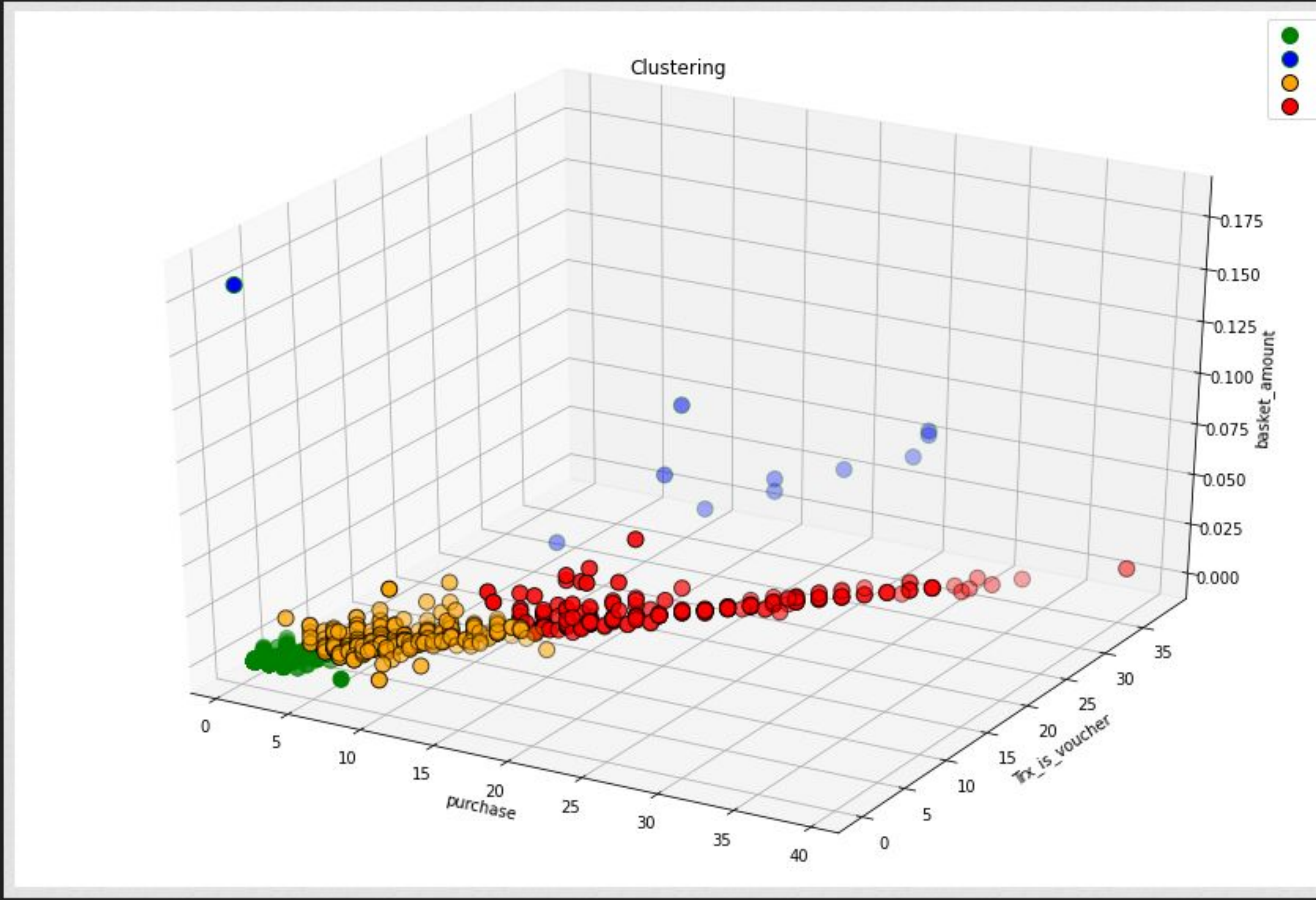
The amount of every cluster (total purchase (1) = 222467 :

0	19809
2	1395
3	316
1	11

The findings :

1. Most users is distributed in cluster 1, which barely use voucher transaction
2. meanwhile cluster 2 is the least distributed customer
3. some few users (0.14% of users who make purchase) still use voucher in purchasing. This group is distributed in cluster 4

Distribution of cluster



Recommendation

From the data that we obtain, There are several findings and recommendation:

- The majority of customers who purchase did not/barely use voucher. But the problem is, this happen to users who did not have prior purchase. So, it’s a little bit risky if don’t give voucher for the, because it can make them stop to make purchase from our platform. So, we need to set up the voucher for new users, to keep them to buy to our platform. We can start from the voucher with small amount based on their purchase amount, and we can add the voucher as the amount of basket increase.
- And the users who make the big purchase (frequently) and have biggest amount of basket amount and have prior purchase, usually make purchase with voucher. So, it can be risky if we remove the voucher from them. As we know the they are can be categorized as loyal customers. But maybe we can change the method of giving the voucher by giving the voucher just for special days or the days that are counted as the long they purchase using our platform

Recommendation (continued from previous slide)

- There are some provinces that have small amount of purchase. and seeing the fact that there are still marketing tier 1 and 2 applied evenly to all province, meanwhile the distribution of purchase is not even. So, we need to change/add the marketing tier in some province with marketing tier 3 (let’s say). marketing tier 3 is an “Air Cover” marketing. So we widely promote our platform and its benefit to as many as possible people. We can make promotion through billboard, radion, etc. Or we can add marketing tier 1 and 2 to be more dense in some provinces.

Additional Insights

Classify the Customer

Actually we can predict whether a user will purchase or not based on their data in e-commerce database.

In this case, I will use 26 variables from the dataset. I will use 3 different ensemble algorithms. I use Ensemble algorithm because of the imbalance dataset of purchase class (0 and 1).

I will compare the 3 algorithm based on their accuracy .

These are the dependent variables I used to predict the class

	variables	VIF
0	is_new	4.059021
1	voucher_type	3.099668
2	voucher_valid	0.000000
3	basket_amount	1.279076
4	voucher_max_amount	1.320449
5	voucher_percentage	1.538253
6	voucher_min_purchase	2.579763
7	voucher_amount	2.808558
8	trx_is_voucher	1.542843
9	is_paid	2.697667
10	is_remitted	2.295364
11	user_purchased_prior	1.623669
12	num_voucher_errors	1.086782
13	province	1.551672
14	user_type	1.056091
15	user_group	3.488716
16	account_type	1.025150
17	referrer_type	3.054999
18	user_register_from	1.557117
19	sessions	2.580160
20	average_session_length	1.086637
21	num_visit_promo_page	1.309533
22	num_product_types	3.112602
23	num_trx	2.055363
24	num_trx_voucher	1.817771
25	gmv	3.397484
26	aov	3.869446

Results

1. XG Boost Classification

   accuracy : 0.983

2. Random Forest Classification

   accuracy : 0.99

3. ADA Boost

   accuracy : 0.99

Conclusion

Why do I use classification to predict a customer will purchase or not, because it is aimed to make the optimum promotion method to attract them to use our e-commerce to make a transaction. The biggest accuracy obtained using Random Forest and ADA Boost algorithm. So, for this kind of dataset with the dependent and independent variable we have, it is highly recommended to use those 2 to predict the tendency of a user to purchase or not.

if users tend to not purchase, we recommend :

- to promote more
- to give them voucher