



Bank Marketing Segmentation

AHMED ALI BAIG

WHAT IS MARKET SEGMENTATION?

In marketing, market segmentation is the process of dividing a broad consumer or business market, normally consisting of existing and potential customers, into subgroups of consumers based on some type of shared characteristics.

OBJECTIVE

A case requires to develop a customer segmentation to give recommendations like saving plans, loans, wealth management, etc. on target customers groups

THE PROBLEM STATEMENT AND BUSINESS CASE

- In this project , I have search and collected data Kaggle which contains extensive data on the bank customers for the past 6 months.
- Data includes transactions frequency, amount, tenure etc.
- I like to leverage AI/ML to launch a targeted marketing ad campaign that is tailored to specific group of customers.
- In order for this campaign to be successful, the bank has to divide its customers into at least 3 distinctive groups.
- This process is known as "marketing segmentation" and it crucial for maximizing marketing campaign conversion rate.

PROJECT OVERVIEW

One of the key pain points for marketers is to `know their customers` and `identify their needs`.



DATA DESCRIPTION

1. CUSTID: Identification of Credit Card holder
2. BALANCE: Balance amount left in customer's account to make purchases
3. BALANCE_FREQUENCY: How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)
4. PURCHASES: Amount of purchases made from account
5. ONEOFFPURCHASES: Maximum purchase amount done in one-go
6. INSTALLMENTS_PURCHASES: Amount of purchase done in installment
7. CASH_ADVANCE: Cash in advance given by the user

DATA DESCRIPTION

8. PURCHASES_FREQUENCY: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)

9. ONEOFF_PURCHASES_FREQUENCY: How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)

10. PURCHASES_INSTALLMENTS_FREQUENCY: How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)

11. CASH_ADVANCE_FREQUENCY: How frequently the cash in advance being paid

12. CASH_ADVANCE_TRX: Number of Transactions made with "Cash in Advance"

DATA DESCRIPTION

13. PURCHASES_TRX: Number of purchase transactions made

14. CREDIT_LIMIT: Limit of Credit Card for user

15. PAYMENTS: Amount of Payment done by user

16. MINIMUM_PAYMENTS: Minimum amount of payments made by user

17. PRC_FULL_PAYMENT: Percent of full payment paid by user

18. TENURE: Tenure of credit card service for user

PERFORMING DATA ANALYSIS

-So, to start with our problem, we will clean the dataset by checking for null values, handling outliers, checking for data consistency.

A) Describing the data

```
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
BALANCE	8950.0	1564.474828	2081.531879	0.000000	128.281915	873.385231	2054.140036	19043.13856
BALANCE_FREQUENCY	8950.0	0.877271	0.236904	0.000000	0.888889	1.000000	1.000000	1.000000
PURCHASES	8950.0	1003.204834	2136.634782	0.000000	39.635000	361.280000	1110.130000	49039.57000
ONEOFF_PURCHASES	8950.0	592.437371	1659.887917	0.000000	0.000000	38.000000	577.405000	40761.25000
INSTALLMENTS_PURCHASES	8950.0	411.067645	904.338115	0.000000	0.000000	89.000000	468.637500	22500.00000
CASH_ADVANCE	8950.0	978.871112	2097.163877	0.000000	0.000000	0.000000	1113.821139	47137.21176
PURCHASES_FREQUENCY	8950.0	0.490351	0.401371	0.000000	0.083333	0.500000	0.916667	1.000000
ONEOFF_PURCHASES_FREQUENCY	8950.0	0.202458	0.298336	0.000000	0.000000	0.083333	0.300000	1.000000
PURCHASES_INSTALLMENTS_FREQUENCY	8950.0	0.364437	0.397448	0.000000	0.000000	0.166667	0.750000	1.000000
CASH_ADVANCE_FREQUENCY	8950.0	0.135144	0.200121	0.000000	0.000000	0.000000	0.222222	1.500000
CASH_ADVANCE_TRX	8950.0	3.248827	6.824647	0.000000	0.000000	0.000000	4.000000	123.00000
PURCHASES_TRX	8950.0	14.709832	24.857649	0.000000	1.000000	7.000000	17.000000	358.00000
CREDIT_LIMIT	8949.0	4494.449450	3638.815725	50.000000	1600.000000	3000.000000	6500.000000	30000.00000
PAYMENTS	8950.0	1733.143852	2895.063757	0.000000	383.276166	856.901546	1901.134317	50721.48336
MINIMUM_PAYMENTS	8637.0	864.206542	2372.446607	0.019163	169.123707	312.343947	825.485459	76406.20752
PRC_FULL_PAYMENT	8950.0	0.153715	0.292499	0.000000	0.000000	0.000000	0.142857	1.000000
TENURE	8950.0	11.517318	1.338331	6.000000	12.000000	12.000000	12.000000	12.000000

Insights

1. Mean balance is \$1564
2. Balance frequency is frequently updated on average ~0.9
3. Purchases average is \$1000
4. one off purchase average is ~\$600
5. Average purchases frequency is around 0.5
6. average ONEOFF_PURCHASES_FREQUENCY, PURCHASES_INSTALLMENTS_FREQUENCY, and CASH_ADVANCE_FREQUENCY are generally low
7. Average credit limit ~ 4500
8. Percent of full payment is 15%
9. Average tenure is 11 years

PERFORMING DATA ANALYSIS

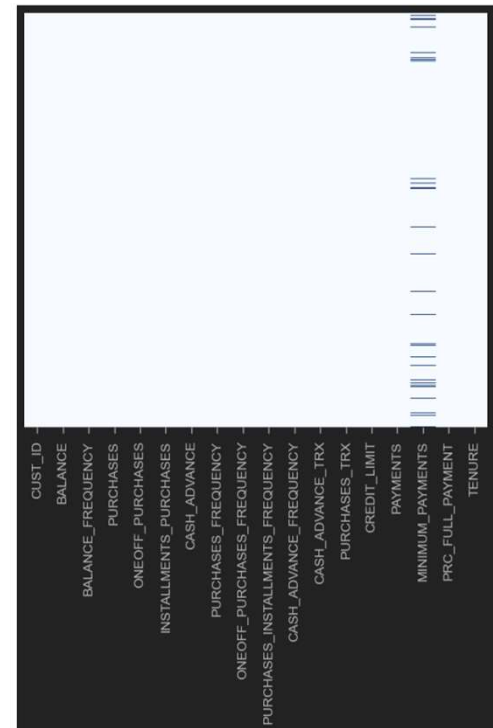
B) Checking for missing values –

Let's see if we have any missing data

Replacing missing value with mean

```
df.isnull().sum()
```

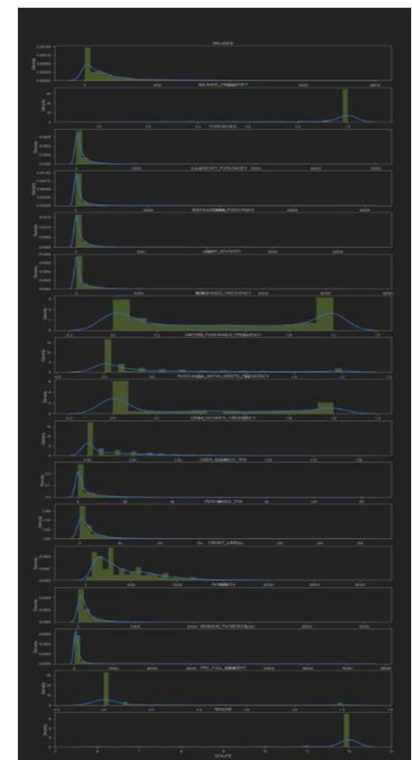
```
CUST_ID      0
BALANCE      0
BALANCE_FREQUENCY  0
PURCHASES    0
ONEOFF_PURCHASES  0
INSTALLMENTS_PURCHASES  0
CASH_ADVANCE  0
PURCHASES_FREQUENCY  0
ONEOFF_PURCHASES_FREQUENCY  0
PURCHASES_INSTALLMENTS_FREQUENCY  0
CASH_ADVANCE_FREQUENCY  0
CASH_ADVANCE_TRX  0
PURCHASES_TRX  0
CREDIT_LIMIT  1
PAYMENTS     0
MINIMUM_PAYMENTS 313
PRC_FULL_PAYMENT  0
TENURE       0
dtype: int64
```



DATA VISUALIZATION

Insights

1. Mean of balance is 1500 dollars
2. 'Balance_Frequency' for most customers is updated frequently ~1
3. For 'PURCHASES_FREQUENCY', there are two distinct group of customers
4. For 'ONEOFF_PURCHASES_FREQUENCY' and 'PURCHASES_INSTALLMENT_FREQUENCY' most users don't do one off purchases or installment purchases frequently
5. Very small number of customers pay their balance in full 'PRC_FULL_PAYMENT'~0
6. Credit limit average is around \$4500
7. Most customers are ~11 or ~12 years tenure



DATA VISUALIZATION

Insights

1. 'PURCHASES' have high correlation between one-off purchases, 'installment purchases, purchase transactions and payments.
2. Strong Positive Correlation between 'PURCHASES_FREQUENCY' and 'PURCHASES_INSTALLMENT_FREQUENCY'



PRINCIPAL COMPONENT ANALYSIS (PCA)

- PCA is an unsupervised machine learning algorithm.
- PCA performs dimensionality reductions while attempting at keeping the original.
- PCA works by trying to find a new set of features called components.
- Components are composites of the uncorrelated given input features.

	PCA1	PCA2
0	-1.682221	-1.076453
1	-1.138292	2.506465
2	0.969685	-0.383512
3	-0.873628	0.043159
4	-1.599435	-0.688581
...
8945	-0.359631	-2.016143
8946	-0.564374	-1.639120
8947	-0.926206	-1.810784
8948	-2.336555	-0.657964
8949	-0.556421	-0.400461

8950 rows × 2 columns

Model Evaluation

K-Mean Clustering

`K-means` is an unsupervised learning algorithm (clustering). K-means works by grouping some data points together (clustering) in an unsupervised. The algorithm groups observations with similar attribute values together by measuring the Euclidian distance between points.

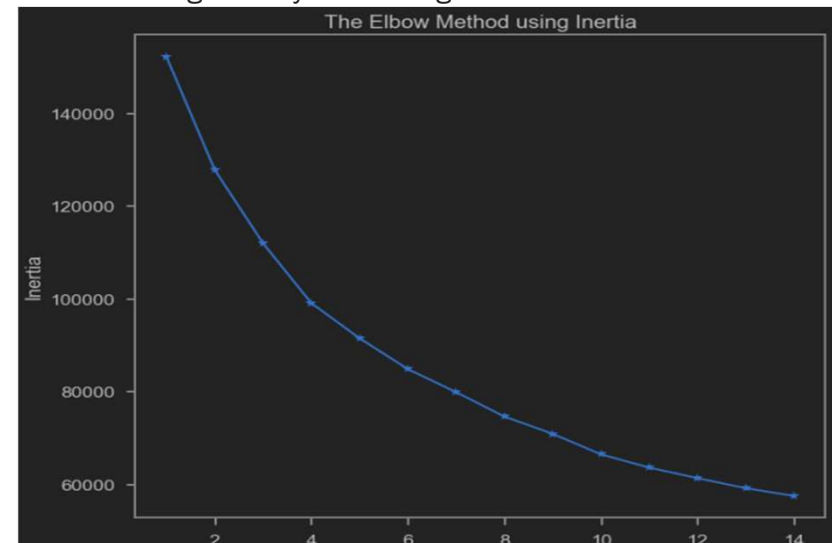
-Elbow method is customer check how many clusters are best for my

Model. Here you can clearly see its 4.

Silhouette Score = 0.4992866715615686

Silhouette Samples = [0.63300434 0.35823128 0.55249751 ... 0.22360298 0.64401382 0.55818347]

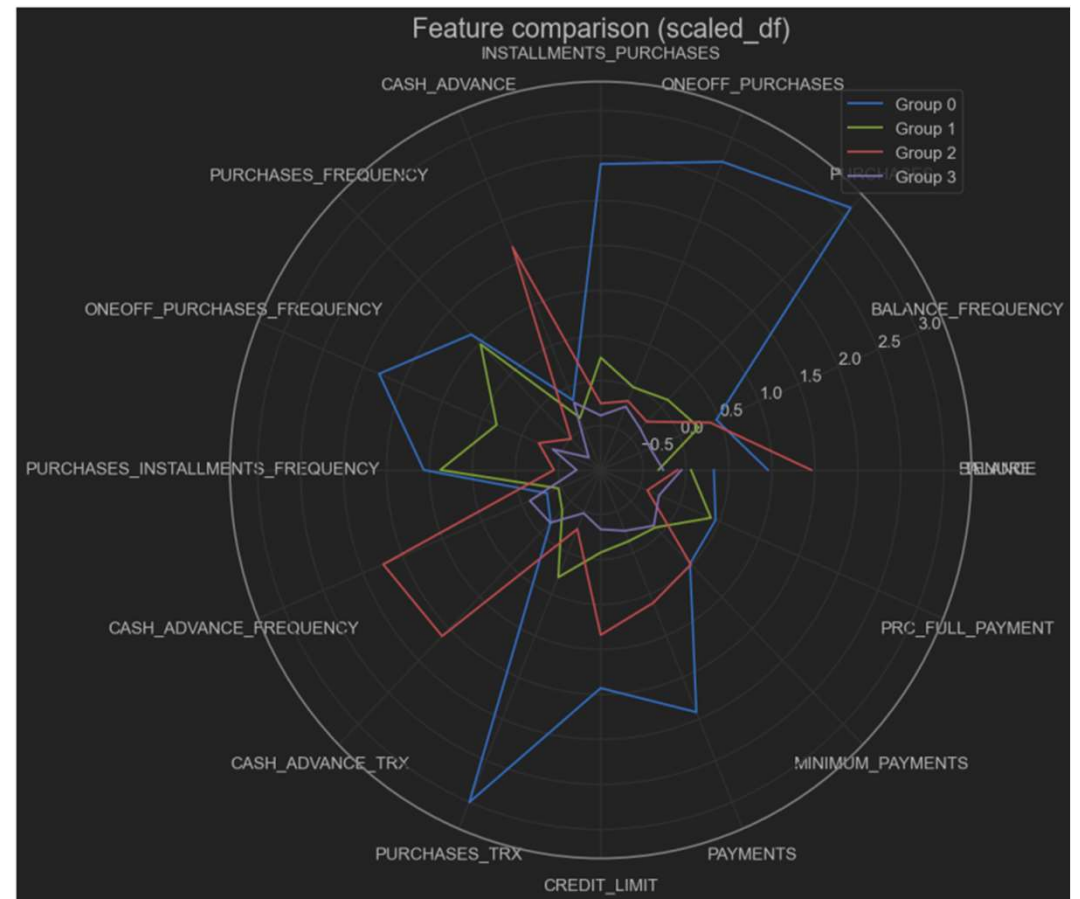
-The score is impressive which means our cluster are neither overlapping nor tight clusters.



the following plot should give us a better understanding on our clusters:

Standard, Imputed Data vs Labels!

The following plot should give us a better understanding on our clusters:



Standard, Imputed Data vs Labels!

Wow! now we are talking. I know it's hard to read, but I just want you to notice these points:

Group 0 contains customers who have a good income (second highest BALANCE) and are enjoying it! they purchase a lot (highest PURCHASESFREQUENCY), both in installments and in one-go. let's call them Bourgeoisie!

Group 1 contains customers who don't make a lot of money (look at their BALANCE , it is the lowest of all) but this doesn't keep them away from purchasing stuff! in terms of PURCHASES , they are the second class. how do they do this? take a closer look: they don't buy stuff in one go (they have lowest ammount of ONEOFF_PURCHASES and ONEOFF PURCHASES FREQUENCY) their key to success is instalments! It's easy: if you don't make enough money to buy stuff in one go, just pay over a period of time. (They have highest values of INSTALLMENTSPURCHASES and PURCHASESINSTALLMENTSFREQUENCY I refer to these people as Dreamers because although they don't make much money, lack of money doesn't prevent them from reaching for their dreams!

Standard, Imputed Data vs Labels!

Group 2 is mysterious. look at them! they have highest BALANCE , but lowest PURCHASE of all! it seems they only use their fortunes when they want pay in advance (highest CASHADVANCE , CASHADVANCEFREQUENCY and CASHADVANCETRX). A cash advance is a service provided by most credit card and charge card issuers. The service allows cardholders to withdraw cash, either through an ATM or over the counter at a bank or other financial agency, up to a certain limit. For a credit card, this will be the credit limit (or some percentage of it). so, these guys don't use their credit cards to buy stuff, instead, they get chash from ATMs to do so. why? it is beause they want to buy something illegal? Let's just call them The Mafia for now.

Group 3 shows customers who are not very rich, and don't take risks. their BALANCE (amount left in their account to make purchases) is below average, and they don't purchase much. (their PURCHASES is below average as well, and their PURCHASE FREQUENCY is very low.) I call these people Economicals. to them, every penny is important.

Let do some more EDA

Evaluating our hypothesis now,

I want to plot our data using only BALANCE and PURCHASES. this is my hypothesis:

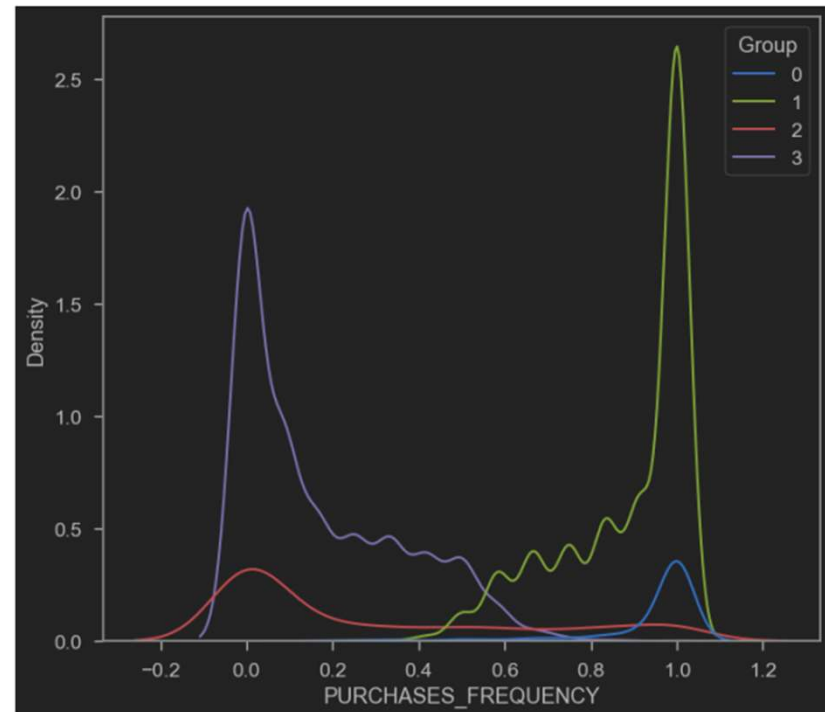
If BALANCE is high and PURCHASES is high → Class 0 (Bourgeoisie 🏠)

If BALANCE is low and PURCHASES is high → Class 1 (Dreamers ✨)

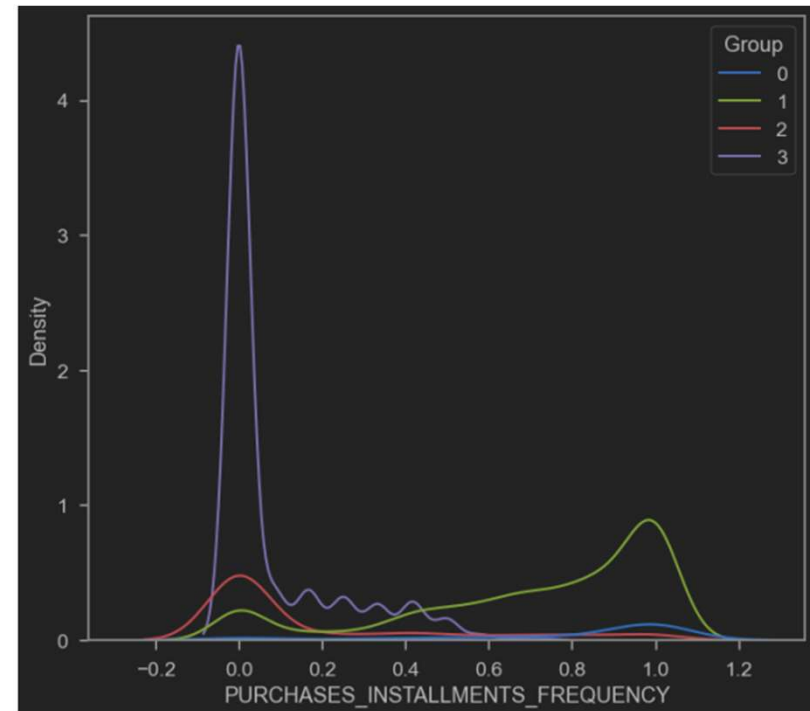
If BALANCE is high and PURCHASES is low → Class 2 (The Mafia 🕶)

If BALANCE is low and PURCHASES is low Class 3 (Economicals 💰)

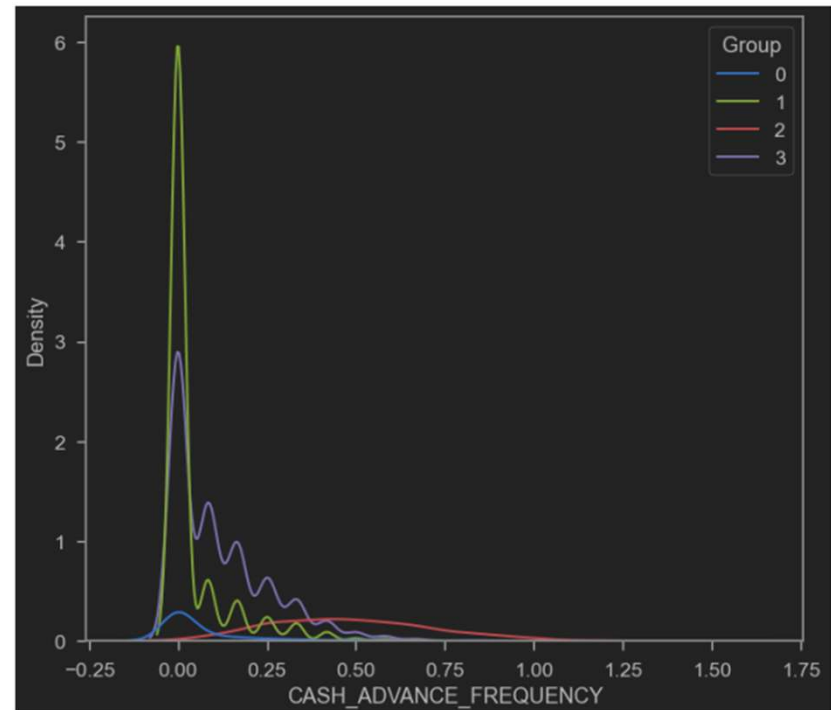
Looking at this plot, it is obvious that the mafia and economicals are purchasing less often than dreamers and Bourgeoisie



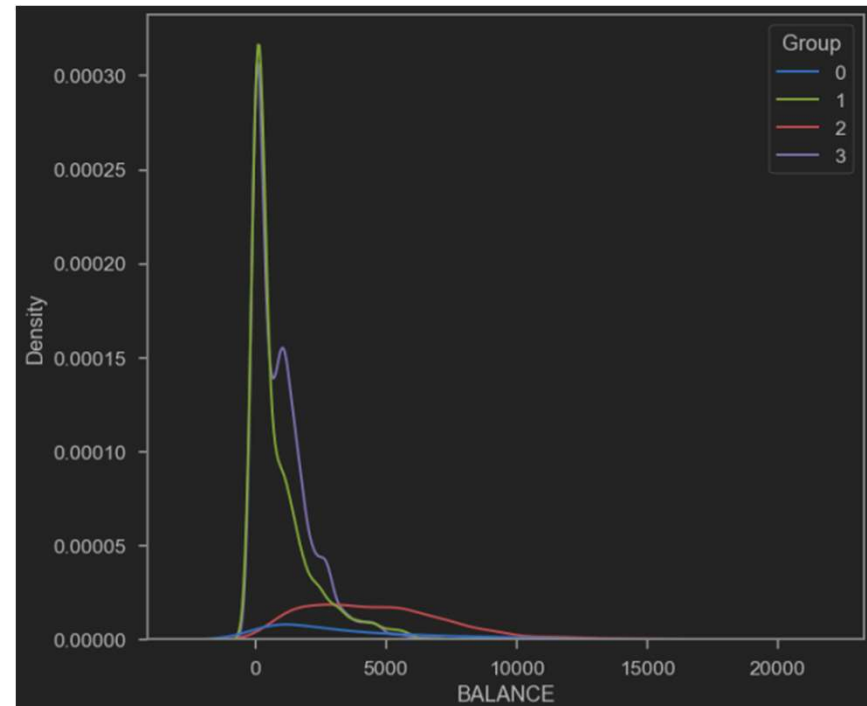
This plot shows how the dreamers are trying to buy whatever they pursue, by buying first and paying later.



Here, this plot shows that the infamous mafia are getting cash from ATMs more often than other groups.



The mafia and the Bourgeoisie
has highest balance then others

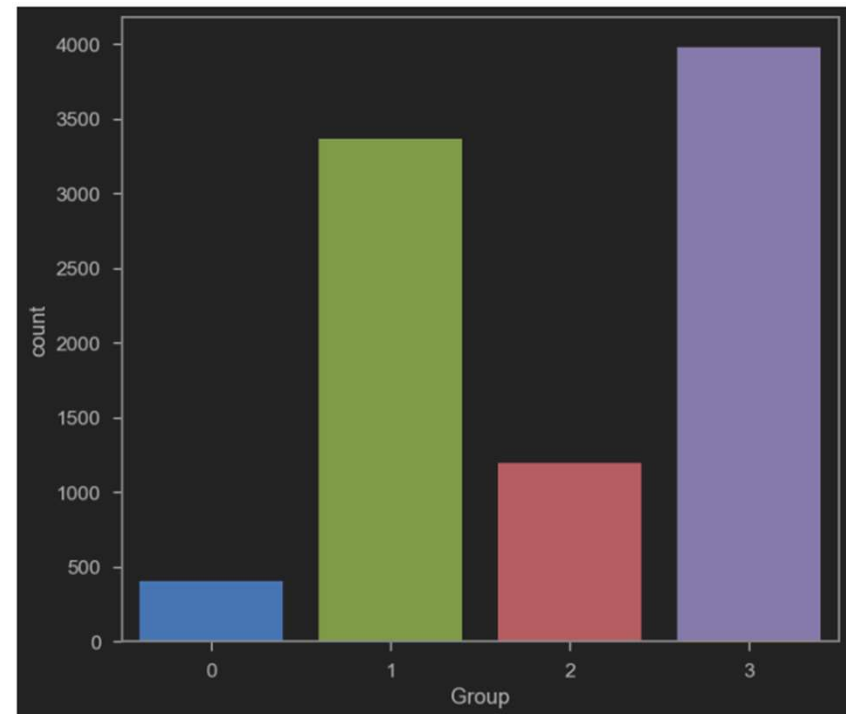


It looks like my hypothesis was quite right.

In this plot, it is clear that:

- People of Class 0 have high balance and purchase a lot
- People of Class 1 spend a lot while they have low balance.
- People of Class 2 have a low balance and spend less than others
- People of Class 3 don't purchase much, although they have lots of money we can further investigate this hypothesis using some plots:

This shows how my data is distributed in the clusters.



Thank you!
