Super Store Segmentation and Discount Analysis

Aliriza Hamonangan Matondang
Email : riza.ali18@gmail.com
LinkedIn : https://www.linkedin.com/in/alirizahm/

Contents of this template

To find the code, you can click here

Table of Content

Background

A superstore is a very large supermarket, often selling household goods, clothes, and electrical goods, as well as food. Superstores typically charge anywhere from 15 to 45 percent less than their smaller counterparts. As a Data Analyst, try to find out the weak areas where you can work to make more profit

Background

discount	year	
0.0	2014	58617.1665
	2015	68870.7566
	2016	89341.9679
	2017	104157.7122
	total	320987.6032
discount_promotion	2014	-9073.1924
	2015	-7252.1529
	2016	-7546.7936
	2017	-10718.4426
	total	-34590.5815

Name: profit, dtype: float64

1. The profit of Super Store increases every year, but Super Store needs to increase more profit.
2. The Super Store has discount promotion that impact to decrease the profit year by year.
3. The discount promotion doesn't improve so well to profit, but the discount promotion from the Super Store is quite good, ranging from 10 - 20% of the profit received

Objective

1. Create customer segmentation to understand the customer profile
2. Do discount analysis by the segmentation of the customer
3. Create the recommendation of discount promotion
4. Issue identification and the weakness of Super Store to increase the Profit

Analysis Scope and Approach

The Dataset that is used in this analysis is Super Store from Kaggle

The method of this analysis are RFM segmentation and K-Means

Dataset Information

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   object
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: datetime64[ns](1), float64(3), int64(3), object(14)
memory usage: 1.6+ MB
```

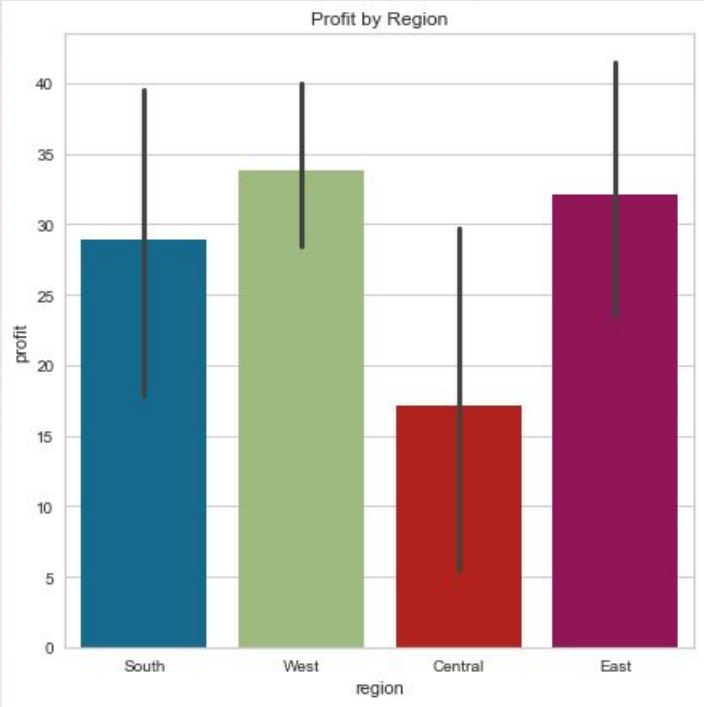
Features

1. Row ID => Unique ID for each row.
2. Order ID => Unique Order ID for each Customer.
3. Order Date => Order Date of the product.
4. Ship Date => Shipping Date of the Product.
5. Ship Mode=> Shipping Mode specified by the Customer.
6. Customer ID => Unique ID to identify each Customer.
7. Customer Name => Name of the Customer.
8. Segment => The segment where the Customer belongs.
9. Country => Country of residence of the Customer.
10. City => City of residence of of the Customer.
11. State => State of residence of the Customer.
12. Postal Code => Postal Code of every Customer.
13. Region => Region where the Customer belong.
14. Product ID => Unique ID of the Product.
15. Category => Category of the product ordered.
16. Sub-Category => Sub-Category of the product ordered.
17. Product Name => Name of the Product
18. Sales => Sales of the Product.
19. Quantity => Quantity of the Product.
20. Discount => Discount provided.
21. Profit => Profit/Loss incurred.

1. The dataset is from 2014 - 2017
2. The dataset contains 21 Features
3. The dataset contains 9994 rows

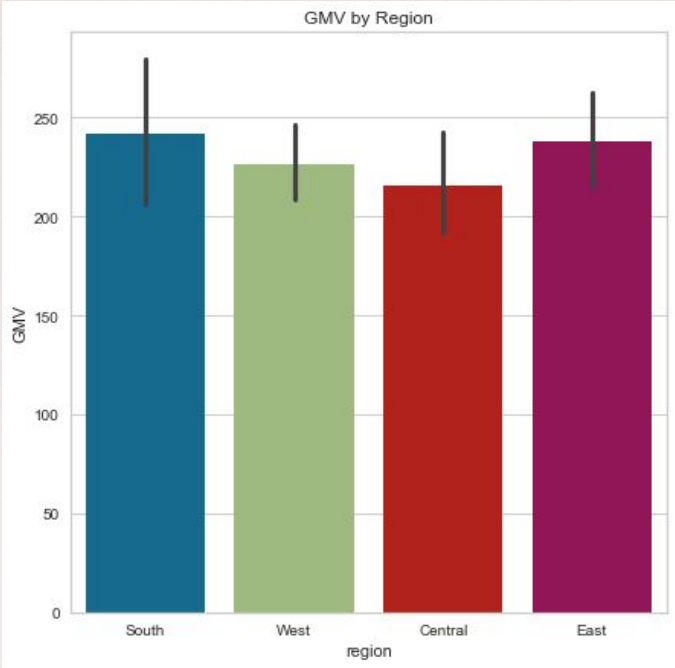
Exploratory Data Analysis

Exploratory Data Analysis



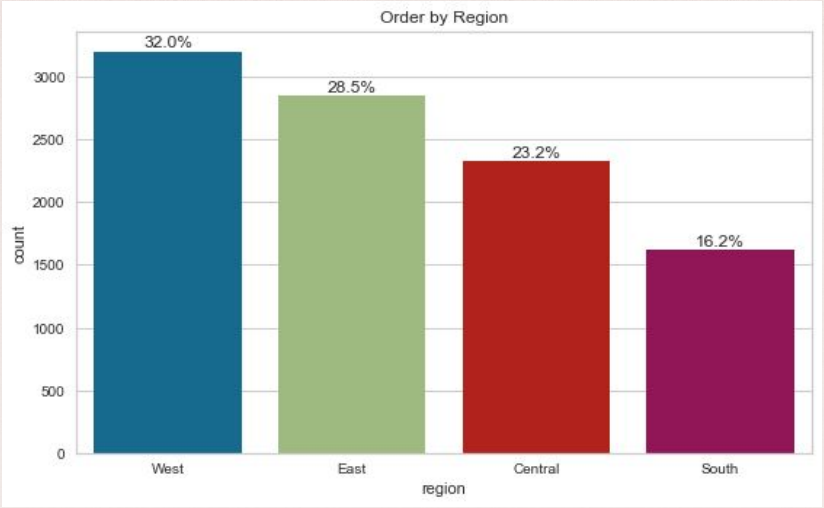
1. The West Region has highest profit
2. The Central Region has lowest profit

Exploratory Data Analysis



1. The South Region has highest GMV
2. The Central Region has lowest profit
3. The GMV differences in each region are not really significant
4. It seems that the Central Region provides a lot of promotions, so that the profit perform is less level than the maximum

Exploratory Data Analysis



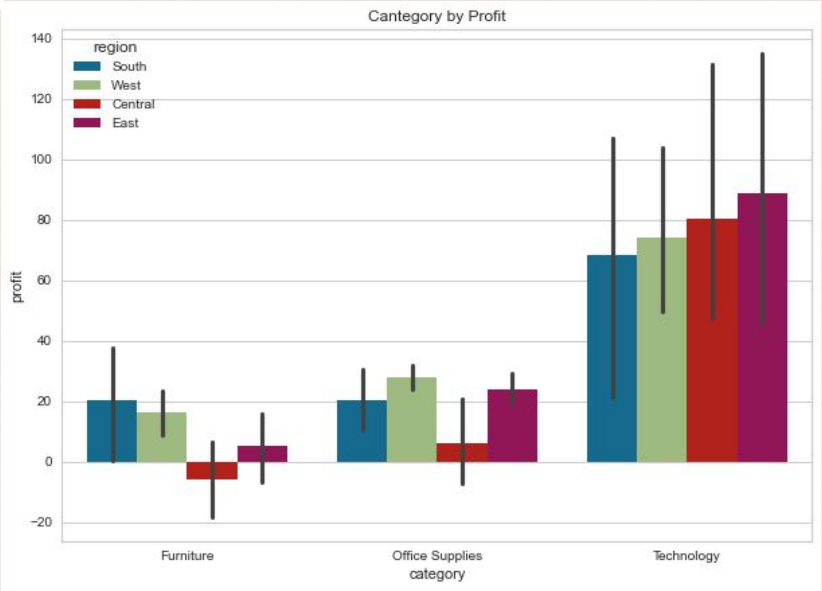
1. The order level South Region is the lowest than others region, but the profit performance is so good and competitive with East and West region
2. At the same time, central region has a lot of orders, but the profit is low.
3. It seems the promotions have significant impact on the region central profit

Exploratory Data Analysis

region	Total_Customer
Central	629
East	674
South	512
West	686

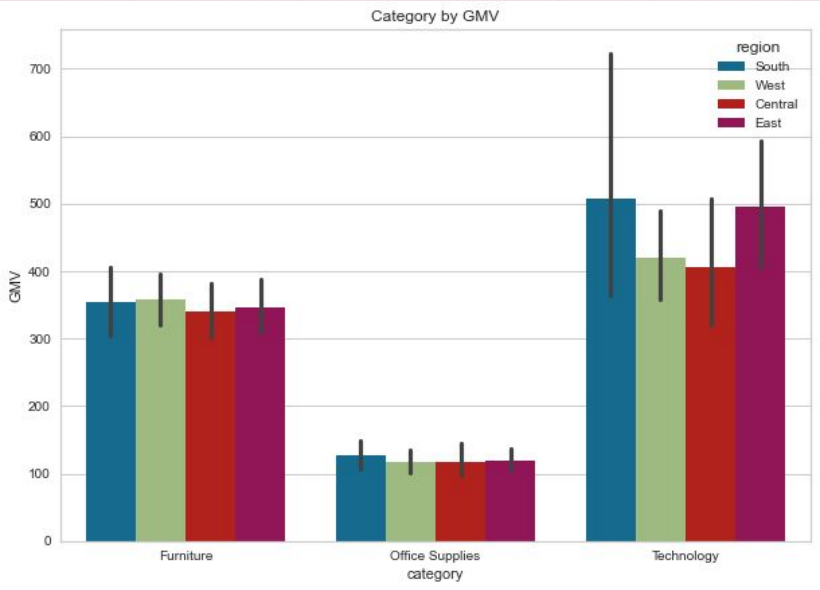
1. The Total Customer differences in each region are not really significant
2. It seems like the Central region did the promotion to increase new customers
3. Apparently, South region has the lowest total customer who are very loyal customers. Therefore, South region has competitive profit.

Exploratory Data Analysis



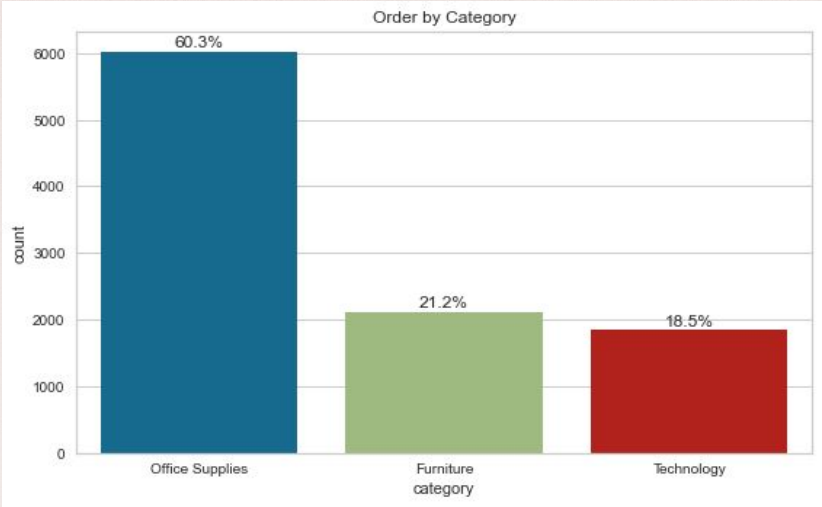
1. The Technology category has the highest profit compared to other categories, because the price for the Technology category is very high and every customer needs a Technology category product
2. It seems like the East and Central regions did promotion for the Furniture category
3. Apparently, Central region did promotion for Office Supplies Category, because the profit is so low

Exploratory Data Analysis



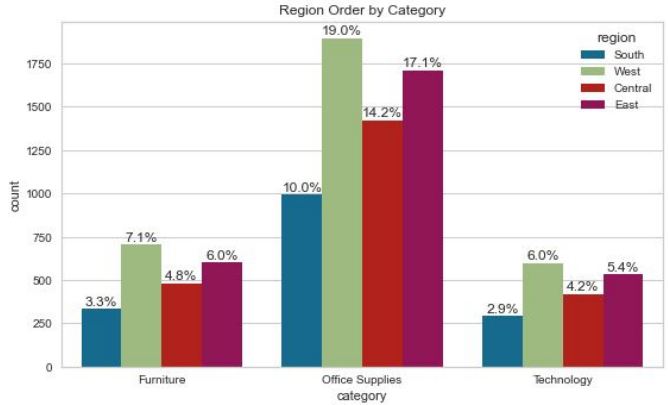
1. Overall, the total GMV every region is so well, there is no significant differentiation total GMV in every region
2. The significant differentiation appears on Office Supplies Category compared others category
3. Apparently, the regional central did promotion on furniture and office supplies category products to increase new customers or provide treatment to old customers

Exploratory Data Analysis



1. The Office Supplies Product Category has the highest order but it has the lowest GMV compare to other Product Categories
2. It looks like the Office Supplies Product Category has low prices and has a lot of promotion
3. The Office Supplies Product Category is the best Product Category for Promotion, because it has so many orders and the demand is really high.

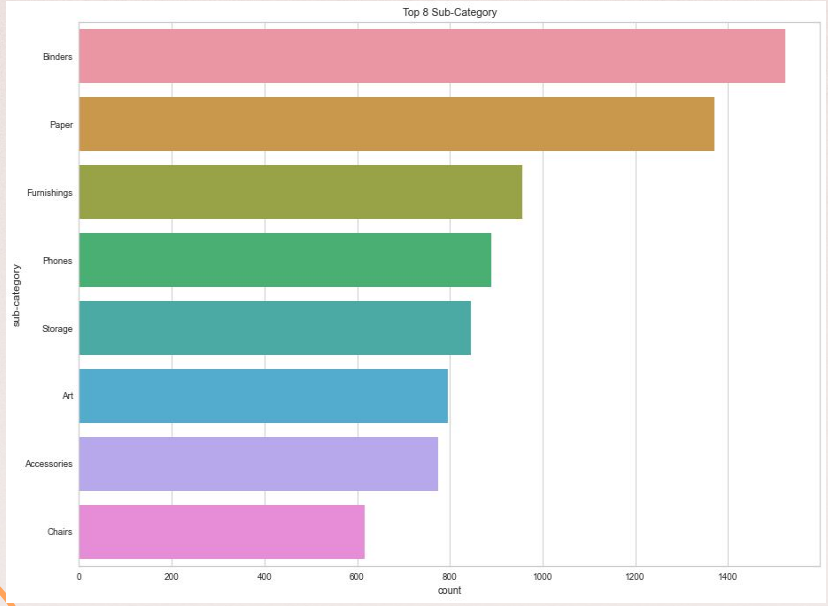
Exploratory Data Analysis



Findings :

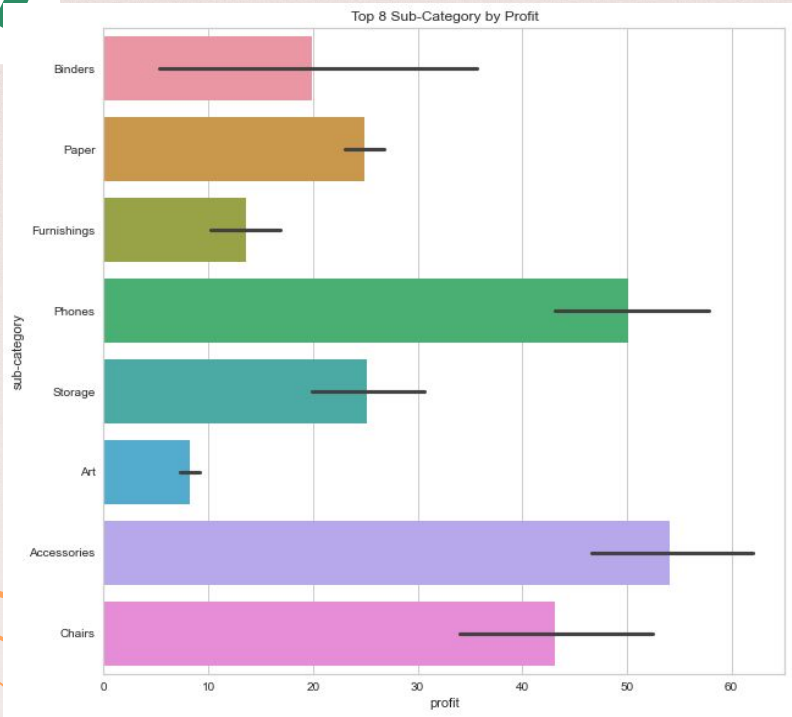
1. Office Supplies has highest order than others category products
2. Office Supplies has so many demand, cheap price, and has promotion. So, Office Supplies has highest order than others category products, but the highest prodit is belong to Technology
3. Super store needs to increase order on technology products category for increasing profit via promotion
4. Because of the Office Supplies demand is really high, therefore Super Store needs to decrease the promotion on Office 5. Supplies, I thought they buy the products its because they need them
5. Super Store has tried to increase order on Furniture product category with promotion, as appropriate action
6. I think Furniture product category is not the priority product category by Super Store, however Furniture product category has so competitive GMV, so Super Store has tried to introduce Furniture product category by promotion

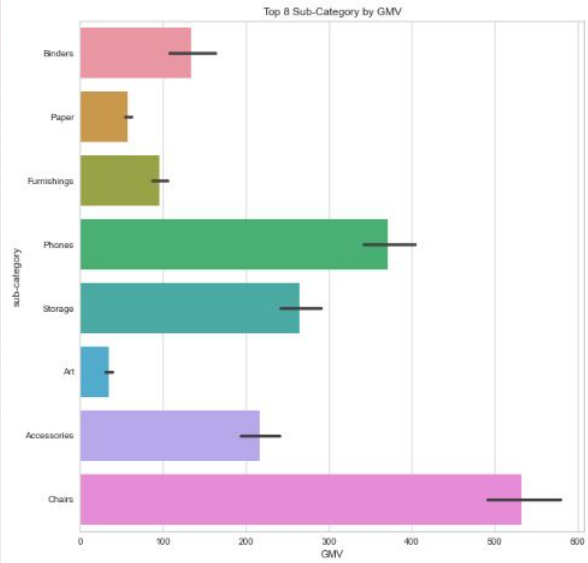
Exploratory Data Analysis



1. The top 8 Sub-Category
2. Binder is from Office Supplies products category and It has the highest order than others
3. Super Store needs to increase order from Sub-Category, especially Furniture and Technology Product Category for increasing Profit

Exploratory Data Analysis

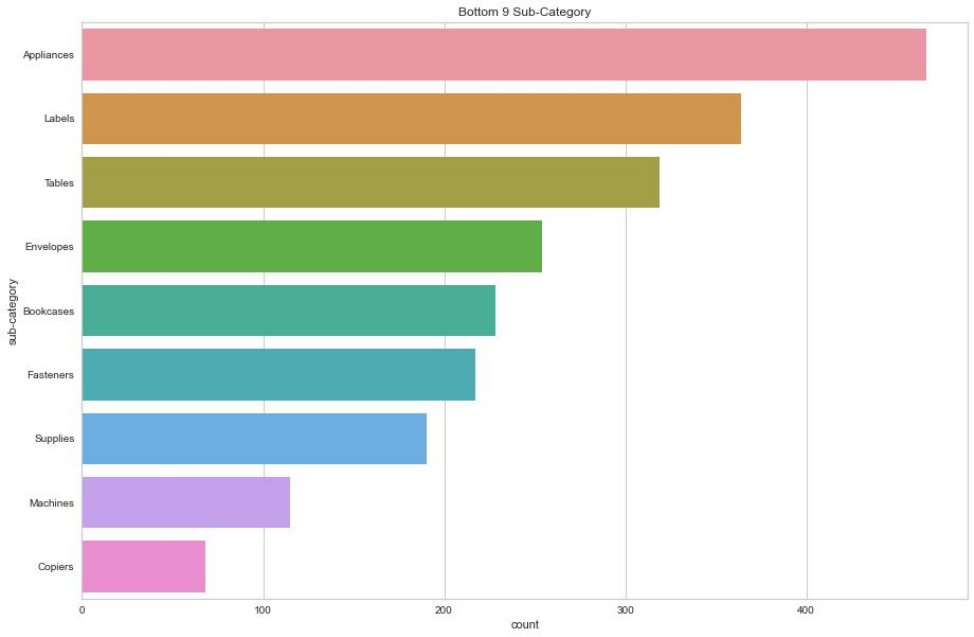




Findings :

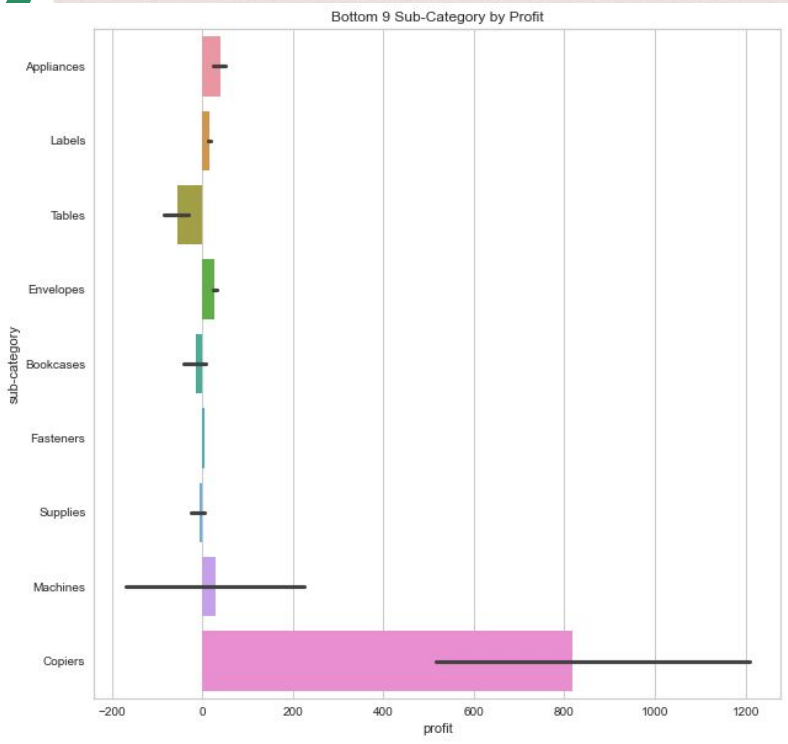
1. If we compare the GMV and Profit on top 8 Sub-Catergory the results id so good, because the difference between GMV and is still normal
2. The top 8 Sub-Catergory has good performance by Order, GMV, and Profit, in case Super Store needs to increase order on the bottom 9 Sub-Catergory to increase profit
3. Super Store needs to decrease promotion on top 8 Sub-Catergory, but for the Technology Products, they still need measurable promotion to increase Profit

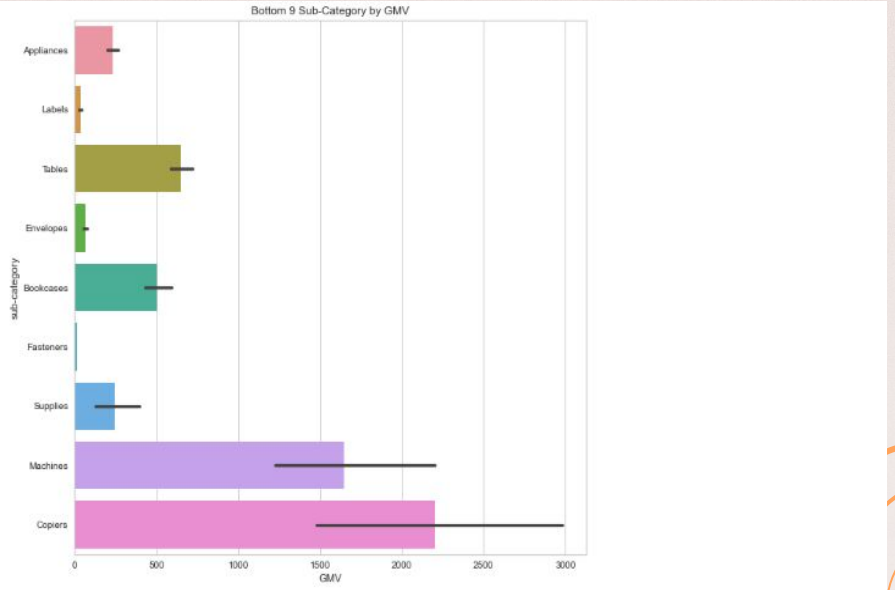
Exploratory Data Analysis



1. Super Store needs to increase Bottom 9 Sub-Category orders to increase profit
2. The marketing team could create either a promotion or advertising program for the Bottom 9 Sub-Category exp : create a storefront promotion for the Bottom 9 Sub-Category and put it in front
3. Create either online ads or collaborate with Influencers to introduce the bottom 9 Sub-Category Products

Exploratory Data Analysis

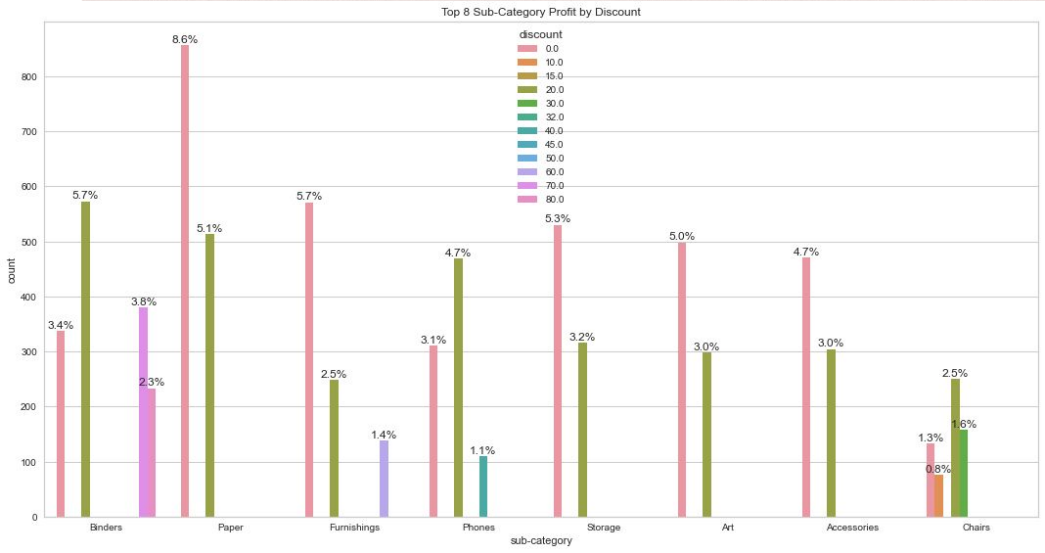




Findings:

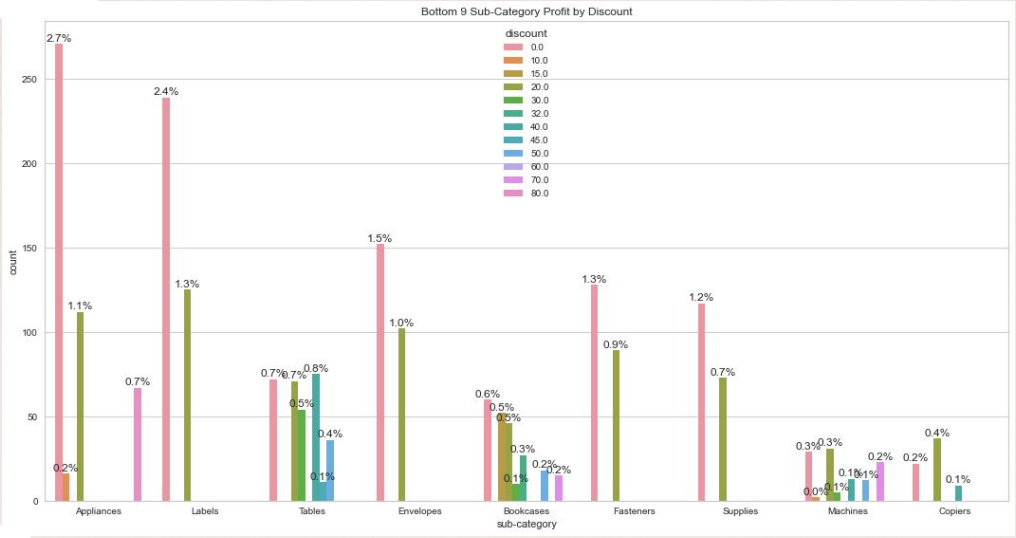
1. If we compare the GMV and profit, this is indecate that Super Store did promotion at the Bottom 9 Sub-Category products, because the difference between GMV and Profit is very high
2. Super Store carried out right promotion at the Bottom 9 Sub-Caregory products, but Super Store needs either to maximise promotions or evaluate the products

Exploratory Data Analysis



1. Overall, the promotion at Top 8 Sub-Category is good but, less than maximum
2. Super Store needs to create efficiently promotion to streamline the promotion budget Top 8 Sub-Category
3. Super Store is able to create a Loyalty program to streamline budget promotion
4. The Customers usually buy the products because they need them, and the percentage discount of 0.0% is very high. So, Super Store is able to undertake efficiently the promotion budget at Top 8 Sub-Category

Exploratory Data Analysis



1. Overall, the discount promotion which is carried out by Super Store is quite good for the Bottom 9 Sub-Categories, but it was less than maximum because the GMV is still low
2. The percentage discount of 0.0% is very high which indicates that customers buy the products without promotions
3. Since Super Store needs to introduce Bottom 9 Sub-Category, they need to make either a good promotion or event based on Bottom 9 Sub-Category products
4. The orders of Bottom 9 Sub-Category is still low but, the promotion user is so high

Exploratory Data Analysis

discount	segment	Total_Quantity	Total_Profit	Avg_Unite_Price
0.0	Consumer	9359	157901.9614	56.029979683055664
0.0	Corporate	5512	102150.7824	63.419285213735115
0.0	Home Office	3396	60934.8594	59.60141758241758
10.0	Consumer	210	4681.5045	156.305
10.0	Corporate	98	2694.6798	194.1577777777778
10.0	Home Office	65	1652.9927	133.6823529411765
15.0	Consumer	102	456.59450000000004	165.22206896551725
15.0	Corporate	62	689.95	158.91071428571428
15.0	Home Office	34	272.447	199.44333333333333
20.0	Consumer	7106	42969.3096	72.07280271398747
20.0	Corporate	4246	26214.485	69.3487632508834
20.0	Home Office	2308	21153.5114	77.97243021346469
30.0	Consumer	464	-6338.4144	188.51924242424244
30.0	Corporate	231	-2543.2759	146.5990909090909
30.0	Home Office	154	-1487.5871	201.79125
32.0	Consumer	50	-1335.2122	240.28769230769234
32.0	Corporate	45	-827.0201	154.281
32.0	Home Office	10	-228.9054	188.48000000000002
40.0	Consumer	377	-8972.3559	255.5091666666667
40.0	Corporate	269	-9768.2636	253.71716216216217
40.0	Home Office	140	-4316.4309	185.27333333333334
45.0	Consumer	23	-1299.9114	256.6083333333333
45.0	Corporate	17	-842.9059	179.6966666666667
45.0	Home Office	5	-350.29380000000003	294.365
50.0	Consumer	123	-9382.1707	302.0016216216216
50.0	Corporate	65	-6598.8014	421.1441176470588
50.0	Home Office	53	-4525.456	845.1716666666666
60.0	Consumer	264	-2859.3468	30.57608108108108
60.0	Corporate	159	-1584.3097	29.326923076923077
60.0	Home Office	78	-1500.9987	39.34
70.0	Consumer	812	-22462.111	90.69809523809525
70.0	Corporate	552	-9183.2091	57.119851851851855
70.0	Home Office	296	-8430.0368	87.73945205479453
80.0	Consumer	631	-19240.6384	67.50065868263472
80.0	Corporate	352	-8422.9775	66.37586206896553
80.0	Home Office	205	-2875.4233	49.00586956521739

Findings :

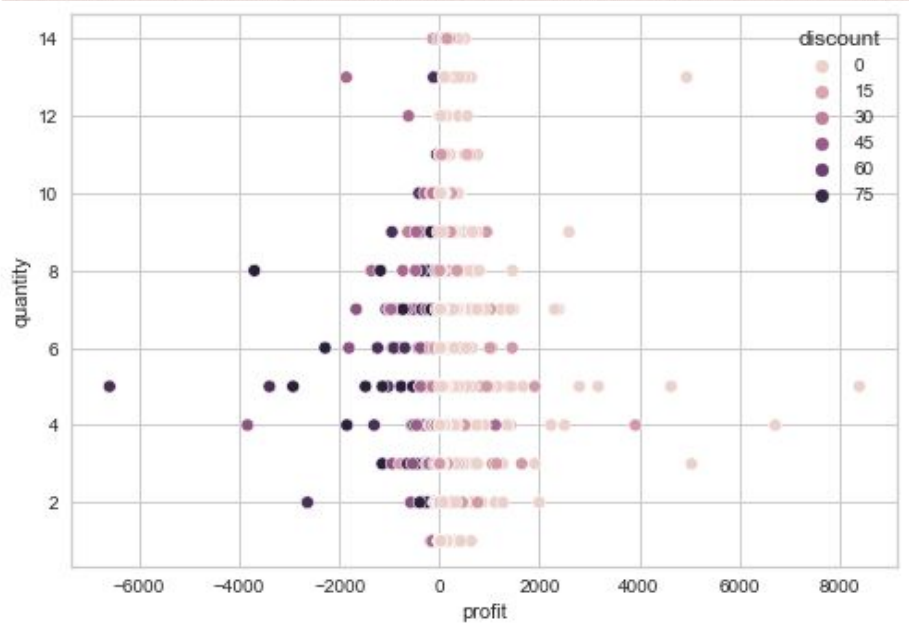
1. From the following table, we can see that the discount promotion given from 30% - 80%. The results is loss profit
2. Super Store needs to change the discount promotion that is on target
3. The existence of a loyalty program is able to provide promotion to the target customers
4. From the segment Corporate and Home Office segments, They are better to offer a good price than give discount promotion
5. It is necessary to have either a contract or make a sales agreement at the Corporate or Home Office with a good price offer in order to increase profit. So Super Store is able do efficiently on budget promotion
6. Discount promotions are also still be given to Corporate or Home Offices, however, in accordance with the existing promotion budget which refers to profit
7. Create a periodic promotion is also very good to streamline the budget promotion
8. Super Store needs to validate the existence of a profit that is 0. There is a concern which is an error in the data, or the product is on the giveaway promotion
9. The Product prices is also be used as a reference for determie the discount promotion, so that the promotion be given are not excessive. Ex: high prices, low discounts, but refer to profit

Exploratory Data Analysis

discount	category	region	Total_Customer	Total_Profit	avg_unite_price
0.0	Furniture	Central	156	16641.3819	113.34544871794871
0.0	Furniture	East	226	12708.6124	62.749159292035394
0.0	Furniture	South	160	14721.2792	104.02349999999998
0.0	Furniture	West	294	14061.8029	50.32360544217687
0.0	Office Supplies	Central	517	34732.6993	50.44295938104449
0.0	Office Supplies	East	916	35205.6872	34.15524017467249
0.0	Office Supplies	South	492	25281.7617	44.356402439024386
0.0	Office Supplies	West	1204	35285.9582	31.925191029900333
0.0	Technology	Central	155	24751.3595	119.68722580645162
0.0	Technology	East	307	57463.2468	149.65273615635178
0.0	Technology	South	153	22978.2969	110.5703267973856
0.0	Technology	West	218	27155.5172	93.05334862385321
10.0	Furniture	East	76	7111.0119	156.96
10.0	Office Supplies	Central	16	1086.0808	98.198125
10.0	Technology	Central	2	832.0843	914.99
15.0	Furniture	West	52	1418.9915	169.44576923076923
20.0	Furniture	East	142	1150.5055	80.74450704225352
20.0	Furniture	South	145	890.7057	75.00220689655173
20.0	Furniture	West	328	4224.7379	159.4661280487805
20.0	Office Supplies	Central	589	3600.239	28.41984719864177
20.0	Office Supplies	East	625	11780.535	43.744127999999996
20.0	Office Supplies	South	371	3109.105	33.79140161725068
20.0	Office Supplies	West	616	19548.8722	41.06209415584416
20.0	Technology	Central	245	10454.7921	98.05122448979591
20.0	Technology	East	95	8191.7888	123.1621052631579
20.0	Technology	South	128	4648.7636	101.64796875
20.0	Technology	West	373	22737.2612	157.61246648793568
30.0	Furniture	Central	142	-6866.8937	170.2473943661972
30.0	Furniture	East	80	-3828.4232	184.06787500000002
30.0	Technology	Central	5	326.0395	423.716
32.0	Furniture	Central	27	-2391.1376999999998	200.75814814814817
40.0	Furniture	East	59	-9839.7291	251.4701694915254
40.0	Furniture	South	16	-6347.6677	261.38
40.0	Technology	Central	13	-2666.8433999999997	566.4915384615385
40.0	Technology	East	118	-4202.8102	199.92016949152543
45.0	Furniture	South	11	-2493.1111	242.49727272727276
50.0	Furniture	Central	18	-4309.7447	202.61055555555555
50.0	Furniture	East	18	-4255.8117	162.65444444444444
50.0	Furniture	West	18	-4305.6426	222.0961111111111
50.0	Technology	South	12	-7635.2291000000005	1491.9225
60.0	Furniture	Central	138	-5944.6552	31.81072463768116
70.0	Furniture	West	15	-3894.9394	137.82133333333334
70.0	Office Supplies	East	171	-5971.6431	42.577660818713454
70.0	Office Supplies	South	132	-8404.4739	63.017196969696975
70.0	Office Supplies	West	77	-2224.9814	28.604155844155844
70.0	Technology	East	15	-13990.1903	658.2046666666666
70.0	Technology	West	8	-5589.1288	427.6
80.0	Office Supplies	Central	300	-30539.0392	64.3386

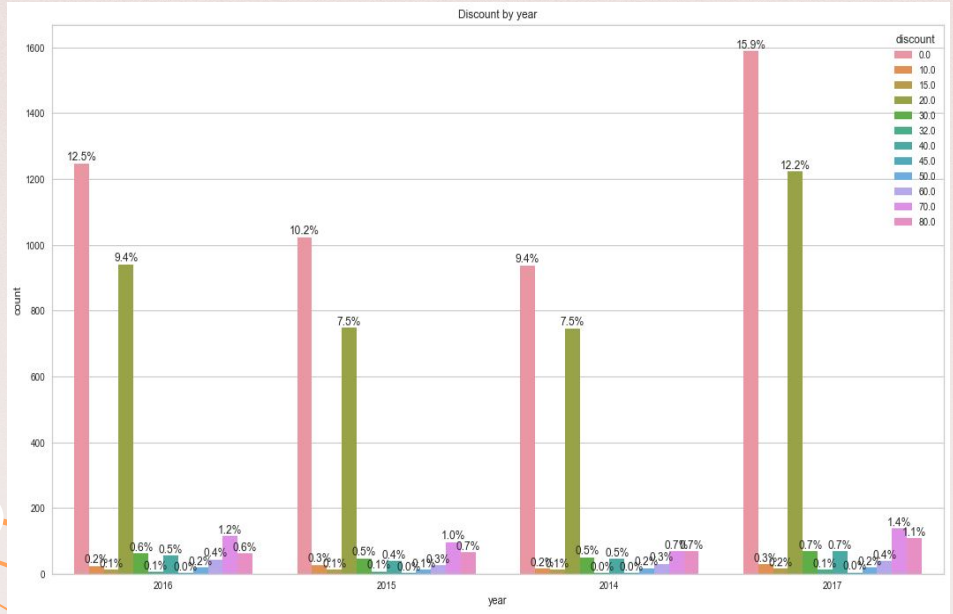
1. Overall, Super Store applied the right promotion, because Super Store undertook promotion on high product price
2. There are a lot of customers who buy product without discount, but only for the cheap products
3. From the table, it is necessary to reduce the discount value in the South, West, East regions on discount value above 20% because the customers usually doesn't buy products discounts
4. Central Region needs to evaluate the discount promotion, because the Central region already has a lot of customers, and the budget of promotion can be used at the south region because the total customers of South Region is the lowest the others region, so south region needs to do promotion to increase number of new customers

Exploratory Data Analysis



1. It can be seen that there is no significant effect of promotion above 30% on profit and quantity
2. The promotion should be able to increase the quantity of purchases, but the low discount promotion and customers who do not buy discount product has so many order quantity
3. Therefore, it is necessary to evaluate the discount promotion

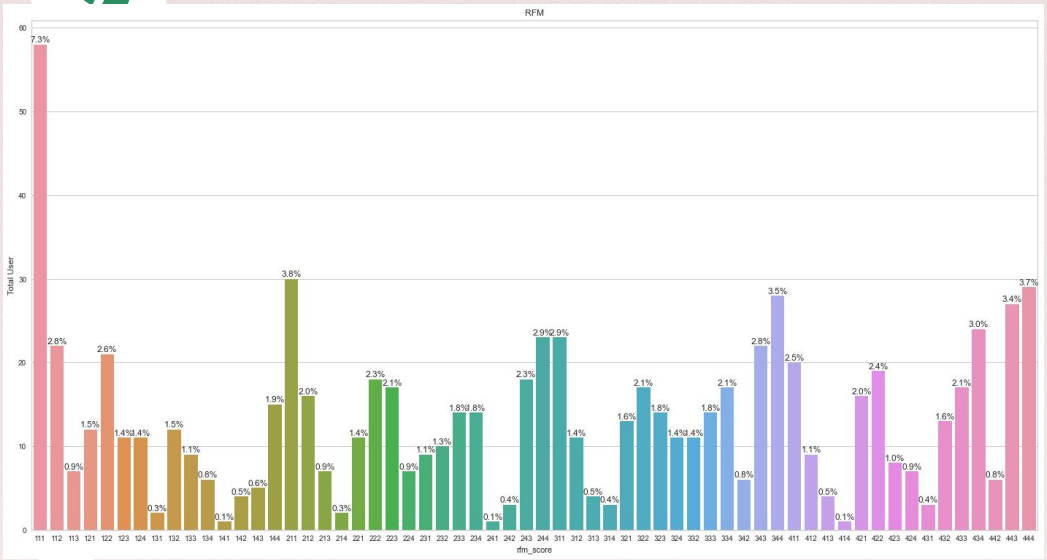
Exploratory Data Analysis



1. Super Store applies the discount promotion year to year
2. There are decline and rise discount promotion and rise again in 2017
3. From year to year, it can be seen, the most effective discount promotion are 20% and 70% - 80%
4. The customers who do not buy discount products are also high every year. It means the customer who interested in buying the products, they need them, also because of the discount promotion

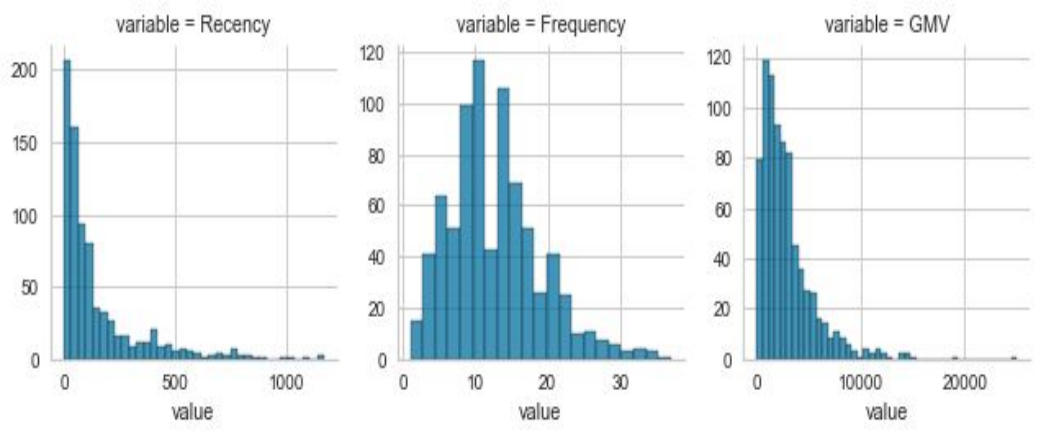
RFM Segmentation

RFM Segmentation



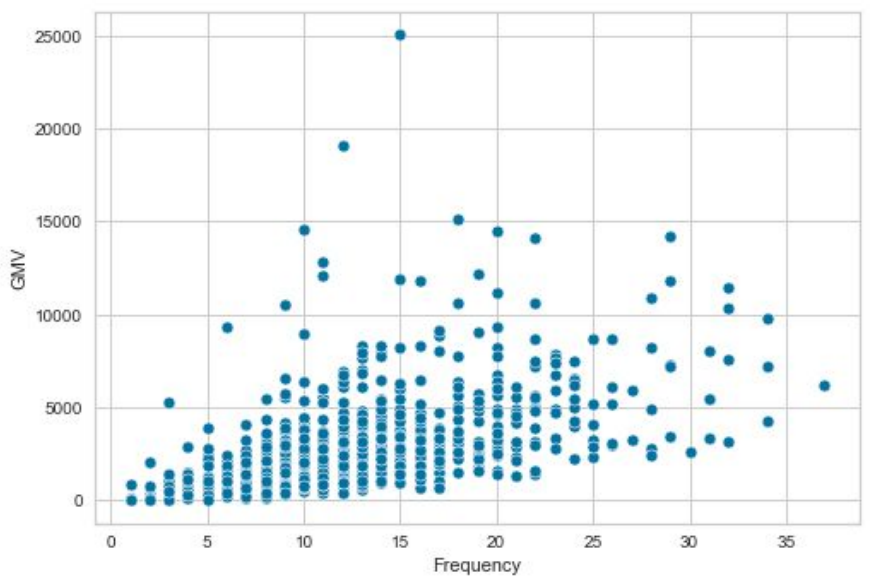
1. The highest percentage of customers is at 111 which indicates a high churn rate
2. The second highest percentage is also found in customers who need attention
3. From the following RFM Segmentation, the business conditions are still quite normal, because there are a lot of Champion, Loyal Customer, and Promising segmentation
4. Super Store needs to be preventive measures so in order to avoid churn customers and it is also seen there are so many customers At Risk segment

RFM Segmentation



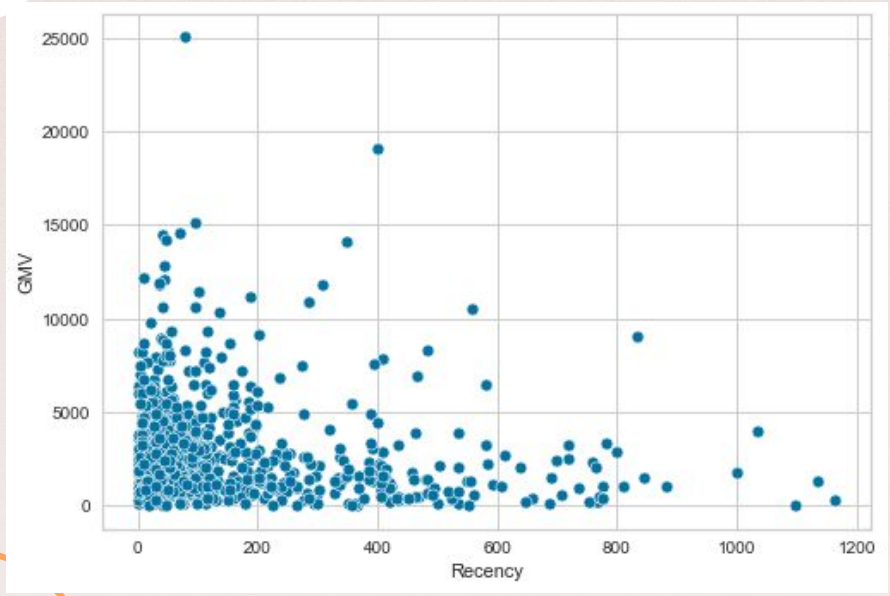
1. Based on the distribution data, the data skew to the right, but the frequency is quite okay
2. Overall, the data distribution still looks normal
3. The Recency value is also quite okay where the Recency value has an average distribution close to 0

RFM Segmentation



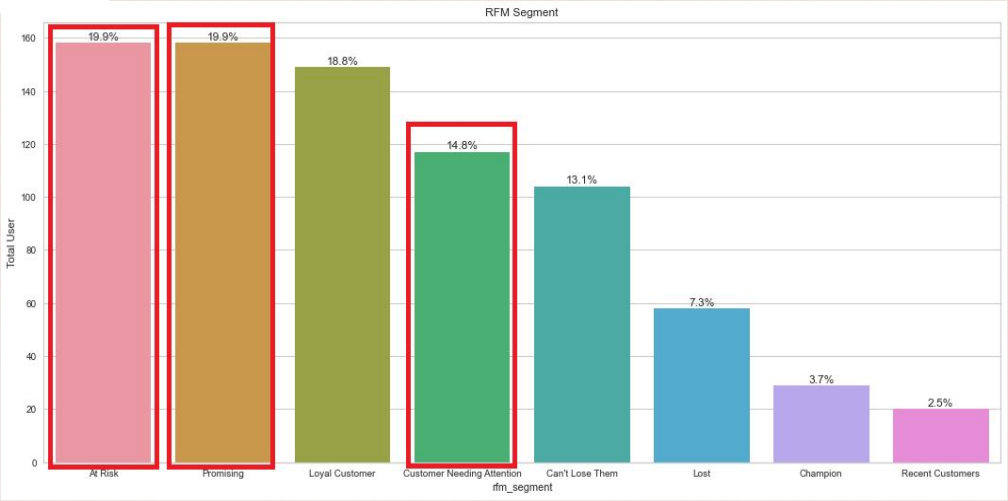
1. Based on the relationship between GMV and Frequency is quite good
2. However, it is necessary to maximize the increase of GMV because the frequency order is already high

RFM Segmentation



1. The relationship between Rencency and GMV is also quite good
2. Recency is low, and GMV is also high at Recency close to 0

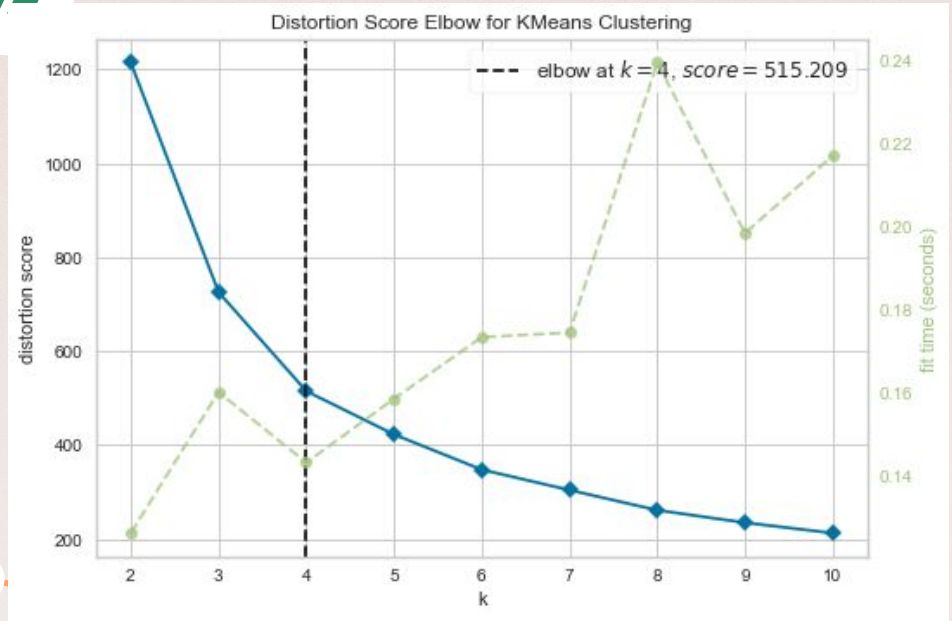
RFM Segmentation



1. The Super Store customer segmentation by RFM Segmentation is very good and the lost segmentation is so low
2. There are so many loyal and promising customers could be the target customers to increase the profit
3. Super Store needs to intervene At Risk segmentation, so that customers remain loyal and comfortable do the transactions at the Super Store. So they do not churn
4. Super Store needs to create Loyalty program to streamline promotion budget by targeting customers based on loyalty levels, RFM Segment can also be used as a reference for making Loyalty programs
5. Super store could focus do promotion on At Risk, Promising and customers need attention segmentation to increase profit and decrease the promotion budget

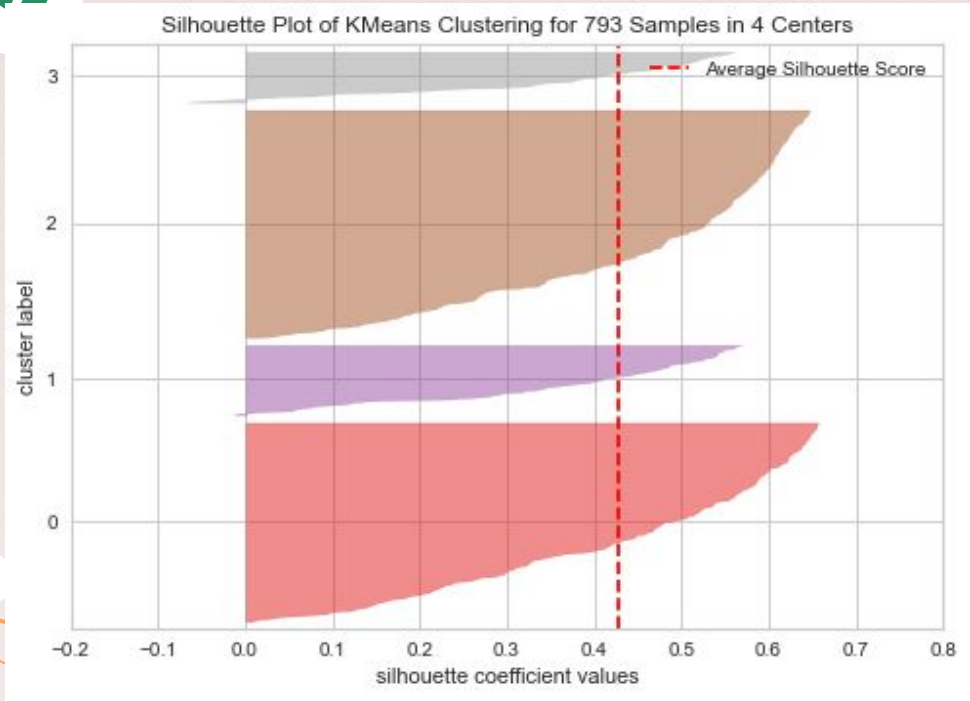
K-Means Cluster

K-Means Cluster



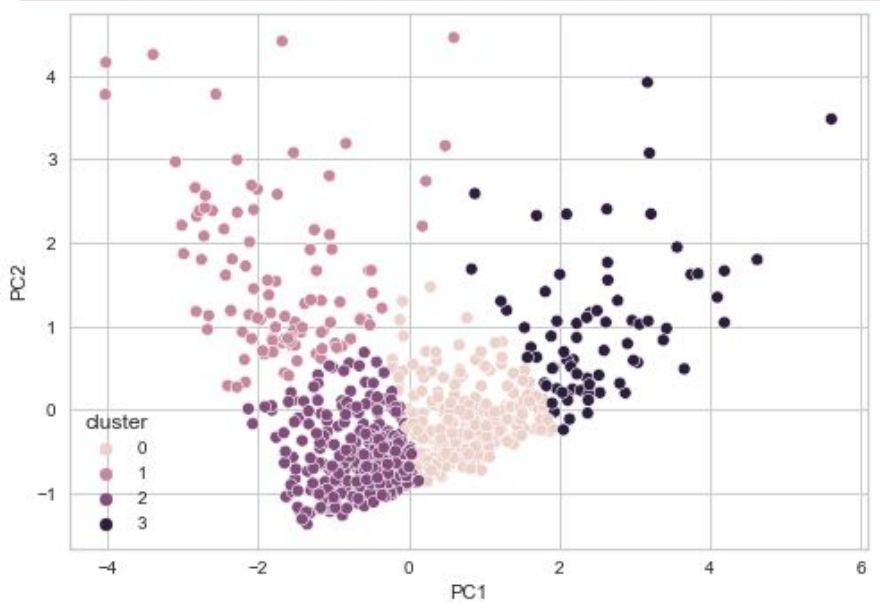
Based on the elbow method, the best K value is 4

K-Means Cluster



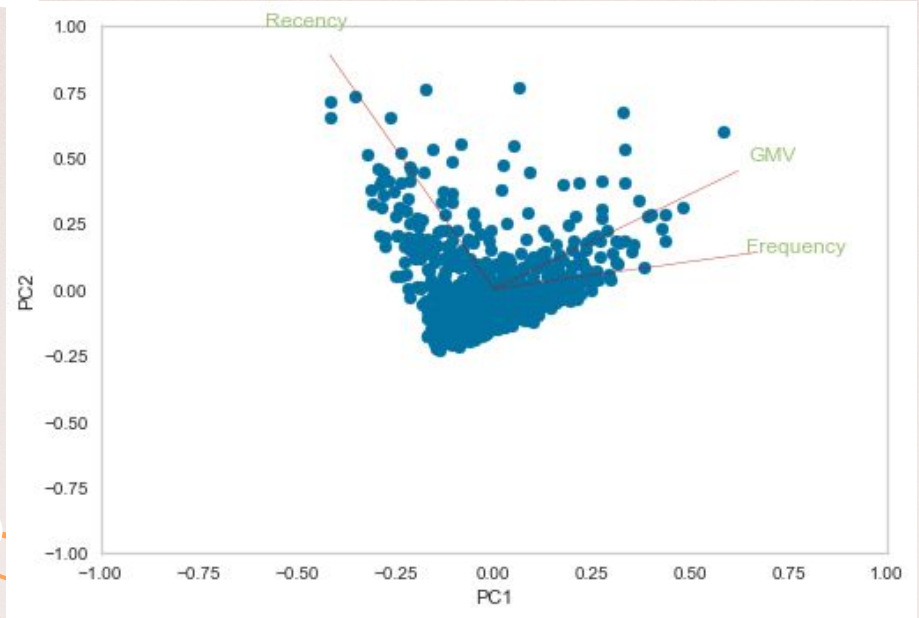
Based on Silhouette Visualization, the lowest error or minus score is K = 4

K-Means Cluster



The clustering looks very good on K = 4

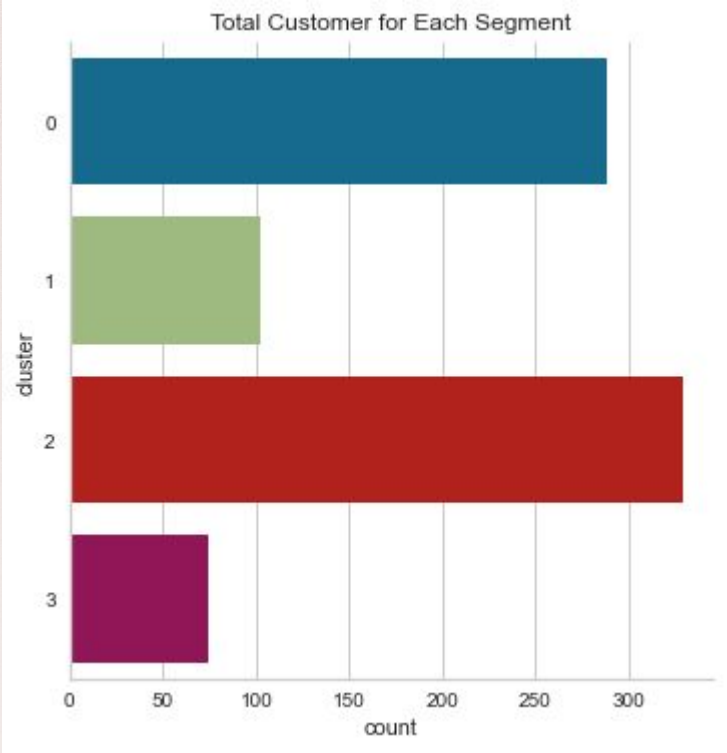
K-Means Cluster



1. Based on Biplot PC1 1 formed based on GMV and Frequency (Spending Behavior)
2. PC2 is formed based on Recency (How often customers do transaction "Recency")

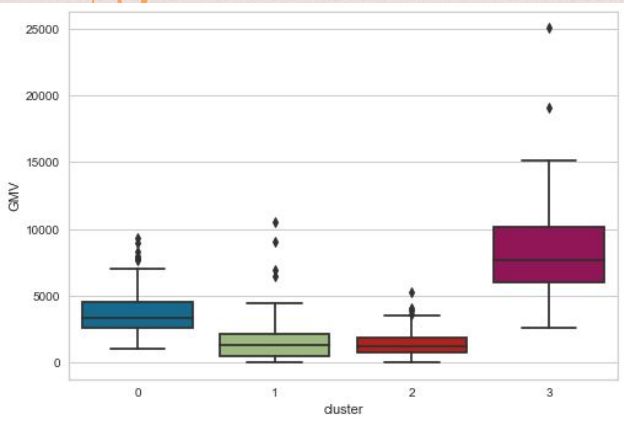
Evaluation

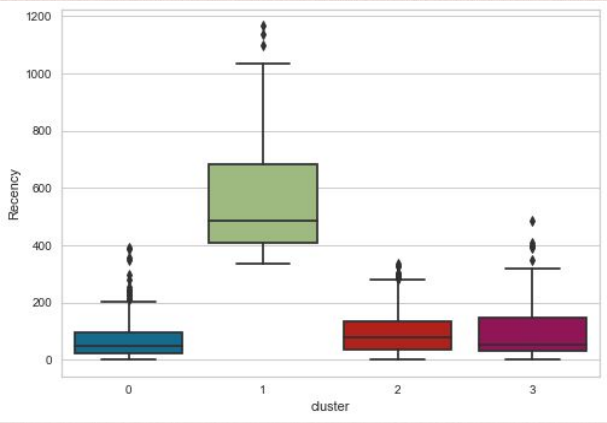
Evaluation

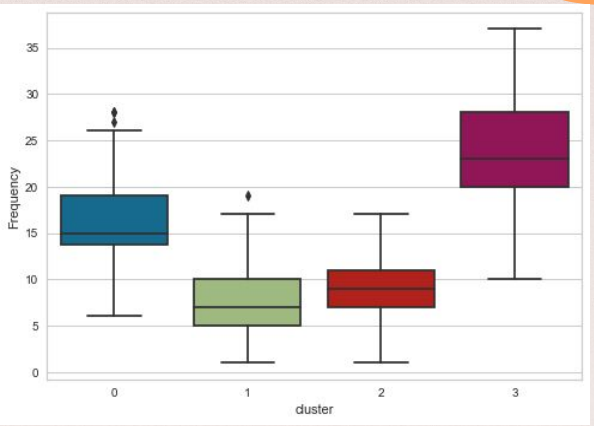


1. The highest clusters are in Clusters 0 and 2
2. The lowest cluster is 3

Evaluation







1. Based on boxplot, Cluster 3 is the superior customers
2. The conclusion of the K-Means Clusters are :
    1. Cluster 3 is platinum customers
    2. Cluster 0 is gold customers
    3. Cluster 2 is silver customers
    4. Cluster 1 is bronze customers

Evaluation

	Recency					Frequency					GMV				
	min	max	sum	mean	count	min	max	sum	mean	count	min	max	sum	mean	count
cluster															
0	0	390	20027	69.538194	288	6	28	4624	16.055556	288	1050.636	9351.212	1.054011e+06	3659.759702	288
1	336	1165	56155	550.539216	102	1	19	778	7.627451	102	4.833	10501.653	1.690439e+05	1657.292654	102
2	0	334	32176	97.799392	329	1	17	2870	8.723404	329	16.739	5253.876	4.550858e+05	1383.239658	329
3	2	484	8056	108.864865	74	10	37	1722	23.270270	74	2580.704	25043.050	6.190604e+05	8365.680646	74

1. Cluster 0 = Loyal (gold)
2. Cluster 1 = Churn (bronze)
3. Cluster 2 = Need Attention (silver)
4. Cluster 3 = Champion (Platinum)

		Recency					Frequency					GMV				
		min	max	sum	mean	count	min	max	sum	mean	count	min	max	sum	mean	count
cluster	rfm_segment															
0	At Risk	90	385	35697	256.812950	139	8	22	2125	15.287770	139	2.286	3930.072	36913.0720	265.561669	139
	Can't Lose Them	76	390	156540	143.878676	1088	13	26	18936	17.404412	1088	1.240	4548.810	245655.2322	225.786059	1088
	Champion	1	30	4684	18.015385	260	18	23	5240	20.153846	260	1.788	1983.968	65711.6640	252.737169	260
	Customer Needing Attention	56	180	28081	108.420849	259	6	21	3621	13.980695	259	1.680	7999.980	71000.5166	274.133269	259
	Loyal Customer	0	75	70911	33.120504	2141	13	28	37559	17.542737	2141	0.556	5399.910	466038.9226	217.673481	2141
	Promising	4	65	23158	31.421981	737	8	24	10991	14.913161	737	0.984	6354.950	168691.3868	228.889263	737
1	At Risk	336	1135	249092	500.184739	498	2	17	4992	10.024096	498	1.448	9099.930	121521.1717	244.018417	498
	Can't Lose Them	391	1035	67433	725.086022	93	13	19	1467	15.774194	93	0.898	4228.704	29115.7660	313.072753	93
	Lost	352	1165	104405	558.315508	187	1	8	1107	5.919786	187	2.025	1004.976	18406.9130	98.432690	187
2	At Risk	76	334	132290	185.020979	715	2	17	6489	9.075524	715	1.112	3347.370	122998.5158	172.025896	715
	Can't Lose Them	210	210	2730	210.000000	13	13	13	169	13.000000	13	4.554	503.960	2367.2770	182.098231	13
	Customer Needing Attention	32	183	87275	97.405134	896	1	16	9300	10.379464	896	0.852	2715.930	147558.0923	164.685371	896
	Lost	189	300	26054	248.133333	105	2	8	649	6.180952	105	1.632	698.352	10504.2510	100.040486	105
	Promising	0	70	30552	29.575992	1033	5	15	10809	10.463698	1033	0.836	2678.940	162703.8077	157.506106	1033
	Recent Customers	0	30	1914	17.722222	108	3	8	628	5.814815	108	1.408	629.958	8953.9038	82.906517	108
3	At Risk	399	399	4788	399.000000	12	12	12	144	12.000000	12	7.312	17499.950	19052.2180	1587.684833	12
	Can't Lose Them	79	484	146150	194.866667	750	14	34	18470	24.626667	750	1.234	22638.480	276458.6720	368.611563	750
	Champion	2	28	6284	15.670823	401	19	37	10447	26.052369	401	1.044	8749.950	113054.0778	281.930369	401
	Loyal Customer	6	64	22415	42.533207	527	15	34	13331	25.296015	527	0.444	9892.740	170897.4100	324.283510	527
	Promising	43	69	1636	51.125000	32	10	11	342	10.687500	32	1.248	11199.968	39597.9900	1237.437187	32

1. Cluster 0 = Loyal (gold)
2. Cluster 1 = Churn (bronze)
3. Cluster 2 = Need Attention (silver)
4. Cluster 3 = Champion (Platinum)
5. The results between RFM Segmentation and K-Means seems like similar but the total of cluster is different

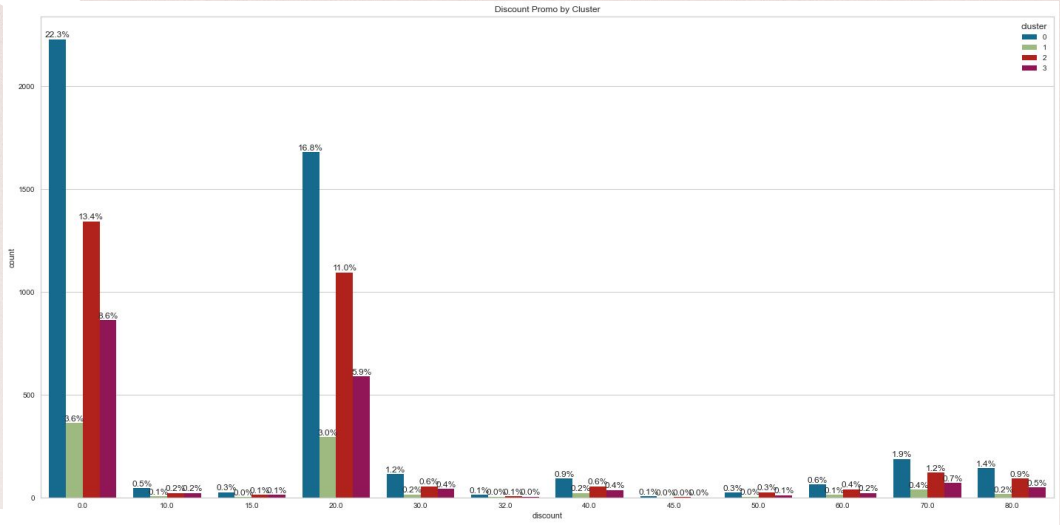
Evaluation

discount	cluster	Total_Customer	Mean_Recency	Mean_Frequency	Total_GMV	Total_Quantity
0.0	0.0	2229.0	70.41318977119785	16.812023328847015	488452.21	8635.0
0.0	1.0	363.0	524.9118457300275	9.776859504132231	74303.24	1373.0
0.0	2.0	1343.0	99.75279225614297	9.798212956068504	210979.54	4971.0
0.0	3.0	863.0	110.8053302433372	24.764774044032446	314173.48	3288.0
10.0	0.0	45.0	78.06666666666666	16.288888888888888	29229.714	173.0
10.0	1.0	5.0	517.6	11.2	1949.571	22.0
10.0	2.0	22.0	106.4090909090909	9.227272727272727	9682.209	98.0
10.0	3.0	22.0	112.72727272727273	26.181818181818183	13507.857	80.0
15.0	0.0	26.0	70.1923076923077	17.423076923076923	12508.914499999999	91.0
15.0	1.0	2.0	620.0	11.5	770.831	7.0
15.0	2.0	12.0	74.91666666666667	9.583333333333334	5556.2119999999995	38.0
15.0	3.0	12.0	66.08333333333333	23.0	8722.564	62.0
20.0	0.0	1679.0	67.71828469326981	17.029184038117926	348645.888	6162.0
20.0	1.0	295.0	555.9762711864407	9.616949152542373	71107.496	1163.0
20.0	2.0	1095.0	96.87214611872146	9.806392694063927	167541.336	4040.0
20.0	3.0	588.0	102.0323129251700 7	25.006802721088434	177299.64800000002	2295.0
30.0	0.0	115.0	65.86086956521739	17.565217391304348	55932.667	431.0
30.0	1.0	15.0	475.46666666666664	10.333333333333334	4850.349	50.0
30.0	2.0	55.0	77.65454545454546	9.290909090909091	18400.214	196.0
30.0	3.0	42.0	84.97619047619048	24.80952380952381	24043.425	172.0
32.0	0.0	13.0	79.38461538461539	15.153846153846153	6352.6212	46.0
32.0	1.0	4.0	711.5	12.75	2031.5952	18.0
32.0	2.0	6.0	98.5	8.166666666666666	2558.5816	19.0
32.0	3.0	4.0	51.75	23.25	3550.6608	22.0
40.0	0.0	93.0	62.763440860215056	17.440860215053764	51682.968	370.0
40.0	1.0	22.0	556.9090909090909	10.272727272727273	6286.398	65.0
40.0	2.0	55.0	87.43636363636364	9.836363636363636	21040.98	195.0
40.0	3.0	36.0	90.58333333333333	24.5	37407.438	156.0
45.0	0.0	6.0	22.666666666666668	16.666666666666668	2718.0285	22.0
45.0	1.0	1.0	424.0	9.0	562.2925	7.0
45.0	2.0	3.0	71.0	10.333333333333334	1271.391	12.0
45.0	3.0	1.0	189.0	20.0	933.262	4.0
50.0	0.0	25.0	58.04	16.2	21645.425	96.0
50.0	1.0	4.0	644.75	8.5	1474.495	17.0
50.0	2.0	26.0	106.3076923076923	9.846153846153847	7779.89	89.0
50.0	3.0	11.0	85.27272727272727	23.909090909090091	28018.73	39.0
60.0	0.0	63.0	68.53968253968254	16.936507936507937	3216.872	223.0
60.0	1.0	13.0	440.61538461538464	8.0	869.948	55.0
60.0	2.0	40.0	85.95	9.325	1440.748	140.0
60.0	3.0	22.0	77.95454545454545	25.818181818181817	1117.132	83.0
70.0	0.0	188.0	72.75531914893617	17.680851063829788	22706.61	729.0
70.0	1.0	38.0	558.1578947368421	8.921052631578947	3797.907	163.0
70.0	2.0	121.0	95.52892561983471	9.950413223140496	6894.366	492.0
70.0	3.0	71.0	106.46478873239437	24.12676056338028	7221.399	276.0
80.0	0.0	142.0	63.78169014084507	17.5	10918.876	603.0
80.0	1.0	16.0	648.4375	11.4375	1039.728	67.0
80.0	2.0	92.0	107.46739130434783	9.391304347826088	1940.38	341.0
80.0	3.0	50.0	98.84	24.5	3064.772	177.0

Findings :

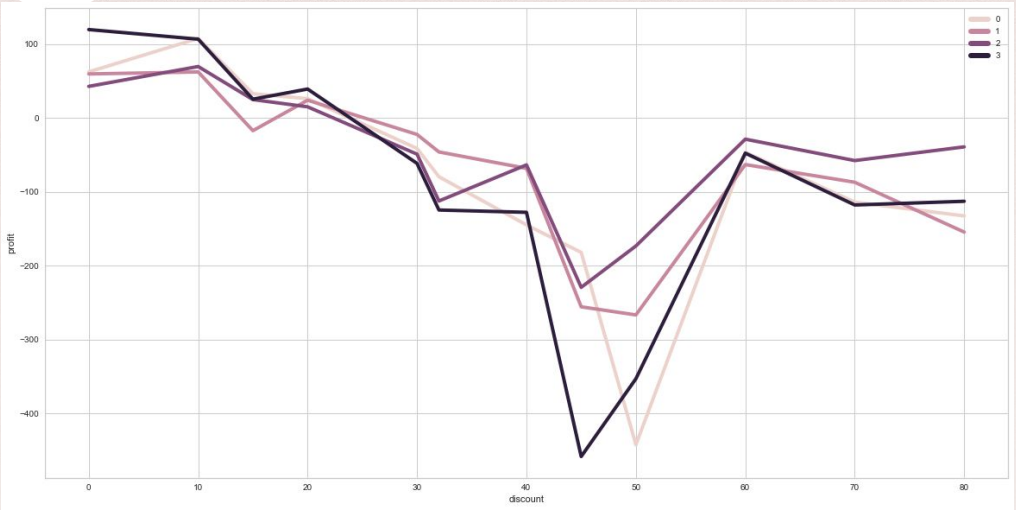
1. All segmentation from RFM and K-Means, have the same discount promotion
2. Therefore, it is necessary to create more efficient discount promotion based on the target of segmentation customers, to increase profit and reduces promotion budget
3. By creating the segmentation customers, Super Store could give promotion based on the target of segmentation customers
4. Super Store could create voucher discount that send to email, WhatsApp, etc before they go shopping, in order to attract customers to shop
5. Provide discounts price on product displays, could be adjusted to the nominal product discounts that will be given based on customer segments (Ex : The discount Price is 5% at the display products, but we give them 5% after they pay the bill, so the total discount is 10%)
6. Super Store could create big discount promotion by periodic with the event that attract customers to buy the products

Evaluation



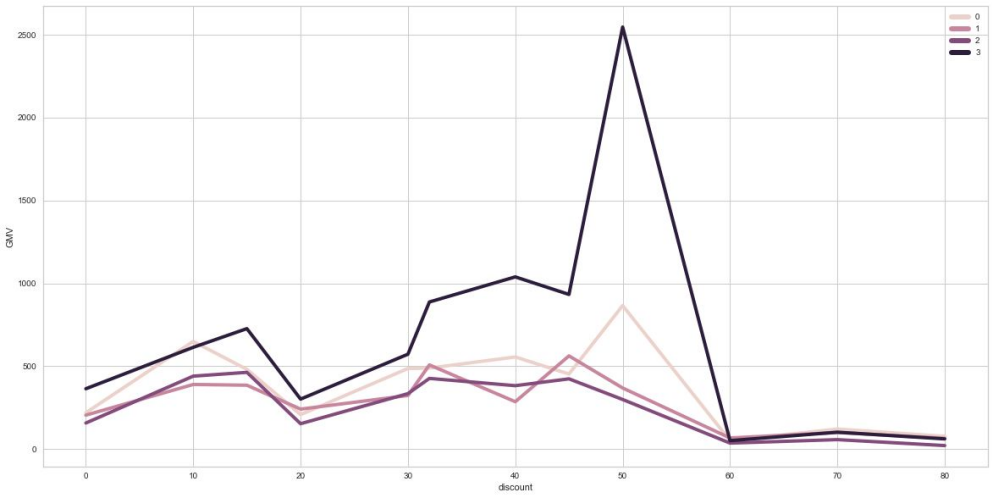
1. The Discount Promotion 20% is so high and transaction without discount promotion is really high too
2. This indicates that customers do the transactions without discount promotion, because they really like shopping at Super Store and need their products. It is not because of high discount promotion.
3. The most popular discount for customers is 20%

Evaluation



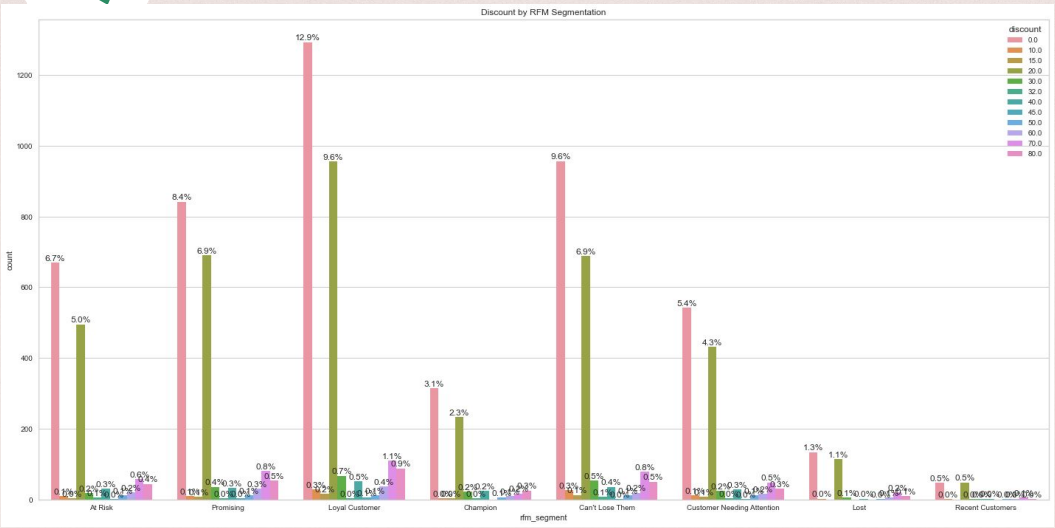
1. Based on the line plot, the discounts promotion that provide profit are 10 - 20%
2. The discounts promotion that provide very high loss are 40-50%
3. The high loss is caused by clusters 0 and 3 which are superior clusters
4. Supposedly, the promotion used by clusters 1 and 2
5. The discount promotion is still not efficient so that the profit generated is still less than maximum

Evaluation



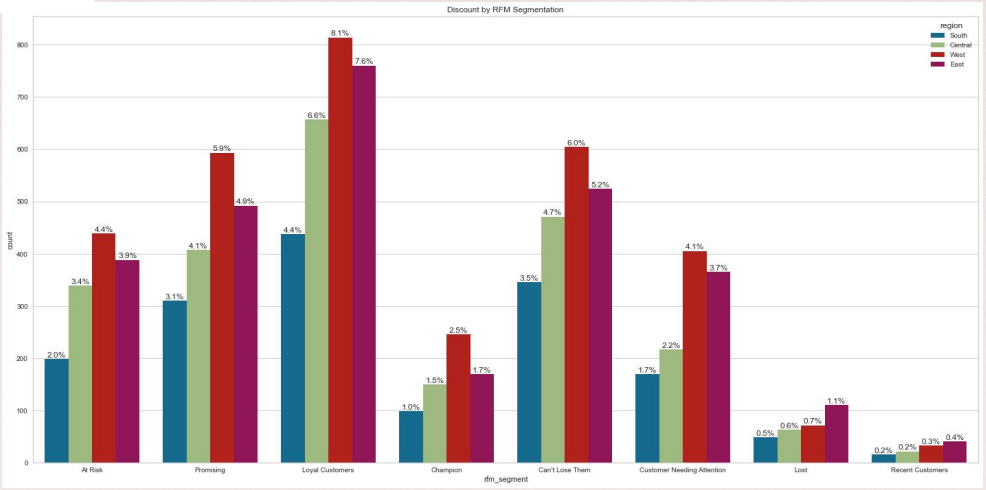
1. Discounts promotion provided an increase GMV
2. The discount 50 % is increase the GMV which is very high, it can be assumed, that the product purchased is from the technology category because the price is expensive
3. The discount 20 % does not provide a significant increase of GMV, but it still provided the profit

Evaluation



1. Based on RFM Segmentation, the customers who do transaction with discount 20% is so high and the customers who do transaction without discount promotion is so high too
2. We assume that except discount 20%, the discount promotion given frequently, but when we combine the nominal value is still high. So there is a high loss profit if the customers do transaction on discounts product
3. It is necessary to arrange re-discount according to the segment or loyalty, to increase profit and decrease promotion budget

Evaluation



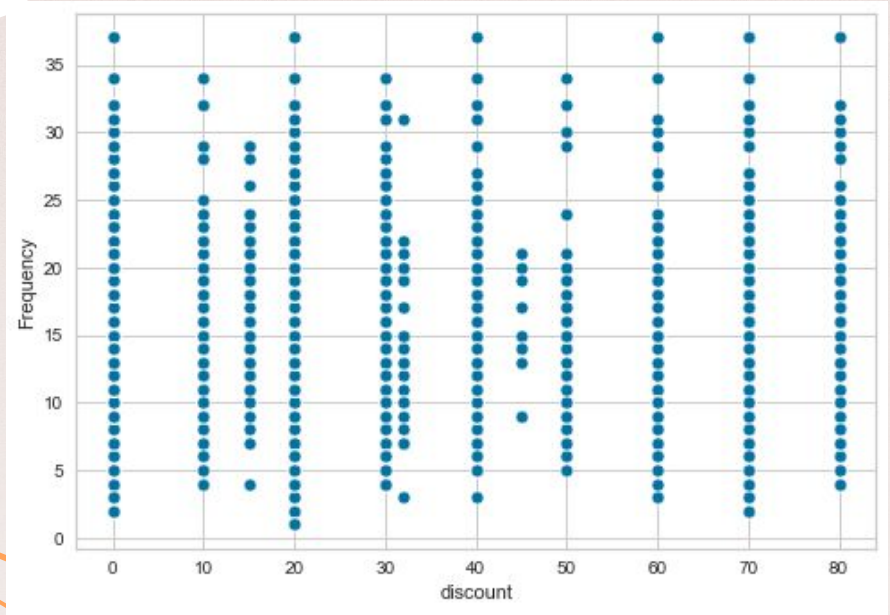
1. It can be seen that the regional center do promotion not to increase new customers
2. There is no significant growth in recent customers

Evaluation

discount	rfm_segment	region	Total_Customer	Total_profit
0.0	At Risk	Central	138	17010.7365
0.0	Can't Lose Them	Central	186	18270.6887
0.0	Champion	Central	53	4903.4082
0.0	Customer Needing Attention	Central	59	3945.3407
0.0	Lost	Central	22	594.6952
0.0	Loyal Customers	Central	246	21094.5852
0.0	Promising	Central	120	10285.7858
0.0	Recent Customers	Central	4	20.200400000000002
10.0	At Risk	Central	2	71.729
10.0	Can't Lose Them	Central	8	1264.9332
10.0	Champion	Central	1	152.0883
10.0	Loyal Customers	Central	6	399.6738
10.0	Promising	Central	1	29.7408
20.0	At Risk	Central	112	1872.6905
20.0	Can't Lose Them	Central	157	3018.9859
20.0	Champion	Central	41	185.7986
20.0	Customer Needing Attention	Central	87	1024.2865000000002
20.0	Lost	Central	24	-5.656100000000004
20.0	Loyal Customers	Central	228	4389.573
20.0	Promising	Central	172	3424.5959
20.0	Recent Customers	Central	13	144.7568
30.0	At Risk	Central	14	-540.7872
30.0	Can't Lose Them	Central	37	-1546.4274
30.0	Champion	Central	15	-805.3349000000001
30.0	Customer Needing Attention	Central	12	-415.4447
30.0	Lost	Central	3	-218.20549999999997
30.0	Loyal Customers	Central	39	-1906.6483
30.0	Promising	Central	26	-1093.1358
30.0	Recent Customers	Central	1	-14.8704
32.0	At Risk	Central	5	-680.0124000000001
32.0	Can't Lose Them	Central	8	-588.4293
32.0	Champion	Central	3	-468.254
32.0	Customer Needing Attention	Central	3	-143.3612
32.0	Loyal Customers	Central	4	-308.07939999999996
32.0	Promising	Central	3	-191.4054
32.0	Recent Customers	Central	1	-11.596
40.0	At Risk	Central	1	-121.2705
40.0	Can't Lose Them	Central	3	-1401.7133999999999
40.0	Champion	Central	1	33.5895
40.0	Customer Needing Attention	Central	3	-299.0367
40.0	Loyal Customers	Central	5	-878.4123
50.0	At Risk	Central	3	-1471.7685000000001
50.0	Can't Lose Them	Central	3	-654.3669
50.0	Champion	Central	1	-300.735
50.0	Customer Needing Attention	Central	6	-917.9123999999999
50.0	Loyal Customers	Central	4	-833.5169000000001
50.0	Promising	Central	1	-131.445
60.0	At Risk	Central	20	-620.2594
60.0	Can't Lose Them	Central	19	-840.8027
60.0	Champion	Central	9	-284.1102
60.0	Customer Needing Attention	Central	16	-280.9892
60.0	Lost	Central	5	-463.2278
60.0	Loyal Customers	Central	37	-2447.0629
60.0	Promising	Central	31	-942.141
60.0	Recent Customers	Central	1	-66.062
80.0	At Risk	Central	44	-1601.5887
80.0	Can't Lose Them	Central	50	-5870.8694000000005
80.0	Champion	Central	25	-693.1464000000001
80.0	Customer Needing Attention	Central	30	-3368.0415000000003
80.0	Lost	Central	9	-689.2559
80.0	Loyal Customers	Central	88	-14277.7212
80.0	Promising	Central	53	-4024.8009
80.0	Recent Customers	Central	1	-13.6152

1. Based on the following table, it is clear that recent customers buy a little discount product
2. Loyal customers and can't lose them are dominant on discount promotion
3. Region Central applies discount promotion to retain existing customers

Evaluation



1. Discounts promotion does not provide a significant frequency transaction
2. This indicates, customers who do transaction, not because of discount promotion, but they need the product and feel comfortable for doing transactions at Super Store

Conclusion

1. The budget of promotion at Super Store is quite good in range 10 - 20% from profit, but it can be more efficient with providing promotion based on customer target to increase the profit and reduce the promotion budget
2. The difference between GMV and Profit values is 2010803.8386 which is very high. So it is necessary to evaluate the promotion and the operational costs
3. The Loyalty Program is needed in order to be able to provide promotion according to target customers and provide maximum profit
4. Central region is the region where is the weakness of the Super Store based on profit, because the promotion that has given is quite high, but the GMV generated is also high
5. The best discount value are 10-20 % based on profit and customers profil
6. Super Store needs to increase the promotion on bottom products to introduce the bottom products and increase orders for these products for increasing the profit

Recommendation

1. Super Store needs to create Loyalty program based on customer segmentation or Customer Lifetime Value. So the promotion will be given to the target customers
2. Super Store needs to re-evaluate the discount promotion value based on product prices to maximize the profit. It will make the customers feel happy as they are getting the best price
3. The discounts price on product display could be adjusted to the nominal product discounts that will be given based on customer segments (Ex : The discount Price is 5% at the display products, but we give them 5% after they pay the bill. So the total discount is 10%)
4. Super Store could create voucher discount that will be sent via email, WhatsApp, etc before they go shopping, in order to attract customers to shop by loyalty program
5. It is necessary to do evaluation on operational perspective to find out more about the weaknesses of the Super Store and the reason of high loss profit
6. Super Store needs to re-evaluate promotion and operational costs at the centre region to increase the profit, because it has been running for 4 years but the profit generated is not optimal
7. A. Diverting promotion budget from top product to bottom 9 products to introduce the bottom 9 products and increase sales of these products.
   B. Re-evaluating the bottom 9 products, whether Super Store should abolish or maintain the products.
8. Super Store could choose K-Means Cluster or RFM Segmentation to do segmentation of the customers as the results of K-Means Cluster and RFM Segmentation are similar. RFM Segmentation has more detailed groups because there are 8 segments by RFM Segmentation

Thank you