



# HDI HOLDINGS

**SEGMENTATIONPRODUCT  
RECOMMENDATION**



# Objective

- 1.Customer Segmentation**
- 2.Segment Movement Analysis**
- 3.Product Recommendation**

# Data understanding

transaction 2021 Jan-Jun

transaction 2022 Jan-Dec

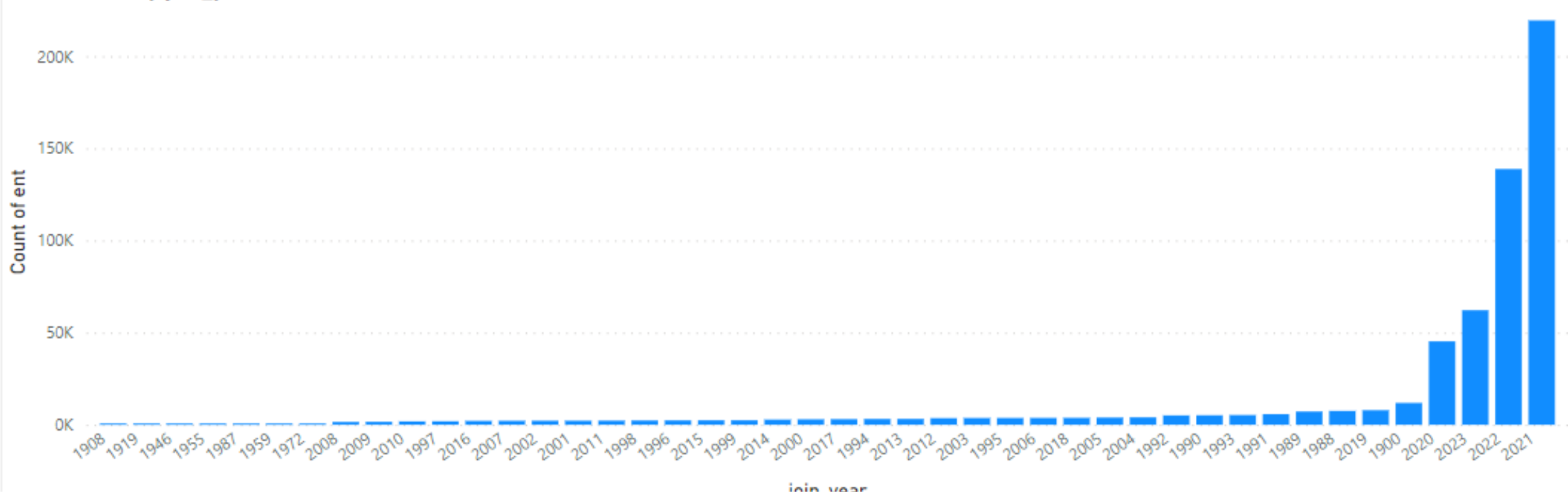
transaction 2023 Jan-July  
data member

transaction

#	Column	Non-Null Count	Dtype
0	payment_date	362109 non-null	object
1	ent	362109 non-null	object
2	center	362109 non-null	object
3	product_json	362100 non-null	object
4	total_amount	362109 non-null	float64
5	discount	361969 non-null	float64
6	paid_amount	362109 non-null	int64
7	trans_origin_type	362109 non-null	object

#	Column	Non-Null Count	Dtype
0	ent	590565 non-null	object
1	original_status	590565 non-null	object
2	join_month	590565 non-null	int64
3	join_year	590565 non-null	int64
4	sponsor	590565 non-null	object

Count of ent by join\_year



transaction

2023

	payment_date	ent	center	product_json	total_amount	discount	paid_amount	trans_origin_type
0	2023-01-01 00:00:00.000	TCC4QCZJZQI	TKC11Z	[{"product":"5C4CCE","qty":1},{"product":"6CQC...	590000.0	2500.0	0	online
1	2023-01-01 00:00:00.000	TCC4QCE34Z7	T2CEQ1	[{"product":"BC4C41","qty":2},{"product":"6CQC...	243000.0	2500.0	240500	online
2	2023-01-01 00:00:00.000	TCC44Q3JQQ2	TFCEZ1	[{"product":"2C1CCW","qty":1},{"product":"2C1C...	1792000.0	0.0	0	online
3	2023-01-01 00:00:00.000	TCC41JJZCRF	TKCC3J	[{"product":"2C1CCW","qty":2},{"product":"2C1C...	3025000.0	0.0	0	online
4	2023-01-01 00:00:00.000	TCC4JZCJ1JK	TKCRCW	[{"product":"4E1J","qty":2}]	198000.0	0.0	0	online

data

member

ent	original_status	join_month	join_year	sponsor
UCC4CC4F	-	5	2014	8844CQCEY
UCC4CCEK	-	5	2014	8844CQCCU
UCC4CC15	-	5	2014	8844CQCEY
UCC4CCR2	-	5	2014	8844CQCEY
UCC4CCJP	-	5	2014	8844CQCEY
UCC4CCQ0	-	5	2014	8CJWC1QW
UCC4CCWU	-	5	2014	8844CQCW5
UCC4CC3B	-	5	2014	8844CQCCU
UCC4CCZY	-	5	2014	UCC4CC15
UCC4C4C7	-	5	2014	UCC4CCJP
UCC4C44I	-	5	2014	UCC4CCQ0
UCC4C4ES	-	5	2014	UCC4CC15
UCC4C41F	-	5	2014	8844CQCCU
UCC4C4RK	-	5	2014	UCC4CC15

# Data Preparation

	payment_date	ent	center	product_json	total_amount	discount	paid_amount	trans_origin_type	qty	product_variety	month	bucket_id	SHOP_MONTH	REV_PER_UNIT
0	2021-01-01	TCC43RQQZ3B	TKC4CJ	[{"product": "1ECZ", "qty": 1}, {"product": "8CEC4Q...	1627000.0	0.0	0	online	2	2	1	2021_0	2021-01	8.135000e+05
1	2021-01-01	TCC4Q4JZ4Q0	TKCEWQ	[{"product": "4E4W", "qty": 2}, {"product": "DC4CC4...	594000.0	0.0	0	online	3	2	1	2021_1	2021-01	1.980000e+05
2	2021-01-01	TCC4RJ11QE7	TKC14E	[{"product": "1C4CC4", "qty": 1}]	468000.0	0.0	0	online	1	1	1	2021_2	2021-01	4.680000e+05
3	2021-01-01	TCC434QRCJY	TKC14R	[{"product": "5C4CC4", "qty": 1}]	341000.0	0.0	0	online	1	1	1	2021_3	2021-01	3.410000e+05
4	2021-01-01	TCC43RRCE10	TKC11Z	[{"product": "5C4C4Q", "qty": 1}]	189000.0	0.0	0	online	1	1	1	2021_4	2021-01	1.890000e+05
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Filter Dataset for segmentation

transaction 2021 Jan-Jun

transaction 2022 Jan-Dec

transaction 2023 Jan-July

—————→ H1\_2022 , H2\_2022

H1\_2021

H1\_2023

# Customer Single View

Feature	Description
ent	Custoemr id
LAST_VISIT	Last date visit
TOTAL_VISIT	Total visit
TOTAL_SPEND	Total spending
AVG_MONTHLY_VISIT	Average monthly visit
AVG_MONTHLY_SPEND	Average monthly spending
AVG_BASKET_SIZE	Average monthly basket size
REV_PER_UNIT	Average total amount/quantity

	ent	LAST_VISIT	TOTAL_VISIT	TOTAL_SPEND	AVG_MONTHLY_VISIT	AVG_MONTHLY_SPEND	AVG_BASKET_SIZE	REV_PER_UNIT	Clusters
0	TCC4RJ11QE7	2021-06-02	15	16723000	3.750000	4.180750e+06	1.114867e+06	357910.476190	0
1	TCC43C3JRJ5	2021-04-23	14	2811500	3.500000	7.028750e+05	2.008214e+05	200821.428571	1
2	TZEEQRRU	2021-06-10	17	13326000	2.833333	2.221000e+06	7.838824e+05	470274.509804	0
3	TCC4QEZQW4I	2021-05-08	28	9670000	5.600000	1.934000e+06	3.453571e+05	262857.142857	0
4	TCC41E3E142	2021-06-09	37	27613000	6.166667	4.602167e+06	7.462973e+05	360878.378378	0
...	...	...	...	...	...	...	...	...	...
325622	TCCECE4E1ZF	2021-06-20	13	6682500	4.333333	2.227500e+06	5.140385e+05	314064.102564	0
325623	TCCECE4C43I	2021-06-11	13	6357500	4.333333	2.119167e+06	4.890385e+05	346282.051282	0
325624	TCC4QR314RY	2021-06-21	13	9102500	6.500000	4.551250e+06	7.001923e+05	396493.589744	0
325625	TCCEC1CCRCK	2021-06-22	13	6560500	6.500000	3.280250e+06	5.046538e+05	240301.282051	0
325626	TCCEC14E1ZK	2021-06-11	13	3327000	6.500000	1.663500e+06	2.559231e+05	208692.307692	1

325627 rows × 9 columns



# Customer Segmentation using K-mean clustering

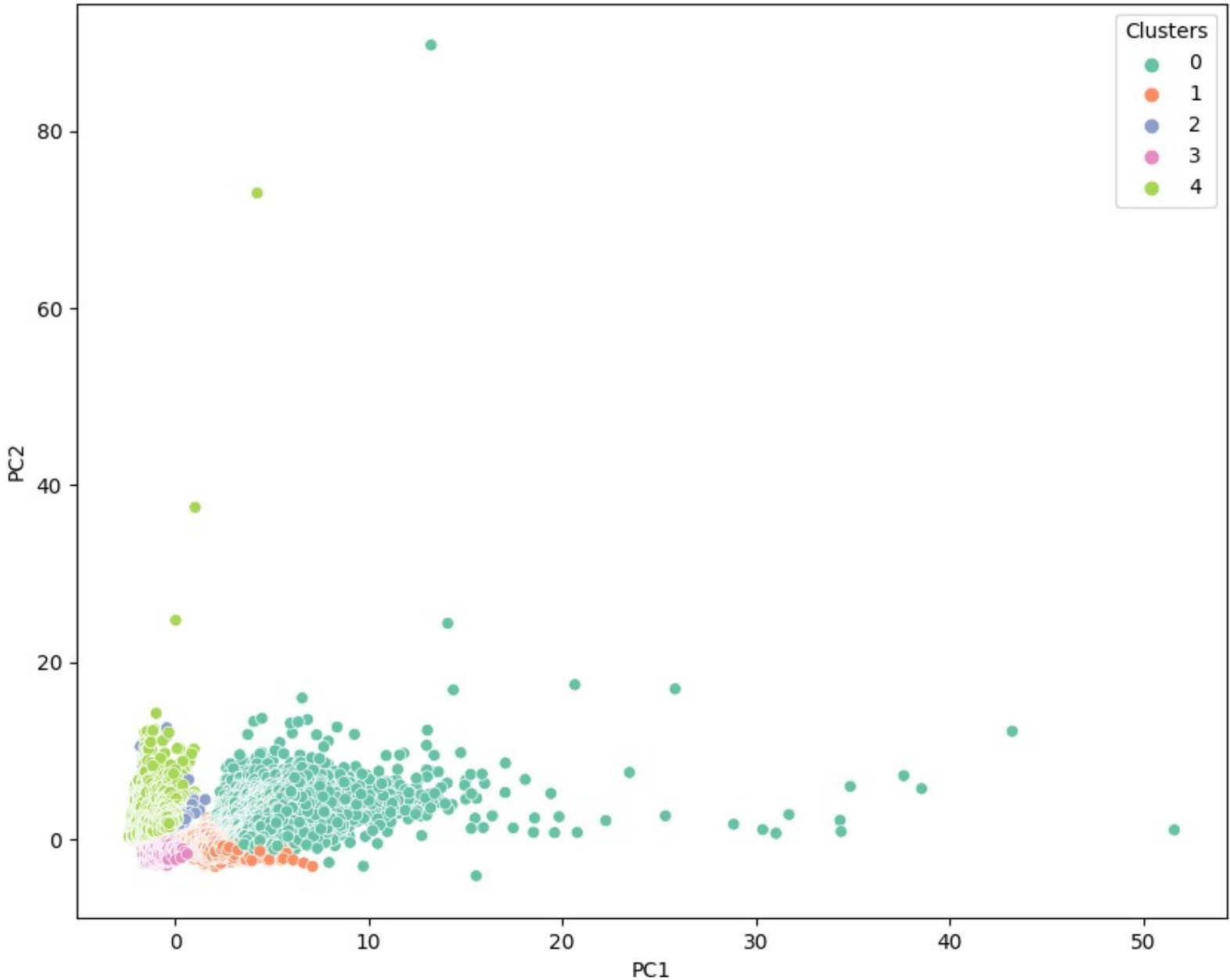
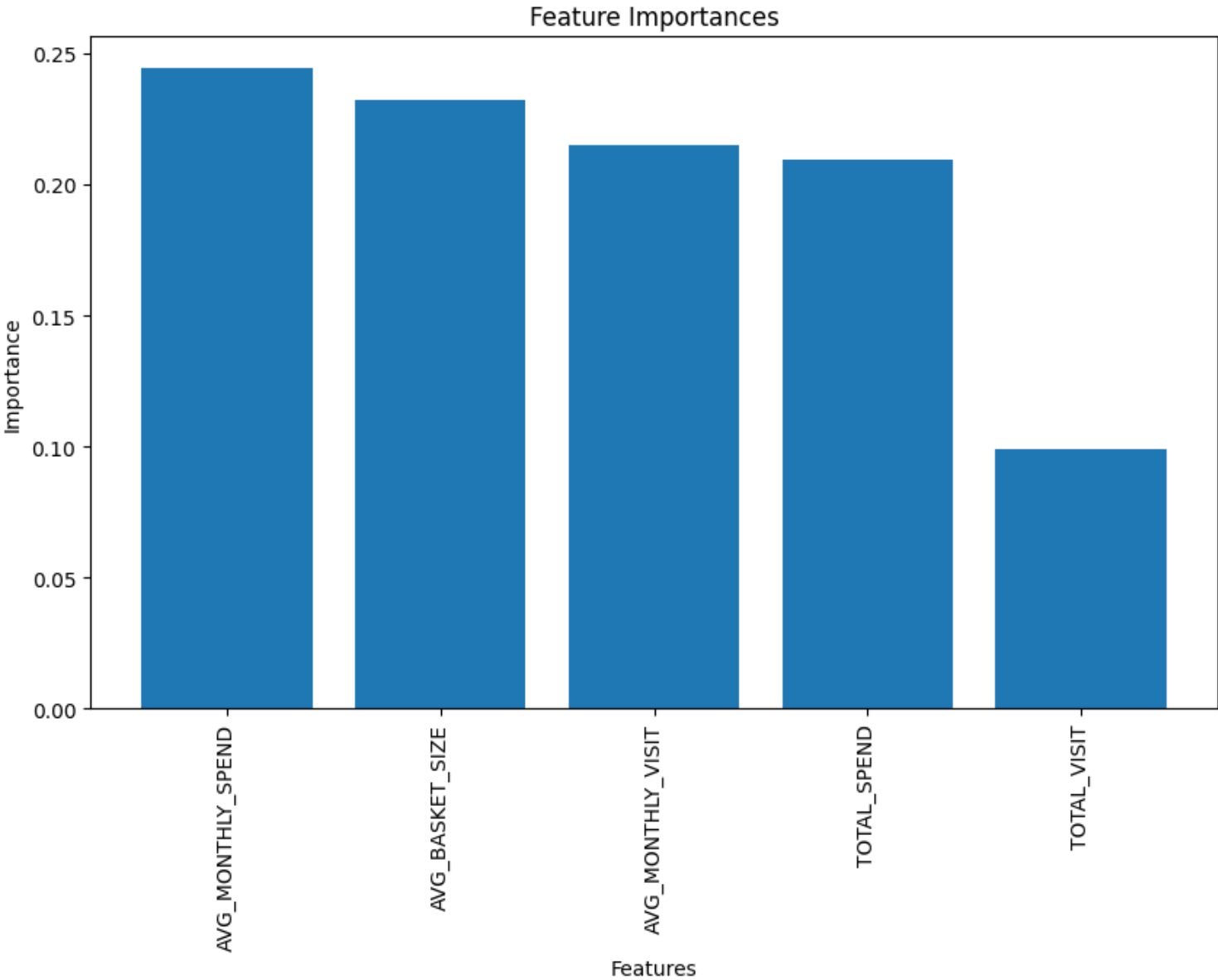
## PCA

	count	mean	std	min	25%	50%	75%	max
PC1	325627.0	-1.452390e-16	1.833751	-2.383425	-1.583696	0.443736	1.037140	51.612213
PC2	325627.0	1.117223e-17	1.490651	-4.147540	-0.951778	-0.292338	1.199381	89.693641

## Feature selection

TOTAL\_VISIT,AVG TOTAL\_SPEND,  
AVG\_MONTHLY\_VISIT ,  
AVG\_MONTHLY\_SPEND,  
AVG\_BASKET\_SIZE,  
AVG\_REV\_PER\_UNIT

## Feature importance



	# ent
1	146324
4	68721
3	66801
2	23315
0	20466

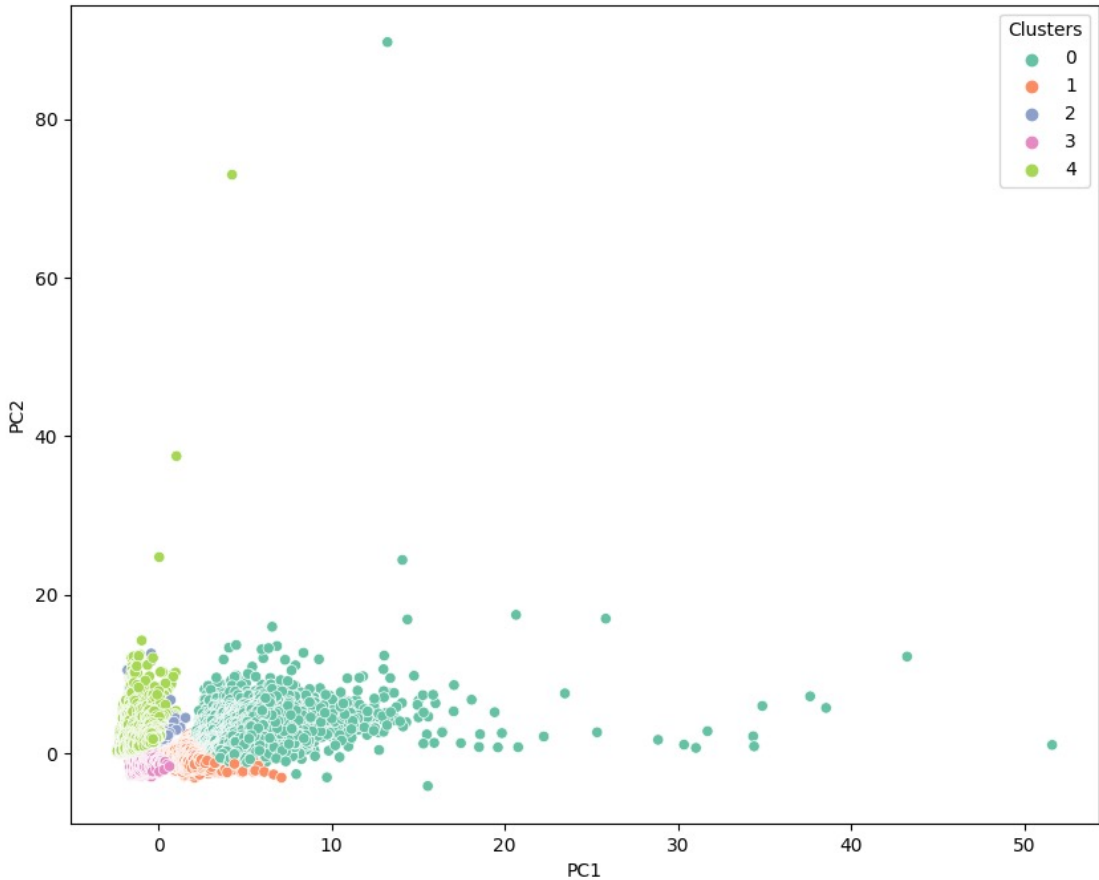
K=5

random\_state=66

6

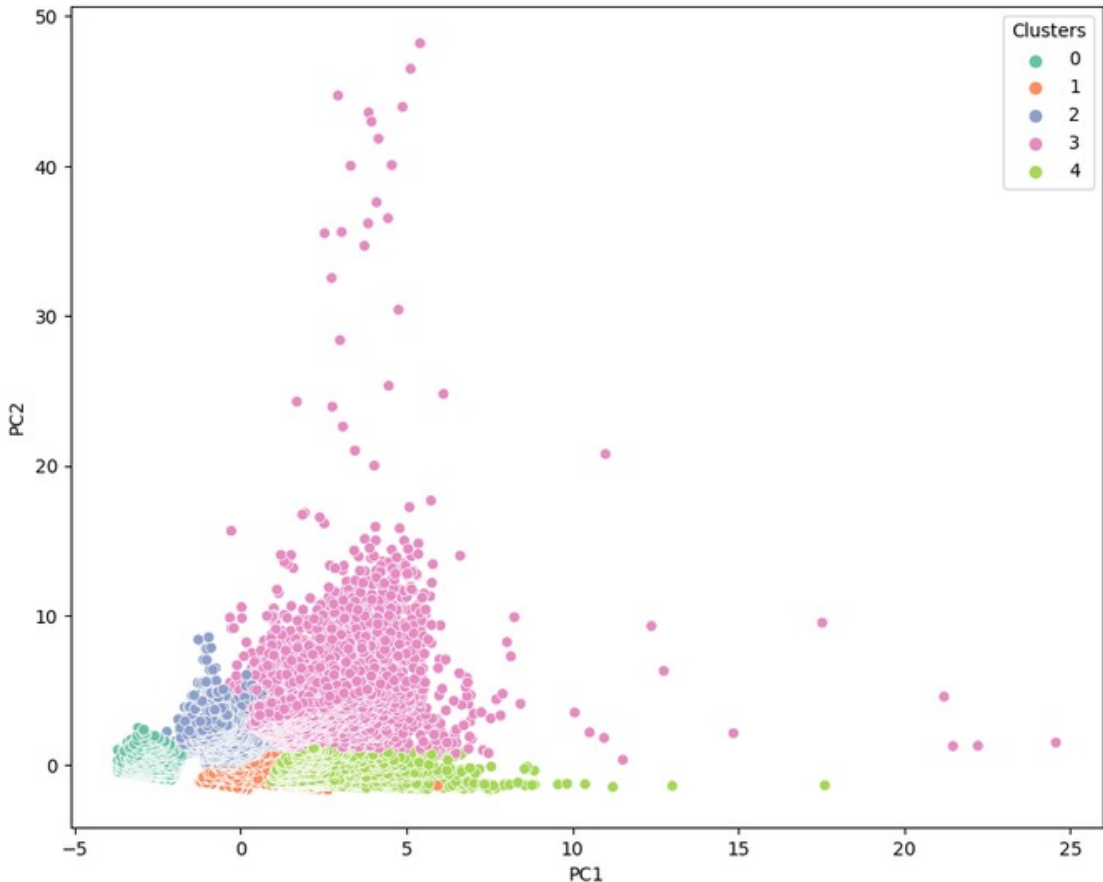
Comparative Analysis of customer segments Over a Six-Month Period

H1\_2021



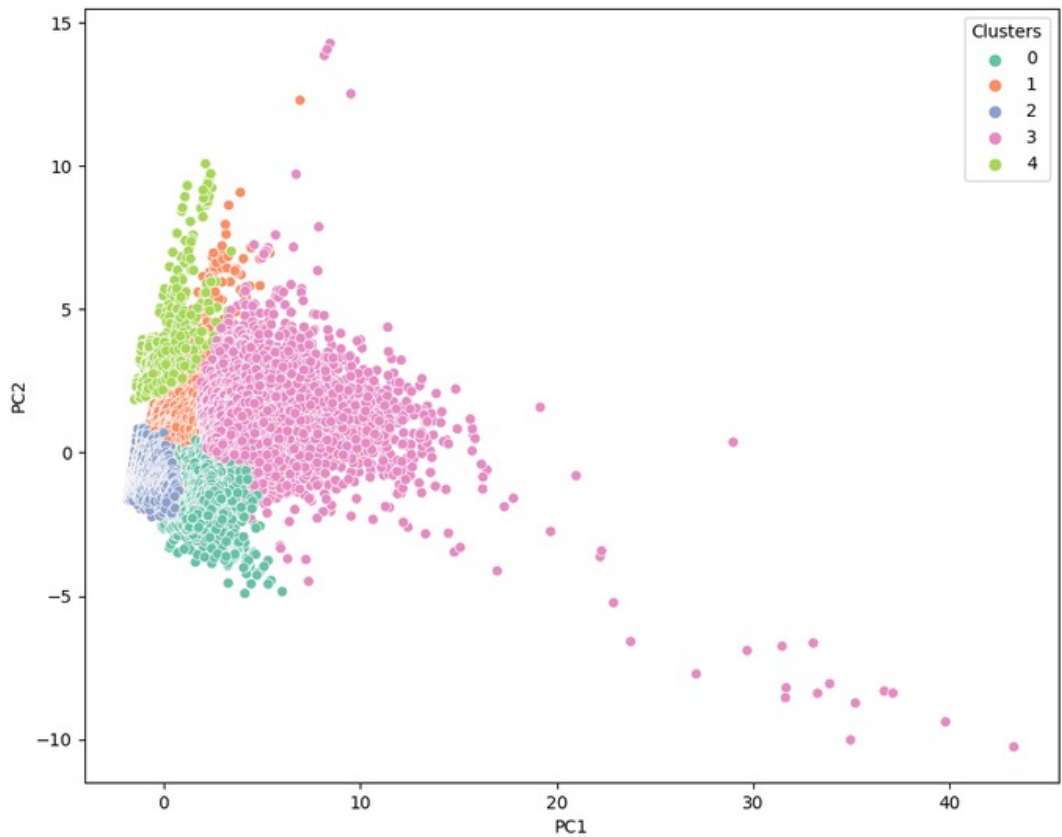
	# ent
1	146324
4	68721
3	66801
2	23315
0	20466

H1\_2022



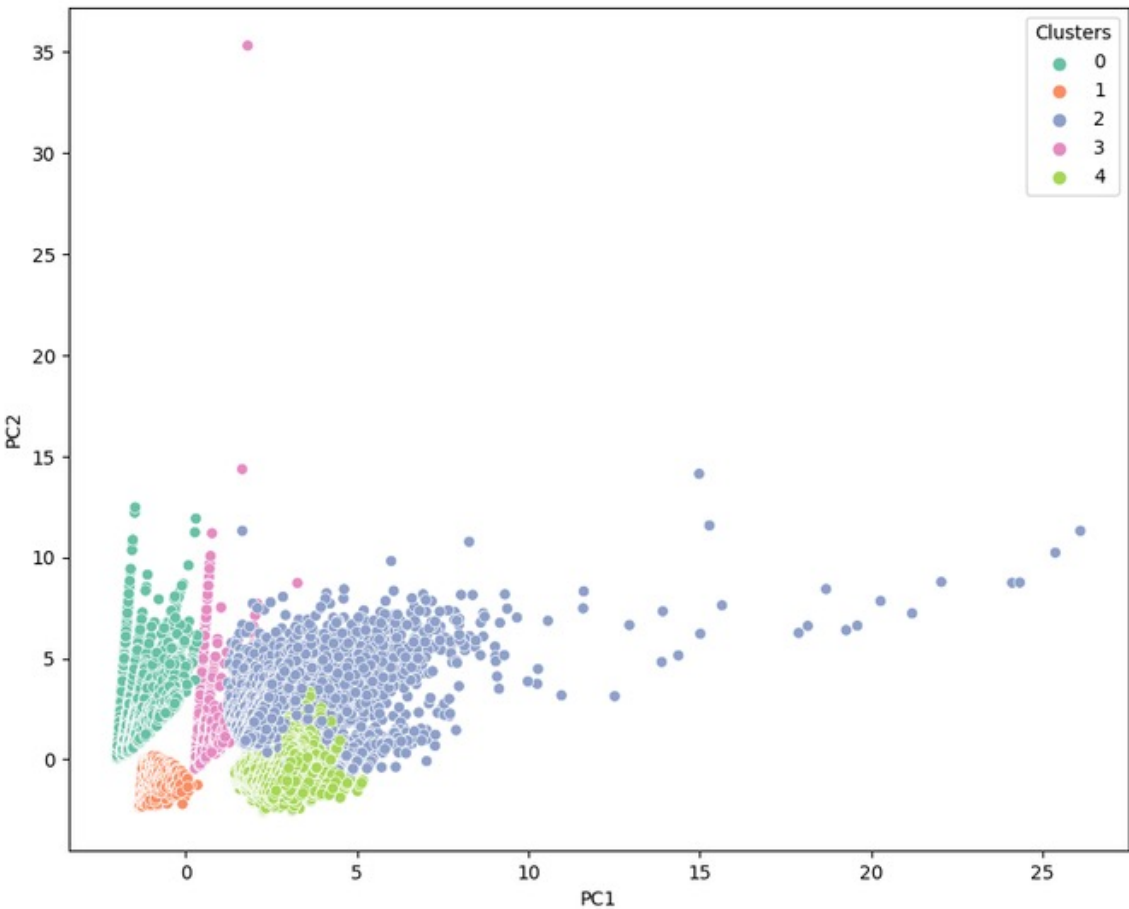
	# ent
4	64958
2	55694
0	40695
1	11768
3	11343

H2\_2022



	# ent
0	45965
1	41961
2	36679
4	10921
3	7489

H1\_2023



	# ent
1	32162
4	29661
0	27428
3	5699
2	5319

random\_state=66



# BEHAVIORAL PATTERNS WITHIN CUSTOMER SEGMENTS

Clusters	Average of TOTAL_SPEND	Average of REV_PER_UNIT	AVG_MONTHLY_SPEND	AVG_MONTHLY_VISIT	Average of TOTAL_VISIT	Count of ent
0	11990K	480K	2,430K	2.79	14.01	20K
1	2257K	395K	1,647K	2.44	3.38	146K
2	1780K	1,539K	1,599K	1.01	1.14	23K
3	783K	285K	464K	1.21	1.84	67K
4	2676K	633K	1,813K	1.11	1.72	69K

- 0 Power Shoppers

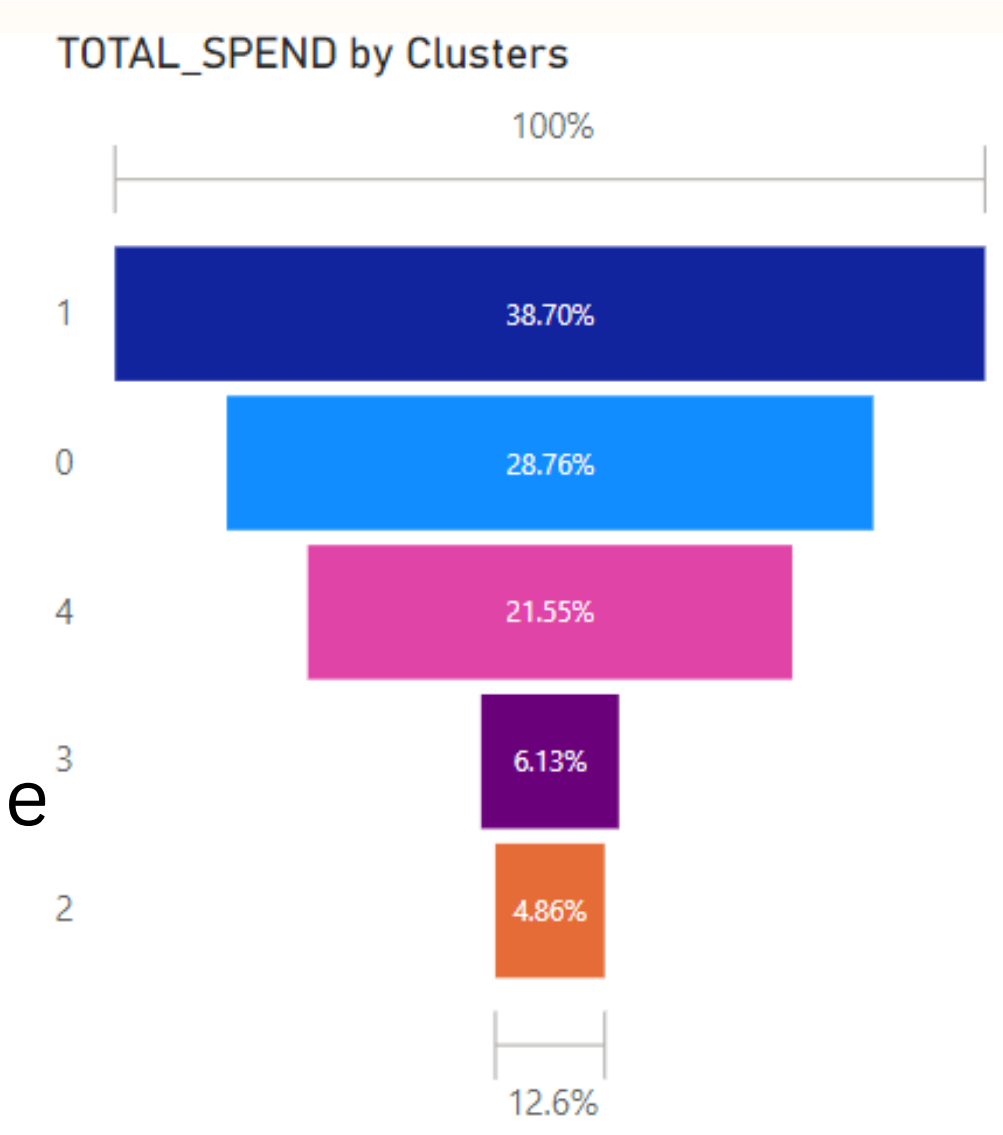
frequently purchase a high volume of your products
- 1 Value Maximizers

moderate overall spend but a higher revenue per unit
- 2 High-Yield Patrons

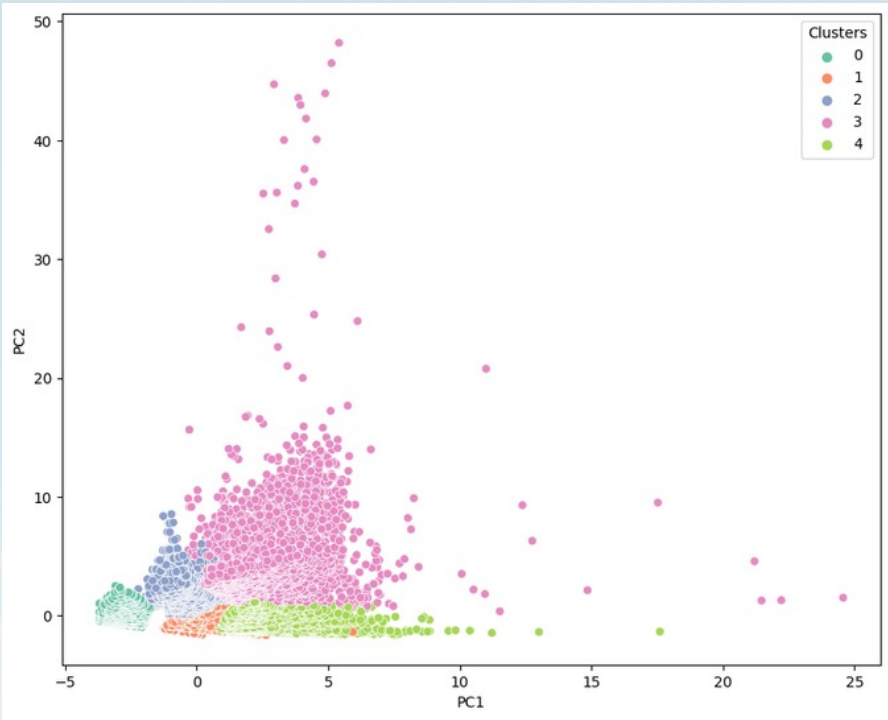
lower overall spend but the highest revenue per unit
- 3 Frugal Shoppers

spend the least overall and have the lowest revenue per unit
- 4 Luxury Enthusiasts

These customers spend a considerable amount, focusing on high-revenue item

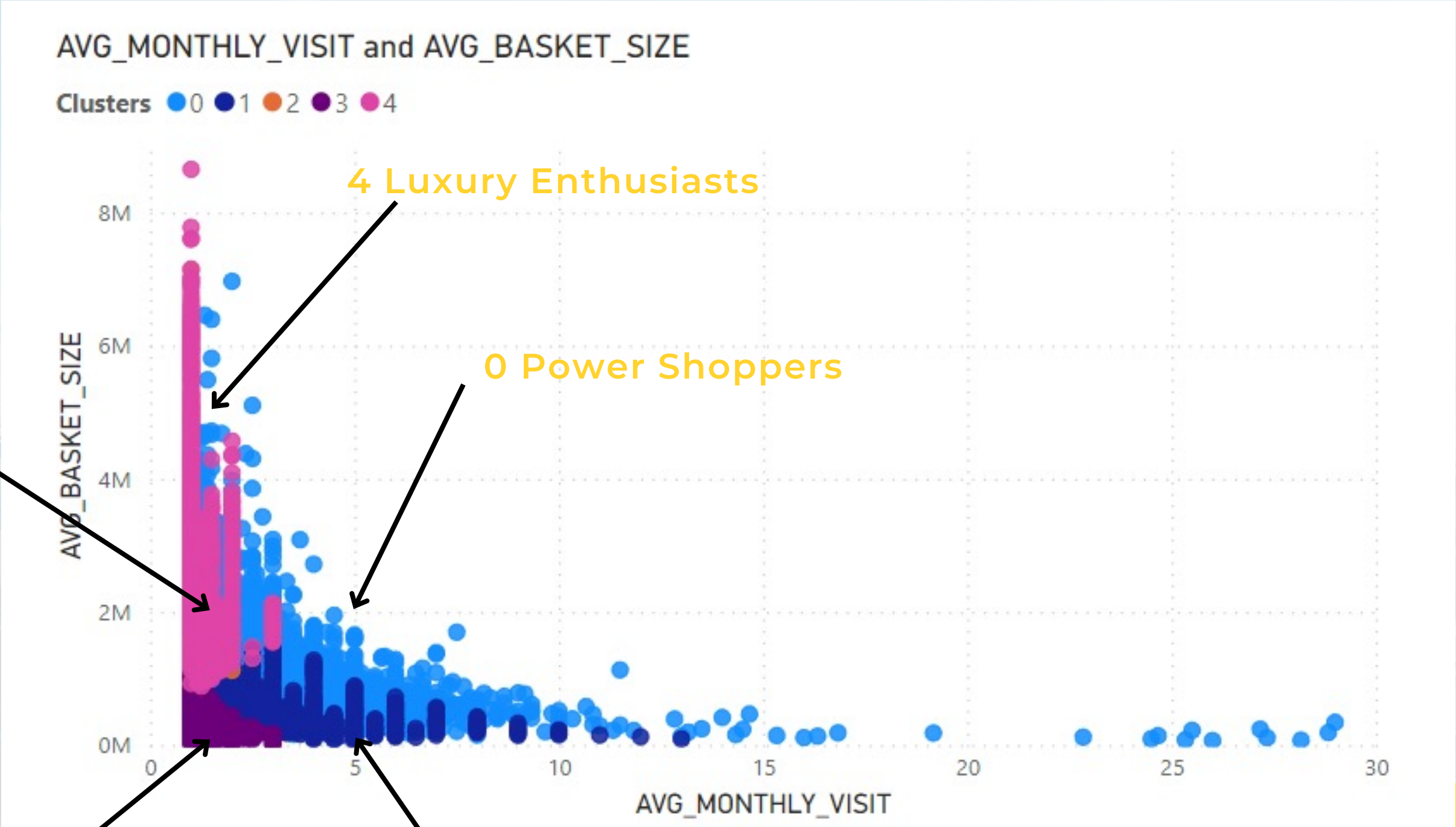
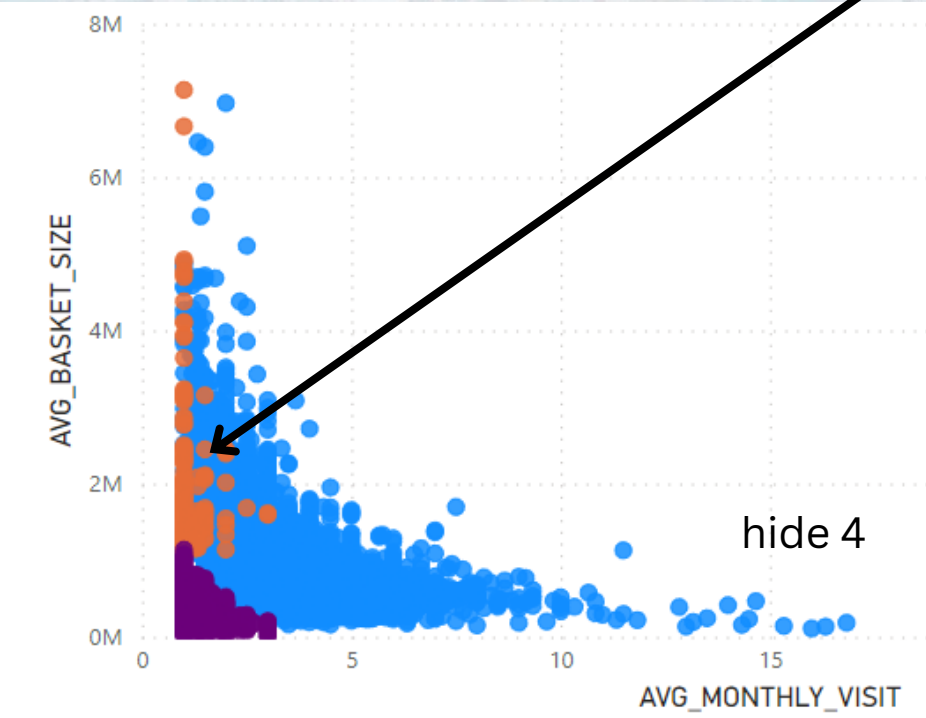


# K-MEAN CLUSTERING (PCA)



# VISUALIZE CLUSTERS

2 High-Yield Patrons



3 Frugal Shoppers

1 Value Maximizers

# CLUSTERS DEFINITION AND ACTIONS

## 0 Power Shoppers

28.76% of Revenue



This group represents customers who frequently purchase a high volume of your products. They may not always buy the most expensive items (average revenue per unit isn't the highest), but their overall spend is significantly higher than other groups.

For these customers, consider offering volume discounts or bundled deals on your products. They seem to buy frequently and in large quantities, so any incentives that encourage bulk buying will likely appeal to them. Also, keep them engaged with regular updates on new products or promotions.



# CLUSTERS DEFINITION AND ACTIONS

## 1 Value Maximizers

38.7% of Revenue



These customers have a moderate overall spend but a higher revenue per unit, indicating a preference for higher-end, potentially premium products, . They shop less frequently than the Honeycomb Hoarders, but when they do, they tend to go for the luxury items.

As the group contributing the highest share of revenues, Market your premium products towards this group. Given their preference for higher-end items, consider offering them exclusive access to new premium product launches, or perhaps a loyalty program that rewards high average spending.

# CLUSTERS DEFINITION AND ACTIONS

## 2 High-Yield Patrons



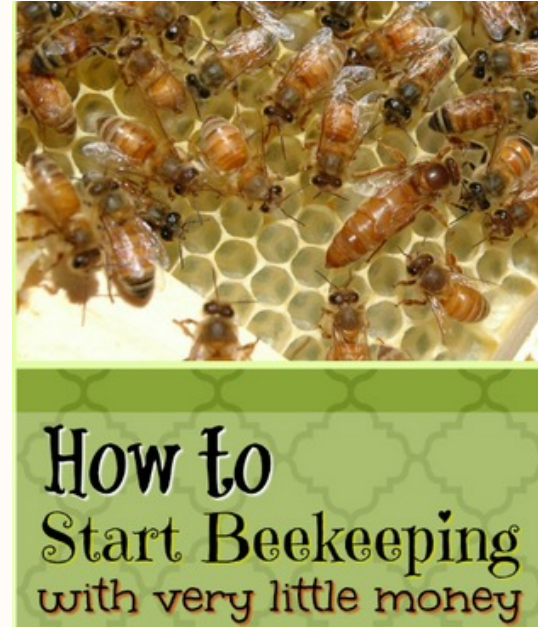
This group shows a lower overall spend but the highest revenue per unit, suggesting a preference for premium, possibly high-end products. They shop less frequently than the other groups but make significant purchases when they do.

Given their tendency to spend big on premium items, personalized service might be a great approach for these customers. Consider offering a VIP membership or personalized recommendations. These customers are also likely to appreciate being the first to know about new premium product arrivals.



# CLUSTERS DEFINITION AND ACTIONS

## 3 Frugal Shoppers



These customers spend the least overall and have the lowest revenue per unit. They don't shop often, but when they do, they tend to go for more affordable or lower-priced items, making them your budget-conscious customers.

Since these customers are more budget-conscious, focus on offering them great value for their money. Promotions, discounts, or sales would be attractive to this group. Also, consider marketing more affordable products to them.

# CLUSTERS DEFINITION AND ACTIONS

## 4 Luxury Enthusiasts



These customers spend a considerable amount, focusing on high-revenue items, suggesting a preference for specific, potentially premium products such as bee pollen. They don't shop as frequently as some other groups, but their purchases are significant when they do.

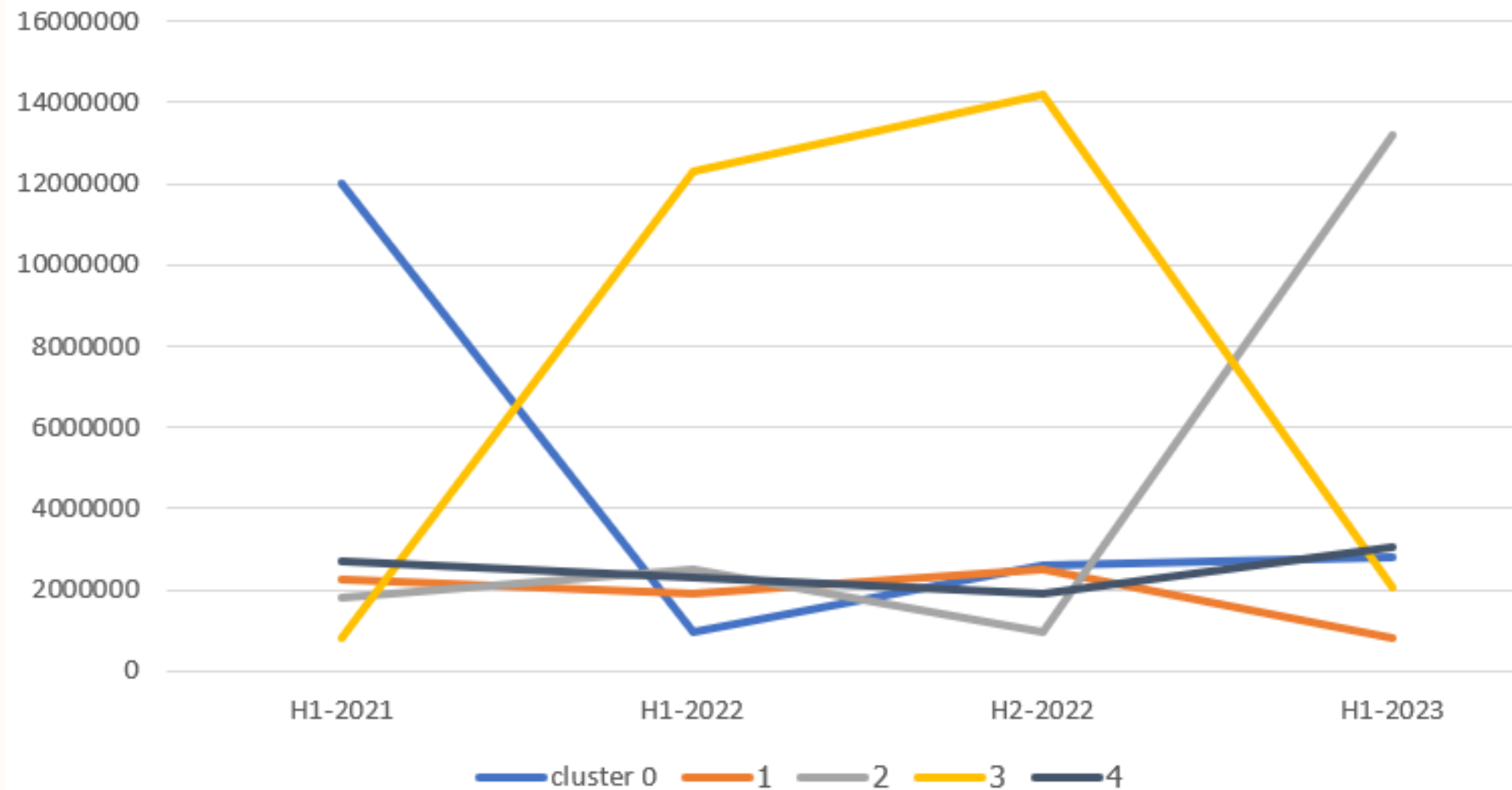
These customers are interested in specific high-revenue products such as bee pollen. Keep them in the loop about new variants of their favorite products, or offer discounts on related products to encourage cross-selling.

The average total spend for each cluster during each half-year period (H1-2021, H1-2022, H2-2022, H1-2023) and the %HoH growth rates.

Clusters	H1-2021	H1-2022	H2-2022	H1-2023	Growth H1-2022 vs H1-2021	Growth H2-2022 vs H1-2022	Growth H1-2023 vs H2-2022
0	11990004.09	945556.31	2600583.79	2811794.3	-92.12%	174.94%	8.12%
1	2257096.68	1925343.3	2475827.53	831868.85	-14.68%	28.58%	-66.39%
2	1780056.17	2477966.33	953706.79	13202615.87	39.21%	-61.51%	1284.00%
3	782687.47	12319351.4	14191338.2	2072804.21	1473.60%	15.15%	-85.39%
4	2676238.08	2324073.63	1884393.65	3046872.39	-13.18%	-18.91%	61.68%

# THE AVERAGE TOTAL SPEND SITUATION

Average Total Spend over period



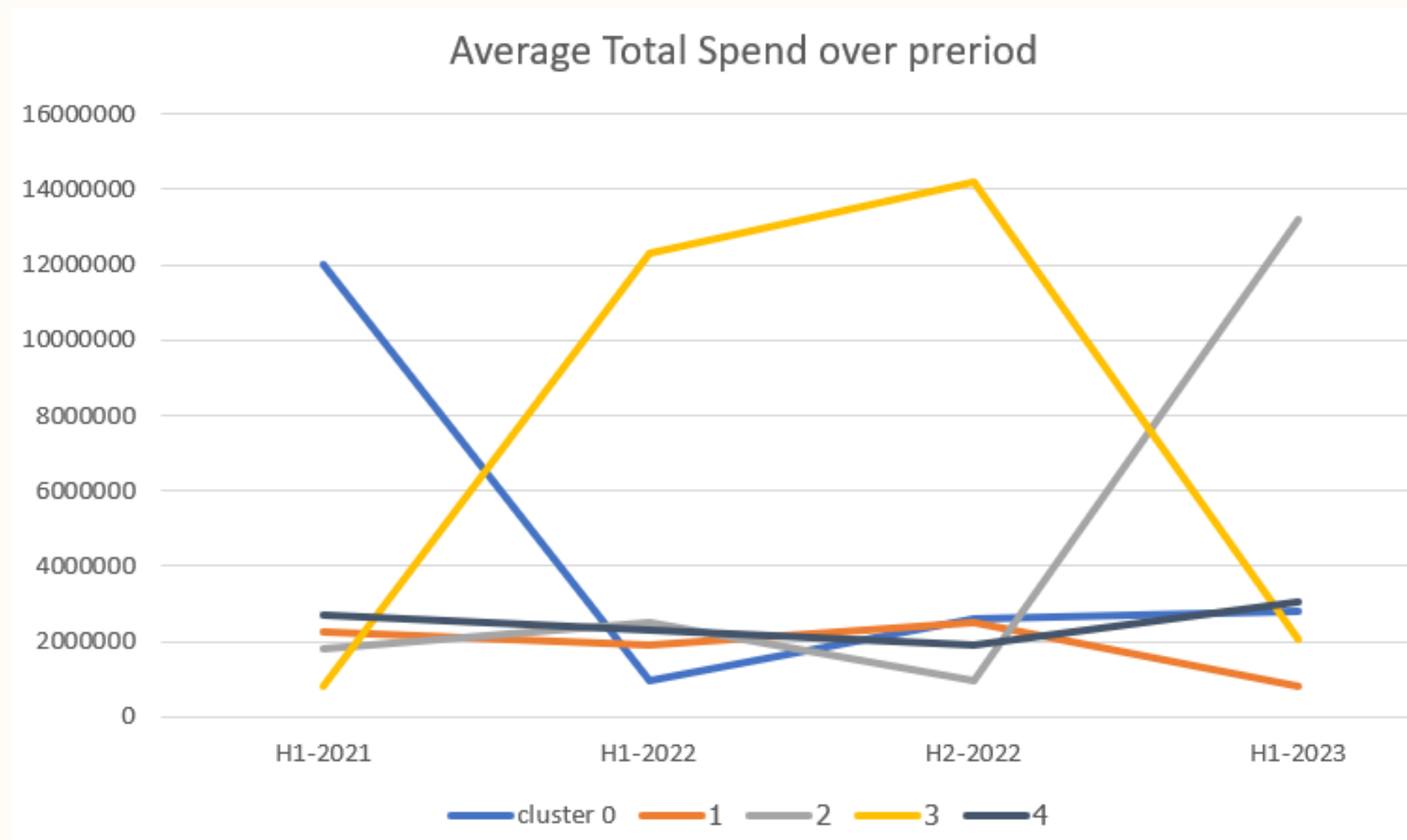
## 0 Power Shoppers

experienced a significant **decrease** in spending from H1-2021 to H1-2022 but **rebounded** considerably in H2-2022 and **continued to grow** in H1-2023. As a marketing manager, you could focus on maintaining this positive trend. This could include loyalty programs, special promotions, or personalized marketing based on their preferences.

## 1 Value Maximizers

This cluster showed a **consistent decrease** in spending over the periods. Consider researching why this group's **spending has decreased** – are they dissatisfied, or have their needs changed? You might survey these customers or use market research to get more data. Then, develop strategies to re-engage this group, such as targeted offers or new products that better meet their needs.

# THE AVERAGE TOTAL SPEND SITUATION



## 4 Luxury Enthusiasts

This cluster has shown **mixed results**, with an initial decrease in spending, followed by another decrease, but a significant increase in H1-2023.

Here, it would be beneficial to understand the reasons behind the **fluctuations** and focus on strategies that led to the positive result in H1-2023.

## 2 High-Yield Patrons

There's a **massive spike in spending** in H1-2023 after a **considerable decrease** in H2-2022. The marketing team should try to understand what caused this spike. Was it a particular marketing campaign or a new product? Understanding this can help replicate this success in the future.

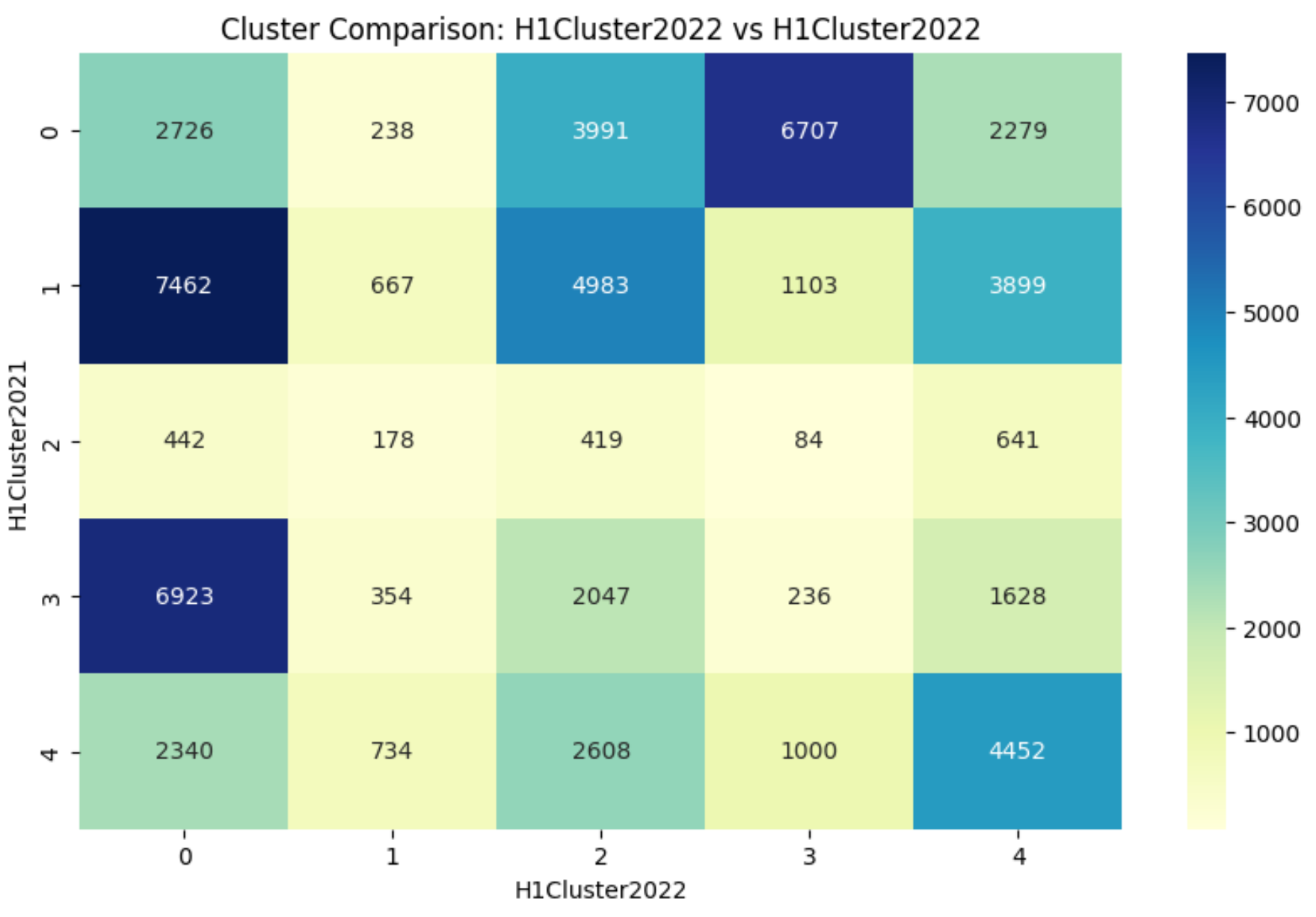
## 3 Frugal Shoppers

This group showed a **sharp increase** in spending from H1-2021 to H1-2022, followed by a **small increase** in H2-2022, but a **huge decrease** in H1-2023. It's crucial to understand what led to this massive drop.

Has the cluster been affected by economic factors, or have you lost these customers to competitors?



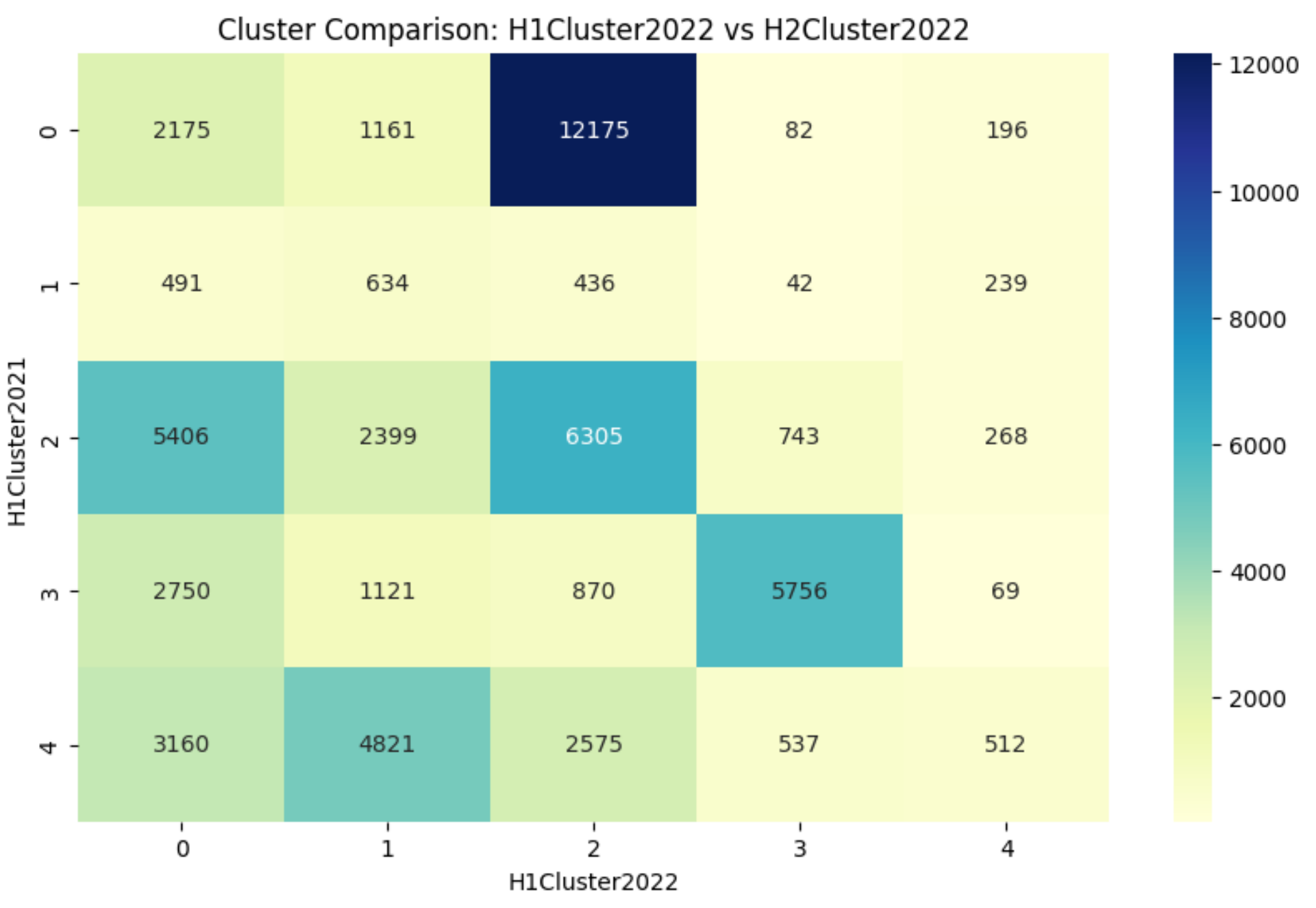
# Customer Movement Analysis



What is insights from analysis?

**In the first haft year we found from cluster power shoppers : Customer 6,707 people have changed to be group frugal shoppers. and they behavior seem they are the new customer because of they buy many product to find the best product from them. and customer from the cluster1 is the loyalty from our product**

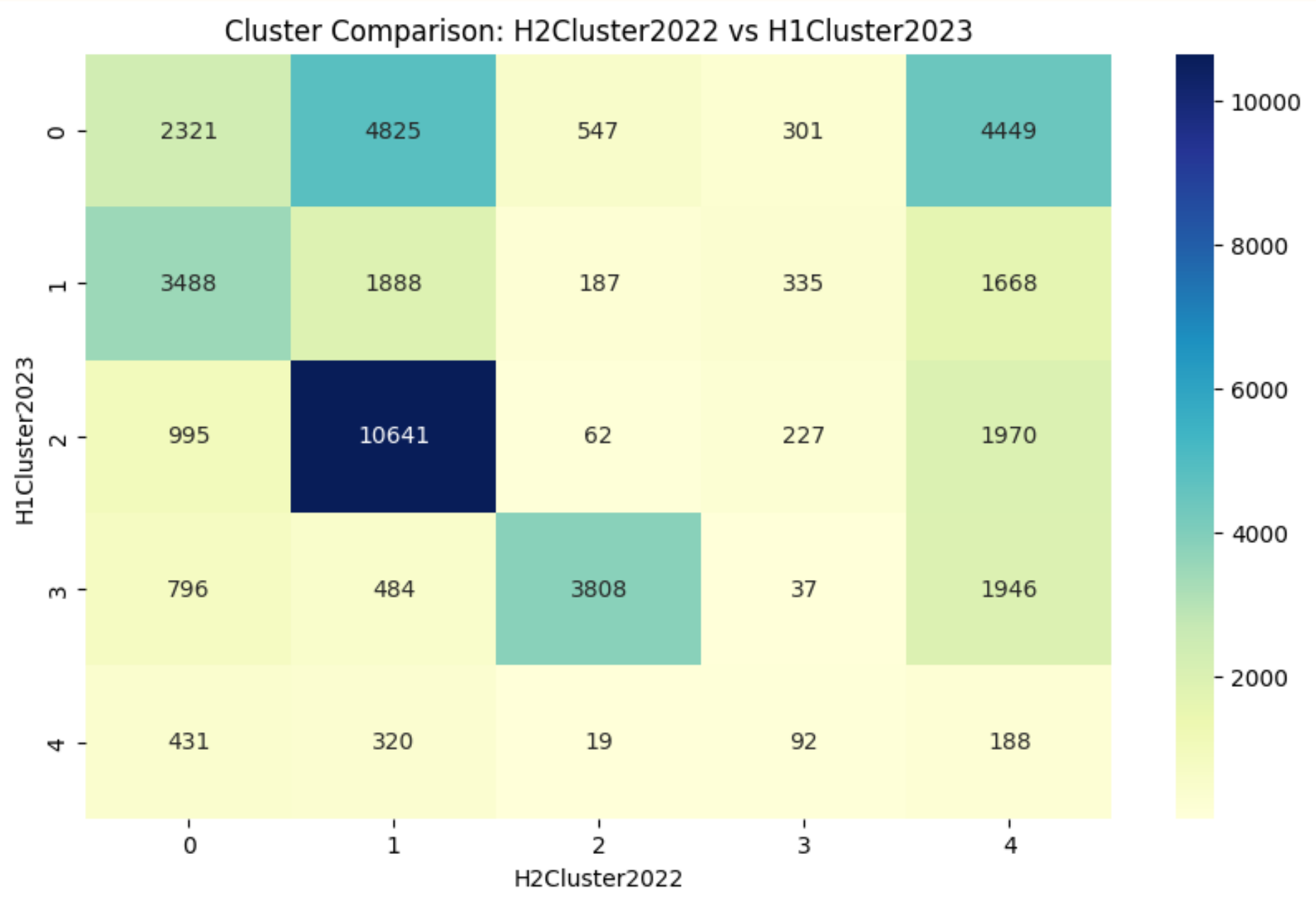
# Customer Movement Analysis



What is insights from analysis?

**In this period the customer from group power shoppers change to be group high-yield patrons Group value maximizers is the lowest come in this Period because of They shop less frequently they tend to go for the luxury items. so in this haft year they will not come along to buy our product much.**

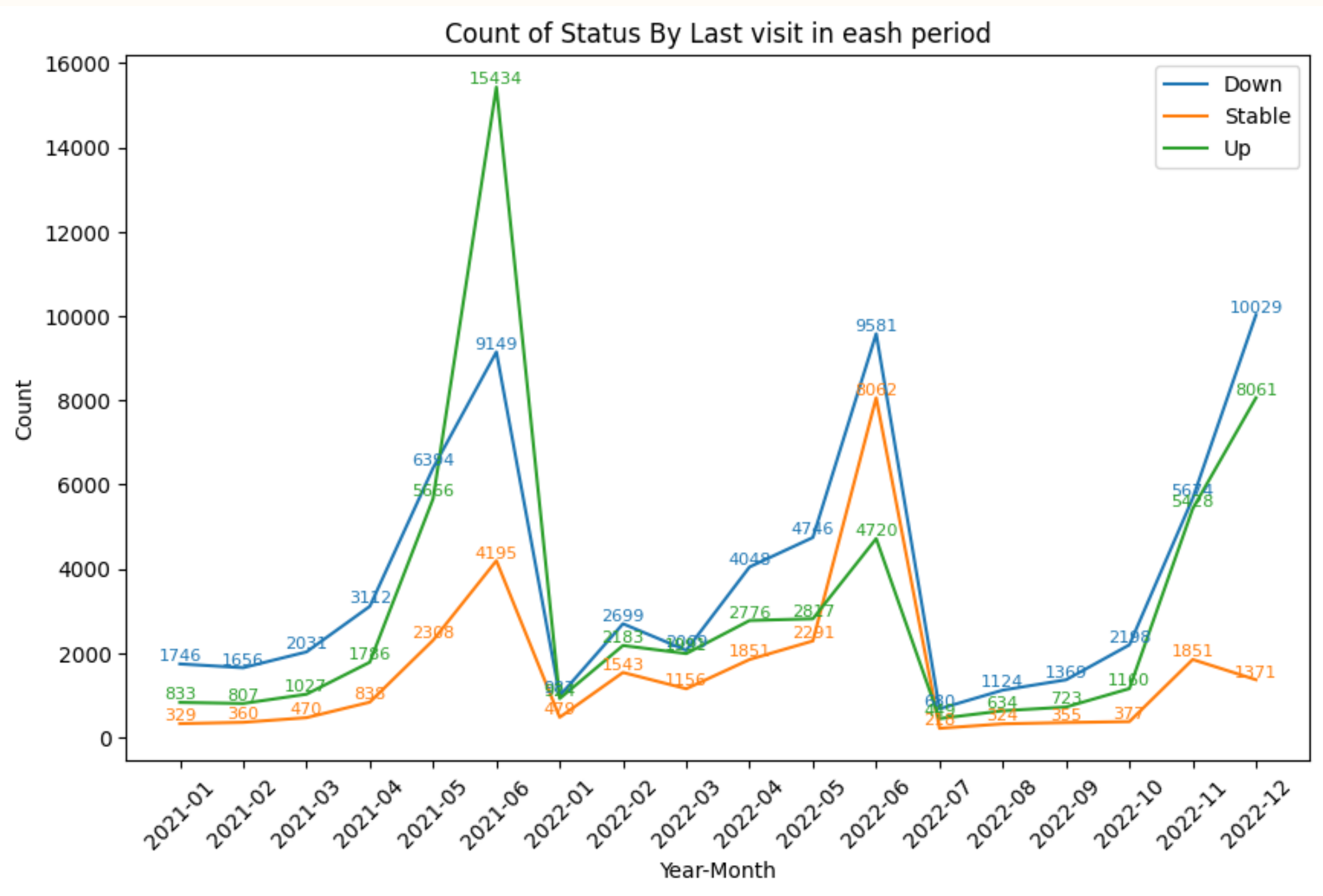
# Customer Movement Analysis

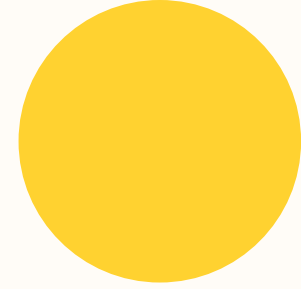


What is insights from analysis?

**This period group high-yield patrons is com back to buy tour product .**  
**And it having big changed on the group high-yield patrons , Many customer had changed to group value maximizers so that mean After they consume our new and our products they need to use the luxury product**

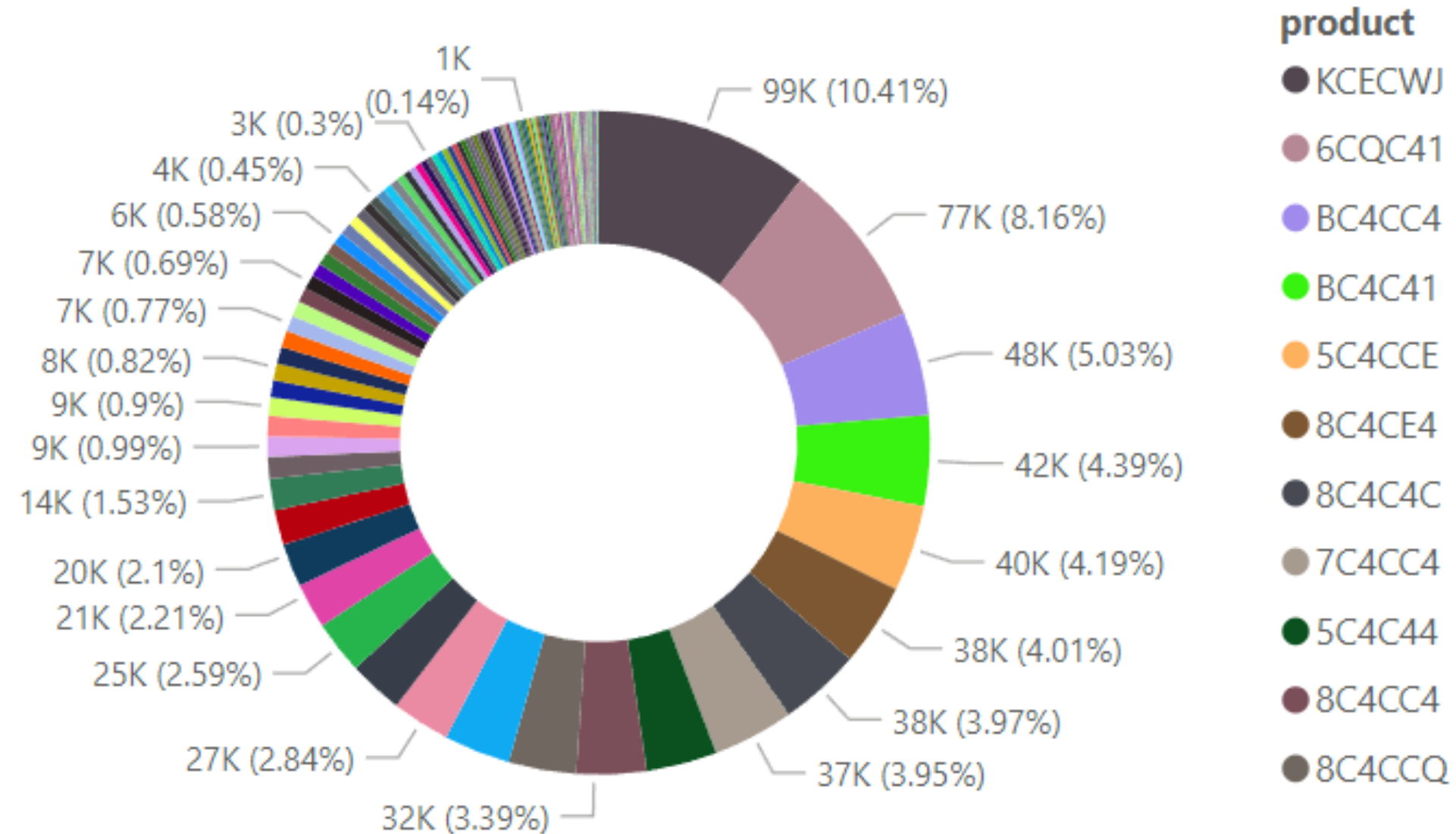
# Customer Movement Analysis





# ABOUT PRODUCT

Sum of qty by product

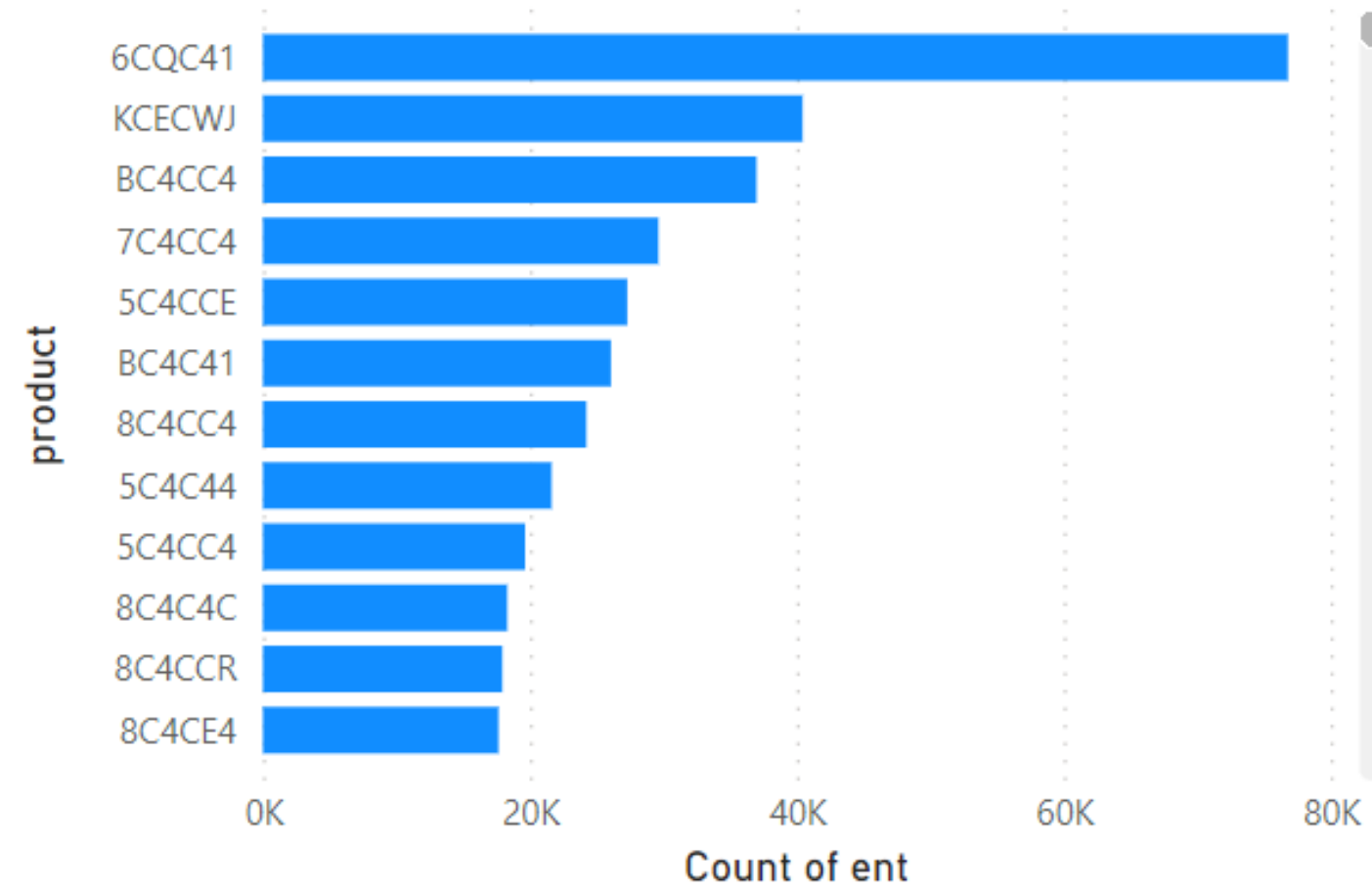


Product KCECWJ in 2023 is the best seller.



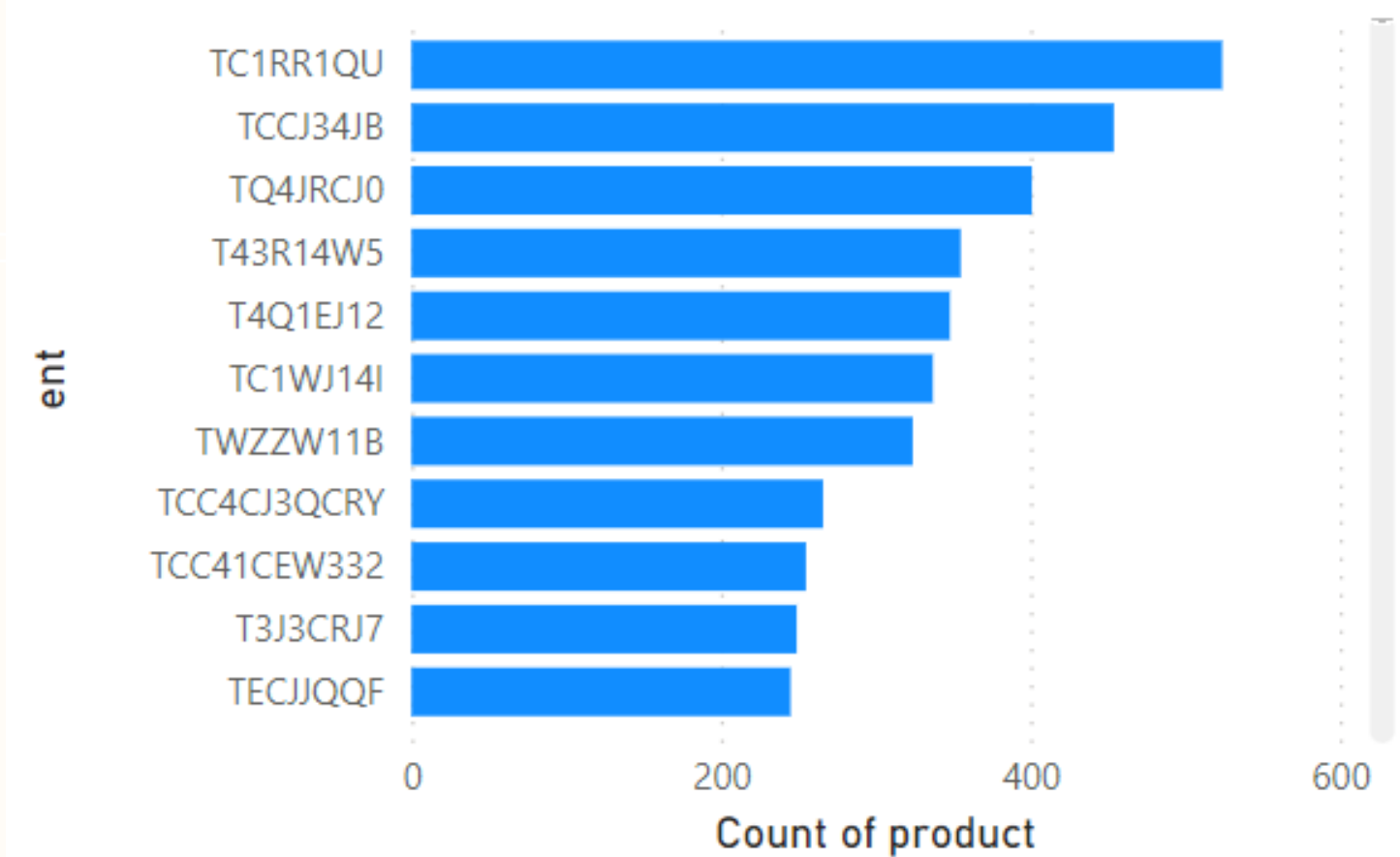
# ABOUT PRODUCT

Count of ent by product



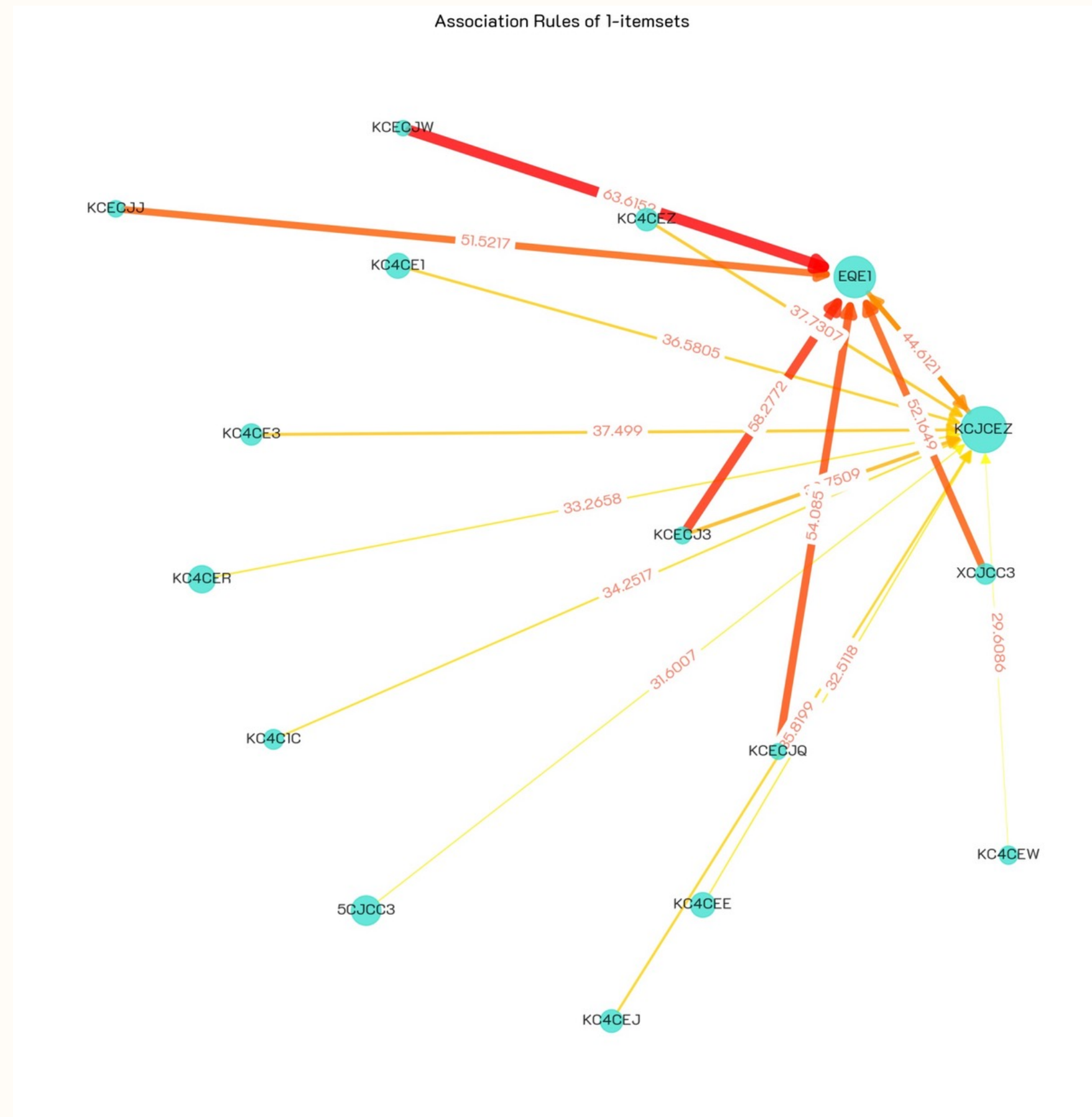
In 2023 , From the graph count of ent by product, that product 6CQC41 has the highest number of purchases by user

Count of product by ent



In 2023 , From the graph count of product by ent, that ent TC1RR1QU has bought the most products

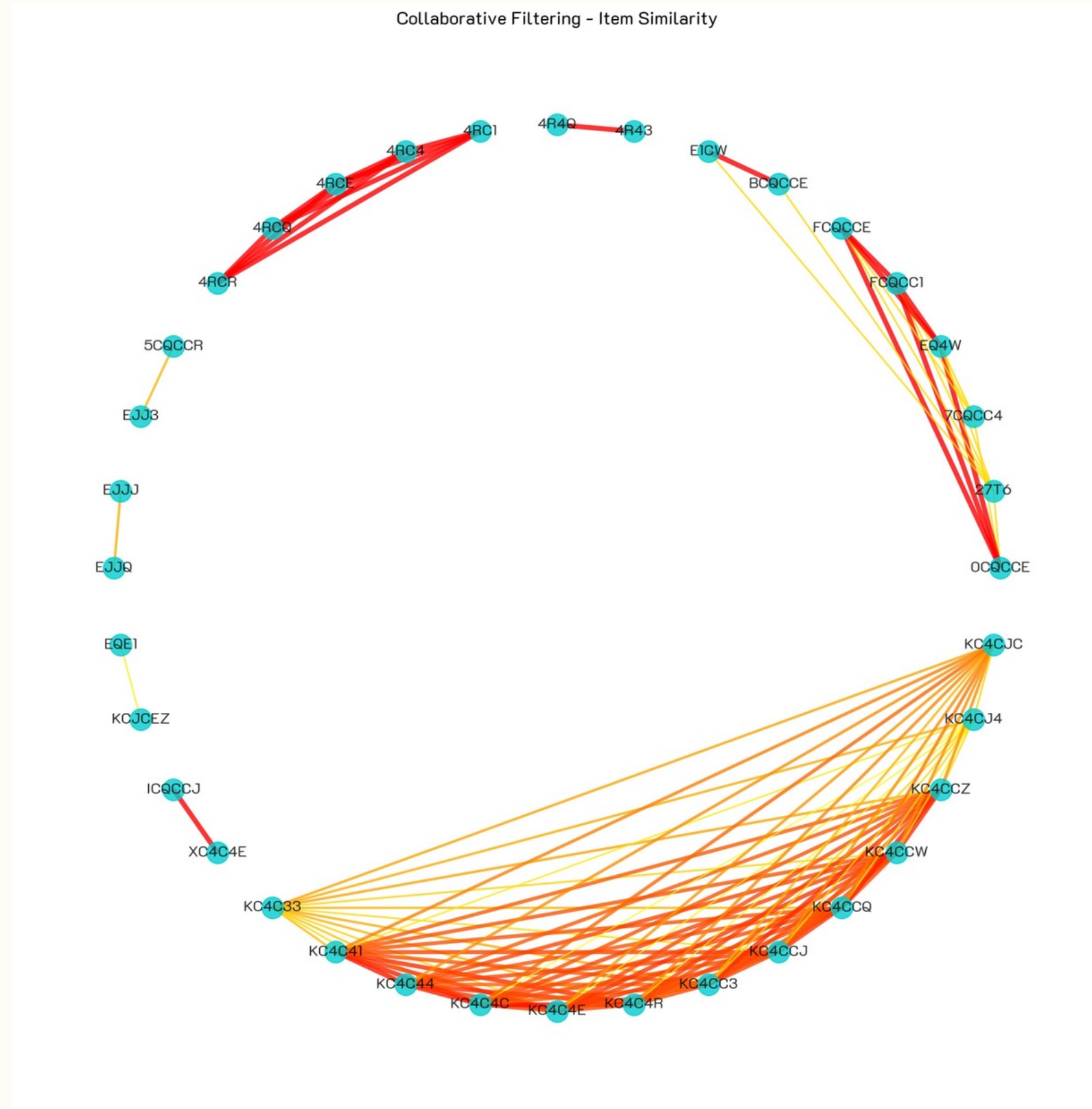
# ASSOCIATION RULES OF 1-ITEMSETS



# NETWORK CHART

Number of nodes: 36

Number of edges : 109



# EXAMPLE PRODUCT RECOMMENDATION

	x	y	type	label
0	-29.195936	15.500793	user	TCC4QCZJZQI
1	-31.034950	12.539783	user	TCC4QCE34Z7
2	-30.895979	12.545423	user	TCC44Q3JQQ2
3	-42.474350	15.750132	user	TCC41JJZCRF
4	-44.844601	16.965530	user	TCC4JZCJ1JK
...	...	...	...	...
338	-45.712360	17.017563	item	XCJC4C
339	-45.303249	18.027248	item	XCJCC3
340	-45.527443	16.479038	item	XCJCCZ
341	-44.848515	15.555982	item	XCQCC4
342	-45.731972	16.577234	item	YC1CC4

	ent	item	r	r_pred
6759	TZW3RE37	8C4CC1	0	0.492314
2414	TWJZ3WRK	44C4	0	0.236019
6209	TECC14RU	7C4CC4	0	0.095685
6999	TCCWJ4ZU	8C4CC4	0	0.138311
7113	TCCERZ1EZCY	8C4CCJ	0	0.322085
...	...	...	...	...
8571	TCC41JJZCRF	BC4C41	0	0.110733
10079	TCC41J33J4I	DCJCC3	0	0.101314
6771	TCC41C1QECI	8C4CC1	0	0.393047
7031	TCC41C1CJCK	8C4CC4	0	0.040927
6762	TC1EJJE2	8C4CC1	0	0.368845

# USER & ITEM ON LATENT SPACE TOP 100

