



# E-COMMERCE BEHAVIOR ANALYSIS

DS Intake 43 - ITI



# E-COMMERCE INDUSTRY

E-commerce or Electronic commerce is a process of buying, selling, transferring, or exchanging products, services, and/or information via electronic networks and computers



# TABLE OF CONTENT

- PROBLEM STATEMENT
- OUR OBJECTIVE
- PREPARARTION INSIGHTS
- ANALYSIS AND INSIGHTS
- RECOMMENDATIONS MODEL

# PROBLEM STATEMENT



The problem statement is to identify the key factors driving customer behavior and loyalty to the online cosmetics store and provide actionable recommendations to improve the business's decision-making process.



# SAMPLE FROM DATA

This file contains behavior data for two months (October and November 2019).

	event_date	event_time	event_type	product_id	category_id	brand	price	user_id	user_session
0	2019-10-01	00:00:00	cart	5773203	1487580005134238553	runail	2.62	463240011	26dd6e6e-4dac-4778-8d2c-92e149dab885
1	2019-10-01	00:00:03	cart	5773353	1487580005134238553	runail	2.62	463240011	26dd6e6e-4dac-4778-8d2c-92e149dab885
2	2019-10-01	00:00:07	cart	5881589	2151191071051219817	lovely	13.48	429681830	49e8d843-adf3-428b-a2c3-fe8bc6a307c9
3	2019-10-01	00:00:07	cart	5723490	1487580005134238553	runail	2.62	463240011	26dd6e6e-4dac-4778-8d2c-92e149dab885
4	2019-10-01	00:00:15	cart	5881449	1487580013522845895	lovely	0.56	429681830	49e8d843-adf3-428b-a2c3-fe8bc6a307c9

# OUR OBJECTIVE

The objective of the e-commerce events data analysis project is to use data analysis techniques to gain insights into user behavior on the e-commerce website and identify areas for improvement that will lead to increased conversion rates and reduced cart abandonment rates.

- Conversion Rate
- Cart Abandonment
- Average Order Value
- Funnel Analysis
- Analyze user behavior
- Segment customers using RFM analysis



# PREPARATION INSIGHTS

# Nulls in the data

- Category\_code = 98.3%
- Brand = 40.3%
- user\_session = 0.01%



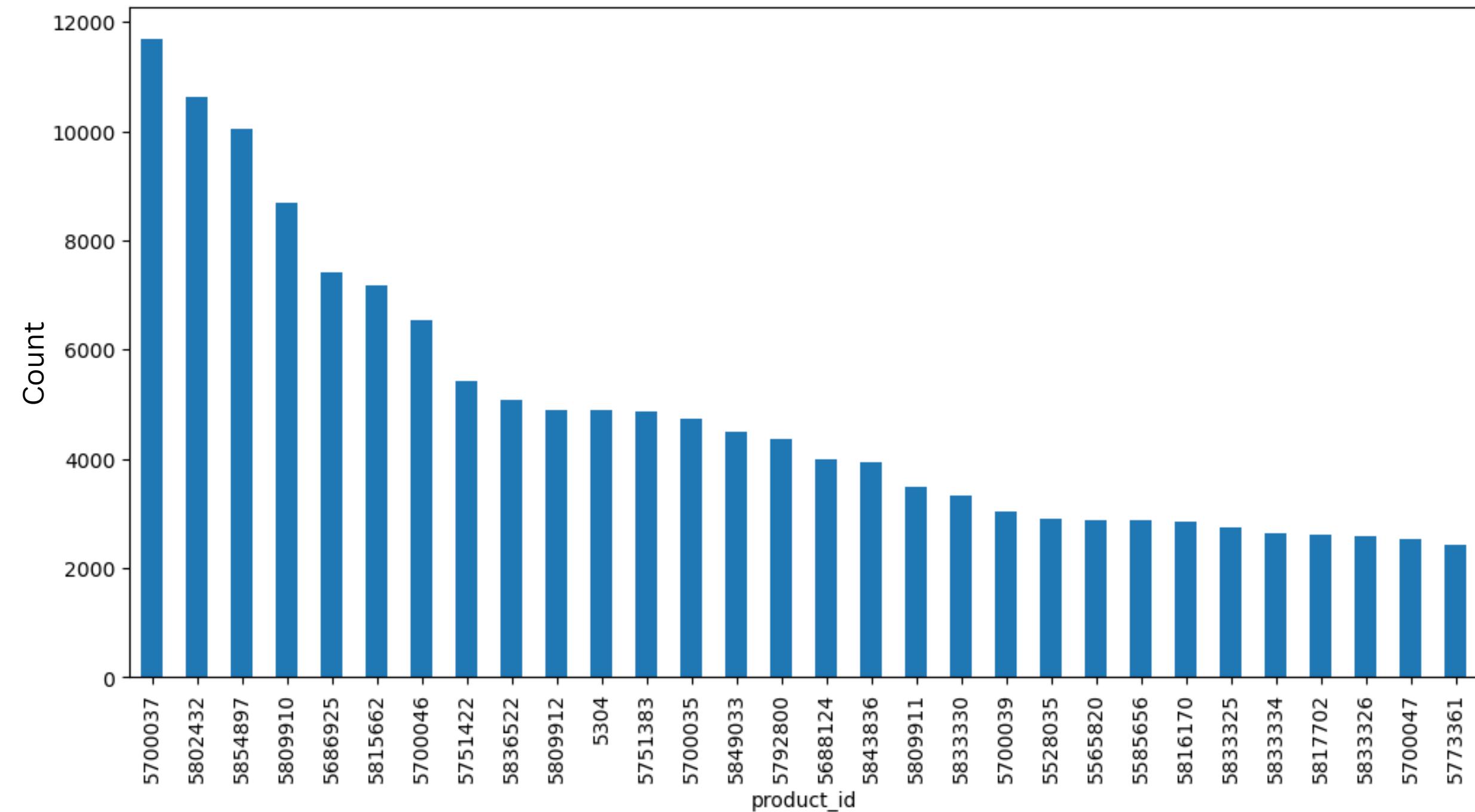
# ANALYSIS AND INSIGHTS

# PRODUCT ANALYSIS

Which product to put  
pormotions on based  
on the top products  
people put in **cart or  
purchase?**

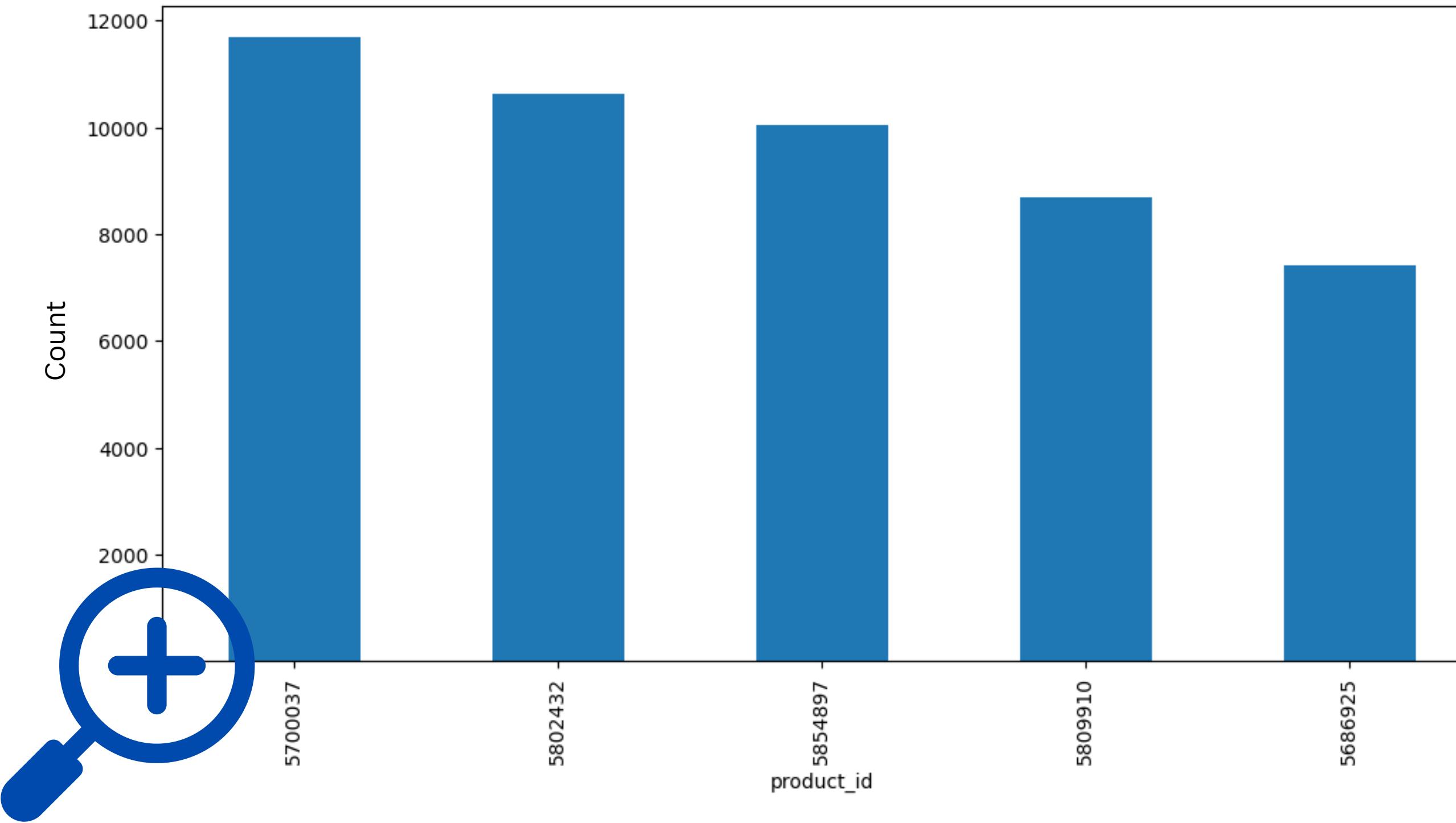
01

Which products to feature in the next advertising campaigns and promotions?



01

Which products to feature in the next advertising campaigns and promotions?

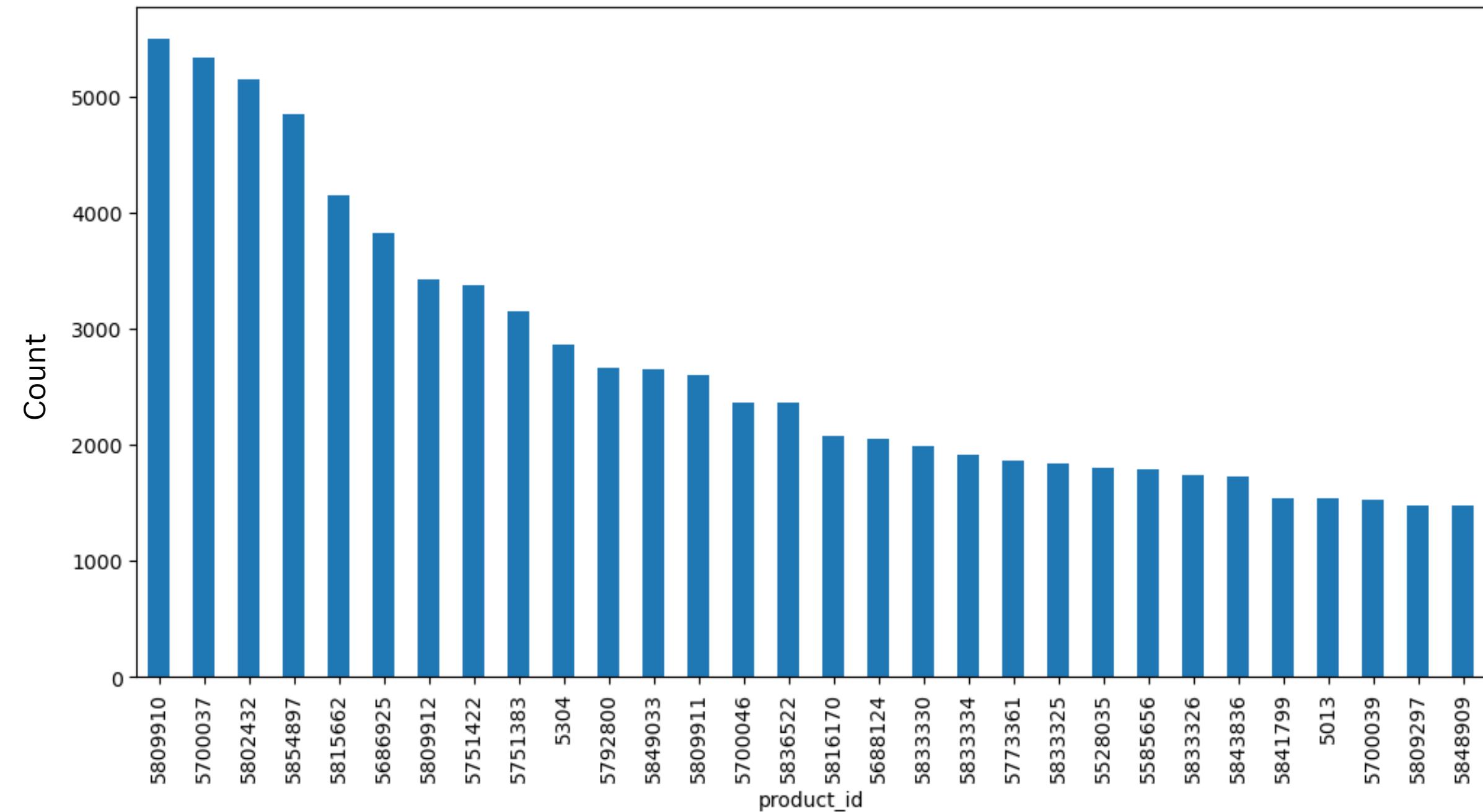


# PRODUCT ANALYSIS

Which product to put promotions on based on the top products people put **in cart and don't purchase?**

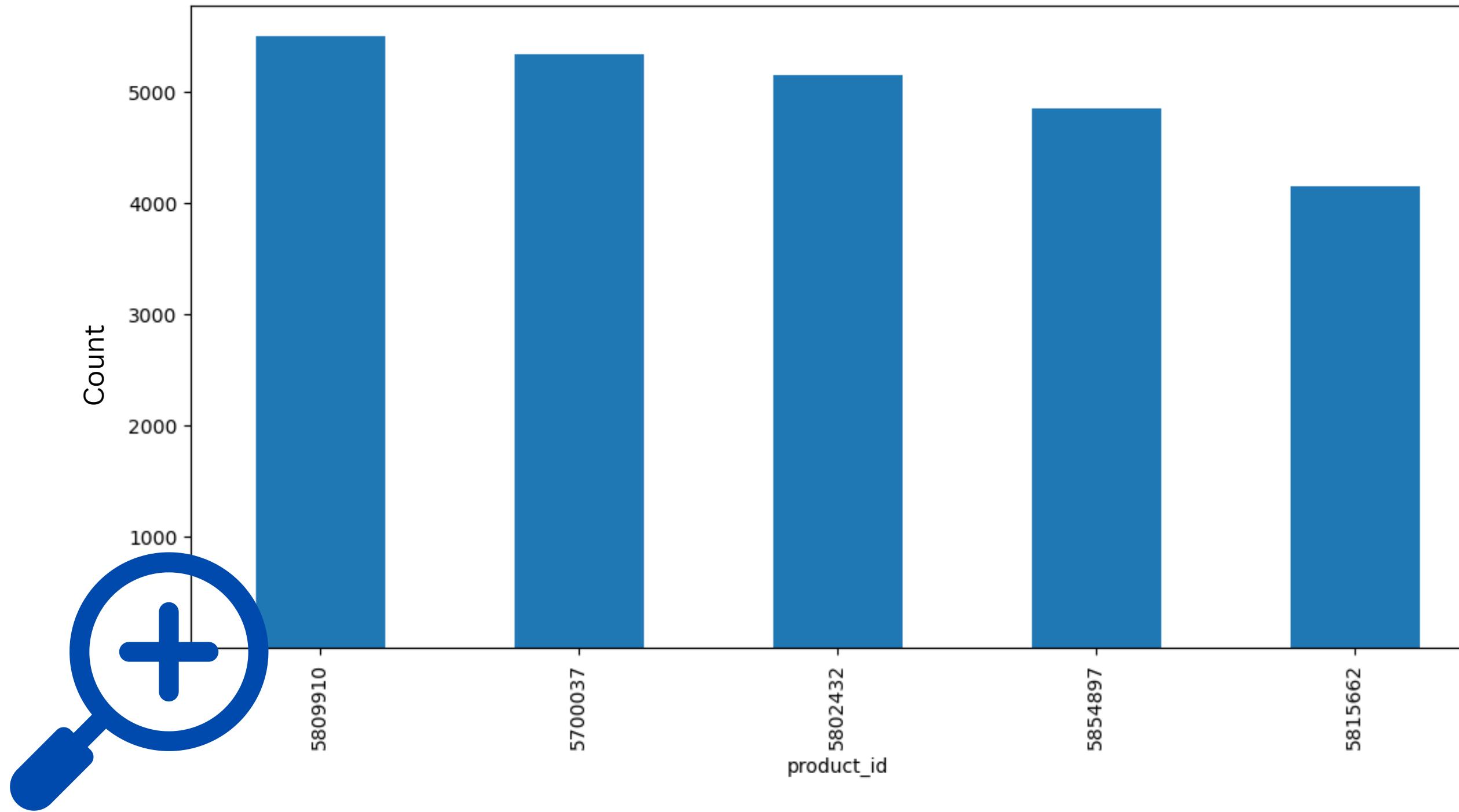
01

Which products to feature in the next advertising campaigns and promotions?



01

Which products to feature in the next advertising campaigns and promotions?



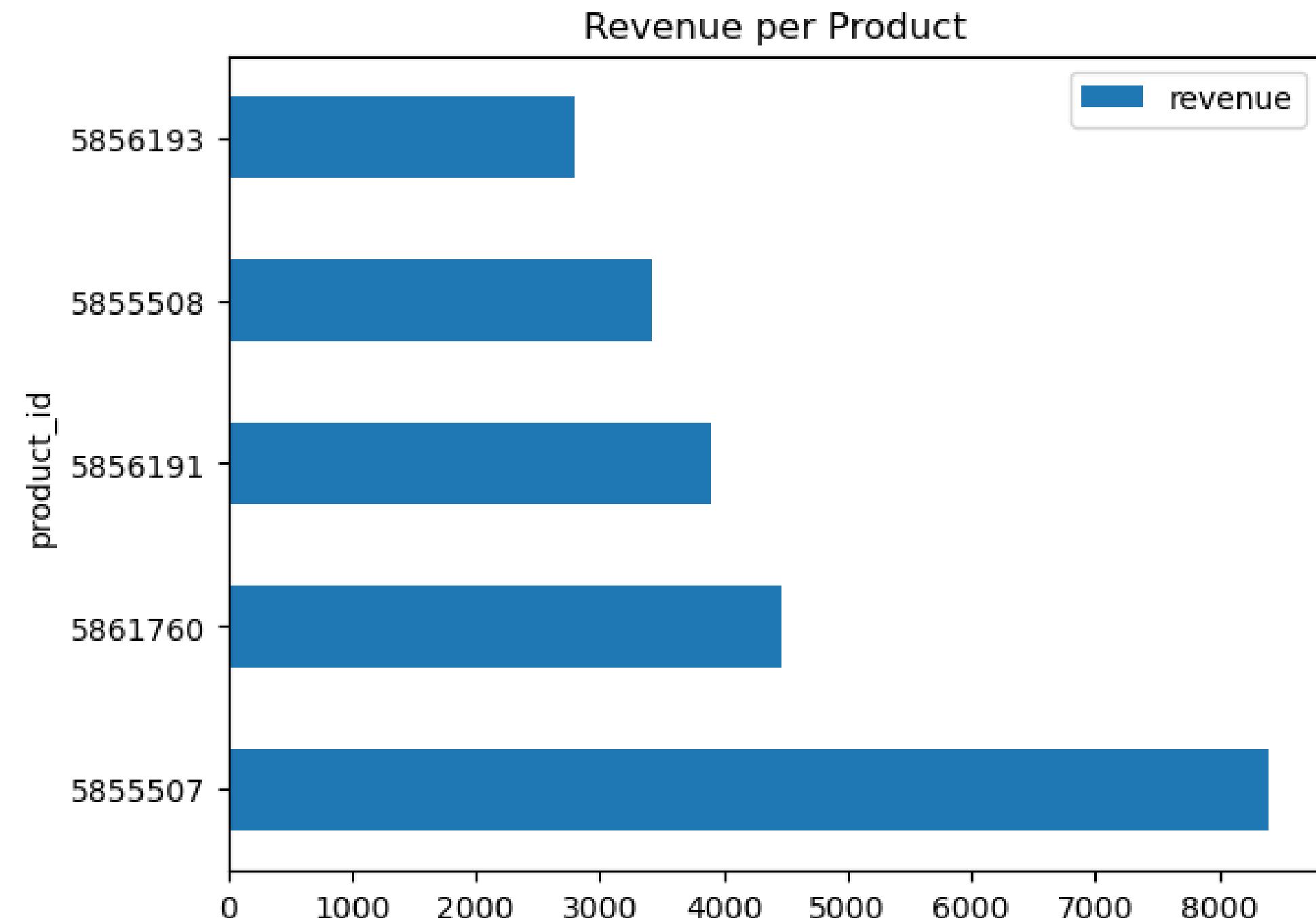
# SALES ANALYSIS

We sum the revenue by product to get the best products that get benefit for the website.

Which can help us to specially include them in the marketing campaign.

02

What are the products with best revenue?



# BRANDS ANALYSIS

03

Which brands are customers  
most loyal to?



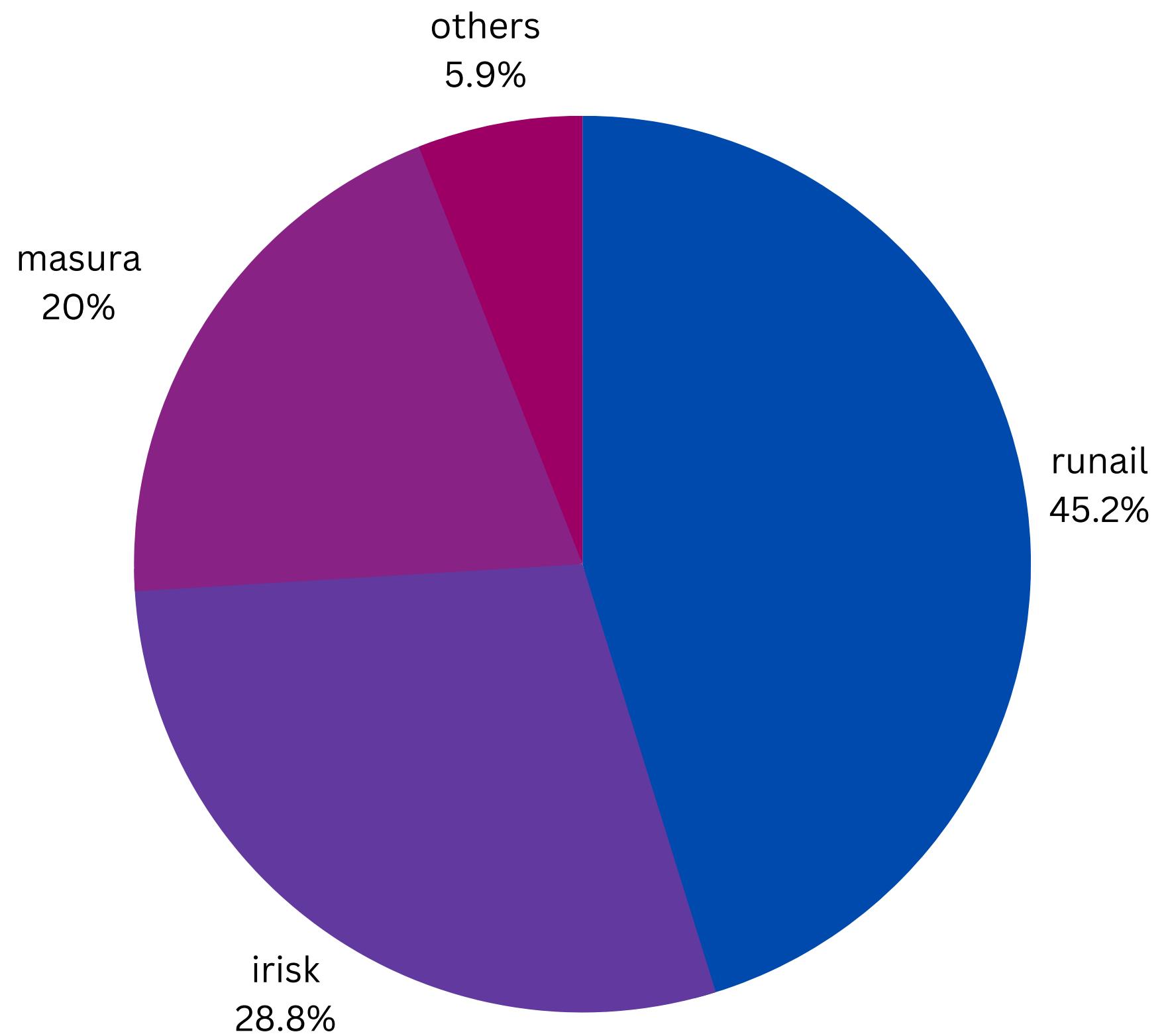
Calculating the most purchased  
brands is different than  
calculating the brands that  
customers are most loyal to.

## 03

### which brands does people buy the most?

When calculating the most purchased brands, you are simply looking at the total number of purchases made for each brand.

This metric may not necessarily indicate customer loyalty since a customer may purchase a brand's product once or twice but may not necessarily be loyal to that brand.

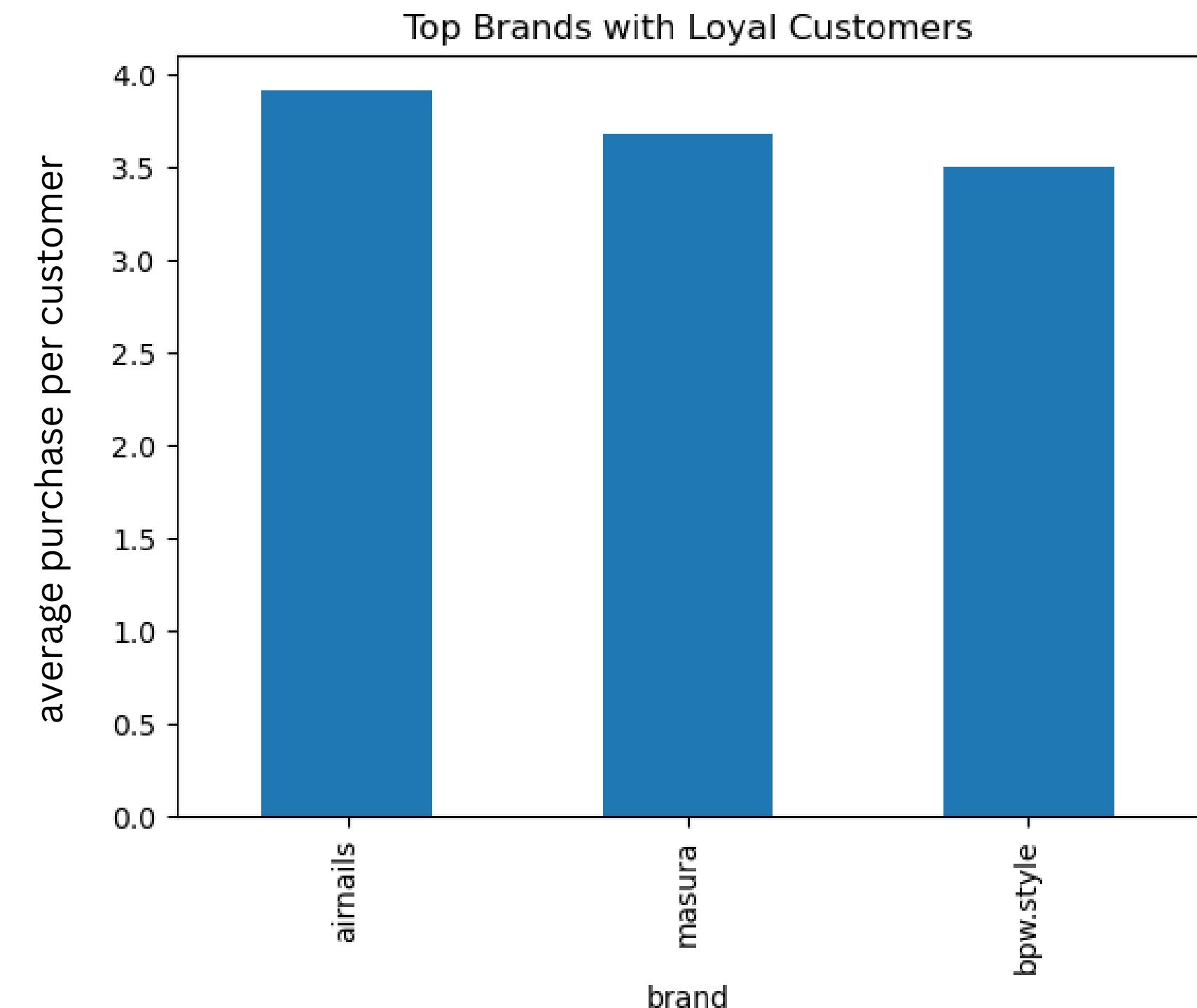


## 03

### Which brands are customers most loyal to?

On the other hand, when calculating the brands that customers are most loyal to, you are looking at the average purchase amount for each brand per customer.

This metric takes into account how much each customer spends on a particular brand over time, which indicates a higher level of customer loyalty.

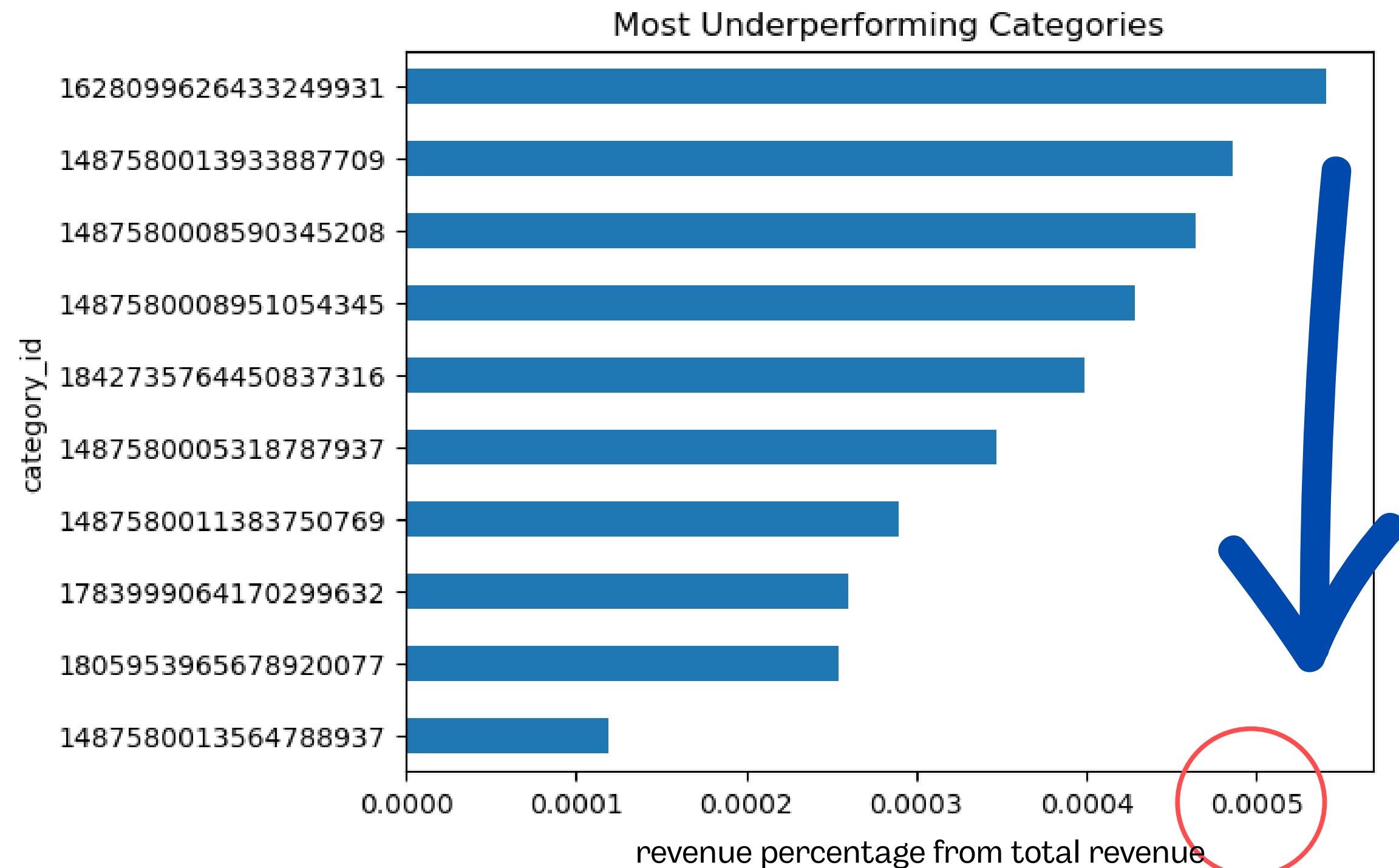


# CATEGORY ANALYSIS

Finding the underperforming categories may help us use our market campaign to fix the revenue for these categories to minimize the loss.

04

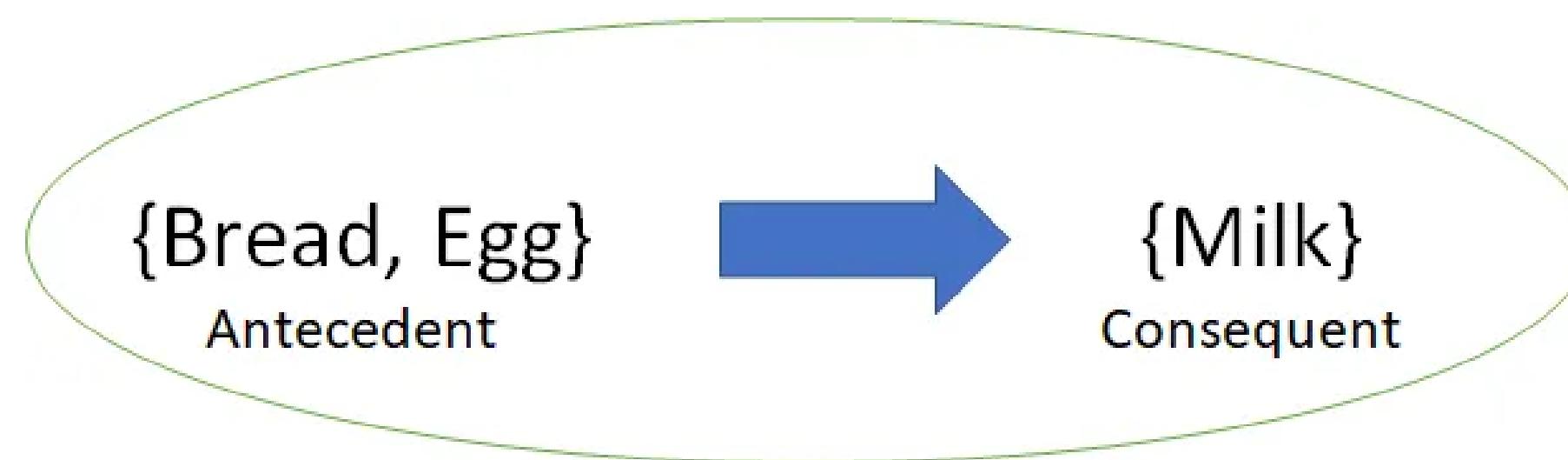
Are there any categories that are underperforming?



# ASSOCIATION ANALYSIS

05

Are there any products that are frequently purchased together?



Itemset = {Bread, Egg, Milk}

We can find out using Apriori algorithm and association rules.

# ASSOCIATION ANALYSIS

$$Support(\{X\} \rightarrow \{Y\}) = \frac{\textit{Transactions containing both } X \textit{ and } Y}{\textit{Total number of transactions}}$$



Support is a measure that gives an idea of how frequent an itemset is in all the transactions.

# ASSOCIATION ANALYSIS

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions containing both } X \text{ and } Y}{\text{Transactions containing } X}$$



Confidence measures the likelihood of items given that the shopping cart already has other items

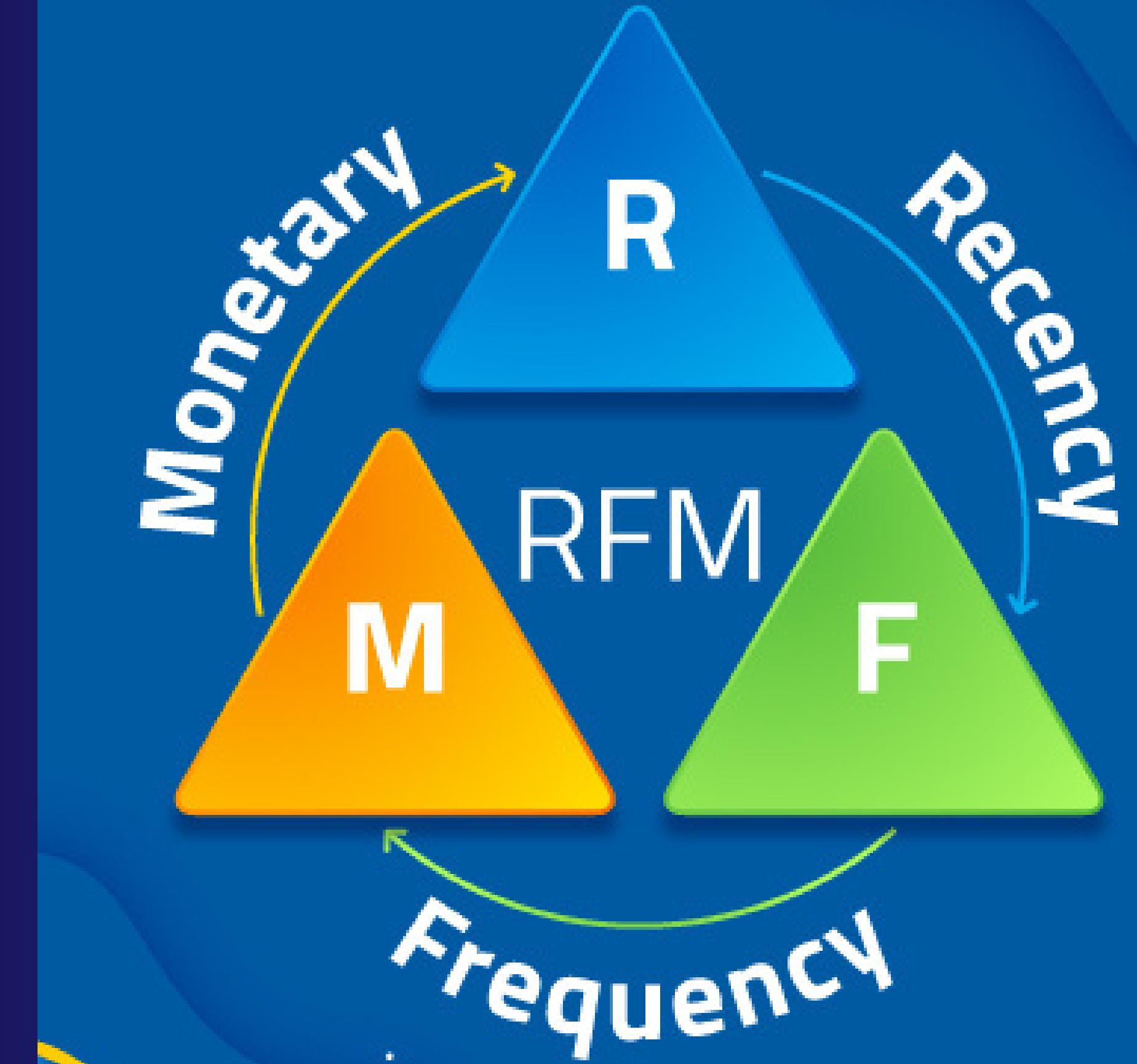
# ASSOCIATION ANALYSIS

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions \ containing \ both \ X \ and \ Y) / (Transactions \ containing \ X)}{Fraction \ of \ transactions \ containing \ Y}$$



Lift controls for the support (frequency) while calculating the conditional probability of occurrence of {Y} given {X}.

# WHO ARE THE MOST VALUABLE CUSTOMERS?



01

# RECENCY



	user_id	last_purchase_date	recency
0	9794320	2019-11-25 07:07:13	1218
1	10079204	2019-11-06 12:43:30	1237
2	10280338	2019-11-21 13:31:09	1222
3	15781728	2019-11-17 01:37:59	1226
4	19762782	2019-11-10 21:22:01	1232

02

# FREQUENCY

---

	user_id	frequency	edit
0	9794320	1	
1	10079204	1	
2	10280338	2	
3	15781728	1	
4	19762782	1	
...			

03

# MONETARY



monetary



user\_id

9794320 12.68

10079204 25.81

10280338 54.83

15781728 20.31

19762782 80.29

04

# RFM SCORE FOR EACH CUSTOMER

	user_id	last_purchase_date	recency	frequency	monetary	recency_score	frequency_score	monetary_score	rfm_score	
5143	384426984	2019-10-14 16:49:32	1260	2	86.93	5143	5	1	5	515
26610	545556413	2019-10-09 18:28:09	1265	1	152.60	26610	5	1	5	515
26588	545468494	2019-10-11 23:52:03	1262	1	260.50	26588	5	1	5	515
9635	439618166	2019-10-08 22:11:59	1265	2	104.68	9635	5	1	5	515
15564	485676003	2019-10-02 22:37:56	1271	1	323.60	15564	5	1	5	515
32831	557633004	2019-10-07 10:25:27	1267	1	88.73	32831	5	1	5	515
32853	557647577	2019-10-08 09:52:48	1266	1	86.43	32853	5	1	5	515
15573	485783522	2019-10-14 14:28:10	1260	3	144.91	15573	5	1	5	515
26544	545334224	2019-10-12 16:48:12	1262	3	153.89	26544	5	1	5	515
26543	545328268	2019-10-03 00:29:26	1271	1	355.00	26543	5	1	5	515

# WHAT ARE THE KPI'S THAT MEASURE CUSTOMER LOYALTY?

- REPEAT PURCHASE RATE (RPR)
- CUSTOMER RETENTION RATE (CRR)

01

# REPEAT PURCHASE RATE (RPR)

$$\text{Repeat Purchase Rate} = \frac{\text{Number of Repeat Purchase Customers}}{\text{Total Number of Customers}}$$

Measures the percentage of customers who return to buy from the same brand.

01

# REPEAT PURCHASE RATE (RPR)

```
repeat_customers = frequency_df[frequency_df['frequency'] > 1]
num_repeat_customers = len(repeat_customers)

num_customers = len(df['user_id'].unique())

# calculate the repeat purchase rate
rpr = (num_repeat_customers / num_customers) * 100

print('Repeat Purchase Rate: {:.2f}%'.format(rpr))
```

Repeat Purchase Rate: 1.32%

02

# CUSTOMER RETENTION RATE (CRR)

$$\text{Customer Retention Rate} = \frac{\text{Ending Customers} - \text{New Customers}}{\text{Beginning Customers}}$$

The percentage of existing customers who remain customers after a given period

02

# CUSTOMER RETENTION RATE (CRR)

```
▶ current_start_date = pd.to_datetime('2019-11-01')
current_end_date = pd.to_datetime('2019-11-30')
previous_start_date = pd.to_datetime('2019-10-01')
previous_end_date = pd.to_datetime('2019-10-30')

# identify the active customers in the current time period
df_purchased=df[df.event_type=='purchase']
active_customers = df[df['event_date'].between(current_start_date, current_end_date)]['user_id'].unique()

# identify the customers from the previous time period
previous_customers = df[df['event_date'].between(previous_start_date, previous_end_date)]['user_id'].unique()
# calculate the retention rate
retention_rate = len(set(active_customers) & set(previous_customers)) / float(len(previous_customers)) * 100

# print the result
print(f"The customer retention rate is {retention_rate:.2f}%")
```

□ The customer retention rate is 13.54%.

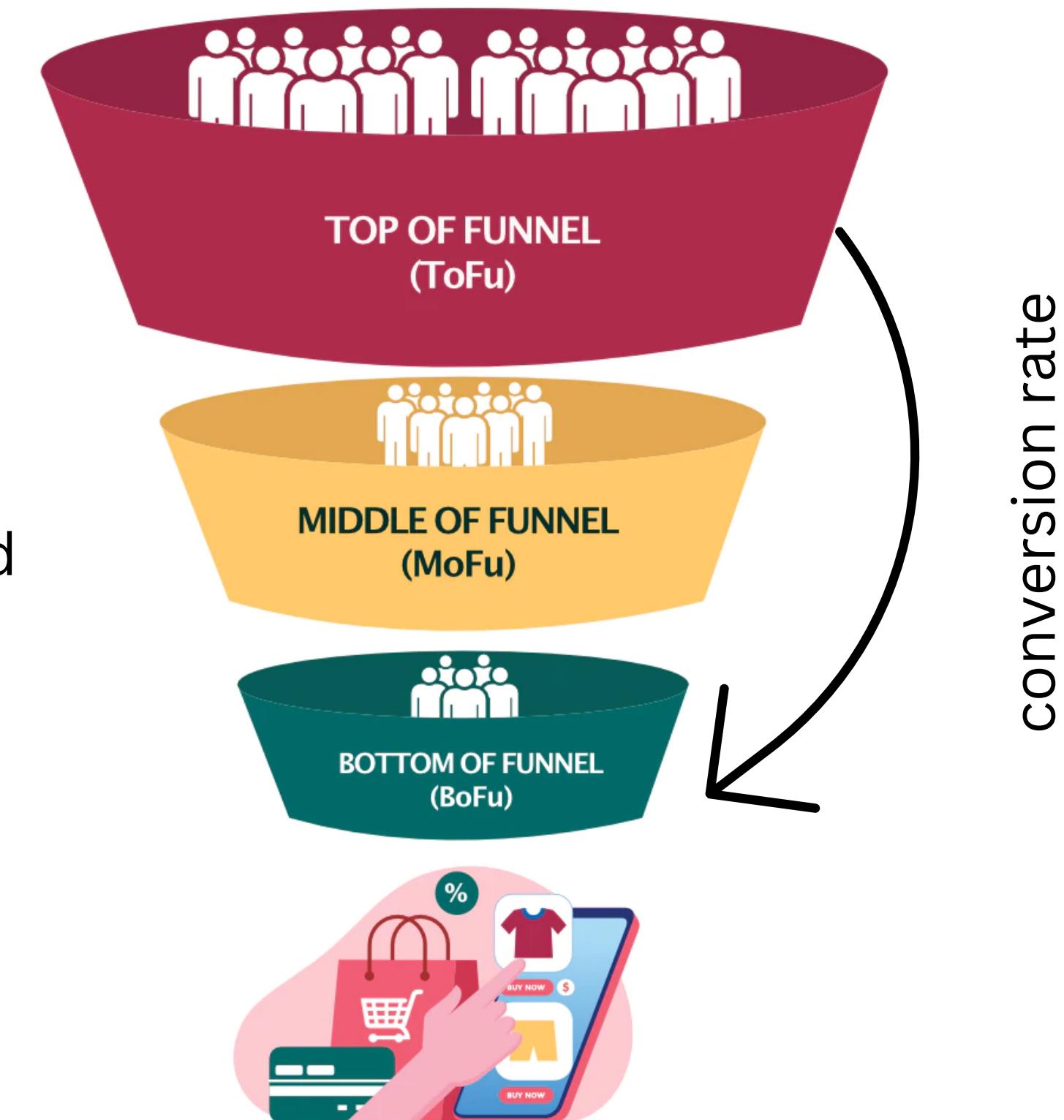
# FUNNEL ANALYSIS

# FUNNEL ANALYSIS

View

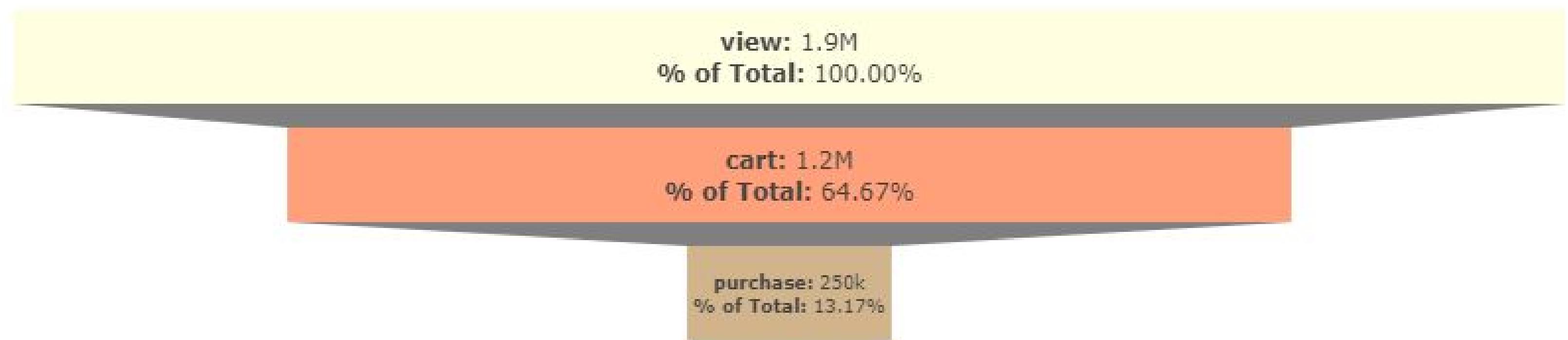
add to card

purchase



# FUNNEL ANALYSIS

Customer Funnel for Purchase Journey

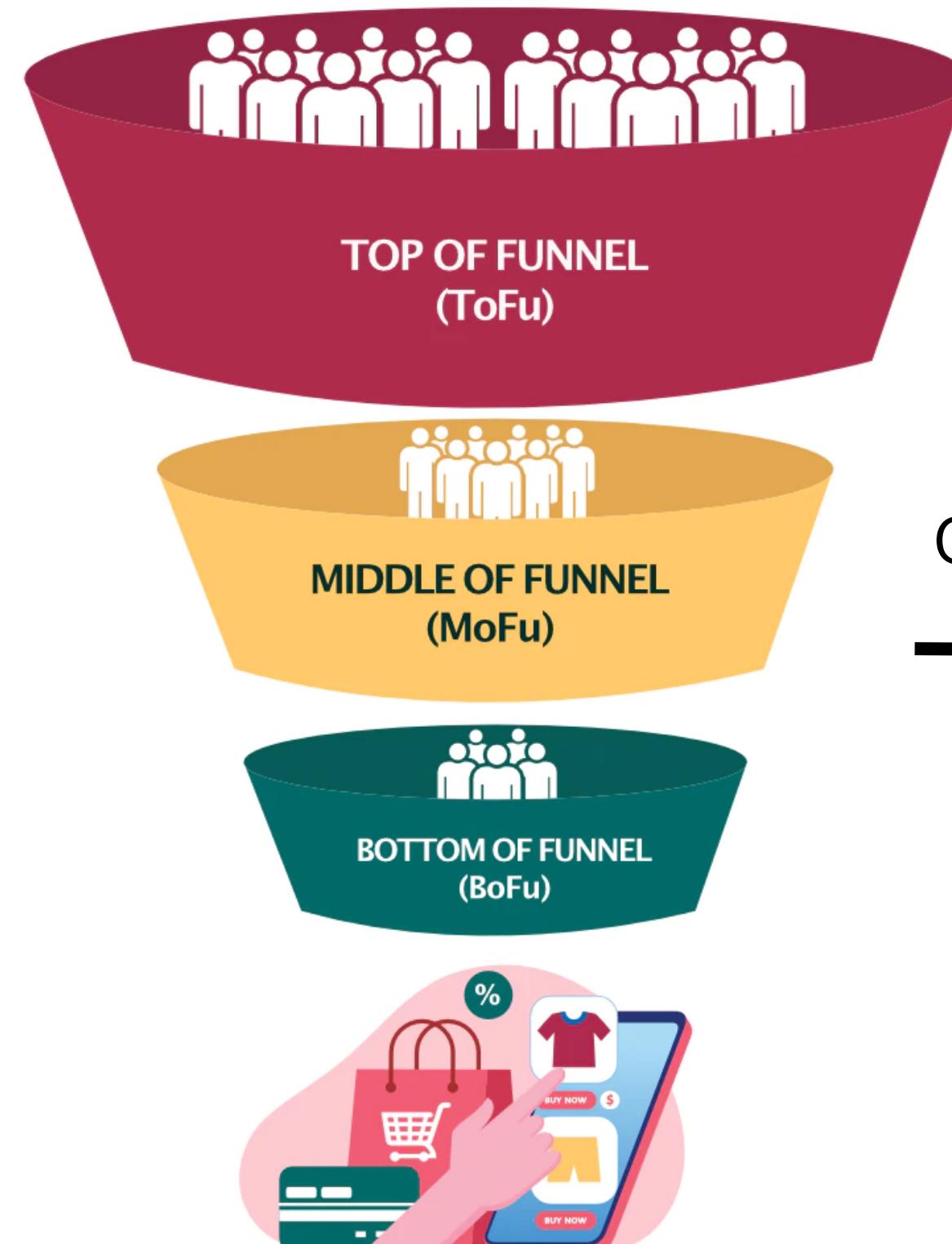


# FUNNEL ANALYSIS

View

add to card

purchase

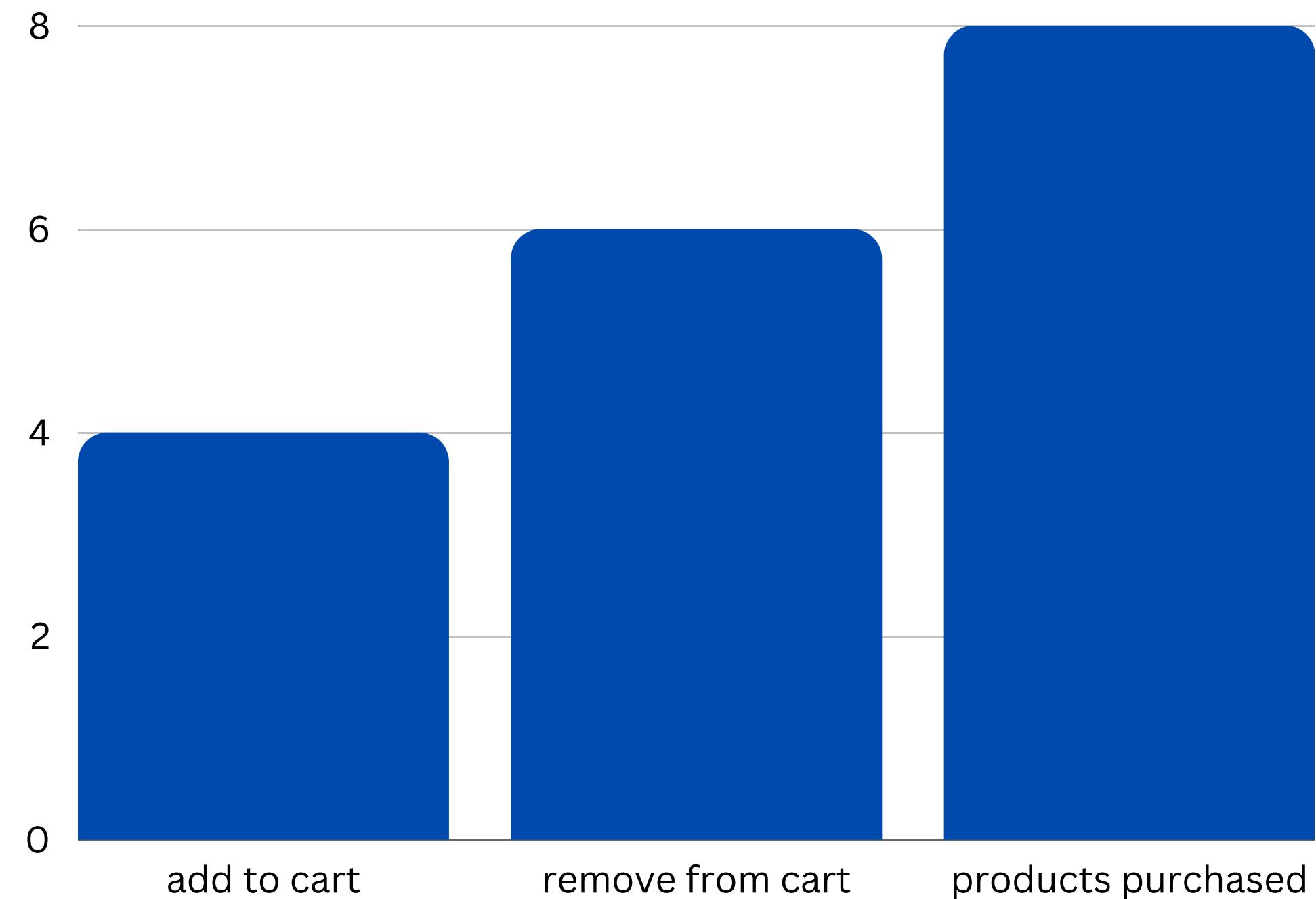


Cart Abandament rate 79%

# SALES ANALYSIS

First, we found the revenue per each product and while doing so we found the **products with best revenue**.

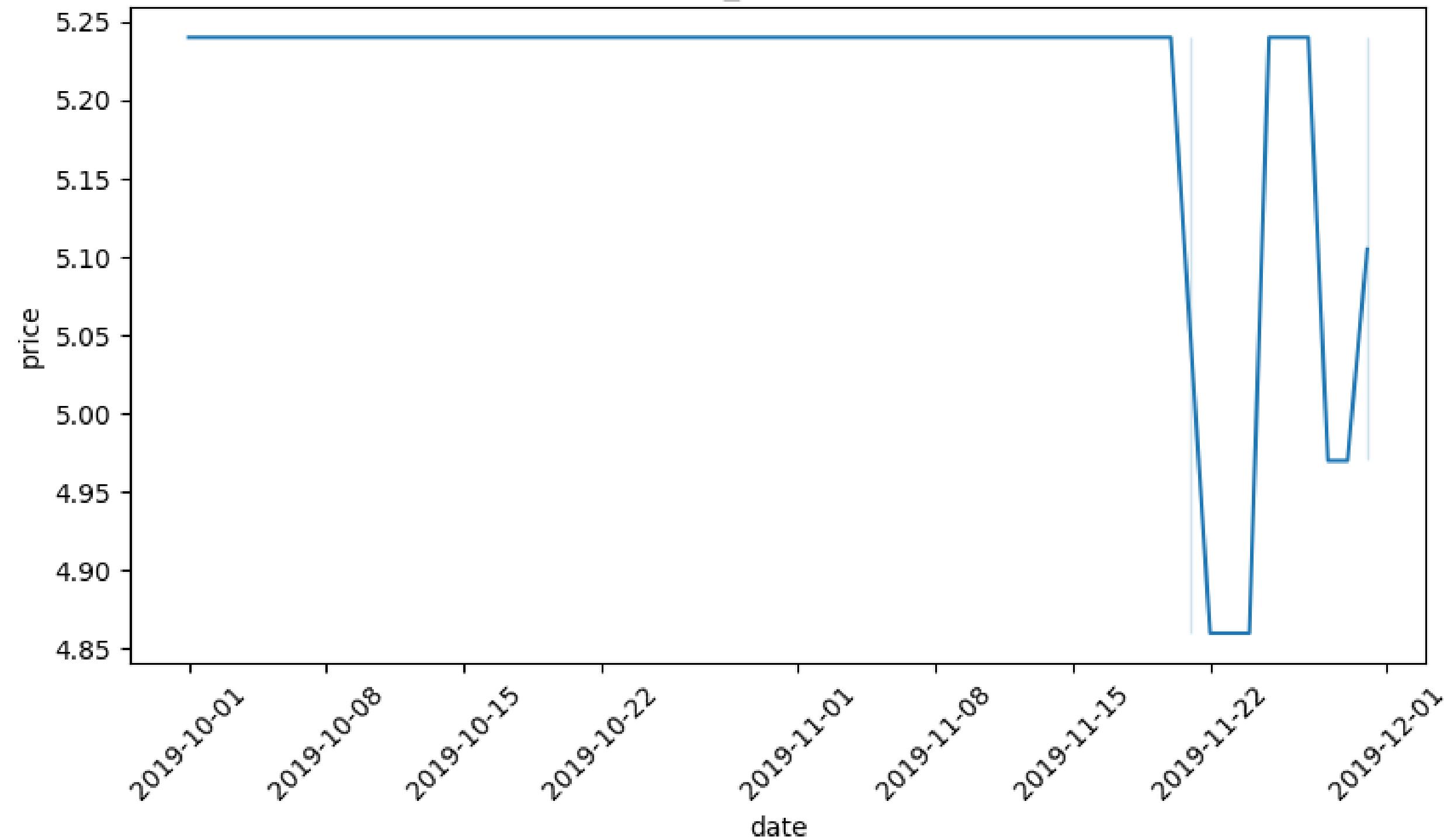
Website Performance on average



PRICE TRENDS FOR A  
PARTICULAR PRODUCT  
OVER TIME?

# PRICE TRENDS OVER TIME

product\_id = 5809910



After finding top 3 products appear in the data, we take a look into the price change over two months specifically october & november. we can see there is a unique pattern.

# PRICE TRENDS OVER TIME

Deeper look ..

December

**5.24\$**



**4.86\$**



**4.97\$**

Oct-1

..

Nov-20

Nov-21

..

Nov-25

Nov-26

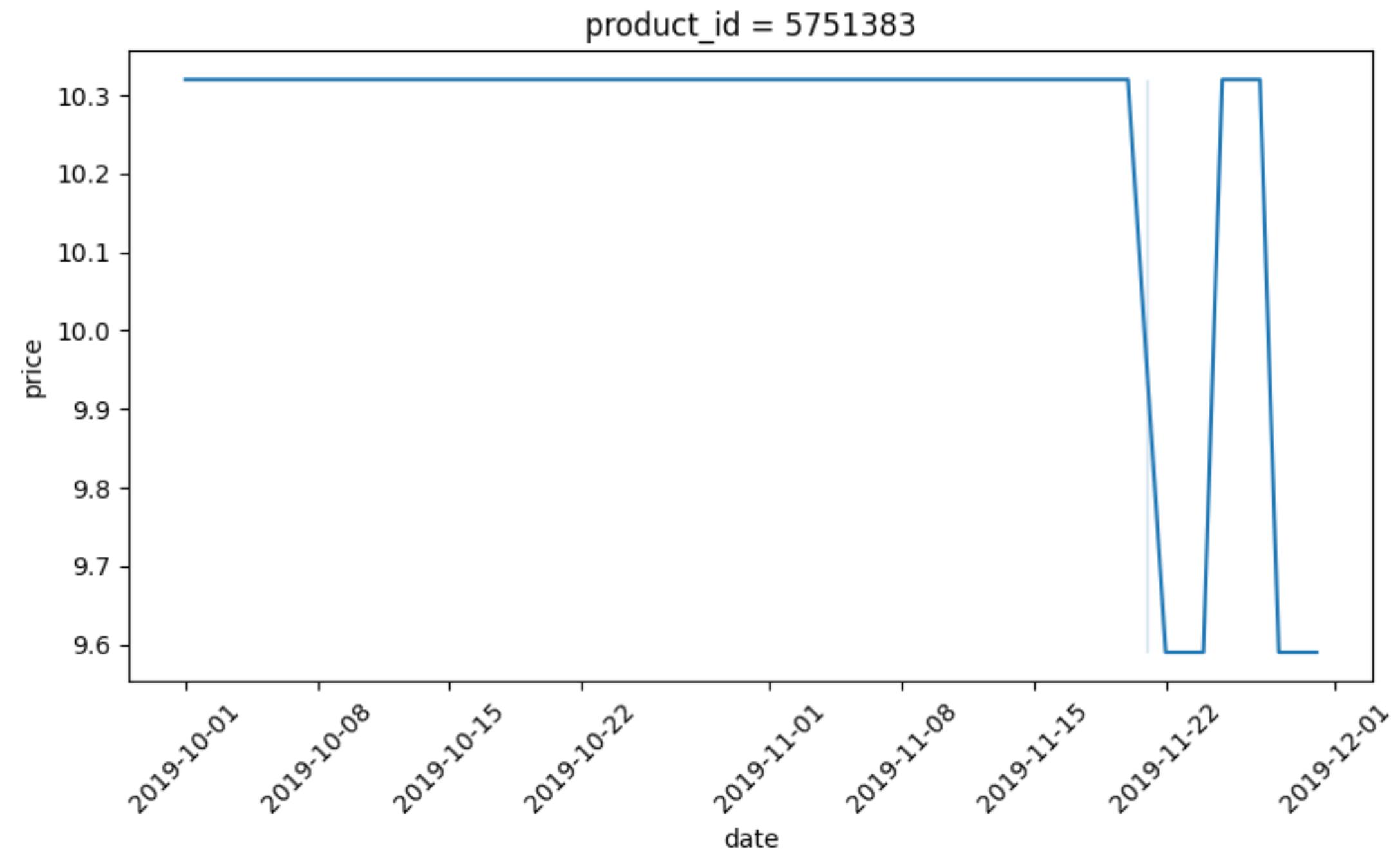
..

Nov-30



# PRICE TRENDS OVER TIME

Almost the same pattern ..

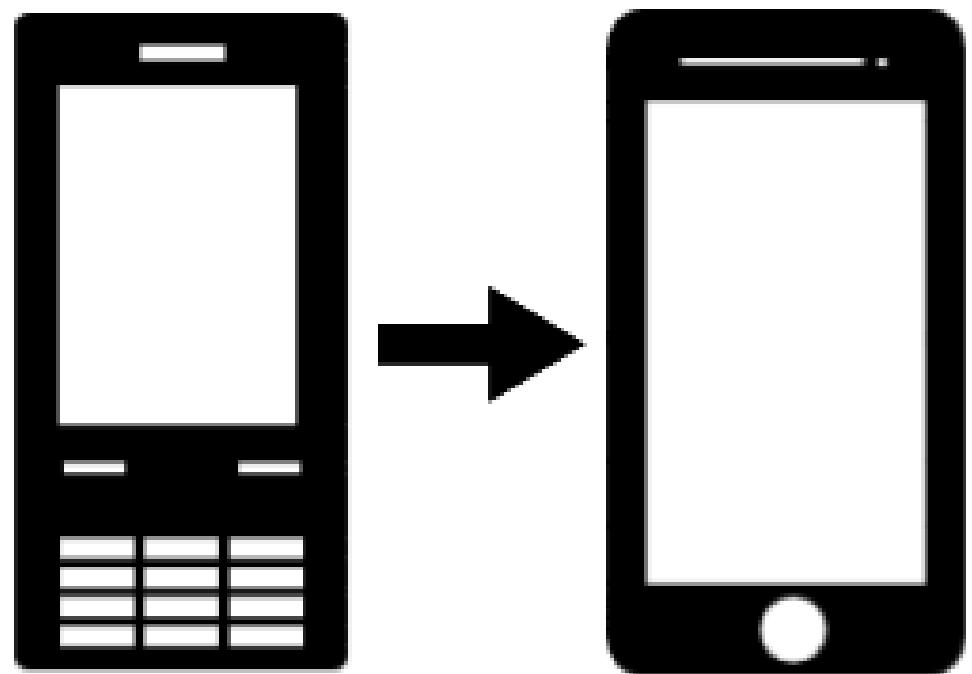


# RECOMMENDATION MODEL

# ASSOCIATION ANALYSIS

antecedents	consequent	antecedents support	consequent support	support	confidence	lift	leverage	conviction
(5743974)	(5749198)	0.041944	0.044757	0.0115	0.2743	6.130662	0.009632	1.316469
(5749198)	(5743974)	0.044757	0.041944	0.0115	0.257143	6.130662	0.009632	1.289691
(5766980)	(5788783)	0.075448	0.029668	0.0112	0.1491	5.027469	0.009015	1.140431
(5788783)	(5766980)	0.029668	0.075448	0.0112	0.3793	5.027469	0.009015	1.489557

# ASSOCIATION ANALYSIS



UPSELL



CROSS SELL

# ASSOCIATION ANALYSIS

Frequently bought together



+



+



Total price: EGP 628.60

Add all three to Cart

# CONCLUSION

- PROBLEM STATEMENT
- OUR OBJECTIVE
- PREPARARTION INSIGHTS
- ANALYSIS AND INSIGHTS
- RECOMMENDATIONS MODEL



# **DO YOU HAVE ANY QUESTIONS?**

## **OUR TEAM**

**SALMA AMR**

**AMR EL SAYEH**

**SALMA EISSA**

**AMIRA AMR**

# THANK YOU

