

Data Analytics: A Customer Segmentation Case Study

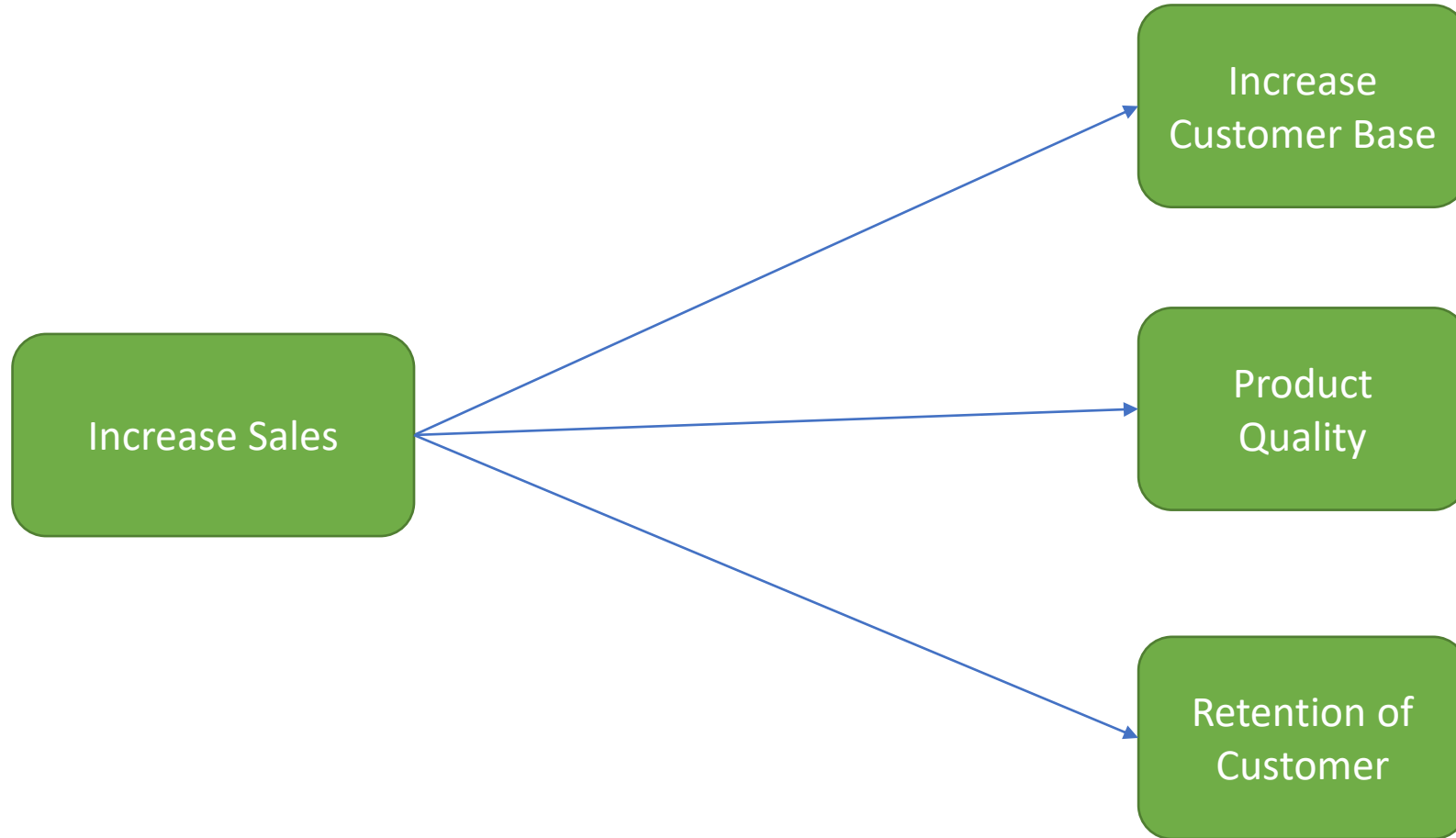
by

Rayala Kartheek

Overview

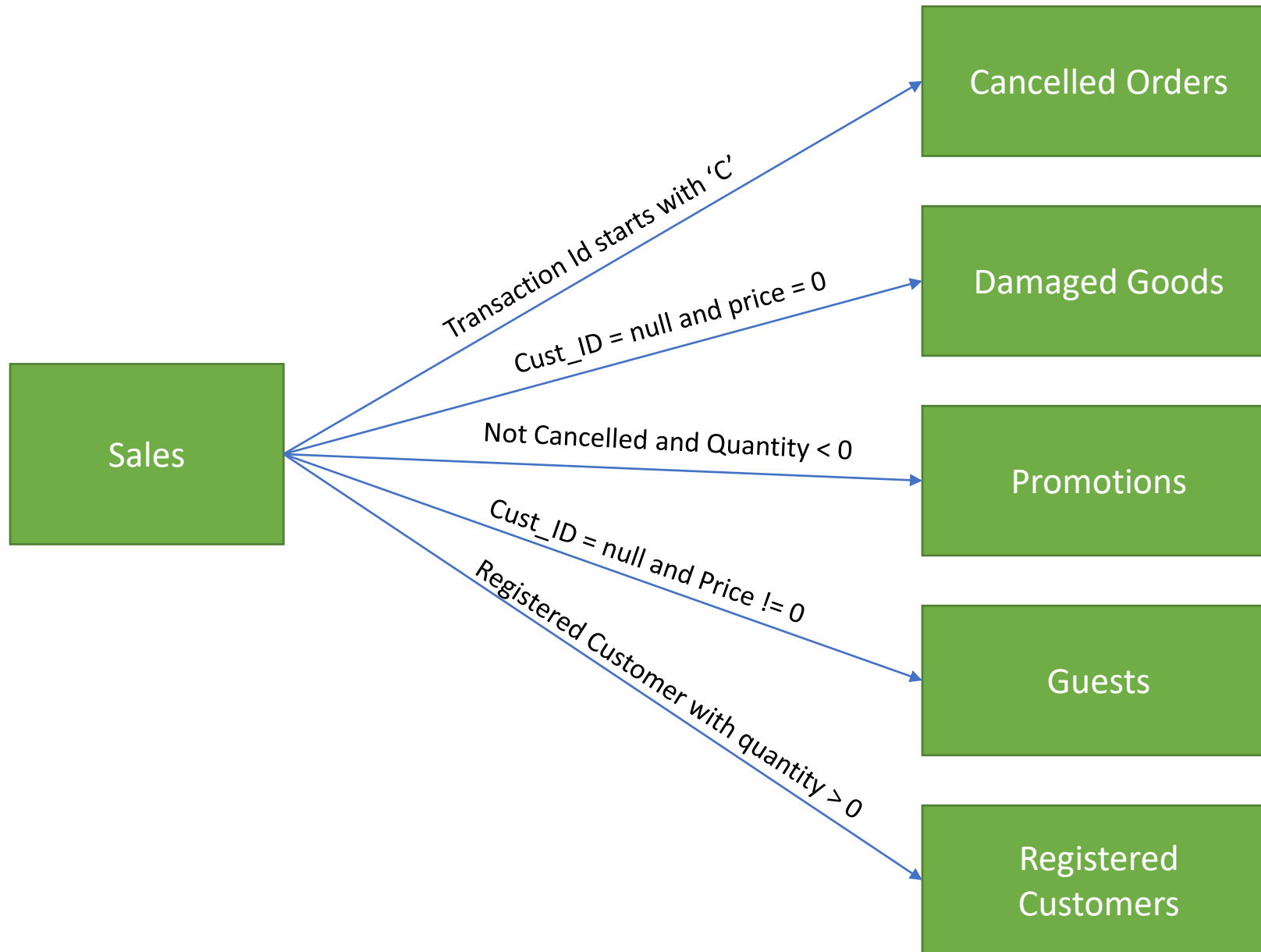
- Total Number of Transactions: 24367
- Number of Registered Customers : 4264
- Number of Products : 4061
- Country With Most Sales : United Kingdom
- Total Cancelled Orders : 3694
- Total Revenue: 10566017

Objective



Assumptions

- If the Customer Id is missing and Unit Price is zero then I assumed them as Damaged goods.
- If the order is not Cancelled order and Quantity is less than zero(which is unusual) I assumed these may be used for Promotions
- If the Customer Id is missing and Unit Price is not zero then I assumed these as Guest(Who is not registered)

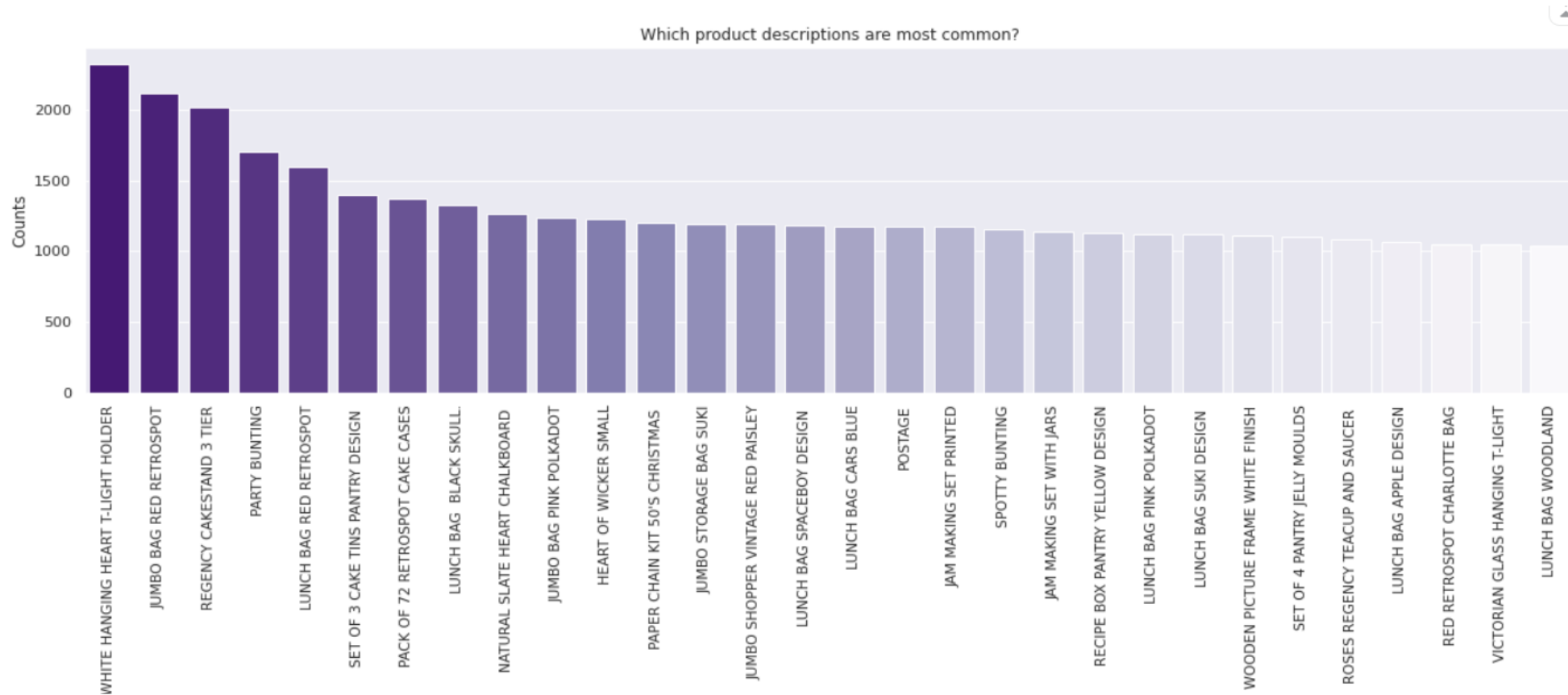


Product Analysis

Overview

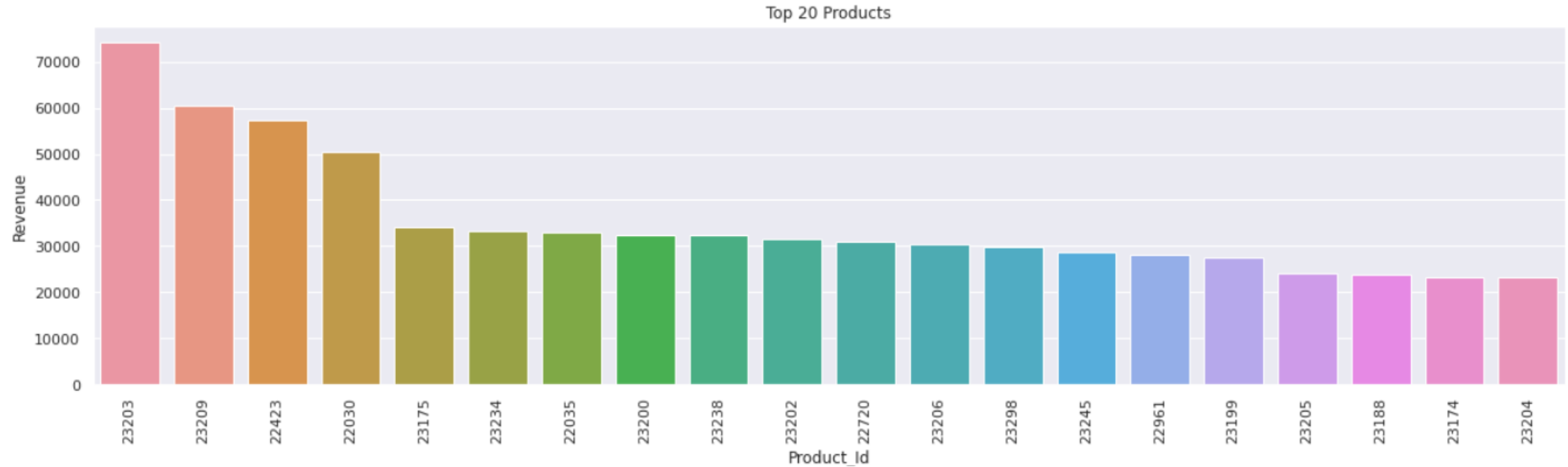
- Total number of Products : 4061
- Number of Products Ordered By registered Customers: 2006
- Number of Products Ordered by Guest: 1556
- Number of Cancelled Products: 2021
- Number of Damaged Products: 89

Top 20 Products on basis of sales



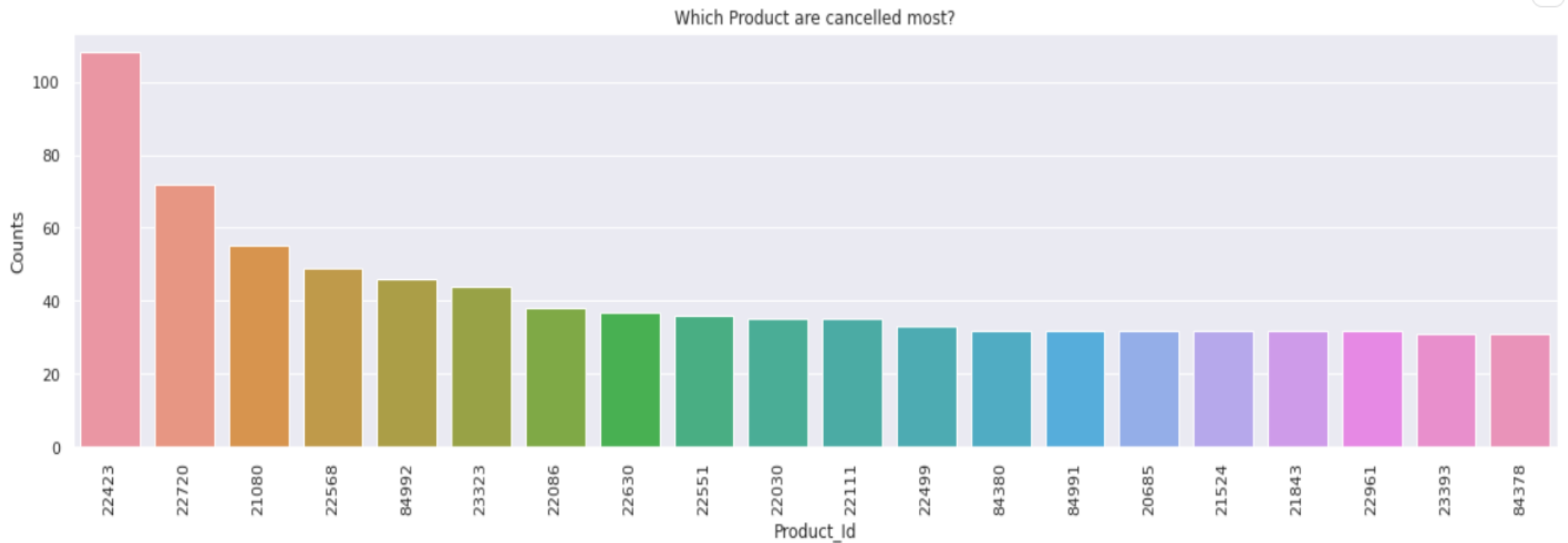
It is Clear from the above bar plot No product is dominating the whole market. Nothing can be inferred from the Bar plot.

Top 20 Products on basis of Revenue



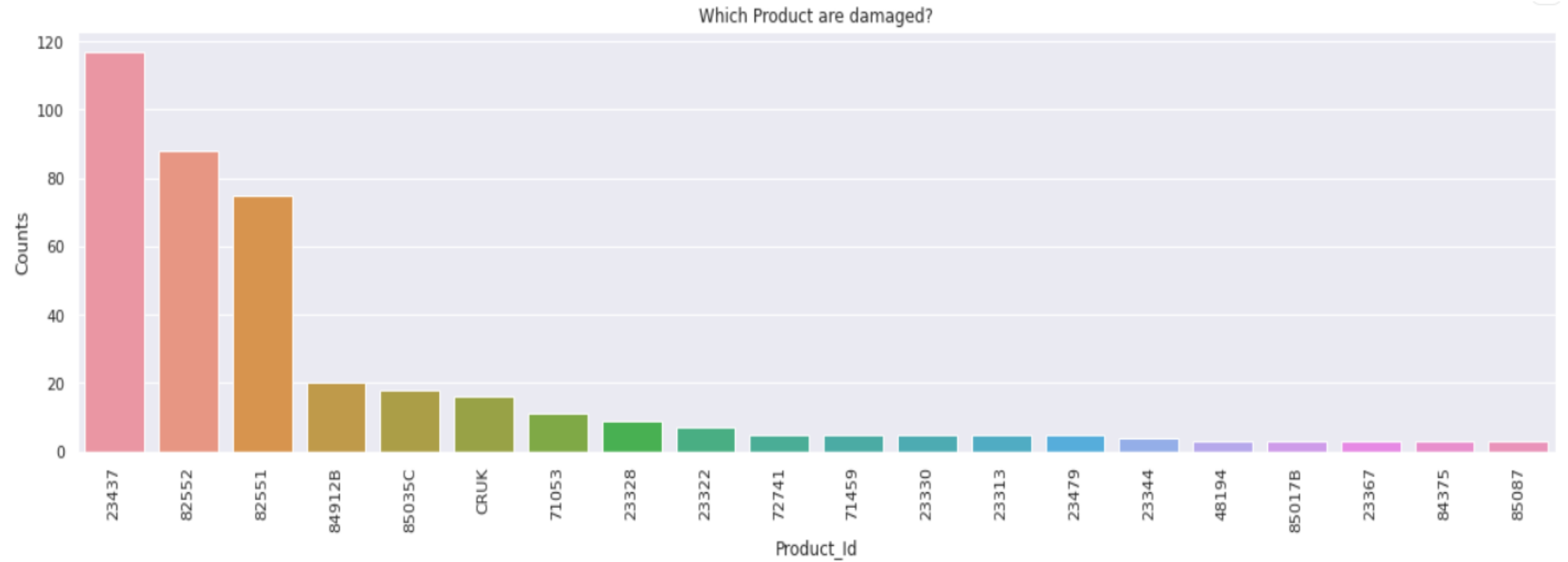
It is clear that top 4 products revenue is more than rest of products. We can Promote by offering discount on these Products

Top 20 Cancelled Products



We need to check the root cause of why the product is cancelled. May be the quality of the product is the issue. We need to check this constantly to get a better idea of product.

Top 20 Damaged Products



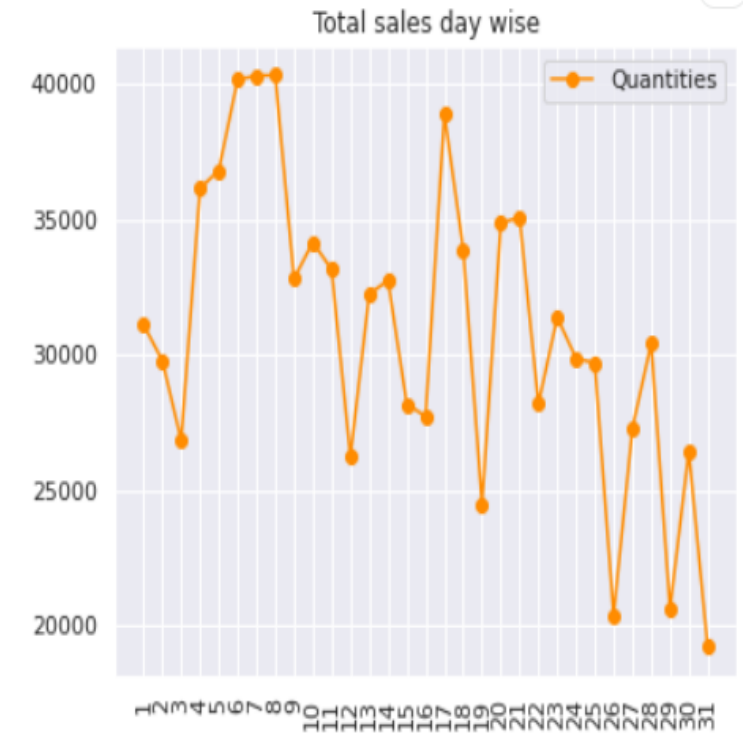
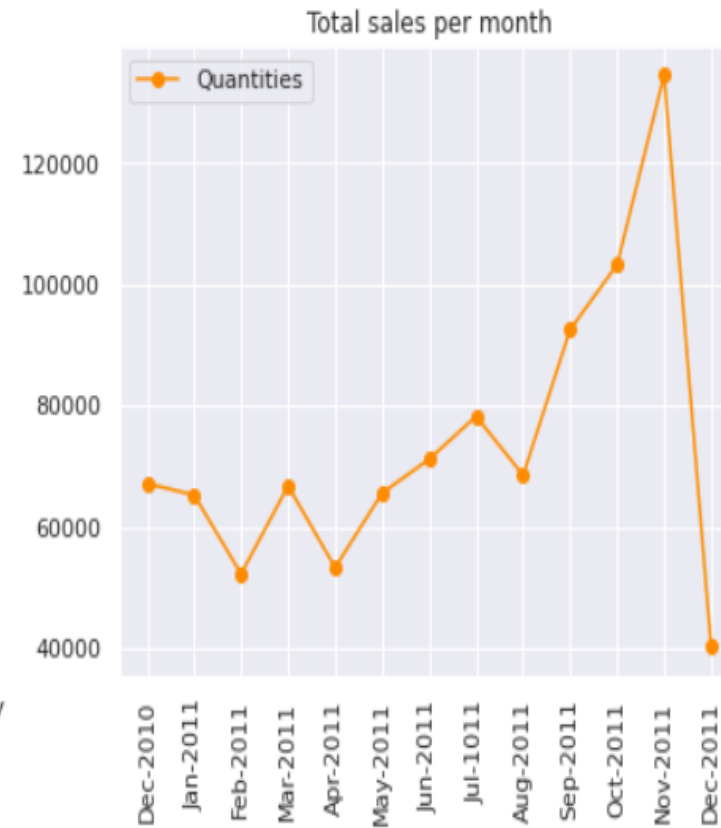
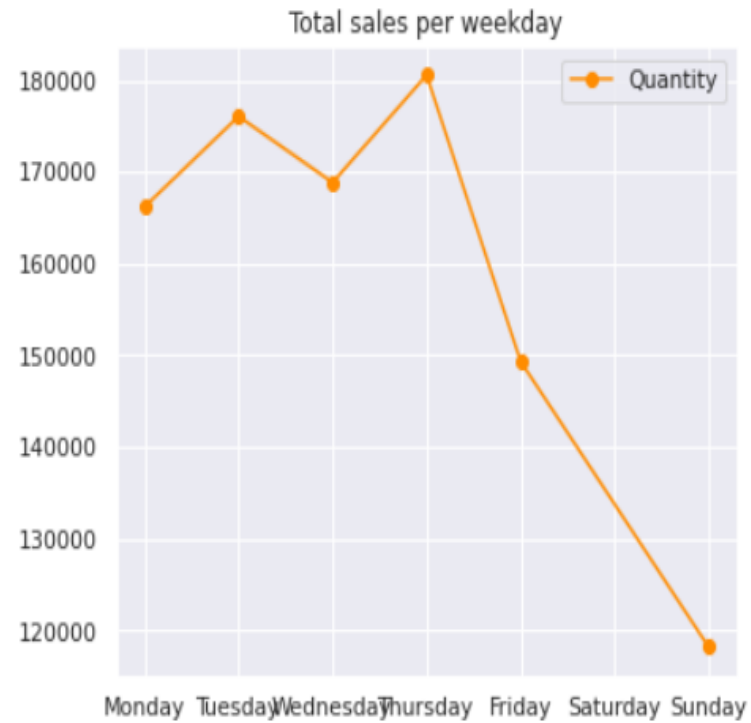
We can drop top 3 damaged products from our product list

Guest Analysis

Overview

- Guest means sales done by unregistered customers.
- Percentage of Guest orders is 24 Percentage. These Guest can be treated as Potential base who can be converted into registered customers.
- Total Number of Transaction is 16174 which is 45% of total transaction ID.
- Revenue from Guest to total revenue ratio is 61 Percentage.

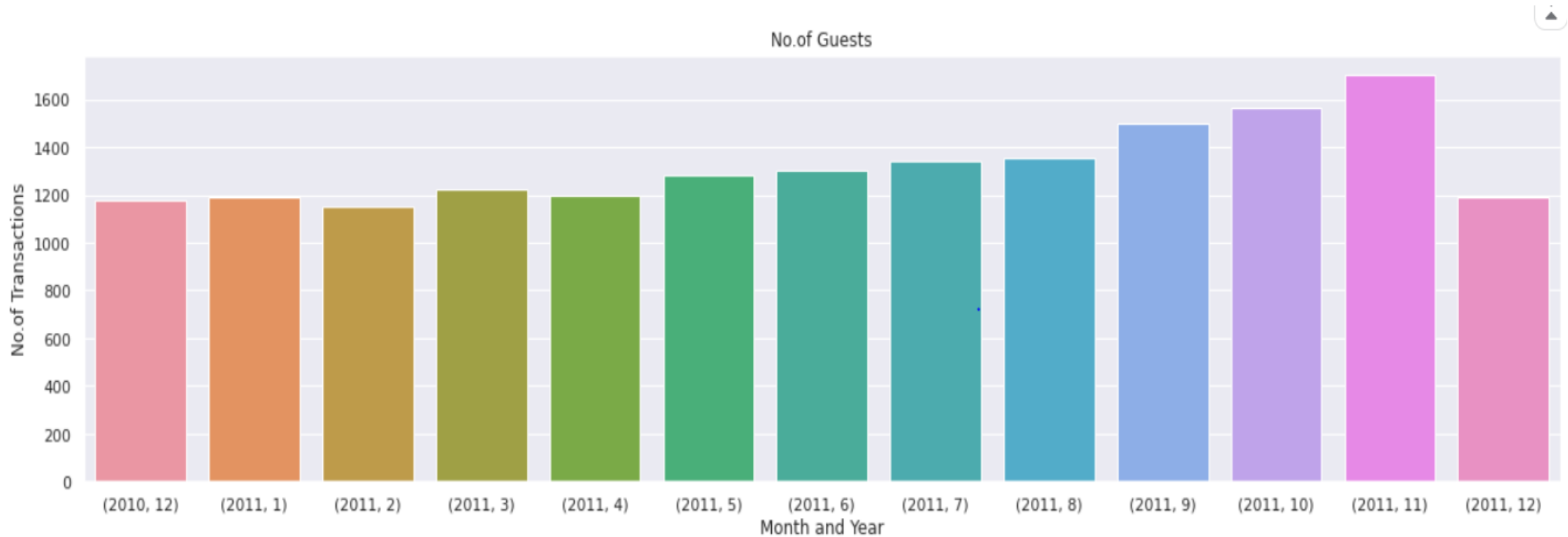
Guest sales vs Time



- In the above graph it is clear that most sales happened on Thursday in weekday.
- Total sales are more in start of month and middle of the month.
- Sales are very high during in November, because of Christmas.

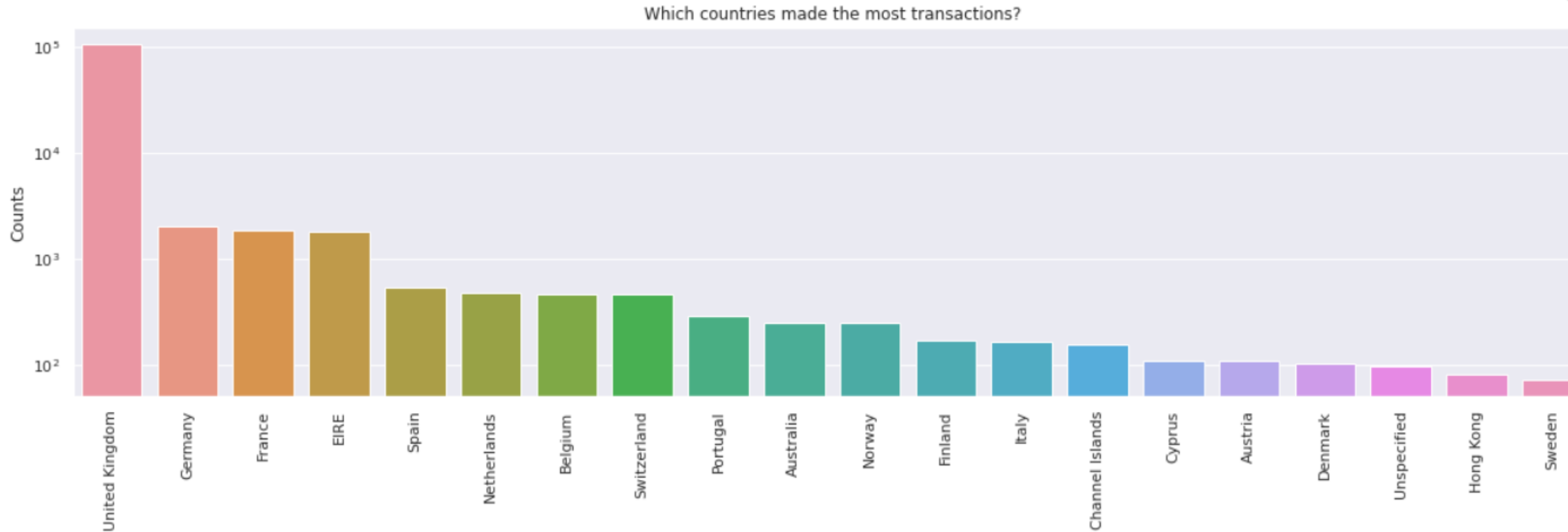
These period should be our target period for promotions to convert Guest into registered customers and to promote the company.

Number of Transactions across Months



There is gradual increase in number of Transactions across the months which is a good indications.

Number of Guest vs Countries



There are lot of Guests from UK. So there is still scope for more registered customers from this area.

Customer Analysis

Overview

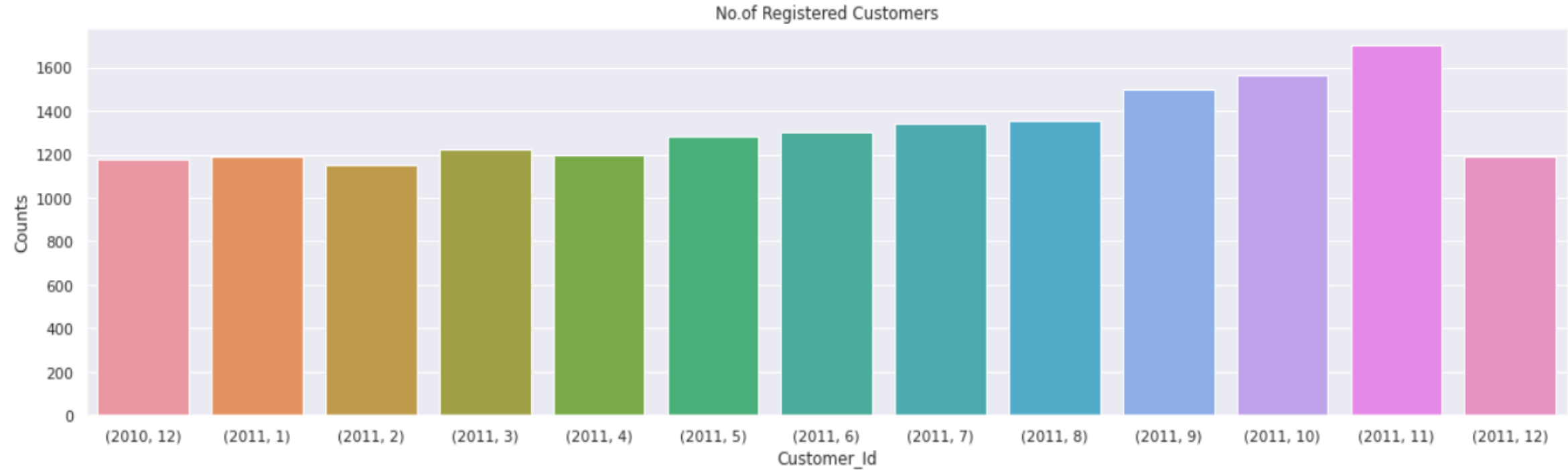
- Number of registered customers: 4264
- Total Number of Transactions is 190309, which contributes about 55 percent of total transactions.
- Revenue from Guest to total revenue ratio is 39 Percentage.

Total number of Products brought by customer



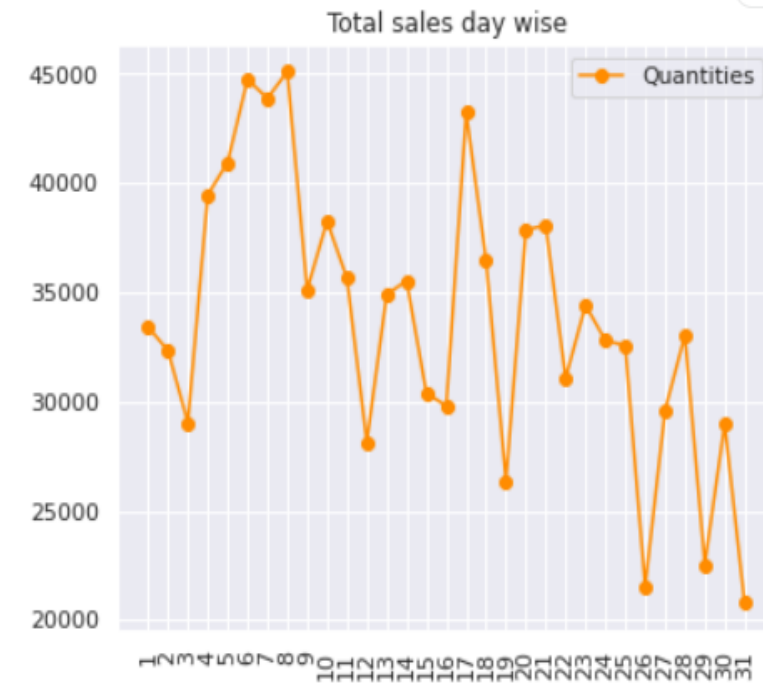
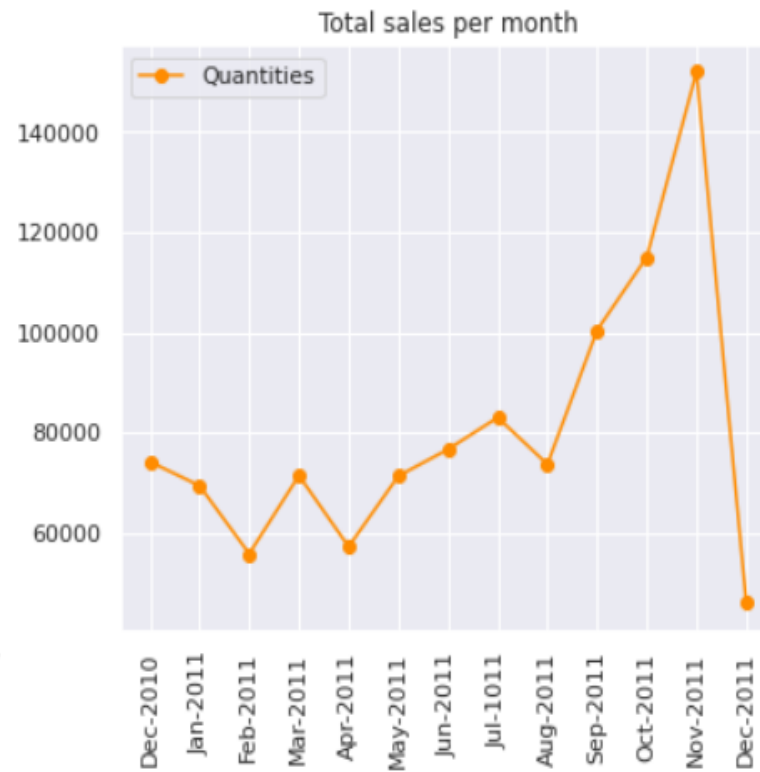
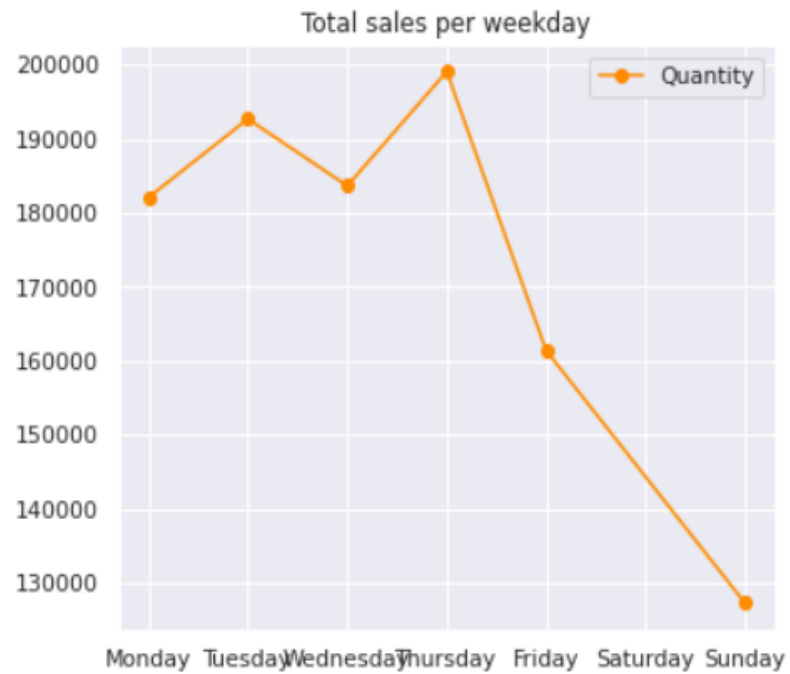
In above plot it is clear that most of customers only brought 1 or 2 products

Number of Customers across months



In the above bar plot it is clear that number of registered customers are Increasing across months.

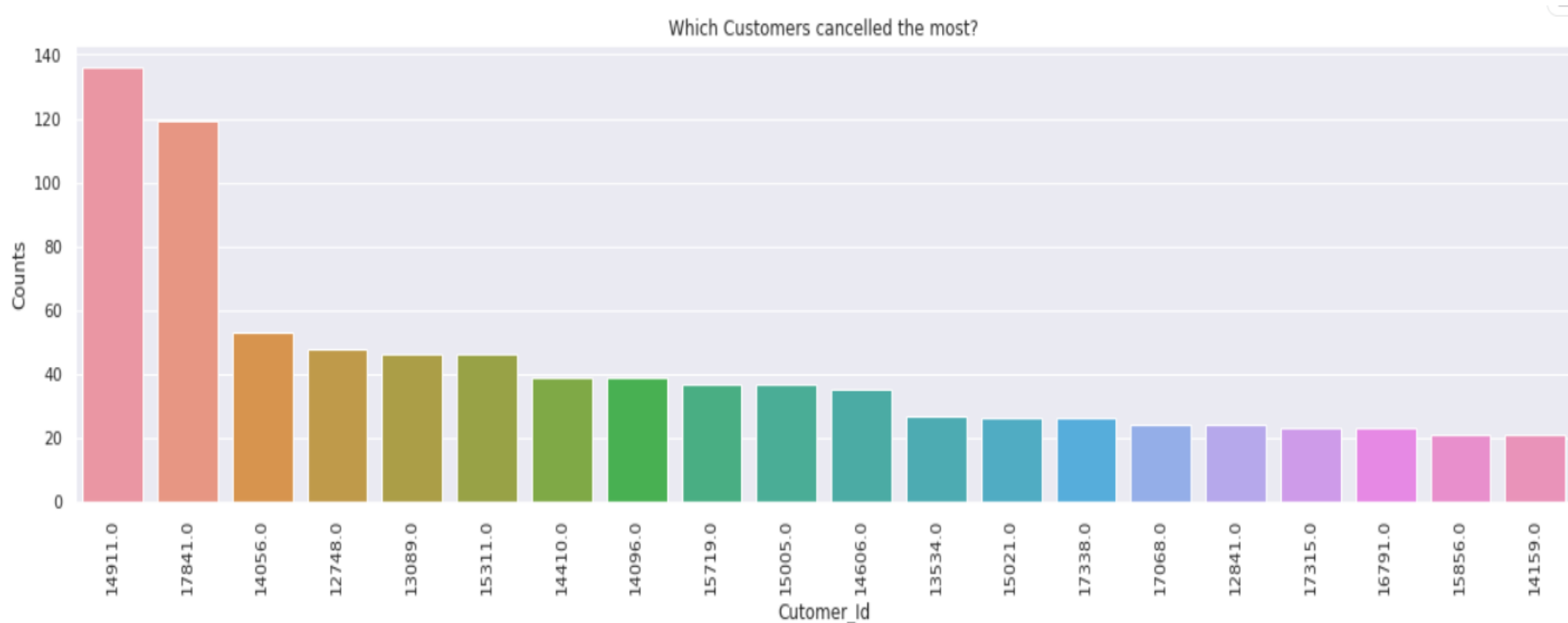
Sales vs Time



These trend is similar to what we have observed for Guest sales. So same strategy is applied for promotion.

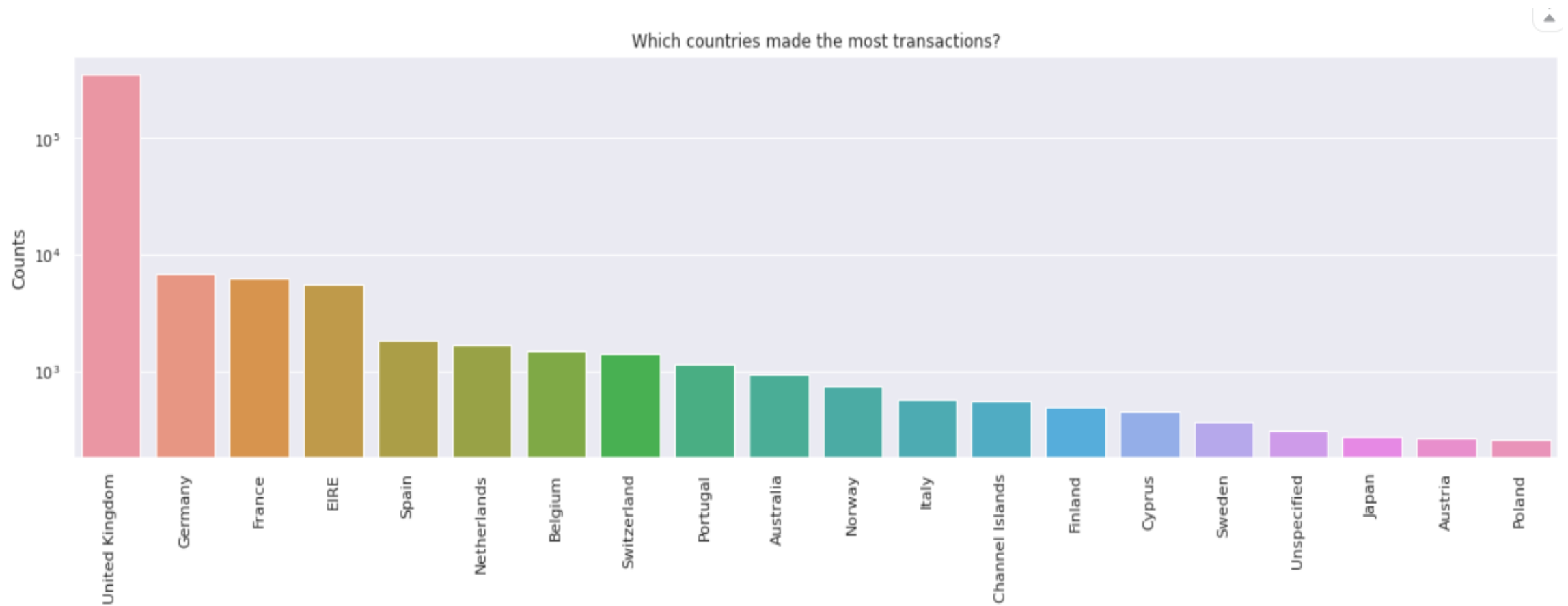
Cancelled orders

- Total percentage of Cancelled Orders : 15%
- It is very high. Lets check for the customers who are cancelling order frequently.



Customer Id 14911, 17841 Cancelled many orders. May be we can call them when they place order or Blacklist them. These metric is very important.

Customer vs Countries



In the above graph it is clear that most customers are mostly from United Kingdom

RFM score for Customer

- Used RFM principle to segregate customers

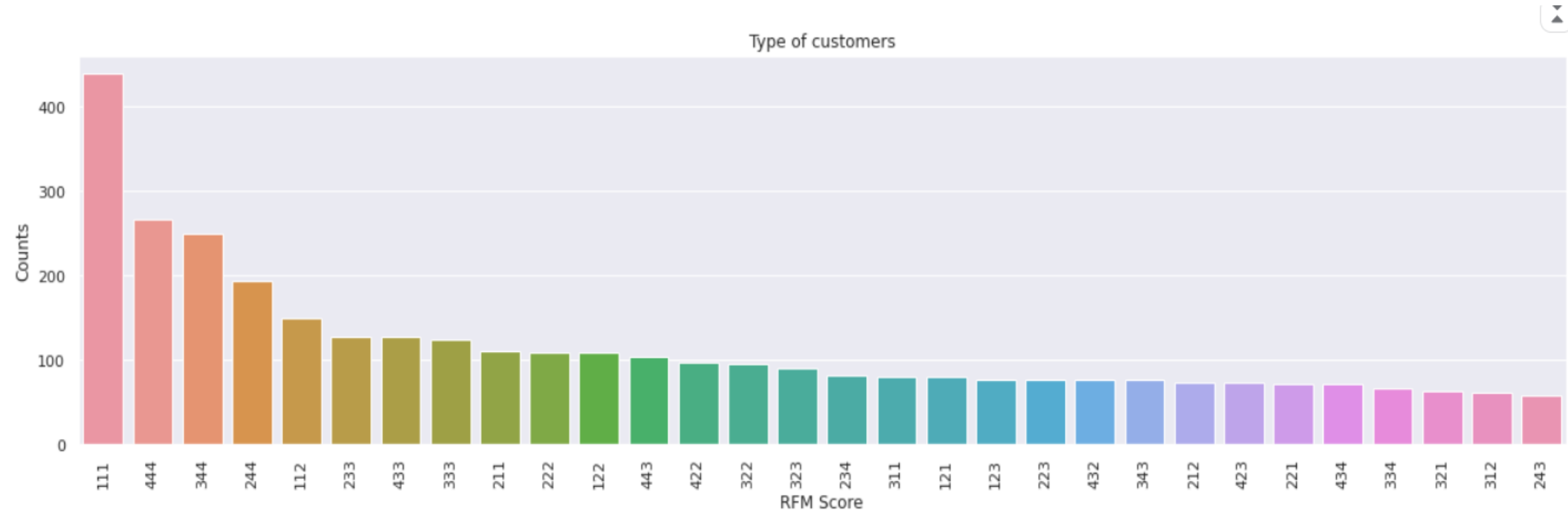
RFM Principle: RFM stands for Recency, Frequency and Monetary. It is a customer segmentation technique that uses past purchase behavior to divide customers into groups.

RFM score consists of three numbers where first number signifies recency, second number signifies frequency and third number signifies Monetary value.

RFM Table

Segment	RFM	Description	Marketing
Best Customers	111	Bought most recently and most often, and spend the most	No price incentives, new products, and loyalty programs
Loyal Customers	X1X	Buy most frequently	Use R and M to further segment
Big Spenders	XX1	Spend the most	Market your most expensive products
Almost Lost	311	Haven't purchased for some time, but purchased frequently and spend the most	Aggressive price incentives
Lost Customers	411	Haven't purchased for some time, but purchased frequently and spend the most	Aggressive price incentives
Lost Cheap Customers	444	Last purchased long ago, purchased few, and spent little	Don't spend too much trying to re-acquire

Number of customers vs RFM Score



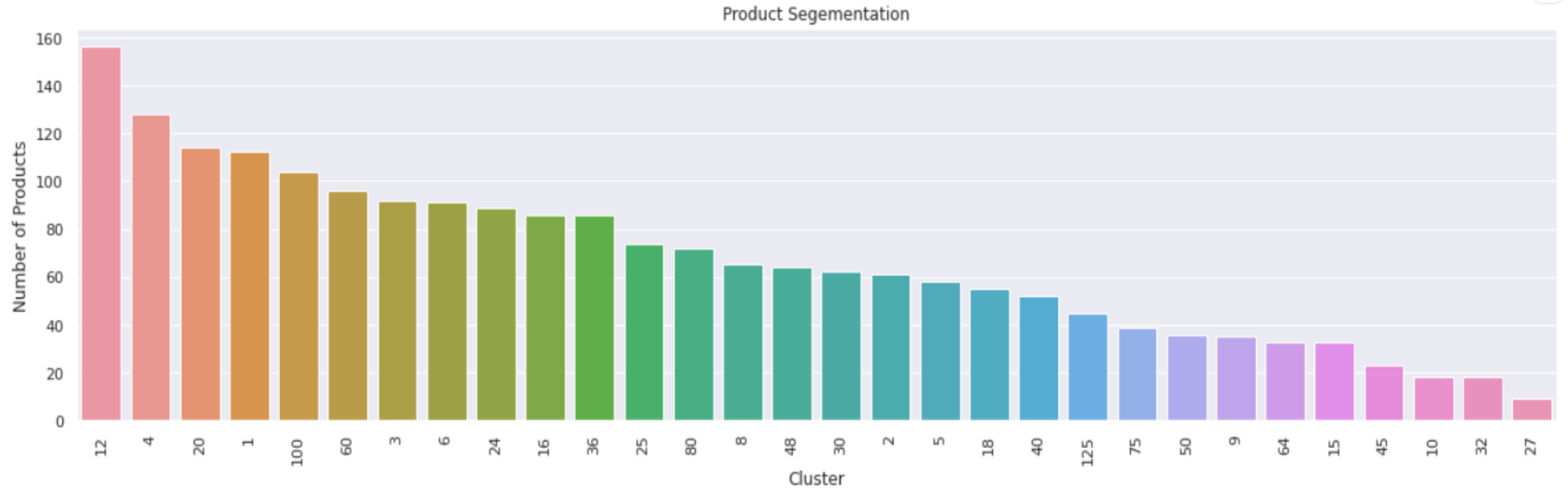
The Best Customers are more compared to all categories.

Product Segmentation

Principle

- Taking similar approach as RFM score I obtained a score called OTM score. Where 'O' stands for number of orders, 'T' stands for total quantity of product ordered, 'M' stands for median of price of product.
- I obtained 'O' score, 'M' score, 'T' score and multiplied each score to obtain OTM value.
- Based on the score I was able to get 30 cluster of products

Number of Products in clusters

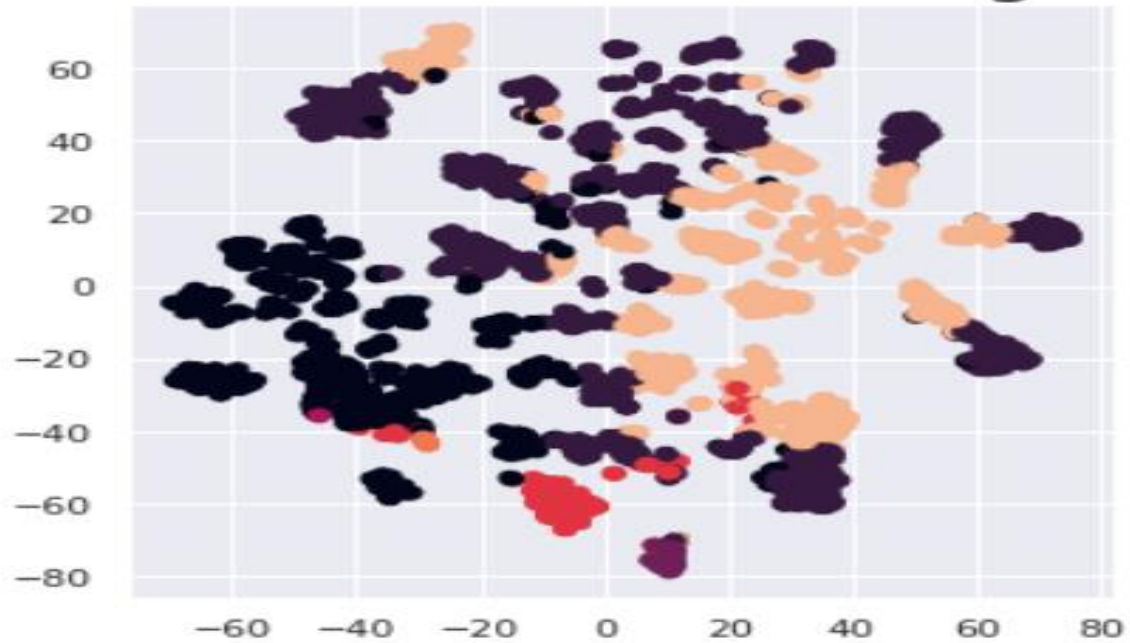


Customer Segmentation

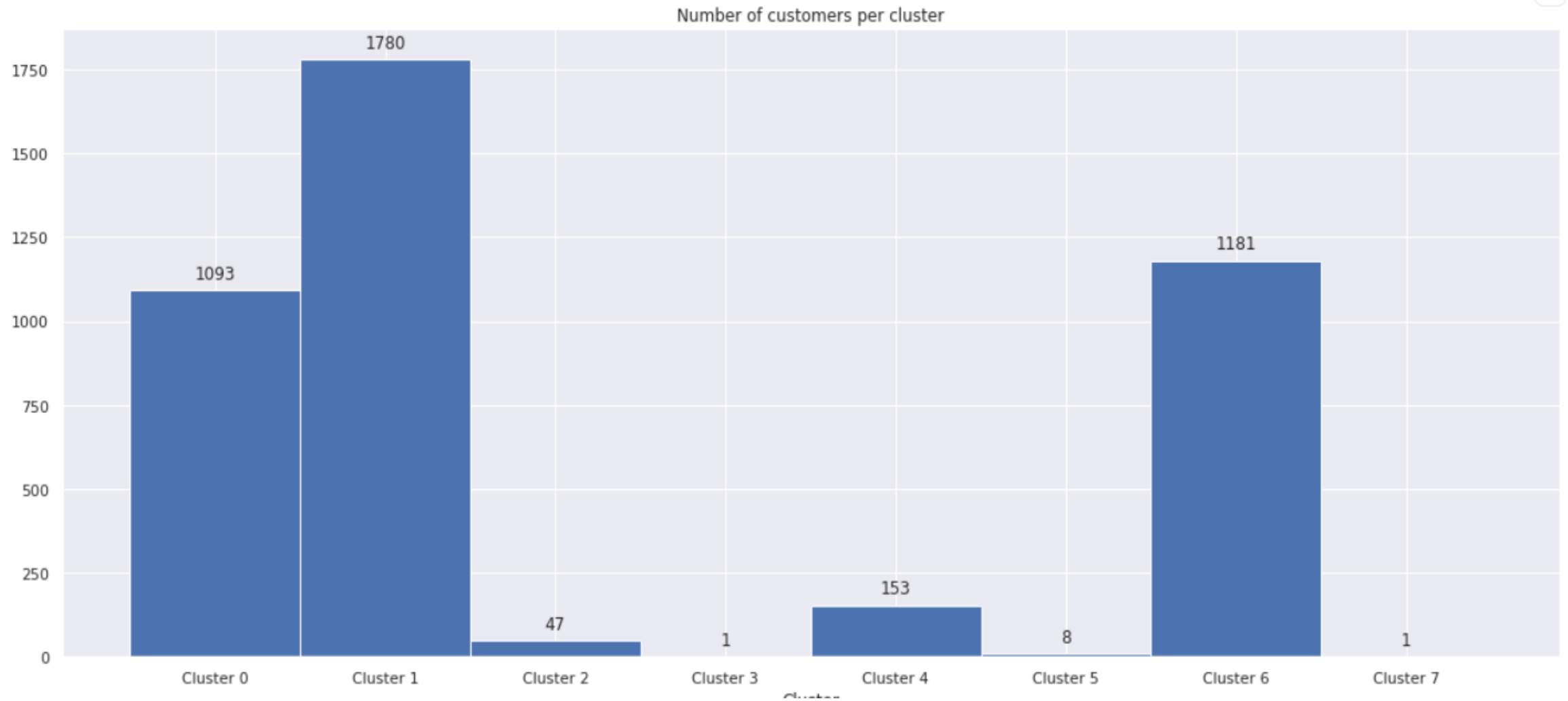
- Used K Mean clustering algorithm for customer segmentation.
- Used Silhouette Score metric to obtain minimum number of clusters.
- We want to have at least 5, 6 clusters so we won't take 2 or 3 clusters even though they have the highest silhouette scores, 8 clusters would fit the best here.

Visualization of Clusters with TSNE

Visualization of the clustering with TSNE

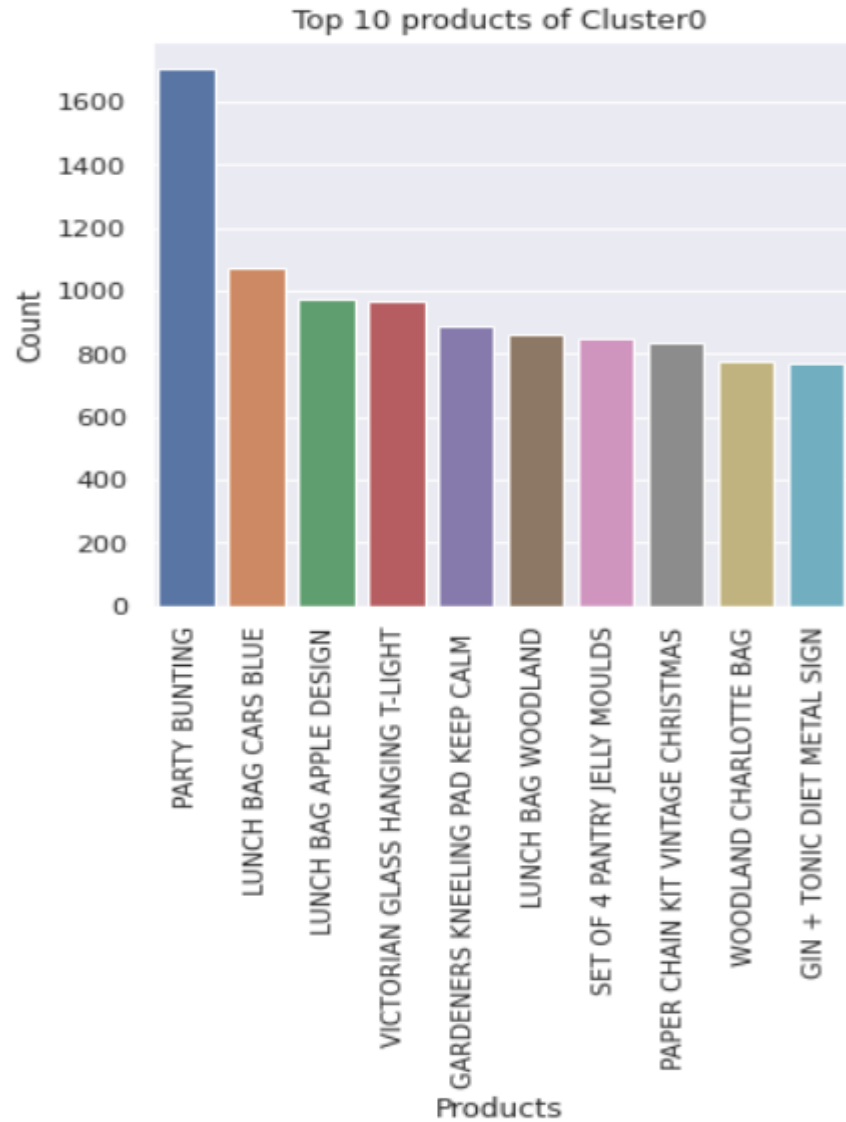


No.of People for clusters



It looks like most people belong to 3 clusters

Top 10 Products of Cluster0

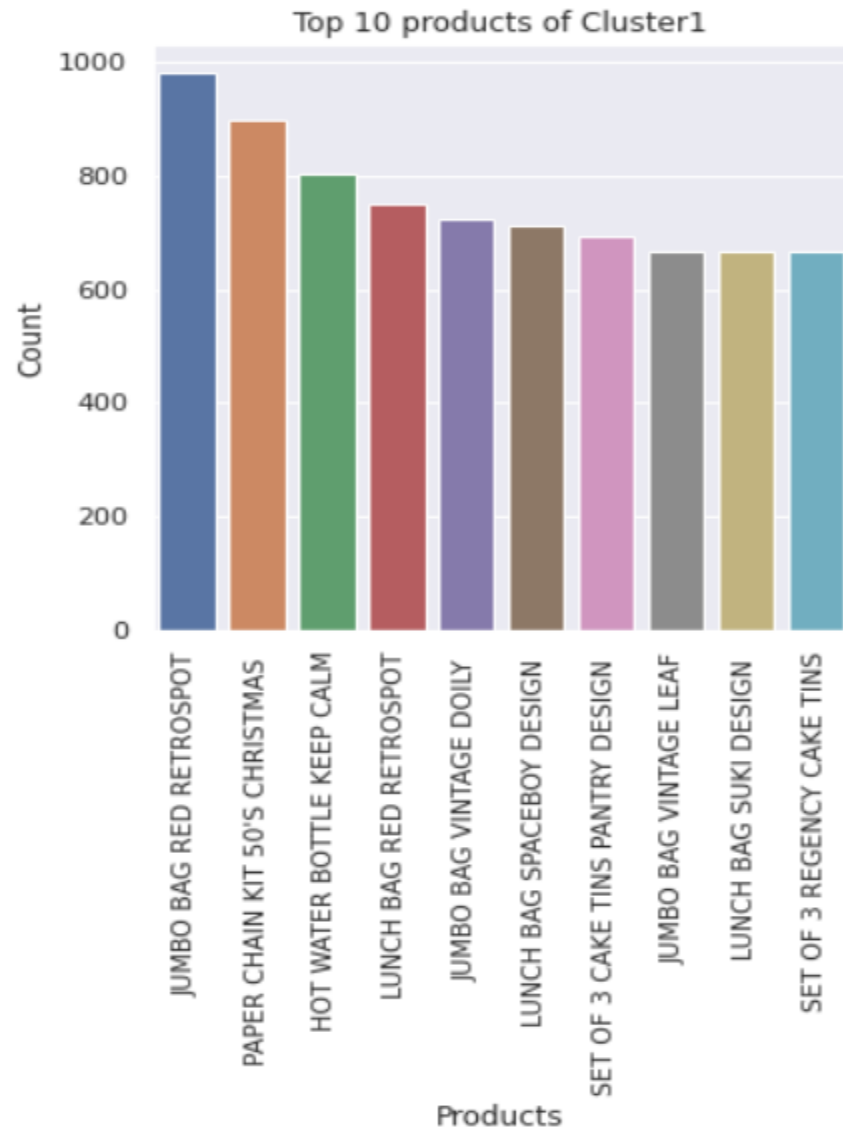


Top 10 RFM Score of Cluster0

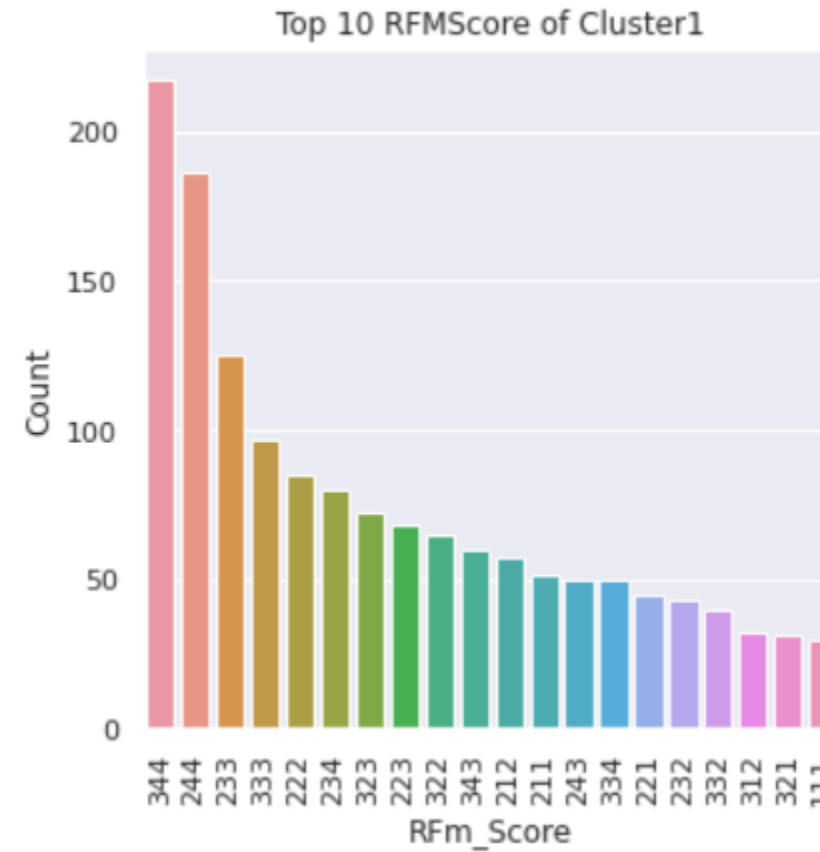


The RFM Score is mostly 111, 112, 122, 123

Top 10 Products of Cluster1

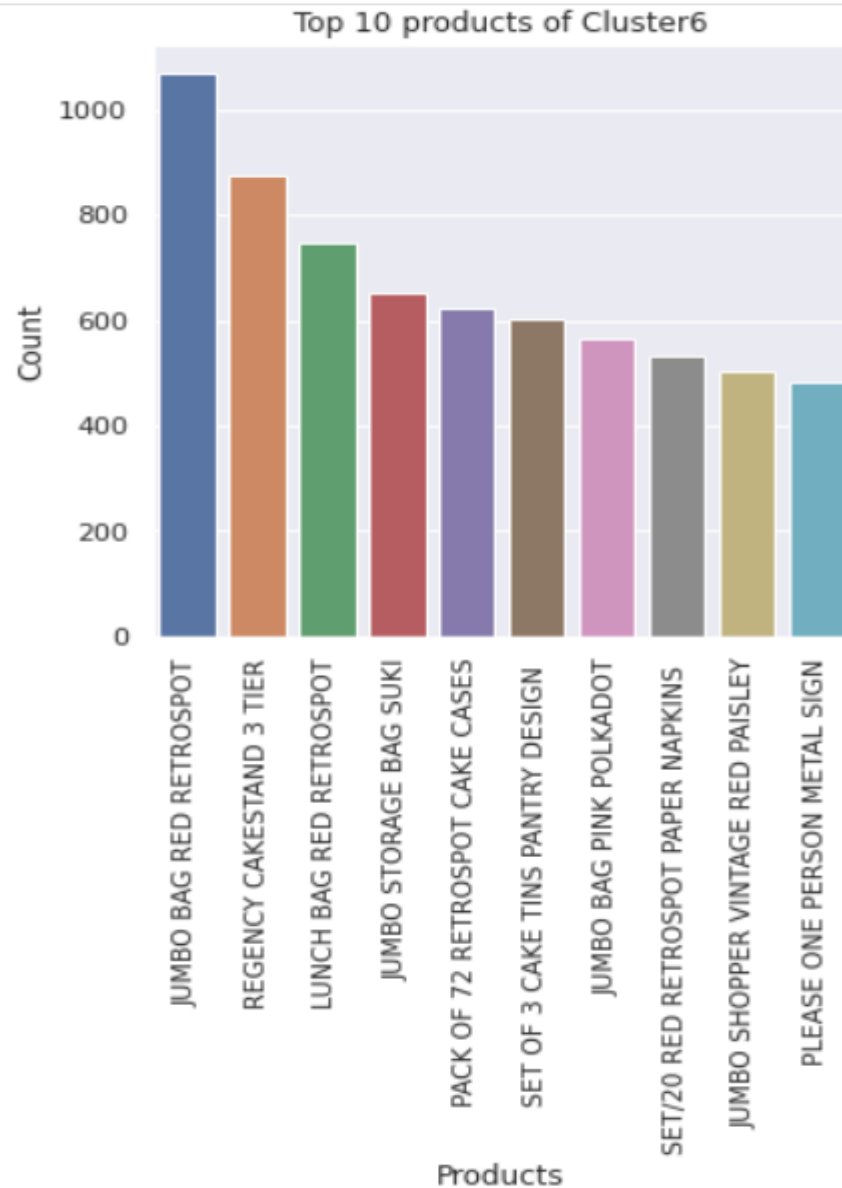


Top 10 RFM Score of Cluster1

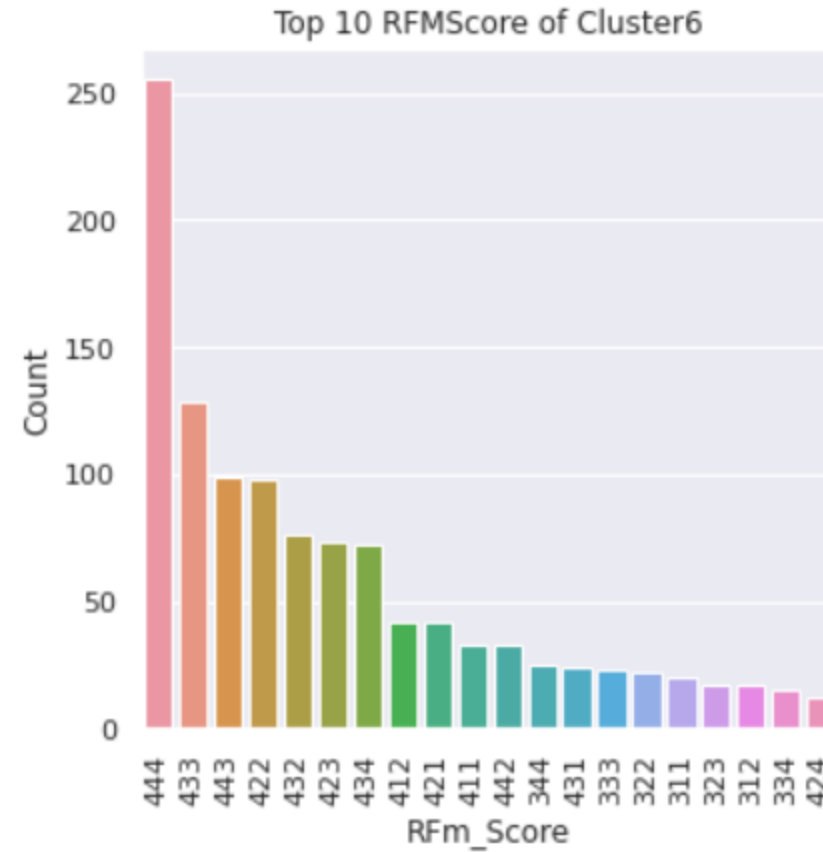


The RFM Score is mostly 344, 244, 233, 333

Top 10 Product of Cluster6



Top 10 RFM Score of Cluster6



The RFM Score is mostly 444, 433, 443, 442

Promotions

- The data with not cancelled orders and quantity less than 0 or unit price is 0.
- The amount spend on promotion is 7 percent of total revenue which is very less. If we spent more money on promotions there may be a chance of guest converting into registered customers.

Conclusion:

Clusters

- Cluster 0: high frequency with a lot of quantity, bought on average and high monetary value (VIP clients)
- Cluster 1: Aggressive market strategy is required for this cluster of customers.
- Cluster 6: Customers whose frequency is Low, Monetary value is low comes in this category. It is better not to concentrate on this cluster of customers.

Promotions

- It is best to when the flow of customers is very high.
- In weekdays it is better to have offers on Thursday. In month wise it is better to have on November for better reach

Here the revenue from guest is high compared to registered customers. So there is lot of scope to convert guest into best customers. With Promotions it is achievable.

Thank You