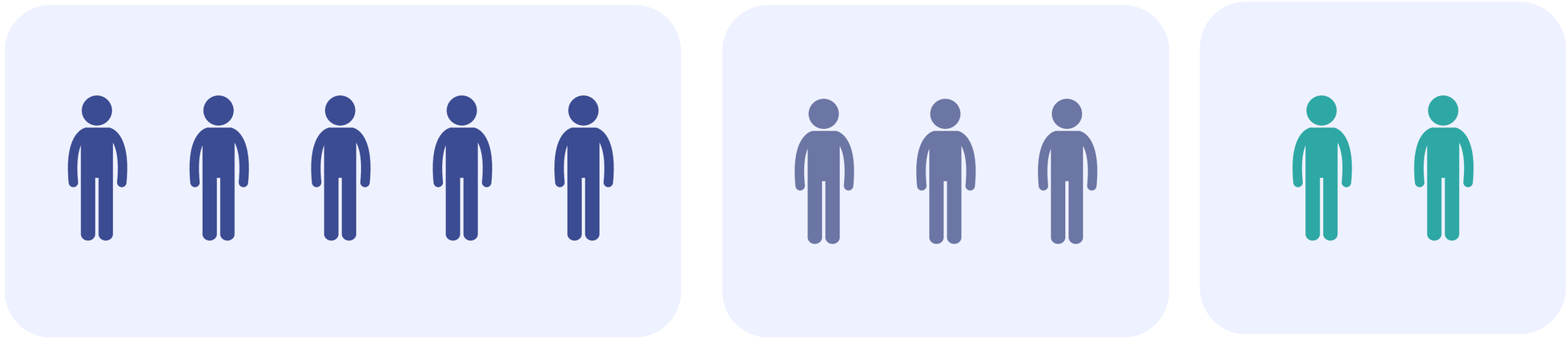
Maximizing Customer Value:

RFM Segmentation and Churn Analysis



Widiareta Safitri

wretasafitri33@gmail.com

Table of Content

01

Problem Research

1 | Problem Research

Problem Background

In today's competitive business landscape, understanding and retaining customers is critical. **Customer churn, or the rate at which customers stop doing business with a company,** poses significant challenges for organizations in various industries such as having to look for new customers instead of having loyal customers which will incur more costs. To overcome this challenge, businesses can utilize data-driven strategies such as RFM Customer Segmentation.

RFM segmentation is a **data-driven technique used to categorize customers based on three key factors: Recency (R), Recency (R), Monetary (M).** By segmenting customers based on these metrics, businesses can gain valuable insights into customer behavior and then can correlate that customer behavior with customer churn.

1 | Problem Research

Objective

Segmenting Customers into Four Groups Using the RFM Approach

Identifying Potential Churn Customers Based on Customer Segmentation

Analyze Churn Probability from Customer

1 | Problem Research

Data Sources

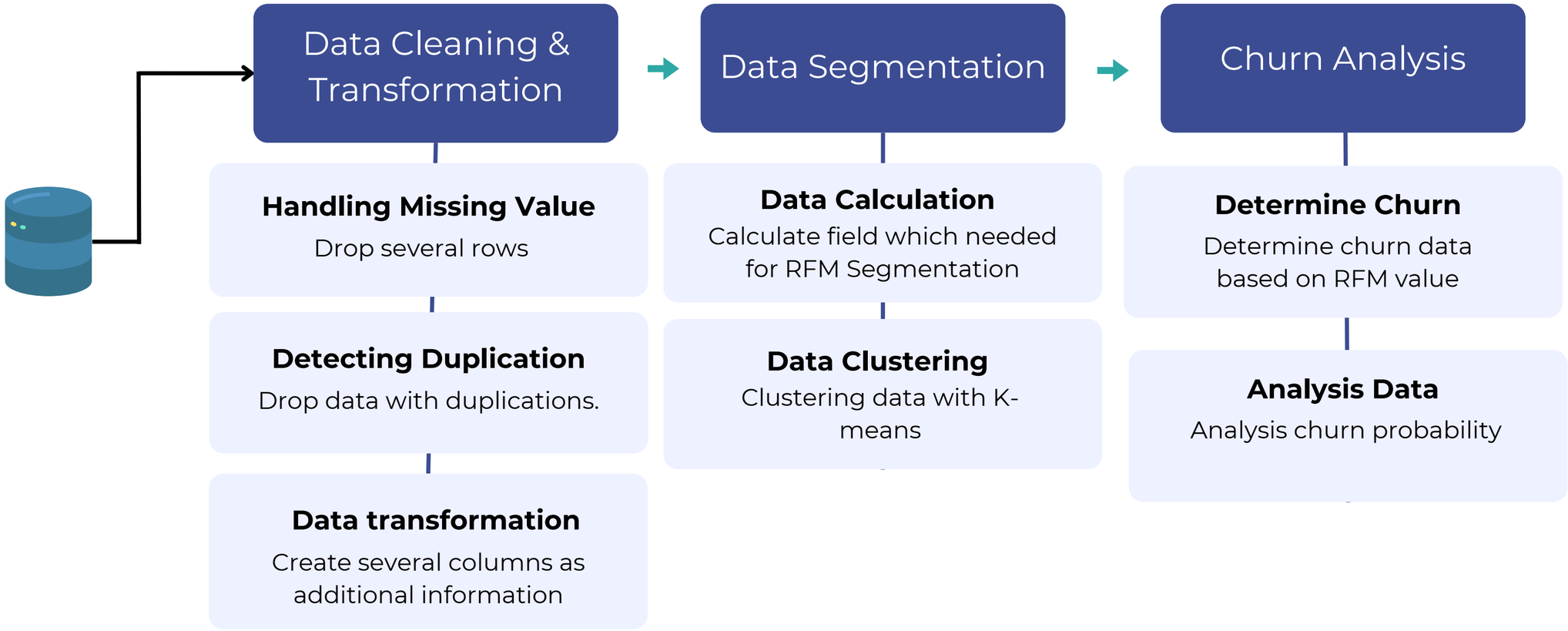
Online Retail Dataset.csv

Column	Explanation
InvoiceNo	Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
StockCode	Product (item) code. a 5-digit integral number uniquely assigned to each distinct product.
Description	Product (item) name. Nominal.
Quantity	The quantities of each product (item) per transaction. Numeric.
InvoiceDate	Invoice Date and time. Numeric, the day and time when each transaction was generated.
UnitPrice	Unit price. Numeric, Product price per unit in sterling.
CustomerID	Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
Country	Country name. Nominal, the name of the country where each customer resides.

02

Data Preprocessing

Data Preprocessing

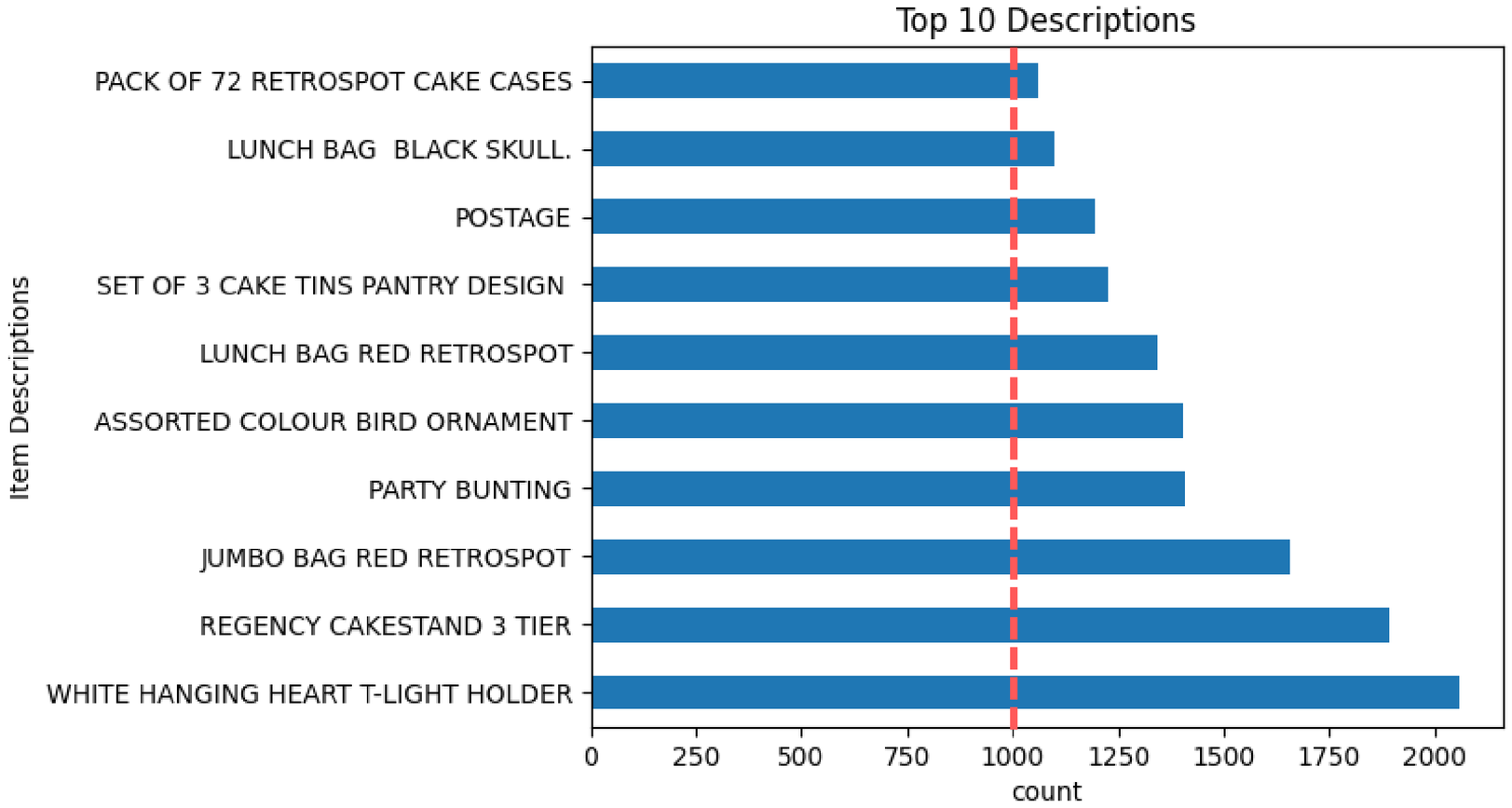


03

Business Insight

3 | Business Insight

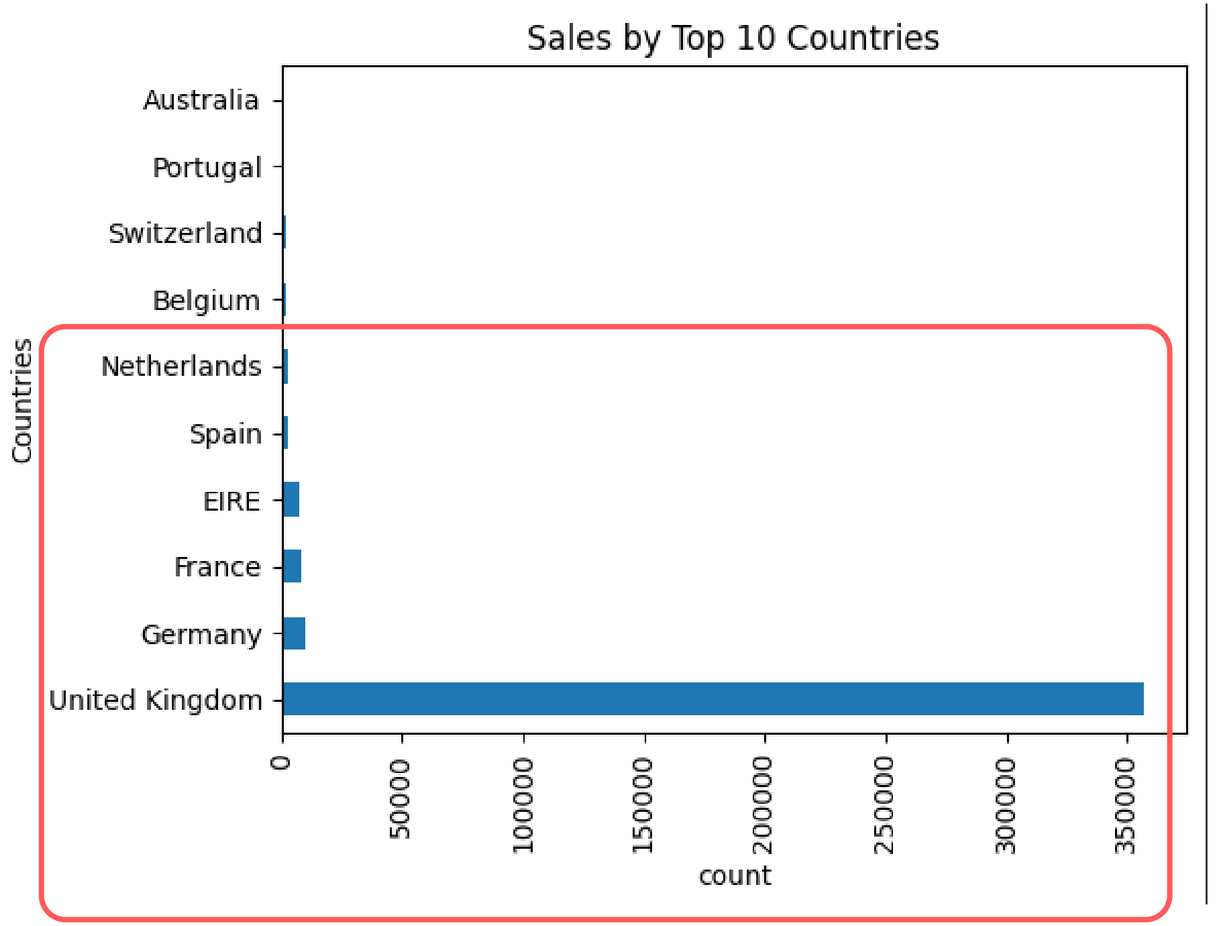
Which items recorded as the best-selling



- The best-seeling item was “**White Hanging Heart T-Light Holder**”
- The 10 best-selling products all have sales of more **than 1000 sales** but not more than 2500 sales.

3 | Business Insight

Which Country had the highest sales?

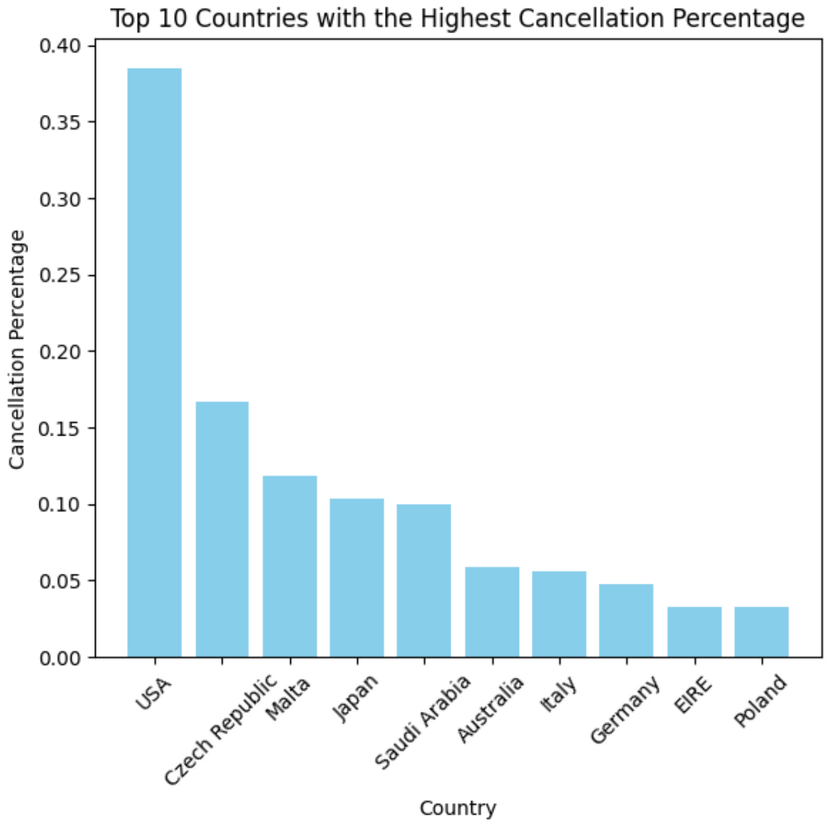


- It can be clearly seen that the main market for sales is in the **United Kingdom** with the total of transaction more than **350000** transactions.
- The most transactions other than in the United Kingdom are in European Union countries such as Germany, France, IERE (Ireland), Spain and the Netherlands

3 | Business Insight

Top 10 Countries with the Highest Cancellation Percentage

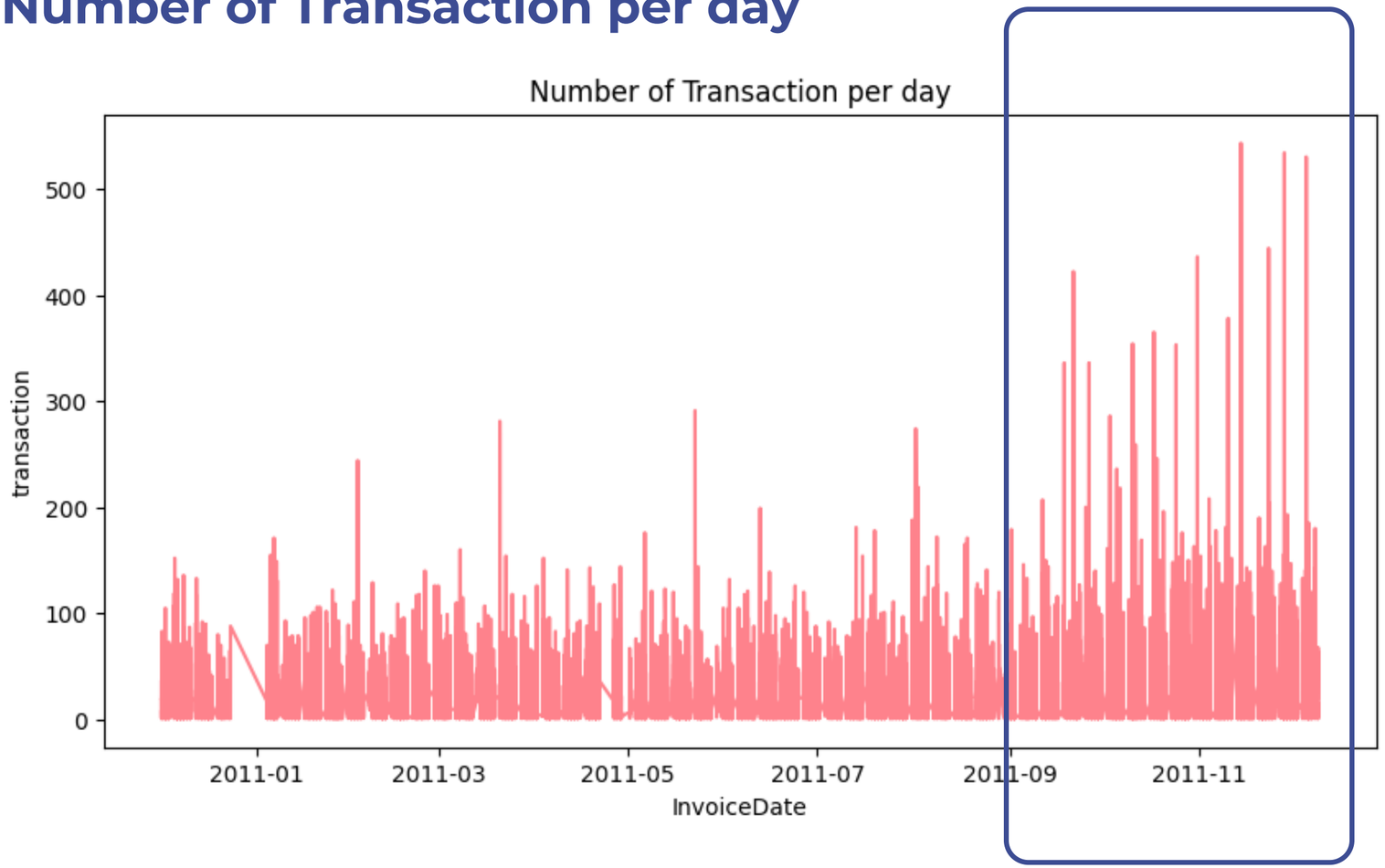
Country	order_canceled		
	sum	count	perc
USA	112	291	0.384880
Czech Republic	5	30	0.166667
Malta	15	127	0.118110
Japan	37	358	0.103352
Saudi Arabia	1	10	0.100000
Australia	74	1258	0.058824
Italy	45	803	0.056040
Germany	453	9480	0.047785
EIRE	247	7475	0.033043
Poland	11	341	0.032258



- In the return data, it was found that not the **USA** is the country with the largest return presentation
- Most cancelled transactions are outside the UK and EU.

3 | Business Insight

Number of Transaction per day



- Transactions per day have **increased in the last 3 months**, from September 2011 to November 2011

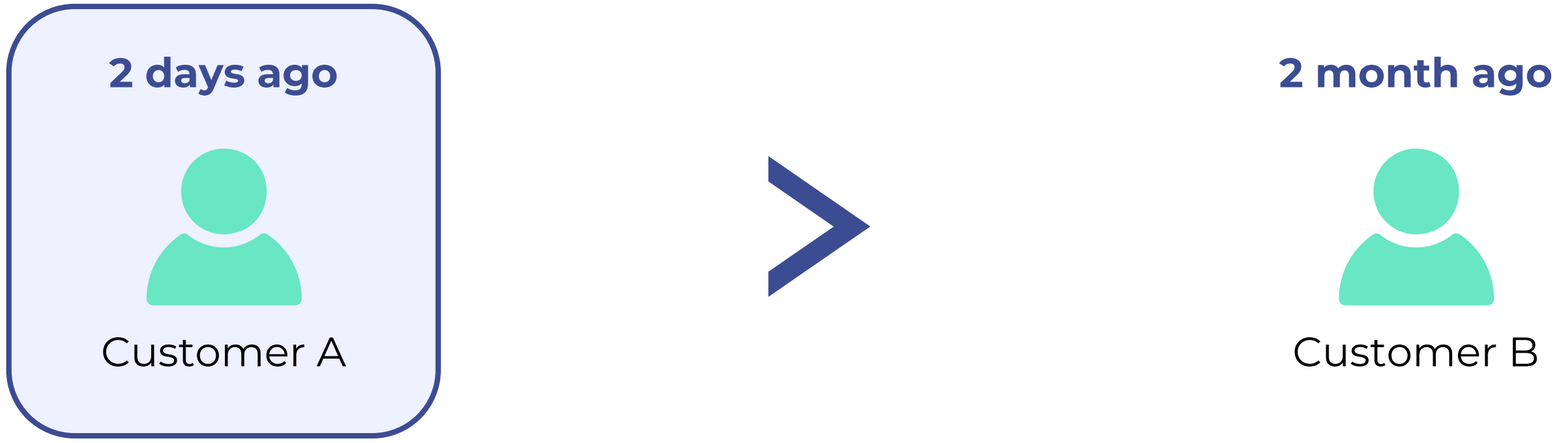
04

RFM Segmentation

4 | RFM Segmentation

Recency : The time elapsed since their last purchase

To determine the recent activity of a customer, we must accurately establish the last time each user was observed making a purchase. The next step is to assess and assign a recency score to each customer based on the time of their last purchase.

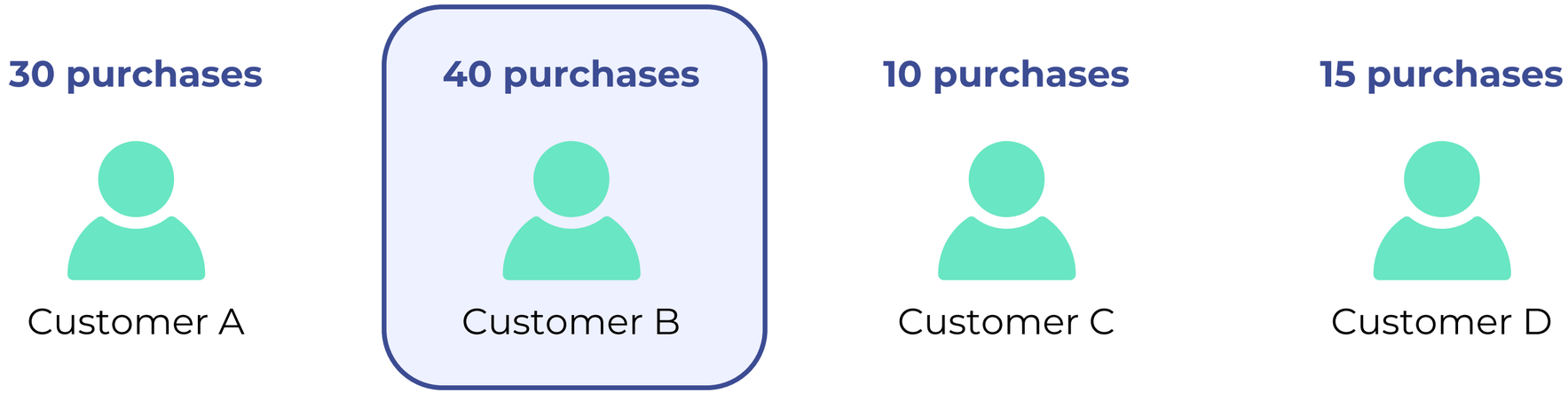


Customer A will **have higher recency value** because he made transaction **2 days ago** compare to customer B who made a **transaction 2 months ago.**

4 | RFM Segmentation

Frequency : How frequently customer make purchases

To calculate Frequency, which represents how many times each customer has made a purchase on the platform, we utilize a group by operation on customerID and Date to obtain the frequency values



Customer B will **have highest frequency value** among others customer because he has the **most purchase history**

4 | RFM Segmentation

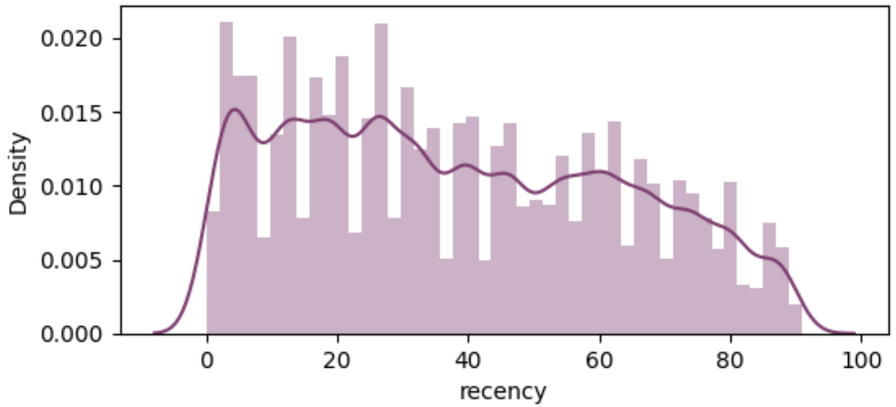
Monetary Value : The average amount they spend during their purchases.

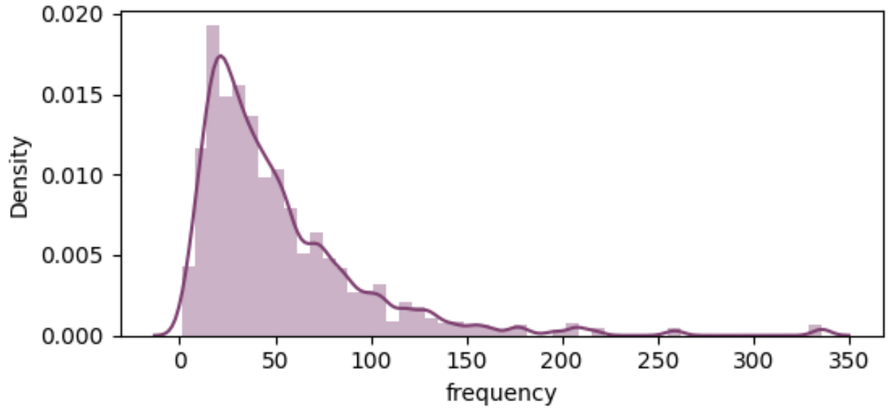
To determine the Monetary value, we need to calculate the total purchases made by each customer by multiplying the quantity and unit price of the items they bought. Then, we group this total sales by customerID to obtain the Monetary value

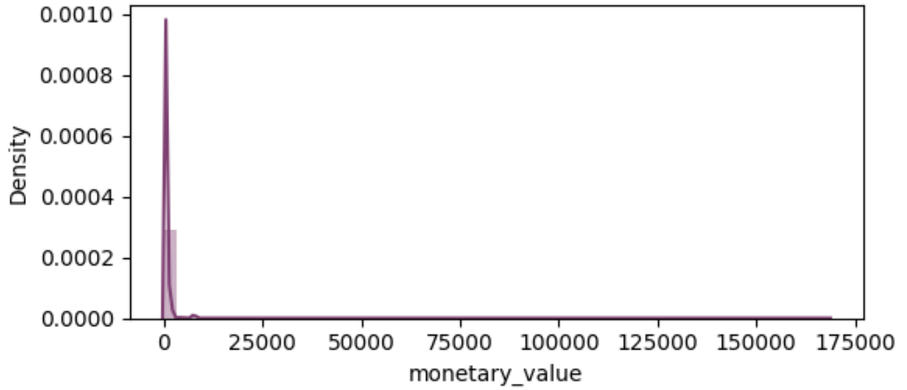
Customer D will **have highest monetary value** among others customer because he has the **most total amount spent**

4 | RFM Segmentation

Visualizing the distribution of feature variables in RFM





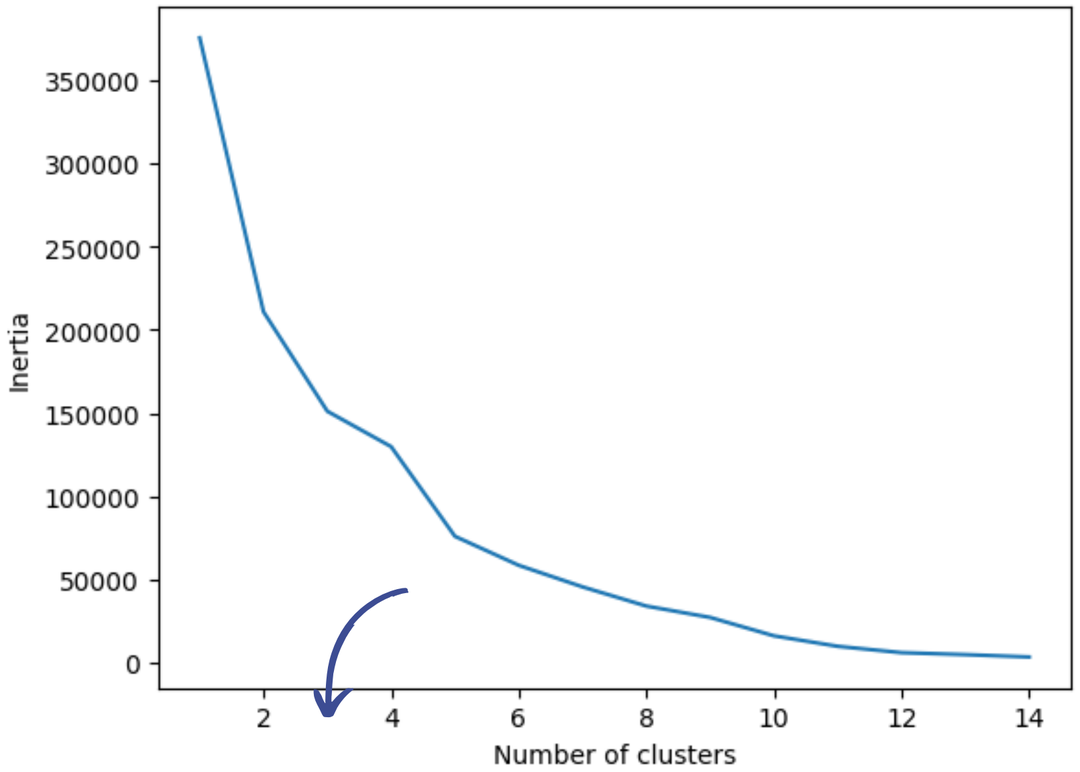


- In recency, the number of recency values with values **0-30 is quite stable**
- While on frequency, the most frequency values are in the **value of 0-50**
- And for monetary value, almost all of the value is **less than 25000**

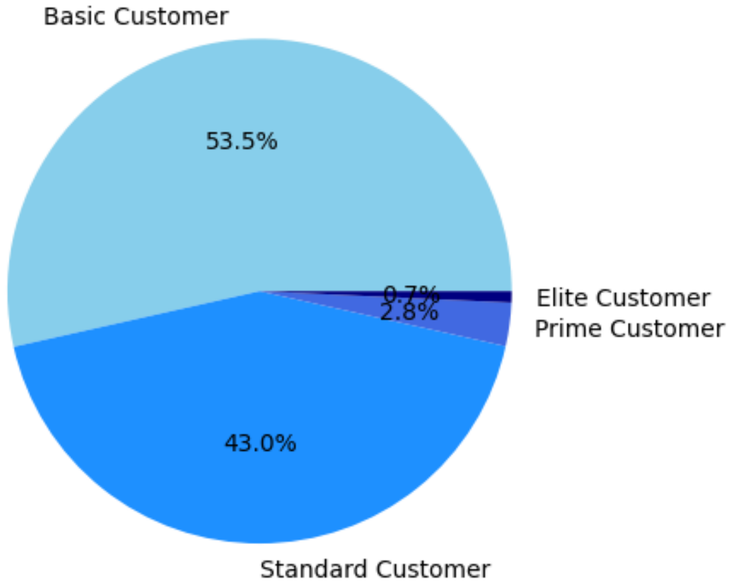
4 | RFM Segmentation

K-Means Clustering & Segmentation Process

K-Means Clustering to find optimal cluster value for the data



Based on the graph, the **optimal cluster value is K=8**



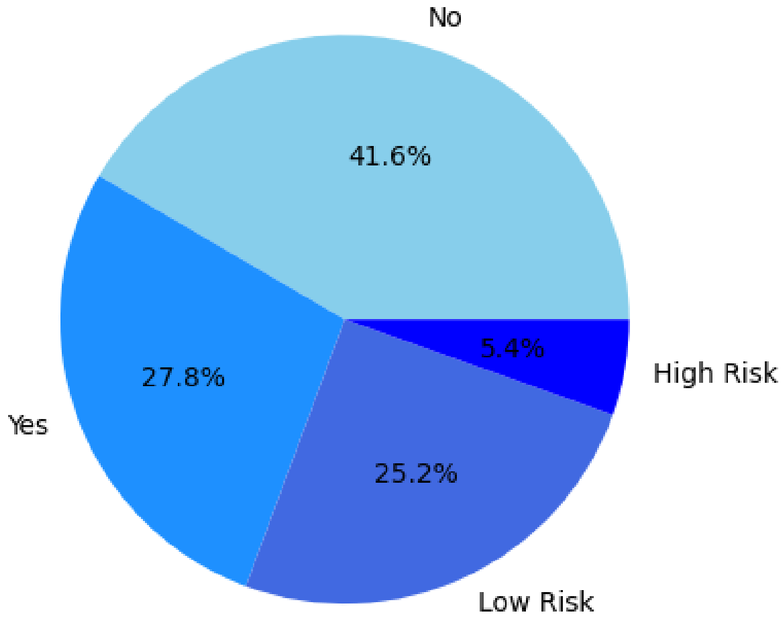
Four customer segments were obtained with an RFM approach where the data showed that most customers tended to **be basic and standard customers**

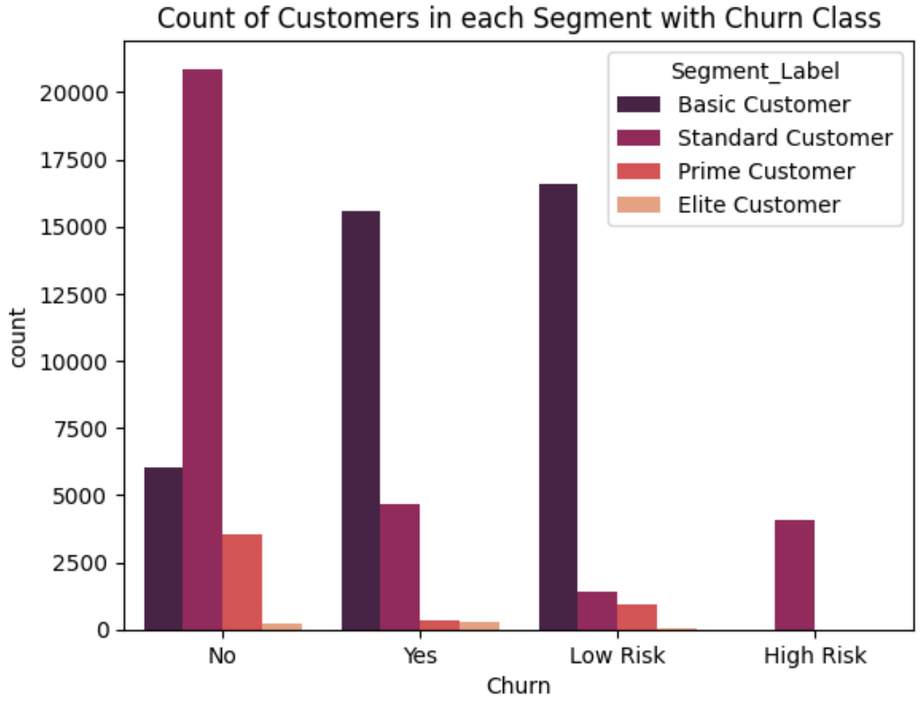
05

Churn Analysis

5 | Churn Analysis

Churn Analysis

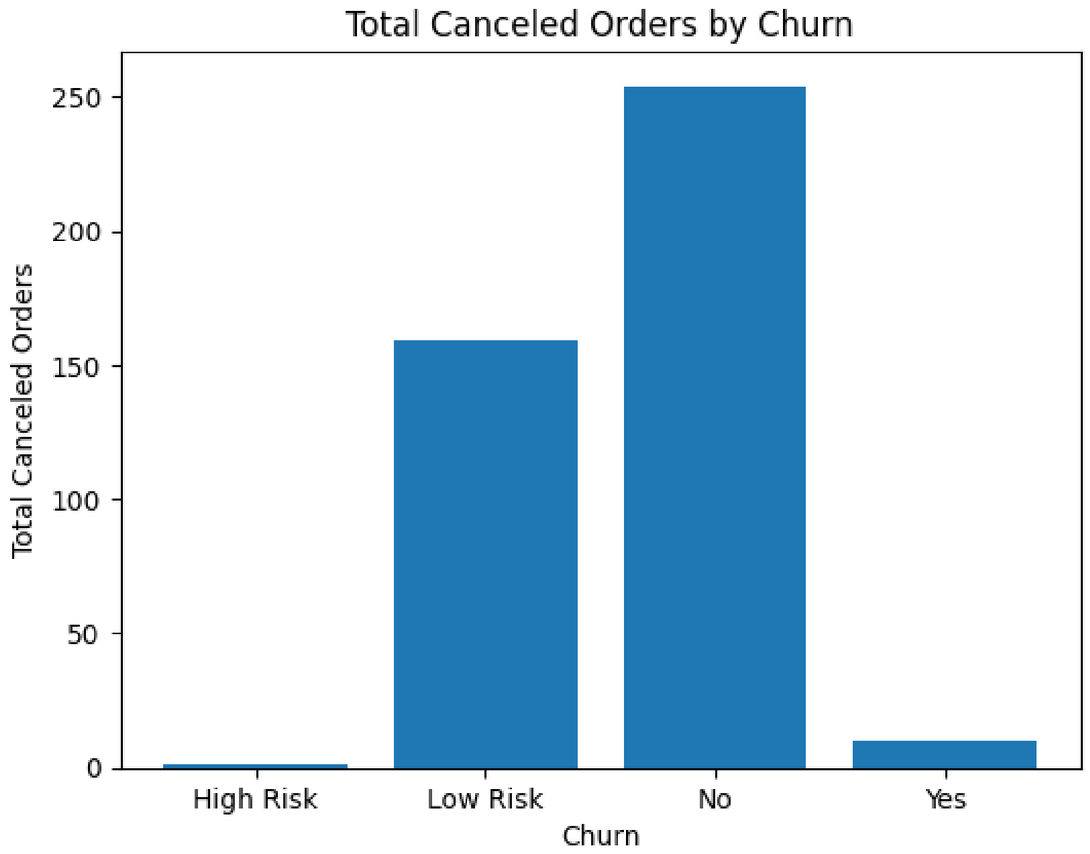




In the graph above, it can be seen that the percentage of "Yes" churn status **reaches 27.8%** and if viewed in more detail, churn is predicted to be mostly done by **Basic Customer and Standard Customer**

5 | Churn Analysis

Churn Analysis



Previously, I saw that there were orders canceled with quite a large amount. I assume that cancelation will have an effect on churn, but after looking at the chart on the side, canceled orders are placed by customers who are not predicted as churn or the churn status is 'yes', **but many canceled orders are done by customers who are not churn customers**

06

Business Recomendations

6 | Bussiness Recommendation

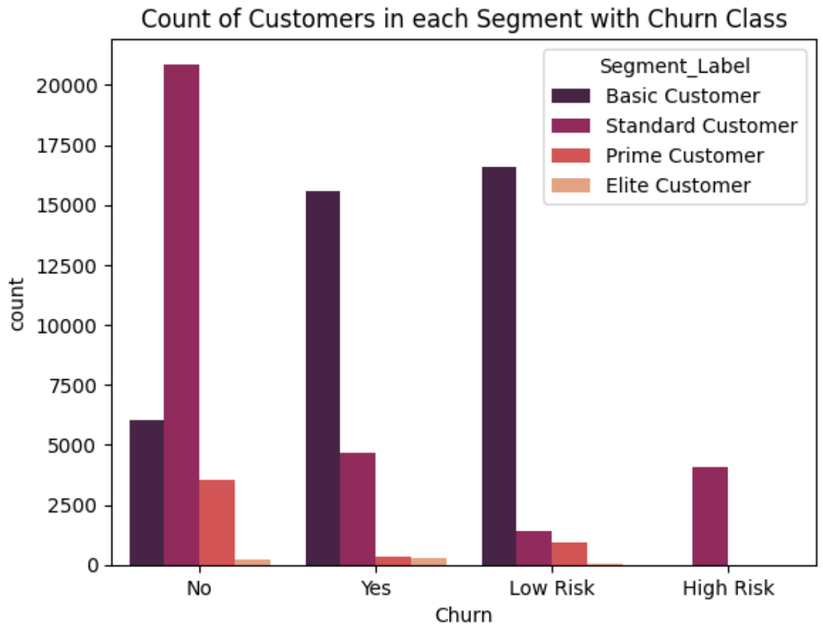
Canceled Order

- Canceled orders are a problem where canceled orders do not depend on the customer churn. In addition, many canceled orders are also carried out by Basic Customers who are the main customers in retail sales.
- The presentation of canceled orders is also widely carried out by customer from countries outside Europe and the UK where sales are still small compared to sales in the UK

Canceled orders are a separate problem where further survey is needed what is the cause of canceled orders, especially those made by customers from outside Europe and the UK, whether shipping costs are large, or there are other reasons.

6 | Bussiness Recommendation

Churn Customer



Basic customers are customers with low recency, frequency, and monetary value and have a fairly high churn possibility. This is quite reasonable and can be said that they are new customers or customers who rarely make transactions in the retail business. However, the number of basic customers is very high so the potential for churn will also be high if you look at the chart.

It is necessary to do more camping on basic customers so that they can rise to standard customers or even prime customers as one step in reducing the churn rate in business.

Thank You

Widiareta Safitri

wretasafitri33@gmail.com

https://github.com/widiarsaf