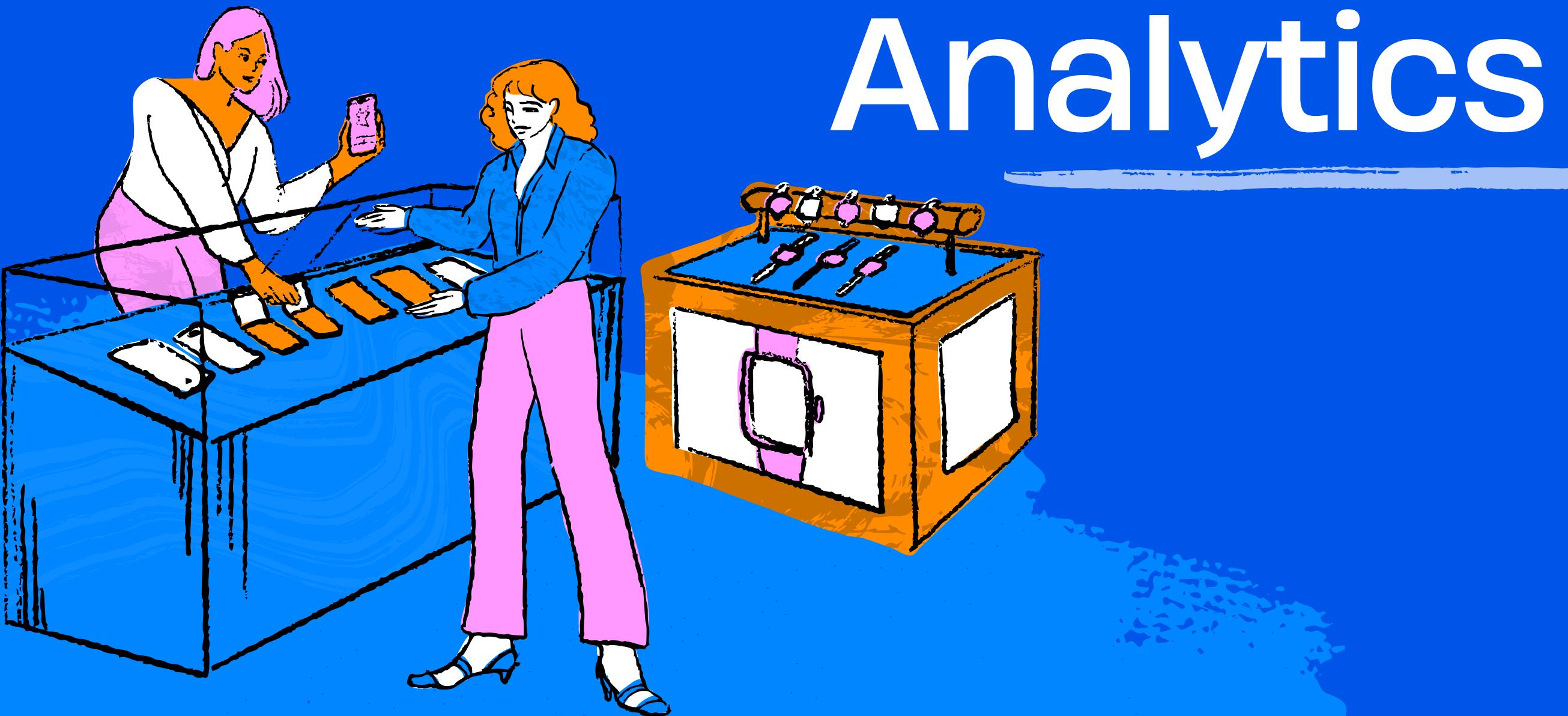
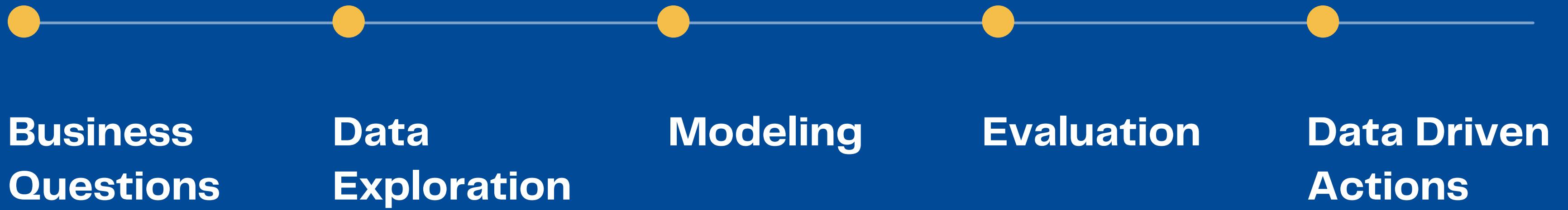


University of Milano-Bicocca  
LM DATA SCIENCE

# Marketing Analytics



# Workflow



# Business Questions

1 WHICH OF OUR CUSTOMERS ARE VALUABLE?

2 How can we personalize marketing efforts based on customers' recent purchases, buying frequency, and spending habits?

3 How can we proactively identify customers who are at risk of leaving and implement strategies to retain them?

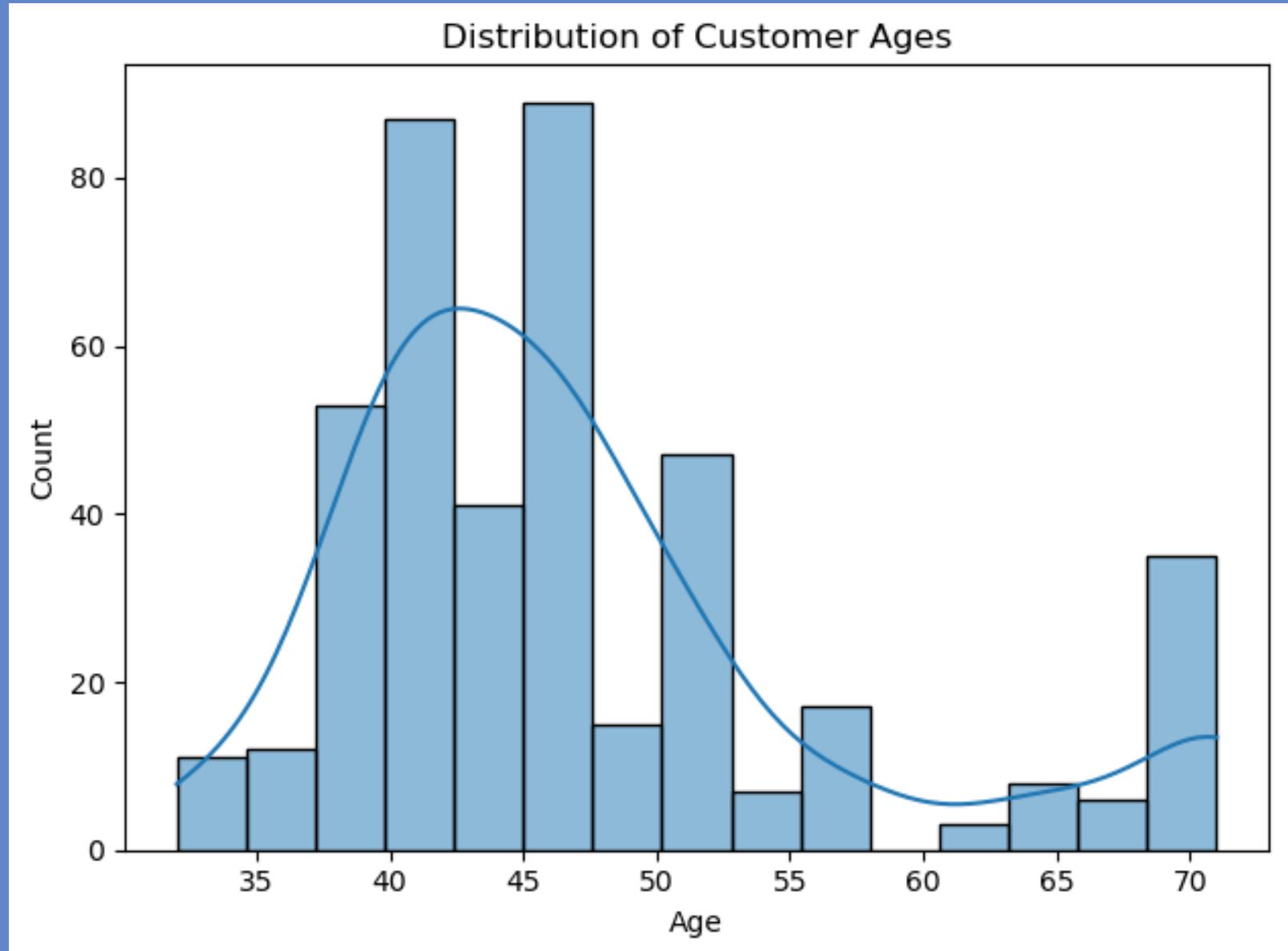
4 What product combinations can we promote to increase sales and enhance customer satisfaction? support your hypothesis.

5 What is the probability of abandonment associated with each customer?



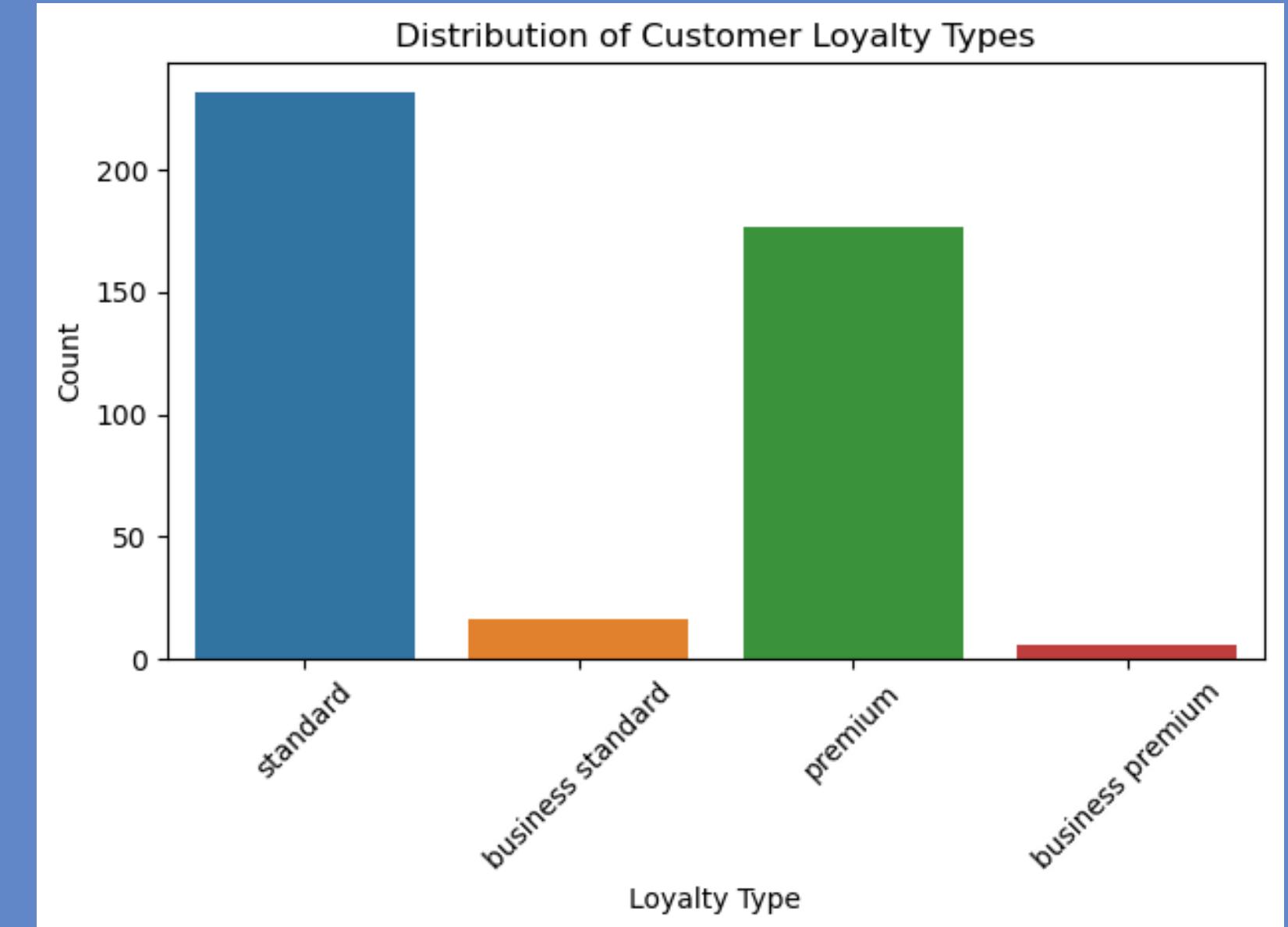
# Data Exploration

## Target Clients



### By Age

The Target audience comprises the dynamic and influential segment of individuals aged 35-50.



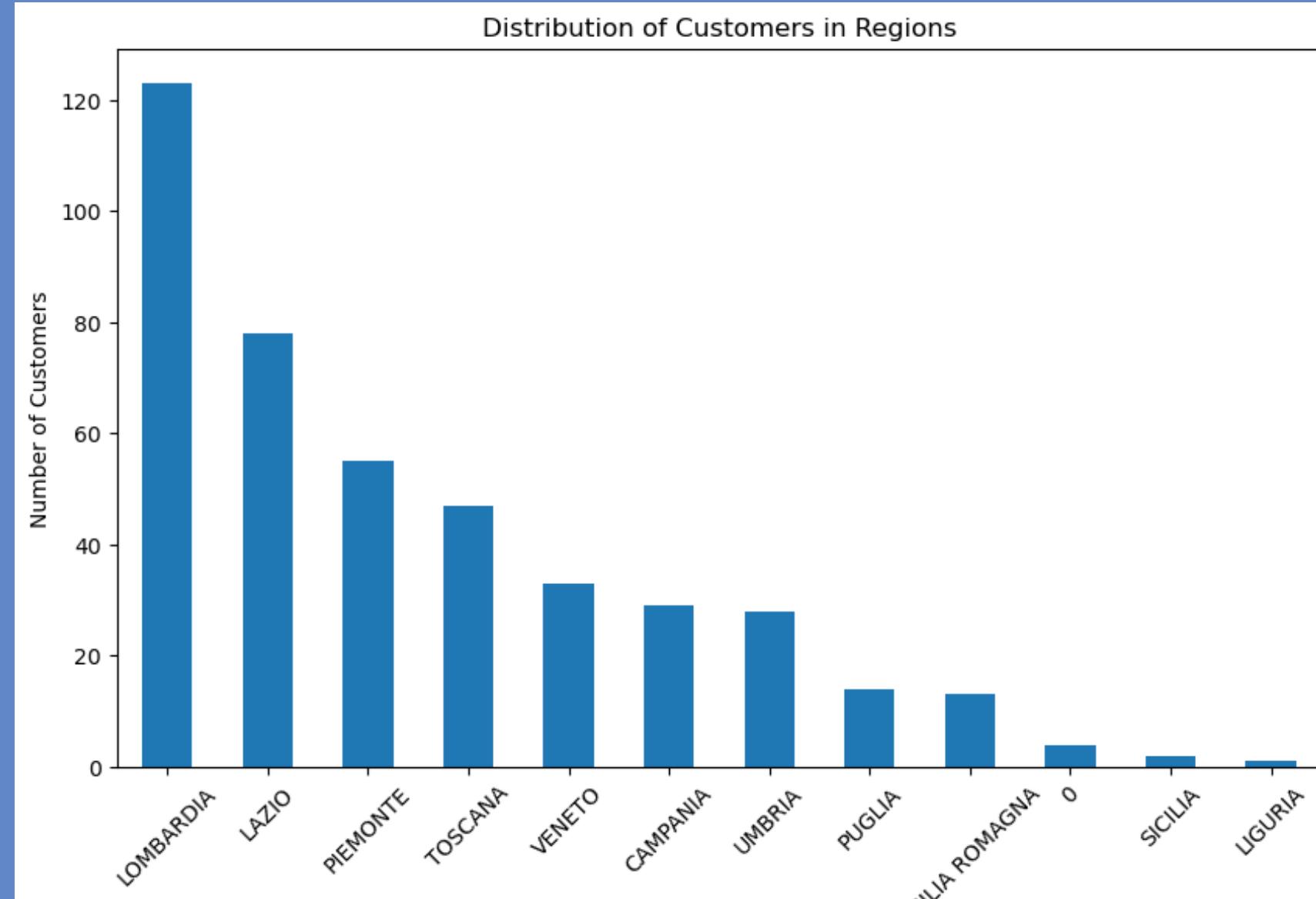
### By Fidelity Types

majority of individuals prefer the standard option, They prefer consistent experience without additional features



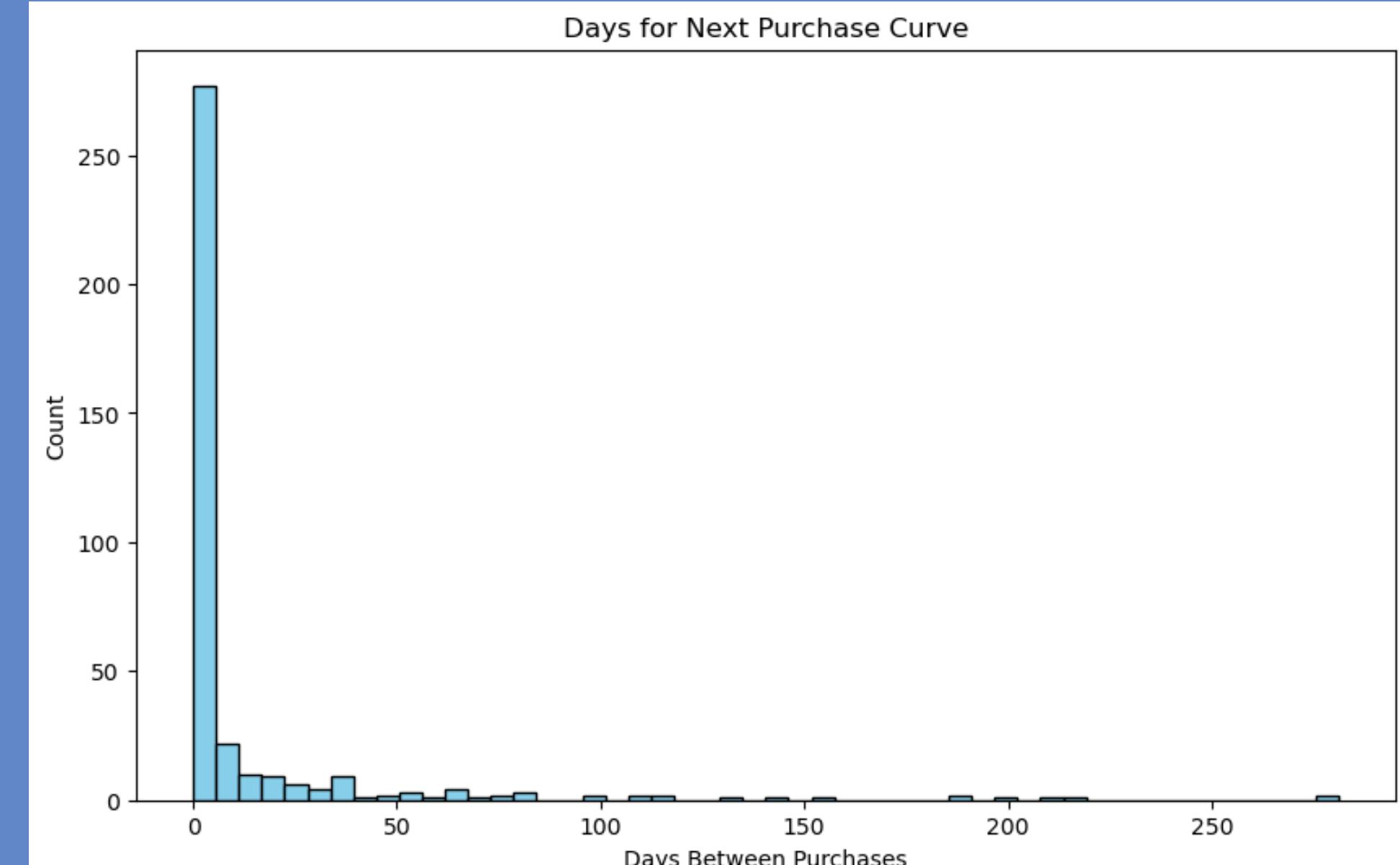


# Target Clients



## By Region Distribution

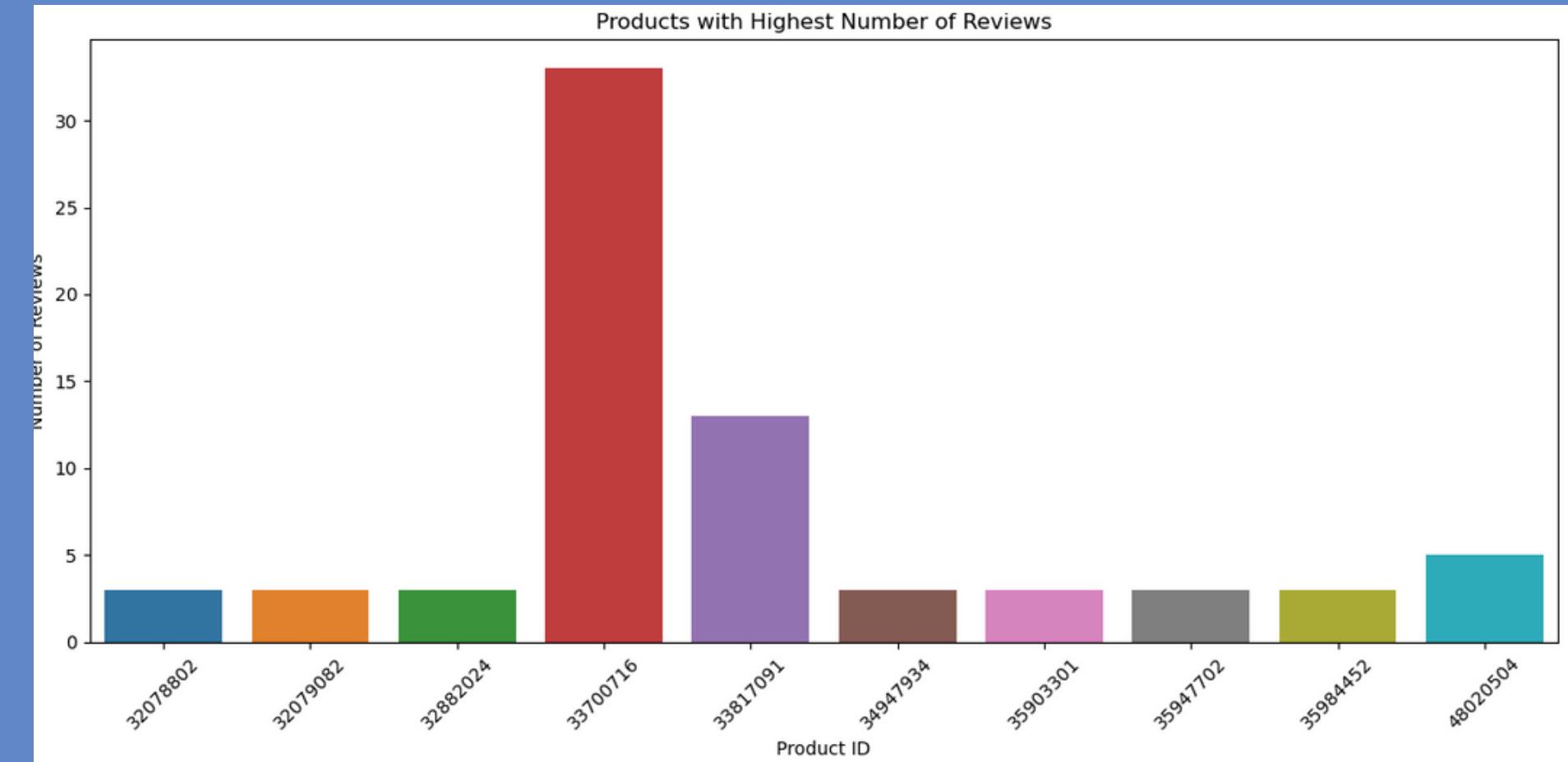
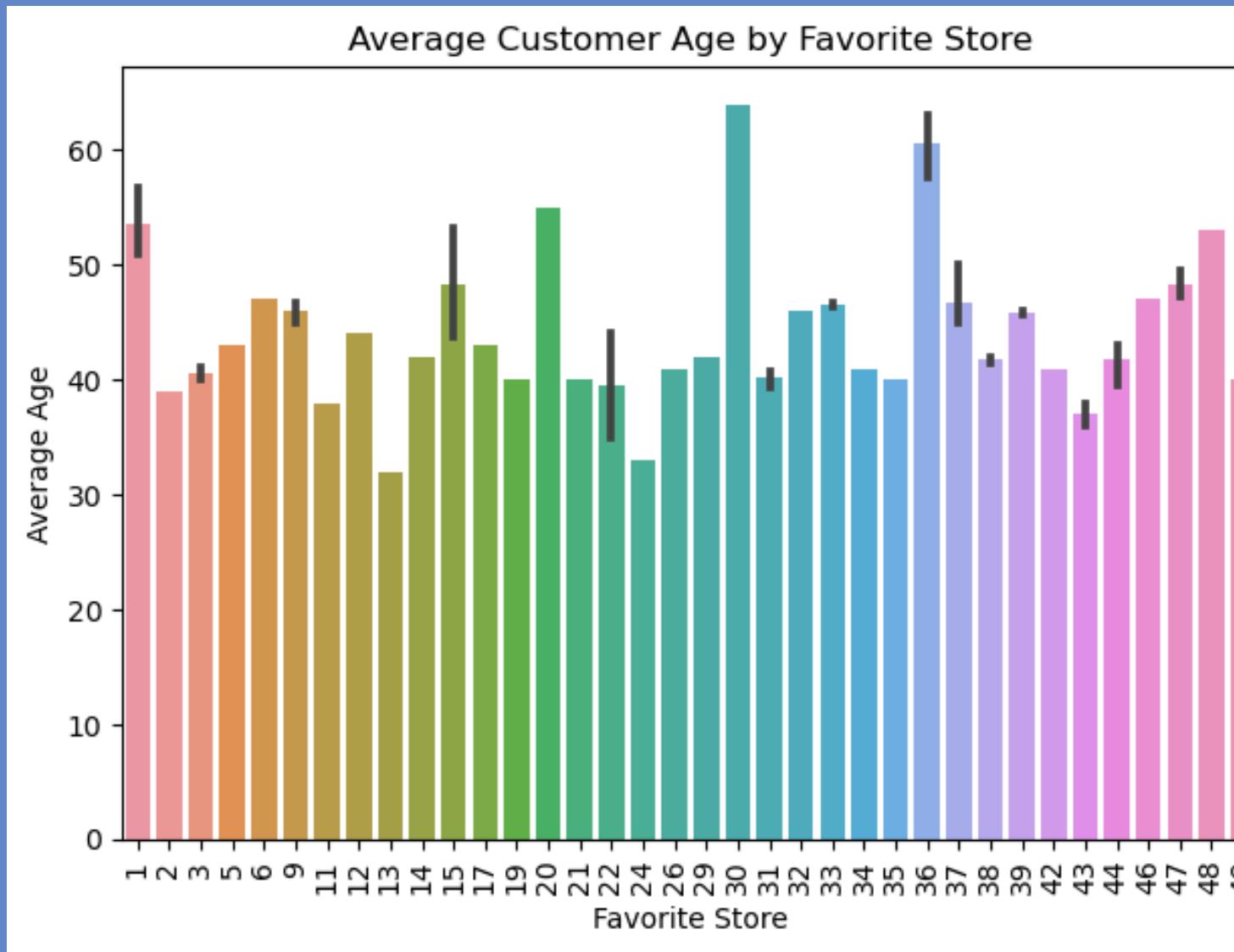
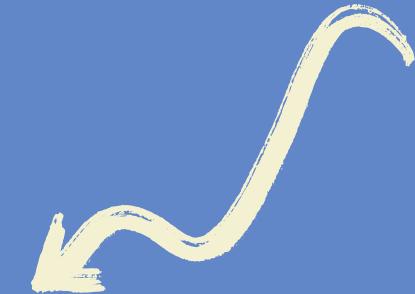
This distribution reveals that the company's customer base is primarily concentrated in the northern and central regions of Italy, with some presence in the southern regions as well.



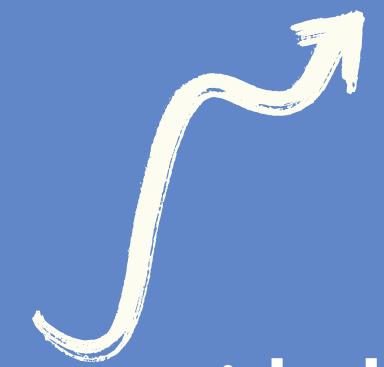
## By the Number of Days for next Purchase

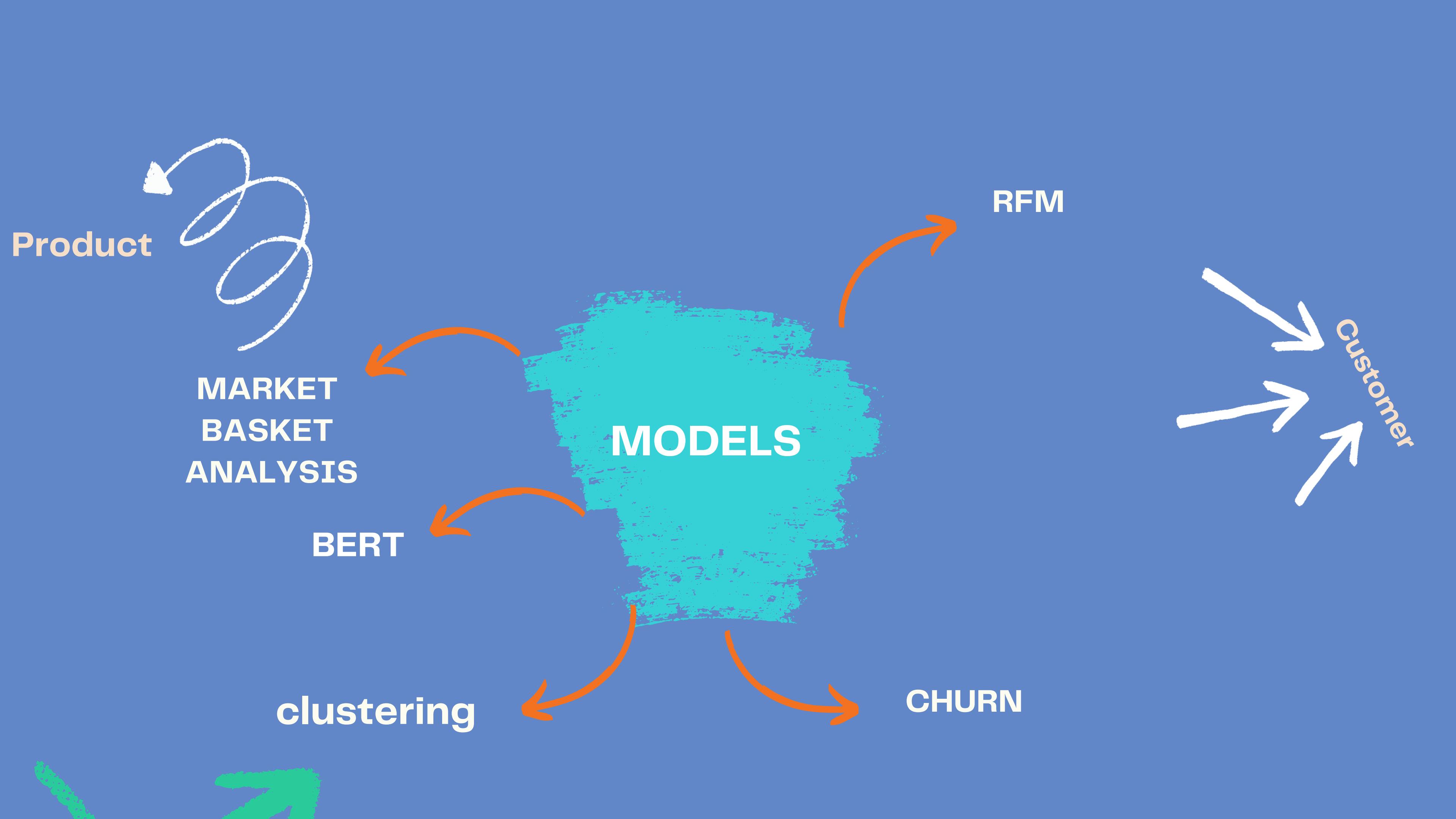
The graph reveals a high count of repurchases occurring within approximately the first 10 days. This indicates a relatively short repurchase cycle, and targeted marketing strategies to drive repeat purchases.

Average customer age varies across different favorite stores, ranging from the early 30s to the mid-60s. This suggests that different stores may attract customers from different age groups.



Product ID 33700716 stands out with the highest number of reviews (33), indicating strong customer engagement. This suggests a significant interest and popularity surrounding this particular product. Monitoring and leveraging customer feedback can optimize customer satisfaction and capitalize on its success.





# RFM OBJECTIVE AND MAIN COMPONENTS

The objective of the RFM (Recency, Frequency, Monetary) model is to segment customers based on their transactional behavior to target them with personalized marketing campaigns.

## RECENCY

When was the last purchase of the customer?

## FREQUENCY

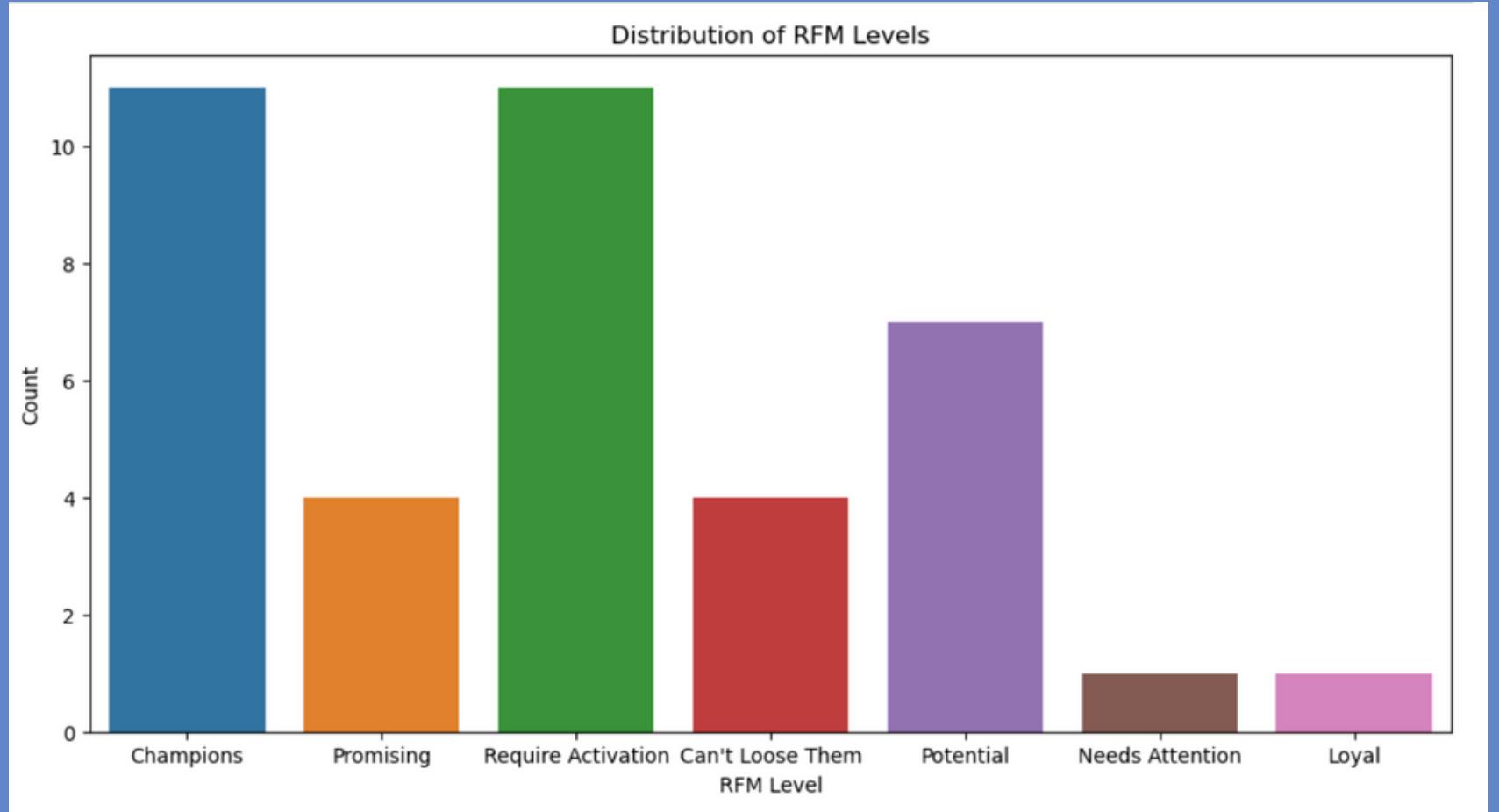
what is the frequency of purchase by the customer?

## MONETARY

How much money does the customer spend?

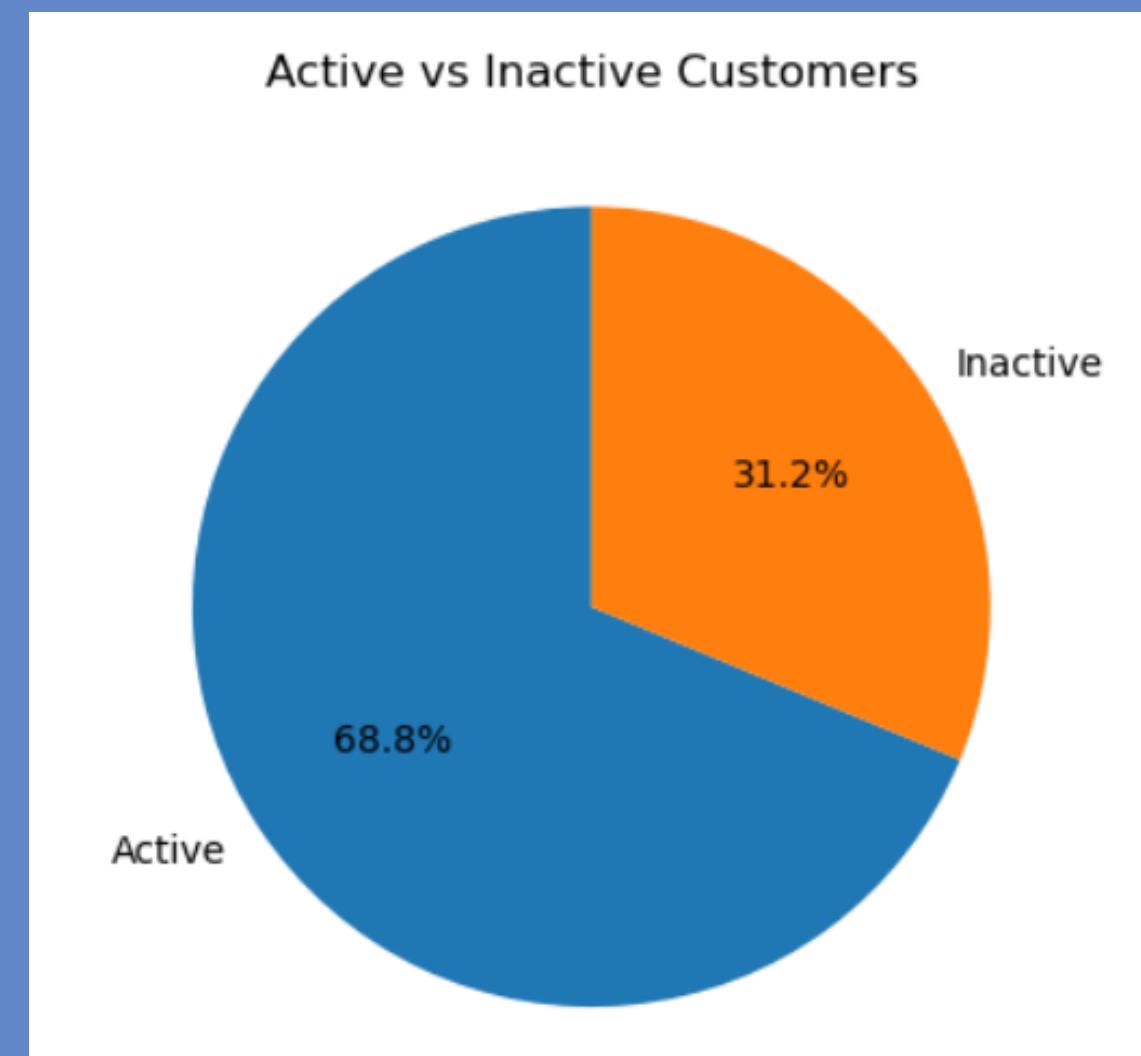


# RFM



"Require Activation" segment highlights the need for targeted strategies to re-engage customers and maximize their potential value.

Based on the churn analysis, where 31.2% of customers are inactive and 68.8% are active, the marketing strategy should focus on re-engaging and reactivating the inactive customer segment.



# CHURN OBJECTIVE AND USED ALGORITHMS

Identify and understand the factors that contribute to customer churn in order to develop effective retention strategies.

**DECISION  
TREE**

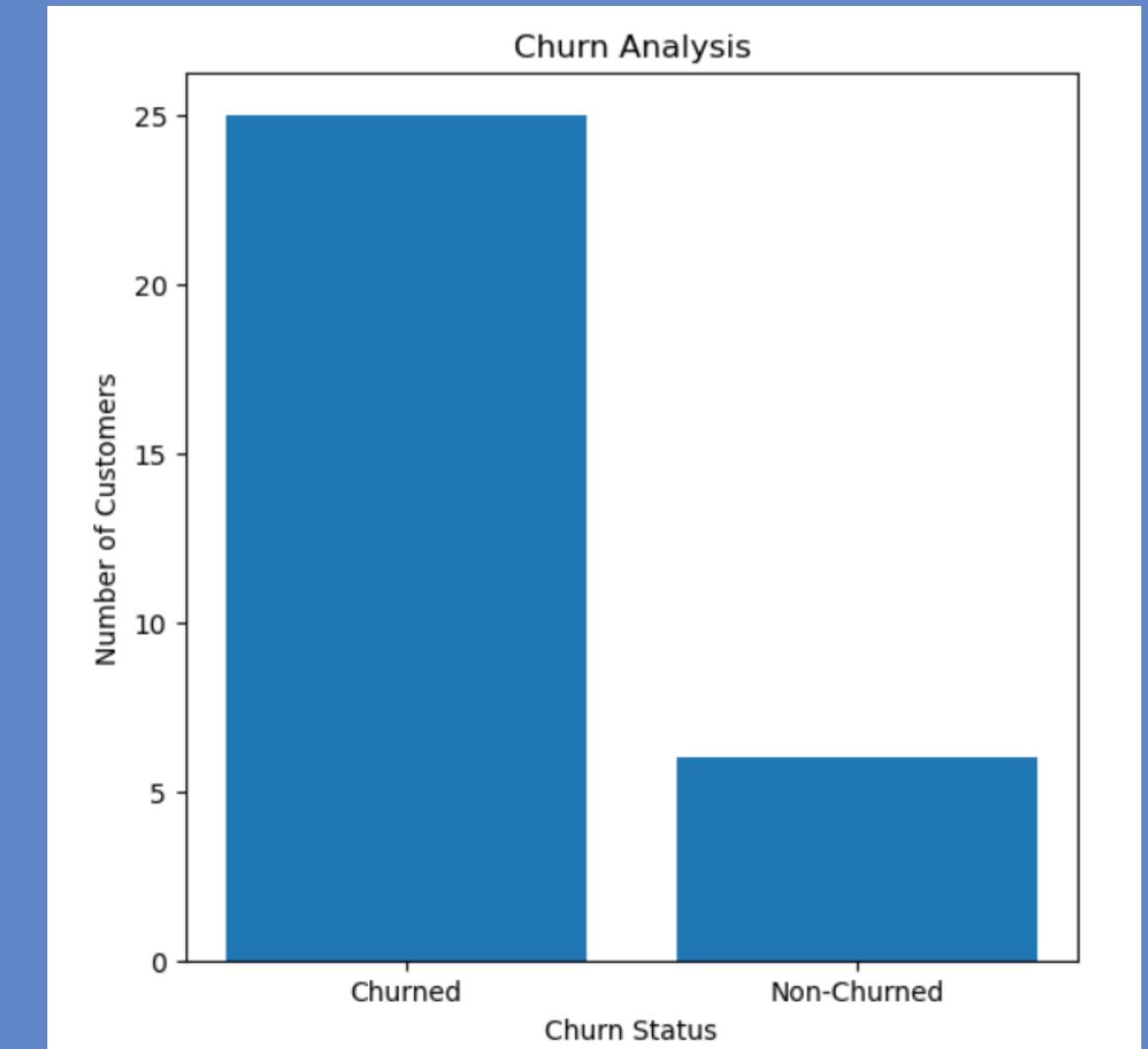
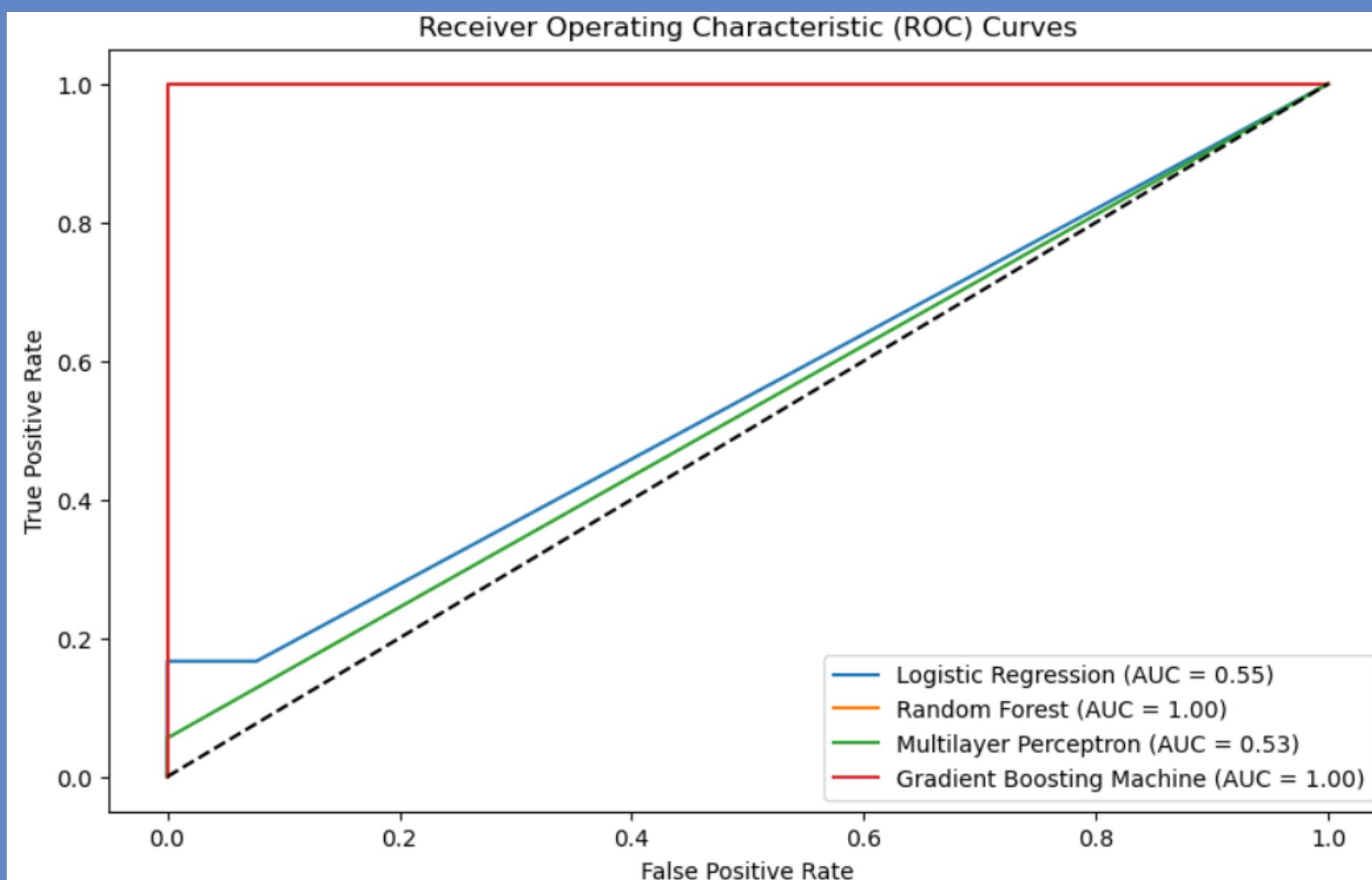
**Random  
Forest**

**Logistic  
Regression**

**Multilayer  
perceptron**



# CHURN



**Logistic Regression with AUC = 0.55 and Multilayer Perceptron with AUC = 0.53** An AUC of 0.55 suggests that the logistic regression model is performing slightly better. **Random Forest and Gradient Boosting Machine with AUC = 1.00:** An AUC of 1.00 indicates perfect separation between positive and negative classes.

Identified churned customers requires personalized and engaging addressing their concerns and providing incentives to bring them back to the business.

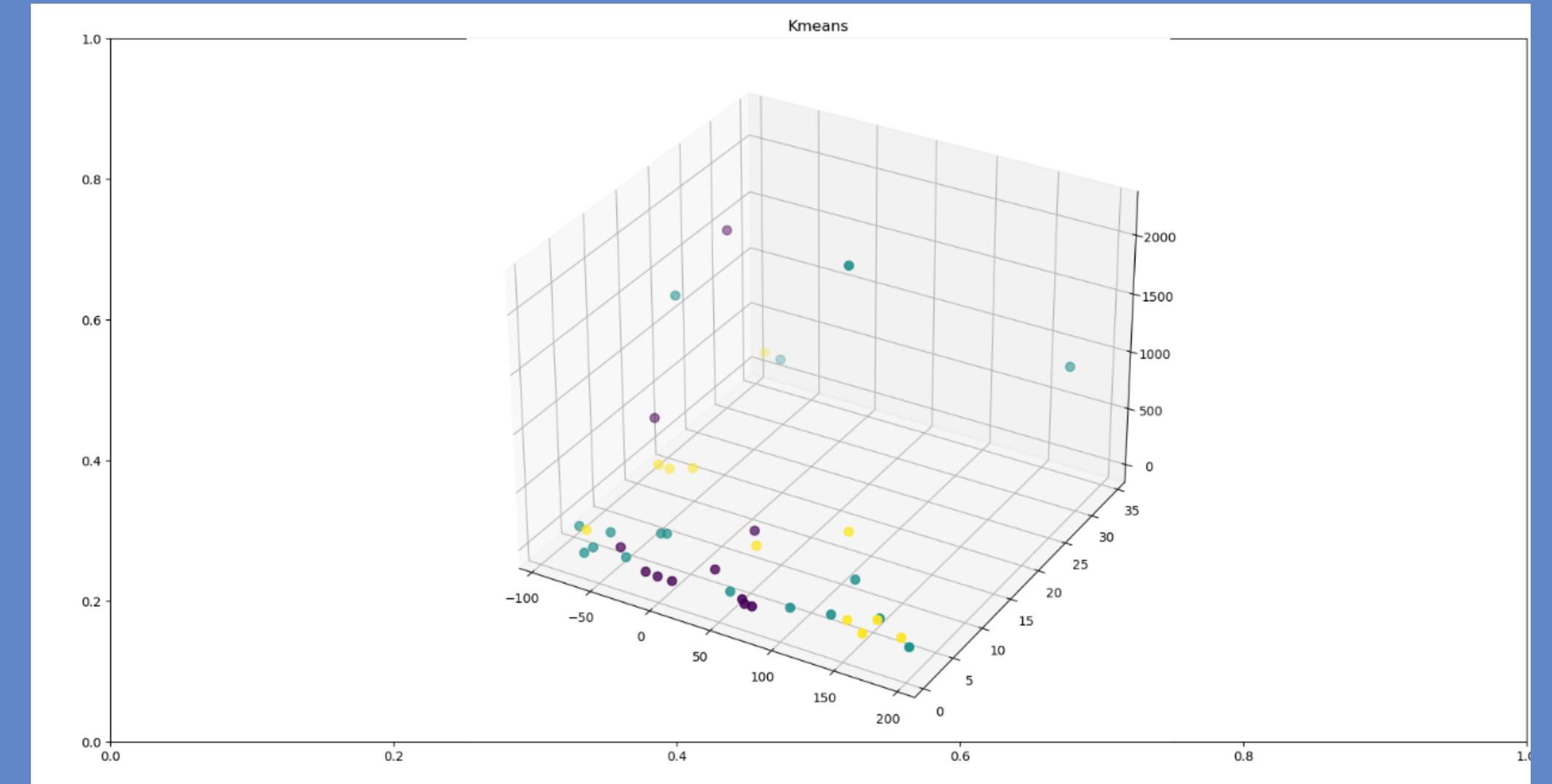
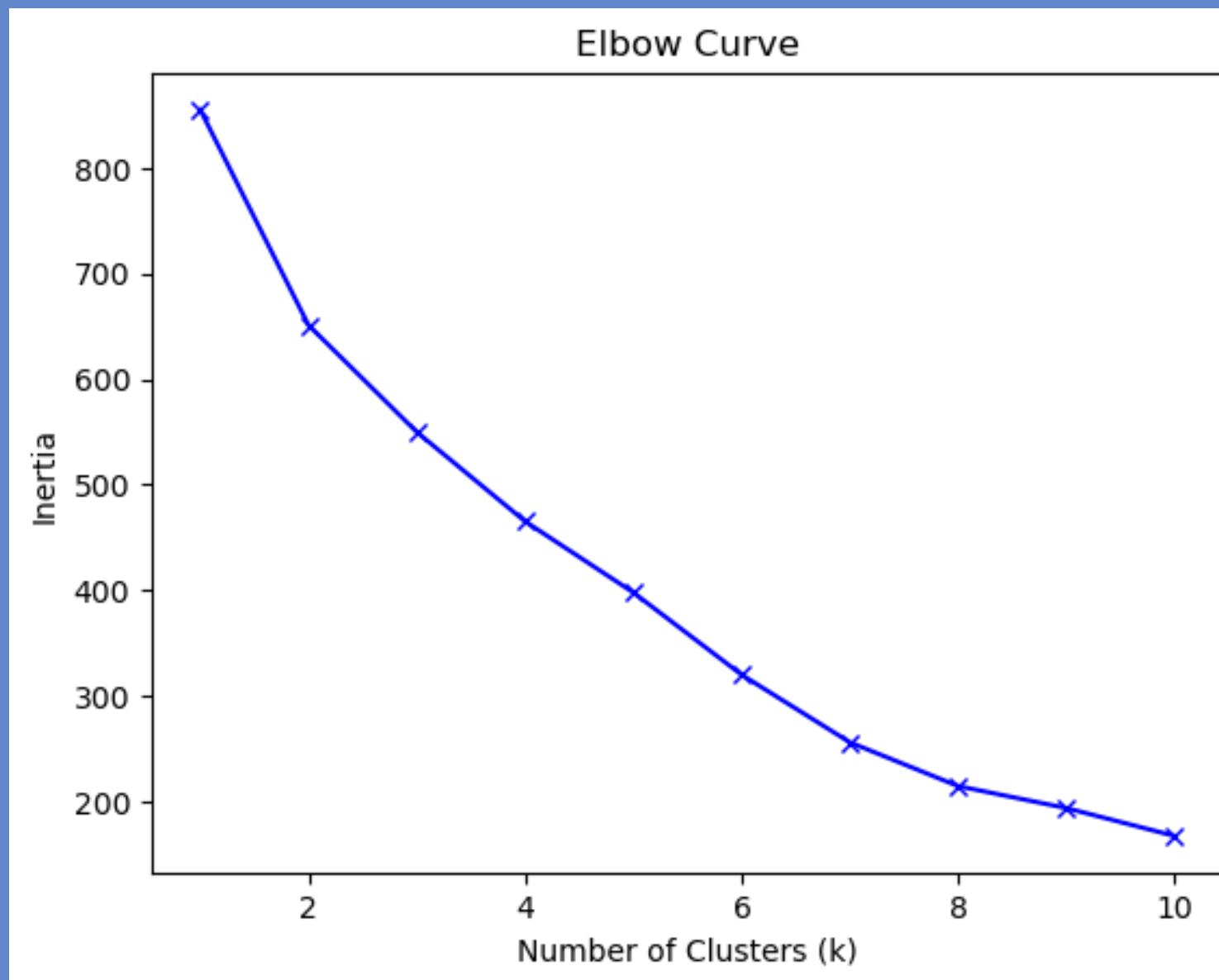
# CLUSTERING OBJECTIVE AND USED ALGORITHMS

The objective of clustering in marketing analytics is to group similar customers together based on their characteristics or behaviors, enabling targeted marketing efforts and personalized experiences.

**k-Means  
with PCA**

**K-Means  
Elbow method**

# Clustering



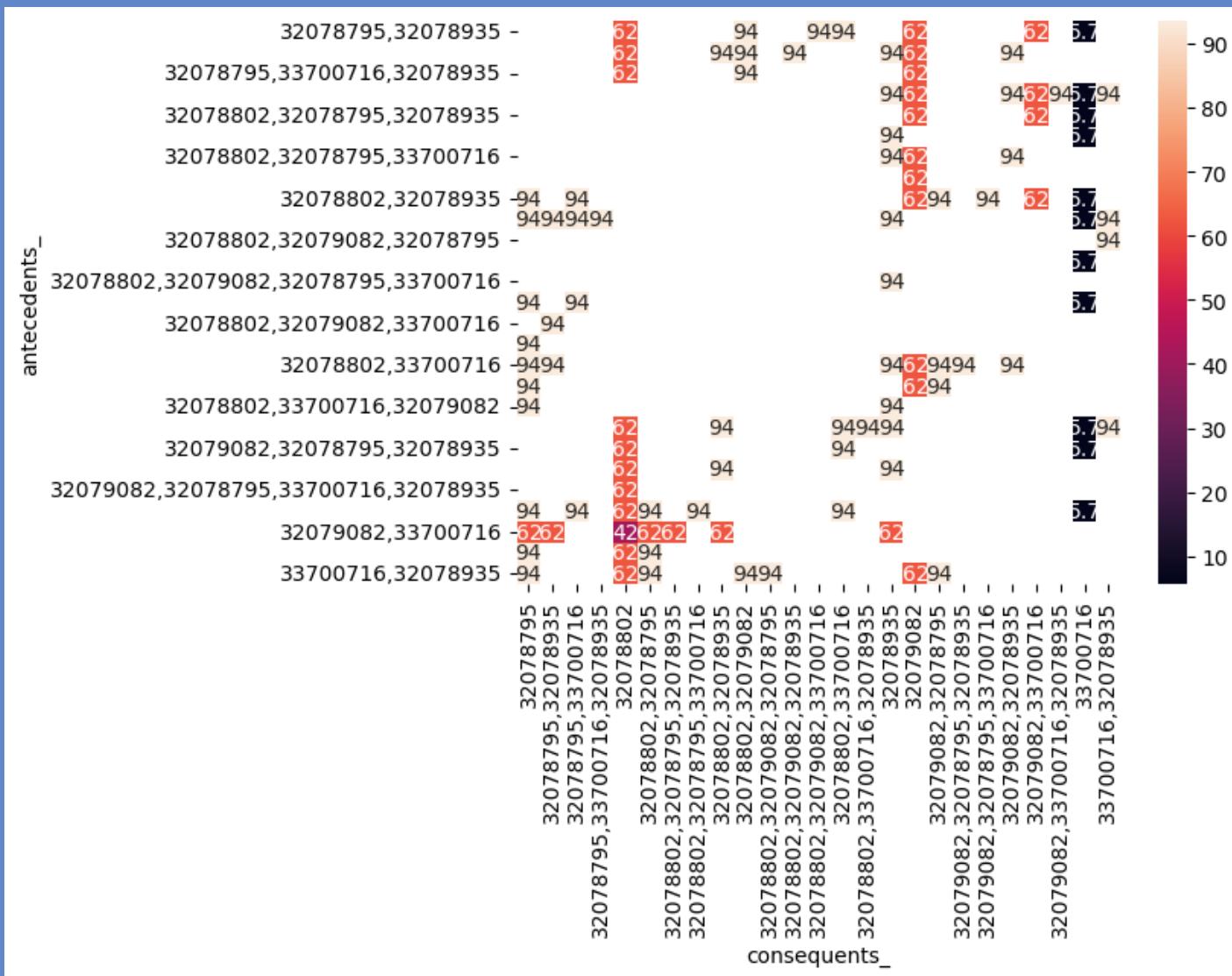
To determine the optimal number of clusters, we have to select the value of k at the “elbow” ie the point after which the distortion/inertia starts decreasing in a linear fashion. Thus for the given data, we conclude that the optimal number of clusters for the data is 3

# BASKET MARKET ANALYSIS OBJECTIVE AND USED ALGORITHM

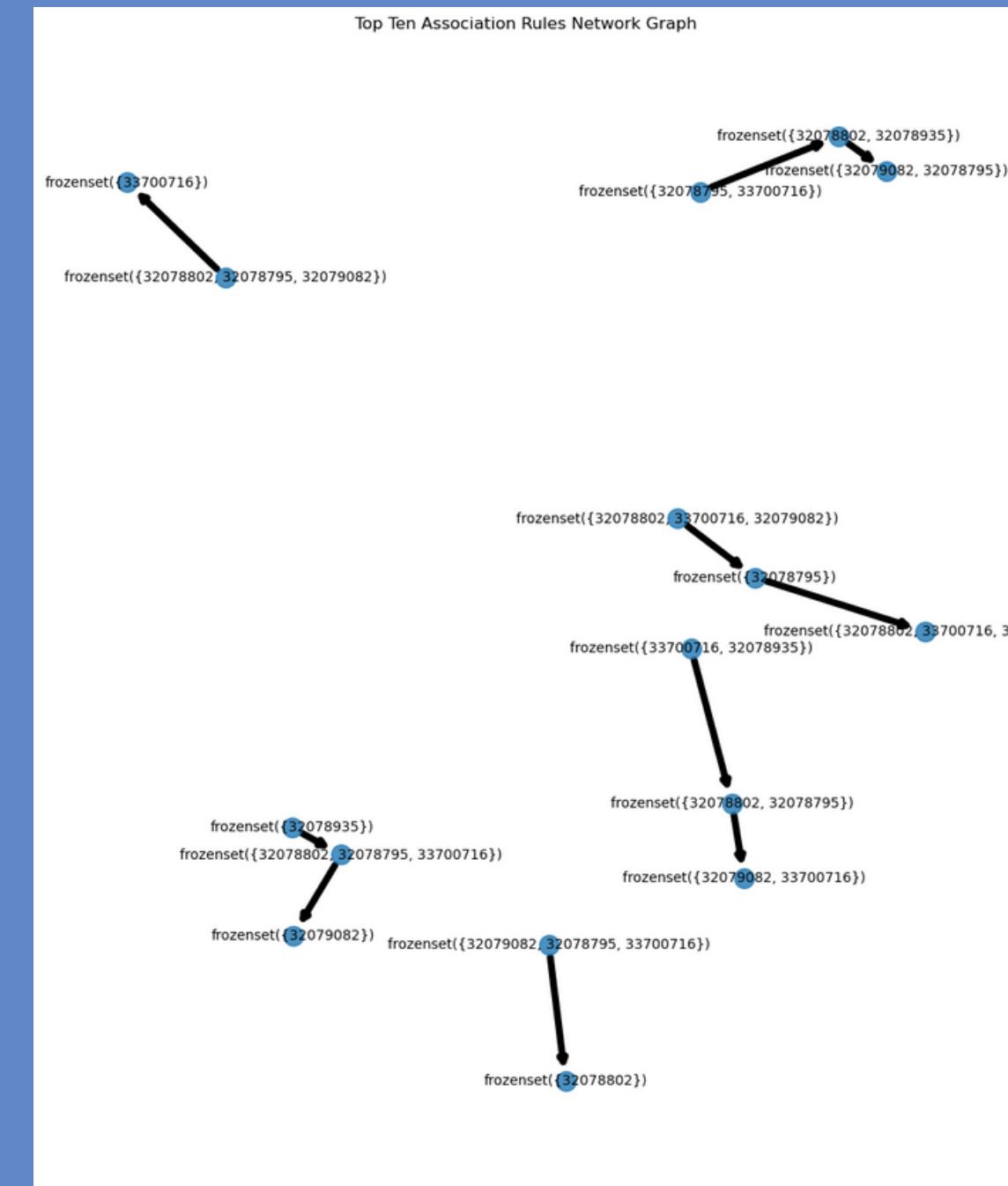
The objective of basket market analysis is to uncover associations and patterns in customer purchasing behavior to optimize cross-selling and upselling strategies.

APRIORI

# BASKET MARKET ANALYSIS



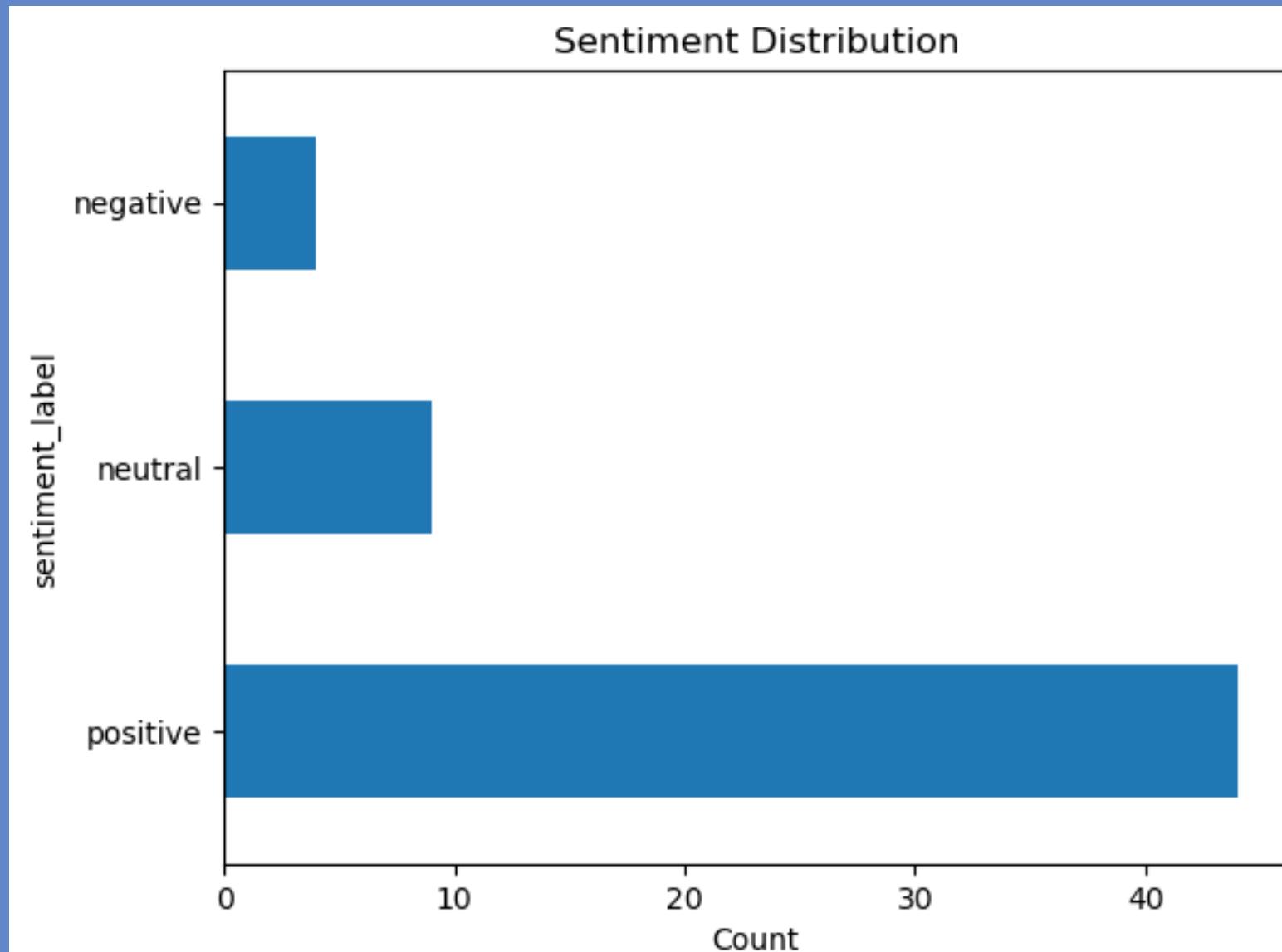
min\_support = 0.01



- The most purchased product is identified as 33700716, with nearly sixty thousand units sold.
  - The top eight purchased products are 33700716, 33817091, 34843564, 32882024, 32079103, 34252904, 35209202, and 316183.

**The network graph provides a clear and intuitive visualization of the top ten association rules in the basket market analysis, allowing us to easily identify the most influential relationships and understand the flow of item dependencies.**

# SENTIMENT ANALYSIS



## Model Training and Evaluation:

Utilized TF-IDF vectorization and a LinearSVC classifier for sentiment analysis.

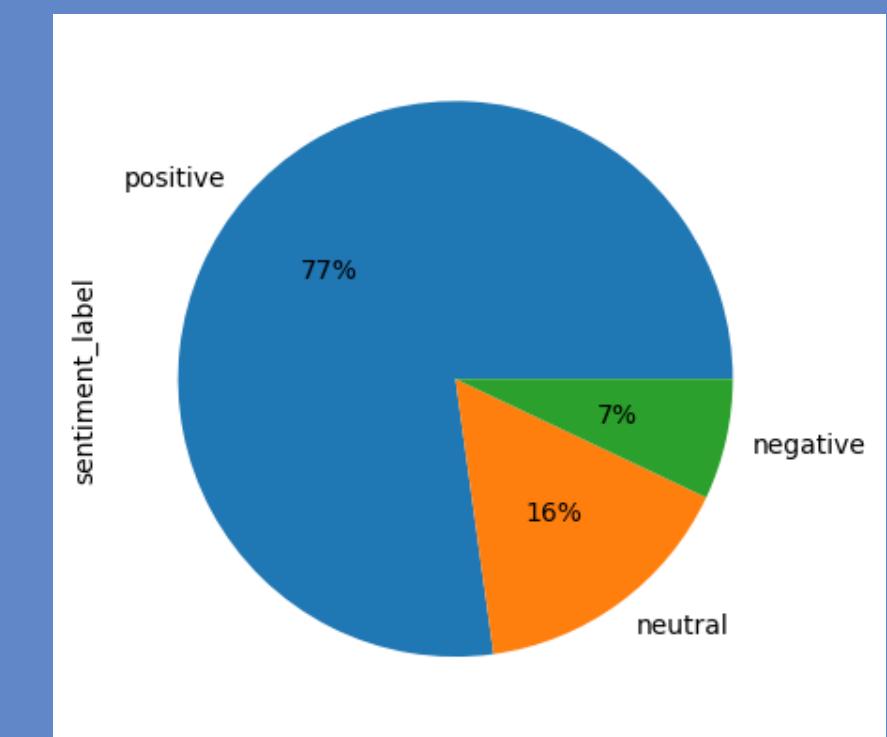
Split the dataset into training and testing sets (80:20 ratio).

Achieved an accuracy of approximately 67% on the test set.

Evaluated the model's performance using a classification report and confusion matrix.

```
accuracy_score(y_test, y_pred)  
# 67% accuracy
```

```
0.6666666666666666
```



# BERT MODEL OBJECTIVE AND USED ALGORITHM

- (Bidirectional Encoder Representations from Transformers): BERT is a pre-trained language model that utilizes the Transformer architecture to understand the contextual representations of words in a sentence. It is being used here for text embedding extraction.
- Logistic Regression: it is used as the classifier to train and predict sentiment labels based on the BERT embeddings.



# large pretrained language model

## BERT Model

```
In [75]: from sklearn.metrics import classification_report

# Generate a classification report
report = classification_report(y_test, y_pred)
print("Classification Report:")
print(report)

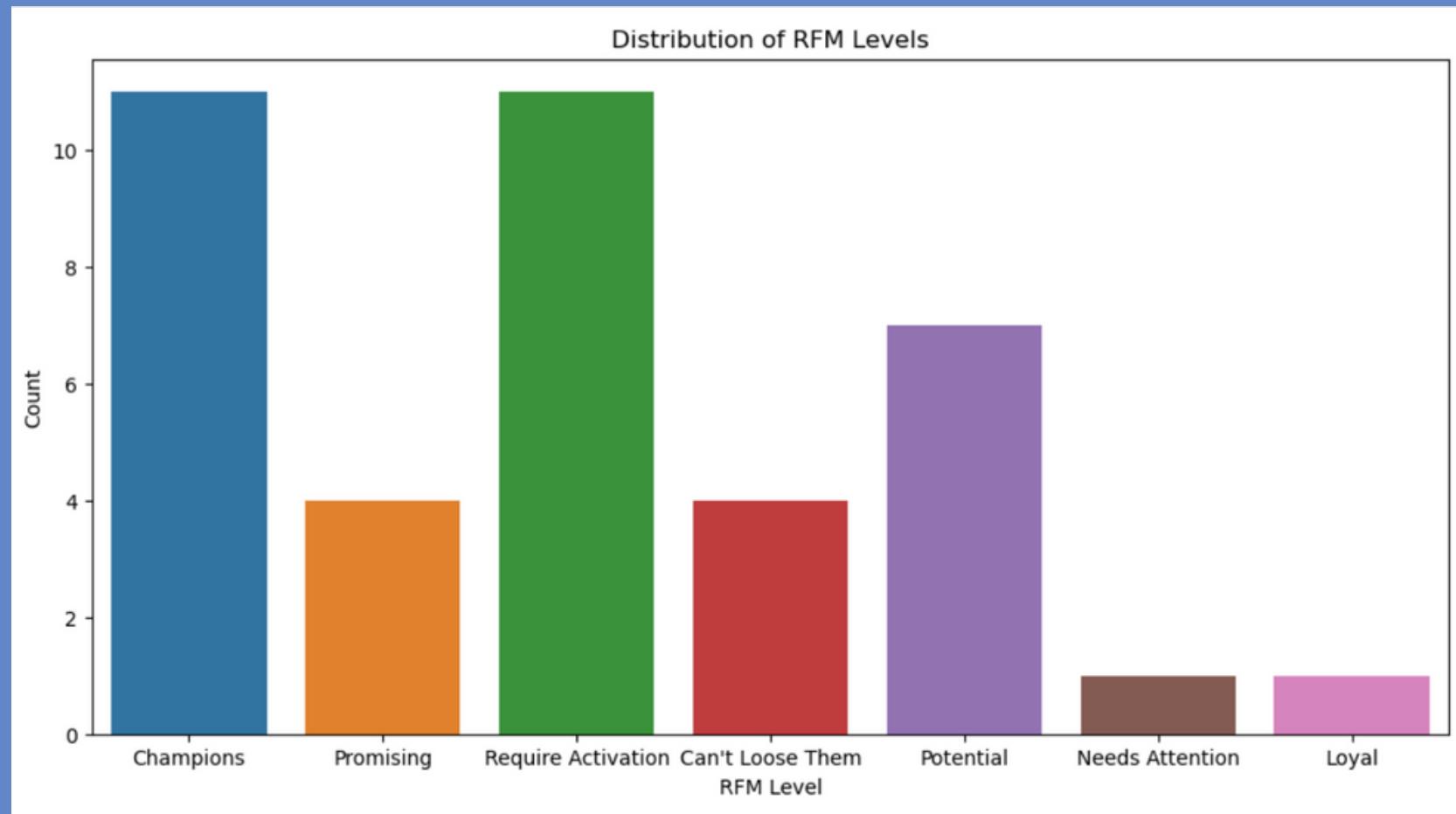
Classification Report:
              precision    recall  f1-score   support

        neutral       0.00     0.00     0.00      2
      positive       0.83     1.00     0.91     10
      accuracy          -         -     0.83     12
    macro avg       0.42     0.50     0.45     12
  weighted avg       0.69     0.83     0.76     12

C:\Users\Admin\anaconda3\lib\site-packages\sklearn\metrics\_classification.py:45: UserWarning: F1 score is ill-defined and being set to 0.0 in labels with no predicted samples.
  warnings.warn(_F1_SCORE_IS_ILL_DEFINED_MESSAGE, UserWarning)
```

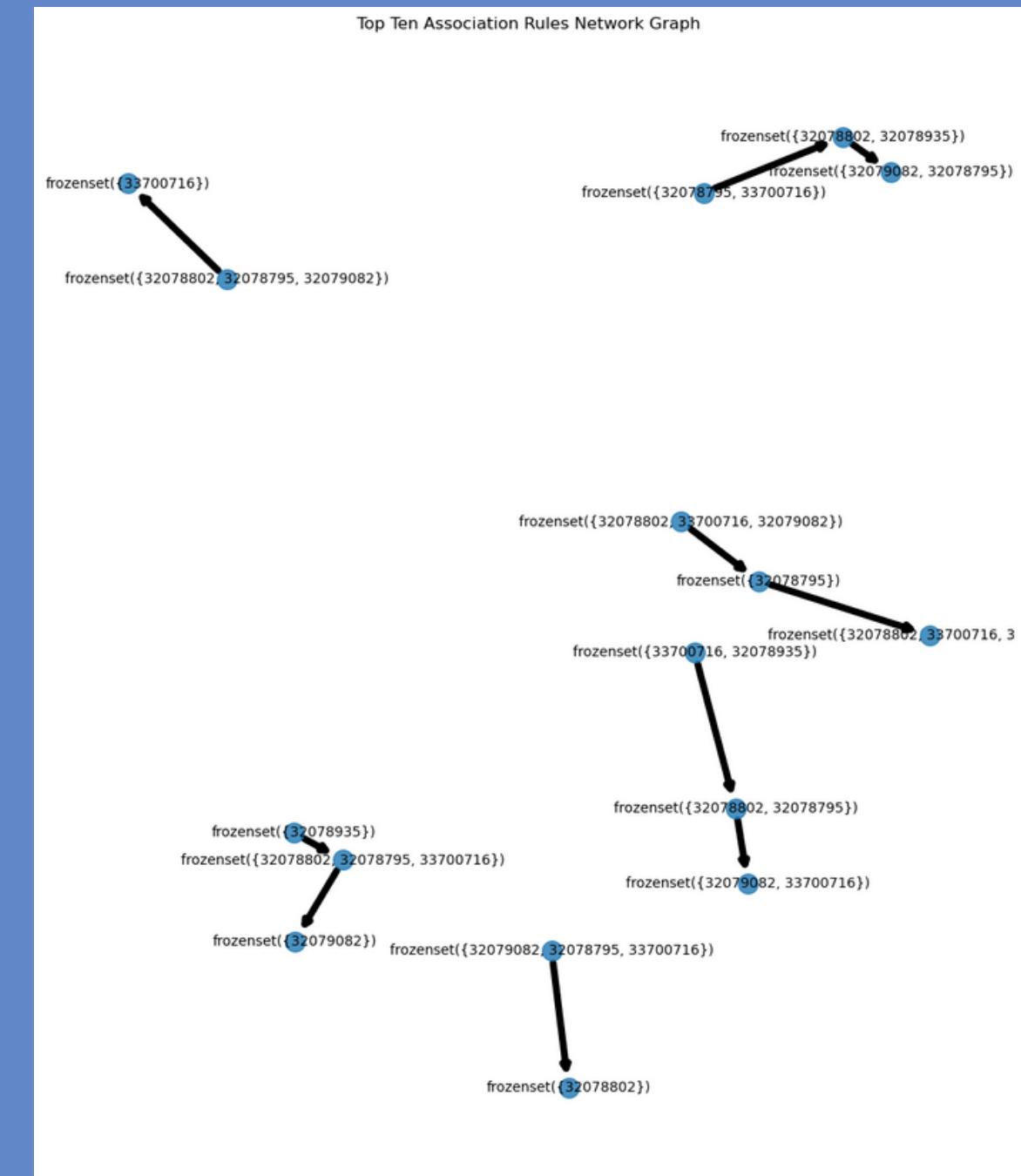
Classification report gives a better picture of the model's performance.

# Detailed Marketing Campaign that combine 2 different aspects RFM and MBA



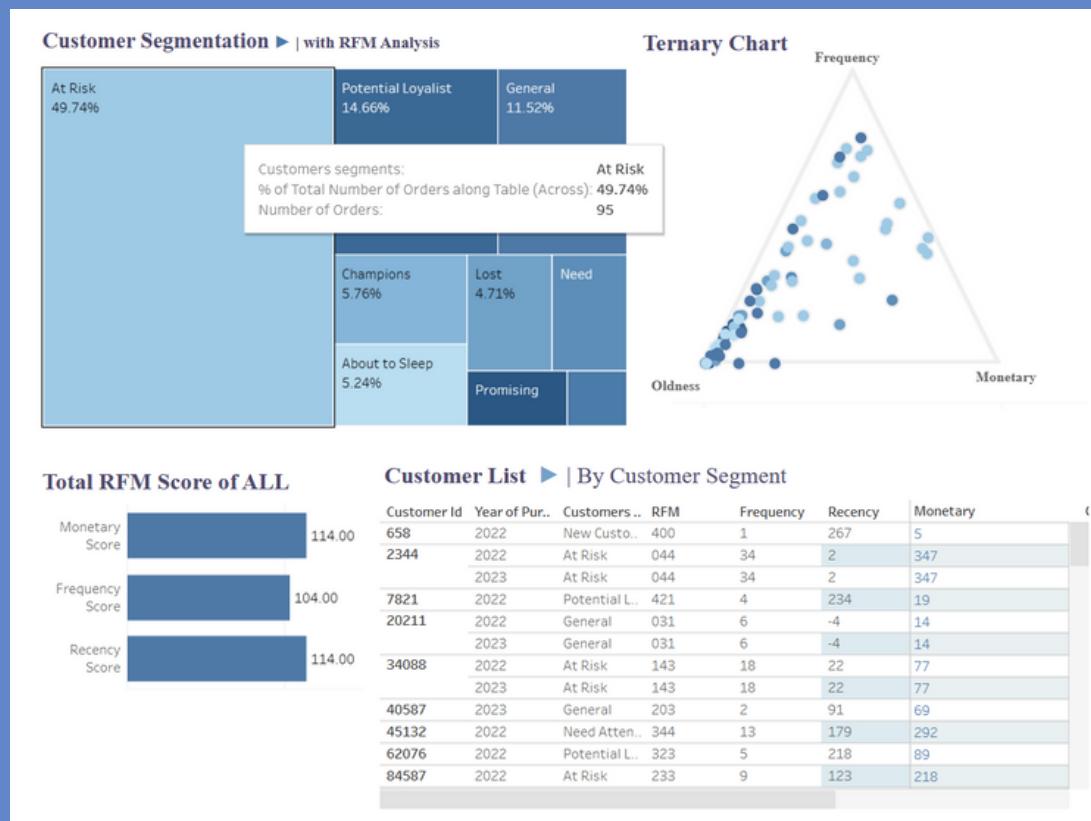
**Segment high value customers based on RFM values.**

**Target Loyal customers interested in products associated with their preferences.**

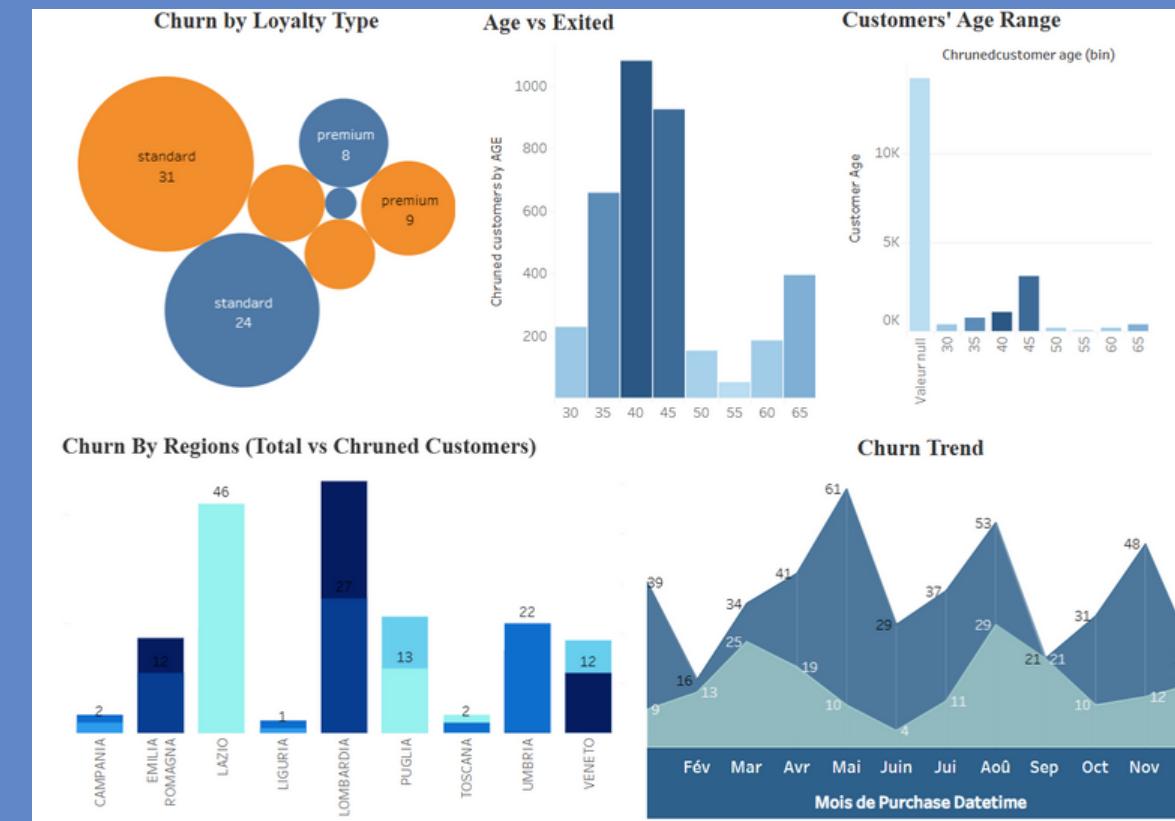


# Click here to Open Interactive Dashboard

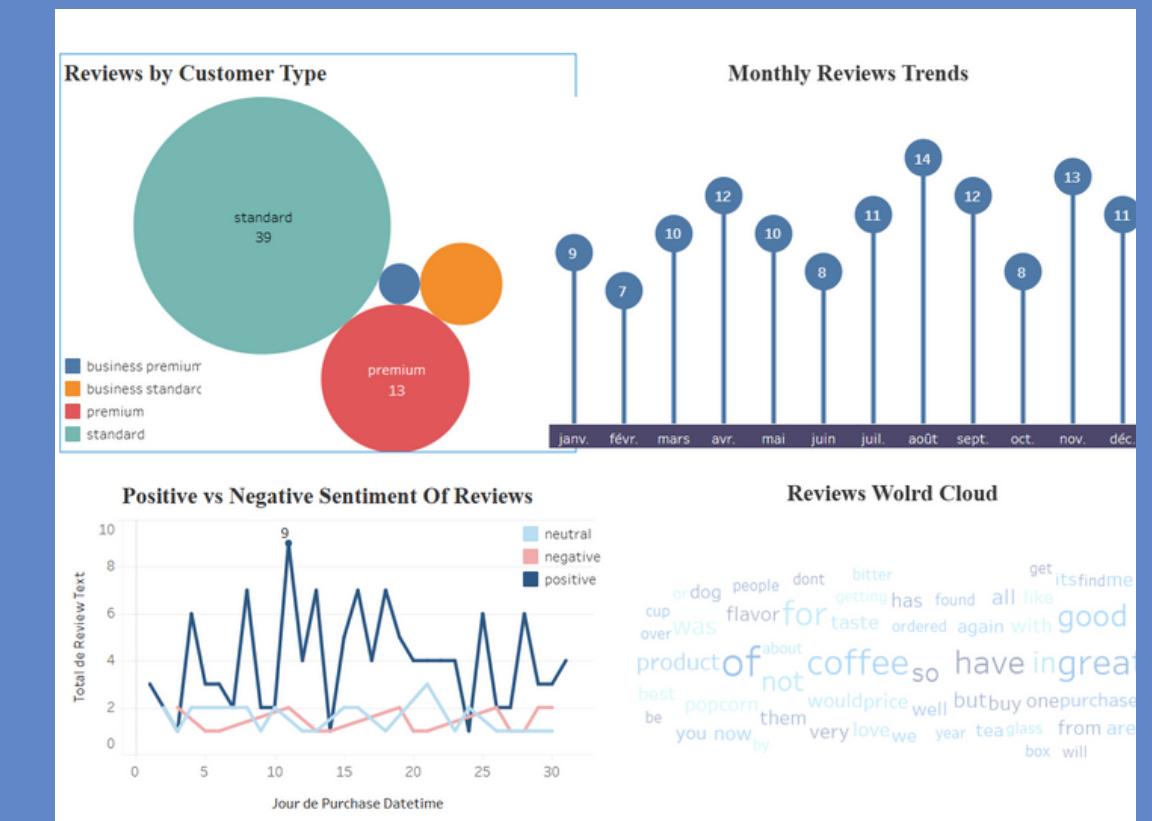
## RFM Analysis



## Churn Analysis



## Sentiment Analysis



# Data Driven Actions

## CRM Optimization

**Optimize customer relationship management practices to cultivate long-lasting, profitable relationships and unlock sales opportunities.**

## Buyer Personas

**Create comprehensive customer profiles to deeply understand their needs and desires, enabling personalized and tailored experiences.**

## Multichannel Engagement

**Strategically leverage multiple communication channels to engage and captivate customers effectively**

## Local Advertising Focus

**Channel resources towards impactful regional advertising campaigns, gaining a competitive edge over national approaches.**



THANK YOU