Marketing Analytics Project

July 2024

Boost your company using ICTs

Valentina Barbera 856780
Vittorio Haardt 853268
Luca Sinanaj 844540

Objectives

1. Churn Analysis

Prevent churn by reminding customers of the company's presence, if they are not actively considering leaving.

2. RFM

Propose a tailor-made campaign for different customer groups to incentivize engagement, reward loyalty, and reengage inactive customers.

3. MBA

Transforming one-shooter customers into repeaters by using insights from their purchases to recommend products, focusing on customer spending behaviour.

4. NLP

Convert detractors' sentiments and engage promoters more by leveraging customer reviews to act both at customer and product level.

Data Driven Strategy

Churn Analyses

RFM

MBA

NLP

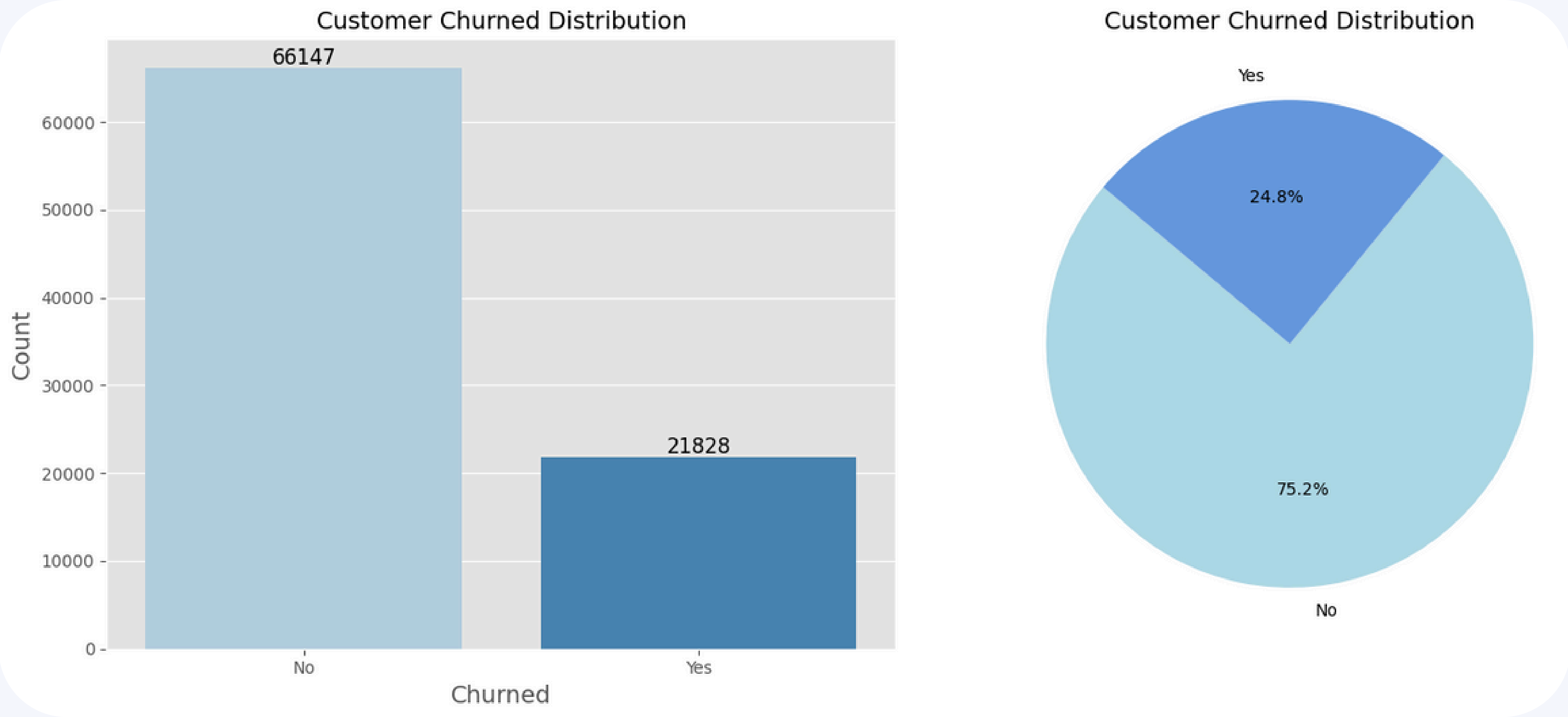
CHURN DETECTION AND ANALYSES

Churn analysis aims to predict and understand reasons for customer attrition, and to develop strategies to improve retention and enhance customer lifetime value.

To select customer churn, we adopted a strategy based on analyzing customer purchasing behaviors over a defined period, excluding one-shooter customers. The primary goal was to identify customers who had not made any further purchases after a specific period, indicating potential churn.

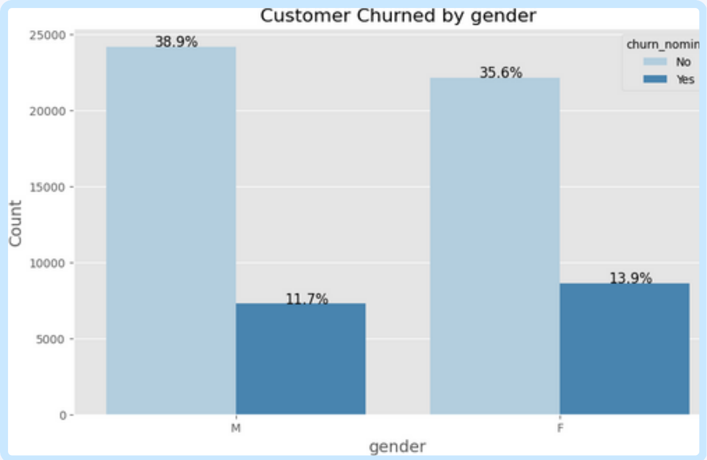
For each customer, we tracked the date of their last purchase. If, after **188 days** from this date, no further purchases were recorded, the customer was labeled as "churned"

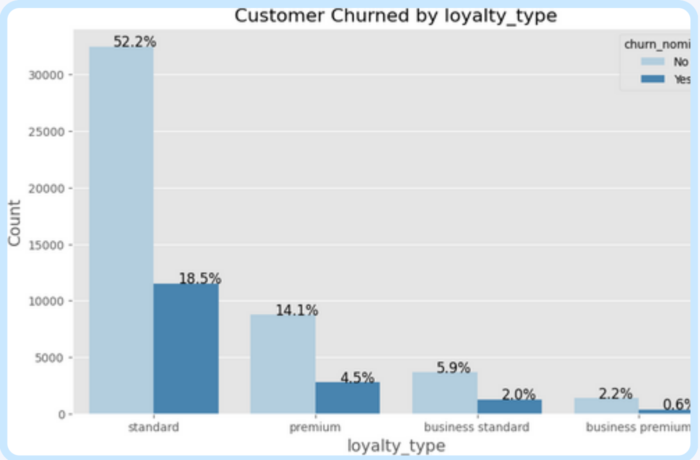
Based on the analysis:
Total Non-Churned Customers: **66'147 (75.2%)**
Total Churned Customers: **21'828 (24.8%)**

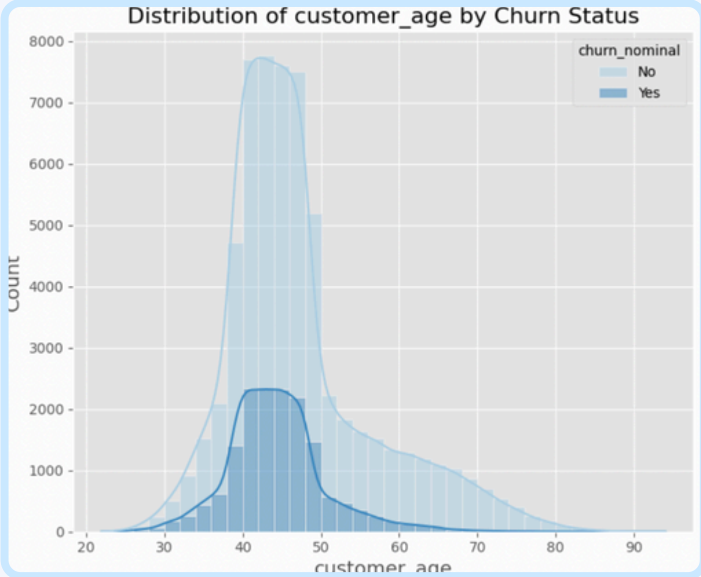


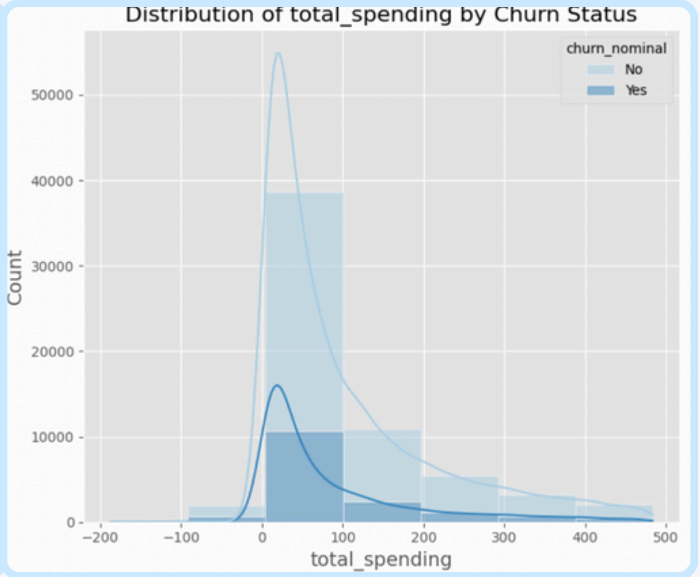
CHURN INSIGHTS

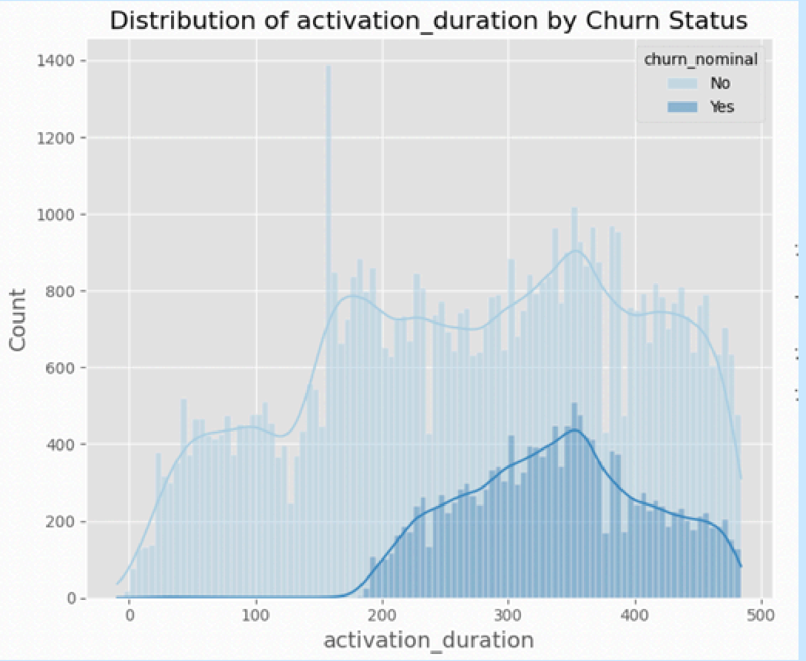
Our **exploratory analysis** of both discrete variables (e.g. gender, loyalty type) and continuous variables (e.g. age, total spending) indicates that while there are some differences in churn rates across these variables, none of them show a strong, decisive influence on customer churn. This suggests that **churn is likely influenced by a combination of multiple factors** rather than any single variable.

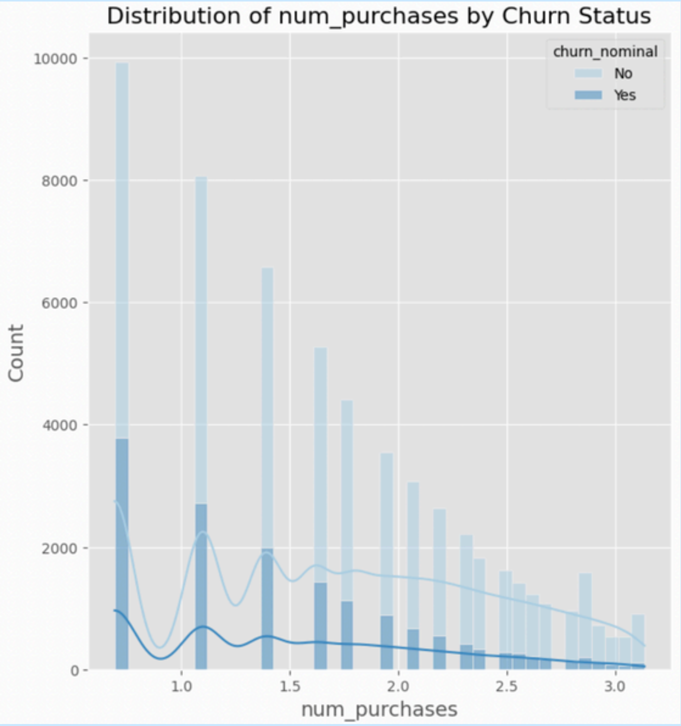












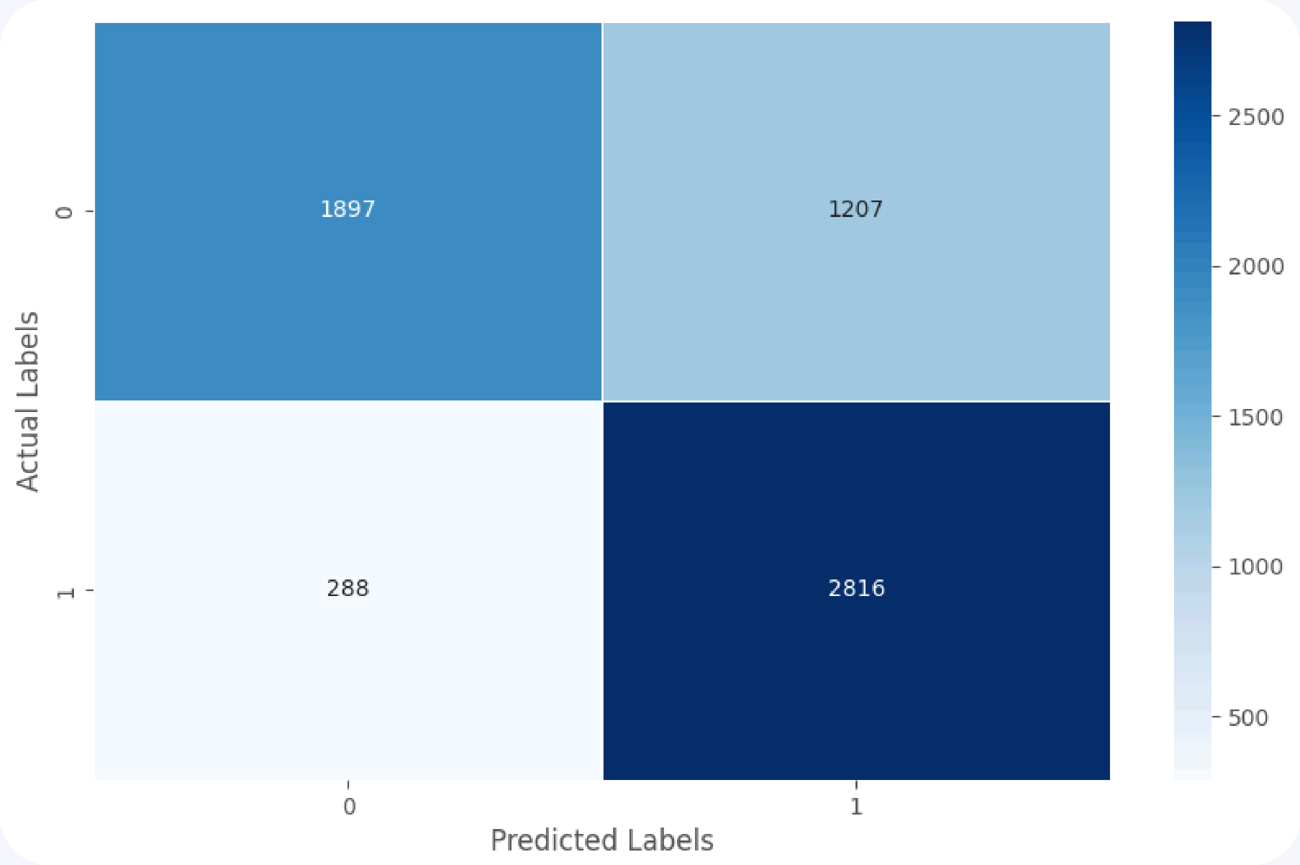
The analysis of **activation duration** and the **number of purchases** provides insights into factors that potentially influence customer churn. **Higher activation durations and fewer purchases are associated with higher churn rates**, suggesting these variables play a significant role in predicting customer churn.

MODELS

The table presents the performance metrics for three different models used to predict customer churn: Decision Tree, Random Forest, and XGBoost.

The **Random Forest model** emerges as the best-performing model for predicting customer churn, with the highest values across all evaluated metrics. It demonstrates a slightly better accuracy and F1-score, and a notably better AUC, indicating its **superior capability to classify churned and non-churned customers accurately**.

	Accuracy	F1-score	AUC
Decision Tree	75,5%	75,1%	82%
Random Forest	75,8%	75,3%	83%
XGBoost	74,6%	74,3%	82%



The confusion matrix provides a **detailed breakdown of the performance** of the **Random Forest** model by showing the number of correct and incorrect predictions for each class, by which we can calculate a **recall of 90.7% means that the model identifies 90.7% of the customers who actually churn.**

MARKETING STRATEGY

Churn Costumers

Maintaining almost more than 21'000 inactive customers on a newsletter list can incur significant costs. As an innovative approach to managing customer engagement and reducing churn, we propose a strategy similar to the one depicted in the image.

- **Personalized Unsubscription Campaign**: During the holiday season, send a friendly and engaging email to inactive subscribers with a message similar to "**Happy Holidays - We're Unsubscribing You!**"
- Content: The email will highlight the benefits of being a part of our newsletter and provide an **easy way for customers to resubscribe if they wish to continue receiving updates.**

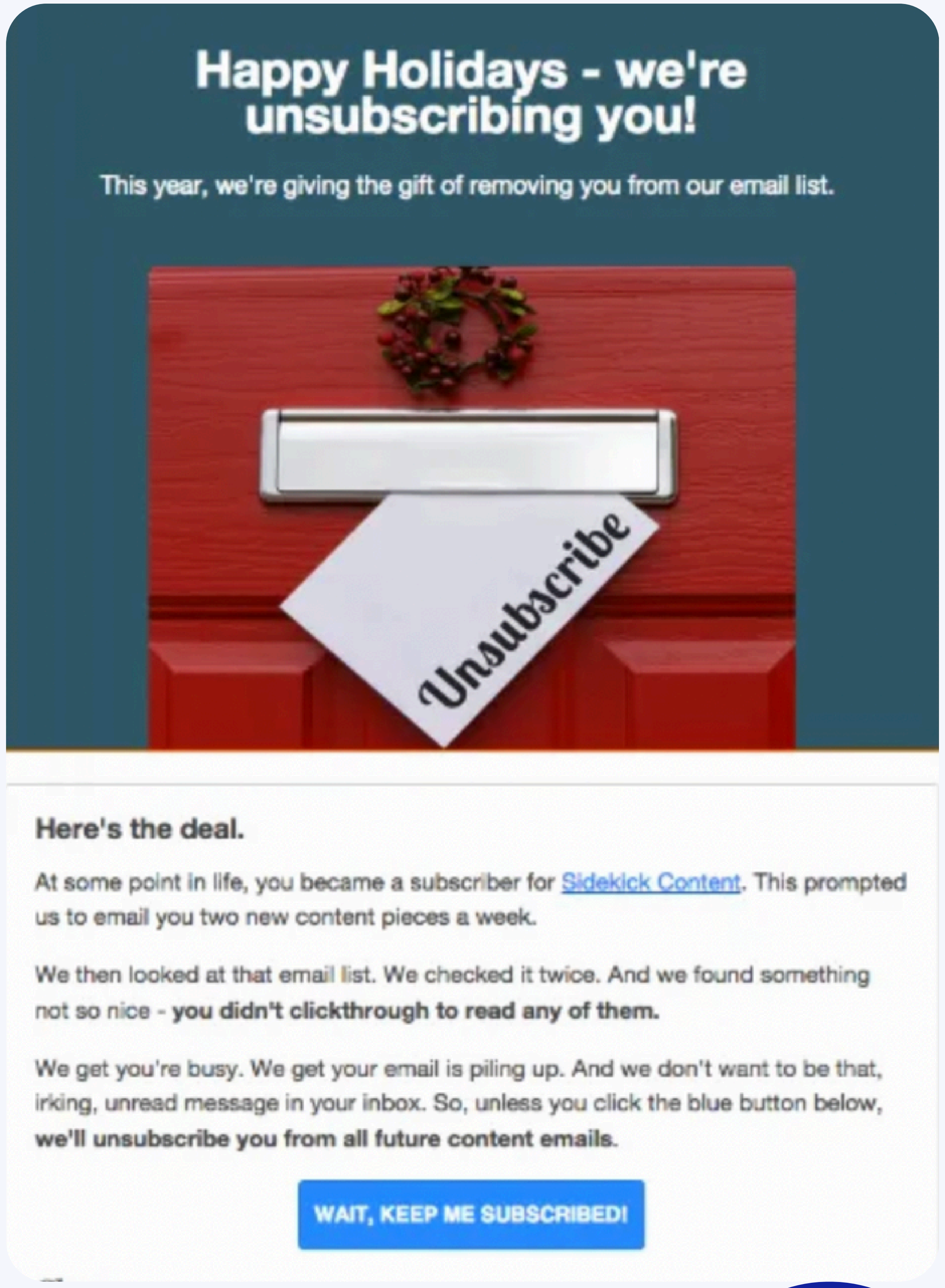
Target

Total Churned Customers: **21'828**

Customers without Privacy Flag: Only 1.3% of the churned customers did not provide the privacy flag. **284 churned customers cannot be contacted via email due to the absence of the privacy flag.**

Benefits

- Cost Reduction
- Improved Engagement
- Customer Respect
- Data Cleanliness



Happy Holidays - we're unsubscribing you!

This year, we're giving the gift of removing you from our email list.

Here's the deal.

At some point in life, you became a subscriber for Sidekick Content. This prompted us to email you two new content pieces a week.

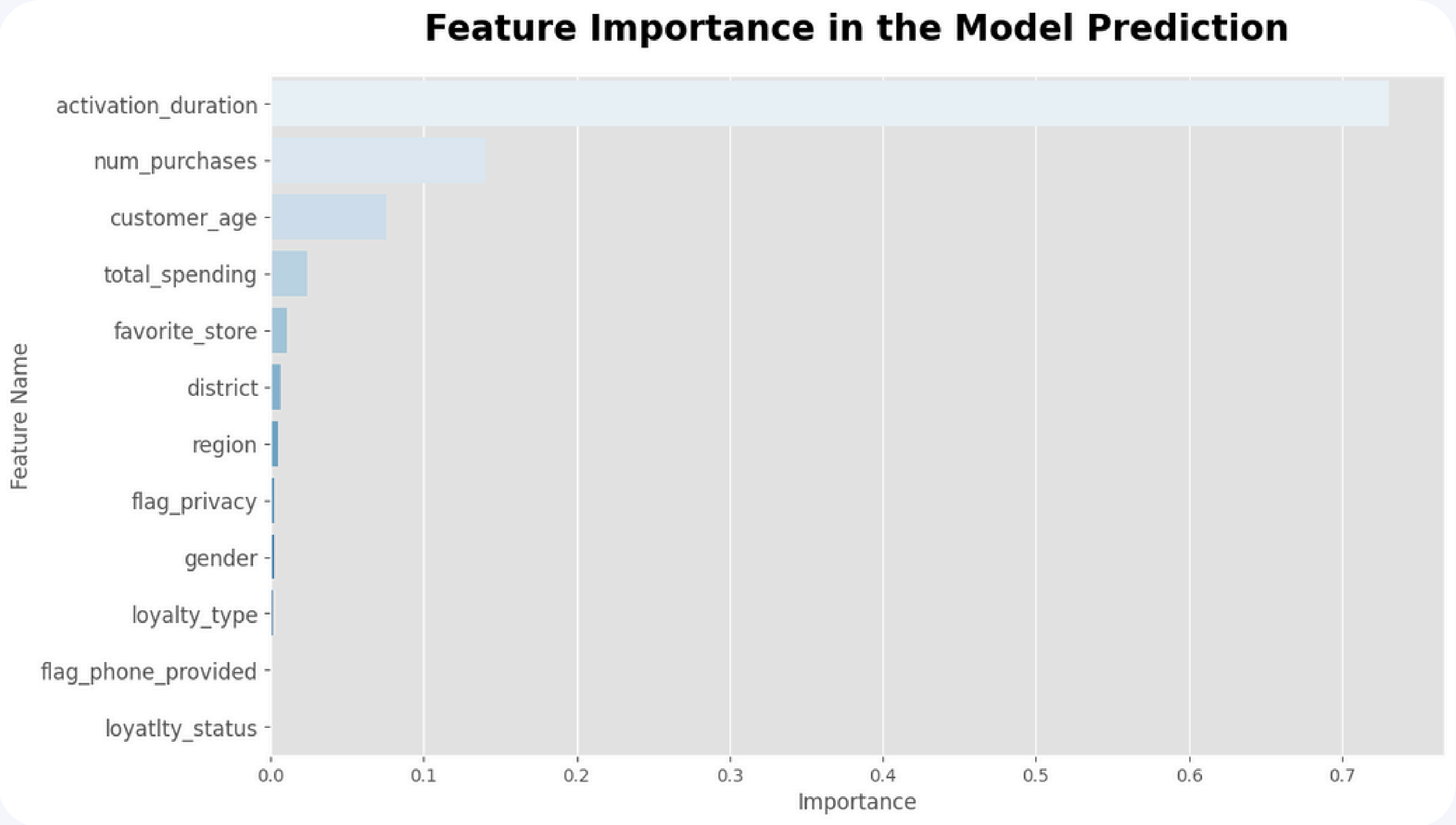
We then looked at that email list. We checked it twice. And we found something not so nice - **you didn't clickthrough to read any of them.**

We get you're busy. We get your email is piling up. And we don't want to be that, irking, unread message in your inbox. So, unless you click the blue button below, **we'll unsubscribe you from all future content emails.**

MARKETING STRATEGY

Churn Prediction

The chart displays the **importance of various features** used in predicting customer churn for Decision Tree, Random Forest, and XGBoost. **The importance of each feature is determined by how much it contributes to the model's predictive power.** This helps us tailor our marketing strategies to address key factors influencing customer churn.



- **Activation Duration**: highest importance, indicating that **the length of time since customer activation is a critical factor in predicting churn.**
- **Number of Purchases**: significantly influences churn prediction. **Customers with fewer purchases are at a higher risk of churning.**

Milestone Rewards Program

- Action: **Implement a rewards program that recognizes and rewards customers at various activation duration milestones.**
- Components:
  - Reward customers at key milestones (e.g., 30 days, 90 days, 180 days) with points, discounts, or gifts.
  - Highlight the benefits of **continued engagement and loyalty.**
  - **Encourage customers to reach the next milestone** with preview offers or sneak peeks at upcoming rewards.

Data Driven Strategy

Churn Analyses

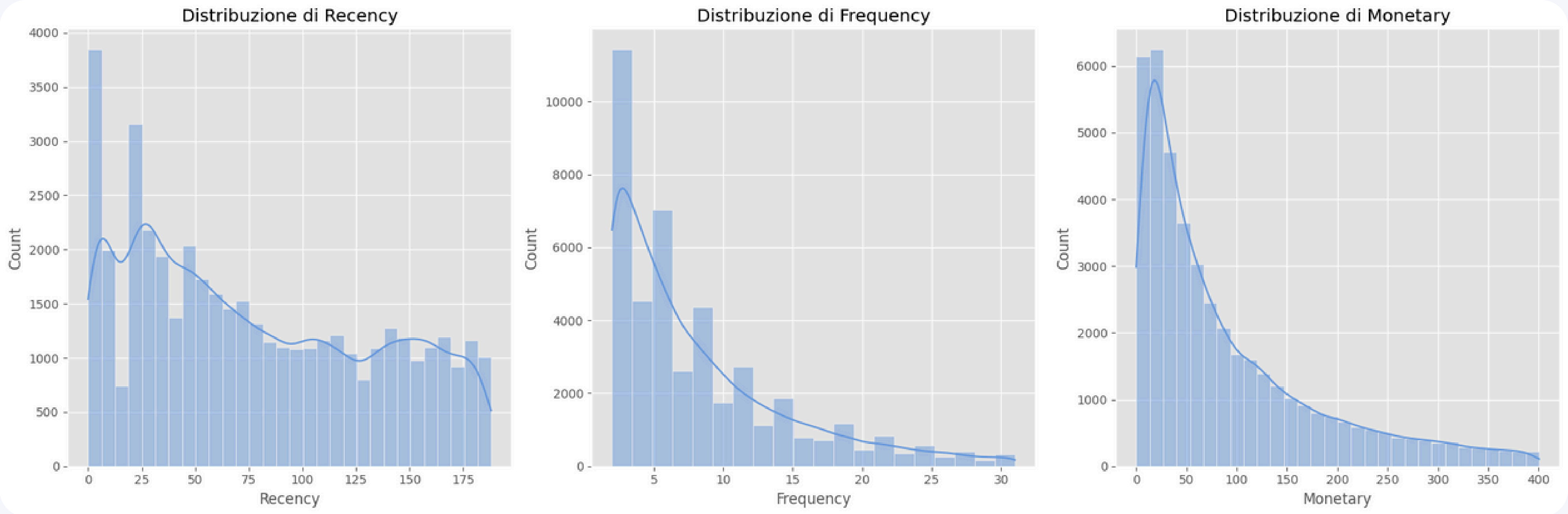
RFM

MBA

NLP

RECENCY, FREQUENCY, MONETARY

After excluding churn customers, we have implemented RFM analysis, in order to make data-driven decisions to enhance our marketing strategies, improve customer retention, and **optimise our resources focusing our efforts on the most valuable customer segments.**



<u>**Recency:**</u> The distribution is right-skewed with **many recent purchasers**, but a long tail indicates customers who haven't purchased in a while, posing a churn risk.

<u>**Frequency:**</u> Most customers have low purchase frequencies, suggesting **opportunities to boost engagement and repeat purchases through targeted promotions.**

<u>**Monetary:**</u> **The majority of customers have lower spending**, while a small segment contributes significantly to revenue. Retention strategies should focus on high-value customers and increasing spending among lower-value ones.

CLUSTERING TECHNIQUES

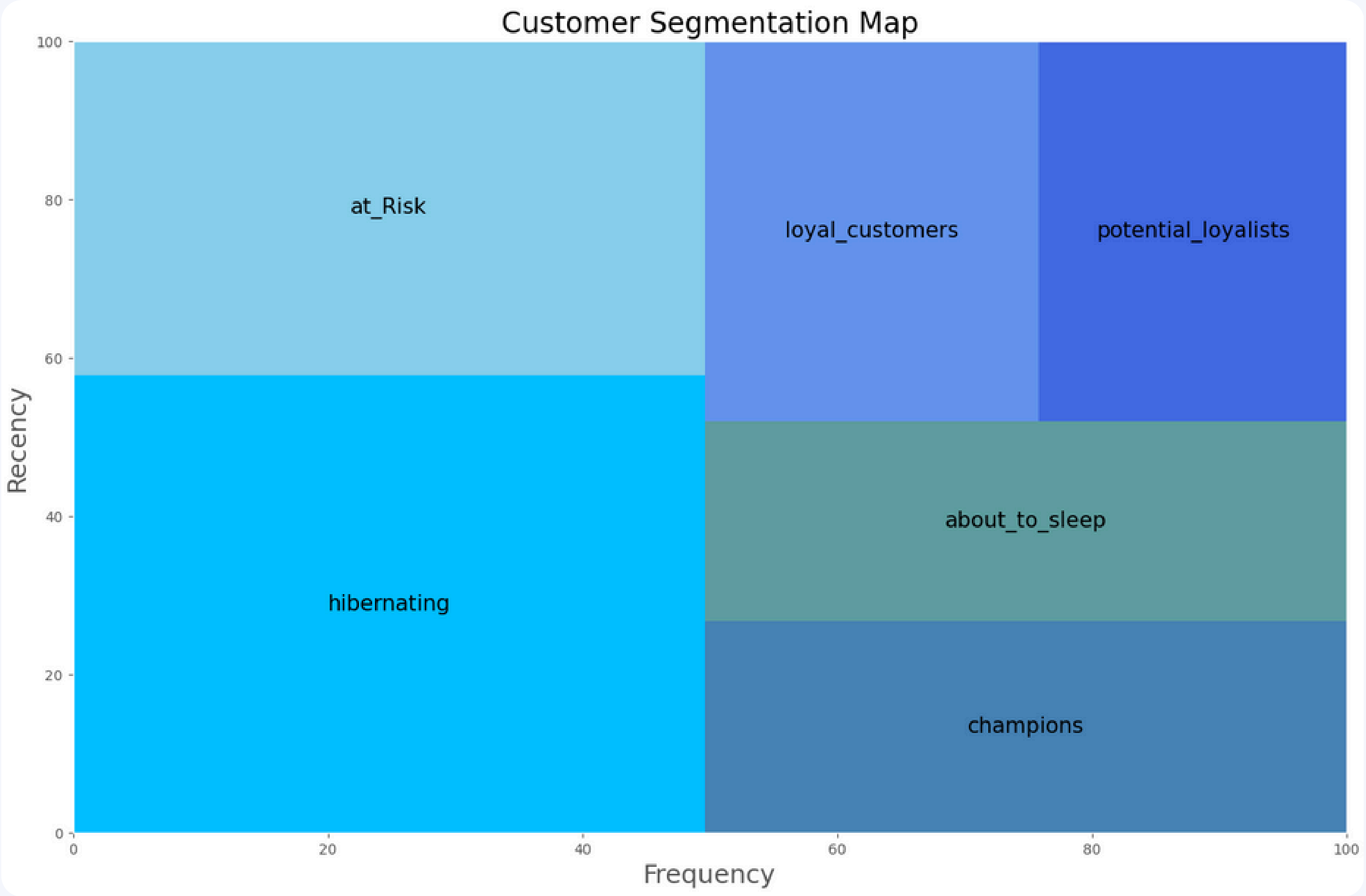
- **Manual**: Manually defining the cluster based on predefined RFM score ranges.
- **K-Means**: unsupervised machine learning algorithm that finds patterns and groups customers into clusters by minimising the variance within each cluster and optimising clusters number.

MANUAL CLUSTERING

We assigned scores to each RFM metric. Scores range from 1 to 4, **where a higher score indicates better performance on that metric.**

We combined the individual scores into a single one. Eg. A customer with R=4, F=4, and M=4 would have an RFM score of 444, indicating a **high-value** customer.

RFM Pattern	Segment Name	Description
[1-2][1-2]	Hibernating	Low R, F, and M values. They are inactive and haven't engaged much recently.
[1-2][3-4]	At Risk	Low R but high F and M values. They used to be valuable but haven't purchased recently.
3[1-2]	About to Sleep	Medium R and low F and M values. They are at risk of becoming inactive.
[2-3][3-4]	Loyal Customers	Medium to high R and high F and M values. They are regularly engaged and valuable.
[3-4][1-2]	Potential Loyalists	High R but low to medium F and M values. They have recently engaged but need encouragement to increase their value.
4[3-4]	Champions	High R, F, and M values. They are the most valuable and loyal customers, making frequent and high-value purchases.



The segmentation map represents the distribution of customer segments based on Recency and Frequency.

Performance Metrics:

- **Silhouette Score:** 0.074 – Clusters need refinement.
- **Calinski-Harabasz Score:** 31,328 – Clusters are well-separated.
- **Davies-Bouldin Score:** 3.329 – Moderate overlap.

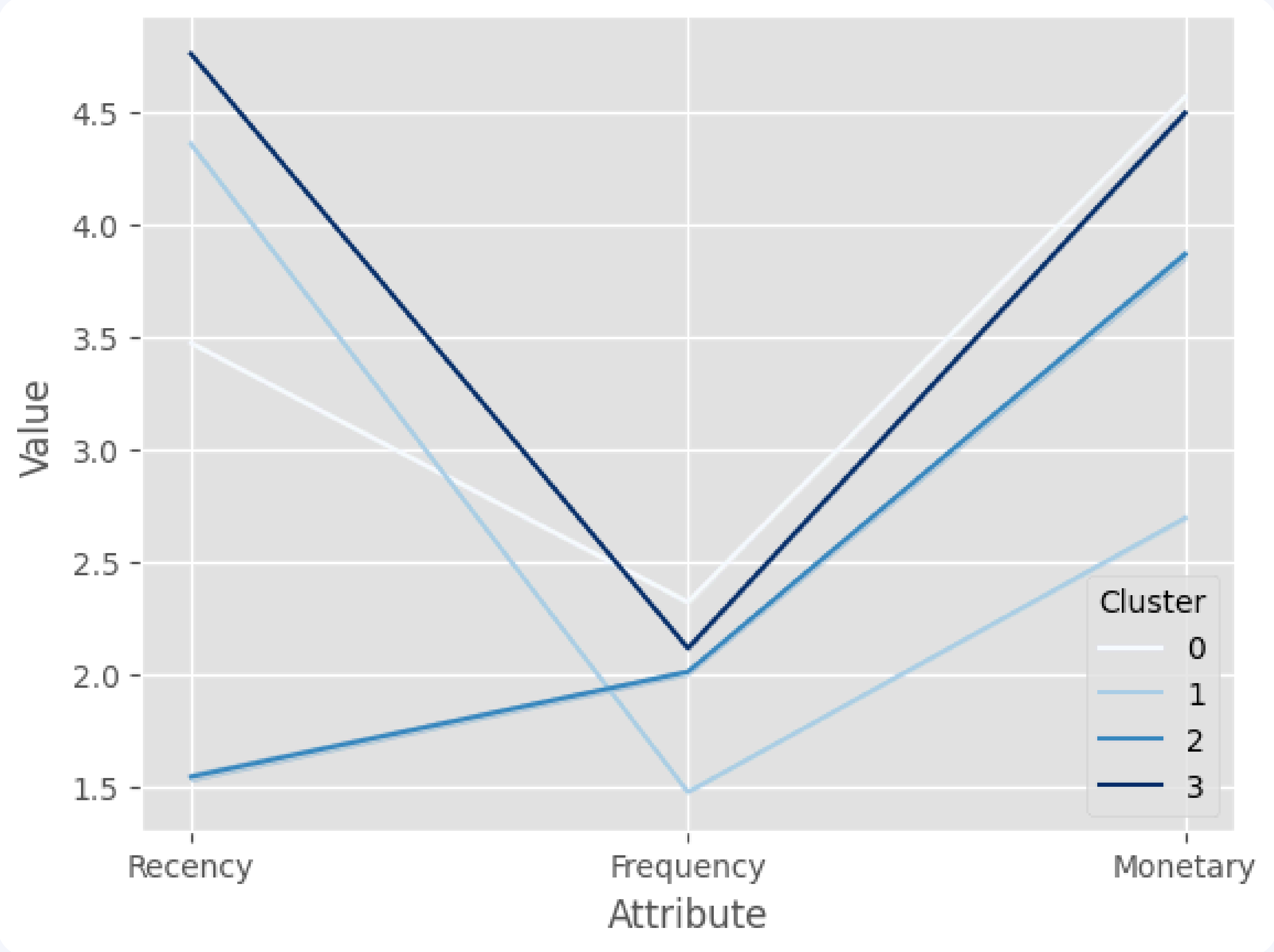
Due to the moderate cluster definition, we plan to apply K-Means clustering.

K-MEANS CLUSTERING

To improve our customer segmentation, we applied K-Means clustering, a popular **unsupervised machine learning algorithm.** Starting by preparing the dataset, focusing on the RFM values for each customer and applying **data normalization** to ensure all features contributed equally to the clustering process.

We used the **Elbow Method** to determine the optimal number of clusters **which is 4.**

Cluster	Recency	Frequency	Monetary	Description
0	Moderate	High	High	Customers have not made recent purchases, have a low frequency of purchases, but have a moderate monetary value. **They might have made few high-value purchases a while ago.**
1	High	Low	Low	These customers have low recency, frequency, and monetary values, indicating low engagement and spending. **They may be at risk of churning.**
2	Low	Moderate	Moderate	Customers make frequent purchases but have lower monetary values. **They are regular buyers but tend to spend less per purchase.**
3	Very Low	High	High	Customers have low recency, high frequency, and high monetary values. **They are highly engaged, frequent purchasers with significant spending.**



The image presents a parallel coordinates plot showing the clustering results based on Recency, Frequency, and Monetary values.
Performance Metrics:

- **Silhouette Score:** 0.307 - Moderate cluster definition, indicating reasonable but improvable segmentation.
- **Calinski-Harabasz Score**: 30'524 High, suggesting well-separated clusters.
- **Davies-Bouldin Score:** 1.121 Low, indicating distinct and well-defined clusters.

MARKETING STRATEGY

We propose a **tailored loyalty program to different customer segments**. The goal is to incentivize engagement, reward loyalty, and re-engage inactive customers effectively.

Champions: Customers with high RFM values. They are the most valuable and loyal customers.

- **Loyalty Type Upgrade**: **Upgrade to Premium loyalty status on their next purchase**. Encourages continuous engagement and rewards top customers for their loyalty.
- **Special Discounts**: **Offer significant discounts during holiday seasons**. Rewards high spenders, encourages substantial purchases, and strengthens customer loyalty.

Loyal Customers / Potential: Customers with medium to high R and high F and M values. They are regularly engaged and valuable.

- **Cross-Selling:** 70% discount on product Y when they purchase product X. Promotes cross-selling, **increases average order value, and enhances customer satisfaction, could be enhanced by following technique of MBA.**

At Risk / About to Sleep: Customers with medium RFM values, indicating they are at risk of becoming inactive.

- **Incentives**: Offer double loyalty points and a discount on the next purchase if they buy within a specific timeframe and meet a minimum purchase amount. **Motivates re-engagement and increases purchase frequency.**

Hibernating: Customers with low RFM values. They are inactive and haven't engaged much recently.

- **Use cookies** to monitor visits to similar sites and send communications about special offers on related products. Recaptures interest, **brings back inactive customers, and increases engagement.**

New Customers: Customers who have made their first purchase recently.

- **Welcome Offer**: Provide **free shipping** on their first purchase if they spend a minimum amount. **Attracts new customers and reduces barriers to their first purchases.**

LOYALTY PROGRAM

Manual Clustering

Segment	Number of Customers
Champions	6,525
Loyal Customers	5,692
Need Attention	6,581
At Risk	9,822
Potential Loyalists	6,520
Hibernating	15,216

K-Means Clustering

Cluster	Number of Customers
0	11,529
1	16,385
2	5,920
3	16,522

Data Driven Strategy

Churn Analyses

RFM

MBA

NLP

MARKET BASKET ANALYSIS (MBA)

Market basket analysis is a strategic data mining technique used by retailers to enhance sales by gaining a deeper understanding of customer purchasing patterns.

OBJECTIVE

- Find the most associated products using the FP Growth algorithm an advancement of the APriori algorithm that is more efficient.
- We are also interested in analysing the correlations between products (Product_Id) and the respective net prices of each product in order to understand if there are patterns with the antecedent net price and cosequent net price.

METRICS

- **SUPPORT** : frequency with which the rule appear in the dataset. We have considered a support equal to 0.002.
- **CONFIDENCE** : measure the frequency with which the rule is true
- **LIFT** :how much greater the probability of finding the antecedent and the consequent together than would be expected if they were independent.

MARKETING STRATEGY

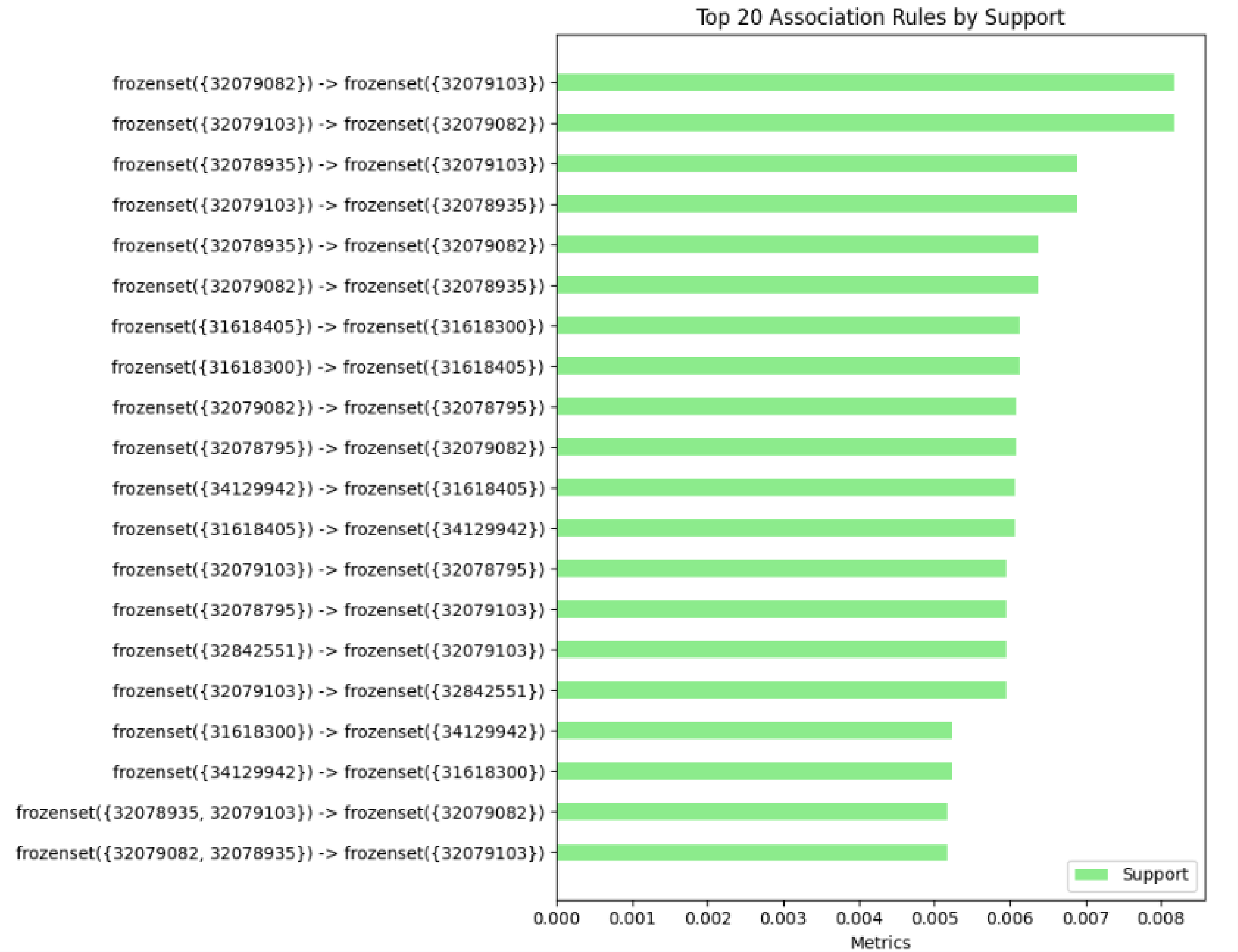
The rules were obtained considering the **customers who had made more than 10 orders** because they represent a more consistent and reliable sample than those who have made only a few purchases.

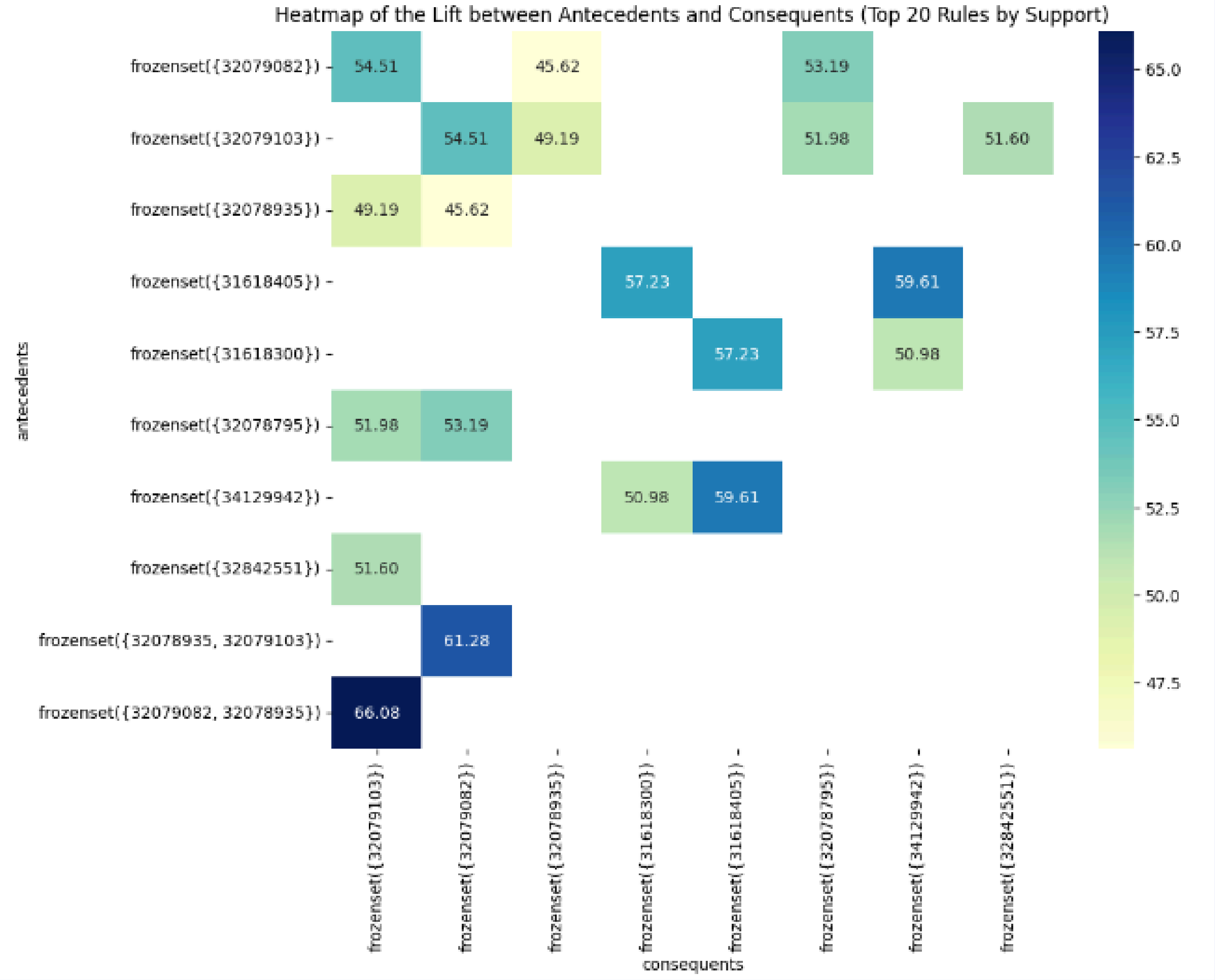
The rules obtained were applied :

- on the one shooter customers, in order to transform them into repeater customers.
- on the **500 customers who had spent more**, in order to lead them to further purchase.
- to suggest products for each customer in the dataset with more than one purchase.

TOTAL RULES

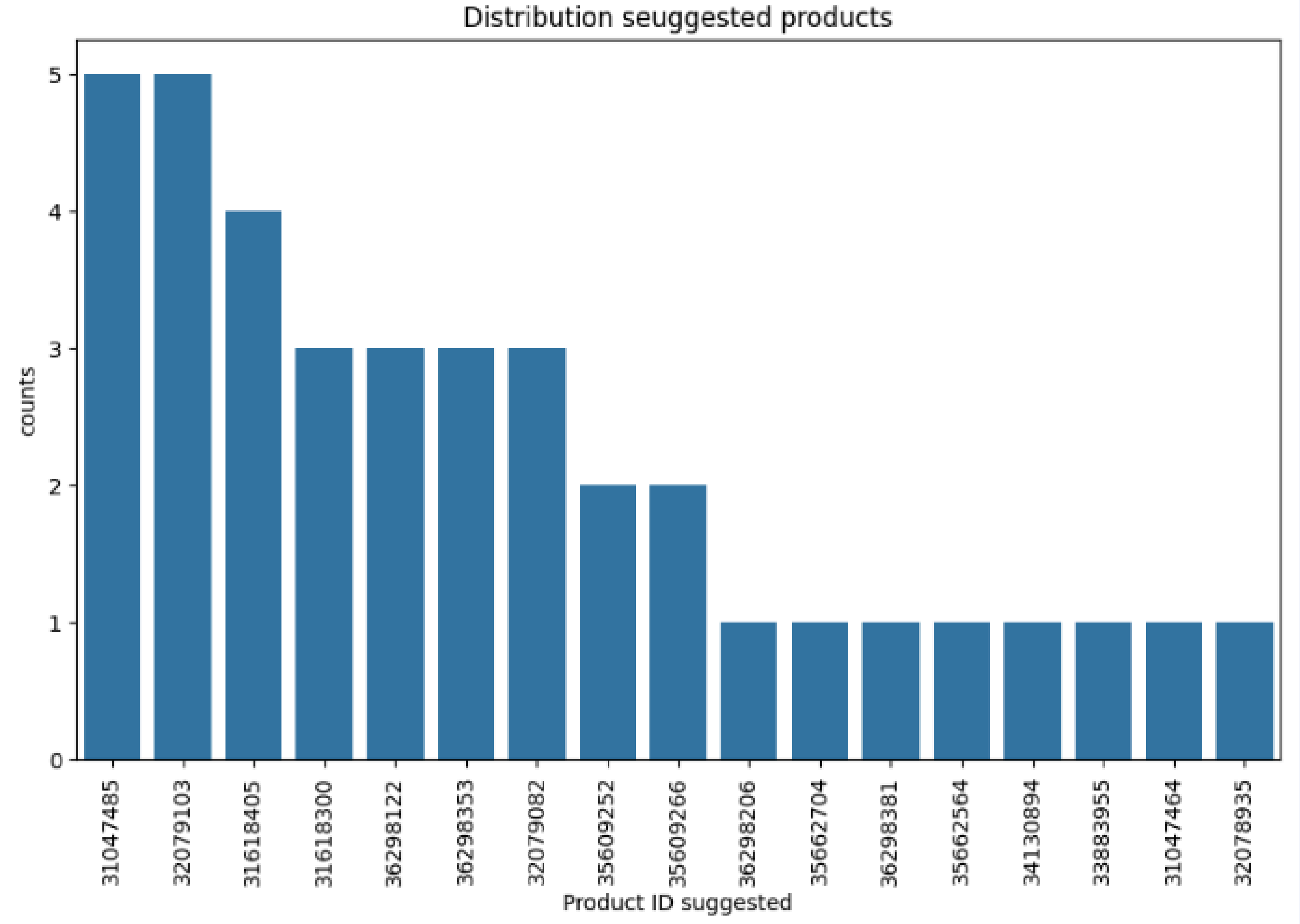
674

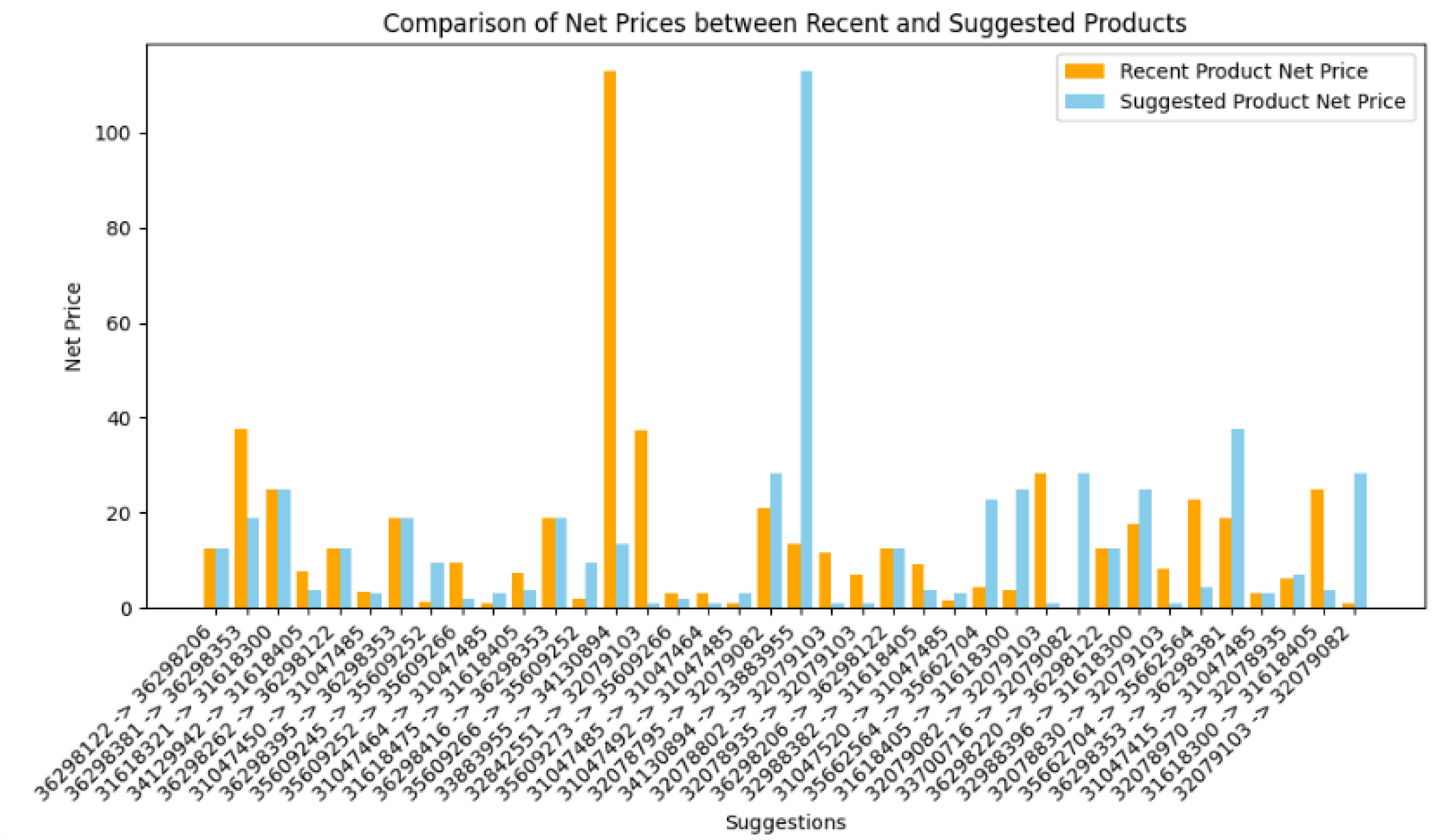




The following plots represent the 20 rules with the highest support value, so the strongest ones.

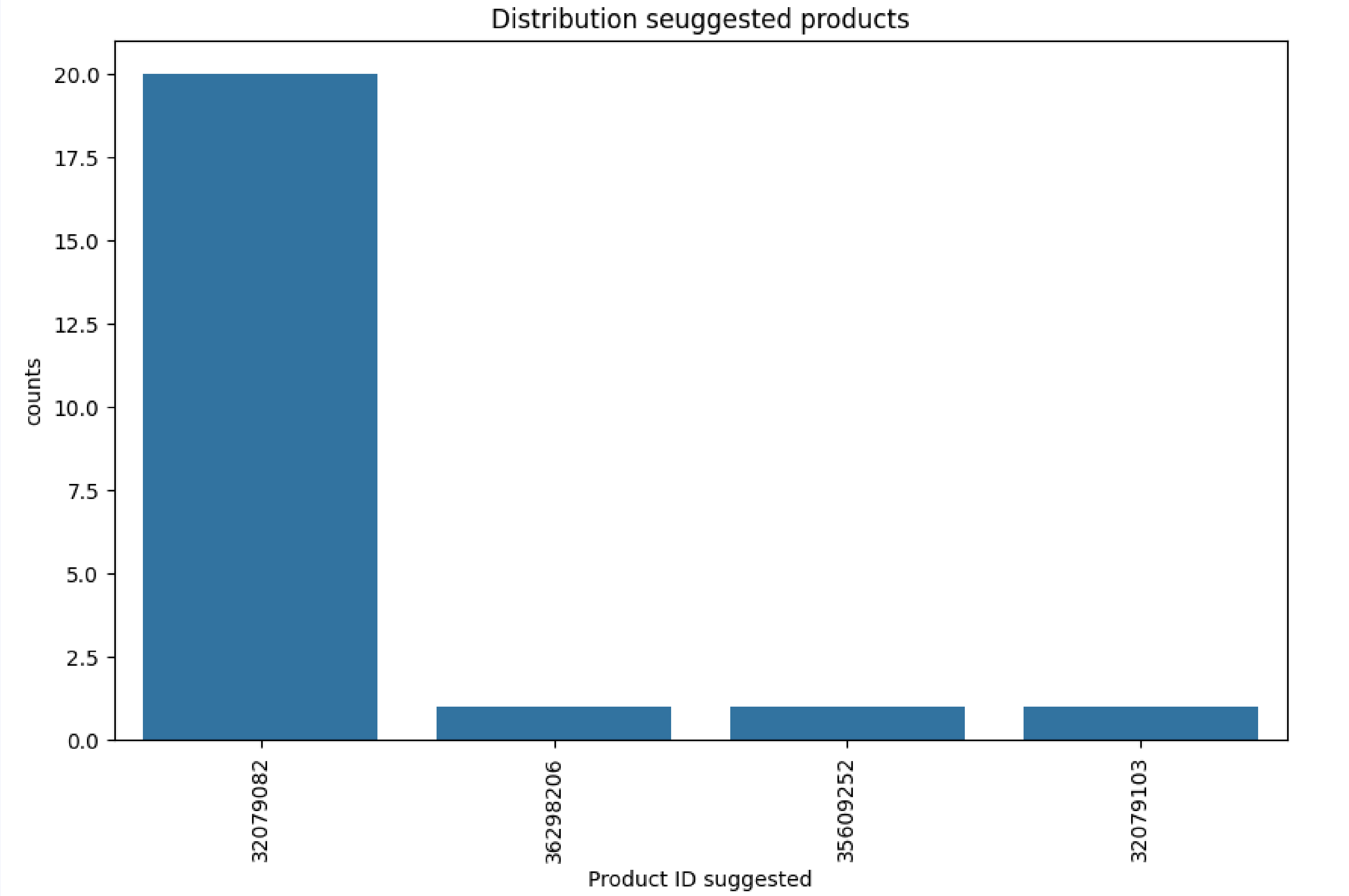
ONE SHOTER CUSTOMERS

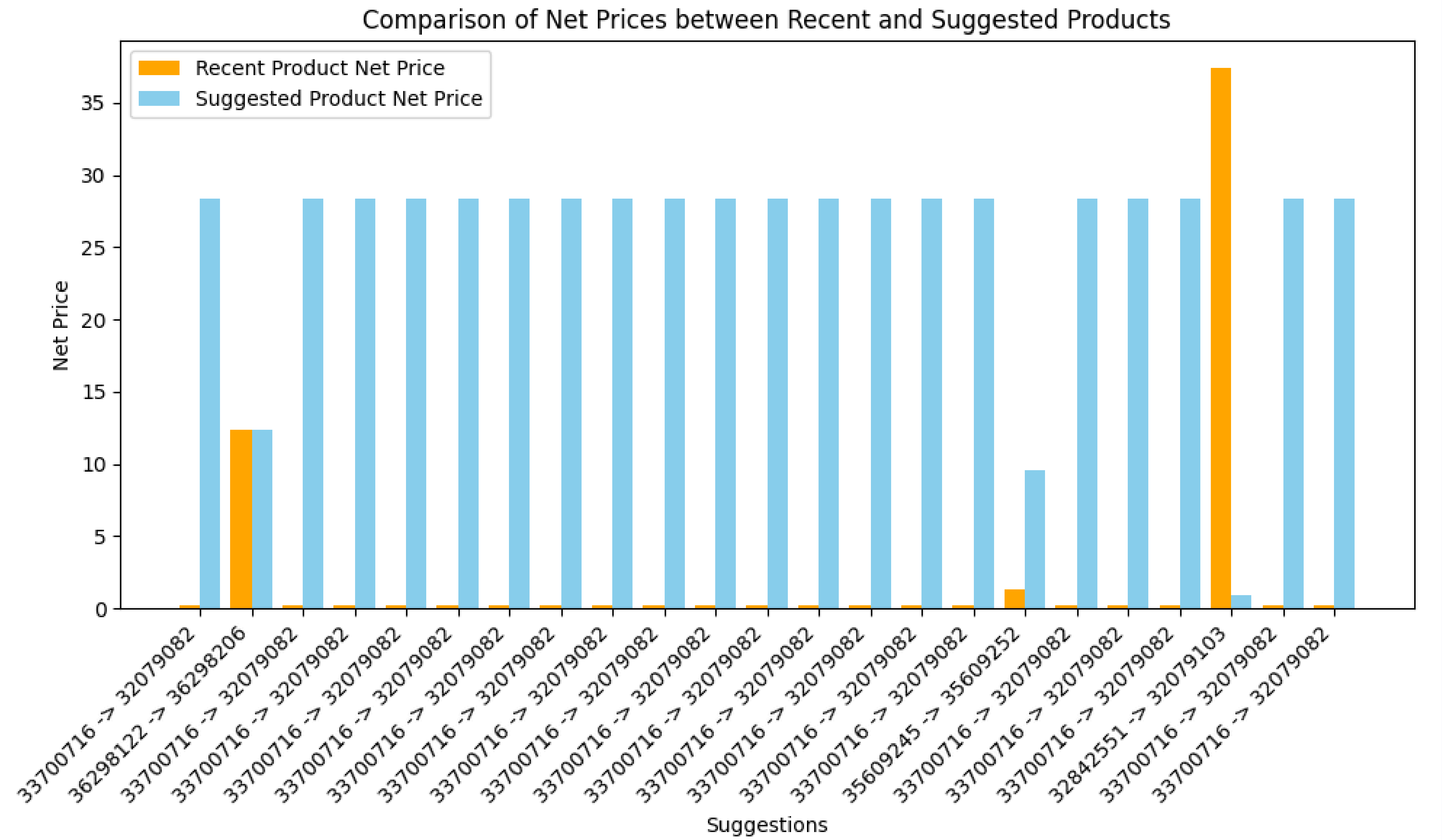




- The first plot represent the distribution of the suggested products for the one shoter customer, the first two products were suggested for 5 customers each of them.
- The second plot represent in orange the net price of the antecedent that represent the last product purchased by the customer and the blue bar the net price of the suggeste product, we can see that only few antecedents had a net price similar or lower to the consequents but we don’t have a particular pattern in the net prices of the products.

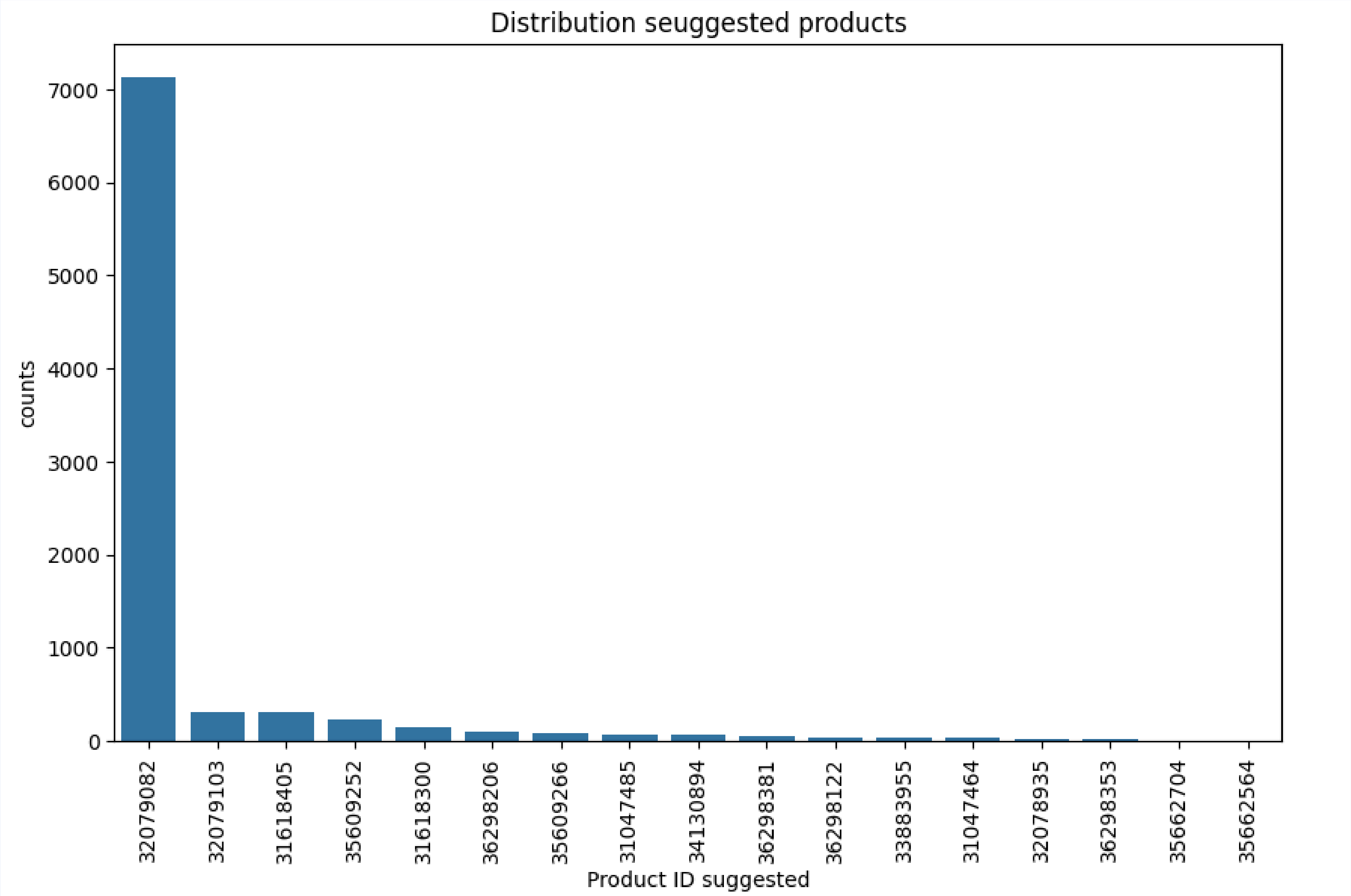
TOP 500 CUSTOMER WHO HAD SPENT MORE

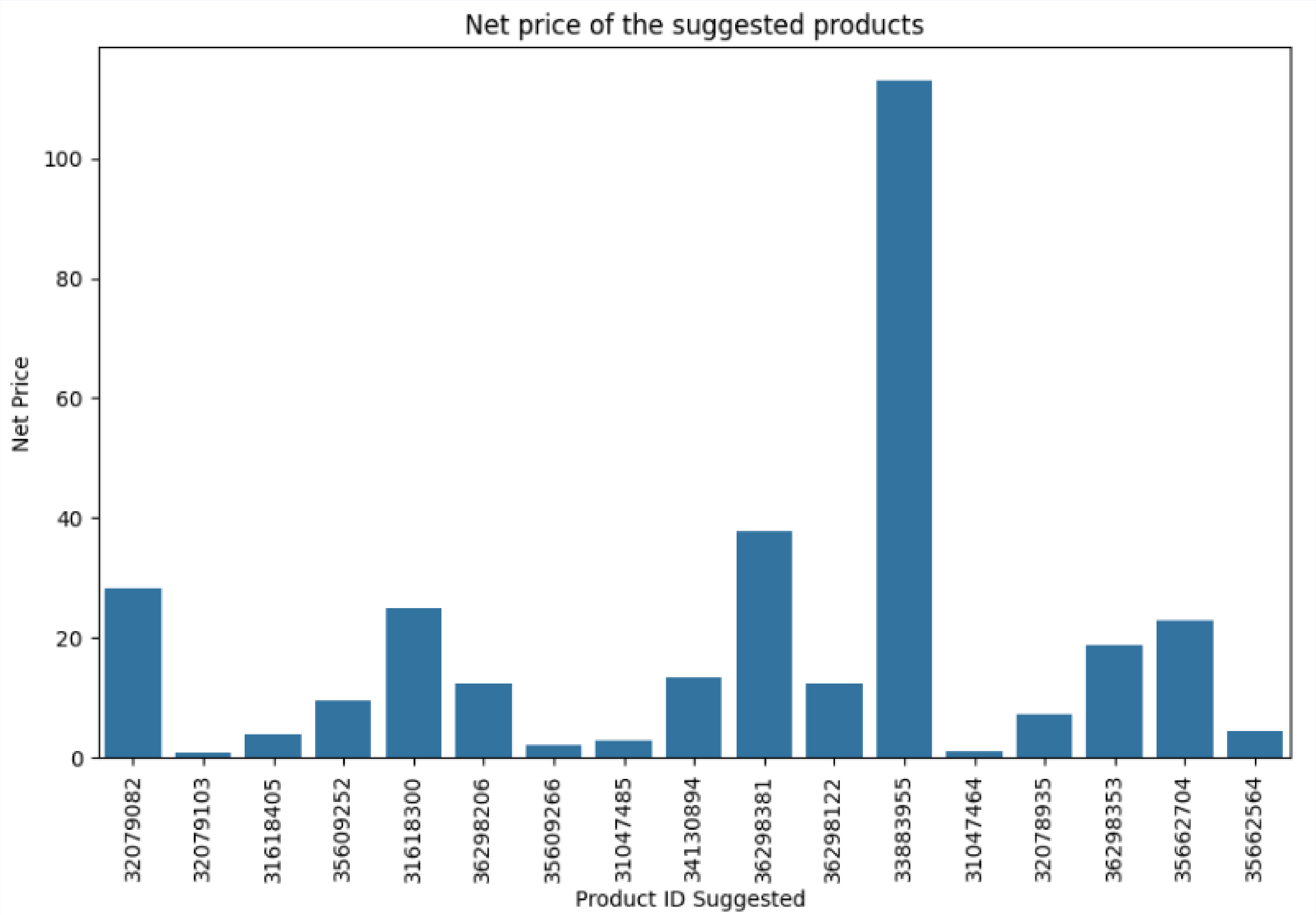




- The most suggested product is also the one with the highest net price, this can be a possible correlation with the type of customers chosed because are the ones that had spent more.
- The diferences between the net price of the suggested products and the recents ones, in this case, are significants.

SUGGESTED PRODUCTS FOR EACH CLIENT





- From the first graph we can see that the most recommended product was selected for no less than 70000 customers.
- The second plot show the net price of each product suggested and the most expensive one was suggested for less than 10 customers.
- We can conclude that there is not a significant correlation between the product more suggested and the respective net price.

Data Driven Strategy

Churn Analyses

RFM

MBA

NLP

NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a field of AI that empowers computers to interpret, process, and generate human language in meaningful way. It is utilised to extract insight form users comments and to monitor strategies based on those insights.

TECHNIQUES

- **Topic Modelling**: discover conversation topics in text, used for the primary analysis.
- **Sentiment Analysis**: prediction model to automatically classify the opinion in comments, monitor the effectiveness of implemented strategies.

OBJECTIVES

Insight form the processed comments are used for two key strategies:

- **Product Focus**: identify leading product to attract clients and identify underperforming products to be individually investigated.
- **Customer Focus**: find promoters to become ambassadors and to find detractors to covert their opinion.

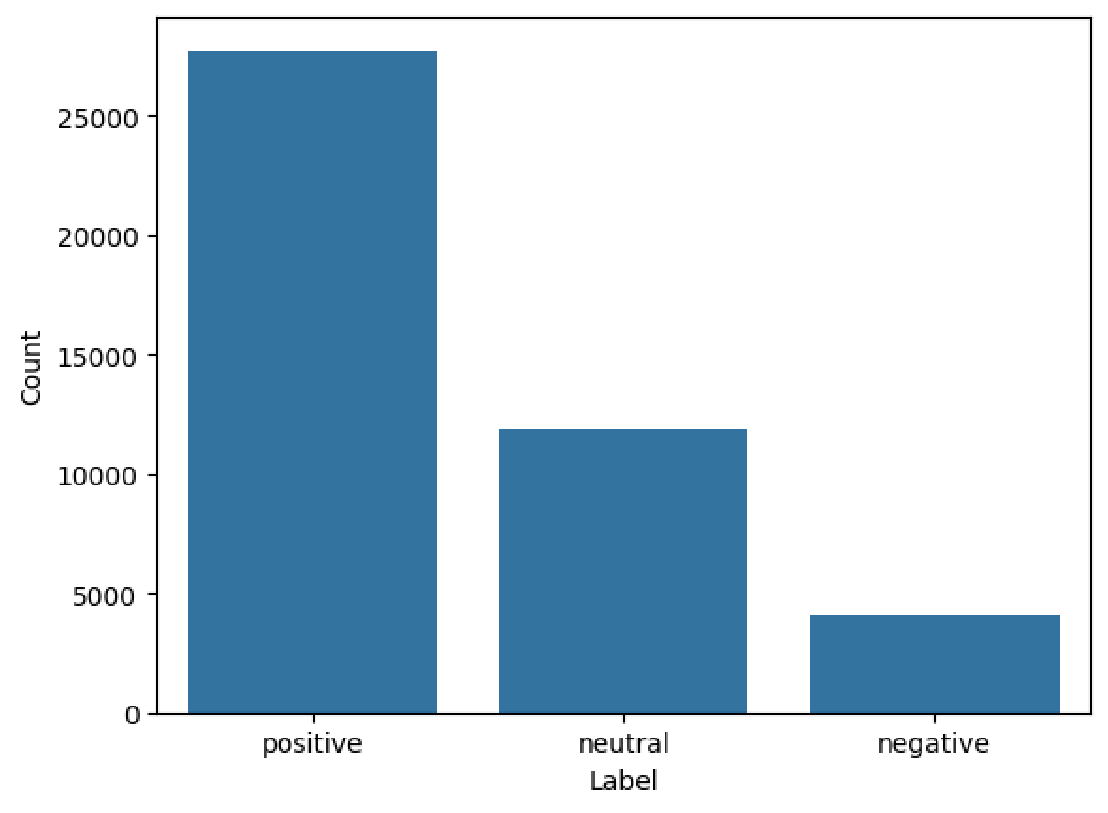
TOPIC MODELLING

After preprocessing 43,682 data points, the topic modeling provided valuable insights into the discussed topics within the comments. We were particularly interested in identifying the negative and positive comments, which totaled 4,076 and 27,727, respectively.

Since each user commented only once, the detractors and promoters were easily identified based on their single negative or positive comment. Consequently, the insights derived from this analysis are utilized for both product-focused and customer-focused strategies.

PREPROCESSING

- Remove duplicate
- Remove punctuation
- Lower case conversion
- Stop words removal



Class unbalance was assessed in SA, with under sampling.

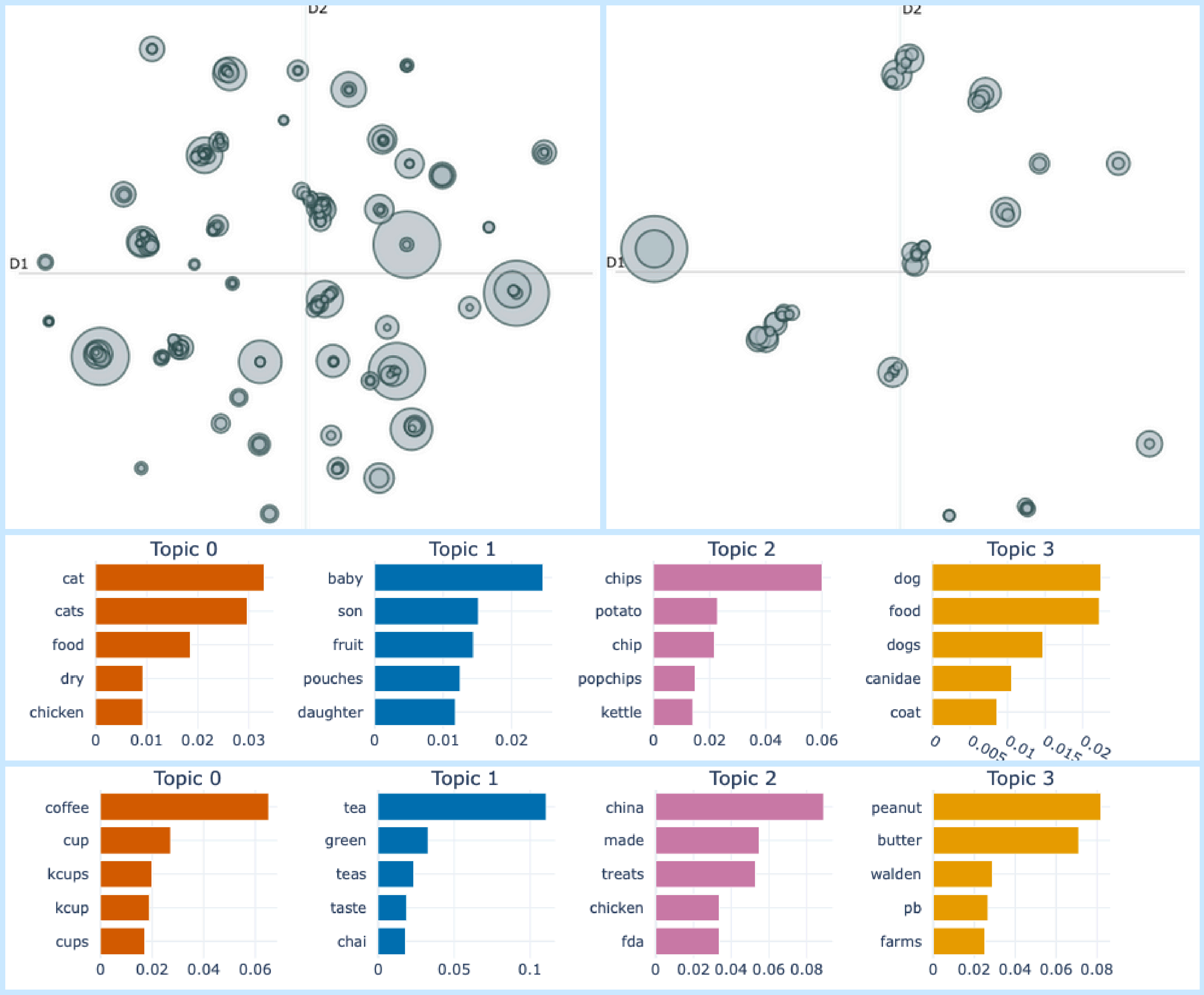
TOPIC ANALYSIS

Positive and negative comments were separated, and the BERTopic model was applied individually to each set to gauge the topics. From this, the 9 most discussed topics (> 80 comments) were considered important. The main topics were highly polarised (in 2D cordites system), indicating that comments covered a wide range of subjects, often including multiple subtopics.

Among these main topics, the ones with both significant positive and negative comments were individually analyzed to determine if they were positive, negative, or neutral. This determination was based on the percentage of each type of comment.

1. Positive

2. Negative



SENTIMENT ANALYSIS

To effectively monitor the strategy developed from the topic modeling results, it is essential to automatically evaluate customer comments. This is achieved through sentiment analysis (SA) trained on labeled data, after under sampling for class unbalance (4076 per class).

RESULTS

MODEL	Accuracy
SEEDOT	0.32
BERT	0.37
FT BERT	0.61

MODELS

- **SEEDOT**: A tool that adapts VADER (a well-known knowledge-based SA) to your specific topic of interest, ensure explainability.
- **Pre trained BERT**: A pre-trained, learning-based sentiment analysis model used for benchmarking purposes.
- **Fine-tuned BERT**: A BERT model fine-tuned for 10 epochs to optimise sentiment analysis results.

PRODUCT FOCUS MARKETING STRATEGY

Since the topics represent a macro category given the comments grouped by topic, we trace back to the client who commented and look at their purchases. This helps in understanding the specific product, given also the product category name.

- Positive judged products will be used for ads, placed in the "best seller" category on the website, and also gifted with other products to enhance sentiment on negative products.
- Negative judged products undergo a manual inspection. It is suggested to change suppliers, improve the product, or sell alternative products instead.

The SA model is used to automatically categorise comments. Sentiment KPI on a single product is monitored every six months (since sentiment is slow to reverse) to take corrective measures if necessary.

TOPICS

POSITIVE	NEGATIVE
cat food	coffee kcups
baby fruit	green tea
potato chip	chicken
dog food	coconut water
popcorn	jerky beef
cookies	baby cereal
protein powder	agave nectar
penut butter	
coffee	

>

CUSTOMER FOCUS MARKETING STRATEGY

- Promoters: mailing campaign is made, where the email thanks them for the review and suggests a list of connected products that the client may like (discovered with MBA rules and client shopping history).
- Detractors: the marketing strategy focuses only on high-value customers (discovered with RFM) that are not churns (discovered with churn analysis) to avoid overlapping with churn marketing strategy. The mailing campaign involves a 20% discount on “best seller” products (previously identified by other customers' reviews) that cost less than 50 euros. This aims to convert customer opinions.

MAIL TO SEND

Type	Number
Promoters	308
Detractors	46

The SA model is used to automatically monitor if promoters continue to comment positively. The discount campaign is sent via email every month if they continue commenting positively. For detractors, it is checked if they have changed their opinion by leaving positive comments or not commenting negatively for 3 months. If so, the conversion campaign is stopped; otherwise, it is continued with an automatic email sent a week after the comment. The sentiment KPI on the two groups is monitored every six months to take corrective measures if necessary.

Thank you