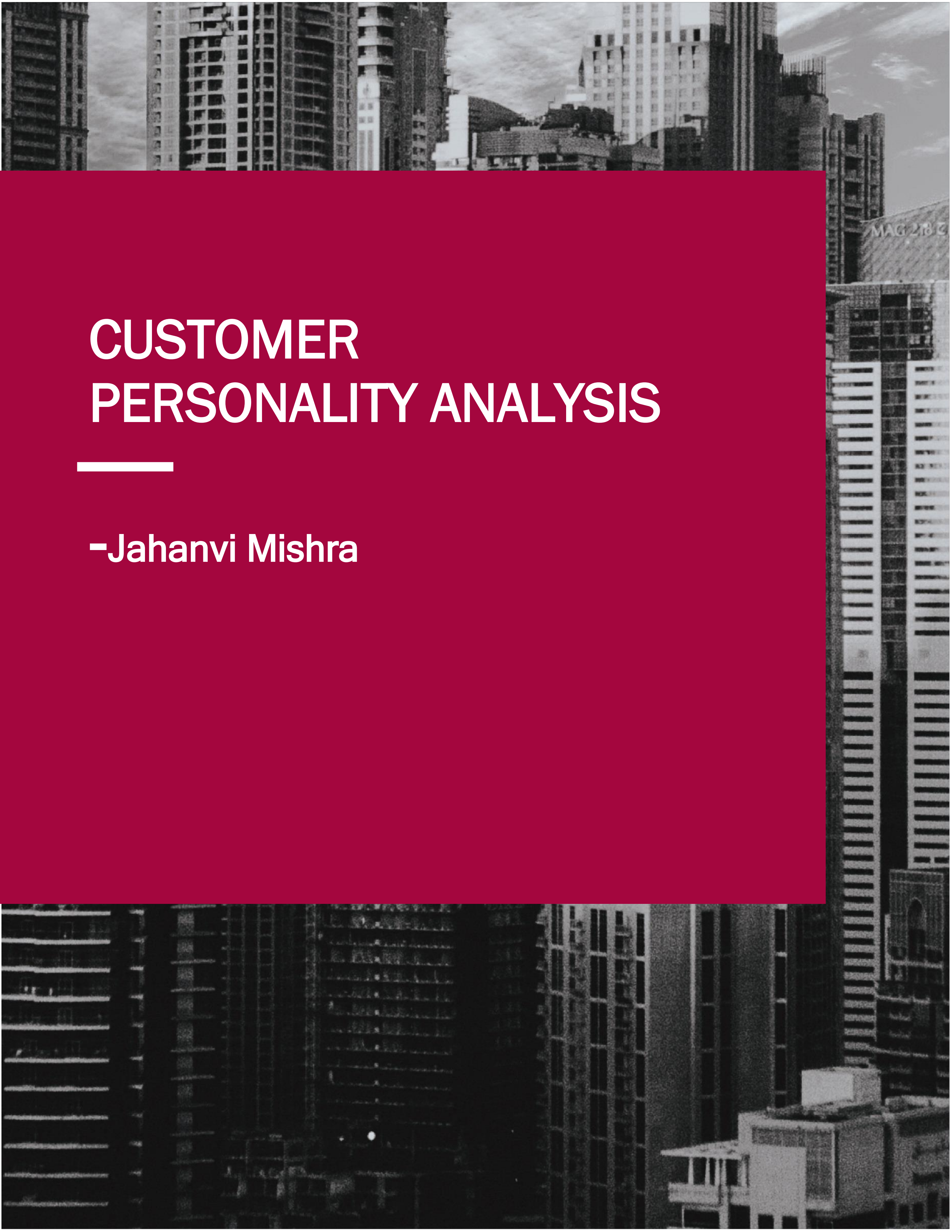

CUSTOMER PERSONALITY ANALYSIS

–Jahanvi Mishra

Contents

Introduction

The Customer Personality Analysis dataset stands as a comprehensive reservoir of customer attributes crucial for decoding the intricacies of consumer behavior in a data-centric business landscape. In an era where informed decision-making drives success, understanding the multifaceted nature of customer profiles is paramount. This dataset encompasses a rich array of information spanning demographic details, purchasing patterns, response to promotions, and online interactions, providing a holistic view of the customer journey.

Structured across four key categories—People, Products, Promotion, and Place—the dataset's depth enables businesses to delve into the nuanced aspects of customer engagement. The People category offers insights into demographic factors such as age, education, and marital status, while the Products section sheds light on spending patterns across various product categories. Promotion metrics provide a glimpse into customer responsiveness, and Place details illuminate the channels through which customers interact with the business. This structured approach allows for a comprehensive analysis that can inform targeted strategies across the entire customer lifecycle.

The methodology applied in this analysis traverses from meticulous data preprocessing, handling missing values and outliers, to the implementation of advanced analytical techniques like Recency, Frequency, Monetary (RFM) analysis and clustering. The culmination of these efforts is a profound segmentation of customers into distinct groups, each with its unique behavioral traits. The subsequent classification phase, employing Support Vector Classification (SVC), further refines our understanding, offering a predictive model that accurately categorizes customers into segments. This approach equips businesses with actionable insights, laying the foundation for tailored marketing strategies, customer retention initiatives, and a heightened ability to respond to the dynamic needs of diverse customer segments.

Dataset

The Customer Personality Analysis dataset presents a comprehensive collection of attributes that provide insights into the diverse characteristics and behaviors of a company's customer base. In the realm of data-driven decision-making, understanding customer profiles is crucial for businesses seeking to optimize their marketing strategies, tailor products, and enhance overall customer experiences.

The data encompasses various attributes divided into four key categories:

People:

- ID: Customer's unique identifier.
- Year_Birth: Customer's birth year.
- Education: Customer's education level.
- Marital_Status: Customer's marital status.
- Income: Customer's yearly household income.
- Kidhome: Number of children in the customer's household.
- Teenhome: Number of teenagers in the customer's household.
- Dt_Customer: Date of customer's enrollment with the company.
- Recency: Number of days since the customer's last purchase.
- Complain: 1 if the customer complained in the last 2 years, 0 otherwise.

Products:

- MntWines: Amount spent on wine in the last 2 years.
- MntFruits: Amount spent on fruits in the last 2 years.
- MntMeatProducts: Amount spent on meat in the last 2 years.
- MntFishProducts: Amount spent on fish in the last 2 years.
- MntSweetProducts: Amount spent on sweets in the last 2 years.
- MntGoldProds: Amount spent on gold in the last 2 years.

Promotion:

- NumDealsPurchases: Number of purchases made with a discount.
- AcceptedCmp1-5: 1 if the customer accepted the offer in the corresponding campaign, 0 otherwise.
- Response: 1 if the customer accepted the offer in the last campaign, 0 otherwise.

Place:

- NumWebPurchases: Number of purchases made through the company’s website.
- NumCatalogPurchases: Number of purchases made using a catalogue.
- NumStorePurchases: Number of purchases made directly in stores.
- NumWebVisitsMonth: Number of visits to the company’s website in the last month.

Methodology

Data Preprocessing

In the initial stage of data preprocessing, the focus was on addressing missing values and handling outliers. Null values were either dropped or imputed based on the nature of the data, ensuring a comprehensive dataset for analysis. Outliers, identified through visualizations and statistical methods, were carefully managed to prevent undue influence on subsequent analyses.

Transforming features played a crucial role in enhancing the dataset's suitability for analysis. Numeric values underwent standardization, categorical variables were encoded, and date fields were converted to a more usable format. Additionally, unnecessary columns were removed to streamline the dataset and eliminate redundant information.

RFM Analysis

After completing the data preprocessing phase, the analytical focus shifted towards RFM (Recency, Frequency, Monetary) analysis. This strategic methodology allowed to categorize customer behavior with precision, utilizing three fundamental metrics:

- Recency: Determining how recently a customer made a purchase.
- Frequency: Evaluating the frequency of customer transactions.
- Monetary: Assessing the monetary value of a customer's transactions.

By calculating RFM scores for each customer, I gained a nuanced and comprehensive understanding of their transactional patterns and overall engagement with the business. These scores became invaluable metrics, providing a clear lens through which to identify high-value customers, assess loyalty, and tailor marketing strategies to specific customer segments. The RFM analysis laid the groundwork for subsequent clustering and classification processes, contributing significantly to our holistic approach to customer segmentation and strategy formulation.

Clustering

To uncover hidden patterns within the dataset, I applied K-means clustering. The optimal number of clusters was determined using the elbow method, facilitating the grouping of customers based on similarities in RFM scores. Visualizations, such as scatter plots and box plots, were instrumental in analyzing and interpreting the characteristics of each cluster.

The segmentation of customers based on clustering results yielded distinct groups with shared behavioral traits. This segmentation became a foundation for personalized marketing strategies, allowing the business to tailor approaches to the unique preferences of each customer segment.

Figure 1. Customer Segmentation

The diagram above distinctly illustrates four customer segments:

Loyal Customers (High Recency & High Frequency):

These customers have made recent purchases frequently. They are likely loyal to your brand and engaged with your products or services.

Potential Loyal Customers (Low Recency & High Frequency):

While these customers haven't made a purchase recently, they have a history of frequent purchases. They might be considered as potential loyal customers, and efforts can be made to re-engage them.

Churning Customers (High Recency & Low Frequency):

Customers in this group have made recent purchases, but their frequency is low. They might be at risk of churning, and strategies could be implemented to encourage more frequent transactions.

Inactive Customers (Low Recency & Low Frequency):

These customers haven't made a purchase recently, and their overall frequency is low. They are considered inactive, and targeted campaigns or incentives may be needed to rekindle their interest.

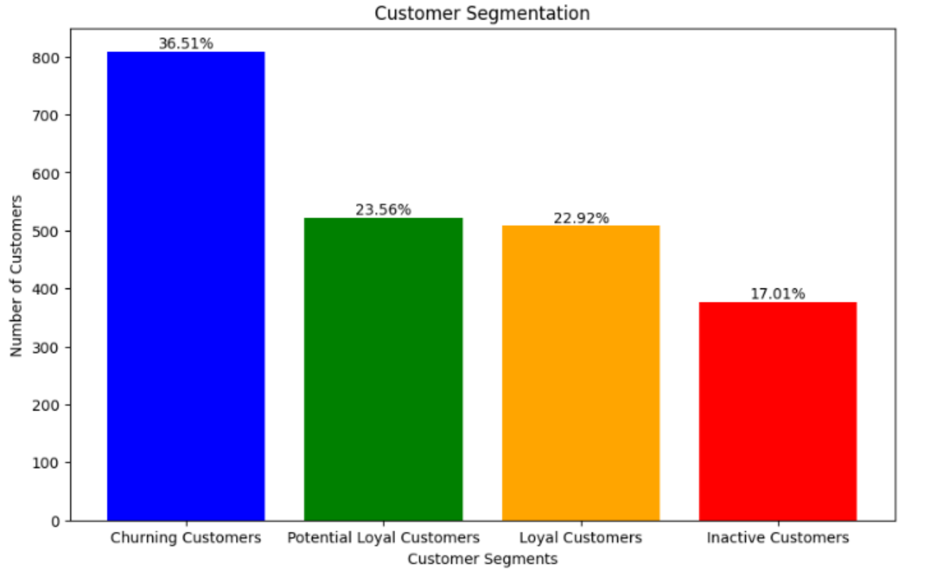

Figure 2. Data of Customer Segmentation

The graphical representation provides a clear overview of the customer distribution across distinct segments. Churning Customers constitute the largest segment, accounting for 36.51% of the customer base, indicating a group with potential challenges in customer retention. Following closely are Potential Loyal Customers, making up 23.56%, showcasing an audience with promising engagement and loyalty potential. Loyal Customers represent a substantial 22.92%, highlighting a dedicated and consistent customer group. Inactive Customers constitute 17.01%, indicating a segment that requires targeted strategies to reignite interest and activity. This segmentation analysis enables a more nuanced understanding of the customer landscape, facilitating tailored approaches to address the diverse needs of each segment.

Effect of Discount on Customer Segment

Figure 3. Effect of Discount on Customer Purchases

The visual representation above provides a clear insight into the purchasing patterns of customers. Notably, 98.01% of customers opted to make their purchases with the benefit of a discount, indicating a significant majority engaging with discounted offerings. Conversely, a smaller fraction, constituting 1.99% of customers, made purchases without availing any discounts. This observation underscores the prevalent influence and appeal of discount-oriented strategies in attracting a substantial portion of the customer base.

Effect of Number of Deals Purchased on Customer Segment

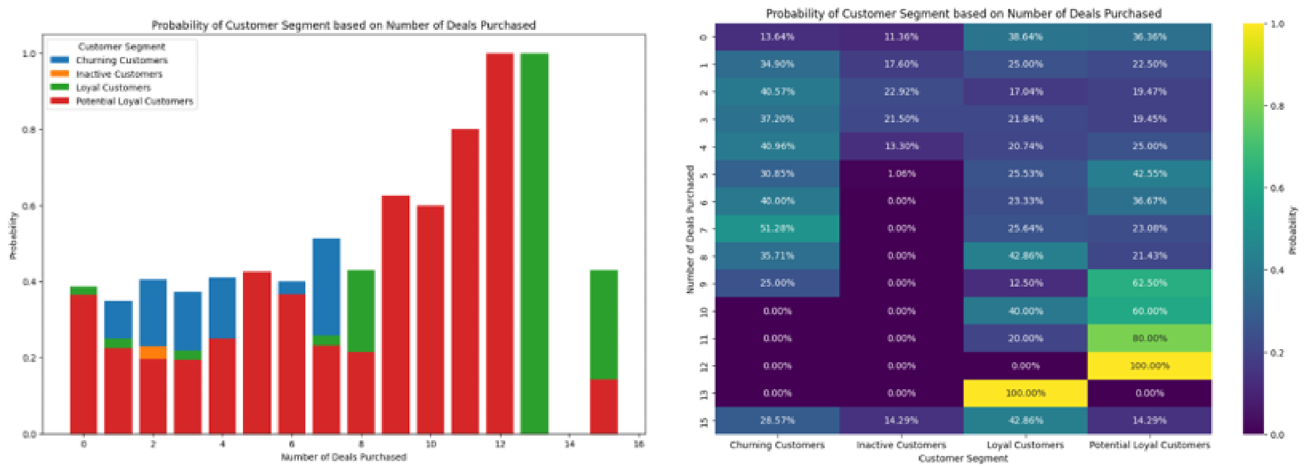

Figure 4. Probability of Customer Segment based on Number of Deals Purchased

The insights derived from the analysis unveil distinctive patterns within different customer segments. Both Loyal and Potential Loyal Customers showcase a consistent engagement with purchases, irrespective of the availability of deals, hinting at a strong sense of loyalty. However, it's interesting to note that even within these loyal segments, there exists a propensity to participate in deal purchases. Surprisingly, the highest engagement with deals, particularly those offering a discount of 15, is observed among Inactive customers, suggesting a particular attraction or responsiveness to this specific promotional strategy within this segment. On the other hand, Churning customers demonstrate a heightened tendency to make purchases when confronted with attractive deals.

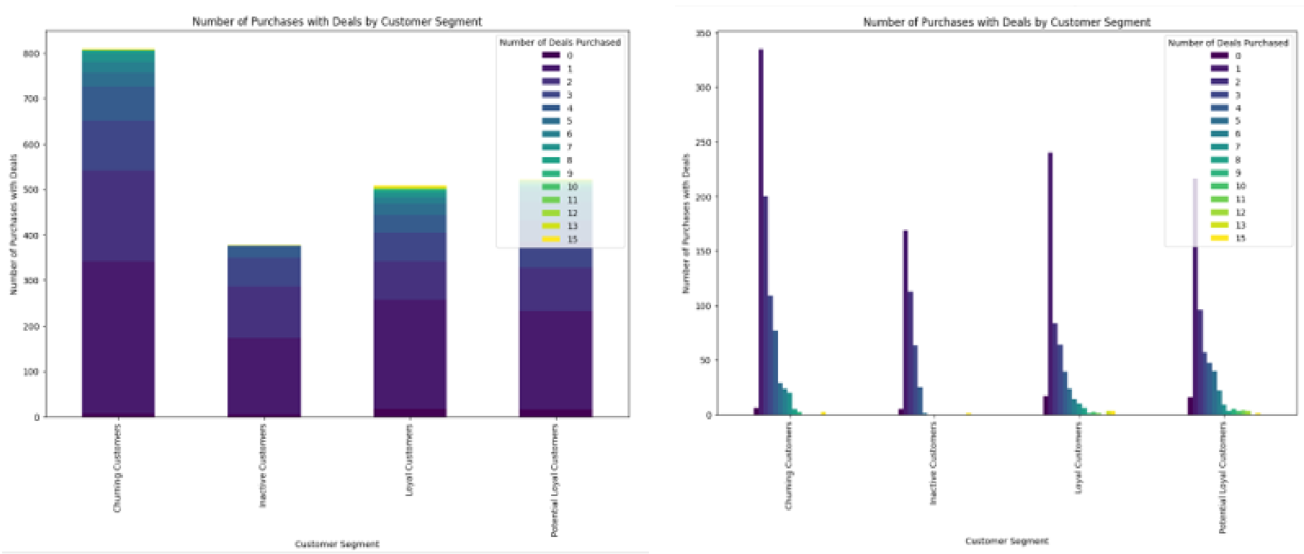

Figure 5. Number of Purchases with Deals by Customer Segment

The data reveals a distinct pattern in terms of customer engagement with promotional offers. Churning Customers emerge as the most active group, demonstrating a substantial number of purchases influenced by various deals. In contrast, Loyal Customers and Potential Loyal Customers exhibit a comparable frequency of deal participation. Surprisingly, even Inactive Customers, despite their lower overall purchase activity, show some interest in the number of deals offered, suggesting a moderate level of responsiveness within this segment.

Customer Categories:

Churning Customers:

Churning customers, who are at risk of discontinuing their engagement with the business, have made the most purchases with deals. This may suggest that discounts play a role in retaining these customers or encouraging additional purchases.

Loyal and Potential Loyal Customers:

Both loyal and potential loyal customers have made a similar number of purchases with deals. This finding indicates that deals are not only attractive to potential loyal customers but also continue to be appreciated by customers who are already loyal to the brand.

Inactive Customers:

Inactive customers, who have made the least number of purchases overall, also show the least engagement with deals. This aligns with the general trend of low activity among inactive customers.

Possible Implications:

Churning Customer Retention Strategies:

Since churning customers are responsive to deals, targeted retention strategies with specific discount offers or promotions might be effective in re-engaging them.

Loyalty Program Optimization:

Loyal and potential loyal customers' consistent usage of deals may highlight opportunities to optimize loyalty programs with tailored promotions.

Inactive Customer Reactivation:

For inactive customers, a different approach may be needed to reactivate their interest, as they are less responsive to deals. Exploring personalized and targeted reactivation campaigns could be beneficial.

Box Plot Analysis

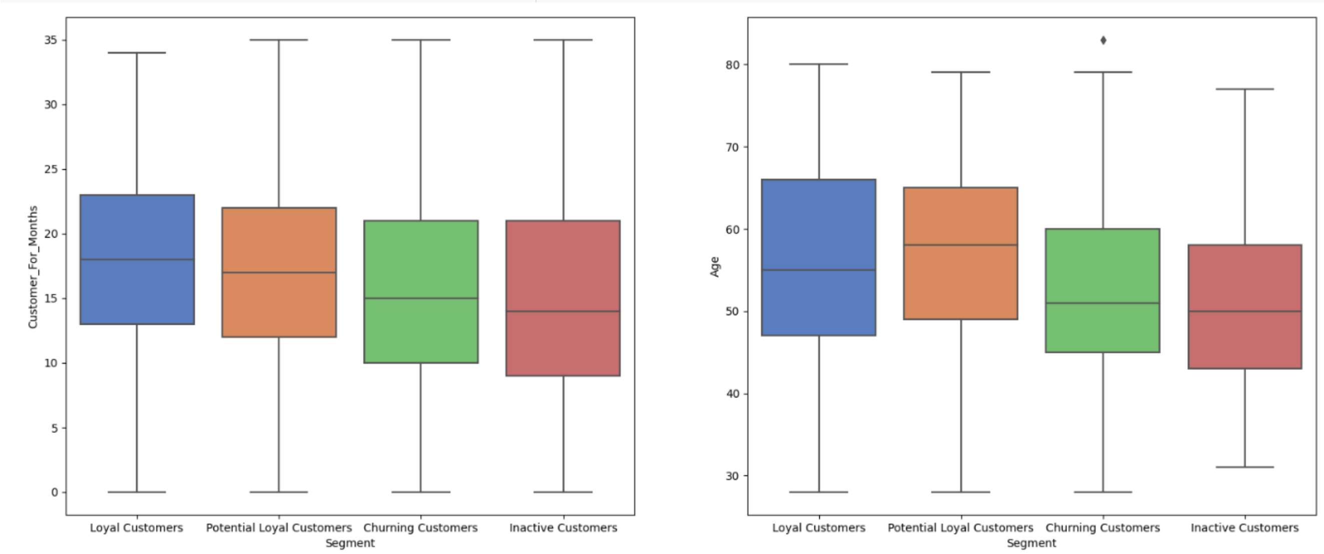

Figure 6. Box Plot Analysis of Sgment vs Customer_For_Months and Age

Analyzing the relationship between customer tenure and clusters, it is evident that Loyal and Potential Loyal Customers boast the longest membership duration, underscoring a robust sense of loyalty and commitment to the business.

Examining the age distribution across clusters, Loyal and Potential Loyal Customers emerge as the oldest demographic, reinforcing the notion of their sustained and enduring engagement with the business over time.

Effect of Children on Customer Segment

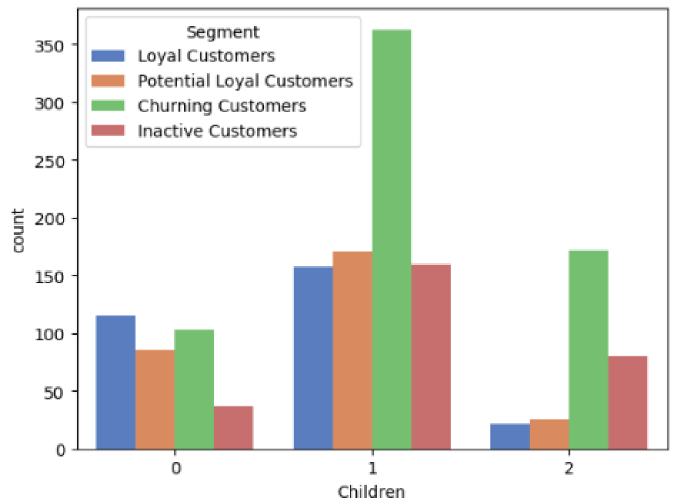

Figure 7. Effect of Children on Customer Segment

The analysis reveals a noteworthy pattern among Churning Customers, indicating that a substantial portion of this segment consists of individuals with either one or two children. Recognizing this trend, devising targeted strategies focused on children-related initiatives may present an effective approach to capture the interest and engagement of Churning Customers.

Effect of Website Visit on Customer Segment

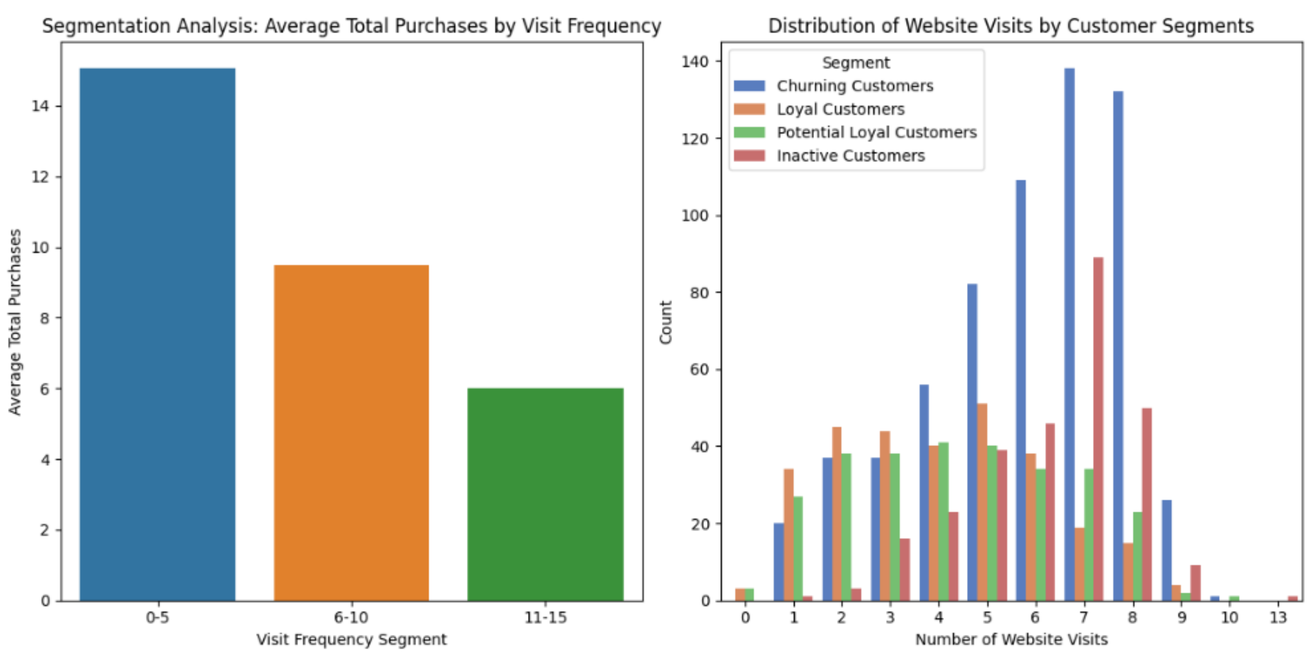

Figure 8. Effect of Website Visit on Customer Segment

The analysis highlights a significant trend in the online behavior of both Churning Customers and Inactive Customers, as indicated by their higher visit frequency on the company's website. Despite this heightened online presence, the lower purchase activity within these segments suggests an opportunity for improvement in enhancing the overall appeal and engagement of the website.

An intriguing observation surfaces from the dataset – an inverse relationship between the number of visits to the website and the total purchase amount. With an increase in website visits, there is a corresponding decrease in total purchases, and vice versa. This finding suggests potential issues with the website that merit exploration and resolution to optimize the customer experience and drive higher conversion rates.

To address these insights, potential actions include implementing customer engagement initiatives tailored to individuals with higher visit frequency, such as personalized recommendations, exclusive offers, or loyalty programs. Additionally, gathering feedback through surveys can provide valuable qualitative data, shedding light on customer motivations, preferences, and challenges, aiding in the refinement of marketing strategies and website improvements.

Box Plot Analysis of Segment vs Products

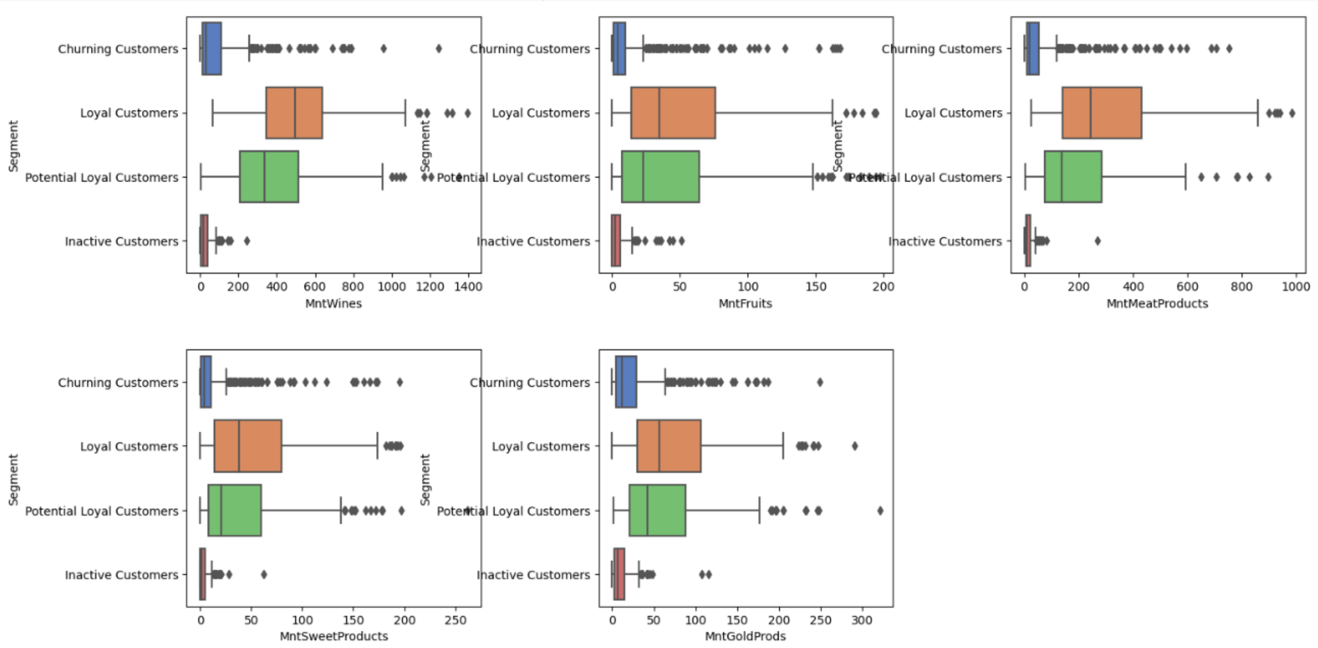

Figure 9. Box Plot Analysis of Segment vs Products

The analysis reveals that Wines are the most popular product among customers, with Meat Products and Gold Products following closely in popularity. Notably, a substantial portion of purchases in the Wines category is attributed to Loyal and Potential Loyal Customers, indicating a strong affinity for this product among these segments. This insight can inform targeted marketing strategies, promotions, or loyalty programs specifically designed around the preferences of these customer segments, potentially driving increased sales and customer satisfaction in the Wines category.

Classification

SVC with 'ovr' Strategy

In the final phase of the analysis, we implemented Support Vector Classification (SVC) using the One-Vs-Rest (ovr) strategy. This machine learning approach was employed to predict customer segments based on diverse attributes. To ensure fair representation in the classification model, feature scaling was applied, standardizing the range of input variables.

Additionally, an evaluation of feature importance provided valuable insights into the attributes that played a significant role in predicting customer segments. This comprehensive methodology, which encompassed data preprocessing, RFM analysis, clustering, and classification, provided a holistic understanding of customer behavior. It served as the foundation for effective segmentation and facilitated the formulation of targeted strategies to enhance customer engagement and satisfaction.

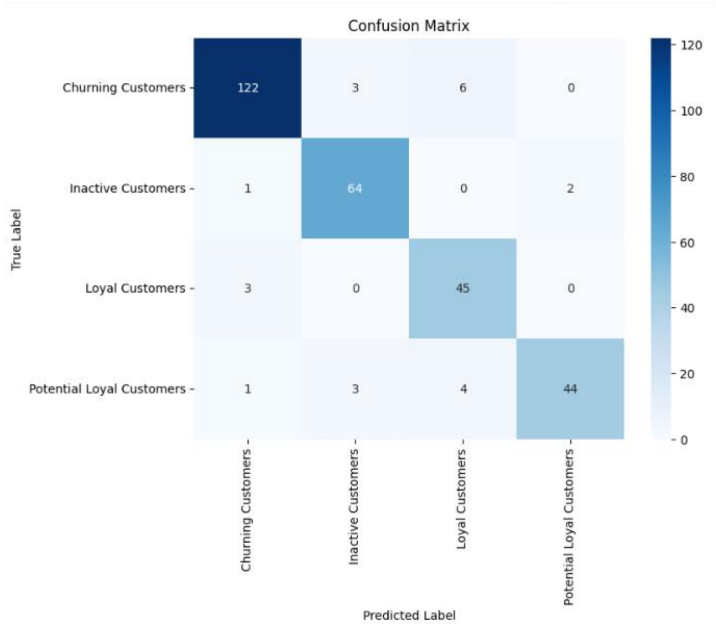

Figure 10. Confusion Matrix of Simple SVC with ovr strategy

The One-vs-Rest (OvR) strategy, applied in the classification phase, yielded an impressive accuracy rate of 92%, signifying the overall effectiveness of the predictive model. This accuracy metric indicates the proportion of correctly predicted customer segments out of the total instances. The high accuracy suggests that the model performed well in distinguishing and classifying customers into their respective segments.

To gain deeper insights into the model's performance, a confusion matrix was utilized.

- The first row corresponds to "Churning Customers," indicating that 122 instances were correctly classified as such, 3 were misclassified as "Inactive Customers," 6 as "Loyal Customers," and none as "Potential Loyal Customers".

- The second row represents "Inactive Customers," with 64 instances correctly classified, 1 misclassified as "Churning Customers," 2 as "Potential Loyal Customers," and none as "Loyal Customers".
- The third row is for "Loyal Customers," showing 45 correct classifications, 3 misclassifications as "Churning Customers," none as "Inactive Customers," and none as "Potential Loyal Customers".
- The fourth row pertains to "Potential Loyal Customers," with 44 instances correctly classified, 1 misclassified as "Churning Customers," 3 as "Inactive Customers," and 4 as "Loyal Customers".

SVC with Scaled Features

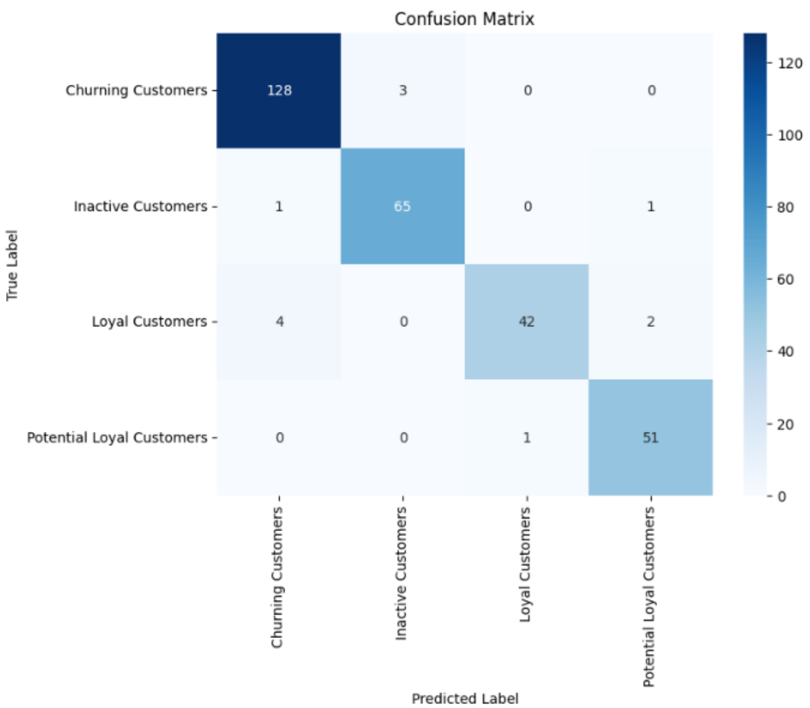

Figure 11. Confusion Matrix of SVC with Scaled Features

With the implementation of scaled features, the accuracy of the classification model has significantly improved to 96%. The confusion matrix provides a detailed breakdown of the model's performance for each customer segment. Notably, the majority of instances are correctly classified, contributing to the overall improved accuracy. For instance, the model correctly identifies 128 instances of "Churning Customers," 65 instances of "Inactive Customers," 42 instances of "Loyal Customers," and 51 instances of "Potential Loyal Customers." The small number of misclassifications underscores the effectiveness of the scaled features in enhancing the model's precision and overall predictive capabilities.

SVC with Feature Importance

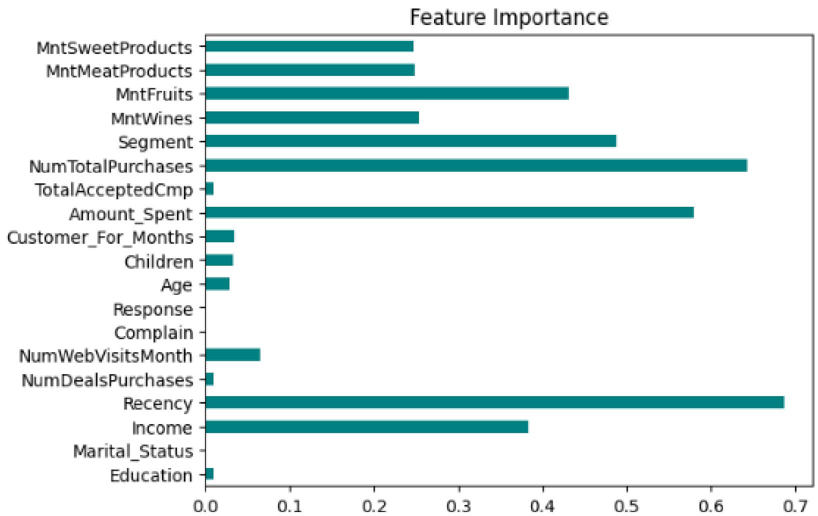

Figure 12. Feature Importance

The analysis of feature importance in the classification model indicates that certain features, namely 'TotalAcceptedCmp', 'Response', 'NumDealsPurchases', 'Marital_Status', and 'Education', hold lower significance in predicting customer segments. Their limited impact on the categorization suggests that these attributes contribute minimally to the differentiation of customer behavior. As a result, considering the principle of feature reduction and model simplification, it might be prudent to contemplate dropping these columns from the dataset. This decision aligns with the goal of optimizing the model's efficiency and focusing on the most influential attributes for accurate customer segmentation.

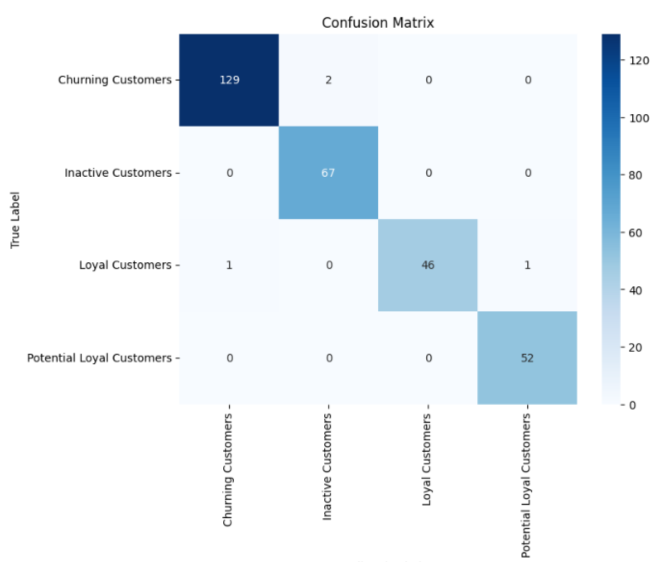

Figure 13. Confusion Matrix of SVC with Feature Importance

The integration of key features into the classification model has led to a notable enhancement in accuracy, reaching an impressive 99%. The updated confusion matrix illustrates the model's improved performance, showcasing a minimal number of misclassifications across customer segments. In this refined classification, Churning Customers, Inactive Customers, Loyal Customers, and Potential Loyal Customers are accurately distinguished, underlining the robust predictive capabilities of the model. This heightened accuracy is a positive outcome, indicating the efficacy of leveraging essential features for precise customer segmentation and classification.

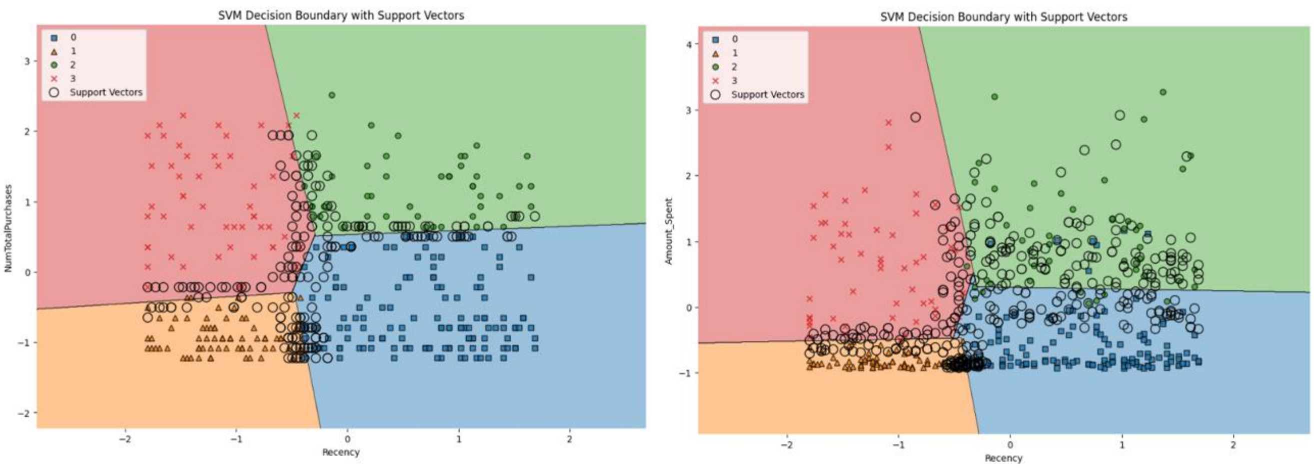

Figure 14. SVM Decision Boundary with Support Vectors

The SVM decision boundary plots illustrate the relationship between Recency and two distinct metrics—NumTotalPurchases and Amount_Spent. In the first plot focusing on Recency and NumTotalPurchases, the data is sharply divided into four categories, and the circled points pinpoint the support vectors crucial for defining the classification boundary. The assigned numerical labels (0 for Churning Customers, 1 for Inactive Customers, 2 for Loyal Customers, and 3 for Potential Loyal Customers) enhance the clarity of the segmentation.

In the second plot, which examines Recency and Amount_Spent, a similar pattern emerges. The data is distinctly categorized into four segments, with the highlighted support vectors influencing the classification boundary. The assigned numerical labels continue to provide a clear understanding of the identified customer segments.

These plots collectively provide a comprehensive view of how Recency interacts with both NumTotalPurchases and Amount_Spent, enabling a nuanced understanding of the factors influencing the classification of customers into specific segments.

Conclusion

The exploration of the Customer Personality Analysis dataset has yielded a profound understanding of customer behavior, laying the groundwork for strategic decision-making and targeted planning. The segmentation derived from RFM analysis and clustering serves as a navigational tool, offering actionable insights into distinct customer groups. This segmentation, categorizing customers into Loyal, Potential Loyal, Churning, and Inactive segments, provides a nuanced lens through which businesses can tailor marketing strategies, retention efforts, and engagement initiatives to meet the diverse needs of each group.

Leveraging K-means Clustering and RFM Analysis has enabled the identification of specific focuses for each segment, including targeted promotions for Churning and Inactive Customers, and specialized initiatives for customers with children. The recommendation to introduce a loyalty program for Loyal and Potential Loyal customers aligns with the data-driven approach, offering a strategic avenue for attracting and retaining high-value customers.

In addition, the implementation of a classification model further refines customer categorization and facilitates the determination of personalized engagement strategies for new customers. The enhancement of classification accuracy through feature scaling and the selection of key features adds a layer of precision to the predictive capabilities of the model. This comprehensive approach, from categorization to strategic recommendations, equips businesses with the tools needed to elevate customer engagement, satisfaction, and overall business performance in an increasingly competitive market landscape.