

Capstone 3: Final Project Report

Problem Statement:

How many distinct customer groups are needed to segment the customers into to allow for targeted marketing campaigns based on purchasing patterns and behavior?

Context:

Customer segmentation helps businesses tailor marketing strategies to different groups of customers based on their behavior. I want to focus on RFM model-based clustering techniques to group customers to better target specific customer segments depending on the product or products we are trying to sell.

Data:

Online Retail Data from UC Irvine Machine Learning Repository - [Retail Transactions](#)

Method:

The approach for this project involved multiple steps, beginning with exploratory data analysis (EDA), followed by segmentation through machine learning clustering, and concluding with predictive modeling to classify new customers.

Project Approach:

The approach for this project involved multiple steps, beginning with exploratory data analysis (EDA), followed by segmentation through machine learning clustering, and concluding with predictive modeling to classify new customers.

- **Data Collection and Cleaning:** The dataset used contained customer transactions, including 'InvoiceNo', 'StockCode', 'Quantity', 'UnitPrice', 'CustomerID', 'Country', and 'InvoiceDate'. I cleaned the data by removing canceled transactions, handling missing values, and correcting data inconsistencies.
- **Feature Engineering:** The RFM model was utilized to derive three main features:
 - **Recency:** The number of days since the customer's last purchase.
 - **Frequency:** The number of transactions a customer made.
 - **Monetary:** The total amount spent by the customer.
- **RFM Clustering:** I applied k-means clustering to segment customers into four clusters based on their RFM scores. The optimal number of clusters was determined using the elbow method.
- **Predictive Modeling:** A Random Forest classifier was trained to predict the cluster assignment for new customers based on their RFM metrics. Further optimization was conducted using Bayesian optimization and ensemble methods, and XGBoost was also evaluated for improved performance.

Exploratory Data Analysis (EDA):

EDA was conducted to understand the distribution and relationships within the dataset. Key findings included:

- **Sales Distribution:** The majority of sales originated from the United Kingdom, with a significant drop-off in sales volume from other countries. Based on this finding I decided to only move forward with analyzing the UK data.
- **Customer Behavior:** The data revealed a mix of high-frequency but low-spending customers, as well as low-frequency, high-spending customers. Identifying these patterns helped inform the segmentation process.
- **Cancellations Analysis:** A portion of transactions were identified as cancellations (8872), which impacted the monetary value calculation and were treated separately in the analysis.

Modeling:

- **RFM Clustering:** K-means clustering was used to segment customers into four clusters based on their RFM scores. The elbow method indicated that four clusters provided an optimal balance between model complexity and interpretability.
 - **Cluster Characteristics:**
 - **Cluster 2:** High-value, loyal customers with high frequency and monetary value.
 - **Cluster 3:** Mid-value customers who make frequent purchases but do not spend as much as Cluster 3.
 - **Clusters 0 and 1:** Low-value customers with infrequent purchases and low monetary value.
- **Predictive Modeling:** To predict cluster assignments for new customers, I trained a Random Forest classifier using the RFM features. The initial model achieved an accuracy of 97%. Subsequent optimization efforts included:
 - **Bayesian Optimization:** Used to fine-tune hyperparameters, which resulted in minor improvements in model performance.
 - **Ensemble Methods:** Combined multiple models, including Random Forest and XGBoost, to improve overall accuracy. The ensemble model achieved an accuracy of 98%, showing a slight improvement over the individual models.
 - **XGBoost:** Evaluated for its ability to handle complex relationships within the data, achieving similar accuracy to the Random Forest model but with increased computational cost.

Conclusion:

The segmentation of customers using RFM and machine learning provided valuable insights into customer behavior and spending patterns. With effective marketing strategies tailored to each customer segment, the client can enhance customer retention, maximize revenue, and allocate resources more efficiently. Moving forward, expanding upon these insights through further analysis of customer behavior and integrating lifetime value prediction can offer additional value.

Future Work:

- **Behavioral Analysis:** Conduct additional research into the purchasing behavior of each cluster, including product preferences and buying triggers. This analysis could support more refined targeting strategies and improve the effectiveness of cross-selling.
- **Seasonality Analysis:** Investigate purchasing patterns based on seasonality to identify peak times for different customer segments and align marketing campaigns with those periods.
- **Customer Lifetime Value (CLV) Prediction:** Develop a model to predict CLV for each customer segment, allowing for a better understanding of the long-term value of each cluster and helping prioritize marketing resources effectively.

Recommendations:

1. **Targeted Campaigns for High-Value Customers (Cluster 2):** Focus marketing efforts on retaining and rewarding high-value customers through loyalty programs, exclusive offers, and early access to sales. Maintaining these customers will provide consistent and substantial revenue.
2. **Personalized Promotions for Mid-Value Customers (Cluster 3):** Implement personalized promotions and cross-selling opportunities for Cluster 3 to encourage them to increase their spending. Bundling products or offering tiered incentives can help transition these customers into higher-value segments.
3. **Re-engagement Strategies for Low-Value Customers (Clusters 0 and 1):** Provide discounts or incentives for repeat purchases and target these customers with educational content on the value of products. This can help increase engagement and potentially convert them into more loyal customers.