Capstone Project – Online Retail

Table of Contents

Problem Statement

An online retail store aims to understand and segment customer purchasing patterns to optimize marketing strategies and enhance overall business performance.

Project Objective

Utilize RFM (Recency, Frequency, Monetary) analysis and K-means clustering to segment customers and provide insights for targeted marketing and inventory management.

Data Description

The dataset 'OnlineRetail.csv' consists of sales data, including information on invoices, products, quantities, prices, customer IDs, and purchase dates.

Data Pre-processing Steps and Inspiration

1.Handling Missing Values:

- Dropped rows with missing values.

2.Filtering Negative Quantities:

- Removed entries with negative quantity values.

3.Converting Date Format:

- Converted the 'InvoiceDate' column to a datetime format.

4.RFM Calculation:

- Grouped data by customer and calculated Recency, Frequency, and Monetary values.

5.Standardization:

- Standardized the features for clustering.

Choosing the Algorithm for the Project

Utilized K-means clustering for customer segmentation based on RFM values.

Motivation and Reasons For Choosing the Algorithm

K-means clustering is a popular and efficient algorithm for segmentation tasks. It allows for the identification of distinct customer groups based on similar purchasing behaviors, aiding in targeted marketing.

Assumptions

1. The chosen number of clusters (4) is suitable for meaningful segmentation.
2. Clusters represent clear and interpretable customer segments.
3. The data provides an accurate representation of customer behavior.

Model Evaluation and Techniques

1.Visualization:

- Utilized pairplots and boxplots to visualize the distribution of RFM variables within clusters.

2.Cluster Labels and Strategies:

- Assigned labels to clusters and defined marketing strategies for each segment.

Inferences from the Same

1. Identified distinct customer segments, including High-Value Customers, Mid-Value Customers, Low-Value Customers, and those at Churn Risk.
2. Developed targeted marketing strategies and product recommendations for each cluster.

Future Possibilities of the Project

1.Integration with Marketing Campaigns:

- Implement and assess the effectiveness of targeted marketing campaigns based on the identified clusters.

2.Continuous Monitoring and Iteration:

- Regularly monitor customer behavior and iteratively refine segmentation and strategies based on evolving patterns.

3.Integration with E-commerce Platform:

- Implement the segmentation model within the e-commerce platform to automate personalized recommendations and promotions.

Conclusion

The RFM analysis and clustering provide valuable insights for the online retail store, enabling personalized marketing strategies and enhanced customer engagement. Continuous monitoring and adaptation will ensure the model remains effective in addressing changing customer behaviors and business dynamics.

References

For k means clustering

MacQueen, J.B. "Some Methods for classification and Analysis of Multivariate Observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, 1967, pp. 281–297..