Project Report on Customer Segmentation

Executive Summary

This report details the process and results of a customer segmentation analysis conducted on the Online Retail II dataset. The goal was to segment customers based on their purchase behavior to tailor marketing strategies effectively. Using K-means clustering, we identified three distinct customer segments, which could help in optimizing marketing efforts and enhancing customer engagement.

Introduction

Problem Statement

In competitive e-commerce environments, understanding customer behavior is crucial for optimizing marketing strategies. Customer segmentation helps in identifying groups of customers with similar characteristics and purchasing behaviors, allowing for targeted marketing and improved customer service.

Importance of Customer Segmentation

Segmenting customers allows businesses to:

- Develop tailored marketing strategies.
- Enhance customer retention.
- Increase overall profitability by focusing on high-value customer groups.

Methodology

Data Collection

The dataset used is the Online Retail II data set, which includes transactions from a UK-based online retail store from December 2009 to December 2011.

Data Preprocessing

The data preprocessing steps included:

- Cleaning: Removed incomplete and duplicate records.
- Feature Engineering: Created new features such as total spend per customer, total orders, and days since last purchase.

Clustering Approach

K-means clustering was applied to segment the customer base. The Elbow method determined the optimal number of clusters, followed by the application of the silhouette score to evaluate the quality of clustering.

Results and Discussion

Exploratory Data Analysis

- **Total Spend**: Found that 20% of customers contribute to 80% of the total revenue.
- **Order Frequency**: Identified that frequent customers often have smaller order values.

Customer Segmentation

- **Cluster 0 (High-Value Customers)**: This segment includes customers who frequently purchase high-value items. They contribute the most to the revenue and have the shortest intervals between purchases.
- **Cluster 1 (Frequent, Low-Value Customers)**: Customers in this segment make purchases frequently but spend less per transaction.
- **Cluster 2 (Infrequent, Low-Value Customers)**: This segment includes customers who shop infrequently and spend the least.

Insights from Clustering

- Cluster 0 should be the primary focus for retention strategies and premium offers.
- Cluster 1 could be targeted with frequent low-cost marketing campaigns to increase their transaction size.
- Cluster 2 may need strategic engagement to boost their purchase frequency and value.

Conclusions and Recommendations

Conclusions

The segmentation revealed distinct behaviors and preferences within the customer base, which can be leveraged to enhance marketing effectiveness and customer satisfaction.

Recommendations

- **Cluster 0**: Implement loyalty programs and exclusive offers to enhance retention.
- **Cluster 1**: Offer bundle deals and promotions to increase basket size.
- **Cluster 2**: Deploy re-engagement campaigns and offer first-purchase incentives.

Appendices

- Appendix A: Code for Data Preprocessing and Clustering
- Appendix B: Additional Graphs and Cluster Visualizations

References

- "Online Retail II Dataset." UCI Machine Learning Repository.
- Scikit-learn Documentation
- Python Data Science Handbook