Introduction

Title: Strategic Customer Segmentation and Predictive Analysis in Marketing

Objective: To apply advanced data analytics techniques for customer segmentation, with the goal of enhancing marketing strategies and customer satisfaction for a UK-based e-commerce company.

Table of Contents

1. Introduction and Background

1.1 Introduction

In the dynamic realm of e-commerce, enterprises must employ data analytics to predict and comprehend consumer behavior, enhancing their marketing tactics and augmenting client pleasure. This study investigates the functioning of a renowned e-commerce corporation located in the United Kingdom. The company is known for providing a wide range of gifts for various occasions and has a substantial customer base.

1.2 Background to the Problem

E-commerce has replaced the direct interpersonal engagement that brick-and-mortar retailers once had with clients. Businesses currently depend on customers' online activities, such as their purchase history and website navigation, to collect vital data. A comprehensive examination of the sales data, along with the demographic and psychographic profiles, is essential for the shop in question to fully comprehend its consumer segments. This analysis covers the timeframe from December 1, 2020, to November 24, 2021.

1.3 Objective of Analysis

The primary objective is to utilize sophisticated analytical techniques to precisely categorize the varied customer base into clearly defined segments. This analysis will offer vital insights into the unique requirements and behavioral tendencies of different segments, differentiating between individual consumers and bulk buyers. The goal is to enhance consumer engagement and loyalty by delivering communication that establishes a deep connection with the specific preferences of each segment.

2. Literature Review

The literature review covers various studies on customer segmentation using different clustering and data mining techniques. Key studies include:

- Visualization method for customer targeting using customer map (Ji Young Woo et al., 2005)
- Buyer Targeting Optimization: A Unified Customer Segmentation Perspective (Jingyuan Yang et al., 2016)
- Identifying patients in target customer segments using a two-stage clustering-classification approach (You-Shyang Chen et al., 2012)
- Customer Segmentation Using Clustering and Data Mining Techniques (Kishan R. Kashwan et al., 2013)
- Segmenting and Targeting Customers Through Clusters Selection & Analysis (Ilung Pranata et al., 2015)
- Application of Clustering Algorithm for Effective Customer Segmentation in E-Commerce (Ritu Punhani et al., 2021)
- Customer Segmentation using K-means Clustering (Tushar Kansal et al., 2018)

3. Methodology

3.1 Data Exploration

The dataset comprises 10,000 rows and 14 columns, including both nominal and numerical data. Descriptive statistics and visualizations were performed on key variables like Quantity, Unit Price, Return Rate, Age, Income, Marital Status, Work, and Education.

3.2 Data Pre-Processing

Data quality issues identified included invalid return rates, missing values in descriptions, and NA values in customer IDs. These were addressed using various techniques like filtering out invalid values, omitting rows with missing customer IDs, and factoring non-factored values.

3.3 Hierarchical Clustering

Hierarchical clustering was performed to construct a hierarchy of data points. An elbow plot was used to estimate the ideal number of clusters for segmenting the data.

3.4 K-means Clustering

K-means clustering was applied to the normalized dataset, resulting in three clusters. The analysis used 1000 random starting points and a maximum iteration limit of 500 to optimize the cluster centroids.

3.5 Linear Discriminant Analysis (LDA)

LDA was used to reduce the dimensionality of the sample set by projecting it onto a lower-dimensional sample space, identifying significant predictors across the identified clusters.

3.6 Analysis of Variance (ANOVA)

ANOVA was utilized to assess the relevance of the discriminant functions obtained from LDA in explaining the variance of the clusters.

3.7 RFM Analysis

RFM analysis was conducted to assess customers' purchasing patterns based on recency, frequency, and monetary factors. The scores of these three variables were combined into a single metric called RFM.

4. Analysis of Segments Using Tableau

The customer segmentation analysis was visualized using Tableau. The dashboard provided a concise overview of client segmentation, presenting crucial facts across three distinct groupings.

5. Conclusion

The analysis identified three well-defined customer segments. Cluster 1, with the smallest customer count, presented the highest average revenue per customer, identifying it as a segment with high-value clients ideal for premium marketing initiatives. Cluster 3's substantial customer volume but lower average revenue per customer suggests a focus on mass-market strategies.

6. References

A comprehensive list of references used throughout the study is provided.