April, 2025

CUSTOMER SEGMENTATION WITH CLUSTERING

Onyeka Muoka

TABLE OF CONTENTS

INTRODUCTION

In the competitive world of e-commerce, analyzing and predicting customer behavior is crucial for businesses looking to improve marketing efficiency and strengthen customer loyalty. Customer segmentation the process of grouping customers based on shared traits enables companies to customize their marketing efforts for each segment, enhancing satisfaction and maximizing resource use for higher profits. Studies show that leveraging in-depth customer data can greatly increase brand recognition and lower cart abandonment rates.

This report aims to build a reliable customer segmentation model for an e-commerce business using clustering methods. The goal is to extract practical insights that guide targeted marketing campaigns, boosting customer engagement and sales growth. Key factors such as purchase frequency, recency, customer lifetime value (CLV), average unit cost (AUC), and customer age are analyzed to define distinct customer profiles. However, certain limitations must be noted: the timeframe assumes observed trends reflect current behavior, though market conditions may have shifted; the analysis prioritizes specific key features, acknowledging that other unexamined factors could influence customer behavior; and the clustering approach relies on assumptions that may not fully represent the intricacies of customer interactions. By addressing these factors, the report seeks to deliver valuable segmentation insights while recognizing the constraints of the methodology.

PROBLEM STATEMENT

The business struggles to identify distinct customer segments due to insufficient insights into its diverse customer base, resulting in generic and ineffective marketing strategies. This lack of personalization reduces marketing efficiency, drives higher customer churn, and limits revenue growth. To address this, the goal is to leverage clustering techniques to segment customers into meaningful groups, enabling data-driven marketing decisions and enhanced customer engagement.

METHODOLOGY

A data-driven approach was employed to segment customers into clusters, utilizing statistical and machine learning techniques. The dataset contained missing values and duplicated rows, which were removed during the preprocessing phase. Exploratory data analysis (EDA) was conducted to examine the distribution and characteristics of the features.

Outliers in the dataset were identified and excluded to ensure data quality. Several methods were applied to determine the optimal number of clusters, including the Elbow method, Silhouette Score method, and Hierarchical Clustering. Based on these methods, six clusters were identified as the optimal solution.

K-Means clustering was then used to segment the customers into six distinct clusters. To visualize the clustering results, dimensionality reduction techniques, PCA and t-SNE, were applied. A 2D plot was generated to display the clusters, each represented by a different color.

DATA STATISTICS

A data-driven approach was employed to segment customers into clusters, utilizing statistical and machine learning techniques. The dataset contained missing values and duplicated rows, which were removed during the preprocessing phase. Exploratory data analysis (EDA) was conducted to examine the distribution and characteristics of the features.

Outliers in the dataset were identified and excluded to ensure data quality. Several methods were applied to determine the optimal number of clusters, including the Elbow method, Silhouette Score method, and Hierarchical Clustering. Based on these methods, six clusters were identified as the optimal solution.

K-Means clustering was then used to segment the customers into six distinct clusters. To visualize the clustering results, dimensionality reduction techniques, PCA and t-SNE, were applied. A 2D plot was generated to display the clusters, each represented by a different color.