INTELLIHACK 5.0

Customer Segmentation Report

Question 02

Date: March 9, 2025

Group : Evora

Abstract

This report analyzes customer behavior on an e-commerce platform using K-Means clustering. It identifies three customer segments: Bargain Hunters, High Spenders, and Window Shoppers. Data preprocessing, feature scaling, and model evaluation are described in detail. Clusters are visualized using Principal Component Analysis (PCA).

Contents

Chapter 1

Introduction

Customer segmentation helps businesses personalize marketing strategies and optimize customer engagement. This study uses K-Means clustering to analyze customer behavior based on total purchases, average cart value, time spent, product clicks, and discount usage.

1.1 Problem Statement

Understanding customer behavior is crucial for targeted marketing. The objective is to segment customers into distinct groups based on their purchasing habits and engagement levels.

Chapter 2

Exploratory Data Analysis (EDA)

The dataset consists of the following features:

- **Total Purchases:** Number of purchases made by a customer.
- **Average Cart Value:** Mean price of items in the cart.
- **Total Time Spent:** Total minutes spent on the platform.
- **Product Clicks:** Number of product views.
- **Discount Usage:** Number of times a discount code was used.
- **Customer ID:** Unique identifier (not used in clustering).

2.1 Data Cleaning and Preprocessing

- **Handling Missing Data:** Rows with NaN values were removed.
- **Feature Scaling:** StandardScaler was applied to normalize numerical values.
- **Customer ID:** Dropped as it is non-numeric.

Chapter 3

Model Selection and Clustering

3.1 Choosing K for K-Means

3.1.1 Elbow Method

The Elbow method was used to determine the optimal number of clusters by plotting inertia values for k=1 to k=10.

3.2 Final Model

Based on the results of the below graph, K=3 was selected as the optimal number of clusters.

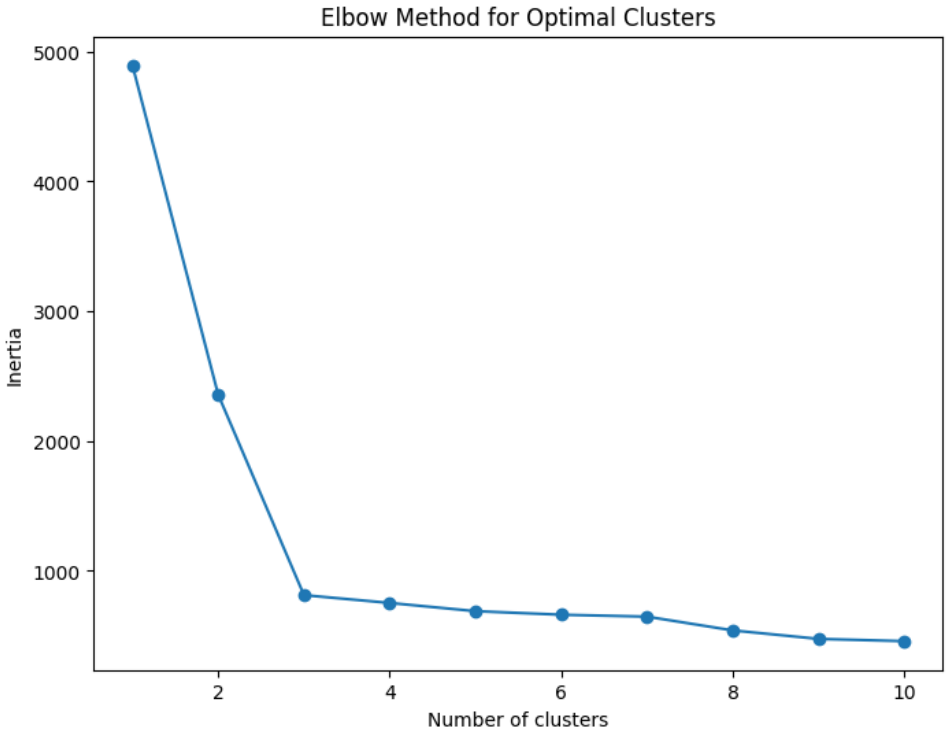

Figure 3.1: Elbow method graph

Chapter 4

Results and Visualization

4.1 Cluster Visualization using PCA

To visualize high-dimensional data, PCA was used to reduce features to 2D space.

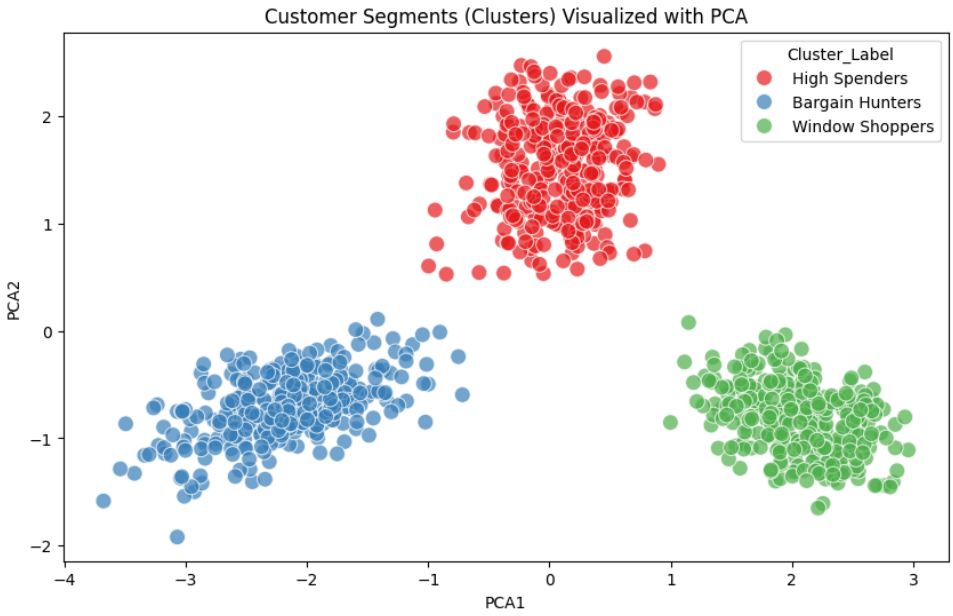

Figure 4.1: Customer Clusters Visualized using PCA

4.2 Cluster Interpretation

- **Bargain Hunters:** Frequent buyers of low-value items, heavy discount users.
- **High Spenders:** Moderate purchase frequency, high-value cart, minimal discount reliance.
- **Window Shoppers:** Low purchases, high browsing time, infrequent discount use.

Chapter 5

Conclusion

This study successfully segmented customers into three groups using K-Means clustering and PCA visualization.

- Using hierarchical clustering for better cluster definition.
- Incorporating additional features like customer demographics.
- Exploring deep learning approaches for enhanced segmentation.