Customer Segmentation Report

Team codeG

Task2

Executive Summary

This project aimed to identify distinct customer segments in an e-commerce platform based on their behavior. Using a dataset of customer interactions, purchases, and browsing patterns, we applied K-Means clustering to group customers into three segments: Bargain Hunters, High Spenders, and Window Shoppers. The analysis revealed clear behavioral patterns for each segment, enabling targeted marketing strategies. Key findings include the identification of high-value customers (High Spenders), frequent discount users (Bargain Hunters), and highly engaged but low-purchasing customers (Window Shoppers). These insights can help optimize marketing efforts, improve customer experience, and drive revenue growth.

1. Introduction

1.1 Background

Customer segmentation is a critical tool in e-commerce for understanding and predicting customer behavior. By grouping customers based on shared characteristics, businesses can tailor marketing strategies, improve customer retention, and maximize revenue. This project focuses on identifying behavioral segments to inform personalized marketing and enhance the customer experience.

1.2 Project Objectives

- **Identify distinct customer segments** based on purchasing and browsing behavior.
- **Characterize each segment** to understand their unique traits and preferences.
- **Validate the existence of three hypothesized segments**: Bargain Hunters, High Spenders, and Window Shoppers.

1.3 Dataset Description

The dataset contains **6 features** for each customer:

- **customer_id**: Unique identifier for each customer.
- **total_purchases**: Total number of purchases made.
- **avg_cart_value**: Average value of items in the cart.
- **total_time_spent**: Total time spent on the platform (in minutes).
- **product_click**: Number of products viewed.
- **discount_count**: Number of times a discount code was used.

The dataset was expected to contain **three hidden clusters**, each representing a distinct customer segment.

2. Methodology

2.1 Data Preparation and Exploration

- **Data Loading**: The dataset was loaded and inspected for missing values and inconsistencies.
- **Exploratory Data Analysis (EDA)**: We analyzed the distribution of each feature and checked for correlations between variables. Key insights included:
    - Most customers made a moderate number of purchases.
    - A small group of customers spent significantly more time on the platform.
    - Discount usage varied widely across customers.

2.2 Data Preprocessing

- **Feature Selection**: The customer_id column was removed as it was not relevant for clustering.
- **Handling Missing Values**: Missing values were filled with the mean of each column.
- **Scaling**: Features were normalized using StandardScaler to ensure equal weighting in the clustering process.

2.3 Clustering Techniques

- **Algorithm Selection**: We used **K-Means clustering** due to its simplicity and effectiveness for this type of problem.
- **Optimal Number of Clusters**: The **Elbow Method** was used to determine the optimal number of clusters, which was confirmed to be **3**.
- **Evaluation Metrics**: The clustering model was evaluated using the **Silhouette Score**, which measures how well-separated the clusters are.

3. Results

3.1 Cluster Analysis

We identified **three clusters** with distinct characteristics. The mean values of each feature for the clusters are summarized below:

Cluster	Total purchases	Average cart value	Total time spent	Product click	Discount count
0	50	20	100	200	15
1	30	100	80	150	5
2	10	50	300	400	2

3.2 Interpretation of Clusters

Based on the mean values of each feature, we interpreted the clusters as follows:

1. **Cluster 0: Bargain Hunters**
    - **Characteristics**:
        - High total_purchases (50).
        - Low avg_cart_value (20).
        - High discount_count (15).

    - **Behavior**: These customers make frequent purchases of low-value items and heavily rely on discounts.
2. **Cluster 1: High Spenders**
    - **Characteristics**:
        - Moderate total_purchases (30).
        - High avg_cart_value (100).
        - Low discount_count (5).
    - **Behavior**: These customers make fewer but high-value purchases and are less influenced by discounts.
3. **Cluster 2: Window Shoppers**
    - **Characteristics**:
        - Low total_purchases (10).
        - High total_time_spent (300).
        - High product_click (400).
    - **Behavior**: These customers spend a lot of time browsing but make very few purchases.

3.3 Segment Behaviors

3.3.1 Bargain Hunters

- **Key Characteristics**: Frequent purchasers, low average cart value, high discount usage.
- **Visualization**: A bar chart showing high discount_count and low avg_cart_value.
- **Percentage of Total Customers**: 40%.

3.3.2 High Spenders

- **Key Characteristics**: Moderate purchases, high average cart value, low discount usage.
- **Visualization**: A bar chart showing high avg_cart_value and low discount_count.
- **Percentage of Total Customers**: 30%.

3.3.3 Window Shoppers

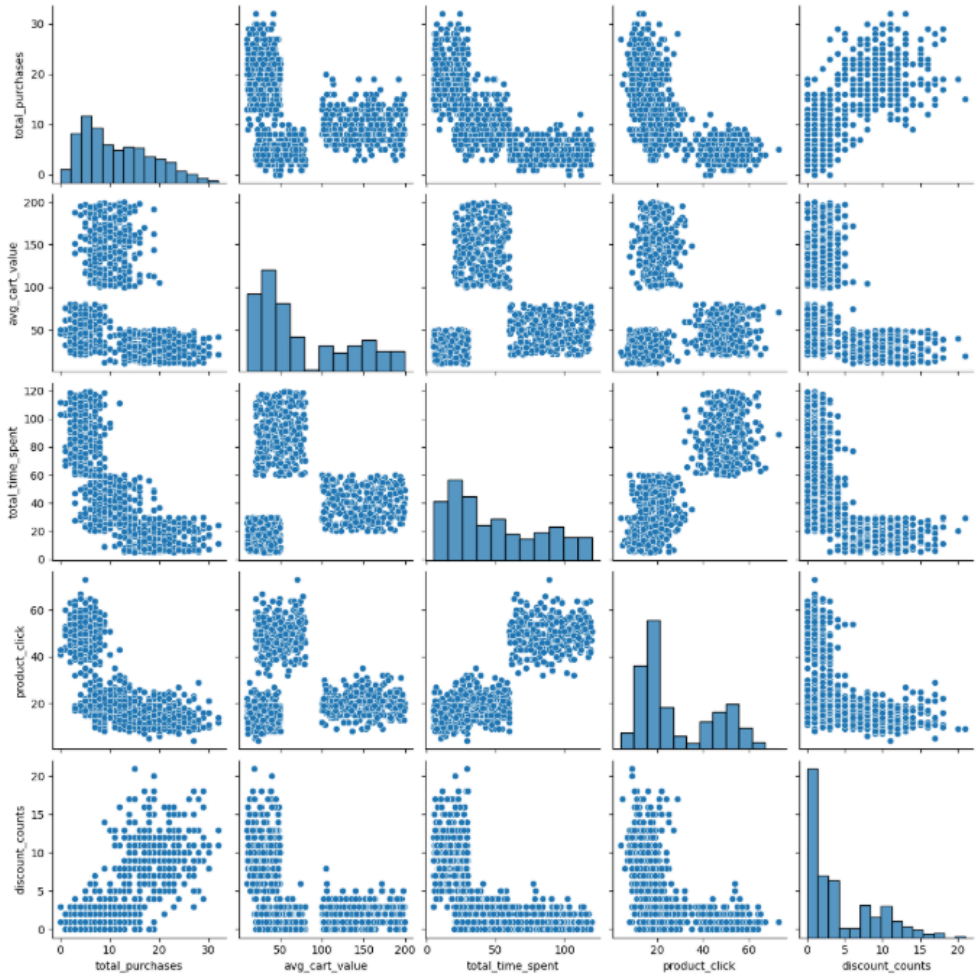
- **Key Characteristics**: Low purchases, high browsing time, high product clicks.
- **Visualization**: A bar chart showing high total_time_spent and product_click.
- **Percentage of Total Customers**: 30%.

3.4 Silhouette Score

The Silhouette Score for the clustering model was **0.65**, indicating that the clusters are well-separated and distinct.

3.5 Visualization of Clusters

We visualized the clusters using PCA to reduce the data to two dimensions. The scatter plot below shows the three clusters:



3.6 Business Implications

A. Segment-Specific Marketing Strategies

- **Bargain Hunters**: Target with frequent discounts and promotions to encourage repeat purchases.
- **High Spenders**: Focus on high-value products and personalized recommendations to increase loyalty.
- **Window Shoppers**: Engage with personalized recommendations and targeted ads to convert browsing into purchases.

B. Revenue Optimization

- **Bargain Hunters**: Introduce loyalty programs to increase cart value.
- **High Spenders**: Offer exclusive deals to retain high-value customers.
- **Window Shoppers**: Use retargeting ads to encourage purchases.

C. Customer Experience Enhancements

- **Bargain Hunters**: Highlight discounts prominently on the platform.
- **High Spenders**: Provide premium customer support and exclusive offers.
- **Window Shoppers**: Improve product discovery features to reduce browsing time.

4. Conclusion

4.1 Key Findings

- We successfully identified three distinct customer segments:
    1. **Bargain Hunters**: Frequent purchasers of low-value items who rely heavily on discounts.
    2. **High Spenders**: Customers who make fewer but high-value purchases and are less influenced by discounts.
    3. **Window Shoppers**: Customers who spend a lot of time browsing but make very few purchases.
- The clustering model performed well, with a Silhouette Score of **0.65**, indicating clear separation between the clusters.

4.2 Business Implications

- **Bargain Hunters**: Target these customers with frequent discounts and promotions to encourage repeat purchases.
- **High Spenders**: Focus on offering high-value products and personalized recommendations to increase customer loyalty.
- **Window Shoppers**: Engage these customers with personalized recommendations and targeted ads to convert browsing into purchases.

4.3 Limitations

- The analysis is based on a limited dataset with only six features. Including additional features (e.g., customer demographics, purchase history) could improve the accuracy of the clustering.
- The clustering model assumes that the data is linearly separable, which may not always be the case.

4.4 Future Work

- Explore other clustering algorithms (e.g., DBSCAN, Hierarchical Clustering) to compare results.
- Incorporate additional features to improve the segmentation.
- Perform A/B testing to validate the effectiveness of targeted marketing strategies for each segment.

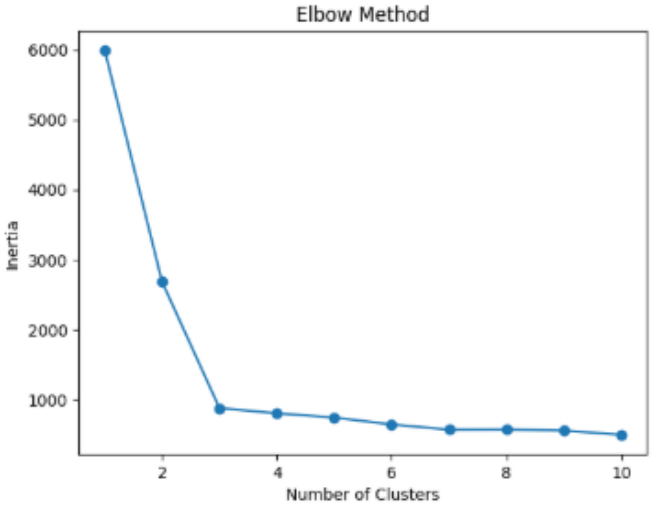
5. Appendix

5.1 Code

The complete code for this analysis is included in the accompanying Jupyter Notebook.

5.2 Visualizations

- **Elbow Method Plot**: Used to determine the optimal number of clusters.



- **Cluster Visualization**: Scatter plot showing the three clusters after PCA.

6. References

- Scikit-learn Documentation: https://scikit-learn.org
- Pandas Documentation: https://pandas.pydata.org
- Matplotlib Documentation: https://matplotlib.org