CDACL-002 – Customer Segmentation Analysis Project

1. Project Overview

This project focuses on **customer segmentation** using U.S. shopping purchase data. The objective is to analyze customer behavior, spending patterns, seasonal trends, and demographic factors to help businesses make **data-driven marketing and discount decisions**.

Business Goals: - Identify product categories suitable for discounts - Analyze card spending behavior across age groups, seasons, and locations - Segment customers into meaningful clusters - Design targeted marketing campaigns for each cluster

---

2. Dataset Description

The dataset contains **3,900 customer purchase records** with demographic, transactional, and behavioral attributes.

Key Attributes

- Customer ID
- Age
- Gender
- Item Purchased
- Category
- Purchase Amount (USD)
- Location
- Season
- Review Rating
- Subscription Status
- Discount Applied
- Promo Code Used
- Previous Purchases
- Payment Method
- Frequency of Purchases

---

3. Data Preprocessing

Steps Performed

1. Removed duplicate records
2. Checked and handled missing values
3. Encoded categorical variables (Gender, Category, Season, Payment Method)
4. Normalized numerical features (Age, Purchase Amount, Previous Purchases)
5. Created derived features such as:

6. Average Spend per Customer
7. Discount Usage Rate

---

4. Exploratory Data Analysis (EDA)

4.1 Product Categories for Discount Application

Categories Recommended for Discounts: - Clothing - Accessories - Footwear

Reasoning: - High purchase frequency but moderate average order value - Strong price sensitivity observed when discounts were applied - Increased repeat purchases with promotional offers

Categories NOT Ideal for Heavy Discounts: - Electronics - Luxury items

Reason: - Lower purchase frequency - Less price sensitivity - Customers prioritize quality over discounts

---

4.2 Card Spending Analysis Based on Age

Age Group Insights: - **18–25 years:** Low to moderate spending, high discount and promo usage - **26–35 years:** Highest average spending, frequent card usage - **36–50 years:** Stable spending, prefers subscriptions - **50+ years:** Lower frequency, higher value per transaction

---

4.3 Impact of Season and Location

Seasonal Trends: - Winter: High spending on clothing and footwear - Summer: Increased accessories and casual wear - Festive seasons: Spike in overall purchase amount

Location Trends: - Urban locations show higher card usage - Suburban regions respond better to promo codes

---

5. Customer Clustering

5.1 Clustering Technique

- Algorithm: **K-Means Clustering**
- Number of clusters: **4** (based on Elbow Method)

Features Used for Clustering

- Age
- Purchase Amount
- Previous Purchases
- Discount Applied
- Frequency of Purchases
- Subscription Status

6. Cluster Analysis

Cluster 1 – Price-Sensitive Shoppers

- Young customers (18–25)
- Low spending
- High discount and promo usage

Strategy: - Flash sales - Student discounts - Limited-time offers

---

Cluster 2 – High-Value Loyal Customers

- Age 26–40
- High purchase amount
- Subscription users

Strategy: - Loyalty programs - Exclusive early access - Premium memberships

---

Cluster 3 – Occasional Seasonal Buyers

- Purchase mainly during festive seasons
- Moderate spending

Strategy: - Seasonal email campaigns - Festival-based discounts

---

Cluster 4 – Premium Shoppers

- Age 40+
- Low frequency but high transaction value

Strategy: - Personalized offers - Premium product recommendations - White-glove customer service

---

7. Key Business Insights

- Discounts significantly increase purchase frequency for young customers
- Subscription status strongly influences loyalty
- Seasonal campaigns drive short-term revenue spikes
- Personalized marketing outperforms generic promotions

---

8. Conclusion

This customer segmentation analysis enables businesses to: - Optimize discount strategies - Improve customer retention - Increase revenue through targeted marketing - Understand seasonal and demographic shopping behaviors

The clustering approach provides actionable insights for marketing teams to design **data-driven, customer-centric campaigns**.

---

9. Tools & Technologies Used

- Python (Pandas, NumPy, Scikit-learn)
- SQL (MySQL – phpMyAdmin)
- Excel / Power BI (Visualization)
- K-Means Clustering

---

10. Future Scope

- Predictive modeling for churn analysis
- Recommendation systems
- Real-time customer segmentation
- Integration with CRM tools

End of Project