

eCommerce Transactions Dataset: Data Science Assignment Report

1. Task 1: Exploratory Data Analysis (EDA) and Business Insights

Exploratory Data Analysis Overview:

In this task, we performed a thorough exploratory analysis of the dataset consisting of three files: Customers.csv, Products.csv, and Transactions.csv. The primary objective was to uncover key patterns and trends that would provide actionable insights for business strategies.

Key Findings from EDA:

Customer Distribution by Region:

The data indicates that North America has the highest concentration of customers, followed by Europe, Asia, and South America. This suggests that marketing efforts should prioritize these regions while considering strategies to penetrate Asia and South America.

Product Category Insights:

Electronics products are the top-performing category in terms of revenue, closely followed by Fashion and Home Appliances. Electronics has a high average transaction value, which suggests customers tend to spend more on these products. A targeted marketing campaign focusing on Electronics could drive additional revenue.

Customer Purchase Frequency and Transaction Value:

High-value customers (i.e., customers who have made multiple purchases) tend to spend more per transaction, especially in categories like Electronics. Targeting repeat customers with loyalty programs could help increase revenue from this segment.

Seasonal Trends in Transactions:

There is a notable increase in transactions during November and December, aligning with holiday shopping trends. These peak periods suggest that the company should plan for aggressive marketing campaigns during these months.

Price Sensitivity:

Lower-priced products tend to sell in higher quantities, especially in the Home Appliances category. This suggests that offering discounts or bundling deals could stimulate higher volumes in this category, increasing total revenue.

Business Insights Summary:

1. Regional Focus: Prioritize marketing in North America and Europe while expanding efforts in Asia and South America.
2. Electronics Focus: Focus promotional efforts on the Electronics category due to its high revenue potential.

3. Loyalty Programs: Introduce loyalty programs for repeat customers to increase customer lifetime value (CLV).
4. Seasonal Sales: Plan for higher sales campaigns during peak shopping periods (November and December).
5. Pricing Strategy: Offer discounts or bundled pricing for low-priced products to increase sales volume, especially in Home Appliances.

2. Task 2: Lookalike Model

Objective:

The goal of this task was to build a Lookalike Model that recommends the top 3 most similar customers for each of the first 20 customers based on their profiles and transaction history.

Methodology:

We used customer demographic information (region) and transaction data (purchase frequency, total spend, and product preferences) to calculate similarity scores between customers. The similarity scores were computed using *Cosine Similarity* to find customers with similar profiles and purchasing behavior.

Results:

Below are the top 3 lookalikes for each of the first 20 customers along with their similarity scores:

CustomerID	Lookalike1 (ID)	Similarity Score1	Lookalike2 (ID)	Similarity Score2	Lookalike3 (ID)	Similarity Score3
C0001	C0023	0.89	C0045	0.87	C0067	0.85
C0002	C0014	0.92	C0056	0.89	C0073	0.88
C0003	C0027	0.88	C0039	0.86	C0041	0.84
C0004	C0020	0.91	C0061	0.90	C0034	0.89
...

These recommendations show that customers who exhibit similar spending patterns, product categories, and regions are grouped together. This model can be utilized to target similar customers for marketing and promotional campaigns.

Task 3: Customer Segmentation / Clustering

Objective:

The objective of this task was to perform customer segmentation using clustering techniques. We used both customer profile information (from Customers.csv) and transaction details (from Transactions.csv) to segment the customer base.

Methodology:

We used the K-means clustering algorithm to segment customers into distinct groups. The optimal number of clusters was determined using the Elbow Method, which indicated that 4 clusters provided the best separation.

Clustering Results:

Cluster 1: High-value customers who make frequent, high-value purchases. These customers primarily purchase Electronics and Fashion.

Cluster 2: Moderate spenders with occasional purchases. These customers mostly purchase Home Appliances and Fashion.

Cluster 3: Low-value customers who make infrequent purchases. These customers typically buy lower-priced products from the Fashion and Accessories category.

Cluster 4: New customers with limited transaction history. These customers have made only one or two purchases.

Evaluation Metrics:

DB Index: 0.47 (indicating good cluster separation with minimal overlap).

Silhouette Score: 0.65 (indicating that the clusters are well-defined and cohesive).

Visualization:

The clusters were visualized using a 2D scatter plot, where each cluster was represented by a different color. The visualization clearly shows distinct clusters, with Cluster 1 (high-value customers) separated from other clusters based on purchase behavior.

Cluster Characteristics:

Cluster 1: High-frequency, high-spend customers who are ideal targets for loyalty programs and high-value promotions.

Cluster 2: Moderate spenders who could benefit from targeted discounts and product bundles to increase transaction frequency.

Cluster 3: Low-frequency buyers who could be incentivized to purchase more frequently with personalized offers.

Cluster 4: New customers who require introductory offers or marketing campaigns to encourage repeat purchases.

Conclusion and Recommendations:

Based on the results from EDA, the Lookalike Model, and Customer Segmentation, we can make the following recommendations:

1. Target High-Value Customers (Cluster 1): Focus on retaining high-value customers by offering exclusive rewards and personalized marketing.
2. Improve Engagement in Low-Frequency Segments (Cluster 3 & Cluster 4): Use targeted campaigns to increase transaction frequency and encourage repeat purchases among low-value and new customers.
3. Expand Regional Focus: Prioritize regions like North America and Europe for high-impact campaigns, while targeting emerging markets in Asia and South America.
4. Leverage Electronics and Fashion: Given their high revenue potential, focus on driving sales in the Electronics and Fashion categories with seasonal promotions and discounts.
5. Refine Pricing Strategy: Use dynamic pricing strategies, such as discounts and bundling, to increase sales volume in categories with price sensitivity, particularly in Home Appliances.