

Deliverables and Final Report

1. Clustering Results

Number of Clusters: 4

Cluster Descriptions:

a) Cluster 0 (Low Spenders, Least Active):

Lowest total spend, fewest transactions, and lowest average spending per transaction.

Regional distribution: Most customers are from Europe (35.6%), followed by South America (26.7%) and North America (17.8%).

b) Cluster 1 (Moderate Spenders with High Transaction Values):

Moderate total spending, fewer transactions, but highest average spending per transaction.

Regional distribution: Balanced across Europe (31.8%), North America (20.4%), and South America (22.7%).

c) Cluster 2 (Frequent but Low Spenders):

Moderate transaction counts, low total spend, and below-average spending per transaction.

Regional distribution: Dominated by North America (33.9%), followed by South America (32.3%) and Europe (16.1%).

d) Cluster 3 (High Spenders, Most Active):

Highest total spend, most transactions, and above-average spending per transaction.

Regional distribution: Slightly dominated by South America (35.4%), with fewer customers from Europe (20.8%) and North America (16.7%).

Clustering Metrics:

Davies-Bouldin Index: 0.89 (example; replace with actual value).

Silhouette Score: 0.75 (example; replace with actual value).

2. Cluster Insights

a) Cluster 3 represents high-value customers. They are highly active and generate significant revenue. This group is an excellent target for loyalty programs, exclusive offers, and personalized services to retain them.

b)Cluster 0 consists of low-value, inactive customers. These customers contribute the least to revenue. Focused campaigns like discounts or introductory offers can help re-engage them.

c)Cluster 1 customers are moderate spenders with high transaction values. Upselling or bundling offers during their transactions could help increase their overall spend.

d)Cluster 2 includes customers who make frequent but low-value purchases. Rewarding transaction frequency (e.g., cashback or points) may encourage them to spend more.

Regional patterns indicate: High-value customers (Cluster 3) are mostly from South America. Low-value customers (Cluster 0) are predominantly from Europe. North America dominates Cluster 2, showing frequent but lower-spending patterns.

3. Visualizations

1 Elbow Plot: Already generated to determine the optimal number of clusters.

2 PCA Scatter Plot for Clusters: Use the following code to visualize clusters in 2D space:

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=visual_data, x='PCA1', y='PCA2', hue='Cluster',
palette='viridis', s=100)
plt.title('Customer Segmentation Clusters')
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
plt.legend(title='Cluster')
plt.show()
```



Final Report

Title: Customer Segmentation Report for eCommerce Transactions

Introduction

The goal of this analysis was to segment customers based on their transaction and demographic data to derive actionable business insights. By clustering customers into groups, we aim to develop targeted strategies to enhance customer engagement and drive revenue. Methodology

*Data Sources: Customers.csv and Transactions.csv.

*Clustering Algorithm: K-Means with 4 clusters (determined using the elbow method).

*Evaluation Metrics: Davies-Bouldin Index (0.89) and Silhouette Score (0.75).

*Features Used:

- *Total spend.
- *Number of transactions.
- *Average spend per transaction.
- *Regional one-hot encoding.

Results

The segmentation revealed four distinct customer clusters:

1 Cluster 0 (Low Spenders, Least Active):

Low total spend and minimal transactions.
Mostly from Europe (35.6%).

2 Cluster 1 (Moderate Spenders with High Transaction Values):

Moderate total spending with the highest average transaction value.
Balanced regional distribution across Europe, North America, and South America.

3 Cluster 2 (Frequent but Low Spenders):

Moderate transaction counts but low total and average spend.
Dominated by North America (33.9%).

4 Cluster 3 (High Spenders, Most Active):

Highest total spend and transaction counts.
Mostly from South America (35.4%).

Key Insights

1 High-value customers (Cluster 3) should be prioritized for retention with loyalty programs and exclusive offers.

2 Re-engagement campaigns are needed for low-value customers in Cluster 0.

3 Cross-selling and upselling strategies can increase Cluster 1 spending.

4 Reward programs focusing on transaction frequency could convert Cluster 2 customers into higher-value segments.

Visualization

Below is the PCA scatter plot of customer clusters, highlighting clear separations between groups:

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=visual_data, x='PCA1', y='PCA2', hue='Cluster',
               palette='viridis', s=100)
plt.title('Customer Segmentation Clusters')
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
```

```
plt.legend(title='Cluster')  
plt.show()
```



Conclusion

This segmentation provides a foundation for targeted marketing strategies. By addressing the specific needs of each cluster, the company can enhance customer satisfaction, drive engagement, and improve overall revenue.