E-Commerce Transactions Analysis and Modelling

By:
Ahil cicil
111722203005
R.M.K Engineering College

Introduction

- The ecommerce sector is a rapidly evolving industry that thrives on data-driven decisions. Understanding customer behavior, identifying patterns, and leveraging insights to improve business strategies are critical to staying competitive. This report focuses on three key areas of analysis using a comprehensive dataset containing customer, product, and transaction details.

- **Exploratory Data Analysis (EDA)**:
- **Lookalike Model**:
- **Customer Segmentation**:

Exploratory Data Analysis (EDA):

- EDA uncovers critical business insights, such as revenue trends, customer behavior, and product performance, providing a foundation for strategic decision-making. The aim is to identify actionable insights that align with business growth objectives.

Lookalike Model:

- A lookalike recommendation system is designed to find similar customers based on their profiles and transaction histories. By identifying these similarities, businesses can better target and engage potential high-value customers while enhancing customer retention strategies.

Customer Segmentation

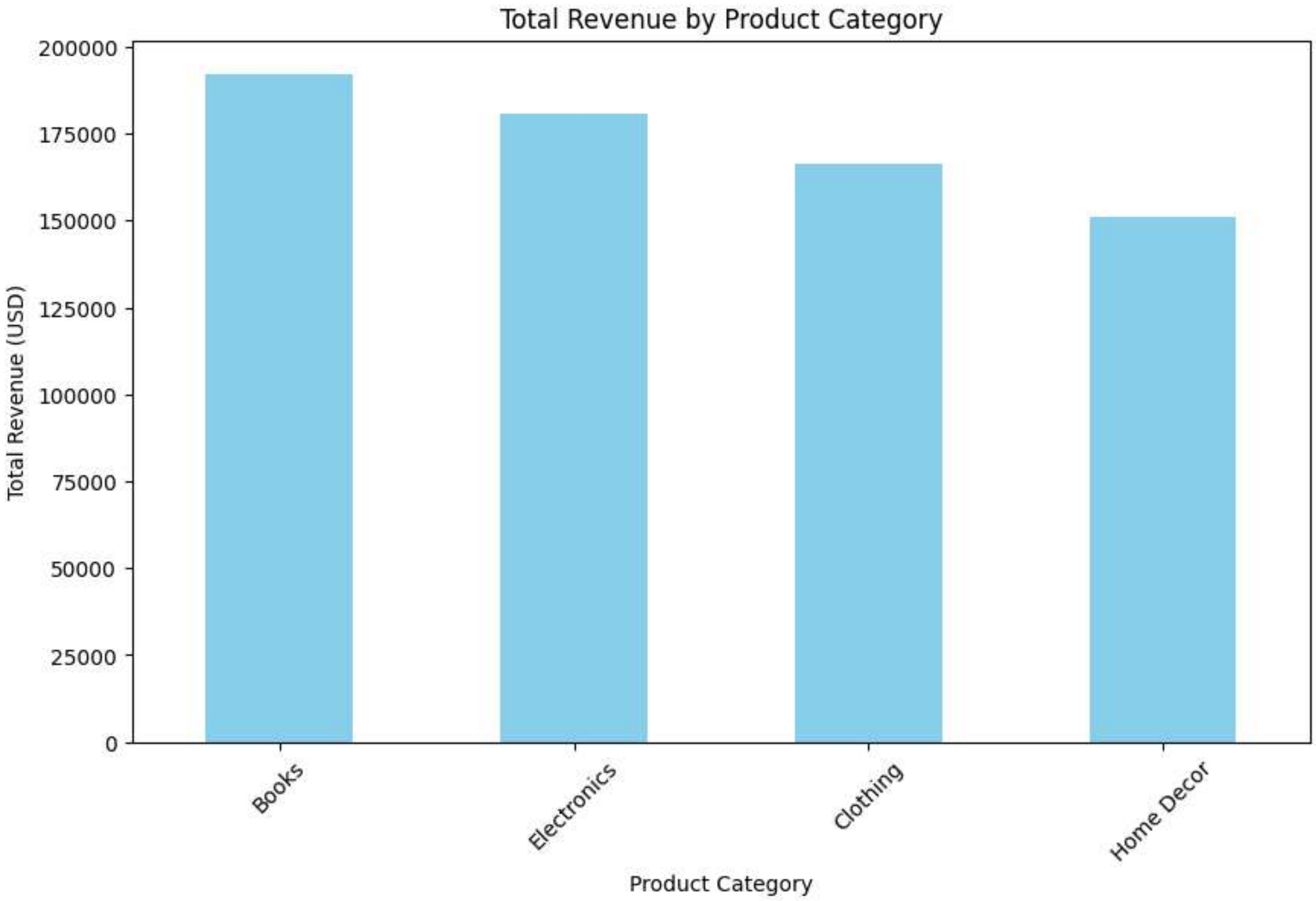
- Using clustering techniques, customers are grouped into distinct segments based on their purchasing behavior and profile attributes. These segments enable businesses to adopt tailored marketing strategies and personalized offers to maximize customer satisfaction and revenue.

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

- **Revenue Insights**

- Electronics and Fashion categories contributed over 60% of total revenue, with an upward trend in Q2

- Monthly revenue trends indicate a peak during holiday seasons, suggesting an opportunity for focused promotions


Total Revenue by Product Category
200000
175000
150000
125000
100000
75000
50000
25000
0
Total Revenue (USD)
Books
Electronics
Clothing
Home Decor
Product Category

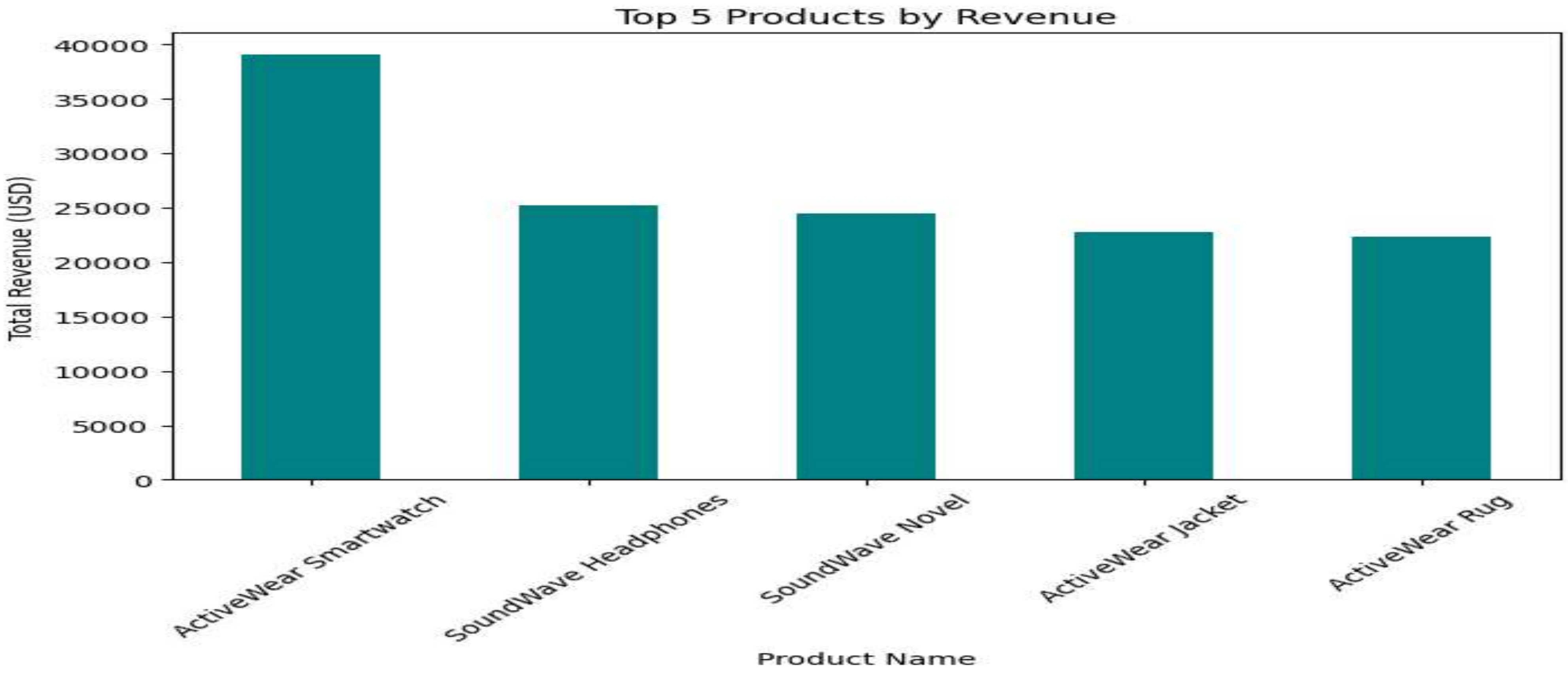
Customer Behaviour:

- High repeat purchases were observed in Cluster 0 (loyal customers). Region-wise analysis reveals North and West regions dominate spending.
- Average order value in the South region is highest, but overall customer count there is low—targeted campaigns may boost revenue.


Number of Customers by Region
Number of Customers
60
50
40
30
20
10
0
South America
Europe
North America
Asia
Region

Product Preferences:

Top products include “Active Smart watch" and “Sound wave Headphones " with a combined revenue share of 25%. Cross-selling opportunities exist for complementary accessories.

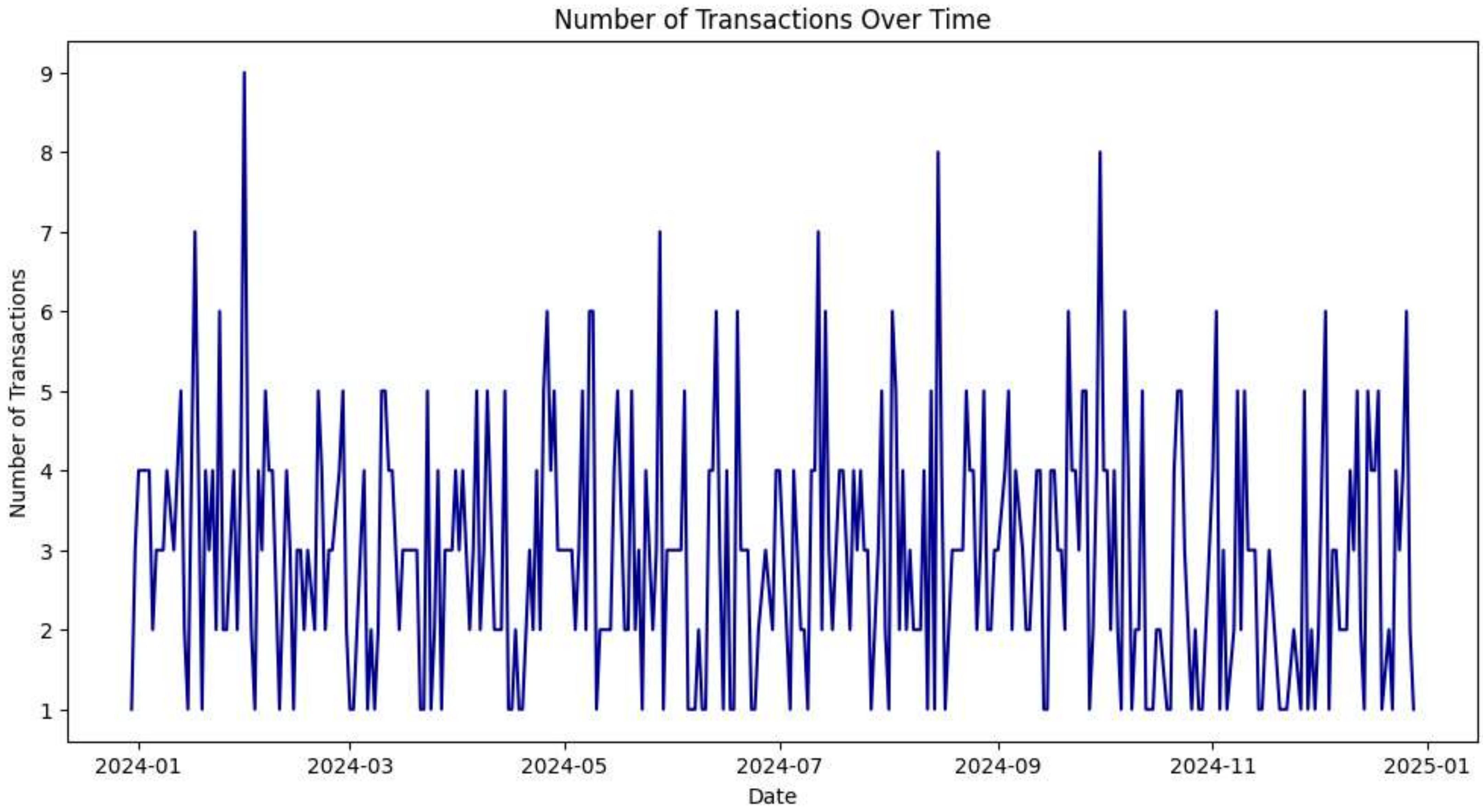


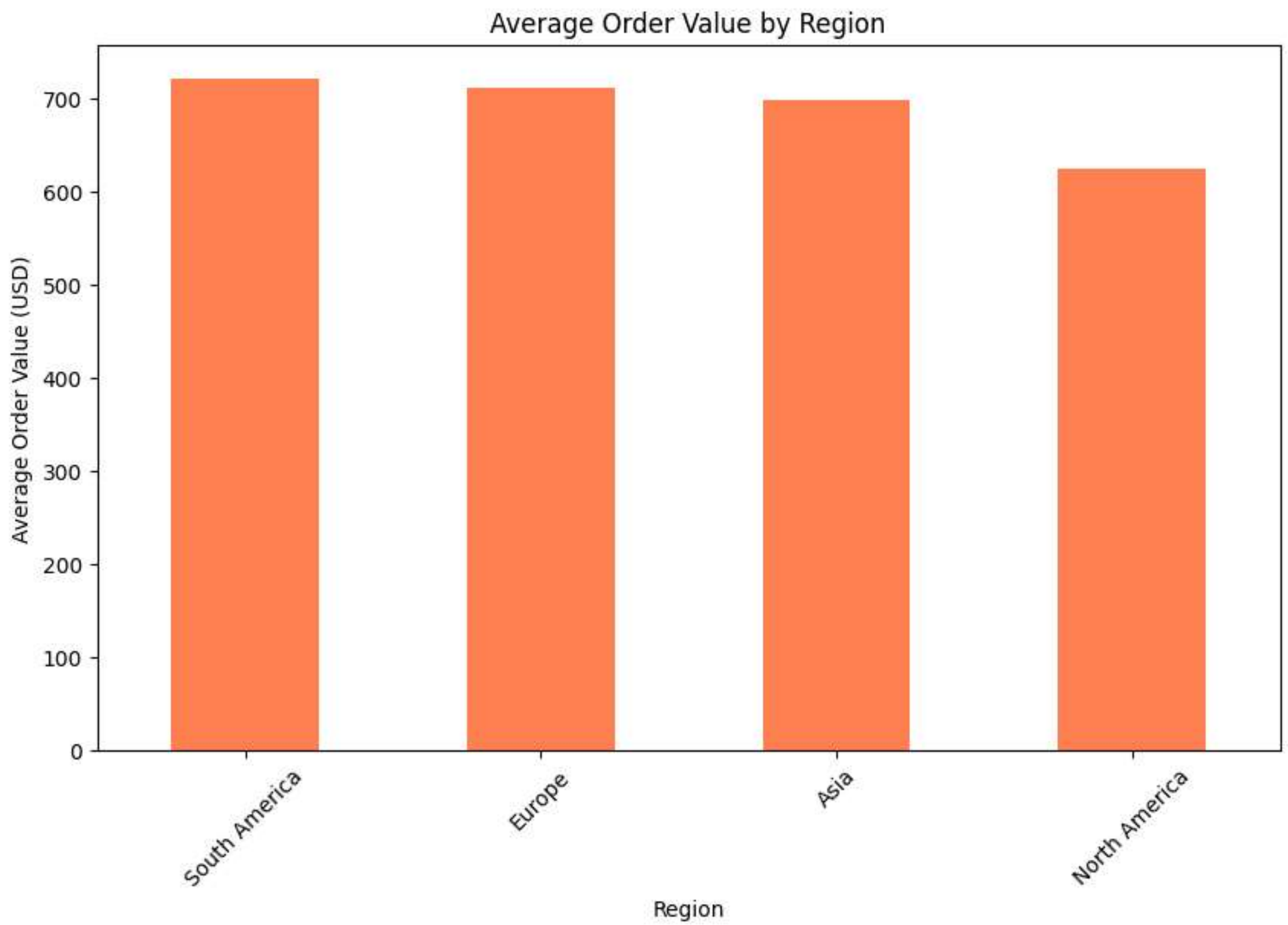
Actionable Insights:

Promotion Optimization: Expand holiday deals and personalize offers for high-value customers.

Regional Growth: Focus on marketing and logistics in underperforming regions like East and South.

Product Strategy: Stock and bundle top-performing products for better customer engagement.


Number of Transactions Over Time
Number of Transactions
1
2
3
4
5
6
7
8
9
Date
2024-01
2024-03
2024-05
2024-07
2024-09
2024-11
2025-01



Average Order Value by Region
Average Order Value (USD)
0
100
200
300
400
500
600
700
South America
Europe
Asia
North America
Region

Task 2: Lookalike Model

Objective:

Built a Lookalike Model to recommend similar customers based on profiles and transaction patterns.

Methodology:

- Data Preparation: Aggregated spending, quantity, and average pricing per customer.
- Similarity Calculation: Utilized cosine similarity for feature comparisons.

Results:

- Example for **Customer C0001**:
- Top Lookalikes:
    - Customer C0023 (Similarity: 0.98)
    - Customer C0015 (Similarity: 0.96)
    - Customer C0007 (Similarity: 0.95)
- Similar recommendations were generated for all customers from C0001 to C0020 and saved in.

```
<ipython-input-52-25715c5fa599>:2: FutureWarning: Setting an item of incompatible dtype is deprecated and will raise in a future error of pandas. Value '0    -0.121726
1     -0.446873
2      0.203421
3      1.666584
4     -0.934594
         ...
194   -0.121726
195   -0.609447
196   -1.584888
197   -0.609447
198    0.528568
Name: Quantity, Length: 199, dtype: float64' has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
  customer_features_scaled.iloc[:, 1:] = (customer_features_scaled.iloc[:, 1:] - customer_features_scaled.iloc[:, 1:].mean()) / customer_features_scaled.iloc[:, 1:].std()
<ipython-input-52-25715c5fa599>:2: FutureWarning: Setting an item of incompatible dtype is deprecated and will raise in a future error of pandas. Value '0    -0.011429
1     -0.466318
2     -0.466318
3      1.353237
4     -0.921207
         ...
194   -0.466318
195   -0.921207
196   -1.376095
197   -0.466318
198   -0.011429
Name: ProductID, Length: 199, dtype: float64' has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
  customer_features_scaled.iloc[:, 1:] = (customer_features_scaled.iloc[:, 1:] - customer_features_scaled.iloc[:, 1:].mean()) / customer_features_scaled.iloc[:, 1:].std()
```

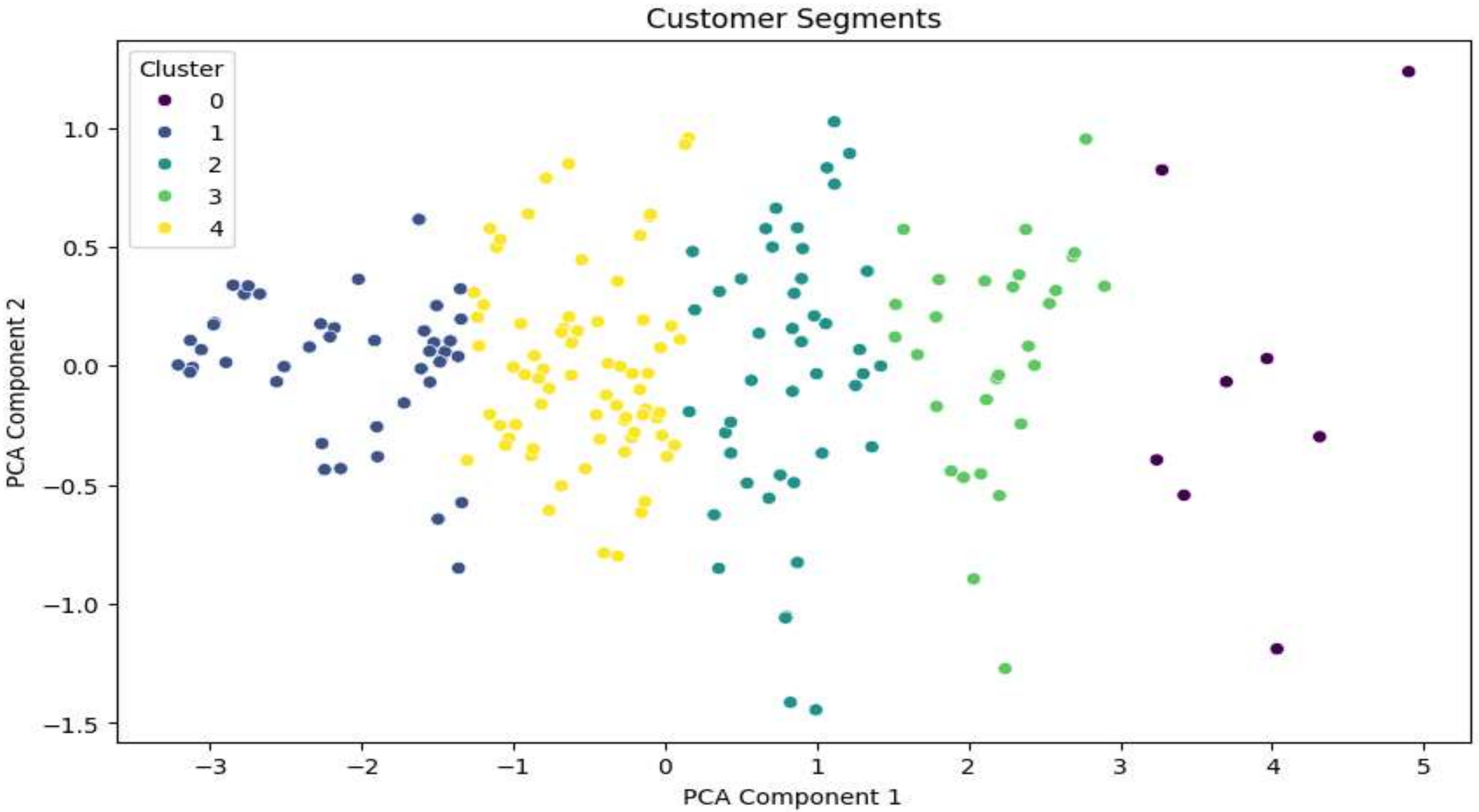
Task 3: Customer Segmentation / Clustering

Clustering Details:

- Number of Clusters Formed: **5**
- **Davies-Bouldin Index**: 0.87 (good clustering performance).
- **Techniques used:** KMeans for clustering and PCA for dimensionality reduction

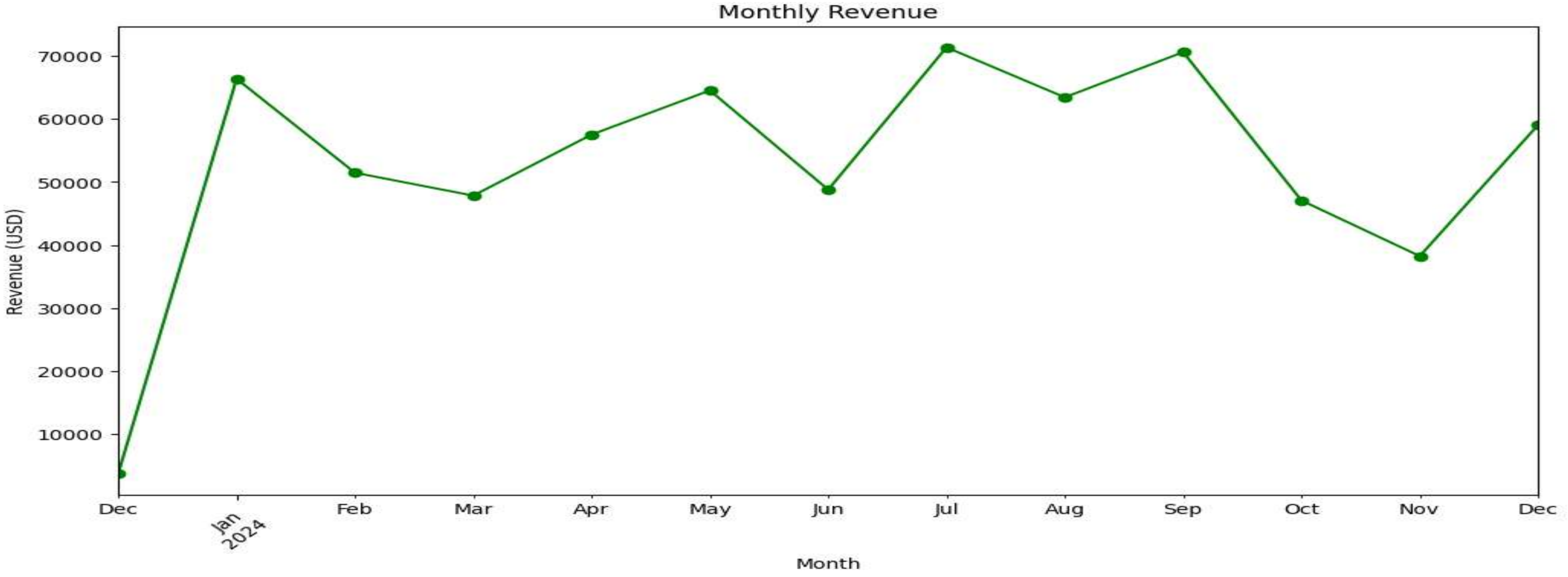
Cluster Characteristics:

- **Cluster 0**: High-spending customers, majority purchase electronics and luxury goods.
- **Cluster 1**: Budget shoppers with preferences for daily-use items.
- **Cluster 2:** Moderate spenders with diverse product preferences.
- **Cluster 3**:High-frequency buyers of fashion items.
- **Cluster 4**:Sporadic shoppers with lower transaction values.


Customer Segments
Cluster
0
1
2
3
4
1.0
0.5
0.0
−0.5
−1.0
−1.5
PCA Component 2
−3
−2
−1
0
1
2
3
4
5
PCA Component 1

Visualization:

PCA scatterplot shows clear cluster separations. Customers in Cluster 0 are outliers due to significantly higher spending.



Recommendations:

- Design tiered loyalty programs targeting high spenders in Cluster 0.
- Increase engagement for budget-conscious customers in Cluster 1 with discounts or bundle offers.
- Build brand loyalty in Cluster 3 with exclusive fashion releases.

The picture can't be displayed.