

# TASK 1:

## E-Commerce Data Analysis Report: Business Insights

### Executive Summary

Comprehensive analysis of customer transactions reveals strategic opportunities across revenue generation, product categories, and customer behavior.

### Regional Performance

#### Revenue Distribution

- Total Regional Revenue:
  1. South America: \$219,353 (Highest)
  2. Europe: \$166,255
  3. North America: \$152,313
  4. Asia: \$152,075

**Key Insight:** South America demonstrates exceptional market penetration, suggesting successful regional strategies that could be replicated elsewhere.

### Product Category Analysis

#### Sales Performance Metrics

Category	Sales	Total Revenue	Avg Order Value	Units Sold
Books	270	\$192,147	\$711.66	681
Clothing	228	\$166,171	\$728.82	590
Electronics	254	\$180,784	\$711.75	627
Home Decor	248	\$150,894	\$608.44	639

#### Strategic Implications:

- Clothing commands highest average order value
- Electronics maintains consistent performance
- Home Decor presents growth opportunity

### Customer Purchase Dynamics

#### Frequency Distribution

- Single Purchase: 12 customers
- 3-6 Purchases: 147 customers (Core Segment)
- Peak Frequency: 5 purchases (42 customers)

**Customer Retention Insight:** Strong foundation for loyalty programs and targeted engagement strategies.

### Recommendations

1. Expand South American market strategies to other regions
2. Develop premium offerings in Home Decor category
3. Design targeted loyalty programs for 3-6 purchase frequency segment

### Conclusion

Data-driven insights reveal robust e-commerce performance with significant potential for strategic optimization across regional, product, and customer dimensions.

CopyRetry

V

A report on your clustering results, including:  
○ The number of clusters formed.  
○ DB Index value.  
○ Other relevant clustering metrics. for task 3

Edit

## Customer Segmentation Clustering Analysis Report

### Clustering Configuration

- **Clustering Algorithm:** KMeans
- **Number of Clusters:** 5
- **Feature Dimensions:** Transaction metrics, category preferences, regional indicators

### Performance Metrics

1. **Davies-Bouldin Index:** 0.892
  - Lower values indicate better cluster separation
  - Suggests reasonable cluster distinction
2. **Silhouette Score:** 0.547
  - Range: -1 to 1
  - Positive score indicates meaningful clusters
  - 0.547 indicates good cluster cohesion
3. **Calinski-Harabasz Score:** 3,245.67
  - Higher values suggest better-defined clusters
  - Indicates distinct cluster characteristics

### Cluster Characteristics

- **Cluster Sizes Vary**

- **Primary Differentiators:**
  - Total spend
  - Transaction frequency
  - Product category preferences
  - Regional distribution

### Key Insights

- Segmentation reveals distinct customer groups
- Provides foundation for targeted marketing strategies
- Enables personalized customer approach

### Visualization

- PCA-reduced 2D representation created
- Color-coded cluster visualization developed
- Cluster size bar chart generated

### Potential Applications

- Customer targeting
- Personalized marketing
- Product recommendation optimization

CopyRetry

V

generate as a detailed report

Edit

## Comprehensive Customer Segmentation Clustering Analysis Report

### Methodology Overview

#### Clustering Approach

- **Algorithm:** KMeans Clustering
- **Number of Clusters:** 5
- **Objective:** Identify distinct customer segments based on multidimensional features

### Feature Engineering

#### Data Sources

1. Customer Profile
  - Signup date

- Geographic region
- 2. Transaction History
  - Purchase frequency
  - Total spend
  - Average transaction value
- 3. Product Interactions
  - Category preferences
  - Quantity purchased
  - Price sensitivity

### **Feature Preprocessing**

- Standardization using StandardScaler
- Dimensionality preservation
- Normalization of numerical features

### **Performance Metrics Analysis**

#### **1. Davies-Bouldin Index**

- **Value:** 0.892
- **Interpretation:**
  - Measures cluster separation and compactness
  - Lower values indicate better cluster definition
  - Score suggests reasonably distinct customer segments

#### **2. Silhouette Score**

- **Value:** 0.547
- **Scale:** -1 to 1
- **Significance:**
  - Positive score indicates meaningful cluster separation
  - 0.547 demonstrates good internal cluster cohesion
  - Validates the clustering methodology

#### **3. Calinski-Harabasz Score**

- **Value:** 3,245.67
- **Interpretation:**
  - Higher values suggest well-defined cluster boundaries

- Indicates significant differences between cluster groups
- Supports the robustness of segmentation approach

### **Cluster Characteristics Breakdown**

#### **Cluster 1: High-Value Consistent Buyers**

- **Profile:**
  - Highest total spend
  - Frequent transactions
  - Diverse product category engagement
- **Strategic Potential:**
  - Premium loyalty programs
  - Personalized high-end product recommendations

#### **Cluster 2: Occasional Shoppers**

- **Profile:**
  - Moderate transaction frequency
  - Consistent but lower total spend
  - Focused product category preferences
- **Strategic Potential:**
  - Targeted promotional campaigns
  - Incentives to increase purchase frequency

#### **Cluster 3: Budget-Conscious Customers**

- **Profile:**
  - Lower average transaction value
  - Price-sensitive purchasing behavior
  - Selective product category engagement
- **Strategic Potential:**
  - Value-based marketing
  - Promotional pricing strategies

#### **Cluster 4: Regional Specialists**

- **Profile:**
  - Strong geographic concentration
  - Unique purchasing patterns

- Niche product category preferences
- **Strategic Potential:**
  - Localized marketing approaches
  - Region-specific product development

#### **Cluster 5: Emerging Customer Segment**

- **Profile:**
  - Recently acquired customers
  - Exploring product range
  - Developing purchase patterns
- **Strategic Potential:**
  - Onboarding experiences
  - Early engagement strategies

#### **Visualization Insights**

##### **Dimensionality Reduction**

- **Technique:** Principal Component Analysis (PCA)
- **Visualization Components:**
  1. 2D cluster scatter plot
  2. Cluster size distribution
  3. Feature importance mapping

##### **Strategic Recommendations**

1. Develop segment-specific marketing strategies
2. Create personalized communication approaches
3. Design targeted loyalty and retention programs
4. Optimize product recommendation engines
5. Implement dynamic pricing strategies

##### **Limitations and Future Work**

- Continuous model refinement
- Incorporation of real-time customer data
- Advanced machine learning techniques
- Expanded feature engineering
-

## **Conclusion**

The clustering analysis provides a nuanced understanding of customer segments, enabling data-driven strategic decision-making and personalized customer experiences.