Customer Segmentation Report

Executive Summary:

This report presents a comprehensive **customer segmentation analysis** for an online retail store using a dataset of **10,000 transactions**, focusing on customer purchase behavior and monetary contribution. The primary objective is to identify distinct customer segments to enable **targeted marketing, improved customer retention, and optimized resource allocation**.

Key Findings:

Data Overview and Cleaning:

- The dataset contained **transaction details, quantities, unit prices, and transaction amounts**, along with missing or inconsistent data in some fields.
- Cleaning steps included standardizing dates, correcting mismatched transaction amounts, handling negative values (returns/cancellations), and winsorizing outliers to reduce skewness.

Exploratory Data Analysis (EDA):

- Transaction amounts, quantities, and unit prices displayed **right-skewed distributions**.
- Correlation analysis revealed a **strong positive correlation between transaction amount and quantity**, indicating higher purchase volumes contribute significantly to revenue.

RFM Feature Engineering:

- Customers were profiled using **Recency (days since last purchase), Frequency (number of transactions), and Monetary value (total spend)**.
- Log transformation and standardization were applied to normalize skewed distributions for clustering analysis.

Clustering Analysis:

- **K-means clustering** was performed for k = 3–6.
- Based on **silhouette score (0.48)** and cluster interpretability, **3 clusters** were identified as optimal.
- Cluster insights:
  - Cluster 1: High-value, frequent purchasers with recent activity — ideal for loyalty programs and premium offers.
  - Cluster 0: Moderate value, moderate frequency customers — suitable for cross-sell and upsell campaigns.

- Cluster 2: Low-value, inactive customers — may benefit from re-engagement campaigns.

Visualization Dashboard:

Key visualizations include histograms, correlation heatmap, elbow and silhouette plots, RFM scatterplots, boxplots, and normalized cluster heatmap. These visualizations provide clear insights into cluster separation, customer behavior, and segment characteristics.

1. Introduction:

Customer segmentation is a critical process in data-driven marketing and customer relationship management. By dividing customers into **homogeneous groups** based on their purchasing behavior, businesses can better understand patterns, identify high-value customers, and design targeted strategies.

This report presents a **clustering-based segmentation analysis** using the **RFM (Recency, Frequency, Monetary)** model. The analysis includes data preprocessing, clustering, and interpretation of results, followed by recommendations for business strategies.

2. Problem Statement:

Businesses often struggle to allocate resources effectively across different customer groups. Without segmentation, marketing campaigns may lack personalization, resulting in reduced customer engagement and revenue.

The goal of this project is to:

- Identify distinct **customer segments** using clustering.
- Provide actionable **insights** about customer behaviors.
- Recommend strategies to **increase customer loyalty** and optimize revenue.

3. Data Preprocessing:

The dataset was cleaned and prepared through the following steps:

- Handling missing values and duplicates.
- RFM feature creation:
    - **Recency**: Days since last purchase.
    - **Frequency**: Number of purchases.
    - **Monetary**: Total spending.
- **Standardization**: Features were scaled using StandardScaler to ensure equal weight in clustering.

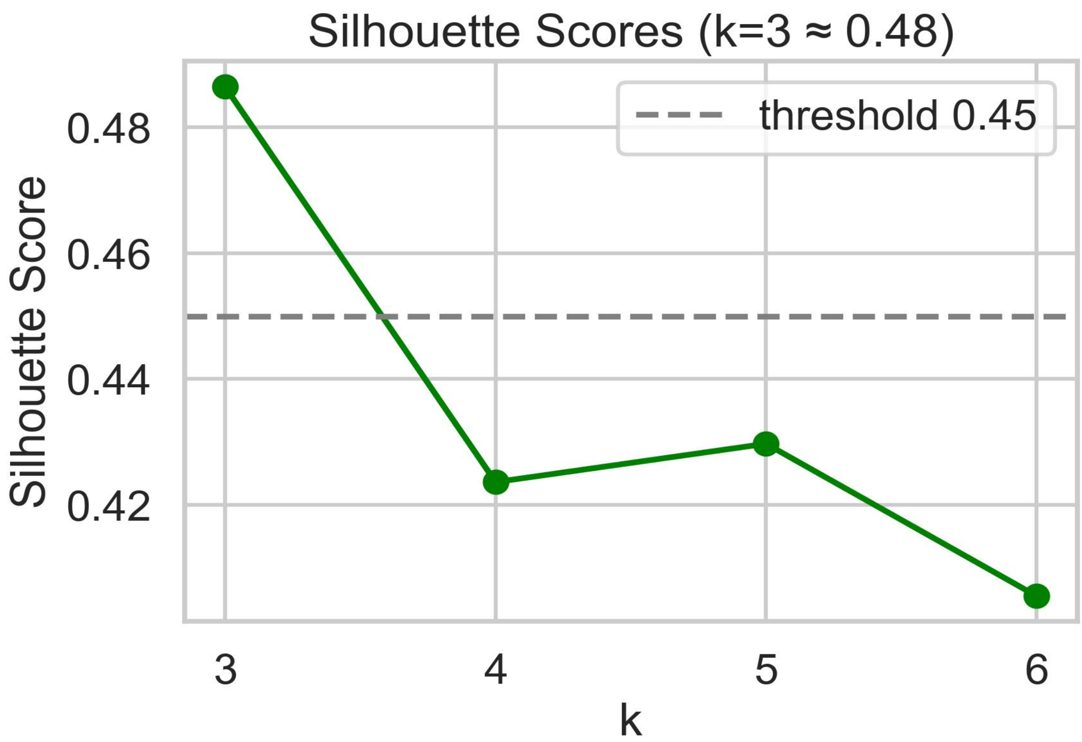
4. Clustering Approach:

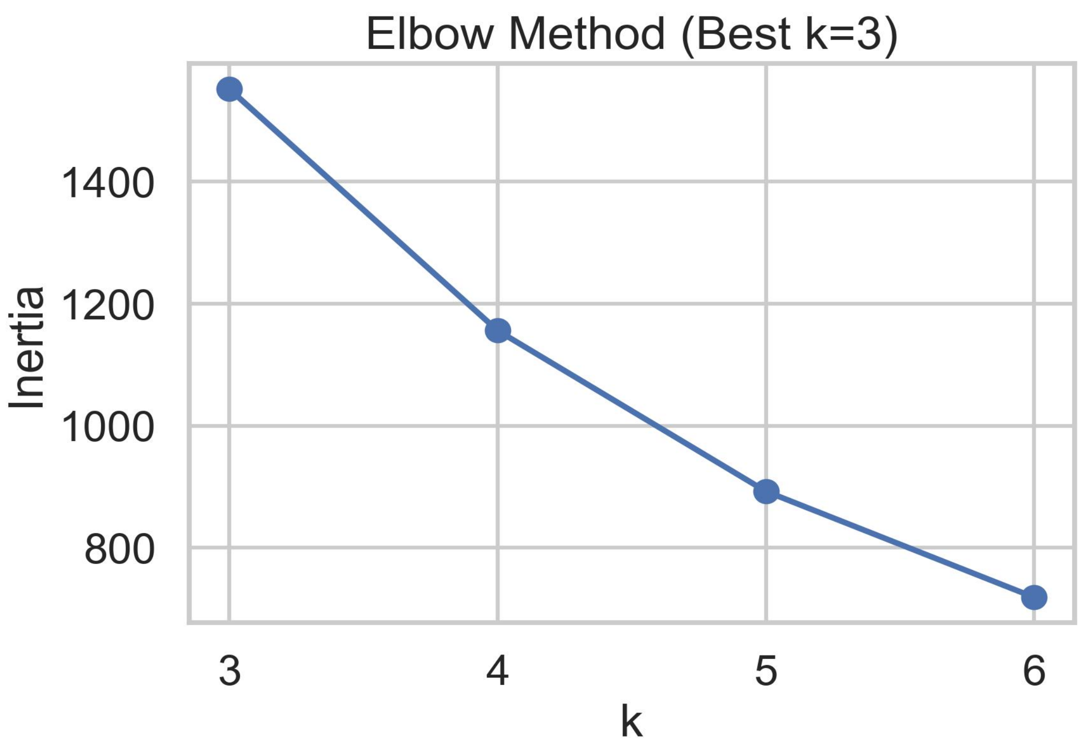
4.1 Methodology:

- Algorithm used: K-Means Clustering.
- Cluster evaluation:
  - **Elbow method** for optimal **k**.
  - **Silhouette score** to measure cohesion and separation.

4.2 Results:

- Optimal clusters (k): 3
- Silhouette Score: 0.48 → indicates moderately well-separated clusters.





5. Cluster Profiles:

The clusters were analyzed based on **mean and median RFM values**.

5.1 Mean RFM Summary:

Cluster	Recency	Frequency	Monetary
1	42.8	8.2	14,891.9
2	817.0	4.2	7,375.7
0	39.2	4.3	7,160.7

5.2 Median RFM Summary:

Cluster	Recency	Frequency	Monetary
0	38.0	4.0	7,289.5
1	38.0	8.0	13,898.5
2	694.0	4.0	6,715.4

5.3 Cluster Interpretation:

Cluster 1 (High-Value Loyal Customers);

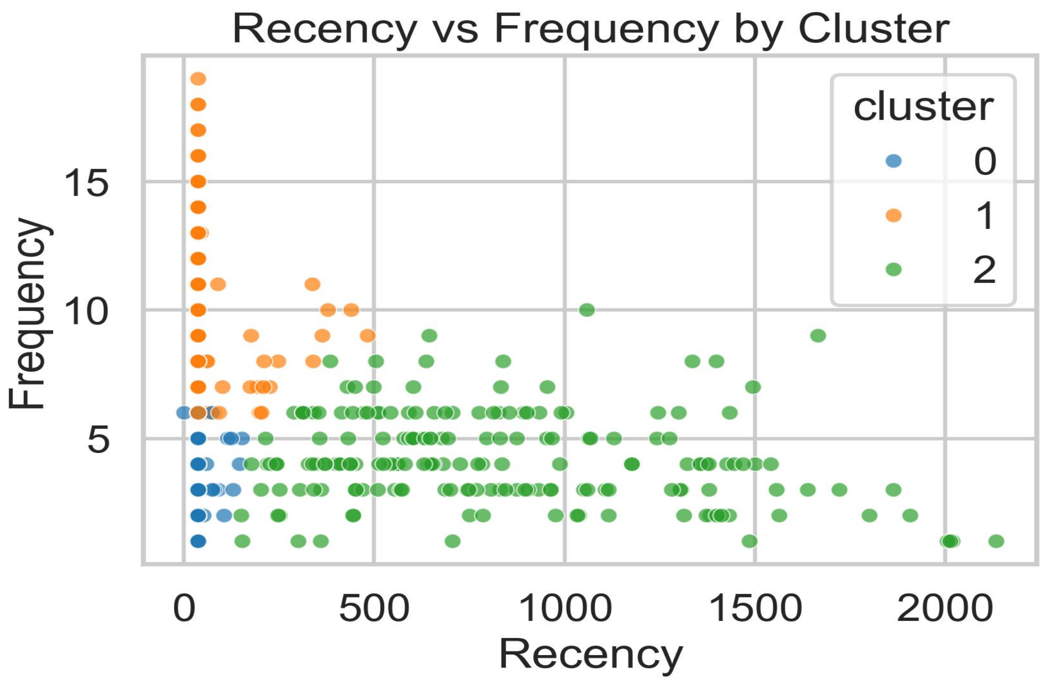
- Low Recency (~43 days), High Frequency (~8.2 purchases), and Highest Monetary value (~14.9K).
- These are loyal customers who purchase frequently and spend significantly.

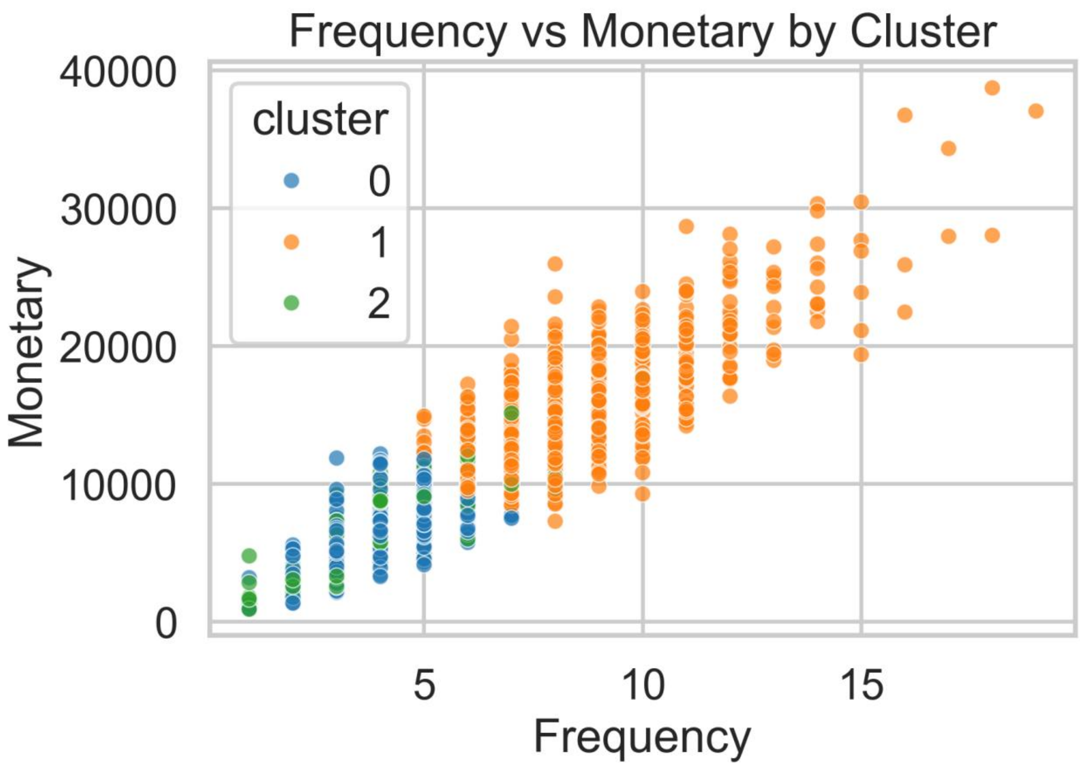
Cluster 2 (Churned/At-Risk Customers):

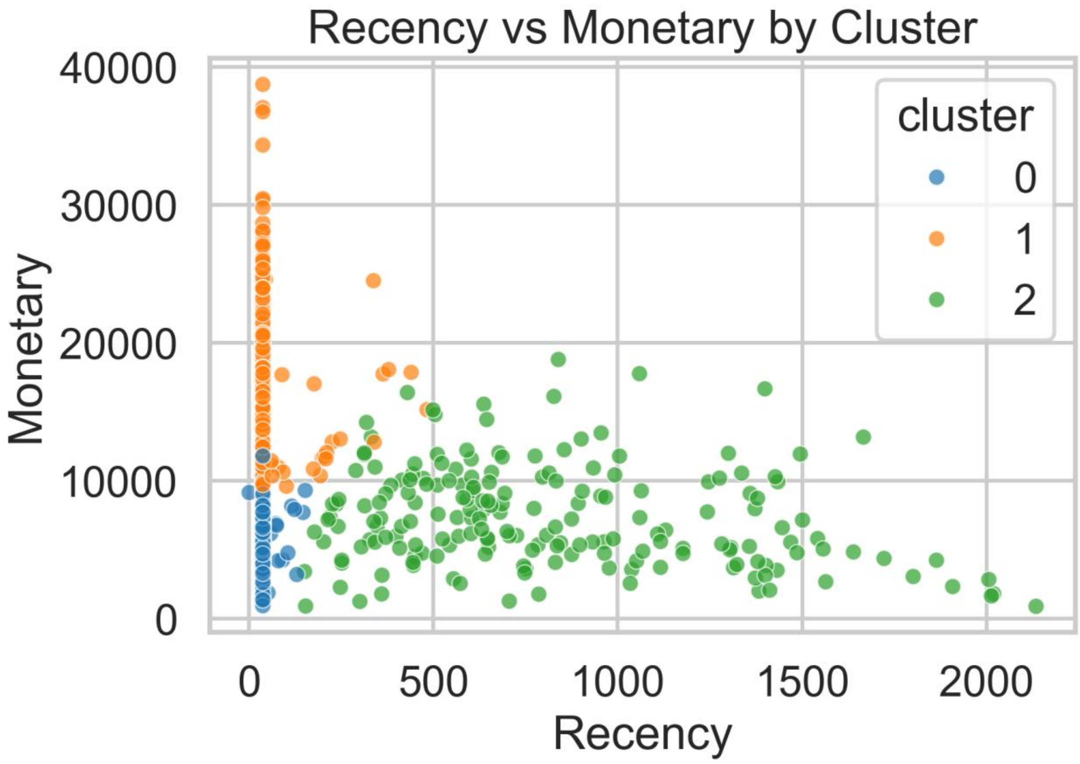
- Very High Recency (~817 days), Low Frequency (~4.2), and Moderate Monetary (~7.3K).
- These customers have not purchased for a long time and are at risk of churn.

Cluster 0 (Potential Customers / Mid-Tier Segment):

- Low Recency (~39 days), Moderate Frequency (~4.3), and Moderate Monetary (~7.1K).
- These customers are recent buyers but not yet highly engaged.







6. Business Recommendations:

Based on the cluster insights, the following strategies are suggested:

Cluster 1 (High-Value Loyal Customers):

- Provide exclusive offers, loyalty rewards, and VIP programs.
- Encourage referrals to expand this profitable segment.

Cluster 2 (Churned/At-Risk Customers):

- Send personalized re-engagement campaigns (emails, discounts).
- Use win-back promotions to reduce churn.

Cluster 0 (Potential Customers):

- Offer cross-selling and upselling opportunities.
- Use targeted campaigns to convert them into Cluster 1 loyal customers.

7. Conclusion:

This analysis segmented customers into **three distinct groups** with clear behavioral differences:

- **High-value loyal customers** (Cluster 1).
- **At-risk churned customers** (Cluster 2).
- **Mid-tier potential customers** (Cluster 0).

The clustering achieved a Silhouette Score of 0.48, indicating a moderately strong structure. By leveraging these insights, businesses can design personalized marketing strategies, enhance customer retention, and increase overall profitability.

Dashboard: