**Transaction Data Analysis Report**

**Introduction**

This report analyzes transaction data to identify patterns and trends using clustering and regression techniques. The data comprises details like transaction amounts, dates, and customer demographics.

**Data Preparation and Analysis Methodology**

The dataset underwent cleaning to handle missing values and standardize formats, specifically converting transaction dates to a consistent datetime format. Analysis involved k-means clustering to discover customer segments and linear regression to understand trends in transaction amounts over time.

**Results and Discussion**

**1. Distribution of Transaction Amounts**

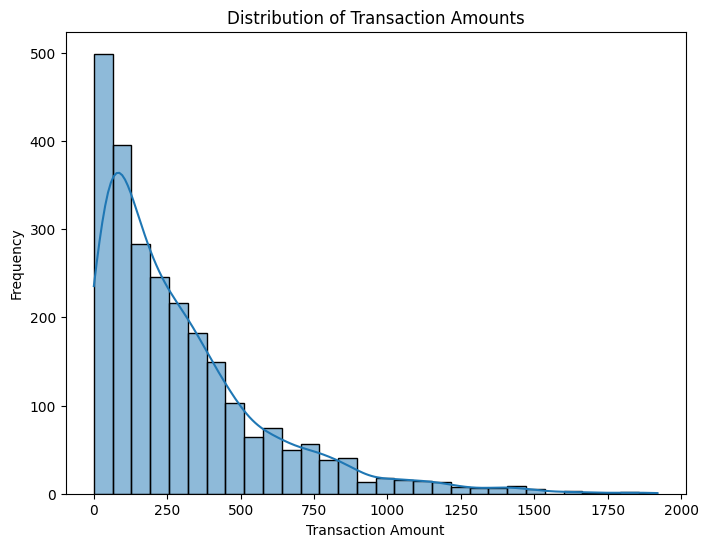


Figure 1Distribution of Transaction Amounts

The histogram reveals a skewed distribution, with most transactions being of lower value, highlighting consumer preference for smaller transactions.

**2. Transaction Amount Over Time**

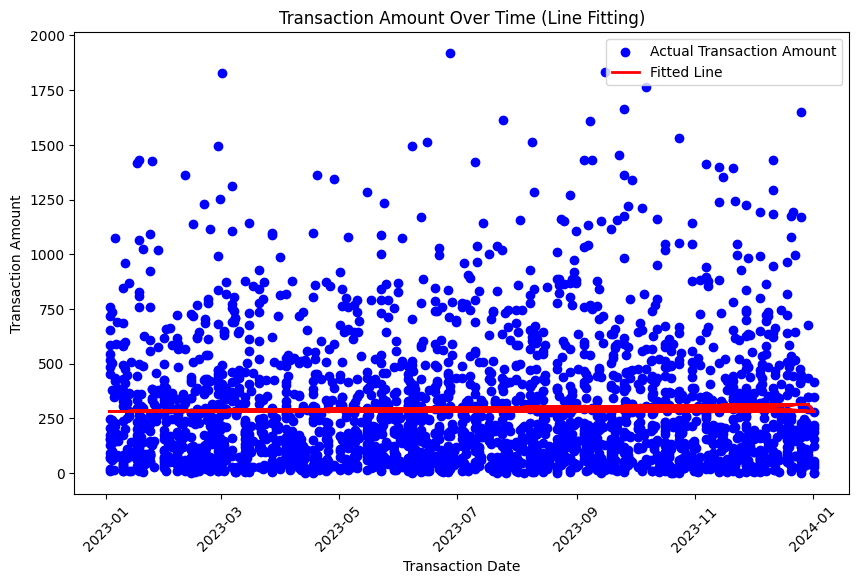
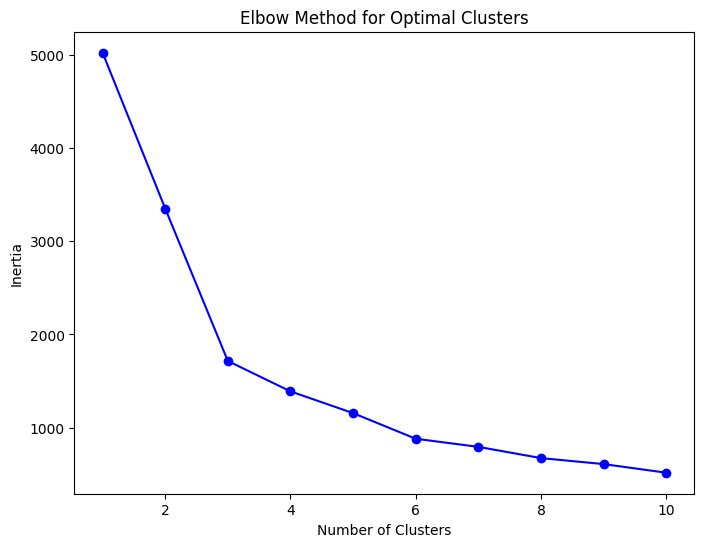


Figure 2Transaction Amount Over Time

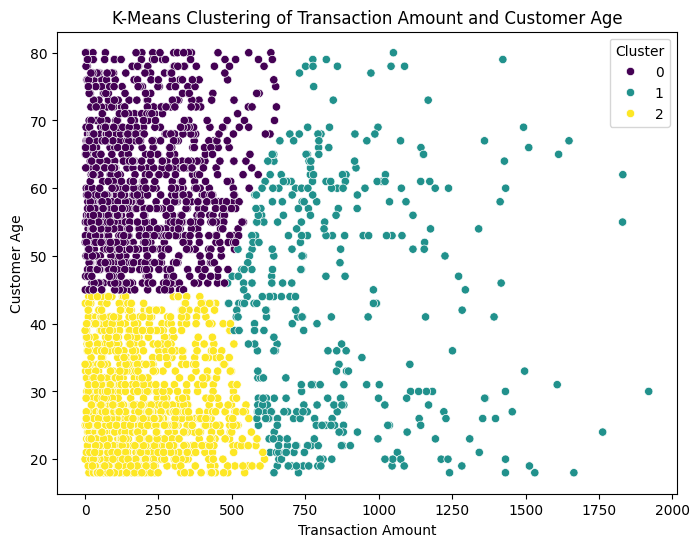
The scatter plot and fitted line show transaction amounts are consistent over time, indicating stability in spending behavior.

**3. Elbow Method for Optimal Clusters**

   
*Figure 3: Elbow Method for Optimal Clusters*

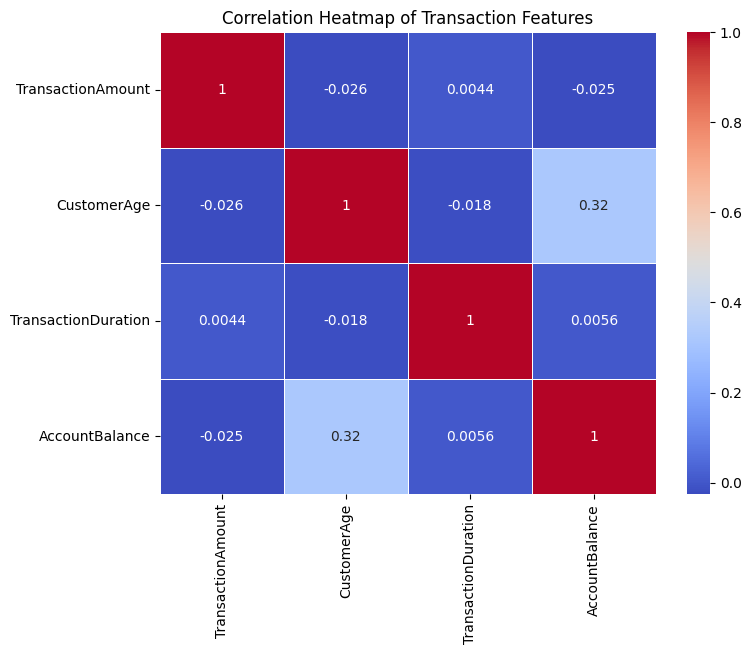
The elbow plot suggests three as the optimal number of clusters, balancing detail and model complexity.

**4. K-Means Clustering of Transaction Amount and Customer Age**

   
*Figure 4: K-Means Clustering of Transaction Amount and Customer Age*

Clusters reveal distinct patterns: Cluster 0 captures older, higher-spending customers; Cluster 1 includes middle-aged, moderate spenders; and Cluster 2 consists of younger, lower-spending customers.

**5. Correlation Heatmap of Transaction Features**

   
*Figure 5: Correlation Heatmap of Transaction Features*

The heatmap shows weak correlations between transaction features, suggesting diverse influences on customer behavior.

**Conclusions**

The linear regression analysis indicates stable spending patterns over the examined period. Clustering has identified customer segments based on age and spending, useful for targeted marketing. The silhouette score of 0.47 confirms the effectiveness of the clustering approach. Weak correlations among features indicate that no single factor dominates in influencing transaction behavior.

**GitHub Repository**

The complete code and dataset are available for review on GitHub to ensure transparency and reproducibility: