Customer Segmentation for Snapdeal Using K-Means Clustering

Data-Driven Insights for Personalized Marketing

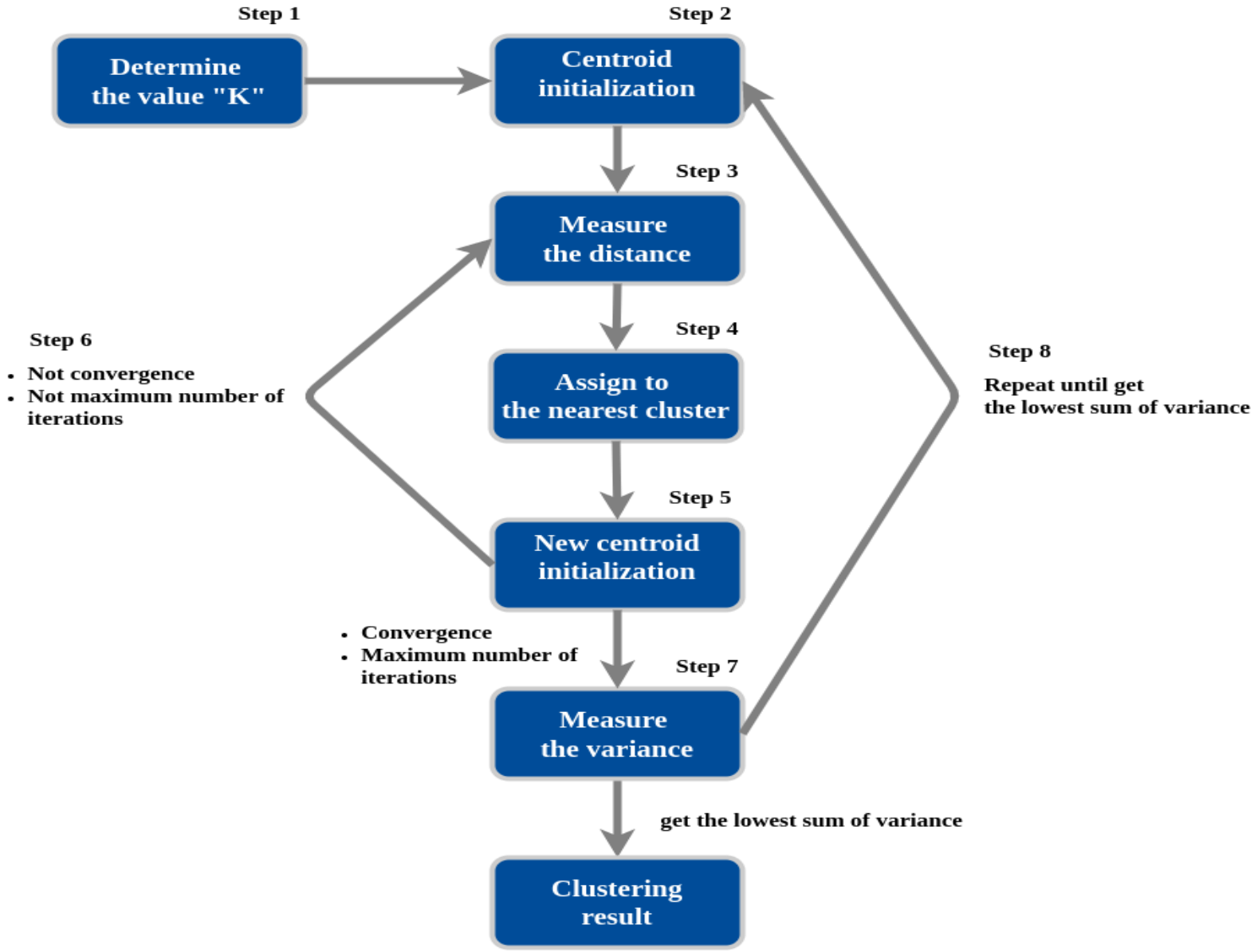
Author: Ravinder | April 2025

Objective

- Segment customers based on purchasing behavior using K-Means clustering.

- Enable personalized marketing, improve retention, and optimize inventory.

Steps of K-Means



Data Overview

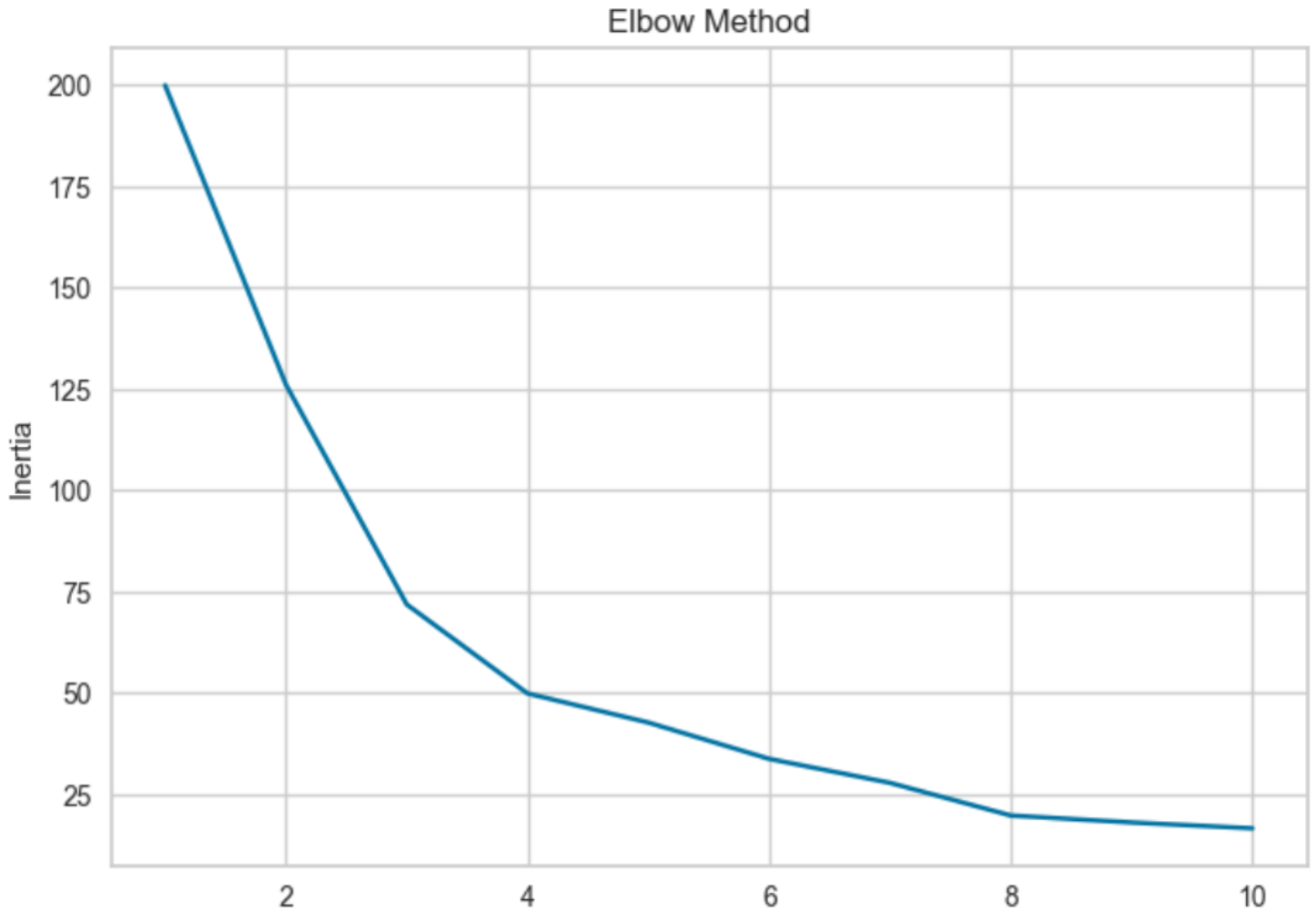
- • Source: Online retail data from Snapdeal
- • Goal: Segmenting Customer by using k-means clustering
- • Key features: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country
- • Cleaned for missing values and returns.

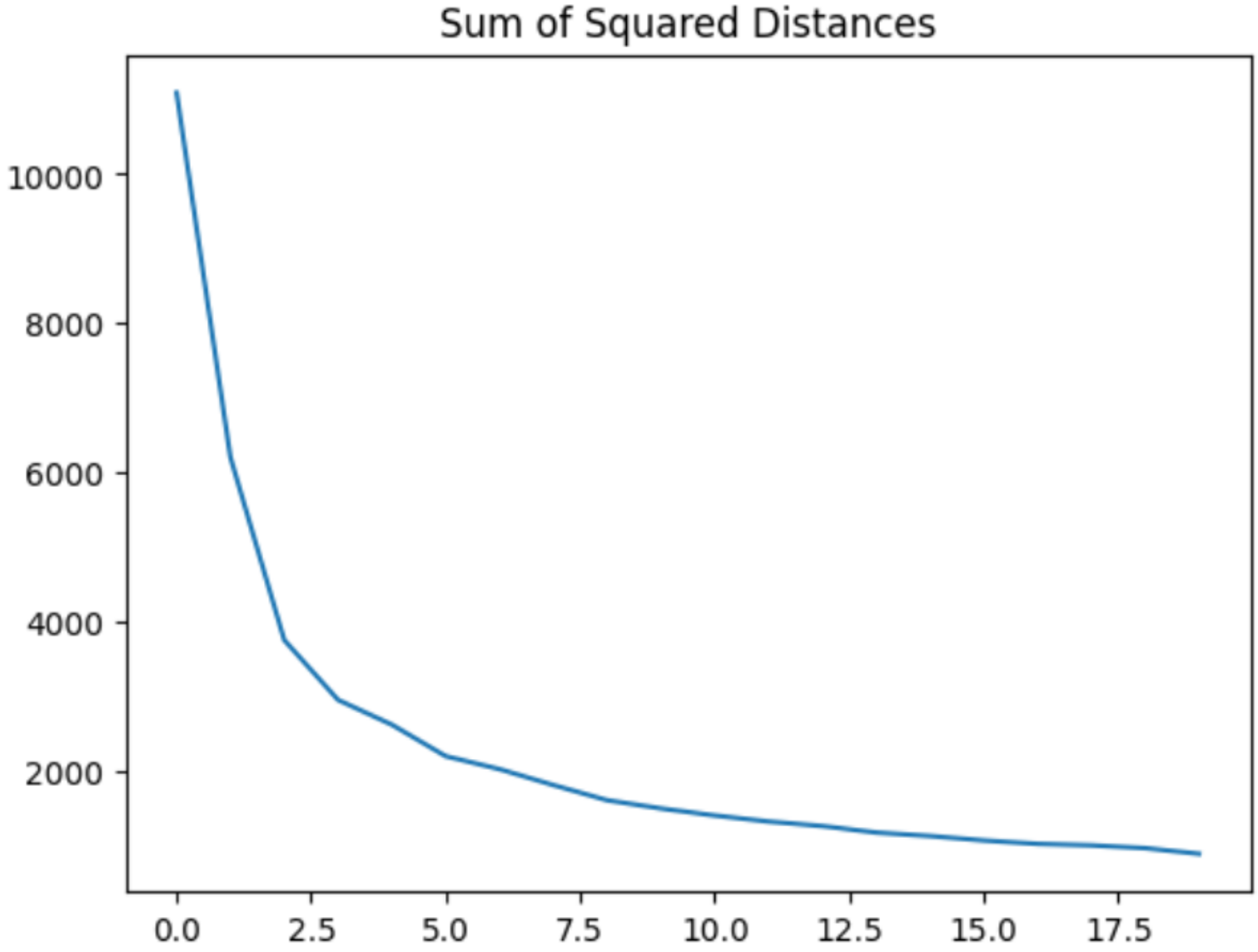
Data Preprocessing

- - Removed missing and negative values
- - Engineered features:
- - Frequency (number of purchases)
- - Monetary Value (total spending)
- - Recency (if available)

Elbow Method

- Elbow method used to identify optimal clusters

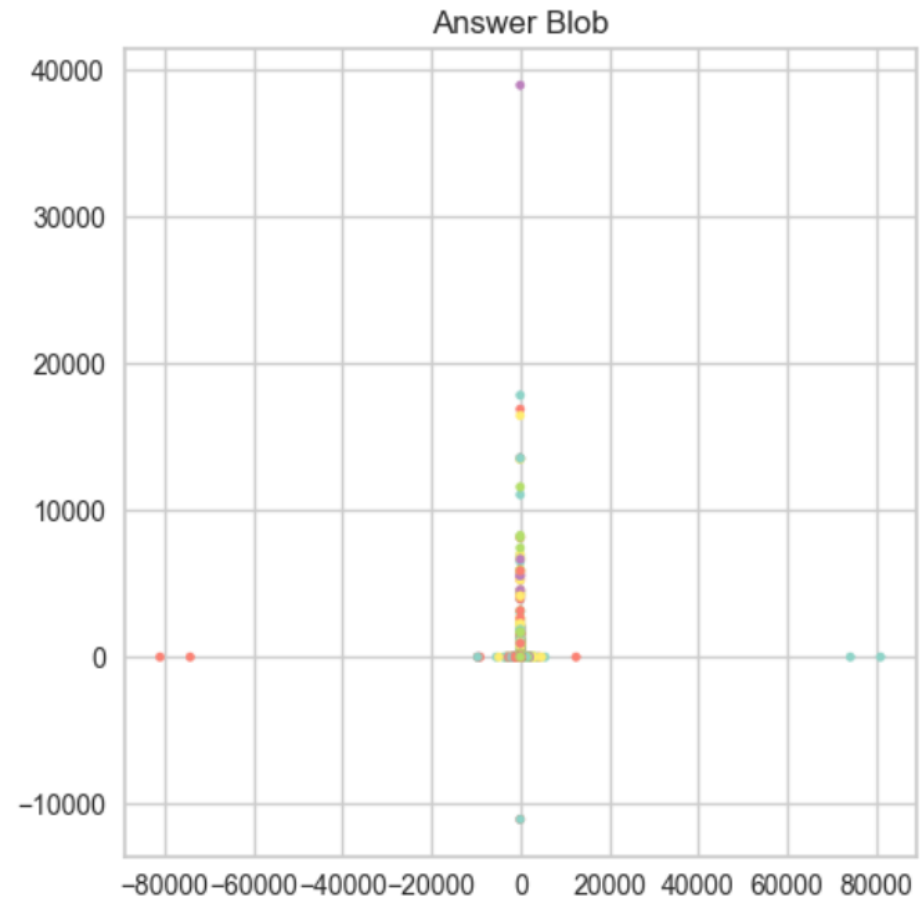


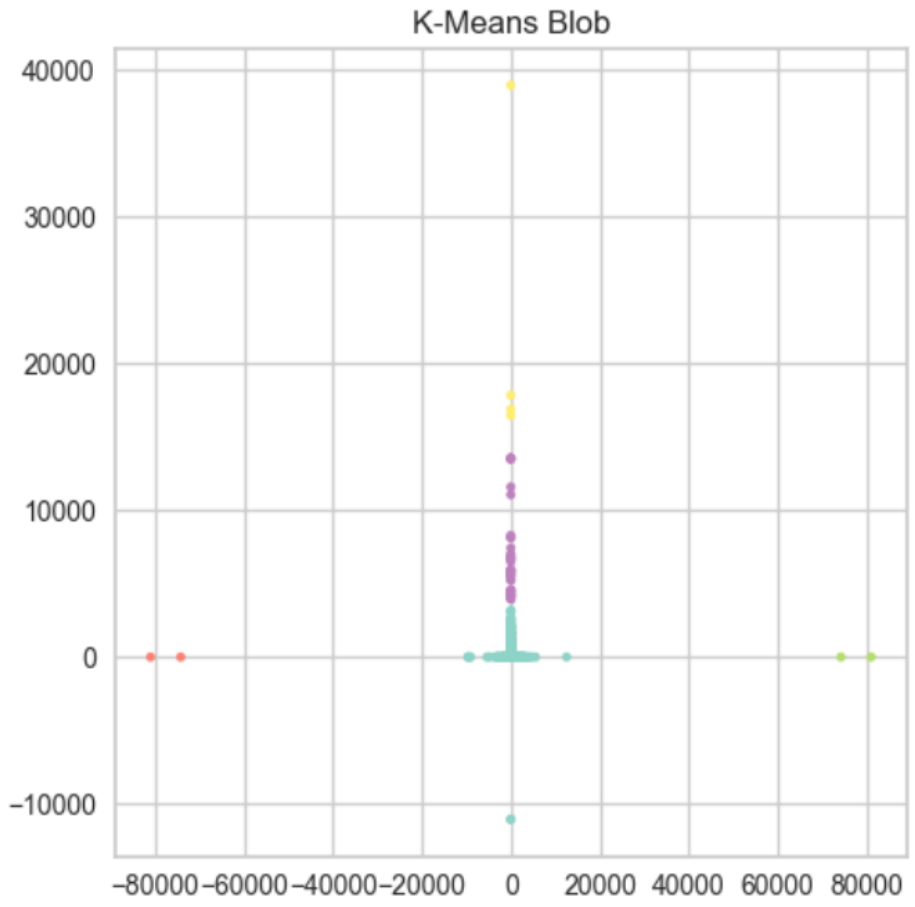

Sum of Squared Distances
10000
8000
6000
4000
2000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

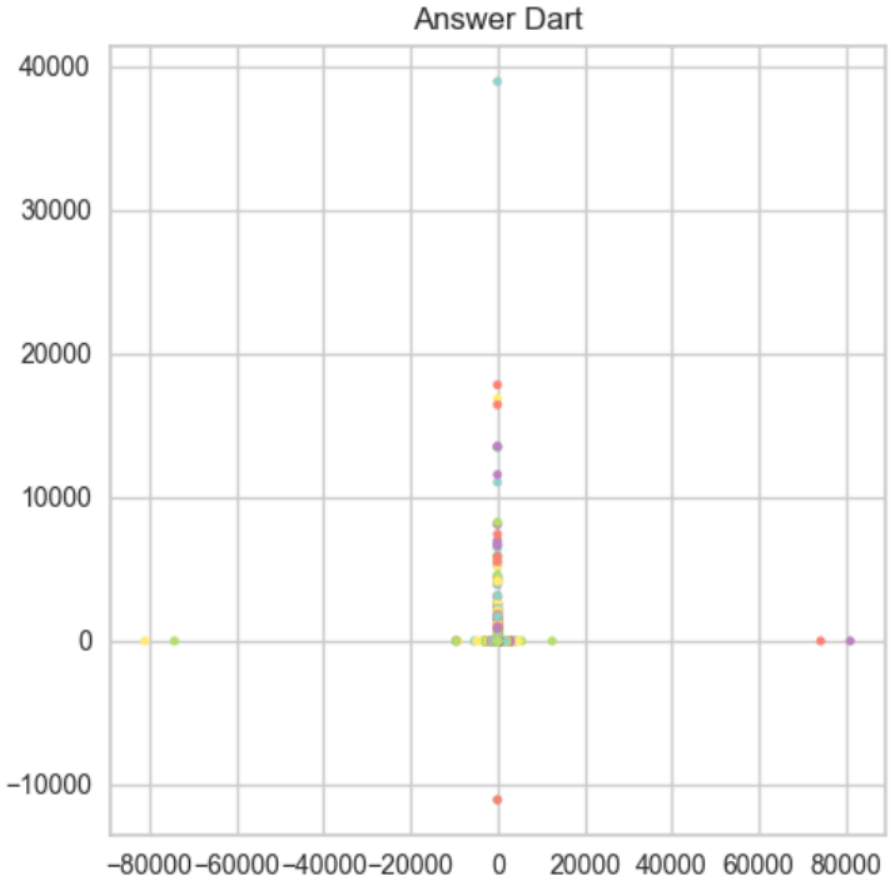
Cluster Visualization

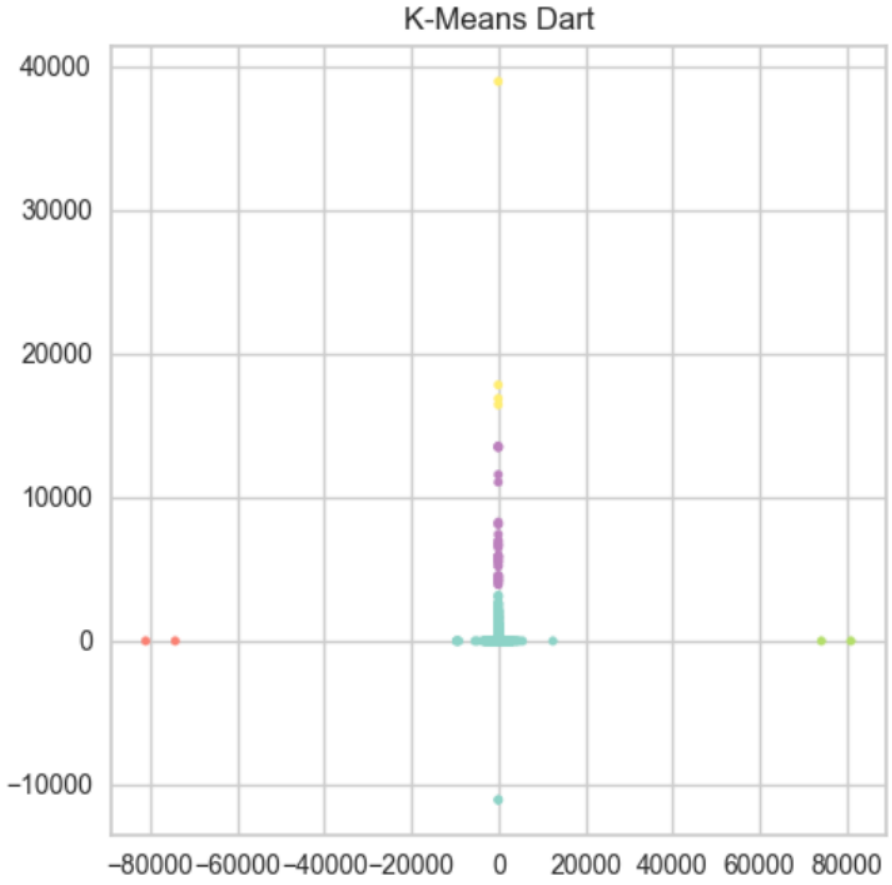
- 2D Scatter plot displaying clusters using frequency and monetary value

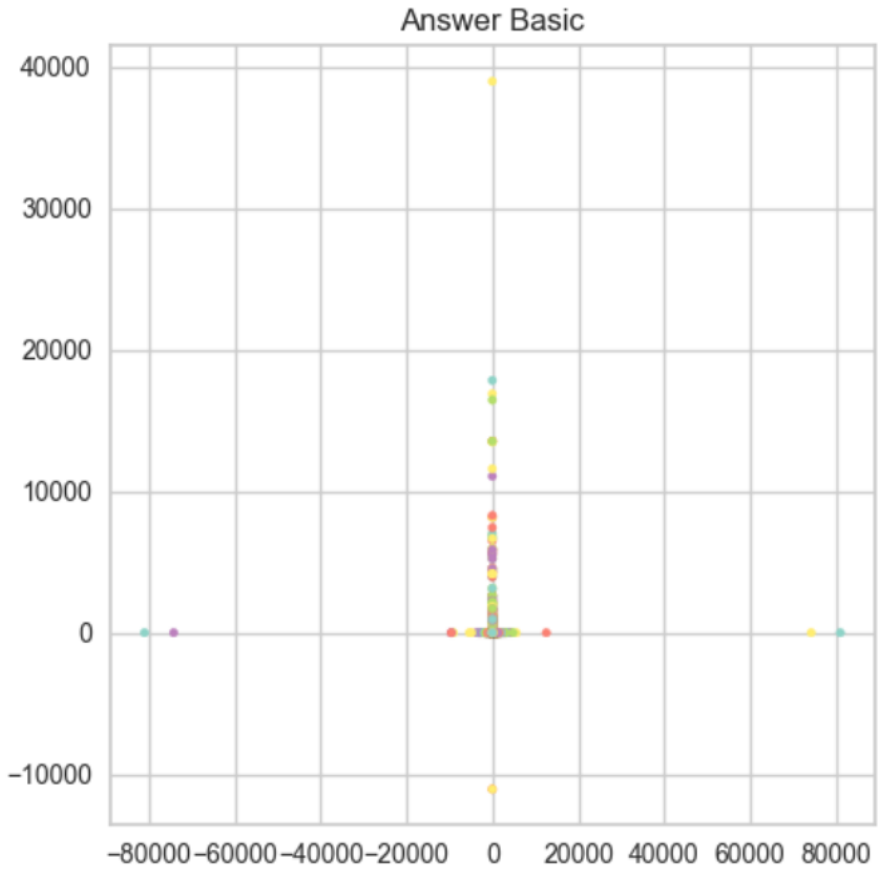
Answer
VS
K-Means

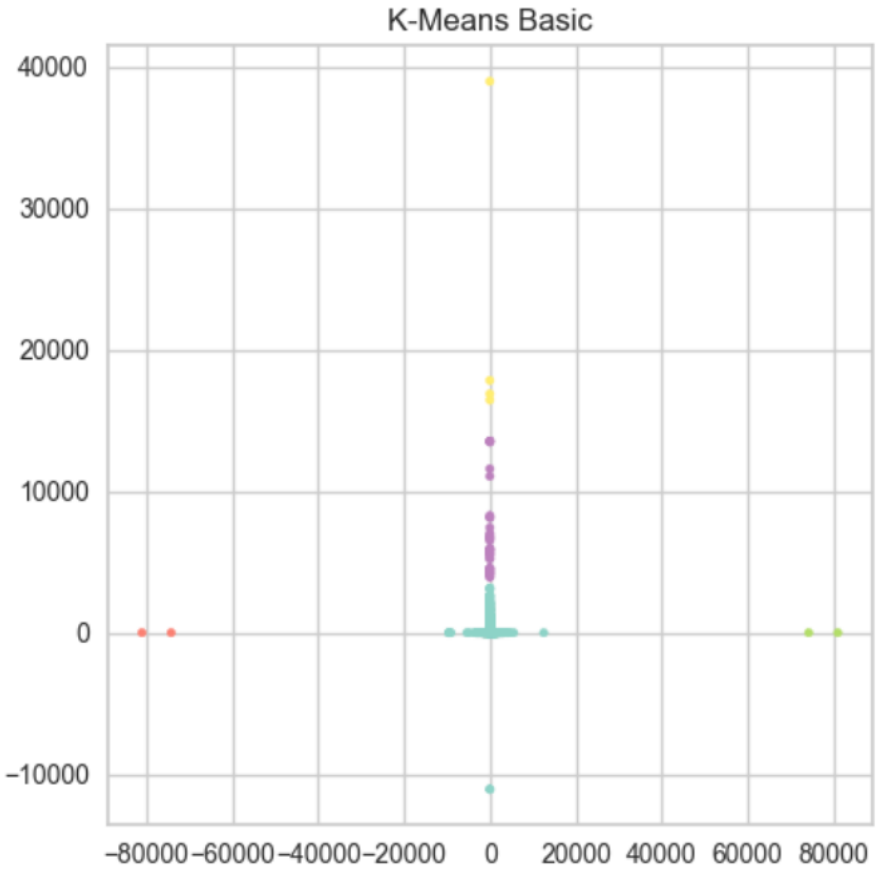

Answer Blob
40000
30000
20000
10000
0
−10000
−80000 −60000 −40000 −20000 0 20000 40000 60000 80000

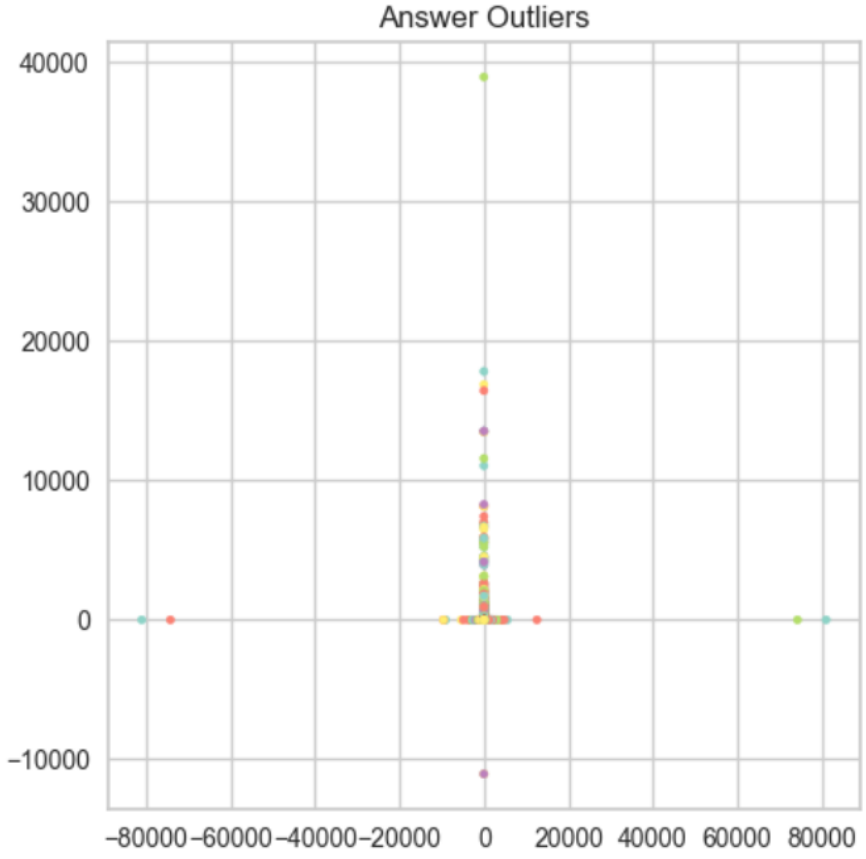

K-Means Blob
40000
30000
20000
10000
0
−10000
−80000 −60000 −40000 −20000 0 20000 40000 60000 80000

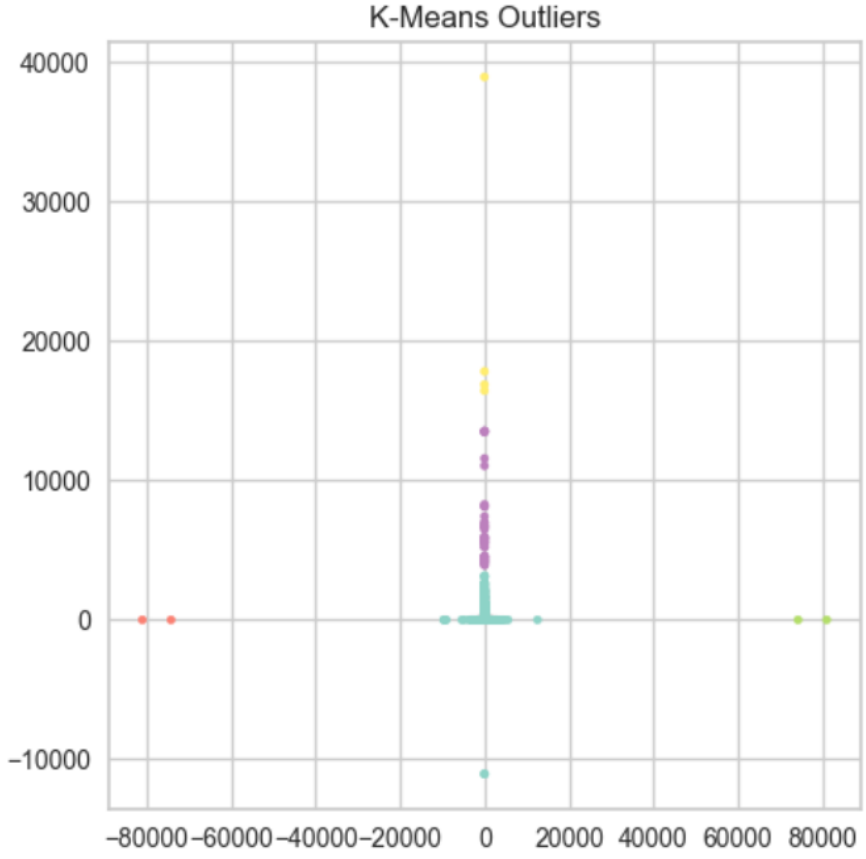

Answer Dart
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000

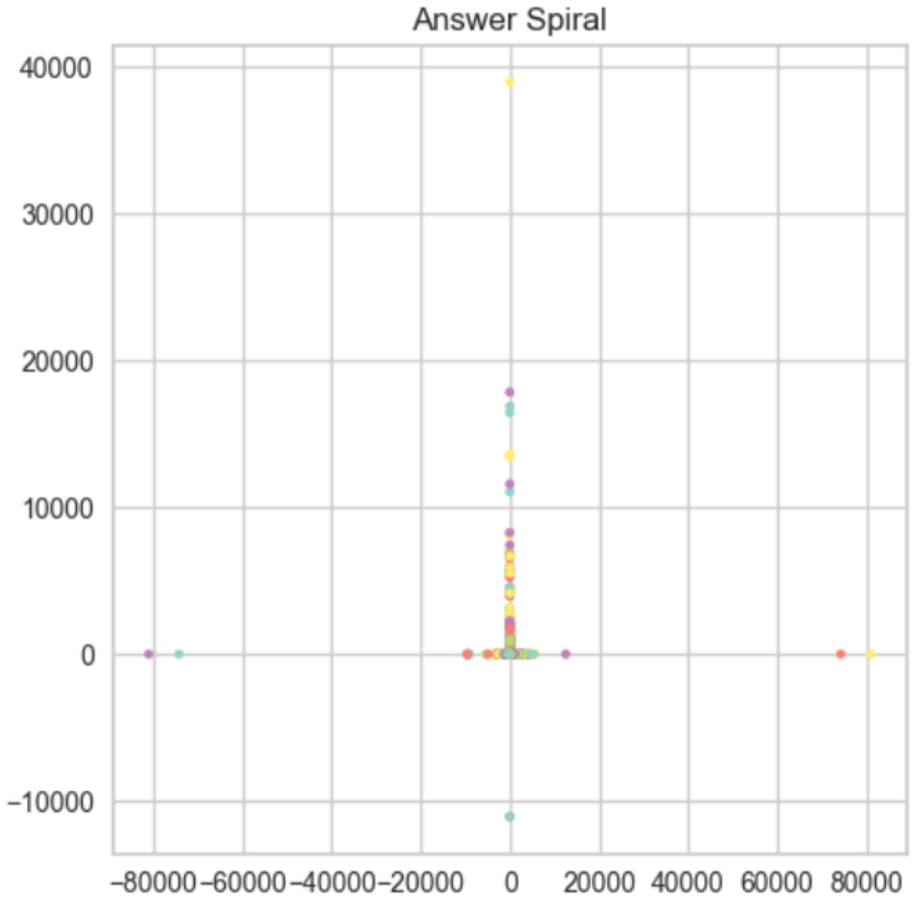

K-Means Dart
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000

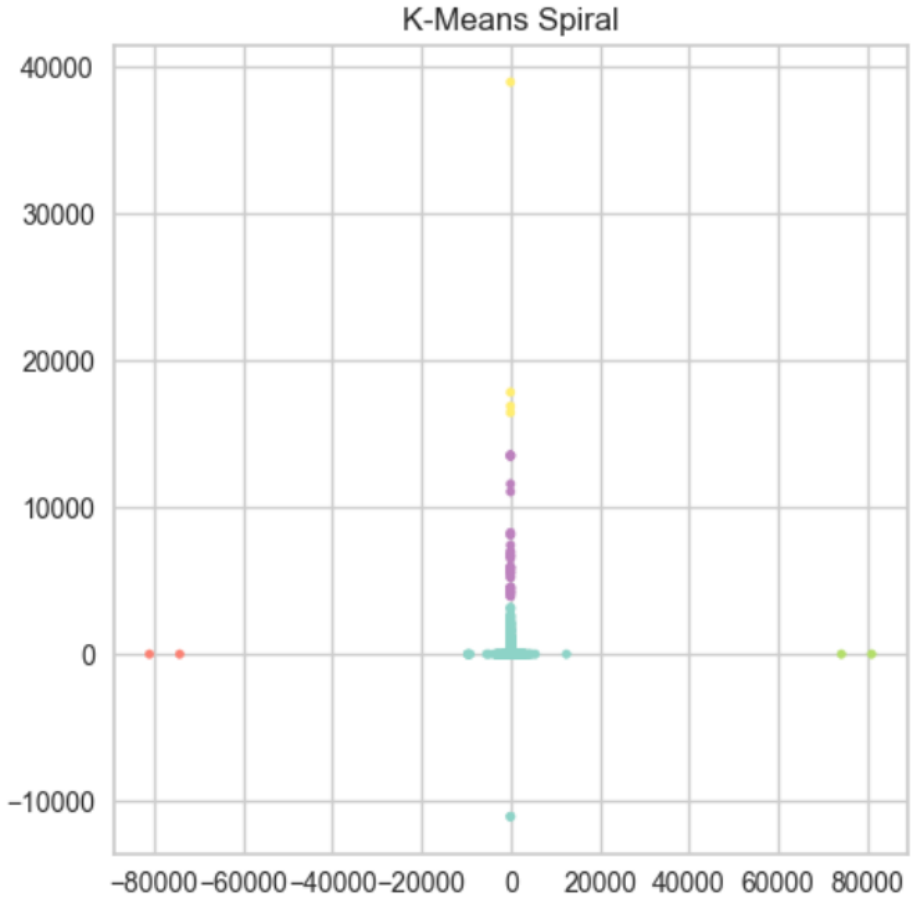

Answer Basic
40000
30000
20000
10000
0
−10000
−80000 −60000 −40000 −20000 0 20000 40000 60000 80000

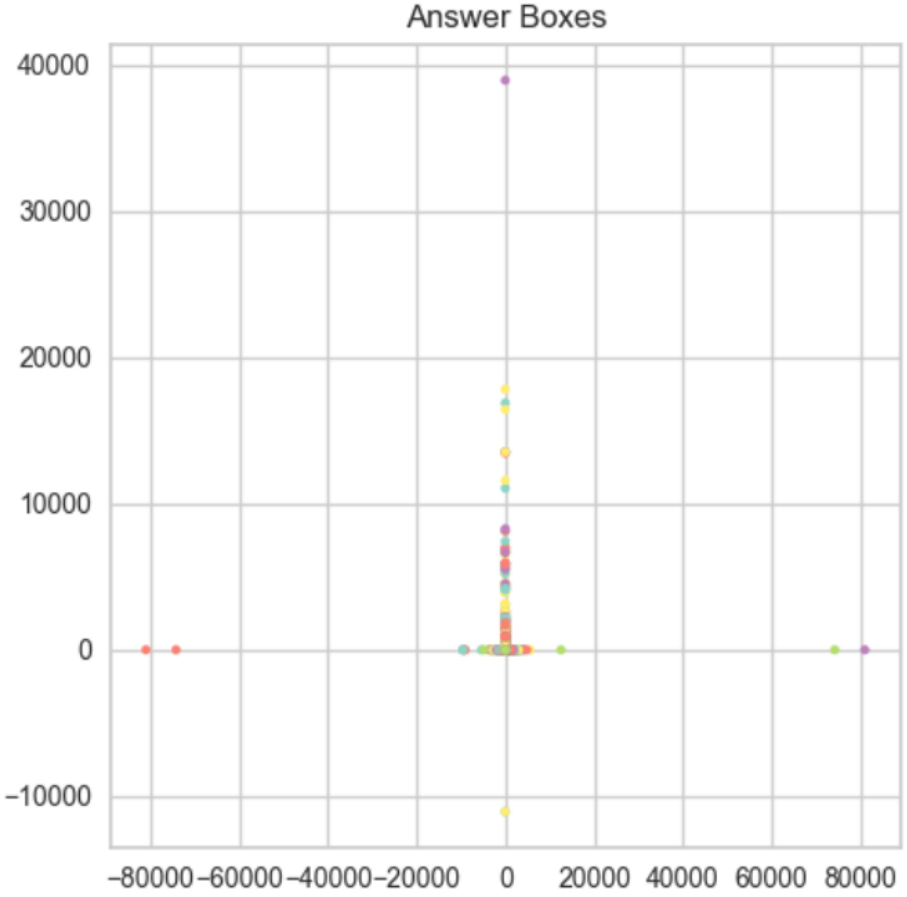

K-Means Basic
40000
30000
20000
10000
0
−10000
−80000 −60000 −40000 −20000 0 20000 40000 60000 80000

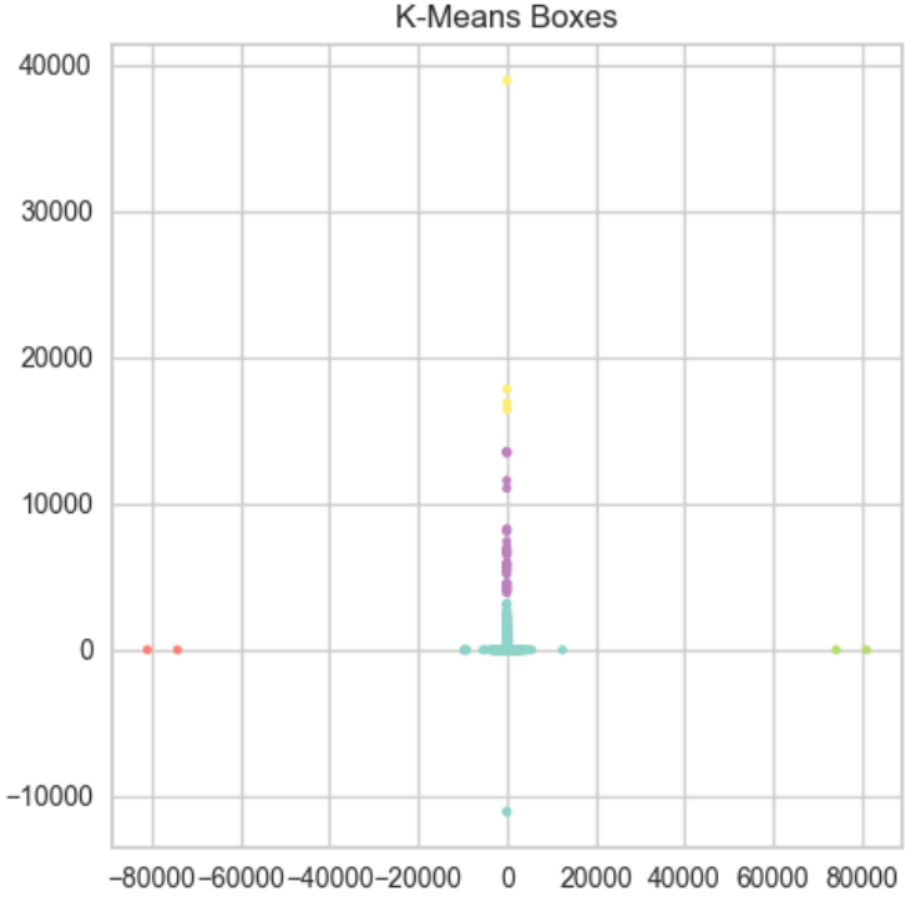

Answer Outliers
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000


K-Means Outliers
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000


Answer Spiral
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000


K-Means Spiral
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000


Answer Boxes
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000

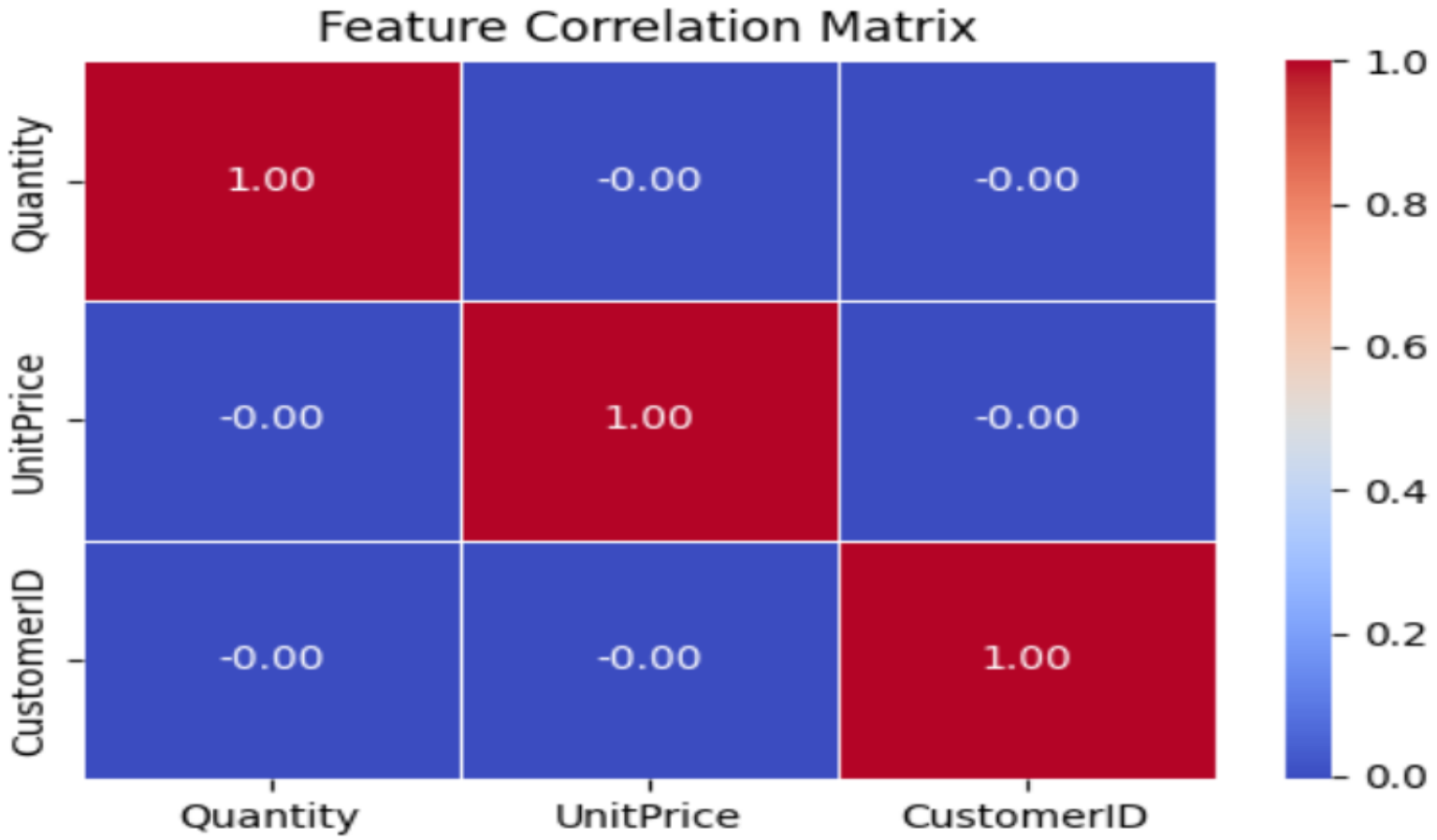

K-Means Boxes
40000
30000
20000
10000
0
−10000
−80000
−60000
−40000
−20000
0
20000
40000
60000
80000

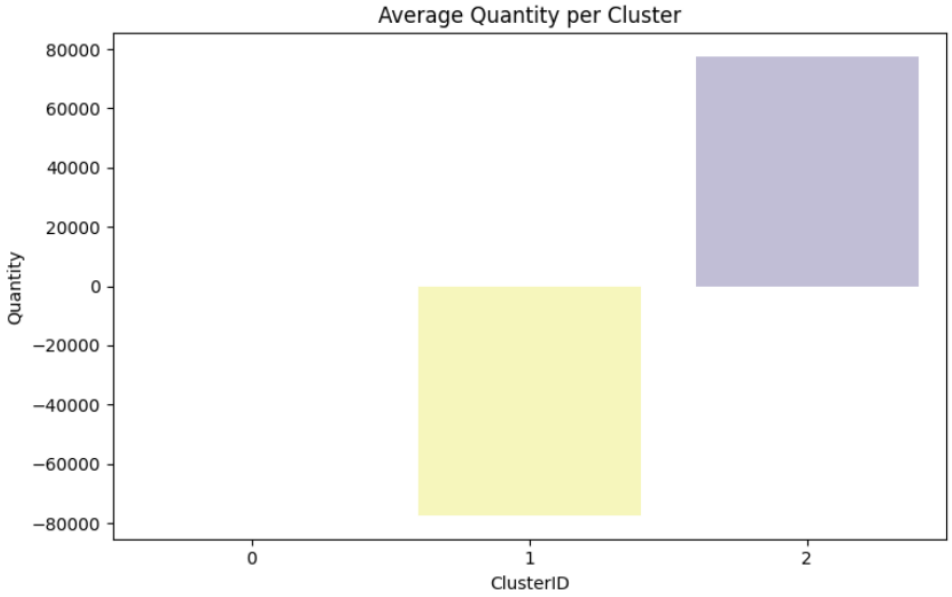
Cluster Profiles

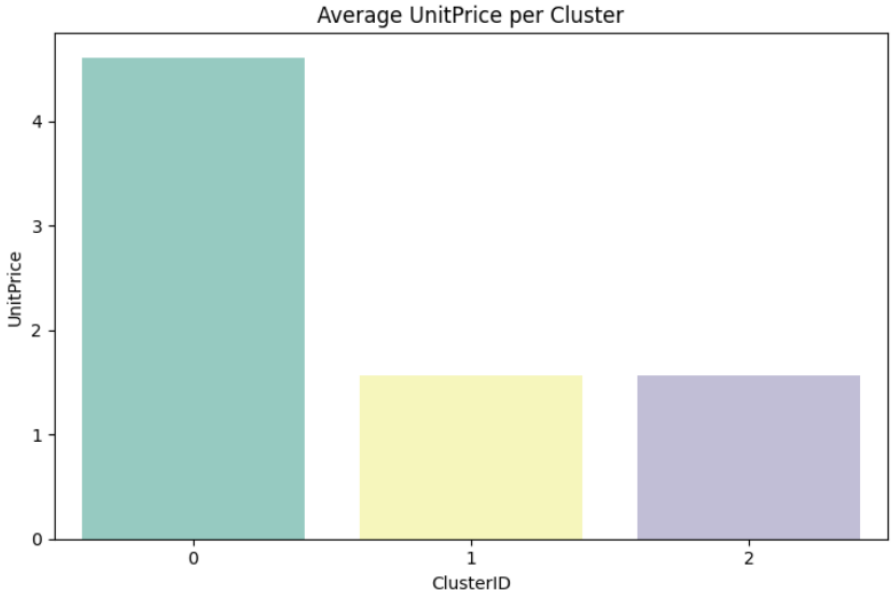
Cluster	Description	Frequency	Spending	Engagement
0	High-Value Customers	High	High	Strong
1	Occasional Shoppers	Medium	Medium	Moderate
2	Price-Sensitive Customers	Medium	Low	Promo-driven
3	One-Time Buyers	Low	Low	Weak
4	Dormant Customers	Previously high	Low	Inactive

Spending Trends by Segment

- Visual representation of spending across clusters



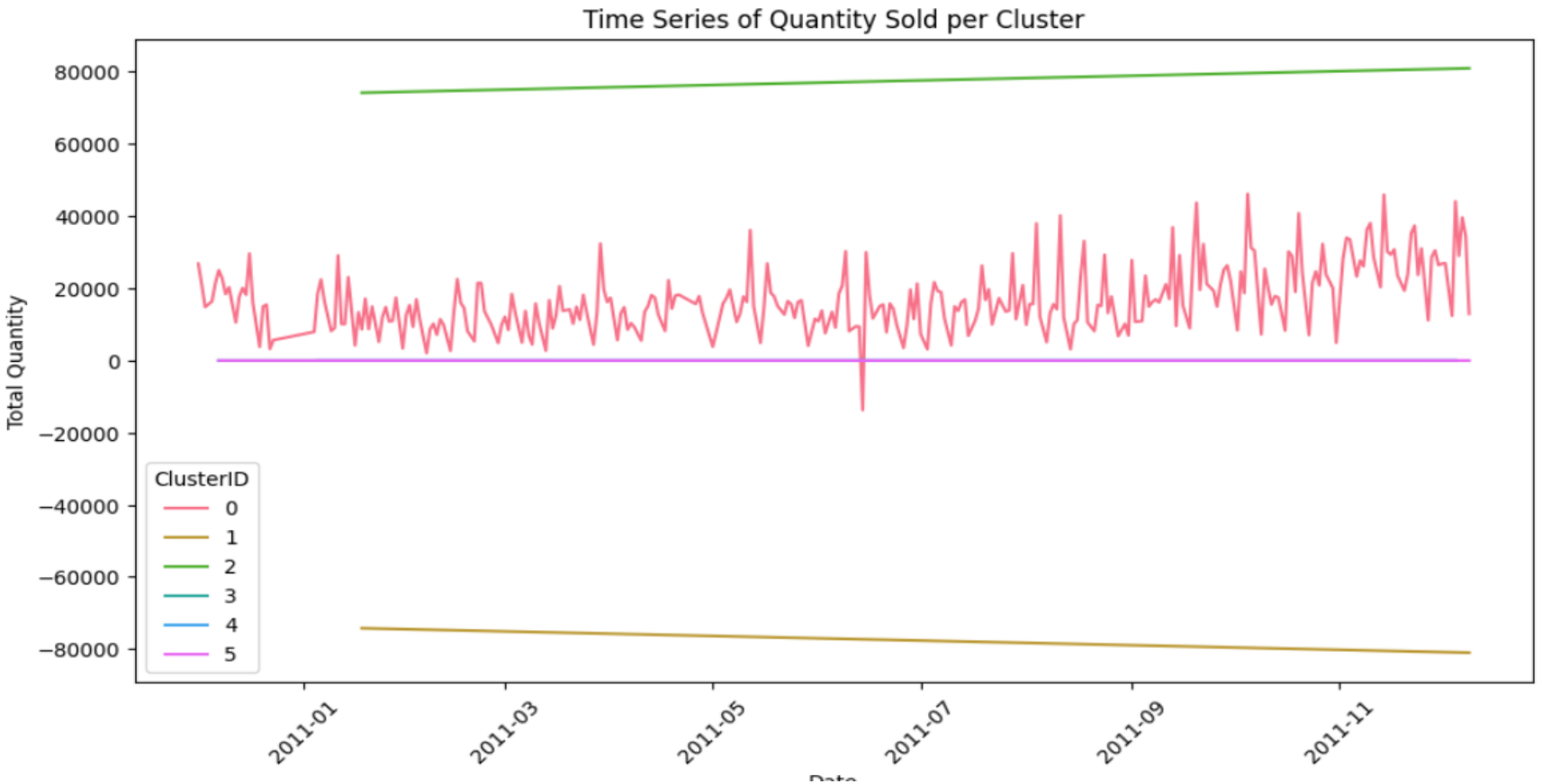




It visualizes the clustering results in a 2x3 grid of scatter plots, each representing one of the clusters. Each plot shows how the Quantity and Unit Price values are distributed within the respective clusters, with distinct colors representing different clusters. If any cluster has no data, it will be marked as "No Data".

Engagement Patterns

- Visualizing purchase frequency and recency to distinguish active vs. dormant users



Final Recommendations

- Personalized marketing campaigns per segment
- Loyalty and retention programs for high-value customers
- Targeted discounts for price-sensitive users
- Re-engagement campaigns for dormant users
- Optimized inventory based on segment behavior
- Regular model updates to track changing behaviors

Final Report

- • Insights from clustering highlight 5 key customer segments
- • Use behavior data to drive engagement and sales
- • Support strategic decisions with data-backed profiles

Conclusion

- K-Means clustering provided clear segmentation based on purchasing behavior.
- Enables targeted strategies for marketing, retention, and revenue growth.
- Continuous monitoring recommended to adapt to market shifts.