# SATESOFT DATASET ANALYSIS

BY: RILEY GOERG, JORGE BERNAL, BEN GOLDBERG, RYAN ALQASEMI, OWEN JOHNSON

OVERVIEW PROBLEM

- Africa's informal economy contributes significantly to GDP but operates with little digital record keeping.

- MSMEs struggle with things like limited visibility into sales trends, inventory levels, and customer behavior. The typical manual processes lead to stockouts, deal stock, or insufficient purchasing.

- Impact: Businesses miss revenue opportunities and suppliers cannot forecast demand.

- Our Team's Mission: Analyze Duqact data to identify the trends in sales, product performance, and shop behaviors, and recommend data driven improvements.
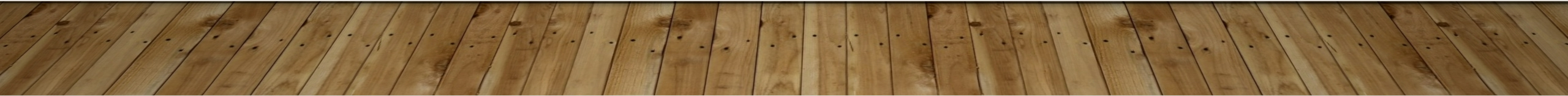
OUR APPROACH

1) Data Preparation and Transformation
   o Cleaned and merged sales-related tables, formed a customer-level dataset, and converted timestamps and removed missing values as needed.

2) RFM Analysis
   o Recency, frequency, monetary values and created RFM scores.

3) Visualization and Pattern Exploration
   o Plotted RFM, visualized clusters, and analyzed revenue contribution by cluster.

FREQUENCY DISTRIBUTION

- Frequency represents how many times each customer has made a purchase.

- Most customers have very low purchase frequency (1–5 purchases), causing a strong right skew.

- A smaller number of customers have moderate frequency (10–20 purchases).

- A few outlier customers make 30–50+ purchases, indicating a very loyal, highly engaged segment.
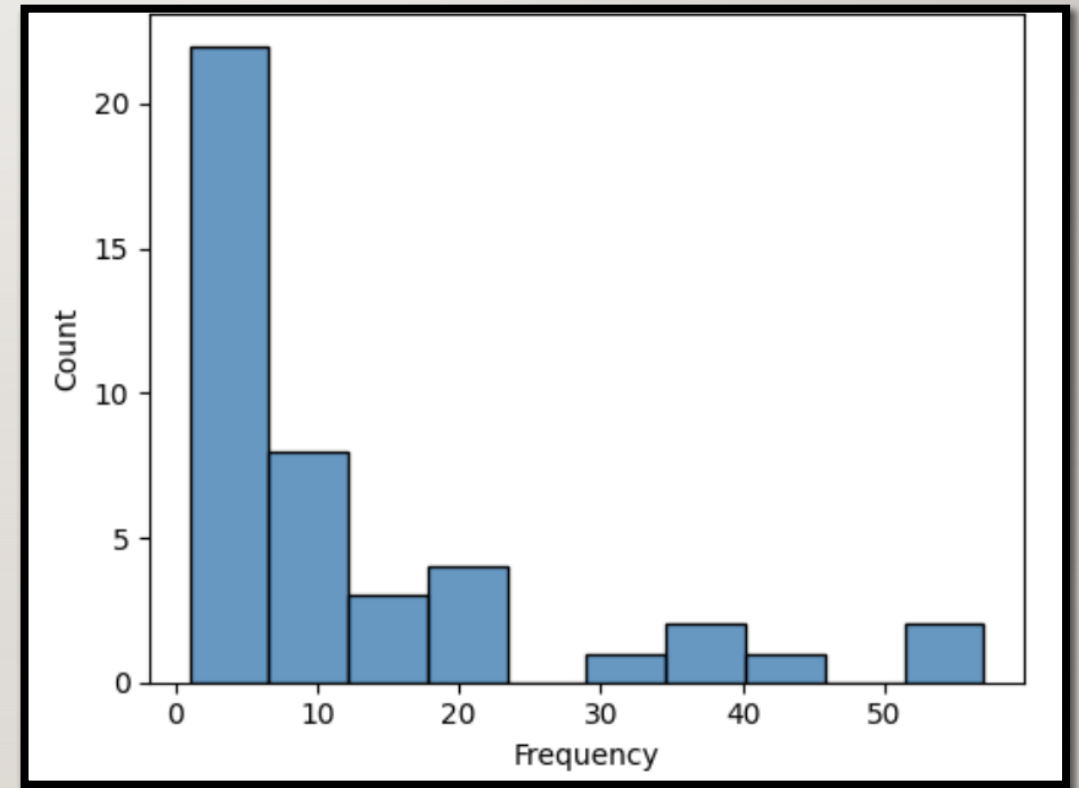
RECENCY DISTRIBUTION

- Recency measures how many days it has been since a customer's last purchase.

- Many customers fall into the 150–200-day range, indicating a sizable group that hasn't purchased recently.

- The distribution suggests opportunities for re-engagement campaigns, especially for customers in higher-recency bins.
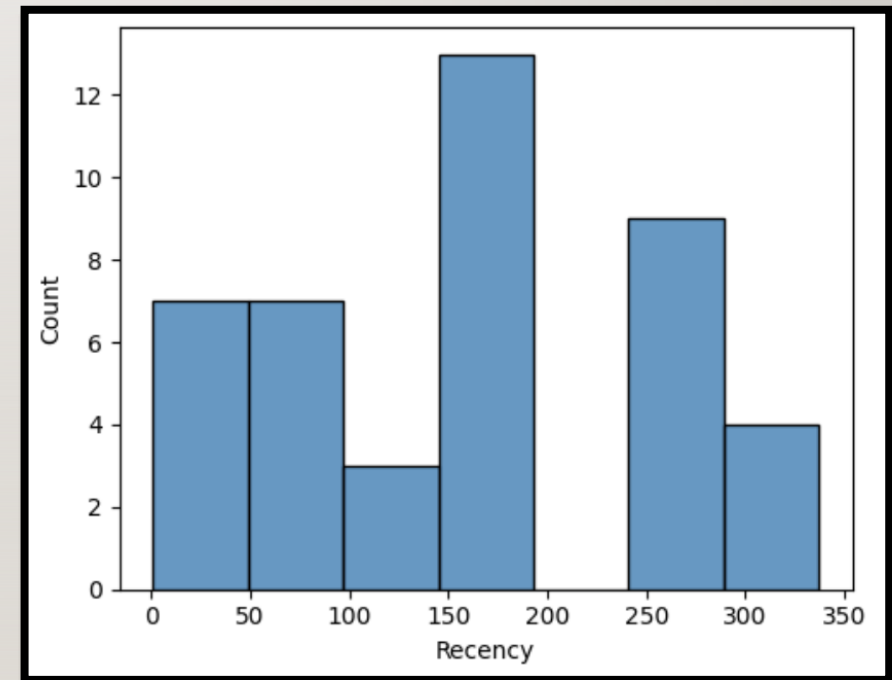
MONETARY DISTRIBUTION
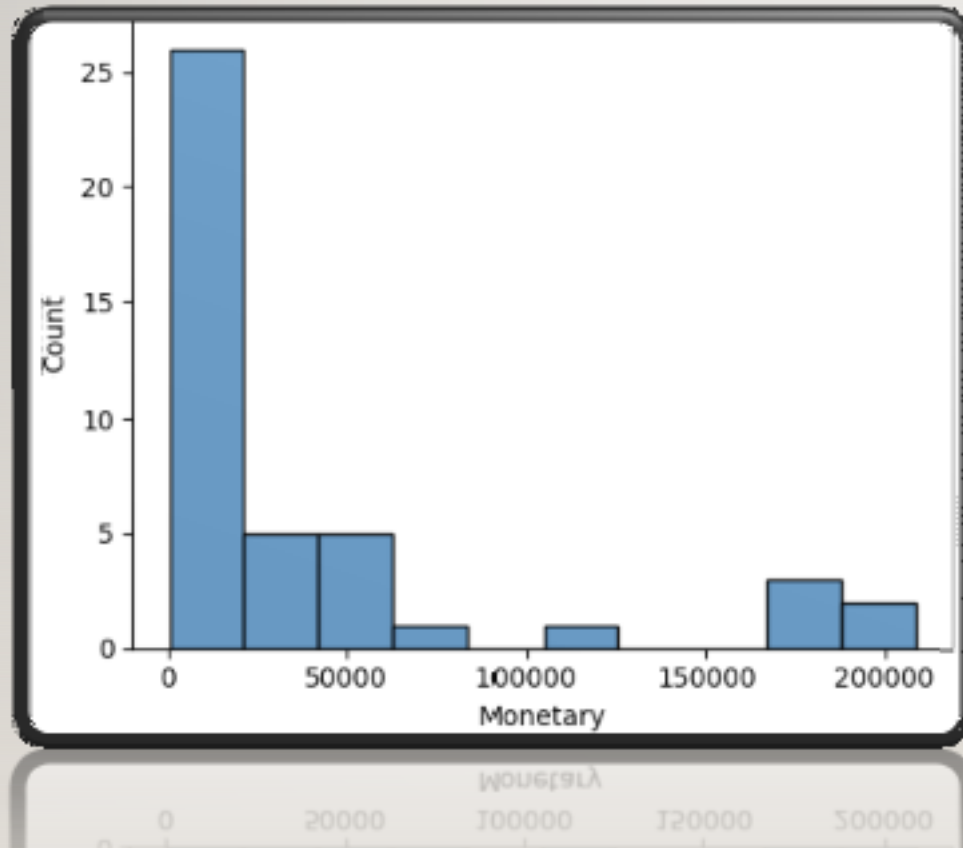
- Monetary value represents the total amount spent by each customer.

- Most customers fall in the low spending range creating a heavily right-skewed distribution.

- A small number of customers account for very high spending ($100k–$200k), indicating the presence of high-value outliers.

- The skewed distribution suggests there are many more small spenders compared to larger ones.

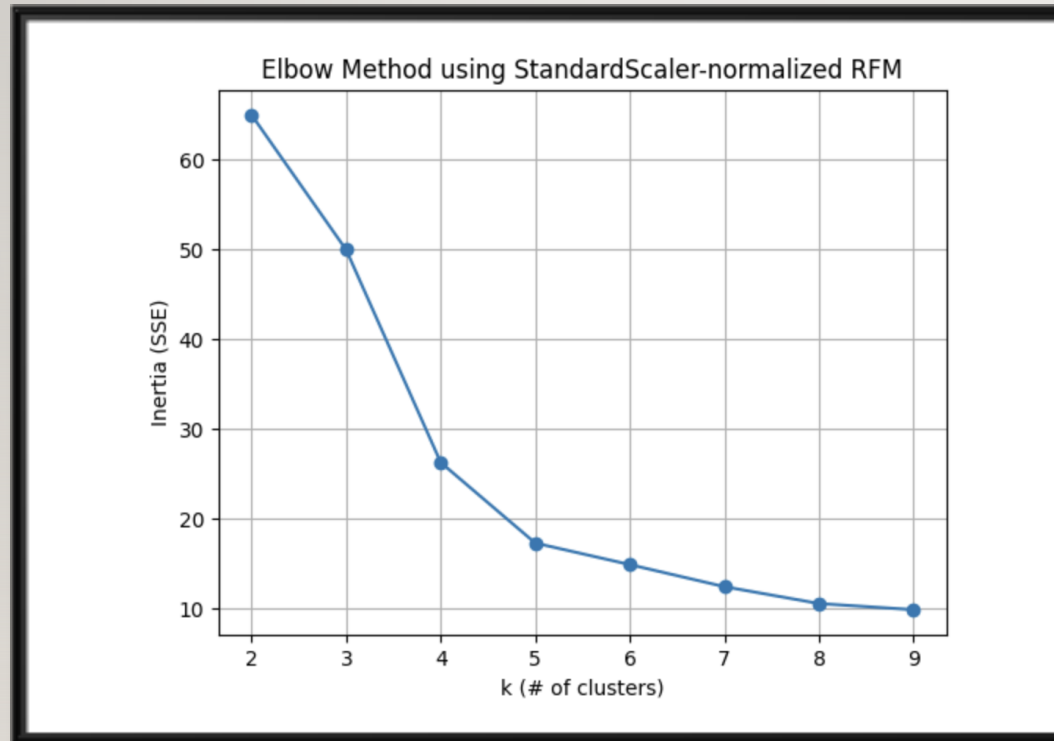Top Revenue-Generating Customers per Store
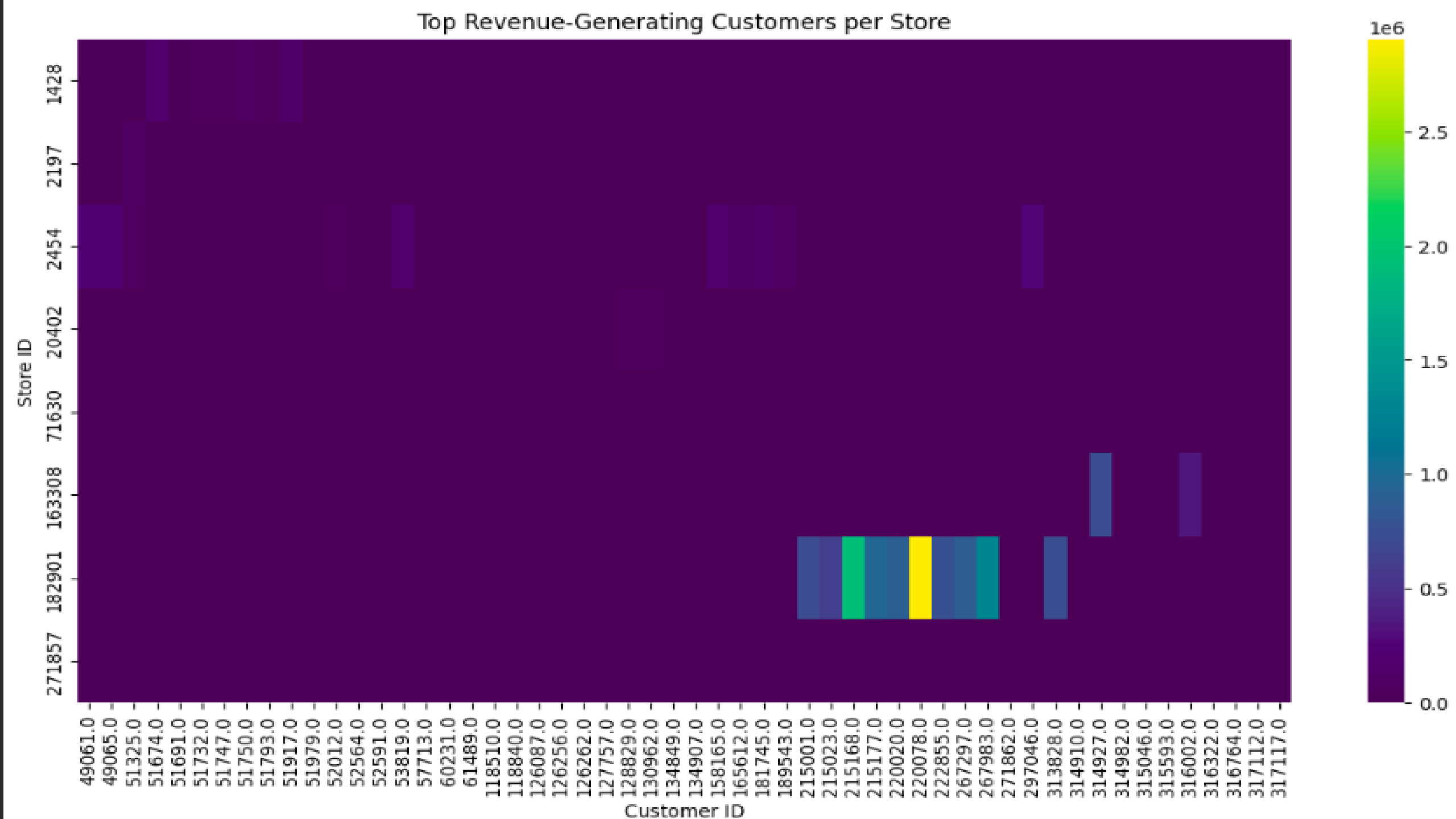
DISTRIBUTION OF CUSTOMER SPEND PER STORE

- The heatmap and box plot both display one store with significantly higher revenue compared to others.

- There are big spenders concentrated into one store.

ELBOW METHOD USING STANDARDSCALER
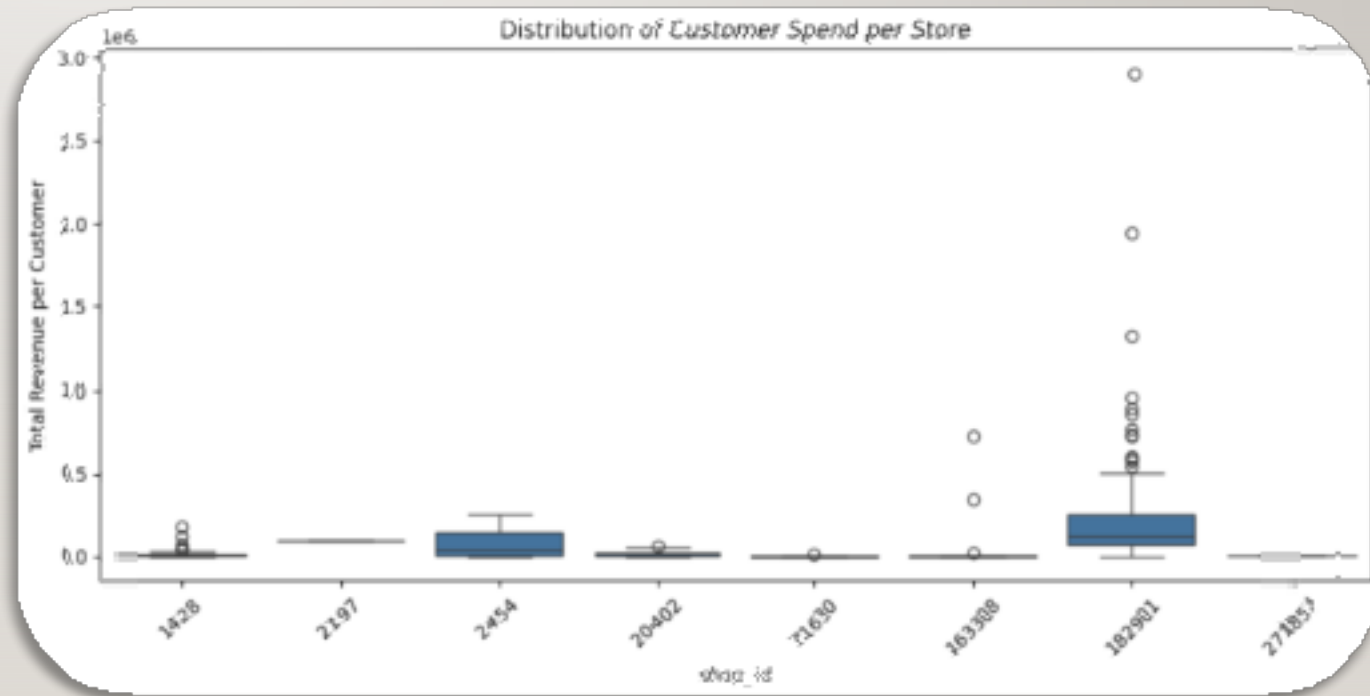
Elbow Method using StandardScaler-normalized RFM

- Elbow at k=5

- There are 5 different customer clusters

CUSTOMER CLUSTERS-FREQUENCY VS MONETARY AND RECENCY VS MONETARY

- Cluster 0: Low spend, semi-frequent, recent visitors

- Cluster 1: Medium-high spend, frequent, recent visitors

- Cluster 2: High spend, semi-frequent, non-often visitors

- Cluster 3: Low spend, non-frequent, non-often visitors

- Cluster 4: Low spend, non-frequent, semi-often visitors
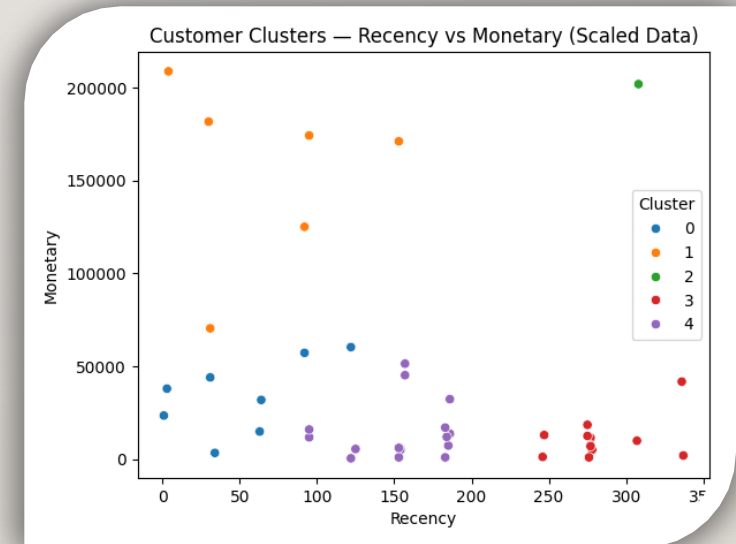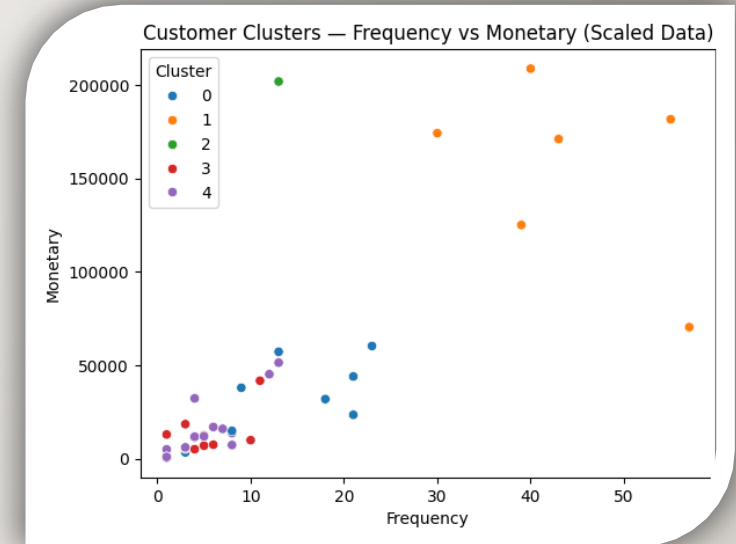
REVENUE BY CUSTOMER SEGMENT

- Segment 1 are the best customers. Segments 0, 2, and 4 follow in revenue by a large margin, with segment 3 trailing behind.

- Customized offers must be made to target each segment correctly.

Revenue Contribution by Customer Segment
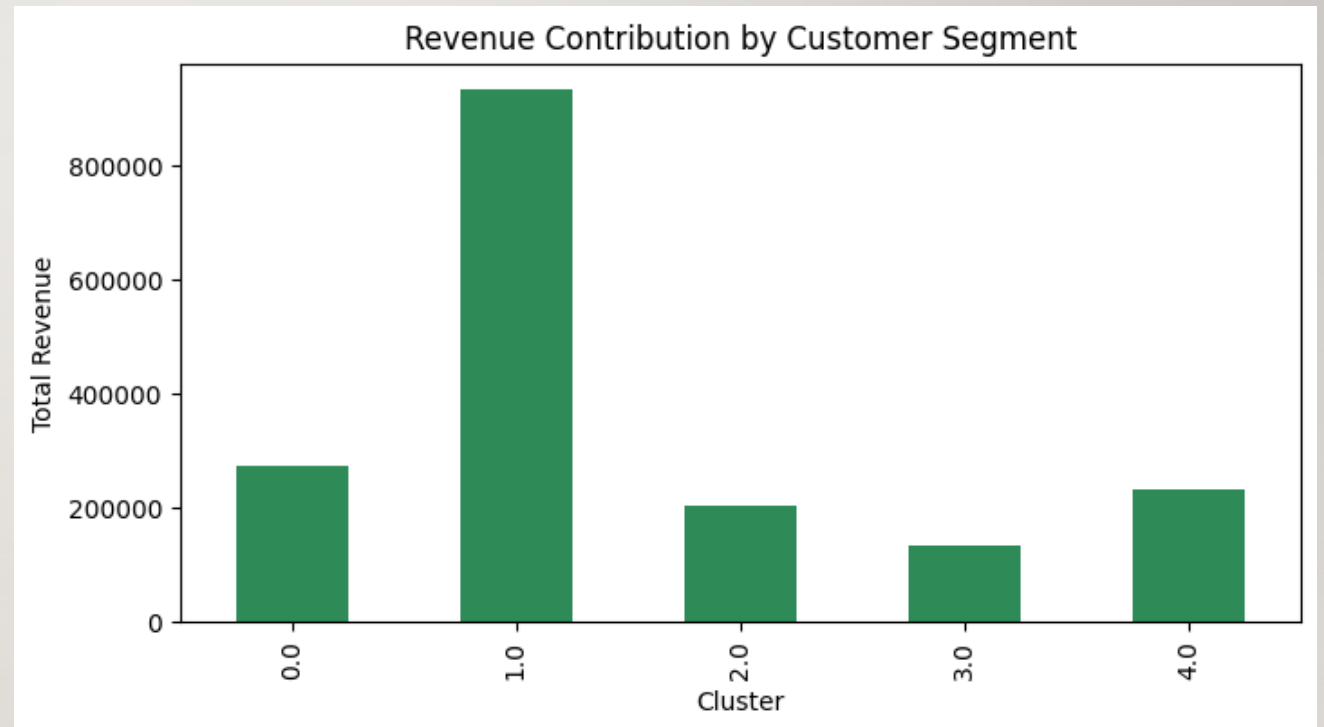
RECOMMENDATIONS BASED ON CLIENT CLUSTER

```
client_id   Cluster                       Recommendation
  65875.0         4       Encourage more frequent visits.
  51750.0         1 Provide VIP perks or bonus rewards.
  51928.0         0       Offer small spend incentives.
  49061.0         2       Send a re-engagement discount.
  49837.0         0       Offer small spend incentives.
 118840.0         4       Encourage more frequent visits.
  51917.0         1 Provide VIP perks or bonus rewards.
 128270.0         3            Send light reminders only.
  52564.0         4       Encourage more frequent visits.
 128829.0         0       Offer small spend incentives.
```
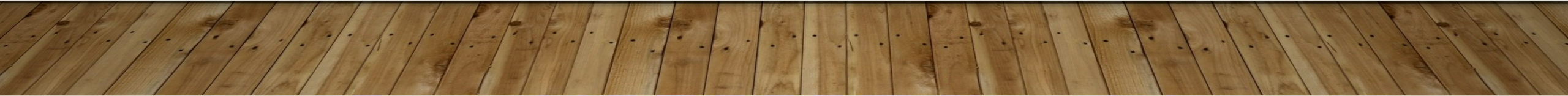
RECOMMENDATIONS BASED ON FREQUENCY AND RECENCY DISTRIBUTION

- Send personalized emails to customers in the 250+ days since purchase range

- Send feedback request emails to customers in the 250+ days since purchase range

- Gather customer acquisition data showing each method used to acquire the customers

- Analyze marketing strategy used to acquire frequent customers and apply those tactics to attract more high engaging customers.
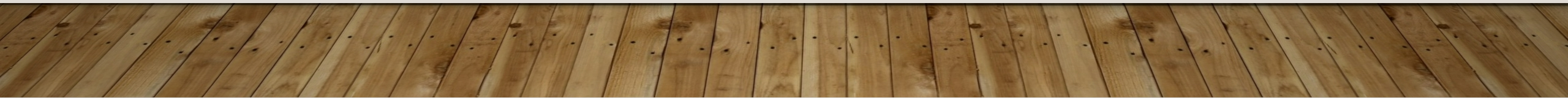
IMPACT OF OUR ANALYSIS

- Improved Customer Understanding: Segments reveal which customers are loyal, which are at risk, and which have high spending potential.

- Better Revenue Targeting: Cluster 1 customers generate the majority of revenue, which could help businesses to use marketing spend more efficiently.

- Actionable Retention Strategies: Recency analysis shows many customers have not purchased in 150-200+ days, revealing a large re-engagement opportunity.

- Resource Optimization: Stores with concentrated high spenders can be prioritized for promotions or inventory allocation.

Q&A