Online Retail II

Capstone 2

By Vanessa Garretson

Problem Statement

Identify three to seven customer segments from the transaction data to focus ongoing marketing efforts while answering the below Key Questions about segment characteristics.

Key Questions:

- How often do each segment make purchases?
- How much have they spent over time?
- How recently have they made a purchase?
- What does their basket look like?
- Are there differences in seasonality of purchases between segments?

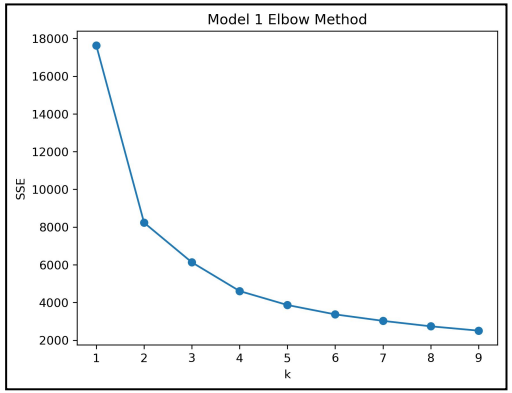
Data Preparation and Model Selection

Data Prep

- Remove cancelled orders
- Remove negative 'Quantity' and 'Price'
- Split known and unknown Customers
- Calculate RFM values
- Normalize and Standardize

Model Selection

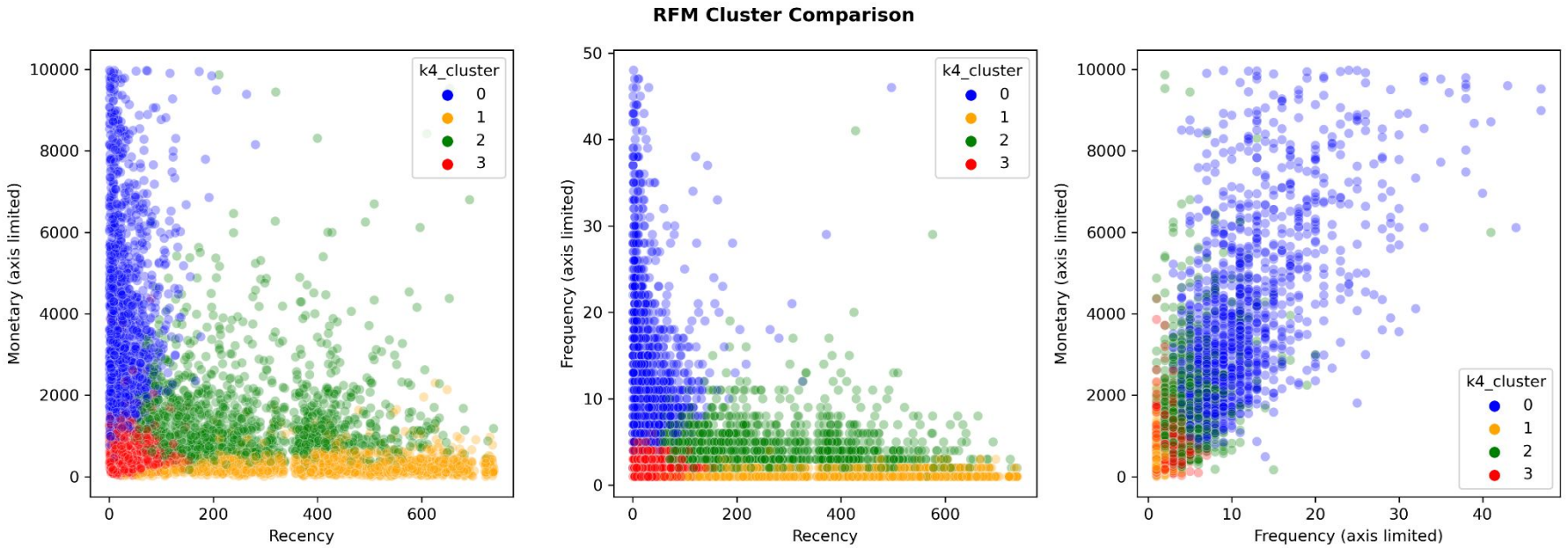
- K-means
  - k=2 vs. k=4




```
Model 1 Silhouette Scores

k=2, Silhouette Score: 0.4378
k=3, Silhouette Score: 0.3436
k=4, Silhouette Score: 0.3820
k=5, Silhouette Score: 0.3448
k=6, Silhouette Score: 0.3341
k=7, Silhouette Score: 0.3166
```

RFM Cluster Comparison



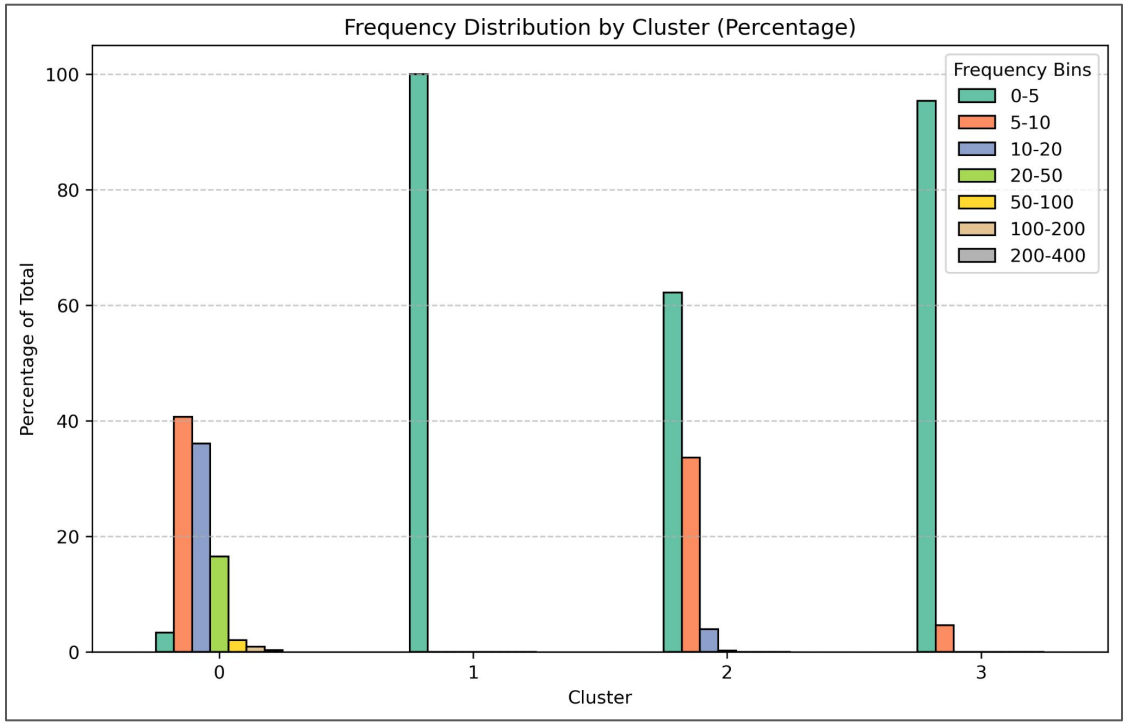
How often do each segment make purchases?

Cluster 0
- 7 to 17 purchases (most skewed)
- 5+ purchases: 96.6% ; 10+: 55.9%

Cluster 1
- 1 purchase: 77.8% of customers
- 2 or less: 98.9% of customers

Cluster 2
- 3 to 5 purchases
- 5+ purchases: 37.8% ; 10+: 2.3%

Cluster 3
- 1 to 3 purchases
- 5+ purchases: 4.6% ; max 7



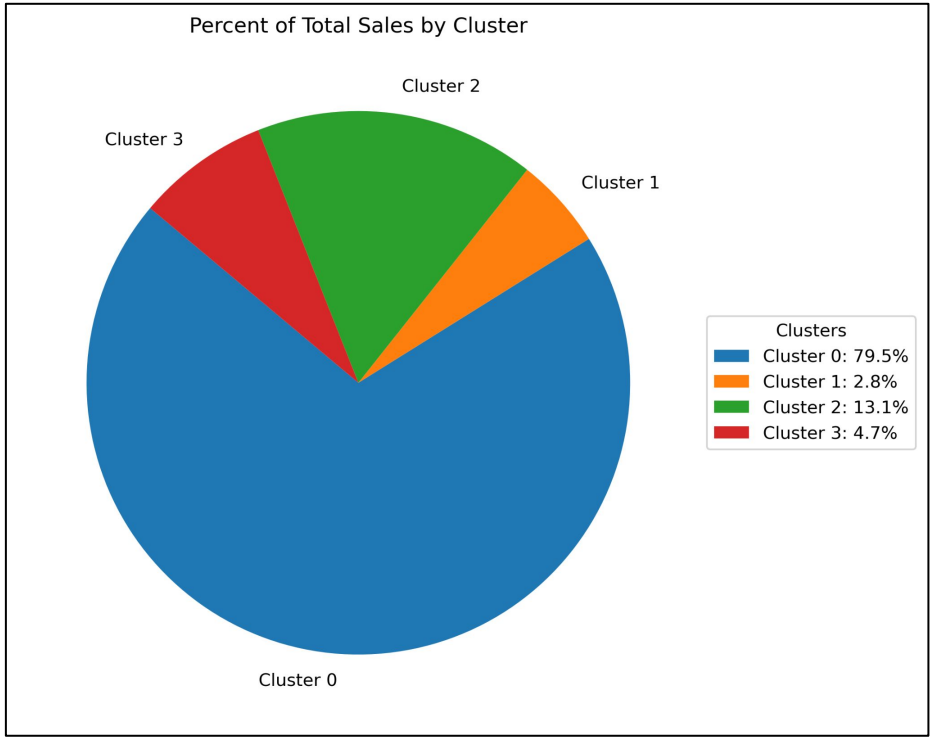
How much have they spent over time?

Cluster 0
- £2,268 to £6,948
- £1k+: 98.4% of customers
- £10k+:15.2%

Cluster 1
- £142 to £374
- £1k+: 2.1% of customers

Cluster 2
- £783 to £1,844
- £1k+: 60.1% of customers;
- £10k+: 0.9%

Cluster 3
- £342 and £887
- £1k+: 18.8% of customers



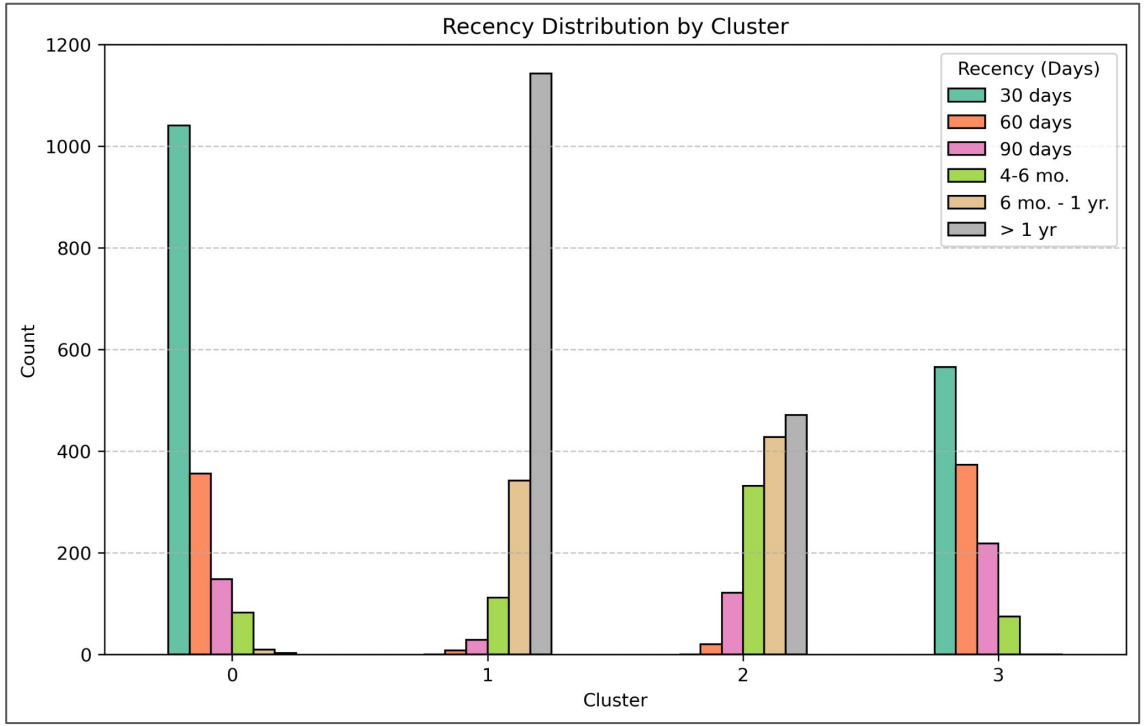
How recently have they made a purchase?

Cluster 0
- 8 to 42 days
- 90 days: 94.3%; 30 days: 64.7%
- 6 months: 99.2%

Cluster 1
- 302 to 573 days
- 90 days: 0.6%;

Cluster 2
- 150 to 293 days
- 90 days: 10.6%; 1 year: 66.2%

Cluster 3
- 150 to 293 days
- 90 day: 94%; 30 days: 47.3%
- 6 months: 100%



What does their basket look like?

'Quantity'

- Nearly identical across all clusters for 80% of customers.
- Higher for top 20% of Clusters 0 and 2 customers

'Price'

- Nearly idetical across all clusters

'AOV' and 'Adjusted AOV'

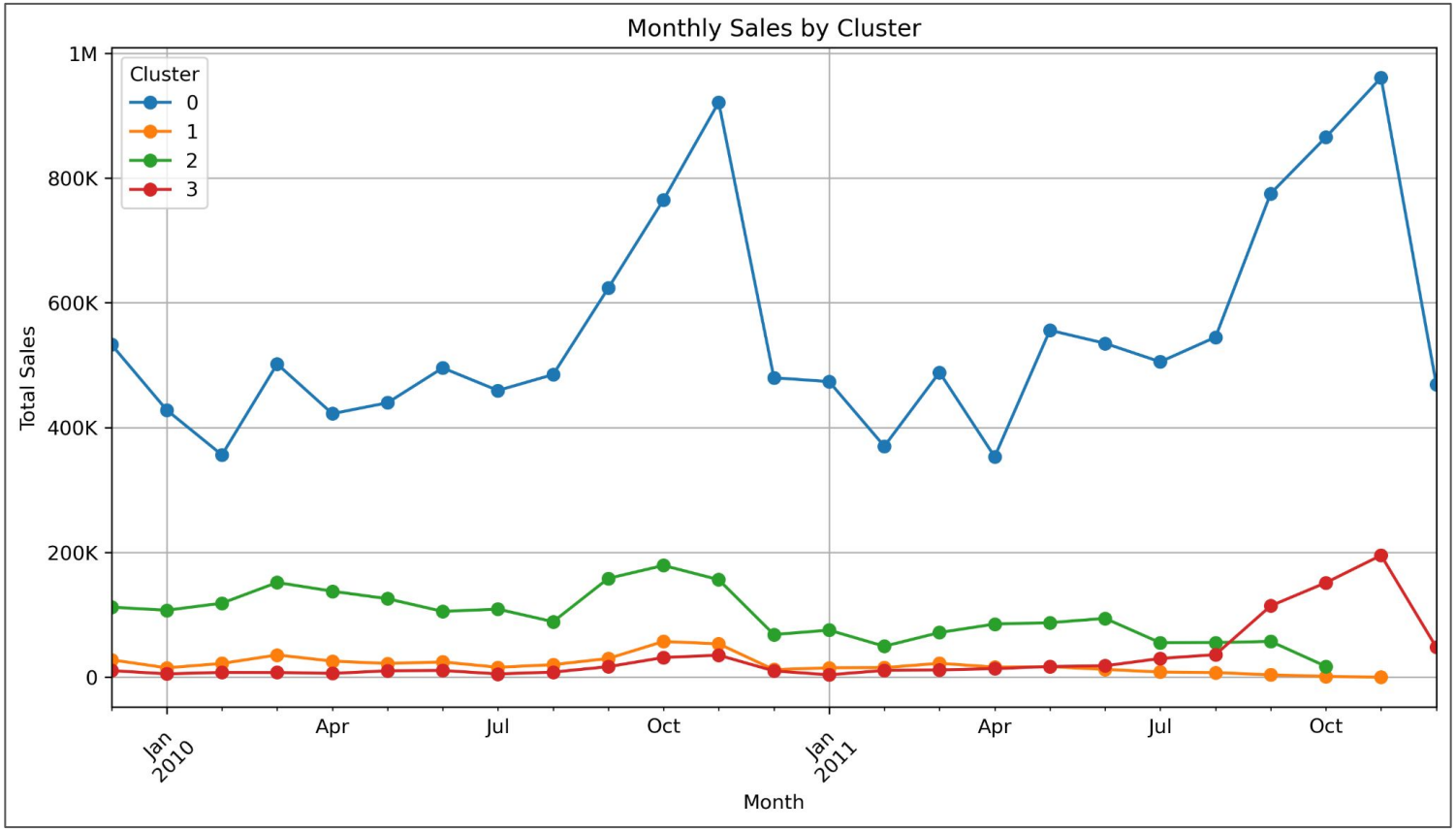
Cluster 0
- Highest 'AOV', lowest 'Adjusted AOV'

Cluster 1
- Lowest 'AOV, highest 'Adjusted AOV'

Clusters 2
- 2nd highest 'AOV', 2nd lowest 'Adjusted AOV'

Cluster 3:
- 2nd lowest 'AOV', 2nd highest 'Adjusted AOV'

Are there seasonal differences between segments?



Customer Segments

Cluster 0: Loyal, Frequent Customers

Cluster 1: One time customers

Cluster 2: Occasional Customers / Churn Risk

Cluster 3: Seasonal and/or New Customers

Recommendations and Follow-Up

Recommendation: Focus on Clusters 0 and 3

- Cluster 0 are the main customers and should be the main focus
- Cluster 3 provides the most opportunity for growth

Follow-up:

- Look deeper into year round customers to mitigate seasonal sales lulls
- Explore DBSCAN model to help account for wide variability of data