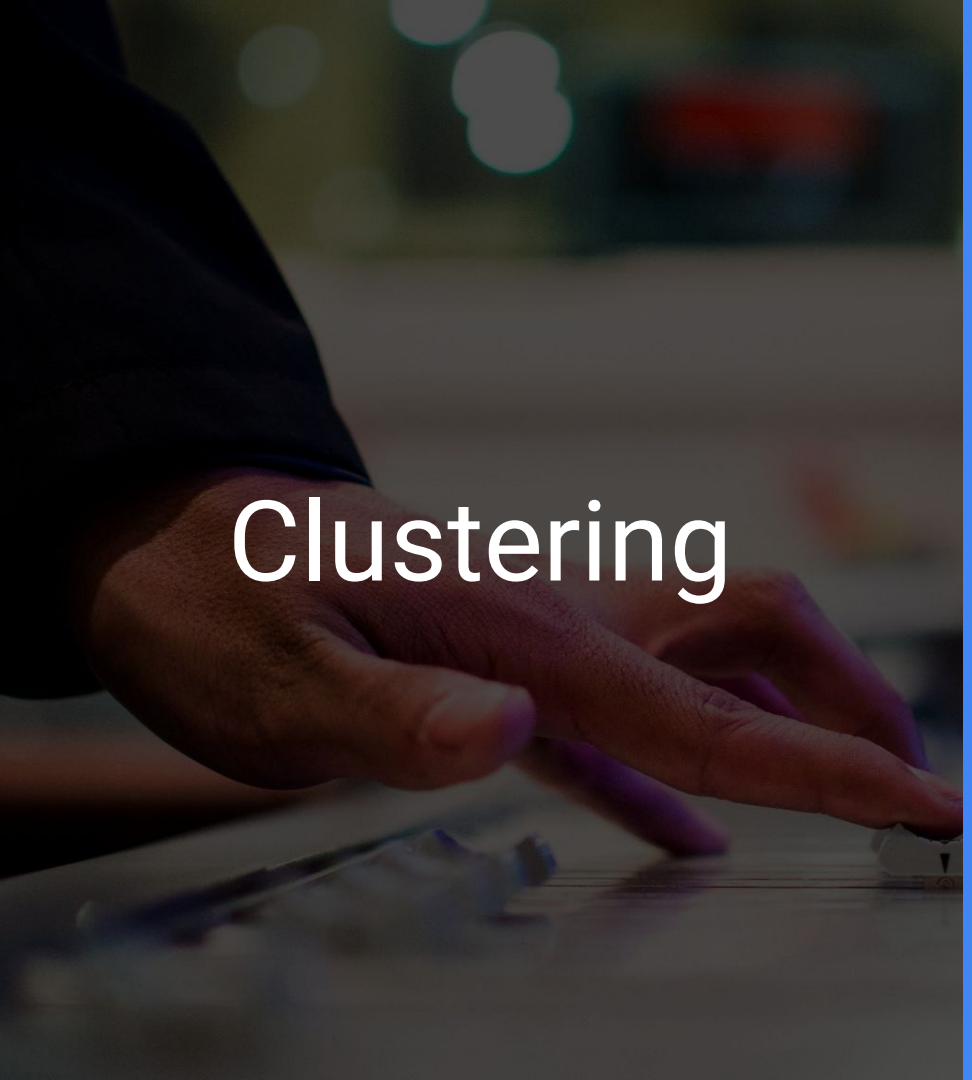
# Market Segmentation

Unsupervised Machine Learning

**Transactional Data:**
1 year, 500K + Observations



# Clustering

## By using clustering algorithms,

We want to identify various customer groups for various types of marketing and advertising campaigns.

# Example Customer Groups

## High Value, High Occurrence

Customers that spend high and Order often, this group is our most prized customer group and we want to protect it

## Low Value, Low Occurrence

This customer group in general should be avoided, and resources should not be allocated towards targeting this group

## High Value, Low Occurrence

We want to target this group in efforts to increase their occurrence since each occurrence is highly valuable

## Low Value, High Occurrence

We want to increase this customer group's value at each occurrence, eg: recommending complimentary items

# Milestones

*Steps to achieve optimal Market Segmentation*

## 1. Transform Data

Transform data to be evaluated at a customer level for clustering

## 2. Feature Engineering RFM

Creating standard marketing metrics from transactional data - (see next slide)

## 3. Elbow Methods

Finding out our optimal number of clusters

## 4. Creating Clusters

Group our customers into clusters based on spending habits/history

## 5. Evaluate

Evaluate the behaviors of our different customer groups

# RFM Variables (Recency, Frequency, Monetary Value)

## Recency

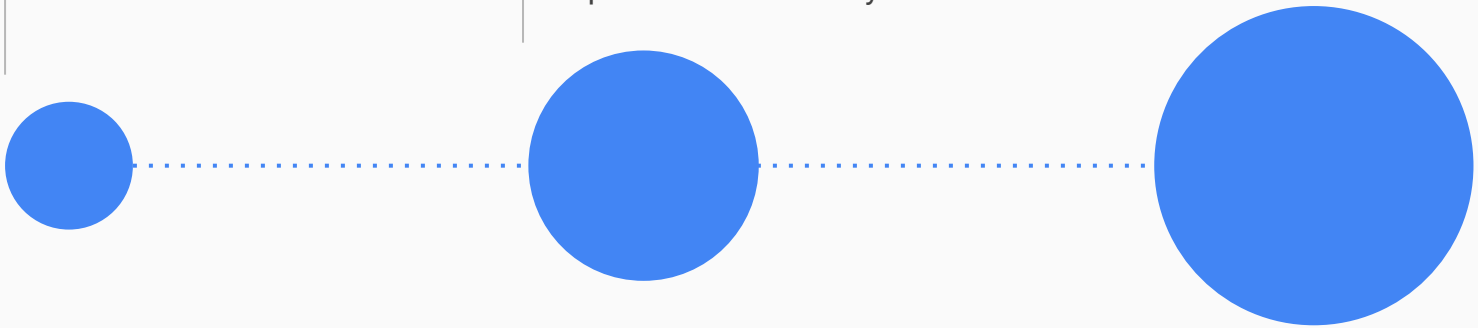
How recently a given customer made a purchase

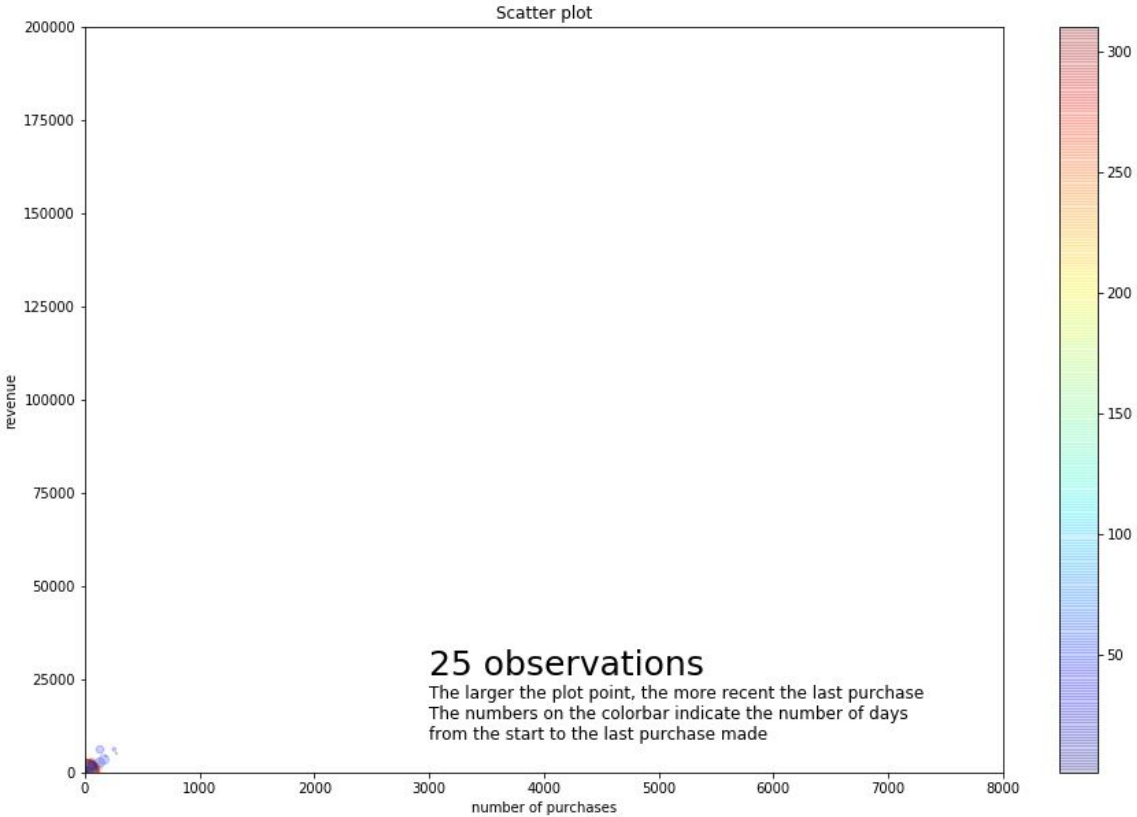
## Frequency

A customer's frequency of purchase, or their total number or purchases over the year

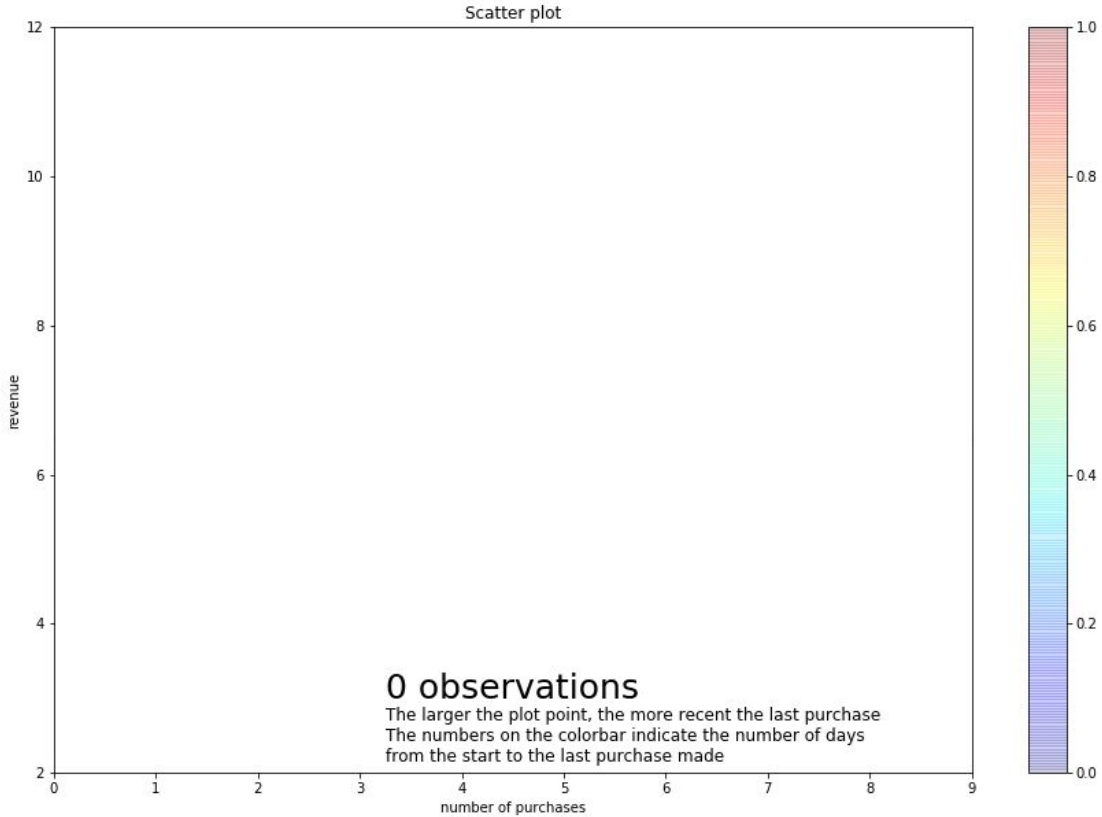
## Monetary Value

The lifetime value (1 year, in our case) of a given customer, ie: their total spend over the year






Visualizing Recency, Frequency, and Monetary Value of the overall dataset at the customer level




Smart GIF Maker

Log Transformed Recency, Frequency, and Monetary Value of the overall dataset at the customer level

# Purchases

***The previous graph is the meat of our transformed data.***

*We will run our clustering algorithms on this data, along with Total Quantity of Items Purchased, to create our different customer groups*

*(Data is Logged & Scaled for optimal results)*

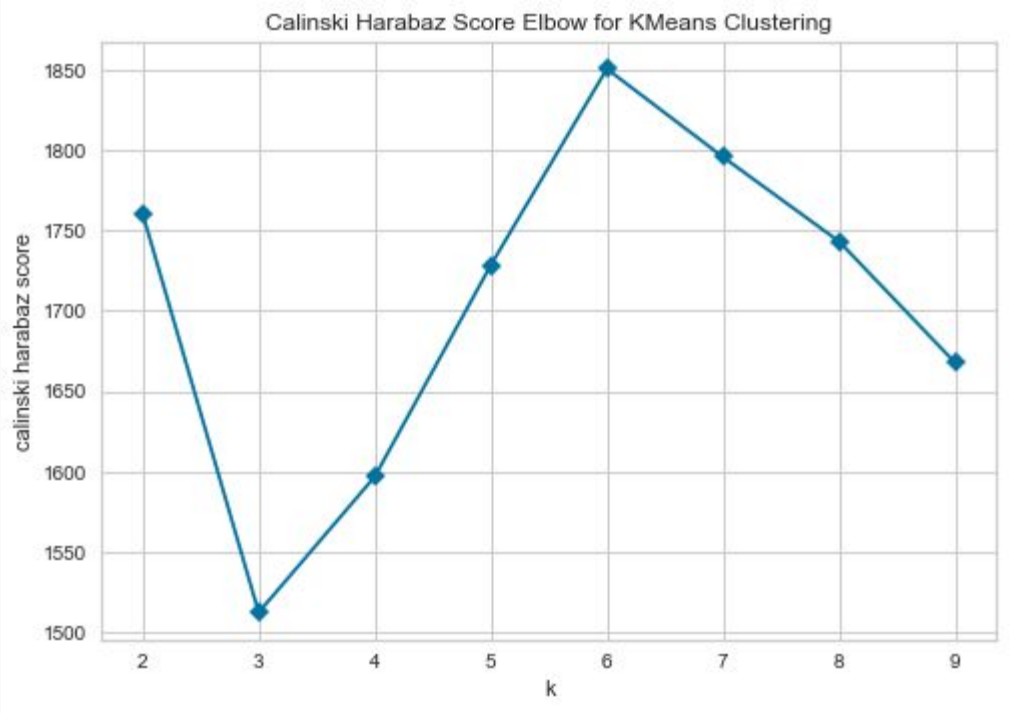
# Elbow Method:

According to our Elbow Method,

**6** would be our optimal number of market segments (clusters).

These clusters would have the highest Calinski-Harabasz (Variance Ratio) Score

We will further fine-tune this parameter when working with our pipelines

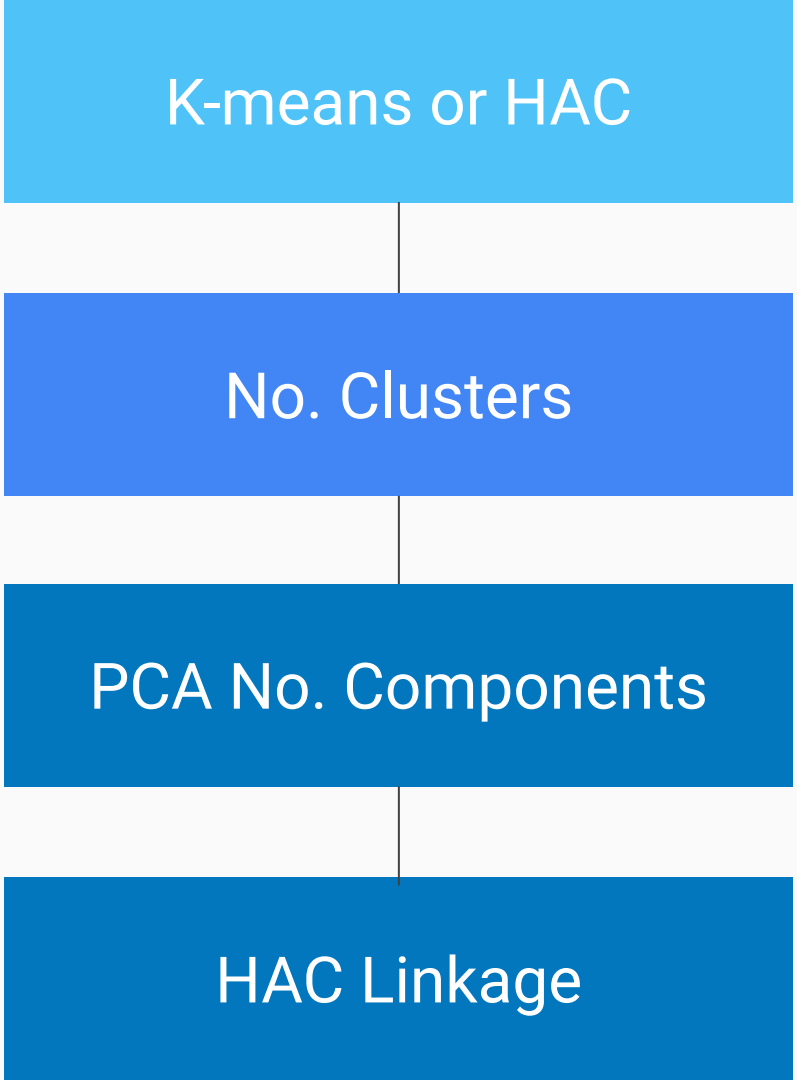


# The Algorithms:

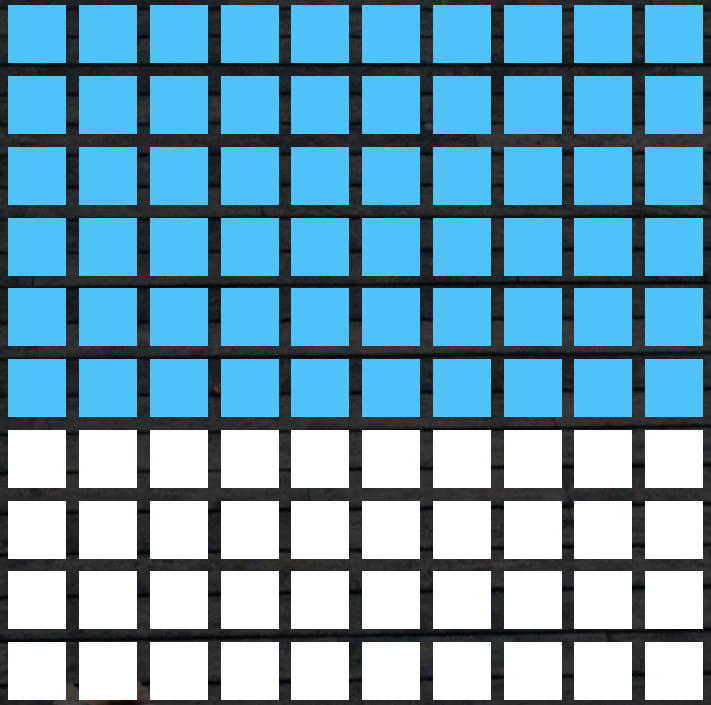
K-means Clustering
Hierarchical Agglomerative Clustering (HAC)

# Pipelines

We fine-tune...

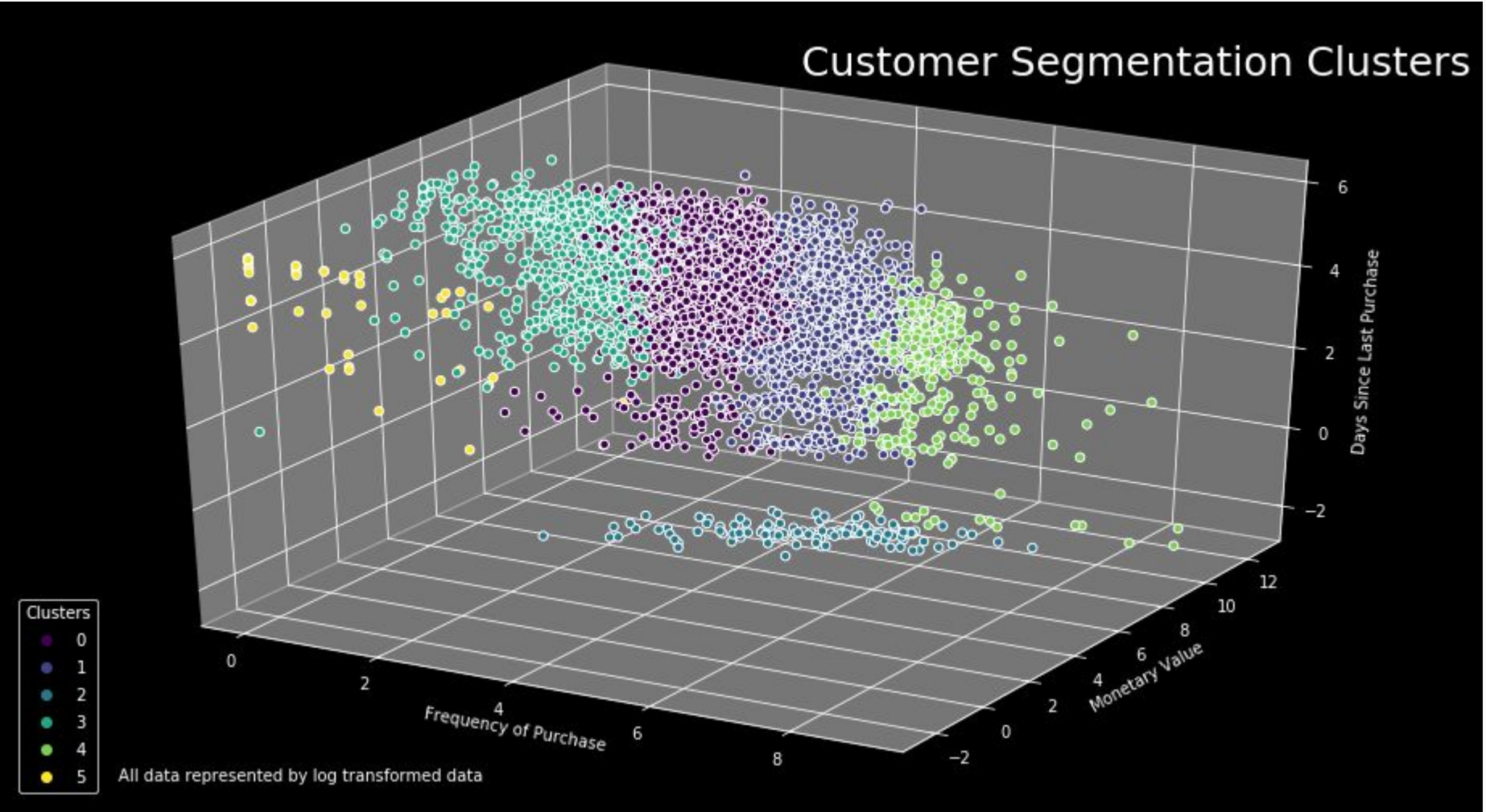


# The Best Solution



# Hierarchical Agglomerative Clustering (HAC)

**“Ward”** linkage
**6** clusters
**4** PCA components


Customer Segmentation Clusters
Clusters
0
1
2
3
4
5
All data represented by log transformed data
Frequency of Purchase
Monetary Value
Days Since Last Purchase

# BOTTOM-TIER CUSTOMERS (CLUSTER #5)

**Summary Stats for cluster # 5**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 51.000000 | 51.000000 | 51.000000 | 51.000000 |
| mean | -35.470588 | 233.588235 | 10.333333 | -258.486078 |
| min | -303.000000 | 9.000000 | 1.000000 | -4287.630000 |
| 25% | -17.500000 | 130.000000 | 1.000000 | -123.575000 |
| 50% | -1.000000 | 293.000000 | 3.000000 | -29.950000 |
| 75% | -1.000000 | 351.500000 | 5.500000 | -2.750000 |
| max | 2.000000 | 364.000000 | 202.000000 | 0.000000 |

**This cluster represents 1.18% of the population and -0.16% of Total Revenue, with a total spend of £-13,182.79**

# MIDDLE-BOTTOM CUSTOMERS (CLUSTER # 0 & 3)

**Summary Stats for cluster # 0**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 1873.000000 | 1873.000000 | 1873.000000 | 1873.000000 |
| mean | 346.876668 | 106.337427 | 36.423385 | 596.160262 |
| min | 42.000000 | 1.000000 | 2.000000 | 77.400000 |
| 25% | 177.000000 | 28.000000 | 21.000000 | 319.610000 |
| 50% | 273.000000 | 66.000000 | 31.000000 | 483.250000 |
| 75% | 440.000000 | 168.000000 | 47.000000 | 740.950000 |
| max | 2924.000000 | 365.000000 | 229.000000 | 7092.060000 |

**This cluster represents 43.22% of the population and 13.96% of Total Revenue, with a total spend of £1,116,608.17**


**Summary Stats for cluster # 3**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 823.000000 | 823.000000 | 823.0000 | 823.000000 |
| mean | 168.669502 | 153.266100 | 8.652491 | 244.173451 |
| min | 0.000000 | 2.000000 | 1.000000 | -43.400000 |
| 25% | 44.000000 | 50.000000 | 5.000000 | 111.150000 |
| 50% | 78.000000 | 146.00000 | 8.000000 | 162.700000 |
| 75% | 128.000000 | 251.50000 | 12.00000 | 233.880000 |
| max | 12540.0000 | 365.00000 | 36.00000 | 6977.040000 |

**This cluster represents 18.99% of the population and 2.51% of Total Revenue, with a total spend of £200,954.75**

# MIDDLE-TOP CUSTOMERS (CLUSTER # 1 & 2)

**Summary Stats for cluster # 1**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 1149.000000 | 1149.000000 | 1149.000000 | 1149.000 |
| mean | 1250.818973 | 43.960836 | 127.000870 | 2045.676 |
| min | 127.000000 | 1.000000 | 7.000000 | 187.690 |
| 25% | 659.000000 | 9.000000 | 82.000000 | 1121.210 |
| 50% | 991.000000 | 22.000000 | 110.000000 | 1661.330 |
| 75% | 1505.000000 | 56.000000 | 156.000000 | 2543.240 |
| max | 17280.000000 | 330.000000 | 548.000000 | 21535.900 |

**This cluster represents 26.51% of the population and 29.38% of Total Revenue, with a total spend of £2,350,481.65**


**Summary Stats for cluster # 2**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 133.000000 | 133.0 | 133.000000 | 133.000000 |
| mean | 1472.375940 | 0.0 | 161.406015 | 2663.839699 |
| min | 32.000000 | 0.0 | 5.000000 | 80.800000 |
| 25% | 518.000000 | 0.0 | 52.000000 | 999.720000 |
| 50% | 1186.000000 | 0.0 | 121.000000 | 2007.440000 |
| 75% | 1978.000000 | 0.0 | 203.000000 | 3642.040000 |
| max | 6211.000000 | 0.0 | 1629.000000 | 11598.050000 |

**This cluster represents 3.07% of the population and 4.43% of Total Revenue, with a total spend of £354,290.68**

# TOP TIER CUSTOMERS (CLUSTER # 4)

**Summary Stats for cluster # 4**

| | quantity | recency | frequency | monetary |
|---|---|---|---|---|
| count | 305.000000 | 305.000000 | 305.000000 | 305.000000 |
| mean | 7567.075410 | 10.852459 | 485.685246 | 13089.081836 |
| min | 710.000000 | 0.000000 | 43.000000 | 1546.910000 |
| 25% | 2410.000000 | 2.000000 | 249.000000 | 4022.460000 |
| 50% | 3489.000000 | 7.000000 | 341.000000 | 5820.160000 |
| 75% | 5635.000000 | 15.000000 | 480.000000 | 10360.040000 |
| max | 188761.000000 | 78.000000 | 7637.000000 | 267761.000000 |

**This cluster represents 7.04% of the population and 49.89% of Total Revenue, with a total spend of £3,992,169.96**

# In Conclusion

We’ve identified our top-tier customers and their ordering habits

We’ve also identified middle-top and middle-bottom tier customers, as well as the bottom tier customers we want to avoid

We’ve can now target our marketing efforts accordingly.