

Customer and Product Segmentation

Using RFM Analysis and k-means Clustering On XYZ Hardware Wholesaler Transactions

Prepared By

Pavan Gorantla

Himaja Barla

Mudit Agrawal

Anusha Karasala

Rohit Doddapaneni

Jen-Hsin Huang

Yun-Han Liu

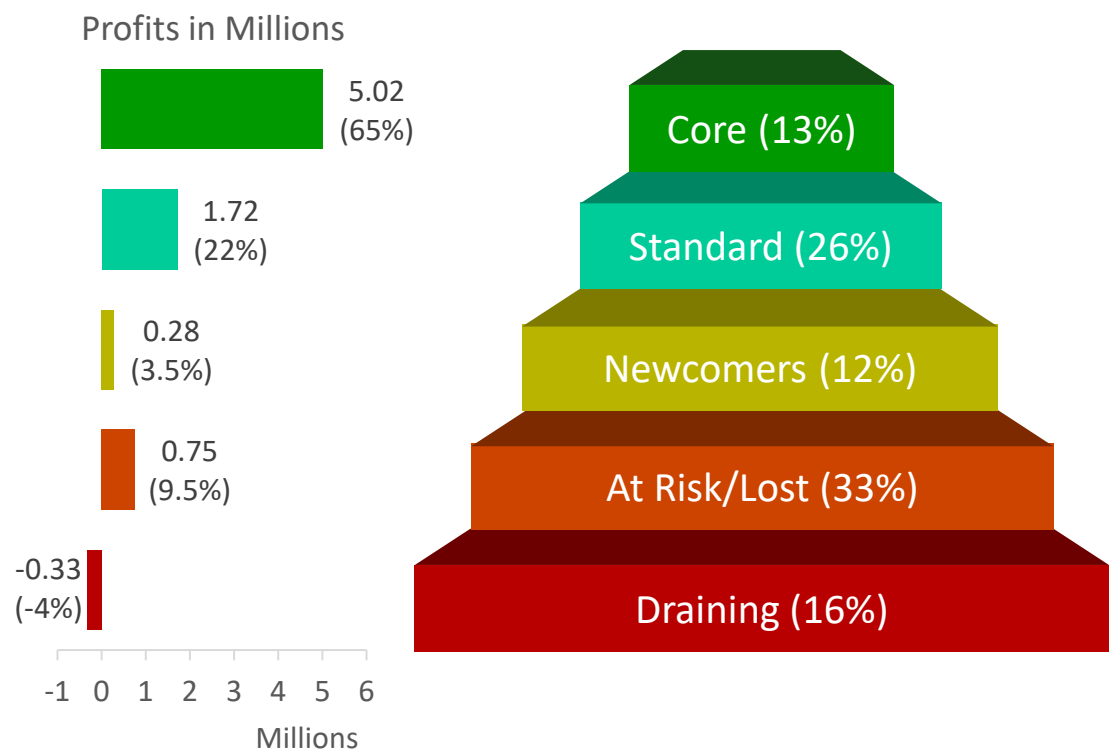
In Association With



Segmentation Pyramids

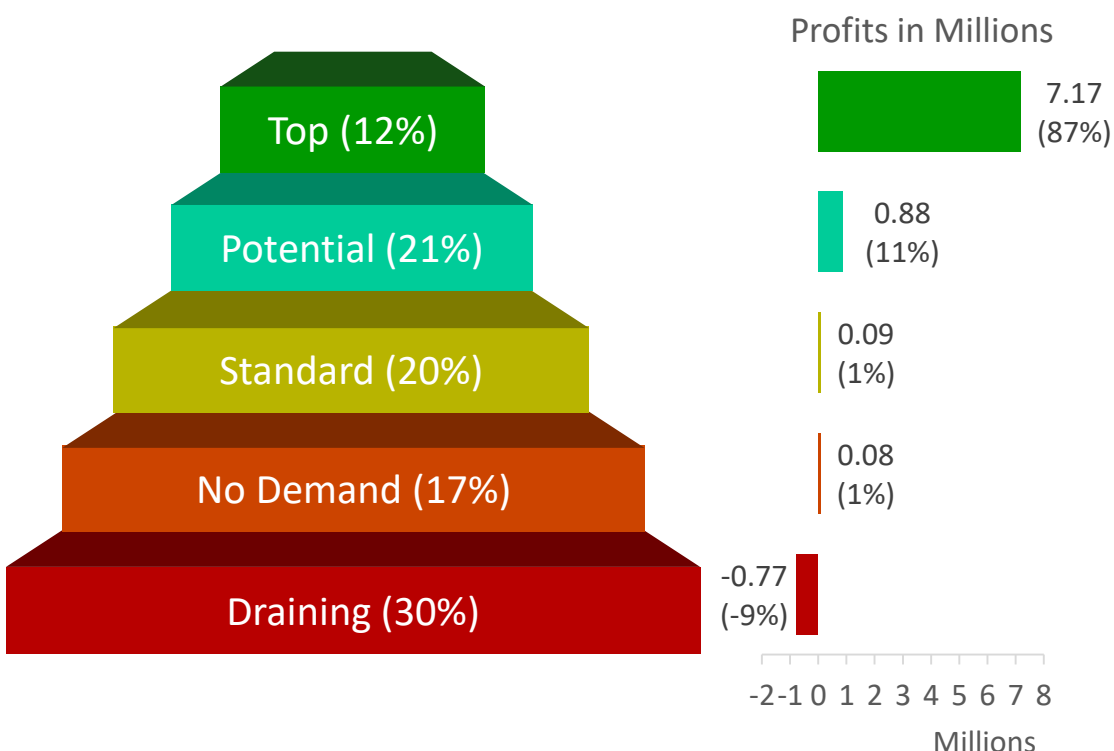
Segmented Groups for Customers

226 (95%) Active Customers out of 4372 Total Customers



Segmented Groups for Products

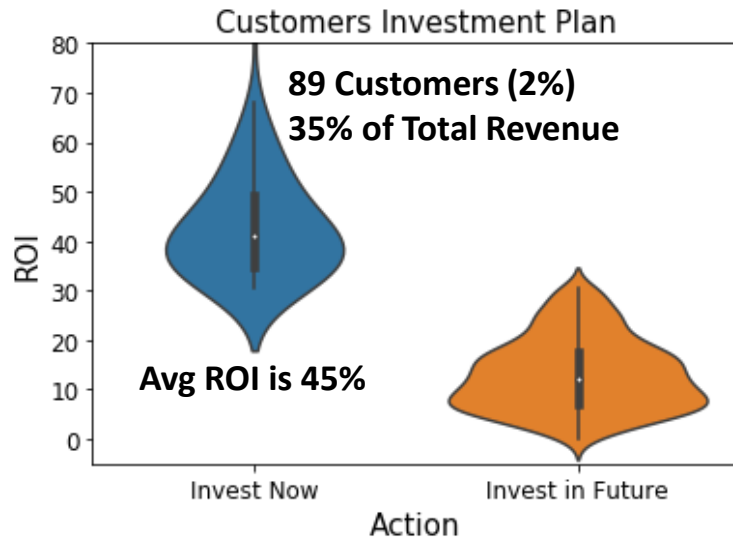
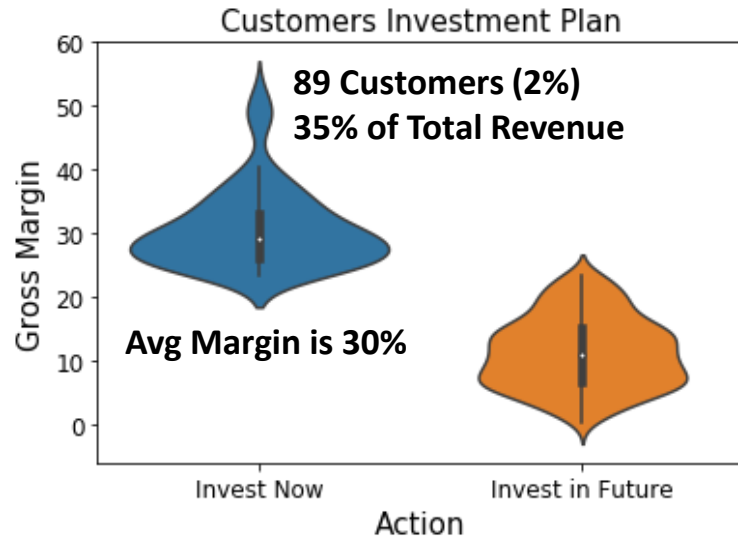
3805 (96%) Shipped Products out of 3958 Total Products



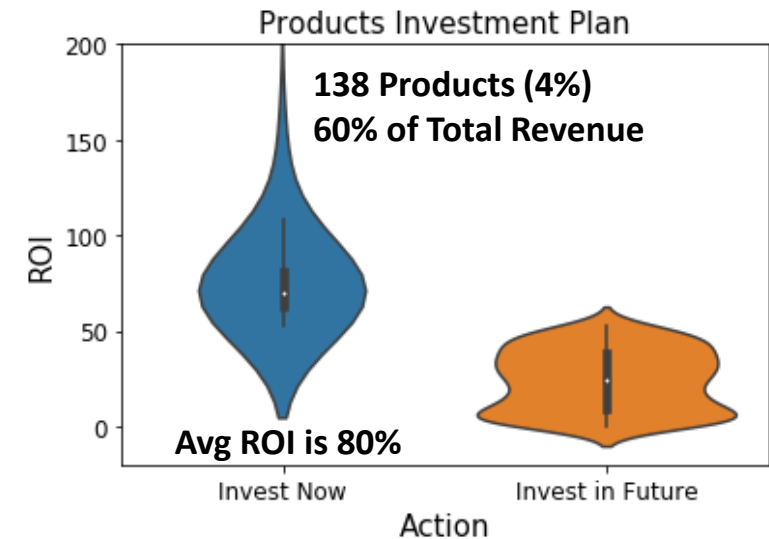
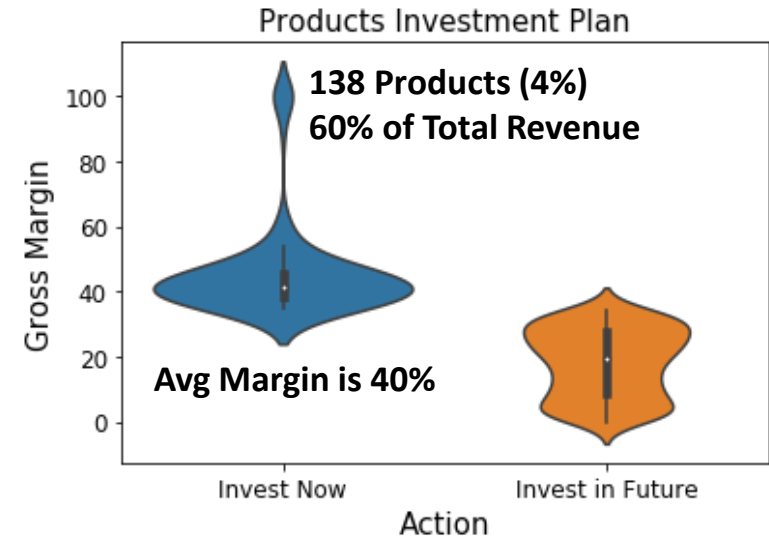
We may want to invest in all customers and products from best groups (green) but they cost more than \$10 million

Where To Invest?

Core Customers with high Margin & ROI costing \$5 million (blue)



Top Products with high Margin & ROI costing \$5 million (blue)



RFM Analysis

RFM Analysis says if a customer or product is good or bad using 3 independent factors

Recency

- When did a customer buy for the last time?
- When was a product sold for the last time?

Measured by: Days since Recent Purchase
(Can use same for Customers and Products)

Frequency

- How frequently does a customer purchase?
- Which products are sold in large quantities?

Measured by: No. of Transactions by Customer
Measured by: Shipped Quantity of the Product

Monetary

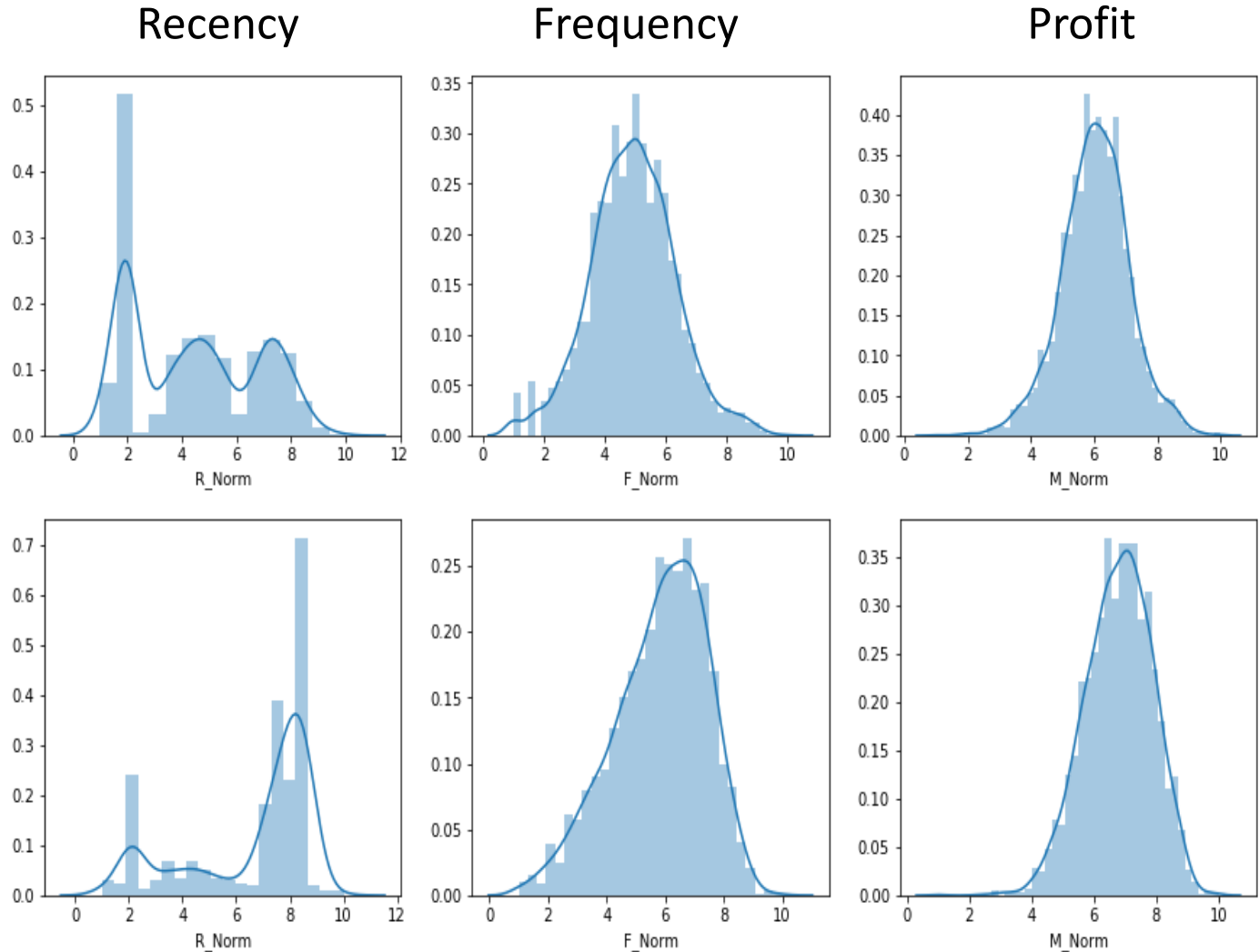
- What's the financial gain from a customer or product?

Customer/Product	Price	Cost	Profit	Margin	ROI
A	50000	40000	\$ 10,000	20%	25%
B	3000	1500	\$ 1,500	50%	100%

1/7th Profit

- If Customer B did 7 similar transactions or Product B is shipped 7 times, then Margin and ROI are best
- Margin and ROI will become better financial metrics provided only if B has higher Frequency than A
- **Profit** is a financial metric that is independent of Frequency (no. of transactions, quantity shipped)

Normalization



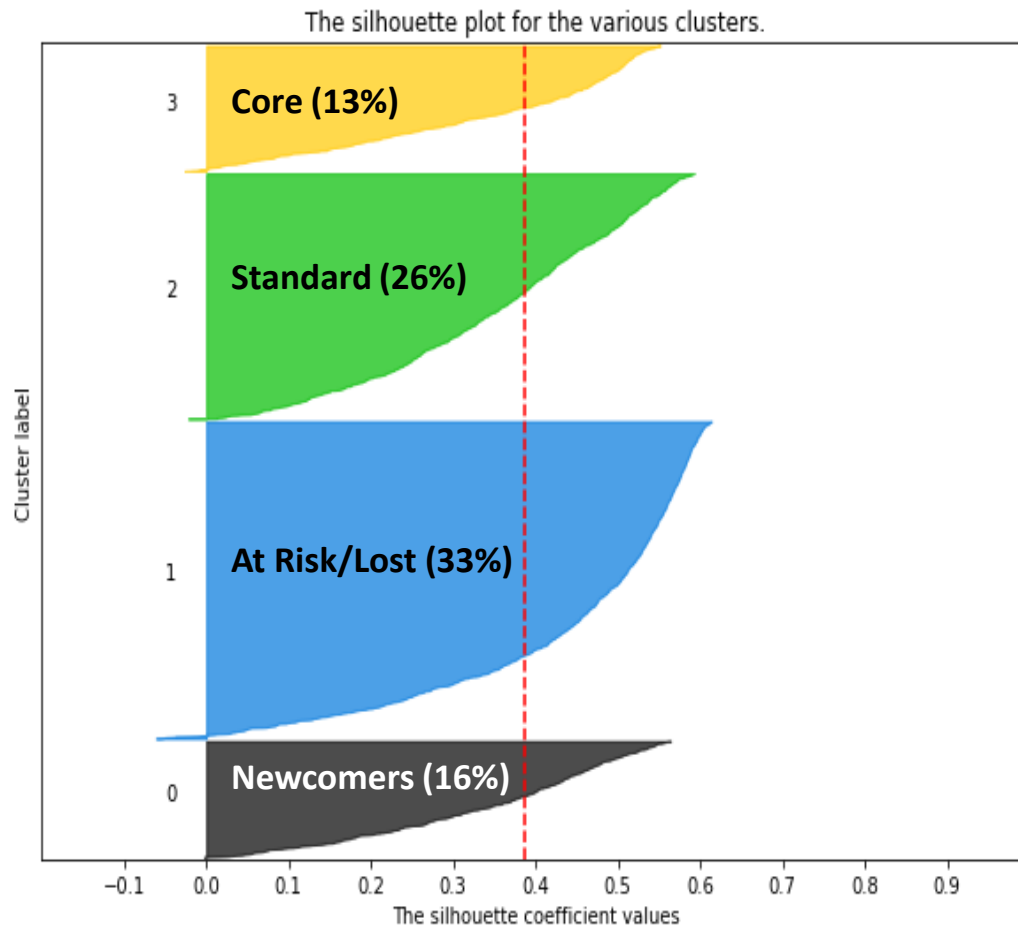
Customers

- 3 features are log transformed
- Normalized to a scale of 1 to 10
- 1 means bad and 10 means good

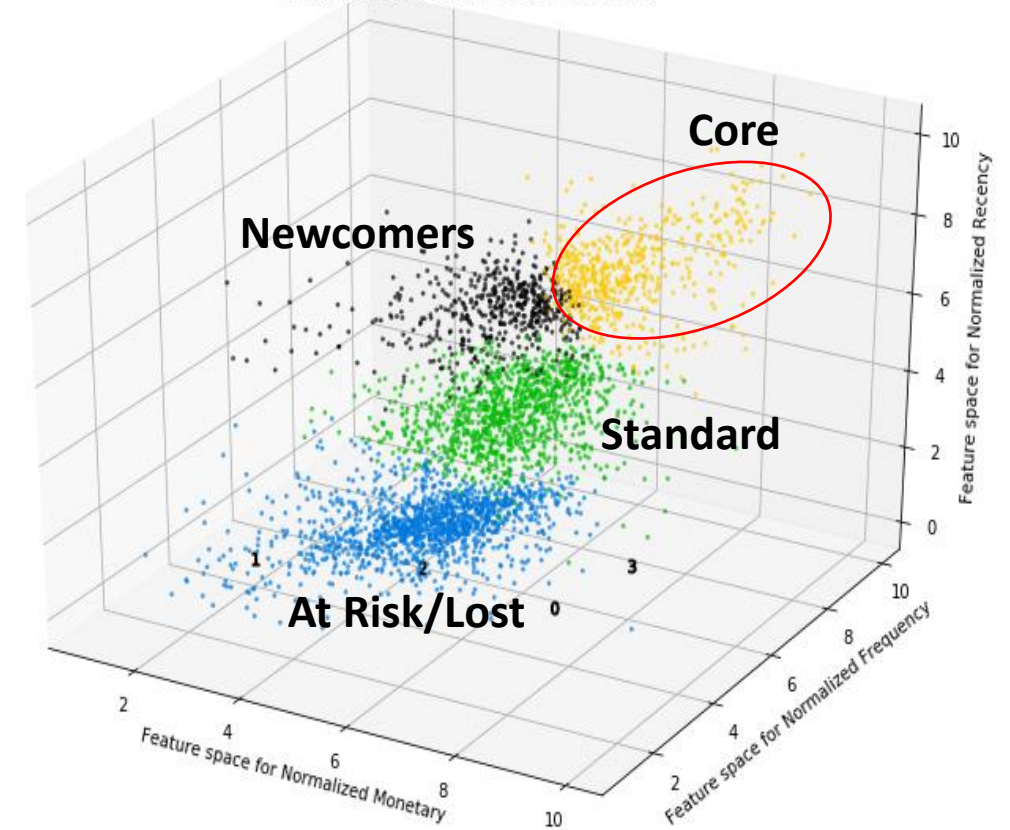
Products

Customers Clustering

Cluster group 3 (**Orange**) has highest Recency, Frequency and Monetary values which we named as our Core group

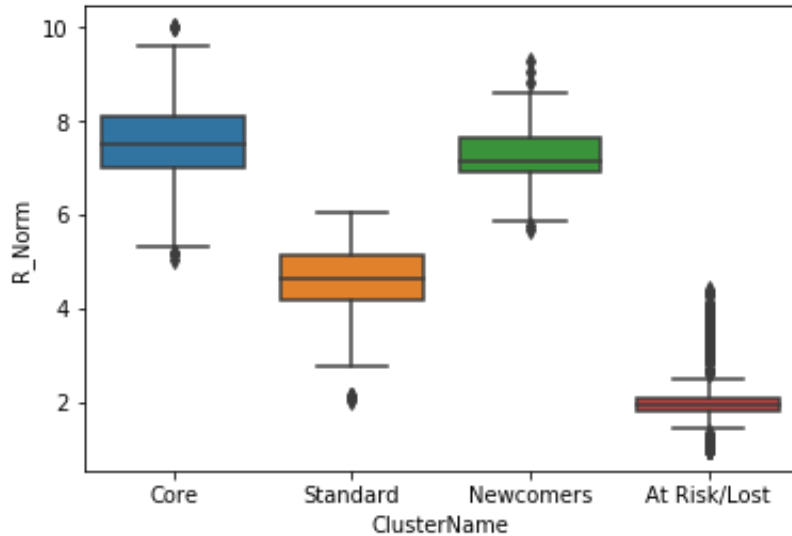


Silhouette analysis for KMeans clustering on Customers RFM data with n_clusters = 4
The visualization of the clustered data.

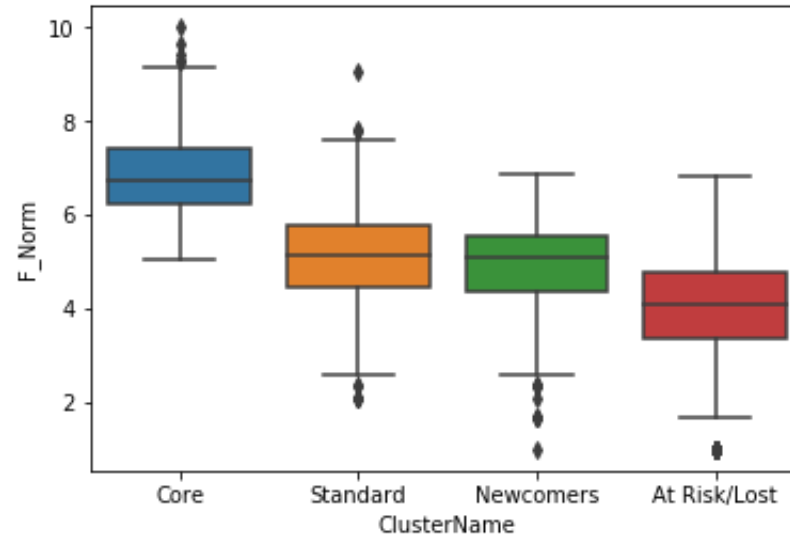


Customers Clustering

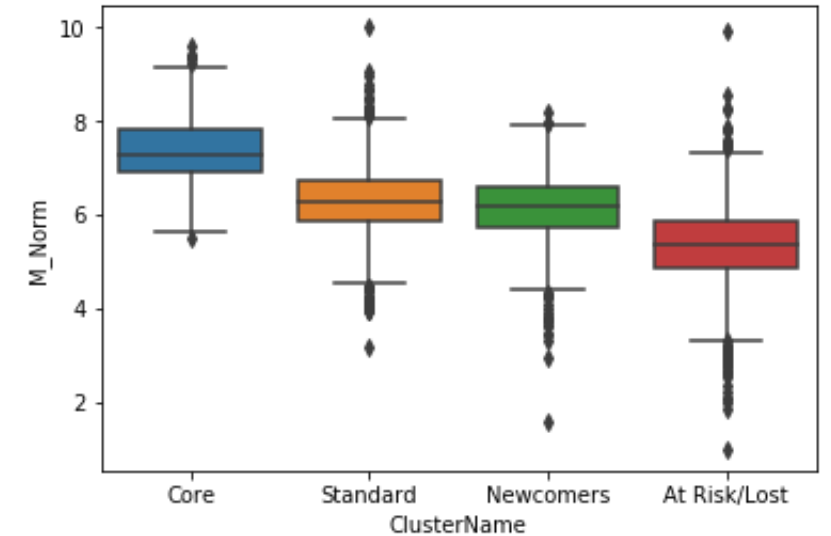
Recency



Frequency



Profit



Core – Very active, purchased most recently, did most transactions, high profits, best to be in regular touch

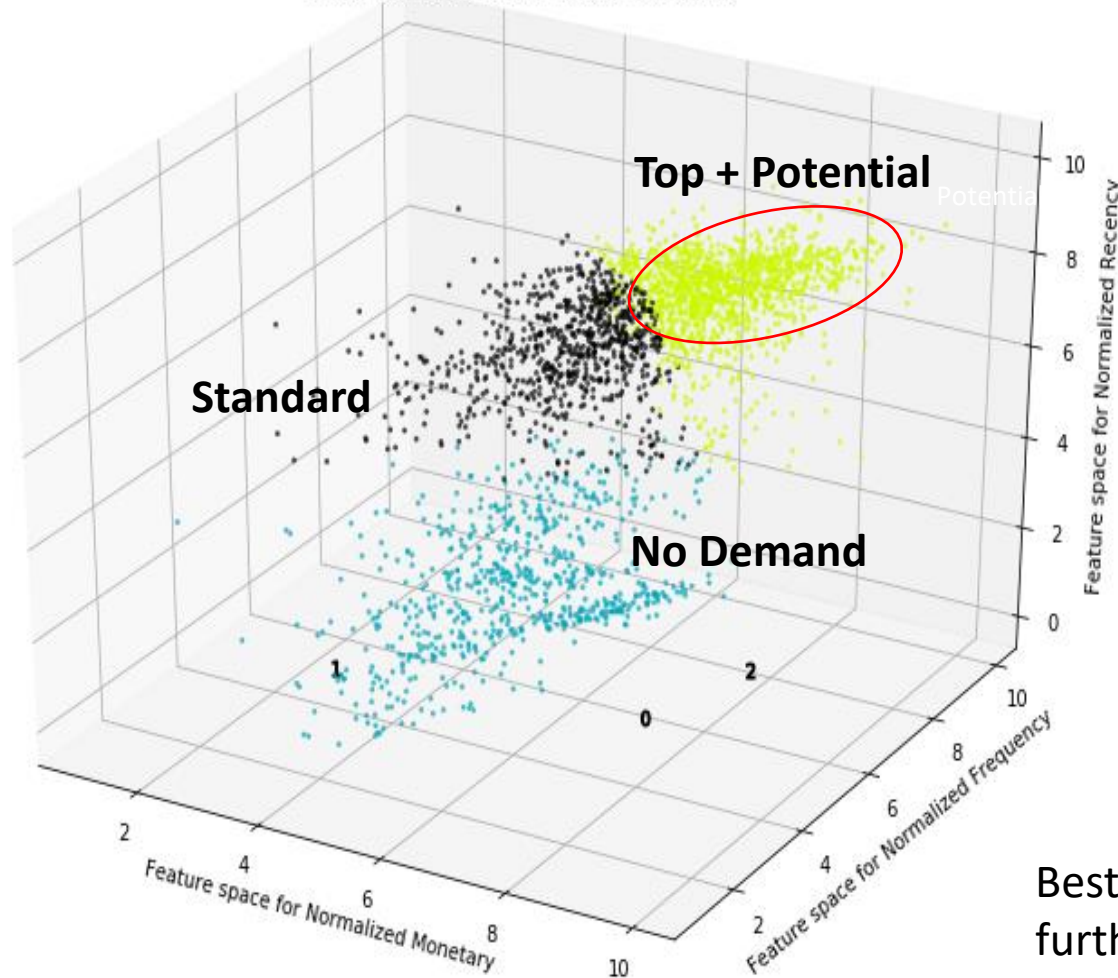
Standard – Moderately Active, purchased a quarter ago, transactions with moderate frequency, decent profits

Newcomers – Very active, started recently, competing with Standard group Customers, observe them closely

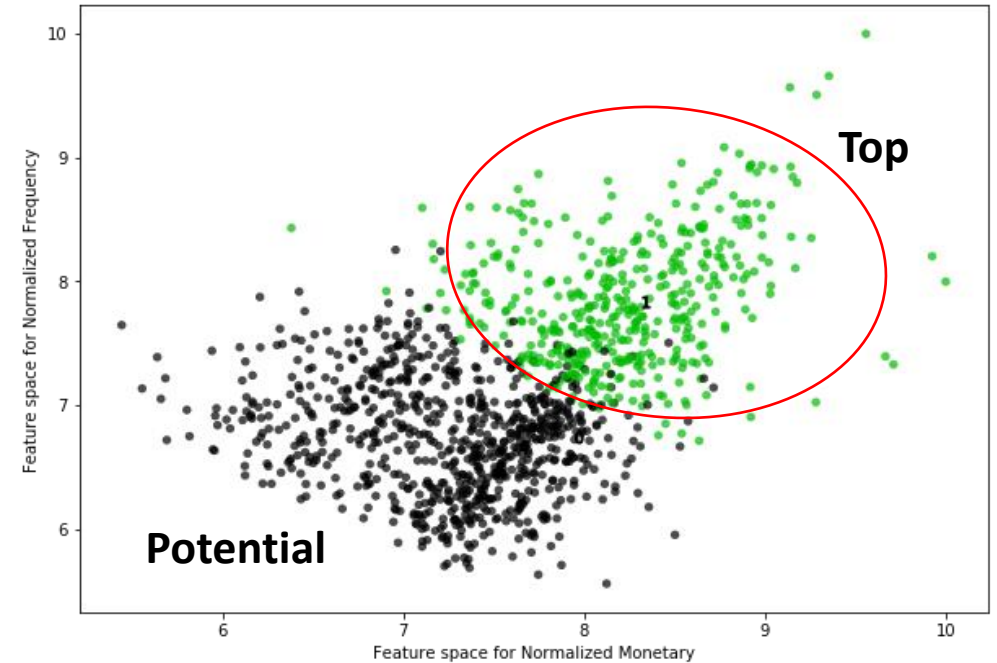
At Risk/Lost – Did transactions over an year ago, good to re-connect and negotiate for new deals and purchases

Products Clustering

Silhouette analysis for KMeans clustering on Customers RFM data with $n_clusters = 3$
The visualization of the clustered data.



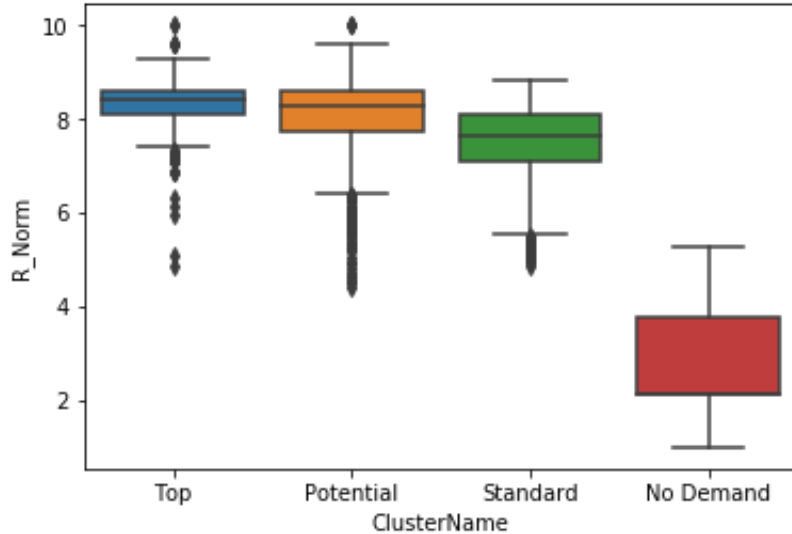
Silhouette analysis for KMeans clustering on Customers RFM data with $n_clusters = 2$
The visualization of the clustered data.



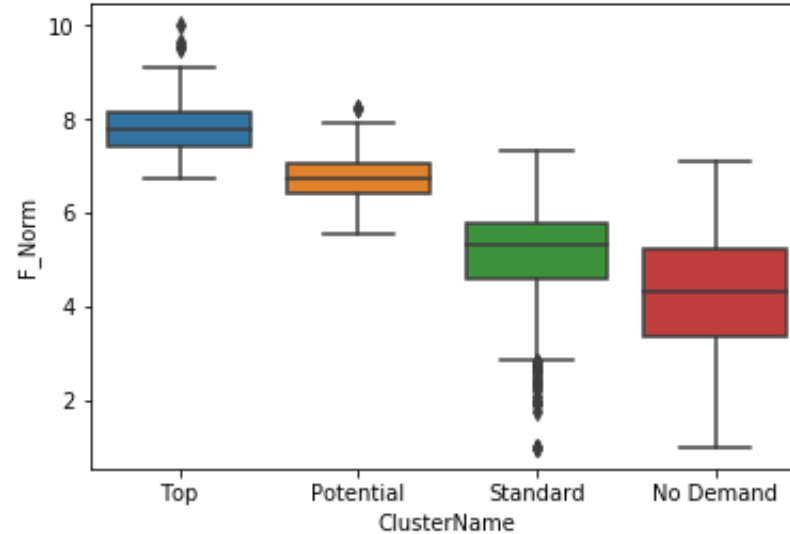
Best Cluster (Yellow) has 1200 products, so only that cluster is further sub-divided to 2 groups giving our Top products (Green)

Products Clustering

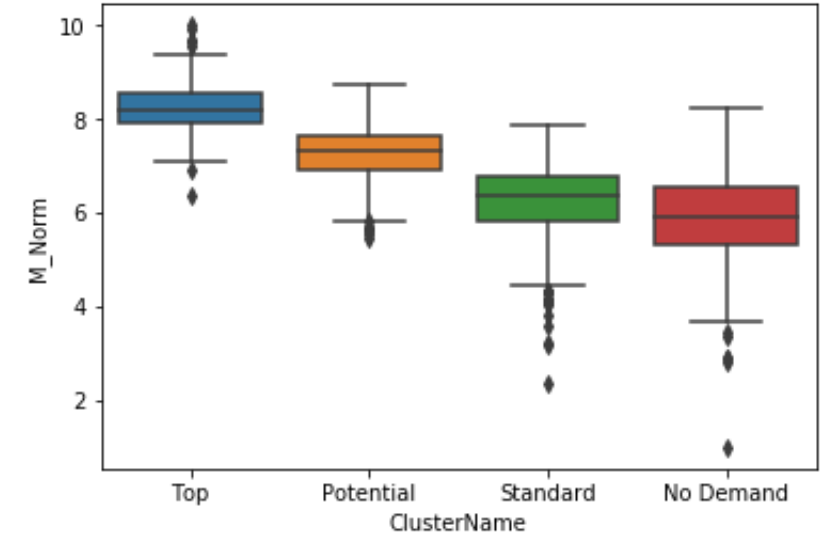
Recency



Frequency



Profit



Top – High demand, Recently purchased, shipped in large quantities, contributed for high profits, best to invest

Potential – Second to Top group, purchased most recently, moderate quantities shipped, decent profit share

Standard – Moderately in demand, medium quantities are shipped, generated below average profits

No Demand – Products almost obsolete, bought over 1 year ago, lowest quantities shipped, lowest profit share



Any Questions?