Online Retail Customer Behavior Analysis and Customer Segmentation

Springboard Data Science Career Track Capstone 3

Yuan Yin • 08.25.2021

Introduction

- **Dataset**:

  Two-year transactional data of an online gift retail, many customers of which are wholesalers

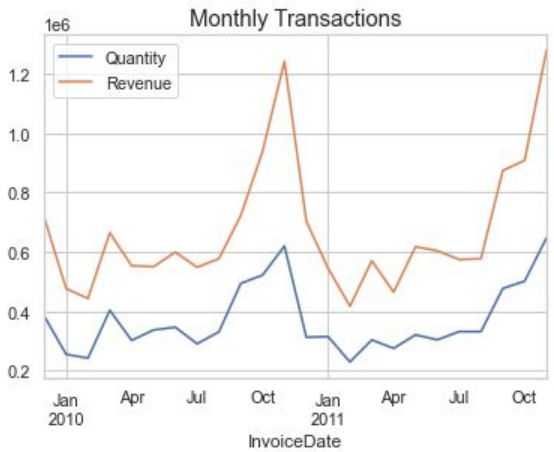
- **Goals of This Project**
  - To **analyze customer behavior** to help the company know the customers well
  - To **cluster customers** for better designing targeted marketing campaigns

Customer Behavior Analysis

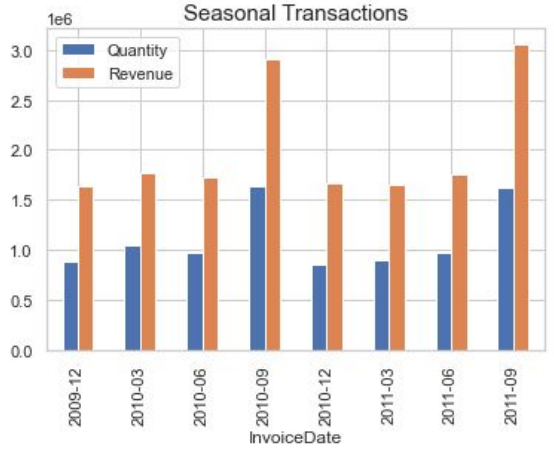
- Monthly/Seasonal Transaction Analysis
- Time Cohort Analysis
- Historical Customer Lifetime Value Calculation

Monthly/Seasonal Transaction Analysis

- **Sales Volume and Revenue by Month**

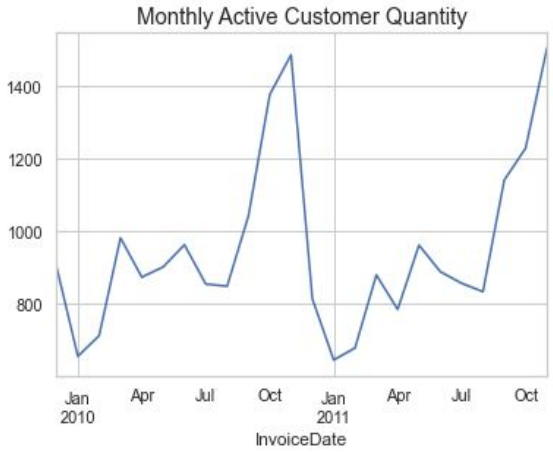


- **Sales Volume and Revenue by Season***

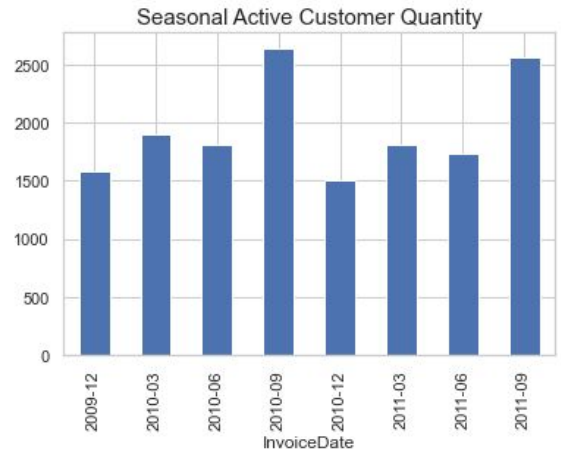


Monthly/Seasonal Customers Analysis

- **Number of Active Customers by Month**



- **Number of Active Customers by Season**



Time Cohort Analysis

- **What is Cohort Analysis**

  Analyse data based on grouped customers(Cohort) rather than looking at the data as one unit.

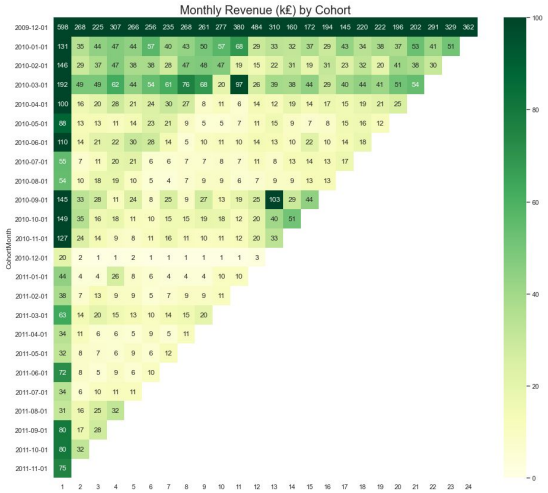
- **Why Cohort Analysis**

  Develop targeted marketing strategies

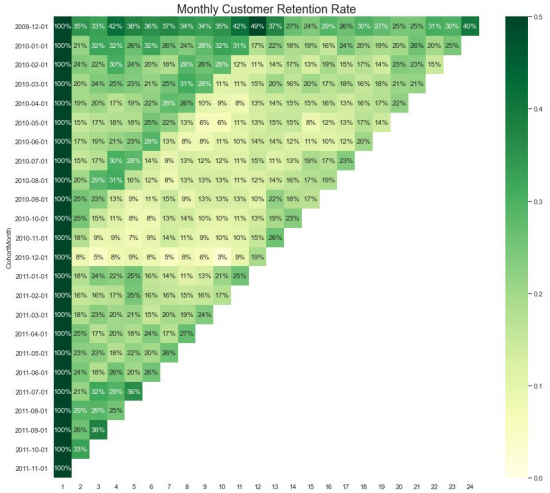
- **Types of Cohort to Analyse**
  - **Time-Based Cohorts**
  - Segment-Based Cohorts
  - Size-Based Cohorts

Time Cohort Analysis

- **Average Revenue of Monthly Cohort**



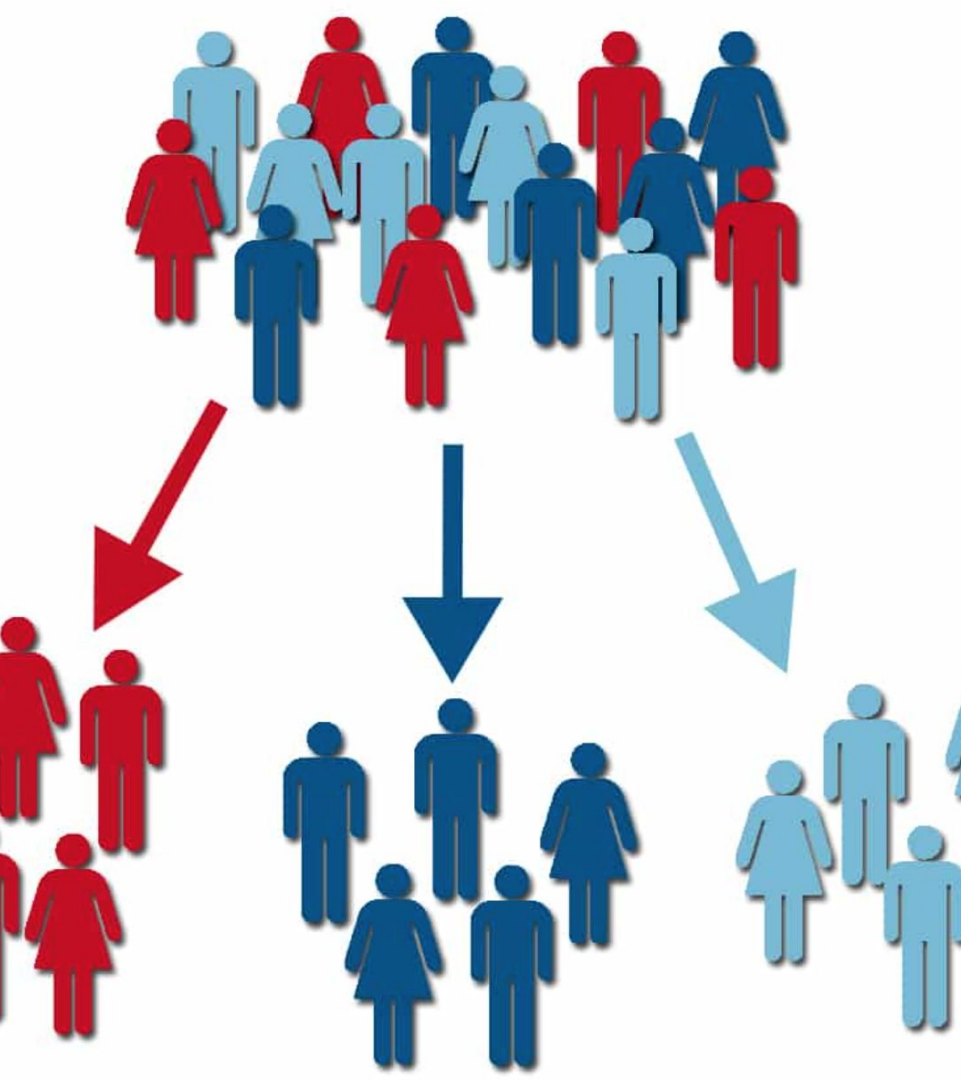
- **Average Retention Rate of Monthly Cohort**



Customer Segmentation

— With KMeans Algorithm

- Metric Selection
- Data Preprocessing
- Number of Clusters Optimizing
- Customer Profiling



- Customer segmentation is the practice of dividing customers into groups based on the similarity of characteristics.
- Each group or segment is related to a significant customer profile so companies can design targeted marketing campaigns.

Metric Selection

- **Recency, Frequency, and Monetary Value (RFM)**

➔ Recency: **how recent** was each customer's last purchase
➔ Frequency: **how many** purchases the customer has done
➔ Monetary Value: **how much** has the customer spent

- **Recency, Monetary Value, and Tenure (RMT)**

➔ Tenure: **how long** the customer has been with the company since their first transaction

Data Preprocessin

➔ **Why Data Preprocessing**

KMeans Assumption:

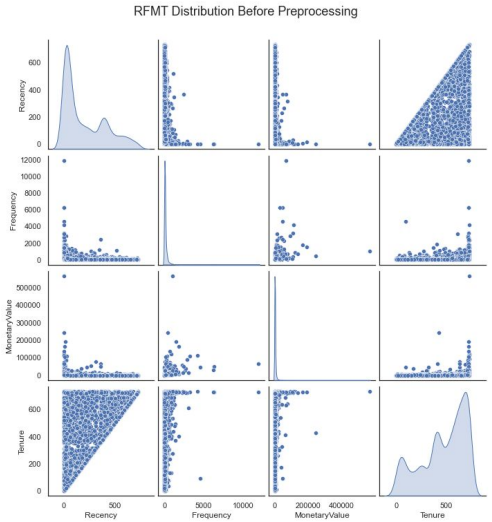
- Numeric Features
- Symmetrical Features
- Same Mean Values and Same Standard Deviation

➔ **How to Preprocess data**

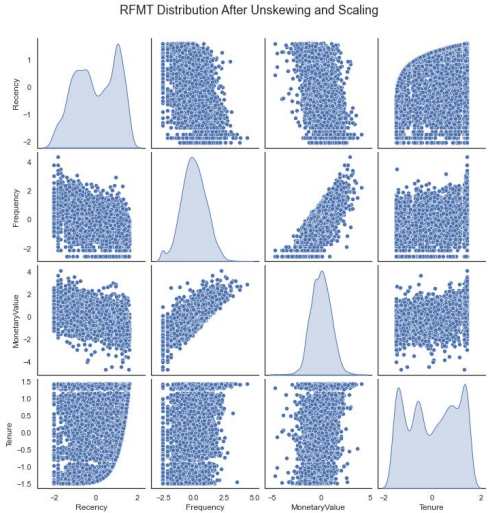
- Unskew Feature
- Scale Feature

Data Pre-processing

- **RFMT Distribution Before Preprocessing**

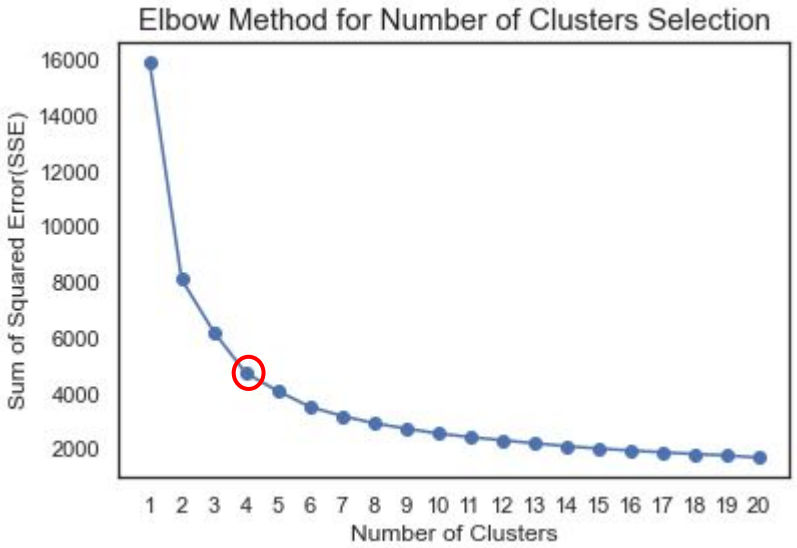


- **RFMT Distribution After Preprocessing**

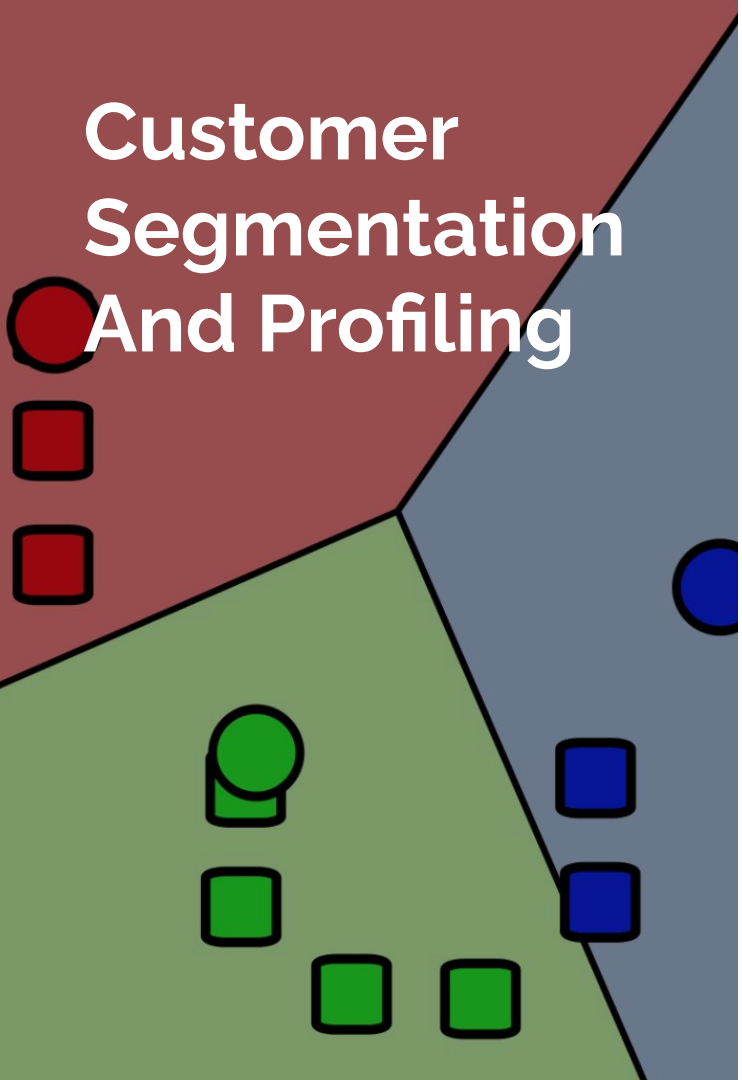


Number of Clusters Optimizing

- **How to Predefine the Number of Clusters, i.e., k**



Customer Segmentation And Profiling



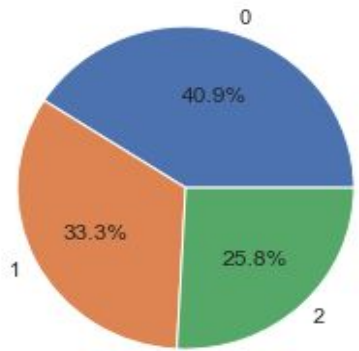
- **How to Do Customer Segmentation**
  - ➢ Customer segmentation is a type of clustering task
  - ➢ The best-known Machine Learning algorithm for clustering is KMeans.
  - ➢ In general, a well-formed set of clusters have two highlights:
    - ✓ good cohesion
    - ✓ good separation
  - ➢ **Business Insights**

Customer Segmentation And Profiling

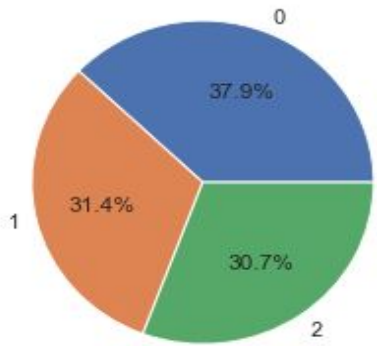
- **How to Identify Customer Profiles**
  - Based on RFM or RMT metrics
  - Summary Statistics
  - Relative Importance
  - Snake Plot
  - 2D kde Plot
  - 3D Scatter Plot

Basic Solution: 3-Segment Model Based on RFM/RMT

- Segment Size Ratio of RFM Model

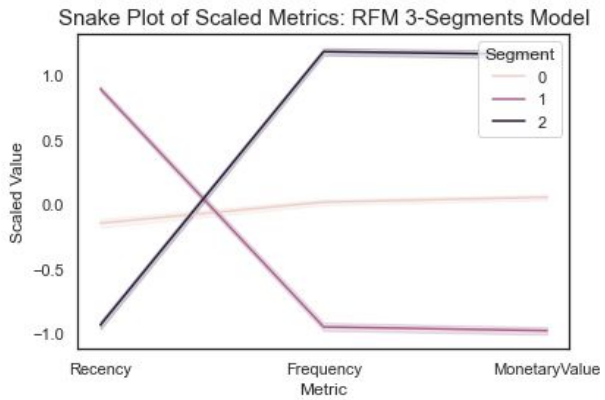


- Segment Size Ratio of RMT Model

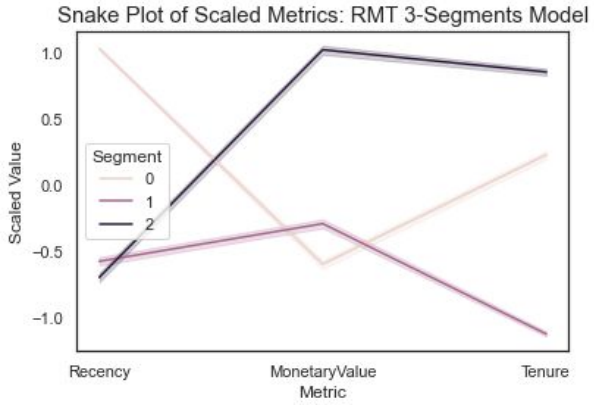


Basic Solution: 3-Segment Model Based on RFM/RMT

- Snake Plot of RFM Model

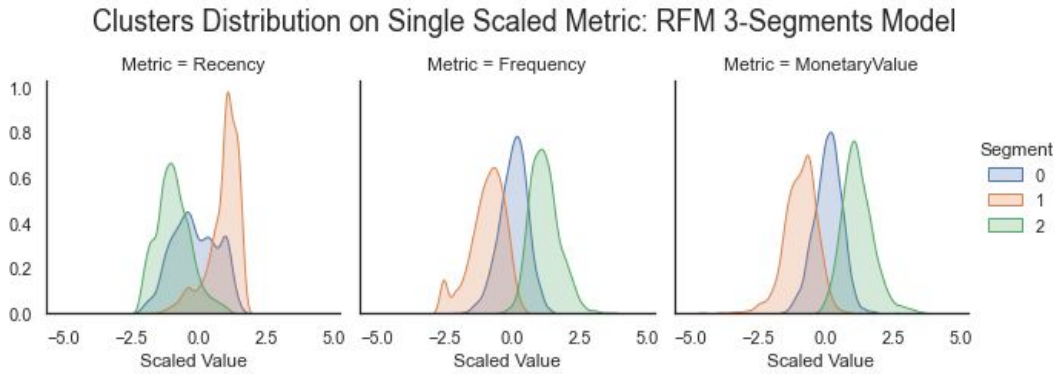


- Snake Plot of RMT Model

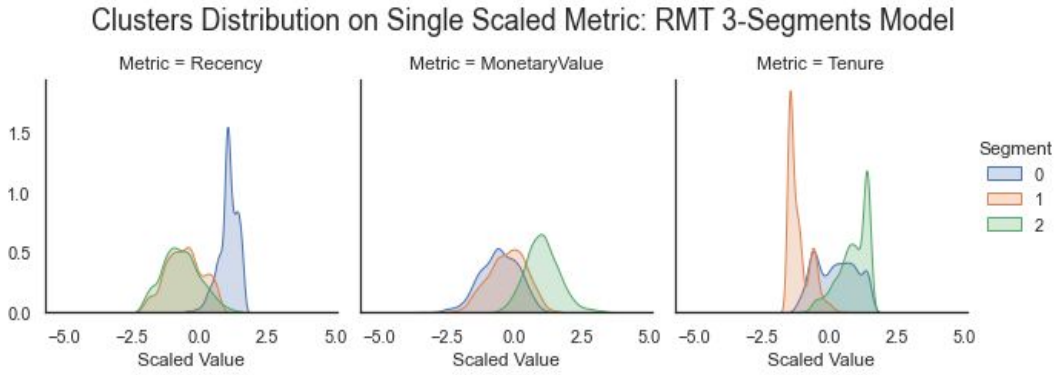


Basic Solution: 3-Segment Model Based on RFM/RMT

- **2D kde Plot of RFM Model**

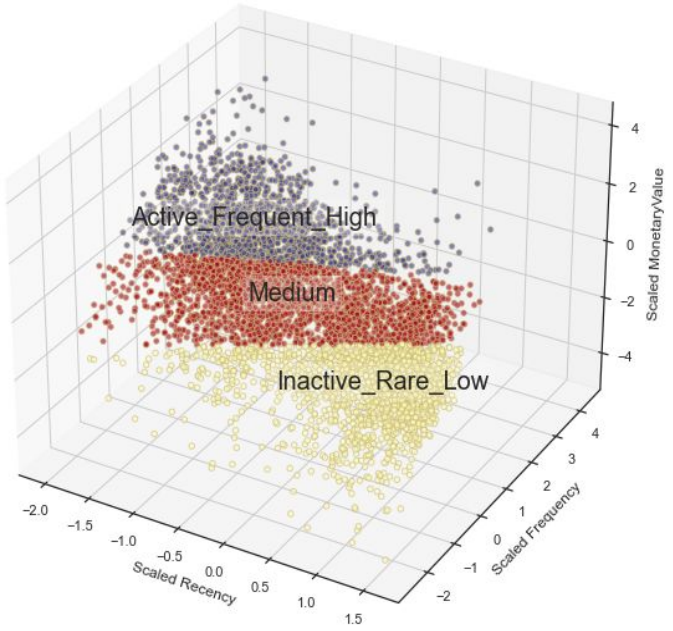


- **2D kde Plot of RMT Model**

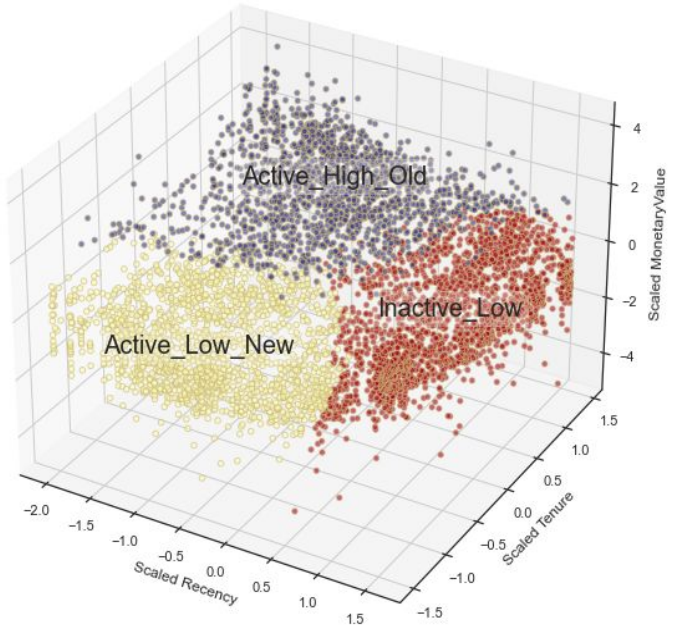


Basic Solution: 3-Segment Model Based on RFM/RMT

- **Scatter Plot of RFM Model**

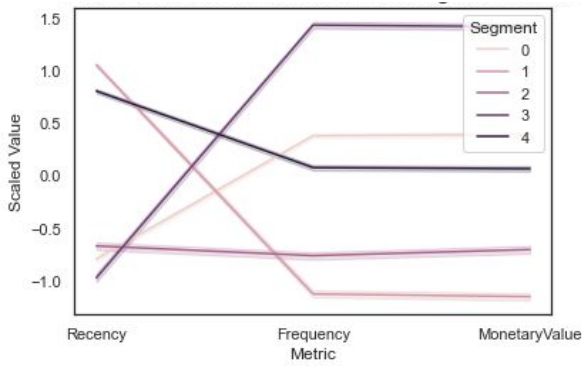


- **Scatter Plot of RFM Model**

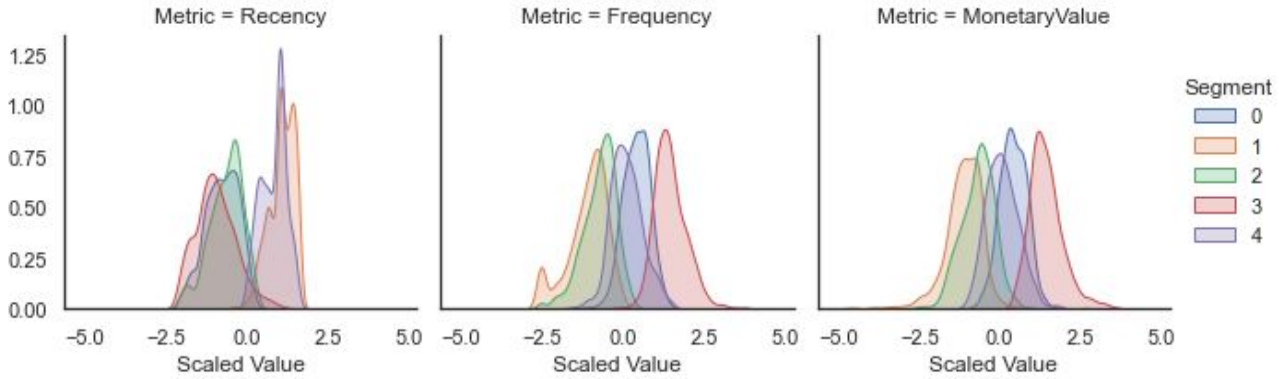


Improved Solution: 5-Segment RFM Model

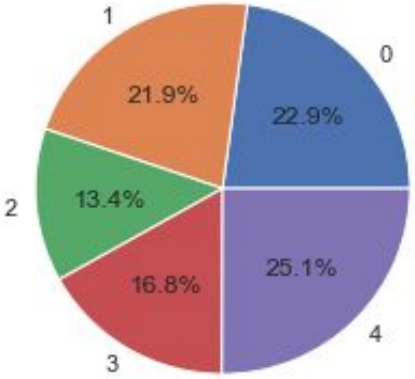
- **Snake Plot**



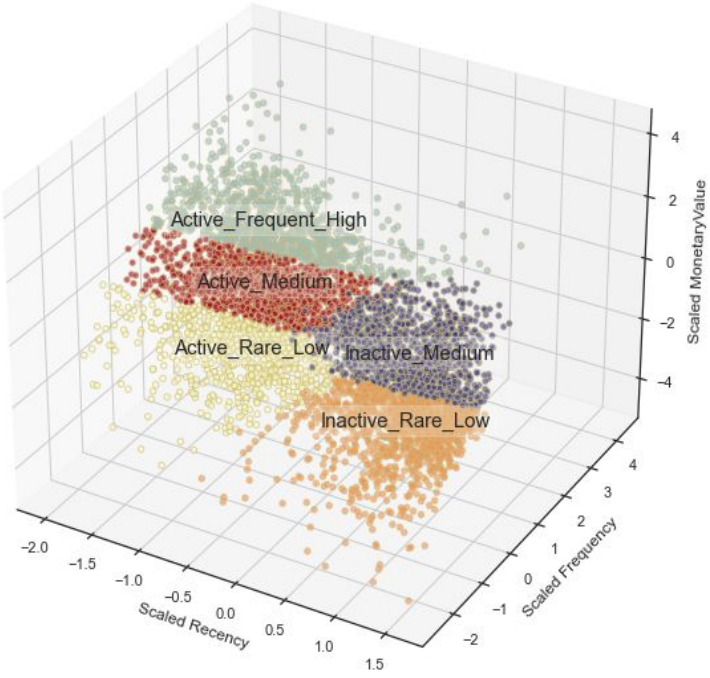
- **2D kde Plot**



Improved Solution: 5-Segment RFM Model



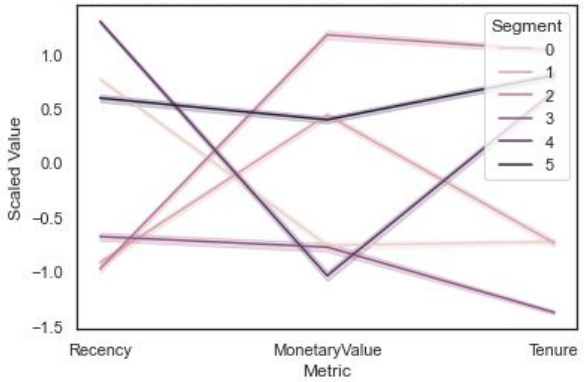
- **Segment Size Ratio**



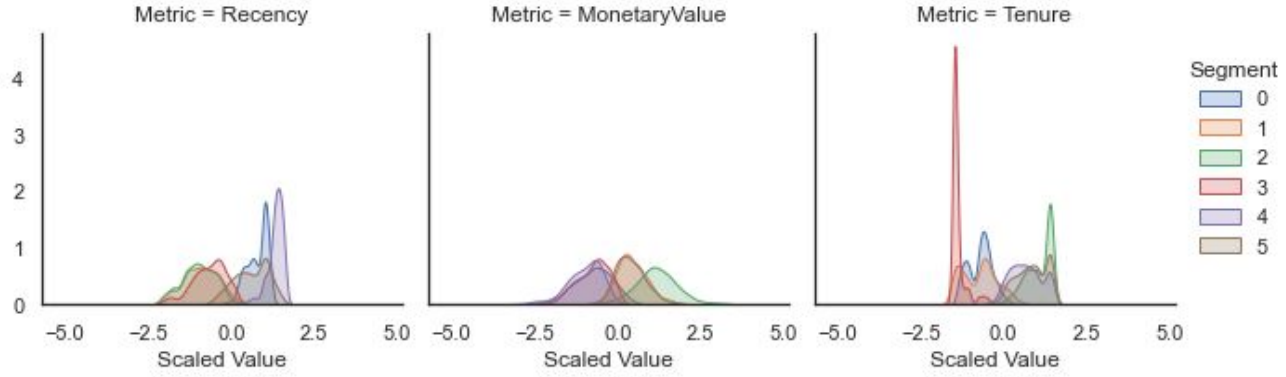
- **3D Scatter Plot**

Best Solution: 6-Segment RMT Model

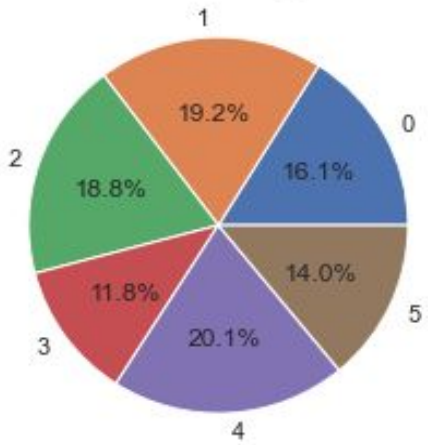
- **Snake Plot**



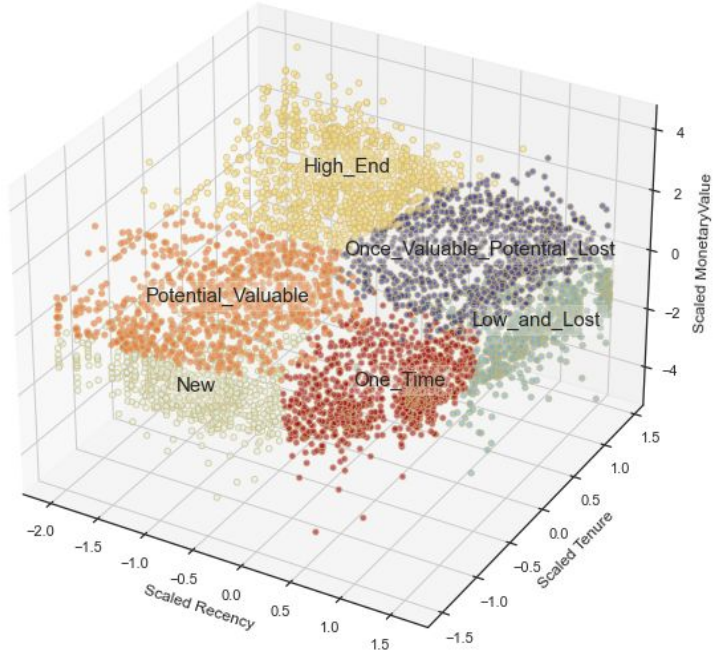
- **2D kde Plot**



Best Solution: 6-Segment RMT Model



- **Segment Size Ratio**



- **3D Scatter Plot**

Best Solution: 6-Segment RMT Model

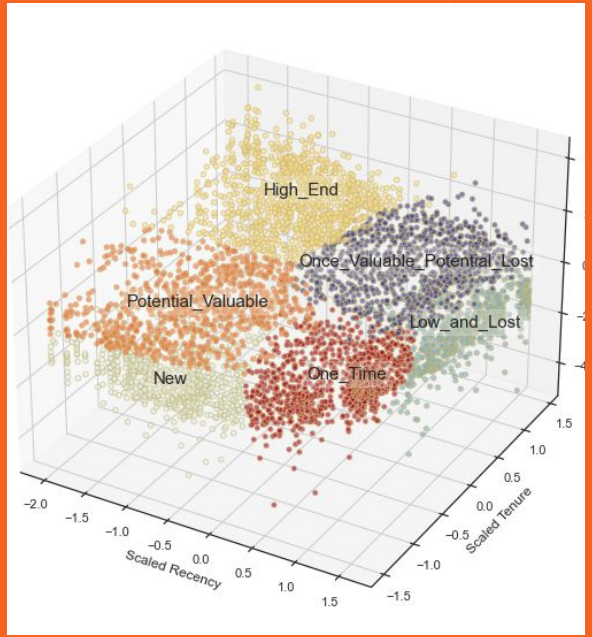
- **Summary Statistics**

	Recency					MonetaryValue			
	mean	median	min	max	std	mean	median	min	max
Segment									
0	315.0	358	37	576	111.0	528.0	432.0	96.0	4512.0
1	31.0	18	1	521	41.0	9286.0	4600.0	462.0	564101.0
2	33.0	23	1	227	32.0	1604.0	1126.0	211.0	44534.0
3	90.0	58	1	326	83.0	262.0	212.0	3.0	1862.0
4	259.0	231	6	730	172.0	1787.0	1276.0	96.0	77347.0
5	538.0	557	42	730	135.0	236.0	195.0	3.0	1437.0

	Tenure					Size
	std	mean	median	min	max	std	
Segment							
0	387.0	388.0	401	105	587	97.0	857
1	24047.0	663.0	686	93	730	79.0	1020
2	2354.0	285.0	297	1	659	170.0	998
3	196.0	133.0	84	1	639	122.0	628
4	2778.0	651.0	658	403	730	64.0	1066
5	169.0	591.0	603	350	730	98.0	744

Best Solution:
6-Segment RMT Model

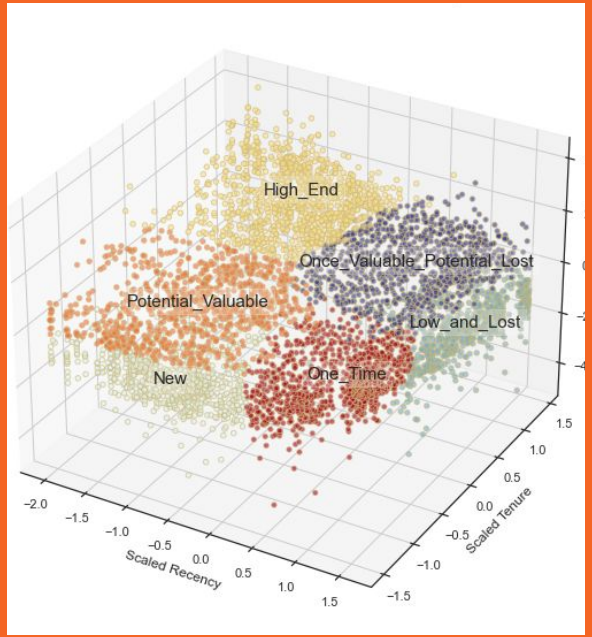
➔ Segment 0: **One_Time.** The customers within this segment purchased products about one year ago, for the first and last time.



Best Solution:
6-Segment RMT Model

➔ Segment 1: **Potential_Valuable.** The customers within this segment are relatively new but have generated considerable revenue. Besides, this segment is pretty active.

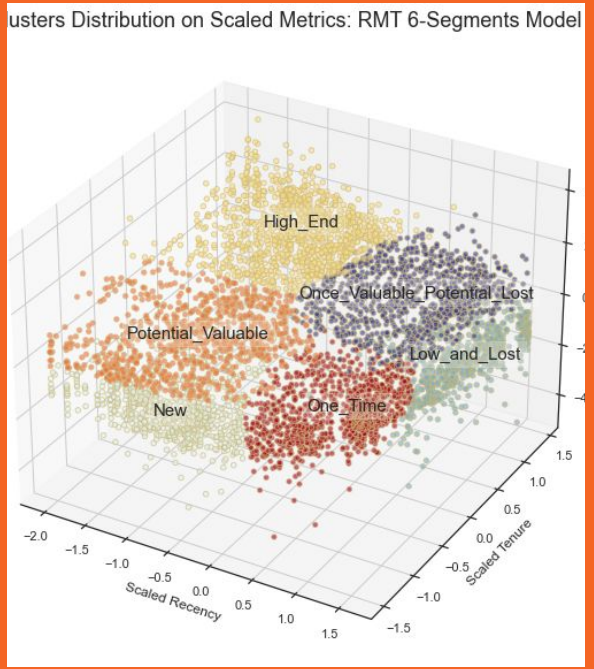
This segment is worthy to highlight



Best Solution:
6-Segment RMT Model

➔ Segment 2: **High_End**. The customers within this segment are active, loyal, and have generated much revenue.

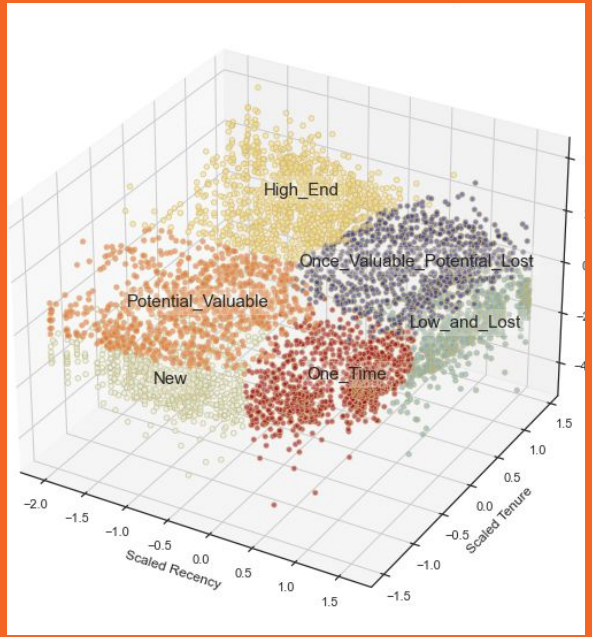
The most important segment for the company



Best Solution:
6-Segment RMT Model

➔ Segment 3: **New**. The customers within this segment are new to the company. The median value of Tenure is only 64 days, which is much lower than the other segments

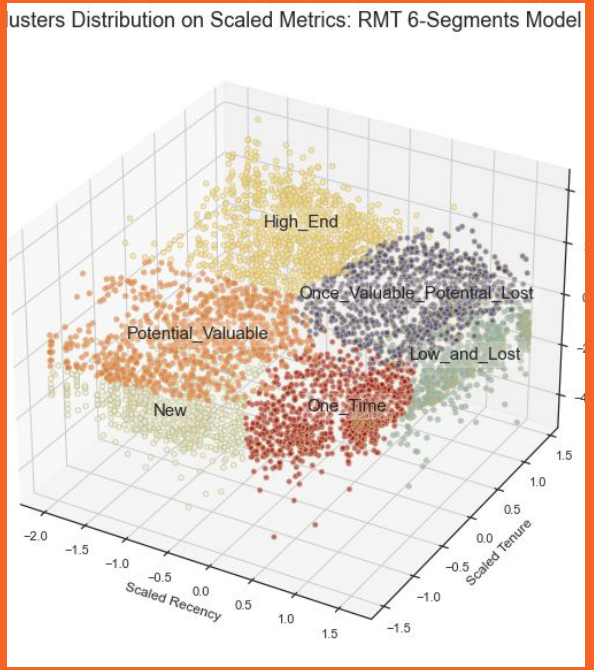
Based on the Time Cohort Analysis, the customers in this segment are easy to churn



Best Solution:
6-Segment RMT Model

➔ Segment 4: **Low_and_Lost.** The customers in this segment generated very low revenue since most of them have not purchased anything for more than one year and a half.

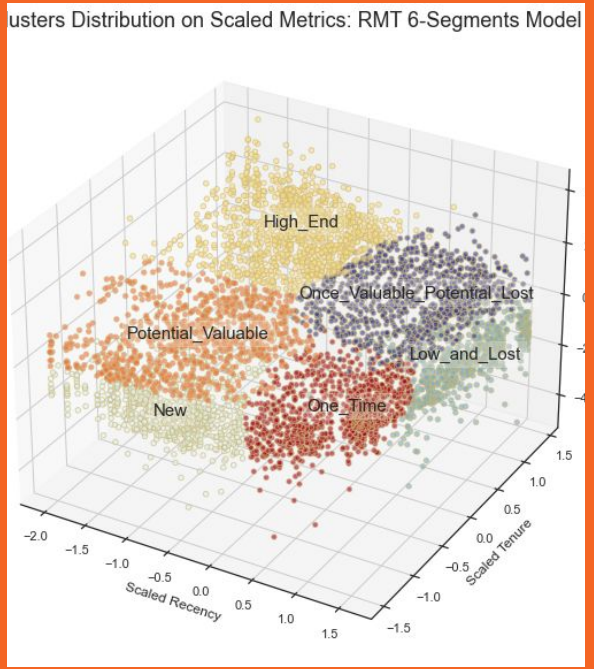
No need to pay close attention to this segment.



Best Solution:
6-Segment RMT Model

- ➔ Segment 5: Once_Valuable_Potential_Lost. The customers within this segment did not purchase recently but they have generated pretty high revenue.

The company may take special actions to prevent these customers from being really lost.



Future Work

- In this project, we segment customers using KMeans. We will try to identify meaningful segments using other clustering algorithms such as Non-negative Matrix Factorization (NMF) or Hierarchical Clustering
- Except for customer segmentation, we will do market basket analysis, product segmentation, Customer Lifetime Value prediction, next transaction prediction, customer churn prediction, and so on.