RFM-AR Customer Segmentation Analysis

IYKRA Data Fellowship Batch 6 - Dataloper

Abiyyu Fathin Derian
Alifia Cesarina Harmadi
Dhea Fajriati Anas
Hendri Prabowo
Nikolas Rakryan Widagdo

Table Of Contents

01

Quality of Dataset

About The Dataset

- The dataset is a record of transaction in Retail
- The dataset records the transaction from 2011 until 2014
- The dataset records day to day transaction of the customer
- The dataset has 23053 rows and 14 columns

Dataset Combination

Customer Dataset

It contains the demographic and geographic of the customer, such as gender, date of birth, and city.

Transaction Dataset

It contains behavioural dataset, such as transaction date, quantity, price, amount, etc.

Dataset Preprocessing

Data Type Conversion

Convert date variable to datetime type

Null Value Dropping

Drop null value due to small number of dataset

Absolute Number Conversion

Convert negative value of numerical dataset into positive

02 Segmentation Processes

Metrics of Customer Segmentation

Recency

Substrate the inspection date to the last transaction date

Frequency

Calculate from count of transaction of each customer

Monetary

Calculate the sum of transaction amount

Age

Substrate the inspection date to the first transaction date

Return

Calculate the combination of customer id and transaction date using diff

RFM Process Activities

01 Metric of Return

Drop duplicate of CustID and Transaction Date

02 Customer Transaction

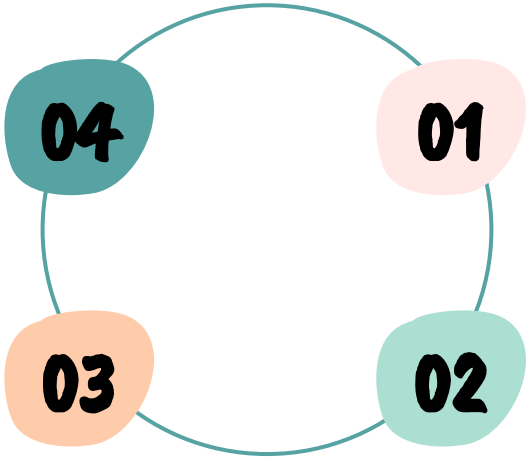
Keep the latest transaction (including the first transaction for new customers)

03 RFM Scoring

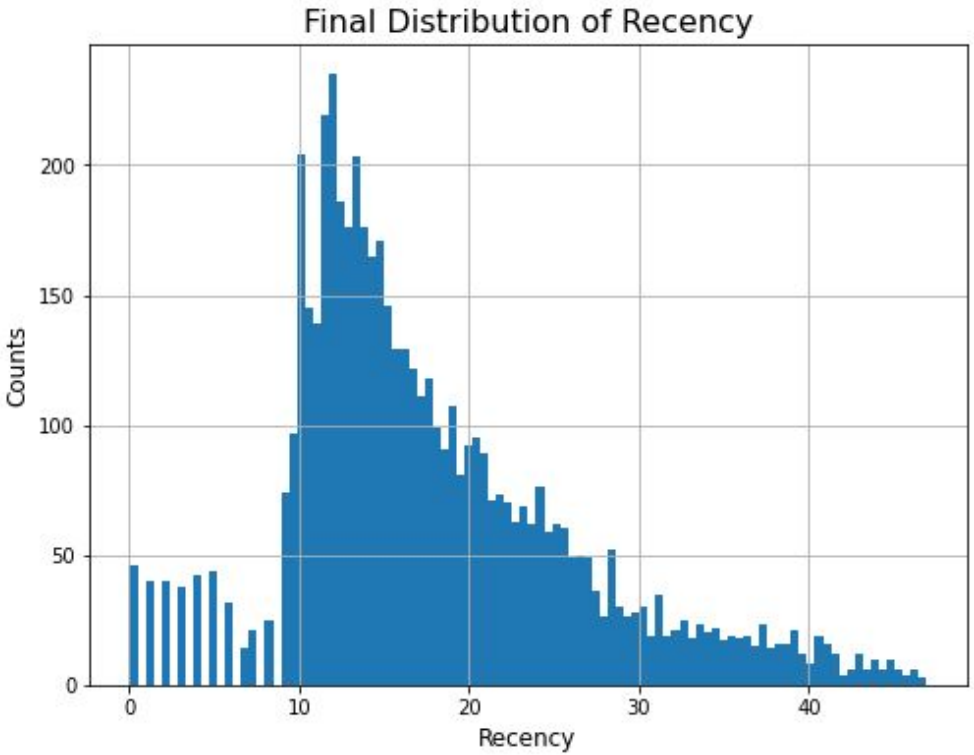
Put label 1-3 by using qcut to return bin in equal variable value, then combine to all metric

04 Customer Labelling

Cluster the customer label based on RFM scoring

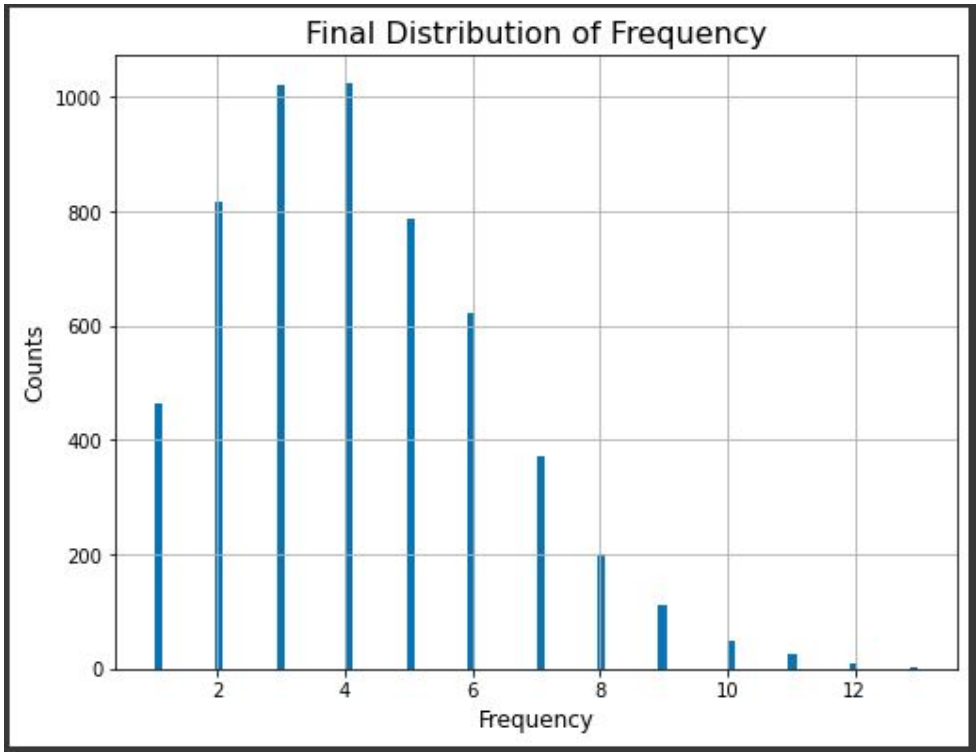


Recency Distribution



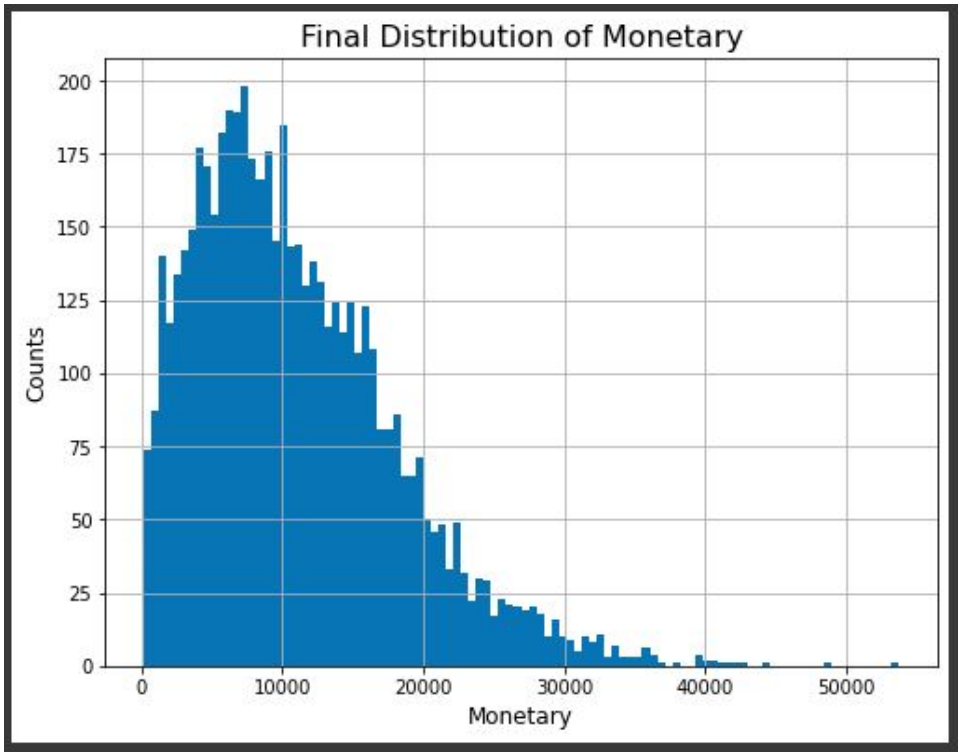
The Recency is accumulated in a range 300 until 800 days.

Frequency Distribution



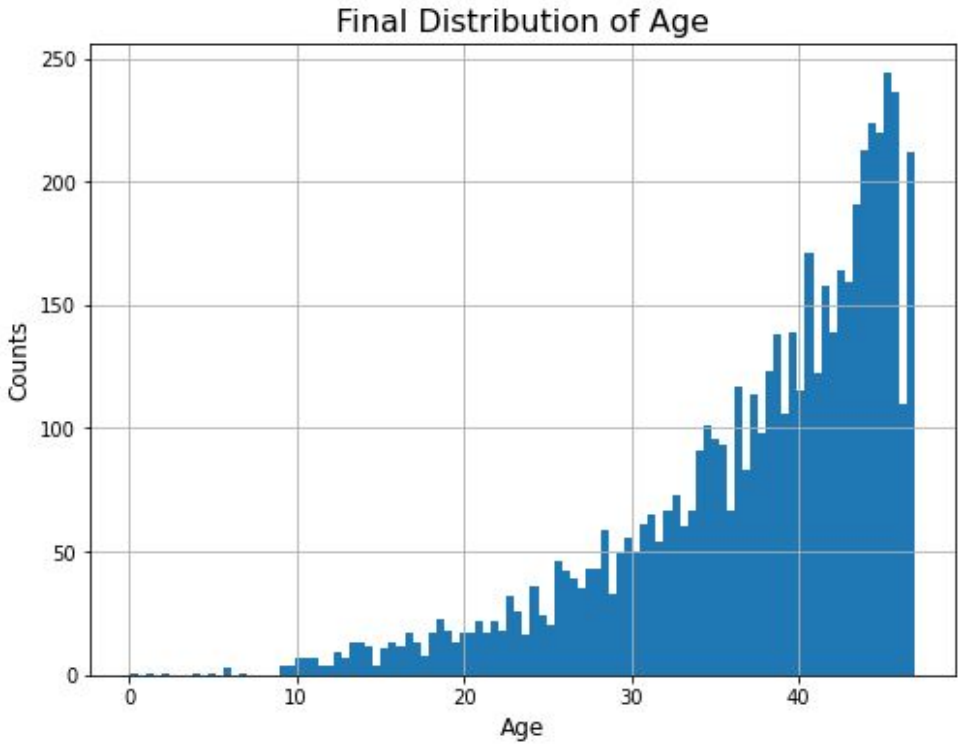
The Frequency is accumulated in a range 1 until 6 times with a right-skewed distribution.

Monetary Distribution



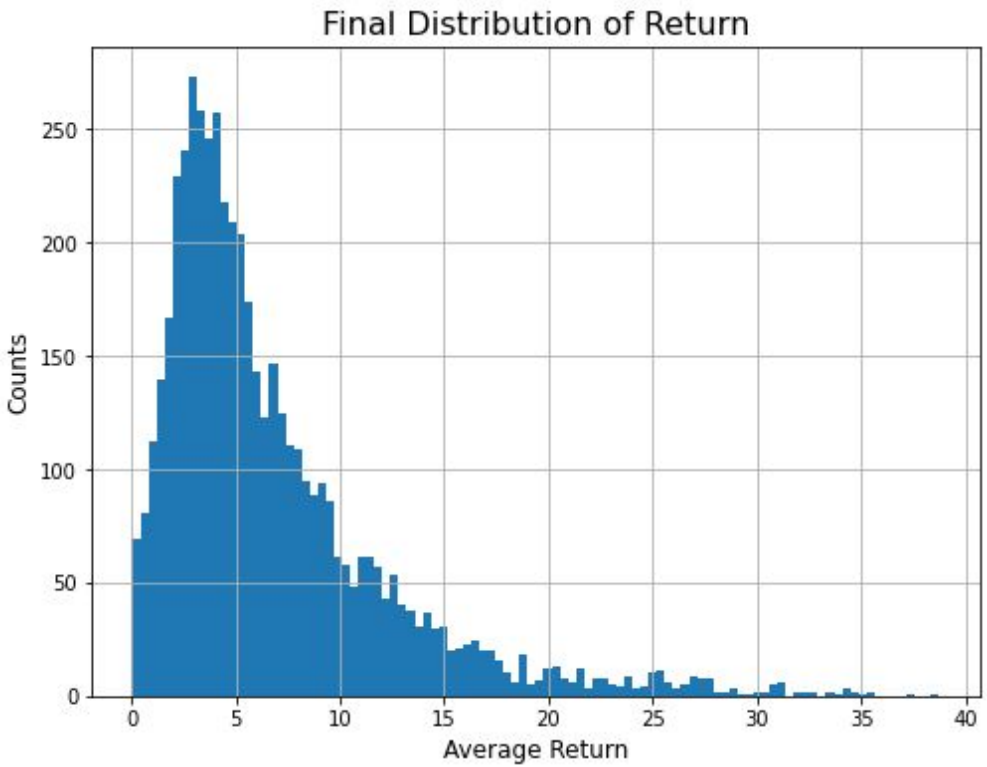
The Monetary is accumulated in a range 0 until less than 2000 dollar with a right-skewed distribution.

Age Distribution



The Age is accumulated in a range 1200 until 1400 days

Return Distribution

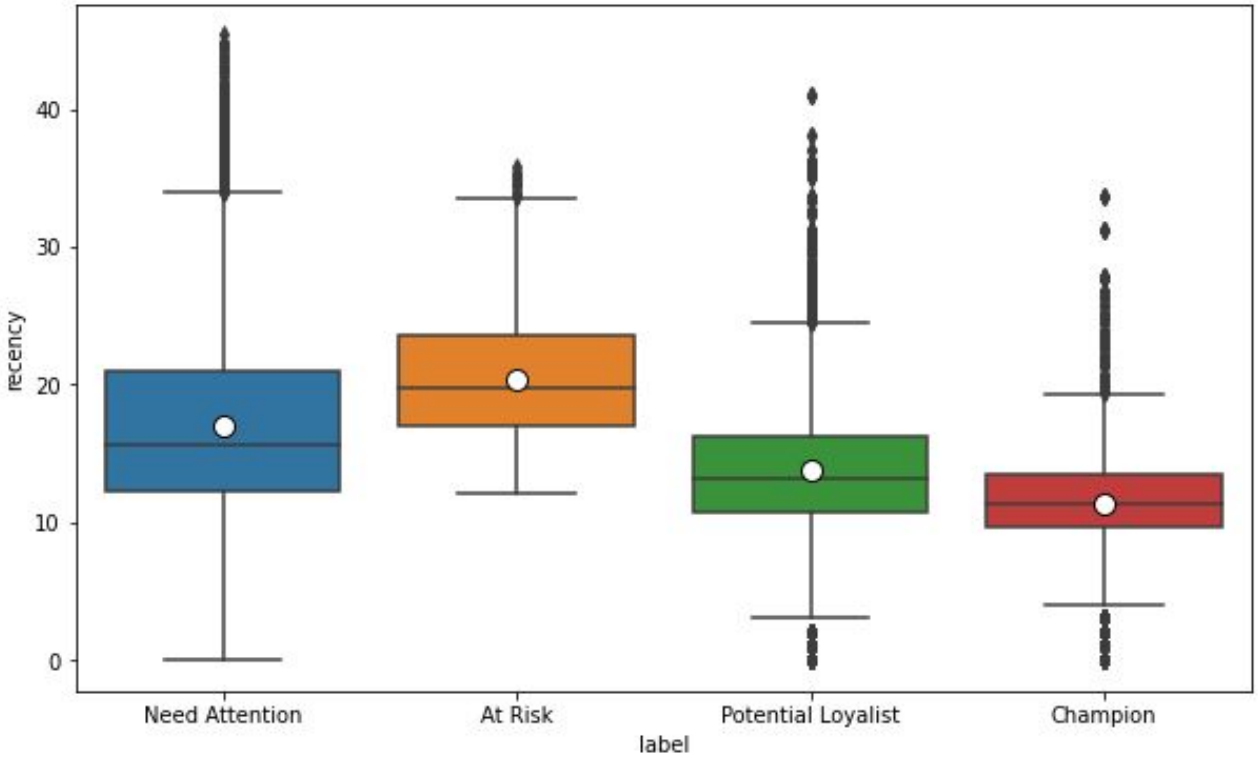


The Return has an impulsive value around 0 until less than 20 days

03

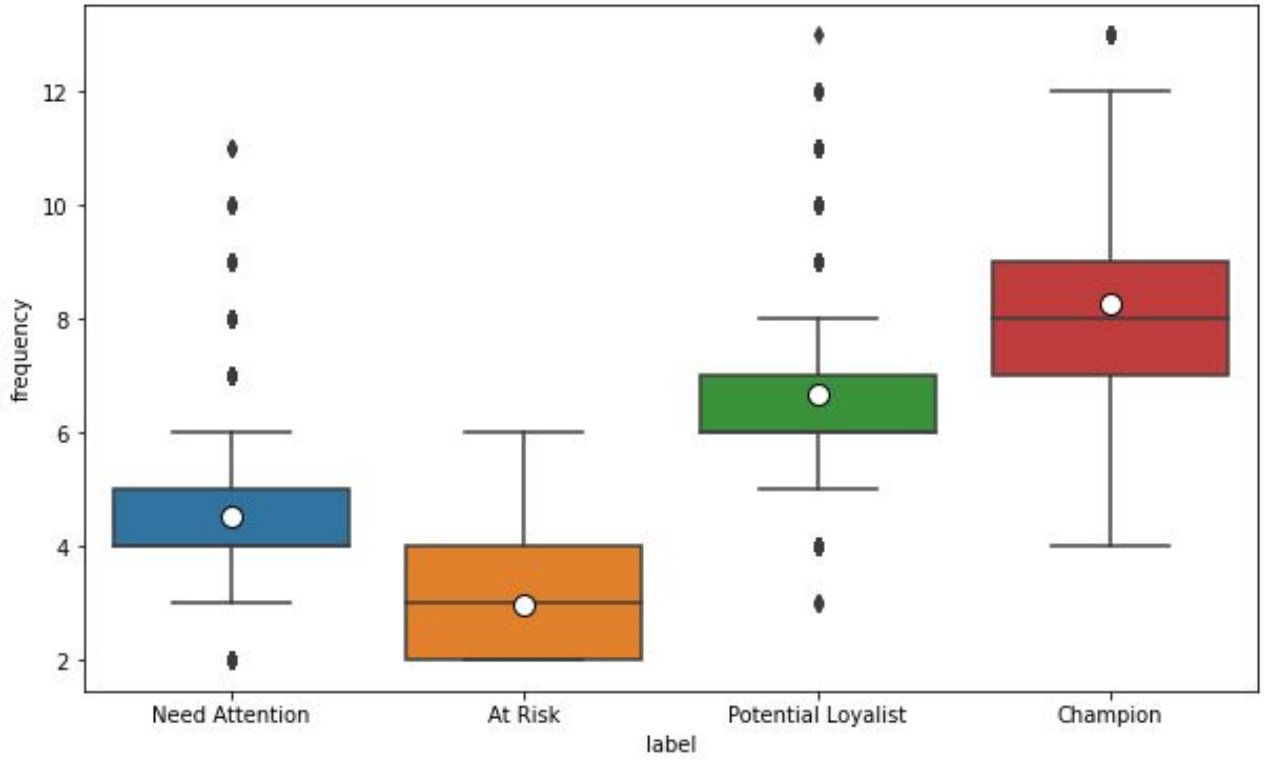
Distinction of the clusters

Recency Distribution in Each Cluster



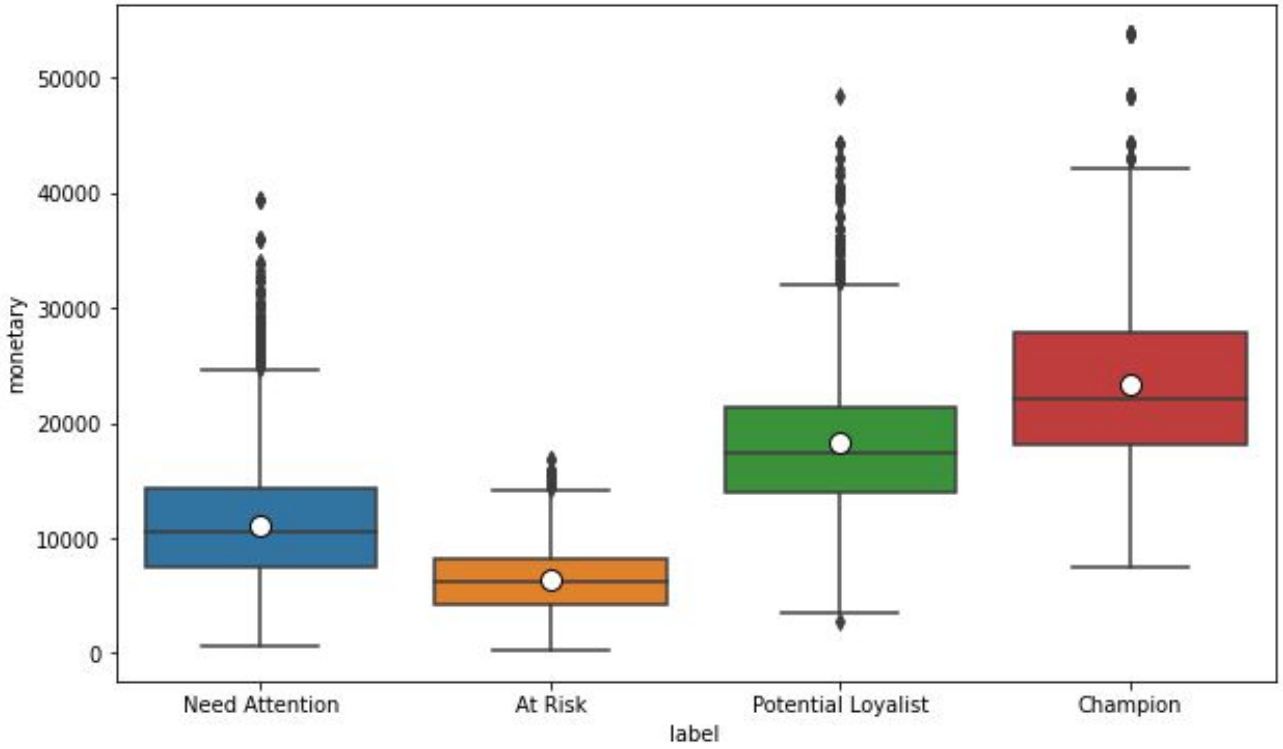
The Recency distribution is quite overlapping each other clusters

Frequency Distribution in Each Cluster



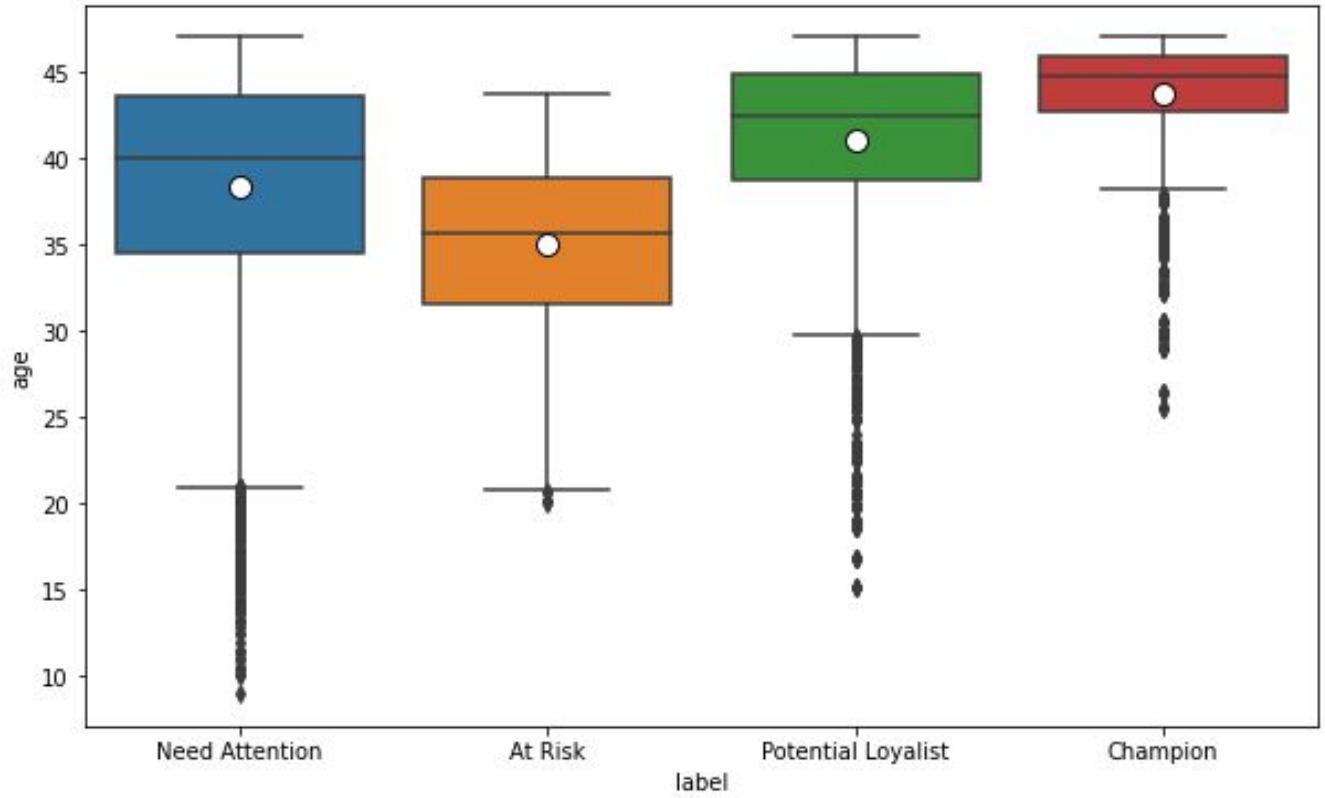
The Frequency distribution is not overlapping each other clusters

Monetary Distribution in Each Cluster



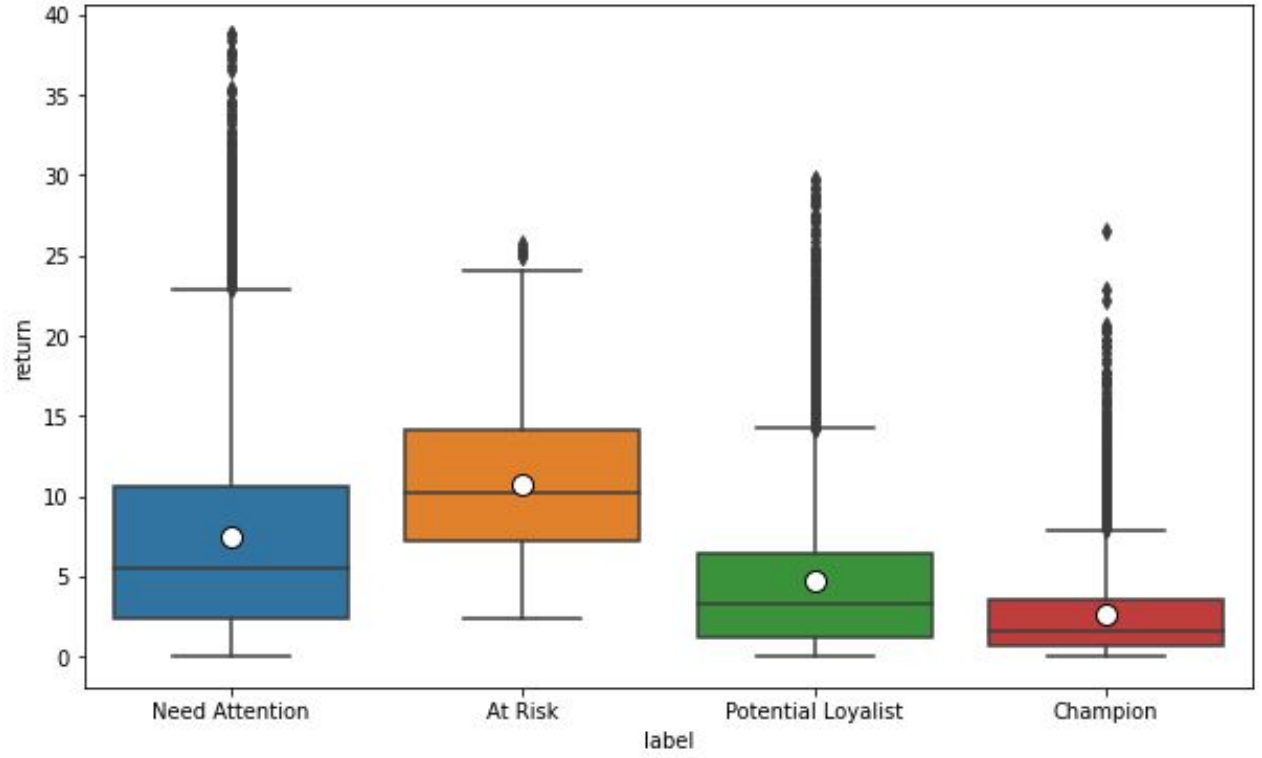
The Monetary distribution is not overlapping each other clusters

Age Distribution in Each Cluster



The Age distribution is quite overlapped each other clusters

Return Distribution in Each Cluster



The Return distribution is quite overlapped each other clusters

04

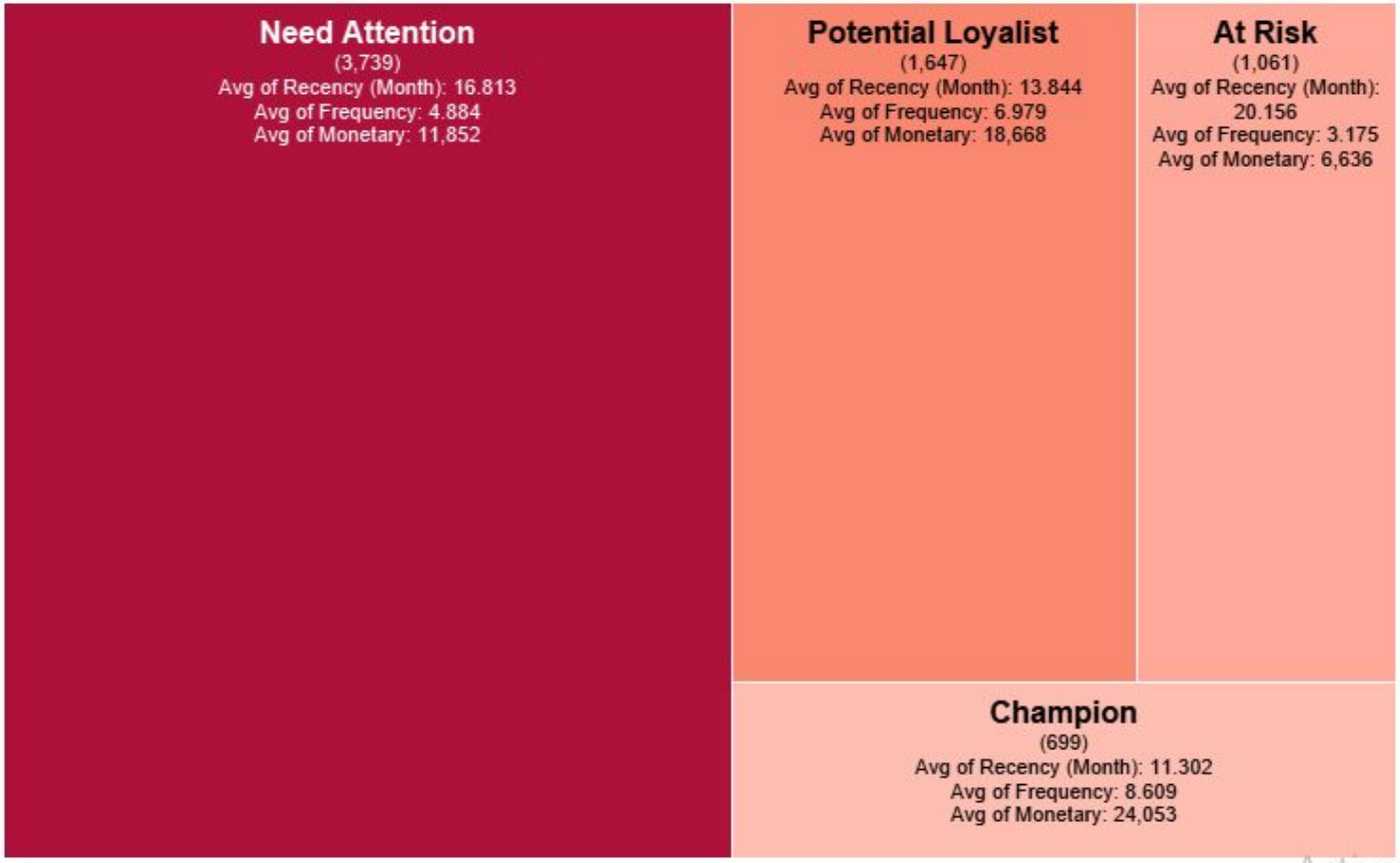
Targeted campaign based on cluster behavior

Characteristic of Each Segment

	label	mean							amax						
		age	frequency	monetary	recency	rfm_score	total_amount	umur	age	frequency	monetary	recency	rfm_score	total_amount	umur
0	At Risk	35.285970	3.174565	6635.510503	20.156379	5.806426	2147.987163	32.569746	43.762706	6.0	16834.675	35.844679	6.0	8265.400	44
1	Champion	43.765553	8.608896	24053.311051	11.302022	13.566702	2818.225845	32.985046	46.982484	13.0	53772.615	33.676256	15.0	8287.500	44
2	Need Attention	38.610209	4.884062	11852.015383	16.813496	8.787958	2470.065812	32.905297	46.982484	11.0	39433.030	45.372595	10.0	8287.500	44
3	Potential Loyalist	40.996239	6.978691	18668.445870	13.844477	11.496919	2731.427336	32.948941	46.982484	13.0	48425.520	40.937186	12.0	8281.975	44

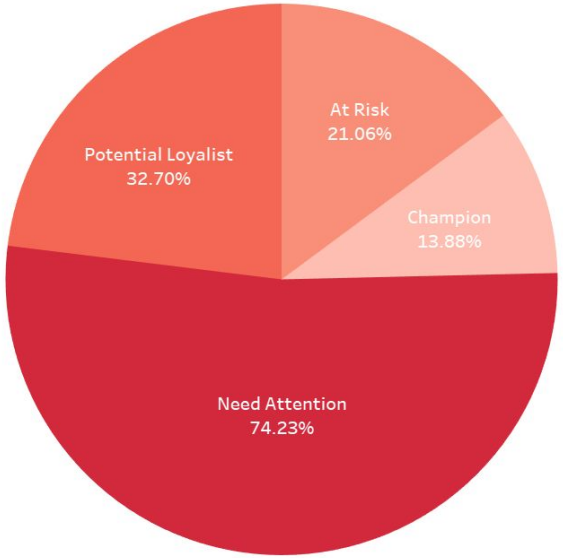
	label	amin							median						
		age	frequency	monetary	recency	rfm_score	total_amount	umur	age	frequency	monetary	recency	rfm_score	total_amount	umur
0	At Risk	20.008624	2.0	269.620	12.156307	5.0	80.665	21	35.778969	3.0	6445.465	19.581511	6.0	1607.775	32
1	Champion	25.495390	4.0	7453.225	0.000000	13.0	83.980	21	44.814062	8.0	22603.880	11.334935	13.0	2375.750	33
2	Need Attention	9.035093	2.0	601.120	0.000000	7.0	77.350	21	40.082959	5.0	11247.795	15.573215	9.0	1907.230	33
3	Potential Loyalist	15.211811	3.0	2718.300	0.000000	11.0	77.350	21	42.349946	7.0	17687.735	13.240518	11.0	2243.150	33

RFM TreeMap


Need Attention
(3,739)
Avg of Recency (Month): 16.813
Avg of Frequency: 4.884
Avg of Monetary: 11,852
Potential Loyalist
(1,647)
Avg of Recency (Month): 13.844
Avg of Frequency: 6.979
Avg of Monetary: 18,668
At Risk
(1,061)
Avg of Recency (Month): 20.156
Avg of Frequency: 3.175
Avg of Monetary: 6,636
Champion
(699)
Avg of Recency (Month): 11.302
Avg of Frequency: 8.609
Avg of Monetary: 24,053

Analyzing RFM Segmentation

- **Champions** are customers who bought most often, highest spenders, and most recently.
- **Potential Loyalist** are recent customers with average frequency, and high spenders.
- **Need Attention** are recent customers with low frequent, and good amount.
- **At Risk** are customers who not frequent shopper, low spenders, and haven't purchased recently.



Proportion of Each Segment

Recommended Campaign for Each Cluster

01

Champion

Offer expensive products, such as electronic

02

Potential Loyalist

Offer membership or loyalty programs or recommend related products to upsell them

03

Need Attention

Offer discounts of products, recommend products based on their behavioural activities

04

At Risk

Send E-newsletter about the product that we sell

Thank You!!!