Customer Segmentation project

for an e-learning platform

ZHUORAN LIU | 2022

Overview

Main analysis :

1. Segment customers
2. List target audience for marketing campaign
3. Identify at-risk customers

METRICS	DESCRIPTION	ASSUMPTION
Recency	Days since last customer transaction	Today = 2022-01-01 Lower recency ≈ higher customer stickiness
Frequency	Number of transactions in the last 24 month	high frequency ≈ high loyalty
Monetary Value	Total spend in the last 24 month	high spend brings higher customer life time value

Data pre-processing

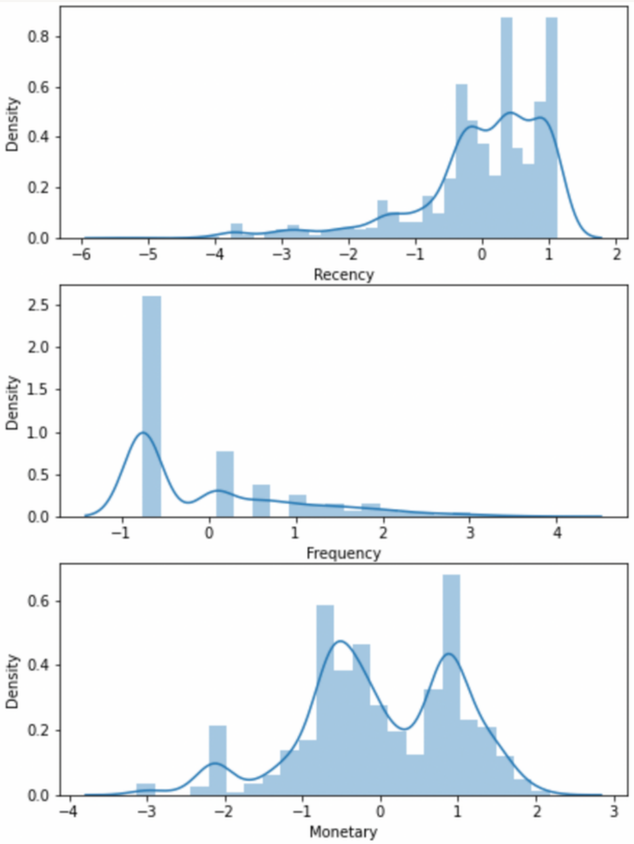
Email	Recency	Frequency	Monetary
il.com	132.0	1	1599
ia.edu	67.0	1	1299
n.edu	451.0	1	1249
il.com	181.0	1	1299
il.com	93.0	1	1499

RFM calculated from transaction data

	Recency	Frequency	Monetary
count	1955.000000	1955.000000	1955.000000
mean	6.067103	0.621148	6.102155
std	0.942053	0.827498	1.286548
min	1.098612	0.000000	2.079442
25%	5.771441	0.000000	5.288267
50%	6.298949	0.000000	5.983936
75%	6.775366	1.098612	7.169350
max	7.126891	3.806662	8.906800

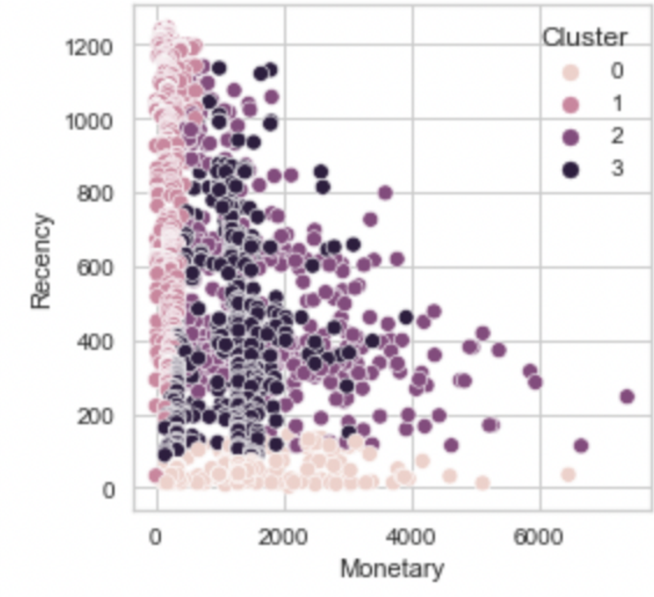
Statistical description (after)

Distribution (after)

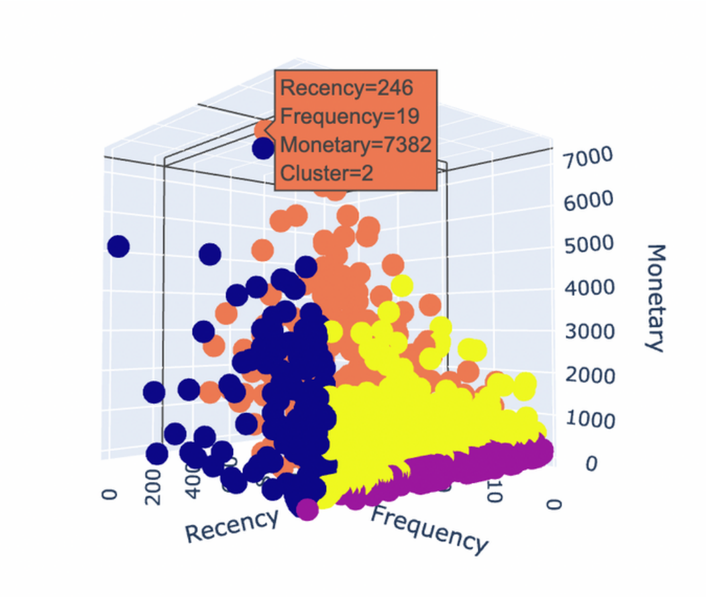


- Retrieve and calculation of RFM metrics using SQL
- Transformation on skewed data
- Data standardisation
- Data normalisation

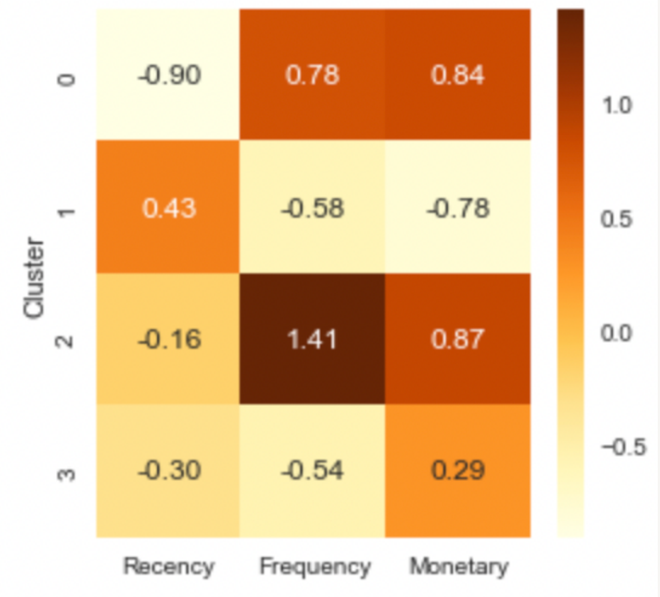
Modeling & Evaluation



Frequency v.s. Monetary Value



3D Scatter Plot of 3 clusters



Relative importance heat map

Customer Cluster Analysis

1 Potential loyal customer (cluster 0)

New customers who spend a lot of money on our platform.
This segment of the customer is the one on which we should **concentrate our marketing efforts.**
Their purchase history reveals a wide range of interests.
We should tailor marketing campaigns to their preferences
in order to entice them to buy more products.

2 Inactive customers

3 Top Customers

4 At-Risk Customers

Customer Cluster Analysis

2 Inactive customers (cluster 1)

This segment is composed of inactive customers :
they showed little interest and the spend per order is quite low.
These people are not a priority for us.
Still, we can push notification of free courses or free trail of star courses in order to keep them.

3 Top Customers

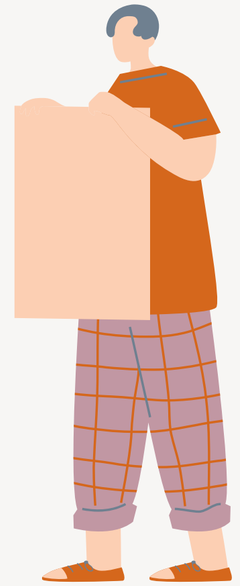
4 At-Risk Customers

1 Potential loyal customer

Customer Cluster Analysis

3 Top Customers (cluster 2)

These people are our long-standing loyal customers :
They buy courses on our platform for a long time.
Based on their interactions with ad campaigns, they are not price sensitive.
Therefore, we should concentrate on providing high-quality products and after-sales service.



4 At-Risk Customers

1 Potential loyal customer

2 Inactive customers

Customer Cluster Analysis

4 At-Risk Customers (cluster 3)

This segment of customers used to spend a lot on our platform, but they haven't repurchased anything in a long time.
Such users are more likely to churn. It's necessary to contact users and identify the causes of user churn first.
In terms of marketing, we can try coupons or discount information to bring them back to our platform.

1 Potential loyal customer

2 Inactive customers

3 Top Customers

Suggestions

For marketing strategy

- Start our new product launch campaigns with top customers and potential loyal customers (segment 1 and 3)
- Send coupons or discount information to at-risk customers (segment 4) to bring them back to our platform.
- Push notification of free courses or free trail of star courses to inactive customers (segment 2)

For customer retention

- Customers in segment 4 are more likely to churn. It's necessary to contact users and identify the causes of user churn, then develop relevant strategies.
- Furthermore, we can conduct cohort analysis to take a closer look at customer churn pattern.

Reflection on the project

Limitation of RFM model

RFM can only be performed on historical data and historical performance. We assume that there is no difference between the before and after behaviour of the user, which is unrealistic.

Include / change metrics

Adding extra variables to RFM model or replace certain metrics depending on business context is necessary.
In our case, we can take customer register time into account.

Data Quality

Dirty data was always the number one challenge, especially the data completeness, which influence both the usability and the reliability of our model.

Thanks for listening!