E-commerce General Customers Clustering

Aldy Budhi Iskandar

ALDY BUDHI ISKANDAR
Data Scientist & Machine Learning

Projects

Build Ideal Model for **Predicting** Loan Status Customers •
E-Commerce General Customers **Clustering** •
Video Game Sales **Exploratory Data Analysis** •
Shipping Data **Exploratory Data Analysis** •
Python **Data Cleansing** •

Educational

Data Science
May – Sept 2022 at **Dibimbing.id**

Mechanical Engineering
2012 - 2016 at **Polman Bandung**

I'm a Data Scientist who interest in Math Logic and Statistics with a background as an Engineer in several manufacturing companies.

For programming skills, I prefer to use Python programming language, SQL query, & Tableau for visualization, because that's related to the software I always use for my projects.

Contents

1 **Background**

Current business problem

2 **Data PreProcessing**

Collection and preparation of required dataset

3 **Modeling**

Do the best modeling to solve existing problems.

4 **Analysis**

Analyze modeling output

5 **Recommendations**

Business insight & recommendations for company

1

Background

Case Study

In this modern era, there are more trade competitors are emerging, therefore service to customers must be improved to retain our customers because creating better relationships with customers will lead to increased company profits. But what efforts should be made to retain as many customers as possible?

Business problem

Why should build good relation with customers?

Customers are assets for all types of companies, especially in the E-commerce field. So it must be maintained.

What strategy to do?

We can provide discounts for new users, attractive price offers, reminders to customers, post products for discounts, buy 2 get 3, etc.

What's the problem?

In order for the efforts to be successful, it requires high funds and effort to provide the best possible treatment to customers.

Current data & problems

Data

There are 68,000+ user data & 56,000+ transaction data recorded only on the web store, not including offline transactions, E-commerce app transactions data, etc. Just only from the web store dataset.

Active user
Must be defended

Not active user
Must be given an attractive offer

Problems

With a large number of customers, it's difficult for E-commerce to divide and determine what treatment is right to offer customers, and surely there are other characteristics of the customer but it can't be defined yet

Objectives

Modeling

Perform customer segmentation (clustering) through the customer dataset. use LRMFC indicators and perform a K-means clustering algorithm.
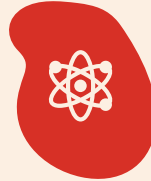
Analysis

Analysis of the characteristics of each cluster resulting from segmentation.

Recommendation

Provide business insight related to the analysis results.

2

Data
PreProcessing

PreProcessing steps

Data Collecting

Data Cleansing

LRFMC Feature

Data Scaling

Dataset information

These datasets consist of:

24,354 rows

24.354 rows of order data where has no missing value in the database

7 features

- user_id
- total_transaction
- total_money_spent
- tran_with_disc
- last_order_date
- created_at
- data_pulled_date

LRFMC Feature

- **L (Length Relation)**

End time of the observation window-time to join (unit: month).

- **R (Recency)**

The time from the last order to the end of the observation window (unit: month).

- **F (Frequency)**

The number of orders in the observation window (unit: times).

- **M (Monetary Value)**

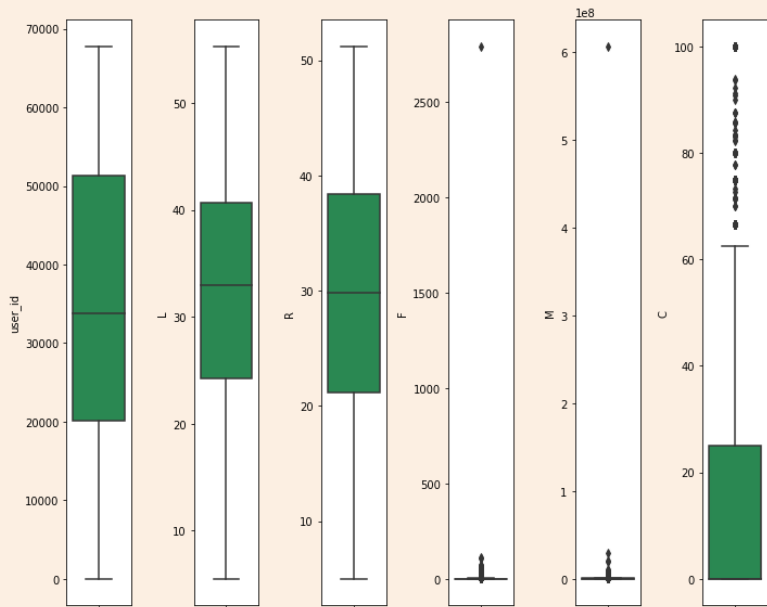The total customer spending money in the observation window (unit: IDR).

- **C (Coefficient Value)**

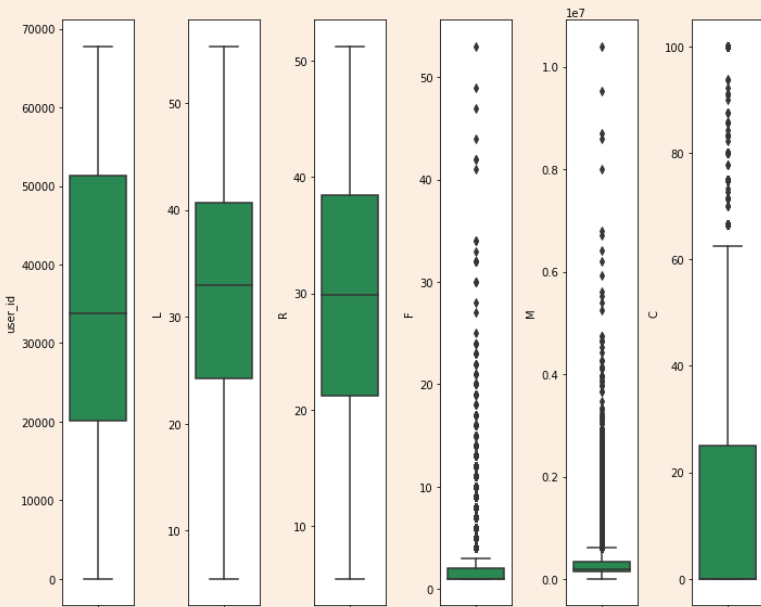Percentage discount rate (unit: percent).

Data Cleansing

Boxplot before process outlier handling

Boxplot after Z-Score process

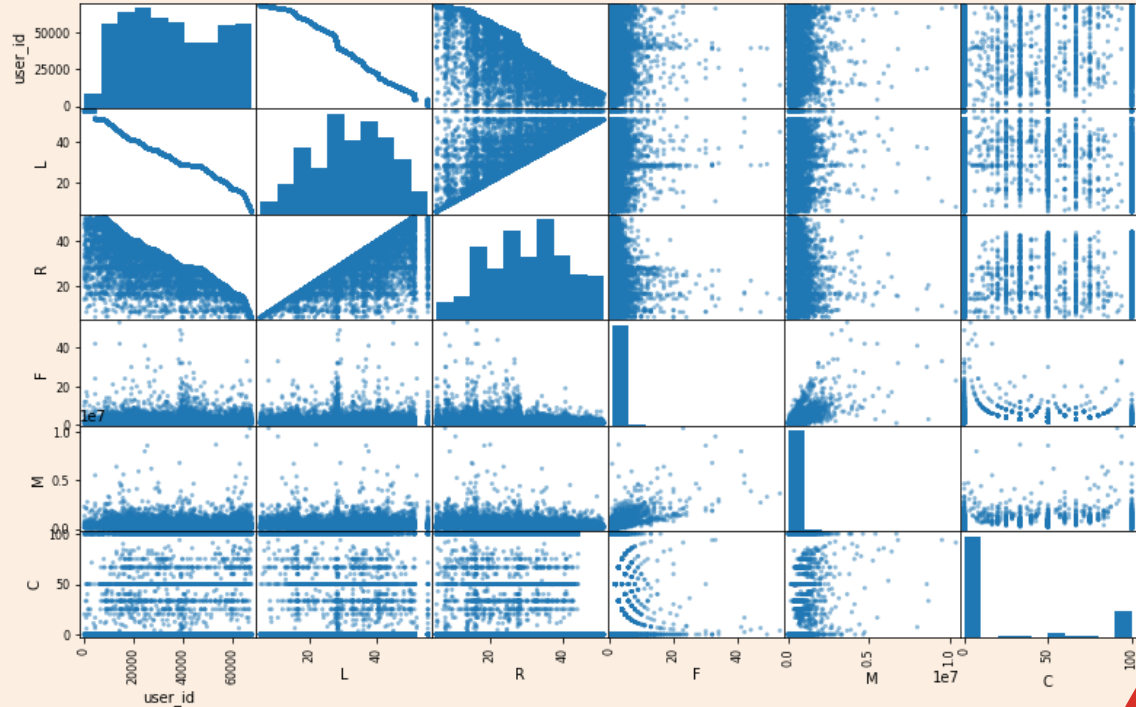**0 Missing value**   **0 Duplicate value**   **13 Outlier value**
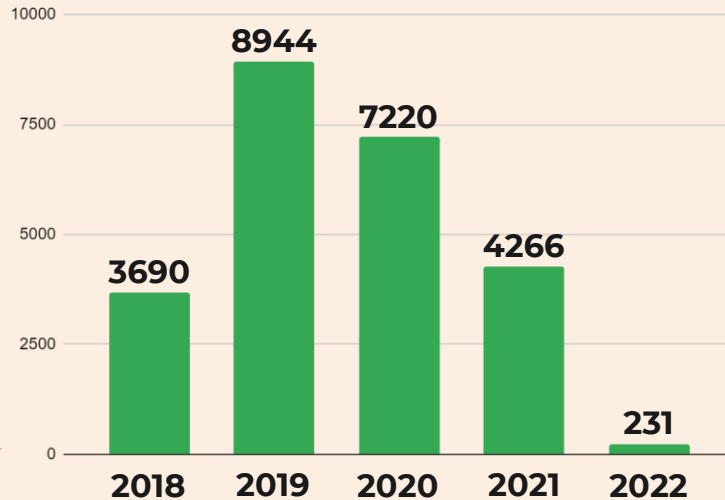
Data Distribution

`L` and `R` columns have a fairly normal distribution, while `F`, `M`, and `C` columns have a positive skew distribution, where more outliers are in the right area.
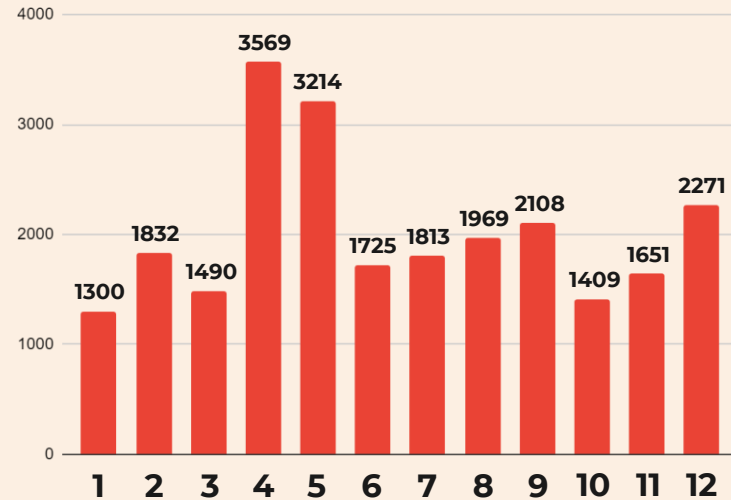
Descriptive & Univariate Analysis

- The average transaction is **1.7 transactions** per user.
- The average money spent is **IDR 332,469** per user.
- Most users join at **April** in every year.
- Most users join in **2019**.

## Year join distribution in the dataset

Year	Count
2018	3690
2019	8944
2020	7220
2021	4266
2022	231

Month join distribution in the dataset

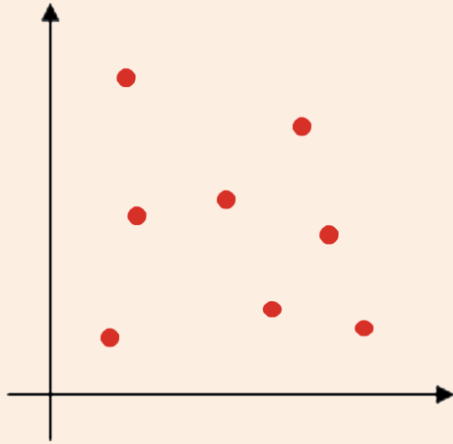| Month	Count
1	1300
2	1832
3	1490
4	3569
5	3214
6	1725
7	1813
8	1969
9	2108
10	1409
11	1651
12	2271

Multivariate Analysis

It can be interpreted that the higher the value of `total transaction`, the higher the value of `total_money_spent` (**directly proportional**).
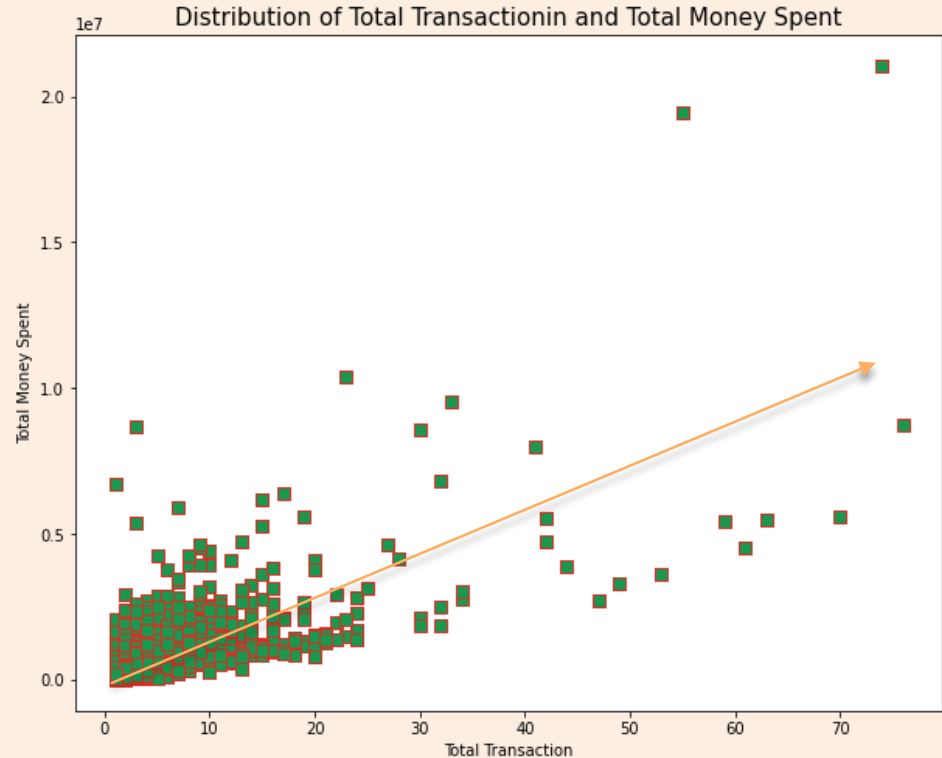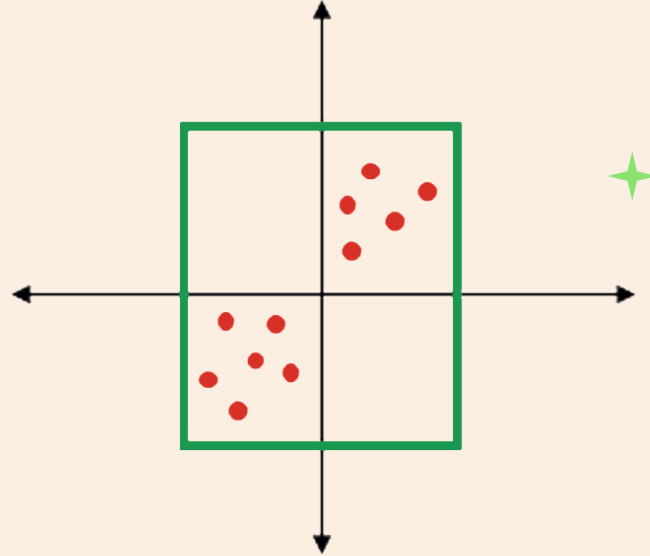
Data scaling

Actual Data

After Standardization

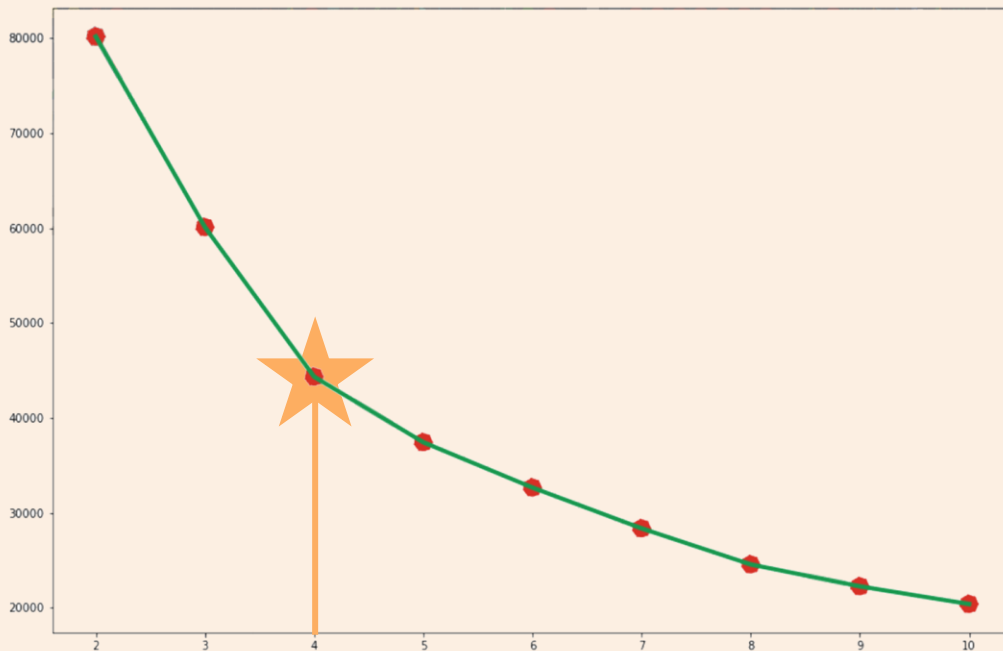StandardScaler to normalize the data so that the data used does not have large deviations.

3

Modeling

Inertia

Before doing clustering, it would be better to determine the best and right number of clusters first.

According to the graphic **(Elbow method)**, the angle change starts to occur at point 4, then the correct K value for K-Means Clustering is **K = 4**.

Score plot

The result of the **score plot** can be taken as the best and ideal value, which is at point 4, or **N = 4**.

Distortion Score Elbow for KMeans Clustering
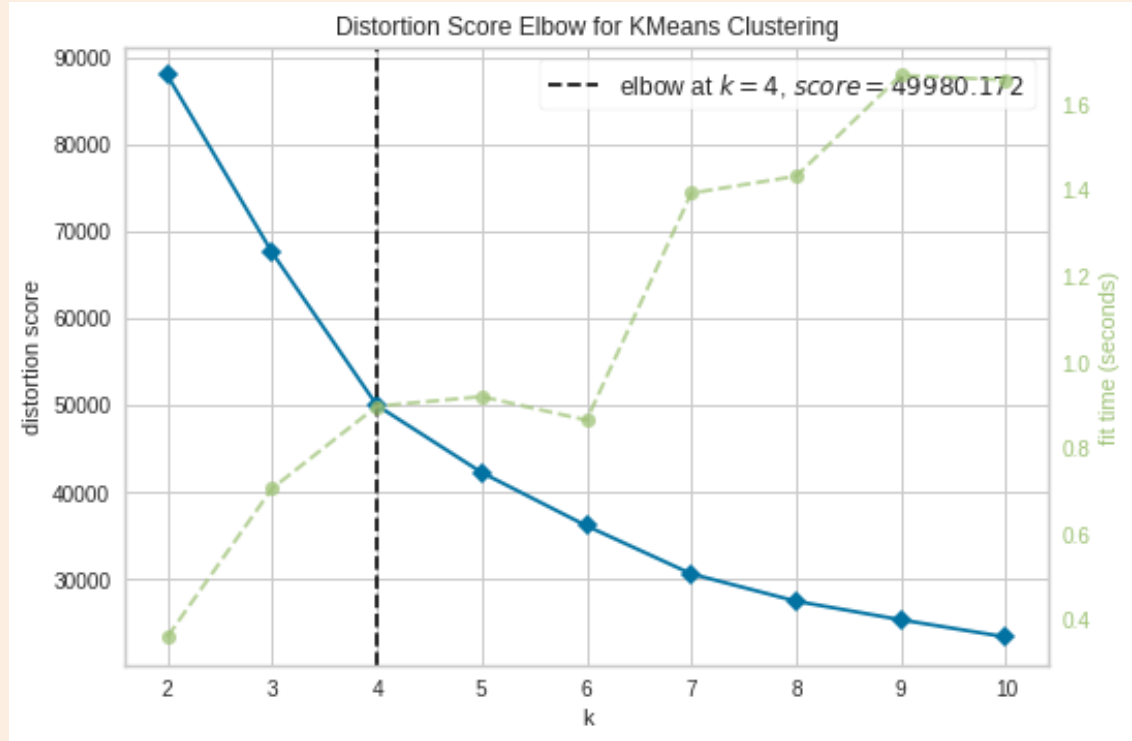
K-Means Clustering

From the results of this customer clustering and visualized with a scatterplot as shown below.

This diagram shows the distribution of customer data which is divided into clusters according to the K-Means Clustering algorithm.
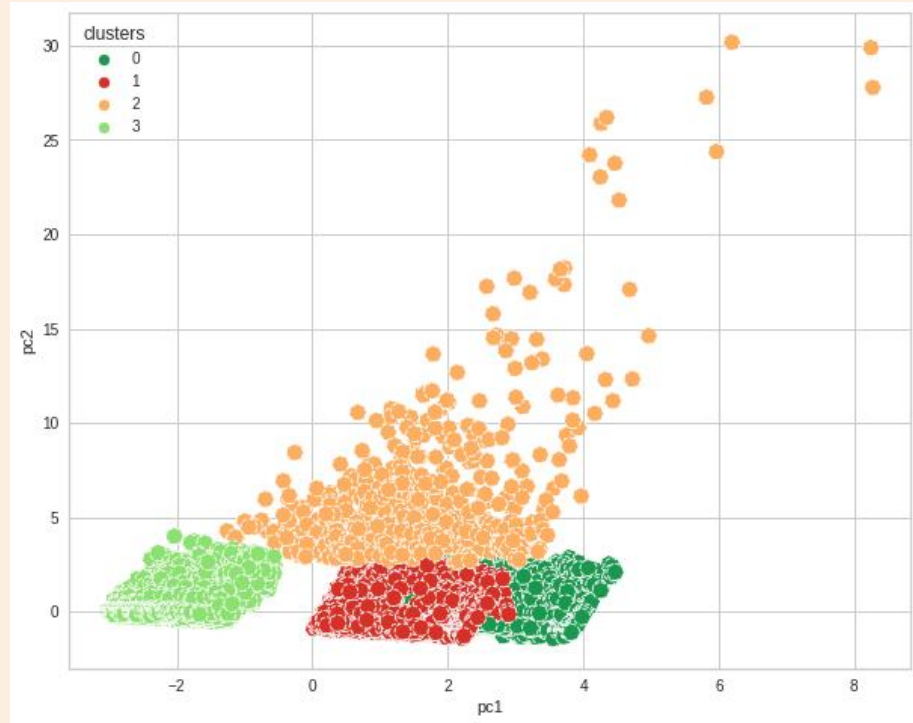
User Cluster Percentage
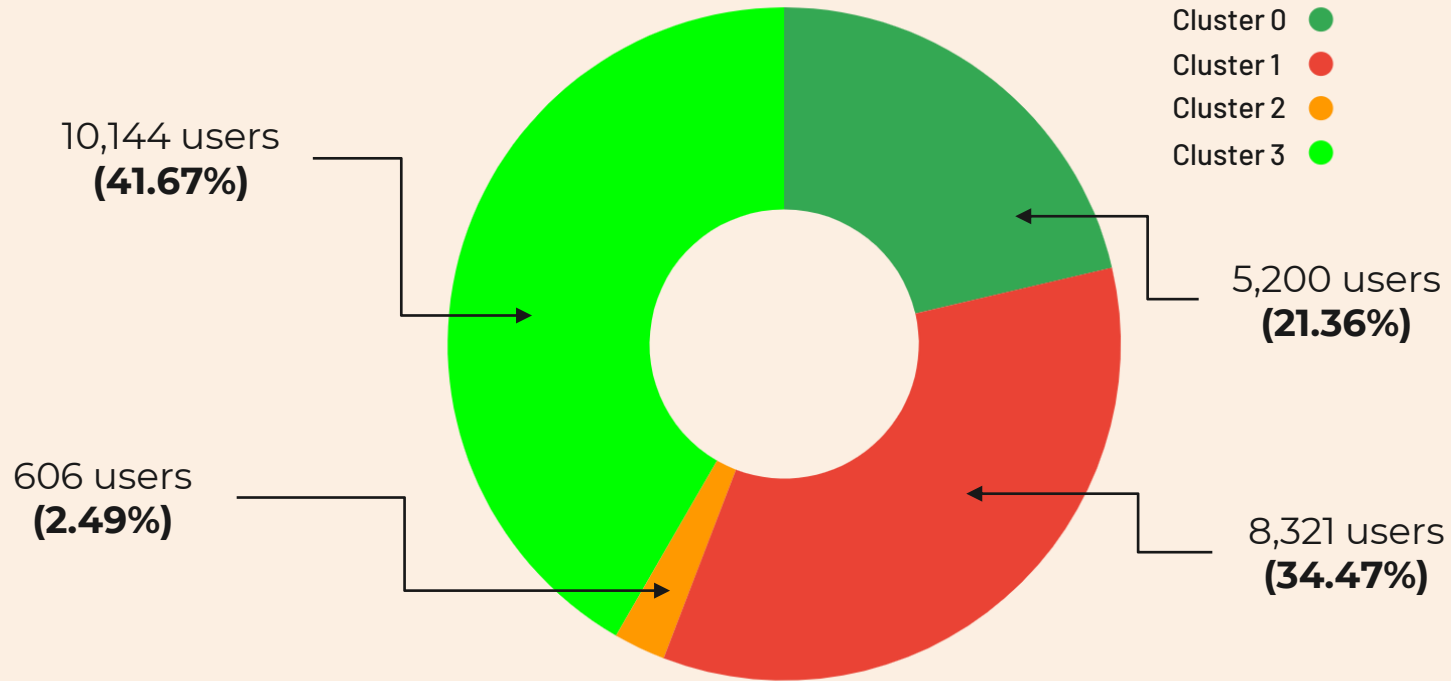
Cluster 0 ●
Cluster 1 ●
Cluster 2 ●
Cluster 3 ●

10,144 users
(41.67%)

5,200 users
(21.36%)

606 users
(2.49%)

8,321 users
(34.47%)

4 Analysis

Radar **chart**

Customers **Scoring**

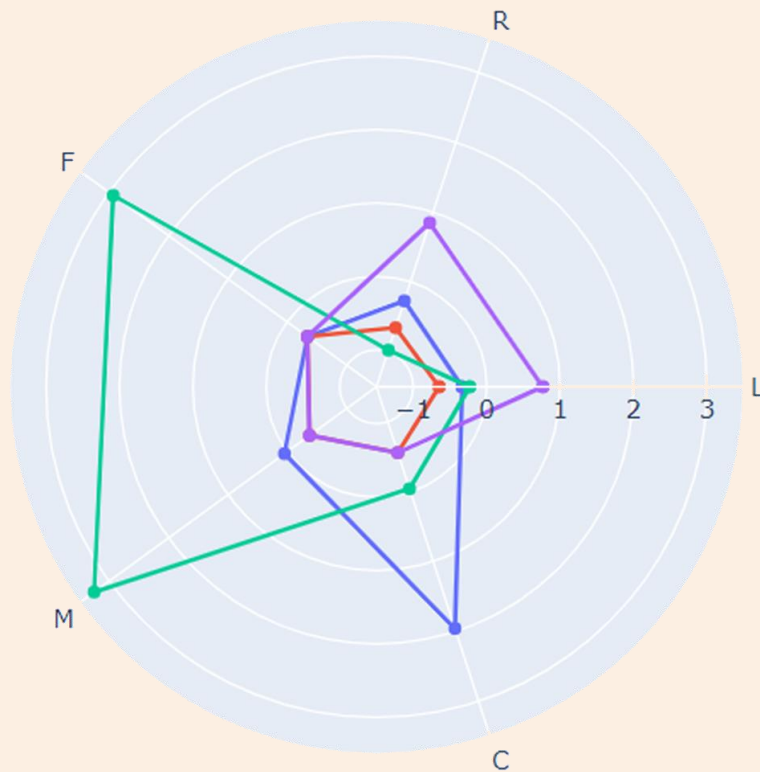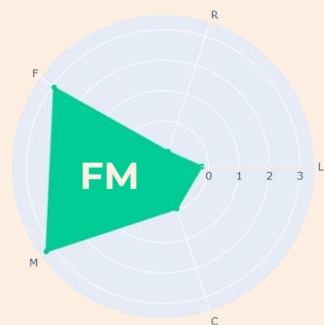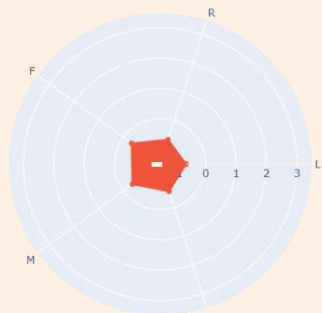| cluster	Length Relation(month)	Recency(month)	Frequency(times)	Monetary(IDR)	Coef_Value	M/F(IDR)	Count	%
0	27.677647	25.805314	1.461538	4.109494e+05	94.654167	281175.871447	5200	21.363132
1	23.450224	22.123096	1.503516	2.192185e+05	1.272196	145803.906706	8391	34.472700
2	31.093894	20.576018	9.470297	1.689753e+06	33.757443	178426.600802	606	2.489627
3	41.632177	39.709286	1.331920	2.348089e+05	1.589610	176293.531123	10144	41.674541

Cluster	Strong points	Average points	Weak points
0	C	M,R	F,L
1	-	-	L,R,F,M,C
2	F,M	C,L	R
3	R,L	-	F,M,C

Customer clusters

The Discount-Hunter

New customers who only make transactions if there is a voucher/promo

The Potential-Loyalist

new customers who have just made their first transaction

The Superior-Buyer

Average customers who frequently make transactions and do not hesitate to spend money on shopping

The Hibernating

No longer active user

Cluster Percentage

The Hibernating
10,144 users
(41.67%)

The Superior-Buyer
606 users
(2.49%)

The Discount-Hunter ●
The Potential-Loyalist ●
The Superior-Buyer ●
The Hibernating ●

The Discount-Hunter
5,200 users
(21.36%)

The Potential-Loyalist
8,321 users
(34.47%)

5
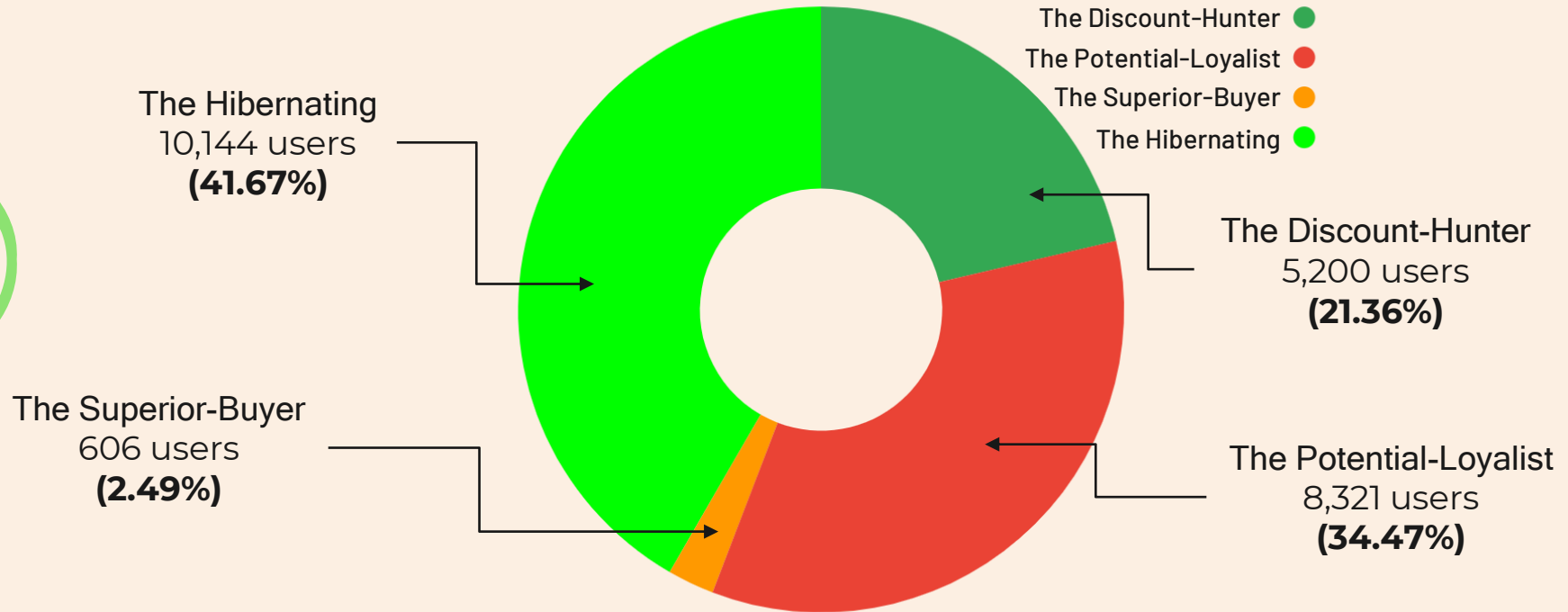
Recommendations

Business recommendations

The Discount-Hunter

This customer cluster can be interpreted as an economical type of customer because they are always waiting for a promo to orders, you should keep this type of customer so that they do not **Hibernate** and even stop subscribing. Business values that E-commerce can give example:

• Give vouchers (max 5%) to attract users' attention with a maximum redemption time limit of 1 week.

• Provide a big sale event voucher (remaining warehouse stock) to impress many discounted items at affordable prices.

• Hold an event to invite your friends to join by providing a referral code to add points that can be exchanged for vouchers, etc.

Business recommendations

The Potential-Loyalist

For this customer cluster, we must lead this user' to become a loyal customer in the future like the **Superior-Buyer** cluster, do not let this customer end up **Hibernating** and move to competitor brands. The business value that can be given to increase the value of transaction frequency and total money spent, for example:

• Provide special discount promos for new users to make transactions.
• Provide free shipping for the first 3 transactions, etc.

Business recommendations

The Superior-Buyer

For this type of customer, we must maintain it, lest its performance decline and even move to competitor brands. Business values that E-commerce can give example:

• Offers VIP members with features providing the latest information about new products etc.

• give promo buy 2 get 3.

• give a discount with a minimum spend of 300,000 IDR, etc.

Business recommendations

The Hibernating

Many possibilities make each user no longer active for example, the price is too high, the transaction process is complicated, the choice of color does not attract the user, and many more. To attract the attention of this type of user, for example:

• Do a reminder via email or WhatsApp at least once a week.

• Provide recommendations for the best-selling items to remind users of E-commerce products.

• Provide information about developments and changes in e-commerce applications such as the ease of online transactions now, the existence of one-day shipping services, better application display, etc.

Reference

The following are references for working on this project:

https://www.programmersought.com/article/63823799496
https://www.kaggle.com/code/felixign/airline-clustering/notebook
https://www.kaggle.com/code/amarmaruf/homework-unsupervised-rakamin-ds8/notebook
https://www.clevertap.com/blog/rfm-analysis/
https://www.moengage.com/blog/rfm-analysis-using-rfm-segments/
https://towardsdatascience.com/rfm-segmentation-in-e-commerce-e0209ce8fcf6

Thank you!

Have any questions?
aldybudhi003@gmail.com
+62 896 5729 1021
www.linkedin.com/in/aldybudhi/
www.github.com/qodym

Let's Check out my python code Jupyter notebook!
Don't hesitate to contact me if you want to do some corrections or discussion!

#DataScience #ClusteringModeling