# Predict Customer Personality to boost marketing campaign by using Machine Learning

Supported by:
**Rakamin Academy**
Career Acceleration School
www.rakamin.com


**Created by:**

## Shanna Yunita Sinaga

**Let's Connect!**

shannasinaga20@gmail.com
https://www.linkedin.com/in/shanna-sinaga-0a5975204/
https://github.com/Shanna2000


As a Chemical Engineering graduate, Shanna has developed a strong foundation in mathematics, statistics, and programming. Through her studies and various internships, She has developed a keen interest in data analysis and its potential to drive business decisions. With her technical background and analytical skills, She is eager to apply her knowledge to a career in data analysis or data science.

# Overview

Rakamin Academy

“A company can develop rapidly when it knows its customer personality behavior, so it can provide better services and benefits to customers who have the potential to become loyal customers. By processing historical marketing campaign data to improve performance and target the right customers so they can transact on the company's platform, from this data insight our focus is to create a cluster prediction model to make it easier for companies to make decisions”

# Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 30 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Unnamed: 0           2240 non-null   int64  
 1   ID                   2240 non-null   int64  
 2   Year_Birth           2240 non-null   int64  
 3   Education            2240 non-null   object 
 4   Marital_Status       2240 non-null   object 
 5   Income               2216 non-null   float64
 6   Kidhome              2240 non-null   int64  
 7   Teenhome             2240 non-null   int64  
 8   Dt_Customer          2240 non-null   object 
 9   Recency              2240 non-null   int64  
 10  MntCoke              2240 non-null   int64  
 11  MntFruits            2240 non-null   int64  
 12  MntMeatProducts      2240 non-null   int64  
 13  MntFishProducts      2240 non-null   int64  
 14  MntSweetProducts     2240 non-null   int64  
 15  MntGoldProds         2240 non-null   int64  
 16  NumDealsPurchases    2240 non-null   int64  
 17  NumWebPurchases      2240 non-null   int64  
 18  NumCatalogPurchases  2240 non-null   int64  
 19  NumStorePurchases    2240 non-null   int64  
 20  NumWebVisitsMonth    2240 non-null   int64  
 21  AcceptedCmp3         2240 non-null   int64  
 22  AcceptedCmp4         2240 non-null   int64  
 23  AcceptedCmp5         2240 non-null   int64  
 24  AcceptedCmp1         2240 non-null   int64  
 25  AcceptedCmp2         2240 non-null   int64  
 26  Complain             2240 non-null   int64  
 27  Z_CostContact        2240 non-null   int64  
 28  Z_Revenue            2240 non-null   int64  
 29  Response             2240 non-null   int64  
dtypes: float64(1), int64(26), object(3)
memory usage: 525.1+ KB
```

## DESCRIPTION

Dataset contains data purchasing from all shop's resources, data accepted campaigns, and the data of shop's customers.

## SHAPE

2.240 rows, 30 features

## DTYPE

Int64 (26 features), float64 (1 features), object (3 features)

## MISSING VALUES

One features has missing values : Income

To see the all the codes in jupyter notebook Click Here

# Feature Engineering

Rakamin Academy

FEATURE ENGINEERING

**Total Campaign Accepted**
- total_acc_campaign

Total Days Customer Start Joined
- - total_days_joined

**Age Classification**
- Age
- Age_group

**Total Spending for Last 2 Years**
- total_spending

**Coversion Rate**
- Total Purchase
- Conversion_rate

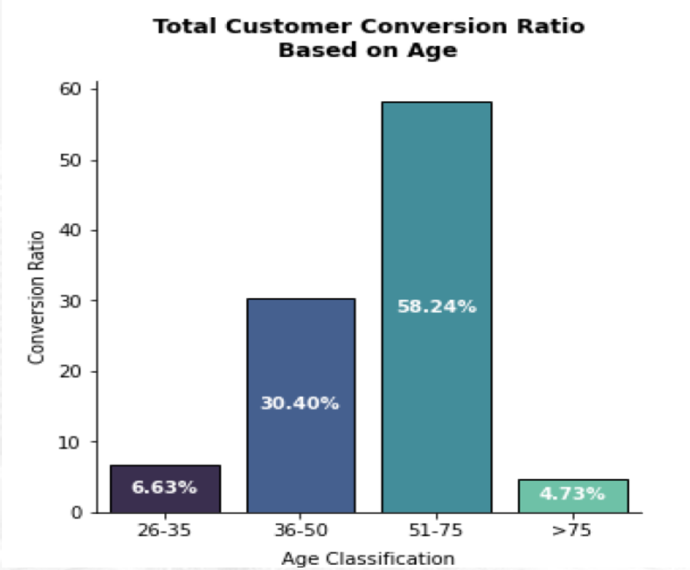
**Total Children**
- total_children

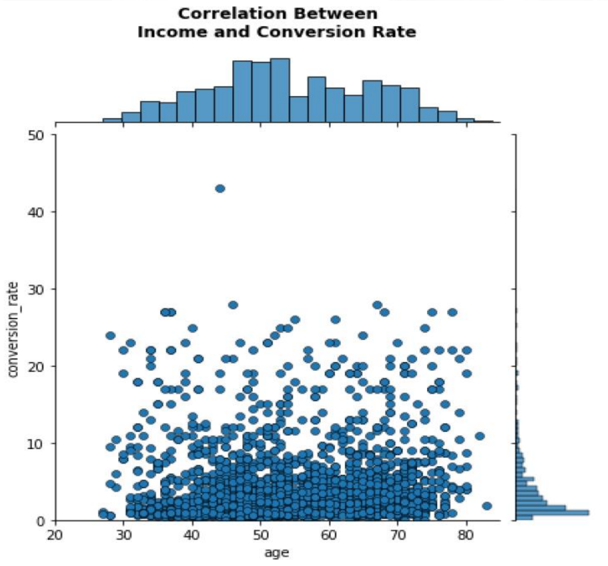
High Correlated Features with Conversion Rate:
1. Income
2. Total spending
3. Total children

To see the all the codes in  jupyter notebook Click Here

# EXPLORATORY DATA ANALYSIS (EDA)

## Correlation Age with Conversion Rate



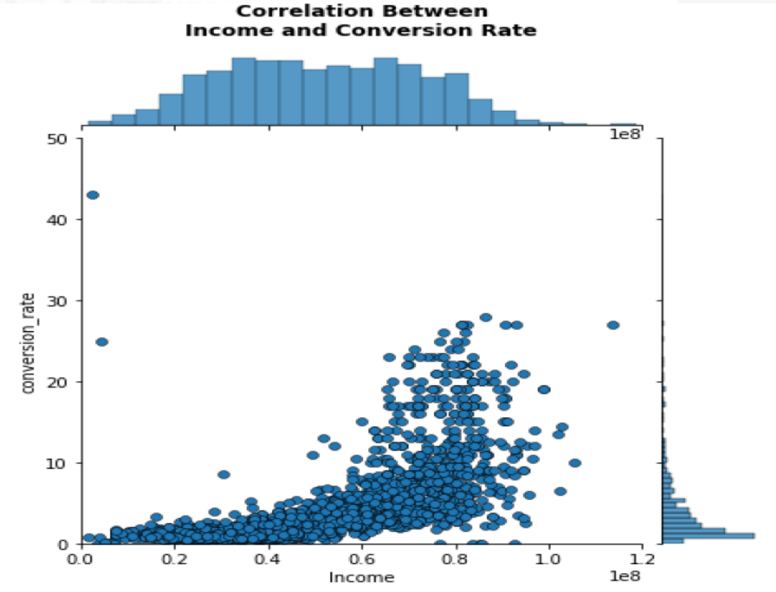




- Based on graphs above, the conversion rate **distribution dominated by 51-75 years old** (58.4%) next the 36 – 50 years old with conversion ratio percentage 30,40%
- On the conclusion **we must engage the 51-75 years old** (more likely middle age citizen) by **provide the supply their age needs**, and giving ads for their concern on the product

To see the all the codes in jupyter notebook Click Here

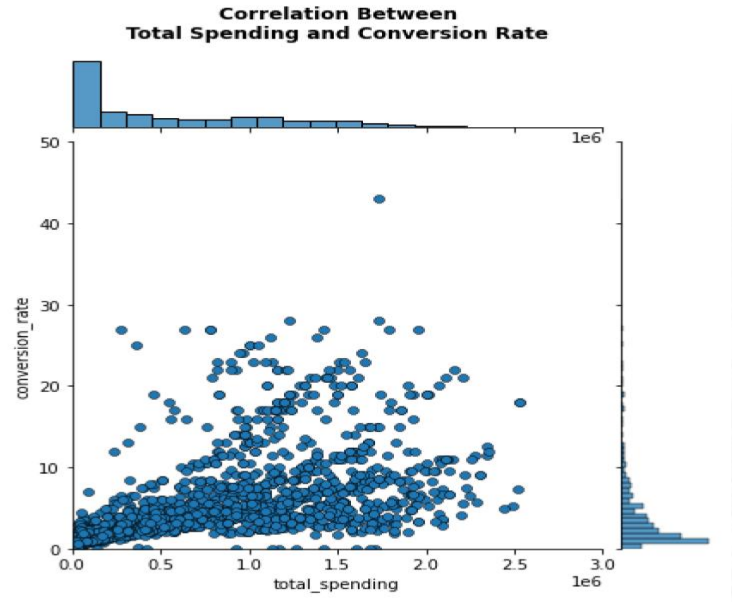
# EXPLORATORY DATA ANALYSIS (EDA)

## Correlation Conversion Rate with Income




Based on the graphs above, higher income could impact higher conversion rate. High conversion rate dominated with customers who have >60 M/year income

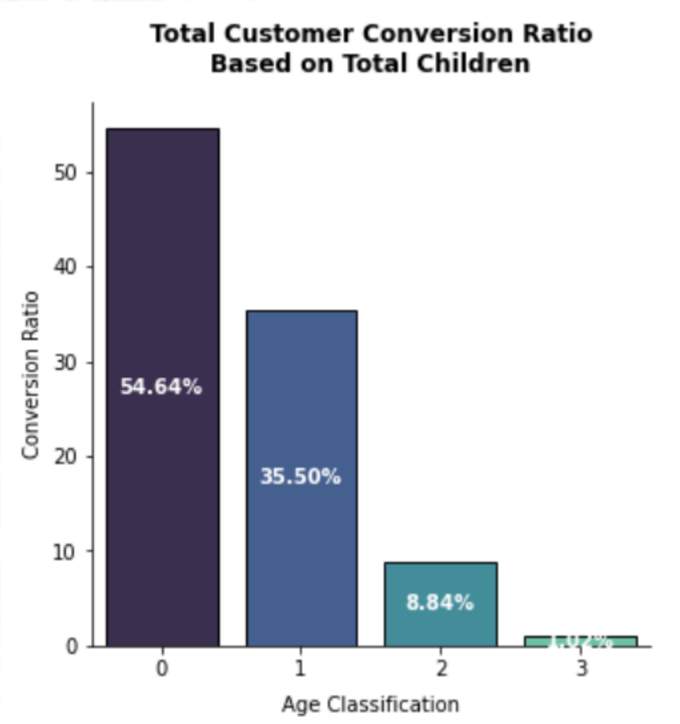
## Correlation Conversion Rate with Total Spending




Based on the graphs above, higher total spending for the last 2 years more likely have higher conversion rate. Total spending >1 Million have higher conversion rate

To see the all the codes in jupyter notebook [Click Here]

## Correlation Conversion Rate with Total Children




Customer with no children dominated the distribution of conversion ratio with percentage 54.64%

To see the all the codes in  jupyter notebook [Click Here]

# DATA CLEANING & DATA PREPROCESSING

Rakamin Academy

```
Education              0
Marital_Status         0
Income                24
Kidhome                0
Teenhome               0
Dt_Customer            0
Recency                0
MntCoke                0
MntFruits              0
MntMeatProducts        0
MntFishProducts        0
MntSweetProducts       0
MntGoldProds           0
NumDealsPurchases      0
NumWebPurchases        0
NumCatalogPurchases    0
NumStorePurchases      0
NumWebVisitsMonth      0
AcceptedCmp3           0
AcceptedCmp4           0
AcceptedCmp5           0
AcceptedCmp1           0
AcceptedCmp2           0
Complain               0
Response               0
age                    0
conversion_rate        0
age_group              0
Total_Purchase         0
total_acc_campaign     0
total_spending         0
Dt_Collected           0
total_days_joined      0
total_children         0
dtype: int64
```

## Handling Missing Values

- 24 null in feature Income (fill with median)

## Duplicated Data

- 183 duplicated data (drop the duplicated data)

## Handling Outliers

- Since the outliers not so far from the distribution and they are not wrong input, then I decided to not handle the outlier

## Feature Encoding

- Since there are no categorical columns being used, there are no feature encoding

## Feature Selection

Using LFRMC (Tao, 2020):

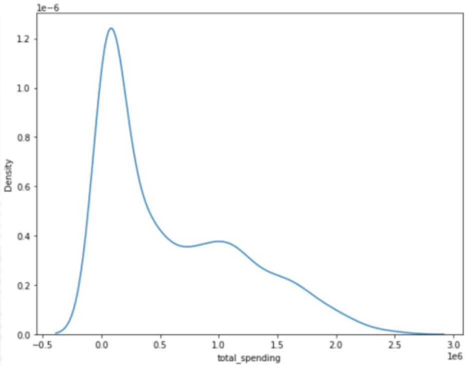
- L : total_days_joined
- F : Total_Purchase
- R : Recency
- M : total_spending
- C : total_acc_campaign

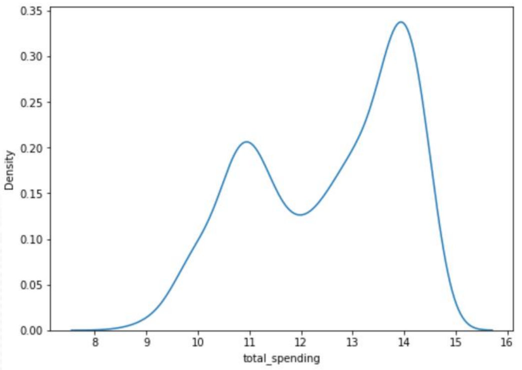
## Feature Transformation

- Log Transformation (total_spending)
- Normalization – MinMaxScaler (all 5 features)

# DATA CLEANING & DATA PREPROCESSING

## Log Transformation

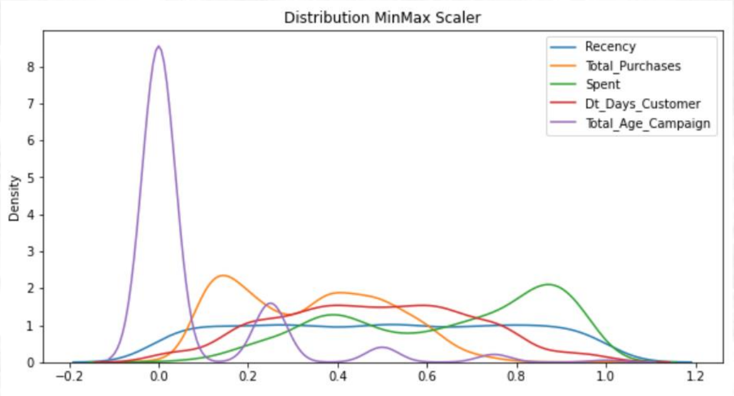



- Features total_spending have positive skew characteristic
- Log transformation to normalize the distribution

c

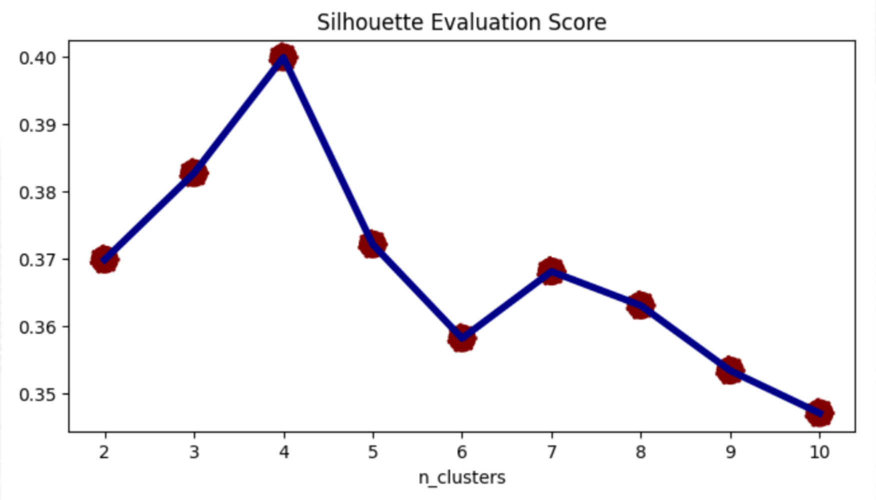
## Normalization - MinMaxScaler

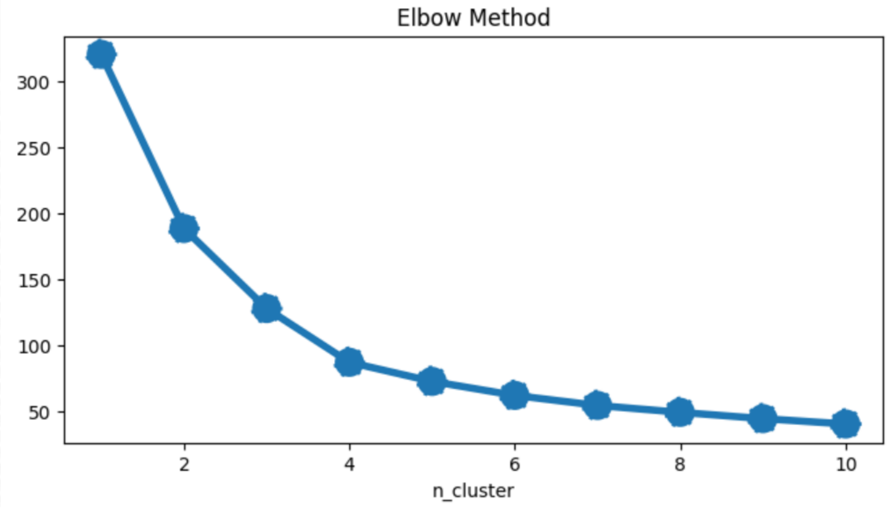



| | total_days_joined | Total_Purchase | Recency | total_spending | total_acc_campaign |
|---|---|---|---|---|---|
| count | 2057.000 | 2057.000 | 2057.000 | 2057.000 | 2057.000 |
| mean | 0.481 | 0.338 | 0.495 | 0.643 | 0.075 |
| std | 0.219 | 0.174 | 0.293 | 0.239 | 0.170 |
| min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 25% | 0.319 | 0.182 | 0.242 | 0.422 | 0.000 |
| 50% | 0.481 | 0.341 | 0.495 | 0.702 | 0.000 |
| 75% | 0.643 | 0.477 | 0.747 | 0.859 | 0.000 |
| max | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

# Data Modeling

## Evaluation Model



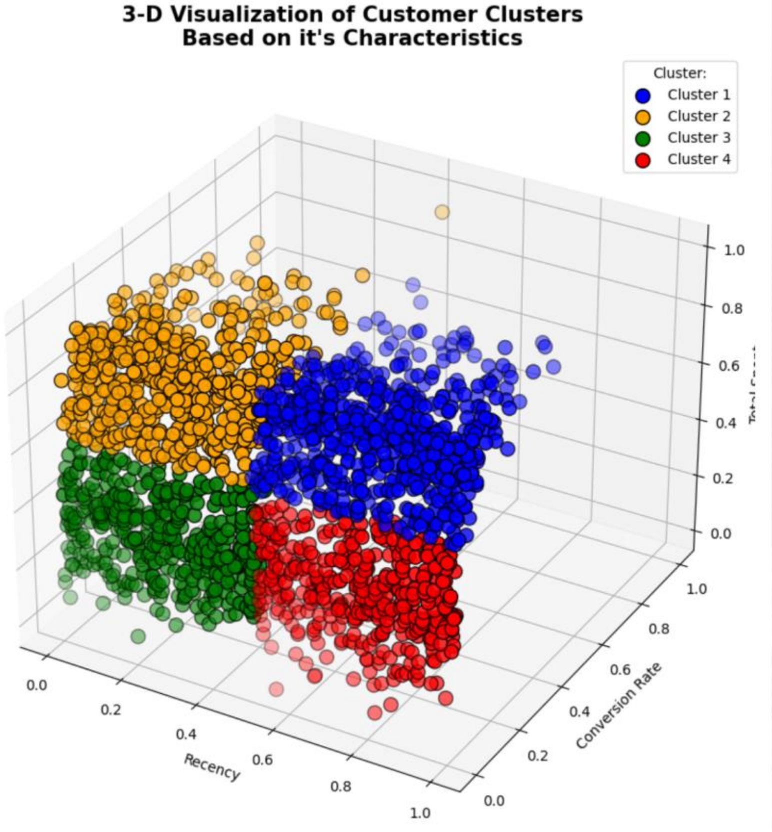



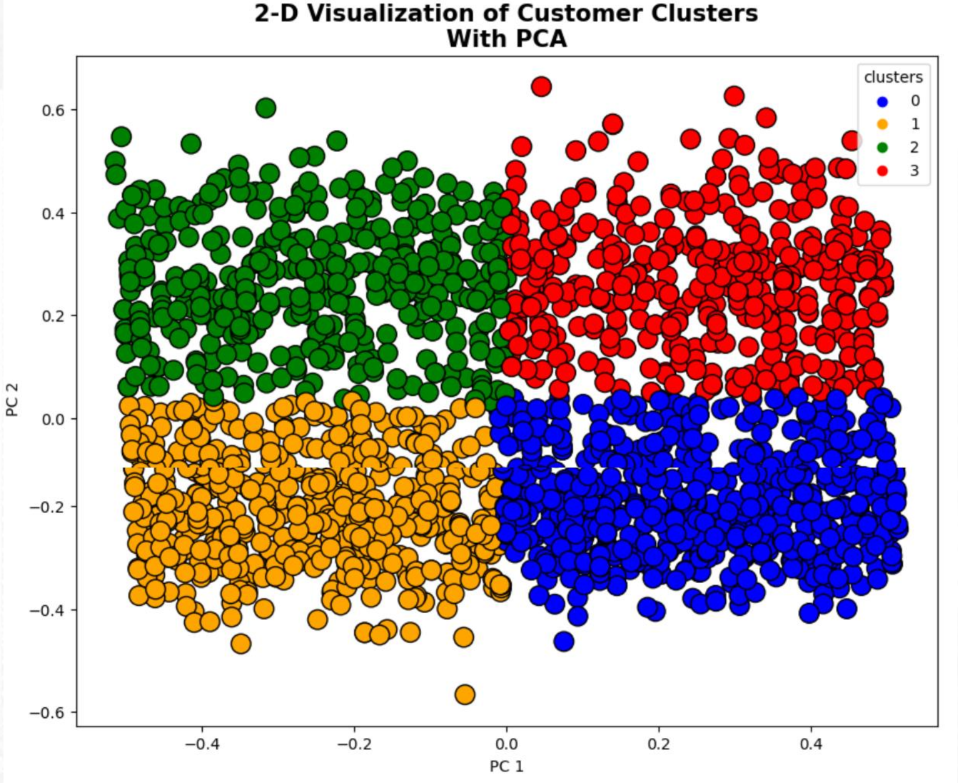

Based on Silhouette Evaluation Score and Elbow Method. The Clustering model would work effectively with four segmentation (cluster)

To see my notebook codes, Click Here

# Data Modeling

## Modelling Result

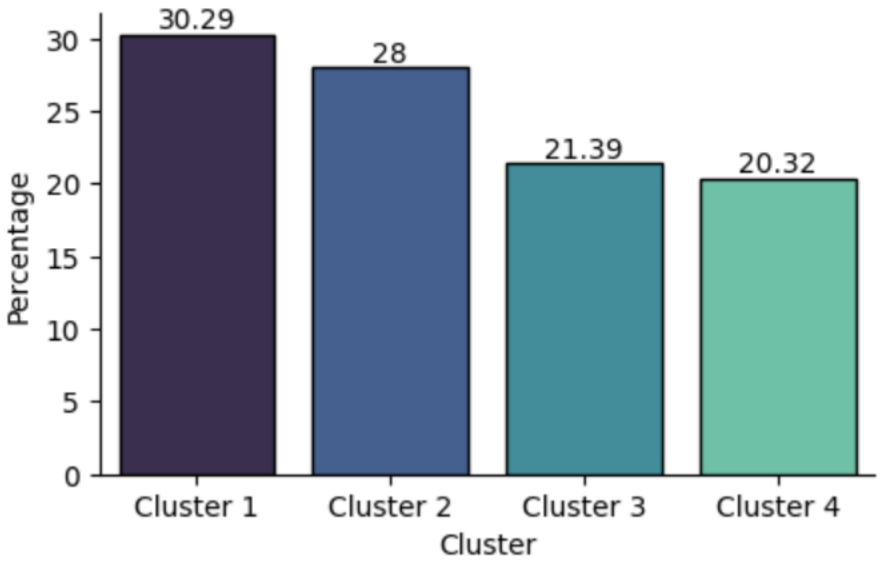



To see my notebook codes, [Click Here]

# Customer Personality Analysis for Marketing Retargeting

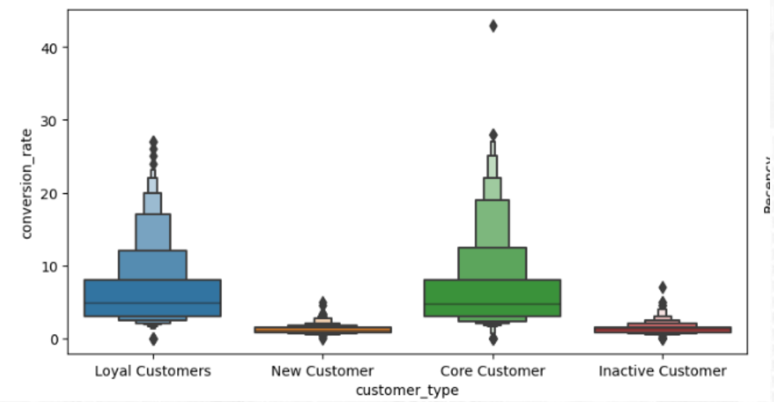
| | conversion_rate | Recency | total_spending |
|---|---|---|---|
| | mean | mean | mean |
| km_labels | | | |
| **0** | 0.152174 | 0.735817 | 0.825211 |
| **1** | 0.156025 | 0.225940 | 0.820800 |
| **2** | 0.027936 | 0.254982 | 0.382517 |
| **3** | 0.030128 | 0.757962 | 0.398990 |

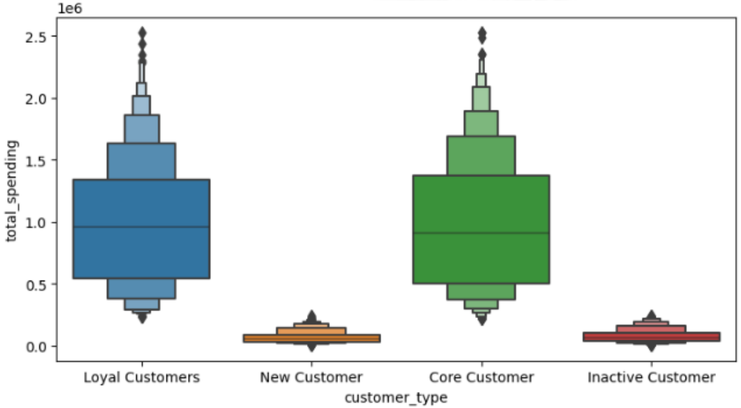



To see my notebook codes, [Click Here]

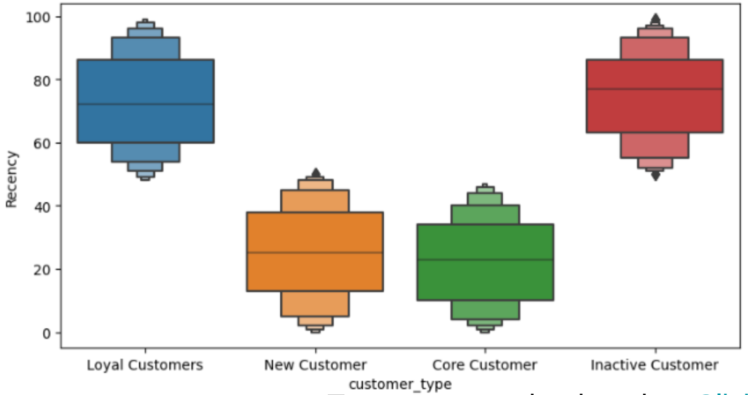
# Customer Segmentation Based on RFM Analysis

Rakamin Academy



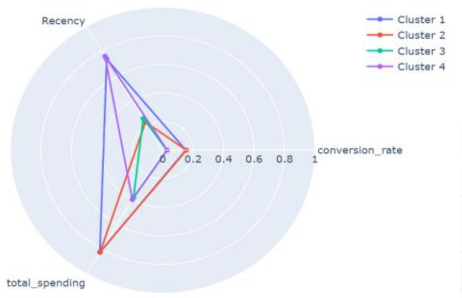

| | |
|---|---|
| **Cluster 1** (Loyal Customers) | Customer with high spending and high freq uent but not a recent customers |
| **Cluster 2** (Core Customers) | Customer with high spending, high frequen t, also recent customer |
| **Cluster 3** (New Customers) | Recent new customer |
| **Cluster 4** (Inactive Customers) | Customer with low spending and not frequ ent customer but some point haven't boug ht again |



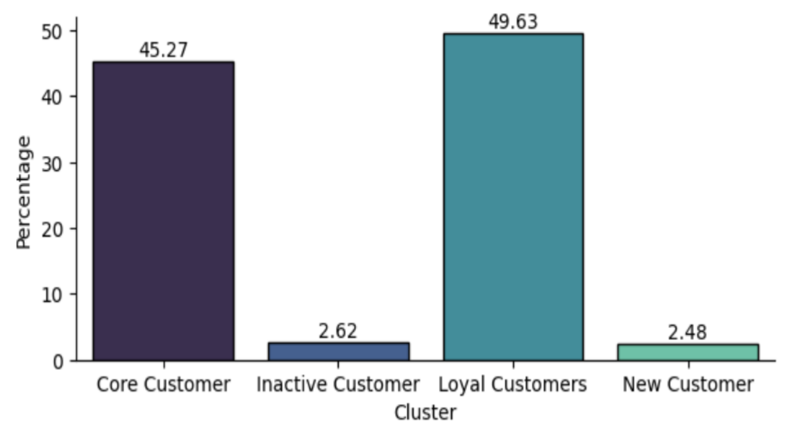

To see my notebook codes, Click Here

# Customer Personality Analysis for Marketing Retargeting

Radar Chart for Recency (R), Conversion Rate (F) , and Total Spending (M),




Percentage Total Spending Based on Customer Segementaion,




**Core Customers**

- There are 28% of total customers from this group
- This customers have a high average spending in our platform (~ 1M/year). They contribute 45.27% income value to our platform. This customers also a frequent customers. Their conversion rate average pretty high (~ 8%) They are also a recent customers, which means they bought a product not too long ago (~ 22 days ago).
- Overall, this type customers giving high value and also a very loyal customers.

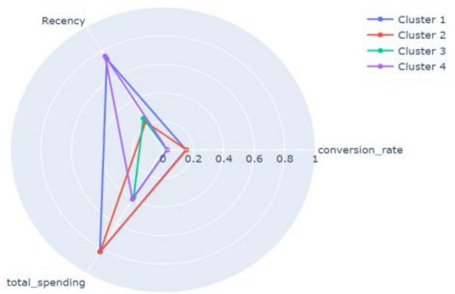
**Loyal Customers**

- There are 30,3% of total customers from this group
- This customers have a high average spending on our platform (~ 1M/year). They contributed for 49.63% income value to our platform. They also identified as a frequent customers (~ 8% conversion rate). But, they are not a recent customers. They last bought was in ~ 75 days ago in average.
- Overall, this customers once a high value and frequent customers, but at some point they stopped to buy in our platform

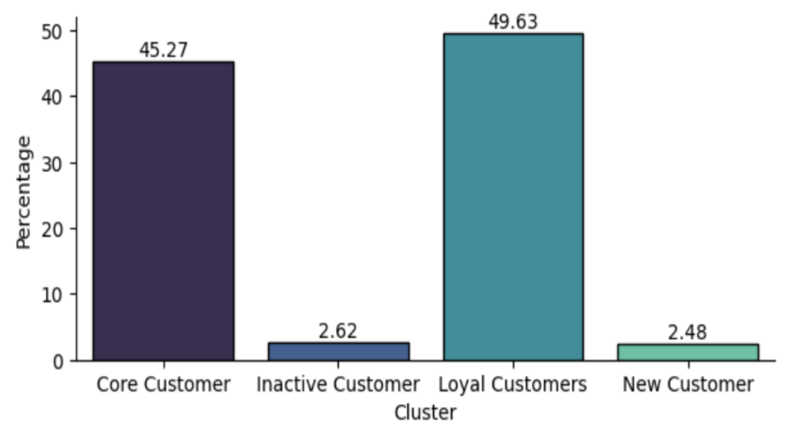
To see my notebook codes, Click Here

# Customer Personality Analysis for Marketing Retargeting

Radar Chart for Recency (R), Conversion Rate (F) , and Total Spending (M),




Percentage Total Spending Based on Customer Segementation,




## New Customers

- There are 21.4% of total customers from this group
- This type customers are new customers. That means they didn't spent much in our platform yet (~70k/year in average). They also not frequent since they're still a new customers. They conversion rate only 1% but they are a recent new customers. That means they bought a product not long ago (25 days ago in average)
- Overall this customers are a new customers. They still spending not a lot but they can be a potential customers to be a core customers

## Inactive Customer

- There are 20.32% of total customers from this group
- This type customers doesn't spent a lot in our website (~ 78k/year in average). They also not a frequent customers, they conversion rates only 1.3%. Last time they bought a product are 75 days ago in average.
- Overall, for some point this customers like inactive customer, they don't spent a lot also not a frequent customers

To see my notebook codes, [Click Here]

## Business Recommendation

| | |
|---|---|
| **Core Customer** | • This customer is loyal and not hesitant to spend money / buy products at a high price. Therefore, for this customer, **discounts should be put aside** and the **focus should be on customer service**, as well as on adding value through offers based on **product recommendations that are based on previous purchases.**<br>• **Identify and engage with these customers** early on in their relationship with the platform to maintain their loyalty and increase their lifetime value. **Provide them with special offers or personalized attention and ensure that they have a seamless experience with the platform** |
| **Loyal Customer** | • Maintain a strong relationship with these customers and continue to provide them with a high level of service and personalized attention. This might include loyalty programs, special discounts or offers, or exclusive access to new products or features. |
| **New Customer** | • Engage with these customers early on in their relationship with the platform and **encourage them to continue engaging. Provide targeted offers or incentives to encourage repeat purchases or usage** |
| **Inactive Customer** | • **Identify why these customers are not engaging with the platform and develop targeted strategies to re-engage them.** This might include personalized offers, improving the user experience, or providing better customer service. |

To see my notebook codes, Click Here

## Recommendation Impact

By giving them the needed treatment, we can prevent their likely to churn and stay on our platform and by applying the business recommendation to give a suitable treatment for particular customers, we could get **1.2B/year** for approximation. And, if we prioritize the loyal customers and core customers, we could still get **94.5%** from all the total revenue.

To see my notebook codes, Click Here