PREDICT CUSTOMER PERSONALITY TO BOOST MARKETING CAMPAIGN BY USING MACHINE LEARNING

BY: DEA DAHLILA

About Me

"I'm a Fresh Graduate of Geophysics Engineering with a keen interest in data science and analysis. I'm passionate about extracting meaningful insights from data and using them to drive informed decision-making."

Let's Connect

✉ ddahlila6@gmail.com

in linkedin.com/in/dea-dahlila/

github.com/Dea-dahlila

WHAT IS THE PROBLEM?

BACKGROUND	Provide better services and benefits to customers who have the potential to become loyal customers by knowing the personality behavior of its customers
GOALS	Determine the right customers to increase marketing reach.
OBJECTIVES	Create a cluster prediction model so that it makes easier for companies to improve marketing campaign.
BUSINESS METRICS	• Clustering • Conversion Rate

DATASET

```
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)
```

Shape

2.240 data rows, 30 features.

Dtype

float64 (1 features), int64 (26 features), object (3 features).

Duplicated

0 data rows.

Missing Value

1 features that has missing value.

Exploratory Data Analysis (EDA)

| Feature Extraction

- Total_Acc_Cmp
- Total_Purchases
- Total Spent
- conversion_rate
- Dt_collected
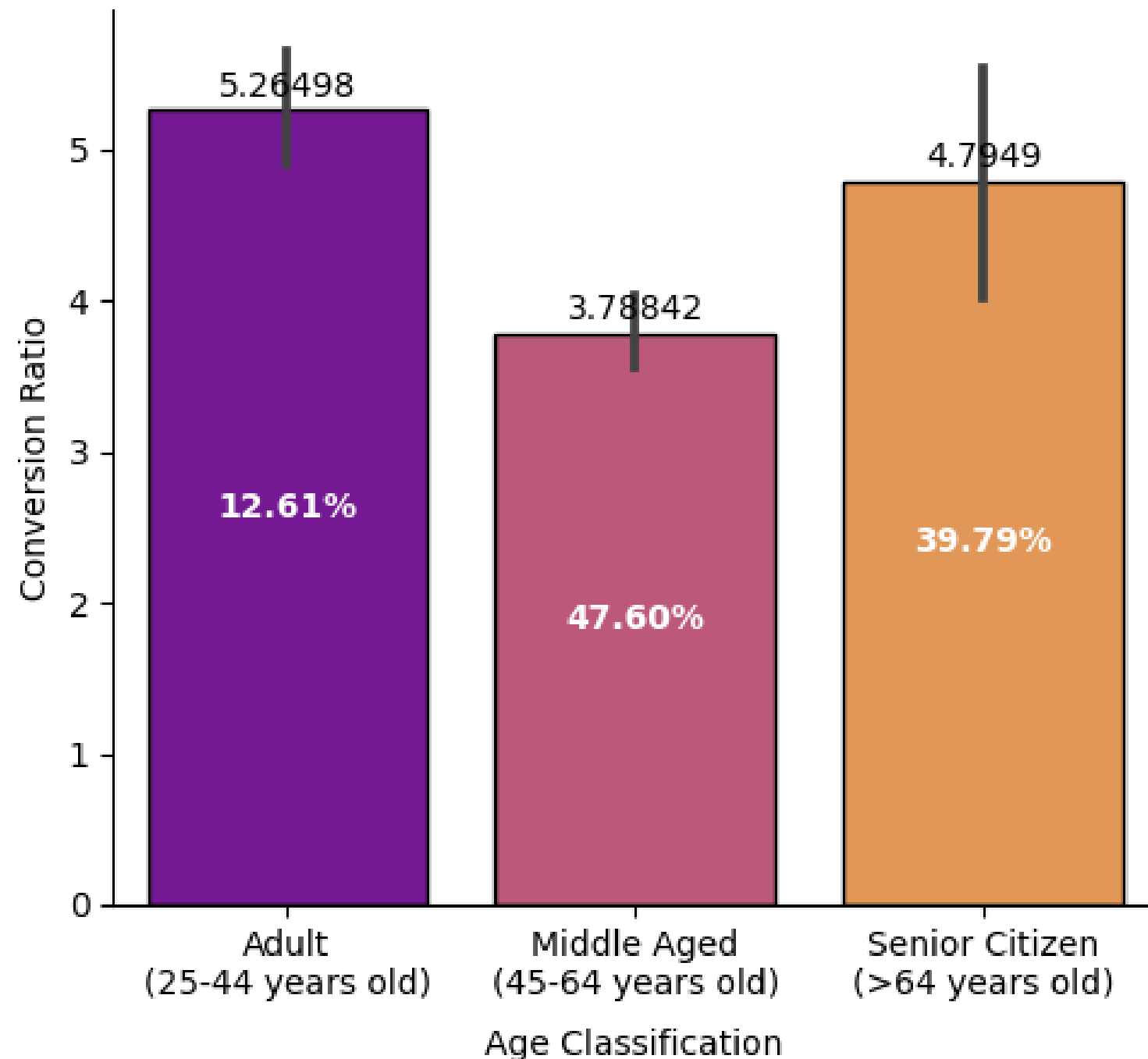- Dt_Days_Customer
- Age_Group

| Correlation

4 Feature highly corelated:

- Age
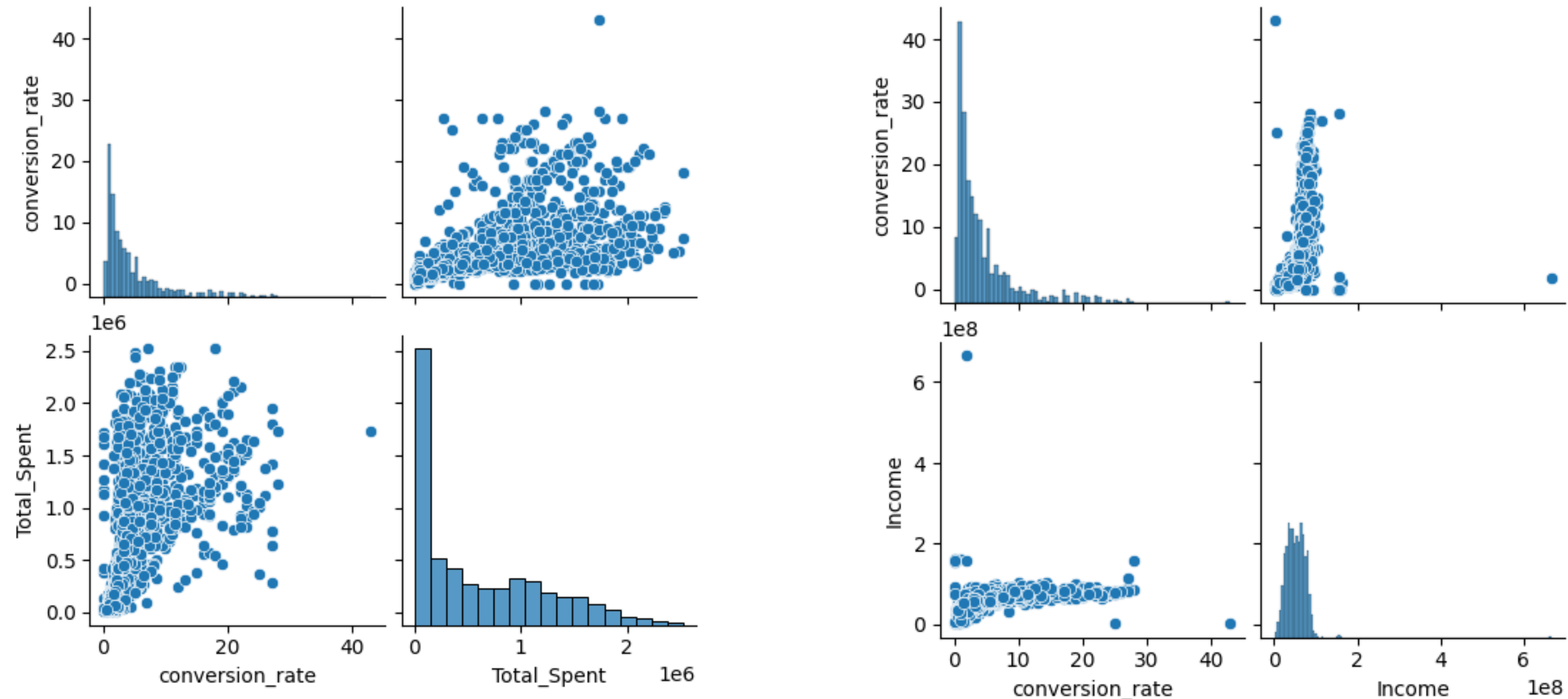- Income
- conversion_rate
- Total Spent

Conversion Rate Based on Age, Income, Spending

Total of Customer Conversion Rate Based on Age

- Middle Aged (45-64 years old) has a higher conversion rate which is around 47.60%, senior citizens (39.79%), and adults (12.61%).
- Therefore we can provide special advertisements/products for more middle aged people.

Conversion Rate Based on Age, Income, Spending

The greater the income, the greater the conversion rate, as well as customer spending, the greater the total spending, the greater the cvr.

Data Cleaning & Preprocesing

1 **Missing Value**

handled missing value in columns Income fillit with median

2 **Duplicated Data**

There is no duplicated data

3 **Outliers**

Handled outliers with IQR to make sure all features are stable for modelling (Q1 = 1% & Q3 = 95%).

4 **Feature Selection**

Using RFMLC methode to reduce dimensionality : Recency (Recency), Frequency (Total Purchases), Monetary (Total Spent), Loyalty (Dt Days Custmoer, C (Age).
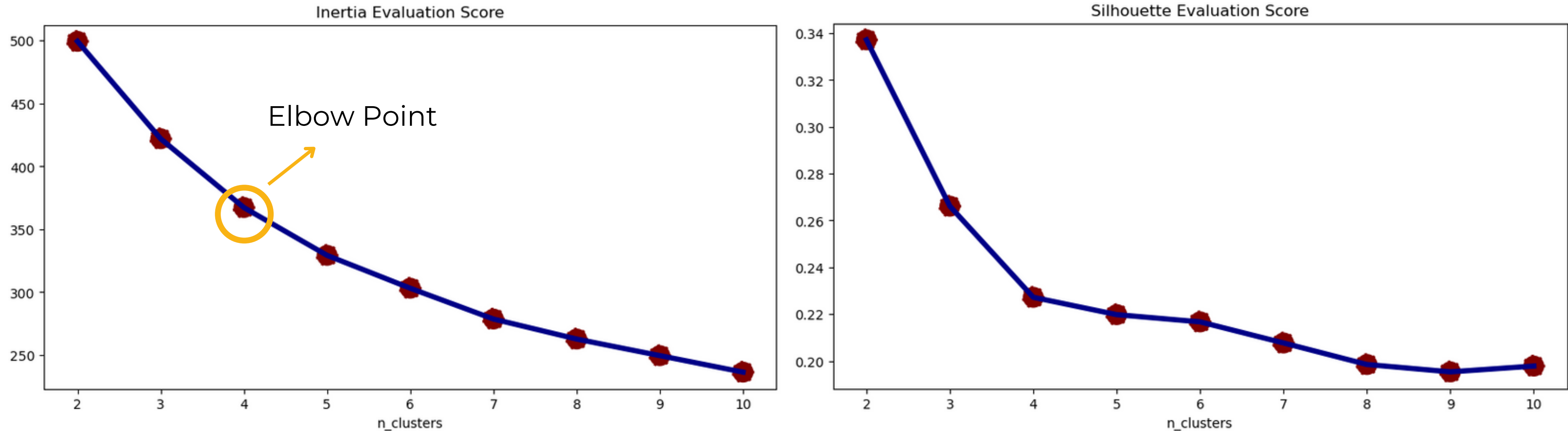
5 **Feature Transformation**

Using MinMaxScaler.

6 **Encoding**

Using Label enc and one hot enc.

Data Modelling

- Elbow Methode analysis and silhouette scores were performed to see the number of clusters obtained from the k-means clustering algorithm.
- In the graph above, the number of clusters = 4 is obtained which is enough to sell each cluster with the most optimal distance.

Customer Personality Analysis for Marketing Retargeting : Cluster Analysis


2-D Visualization of Customer Clusters Wih PCA
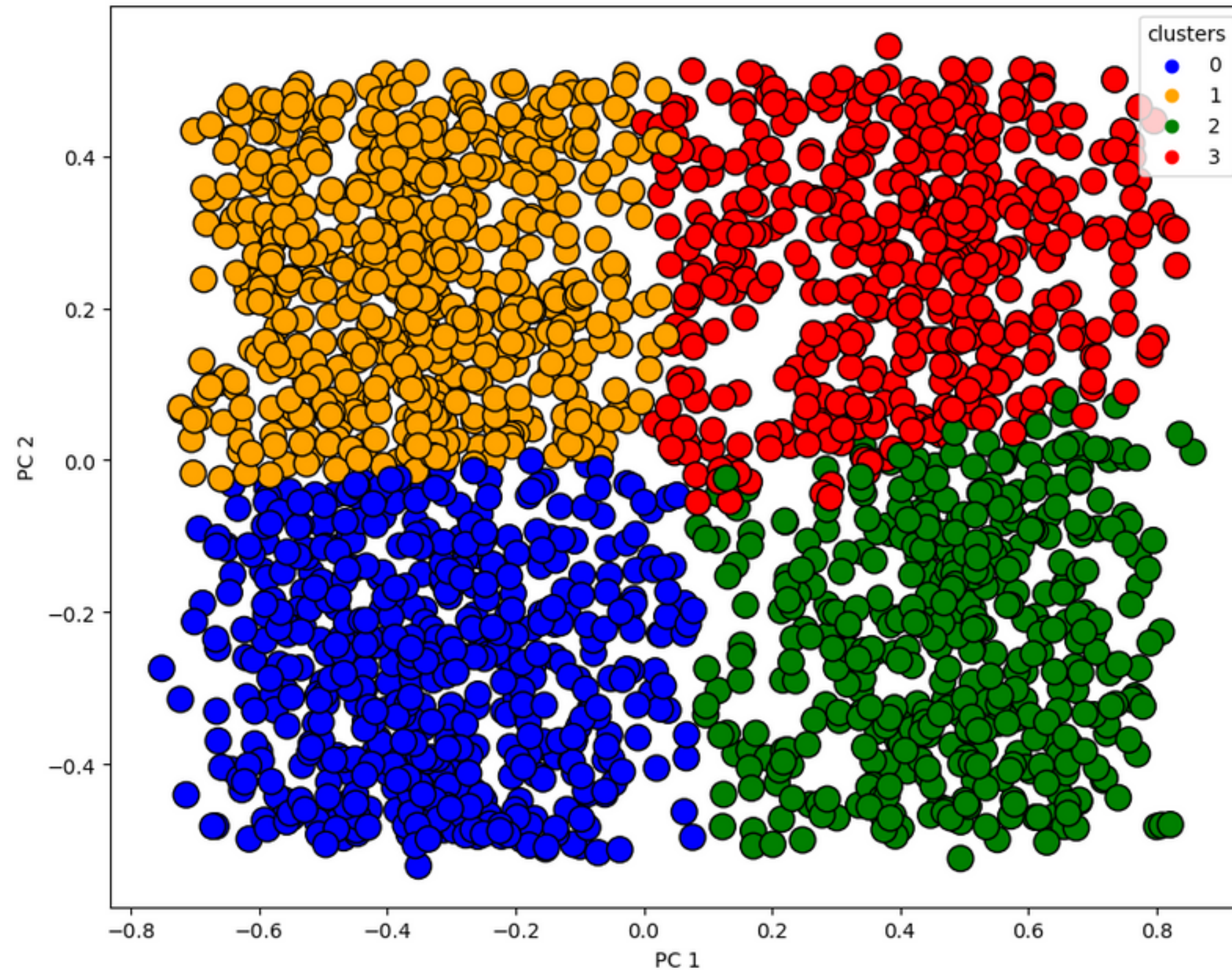
- According to visualization using PCA with 2 main PC's, the clusters are perfectly separated.
- There's clearly 4 customer clusters that generated by K-Means Clustering algorithm using RFM Method for this dataset.

# Customer Personality Analysis for Marketing Retargeting : Cluster Analysis

**3-D Visualization of Customer Clusters
Based on it's Characteristics**

clusters:
- High Spender
- Mid of Churn
- Risk of Churn
- Low Spender
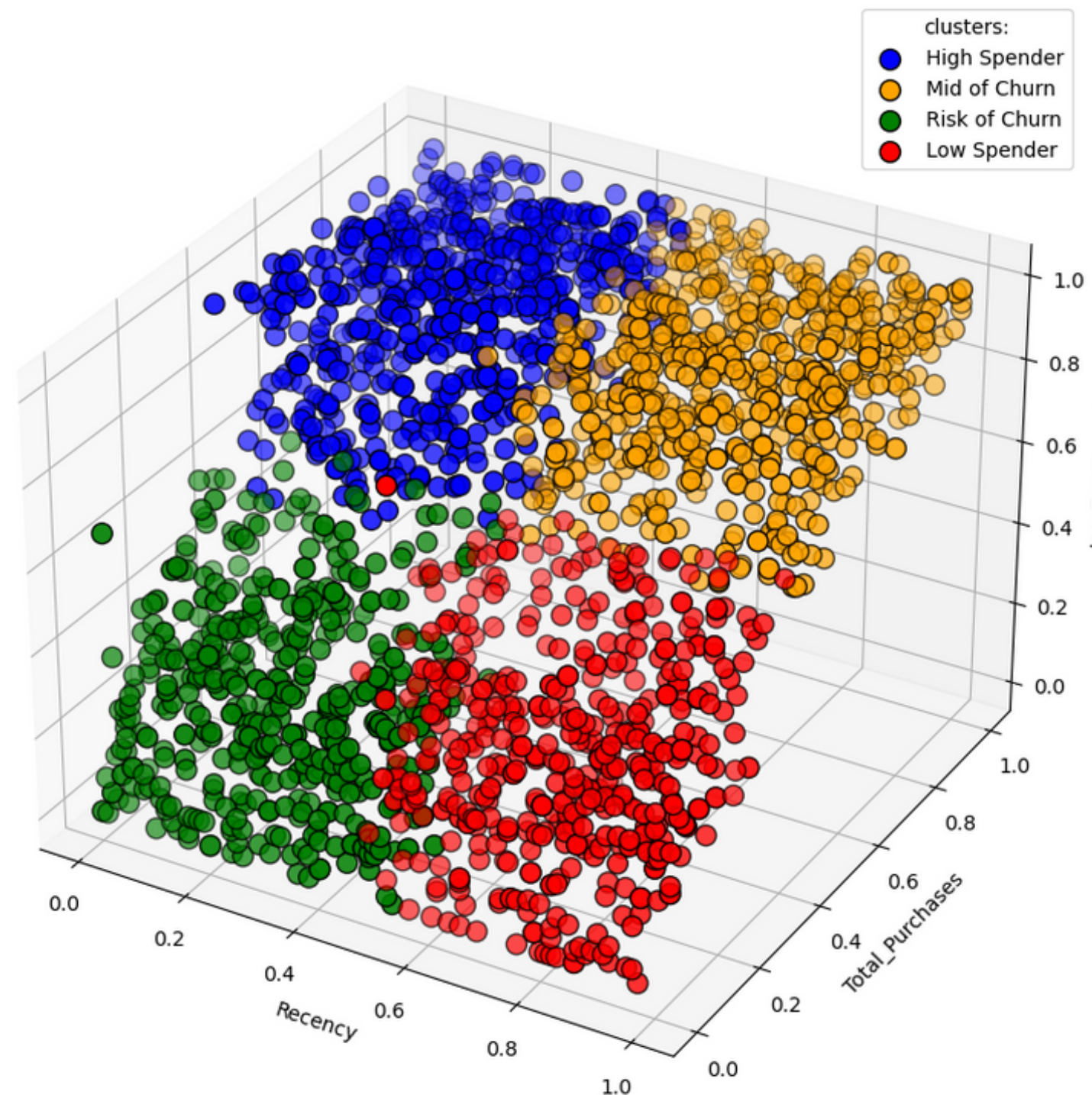
1. High Spender :

- This group is dominated by middle aged (45-54 years old) who are predominantly married and have 1 child.
- There are 563 customers (27.65% of total customers) on this group.
- Customers on this group have `high average recency (25 days)` and `high average of total purchases (70 items)` it means they are frequent shoppers and `they spend a lot on our platform (around IDR 9M/year)`
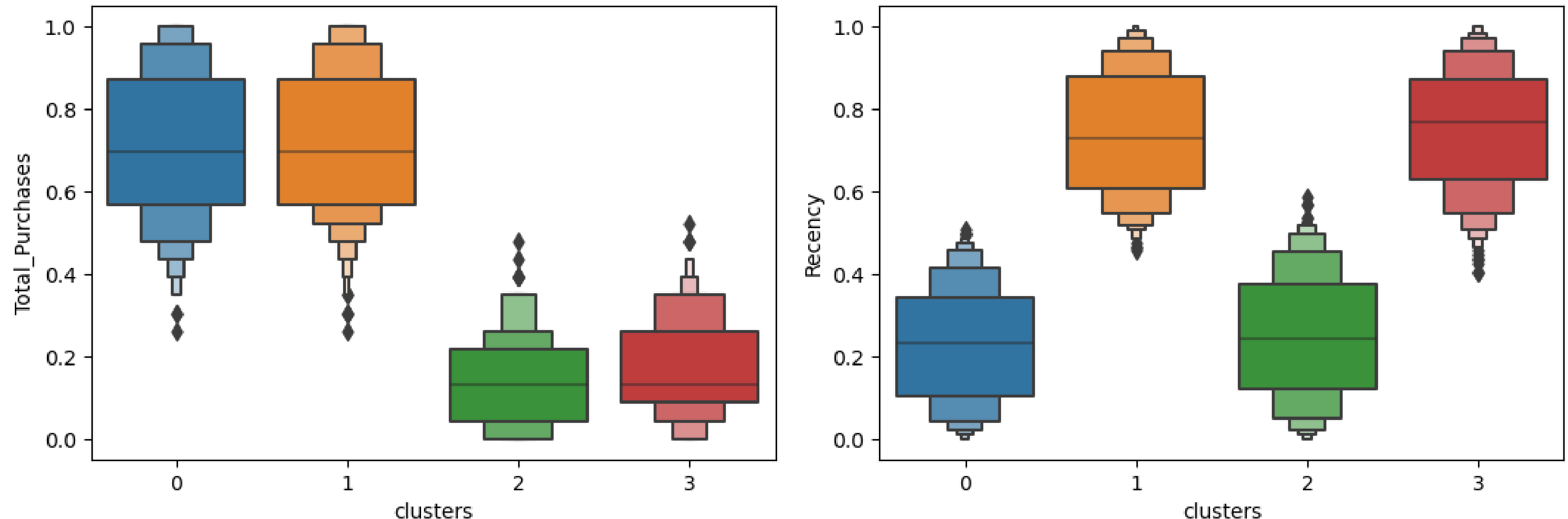- This group has a high conversion rate.

Customer Personality Analysis for Marketing Retargeting : Cluster Analysis
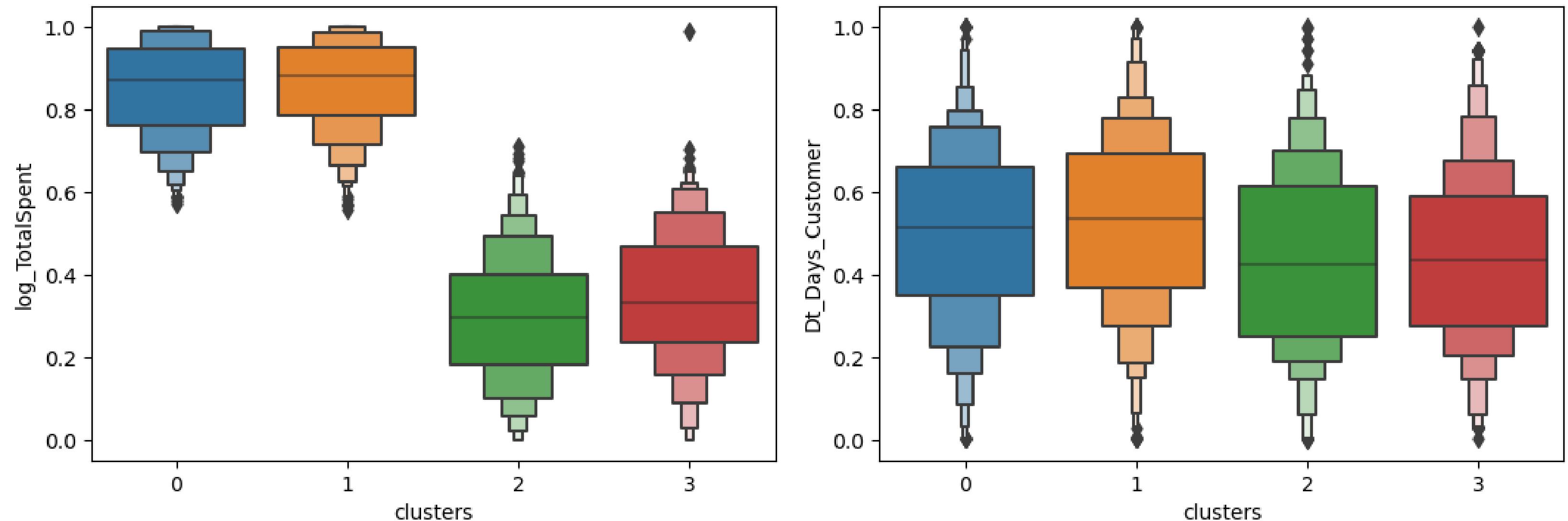
2. Mid Spender :

- Dominated by middle aged (13.85%) who are predominantly married and have 1 child.
- -There are 581 customers (28.54% of total customers) on this group
- Customers on this group have `high average recency (78 days)` and `high average of total purchases (70 items)` and `they spend a lot on our platform (around IDR 9M/year)`

Customer Personality Analysis for Marketing Retargeting : Cluster Analysis
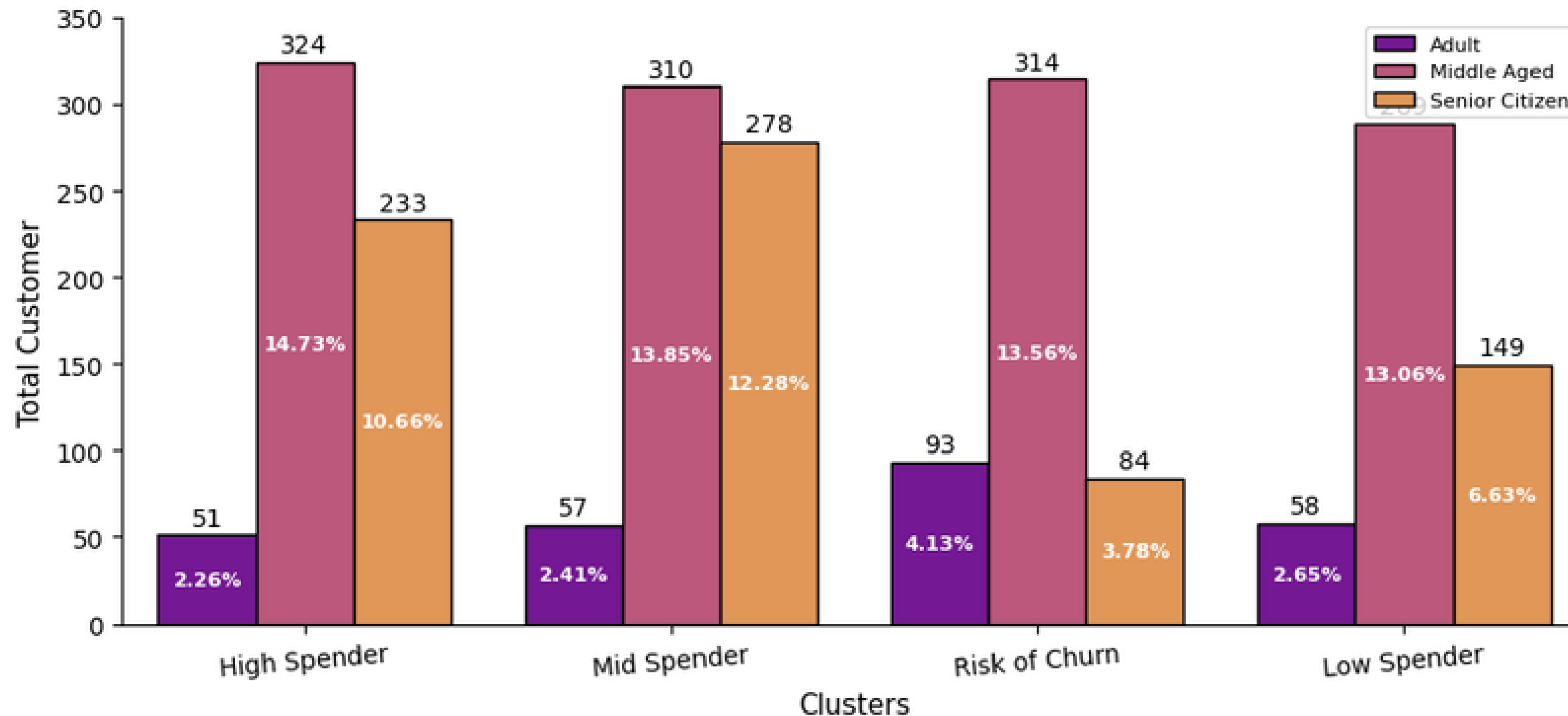

3. Risk of Churn :

- There are 437 customers (22.35% of total customers) on this group
- Dominated by middle aged (13.85%) who are predominantly married and have 1 child.
- Customers on this group have `high average recency (25 days)` and `high average of total purchases (15 items)` and `they spend a lot on our platform (around IDR 2.7M/year)`.

Customer Personality Analysis for Marketing Retargeting : Cluster Analysis

Total of Customers Each Cluster Based on Age

Middle Aged Customer dominated on each cluster (>13% of total customer).

Legend: Adult, Middle Aged, Senior Citizen

High Spender: Adult 51 (2.26%), Middle Aged 324 (14.73%), Senior Citizen 233 (10.66%)
Mid Spender: Adult 57 (2.41%), Middle Aged 310 (13.85%), Senior Citizen 278 (12.28%)
Risk of Churn: Adult 93 (4.13%), Middle Aged 314 (13.56%), Senior Citizen 84 (3.78%)
Low Spender: Adult 58 (2.65%), Middle Aged (13.06%), Senior Citizen 149 (6.63%)

4. Low Spender :

- There are 455 customers (21.46% of total customers) on this group
- Dominated by middle aged (13.85%) who are predominantly married and have 1 child.
- Customers on this group have `high average recency (70 days)` and `high average of total purchases (16 items)` and `they spend a lot on our platform (around IDR 3M/year)`

Recommendations & Impact

Recommendations

1. Keep monitoring transactions and retention of the High Spender group, Focus on improving service so that this group group does not churn.
2. For the Mid Spender group, further analysis can be done on how to increase transactions Fons by providing more personalized recommendations, as well as deeper analysis on how to optimize promos in this segment and keep shopping on our platform.
3. For the Low Spender group, further analysis can also be done on how to increase the conversion rate of visits to transactions, They have a fairly high recency but do not make transactions. This can be caused by products or prices that do not match.

Potential Impact (Quantitative):

- If we keep prioritize on Customer Groups/Clusters and they do not turn to churn, we still have potential GMV around `IDR 1.2B/year` (High Spender: 545m/year; Mid Spender: 588M/year; Risk of Churn: 32M/year; Low Spender: 44M/year).

THANK YOU