Predict Customer Personality to boost marketing campaign by using Machine Learning

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

Rakamin Academy

Created by:
Tsaniya Nur Sukma

Let's Connect!!

tsaniyanurs00@gmail.com
https://www.linkedin.com/in/tsaniyans/
https://github.com/Tsaniyans

A bachelor with problem solving and data analysis skills in data-driven decision making so that also make her proficient in SQL, Python Programming, Machine Learning, Statistics, Data Visualization, Data Warehouse and has contributed to several intern-based projects related to Data Scientist and Data Engineer "

Overview

“Sebuah perusahaan dapat berkembang dengan pesat saat mengetahui perilaku customer personality nya, sehingga dapat memberikan layanan serta manfaat lebih baik kepada customers yang berpotensi menjadi loyal customers. Dengan mengolah data historical marketing campaign guna menaikkan performa dan menyasar customers yang tepat agar dapat bertransaksi di platform perusahaan, dari insight data tersebut fokus kita adalah membuat sebuah model prediksi kluster sehingga memudahkan perusahaan dalam membuat keputusan ”

Data Info

- From the data, they have 2240 total row and 30 total columns.
- Have 24 missing values in Income columns.
- Adjust data type Dt_Customer to datetime

```
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
...
 28  Z_Revenue            2240 non-null   int64
 29  Response             2240 non-null   int64
dtypes: float64(1), int64(26), object(3)
```

Untuk selengkapnya, dapat melihat jupyter notebook disini

Feature Engineering

Rakamin Academy

Age Columns → Age Grouping :
- Infant
- Toodler
- Child
- Teen
- Adult
- Middle Aged
- Senior Citizen

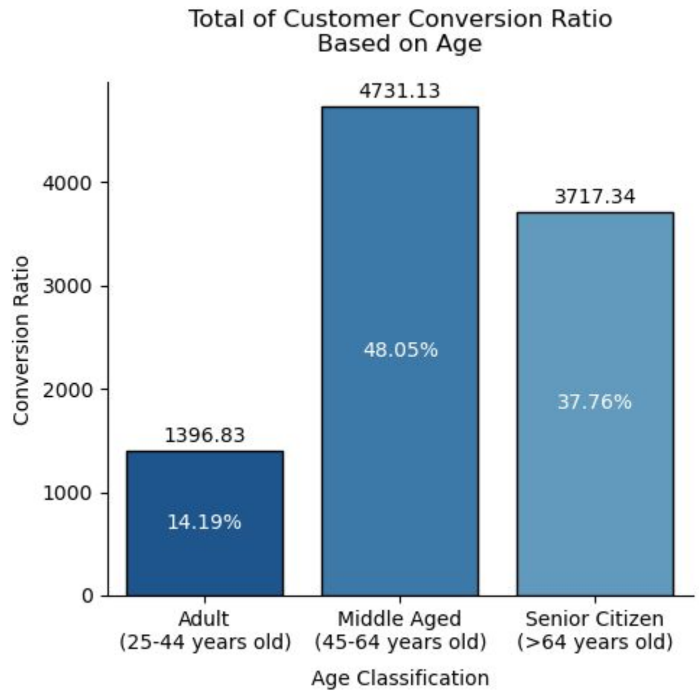
Amount Children Columns → Join columns like :
- Kidhome
- Teenhome

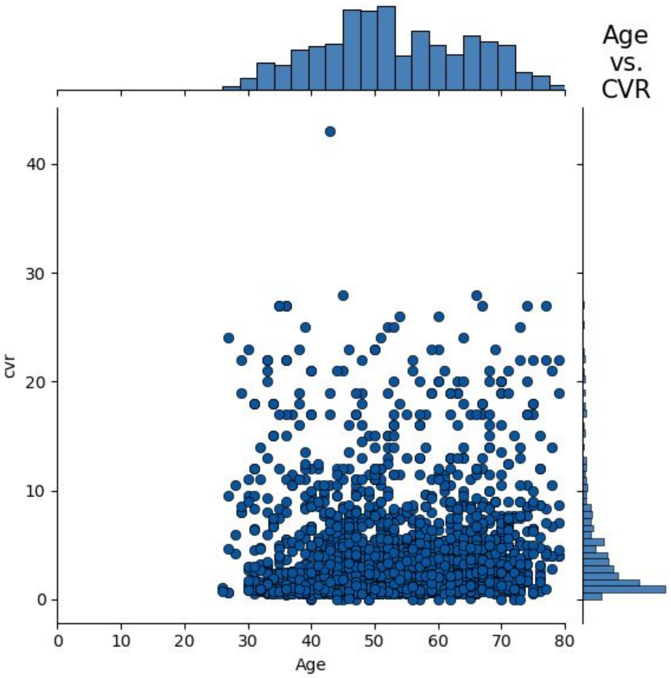
Total Spent Columns → Join total columns like :
- MntCoke
- MntFishProducts
- MntFruits
- MntMeatProducts
- MntSweetProducts
- MntGoldProds

Total days joined → Join columns for integrate to data datetime type like :
- Dt_Collected
- Dt_Customer

Untuk selengkapnya, dapat melihat jupyter notebook [disini]()

Conversion Rate Analysis Based on Income, Spending and Age



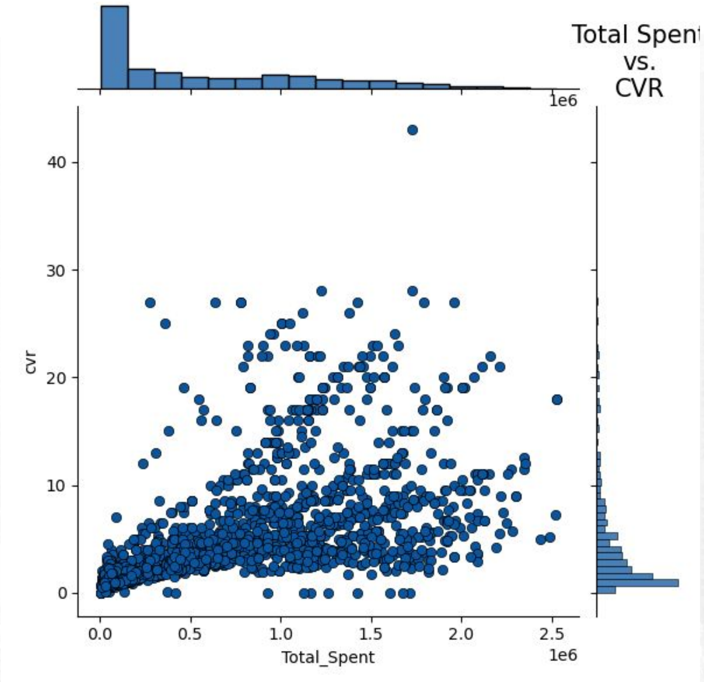


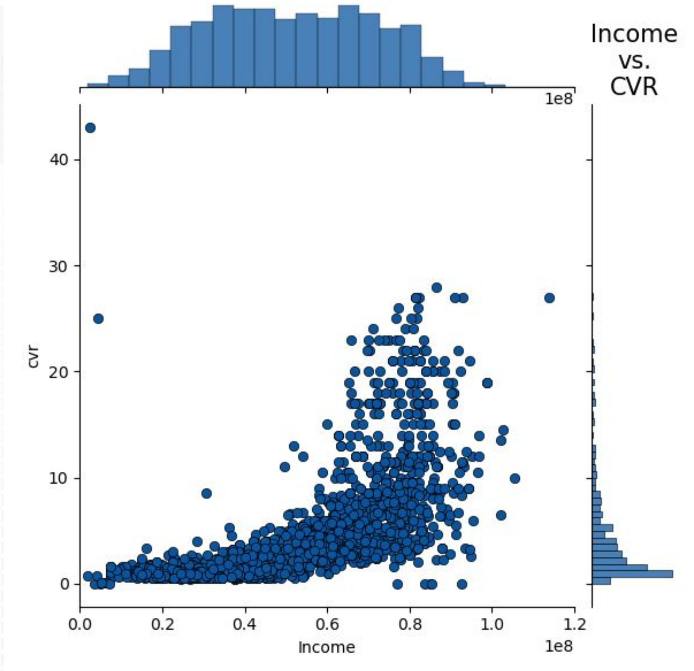
From the data visualization above, the most dominating value is Middle Aged at 48.05% with a distribution like in the picture where Age vs CVR.

With this, you have to pay attention to how Middle Aged can improve their interest in shopping on retail platforms.

Untuk selengkapnya, dapat melihat jupyter notebook disini

Conversion Rate Analysis Based on Income, Spending and Age





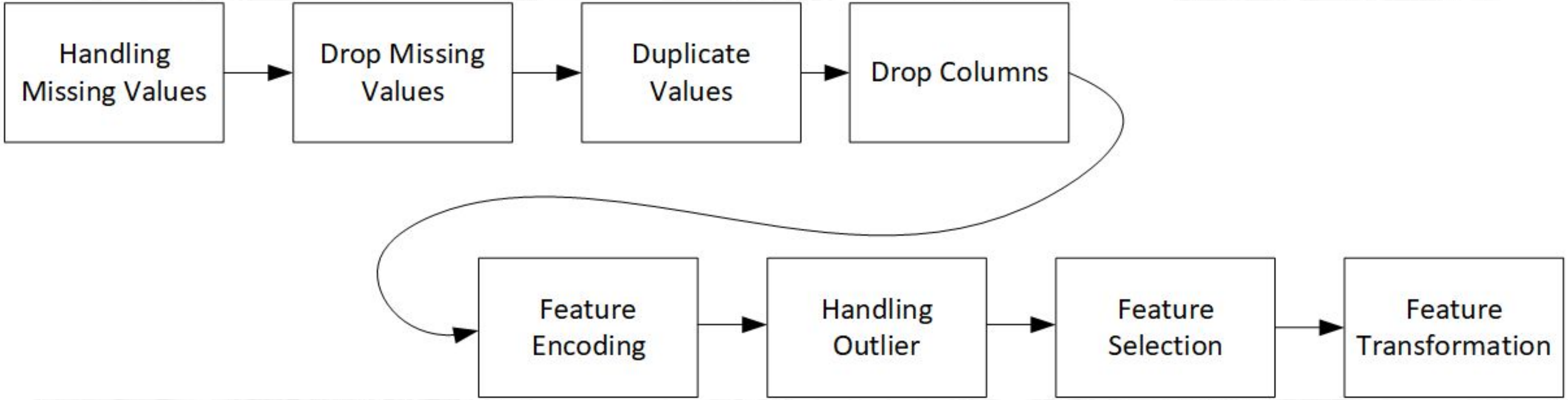
The total spent is known based on the distribution chart, namely between 5-40 conversion rates with an average total spent of more than 1 million / year.
The highest income based on the distribution shown from Visualization Data is more than 60 million / year

Untuk selengkapnya, dapat melihat jupyter notebook [disini]

Data Cleaning & Preprocessing

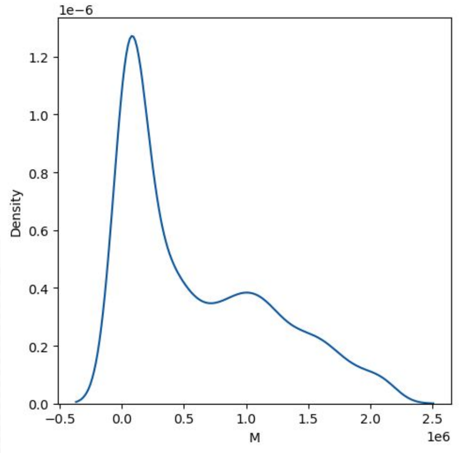
Rakamin Academy

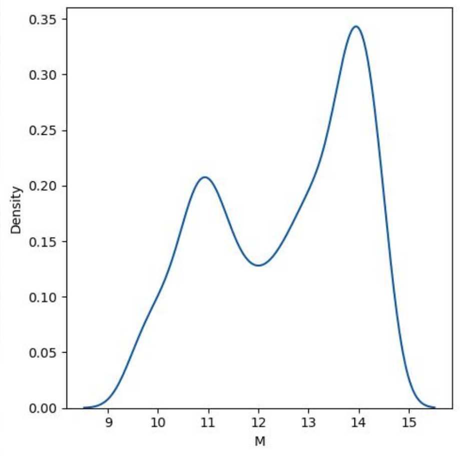
Workflow for Data Cleaning and Data Preprocessing

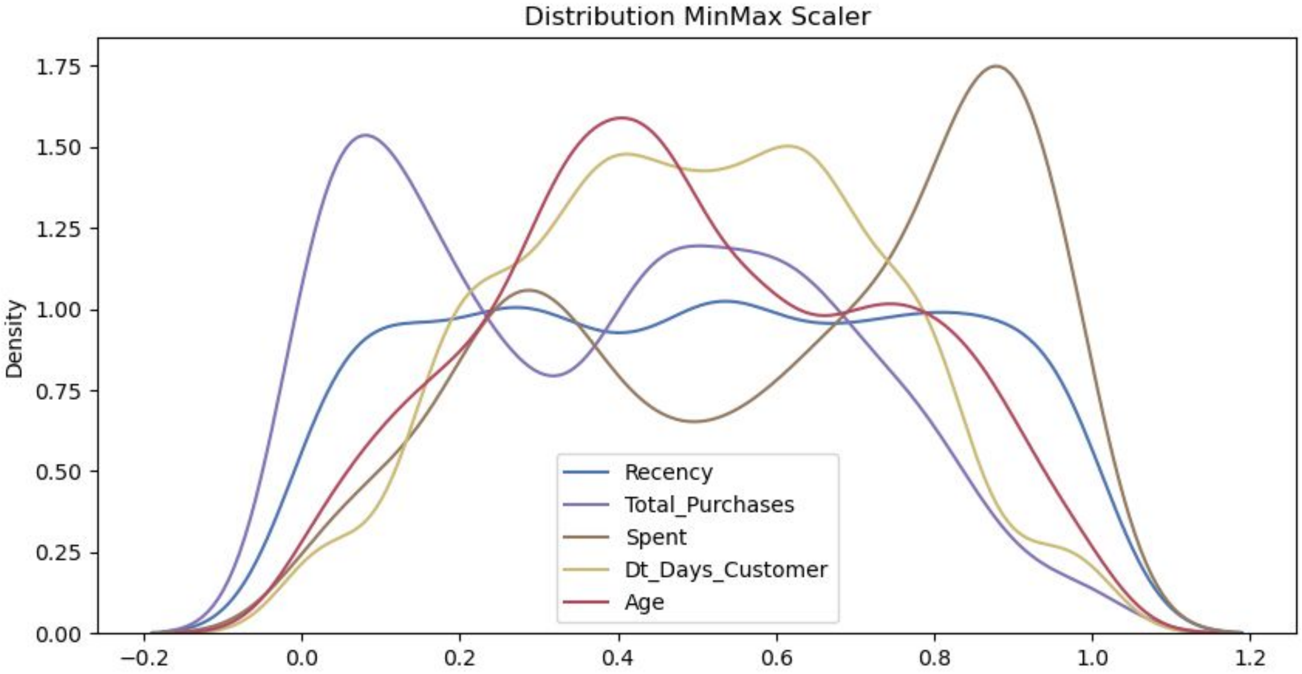


Untuk selengkapnya, dapat melihat jupyter notebook [disini]

Data Cleaning & Preprocessing





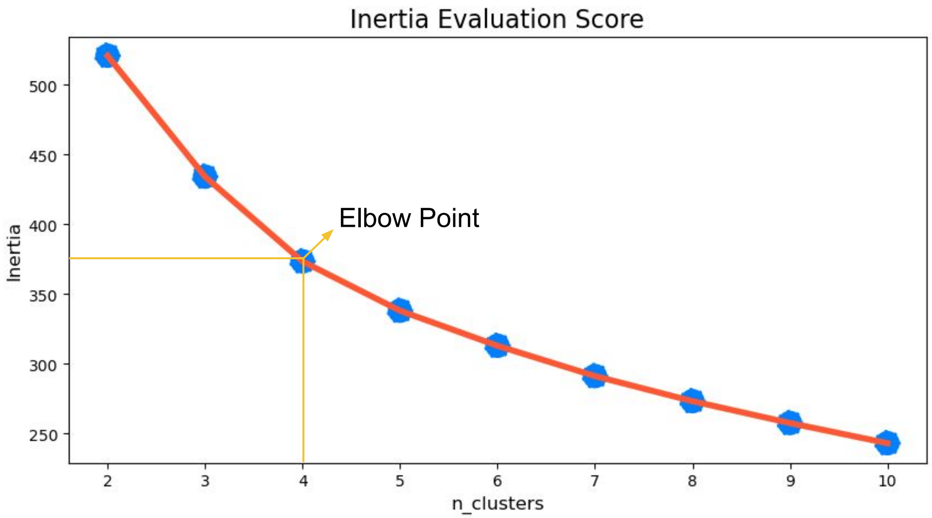


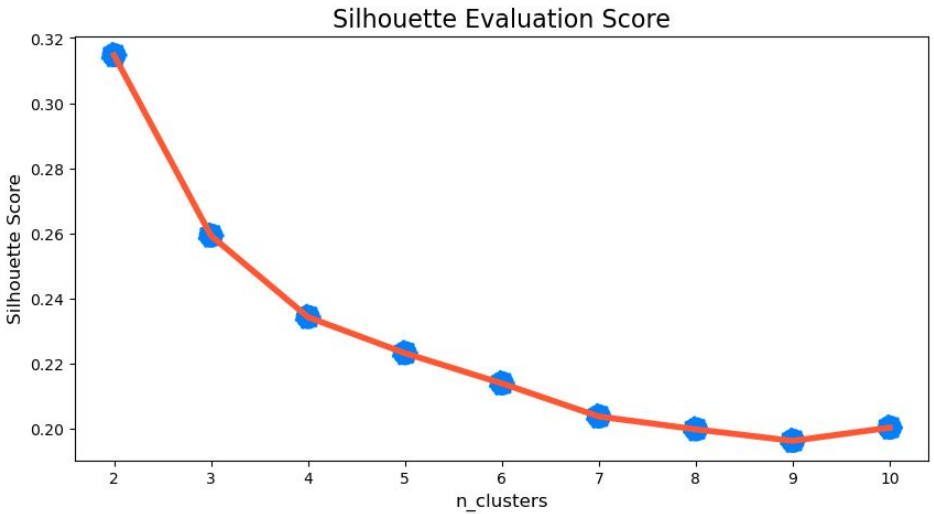
The "Total_Spent" column has a right-skewed distribution, making it unsuitable for K-Means. To adjust it, we change it using the log method so that the distribution becomes more normal. After the transformation, the data distribution is expected to be closer to the normal form, making it easier to analyze with K-Means.

Untuk selengkapnya, dapat melihat jupyter notebook [disini]

Data Modeling

Rakamin Academy

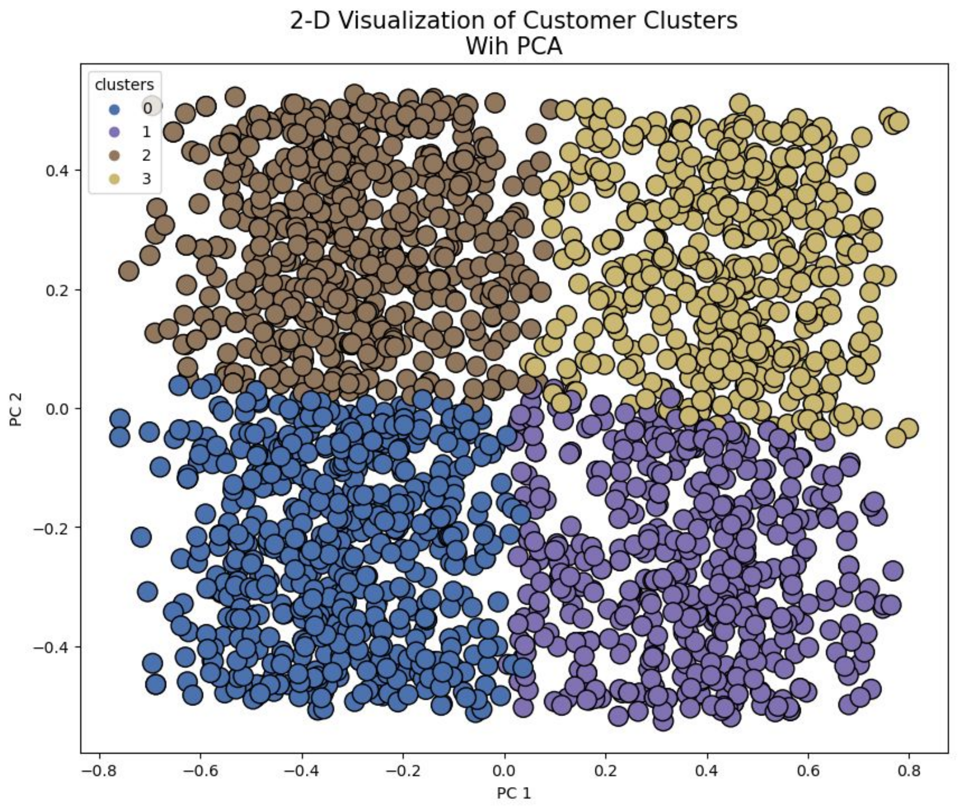




In searching for the optimal number of clusters, I used the elbow method to calculate the inertia score and silhouette score. Based on the evaluation, I found that the best number of clusters is 4. After this point, there is not much significant reduction in the inertia score, and the silhouette score is also better than using 5 clusters. So, n_clusters = 4 is the optimal number for the K-means Clustering model in this dataset.

Untuk selengkapnya, dapat melihat jupyter notebook [disini]

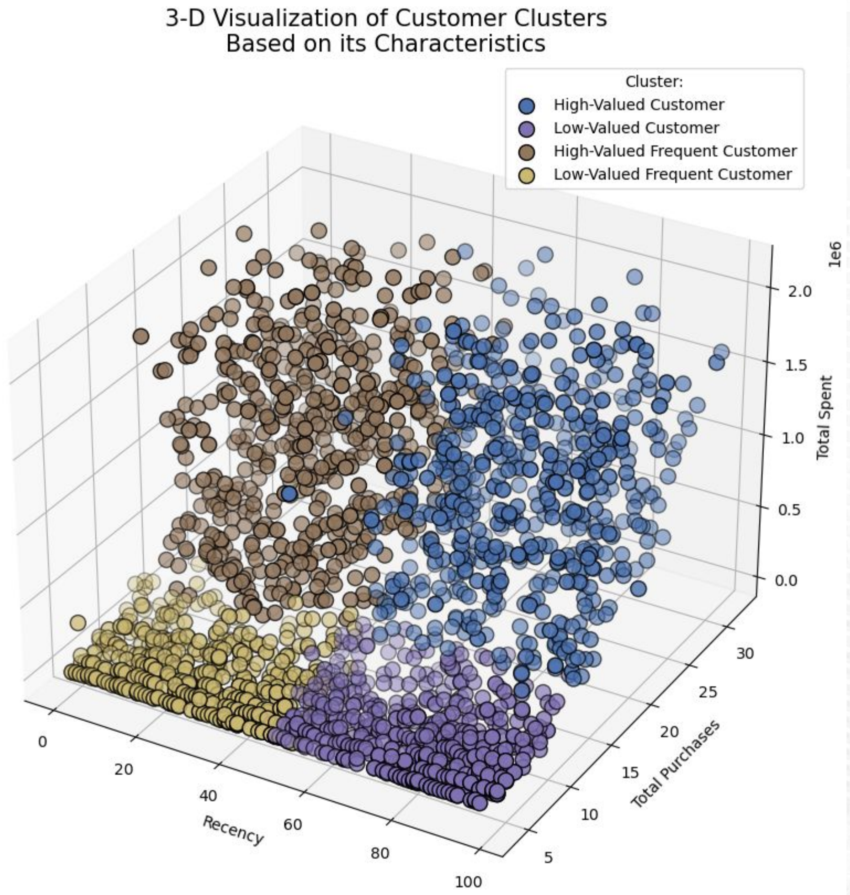
Data Modeling



Visualization results using PCA with 2 main PCs show that the customer clusters are perfectly separated. The K-Means Clustering Algorithm using the RFMLC Method produces 4 clear customer clusters in this dataset.

Untuk selengkapnya, dapat melihat jupyter notebook disini

Customer Personality Analysis for Marketing Retargeting



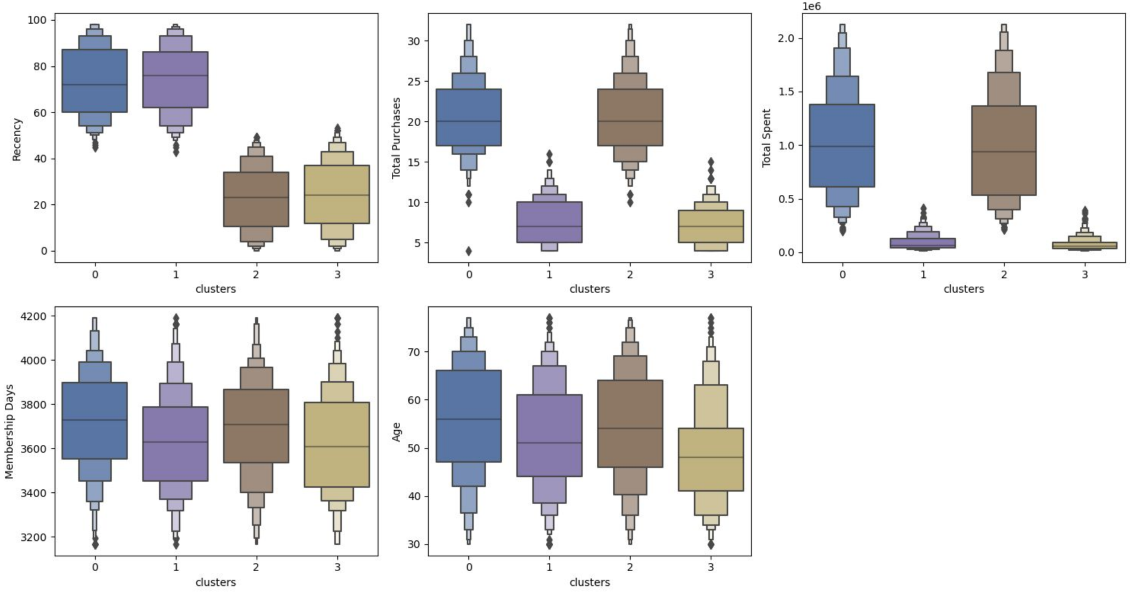
High-Valued Cluster 0 :

has 648 customers (28.93% of the total subscribers). They have high novelty (73 days on average) and high total purchases (21 items on average), indicating high spending on our platform (about 1 million per year). The majority of customers in this group are middle-aged customers (45-64 years) of 48.46%, most have 1 child, and have the highest average income (around IDR 65 million per year) with low web visits per month (average -average 4 times).

Untuk selengkapnya, dapat melihat jupyter notebook disini

Customer Personality Analysis for Marketing Retargeting

Rakamin Academy

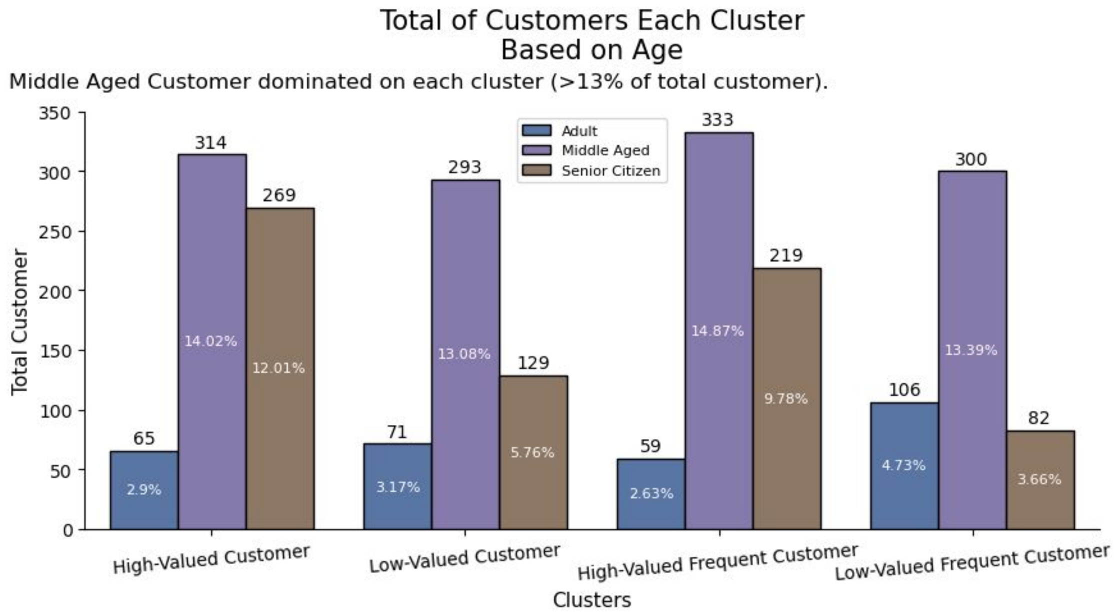


Low-Valued Customers (Cluster 1):

- 493 customers (22.01% of the total) in this group.
- Highest average novelty (74 days) and low purchases (8 items on average), meaning they spend less and less on our platform (around 92k per year).
- Domination by 59.43% middle aged customers (45-64 years) with 1 child and average income (around 36 million per year) and high monthly web visits (6 times on average).

Untuk selengkapnya, dapat melihat jupyter notebook [disini]

Customer Personality Analysis for Marketing Retargeting

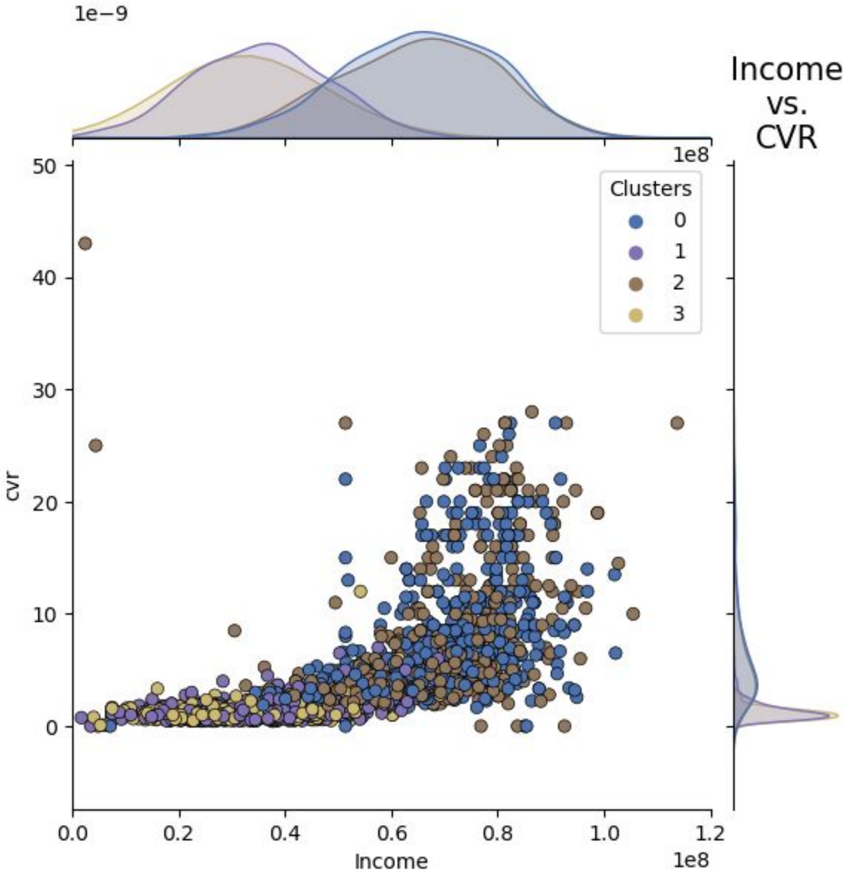


High-Valued Frequent Customers (Cluster 2):

- 611 customers (27.28% of the total) in this group.
- Low average novelty (23 days) and high purchases (21 items on average), meaning they shop frequently and a lot on our platform (around 989k per year).
- Domination by 54.5% middle aged customers (45-64 years) with 1 child and average income (about 65 million per year) with low monthly web visits (4 times average).

Untuk selengkapnya, dapat melihat jupyter notebook disini

Customer Personality Analysis for Marketing Retargeting



Income vs. CVR

Low-Valued Frequent Customers (Cluster 3):

• 488 customers (21.79% of the total) in this group.
• High average recency (24 days) and lowest purchases (average 7 items), meaning they spend often but little on our platform (around 75 thousand per year).
• Domination by 61.48% middle aged customers (45-64 years) with 1 child and average income (around 35 million per year) with high monthly web visits (6 times on average).

Untuk selengkapnya, dapat melihat jupyter notebook [disini]

Recommendation

Insights:

Create a membership tier program (Platinum, Gold, Silver, Bronze) with different privileges for each customer group (High Rated Customer, High Rated Frequent Customer, Low Rated Frequent Customer, Low Rated Customer).

Prioritize focusing on a group of High-Valued Customers to prevent churn. Improve service, after-sales maintenance and product quality. Provide Platinum membership with discounts, promotions and free shipping to encourage more frequent shopping.

Untuk selengkapnya, dapat melihat jupyter notebook disini