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:
Muhammad Miftah Thaha
muhammadmiftaht@gmail.com
https://www.linkedin.com/in/miftahthaha/
https://github.com/miftahthaha

Miftah is a recent graduate with a Bachelor's degree Electrical Engineer who have interest in Data Analytics and Science and have a strong foundation in statistical modeling, data analysis, and programming. As a Junior Data Scientist, he has experience through his final project in building and implementing machine learning models, analyzing complex data sets, and creating visualizations to communicate insights. He is a fast learner with excellent problem-solving skills and a passion for using data to drive business decisions. In addition, he possess strong communication and collaboration skills, having worked on multiple team projects during his studies. With a drive to excel in his field, Miftah is seeking an opportunity to contribute his skills and knowledge to a dynamic and innovative organization as a Junior Data Scientist.

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)
```

- **Description**
  Dataset that contains information related to marketing campaign made by Store or E-Commerce.

- **Shape**
  2240 Row and 30 Columns (Feature)

- **Datatypes**
  Float64 (1 Feature), Int64 (26 Feature), object (3 Feature)

- **Missing Values**
  Income (24 Values)

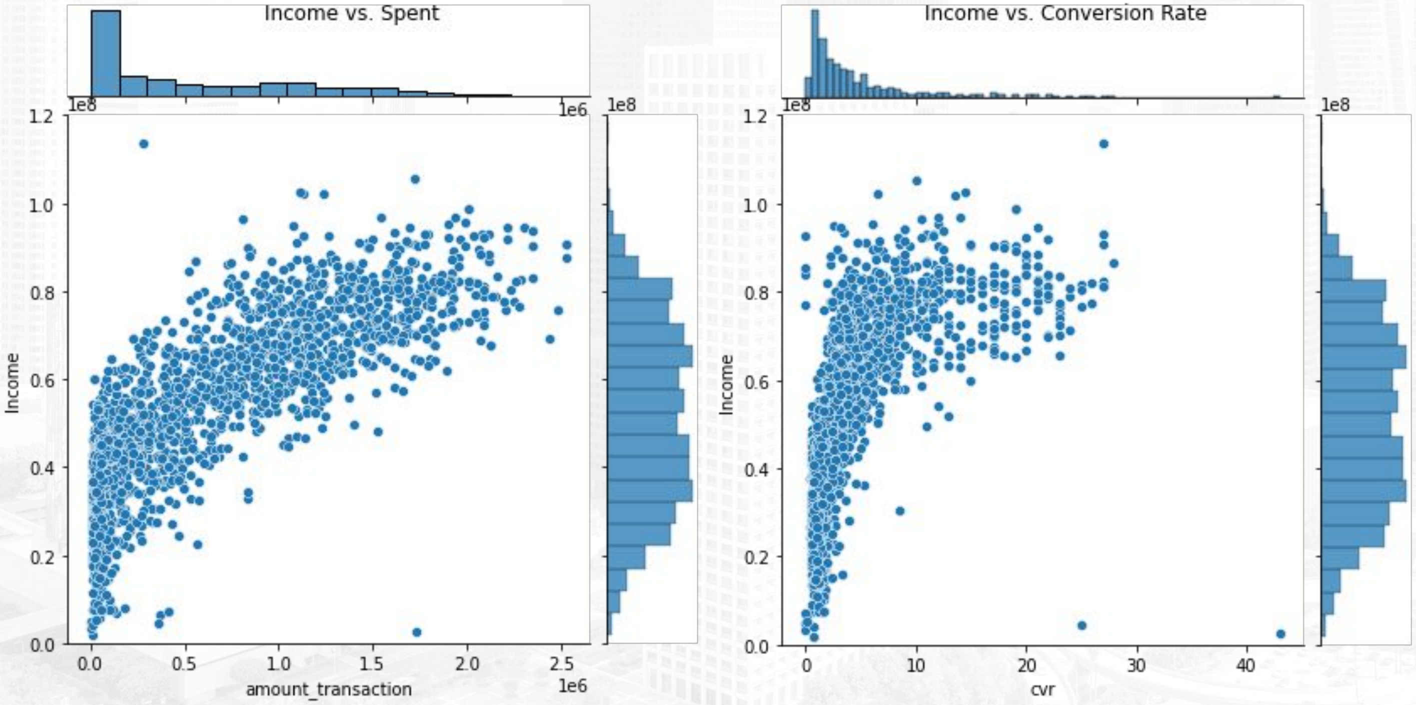
For more details, you can see all file here and code here

Exploratory Data Analysis

1. Feature Engineering:
   - Age & Age Grouping
   - Dependent
   - Total Purchases
   - Amount Transaction
   - Total Accepted Campaign
   - Conversion Rate
2. Multivariate Analysis
3. Conversion Rate Analysis Based on Income, Spending and Age

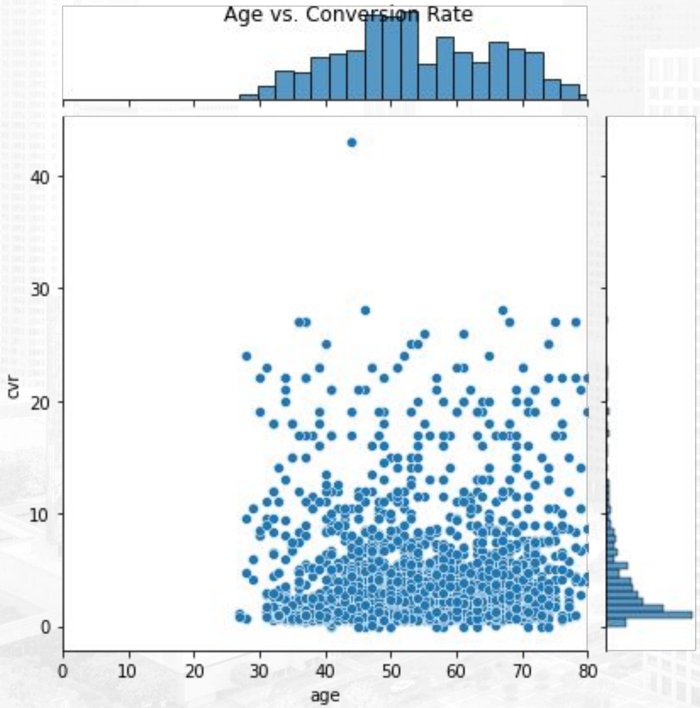
For more details, you can see all file here and code here

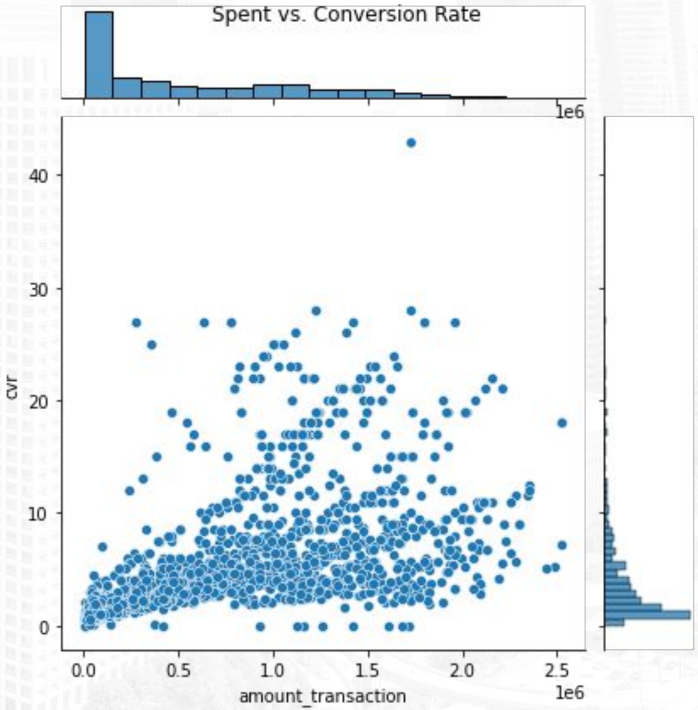
Conversion Rate Analysis Based on Income, Spending and Age



For more details, you can see all file here and code here

Conversion Rate Analysis Based on Income, Spending and Age





For more details, you can see all file here and code here

Conversion Rate Analysis Based on Income, Spending and Age

Rakamin Academy

Based on the analysis, it appears that customers with higher income tend to spend more and have a higher total expenditure on our store/platform. This trend, however, does not seem to apply to the age feature. In other words, while income seems to positively correlate with conversion rate and spending , age does not show a significant correlation with conversion rate. This information could potentially be useful for businesses looking to target specific customer segments and tailor their marketing strategies accordingly.

For more details, you can see all file here and code here

Data Cleaning & Preprocessing

Rakamin Academy

- Handling Null Value

1 Column Had 24 Value (1.0714%) Null Values on Income Column and fill it with Income Median.

Income	24
Income	1.0714

- Handling Duplicated Value

No Duplicated Values

```
df1.duplicated().sum()
0
```

- Handling Unnecesary Values

Drop Column which are not needed.

```
# Drop Unnecesary Column
df1 = df1.drop(columns=['Unnamed: 0', 'Kidhome', 'Teenhome'])
df1.sample(5)
```

```
# label encoder
map_edu = {
    'SMA' : 0,
    'D3' : 1,
    'S1' : 2,
    'S2' : 3,
    'S3' : 4
}

df1['edu_map'] = df1['Education'].map(map_edu)
```

- Feature Encoding

Label encoding will be applied to the "education" column as it will be used in the modeling process.

For more details, you can see all file here and code here

Data Cleaning & Preprocessing

- Feature Selection

  RFMLECA analysis is an extended version of RFM analysis that used to divide customers into several segments. Based on RFMLECA analysis, we will need 7 variables:

  1. R (Recency)
  2. F (Frequency)
  3. M (Monetary)
  4. L (Length Joining)
  5. E (Education)
  6. C (Campaign)
  7. A (Age)

- Outlier Handling

  Handling Outlier with IQR Method (Q1: 0.01; Q3: 0.99)

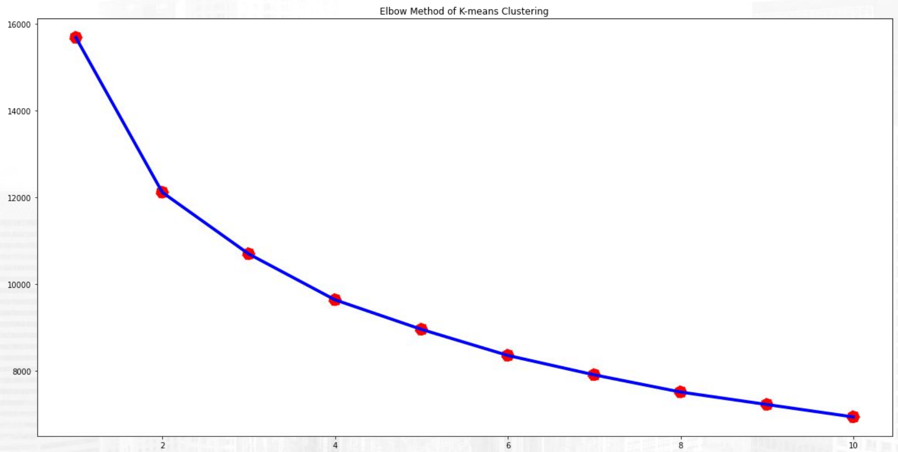
- Feature Standarization

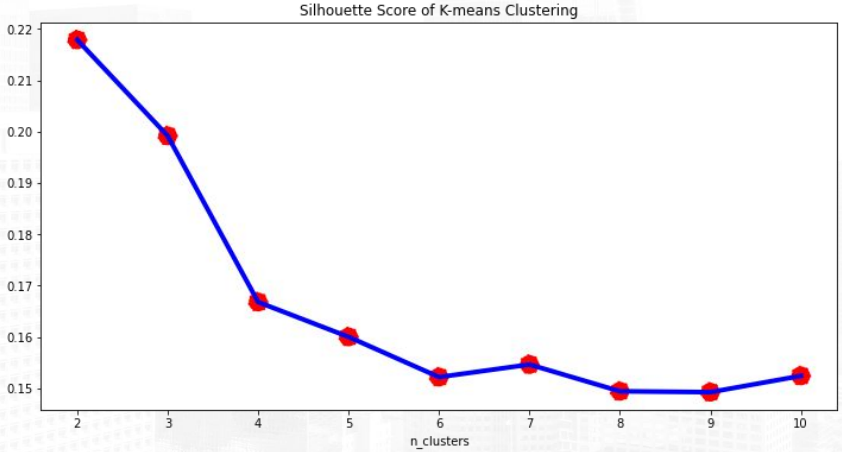
  Standarization to selected column using StandardScaler.

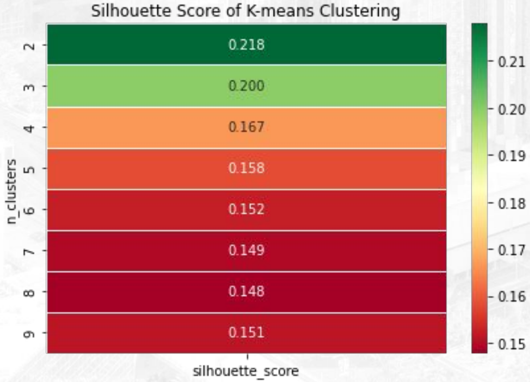
```
std = StandardScaler()
custvalue_std = std.fit_transform(df2)
custvalue_std
```

For more details, you can see all file here and code here

Data Modeling



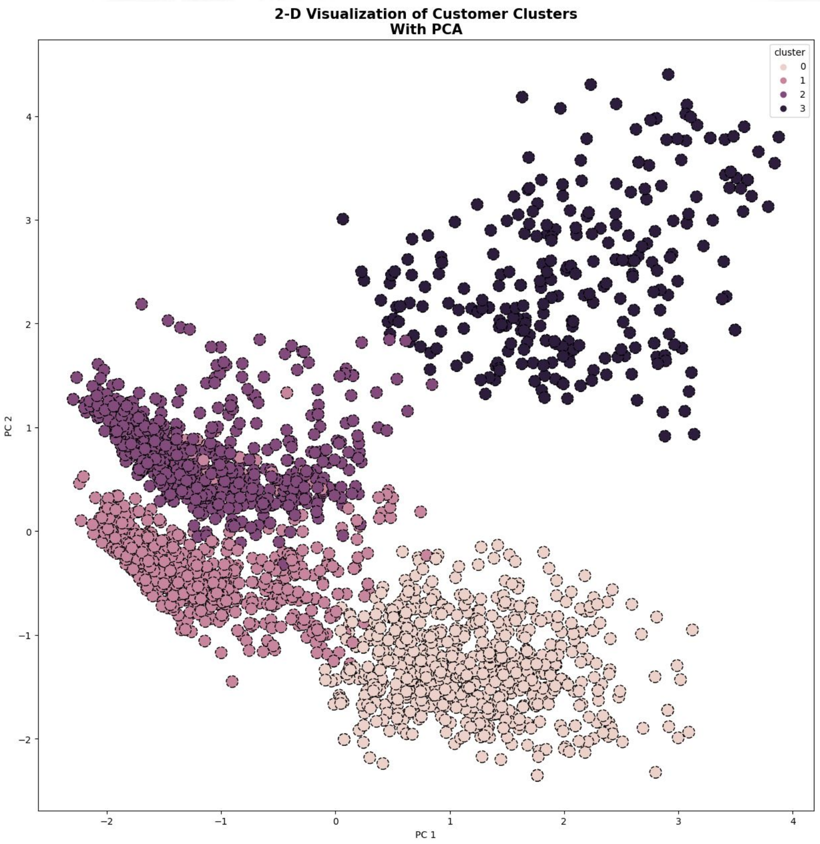




The optimal n_clusters for the K-means Clustering Model on this dataset is 4. This was determined by using the elbow method to evaluate the optimal n_clusters by examining the inertia score and then validated it using the silhouette score. The evaluation revealed that the elbow point is at n_clusters = 4 because there is no significant decrease in the inertia score after this point. Furthermore if we look into elbow methods score, the silhouette score indicates that n_clusters = 4 is better than n_clusters > 4. Therefore, it can be concluded that the optimal n_clusters for this dataset is 4.

For more details, you can see all file [here]() and code [here]()

Customer Personality Analysis for Marketing Retargeting



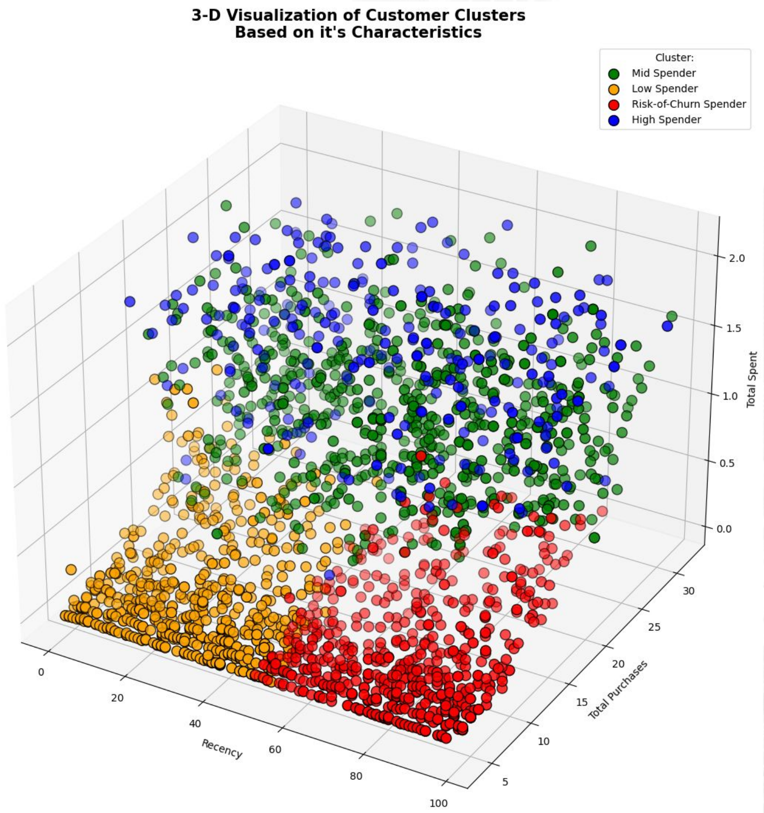
The visualization of the dataset using PCA with 2 primary components indicates that the clusters are separated. The K-Means Clustering algorithm applied with RFMLC method generated four customer clusters for this dataset.

For more details, you can see all file here and code here

Customer Personality Analysis for Marketing Retargeting

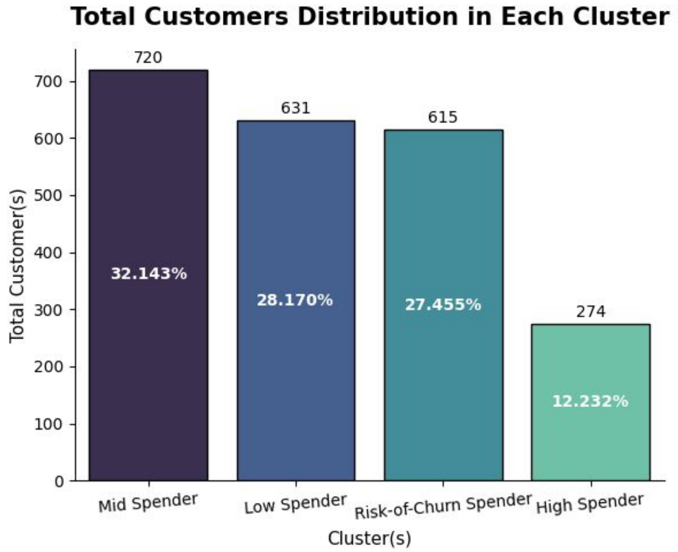
There are 4 Customer Cluster based on RFMLCA metrics:

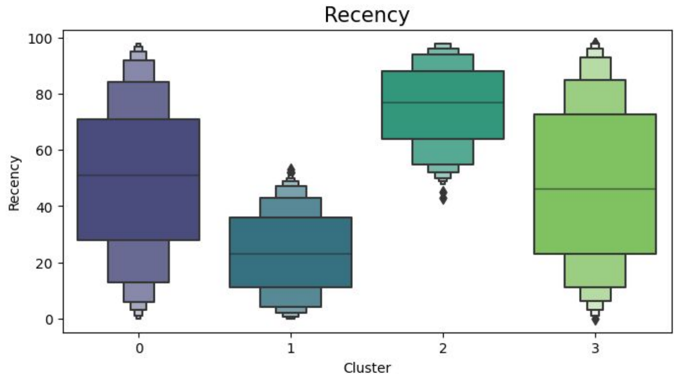
1. Cluster 0: Mid Spender
2. Cluster 1: Low Spender
3. Cluster 2: Risk-of-Churn Spender
4. Cluster 3: High Spender



For more details, you can see all file here and code here

Customer Personality Analysis for Marketing Retargeting





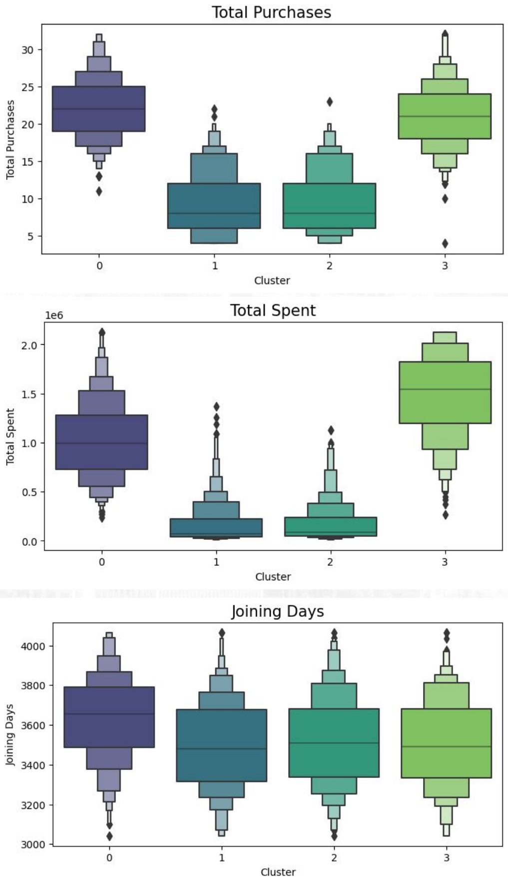
Cluster 0: Mid-High Spender

- Largest group with 720 users
- Dominated by late adults (36-65 years old), mostly married, and without dependents at home
- Second-highest average income and expenditure positions (IDR 65.2M/year and IDR 1.02M/year respectively)
- Relatively low average NumWebVisitMonth (5 times a month)
- Most recent joined days (3635 days joined)
- Highest average total purchases (22 items)
- Second-highest average recency (49 days)
- Not frequent shoppers but big spenders
- Not very responsive to campaigns (Organic customer acquisition)

For more details, you can see all file here and code here

Cluster 1: Low Spender

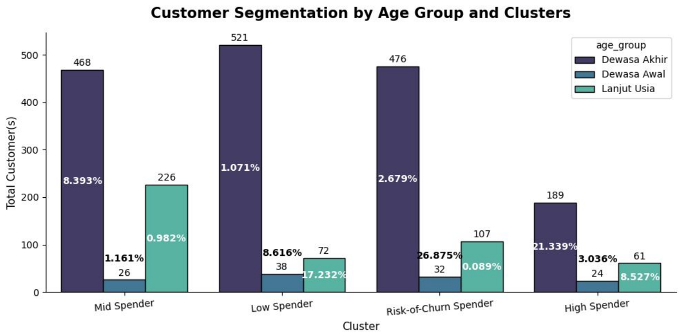
- Second-largest group with 631 users
- Dominated by late adults (36-65 years old), mostly married, and without dependents at home
- Lowest average income and expenditure positions (IDR 39.2M/year and IDR 161k/year respectively)
- Relatively high average NumWebVisitMonth (6 times a month)
- Most recent average joined days (3498 days joined)
- Lowest average total purchases (9 items)
- Lowest average recency (23 days)
- Frequent shoppers with small purchases
- Not very responsive to campaigns (Organic customer acquisition)

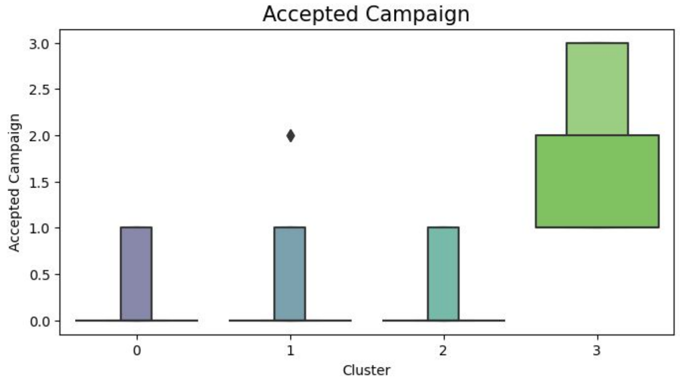


For more details, you can see all file here and code here

Customer Personality Analysis for Marketing Retargeting

Rakamin Academy





Cluster 2: Risk of Churn

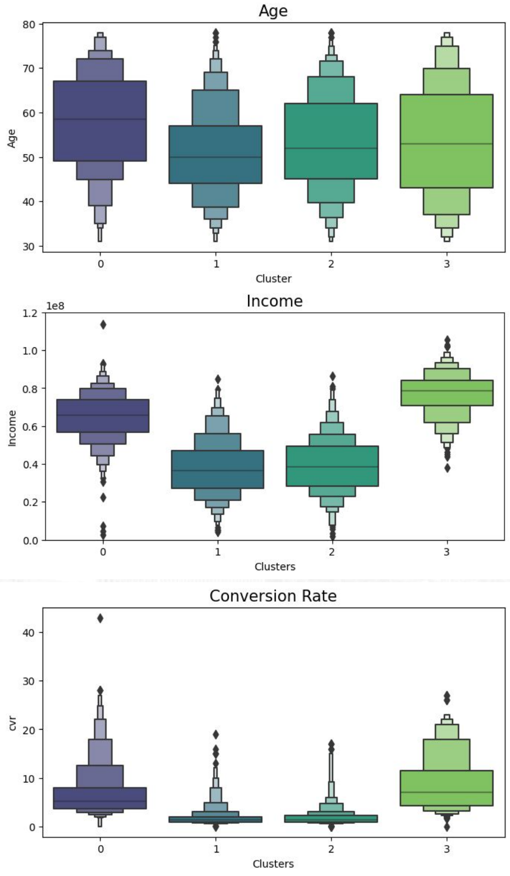
- Third-largest group with 615 users
- Dominated by late adults (36-65 years old), mostly married, and with dependents at home
- Second-lowest average income and expenditure positions (IDR 39.3M/year and IDR 170k/year respectively)
- Relatively high average NumWebVisitMonth (6 times a month)
- Second-oldest average joined days (3518 days joined)
- Lowest average total purchases (9 items)
- Highest average recency (75 days)
- Not frequent shoppers and small purchases
- Not very responsive to campaigns (Organic customer acquisition)

For more details, you can see all file here and code here

Customer Personality Analysis for Marketing Retargeting

Cluster 3: High Spender

- Smallest group with 274 users
- Dominated by Late adults (36-65 years old), mostly married, and with dependents at home
- Highest average income and expenditure positions (around IDR 76.9M/year and IDR 1.51M/year respectively)
- Lowest average NumWebVisitMonth (3 times a month)
- Second-recent average joined days (3510 days joined)
- Second-highest average total purchases (21 items)
- Second-lowest average recency (47 days)
- Frequent shoppers and big spenders
- Very responsive to campaigns (Non-Organic customer acquisition)



For more details, you can see all file here and code here

Recommendation

Implement targeted marketing campaigns by membership program:

To further increase customer retention and attract more customers to shop on our platform, it is recommended to create a membership tier program. The program can have four membership tiers (Platinum, Gold, Silver, and Bronze) based on the customer clusters identified in the analysis (Platinum: High Spender, Gold: Mid Spender, Silver: Low Spender, Bronze: Risk of Churn).

Each membership tier can have different privileges for customers, with the highest membership tier receiving the greatest privileges. For example, Platinum members can receive exclusive access to high-end products, personalized promotions, and free shipping on all orders, while Gold members can receive early access to sales and discounts, and personalized product recommendations. Silver members can receive limited-time promotions and early access to new products, while Bronze members can receive discounts on select products.

For more details, you can see all file here and code here

Recommendation

Rakamin Academy

Improve website user experience: Given that the website visit frequency is an important factor in predicting customer behavior, it is recommended to improve the website user experience to encourage customers to visit more often. This can be achieved by optimizing the website design, improving site speed, and making it more user-friendly.

Increase product offerings: Since customers in the Low Spender and Risk of Churn clusters tend to make smaller purchases, it may be beneficial to expand the product offerings to include more affordable options. This can help attract more customers and encourage them to make more frequent purchases.

Focus on customer retention: The Risk of Churn cluster is particularly at risk of leaving, so it is important to focus on customer retention efforts for this segment. This can be achieved by offering personalized promotions or special deals, providing excellent customer service, and addressing any complaints or issues promptly.

For more details, you can see all file here and code here

Potential Impact

If we focus on continuously monitoring the High Spender group, we will still get potential GMV of IDR 413 million, GMV for the Mid Spender group will be IDR 736 million, while the GMV for the Low Spender Group will be IDR 101 million.

For more details, you can see all file here and code here

Thank You!

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