Predict Customer Personality to boost marketing campaign by using Machine Learning

Created by:
Yasmin Fauziah
yasminfauziah63@gmail.com
https://www.linkedin.com/in/yasmin-fauziah-85b738239/

"Bachelor of Physics from Padjadjaran University> Someone who enjoys learning new things, has good analytical and planning skill. Enjoy to solve problem related to data analysis using Excel, SQL, Phython and Looker Studio. Have a high interest in a career in the data field."

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

Overview

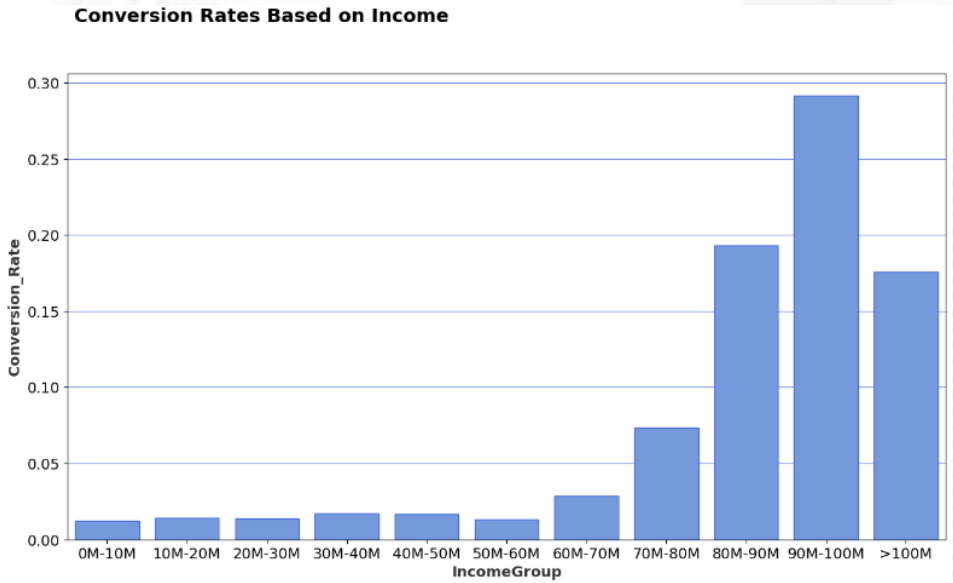
"A company can grow rapidly when it knows its customer personality behavior, so that it can provide better services and benefits to customers who have the potential to become loyal customers. By processing historical marketing campaign data to improve performance and target the right customers so that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

Feature Engineering

No	New Column	Description
1	Convension_Rate	Total purchases divided by number of web visit per month
2	Children	Sum of Kid Home and Teen home
3	Age	Difference between current year and birth year
4	Age_Group	Segmentation of several age ranges ranging from age 20 to age 80 with a difference of every 10 years
5	Total Spending	Sum from column that have "Mnt" at the column name
6	Spending_Group	Segmentation of several spending ranges ranging from 0 M to > 2,5 M with a difference of every 0,5 M
7	Total_Transaction	Sum from column that have 'Num' at the column name
8	Transaction_Group	Segmentation of several transaction ranges ranging from 0 to > 40 with a difference of every 10 purchases
9	Income_Group	Segmentation of several income ranges ranging from 0 M to > 100 M with a difference of every 10 M

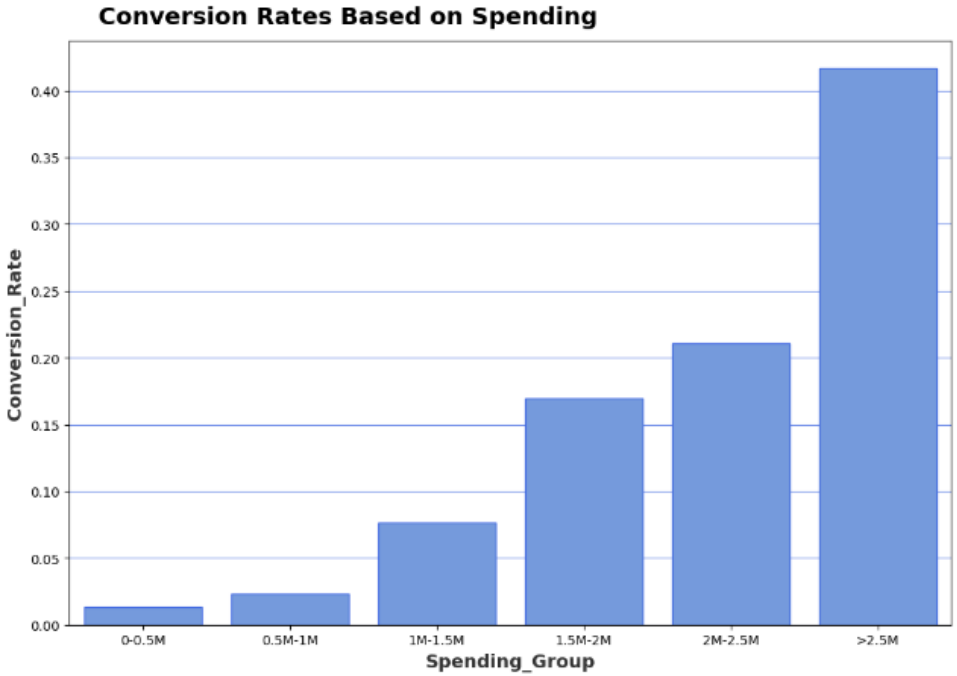
Untuk selengkapnya, dapat melihat jupyter notebook disini

Conversion Rate Analysis Based on Income



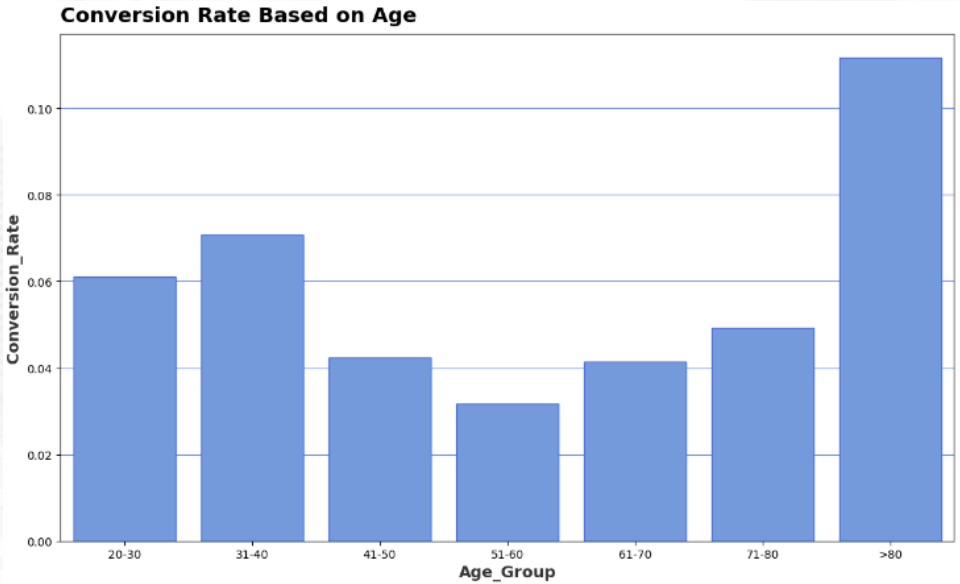
- The difference in income from the customer affectsthe conversion rate value, where for theconversion rate generally gets higher as income gets higher. The highest conversion rate value occurs in the income range of 90-100 million.
- To increase the value ofconversion rate, it would be better if the marketing teamprioritize target customers who fall intoin the high income category, with income >= 80 million.

Conversion Rate Analysis Based on Spending



- The difference in the number of transactions from customersaffect the conversion rate value, wherethe higher the amount of spending, the higher the conversion rate.
- The marketing team can maintain or improve its good performance, by providing rewards in the form of discounts for customers who spend above 2.5 million so that spending in the highest category can continue to increase.

Conversion Rate Analysis Based on Age



- The age difference of the customer is not too differentsignificantly for the conversion rate value in the 20-80 age range, the highest conversion rate value is in the >80 age range.
- The difference in age range has a considerable influence on the conversion rate. Recommendations that can be made include segmenting the market and providing content that is suitable for each segmentation based on age.

Data Cleaning & Preprocessing

```
df2.isna().sum().sort_values(ascending = False)
```

```
Income_Group          24
Income                24
Conversion_Rate       11
```

```
missing_value = df2.isna().sum() *100/len(df)
print(round(missing_value, 4).sort_values(ascending=False))
```

```
Income_Group          1.0714
Income                1.0714
Conversion_Rate       0.4911
```

```
df2.duplicated().sum()
```

```
0
```

```
Before Filtered are 38 columns
After Filtered are 16 columns
```

Missing Value

- Features that have missing values include Income_Group, Income, Converson_Rate.
- Since the number of Missing values is less than 2.5% and will not significantly affect the data, we drop them.

Data Duplicate

- There is no duplicate data

Remove Data

- Remove unnecessary feature for further feature encoding

Untuk selengkapnya, dapat melihat jupyter notebook disini

Data Cleaning & Preprocessing

Rakamin Academy

```
cat_cols = df2.select_dtypes(include='object').columns.tolist()
for col in cat_cols:
    print(f'Number of Unique Value {col} is {df2[col].nunique()}:')
    print(sorted(df[col].unique().tolist()))
    print('\n')
```

```
Number of Unique Value Education is 5:
['D3', 'S1', 'S2', 'S3', 'SMA']


Number of Unique Value Marital_Status is 6:
['Bertunangan', 'Cerai', 'Duda', 'Janda', 'Lajang', 'Menikah']


Number of Unique Value Age_Group is 7:
['20-30', '31-40', '41-50', '51-60', '61-70', '71-80', '>80']


Number of Unique Value Spending_Group is 6:
['0-0.5M', '0.5M-1M', '1.5M-2M', '1M-1.5M', '2M-2.5M', '>2.5M']
```

Feature Encoding

- The feature encoding stage is to convert the object data type into an integerbe data type with the aim of facilitating the modeling process.
- we perform feature encoding for several features including Eduacation, Marital_Status, Age_Group, Spending_Group.

	Education	Marital_Status	Income	Recency	NumWebVisitsMonth	Complain	Response	Conversion_Rate	Age	Children	Total_Spending	Total_Transaction	Age_Group	Incon
0	2	Lajang	58138000.0	58	7	0	1	0.14	67	0	1617000	25	4.0	
1	2	Lajang	46344000.0	38	5	0	0	0.00	70	2	27000	6	4.0	
2	2	Bertunangan	71613000.0	26	4	0	0	0.00	59	0	776000	21	3.0	
3	2	Bertunangan	26646000.0	26	6	0	0	0.00	40	1	53000	8	1.0	
4	4	Menikah	58293000.0	94	5	0	0	0.00	43	1	422000	19	2.0	

me_Group	Spending_Group	Transaction_Group	MaritalStatus_Bertunangan	MaritalStatus_Cerai	MaritalStatus_Duda	MaritalStatus_Janda	MaritalStatus_Lajang	MaritalStatus_Menikah
5	3	2.0	0	0	0	0	1	0
4	0	0.0	0	0	0	0	1	0
7	1	2.0	1	0	0	0	0	0
2	0	0.0	1	0	0	0	0	0
5	0	1.0	0	0	0	0	0	1

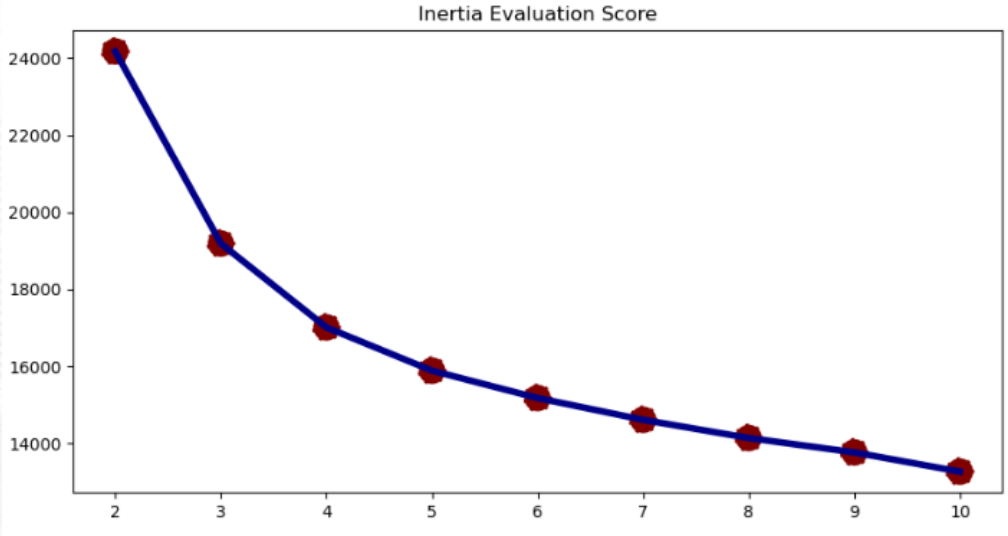
Data Cleaning & Preprocessing

	Income	Recency	NumWebVisitsMonth	Age	Total_Spending	Total_Transaction
0	0.247791	0.312501	0.689304	0.988076	1.679389	1.330024
1	-0.228623	-0.379228	-0.142764	1.238443	-0.960344	-1.161223
2	0.792109	-0.794266	-0.558798	0.320431	0.283153	0.805551
3	-1.024317	-0.794266	0.273270	-1.265228	-0.917179	-0.898986
4	0.254052	1.557614	-0.142764	-1.014861	-0.304561	0.543315
...	...	...	...	...	...	...
2235	0.372409	-0.102536	-0.142764	0.153519	1.221172	0.412196
2236	0.485150	0.243328	0.689304	1.906089	-0.268037	0.936670
2237	0.201055	1.453854	0.273270	-1.014861	1.055151	0.543315
2238	0.696454	-1.416822	-0.974832	1.071532	0.394387	1.067788
2239	0.034952	-0.310055	0.689304	1.238443	-0.719614	-0.505632

Standardization

- Standardizaion is the process of changing the scale of the data so that the values have the same scale to facilitate the process at the next stage.
- Features that are standardized include Income, Recency, NumWebVisitsMonth, Age, Total_Spending, Total_Transaction

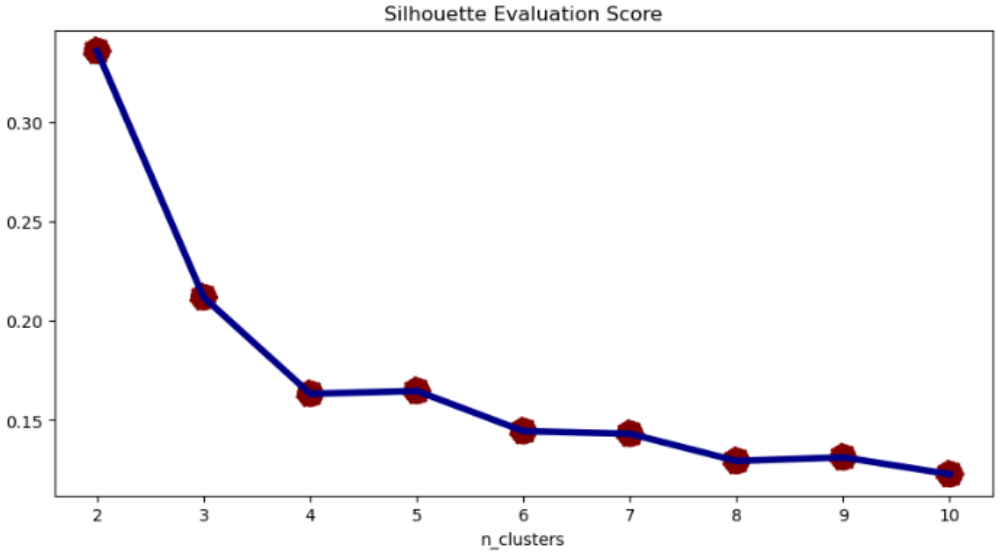
Data Modeling



The graph above is the elbow method in terms of inertia where the size describes how far the data points in 1 cluster are from the cluster center. A significant decrease in distortion score value occurs from point 2 to point 3, after which the decrease in distortion score value is not too significant. This indicates that the elbow point of the plot is at k = 3, so the cluster recommendation from the dataset is divided into 3 clusters.

Untuk selengkapnya, dapat melihat jupyter notebook disini

Data Modeling



Comparison between silhouette score metrics and number of clusters. Silhouette Score measures how close each data point is to the cluster they belong to compared to other clusters. The range of Silhouette Score values is -1 to 1, where a positive value indicates that the object is in the right cluster, while a negative value indicates that the object may be placed in the wrong cluster. In the Silhouette Score plot, the highest score is in cluster range 2 with a score value > 0.3.

Untuk selengkapnya, dapat melihat jupyter notebook disini

Customer Personality Analysis for Marketing Retargeting

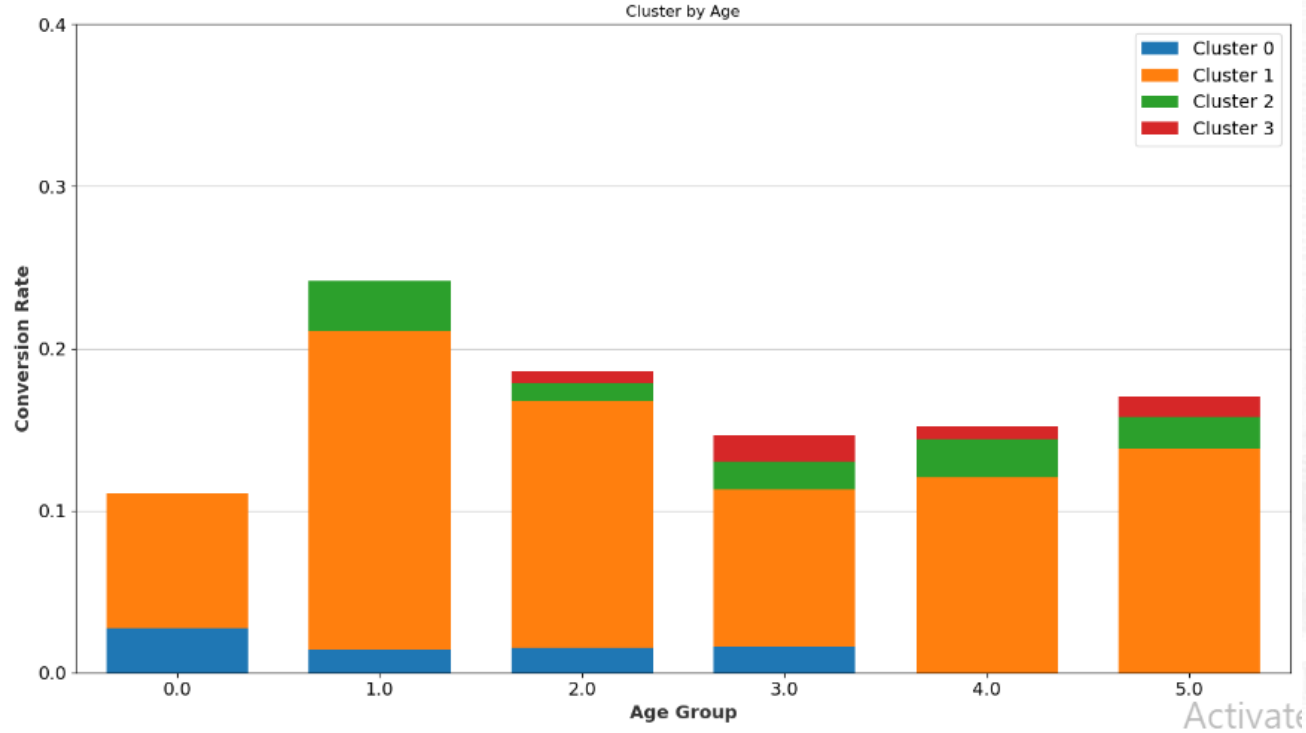
The following is a visualization of some of the features for Cluster 0, 1, 2 and 3:

1. **Income per Cluster**: This plot shows the income profile for each cluster. Cluster 1 has the largest income, while Cluster 3 has the lowest income.
2. **Age per Cluster**: This plot shows the age profile for each cluster. Cluster 1 tends to have a higher age, while Cluster 3 has a lower age.
3. **Spending per Cluster**: This plot shows the average product purchase amount within each cluster. Cluster 1 stands out with a higher number of web purchases.

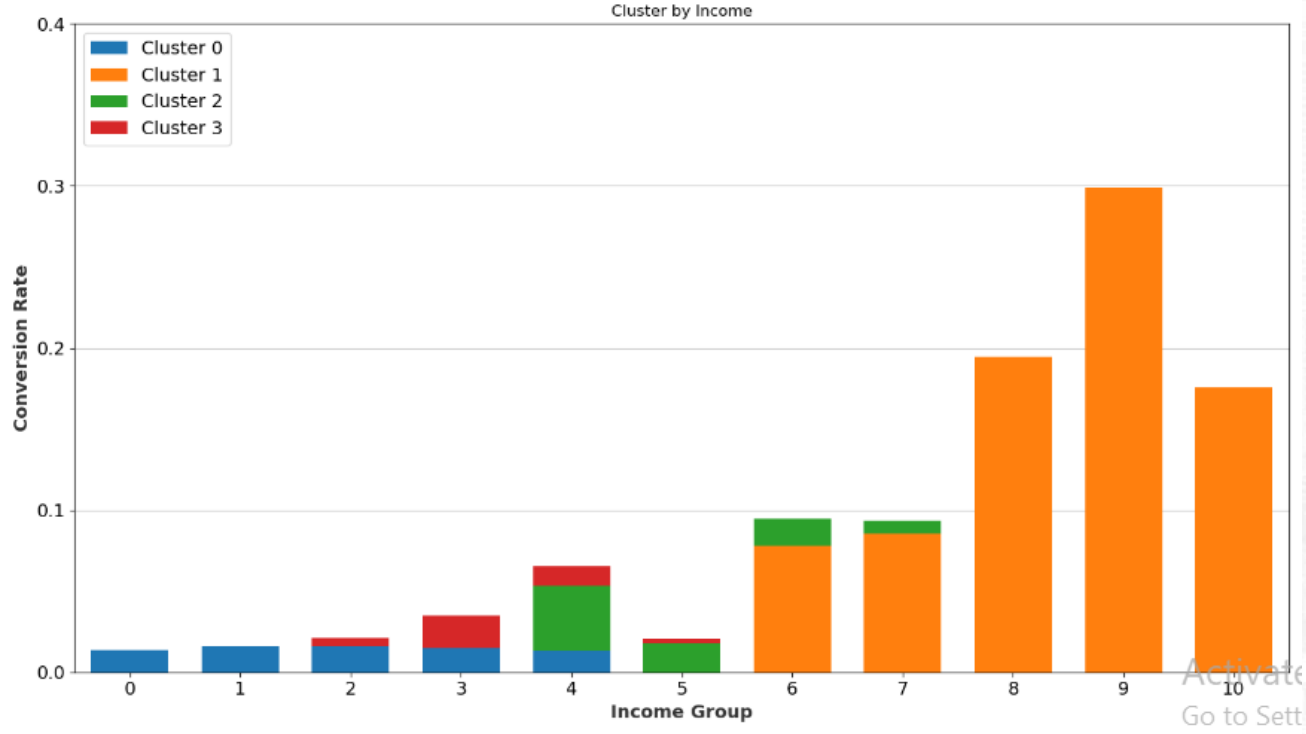
Cluster 0
Income: Low, well below the overall average.
Product Spending (`Mnt`, etc.): Low spending on all products.
Age (`Age`, `Age_Group`): Younger than average.
Cluster 1
Income (`Income`): Medium, overall average.
Product Spending (`Mnt`, etc.): Average spending on all products.
Age (`Age`, `Age_Group`): Exactly average.
Cluster 2
Income (`Income`): Medium, above the overall average.
Product Spending (`Mnt`, etc.): Above average spending on all products.
Age (`Age`, `Age_Group`): Above average.
Cluster 3
Income (`Income`): High, well above the overall average.
Product Spending (`Mnt`, etc.): High spending on all products.
Age (`Age`, `Age_Group`): Well above average.

Untuk selengkapnya, dapat melihat jupyter notebook disini

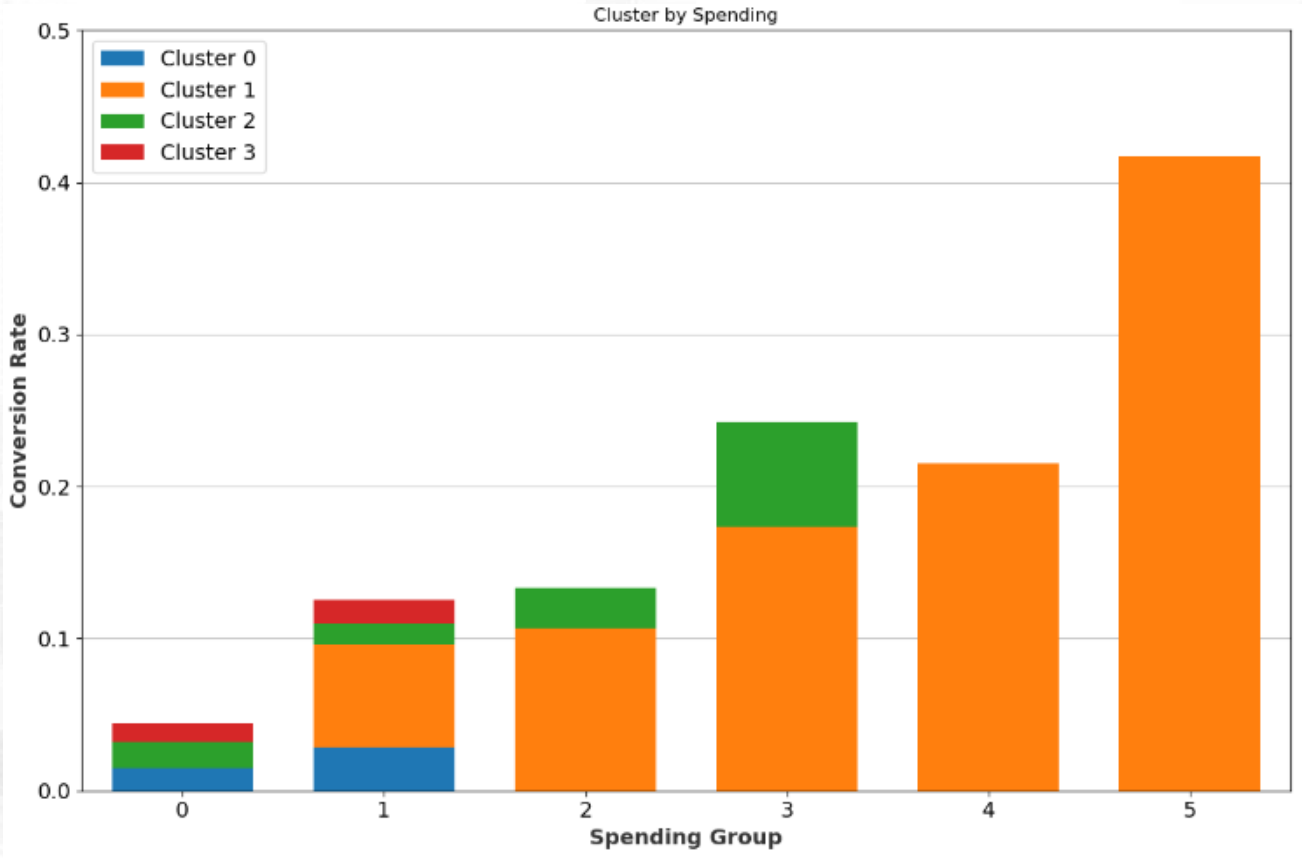
Customer Personality Analysis for Marketing Retargeting



Customer Personality Analysis for Marketing Retargeting



Customer Personality Analysis for Marketing Retargeting



Customer Personality Analysis for Marketing Retargeting

Rakamin Academy

Business Recommendations:

- By knowing the characteristics of each of these clusters, it is possible to adjust the marketing strategy to meet the needs and preferences of each group of customers. For example, for cluster 2 consisting of customers with high income and large total expenditure, the marketing team may consider offering premium products or services to them while cluster 1 with the opposite characteristics may be more responsive to discount offers or special promotions.
- To reduce the number of very low customers and low value customers, we can inform them about our limited discount products and create cheap packages (such as buy 1 get 1 free) because they have the lowest total amount in our store.
- To retain medium and high value customers, we can give them 'special treatment' such as providing bonuses and gifts.