Predict Customer Personality to boost marketing campaign by using Machine Learning

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

Created by:
Kenneth Wahyudi, S.Si.

kennethwahyudi48@gmail.com
https://www.linkedin.com/in/kenneth-wahyudi-80b886209/

I am an aspiring data scientist with interest in machine learning and paddling through the data lake. I have experiences as a research intern in BATAN, and several data science and machine learning projects under my belt. Specifically intermediate to advanced knowledge in Exploratory Data Analysis, Data Pre-processing, Supervised & Unsupervised Learning, as well as Data Visualization and Storytelling. And I am always looking forward to exploring new and uncharted territories that might allow myself to grow and improve.

Overview

Overview

A company can experience rapid growth when it comprehensively understands the intricacies of its customer personalities. This understanding empowers the organization to deliver superior services and benefits to potential loyal customers. By harnessing historical marketing campaign data to enhance performance and precisely target the right customers for transactions within the company's platform, our central objective revolves around the development of a predictive cluster model. This invaluable insight facilitates informed decision-making within the organization.

All results are from outputs of codes written in Python and its packages. JupyterLab was the GUI used in this project.

Link to the dataset used in this project can be found in the following link :
https://drive.google.com/drive/folders/1eFIh4E5KY9fI62rykqwcW8W9IlK9p0yH?usp=sharing

Conversion Rate Analysis Based on Income, Spending and Age

Conversion Rate Analysis Based on Income, Spending and Age

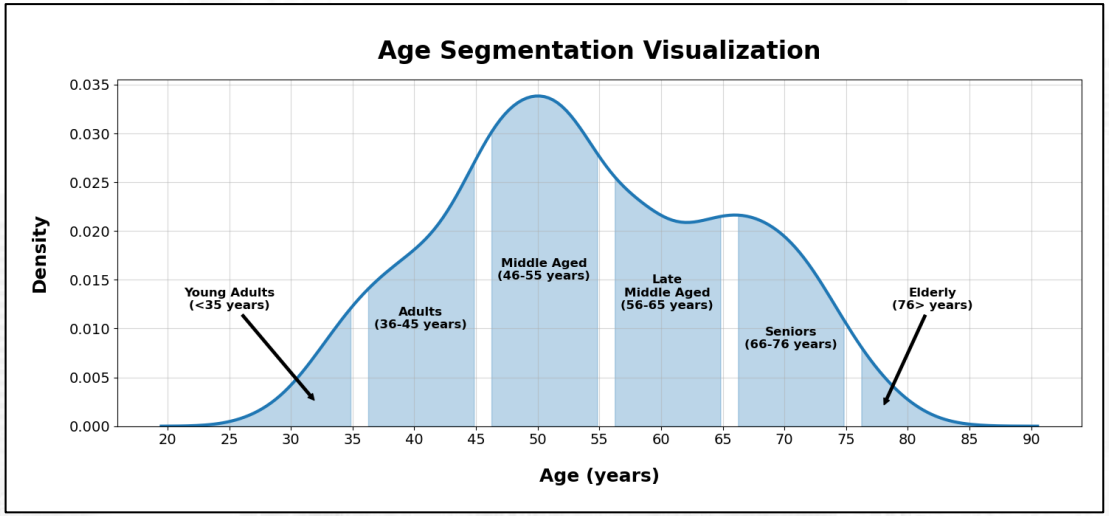
In this analysis, we're going to analyze the conversion rates based on income, total spending, and age. Conversion rate is defined as the total number of accepted campaigns divided by the total amount of web visits. The outline of the process of these analyses are as follows :

1. Creating a conversion rate column.
2. Creating an age column from 2023 minus the birth year.
3. Creating a total spending column from the sum of spendings in all product categories.
4. Creating the segment columns for the age, income, and spending values.
5. Grouping the conversion rates by the age, income and spending segments.
6. Plot the values and interpret them.

Now, let's dive into the details of the processes outlined above.

The complete code and .ipynb file for this analysis can be accessed in this link.

Conversion Rate Analysis Based on Income, Spending and Age



For the segmentation we're going to divide the ages into 6 segments, and the reasoning is as follows :

Young Adults (<35): This segment includes individuals who are in their late twenties to mid-thirties, typically early in their careers or education.

Adults (36-45): This group encompasses individuals in their mid-thirties to mid-forties, many of whom may be established in their careers and possibly starting families.

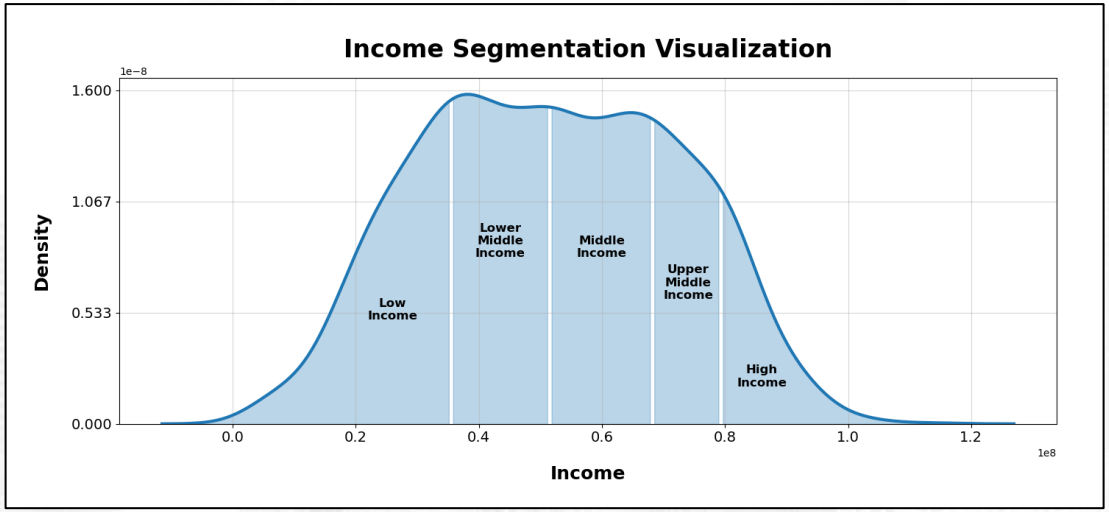
Middle-Aged (46-55): People in this age range are often considered middle-aged, and they may be at the peak of their careers or dealing with the challenges of raising children and planning for retirement.

Late Middle-Aged (56-65): This segment includes individuals in their late fifties to mid-sixties, some of whom may be approaching retirement and others who are already retired.

Seniors (66-75): This group consists of seniors in their late sixties to mid-seventies, many of whom are enjoying retirement or still actively engaged in their communities.

Elderly (76>): The oldest segment, comprising individuals in their late seventies to early eighties, often face health-related challenges and may be enjoying their retirement years.

Conversion Rate Analysis Based on Income, Spending and Age



For the segmentation we're going to divide the ages into 5 segments, and the reasoning is as follows :

Low Income: Approximate Percentile: 0-25th percentile. Low-income individuals typically fall within the bottom quartile of income earners.

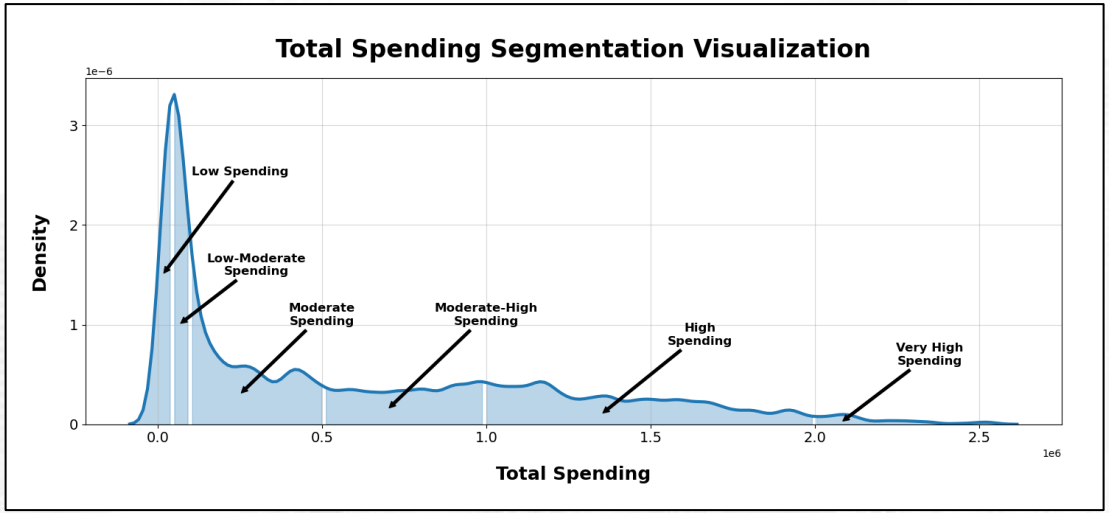
Lower-Middle Income: Approximate Percentile: 25th-50th percentile. Lower-middle-income individuals fall within the second quartile, below the median but above the lowest income earners.

Middle Income: Approximate Percentile: 50th-75th percentile. Middle-income individuals are around the median income level, falling within the third quartile.

Upper-Middle Income: Approximate Percentile: 75th-90th percentile. Upper-middle-income individuals earn significantly more than the median but are not yet among the highest income earners.

High Income: Approximate Percentile: 90th+ percentile. High-income individuals are among the top 10% of earners in the population.

Conversion Rate Analysis Based on Income, Spending and Age



For the segmentation we're going to divide the total spending into 6 segments, and the reasoning is as follows :

Low Spending (< 50,000): This segment includes spending amounts less than 50,000, indicating very minimal expenses.

Low-Moderate Spending (50,000 - 100,000): This segment encompasses spending between 50,000 and 100,000, representing low to moderate expenses.

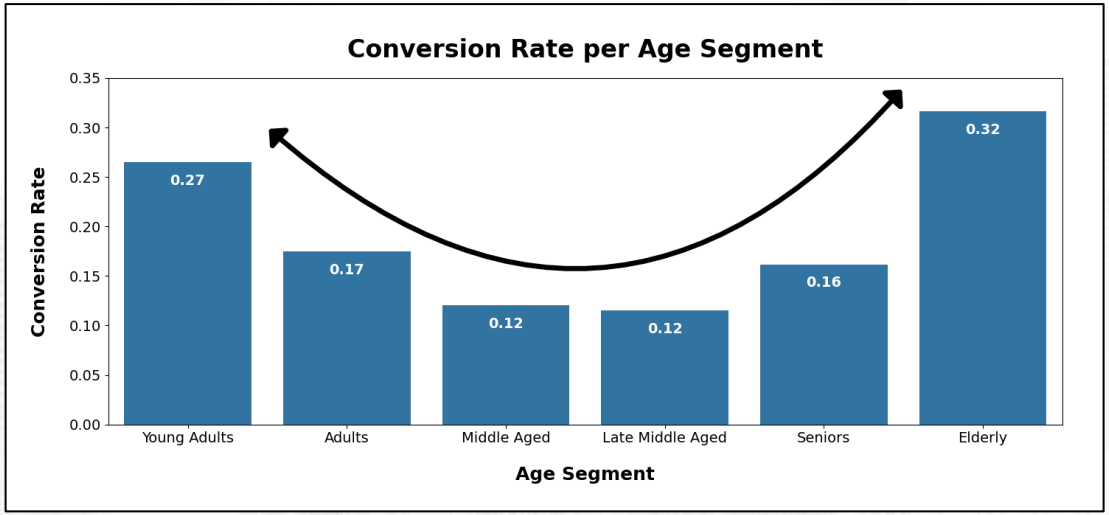
Moderate Spending (100,000 - 500,000): This segment includes spending between 100,000 and 500,000, indicating moderate expenses.

Moderate-High Spending (500,000 - 1,000,000): This segment covers spending between 500,000 and 1,000,000, representing moderate to relatively high expenses.

High Spending (1,000,000 - 2,000,000): This segment includes spending between 1,000,000 and 2,000,000, indicating high expenses.

Very High Spending (2,000,000+): This segment encompasses spending above 2,000,000, representing very high expenses.

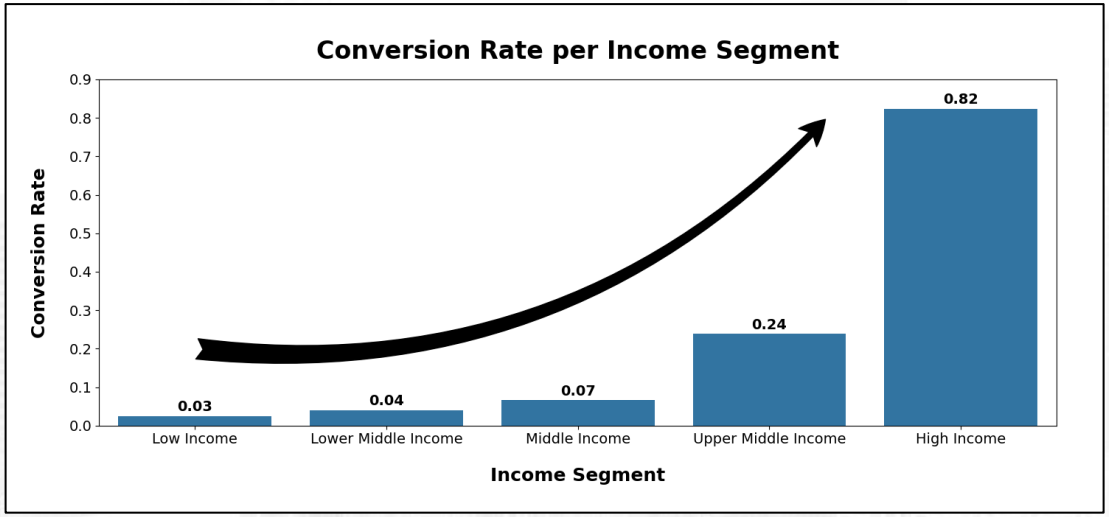
Conversion Rate Analysis Based on Income, Spending and Age



Looking at the chart to the left, we can see that the chart resembles a bimodal distribution (meaning a distribution with two peaks). With the lowest conversion rates are in the middle and late middle aged customers (at around 0.12), and the highest conversion rates are in the far end of the spectrum. With young adult and elderly customers having 0.27 and 0.32 conversion rates respectively. Followed by adult and senior customers at around 0.17 and 0.16 conversion rates respectively.

This could be because young adults usually have no dependents, therefore have way less to think about when accepting campaigns. Same goes for the elderly, they usually no longer have any dependents, and live off of their retirement funds. The middle and late middle aged however probably has their children's education (college or any other higher education) and retirement to think about, therefore have way more to think about when accepting campaigns. Whilst adult and senior customers are there to fill in the gradient between the lowest and highest conversion rates.

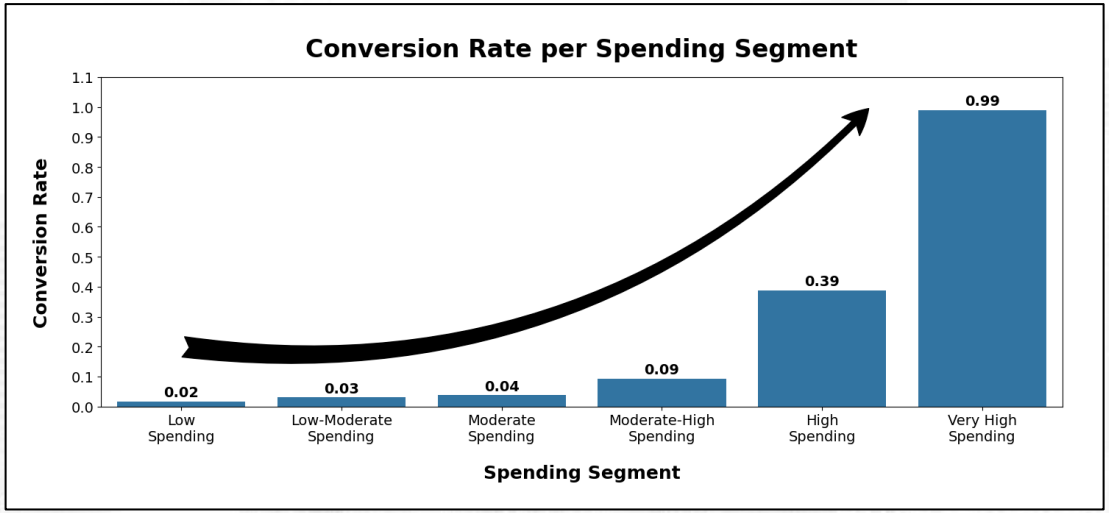
Conversion Rate Analysis Based on Income, Spending and Age



Looking at the chart to the left, we can see that the chart resembles an exponential growth. With the lowest conversion rate being the low income group (at around 0.03), followed closely by the lower middle income group (at around 0.04), followed slightly further by the middle income group (at 0.07), followed even further by the upper middle income group (at around 0.24), before finally shooting off towards 0.82 by the high income group.

This is actually quite self-explanatory. The higher the income, the higher the chance they're going to accept the campaigns, since they simply have more money to spend. But the difference is actually quite interesting, since it resembles an exponential growth. Further analysis should be done as to why the difference are so staggering between the high income group compared to the other groups.

Conversion Rate Analysis Based on Income, Spending and Age



Looking at the chart to the left, we can see that the chart resembles an exponential growth. With the lowest conversion rate being the low spending group (at around 0.02), followed closely by the low-moderate spending group (at around 0.03), followed by the moderate spending group (at 0.04), followed slightly further by the moderate-high spending group (at around 0.09), followed even further by the high spending group (at around 0.39), before finally shooting off towards 0.99 by the very high spending group.

This is actually quite self-explanatory. Much like the conversion rates per income segment, the higher the total spending, the higher the chance they're going to accept the campaigns, since they simply spend more in the first place. But the difference is quite interesting as well, since it also resembles an exponential growth. Especially the very high spending group with a 0.99 conversion rate. This basically means that on average, every time they visit the website, they accept a campaign. Further analysis should be done as to why the difference are so staggering between the high income group compared to the other groups.

Data Cleaning & Preprocessing

Data Cleaning & Preprocessing

In this section, we're focusing on the cleaning and preprocessing aspect of the dataset. And the outlines of the processes involved are as follows :

1. Checking for and handling NULL values if present.
2. Checking for and handling duplicated values if present.
3. Removal of columns/features that won't be used in the modelling stage later on.
4. Encoding the categorical features.
5. And finally, transforming the features.

With that being said, let's move on to the details of those outlined above.

The complete code and .ipynb file of the data cleaning & preprocessing can be found in this link.

Data Cleaning & Preprocessing

Rakamin Academy

NULL Value Checking

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2229 entries, 0 to 2239
Data columns (total 40 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Unnamed: 0              2229 non-null   int64
 1   ID                      2229 non-null   int64
 2   Year_Birth              2229 non-null   int64
 3   Education               2229 non-null   object
 4   Marital_Status          2229 non-null   object
 5   Income                  2229 non-null   float64
 6   Kidhome                 2229 non-null   int64
 7   Teenhome                2229 non-null   int64
 8   Dt_Customer             2229 non-null   object
 9   Recency                 2229 non-null   int64
 10  MntCoke                 2229 non-null   int64
 11  MntFruits               2229 non-null   int64
 12  MntMeatProducts         2229 non-null   int64
 13  MntFishProducts         2229 non-null   int64
 14  MntSweetProducts        2229 non-null   int64
 15  MntGoldProds            2229 non-null   int64
 16  NumDealsPurchases       2229 non-null   int64
 17  NumWebPurchases         2229 non-null   int64
 18  NumCatalogPurchases     2229 non-null   int64
 19  NumStorePurchases       2229 non-null   int64
 20  NumWebVisitsMonth       2229 non-null   int64
 21  AcceptedCmp3            2229 non-null   int64
 22  AcceptedCmp4            2229 non-null   int64
 23  AcceptedCmp5            2229 non-null   int64
 24  AcceptedCmp1            2229 non-null   int64
 25  AcceptedCmp2            2229 non-null   int64
 26  Complain                2229 non-null   int64
 27  Z_CostContact           2229 non-null   int64
 28  Z_Revenue               2229 non-null   int64
 29  Response                2229 non-null   int64
 30  Conversion_Rate         2229 non-null   float64
 31  Age                     2229 non-null   int64
 32  Age_Segment             2229 non-null   object
 33  Income_Segment          2229 non-null   object
 34  Total_Spending          2229 non-null   int64
 35  Spending_Segment        2229 non-null   object
 36  Total_Dependents        2229 non-null   int64
 37  Total_Purchases         2229 non-null   int64
 38  Total_Accepted_Campaigns  2229 non-null   int64
 39  Is_Parents              2229 non-null   object
dtypes: float64(2), int64(31), object(7)
memory usage: 714.0+ KB
```

As we can see from the figure to the left, there are no NULL values anymore in the dataset.

Duplicated Data Checking

```
df.duplicated().sum()
```

```
0
```

And from the figure above, we can see that the dataset also contains no duplicated values.

Data Cleaning & Preprocessing

Rakamin Academy

	Income	Age	Total_Spending	Total_Dependents	Total_Purchases	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	Education	Has_Partner
0	58138000.0	66	1617000	0	25	635000	88000	546000	172000	88000	88000	3	8	10	4	2	0
1	46344000.0	69	27000	2	6	11000	1000	6000	2000	1000	6000	2	1	1	2	2	0
2	71613000.0	58	776000	0	21	426000	49000	127000	111000	21000	42000	1	8	2	10	2	1
3	26646000.0	39	53000	1	8	11000	4000	20000	10000	3000	5000	2	2	0	4	2	1
4	58293000.0	42	422000	1	19	173000	43000	118000	46000	27000	15000	5	5	3	6	4	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2235	61223000.0	56	1341000	1	18	709000	43000	182000	42000	118000	247000	2	9	3	4	2	1
2236	64014000.0	77	444000	3	22	406000	0	30000	0	0	8000	7	8	2	5	4	1
2237	56981000.0	42	1241000	0	19	908000	48000	217000	32000	12000	24000	1	2	3	13	2	0
2238	69245000.0	67	843000	1	23	428000	30000	214000	80000	30000	61000	2	6	5	10	3	1
2239	52869000.0	69	172000	2	11	84000	3000	61000	2000	1000	21000	3	3	1	4	4	1

2229 rows × 17 columns

The features we'll be using for the modelling are 'Income', 'Age', 'Total_Spending', 'Total_Dependents', 'Total_Purchases', 'MntCoke', 'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts', 'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases', 'NumCatalogPurchases', 'NumStorePurchases', 'Education', and 'Has_Partner'. We're going to drop the rest for the modelling stage later on.

Data Cleaning & Preprocessing

Education	Has_Partner
2	0
2	0
2	1
2	1
4	1
...	...
2	1
4	1
2	0
3	1
4	1

	Education	Has_Partner
count	2229.000000	2229.000000
mean	2.458502	0.644235
std	1.003005	0.478852
min	0.000000	0.000000
25%	2.000000	0.000000
50%	2.000000	1.000000
75%	3.000000	1.000000
max	4.000000	1.000000

All of the categorical features (‘Education’ and ‘Has_Partner’) we have are able to be encoded by means of label encoding. Because they have some degree of ordinality within them.

And from the descriptive statistics, it seems like all of the features have been encoded properly. Now, let us move on to the next part, the feature transformation.

Data Cleaning & Preprocessing

Rakamin Academy

	Income	Age	Total_Spending	Total_Dependents	Total_Purchases	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	Education	Has_Partner
0	0.316489	1.015715	1.683247	-1.266277	1.329444	0.980166	1.550778	1.736151	2.456789	1.472029	0.842690	0.358938	1.405471	2.633984	-0.559160	2	0
1	-0.256130	1.272020	-0.962795	1.393978	-1.163838	-0.873191	-0.637618	-0.726371	-0.651738	-0.632996	-0.731824	-0.168815	-1.117198	-0.585437	-1.176175	2	0
2	0.970723	0.332234	0.283674	-1.266277	0.804543	0.359410	0.569773	-0.174584	1.341376	-0.149082	-0.040574	-0.696567	1.405471	-0.227723	1.291882	2	1
3	-1.212502	-1.291031	-0.919526	0.063851	-0.901387	-0.873191	-0.562156	-0.662528	-0.505455	-0.584604	-0.751025	-0.168815	-0.756817	-0.943150	-0.559160	2	1
4	0.324014	-1.034726	-0.305445	0.063851	0.542092	-0.392031	0.418849	-0.215626	0.152822	-0.003908	-0.559011	1.414444	0.324327	0.129990	0.057854	4	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2235	0.466271	0.161364	1.223934	0.063851	0.410867	1.199955	0.418849	0.076229	0.079680	2.197900	3.895711	-0.168815	1.765853	0.129990	-0.559160	2	1
2236	0.601778	1.955500	-0.268833	2.724106	0.935768	0.300008	-0.662772	-0.616926	-0.688309	-0.657191	-0.693421	2.469949	1.405471	-0.227723	-0.250653	4	1
2237	0.260314	-1.034726	1.057517	-1.266277	0.542092	1.791009	0.544619	0.235837	-0.103175	-0.366843	-0.386199	-0.696567	-0.756817	0.129990	2.217403	2	0
2238	0.855752	1.101150	0.395174	0.063851	1.066994	0.365350	0.091848	0.222156	0.774527	0.068679	0.324253	-0.168815	0.684708	0.845417	1.291882	3	1
2239	0.060670	1.272020	-0.721489	1.393978	-0.507711	-0.656372	-0.587310	-0.475559	-0.651738	-0.632996	-0.443803	0.358938	-0.396436	-0.585437	-0.559160	4	1

2229 rows × 17 columns

	Income	Age	Total_Spending	Total_Dependents	Total_Purchases	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	Education	Has_Partner
count	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2.229000e+03	2229.000000	2229.000000
mean	3.825264e-17	2.151711e-16	3.028334e-17	4.462808e-17	1.179456e-16	-2.868948e-17	4.223729e-17	-3.506492e-17	-6.056668e-17	-2.390790e-17	6.853597e-17	-9.244387e-17	-7.331755e-17	5.100352e-17	-6.176207e-17	2.458502	0.644235
std	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.000224e+00	1.003005	0.478852
min	-2.422216e+00	-2.316251e+00	-9.994070e-01	-1.266277e+00	-1.951190e+00	-9.058622e-01	-6.627718e-01	-7.537323e-01	-6.883091e-01	-6.571915e-01	-8.470319e-01	-1.224320e+00	-1.477580e+00	-9.431503e-01	-1.793189e+00	0.000000	0.000000
25%	-7.867034e-01	-6.929856e-01	-8.928996e-01	-1.266277e+00	-9.013871e-01	-8.345792e-01	-6.124639e-01	-6.807687e-01	-6.334528e-01	-6.329958e-01	-6.742194e-01	-6.965674e-01	-7.568171e-01	-9.431503e-01	-8.676676e-01	2.000000	0.000000
50%	-1.196376e-02	-9.494056e-02	-3.470494e-01	6.385089e-02	1.719058e-02	-3.831205e-01	-4.615400e-01	-4.481972e-01	-4.688837e-01	-4.636259e-01	-3.861986e-01	-1.688146e-01	-3.605431e-02	-2.277234e-01	-2.506534e-01	2.000000	1.000000
75%	8.010347e-01	8.448445e-01	7.296736e-01	6.385089e-02	8.045429e-01	5.940499e-01	1.673095e-01	2.996799e-01	2.259636e-01	1.654621e-01	2.282458e-01	3.589382e-01	6.847085e-01	4.877036e-01	6.748679e-01	3.000000	1.000000
max	3.015770e+00	2.468110e+00	3.194321e+00	2.724106e+00	3.691501e+00	3.528531e+00	4.342870e+00	7.112657e+00	4.047624e+00	5.706276e+00	6.103870e+00	6.691972e+00	8.252718e+00	9.072827e+00	2.217403e+00	4.000000	1.000000

In this project, we are going to segmentize the customer's "personalities". So therefore, we're going to use a clustering algorithm. And it is imperative that we scale/transform our numerical features, because it is a distance based algorithm. We therefore need to standardize the values in our numerical features. Now, from the figures above, we can see that our numerical features have been successfully scaled with a standard scaler.

Data Modelling

Data Modelling

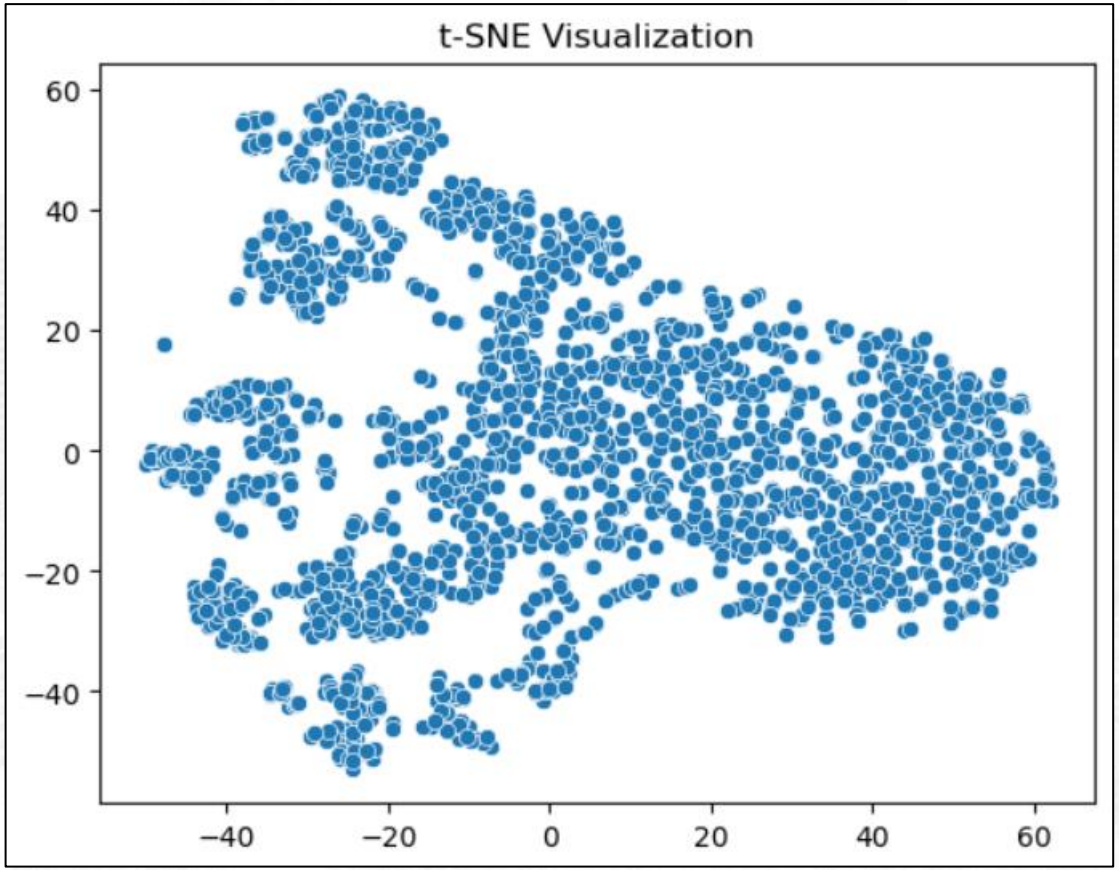
In this section, we are focusing on the making of the clustering model. And the outlines are as follows :

1. Reducing the dimension of the dataset.
2. Use the elbow method to determine the ideal number of clusters.
3. Use the silhouette score to determine the ideal number of clusters.
4. Review the clusters made.

Now, let us delve deeper into the details.

The complete code and .ipynb file of the clustering model can be found in this link.

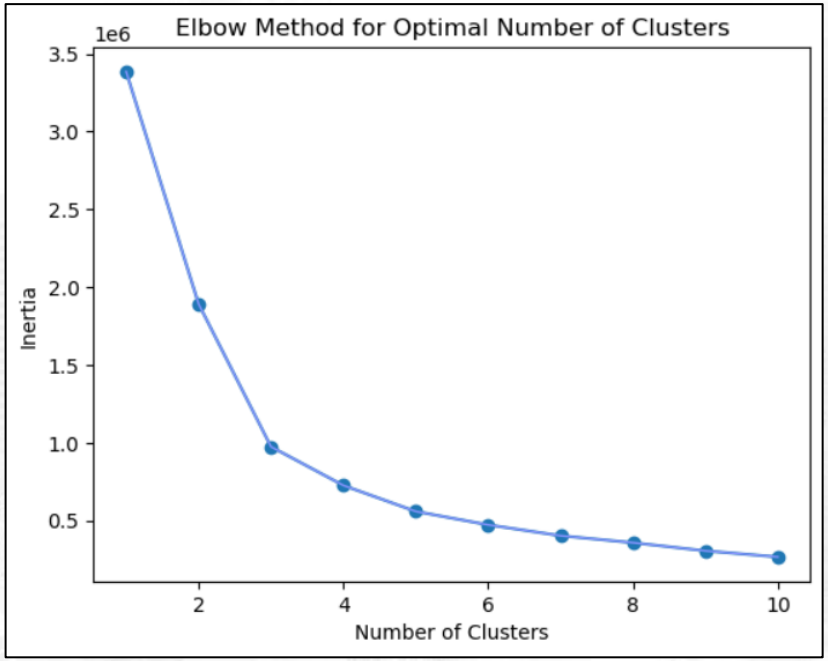
Data Modelling

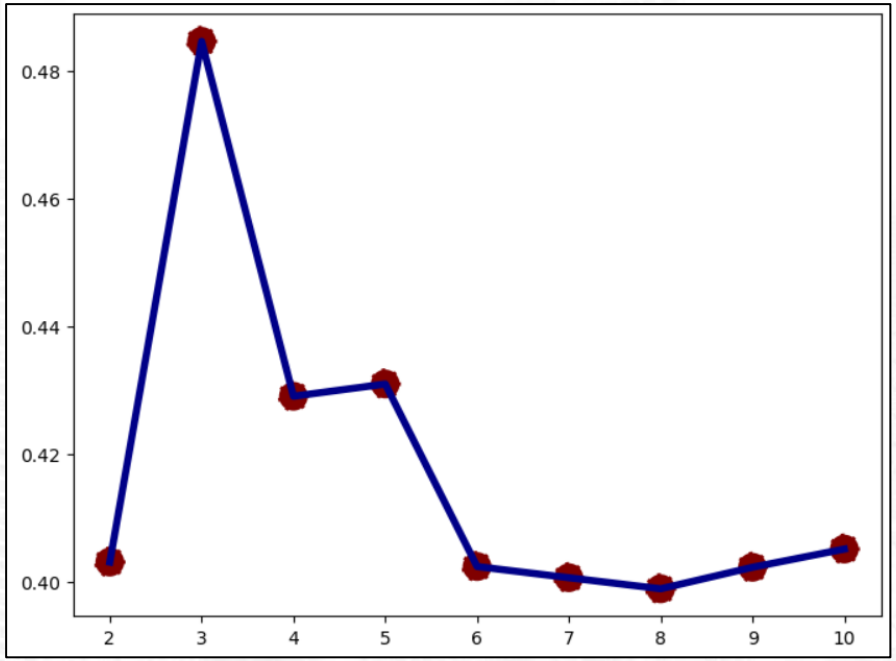


We first are going to reduce the dimension of the dataset using t-SNE. Before finally using the reduced data to do the clustering model.

And to the left, is the visualization of the first 2 components of the t-SNE reduced data.

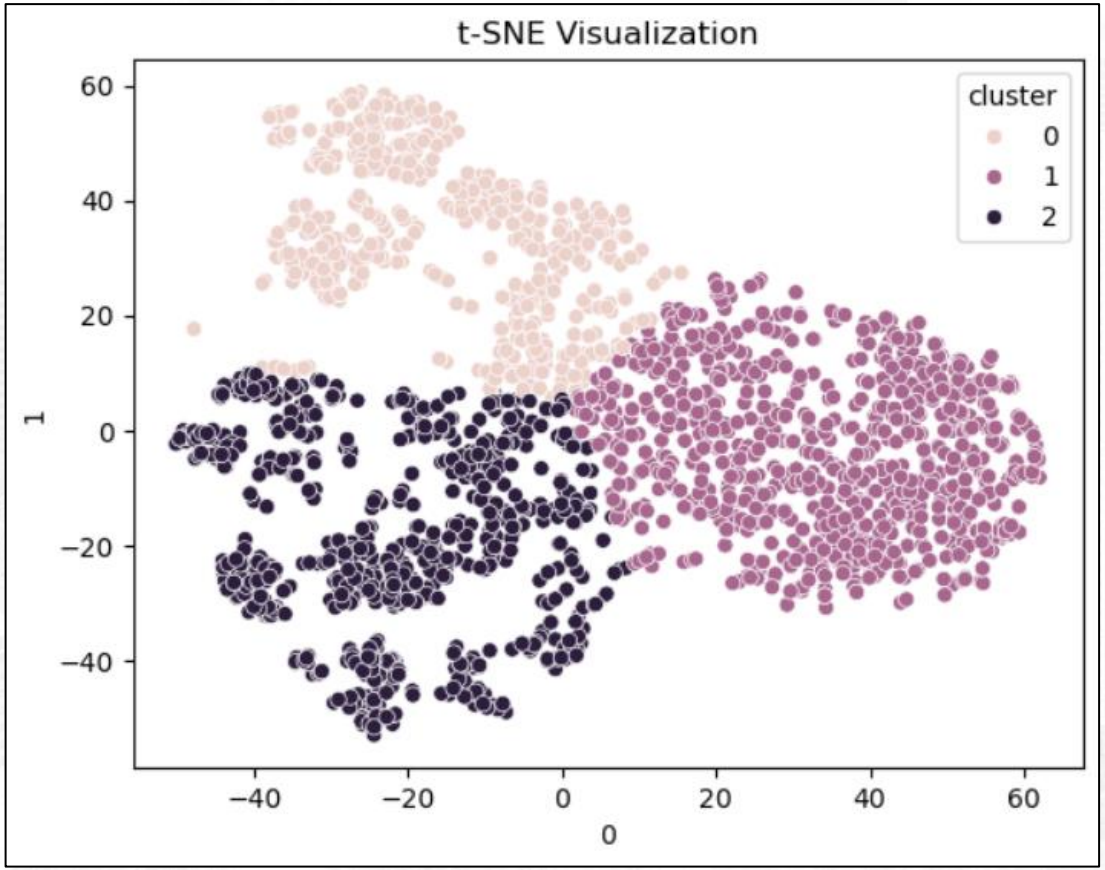
Data Modelling





From the elbow curve and the silhouette score plot, it is visible that the ideal number of clusters is 3 clusters.

Data Modelling



For evaluational purposes, let's plot the t-SNE reduced dataset again, but this time, set the hue as the clusters.

From the plot to the left, we can see that the separations are relatively clean.

Customer Personality Analysis for Marketing Retargeting

Customer Personality Analysis for Marketing Retargeting

Rakamin Academy

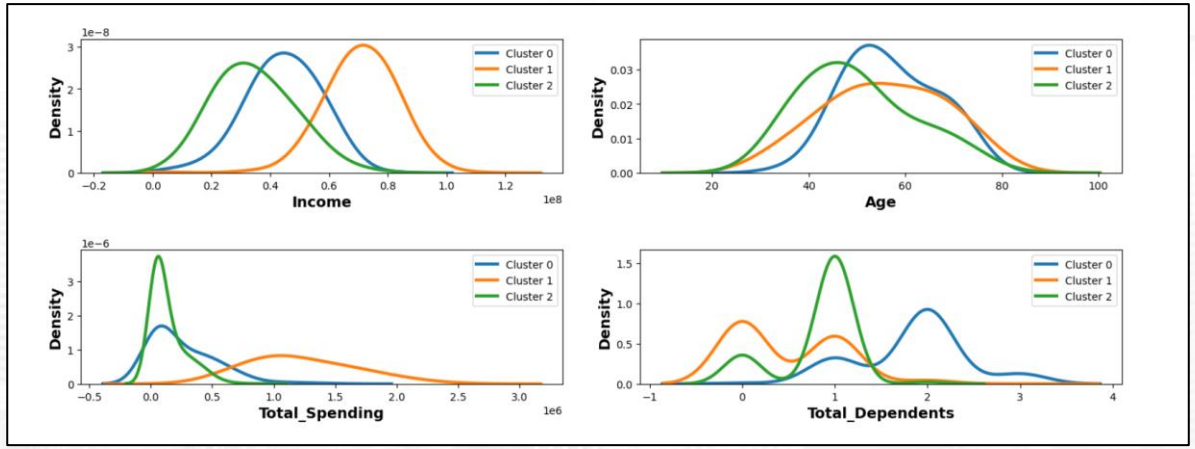
In this section, we're going to focus on the cluster analysis for marketing retargeting. The outlines of the process are as follows :

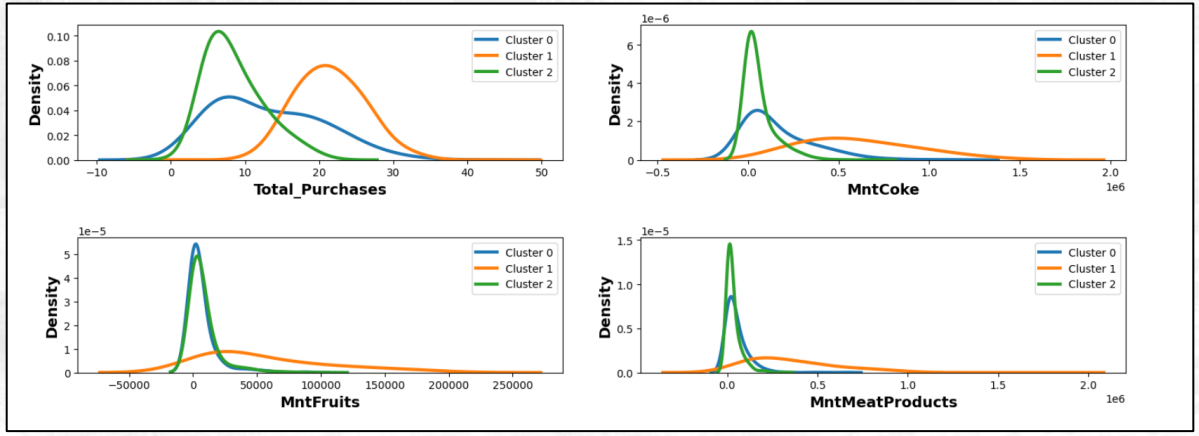
1. Append clusters to the original dataset.
2. Do EDA with the clusters present this time.
3. Creating a new dataframe that only houses the total accepted campaigns and conversion rate columns.
4. Identifying the best cluster to retarget the marketing.
5. Calculate the potential impact/revenue.

Now, lets look into the details even further.

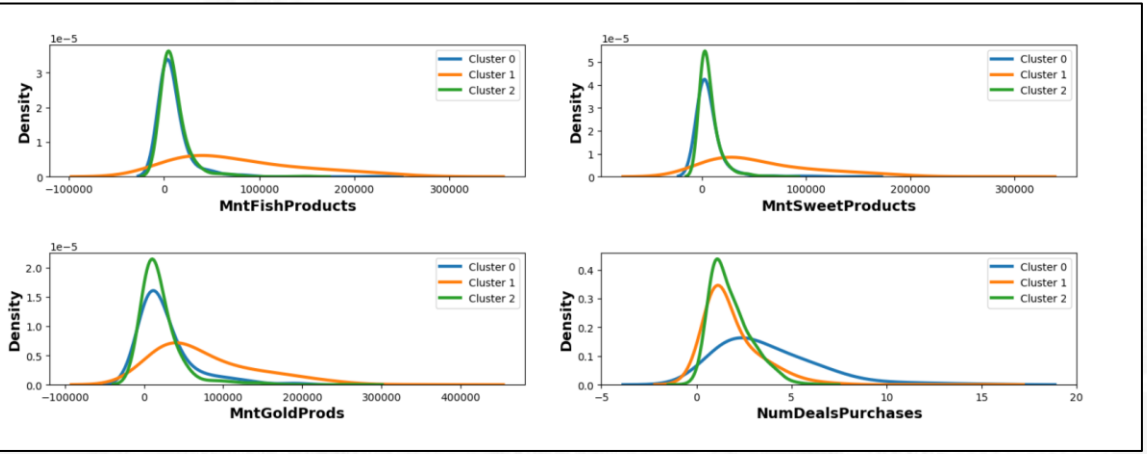
The complete code and .ipynb file for the analysis for marketing retargeting can be accessed in the following link.

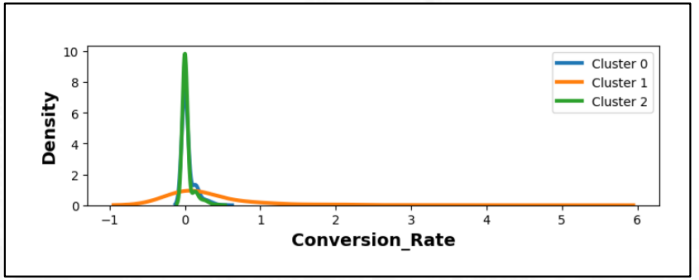
Customer Personality Analysis for Marketing Retargeting

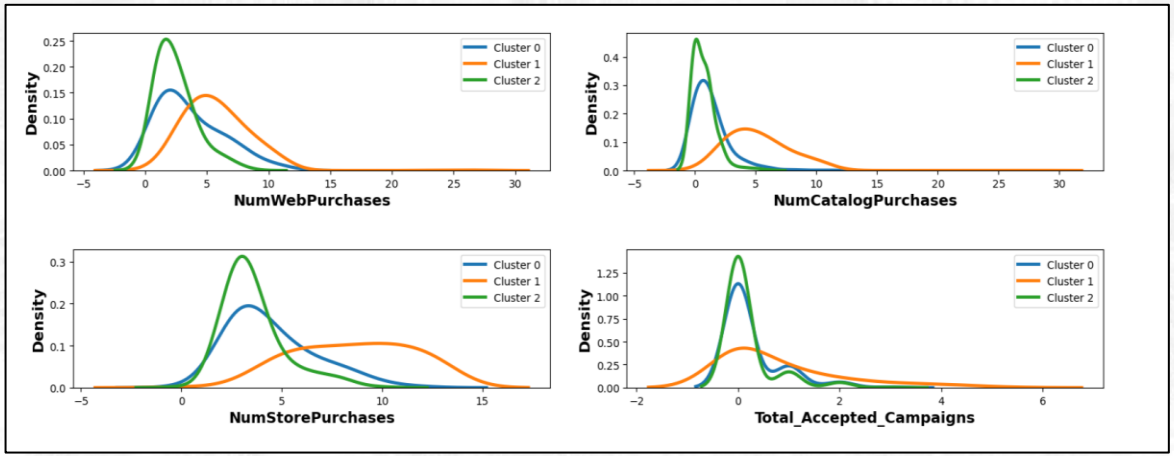




Customer Personality Analysis for Marketing Retargeting







From the numerical features in this and the previous slide, we can see the characteristics of each cluster. And they are as follows (the next slides) :

Cluster 0 :

1. The middle income between the other two clusters.
2. On average is the oldest cluster by a bit.
3. The middle spender when it comes to total spending and purchases.
4. The cluster with the most children/dependents.
5. Predominantly spends on coke, but also spend on meat, gold, sweets on a much lesser amount.
6. The cluster that purchases products through discounts the most. But also purchases through web an store on a much lesser amount.
7. The middle cluster when it comes to accepting marketing campaigns. Slightly higher than cluster 2.

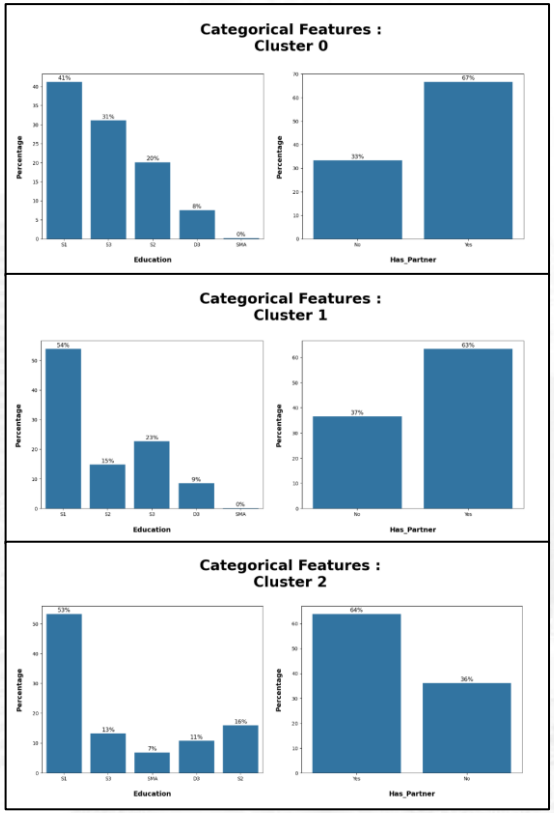
Cluster 1 :

1. The highest income between the other two clusters.
2. Slightly younger than cluster 0.
3. The highest spender when it comes to total spending and purchases.
4. The cluster with the middle amount of dependents compared to the other two.
5. Predominantly spends on coke, but also spends on all of the other product categories the most compared to other clusters.
6. The cluster that purchases products through the store the most. But also purchases through other means on a lesser amount.
7. The best cluster when it comes to accepting marketing campaigns, and it's not even close.

Cluster 2 :

1. The lowest income between the three clusters.
2. On average is the youngest.
3. The lowest spender of them all.
4. Has the least amount of dependents compared to the other two clusters.
5. Although they are the lowest spender, they spend on fruits more than cluster 0.
6. Tends to make purchases through store and website, although on a much lesser amount than the other two clusters.
7. The worst cluster when it comes to accepting marketing campaigns.

Customer Personality Analysis for Marketing Retargeting



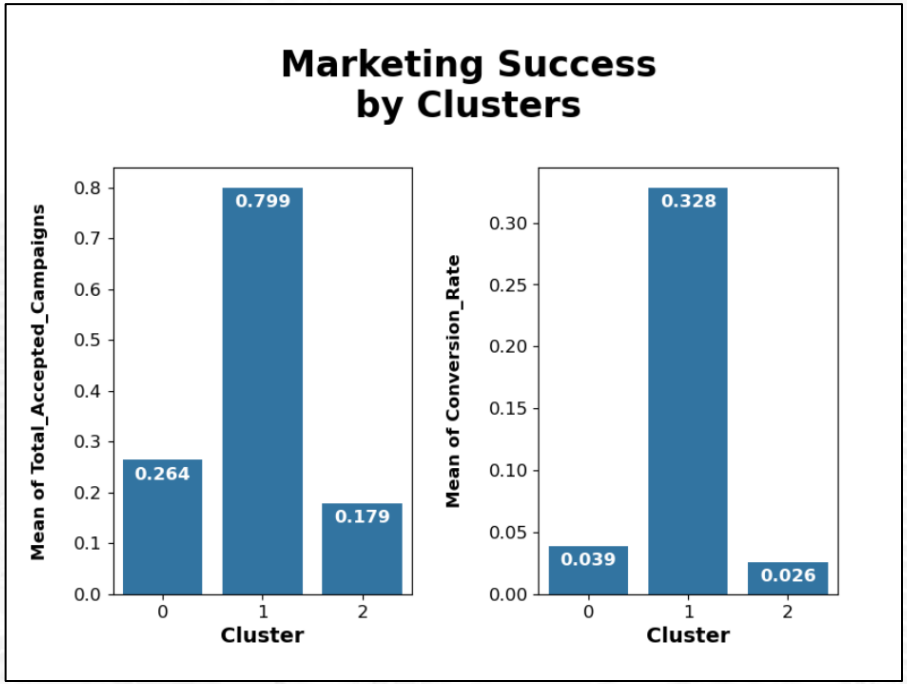
From the categorical features per cluster above, we can see that the 'Has_Partner' feature does not differ between the clusters. The 'Education' feature however, does.

Cluster 0 : The most well educated cluster compared to the rest. With nearly 0% being only high school graduates. Whilst a whopping 51% (majority) are graduate degree holders (degrees above bachelor like masters and doctoral degrees).

Cluster 1 : The middle cluster when it comes to education. Also having nearly 0% of its members who are only high school graduates. But only 38% are graduate degree holders. Whilst the majority are bachelor degree holders (at 51%) and diploma degree holders (at 9%).

Cluster 2 : The least educated cluster compared to the rest. As 7% of them are only high school graduates, and only 29% of them are graduate degree holders. Whilst the majority are bachelor degree holders (at 53%) and diploma degree holders (at 11%).

Customer Personality Analysis for Marketing Retargeting



From the mean of total accepted campaigns and mean of conversion rate above, we can see that cluster 1 is the best when it comes to the total accepted campaigns and conversion rate with a mean of around 0.8, and 0.33 respectively. Whilst cluster 0 is only 0.26 & 0.039 respectively, and cluster 2 (being the worst) is only 0.18 & 0.026 respectively.

Therefore, as a conclusion, cluster 1 is objectively the best cluster to perform a retargeted marketing. In layman terms, we should focus to target cluster 1 for our marketing campaigns in the future.

Potential Impact/Revenue :

Now, the total accepted campaigns are over a course of 6 campaigns. So therefore, we need to divide the mean by 6 in order to get the average accepted campaign. The result is around 13.33%. So this means that, on average, 13.33% of cluster 1 customers will accept the campaign. Compare that to the overall average accepted campaign, which is 0.45/6 = 7.5%.

This means that there would be a 77.33% increase in marketing campaign acceptance. From 7.5%, to 13.33%. Now, we can see that the dataset contains the 'Z_CostContact' and 'Z_Revenue' columns. 'Z_CostContact' is the column that houses the cost to deliver a campaign to the customer, whilst 'Z_Revenue' is the column that houses the revenue gained if the customer accepts the campaign. They both have a constant value of 3 and 11 respectively.

Customer Personality Analysis for Marketing Retargeting

Rakamin Academy

Before the clustering, the cost to deliver a campaign to all of the customers is 3 x 2229 = 6687, and the revenue is 11 x 167 (amount of customers that accept the campaign) = 1837. Now, if we only target cluster 1 customers for our campaign, this means that the cost to deliver a campaign to all of the cluster 1 customers is 3 x 891 (total number of cluster 1 customers) = 2673, and the revenue is 11 x 119 (13.3% of all the cluster 1 customers) = 1309.

This means that the return of investment before the clustering model is 1837/6687 = 27.5%, whilst the return of investment after the retargeting is 1309/2673 = 49%.

Therefore, the recommendation is to retarget the marketing campaigns to the cluster 1 customers in order to get a 78.3% increase in return of investment, and save up in marketing costs since we're basically targetting much less people, and use the spare resources in other areas. The amount of cost saved is 6687 - 2673 = 4014. That is a 60% reduction in marketing costs.

So as a conclusion, the return of investment increases by 78.3%, and the marketing cost decreases by 60%.