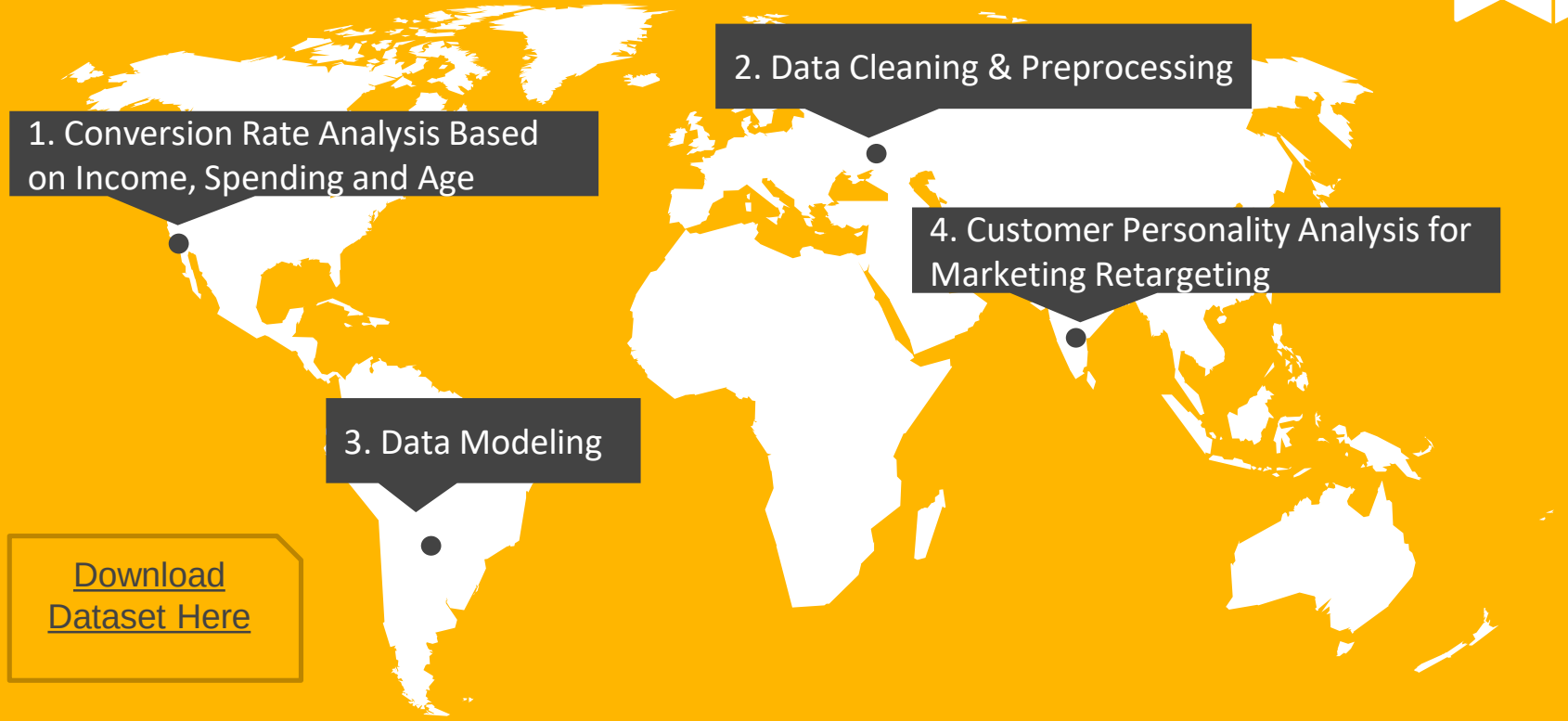
Predict Customer Personality to boost marketing campaign by using Machine Learning

Table of Contents



Download Dataset Here

1

Conversion Rate Analysis Based on Income, Spending and Age

Feature Engineering

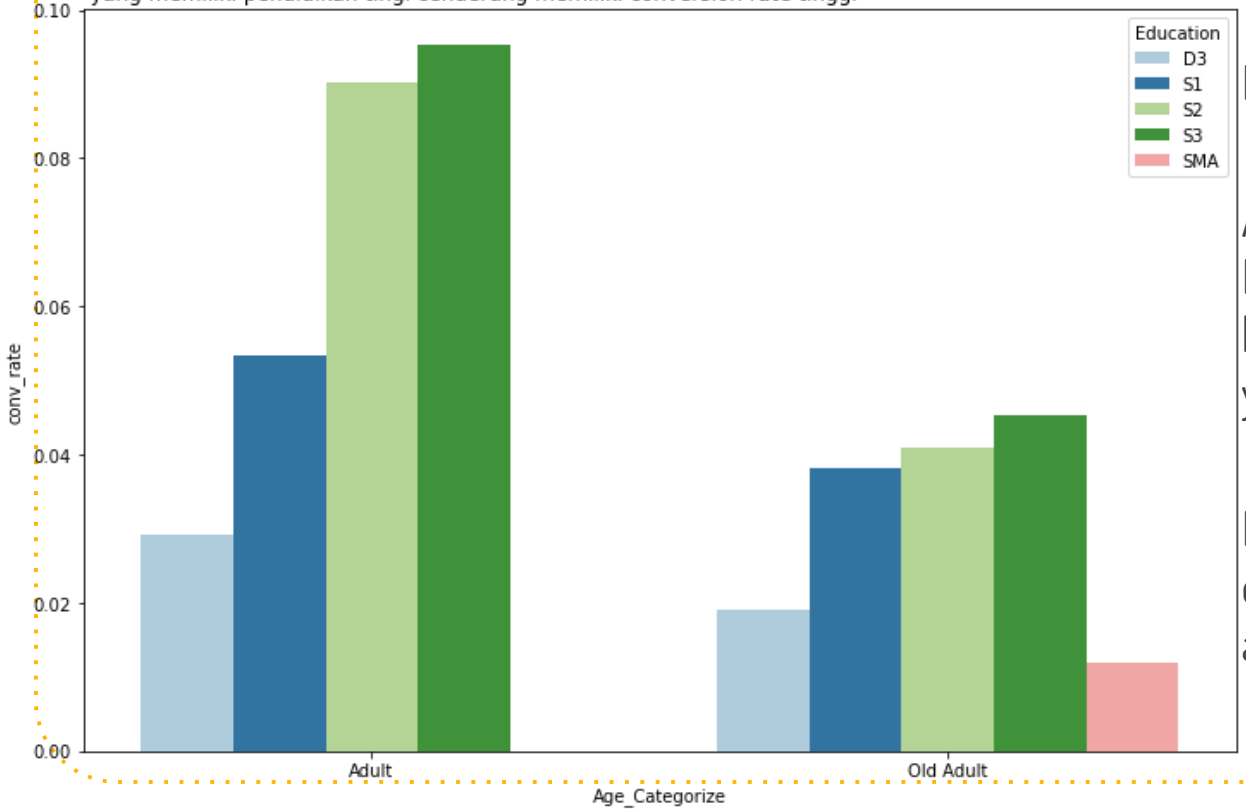
In this project the engineering features carried out are:

- Making Feature Engineering in the form of Conversion rate, age, group grouping age, number of children, whether parents, total spending, and total transactions
- Conducting EDA to draw initial insights to be used as recommendations
- See the correlation between features using heatmap correlation

For details, see jupyter notebook here

Perbedaan tingkat pendidikan terhadap conversion rate berdasarkan kelompok umur

Secara umum, semakin tinggi pendidikan mempengaruhi tingkat conversion rate dalam merespon campaign yang dibuat Hal ini bisa jadi karena produk fit dengan market fit sehingga terdapat kecenderungan pelanggan yang memiliki pendidikan tingi cenderung memiliki conversion rate tinggi

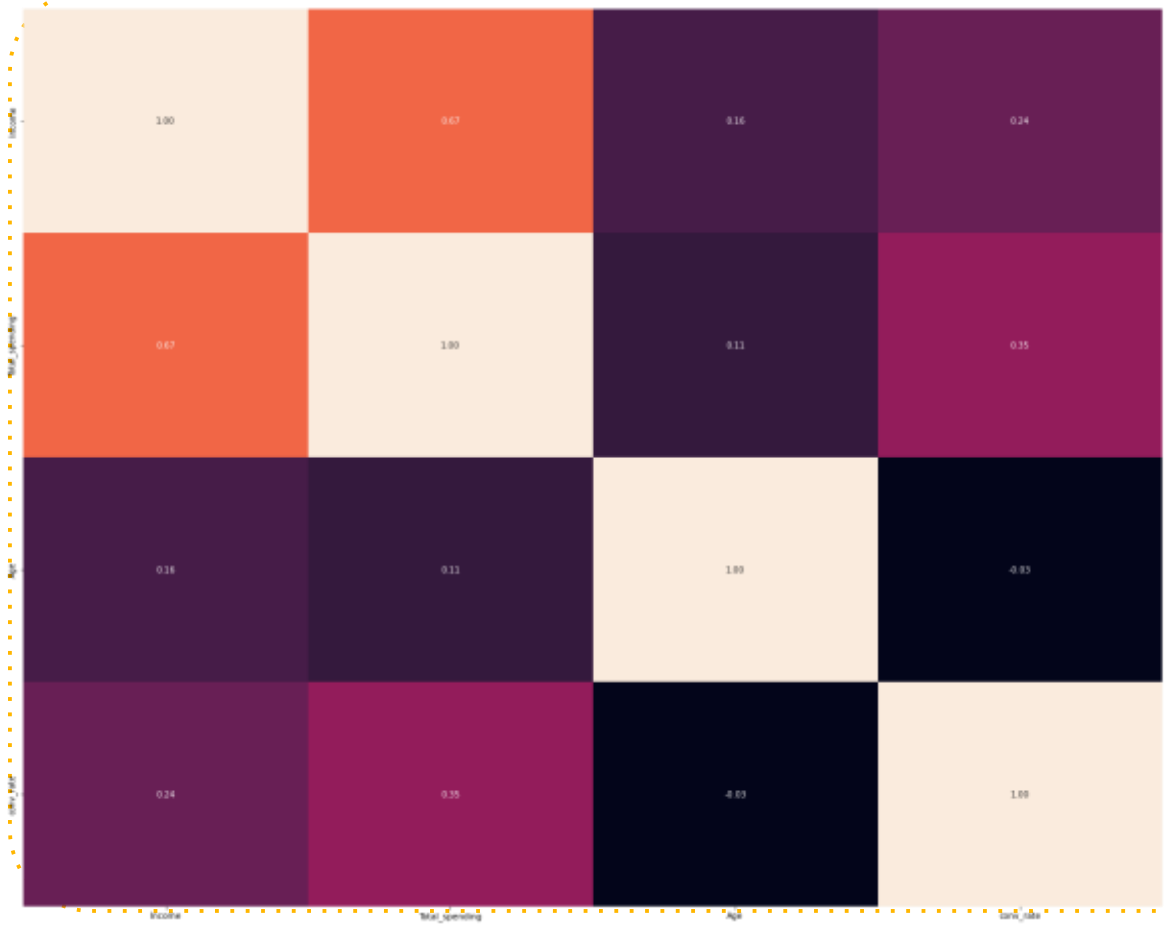


Insights:

Adult group (aged 25 to 45) Have a tendency to convert higher rate than the aged over 45 years.

Education level has potential conversion rate is high at each age

For details, see jupyter notebook here



Heatmap Corelation

Income has a relationship the correlation is quite strong, namely 0.66 to Total Spending.

Insights :

The higher the income customers, then the total spending spent also getting bigger.

For details, see jupyter notebook here

2

Data Cleaning & Preprocessing

- Duplicate data. There are no duplicate data missing values.
- There are 24 columns of missing values in the income column and 11 columns in the conversion rate
- The treatment for both data handles uses the Median on the grounds that the data is numerical and not robust against outliers.

For details, see jupyter notebook here

Feature Encoding and Standarization

	Income	Recency	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases
0	0.235696	0.307039	0.983781	1.551577	1.679702	2.462147	1.476500	0.843207	0.349414
1	-0.235454	-0.383664	-0.870479	-0.636301	-0.713225	-0.650449	-0.631503	-0.729006	-0.168236
2	0.773999	-0.798086	0.362723	0.570804	-0.177032	1.345274	-0.146905	-0.038766	-0.685887
3	-1.022355	-0.798086	-0.870479	-0.560857	-0.651187	-0.503974	-0.583043	-0.748179	-0.168236
4	0.241888	1.550305	-0.389085	0.419916	-0.216914	0.155164	-0.001525	-0.556446	1.384715

NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1
1.409304	2.510890	-0.550785	0.693904	-0.28014	-0.283830	-0.28014	-0.262111
-1.110409	-0.568720	-1.166125	-0.130463	-0.28014	-0.283830	-0.28014	-0.262111
1.409304	-0.226541	1.295237	-0.542647	-0.28014	-0.283830	-0.28014	-0.262111
-0.750450	-0.910898	-0.550785	0.281720	-0.28014	-0.283830	-0.28014	-0.262111
0.329427	0.115638	0.064556	-0.130463	-0.28014	-0.283830	-0.28014	-0.262111

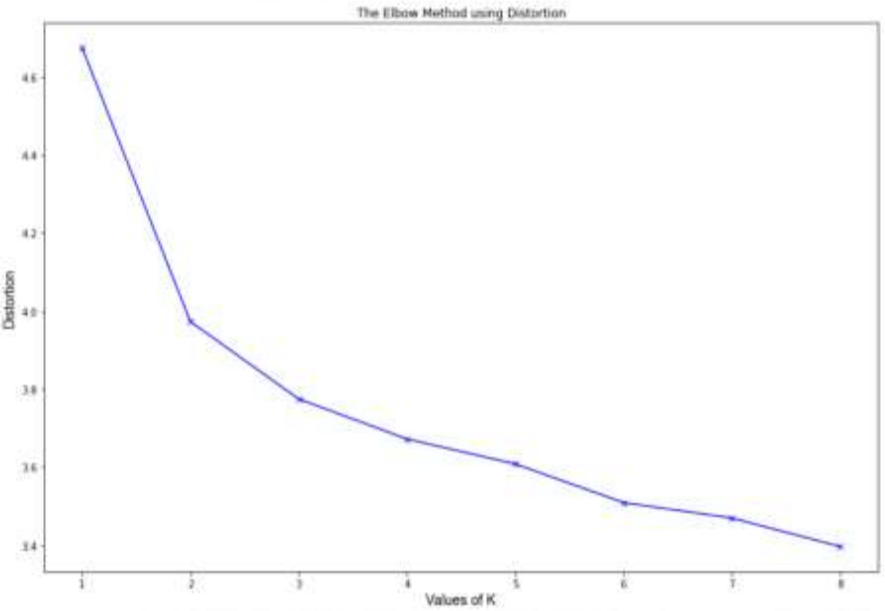
AcceptedCmp2	Complain	Z_CostContact	Z_Revenue	Age	Total_kid	Total_spending	Total_transaction	education_lvl	conv_rate	is_parent
-0.11651	-0.097282	0.0	0.0	0.985345	-1.264505	1.679417	1.320826	-0.458383	0.142857	0
-0.11651	-0.097282	0.0	0.0	1.235733	1.396361	-0.961275	-1.154596	-0.458383	0.000000	1
-0.11651	-0.097282	0.0	0.0	0.317643	-1.264505	0.282673	0.799685	-0.458383	0.000000	0
-0.11651	-0.097282	0.0	0.0	-1.268149	0.065928	-0.918094	-0.894025	-0.458383	0.000000	1
-0.11651	-0.097282	0.0	0.0	-1.017761	0.065928	-0.305254	0.539114	1.533425	0.000000	1

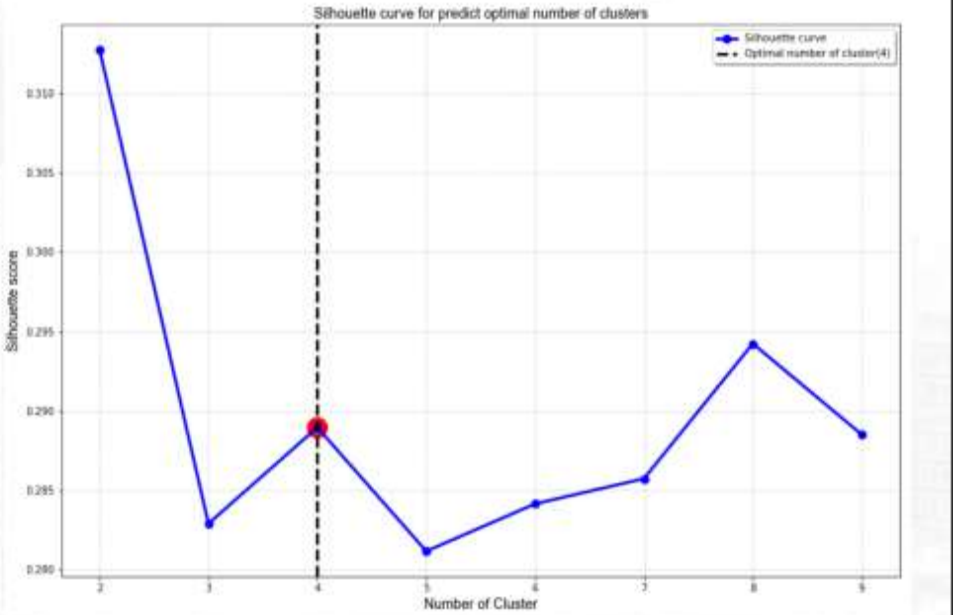
Bertunangan	Cerai	Duda	Janda	Lajang	Menikah	Adult	Old Adult
0	0	0	0	1	0	0	1
0	0	0	0	1	0	0	1
1	0	0	0	0	0	0	1
1	0	0	0	0	0	1	0
0	0	0	0	0	1	1	0

For details, see jupyter notebook here

3

Data Modelling





The optimal value for determining the type of cluster based on the two plots is 4 clusters:
In the elbow method, the value of k is taken after experiencing a deflection
In the silhouette score, the maximum value is taken, even though there is a maximum score at n = 8. But n clusters = 8 is considered too much for customer grouping

For details, see jupyter notebook here

Evaluation Silhouette

Cluster	Silhouette Score
2	0.3127516474151014,
3	0.28289565400020106,
4	0.28895486066913784,
5	0.2811570328004259,
6	0.2841429107017782,
7	0.2857304059651237,
8	0.2885100118187077

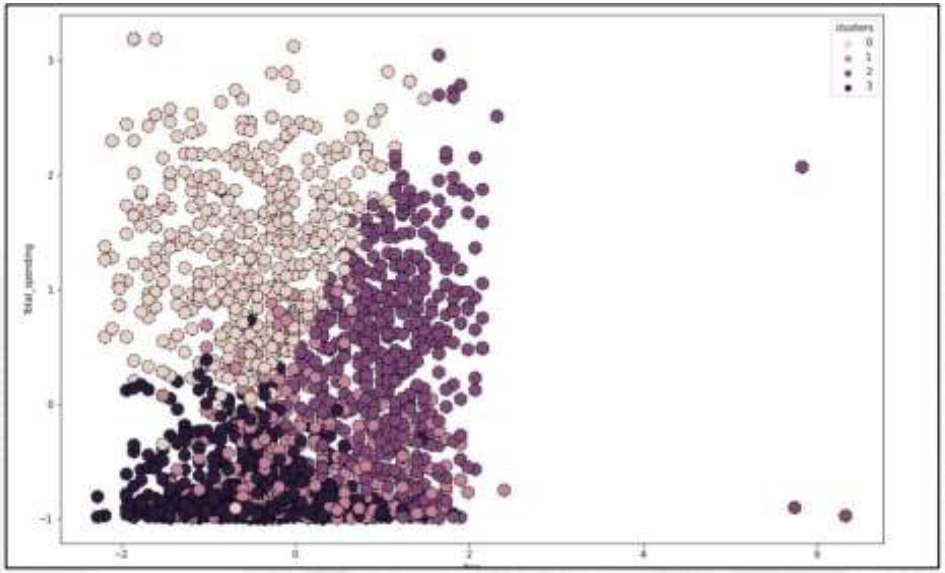
OPTIMAL

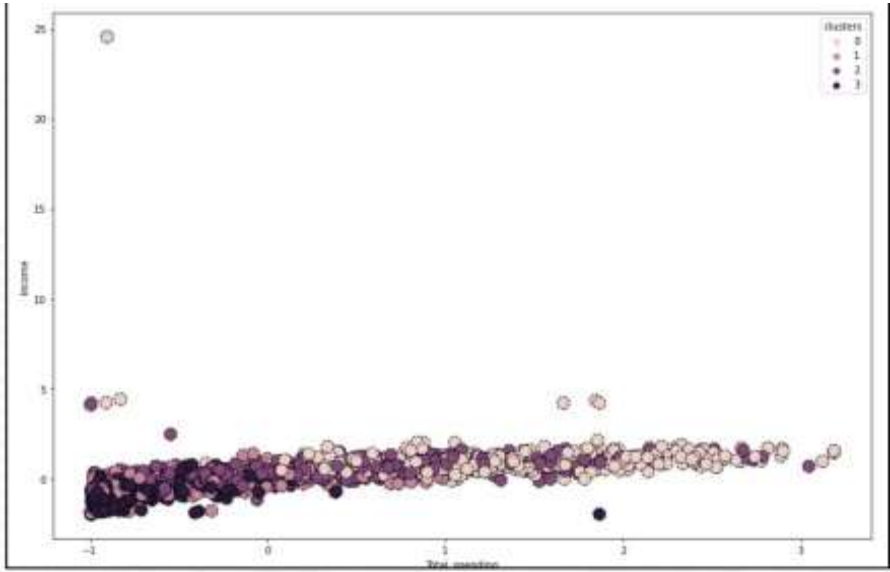
Silhouette score value in cluster 4Has the highest score.
So that using 4 clusters is considered sufficient for the selection of customer clusters

For details, see jupyter notebook here

4

Customer Personality Analysis for Marketing Retargeting





Macam – macam cluster

High Spender	: cluster 0
Mid Spender	: cluster 2
Low Spender	: cluster 1
Risk Churn	: cluster 3

For details, see jupyter notebook here

Summary Per Cluster

Cluster	Jenis Cluster	Total visit per bulan	Rata –rata conv rate	Jumlah Pelanggan	Total Spending
0	High Spender	1648	0.113252	475	660.567.000
1	Low Spender	2958	0.020568	484	135.307.000
2	Mid Spender	2416	0.045262	540	454.259.000
3	Risk Churn	4887	0.012395	741	106.855.000

For details, see jupyter notebook here

Insight dan Business Recommendation

1. High spenders can be said to be loyal customers. The conversion rate has the highest average so it is deep respond to campaigns. Customers like this do not need special treatment because they are looking for products what they want. The most important thing is to maintain stock availability so they keep shopping.
2. Mid Spender has a second number of subscribers with 4% conversion. This type of customer is necessary a more personal treatment such as a birthday promo, so they can go up to a high spender.
3. Low spenders need to do retargeting paid. The number of customers exceeds the high spender, but the power buy less. So it is necessary to do retargeting on this type of cluster.
4. Risk Churn has the highest number of visitors, but with the smallest spending. Recommendations for This type of cluster is a promo, either a discounted price or a postage voucher. Besides saving costs promo,

For details, see jupyter notebook here

Thanks!

Let's Connect

 **Mustiadi Zakki**

 **Mustiadi Zakki**