

Predict Customer Personality to boost marketing campaign by using Machine Learning



Created by:

Ahya Ramdhanitasari

ahyaramdha02@gmail.com

www.linkedin.com/in/ahya-ramdhanitasari

“Bachelor degree of Science from Geography Majoring, University of Indonesia with spatial data (geographic information system) analysis specialization and want to expand my career to data science and business analyst. Equipped with relevant trainings or courses to support my career. Good to operate python data science, SQL query, and BI tools. Likes to explore, problem solving, and analysis activities. Detailed and organized person.”

“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 ”

Fix Data Type

- Membetulkan tipe data yang memiliki kesalahan tipe data, contohnya kolom Dt_Customer memiliki tipe object yang seharusnya tipe Datetime.

Fill NaN

- Mengisi kolom yang memiliki nilai NaN yang terdapat pada kolom 'Income' dengan nilai median dari kolom tersebut.

Feature Engineering

- Membuat kolom baru dari kolom yang telah ada dengan algoritma atau digrupkan berdasarkan range/kelompoknya, yaitu Accepted campaign total, total pembelian, total konsumsi, conversion rate, umur, hingga jumlah anggota keluarga.

Drop Columns

- Drop kolom atau variable yang tidak penting, seperti kolom

Drop Duplicate Values

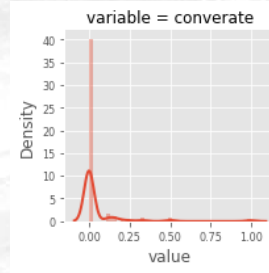
- Mengecek dan membuang value yang duplikat pada dataframe. Ternyata tidak ada value yang dupliasi

Conversion Rate Analysis Based on Income, Consumption, and Age

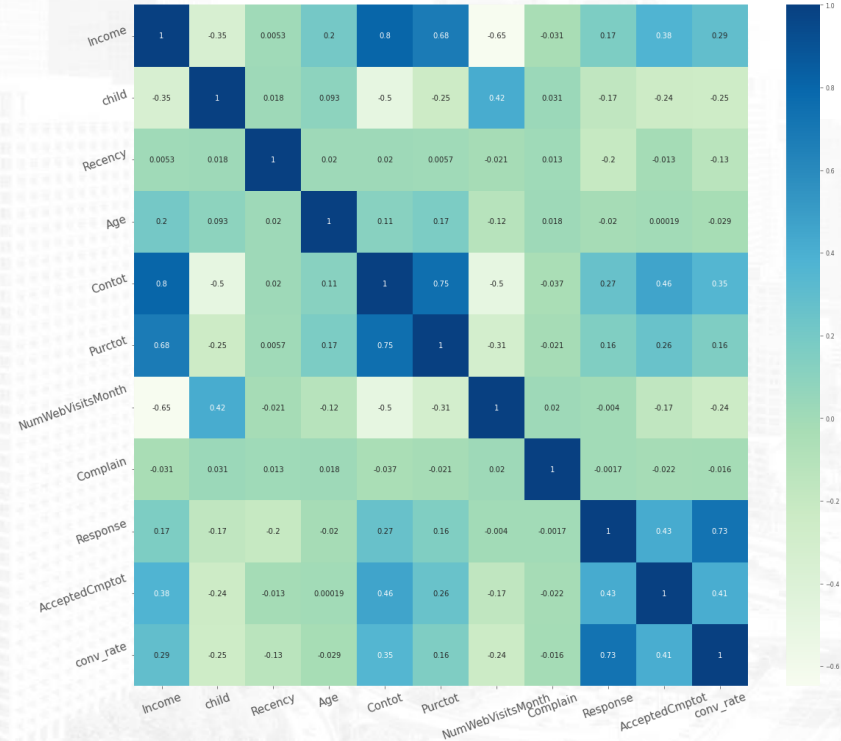
Dari heatmap disamping didapatkan bahwa variabel yang tidak berlawanan dengan variabel conversion rate, yaitu:

- Income,
- Konsumsi total,
- Respon,
- Accepted Campaign.

Selain dari itu berlawanan yang berarti semakin besar nilainya, maka semakin kecil conversion ratenya.

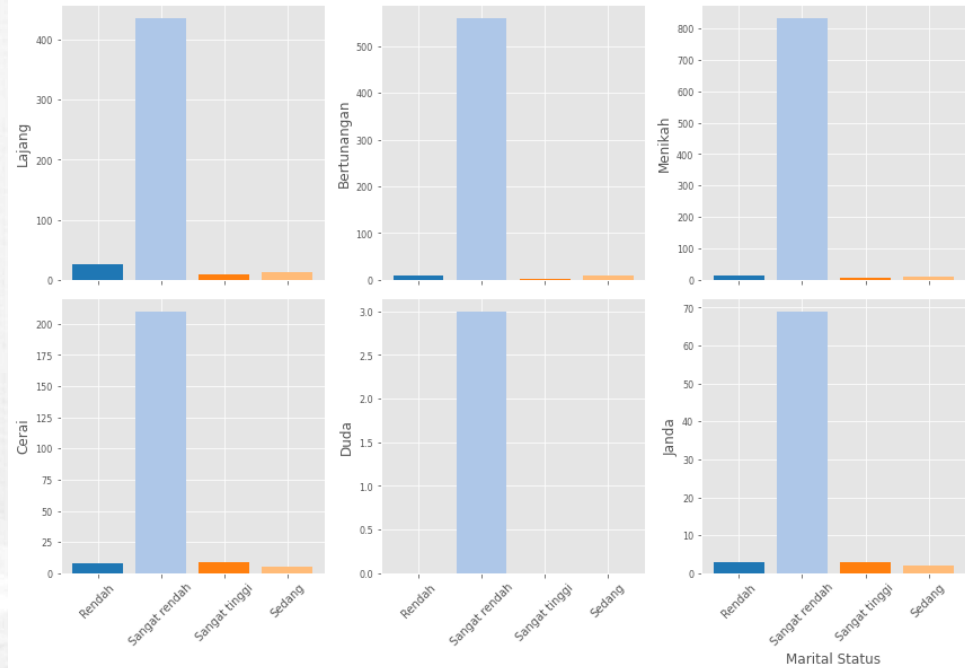


Heatmap Korelasi antar Variabel



Dari visualisasi disamping, didapatkan bahwa status lajang, cerai, dan janda umumnya memiliki kenaikan conversion rate dan memiliki arti bahwa customer dengan status ini lebih banyak mengunjungi website toko.

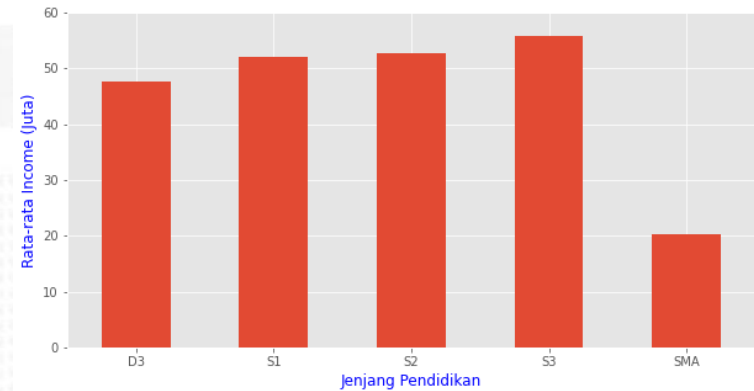
Conversion Rate berdasarkan Status



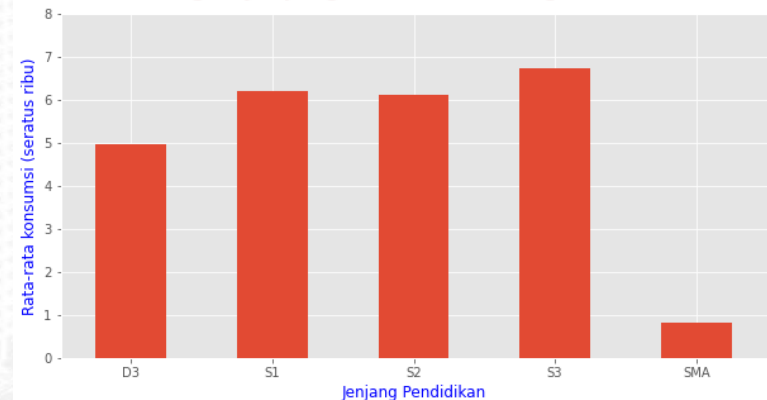
Relation Between Education with Income and Consumption

Berdasarkan visualisasi disamping, didapatkan bahwa rata-rata income akan semakin naik untuk kenaikan setiap jenjang dan hal tersebut cukup berlaku pula pada rata-rata jumlah konsumsi. Sehingga dapat disimpulkan bahwa semakin tinggi pendapatan, umumnya semakin tinggi pula total konsumsinya.

Hubungan Jenjang Pendidikan dengan Income

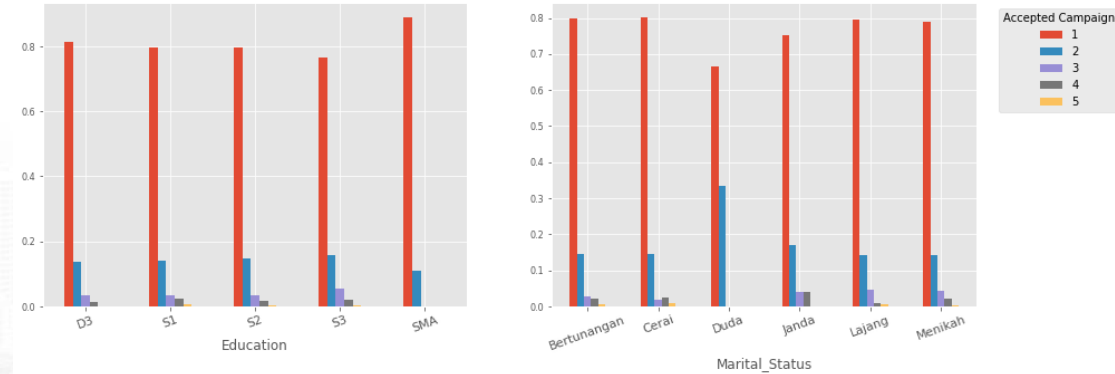


Hubungan Jenjang Pendidikan dengan Konsumsi

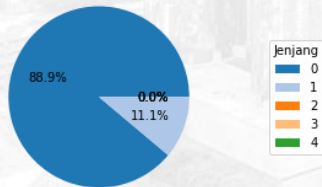


Accepted Campaign Analysis Based on Education and Marital Status

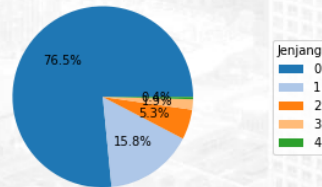
Dari visualisasi didapatkan bahwa semakin tinggi jenjang pendidikan, semakin tinggi pula campaign yang diterima (dibaca).
Kustomer yang lajang, bertunangan, dan cerai menunjukkan memiliki peluang untuk menerima campaign.



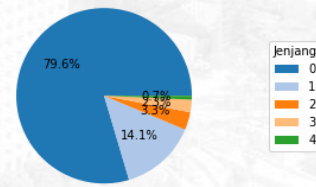
Jumlah Accepted Campaign dengan Edukasi Jenjang SMA



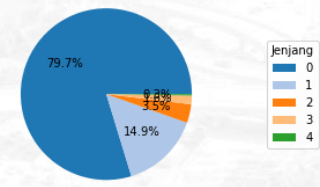
Jumlah Accepted Campaign dengan Edukasi Jenjang S3



Jumlah Accepted Campaign dengan Edukasi Jenjang S1

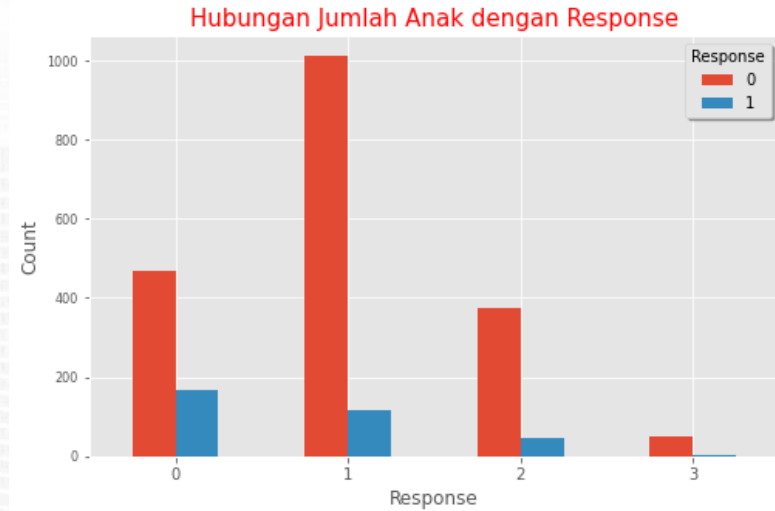


Jumlah Accepted Campaign dengan Edukasi Jenjang S2

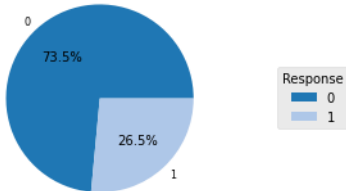


Response Analysis Based on Child in Home

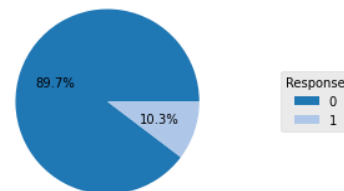
Dari visualisasi didapatkan bahwa semakin banyak anak dalam keluarga customer, semakin kecil persentase customer untuk merespon campaign.



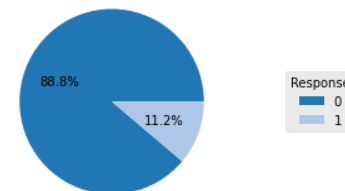
Response dengan Tidak Punya Anak



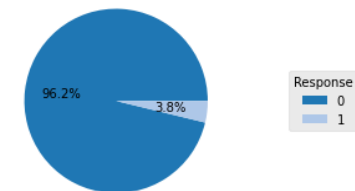
Response dengan Jumlah 1 Anak



Response dengan Jumlah 2 Anak



Response dengan Jumlah 3 Anak

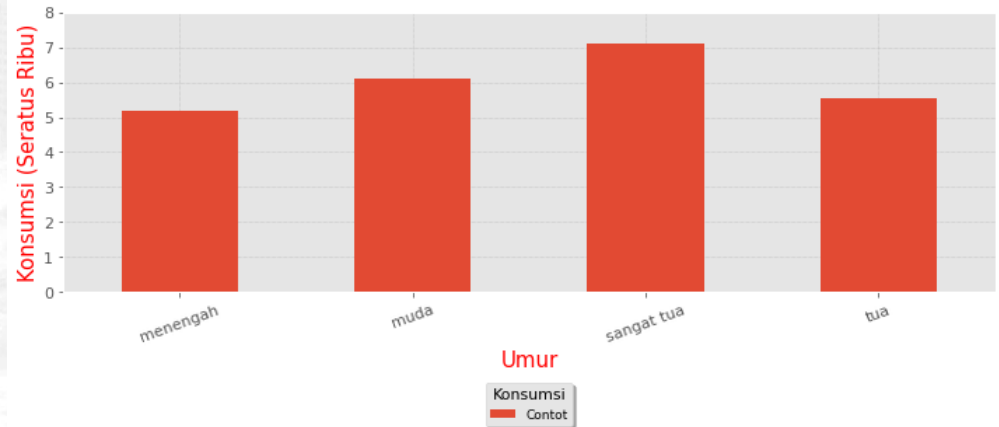


Keterangan:

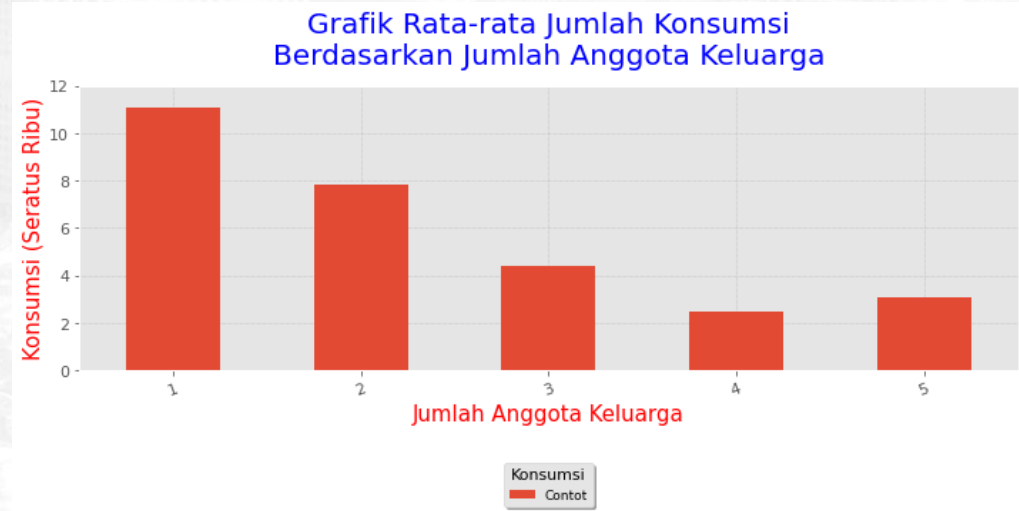
- < 30 tahun (muda)
- 30 – 39 tahun (menengah)
- 40 – 50 tahun (tua)
- > 50 tahun (sangat tua)

Kustomer yang memiliki jumlah konsumsi yang tinggi ialah customer dengan rentang kategori umur muda dan sangat tua.

Grafik Rata-rata Jumlah Konsumsi Berdasarkan Umur



Kustomer yang memiliki jumlah konsumsi yang tinggi ialah customer yang memiliki anggota keluarga hanya sendiri – 3 orang.



1. Customer dengan status janda, cerai dan lajang dapat dapat ditargetkan untuk diiklankan campaign melalui website karena mengalami peningkatan conversion rate.
2. Campaign lebih ditargetkan untuk jenjang pendidikan yang lebih tinggi seperti S1 keatas karena berpeluang lebih tinggi untuk Accepted/dibaca/diterima dan memiliki rata-rata jumlah konsumsi yang semakin meningkat seiring meningkatnya jenjang pendidikan. Hal tersebut dapat disebabkan oleh jumlah income yang semakin tinggi yang mempengaruhi jumlah pendapatan pula.
3. Campaign produk lebih ditargetkan pada customer yang memiliki umur yang masih muda (< 30 tahun) dan sudah sangat tua (> 50 tahun) karena memiliki jumlah konsumsi yang lebih besar dibandingkan dengan jumlah konsumsi pada kategori umur lainnya.
4. Campaign produk juga sebaiknya lebih ditargetkan pada customer yang memiliki jumlah anggota keluarga sendiri sampai 3 orang saja karena memiliki jumlah konsumsi yang tinggi.

Feature Encoding

Dilakukan pada kolom non-numeric, seperti kolom 'Education' dengan label encoder karena merupakan data yang memiliki level, kolom 'Conversion rate' yang telah dikelompokkan karena merupakan data yang memiliki level, dan one-hot encoder pada kolom 'Marital Status' karena bukan merupakan data yang memiliki level atau tingkatan.

```
1 # label encoder pada Education
2 df['Education'] = df['Education'].astype('category').cat.codes
3 # label encoder pada conv_rate
4 df['conv_rate'] = df['conv_rate'].astype('category').cat.codes
5
6 # one hot encoding pada marital status karena tidak memiliki urutan
7 status_onehot = pd.get_dummies(df['Marital_Status'], prefix='Status')
8 df = df.join(status_onehot)
9 status_onehot
```

	Status_Bertunangan	Status_Cerai	Status_Duda	Status_Janda	Status_Lajang	Status_Menikah
0	0	0	0	0	1	0
1	0	0	0	0	1	0
2	1	0	0	0	0	0
3	1	0	0	0	0	0
4	0	0	0	0	0	1
...

Standarisasi

Dilakukan standarisasi pada dataframe menggunakan standardscaler.

```
1 scaler = StandardScaler()
```

```
1 # Method 2: Modify the existing DataFrame in place  
2 dfnew[dfnew.columns] = scaler.fit_transform(dfnew[dfnew.columns])
```

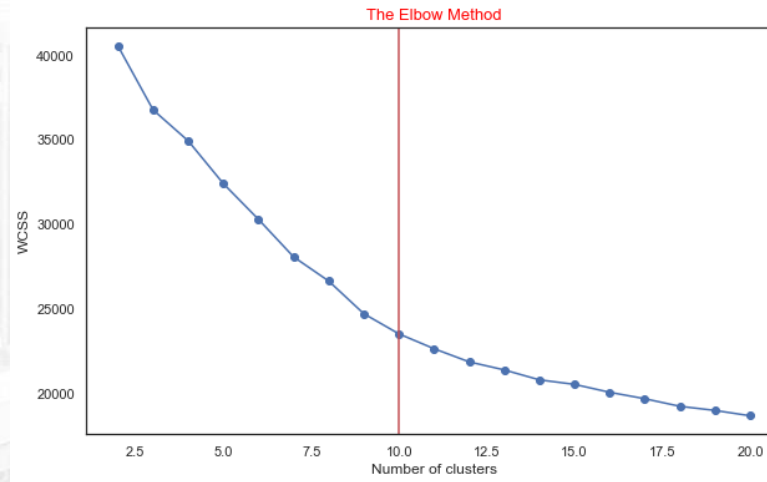
```
1 dfnew
```

	Unnamed: 0	ID	Education	Income	Recency	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	...	Customer_Months
0	-1.731278	-0.020999	-0.579850	0.299651	0.307039	0.983781	1.551577	1.679702	2.462147	1.476500	...	1.944060
1	-1.729731	-1.053058	-0.579850	-0.263808	-0.383664	-0.870479	-0.636301	-0.713225	-0.650449	-0.631503	...	-1.718864
2	-1.728185	-0.447070	-0.579850	0.943420	-0.798086	0.362723	0.570804	-0.177032	1.345274	-0.146905	...	-0.149040
3	-1.726638	0.181716	-0.579850	-1.204881	-0.798086	-0.870479	-0.560857	-0.651187	-0.503974	-0.583043	...	-1.980502
4	-1.725092	-0.082614	1.416866	0.307056	1.550305	-0.389085	0.419916	-0.216914	0.155164	-0.001525	...	-0.803133
...
2235	1.725092	1.625983	-0.579850	0.447037	-0.107383	1.203678	0.419916	0.066692	0.081926	2.203398	...	0.112598
2236	1.726638	-0.490200	1.416866	0.580377	0.237969	0.303291	-0.661449	-0.606873	-0.687068	-0.655733	...	-1.980502
2237	1.728185	0.516905	-0.579850	0.244375	1.446700	1.795020	0.545656	0.221789	-0.101168	-0.364974	...	-0.803133
2238	1.729731	0.814199	0.418508	0.830288	-1.419719	0.368666	0.092992	0.208495	0.777683	0.071165	...	-0.803133
2239	1.731278	1.174650	1.416866	0.047924	-0.314594	-0.653555	-0.586005	-0.469501	-0.650449	-0.631503	...	1.159147

2240 rows × 29 columns

Pemilihan Jumlah Cluster

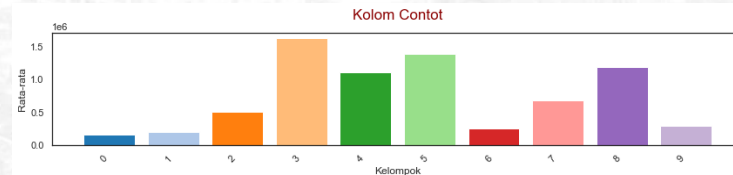
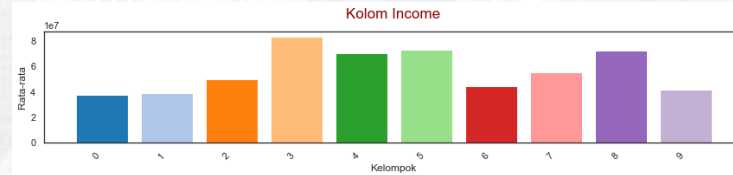
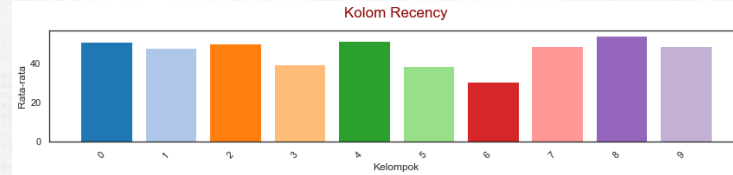
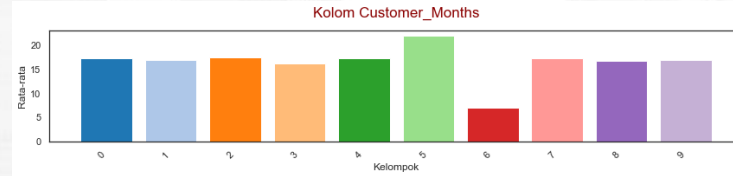
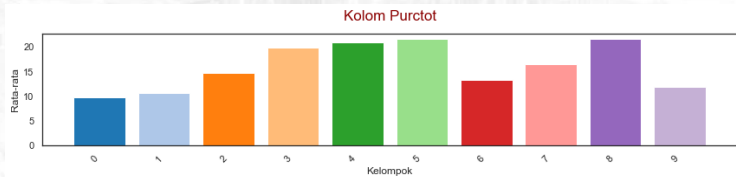
Jumlah *cluster* dipilih menggunakan metode elbow yang akhirnya didapatkan sebanyak 10. Jumlah tersebut adalah jumlah yang paling cocok karena memiliki *silhouette score* sebesar 0,17585 yang paling mendekati 1 daripada jumlah *cluster* yang lain.



Untuk selengkapnya, dapat melihat jupyter notebook disini:
https://drive.google.com/drive/folders/1cgWaY9gReAX5AAfQKFXewq4_gAiaYlii?usp=sharing

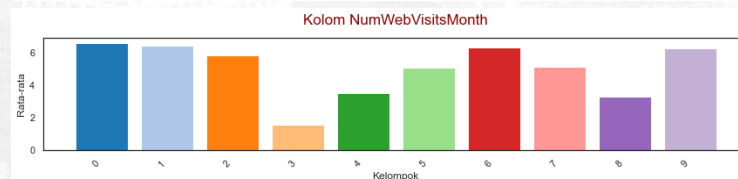
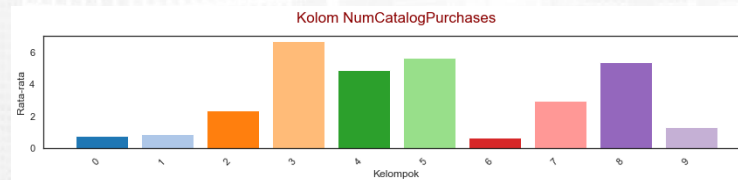
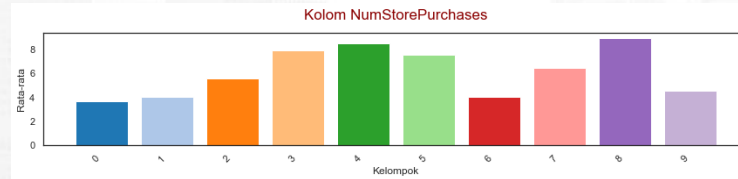
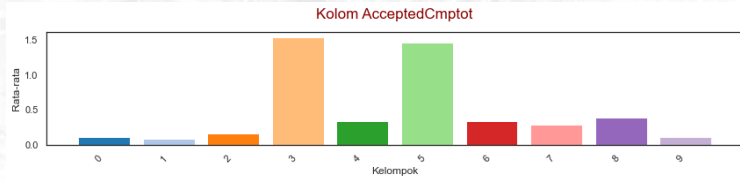
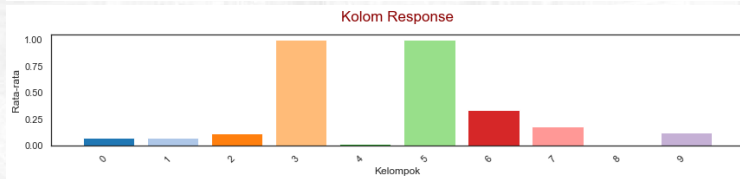
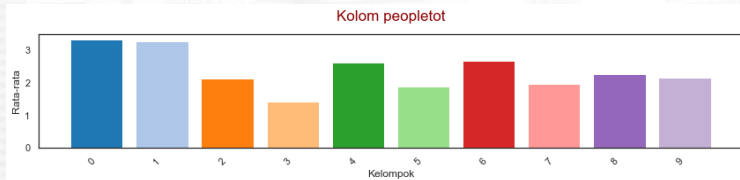
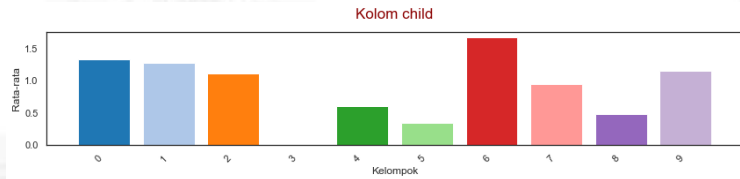
Melakukan *Clustering* Menggunakan KMeans

	Education	Marital_Status	age_group	kmeans_10cluster
kmeans_10cluster				
0	S1	Bertunangan	sangat tua	0.0
1	S1	Menikah	tua	1.0
2	S1	Cerai	sangat tua	2.0
3	S1	Lajang	sangat tua	3.0
4	S1	Bertunangan	sangat tua	4.0
5	S1	Lajang	sangat tua	5.0
6	S1	Duda	muda	6.0
7	S1	Janda	sangat tua	7.0
8	S1	Menikah	sangat tua	8.0
9	S1	Lajang	tua	9.0



Untuk selengkapnya, dapat melihat jupyter notebook disini:
https://drive.google.com/drive/folders/1cgWaY9gReAX5AAfQKFXewq4_gAiaYlii?usp=sharing

Melakukan *Clustering* Menggunakan KMeans



Untuk selengkapnya, dapat melihat jupyter notebook disini:
https://drive.google.com/drive/folders/1cgWaY9gReAX5AAfQKFxewq4_gAiaYlii?usp=sharing

Interpretasi tiap kelompok

0. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status bertunangan dan usianya sangat tua, memiliki anak yang banyak, resensi yang tinggi, pendapatan yang kecil, serta pembelian dan jumlah konsumsi yang sedikit. Sering Mengunjungi website toko.
1. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status menikah dan usianya tua, memiliki anak yang banyak, resensi yang tinggi, pendapatan yang kecil, serta pembelian dan jumlah konsumsi yang sedikit. Hampir sama dengan kelompok pertama. Sering Mengunjungi website toko.
2. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status cerai dan usianya sangat tua, memiliki anak yang cukup, resensi yang sedang, pendapatan menengah, pembelian yang sedang, dan jumlah konsumsi yang sedikit. Sering Mengunjungi website toko.
3. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status lajang dan usianya sangat tua, tidak memiliki anak, resensi menengah, pendapatan yang tinggi, serta pembelian, jumlah konsumsi yang tinggi, serta memiliki peluang respon dan *campaign accepted* yang besar. Sering Membeli melalui katalog dan toko fisik.
4. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status bertunangan dan usianya sangat tua, memiliki anak yang sedikit, resensi yang tinggi, pendapatan yang tinggi, pembelian yang tinggi, dan jumlah konsumsi menengah. Sering membeli melalui toko fisik.

Untuk selengkapnya, dapat melihat jupyter notebook disini:

https://drive.google.com/drive/folders/1cgWaY9gReAX5AAfQKFXewq4_gAiaYlii?usp=sharing

Interpretasi tiap kelompok

5. Didominasi oleh kustomer yang sudah lama berlangganan dengan status lajang dan usianya sangat tua, memiliki anak yang sedikit, resensi menengah, pendapatan yang tinggi, pembelian dan jumlah konsumsi yang tinggi, serta memiliki peluang respon dan *campaign accepted* yang besar. Sering Membeli melalui katalog dan toko fisik.
6. Didominasi oleh kustomer yang baru berlangganan dengan status duda dan usianya muda, memiliki anak yang banyak, resensi yang rendah, pendapatan yang kecil, pembelian menengah, dan jumlah konsumsi yang sedikit. Sering Mengunjungi website toko.
7. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status janda dan usianya sangat tua, memiliki anak yang cukup, resensi menengah, pendapatan menengah, pembelian menengah, dan jumlah konsumsi yang sedikit. Cukup sering membeli di toko fisik dan mengunjungi website toko.
8. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status menikah dan usianya sangat tua, memiliki anak yang sedikit, resensi yang tinggi, pendapatan yang tinggi, pembelian yang tinggi, dan jumlah konsumsi menengah. Sering membeli melalui toko fisik.
9. Didominasi oleh kustomer yang sudah cukup lama berlangganan dengan status lajang dan usianya tua, memiliki anak cukup, resensi yang rendah, pendapatan yang rendah, serta pembelian dan jumlah konsumsi yang rendah. Sering Mengunjungi website toko.

Untuk selengkapnya, dapat melihat jupyter notebook disini:

https://drive.google.com/drive/folders/1cgWaY9gReAX5AAfQKFXewq4_gAiaYliy?usp=sharing

1. **Kustomer duda** umumnya masih muda dan memiliki anak dengan jumlah pendapatan yang sedikit sehingga **tidak direkomendasikan untuk diberikan campaign** terus-menerus terutama tentang hal yang tidak berkaitan dengan kehidupan kustomer dan **jikalau ingin share campaign**, sebaiknya **dilakukan di website toko** karena kustomer jenis ini sering membuka website toko.
2. **Kustomer yang sudah menikah dan memiliki anak** yang **sedikit** umumnya memiliki pendapatan dan pembelian yang besar, sehingga **direkomendasikan untuk diberikan campaign melalui toko fisik** karena lebih sering berkunjung ke toko fisik, **berbeda dengan kustomer yang sudah menikah namun memiliki anak** yang **banyak** yang memiliki pembelian dan konsumsi yang sedikit, namun jikalau ingin memberikan campaign, sebaiknya **melalui website toko** karena kustomer jenis ini sering mengunjungi website toko.
3. **Kustomer yang masih lajang** umumnya memiliki income yang tinggi, sehingga total pembelian dan konsumsi menjadi tinggi pula, sehingga **dapat direkomendasikan menjadi kustomer potensial untuk diberikan campaign melalui toko fisik dan katalog**.
4. **Kustomer yang sudah berpasangan** umumnya **memiliki anak** dan peluang response dan accepted campaign sangat kecil walaupun dengan pendapatan yang tinggi, sehingga **tidak dapat menjadi kustomer potensial untuk selalu diberi campaign** dan **jika ingin menyebarkan campaign** sebaiknya **melalui website toko** karena kustomer jenis ini sering mengunjungi website toko.

Thank You!

Let's Connect!

Portfolio Link
Github project

: <https://bit.ly/portoahya>

: https://github.com/ahyaramdha/predict_customer_personality