

# SAS Data Science Competition

Marketing Campaign  
Analytics

# Biodata



Nama : Iqbal Hanif  
Usia : 26 tahun  
Pekerjaan : Big Data Analytics Officer  
Institusi : PT. Telkom Indonesia  
Domisili : Jakarta

Ketertarikan :

- Machine Learning
- Business Intelligence
- Data Engineering

Riwayat proyek :

- Prediksi Skor/Rating Program TV (Time Series) dengan Decision Tree Regressor ([link](#))
- Langkah Dasar Prediksi Churn Menggunakan Regresi Logistik ([link](#))
- A/B Testing untuk Menguji Efektivitas antara Dua Digital Ads ([link](#))
- Twitter Crawling: Sumber Data untuk Melakukan Social Media Monitoring ([link](#))
- Analisis Biplot: Penyederhanaan Interpretasi Data Multivariat ([link](#))

# Background

SAS-MART telah mengadakan marketing campaign berupa promo “khusus member SAS-MART, dapatkan cashback Rp 50K, setiap belanja di SAS-MART minimal Rp 800K” kepada 5.000 member. SAS-MART berharap agar hasil dari campaign awal ini akan dijadikan dasar untuk **menentukan member yang nantinya akan diberikan promo agar campaign ini efektif.**

Saya menjabarkan problem efektivitas campaign ke dalam 3 pertanyaan utama:

1. Bagaimana memprediksi pelanggan yang akan menggunakan promo tersebut? Faktor apa saja yang mempengaruhinya?
2. Berasal dari segmen apakah pengguna promo terbanyak? Bagaimana karakteristik segmen terbaik, yaitu segmen pelanggan yang berpotensi meningkatkan pendapatan SAS-MART?
3. Bagaimana memanfaatkan data campaign yang lalu untuk membuat campaign berikutnya yang lebih menarik dan personalized?



# Solutions

Saya menawarkan solusi yang dapat meningkatkan efektivitas dari marketing campaign yang akan dijalankan oleh SAS-MART kedepannya, yaitu:

## 1. **Promo Redemption Model with SHAP value**

Model yang akan memprediksi apakah seorang pelanggan akan menggunakan promo yang diberikan, beserta faktor-faktor yang mempengaruhi keputusan tersebut

## 2. **Customer Segmentation with RFM**

Membagi pelanggan yang mengikuti campaign ke dalam beberapa segmen, kemudian menilai prospek dari masing-masing segmen, dan memberikan rekomendasi segmen manakah yang berpotensi meningkatkan pendapatan SAS-MART.

## 3. **Customer Profiling Dashboard**

Dashboard yang memberikan profil lengkap pelanggan yang mengikuti campaign, sebagai referensi dalam membangun gimmick campaign yang lebih menarik dan personalized sehingga memiliki tingkat keberhasilan yang lebih tinggi.



# Promo Redemption Model



# Promo Redemption Model

Pemberian promo merupakan strategi marketing yang bisa memberikan banyak keuntungan bagi pelaku bisnis. Pemberian promo kepada member akan mengundang pelanggan non-member untuk menjadi member. Disamping itu, promo juga bisa meningkatkan engagement dengan member existing. Promo bahkan bisa menjadi daya tarik yang luar biasa, seperti yang dilakukannya Amazon di Amazon Day 2019, yang mampu meningkatkan penjualan total mereka melebihi gabungan event Black Friday dan Cyber Monday 2018 (Business Insider).

Walaupun begitu, keputusan pemberian promo tetap harus melalui analisa data agar berjalan secara efektif dan tidak mubazir. Model prediktif dapat membantu meningkatkan efektifitas promo dengan beberapa cara:

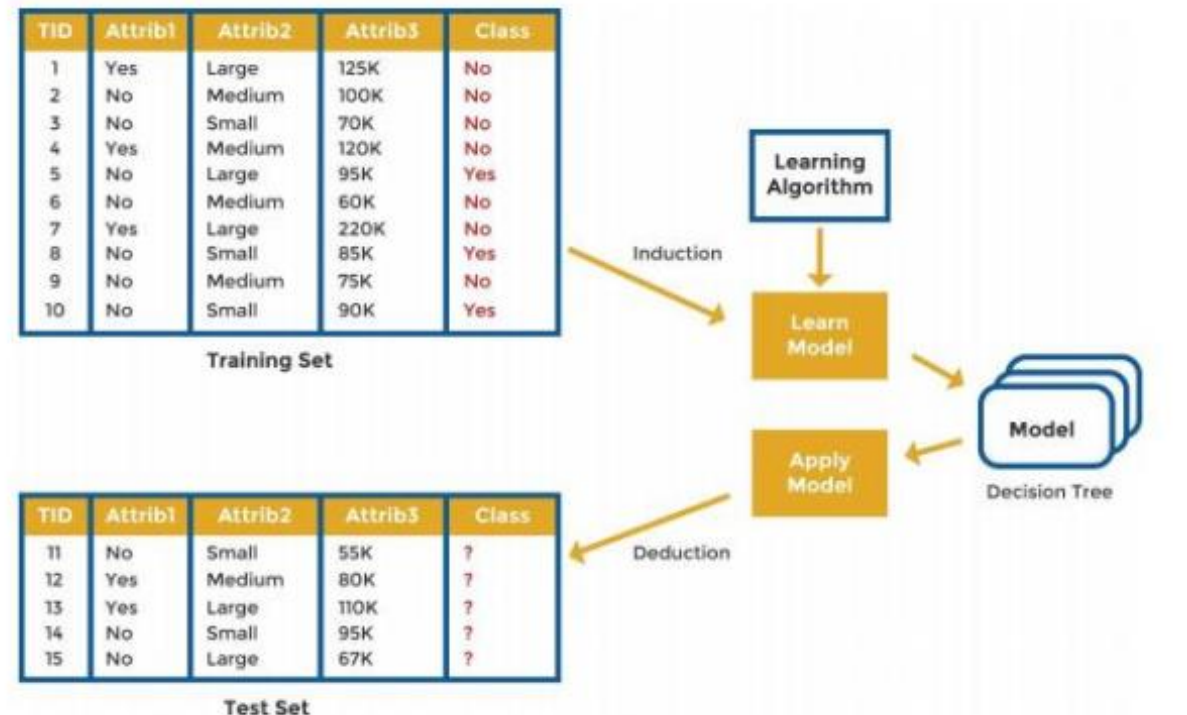
1. Memprediksi pelanggan (dan total pelanggan) yang akan menggunakan promo tersebut.
2. Menganalisa faktor-faktor yang mempengaruhi keputusan pelanggan dalam menggunakan promo tersebut.
3. Membuat rekomendasi campaign lanjutan berdasarkan hasil prediksi dan analisa faktor-faktor yang berpengaruh, agar campaign lebih baik dan lebih efektif kedepannya.



# Classification Machine Learning

Classification machine learning merupakan metode machine learning yang berupaya untuk mengidentifikasi kategori pada amatan yang belum diketahui kategorinya (test set), berdasarkan pada data yang setiap amatannya sudah diketahui kategorinya (train set). Output dari machine learning tersebut adalah sebuah model yang dapat digunakan untuk memprediksi kategori untuk setiap observasi/amatan baru di masa mendatang.

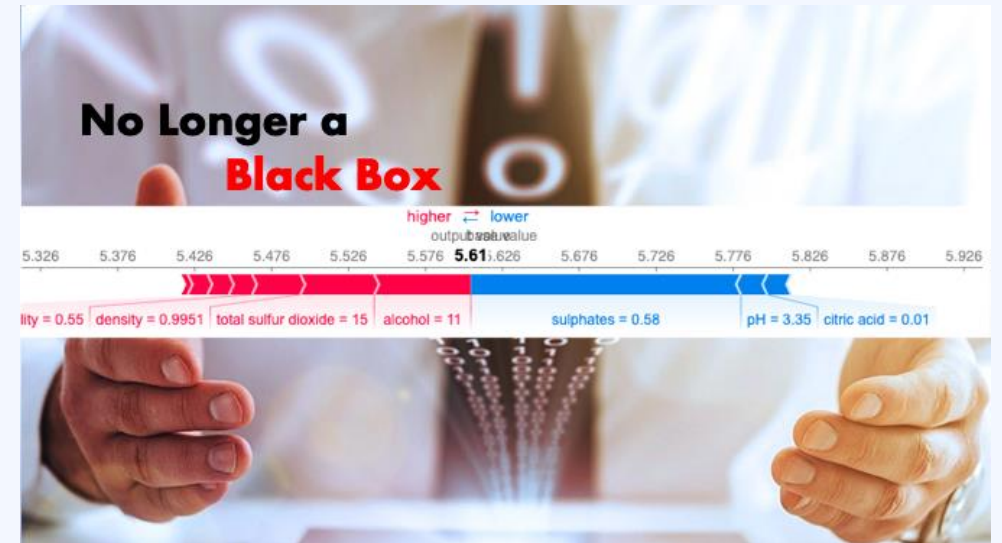
Algoritma yang bisa digunakan untuk klasifikasi: Linear Classifiers (Logistic Regression, Naive Bayes Classifier), Nearest Neighbor, Support Vector Machines, Decision Trees, Boosted Trees, Random Forest, Neural Networks, etc



# SHAP Value

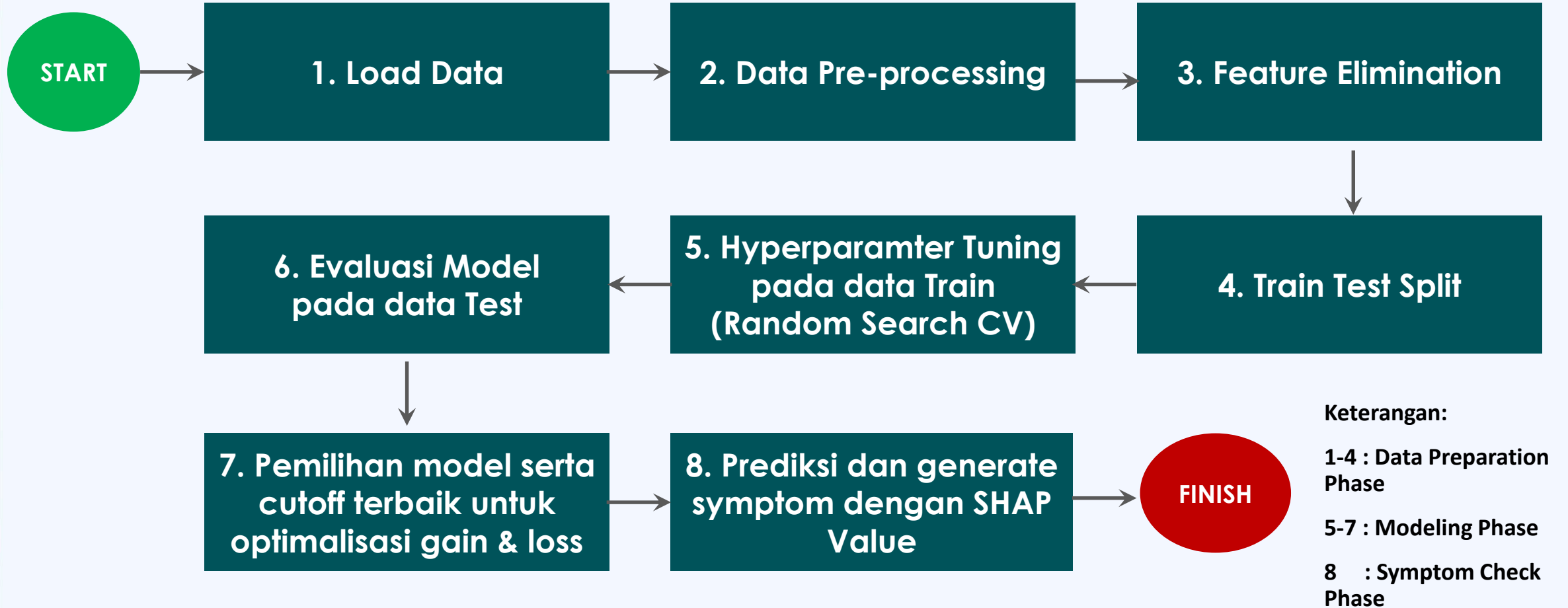
SHAP Value dikembangkan oleh Lundberg dan Lee (2016), yaitu metode untuk menjelaskan prediksi baik secara global (keseluruhan) maupun lokal (individu). SHAP didasarkan pada permainan Terinspirasi oleh beberapa metode interpretasi model, Lundberg dan Lee (2016) mengusulkan nilai SHAP sebagai pendekatan terpadu untuk menjelaskan output dari model machine learning apa pun. Ada tiga keuntungan dari SHAP value:

- *global interpretability* (dapat menunjukkan seberapa berkontribusi prediktor ke variabel target).
- *local interpretability* (interpretabilitas lokal yang memungkinkan kita untuk menunjukkan dan membedakan dampak dari faktor-faktor prediksi ke masing-masing amatan).
- SHAP values bisa diukur untuk tree-based model apapun.





# Process Flow



# Data Preparation Phase

Beberapa langkah yang dilakukan pada tahap ini adalah

1. Melakukan upload data pada analytics platform yang kita gunakan,
2. Melakukan pemeriksaan Null values, jika ditemukan Null Values maka akan dilakukan imputasi menggunakan ukuran pemusatan (mean, median, atau modus).
3. Melakukan outlier handling dengan memeriksa pencilan dan menggantinya dengan nilai  $\mu+3\sigma$  (batas atas) dan  $\mu-3\sigma$  (batas bawah).
4. Melakukan standarisasi dengan menggunakan pendekatan normal baku.

$$z = \frac{x - \mu}{\sigma}$$

5. Melakukan variable selection dengan menggunakan VIF (Variance Inflation Factor) score untuk mencegah multikolinearitas. Variabel dengan  $VIF > 5$  akan di reject.
6. Melakukan train-test split dengan proporsi 70% train dan 30% test.

# Data Preparation Result

## Variabe Selection

No	Variable	VIF	Decision
1	is_single	Inf	Rejected
2	is_cash	Inf	Rejected
3	is_male	Inf	Rejected
4	spending_last_2mo	5.861930	Rejected
5	visit_last_1mo	<5	Accepted
6	visit_last_2mo	<5	Accepted
7	visit_last_3mo	<5	Accepted
8	spending_last_1mo	<5	Accepted
9	spending_last_3mo	<5	Accepted
10	age	<5	Accepted
11	monthly_income	<5	Accepted
12	buy_groceries	<5	Accepted
13	buy_toiletries	<5	Accepted
14	buy_food	<5	Accepted
15	buy_electronic	<5	Accepted
16	buy_clothes	<5	Accepted
17	buy_home_appliances	<5	Accepted
18	recency_last_visit	<5	Accepted
19	is_married	<5	Accepted
20	is_debit	<5	Accepted
21	is_credit	<5	Accepted
22	is_female	<5	Accepted

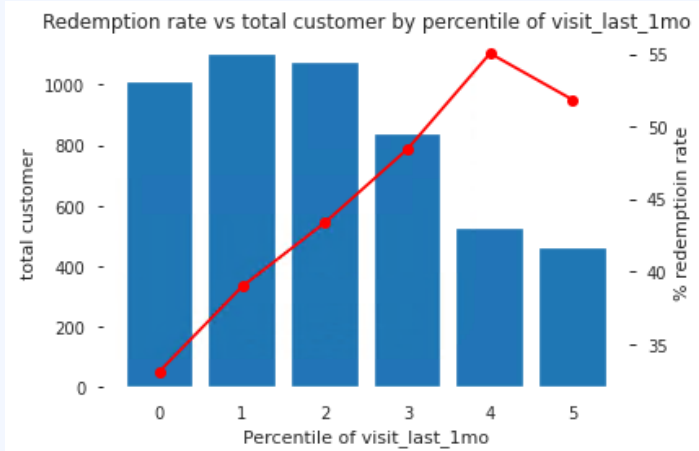
## Train Test Split

Data set	Proportion	Total Rows	Total Columns
Train	70%	3,500	18
Test	30%	1,500	18

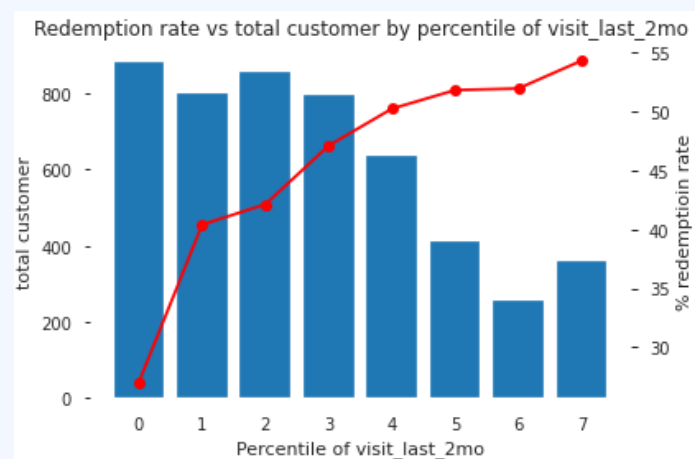
## Penjelasan:

- 4 Variabel ditolak karena memiliki skor VIF > 5
- Tersisa 18 variabel yang akan masuk ke tahap modelling
- Hasil train-test split menghasilkan 3.500 row untuk data train dan 1.500 row untuk data test
- Selanjutnya akan dilakukan exploratory data analysis (EDA) untuk melihat linieritas hubungan antara variabel terpilih dengan tingkat penggunaan promo (redemption rate).

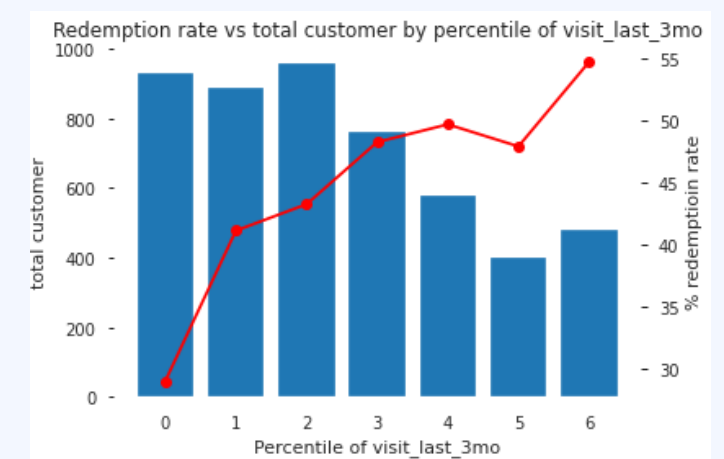
# Exploratory Data Analysis (based on Selected Variables)



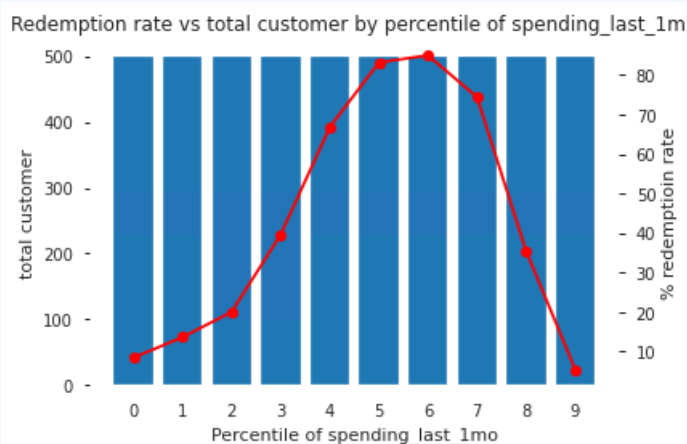
Visit\_last\_1mo berkorelasi positif dengan redemption\_rate



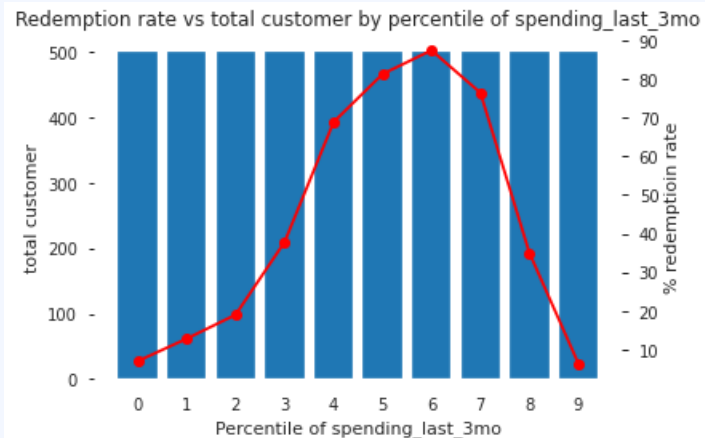
Visit\_last\_2mo berkorelasi positif dengan redemption\_rate



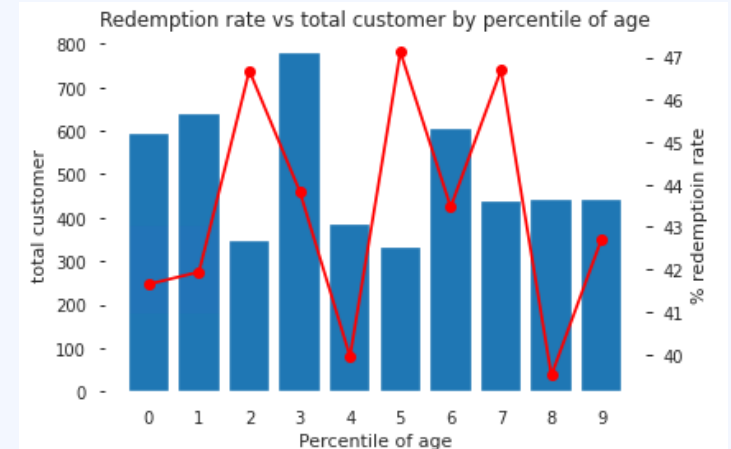
Visit\_last\_3mo berkorelasi positif dengan redemption\_rate



Spending\_last\_1mo berkorelasi positif dengan redemption\_rate untuk nilai dibawah median, negatif untuk nilai diatas median

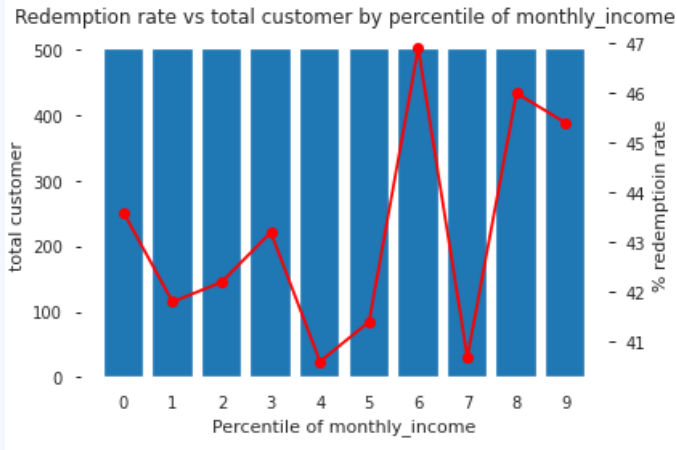


Spending\_last\_3mo berkorelasi positif dengan redemption\_rate untuk nilai dibawah median, negatif untuk nilai diatas median

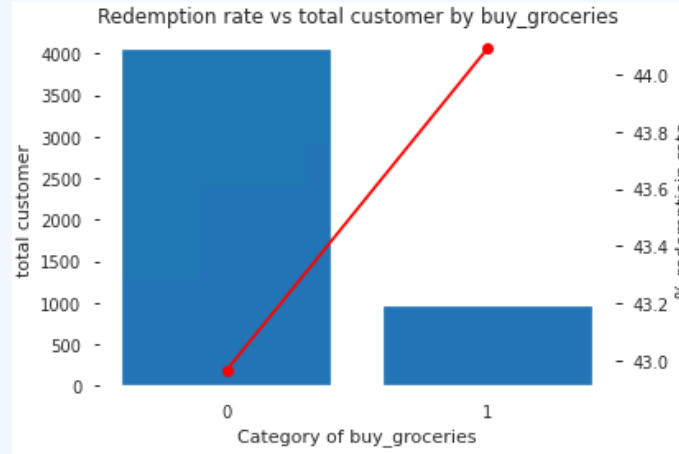


Belum ada relasi yang bisa dijelaskan

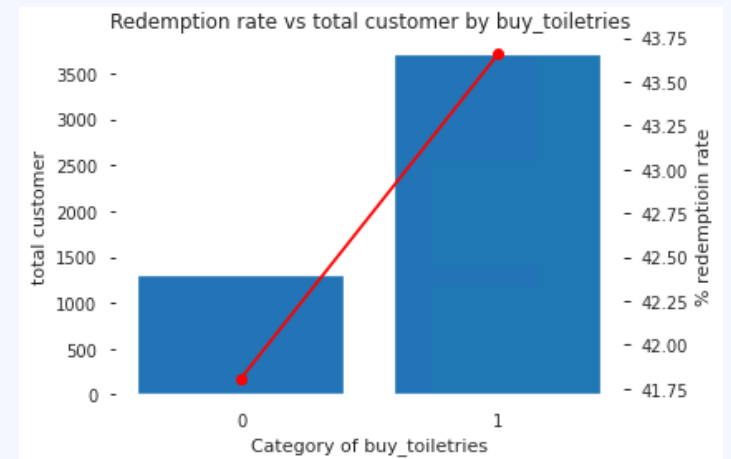
# Exploratory Data Analysis (based on Selected Variables)



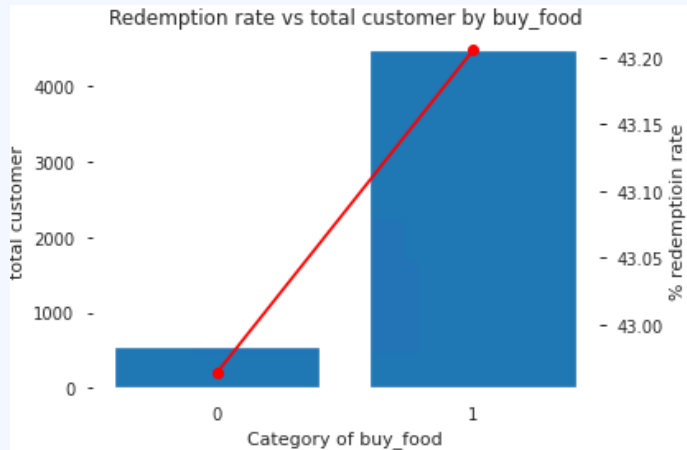
Belum ada relasi yang bisa dijelaskan



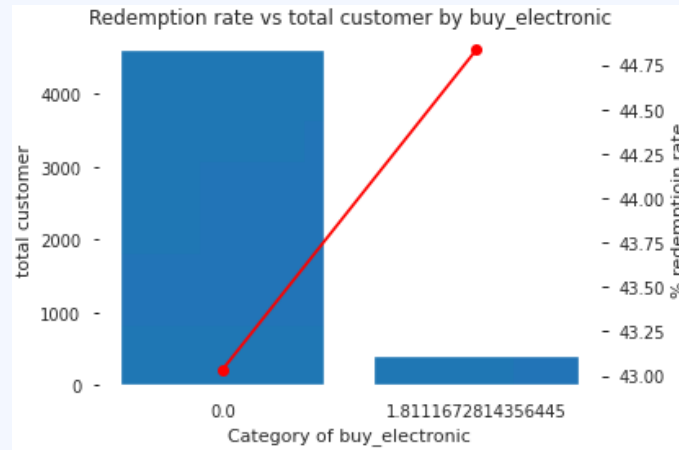
Buy\_groceries berkorelasi positif dengan redemption\_rate



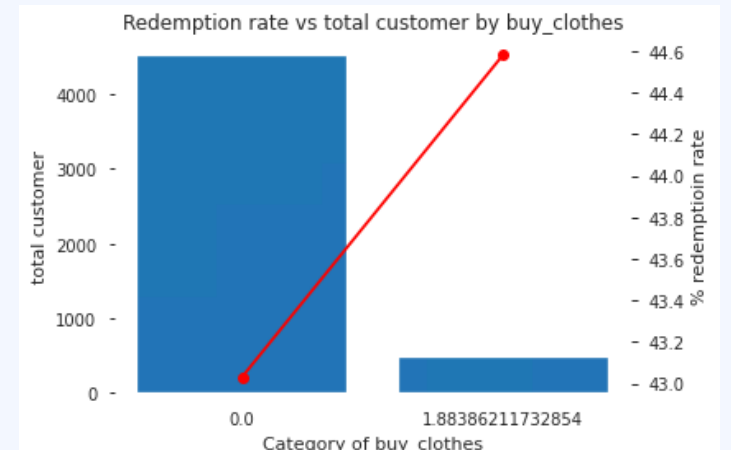
Buy\_toiletries berkorelasi positif dengan redemption\_rate



Buy\_food berkorelasi positif dengan redemption\_rate



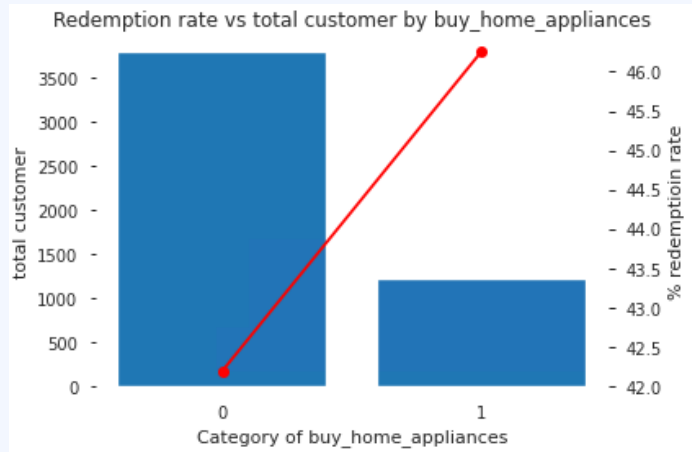
Buy\_electronics berkorelasi positif dengan redemption\_rate



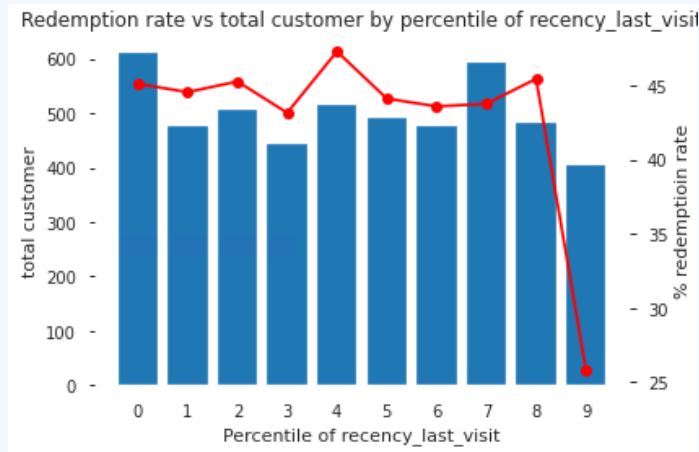
Buy\_clothes berkorelasi positif dengan redemption\_rate



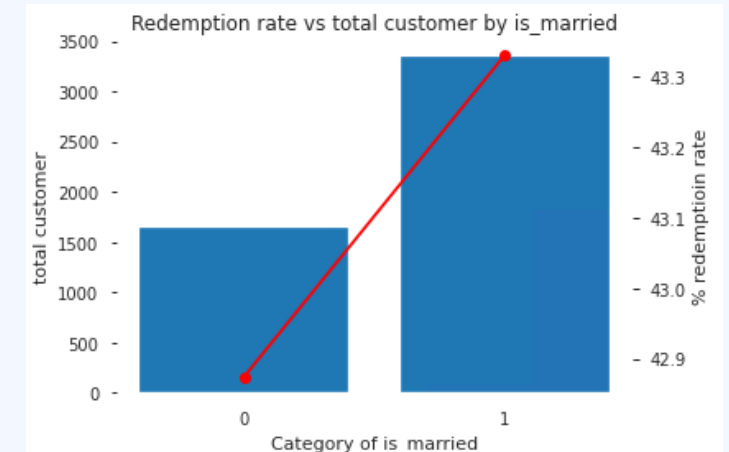
# Exploratory Data Analysis (based on Selected Variables)



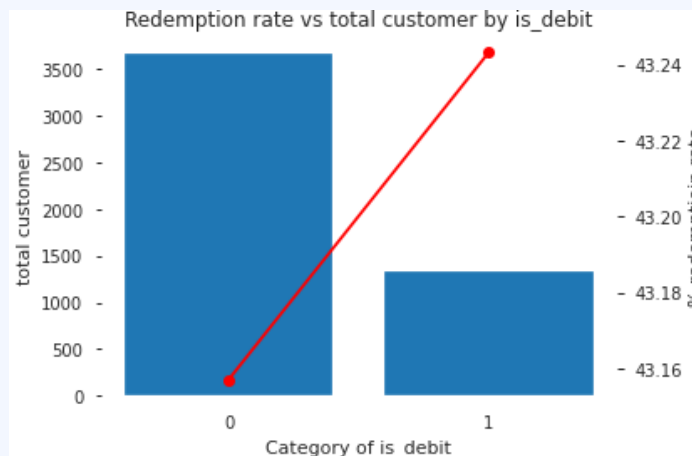
Buy\_home\_appliances berkorelasi positif dengan redemption\_rate



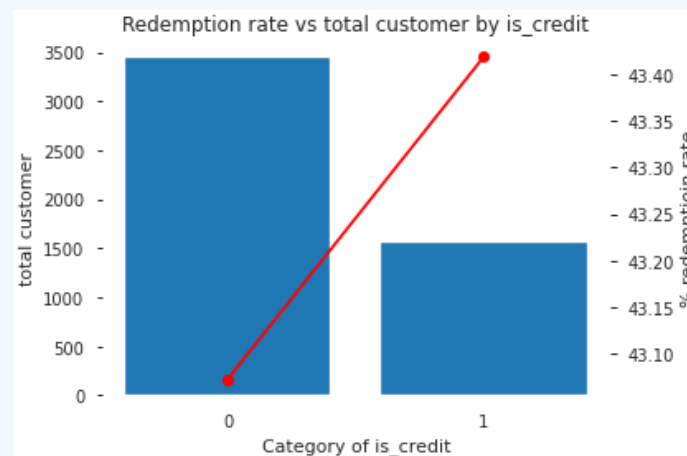
Belum ada relasi yang bisa dijelaskan



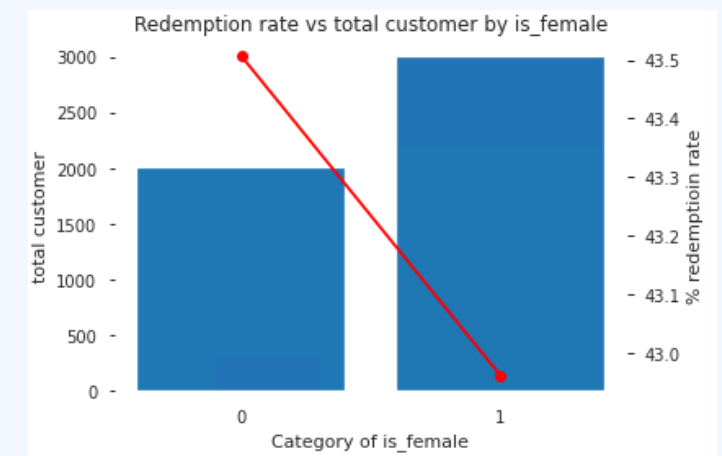
Is\_married berkorelasi positif dengan redemption\_rate



Is\_debit berkorelasi positif dengan redemption\_rate



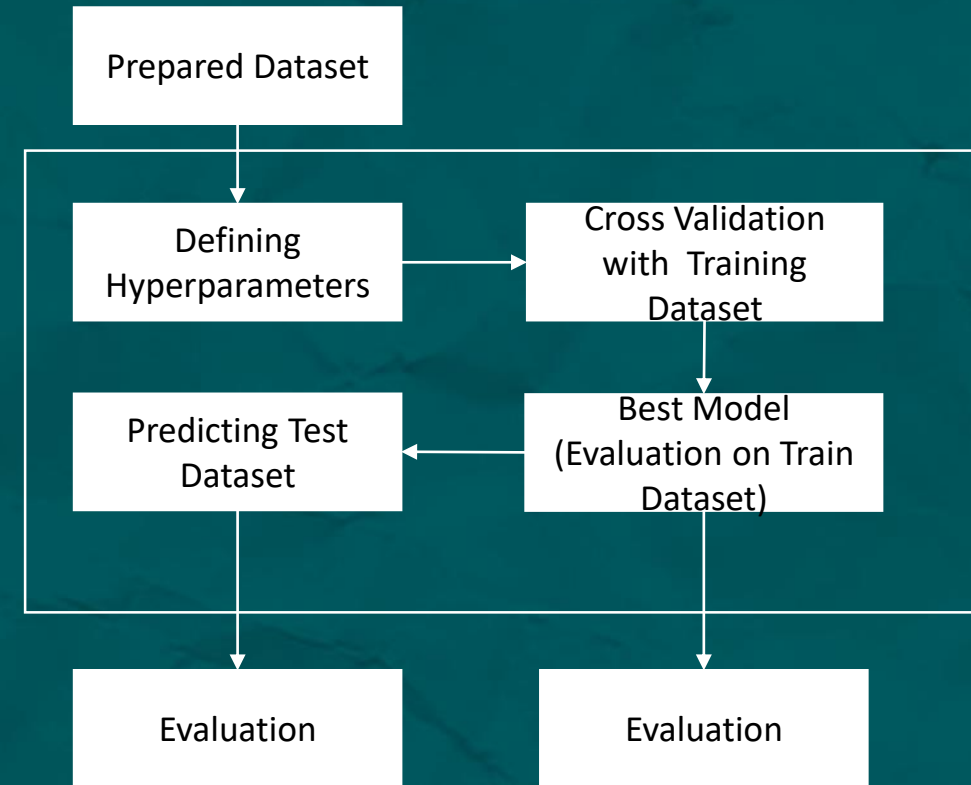
Is\_credit berkorelasi positif dengan redemption\_rate



Is\_female berkorelasi negatif dengan redemption\_rate

# Modeling Phase

- Modeling dilakukan dengan menentukan hyperparameter terbaik melalui metode randomized search k-fold cross validation dengan fold = 10 sebanyak 100 kali iterasi (berarti 100 kombinasi hyperparameter untuk setiap algoritma).
- Algoritma yang diujicobakan adalah:
  1. Decision Tree (Single Learner)
  2. Logistic Regression (Single Learner)
  3. Extreme Gradient Boosting/XGBoost (Ensemble Learner)
  4. Random Forest (Ensemble Learner)
- Evaluasi dilakukan dengan beberapa metrics:
  1. Akurasi
  2. Recall
  3. Precision
  4. ROC-AUC
  5. Lama Proses



# Modeling Result - Evaluation

## Best Hyperparameter

### Decision Tree

Hyperparameter	Value
min_samples_split	5
min_samples_leaf	4
max_features	sqrt
max_depth	10
ROC AUC Validation	0.614

### XGBoost

Hyperparameter	Value
reg_alpha	1
n_estimators	750
min_child_weight	1
max_depth	3
learning_rate	0.015
Gamma	0.5
ROC AUC Validation	0.9247

### Logistic Regression

Hyperparameter	Value
penalty	l2
C	0.0001
ROC AUC Validation	0.9199

### Random Forest

Hyperparameter	Value
n_estimators	1200
min_samples_split	10
min_samples_leaf	4
max_features	sqrt
max_depth	100
Bootstrap	TRUE
ROC AUC Validation	0.9767

## Model Evaluation

### Train Data Set

Algorihm	Accuracy	Recall	Precision	ROC AUC	Time (Cross Validation)	Time (Fitting Best Model)
Decision Tree	0.86629	0.8319	0.85452	0.92973	7.39 s	18.1 ms
Logistic Regresson	0.59029	0.37591	0.53636	0.61403	1.22 s	11.6 ms
XGBoost	0.85771	0.81469	0.84955	0.92475	2 min 46 s	1.87 s
Random Forest	0.89629	0.8544	0.90028	0.97657	11 min 27s	7.42 s

### Test Data Set

Algorihm	Accuracy	Recall	Precision	ROC AUC	Time (Predicting)
Decision Tree	0.81667	0.76698	0.80032	0.85597	608 ms
Logistic Regresson	0.58867	0.39198	0.53249	0.62585	594 ms
XGBoost	0.84133	0.79938	0.82748	0.88937	1.1. s
Random Forest	0.84467	0.79784	0.83522	0.88743	1.37 s

# Modeling Result

## Penjelasan:

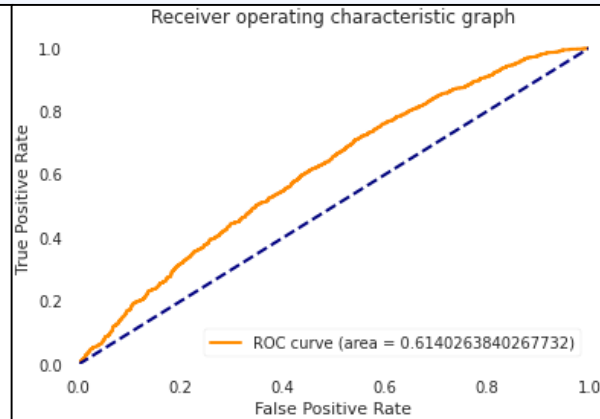
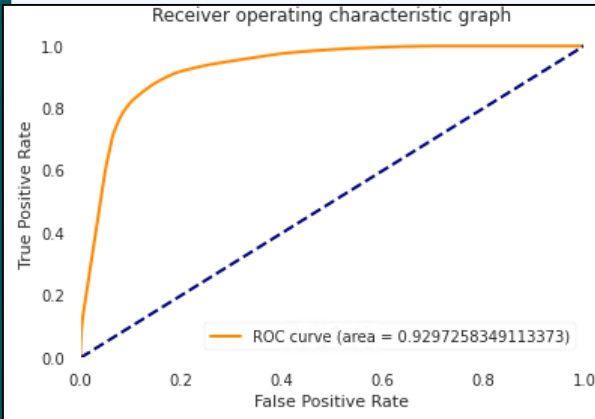
- Setiap algoritma menghasilkan hyperparameter terbaik berdasarkan nilai ROC AUC tertinggi dari proses Randomized K-Fold Cross Validation dengan fold = 10 dan iteration = 100 (total 100 kombinasi hyperparameter yang diuji coba).
- Metode Decision Tree dan Random Forest mengalami overfitting, dimana akurasi turun cukup signifikan dari tahap train ke tahap test.
- Metode Logistic Regression menunjukkan hasil yang kurang baik karena hanya menghasilkan akurasi di sekitar 58% - 60%.
- Metode XGBoost memberikan hasil yang cukup baik dan robust karena tidak ada perbedaan yang signifikan antara akurasi dari data train dengan data test.
- Dikarenakan hasil yang baik ditambah dengan waktu pemrosesan yang lebih cepat dibanding Random Forest, maka ditetapkan bahwa algoritme XGBoost merupakan terbaik dalam membangun model promo redemption.

Cutoff (Train)	Accuracy	Sensitivity	Spesificity		Cutoff (Test)	Accuracy	Sensitivity	Spesificity
0.48	0.855429	0.819325	0.882856		0.48	0.84	0.807099	0.865023
0.49	0.856286	0.816016	0.886878		0.49	0.84	0.800926	0.869718
0.5	0.857714	0.814692	0.890397		0.5	0.841333	0.799383	0.873239
0.51	0.856857	0.811383	0.891403		0.51	0.843333	0.796296	0.879108
0.52	0.857143	0.808736	0.893917		0.52	0.844	0.79321	0.882629
0.53	0.856857	0.804765	0.89643		0.53	0.843333	0.790123	0.883803

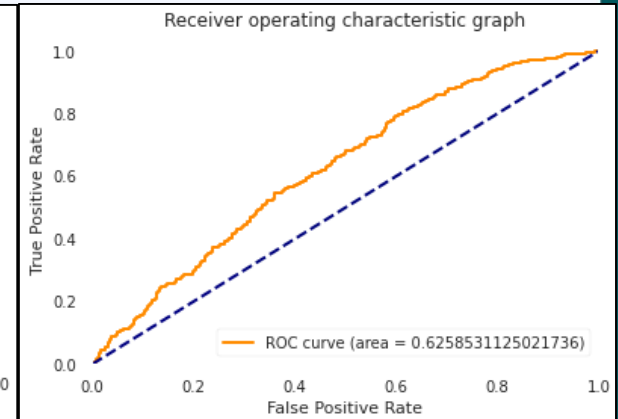
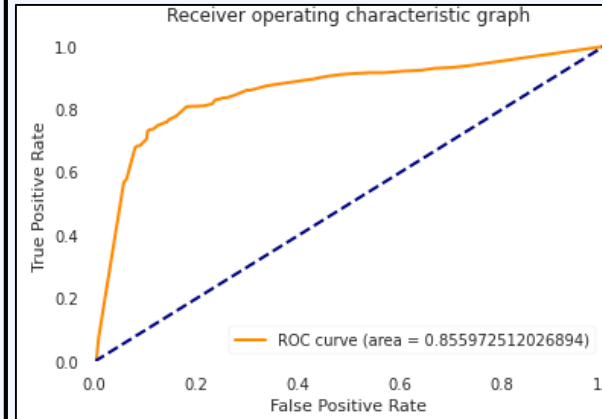
- Cut off terbaik pada data train adalah 0.5, dan cut off terbaik pada data test adalah 0.52. Terdapat trade-off antara akurasi dan recall/sensitivitas dalam menentukan cut off, namun karena perbedaan yang tidak signifikan maka cut off yang dipakai adalah cut off default (0.5)

# Modeling Result – ROC Curve

Data Train



Data Test

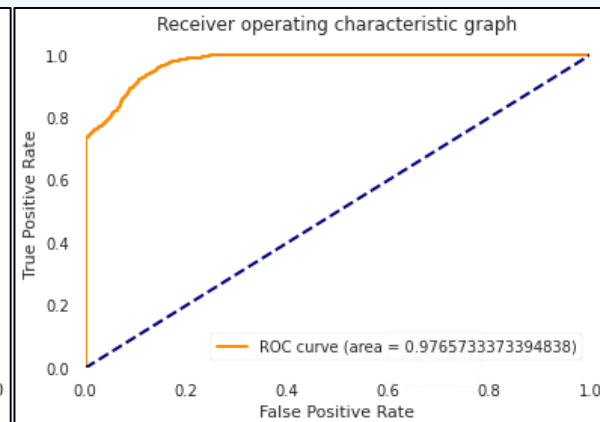
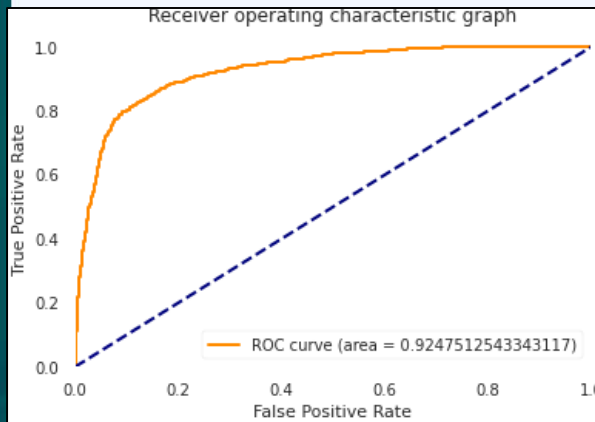


Decision Tree

Logistic Regression

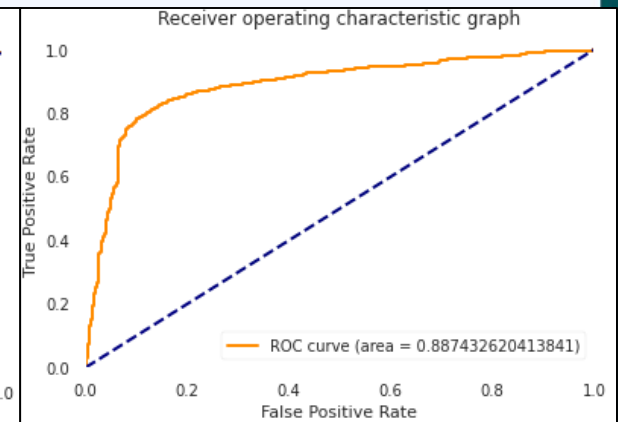
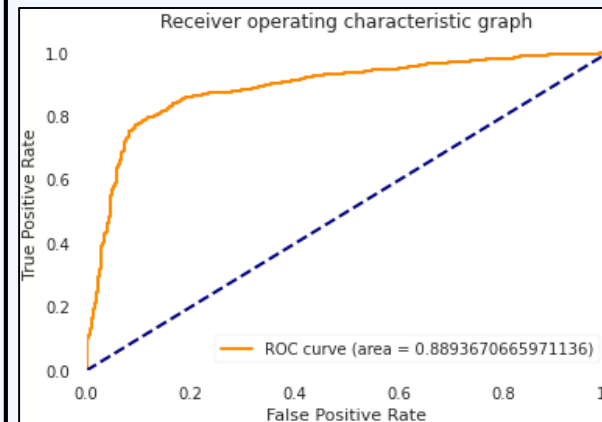
Decision Tree

Logistic Regression



XGBoost

Random Forest



XGBoost

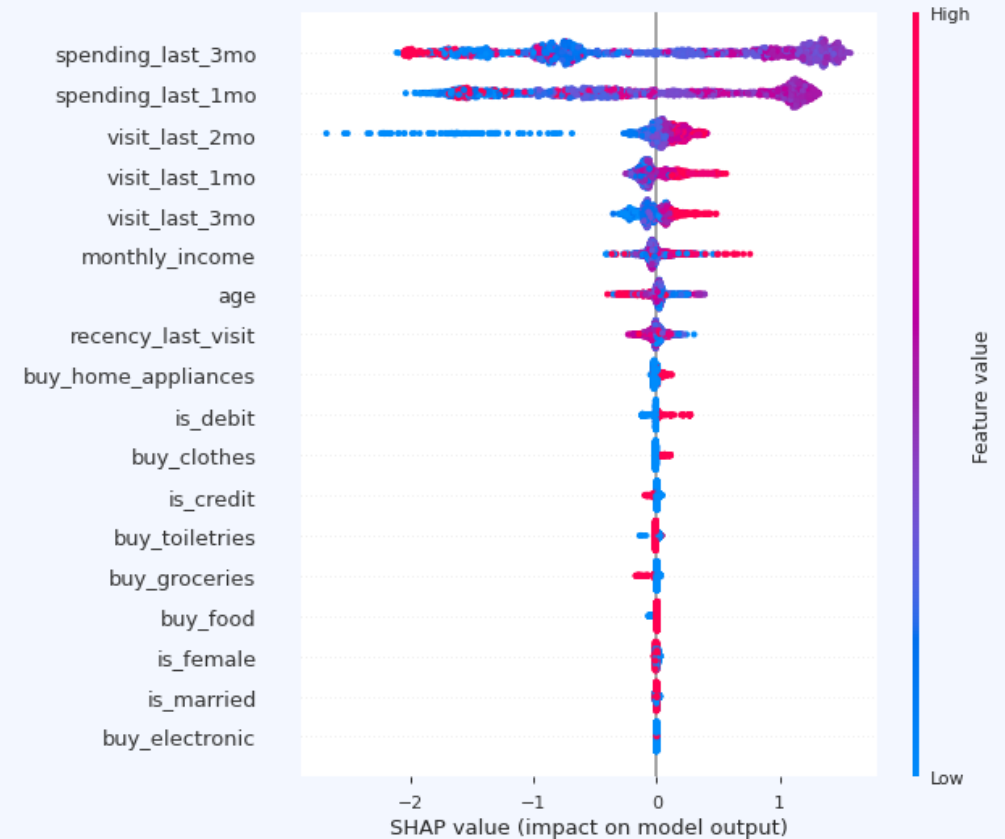
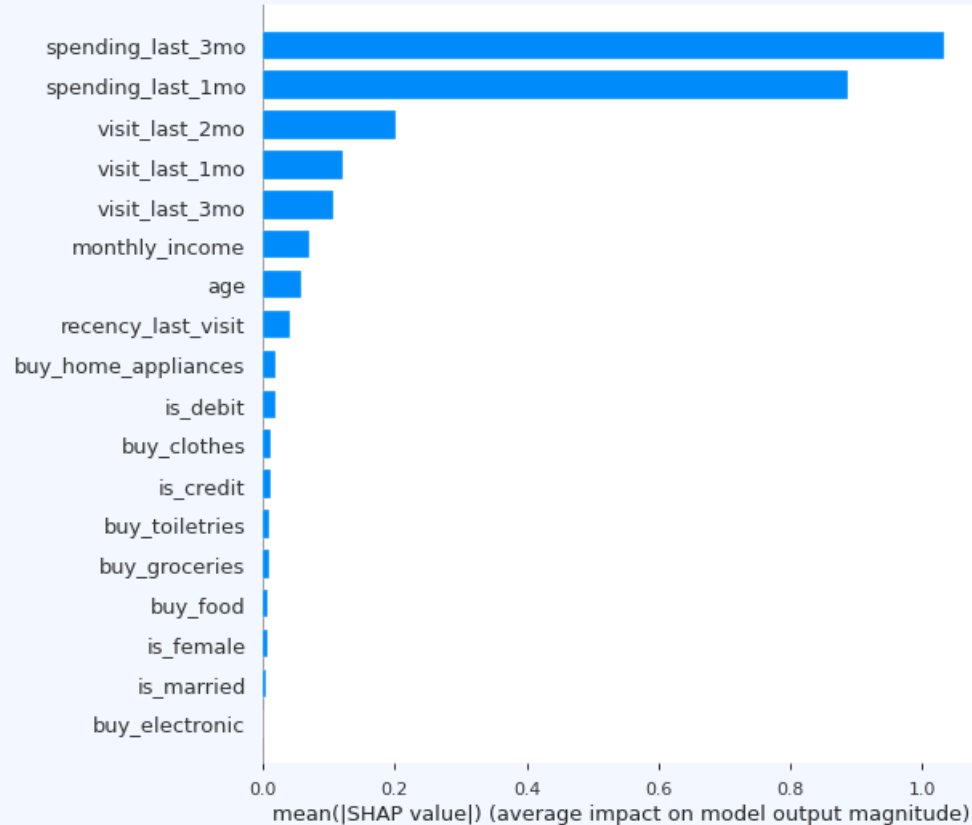
Random Forest



# Symptom Check Phase

- Global interpretation dengan mengeluarkan feature importance plot dan SHAP plot
- Local interpretation dengan mengeluarkan SHAP value untuk setiap row dan setiap variable
- Mengeluarkan 3 top symptom untuk setiap row, yaitu 3 variabel yang memiliki SHAP value tertinggi untuk setiap row
- \*catatan: SHAP value tidak mengukur kausalitas (sebab akibat).

# Symptom Check Result – Global Interpretability



## Penjelasan:

SHAP mengeluarkan general feature importance plot dan juga SHAP plot. SHAP plot lebih informatif dari feature importance plot biasa karena dia juga memberikan impact untuk kisaran nilai di setiap feature (Sebagai contoh, probabilitas untuk redemption tertinggi pada feature `spending_last_3mo` berada pada nilai tengah (berwarna ungu), sesuai dengan apa yang ditemukan pada exploratory data analysis).

# Symptom Check Result – Local Interpretability

member_id	visit_last_1mo	visit_last_2mo	visit_last_3mo	spending_last_1mo	spending_last_3mo	age	monthly_income	buy_groceries	buy_toiletries	buy_food	buy_electronic
165543	-0.107778	0.082715	0.085825	0.520151	1.102302	0.124376	-0.04585	-0.009486	0.017446	0.004448	-0.000032
605202	-0.088039	-0.001866	-0.111196	-0.746525	-1.012783	-0.00477	-0.01239	0.001774	-0.011111	0.003927	-0.000032
22272	-0.107418	-0.002228	-0.090797	-0.740171	-0.985211	-0.00847	-0.04175	0.001774	-0.011111	0.004448	-0.000032

member_id	buy_clothes	buy_home_appliances	recency_last_visit	is_married	is_debit	is_credit	is_female	BEST_VAR1	BEST_VAR2	BEST_VAR3	Prediction
165543	-0.00738	0.007547	-0.018436	0.003838	0.020307	0.007006	0.008642	spending_last_3mo	spending_last_1mo	age	1
605202	0.056064	-0.022925	-0.009945	0.000645	0.026029	0.006526	-0.00031	buy_clothes	is_debit	is_credit	0
22272	-0.00738	-0.011511	-0.021717	-0.000322	0.033129	0.006408	0.006489	is_debit	is_female	is_credit	0

## Penjelasan:

SHAP value juga dapat di generate untuk setiap amatan. Sebagai contoh diambil 3 amatan secara acak, dimana setiap amatan akan ditentukan top 3 variable dengan SHAP tertinggi. Sesuai dengan hasil pada tabel diatas, setiap row memiliki komposisi top 3 SHAP value yang berbeda-beda. Informasi ini dapat digunakan untuk menyusun strategi campaign yang lebih spesifik kedepannya.

# Kesimpulan

- Dengan algoritma XGBoost, diperoleh model terbaik untuk memprediksi pelanggan yang akan melakukan promo redemption, dengan akurasi berkisar pada angka 84%-86% dan recall pada angka 80%-82%, serta waktu yang lebih efisien.
- Identifikasi fitur yang berpengaruh dapat dilakukan dengan tiga pendekatan, yaitu melalui EDA dengan diskretisasi, SHAP plot, dan SHAP value.
- Informasi ini dapat digunakan untuk mengembangkan metode campaign berikutnya. Sebagai contoh: semua variable visit memiliki korelasi positif dengan redemption rate (baik dari hasil EDA maupun SHAP plot). Untuk itu, promo campaign dapat dikaitkan dengan jumlah visit pelanggan, seperti jumlah diskon yang semakin besar seiring dengan seringnya pelanggan datang berbelanja ke SAS-MART.
- Personalized campaign juga dapat dibangun dengan memanfaatkan SHAP Value. Sebagai contoh pelanggan dengan member ID 605202 memiliki SHAP Value tertinggi pada variable buy\_clothes. Promo personal dengan menawarkan diskon menarik untuk produk clothes bisa ditawarkan pada member tersebut.



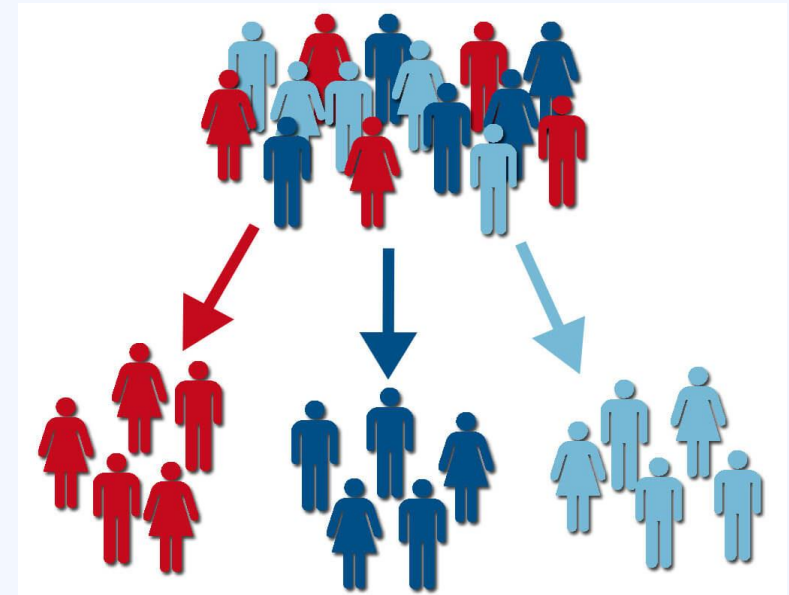
# Customer Segmentatoin



# Customer Segmentation

Customer segmentation adalah upaya membuat segmentasi customer berdasarkan value dari masing-masing customer untuk menjawab beberapa kebutuhan bisnis seperti:

- Siapa saja yang berpotensi memberikan profit lebih kepada perusahaan.
- Siapa saja yang harus dipertahankan.
- Siapa saja customer yang loyal.
- Siapa saja yang cenderung responsif apabila diberikan penawaran.
- dan lain-lain.



# Recency, Frequency, & Monetary (RFM) Analysis

RFM Analysis merupakan metode analisis yang membagi segmentasi pelanggan berdasarkan value suatu pelanggan yang diperoleh dari 3 faktor, yaitu:



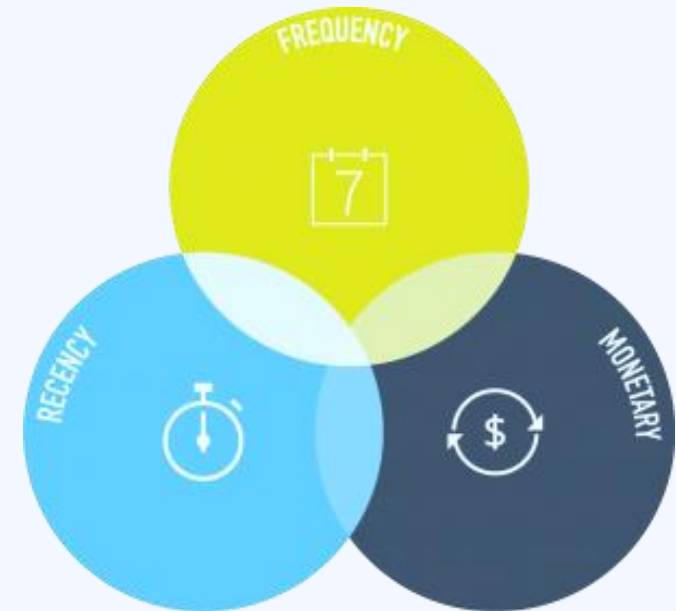
**Recency:** Seberapa lama jeda waktu sejak transaksi terakhir. Nilai Recency dihitung berdasarkan tanggal transaksi terakhir atau interval waktu transaksi terakhir dengan saat ini.



**Frequency:** Seberapa sering pelanggan melakukan transaksi. Nilai Frequency dihitung berdasarkan total keseluruhan transaksi pelanggan hingga saat ini.

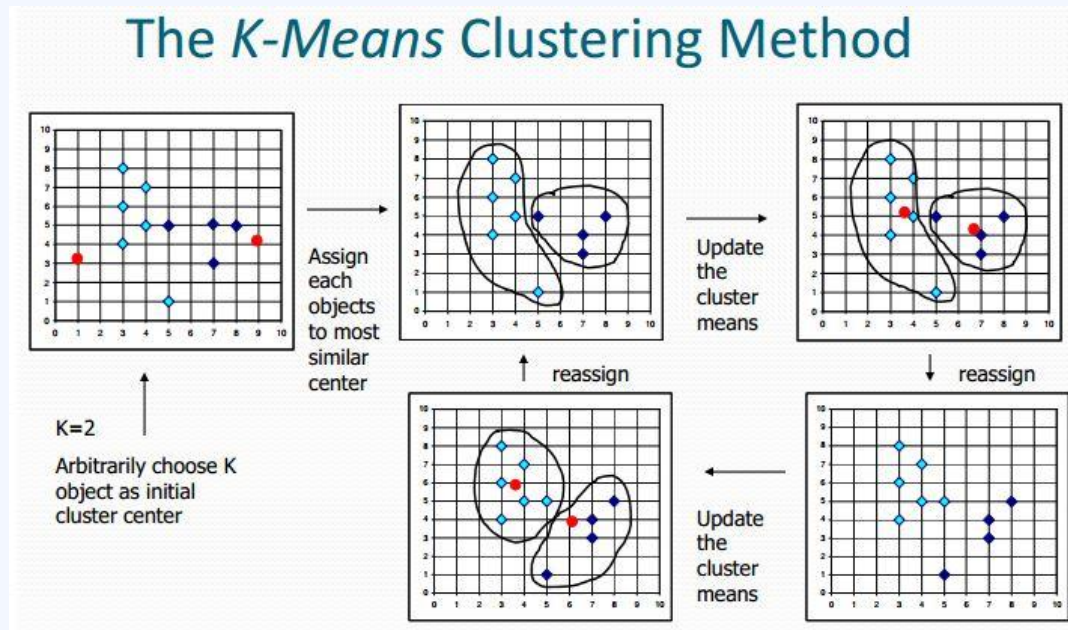


**Monetary:** Seberapa besar pengeluaran yang dihasilkan oleh pelanggan untuk meningkatkan pendapatan perusahaan. Monetary dihitung berdasarkan total pengeluaran pelanggan yang menghasilkan revenue perusahaan hingga saat ini.

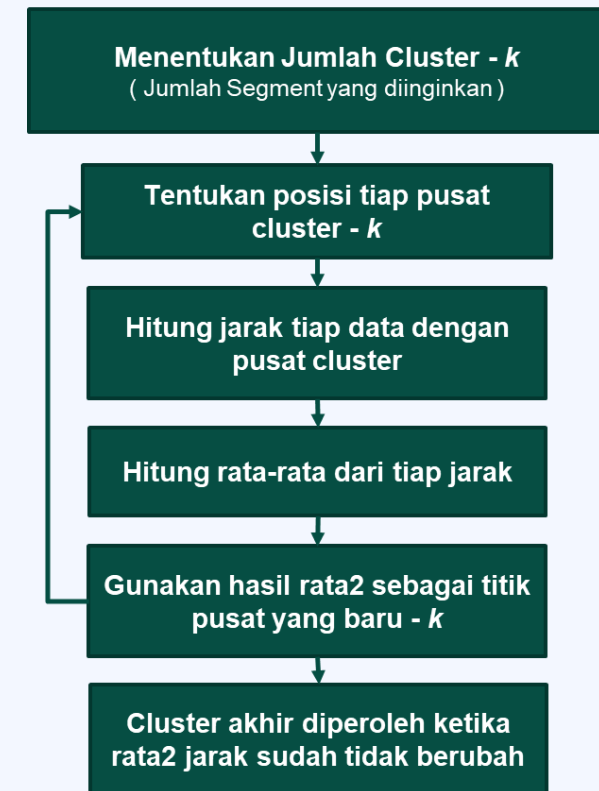


# K-Means Clustering

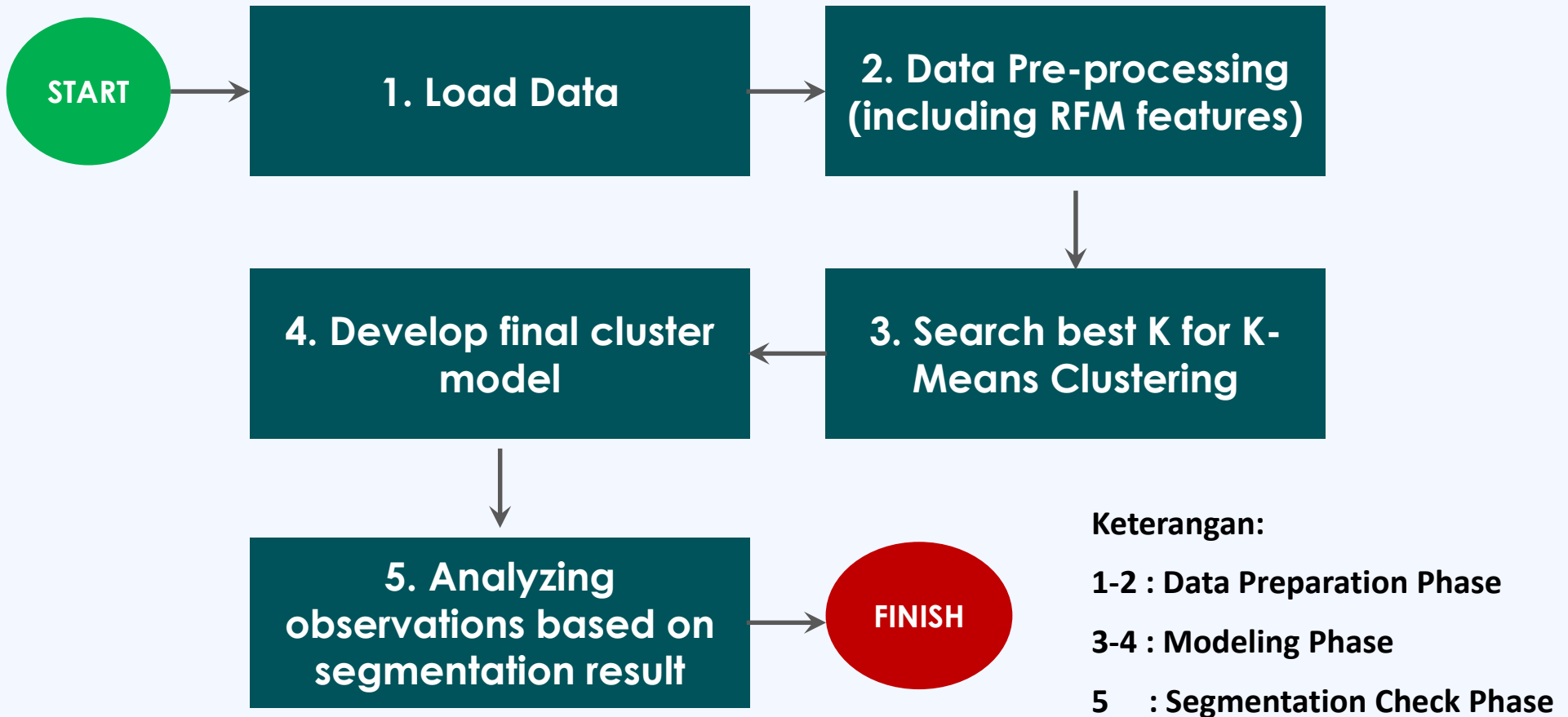
K-Means Clustering adalah metode pengelompokan suatu kumpulan data menjadi beberapa cluster berdasarkan kedekatan data dengan pusat cluster. Data dengan jarak terdekat akan dikelompokkan menjadi 1 cluster.



## K-Means Process



# Process Flow



# Data Preparation Phase

Beberapa langkah yang dilakukan pada tahap ini adalah

1. Melakukan upload data pada analytics platform yang kita gunakan,
2. Melakukan pemeriksaan Null values, jika ditemukan Null Values maka akan dilakukan imputasi menggunakan ukuran pemusatan (mean, median, atau modus).
3. Melakukan outlier handling dengan memeriksa pencilan dan menggantinya dengan nilai  $\mu+3\sigma$  (batas atas) dan  $\mu-3\sigma$  (batas bawah).
4. Menghitung Recency, Frequency, dan Monetary dengan formula masing-masing variabel adalah sebagai berikut:
  - Recency = recency\_last\_visit
  - Frequency = visit\_last\_1mo + visit\_last\_2mo + visit\_last\_3mo
  - Monetary = spending\_last\_1mo + spending\_last\_2mo + spending\_last\_3mo



# Data Preparation Result

## RFM Features

	frequency	monetary	recency
<b>count</b>	5,000.00	5,000.00	5,000.00
<b>mean</b>	10.28	1,351,763.19	16.10
<b>std</b>	5.11	771,133.28	11.51
<b>max</b>	35.00	8,917,156.00	117.00
<b>min</b>	0	0	0

## RFM Scores

	frequency	monetary	recency
<b>Largest</b>	4	4	1
<b>Second Largest</b>	3	3	2
<b>Second Lowest</b>	2	2	3
<b>Lowest</b>	1	1	4

## Penjelasan:

- Frequency adalah variabel yang kita harapkan nilainya tinggi, semakin tinggi nilainya maka akan semakin baik (berarti pelanggan sering mengunjungi SAS-MART)
- Monetary juga merupakan variabel yang kita harapkan nilainya tinggi, semakin tinggi nilainya maka akan semakin baik (berarti semakin banyak uang yang dibelanjakan pelanggan di SAS-MART)
- Recency adalah variabel yang kita harapkan nilainya semakin kecil, karena semakin besar nilainya maka akan semakin buruk (berarti semakin lama pelanggan sudah tidak mengunjungi SAS-MART)
- Selanjutnya akan dilakukan scoring RFM seperti pada tabel disamping, dan segmen dengan skor tinggi dinyatakan sebagai segmen dengan prospek yang tinggi dan dapat menjadi target dari marketing campaign berikutnya (rentang score: 3-12).

9.75-12 : High

5.25-7.5 : Medium Low

7.5-9.75 : Medium High

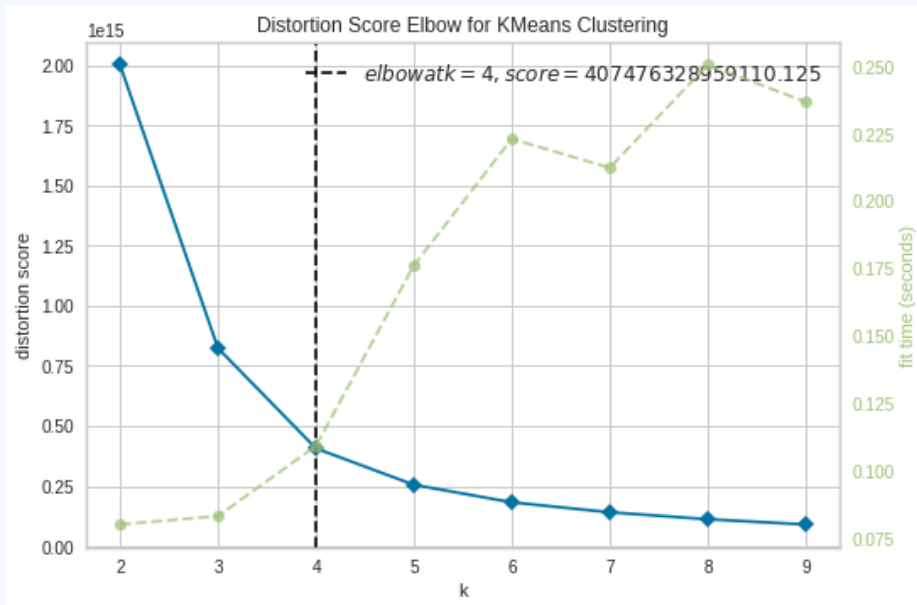
3-5.25 : Low

# Clustering Phase

- Penentuan jumlah cluster terbaik ( $k$ ) akan dilakukan dengan dua metode:
  - Elbow method, dengan jumlah  $k$  terbaik berada di titik belok di kurva elbow method (berdasarkan penghitungan Within-Cluster-Sum-of-Squares (WCSS)).
  - Silhoutte analysis, dengan menghitung koefisien di setiap  $k$  yang dicobakan. Semakin mendekati 1 maka akan semakin baik  $k$  tersebut.
- Final cluster menggunakan  $k$  terbaik, dimana semua row akan mendapatkan nilai kelompok.
- Model clustering juga bisa disimpan untuk memberikan kelompok pada row (pelanggan) baru.

# Clustering Result

## Elbow Method

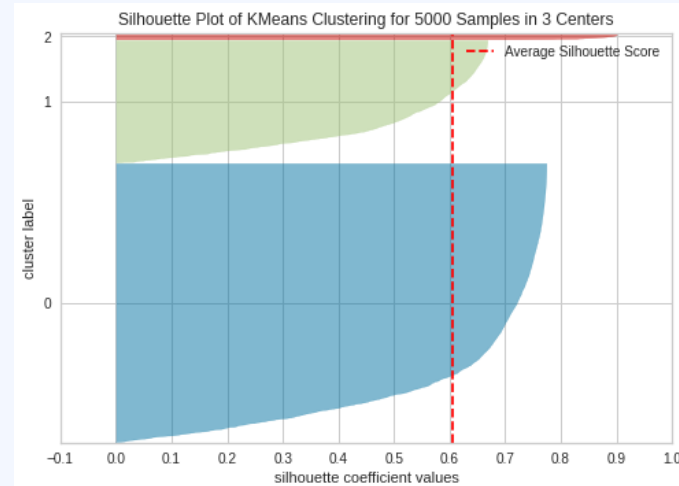


## Penjelasan:

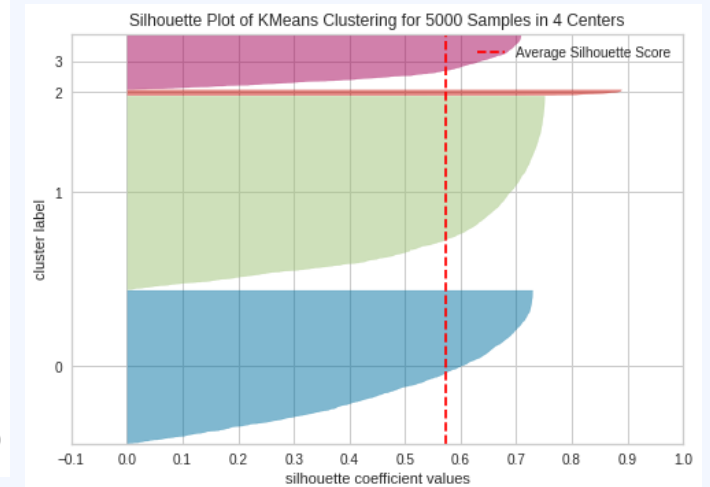
- Hasil dari elbow method merekomendasikan  $k=4$ . Kemudian akan dikonfirmasi melalui silhouette analysis dengan menggunakan  $k$  terdekat lainnya ( $k=3,4,5$ )
- Silhouette analysis menghasilkan koefisien terbaik pada  $k=3$  yaitu 0.6. namun tidak jauh berbeda dengan koefisien di  $k$  lainnya (masih dalam selang 0.55-0.6). Dengan menggabungkan kedua hasil analisis maka ditetapkan jumlah cluster terbaik adalah  $k=4$ .

## Silhouette Analysis

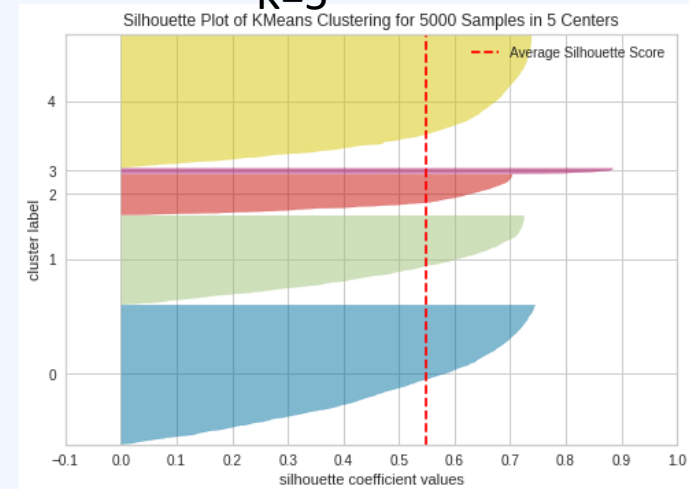
K=3



K=4



K=5

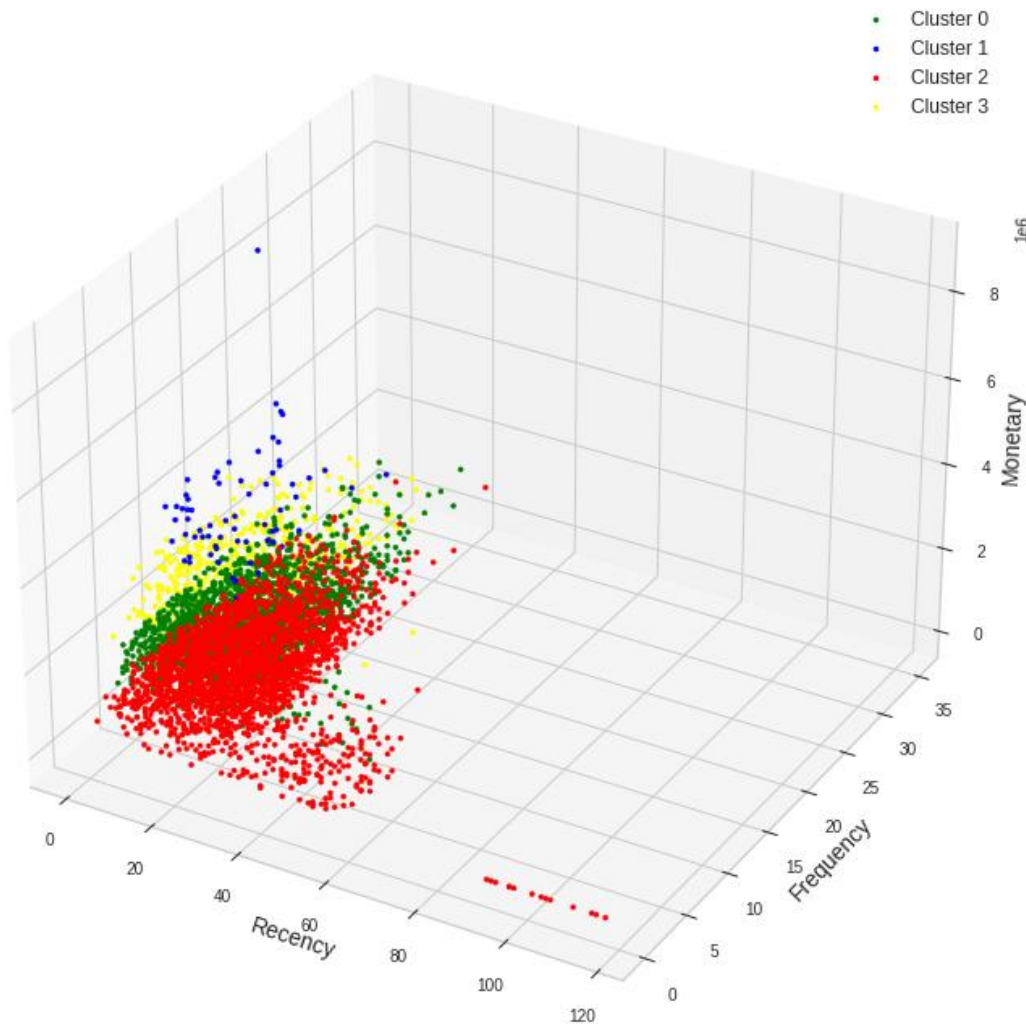


# Segmentation Check Phase

- Membuat plot 3D dari hasil clustering
- Menghitung skor setiap segmen
- Menentukan prospek segmen beserta rekomendasi berdasarkan hasil clustering.

# Segmentation Check Result

3D Clustering Plot



RFM Average for each segment

cluster	count	frequency	monetary	recency	response rate
0	1,887	10.76	1,545,684.07	15.08	0.790
1	66	10.52	4,315,036.21	14.20	0.015
2	2,384	9.70	760,899.88	17.32	0.250
3	663	10.99	2,629,460.59	14.83	0.109

Segement Score

cluster	frequency	monetary	recency	score	prospect
0	3	3	3	9	medium-high
1	2	4	4	10	high
2	1	1	1	3	low
3	4	3	2	9	medium-high

## Penjelasan:

- Plot 3D menunjukkan sebaran setiap segmen di ruang 3 dimensi dengan sumbu berupa unsur-unsur dari RFM.
- Segmen terbaik (sekalius paling minoritas) adalah cluster nomor 1 dengan total skor 10 (terbaik dari segi monetary dan recency)
- Segmen terburuk (sekalius mayoritas) adalah cluster nomor 2, dengan total skor 3 (terburuk disemua sisi RFM)



# Kesimpulan

- Metode K-Means Clustering dengan memanfaatkan RFM features menghasilkan cluster dengan jumlah cluster terbaik yaitu  $k=4$ .
- Statistik dari masing-masing segmen/cluster menunjukkan bahwa segmen terbaik (sekalius paling minoritas) adalah cluster nomor 1 (terbaik dari segi monetary dan recency), sedangkan segmen terburuk (sekalius mayoritas) adalah cluster nomor 2 (terburuk disemua sisi RFM).
- Hasil segmentasi tersebut dapat digunakan untuk menetapkan skala prioritas dalam campaign selanjutnya. Segmen high prospect perlu mendapat perhatian khusus karena berpotensi meningkatkan revenue SAS-MART, disusul segmen medium-high. Untuk segmen dengan prospek low, campaign tetap bisa dilaksanakan namun tetap memperhatikan efektivitas dan efisiensi (salah satunya dengan melakukan prediksi melalui promo redemption model).





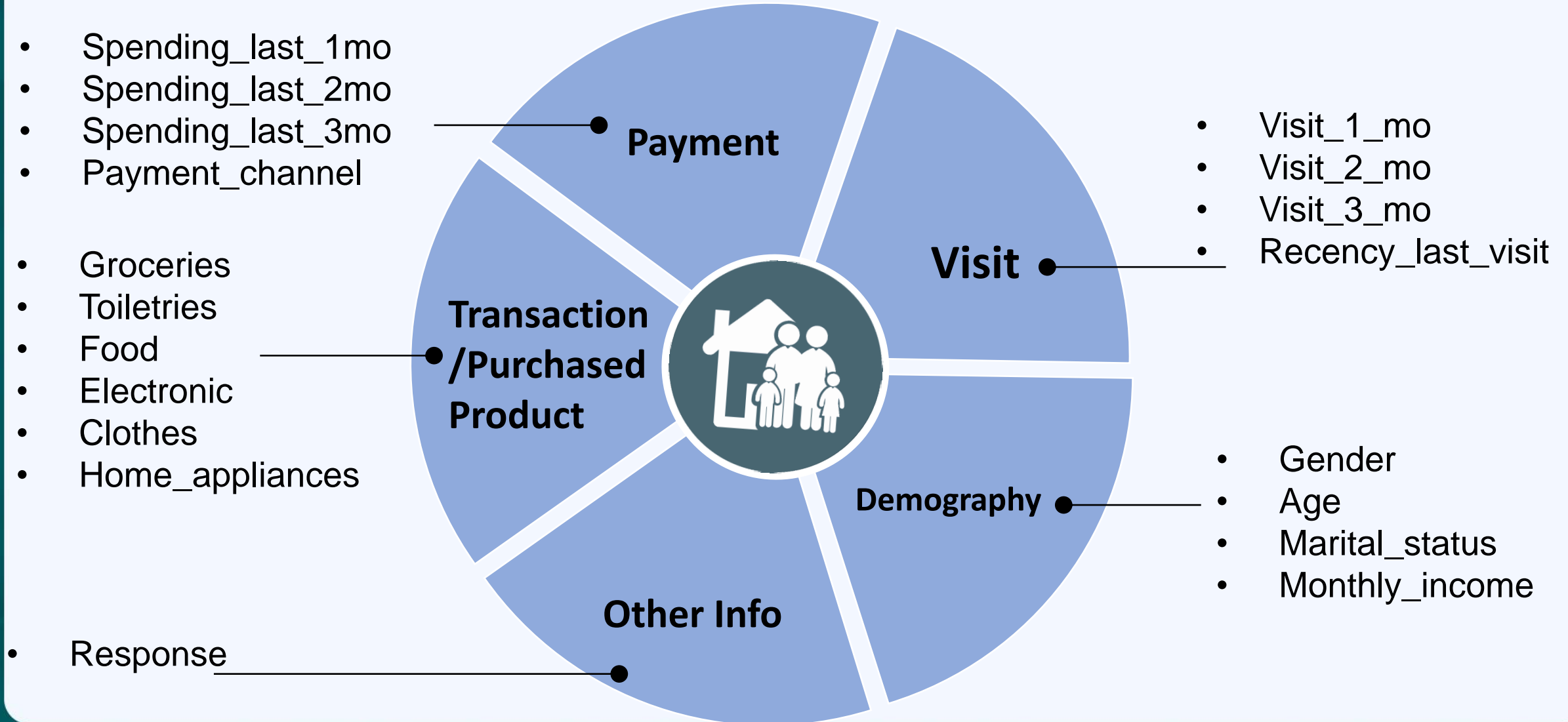
# Customer Profiling Dashboard

# Customer Profiling

- Customer profiling adalah suatu cara untuk membuat *portrait* pelanggan untuk membantu perusahaan dalam membuat keputusan terkait proses penyampaian layanan (service delivery) - *Experience UX*
- Teknologi Customer profiling mendorong perusahaan untuk memahami customer lebih dalam sehingga mampu membuat program marketing yang lebih efektif



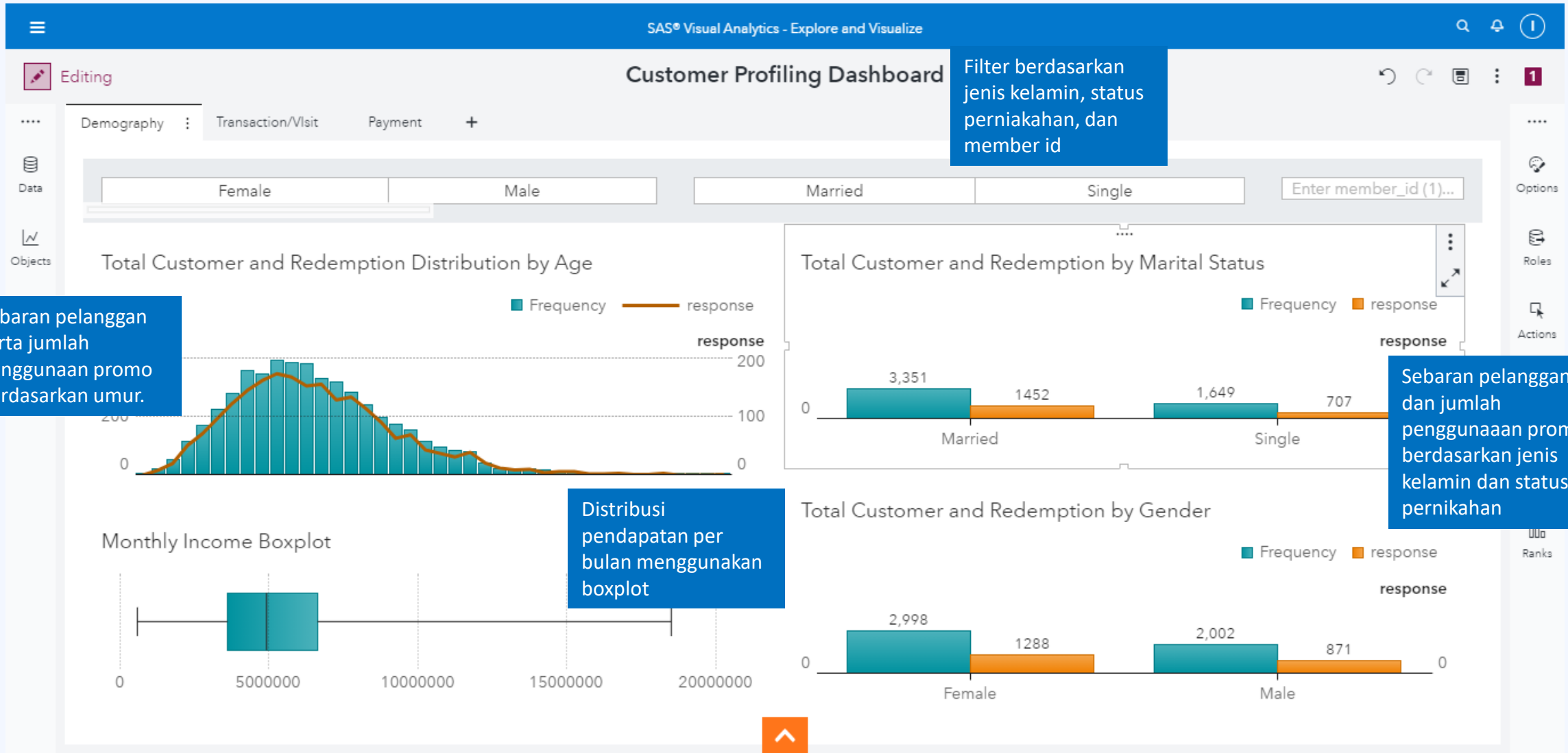
# Utilized Data



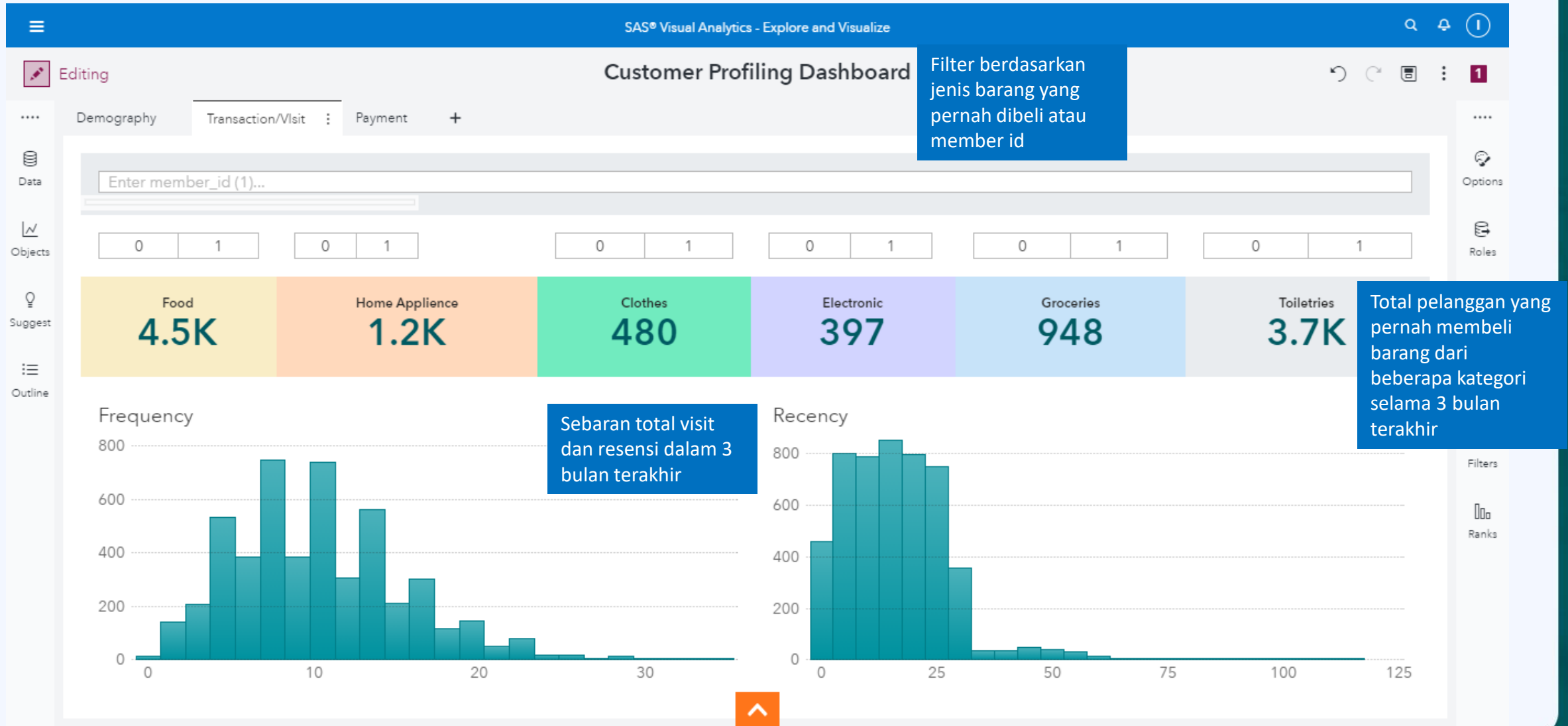
# Create Customer Profiling Dashboard

- Menggunakan SAS Visual Analytics
- Menampilkan informasi baik secara summary (keseluruhan) maupun granular (per Member ID)
- Dibagi menjadi 3 page utama: Demography, Transaction/Visit, dan Payment
- Page Demography menjelaskan profil pelanggan dari sisi gender, age, marital status, dan monthly income
- Page Transaction/Visit menampilkan frequency (total visit), recency, serta item-item yang pernah dibeli pelanggan (food, home appliance, electronics, toiletries, clothes, groceries)
- Page Payment menampilkan payment method, monetary (total spending), dan monthly income

# Demography

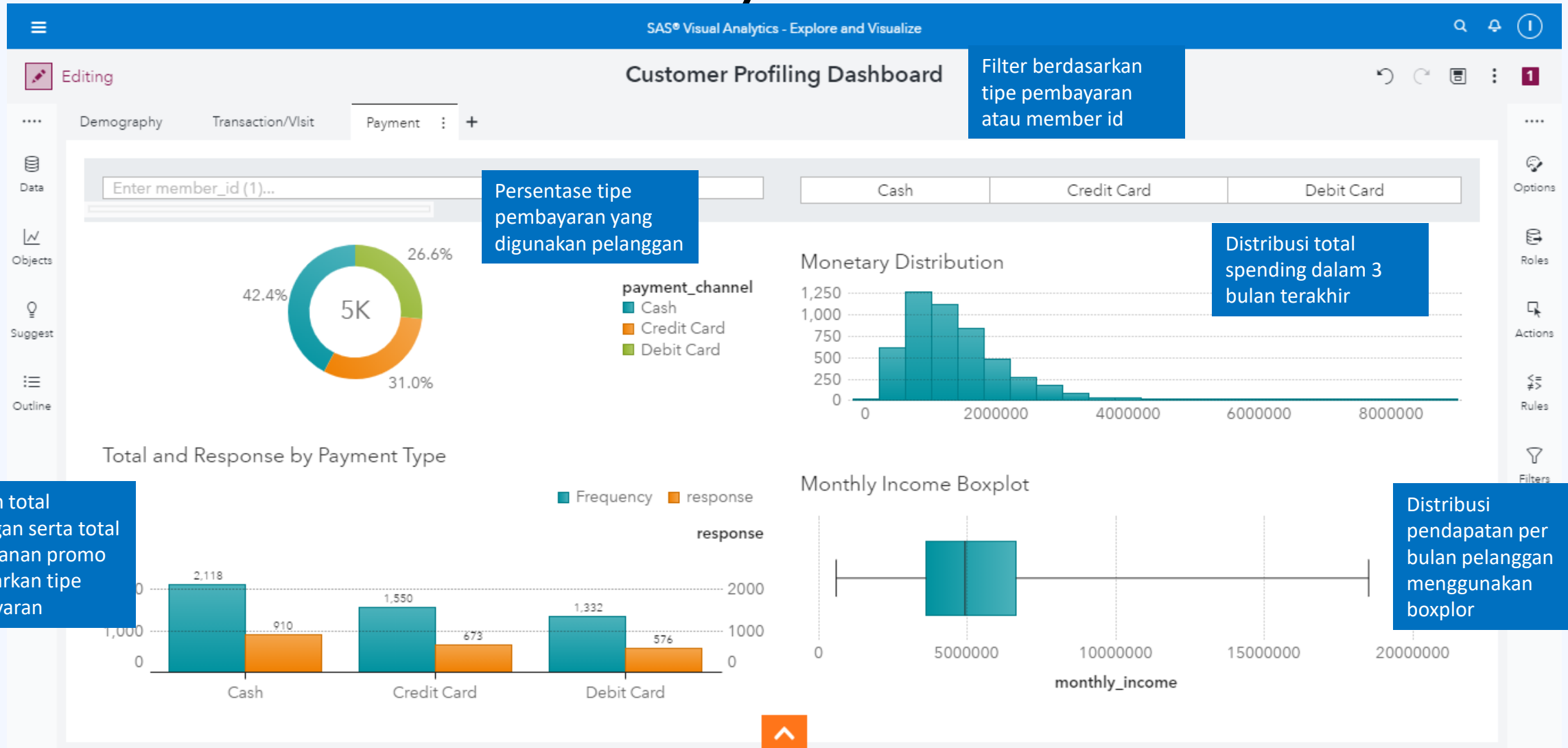


# Transaction/Visit





# Payment



# Kesimpulan

- Dashboard demography dapat digunakan untuk eskplorasi attribut personal dari pelanggan (jenis kelamin, umur, status perkawinan, dll.).
- Dahboard transaction/visit dapat digunakan untuk eksplorasi behaviour pelanggan (jumlah kedatangan dan barang apa saja yang dibeli).
- Dashboard payment dapat digunakan untuk ekplorasi jumlah uang yang ditransaksikan pelanggan serta bagaimana mereka membayarnya.
- Semua dashboard tersebut dapat dieksplor lebih lanjut untuk mencari temuan-temuan menarik baik secara summary maupun granular (per individu). Dari temuan-temuan tersebut dapat disusun campaign yang lebih menarik dan lebih efektif.
- Contoh temuan menarik akan dijelaskan lebih lanjut di video.



# Final Conclusion

# Final Conclusion

1. Promo redemption model dengan SHAP yang saya buat mampu memprediksi pelanggan yang akan menggunakan promo dengan akurasi sebesar 84%-86%. Model dapat digunakan untuk memprediksi apakah pelanggan akan menggunakan promo, total pelanggan yang menggunakan promo, serta faktor-faktor yang berpengaruh dalam keputusan menggunakan promo tersebut baik secara global maupun local.
2. Customer segmentation dengan RFM berhasil menemukan segmen dengan prospek tertinggi berdasarkan nilai RFM, dua segmen medium-high, serta segmen low. Segmen high yang minoritas perlu mendapat perhatian khusus karena berpotensi tinggi meningkatkan revenue perusahaan (berdasarkan Analisa RFM), kemudian diikuti segmen medium-high. Segmen low juga tetap mendapatkan campaign namun perlu memperhatikan efektivitas dan efisiensi.
3. Customer profiling dashboard terdiri dari tiga menu utama: Demography, Transaction/Visit, dan Payment. Dengan dashboard tersebut dapat dilakukan eksplorasi lebih lanjut terhadap data pelanggan baik secara summary maupun granular. Dari insight menarik di dalamnya dapat disusun campaign yang lebih menarik dan lebih efektif.





# Terimakasih!