

Final Project Data Science

Churn Analysis

By Tarfa

Bootcamp DiBimbing DS29

Hello,

I'm Tarfa Gaida H, *A Data Enthusiast*

mechanical Engineering graduated specializing in Aeronautical Engineering with over one year of professional experience as a Project Manager.

Currently transitioning into the field of data science with a focus on data analyst, machine learning, and data visualization. Dedicated to continuous learning and growth, bringing enthusiasm, adaptability, and technical expertise to every opportunity.

Contact:



[LinkedIn](#)



[Gmail](#)



[GitHub](#)

Education

2016-2019

BANDUNG STATE POLYTECHNIC

Associate Degree of Engineering: Aeronautics engineering

2021-2023

BANDUNG NURTANIO UNIVERSITY

Bachelor of Engineering: Aviation engineering

October 2024- Present

Dibimbing.id

Bootcamp Data Science

- Related Coursework: Python, SQL, Data Analysis and Visualization, Data Cleaning, EDA, Probability and Statistics

Experience

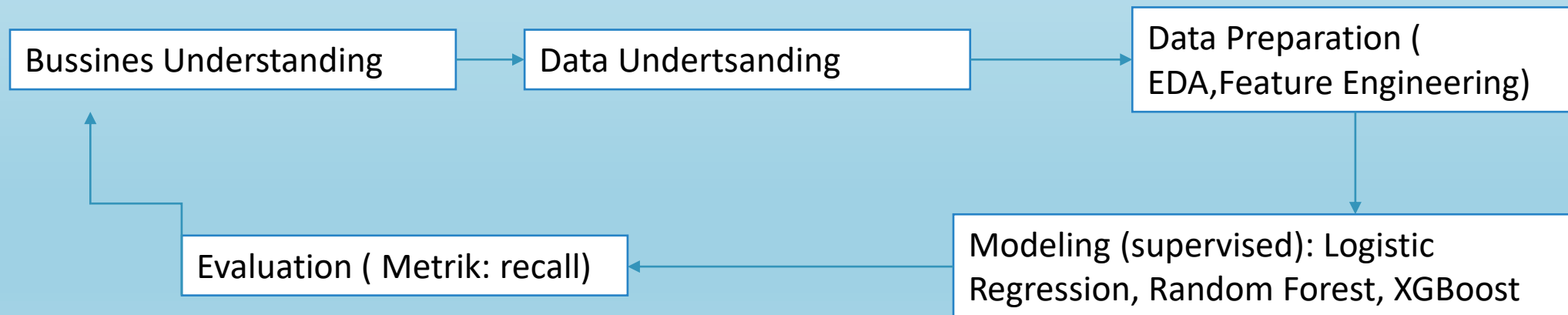
Project Manager

at PT Andala Valvo Flokontrol

Introduction

Dalam industri ritel dan e-commerce, memahami perilaku pelanggan adalah salah satu kunci untuk meningkatkan penjualan dan mempertahankan loyalitas pelanggan. Dengan data yang semakin melimpah, analisis mendalam terhadap pola pembelian menjadi semakin penting untuk memberikan wawasan strategis.

Churn Analysis adalah proses menganalisis data pelanggan untuk siapa yang berhenti menggunakan layanan (churn), kenapa mereka berhenti, dan bagaimana cara mencegahnya di masa depan.



Business Understanding

- **Latar Belakang**
 - Angka pelanggan churn yang cukup tinggi
 - mengidentifikasi pelanggan berisiko churn
- **Tujuan proyek**
 - Membangun model prediksi churn(Prediksi churn pelanggan)
 - Mengetahui fitur yang berpengaruh
 - Memberikan insight actionable

Bussiness Questions

1. apakah pelanggan dengan satification score rendah berpotensi churn?
2. apakah pelanggan dengan total spending yang rendah berpotensi churn?
3. apakah pelanggan dengan numer of return tinggi berpotensi churn?
4. bagaimana pengaruh respon promosi pelanggan terhadap churn rate?
5. gender mana yang memiliki rata-rata total spending tinggi, dan bagaimana churn rate nya?

Data Understanding

- Dataset: [Online Retail Customer Churn Dataset](#)

Dataset yang digunakan dalam proyek ini memberikan gambaran tentang perilaku pembelian pelanggan dan wawasan penjualan, yang bertujuan untuk memprediksi churn pelanggan

Struktur Dataset

Jumlah baris: 1000

Jumlah kolom: 15

Tidak ada nilai duplikat (cek `df.duplicated().sum()`)

Tidak ada missing value

Tipe Data:

Numerik: Age, Annual_Income, Total_Spend, Num_of_Purchases, etc.

Kategorikal: Gender, Email_Opt_In, Promotion_Response, Target_Churn

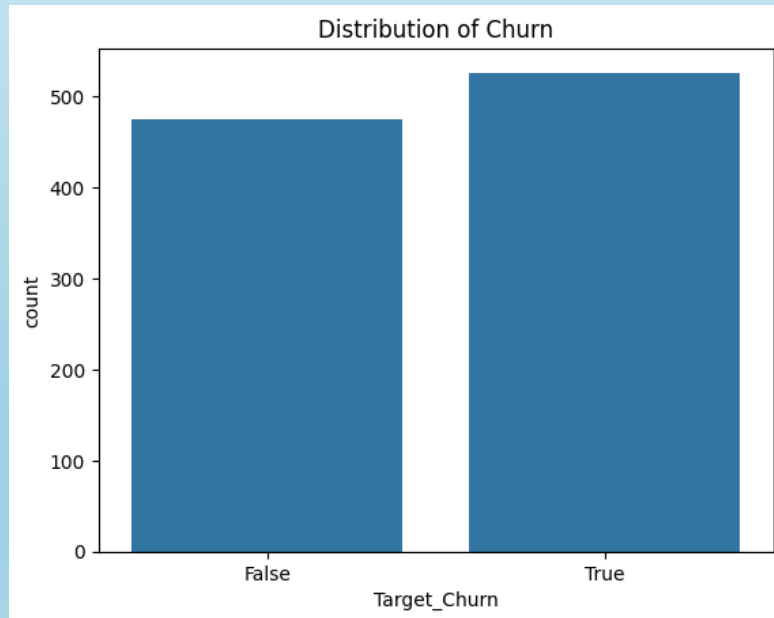
Deskripsi Kolom

- Customer_ID: A unique identifier for each customer.
- Age: The customer's age.
- Gender: The customer's gender (Male, Female, Other).
- Annual_Income: The annual income of the customer in thousands of dollars.
- Total_Spend: The total amount spent by the customer in the last year.
- Years_as_Customer: The number of years the individual has been a customer of the store.
- Num_of_Purchases: The number of purchases the customer made in the last year.
- Average_Transaction_Amount: The average amount spent per transaction.
- Num_of>Returns: The number of items the customer returned in the last year.
- Num_of_Support_Contacts: The number of times the customer contacted support in the last year.
- Satisfaction_Score: A score from 1 to 5 indicating the customer's satisfaction with the store.
- Last_Purchase_Days_Ago: The number of days since the customer's last purchase.
- Email_Opt_In: Whether the customer has opted in to receive marketing emails.
- Promotion_Response: The customer's response to the last promotional campaign (Responded, Ignored, Unsubscribed).
- Target_Churn: Indicates whether the customer churned (True or False).

Data Understanding

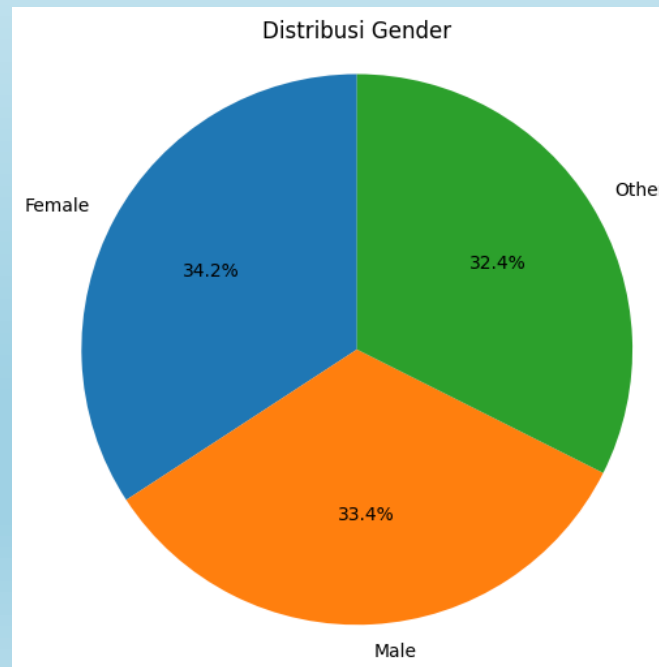
Distribusi churn rate

- True(Churn):526 pelanggan
- False (no Churn): 474 pelanggan



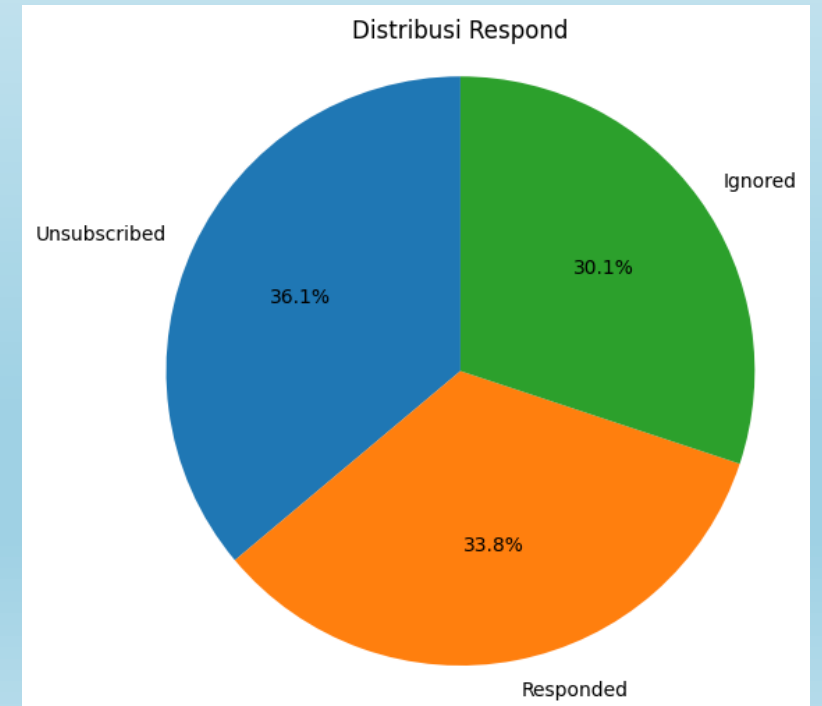
Distribusi Gender

- Female: 342
- Male : 334
- Other : 324



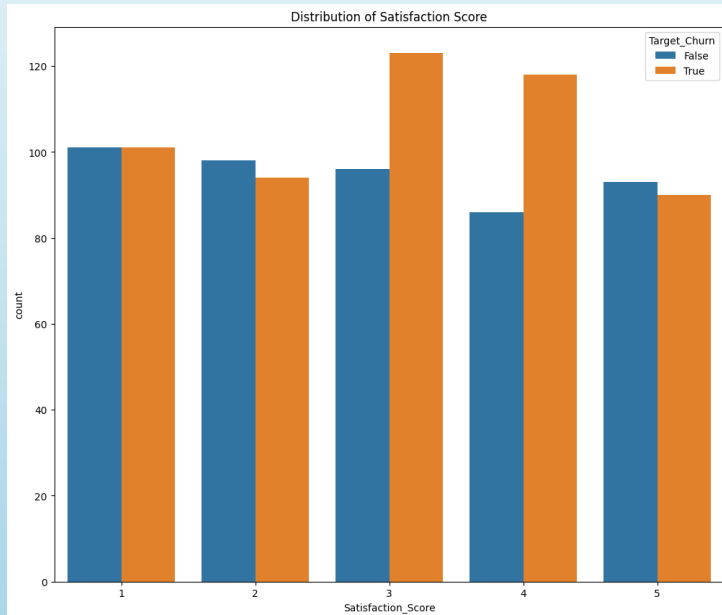
Distribusi respon promosi

- Unsubscribed : 361
- Responded : 338
- Ignored : 301



Statisfication Score terhadap Churn rate

apakah pelanggan dengan satifcation score rendah berpotensi churn?

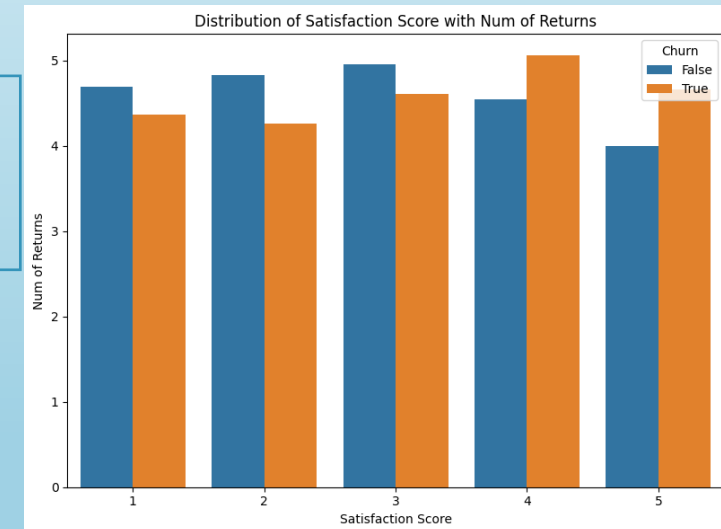


Churn tinggi pada score
3 - 4



Produk atau pelayanan ?

Jumlah pengembalian
barang, di score 3-4 cukup
tinggi



Insight : churn rate tinggi pada score 3-4 , berhubungan dengan jumlah return barang pada score kepuasan tersebut. Hal ini dapat diindikasikan bahwa ada permasalahan pada kualitas barang atau bisa jadi barang yang dikirimkan salah.

Rekomendasi: Evaluasi kualitas produk yang dijual dan proses penanganan pesanan.

Number Of Return terhadap Churn Rate

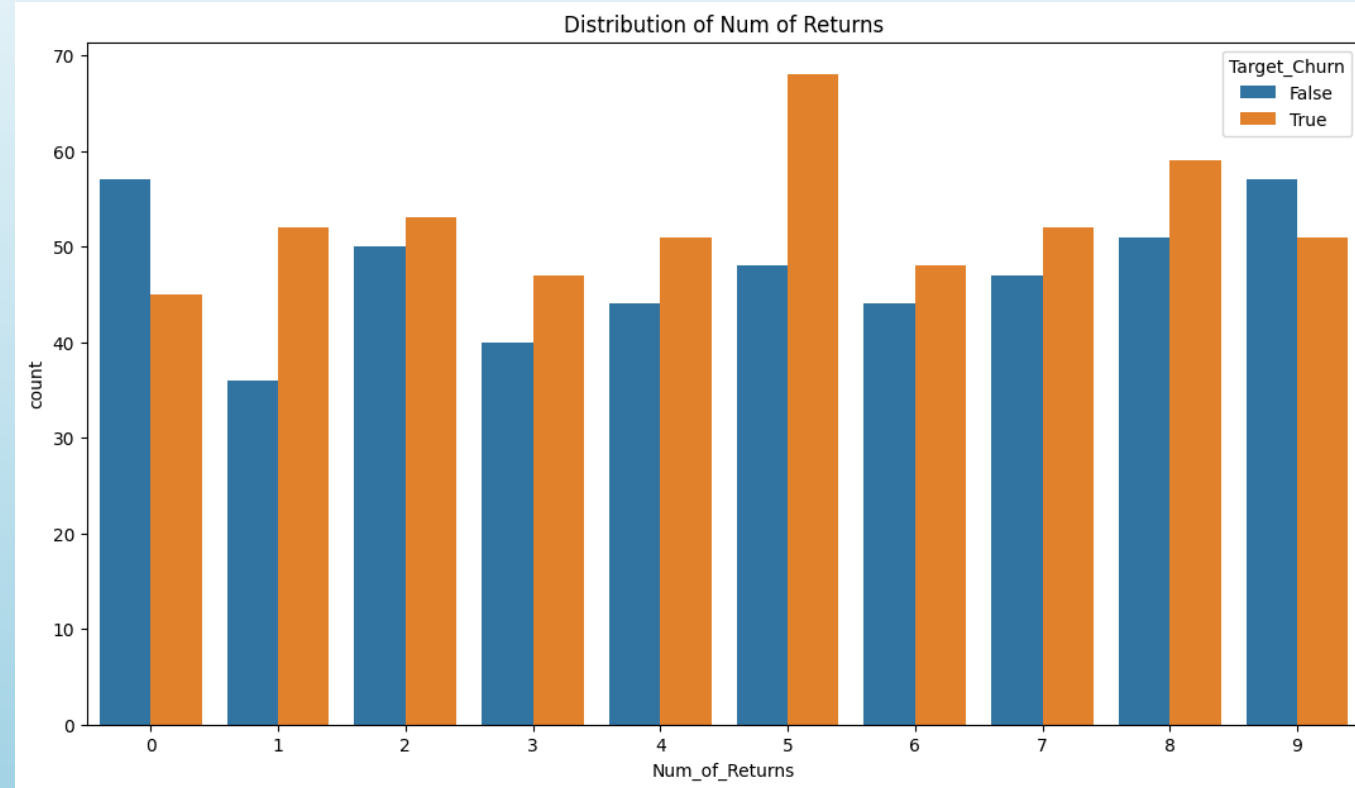
Apakah pelanggan dengan number of return tinggi berpotensi churn?

Return 0: pelanggan bertahan lebih tinggi

Return 1--4: pelanggan churn mulai lebih tinggi

Return 5: pelanggan churn paling tinggi

Return 6--9: pelanggan bertahan mulai mengimbangi tetapi pelanggan churn masih tinggi



Insight : Bisa jadi karena pengalaman yang buruk selama proses pengembalian atau kualitas barang yang tidak konsisten, menjadi penyebab churn yang tinggi.

Recomendasi: Evaluasi kualitas produk yang di jual dan proses penanganan pesanan.

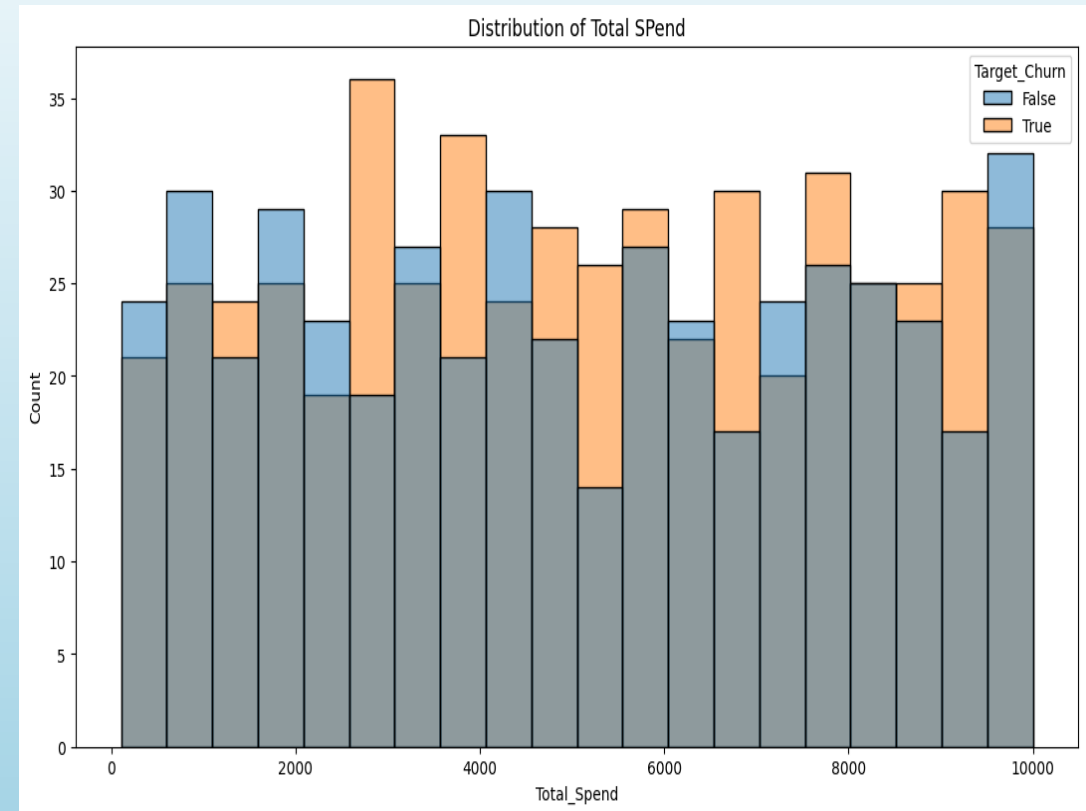
Total Spend terhadap Churn Rate

apakah pelanggan dengan total spend yang rendah berpotensi churn?

0 – 2000 : Churn rendah (awal transaksi)

2000 – 7000 : Churn tinggi, mulai menurun saat spend bertambah

7000 – 10000: churn lebih rendah, banyak pelanggan yang bertahan



Insight : Chart dapat diartikan bahwa semakin tinggi Total_Spend, kemungkinan churn lebih rendah.

Recomendasi: memberikan penawaran khusus, seperti harga discount atau penawaran voucher tertentu untuk pelanggan yang spendingnya masih rendah, dan pelanggan yang masih baru untuk meningkatkan transaksi. Dan untuk mempertahankan loyalitas pelanggan dengan spending tinggi, bisa diberikan reward berupa voucher dengan keuntungan khusus bagi top spender

Respon Promosi terhadap Churn Rate

Bagaimana pengaruh respon promosi pelanggan terhadap churn rate?

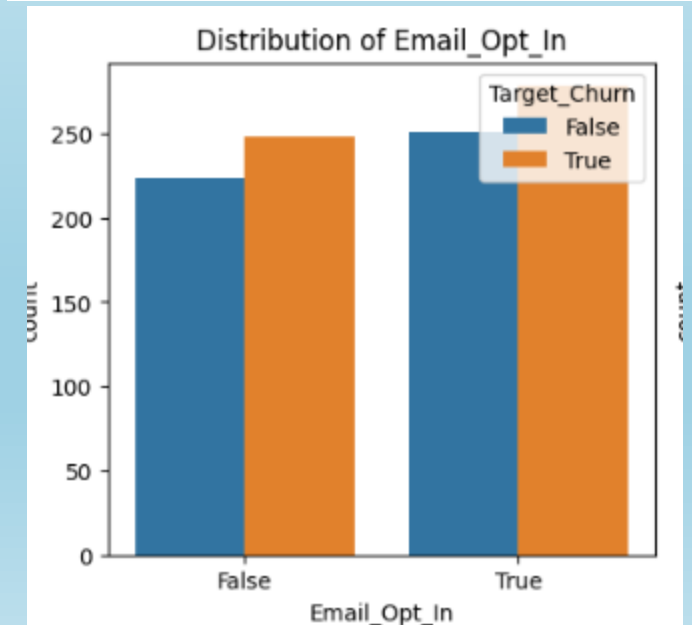
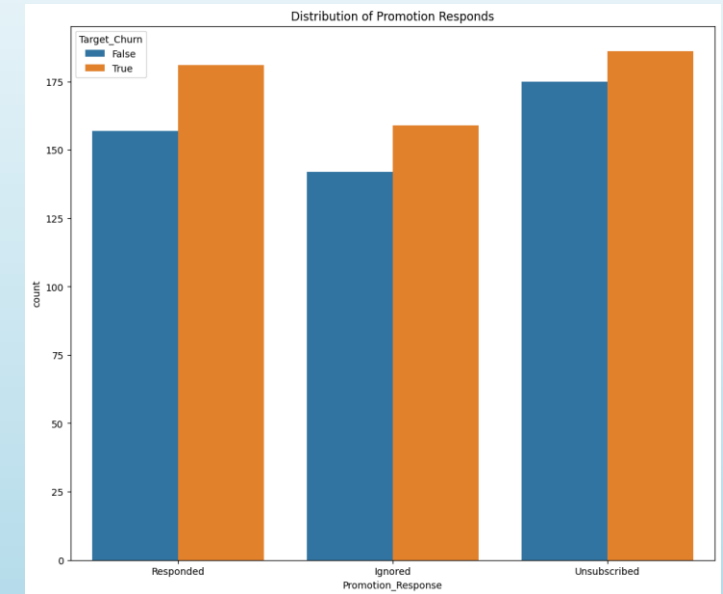
Responded : Churn tinggi dari pada yang bertahan

Ignored : Churn paling rendah, lebih tinggi dari no churn gap lebih kecil dari responded

Unsubscribed : Churn paling tinggi, jumlah keseluruhan terbanyak

Insight : Pelanggan yang tidak menerima promosi cenderung churn. Plot data pelanggan yang merespon promosi lebih tinggi churn rate nya dibandingkan yang mengabaikan. Promosi tidak efektif : terlalu generik , kurang menarik

Rekomendasi : perlu ada strategi agar pelanggan menerima promosi(subscribe) , contoh penyebaran informasi promosi penawaran khusus yang hanya bisa dilihat jika men-subscribe promosi lewat media yang tersedia , seperti email dll.
improvement promosi agar lebih menarik lagi. Bisa mengadakan penawaran khusus atau potongan harga,



Rata-Rata total Spending dan churn rate terhadap Gender

Bagaimana rata-rata total spending tinggi berdasarkan gender, dan bagaimana churnrate nya?

Pelanggan perempuan, spending tertinggi tetapi juga churn tertinggi

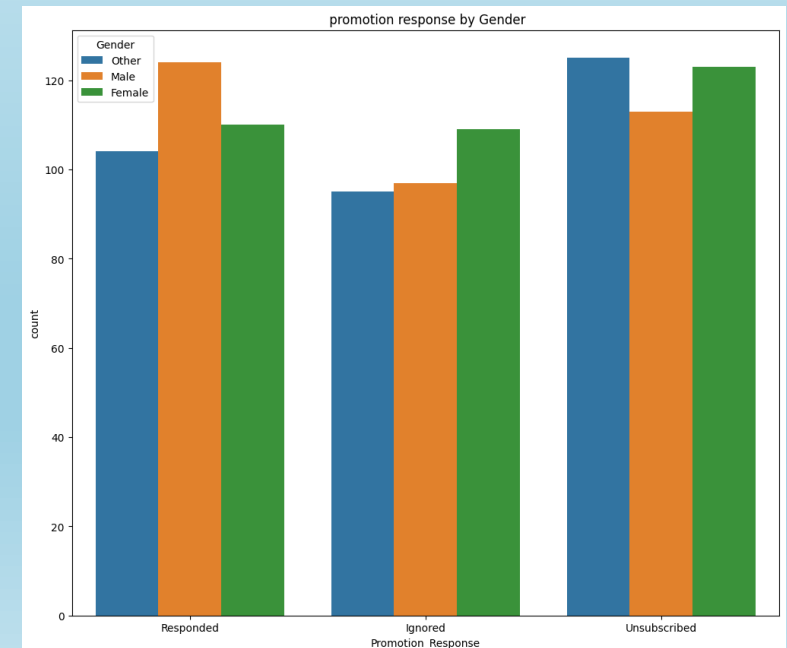
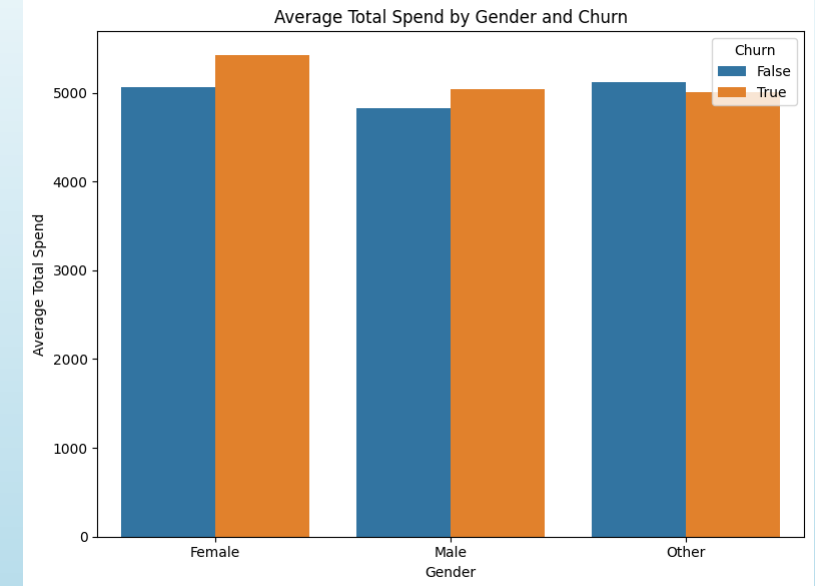


Tidak ada reward atau promosi kurang menarik ?

Respon promosi pelanggan perempuan tinggi, kategori respon termasuk yang churn tinggi

Insight : Ada kemungkinan churn tinggi pada pelanggan perempuan disebabkan oleh promosi yang tidak menarik atau bisa jadi sering kali tidak sesuai dengan kebutuhan mereka. Kemudian tidak ada reward untuk transaksi tinggi

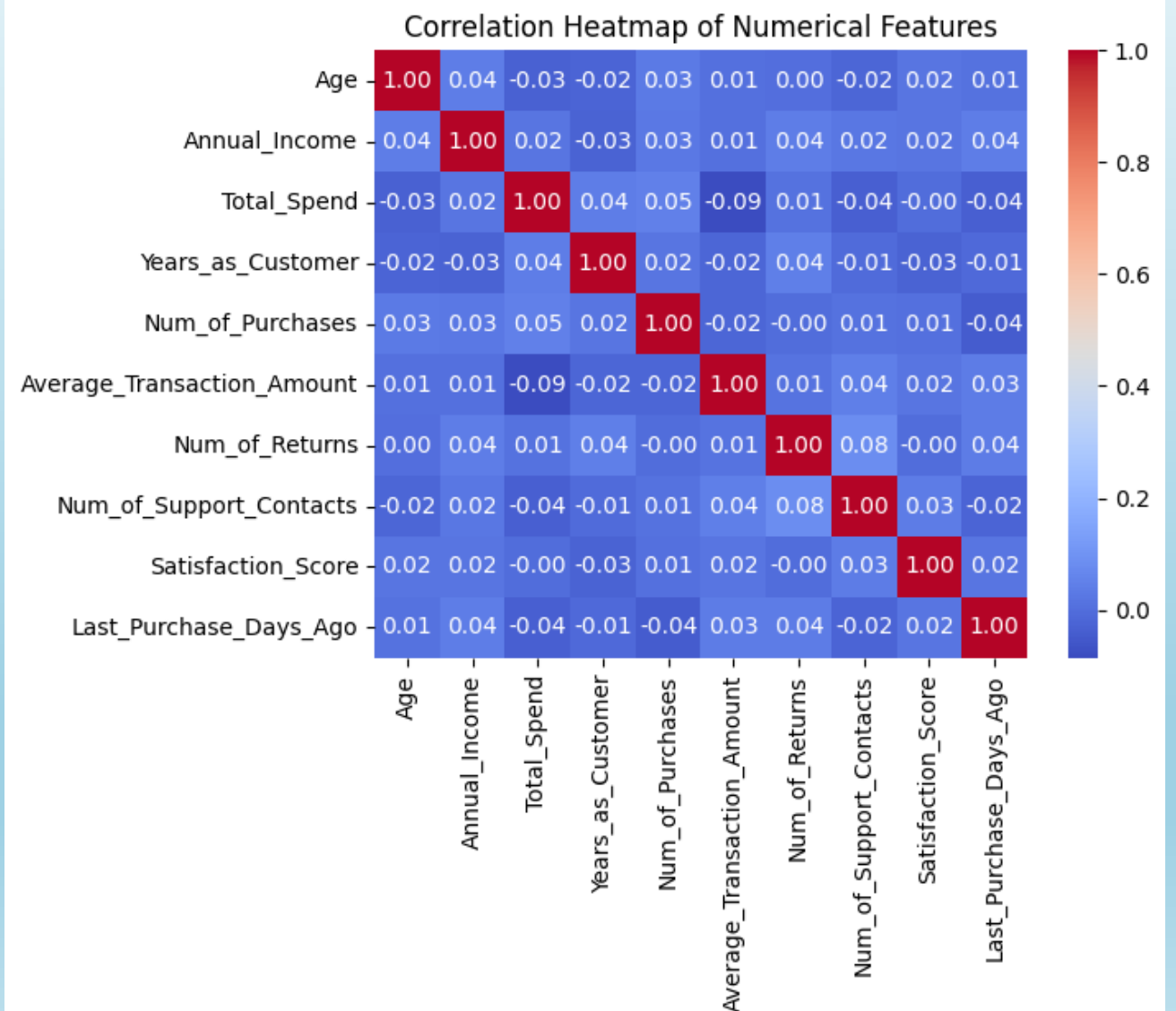
Rekomendasi : improvement di bagian promosi terutama berdasarkan gender contoh untuk perempuan promosi produk kecantikan seperti penawaran khusus atau potongan harga pada waktu waktu tertentu atau terjadwa, bisa juga Reward untuk jumlah transaksi tinggi seperti potongan harga untuk jumlah transaksi tertentu.



Corelalatlon Map Feature

- Correlation map

Tidak ada korelasi yang signifikan antar fitur. Tertinggi hanya 0.09



Feature Engineering

- Penambahan fitur dilakukan untuk meningkatkan perform machine learning.
- Return rate : jumlah return / jumlah transaksi
- Support contact rate : rate pelanggan yang menghubungi CS dalam 1th
- Spending/year : total spending pelanggan pertahun
- Purchase frequency : jumlah transaksi pertahun
- Income spend ratio: perbandingan berapa banyak spending pelanggan terhadap penghasilan annual nya

- Melakukan Encoding pada data
 - kategorikal:
 - Gender,
 - Email_Opt_In,
 - Promotion_Response,
 - Target_Churn

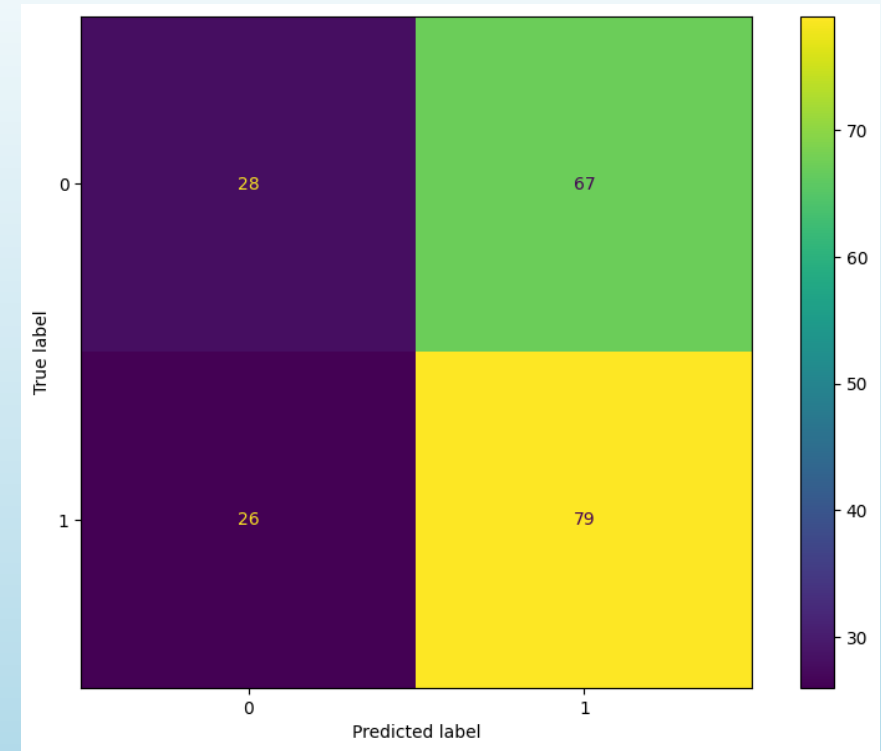
Model Machine Learning

- Pada analysis ini akan dilakukan model prediksi supervised, karena bertujuan untuk memprediksi label untuk data baru.
- menggunakan algoritma Classification (logistic Regressoin(baseine) Random Forest & Xgboost) . Target merupakan label yaitu churn &no churn berupa pengelompokan atau clasifikasi bukan berupa angka.
- Matrix : Recall (berapa persen churn yang berhasil diprediksi dari semua yang benar-benar churn) ,
 - **mengurangi False Negative (FN)** → Semakin tinggi Recall → semakin sedikit pelanggan yang lolos pergi tanpa retensi

Model Machine Learning

Logistic Regression

- model dapat memprediksi 75,2%% pelanggan yang akan churn, Namun kurang baik meprediksi pelanggan yang tidak churn hanya 29,4%
- Recall rata-rata : 52,3%
- precision: 54,1% dari total pelanggan yang diprediksi yang benar benar churn

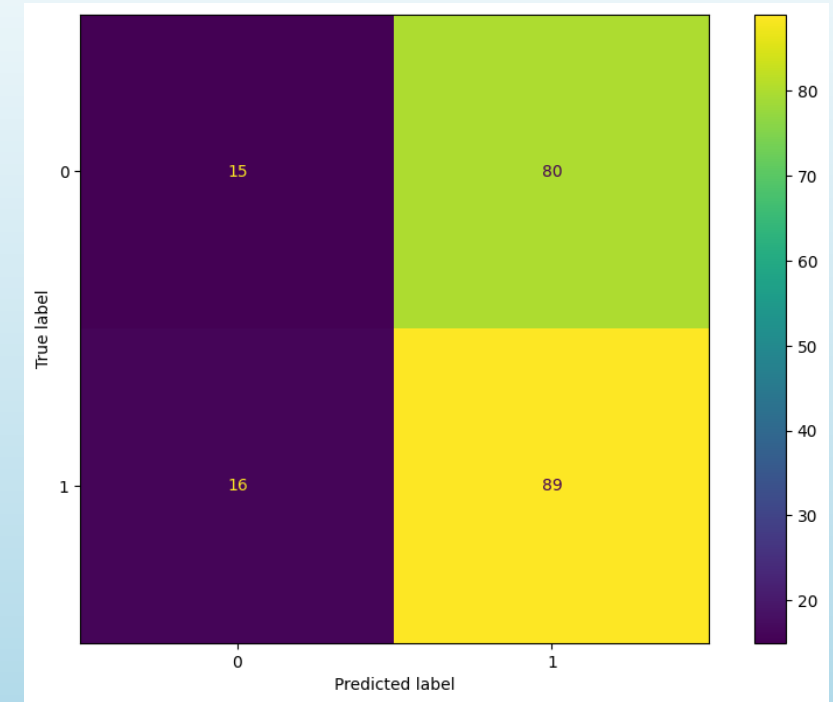


	Not Churn	Churn	accuracy	macro avg	weighted avg
precision	0.518519	0.541096	0.535	0.529807	0.530372
recall	0.294737	0.752381	0.535	0.523559	0.535000
f1-score	0.375839	0.629482	0.535	0.502660	0.509002
support	95.000000	105.000000	0.535	200.000000	200.000000

Model Machine Learning

Random Forest

- model dapat memprediksi 84,7% pelanggan yang akan churn, Namun kurang baik memprediksi pelanggan yang tidak churn hanya 15,7%
- Recall rata-rata : 50,2%
- precision; 52,6% dari total pelanggan yang diprediksi yang benar benar churn

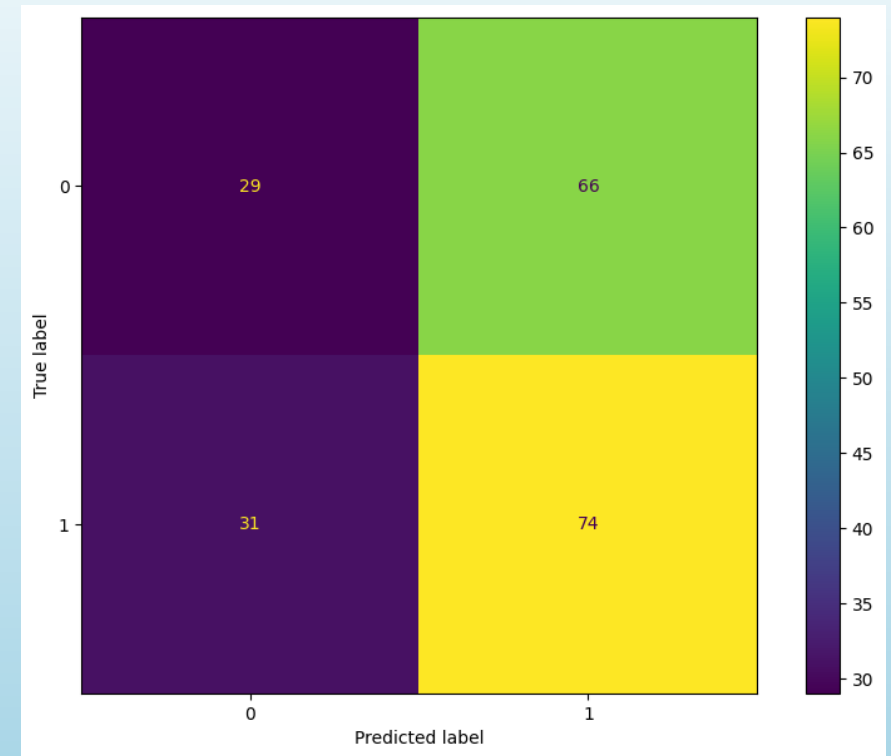


	Not Churn	Churn	accuracy	macro avg	weighted avg
precision	0.483871	0.526627	0.52	0.505249	0.506318
recall	0.157895	0.847619	0.52	0.502757	0.520000
f1-score	0.238095	0.649635	0.52	0.443865	0.454154
support	95.000000	105.000000	0.52	200.000000	200.000000

Model Machine Learning

XgBoost

- Model ini memprediksi pelanggan churn 70,4% lebih baik dari model Random Forest. namun prediksi pelanggan tidak churn lebih tinggi dari model RF. 30,5%
- Recall rata-rata : 50,5%
- precision: 52,8% dari total pelanggan yang diprediksi yang benar benar churn

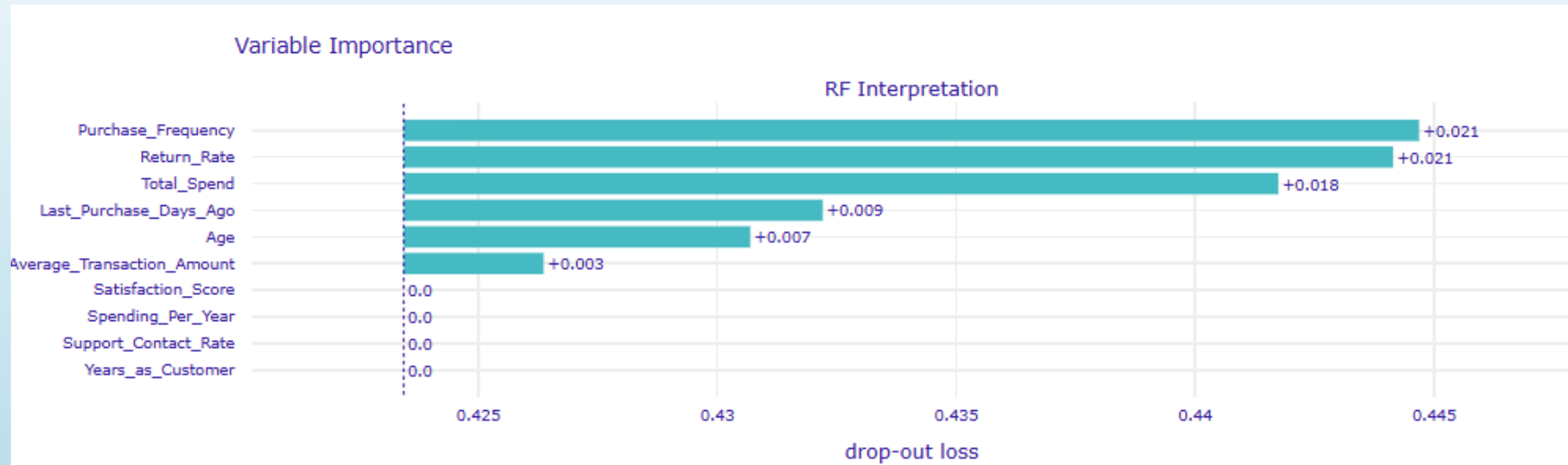


	Not Churn	Churn	accuracy	macro avg	weighted avg
precision	0.483333	0.528571	0.515	0.505952	0.507083
recall	0.305263	0.704762	0.515	0.505013	0.515000
f1-score	0.374194	0.604082	0.515	0.489138	0.494885
support	95.000000	105.000000	0.515	200.000000	200.000000

Insight Rekomendasi Model

- Tujuan memprediksi pelanggan yang churn, model **Random forest** pilihan terbaik dalam project ini.
- Tetapi dengan catatan prediksi false positif tinggi. **Trade off** model ini : lebih menggunakan **biaya promosi (retensi) yang cukup besar**, dibandingkan kehilangan pelanggan.
 - **High Recall, Low Precision:**
Banyak pelanggan yang di"tahan" → beberapa sebenarnya tidak perlu → biaya promo sedikit naik, **tapi churn turun**.
- False Negatif(FN) yang paling merugikan, pelanggan yang lolos(churn) tanpa ditahan.FN pada model ini paling rendah dari 2 model lainnya

Feature Importance Model Random Forest



Interpretasi:

- **Purchase_Frequency** → Kalau pelanggan jarang belanja, mereka berpotensi lebih besar untuk churn.
- **Return_Rate** → Pelanggan yang sering melakukan return (pengembalian) punya kecenderungan churn lebih tinggi, mungkin karena ketidakpuasan.
- **Total_Spend** → Makin besar total belanja, biasanya makin loyal, atau sebaliknya kalau turun drastis.
- **Last_Purchase_Days_Ago** → Semakin lama tidak belanja, makin besar kemungkinan churn.
- **Age** → Umur sedikit berpengaruh; mungkin kelompok umur tertentu lebih rentan churn.

Actionable Insight:

- Fokus meningkatkan **frekuensi pembelian**: reminder, promo rutin, membership.
- Perhatikan pelanggan dengan **tinggi return rate**: bisa buat program service improvement.
- Monitor **pelanggan yang lama tidak belanja**: kasih special reactivation offer.
- Segmen umur tertentu bisa dibuatkan campaign khusus (kalau dari analisis lebih lanjut ada perbedaan signifikan).

Terima Kasih