

# Customer Churn Analysis

1. Maulana Malik Ibrahim
2. Bintang Ary Pradana
3. Maulana Yusuf Ikhsan Robbani

# Table of contents

## 01 Introduction

Pengenalan terhadap Churn Analysis.

## 02 Business Understanding

Pemahaman terkait konteks dan kebutuhan bisnis perusahaan.

## 03 Data Understanding

Pemahaman terkait dataset untuk persiapan analisis.

## 04

## Data Preparation, Analysis, dan Modelling

Melakukan persiapan, analisis data, dan pembuatan modelling untuk memprediksi churn.

## 05

## Conclusion

Kesimpulan dari hasil analisis dan model.

## 06

## Recommendation

Rekomendasi hal yang harus dilakukan perusahaan selanjutnya.

01

# Introduction

---

Apa itu Customer Churn Analysis?

# Customer Churn Analysis and Prediction

**Customer churn** merupakan pelanggan yang tidak melakukan transaksi dalam jangka waktu yang cukup lama, biasanya sekitar 3 bulan yang menandakan bahwa pelanggan tersebut kehilangan keaktifan dalam penggunaan layanan,

Sehingga **perlu dilakukan customer churn analysis dengan tujuan untuk mengidentifikasi pelanggan yang berpotensi (churn) dan mengambil langkah-langkah yang tepat** baik sebelum maupun setelah terjadinya churn, melalui prediksi hasil yang akurat



02

# Business Understanding

---

Apa yang diharapkan perusahaan terhadap project ini?

# Kebutuhan Business Perusahaan



**PT TIGA PILAR FAMILIA**

Saat ini, perusahaan ingin mendapatkan pemahaman yang mendalam tentang pelanggan yang berpotensi churn dan hal yang harus dilakukan untuk mencegah kehilangan pelanggan.

Solusi yang diinginkan harus mampu menganalisis data historis dan perilaku pelanggan untuk mengenali pola-pola yang mengindikasikan kemungkinan churn di masa depan.

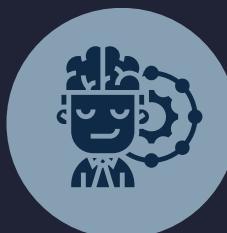
# Goals



Mengidentifikasi pelanggan yang berpotensi (churn) berhenti menggunakan layanan atau produk.



Menganalisis karakteristik pelanggan yang cenderung melakukan churn berdasarkan data historis penjualan.



Mengembangkan dan mengevaluasi model machine learning yang dapat dengan akurat memprediksi kemungkinan churn pada pelanggan.



Menyusun strategi untuk menangani churn dan tindakan yang harus dilakukan.

03

# Data Understanding

---

Data apa saja yang tersedia sebagai bahan analisis?

# Overview Dataset



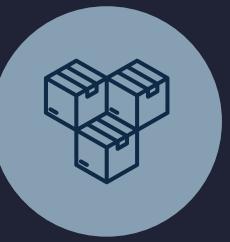
## Sales 2022

Berisikan data transaksi penjualan dari Jan - Apr 2022



## Sales 2022 - 2024

Berisikan data transaksi penjualan dari Mei 2022 hingga Apr 2024.



## Items

Berisikan data detail produk yang ada di perusahaan



## City

Berisikan data daerah atau willayah tujuan pengiriman produk.



## Reward

Berisikan data customer yang mendapatkan reward tertentu



## Customer

Berisikan data detail customer yang telah bertransaksi di perusahaan.

04

# Data Preparation, Analysis & Modeling

---

Data apa saja yang tersedia sebagai bahan analisis?

# Table of contents

## 01 Exploratory Data Analysis (EDA)

Memahami kondisi data secara keseluruhan

## 02 Feature Engineering

Melakukan proses ekstraksi fitur-fitur menjadi variabel yang baru

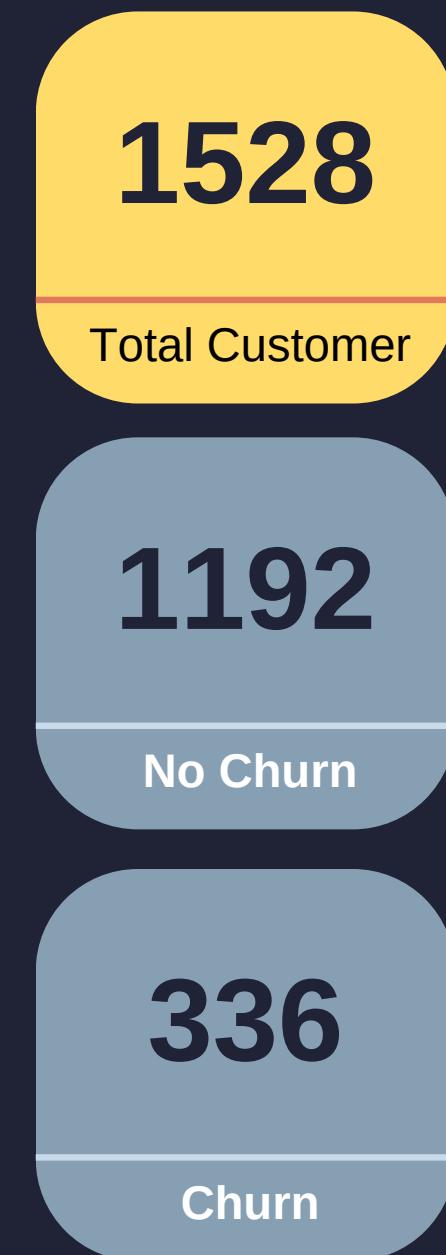
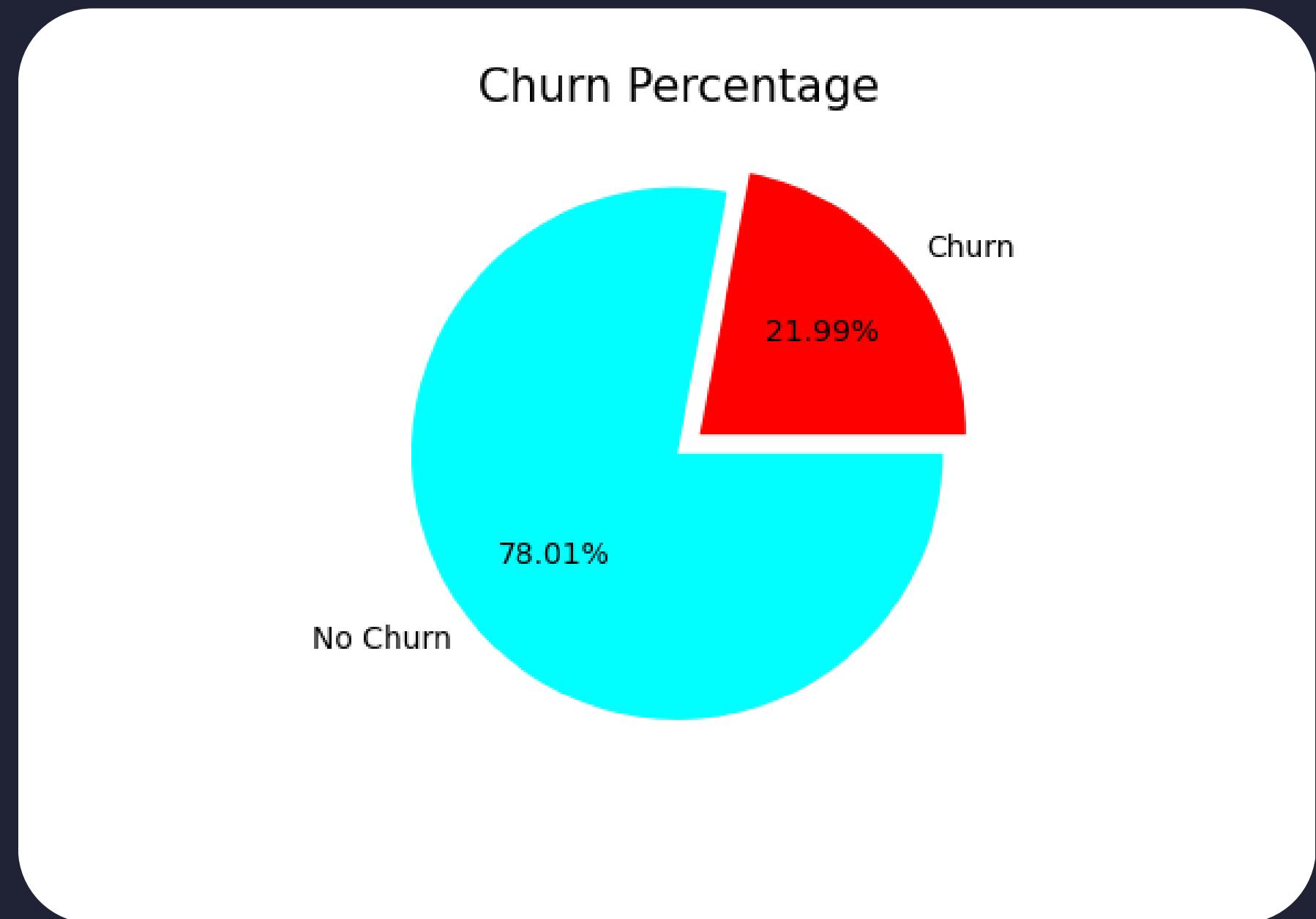
## 03 Scalling & Oversampling Data

Meningkatkan densitas data dan menyeimbangkan data yang imbalance

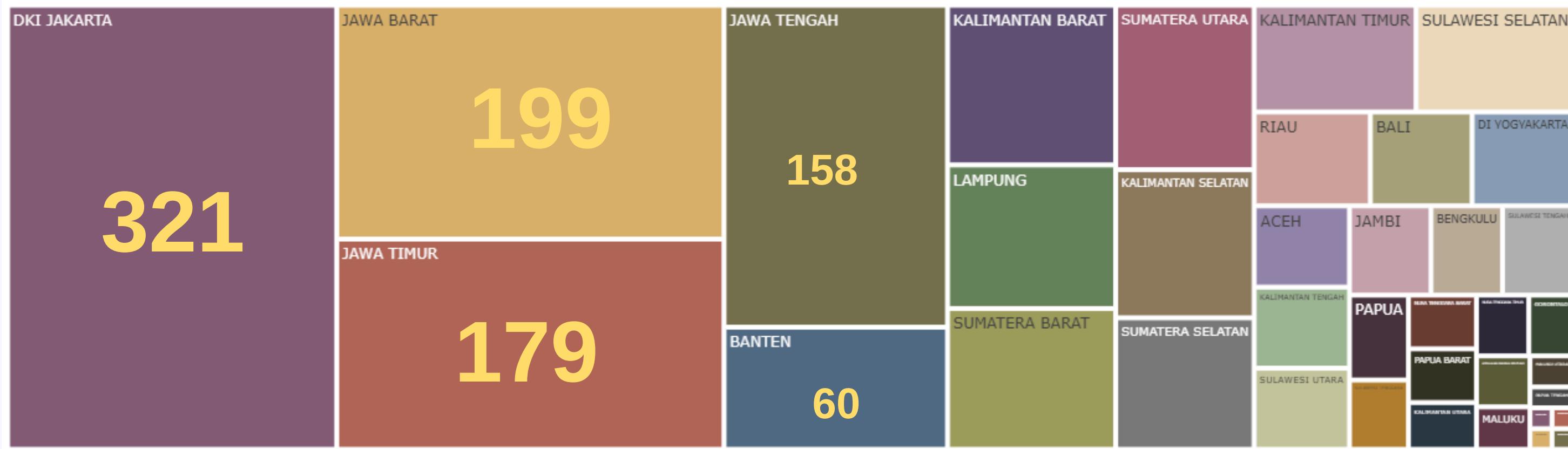
## 04 Modelling

Melatih model dengan data yang kita miliki agar dapat memprediksi customer churn

# Churn Analysis from the Data



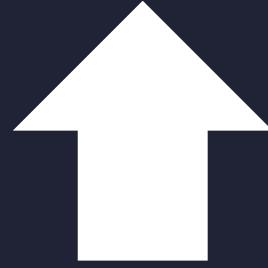
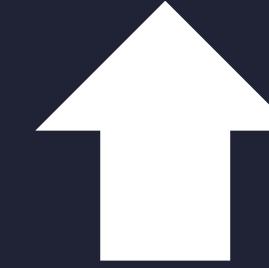
# Customer Transaction by Province



# Feature Engineering

Tujuannya adalah membuat fitur (kolom) baru dari data yang sudah ada. Dengan ini, model akan mendapat lebih banyak wawasan sehingga model lebih baik dalam menentukan customer mana yang churn atau yang tidak.

Jumlah  
kolom



Performa \*  
Model

\*Ada beberapa kasus memperburuk performa

# Feature Engineering

payments	credit_limit	churn	return_barang	total_return_barang	potongan_penjualan	total_potongan_penjualan	total_unique_item	top_category	kategori_count	ship_to_province	reward_2022	reward_2023_p1	reward_2023_p2	create_to_first_sales	tenure
0	0.00	Y	0	0.00	0	0.00	71.00	AKSESORIES CAMPUR	23.00	ACEH	0.00	0.00	0.00	0.00	26.63
14	30000000.00	N	1	51000.00	0	0.00	724.00	TAS & DOMPET	258.00	ACEH	0.00	0.00	1.00	0.00	26.57
0	0.00	N	1	330000.00	0	0.00	290.00	AKSESORIES CAMPUR	115.00	ACEH	0.00	0.00	0.00	0.00	25.70
14	20000000.00	N	0	0.00	0	0.00	250.00	TAS & DOMPET	153.00	ACEH	0.00	0.00	0.00	0.00	26.63
0	0.00	N	0	0.00	0	0.00	147.00	TAS & DOMPET	75.00	ACEH	0.00	0.00	0.00	0.00	25.40
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
0	0.00	Y	0	0.00	0	0.00	2.00	AKSESORIES CAMPUR	2.00	DKI JAKARTA	0.00	0.00	0.00	0.00	19.17
7	0.00	Y	0	0.00	0	0.00	37.00	AKSESORIES CAMPUR	13.00	SUMATERA BARAT	0.00	0.00	0.00	0.00	26.93
7	0.00	Y	0	0.00	0	0.00	1.00	ALAT MAKE UP	1.00	SUMATERA BARAT	0.00	0.00	0.00	0.00	25.60
0	0.00	Y	0	0.00	0	0.00	26.00	AKSESORIS RAMBUT	32.00	SUMATERA BARAT	0.00	0.00	0.00	0.00	26.47
0	0.00	Y	0	0.00	0	0.00	51.00	TAS & DOMPET	39.00	KALIMANTAN SELATAN	0.00	0.00	0.00	0.00	25.87

Hal yang ditambahkan sebagai fitur baru yaitu:

1. '**top\_category**', yaitu kategori yang paling sering dibeli oleh *customer*.
2. '**number\_of\_month**', yaitu jumlah unik bulan dari *customer* itu bertransaksi. (Dipakai avg\_amount)
3. '**kategori\_count**', yaitu jumlah kategori unik yang beli oleh *customer* selama bertransaksi.
4. '**tenure**', yaitu lamanya waktu *customer* sejak pembuatan akun hingga tanggal referensi saat ini, diukur dalam bulan.
5. '**avg\_amount**', yaitu rata-rata jumlah uang yang dikeluarkan oleh *customer* saat bertransaksi.

## Proses Droping Fitur

1. Drop fitur *number\_of\_month* karena datanya redundant

# Feature Engineering

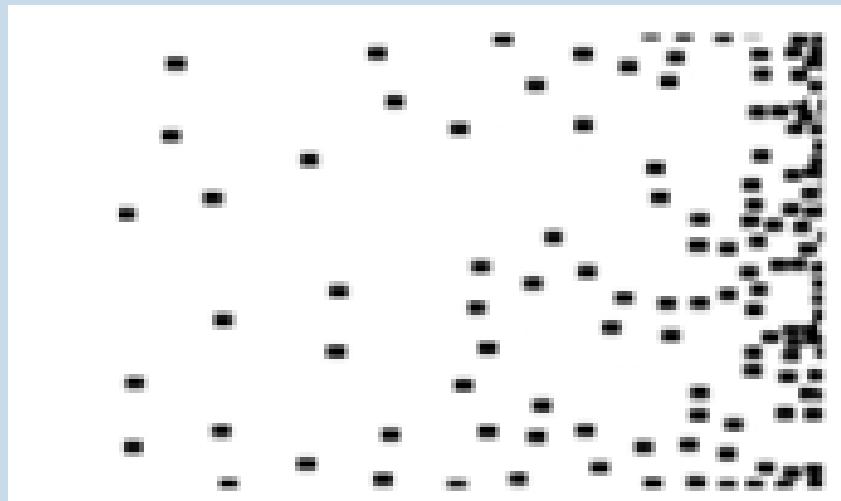
## Encoding

payments	credit_limit	churn	return_barang	total_return_barang	potongan_penjualan	total_potongan_penjualan	total_unique_item	top_category
0	0.00	1	0	0.00	0	0.00	71.00	AKSESORIES CAMPUR
14	30000000.00	0	1	51000.00	0	0.00	724.00	TAS & DOMPET
0	0.00	0	1	330000.00	0	0.00	290.00	AKSESORIES CAMPUR
14	20000000.00	0	0	0.00	0	0.00	250.00	TAS & DOMPET
0	0.00	0	0	0.00	0	0.00	147.00	TAS & DOMPET
kategori_count	ship_to_province	reward_2022	reward_2023_p1	reward_2023_p2	create_to_first_sales	tenure	avg_amount	
23.00	ACEH	0.00	0.00	0.00	0.00	26.63	6685875.00	
258.00	ACEH	0.00	0.00	1.00	0.00	26.57	12032447.00	
115.00	ACEH	0.00	0.00	0.00	0.00	25.70	7452478.36	
153.00	ACEH	0.00	0.00	0.00	0.00	26.63	4786600.00	
75.00	ACEH	0.00	0.00	0.00	0.00	25.40	8414307.69	

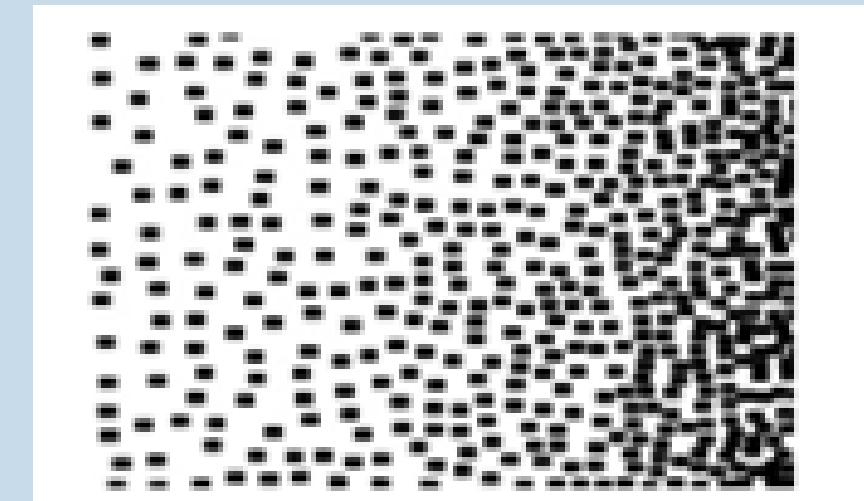
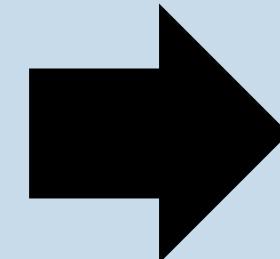
- Encoding adalah pengubahan nilai kategorik menjadi numerik agar model lebih memahami data yang diberikan.
- Kolom seperti ‘*top\_category*’ & ‘*ship\_to\_province*’ dilakukan one-hot encoding. Sehingga dari satu kolom menjadi banyak kolom.
- Contoh ‘*top\_category*’ ada 6 nilai unik sehingga one hot encoding ada 6 kolom baru sesuai nilai kolom itu

# Handling High Density Data

## Data Scalling



Densitas Rendah



Densitas Tinggi

- Density adalah seberapa rapat jarak antara data dengan data lain
- Karena data densitasnya rendah maka perlu ditingkatkan dengan **Robust Scaller**

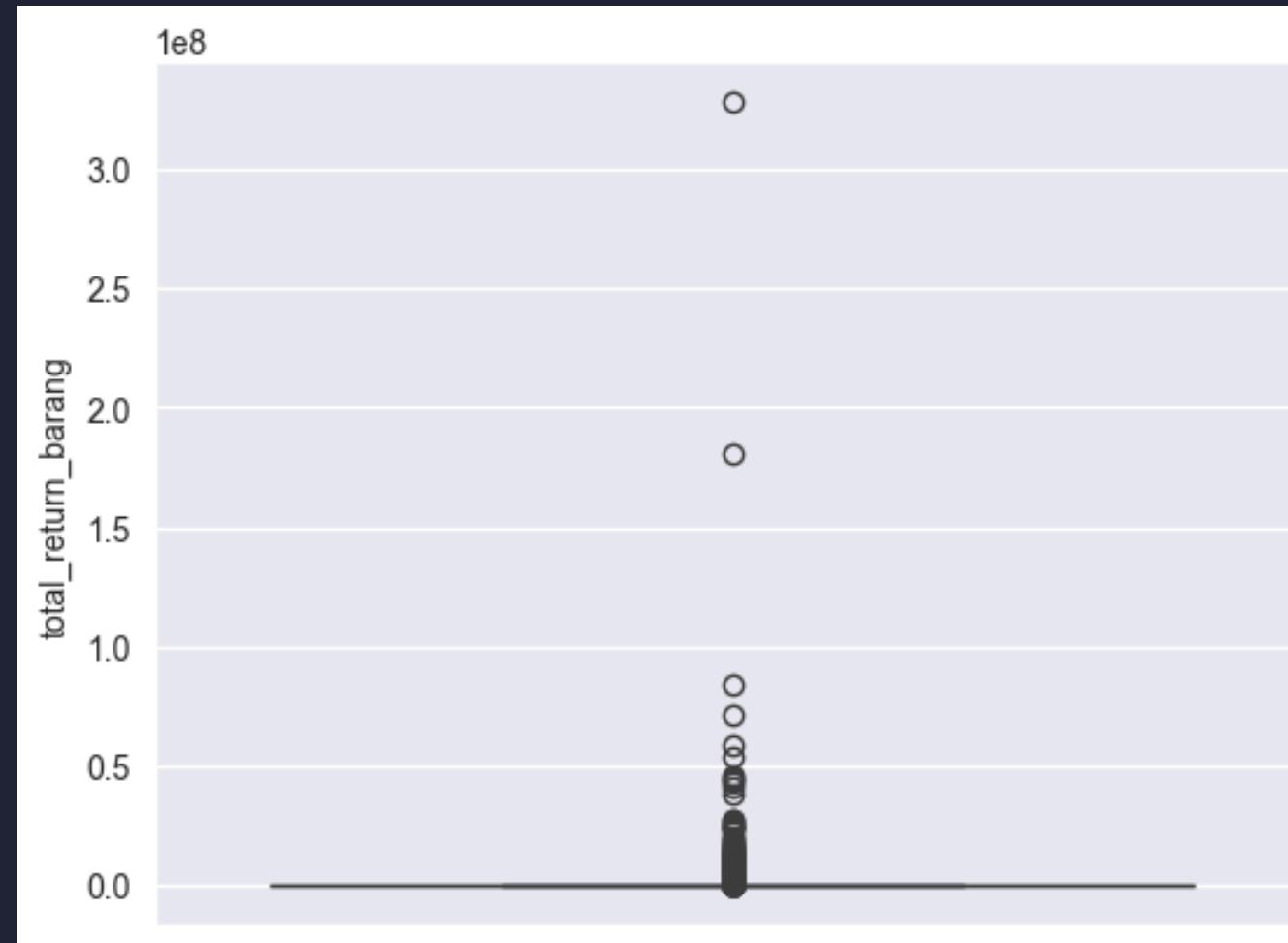
Source: [Open BI](#)



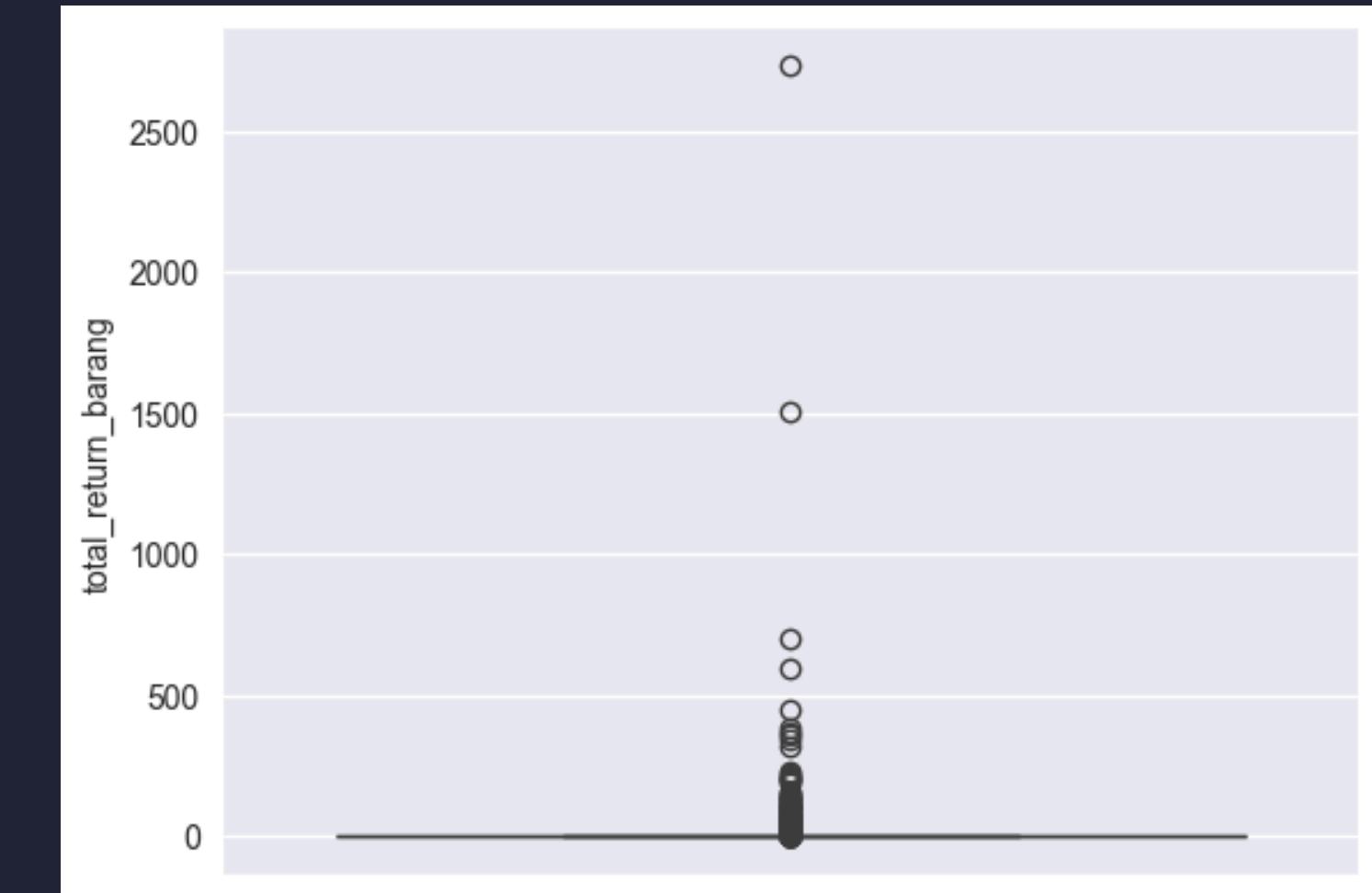
# Handling High Density Data

## Data Scalling

Sebelum Scalling

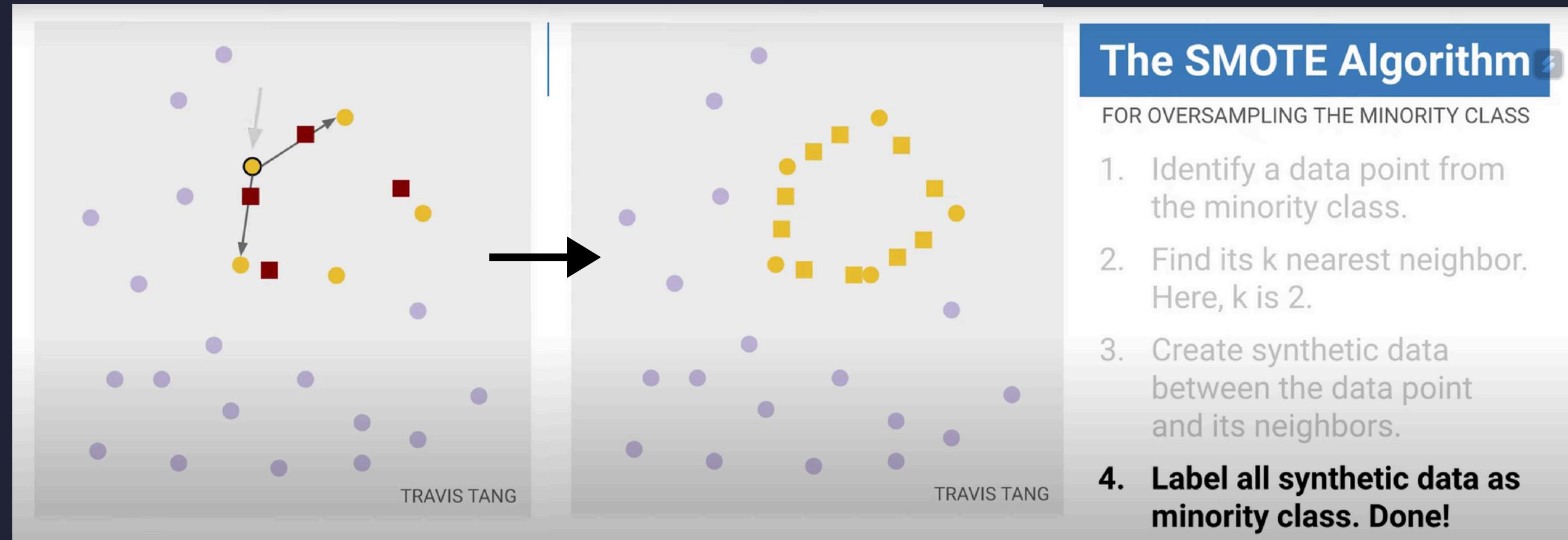


Setelah Scalling



# Handling Imbalance Data

## Oversampling



- Imbalance data adalah kondisi ketika data suatu kolom ada satu nilai jumlahnya mendominasi dibanding nilai lain
- Penanggulangannya dengan Oversampling SMOTE Algorithm

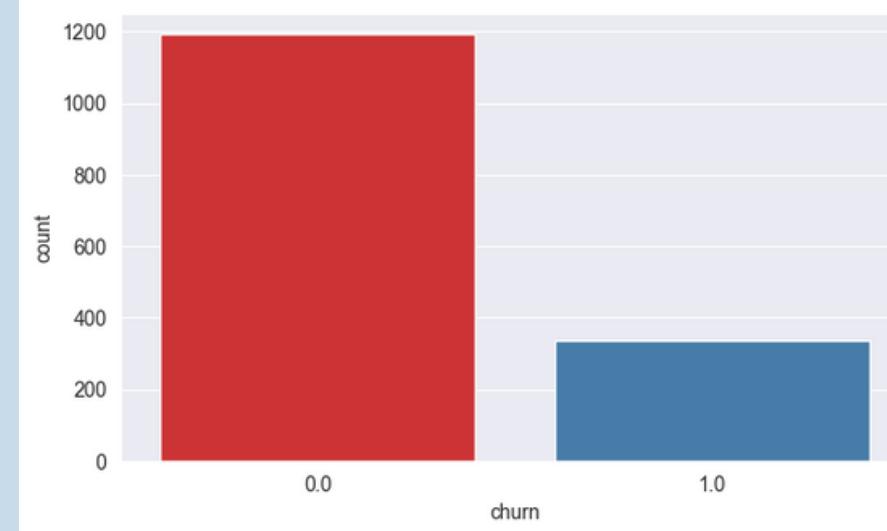
Source: Travis Tang



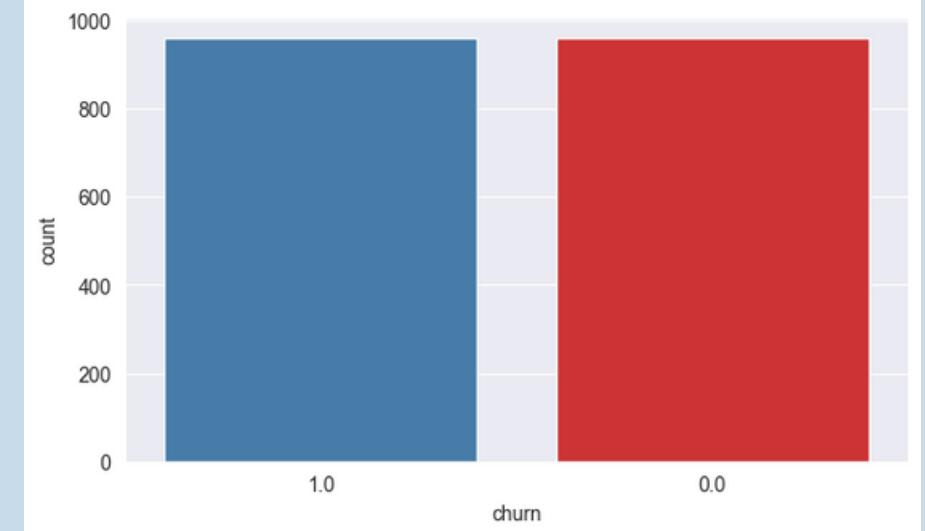
# Handling Imbalance Data

## Oversampling

Sebelum Oversampling



Sesudah Oversampling



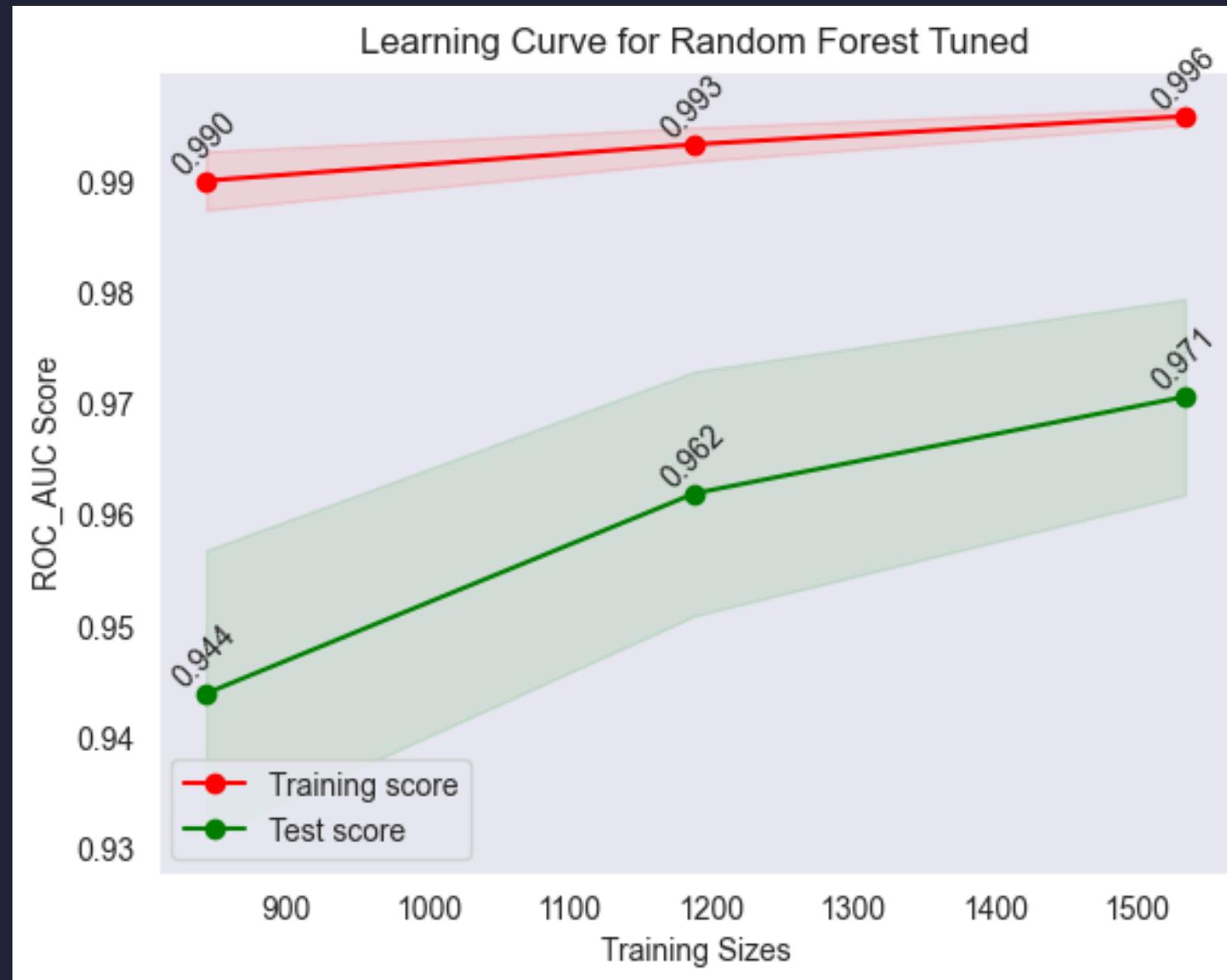
Imbalance

Balance

0.0 untuk No, 1.0 untuk Yes



# Modeling



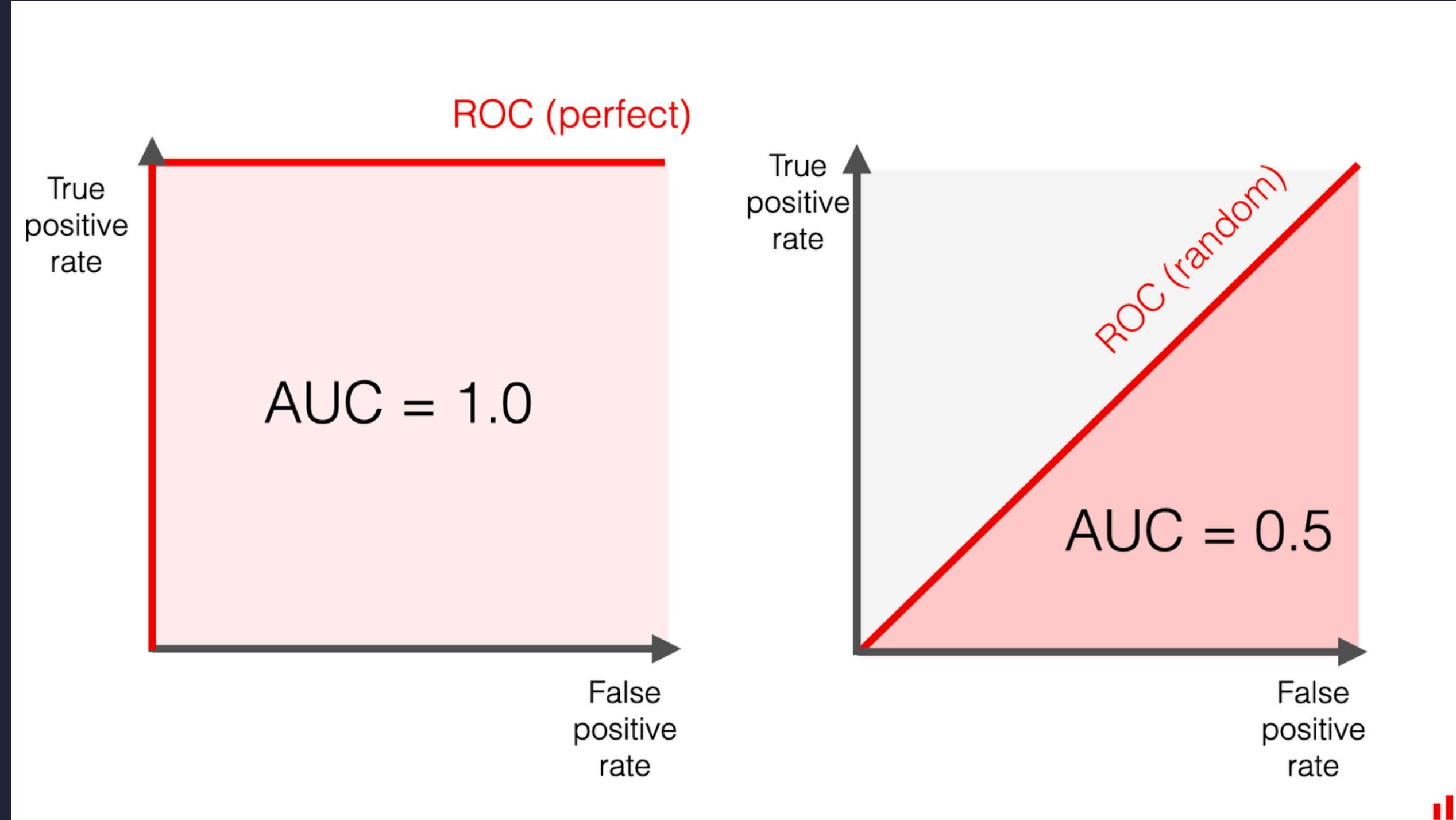
Modelling adalah proses melatih model dengan diberi data yang sudah kita transformasi.

Dari beberapa jenis pemodelan yang sudah dicoba. Maka, **dipilih model Random Forest classifier** yang sudah di hyperparameter tuning dikarenakan:

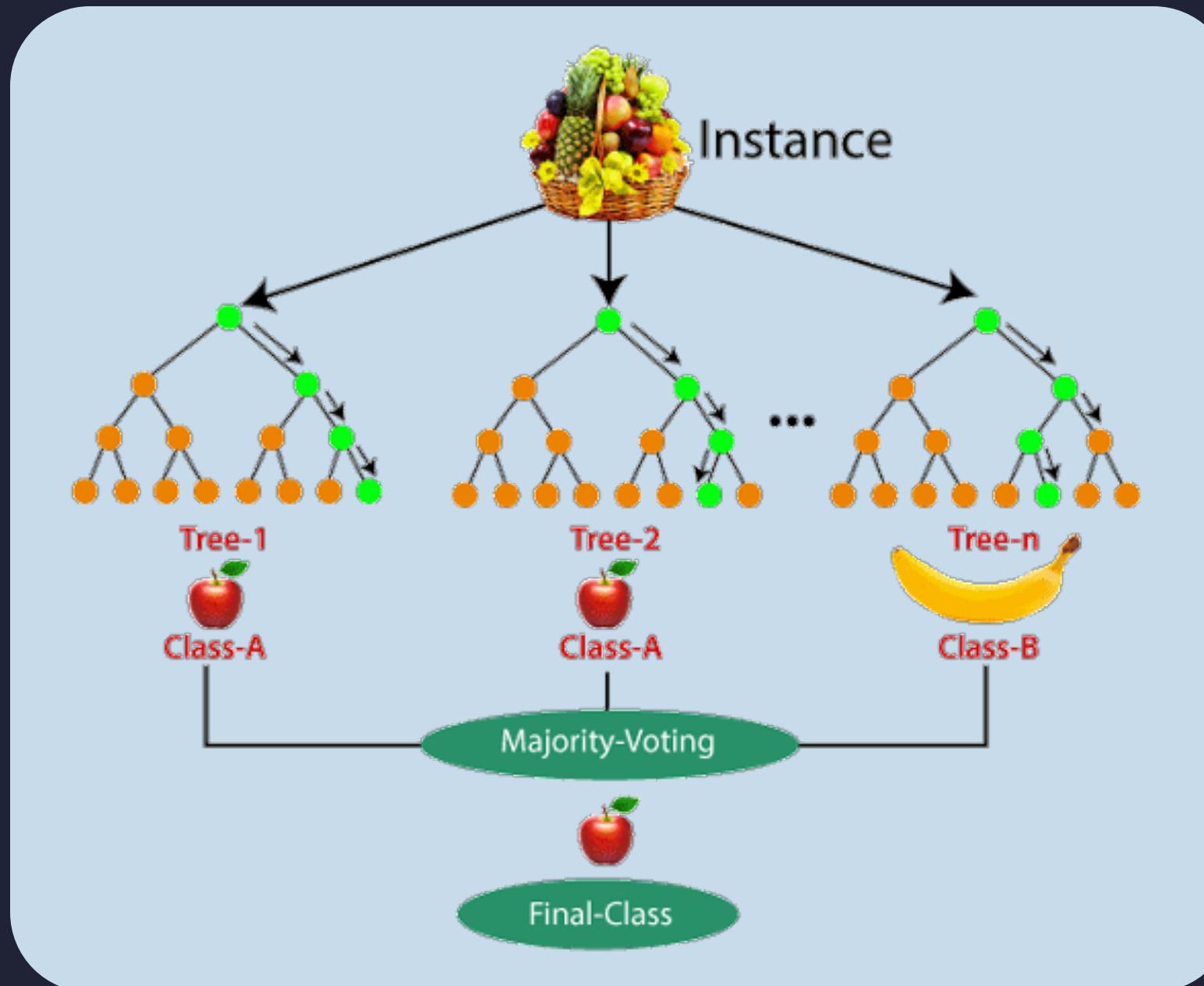
1. Saat ukuran data pembelajaran ditambah, model memiliki peningkatan skor train & test
2. Selisih train & test score lebih rendah daripada model lain, **sehingga tingkat keakuratannya lebih baik dibandingkan dengan jenis model yang lain**

# Modeling

## Penjelasan ROC AUC



# Modeling



## Bagaimana Random Forest bisa memprediksi churn?

**Random Forest** adalah algoritma *machine learning* yang menggunakan kumpulan pohon keputusan (decision tree) untuk membuat prediksi yang lebih akurat.

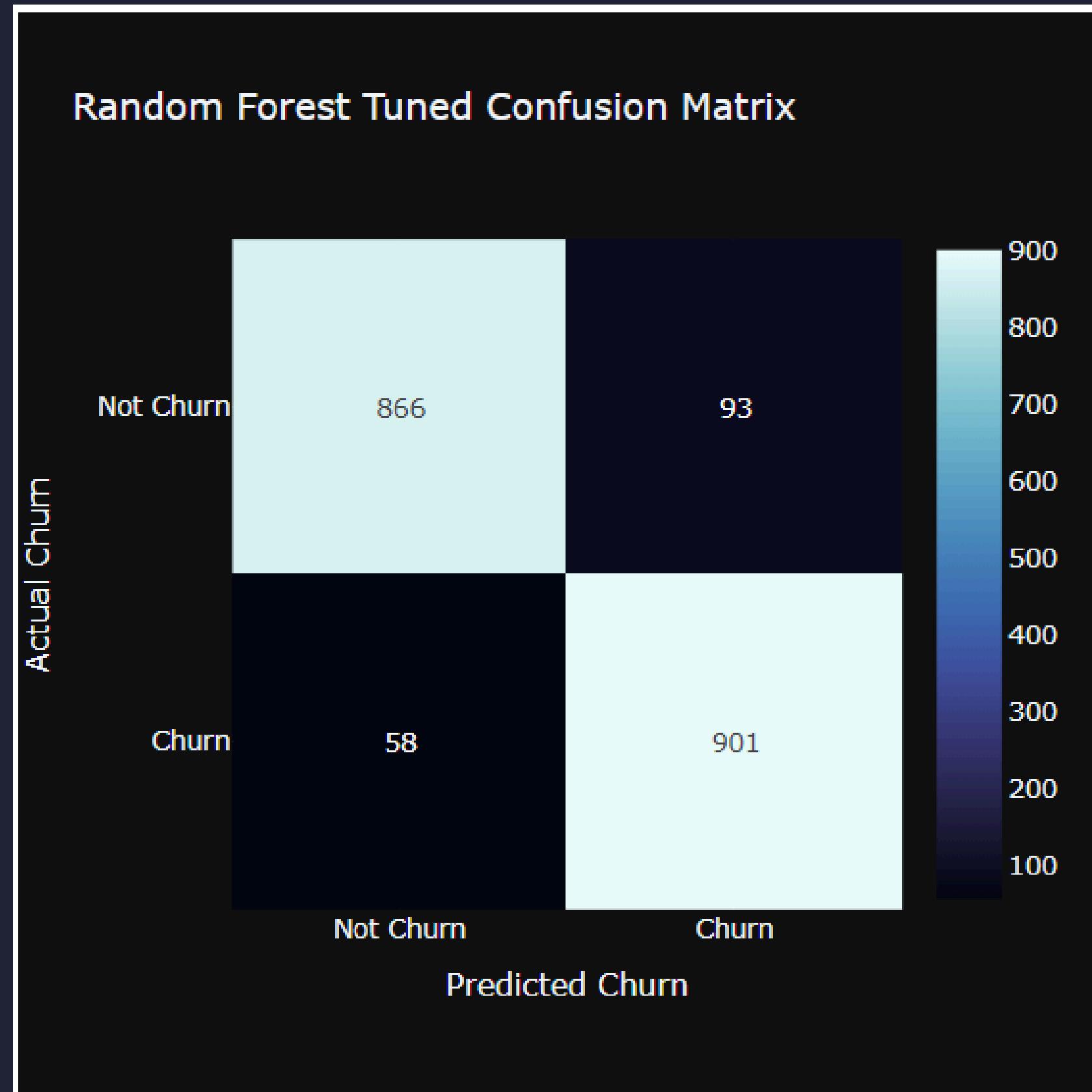
Dimana setiap pohon keputusan memberikan prediksi dan hasil akhir ditentukan berdasarkan mayoritas suara dari semua pohon. Teknik ini membantu mengurangi kesalahan prediksi dan membuat hasilnya lebih konsisten.



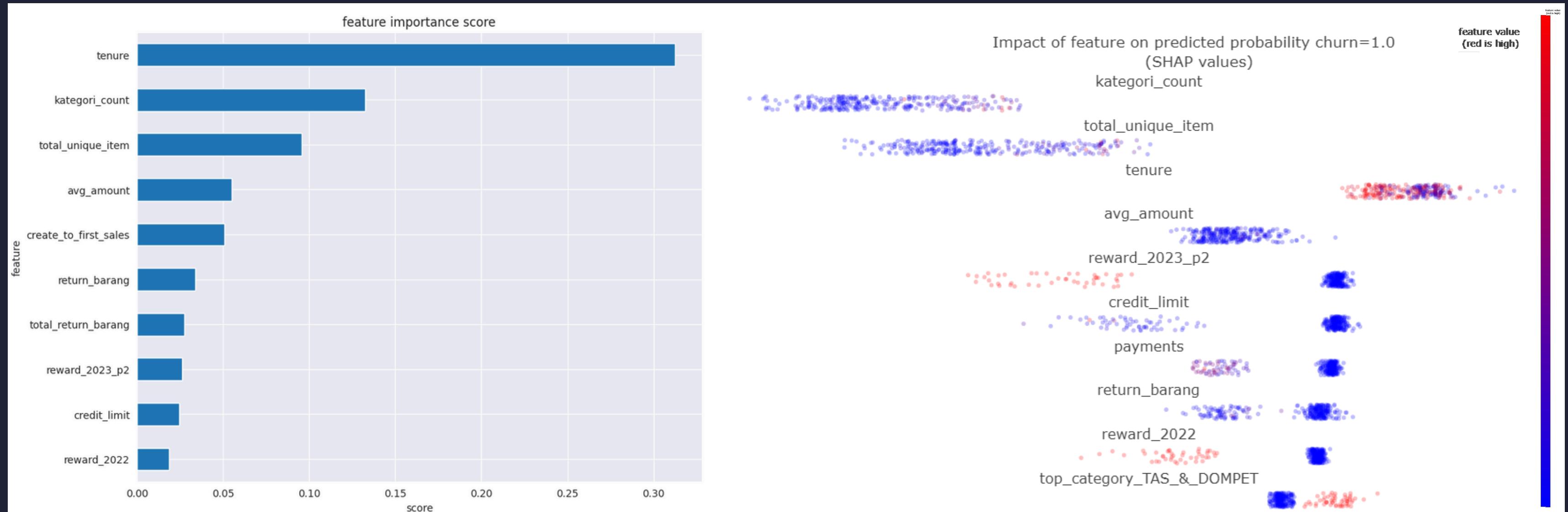
# Modeling

Model	Precision (%)		ROC-AUC (%)	
	Data Train	Data Test	Data Train	Data Test
Random Forest	88,1	87,2	95,6	93,9
Random Forest (Tuned)	98,1	90,7	100	97,6

# Hasil Prediksi



# Feature Importance



# 05

# Conclusion

---

Kesimpulan dari hasil analisis dan model.

# Kesimpulan

Berdasarkan hasil analisa, **model Random Forest Tuned menunjukkan kinerja yang sangat baik dalam memprediksi customer churn.**

Model ini memiliki presisi yang tinggi pada data train dan data test, dan nilai ROC-AUC yang menunjukkan kemampuan model yang baik untuk membedakan antara pelanggan churn dan yang tidak. Feature importance menunjukkan bahwa fitur seperti tenure, kategori\_count, total\_unique\_item, avg\_amount, dan create\_to first\_sales memiliki pengaruh yang besar terhadap prediksi model.

06

# Recommendation

---

Rekomendasi hal yang harus dilakukan perusahaan selanjutnya.

# Recommendation

Berdasarkan interpretasi feature importance dan kesimpulan model, berikut adalah beberapa rekomendasi bisnis yang dapat diambil untuk meningkatkan retensi pelanggan dan mengurangi churn:

## 1. Fokus pada pelanggan baru:

- Berikan perhatian khusus kepada pelanggan baru dan tawarkan program loyalitas atau diskon untuk meningkatkan keterlibatan dan retensi mereka.
- Tingkatkan onboarding dan proses aktivasi pelanggan untuk mendorong pembelian pertama yang lebih cepat dan meningkatkan retensi.

## 2. Dorong pembelian lintas kategori dan item unik:

- Diversifikasikan penawaran produk Anda dan tawarkan rekomendasi produk yang dipersonalisasi untuk mendorong pembelian lintas kategori.
- Tawarkan program loyalitas yang memberi penghargaan kepada pelanggan untuk pembelian item unik, dan tingkatkan variasi produk Anda untuk menarik minat pelanggan.



# Recommendation

## 3. Tingkatkan kepuasan pelanggan:

- Tingkatkan kualitas produk dan layanan pelanggan untuk mengurangi tingkat pengembalian produk.
- Pantau alasan pengembalian produk untuk mengidentifikasi dan mengatasi masalah yang mendasarinya.
- Evaluasi efektivitas program reward dan pertimbangkan untuk memperluas atau meningkatkan program reward untuk meningkatkan retensi pelanggan.

## 4. Manfaatkan data untuk pengambilan keputusan:

- Terus pantau kinerja model churn dan perbaruiinya dengan data baru secara berkala untuk memastikan akurasinya.
- Gunakan data churn analysis untuk mengidentifikasi segmen pelanggan yang berisiko churn dan menargetkan intervensi yang tepat.
- Analisis pola churn untuk mengidentifikasi dan mengatasi penyebab churn yang mendasarinya.



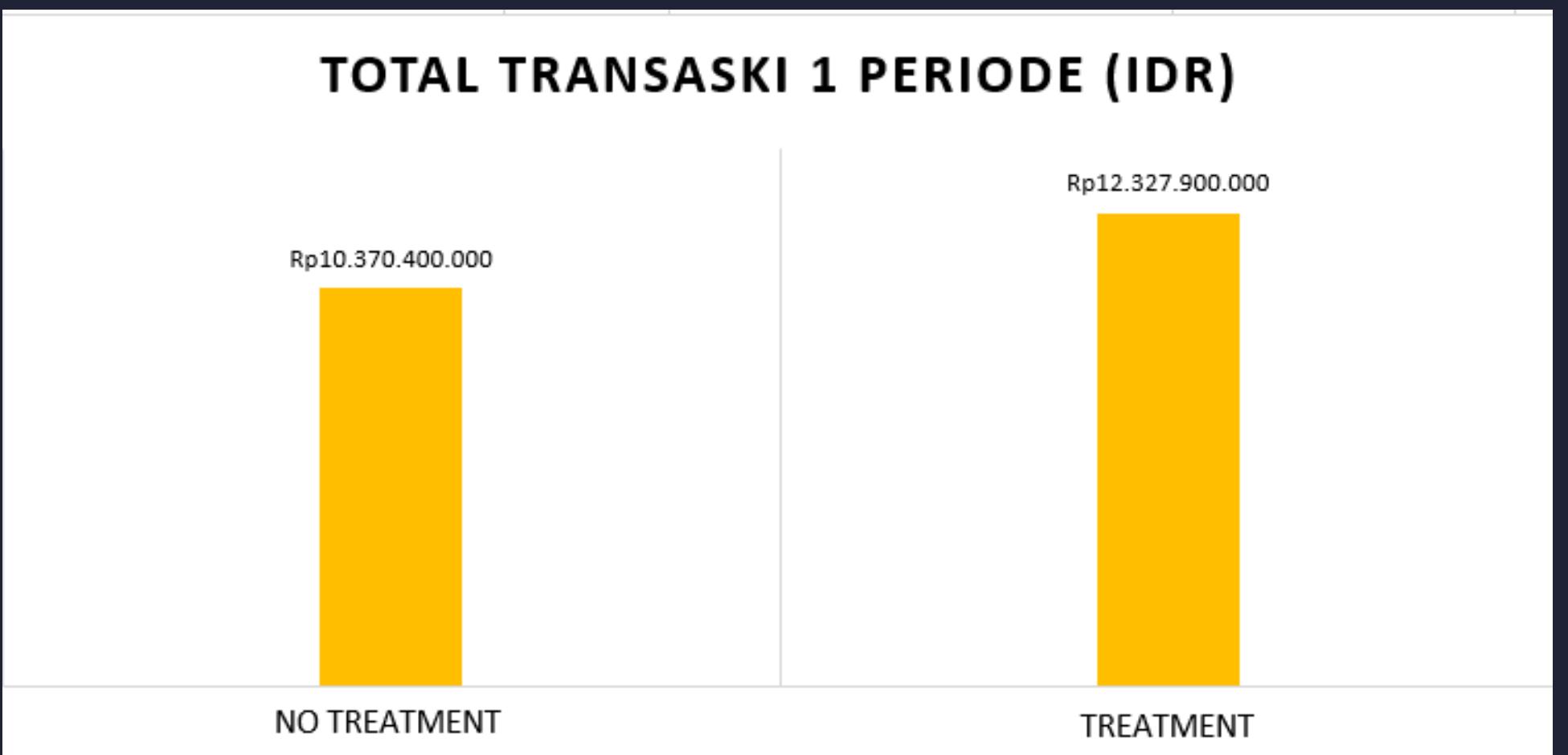
# Impact Analysis

## Asumsi Dasar

1. Nilai Transaksi Rata-Rata IDR 8700.000
2. Succes Ratio Churn Treatment 67%

Parameter	No Treatment	Treatment
Jumlah User	1528 Orang	1528 Orang
Jumlah User Churn	336 Orang	111 Orang
Nilai Transaksi @User	Rp8.700.000	Rp8.700.000
Total Transaksi dalam 1 Periode	Rp10.370.400.000	Rp12.327.900.000
Churn Rate	21,99 %	7,26 %

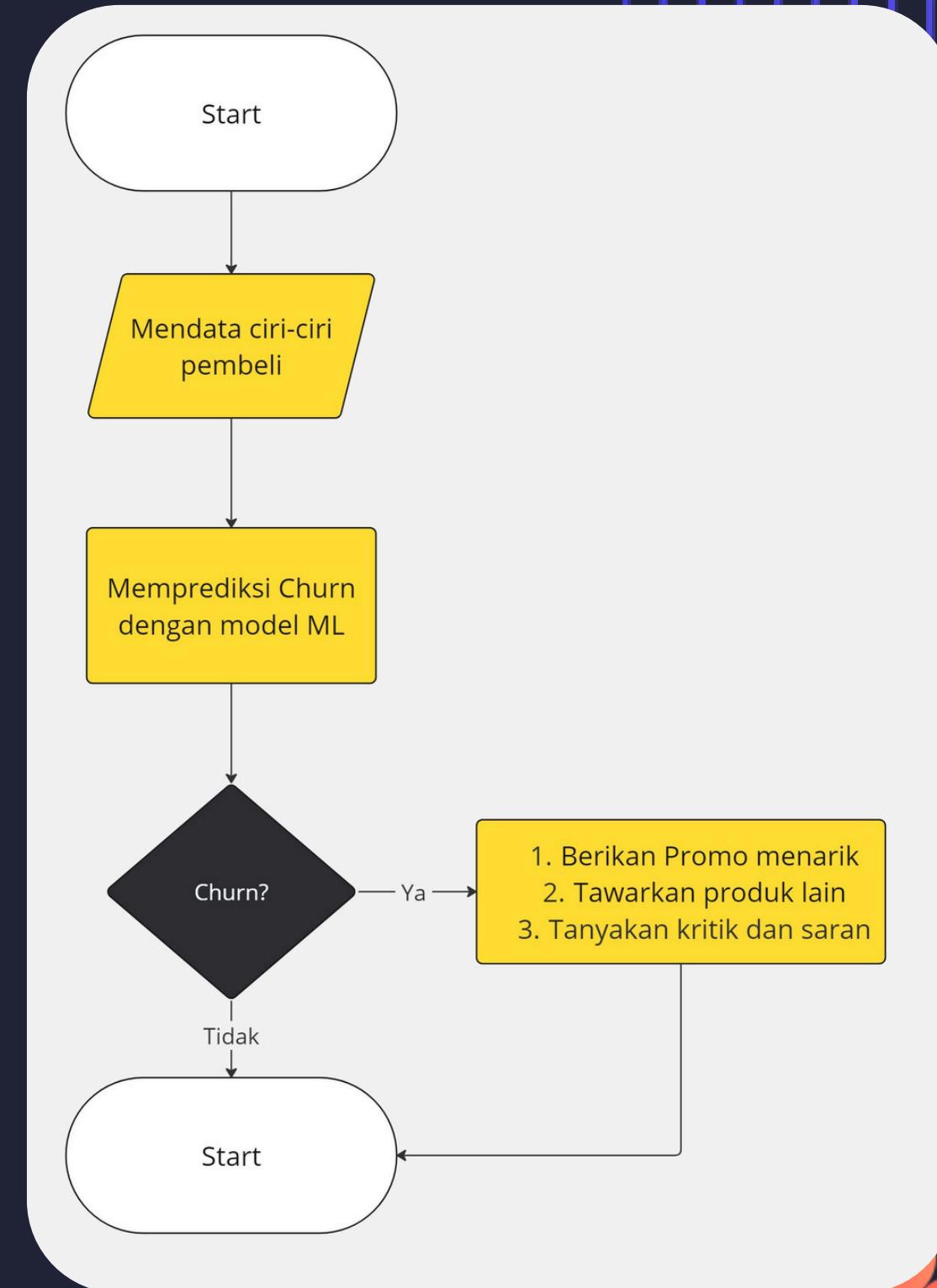
Increase Revenue by 18,9%



# Proses Penanggulangan Customer Churn

Ketika pelanggan sudah tidak bertransaksi 2 - 3 bulan maka perlu dilakukan langkah penanganan agar pelanggan kembali bertransaksi. Berikut langkah-langkahnya:

1. Melakukan pendataan ciri-ciri pembeli
2. Gunakan data tersebut untuk dibaca model
3. Jika model memprediksi pelanggan tersebut churn maka lakukan penanganan



# Terima Kasih



# Customers Buying Patterns Analysis

01

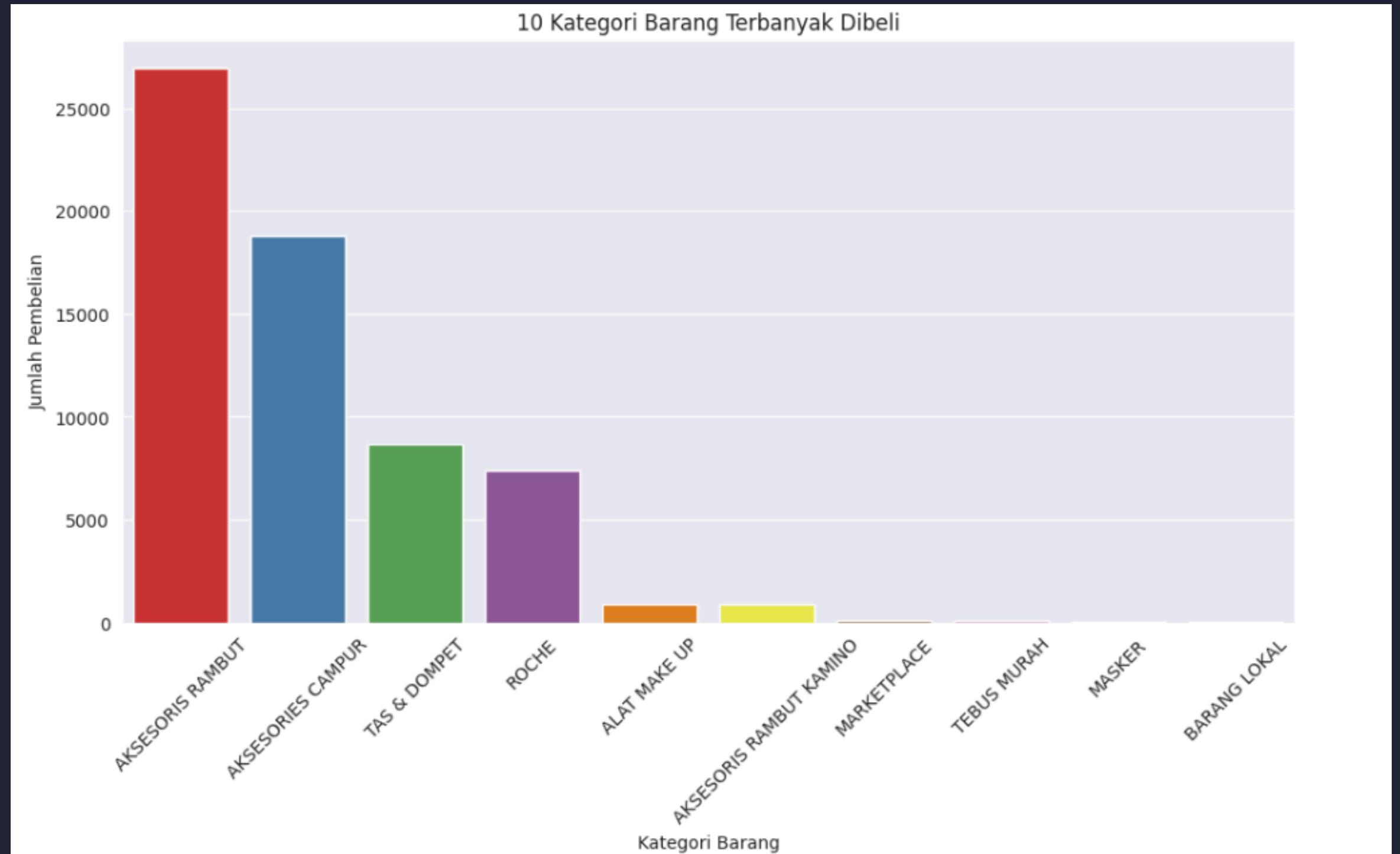
# Exploratory Data Analysis

---

Gambaran umum kondisi penjualan

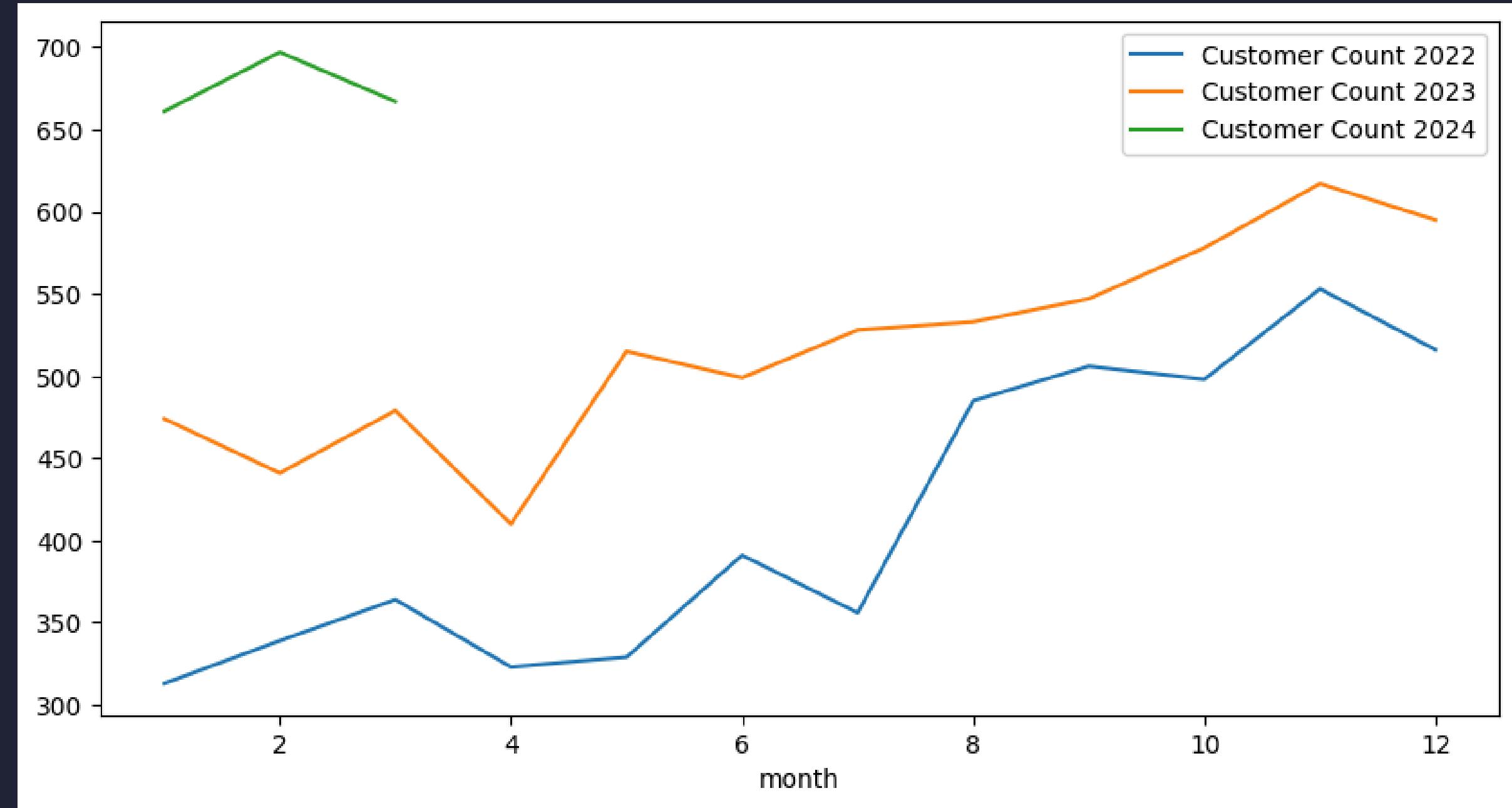
# EDA

10 Kategori Barang Terbanyak Dibeli



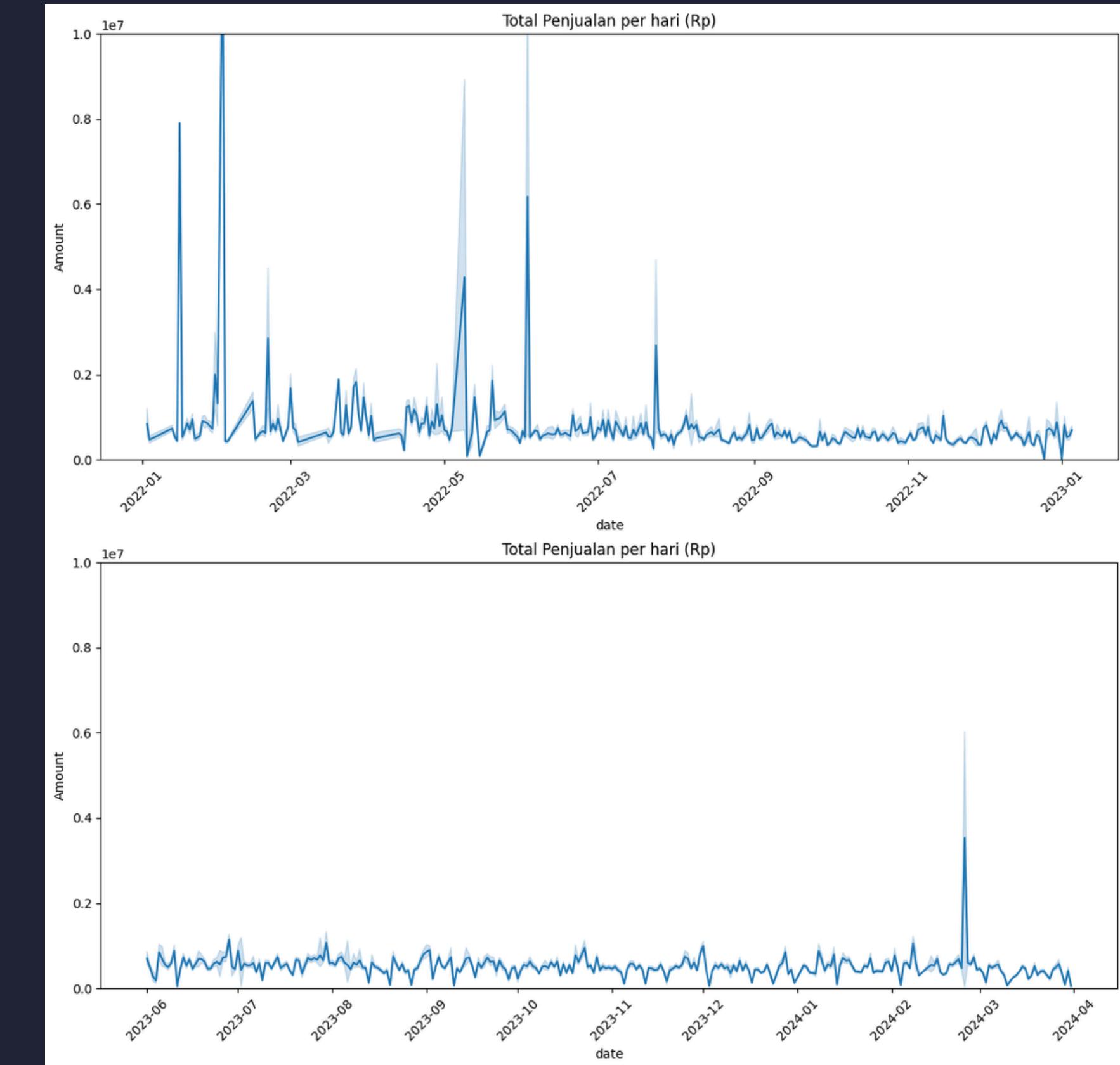
# EDA

## Banyak Customer per bulan

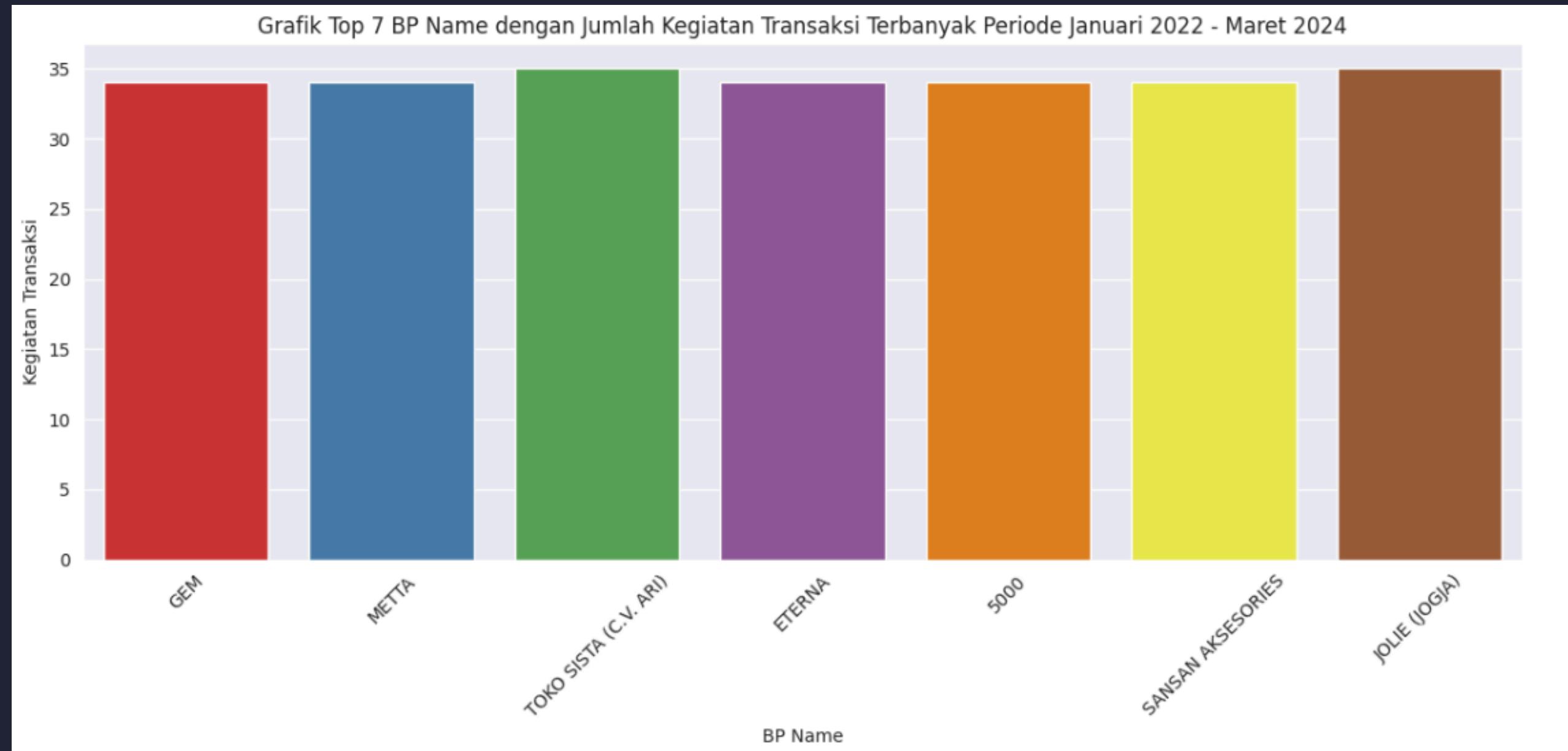


# EDA

## Total Penjualan Terhadap Tanggal Penjualan

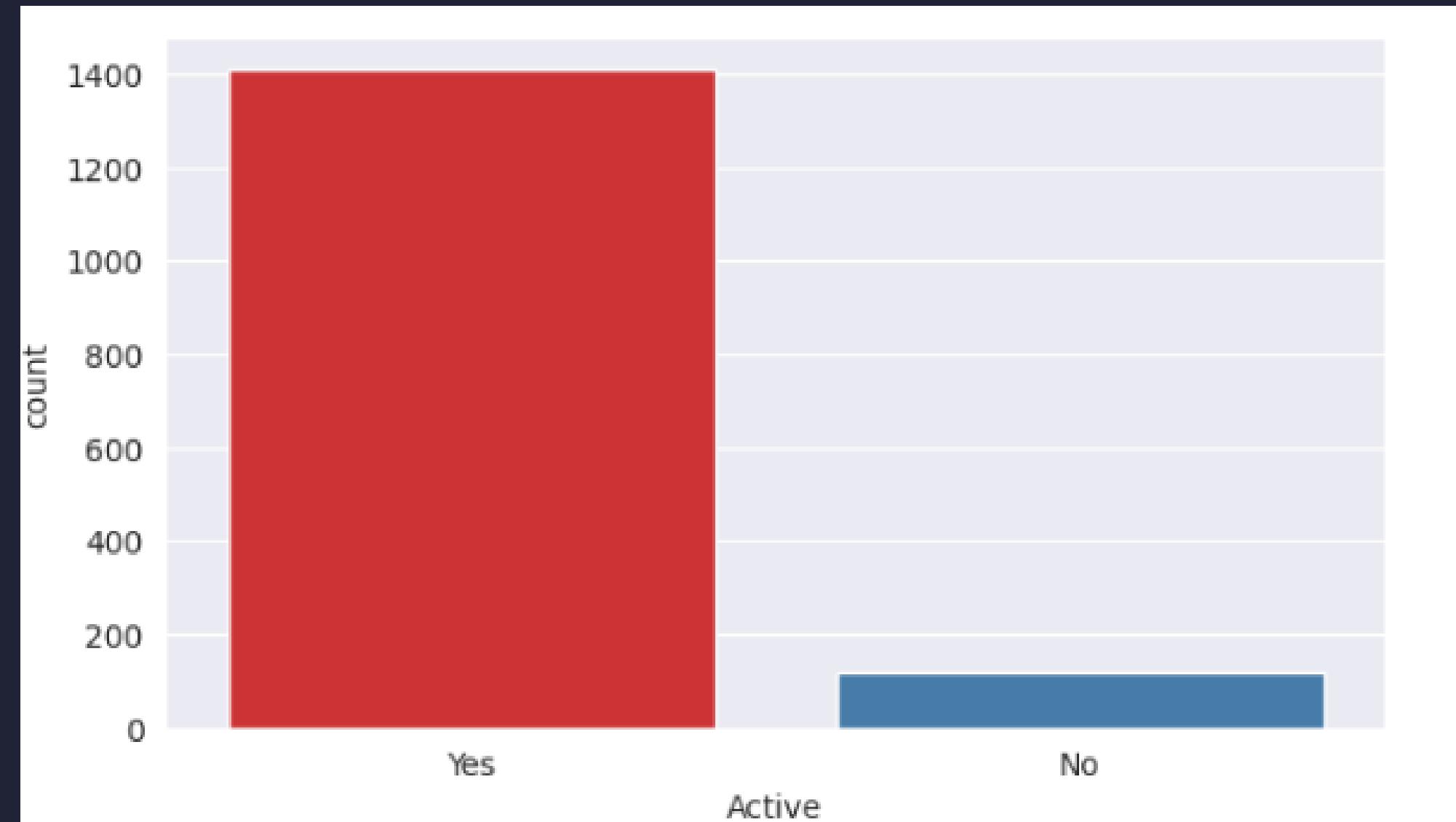


# EDA



# EDA

## Customers Active



# EDA

## Customers Shipping Type

