

# E-COMMERCE CUSTOMER CHURN

► **BY ADHITIA SEPTIAN**

Capstone Project Module 3

# TABLE OF CONTENTS

**OUTLINE****BUSINESS PROBLEM****DATA UNDERSTANDING****DATA CLEANING****DATA ANALYSIS****DATA PREPARATION****MODELING****CONCLUSION AND RECOMM.**

# OUTLINE

**BUSINESS  
PROBLEM**

**DATA  
UNDERSTANDING**

**DATA  
CLEANING**

**DATA  
ANALYSIS**

**DATA  
PREPARATION**

**MODELING**

**CONCLUSION &  
RECOMMENDATION**

# BUSSINESS PROBLEM

## ➤ Background

Customer churn atau tingkat keluar pelanggan adalah salah satu indikator kunci dari kesuksesan bisnis ecommerce. Tingkat keluar pelanggan yang tinggi dapat mengakibatkan kerugian finansial yang signifikan dan mengurangi citra merek.

## ➤ Analytic Approach

Pendekatan analitik yang dilakukan berupa pembuatan, evaluasi, dan implementasi model machine learning klasifikasi yang dapat memprediksi apabila pelanggan akan churn atau tidak berdasarkan riwayat data sebelumnya.

## ➤ Metrics Evaluation

Type 1 error : Konsekuensi : Biaya penawaran khusus yang dikeluarkan oleh perusahaan menjadi sia-sia.

Type 2 error : Konsekuensi : Hilangnya pelanggan loyal

Metrics Evaluation : F1 Score

## ➤ Problem Statements

Pelanggan meninggalkan atau menghentikan hubungan mereka dengan bisnis ecommerce, yang mengakibatkan hilangnya pendapatan dan pada akhirnya menurunnya pertumbuhan bisnis.

Faktor yang berkontribusi terhadap keputusan pelanggan untuk meninggalkan adalah pelayanan pelanggan yang buruk, kualitas produk yang rendah, harga yang tinggi, opsi pengiriman yang tidak nyaman atau tidak dapat diandalkan, dan persaingan dari bisnis ecommerce lainnya.

## ➤ Goals

Untuk meningkatkan retensi pelanggan dan mengurangi kehilangan pelanggan dalam bisnis ecommerce, sehingga bisnis dapat tumbuh dan berkembang dengan baik dalam jangka panjang. Dan juga untuk memahami penyebab churn atau kehilangan pelanggan dalam bisnis ecommerce dan menemukan cara untuk mengurangi tingkat churn tersebut.

# DATA UNDERSTANDING

DATASET 3941 DATA 11 KOLOM

ATTRIBUT	DATA TYPE	DESCRIPTION	VAR. TYPE
TENURE	FLOAT	MASA TENURE PELANGGAN	QUANT. KONTINU
WAREHOUSE TO HOME	FLOAT	JARAK ANTARA WAREHOUSE KE RUMAH PELANGGAN	QUANT. KONTINU
NUMBER OF DEVICE	INT	JUMLAH PERANGKAT YANG TERDAFTAR	QUANT. DISKRIT
PREFERED ORDER CAT	OBJECT	KATEGORI DISUKAI PELANGGAN BULAN LALU	QUAL. NOMINAL
SATISFACTION SCORE	INT	SKOR KEPUASAN PELANGGAN	QUAL. ORDINAL
MARITAL STATUS	OBJECT	STATUS PERNIKAHAN PELANGGAN	QUAL. NOMINAL
NUMBER OF ADDRESS	INT	JUMLAH ALAMAT YANG TERDAFTAR	QUANT. DISKRIT
COMPLAIN	INT	0 - TIDAK ADA COMPLAIN, 1 - ADA COMPLAIN	QUAL. NOMINAL
DAY SINCE LAST ORDER	FLOAT	HARI SEJAK PEMESANAN TERAKHIR	QUANT. KONTINU
CASHBACK AMOUNT	FLOAT	RATA RATA CASHBACK BULAN LALU	QUANT. KONITNU
CHURN	INT	0 - TIDAK CHURN, 1 - CHURN	QUAL. NOMINAL

# DATA CLEANING

## ➤ CEK DATA DUPLIKAT

Data duplikat : 672

Persentase Data Duplikat: 17.05

## ➤ CEK MISSING VALUE

1. Tenure sebanyak 5%
2. WarehouseToHome sebanyak 4%, dan
3. DaySinceLastOrder sebanyak 5%

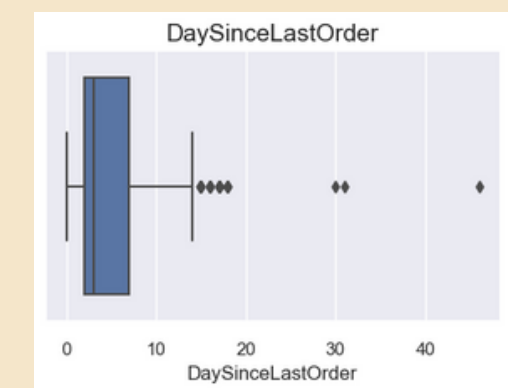
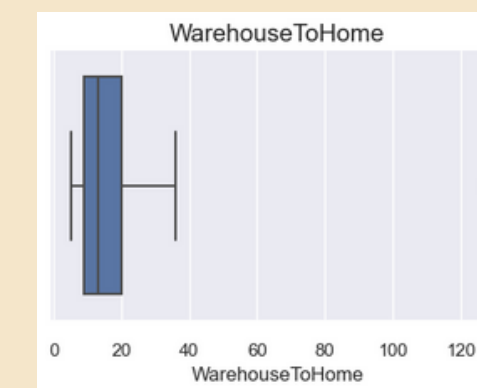
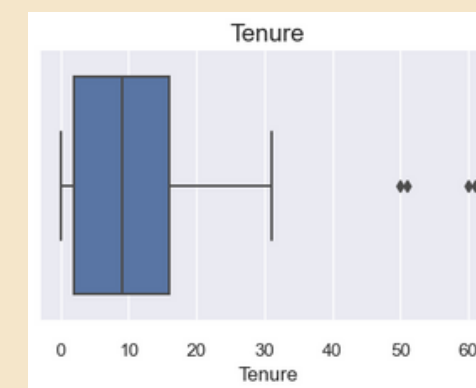
## ➤ CEK TIPE DATA

Tenure : float64

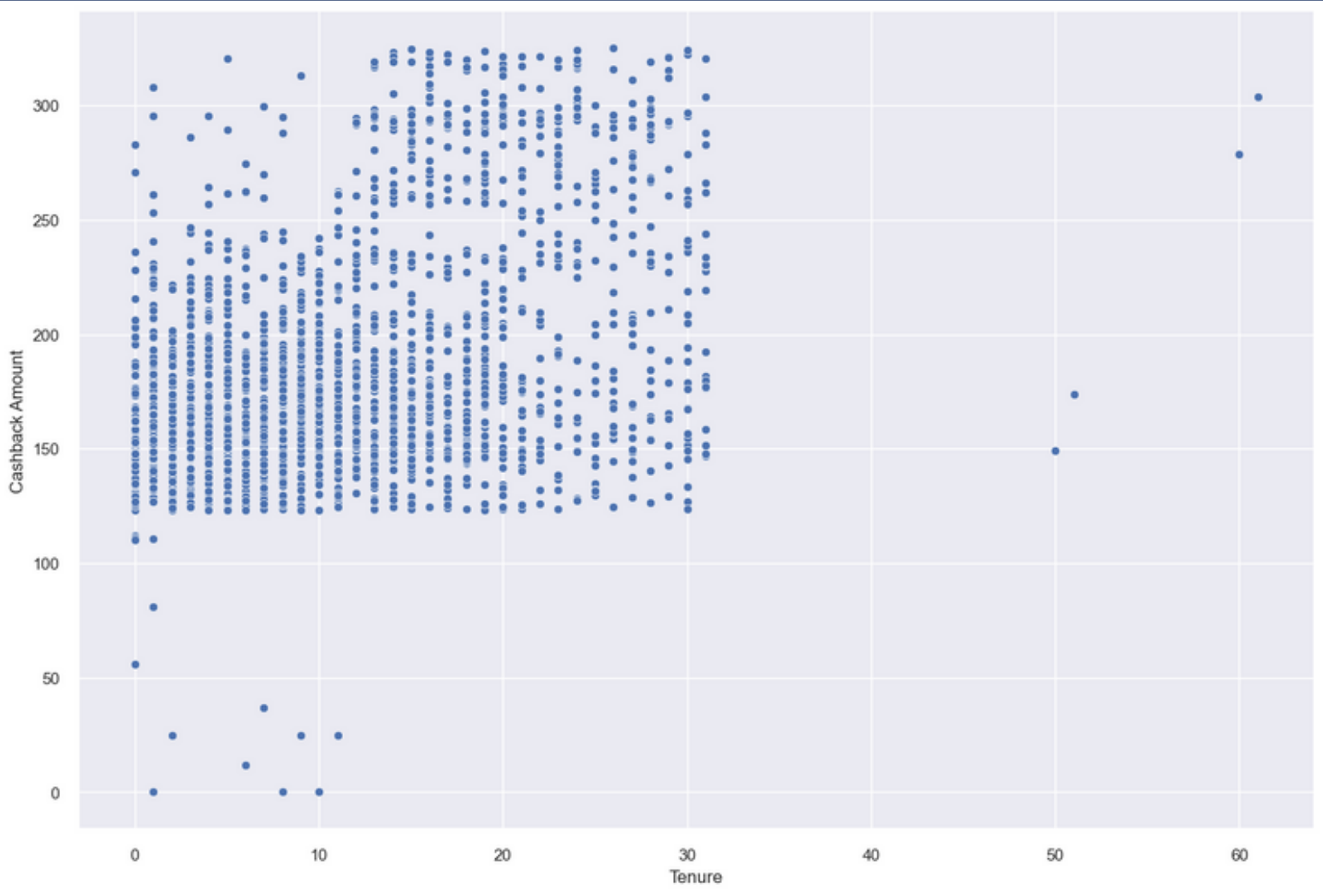
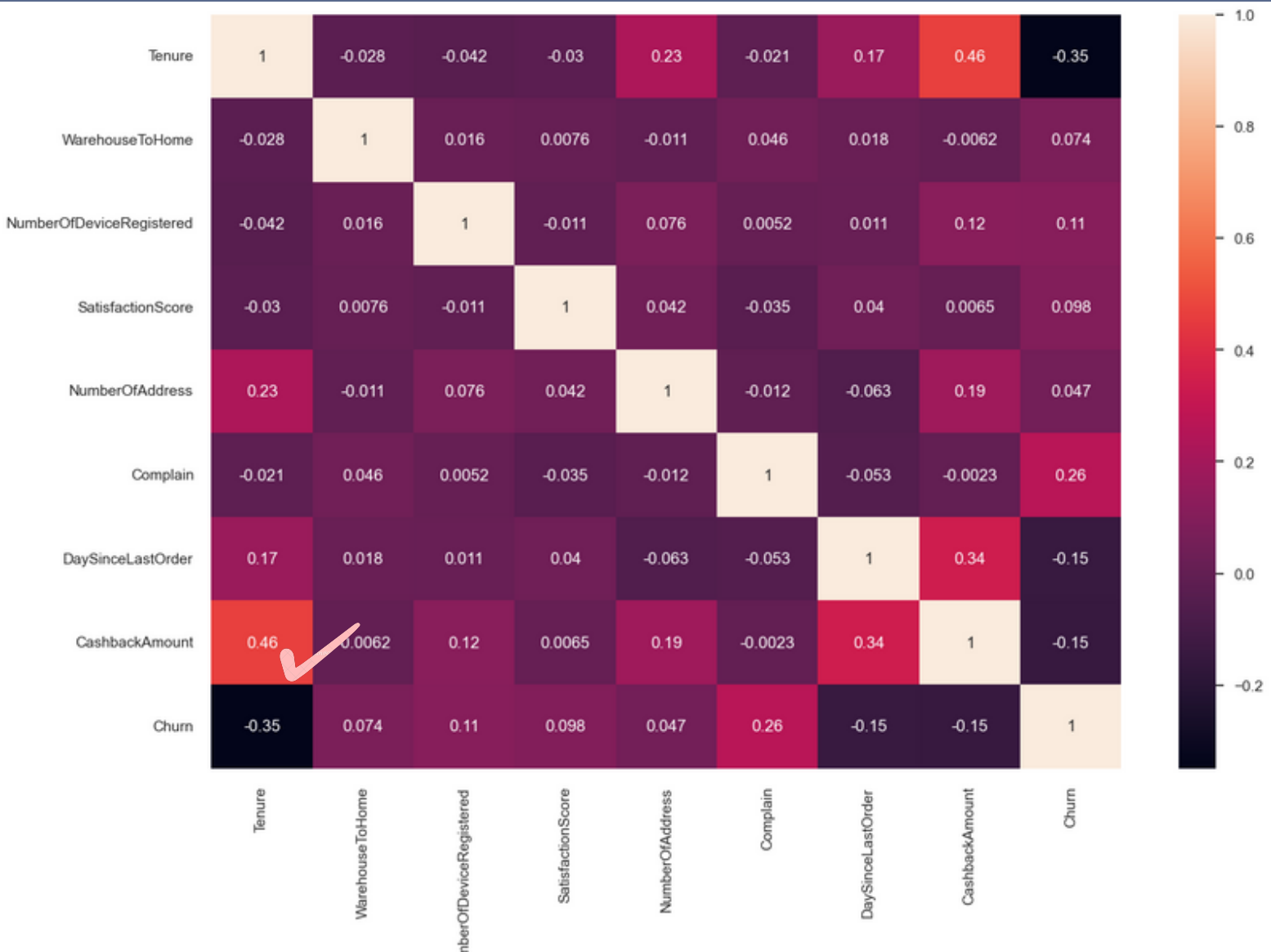
WarehouseToHome : float64

DaySinceLastOrder : float64

## ➤ CEK OUTLIERS

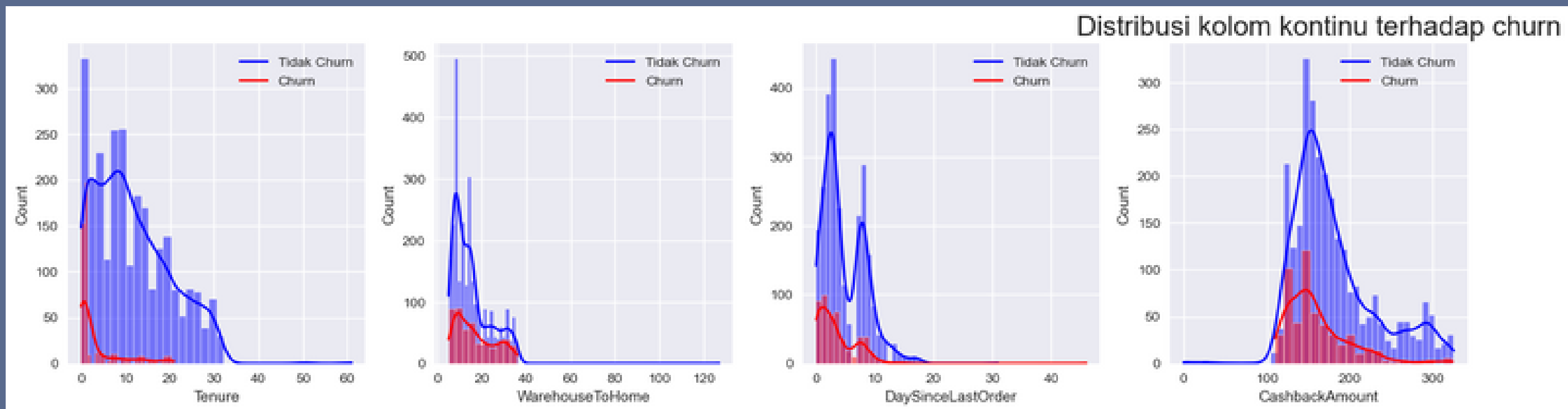
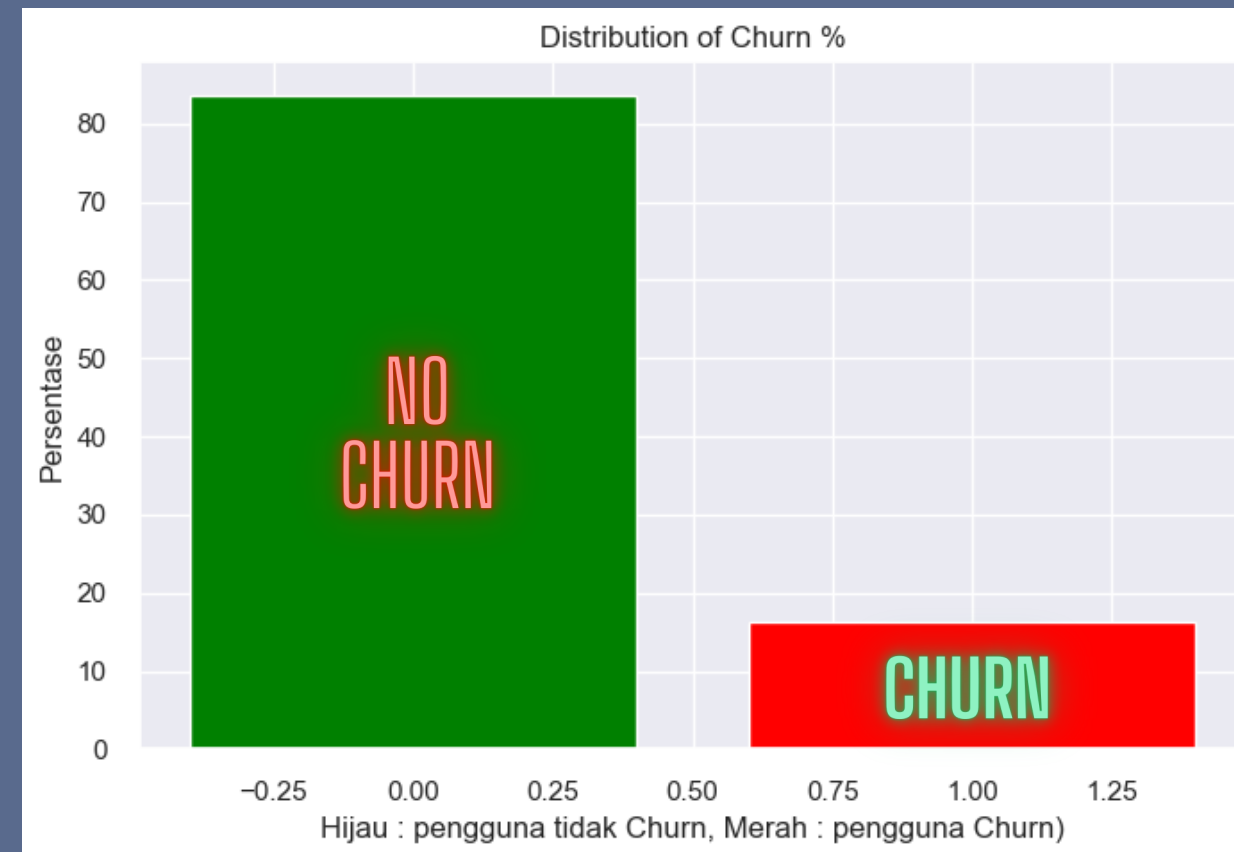


Heatmap



Scatterplot

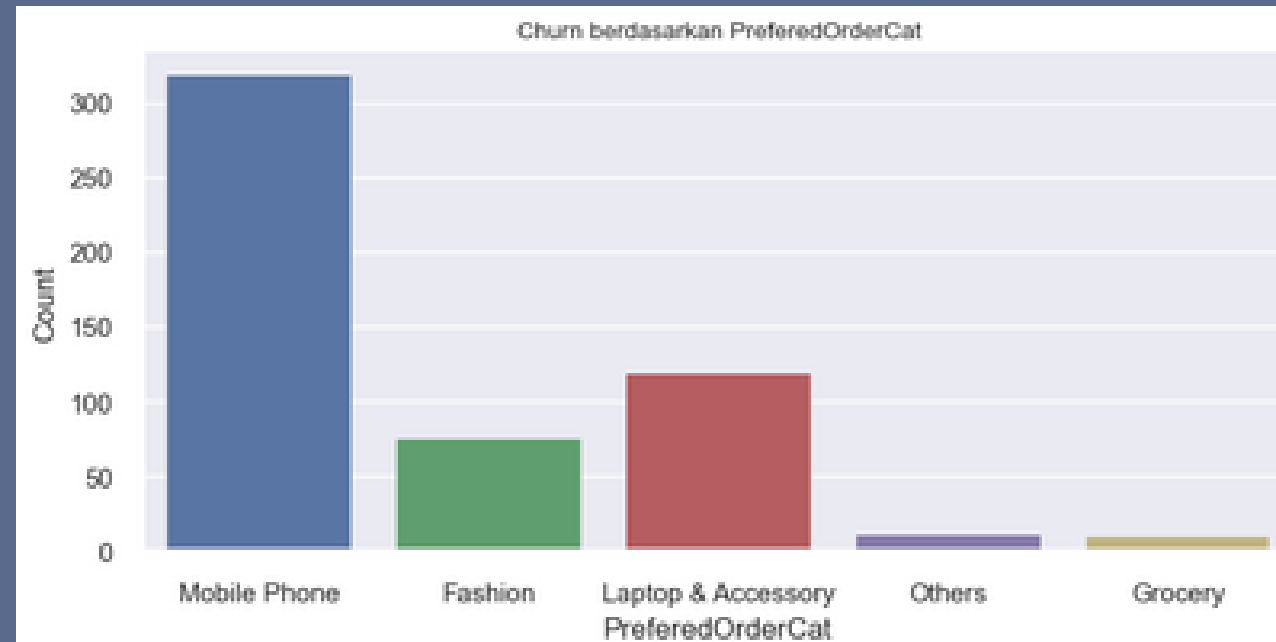
# DATA ANALYSIS



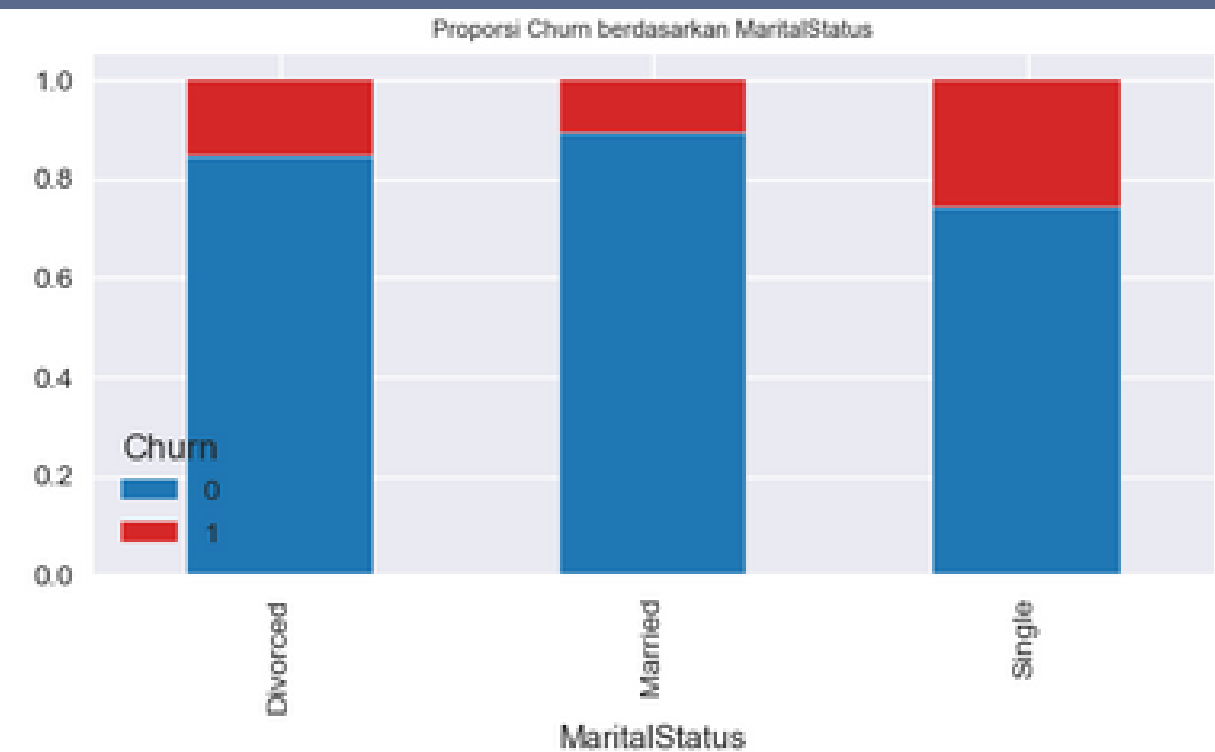
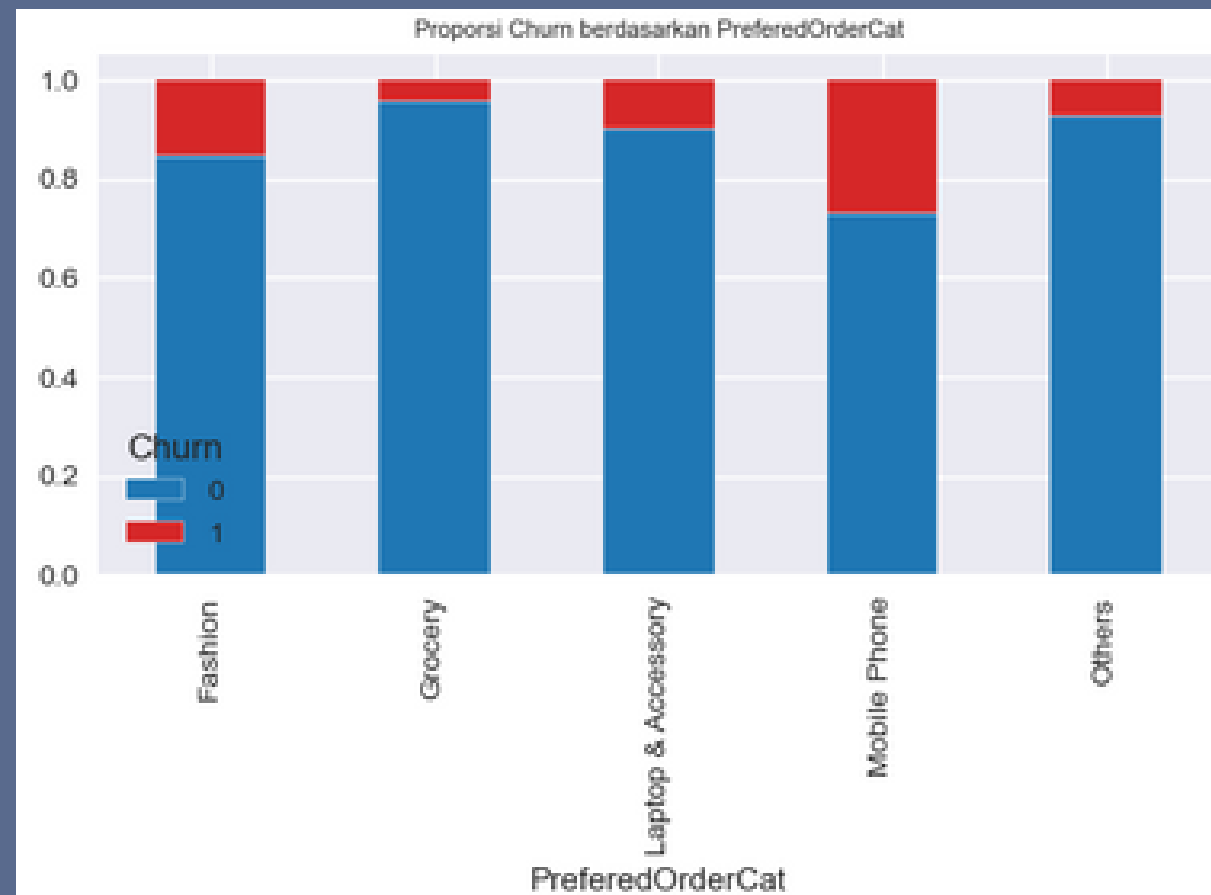
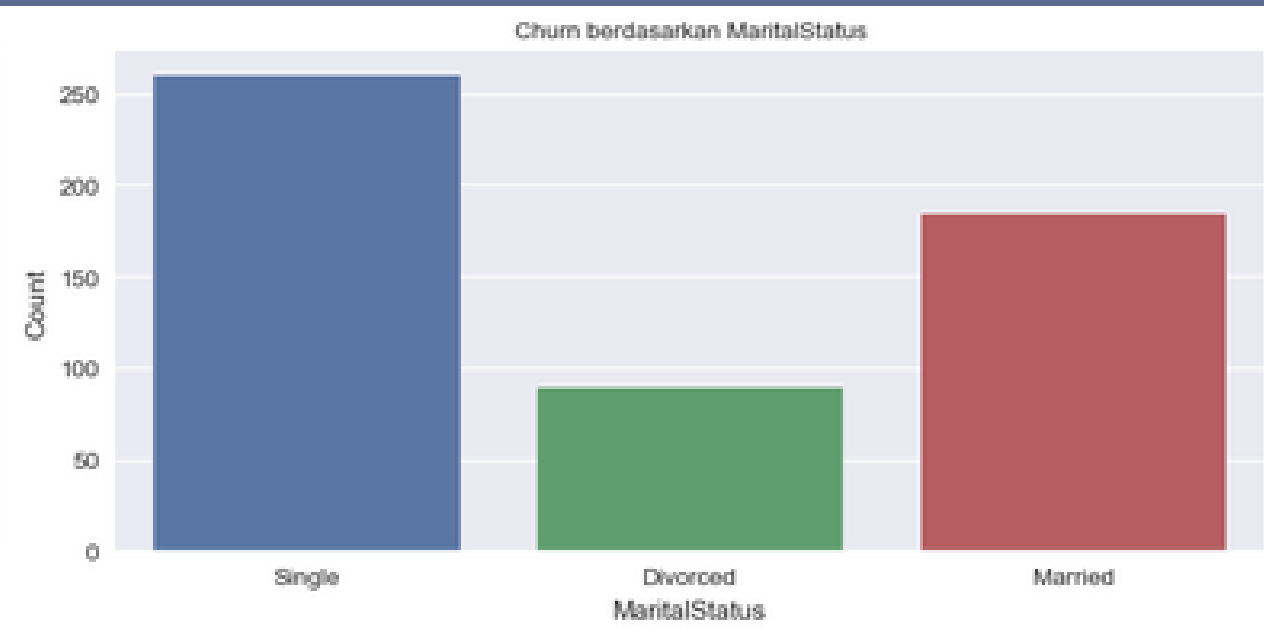


# DATA ANALYSIS

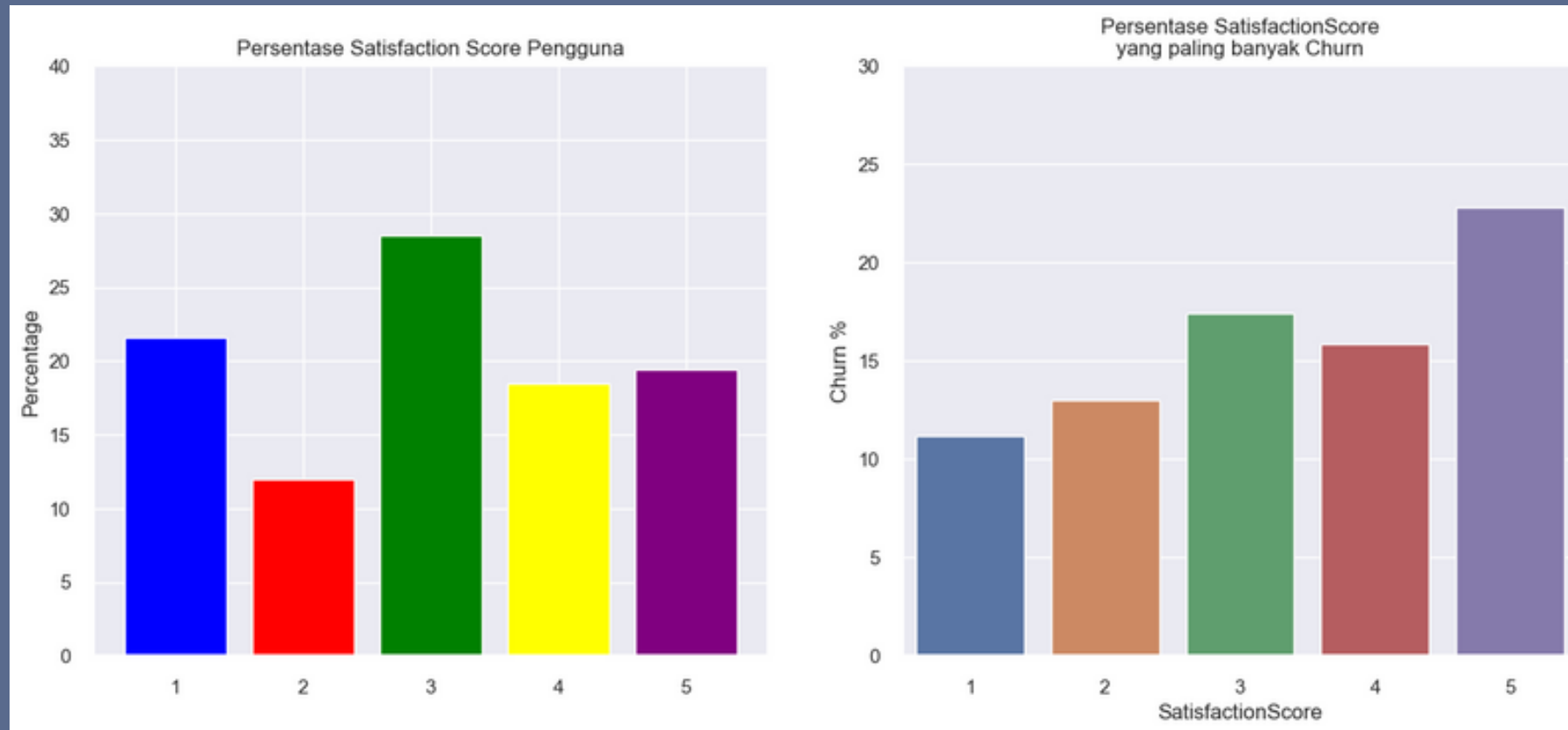
## Preferred Order Cat



## Marital Status

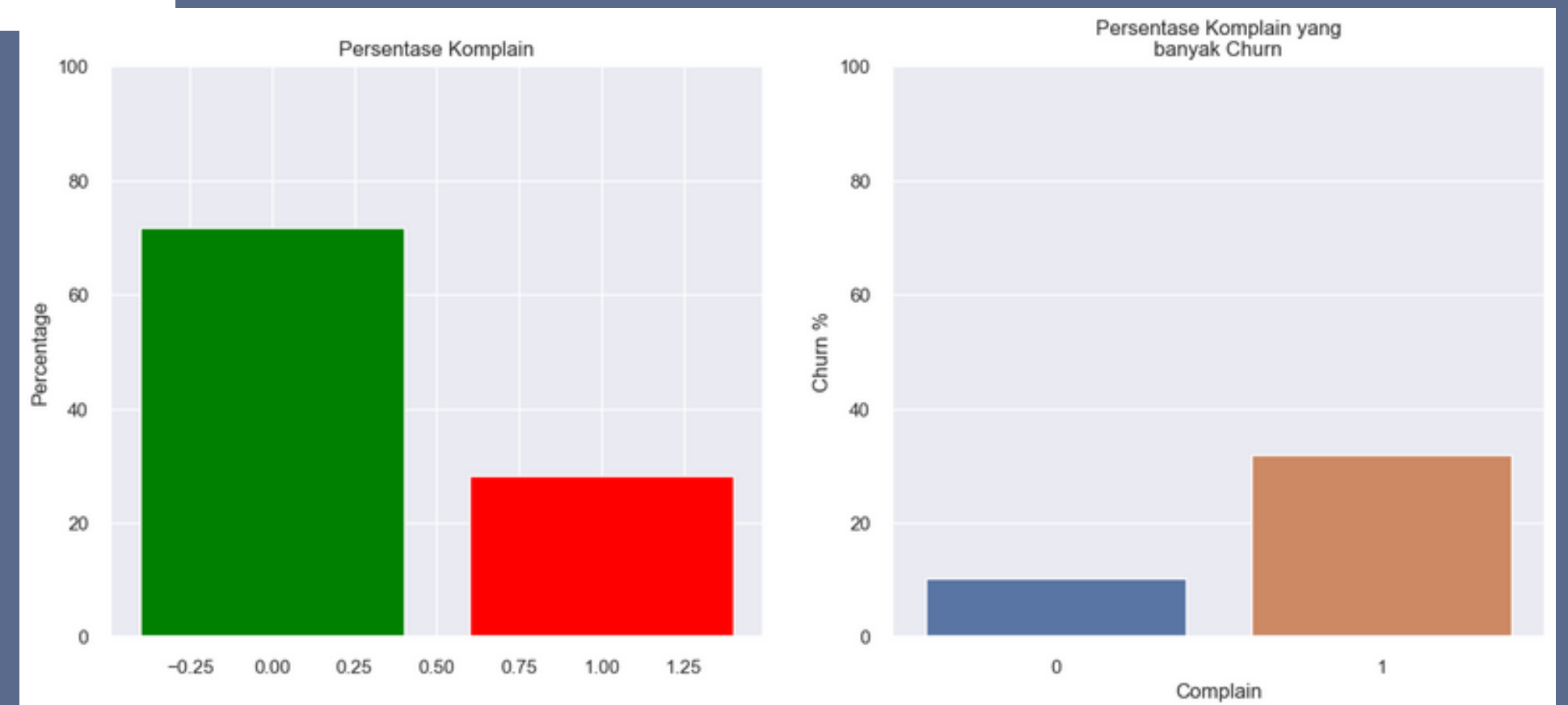


# DATA ANALYSIS



← Satisfaction Score

Komplain



# DATA PREPARATION

Preprocessing	Data
Missing Value	Simple Imputer Tenure, WarehouseToHome, DaySinceLastOrder
One Hot Encoder	Marital Status
Binary Encoder	PreferredOrderCat
Standard Scaler	Tenure, WarehouseToHome, NumberofAddress, DaySinceLastOrder, CashbackAmount

# BENCHMARK MODELING

ALGO	F1 MEAN	STD
XGBOOST CLASSIFIER	0.704	0.043
KNEIGHBORSCLASSIFIER	0.502	0.056
SVC	0.390	0.017
RANDOM FOREST CLASSIFIER	0.671	0.047
LOGISTIC REGRESSION	0.486	0.012

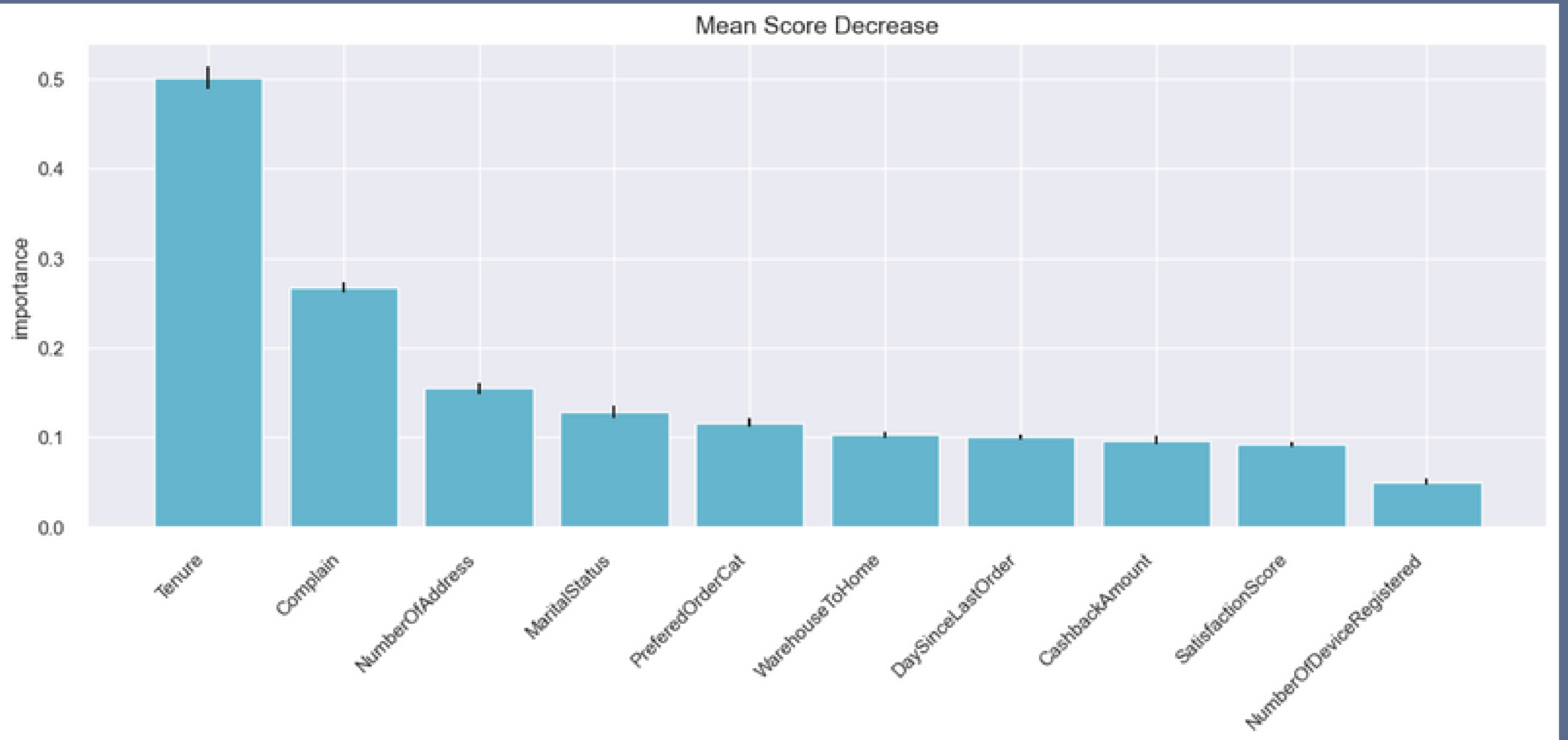
# HYPERPARAMETER TUNING

ALGO	F1 BEFORE	F1 AFTER
XGBOOST CLASSIFIER	0.704	0.705
KNEIGHBORSCLASSIFIER	0.502	0.570
SVC	0.390	0.566
RANDOM FOREST CLASSIFIER	0.671	0.684
LOGISTIC REGRESSION	0.486	0.550

# CLASSIFICATION REPORT

	PRECISION	RECALL	F1 SCORE	SUPPORT
0	0.96	0.95	0.95	547
1	0.75	0.79	0.77	107
ACCURACY			0.92	654
MACRO AVG	0.86	0.87	0.86	654
WEIGHTED AVG	0.93	0.92	0.92	654

# FEATURE IMPORTANCES



# CONCLUSION

Berdasarkan hasil diatas dapat disimpulkan bahwa dengan prediksi menggunakan model yang sudah dibuat, model dapat mengetahui 95% dari pelanggan yang tidak churn dan 79% pelanggan yang churn berdasarkan recall. selain itu model memiliki kemungkinan prediksi benar untuk pelanggan yang akan churn sebesar 75%. Maka masih ada pelanggan yang churn dan di prediksi sebagai tidak churn sebesar 4%.

Dengan adanya bantuan model, perusahaan e-commerce dapat menghemat biaya yang cukup besar dengan melakukan pemasaran yang lebih tepat sasaran.

Selain itu, berdasarkan analisa faktor yang berpengaruh dengan customer churn adalah complain dan masa tenure pelanggan. Hal tersebut harus diperhatikan, apabila pelanggan melakukan complain dan semakin kecil masa tenure pelanggan, ada potensi customer churn.



# RECOMMENDATION

## Business:

- Memberikan penawaran
- Memperhatikan complain yang diberikan pelanggan
- Memberikan penawaran pemotongan ongkos kirim
- Menggunakan model machine learning yang telah dibuat sebagai solusi untuk menentukan strategi pemasaran yang tepat sasaran

## Model:

- Memiliki kolom customer ID untuk memastikan tidak adanya data yang duplikat, serta memastikan data yang didapat tidak ada missing value
- Adanya kolom dengan fitur lain yang berhubungan langsung dengan servis yang perusahaan e-commerce berikan dan transaksi pelanggan
- Model dapat ditingkatkan dengan melakukan tuning hyperparameter
- Meningkatkan nilai recall dengan tuning threshold F1 Score
- Memilah fitur sesuai dengan feature importances dan melakukan trial kombinasi fitur

# THANK YOU