

Logistic Delivery Prediction

Kelompok 3

Cintana Dharma Kirana

(2012310004)

Mazarina Putri Maulidya I S

(2012310010)

Ria Utari Putri M Jaya

(2012310019)

Table of Contents

 **Current Problems**

 **Our Goals**

 **Our Findings**

Why Logistic Delivery Prediction?

Departemen logistik berupaya membangun model prediktif untuk mengklasifikasikan apakah pengiriman akan tepat waktu atau terlambat, berdasarkan data historis seperti metode pengiriman, lokasi gudang, panggilan pelanggan, tingkat kepentingan produk, diskon, dan berat barang. Model ini membantu perusahaan mengidentifikasi potensi keterlambatan lebih awal, meningkatkan efisiensi pengiriman, serta mendorong kepuasan pelanggan melalui layanan yang lebih tepat waktu dan andal.



← → Q What is the current problem?

What is the current problem?

- Perusahaan belum memiliki sistem yang mampu memprediksi keterlambatan pengiriman secara akurat.
- Keterlambatan sering terjadi tanpa terdeteksi lebih awal, sehingga mengganggu efisiensi dan kepuasan pelanggan.
- Diperlukan model prediksi untuk membantu mengantisipasi pengiriman yang berisiko terlambat.



Our Goals

- Mengetahui karakteristik pengiriman
- Menentukan variabel-variabel yang berpengaruh signifikan terhadap ketepatan waktu pengiriman
- Memprediksi apakah pengiriman akan sampai tepat waktu atau mengalami keterlambatan



← → Q Our Strategies to achieve goals

Our Strategies to achieve goals

Exploratory Data Analysis (EDA)

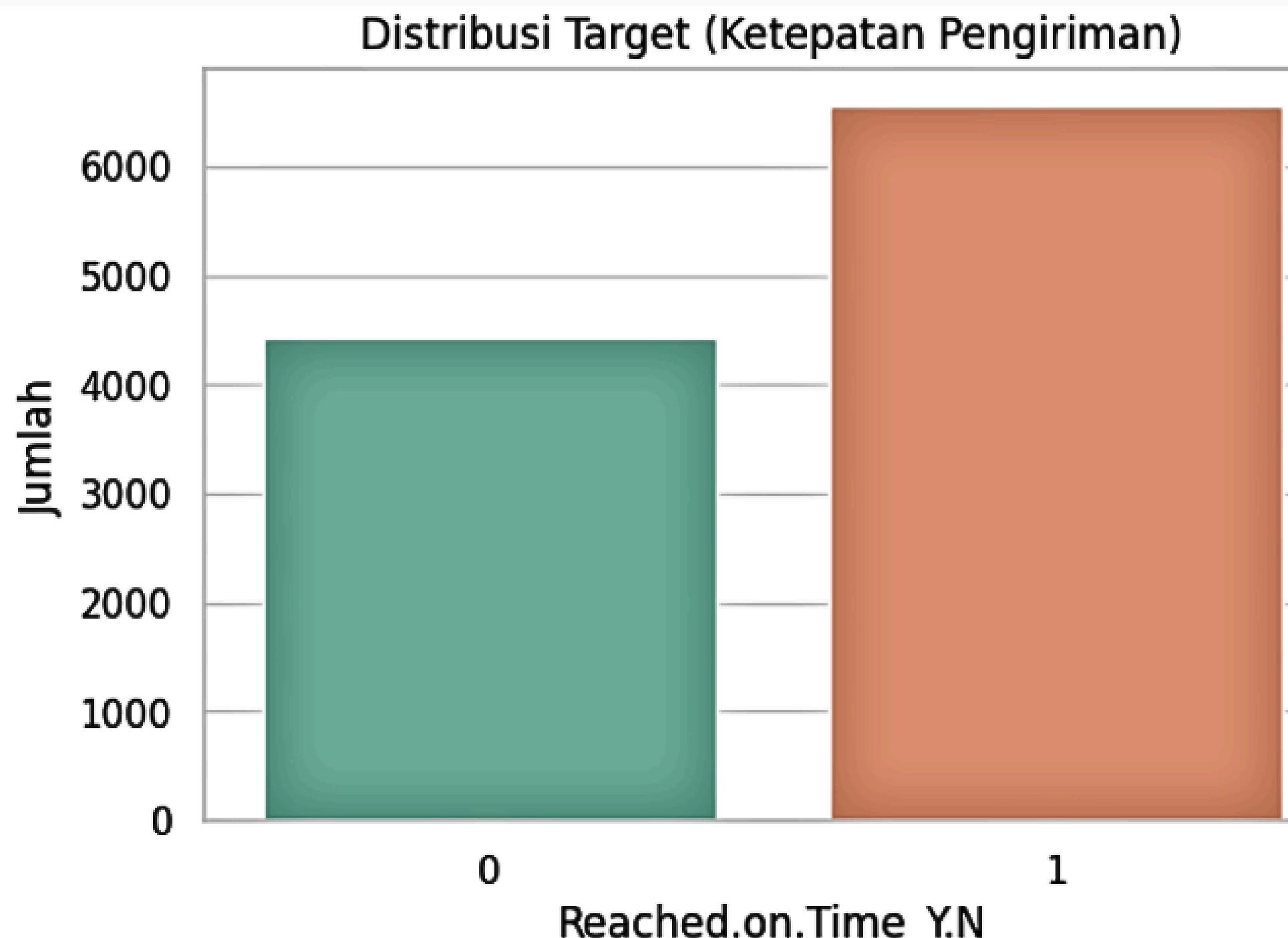
Data Preprocessing

Modelling

→ Q DATASET

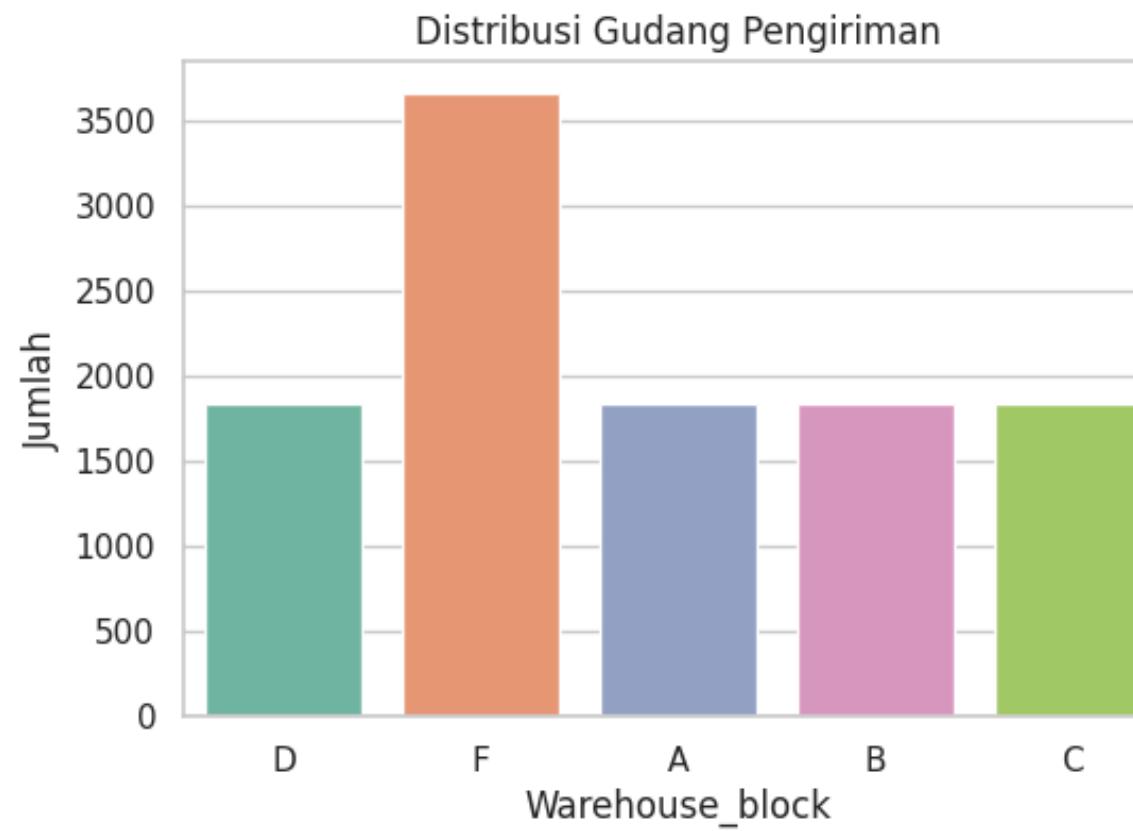
| Nama Kolumn | Tipe Data | Deskripsi |
|----------------------------|---------------|---|
| ID | String / Int | ID unik untuk setiap pengiriman. |
| Warehouse_block | Kategorikal | Blok gudang asal pengiriman (contoh: A, B, C, D, F). |
| Mode_of_Shipment | Kategorikal | Moda pengiriman: Road, Ship, atau Flight. |
| Customer_care_calls | Numerik (Int) | Jumlah panggilan dari pelanggan ke layanan pelanggan. |
| Customer_rating | Numerik (Int) | Penilaian pelanggan (skala 1–5). |
| Cost_of_the_Product | Numerik | Harga produk yang dikirim. |
| Prior_purchases | Numerik (Int) | Jumlah pembelian sebelumnya oleh pelanggan. |
| Product_importance | Kategorikal | Tingkat kepentingan produk: Low, Medium, atau High. |
| Gender | Kategorikal | Jenis kelamin pelanggan: Male atau Female. |
| Discount_offered | Numerik | Besarnya diskon yang diberikan pada produk. |
| Weight_in_gms | Numerik | Berat produk dalam satuan gram. |
| Reached.on.Time_Y.N | Target (0/1) | Status pengiriman: 1 = Tepat waktu, 0 = Terlambat. |

EDA -Univariate Data Analysis



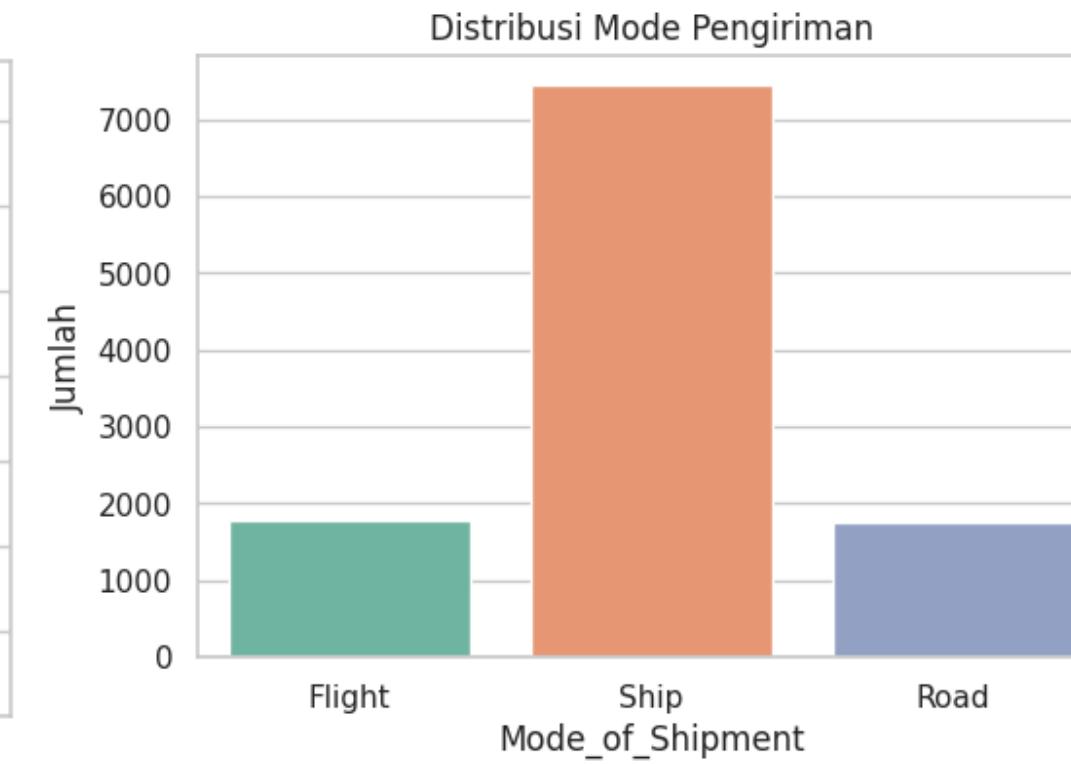
Sekitar 59% pengiriman berhasil sampai tepat waktu, sedangkan 41% sisanya mengalami keterlambatan.

EDA -Univariate Data Analysis

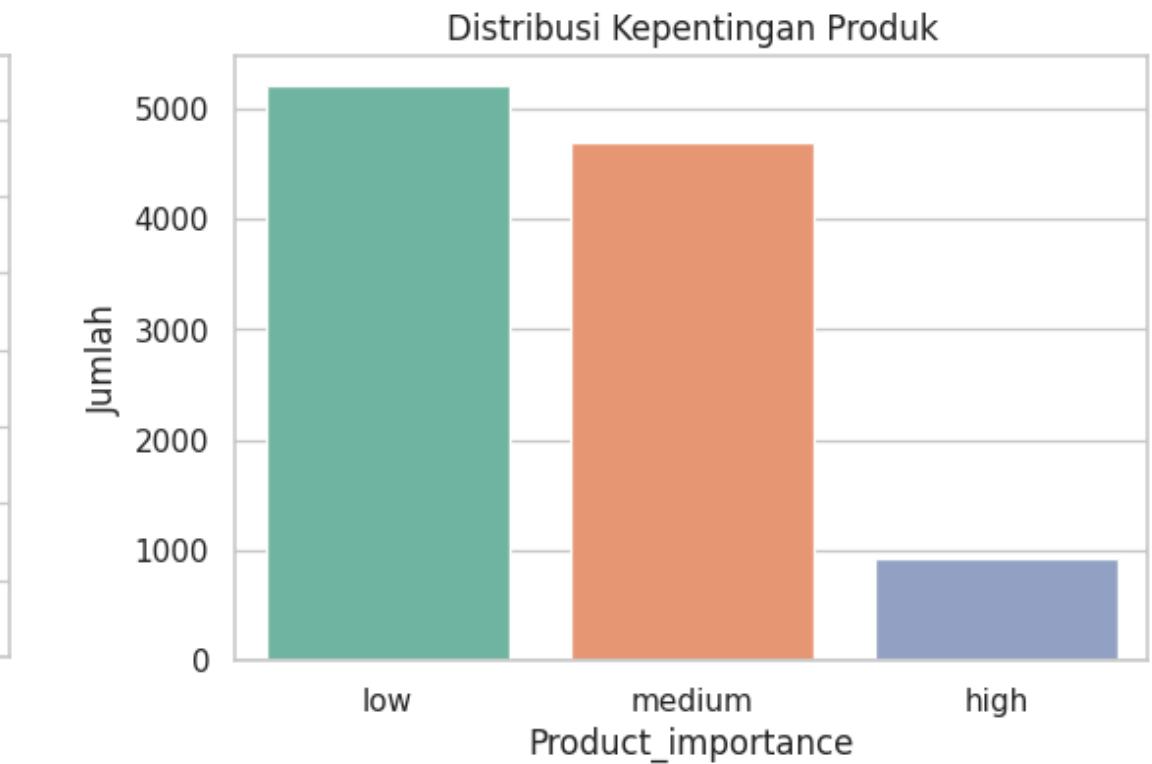


Bar chart menunjukkan distribusi pengiriman dari gudang blok A, B, C, D, E, F.

Blok F paling banyak digunakan.

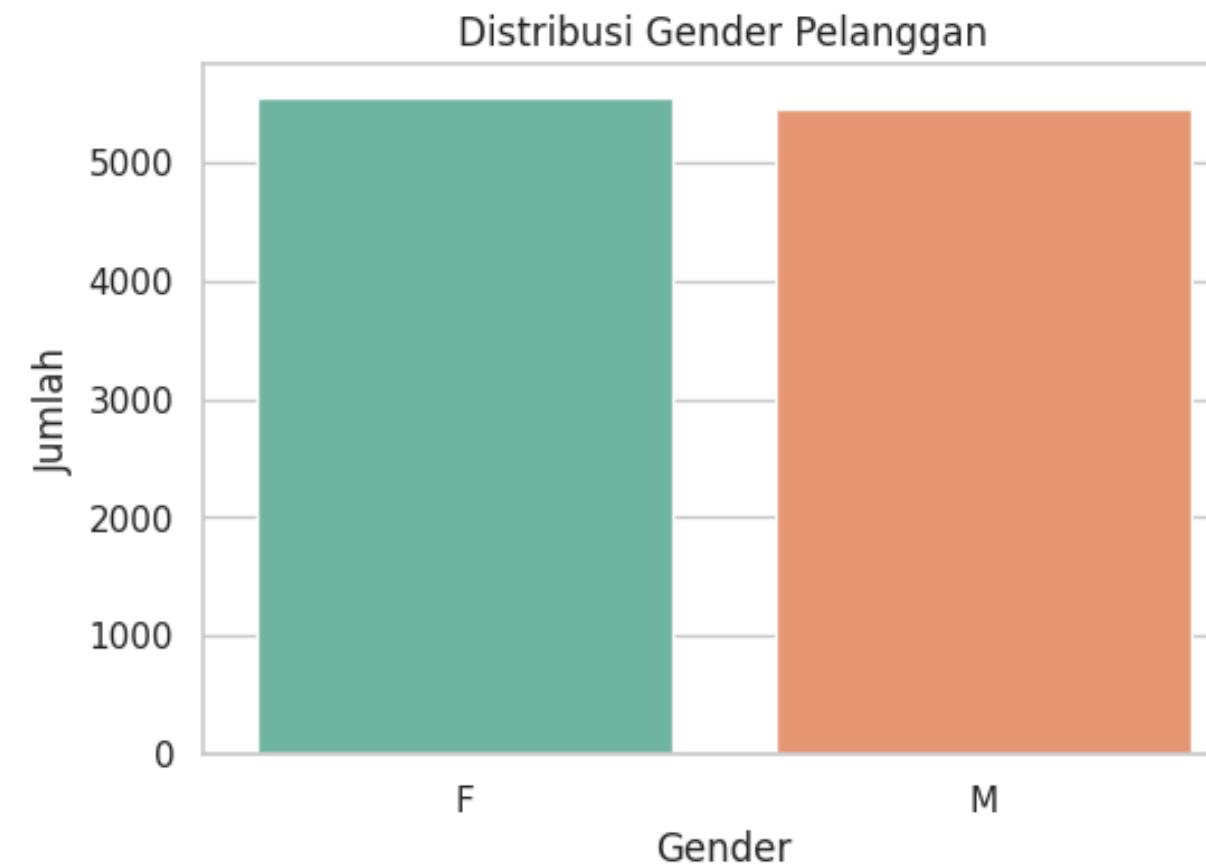


Bar chart yang menunjukkan mayoritas pengiriman menggunakan kapal (Ship).

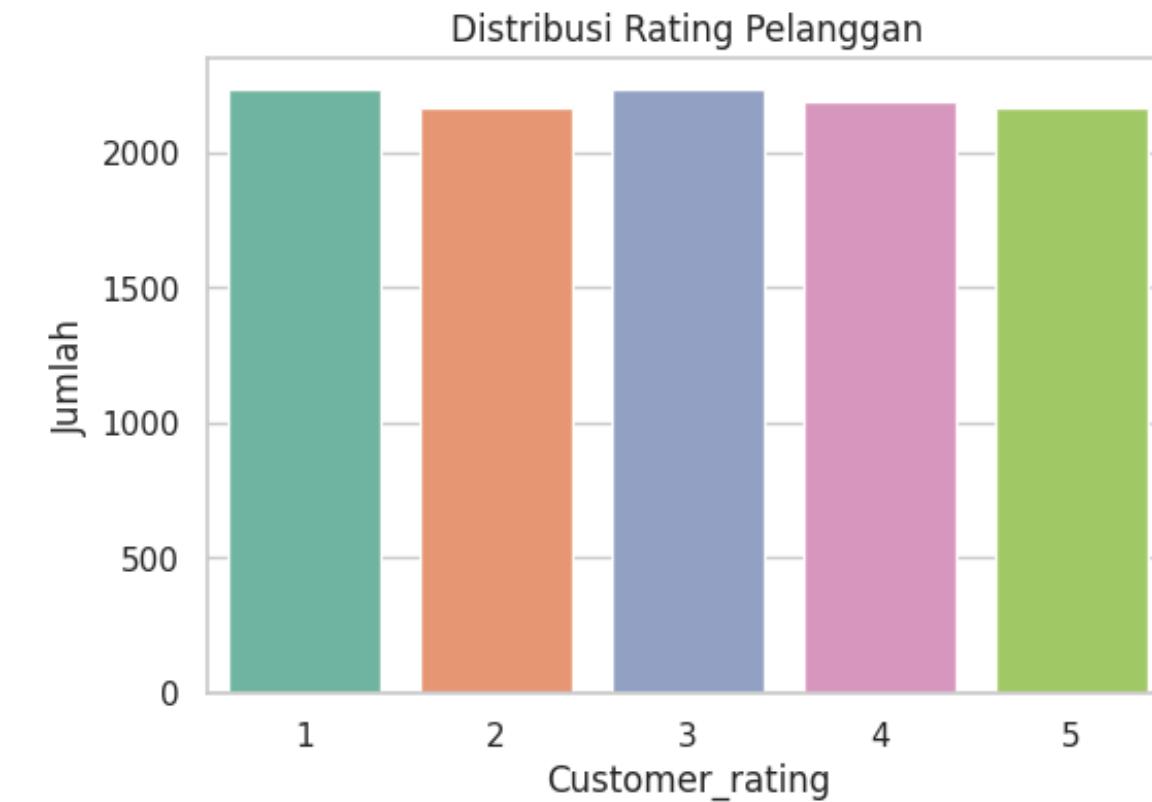


Produk dengan low importance paling banyak.

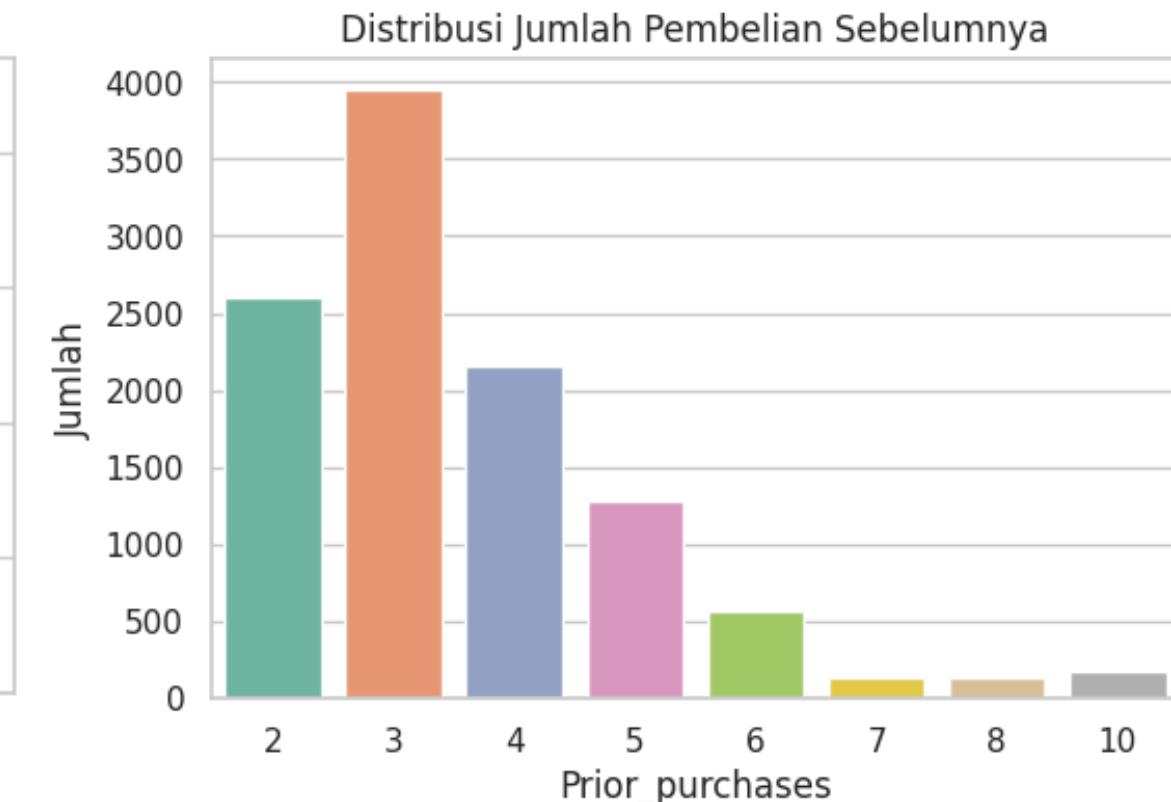
EDA -Univariate Data Analysis



Jumlah pelanggan perempuan dan laki-laki hampir seimbang, dengan selisih yang sangat kecil di antara keduanya.

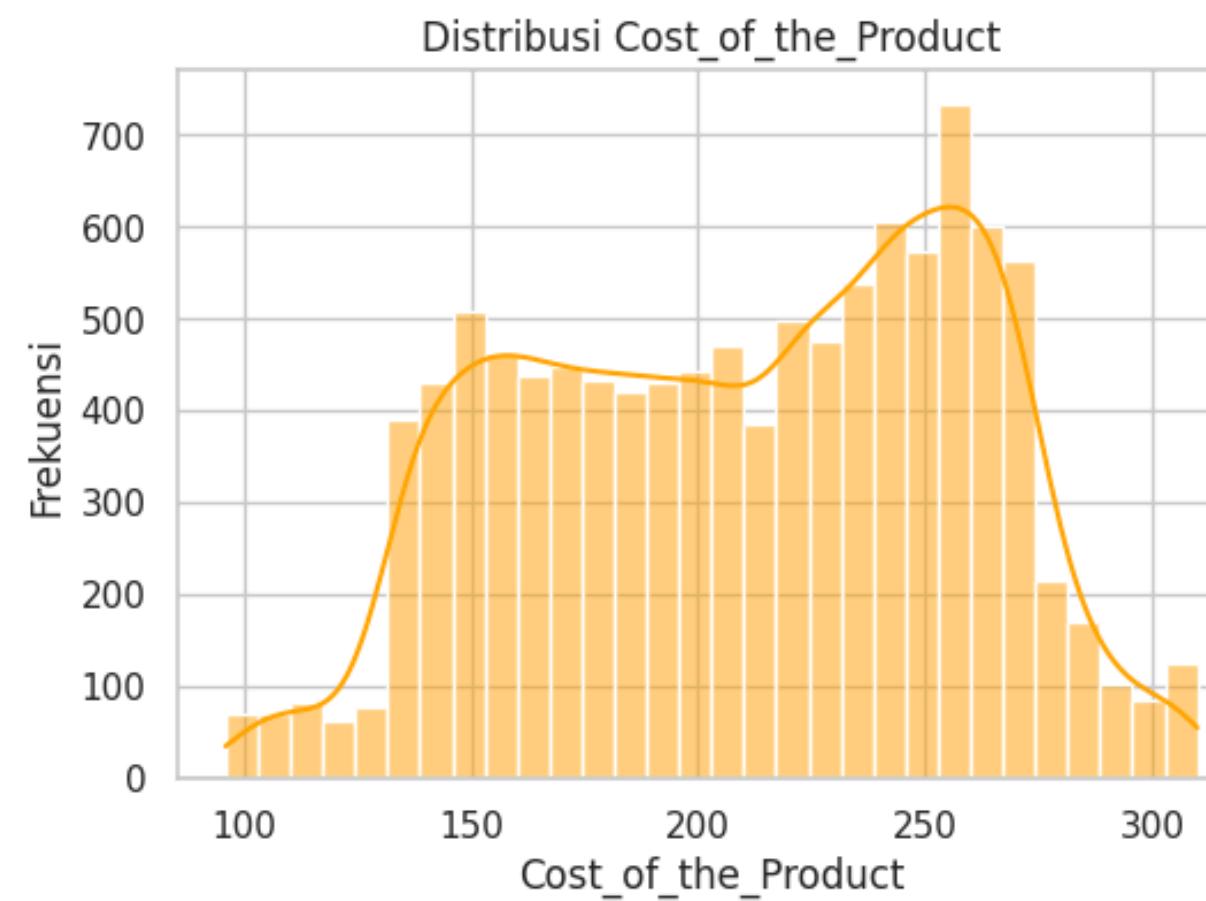


Rating dari 1 sampai 5 relatif merata, sedikit lebih banyak rating 1 dan 3.

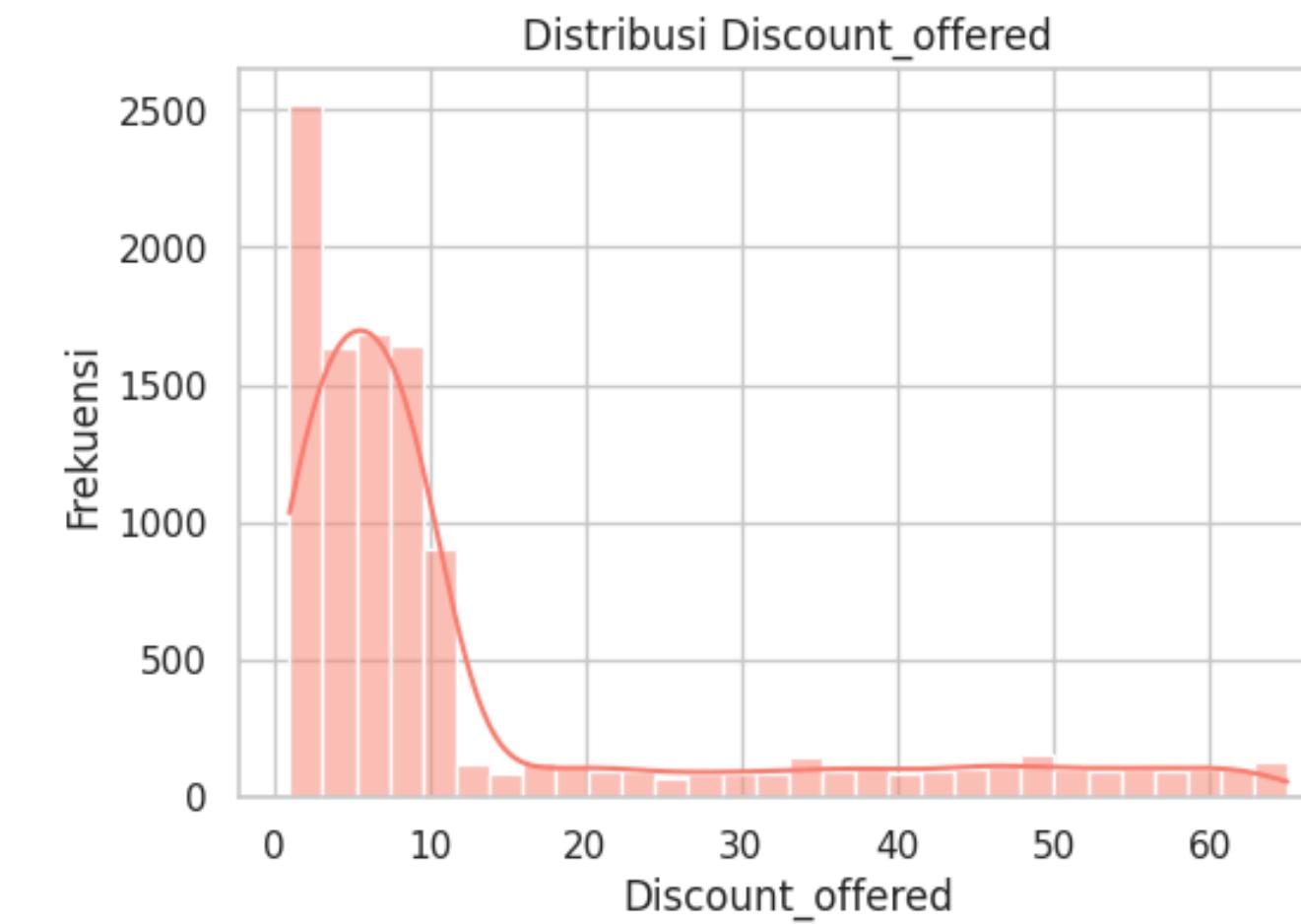


Banyak pelanggan sebelumnya melakukan 2 atau 3 pembelian.

EDA -Univariate Data Analysis

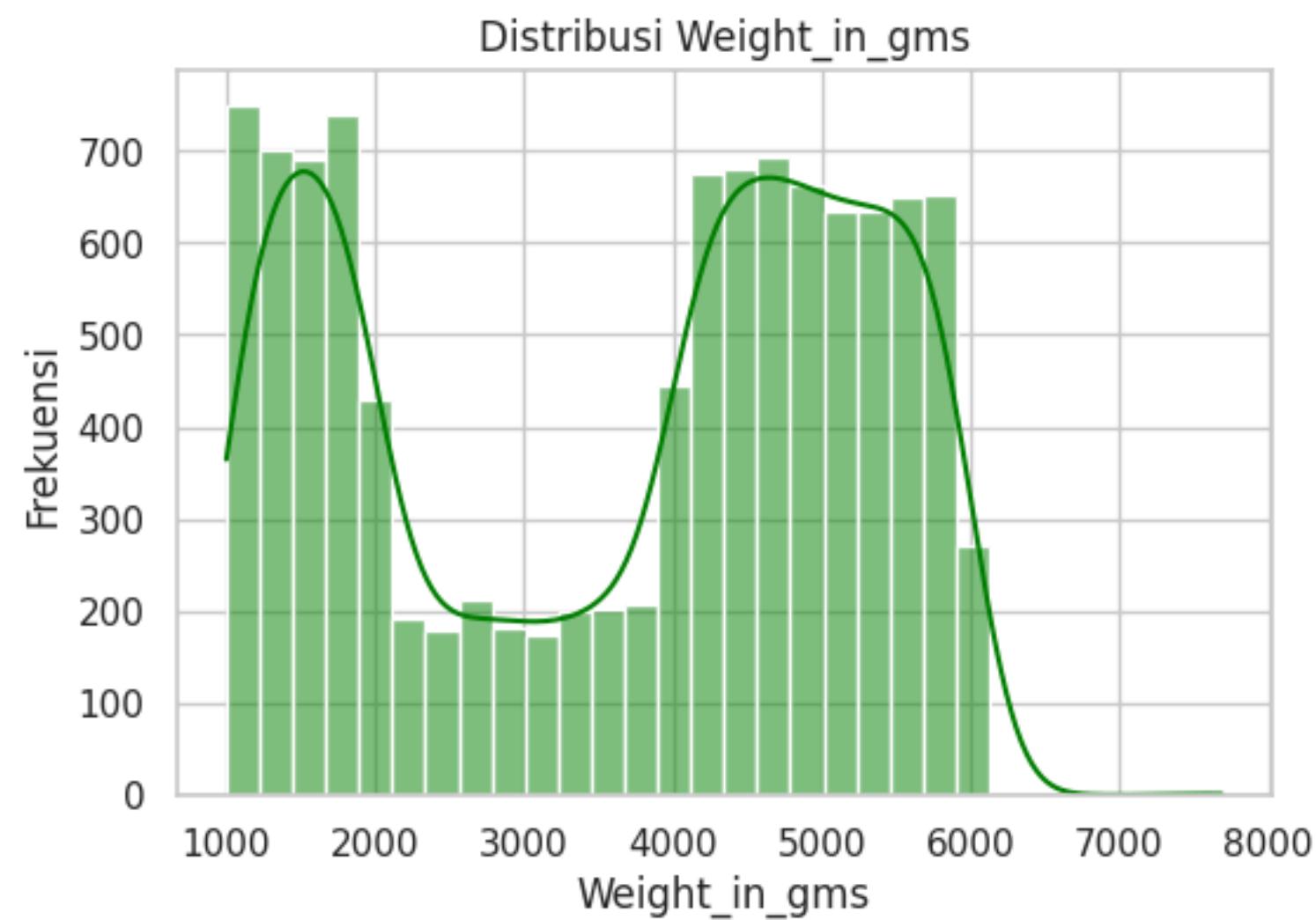


Biaya produk paling banyak berkisar antara 150-275.

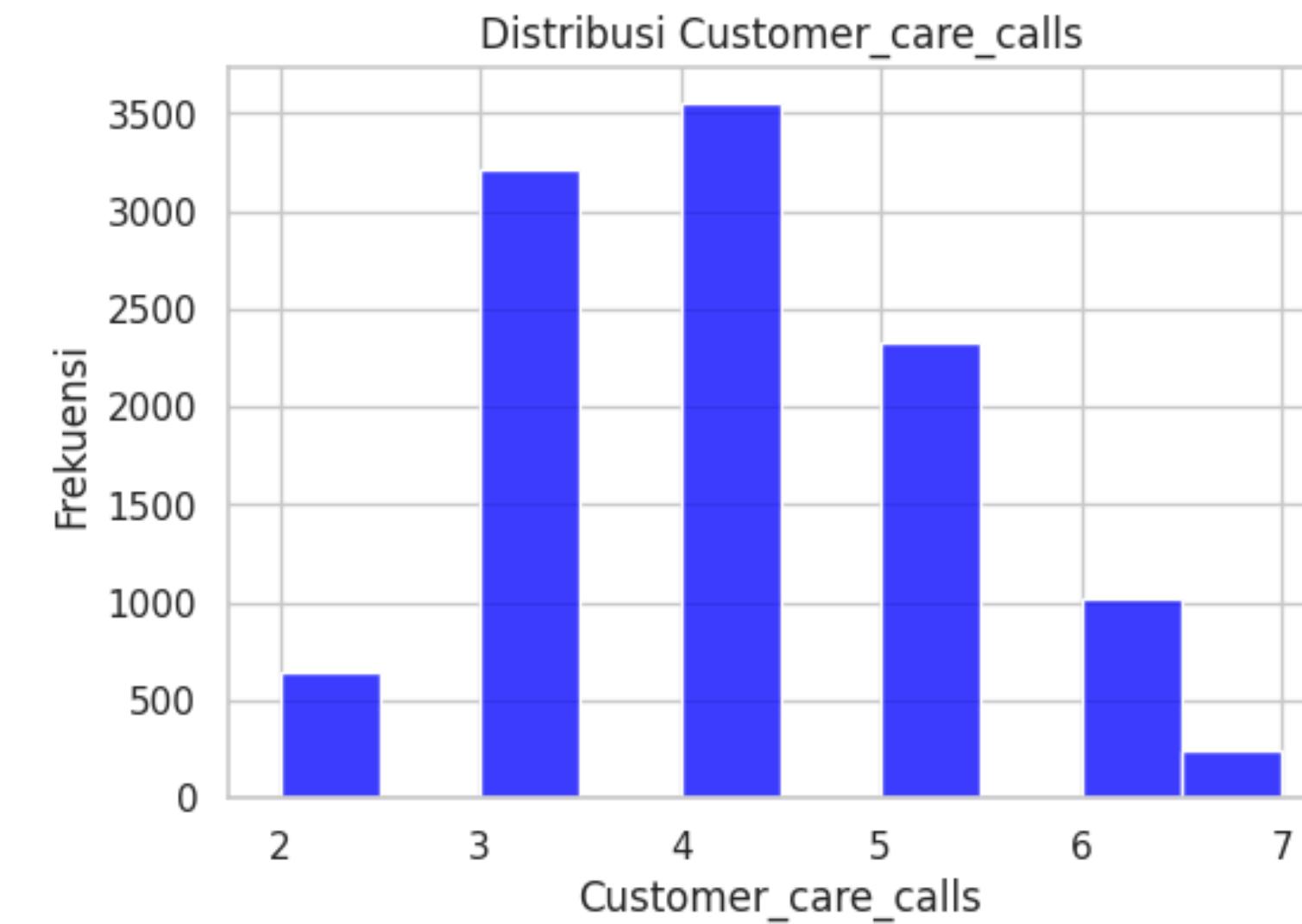


Histogram diskon yang ditawarkan, kebanyakan berada di bawah 10%.

EDA -Univariate Data Analysis

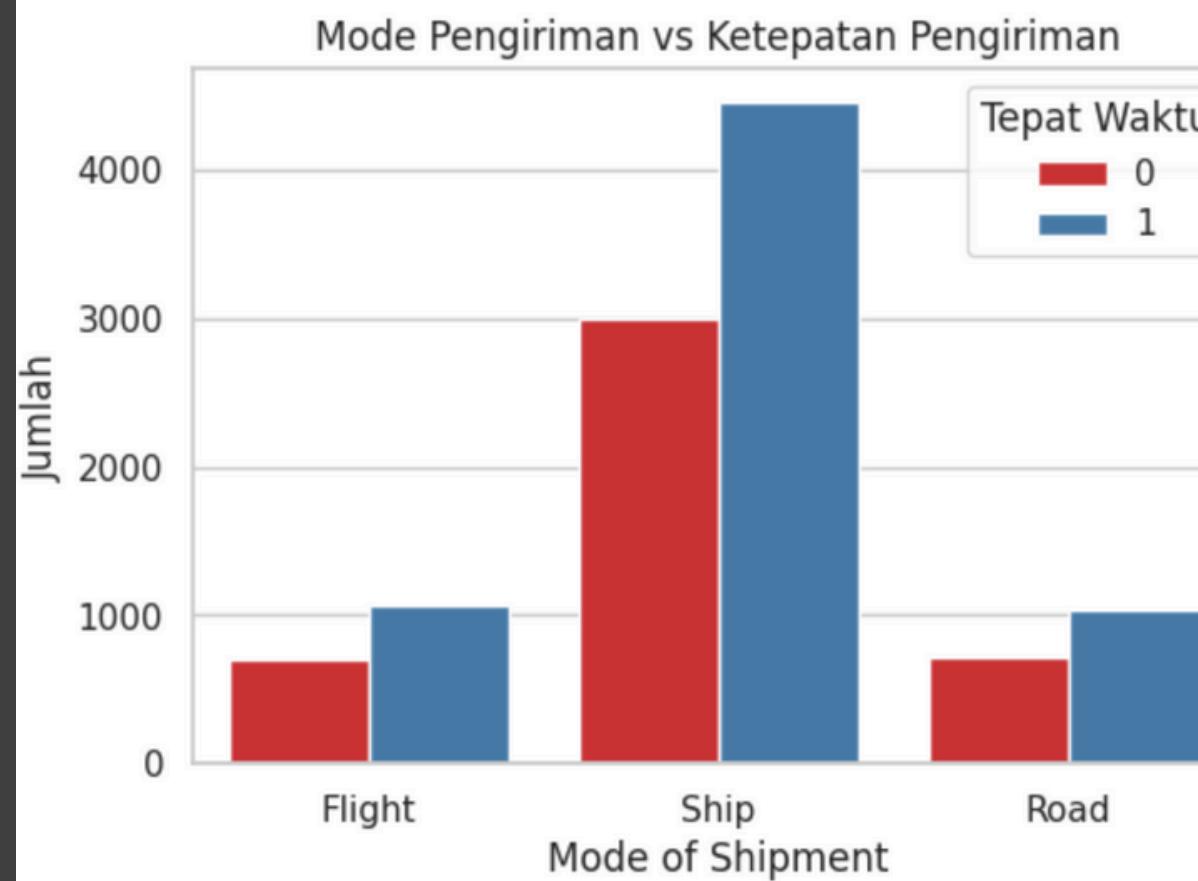


Ada 2 tipe produk utama: ringan (sekitar 1.000 - 2.000 gram) dan berat (sekitar 4.000-6.000 gram)

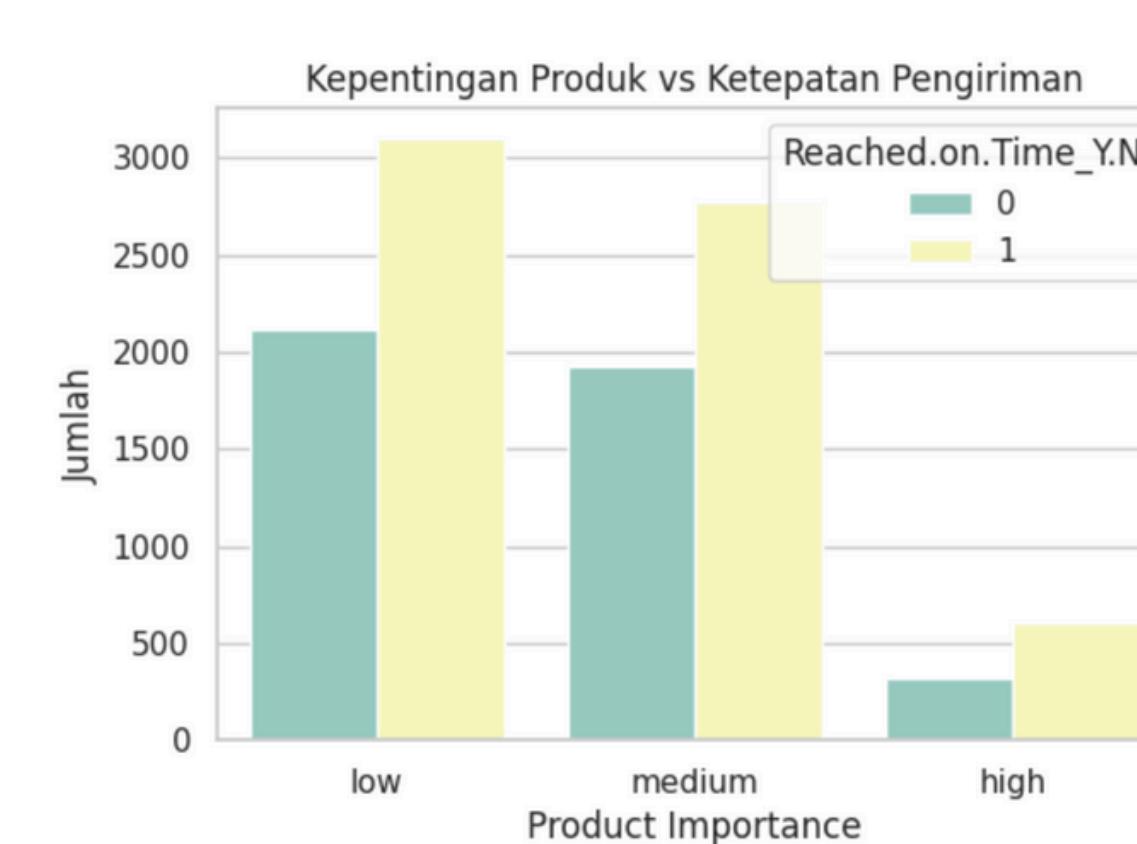


Sebagian besar pelanggan melakukan 3-4 panggilan ke layanan pelanggan.

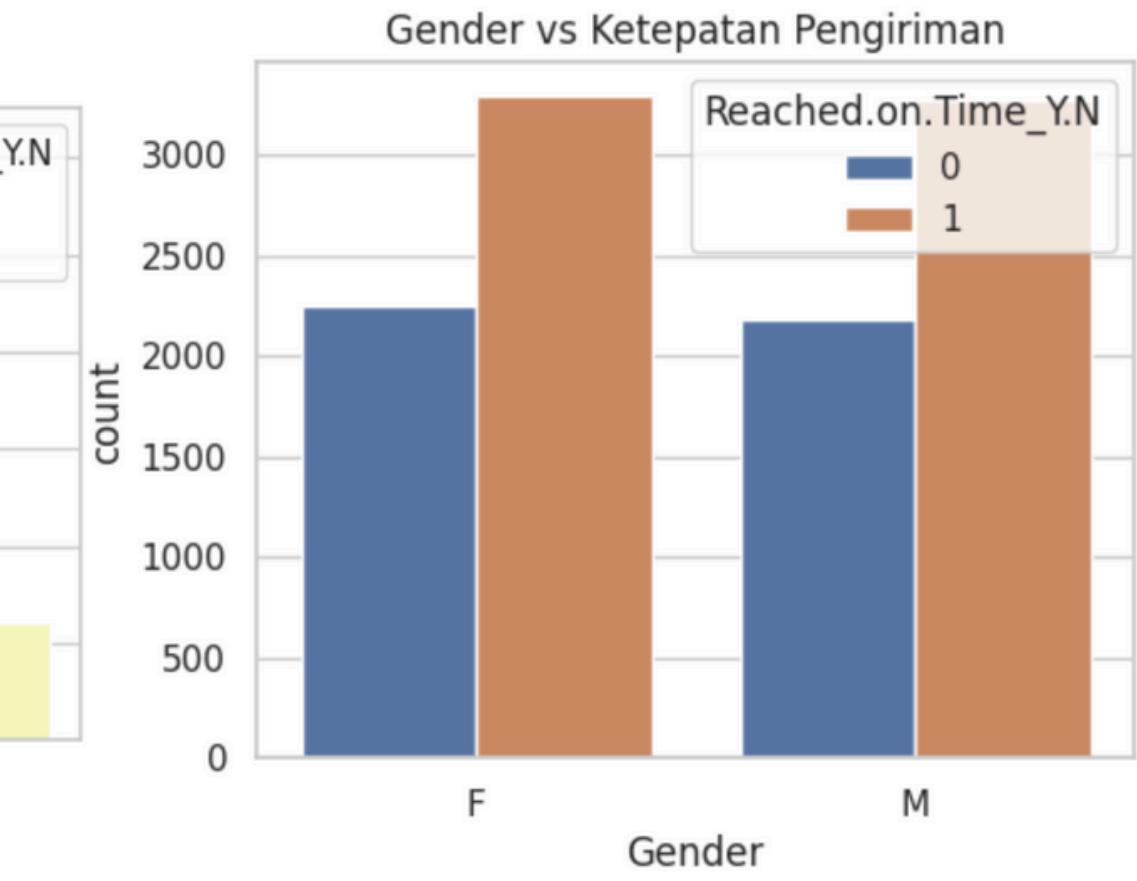
EDA Bivariate Data Analysis



Pengiriman dengan Ship paling banyak digunakan dan memiliki ketepatan waktu yang lebih baik dibanding Road dan Flight.

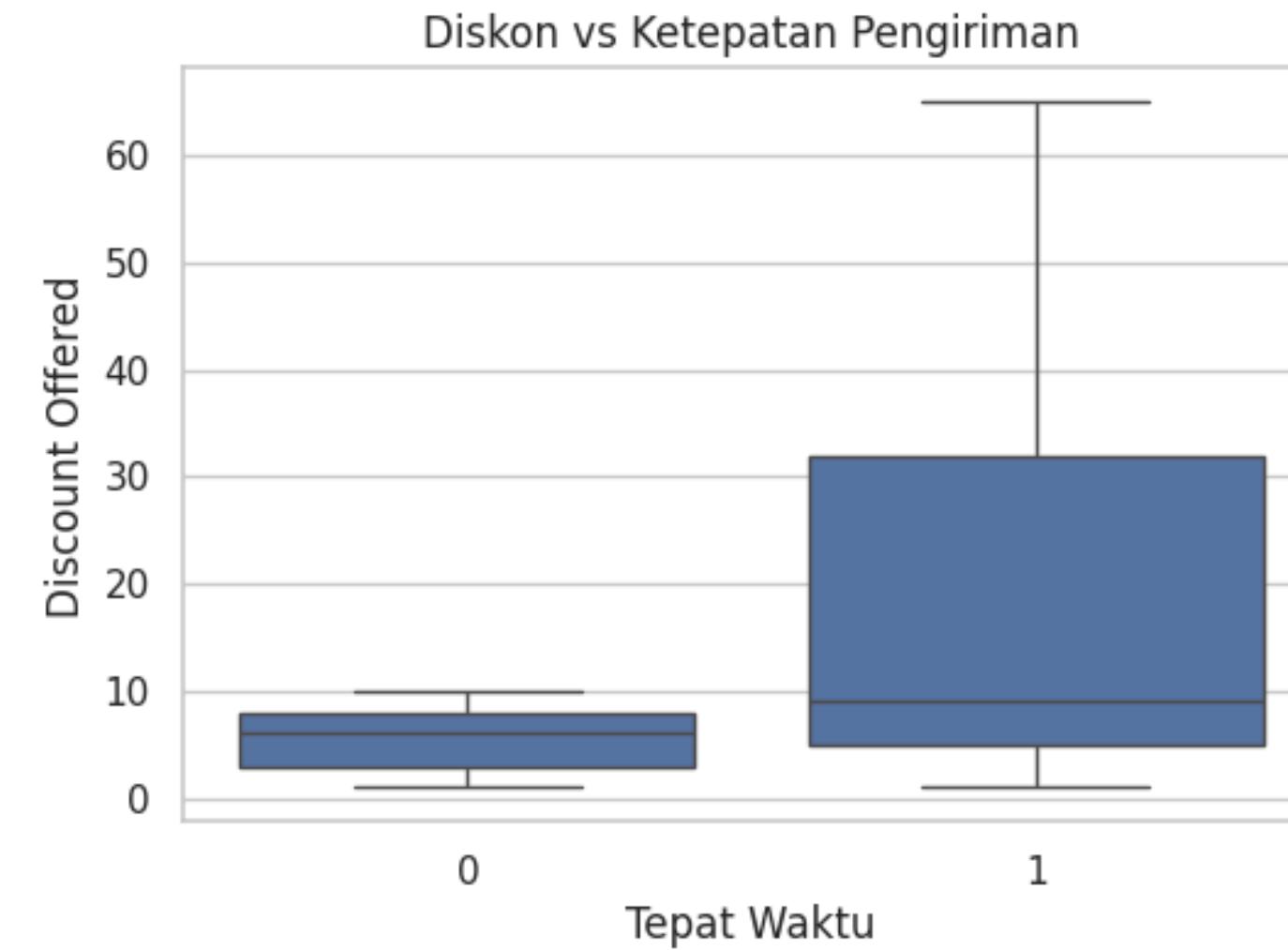


Produk dengan kepentingan tinggi tetap memiliki kasus keterlambatan.

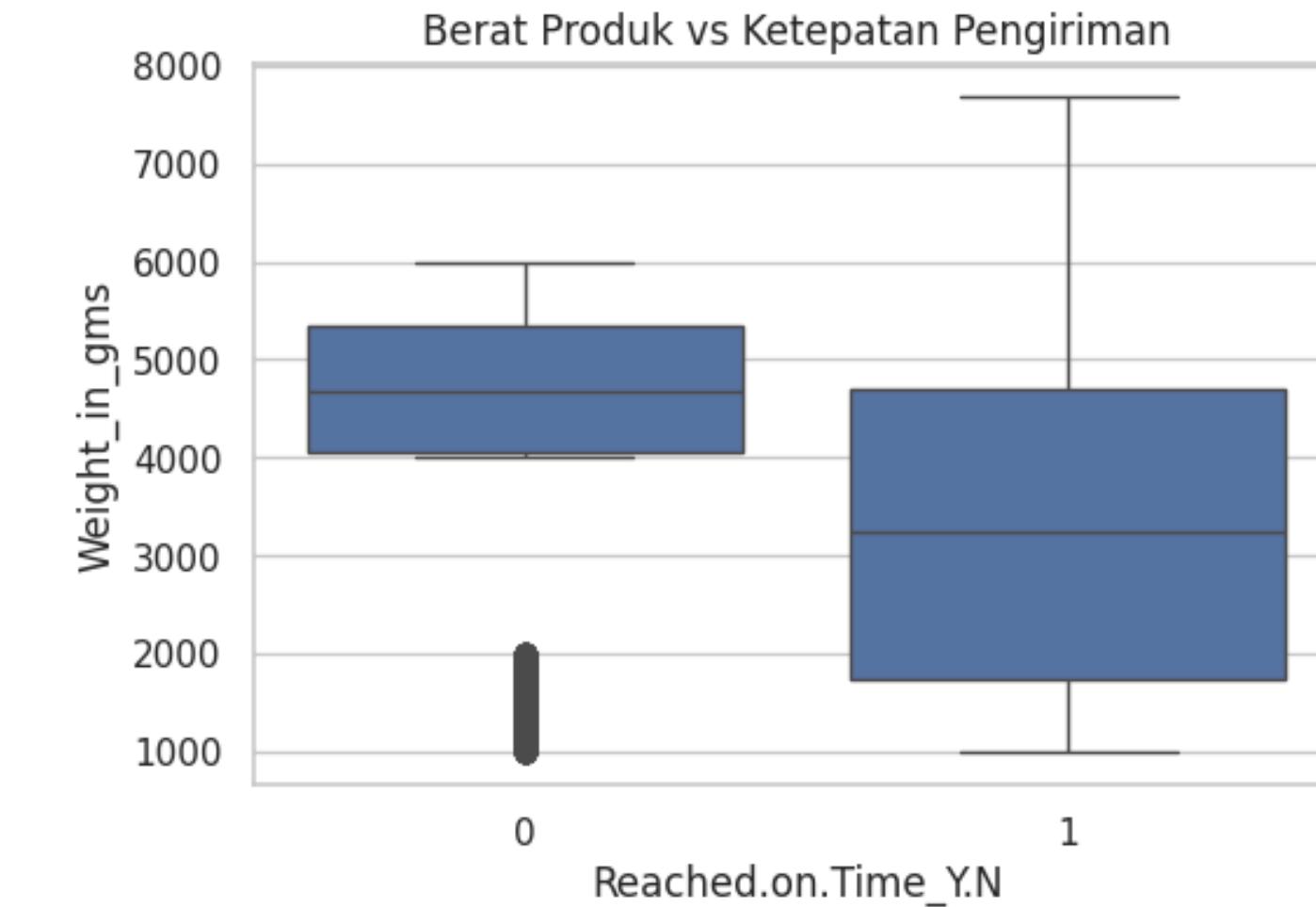


Gender tidak terlalu berpengaruh terhadap ketepatan pengiriman.

EDA Bivariate Data Analysis

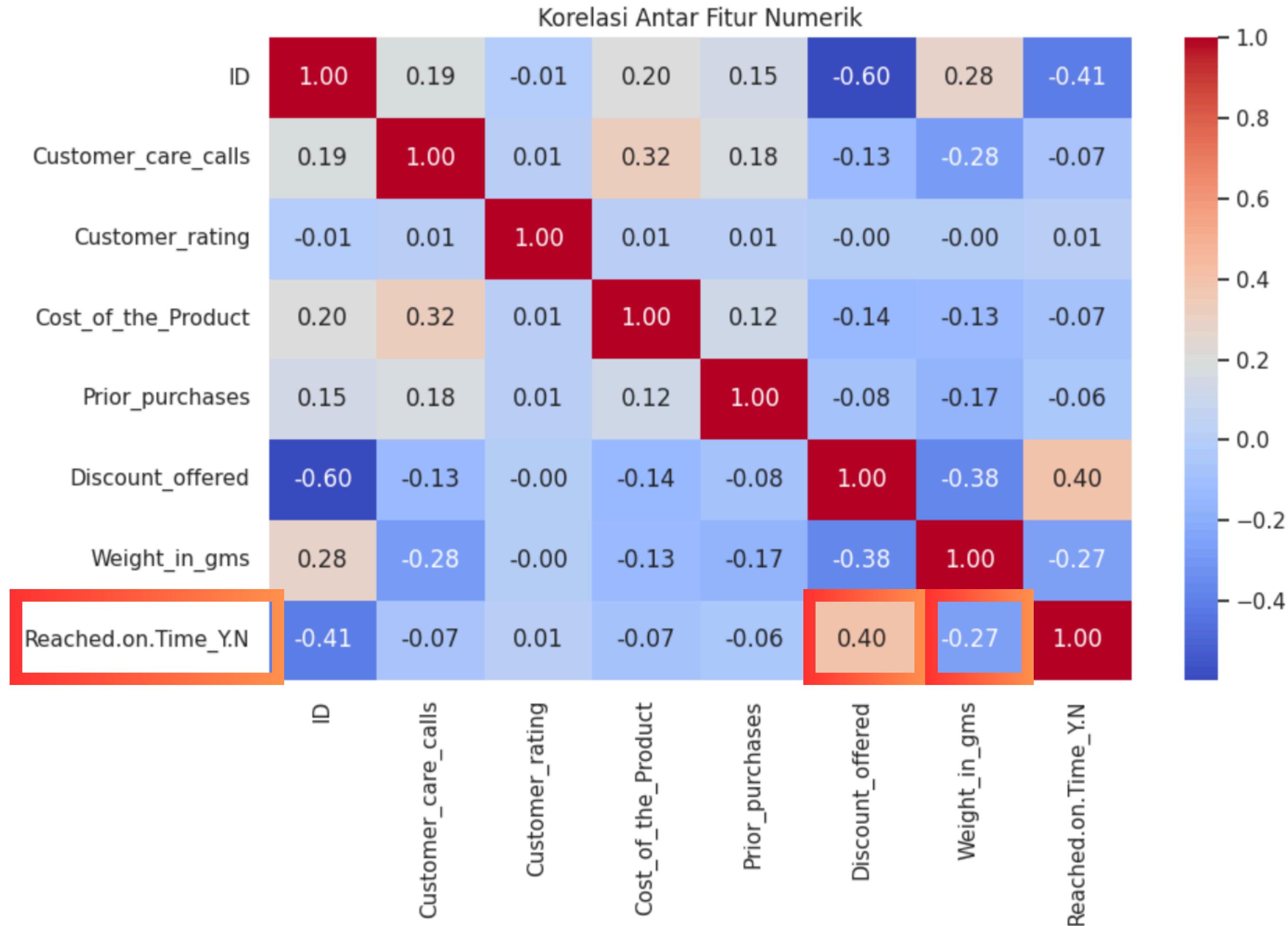


Diskon tinggi cenderung berasosiasi dengan pengiriman tepat waktu

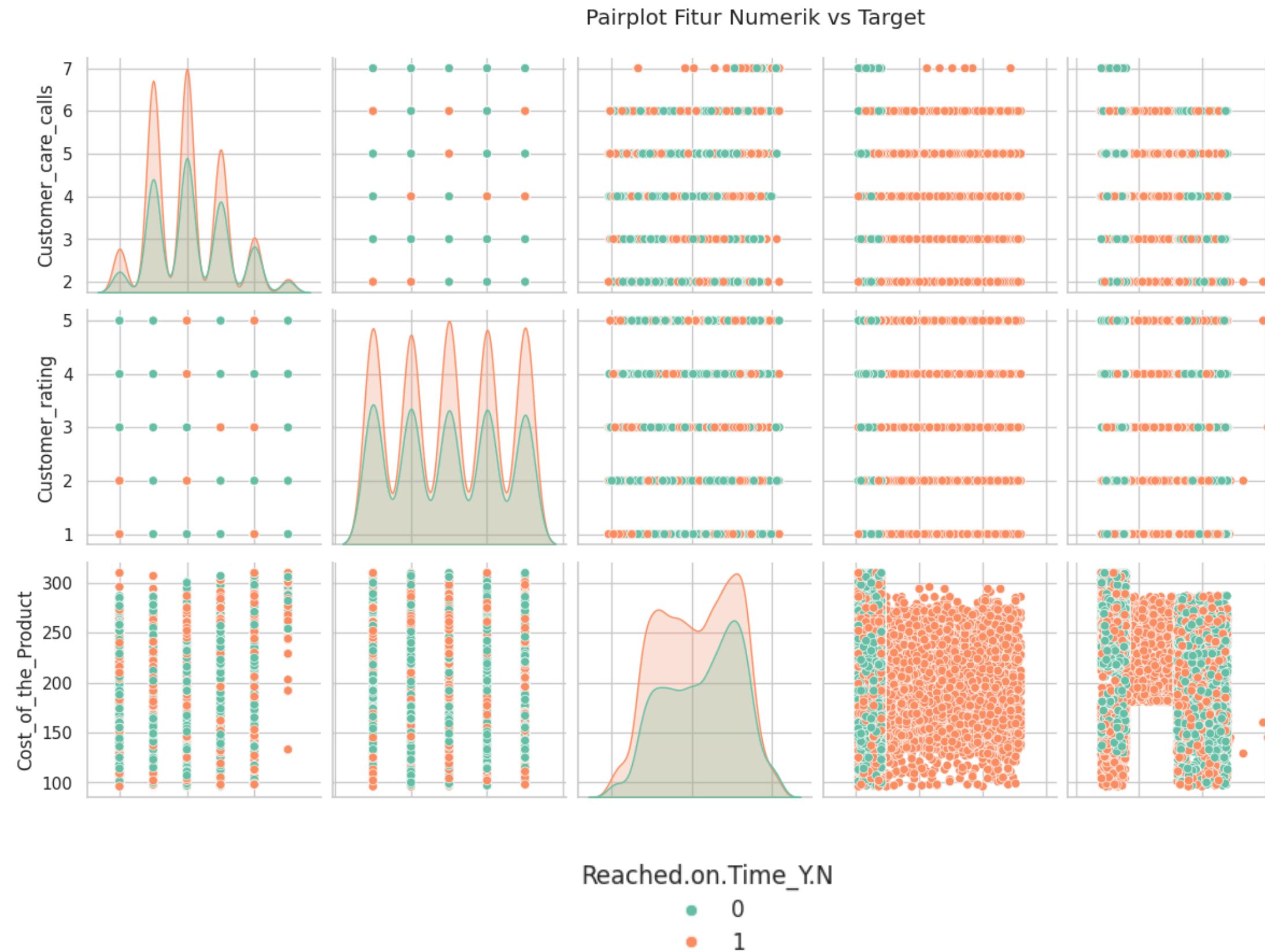


Produk yang lebih berat cenderung tidak dikirim tepat waktu.

EDA Bivariate Data Analysis



EDA - Multivariate Data Analysis -



Tidak ada pola pemisahan yang jelas antara pengiriman tepat waktu (1) dan terlambat (0)

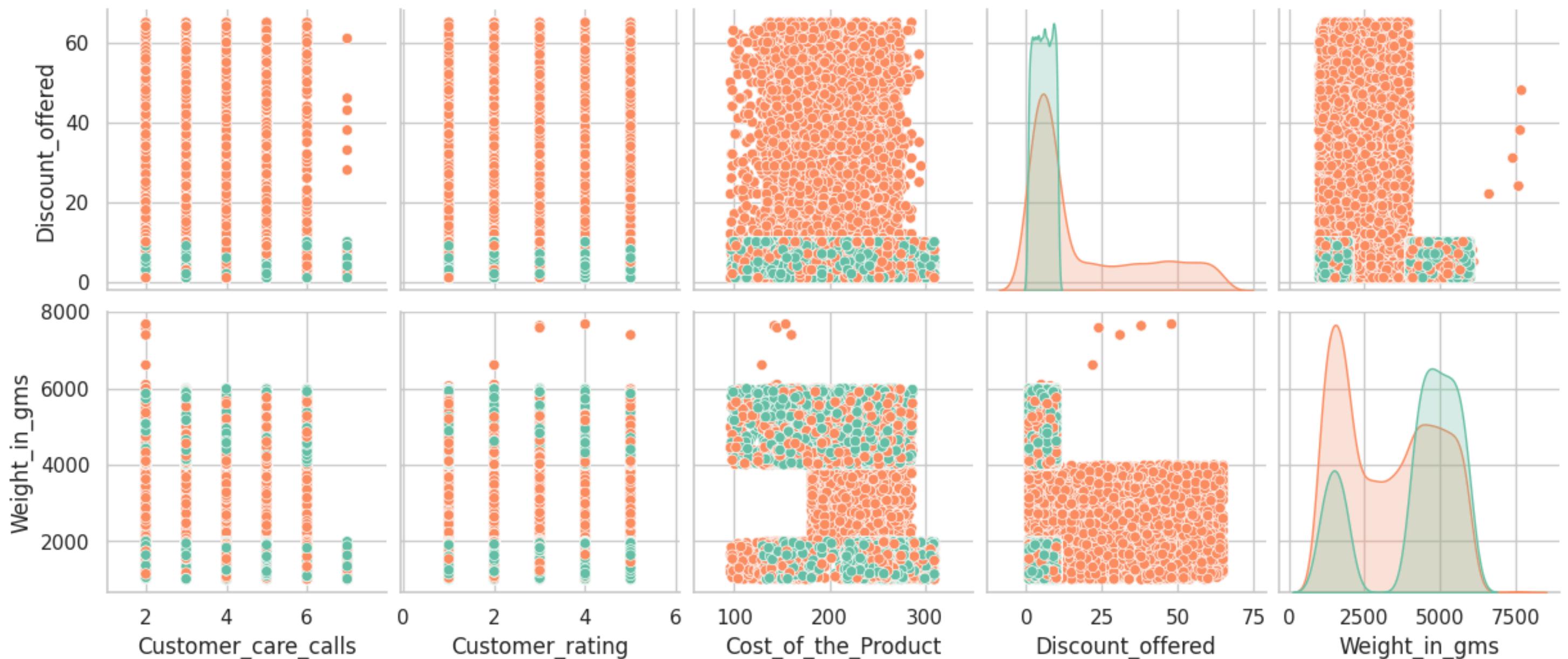
Jumlah panggilan ke customer care tidak memberikan pengaruh signifikan terhadap ketepatan pengiriman.

Customer rating tidak memberikan pemisahan yang bermakna. Rating tinggi maupun rendah tidak menunjukkan pola keterlambatan atau ketepatan.

Distribusi hampir tumpang tindih antara tepat waktu (1) dan terlambat (0)

Harga produk bukan faktor dominan dalam mempengaruhi ketepatan pengiriman.

EDA - Multivariate Data Analysis -



Diskon tinggi lebih sering muncul pada kelas 1
 Diskon besar cenderung dikirim cepat waktu.
 Produk berat lebih banyak di kelas 0 (terlambat)

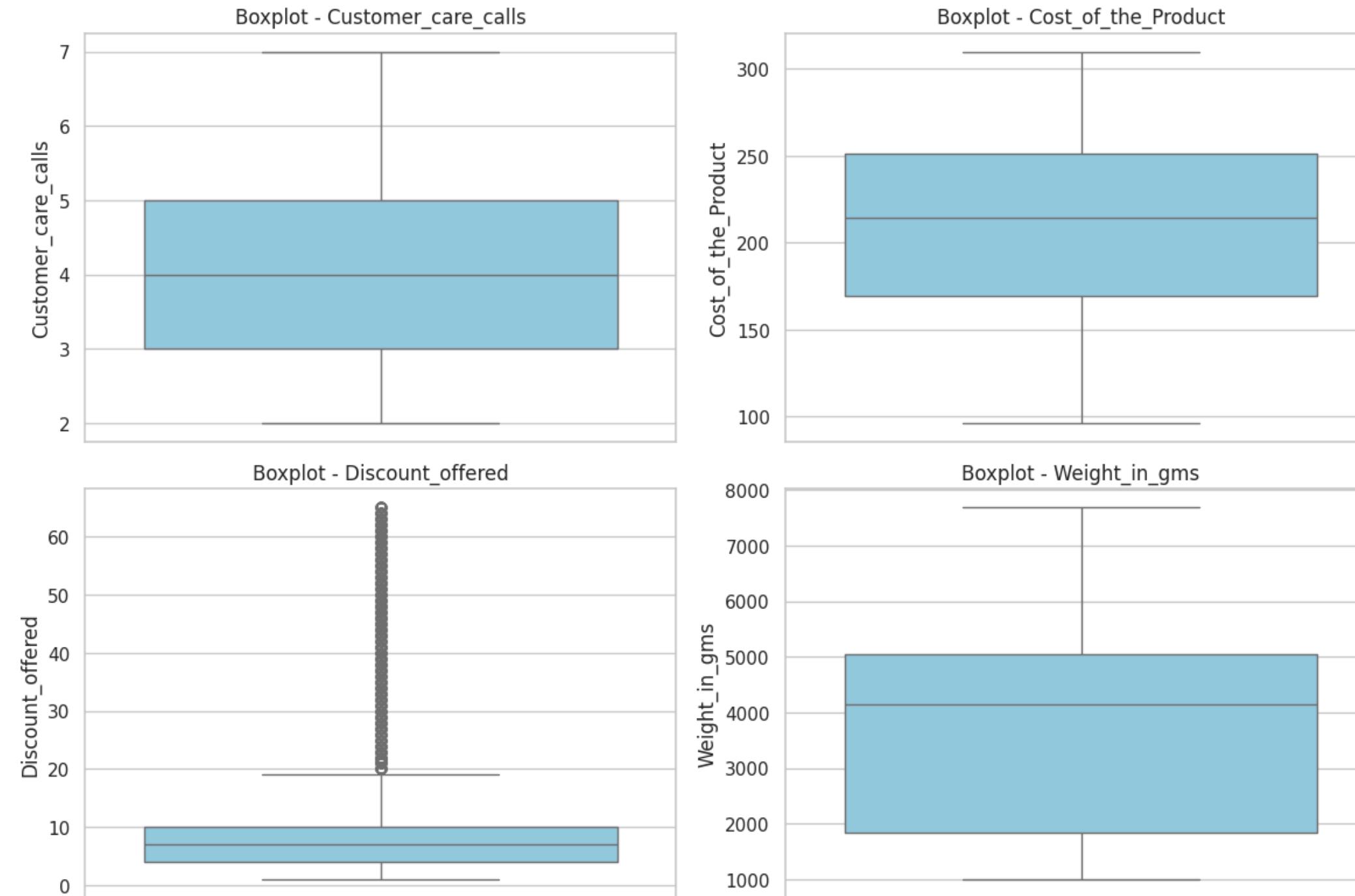
Data Preprocessing

Outliner Handling

Normalization

Imblance Data Handling
(SMOTE)

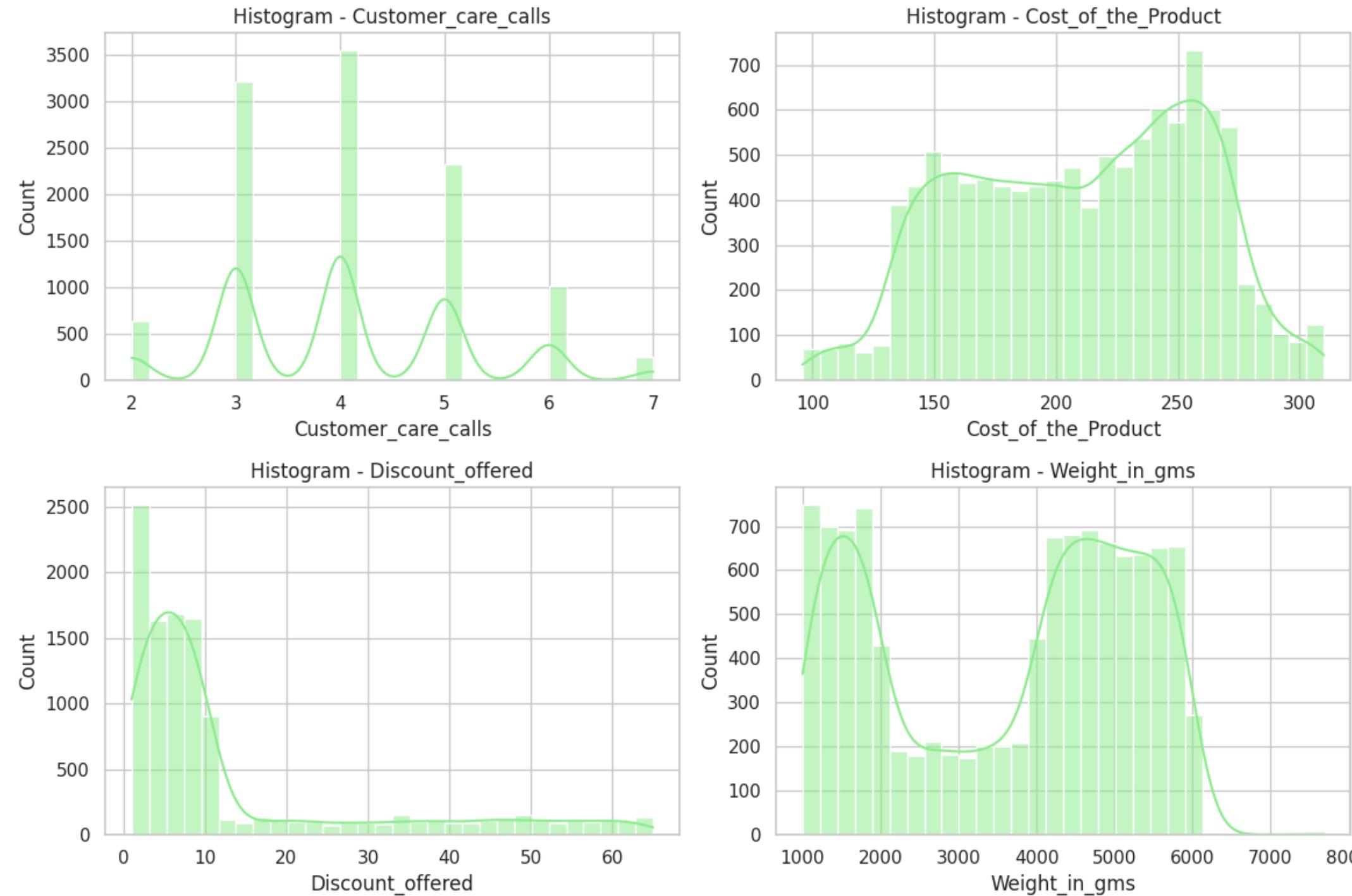
Outlier Handling (numerical data)



Mengidentifikasi dan menangani nilai outlier pada data numerik (misalnya berat, harga, diskon) agar tidak mengganggu hasil analisis dan model.

- Customer_care_calls memiliki beberapa outlier di atas
- Cost_of_the_Product sebaran data harga produk yang cukup stabil.
- Discount_offered banyak outlier di atas kotak
- Weight_in_gms 2 kelompok data berbeda : produk ringan & berat.

Outlier Handling (numerical data)



- Menunjukkan distribusi frekuensi data.
- Membantu melihat apakah data simetris, miring, atau ada lonjakan/puncak tajam (indikasi outlier).
- Menampilkan seberapa banyak nilai yang muncul dalam rentang tertentu.

← → Q Insert your topic here

Normalization

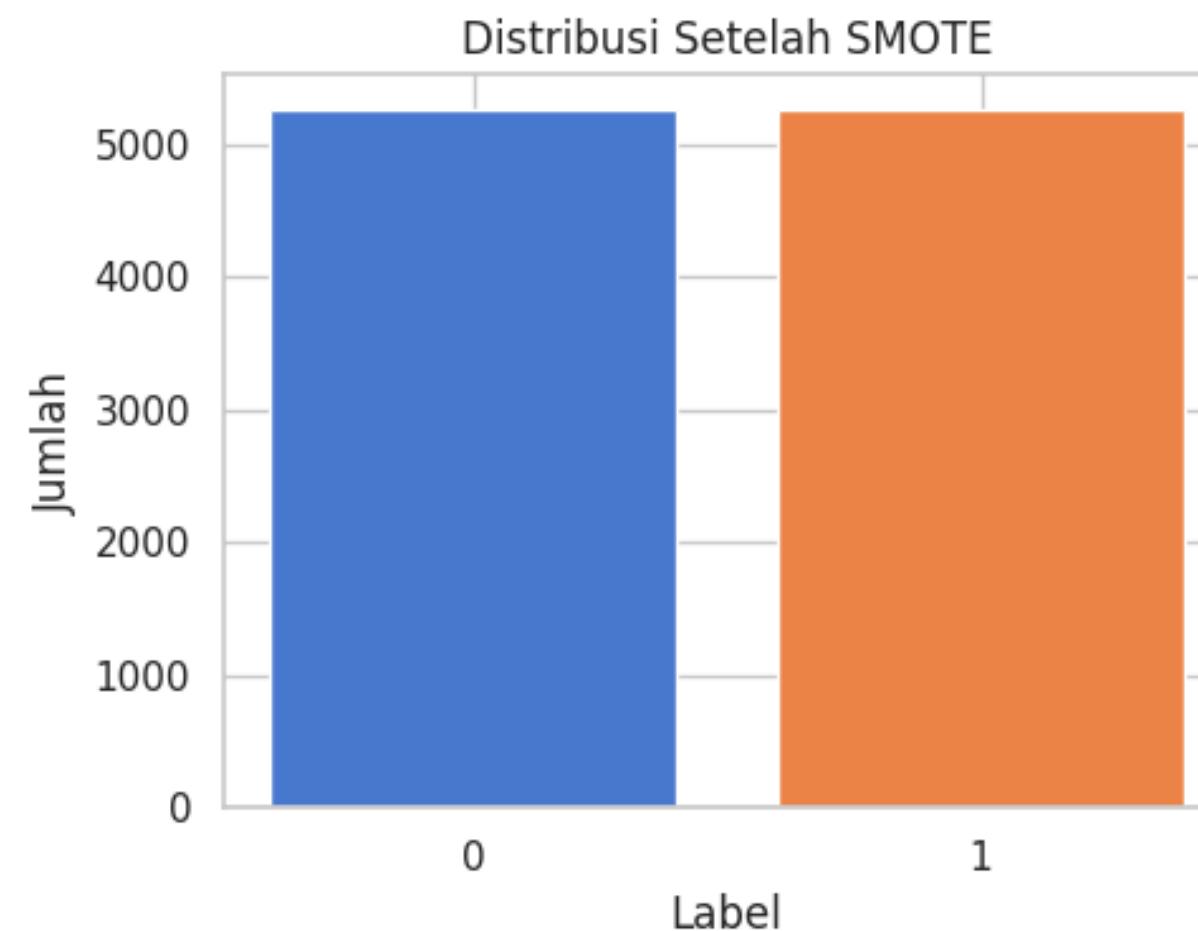
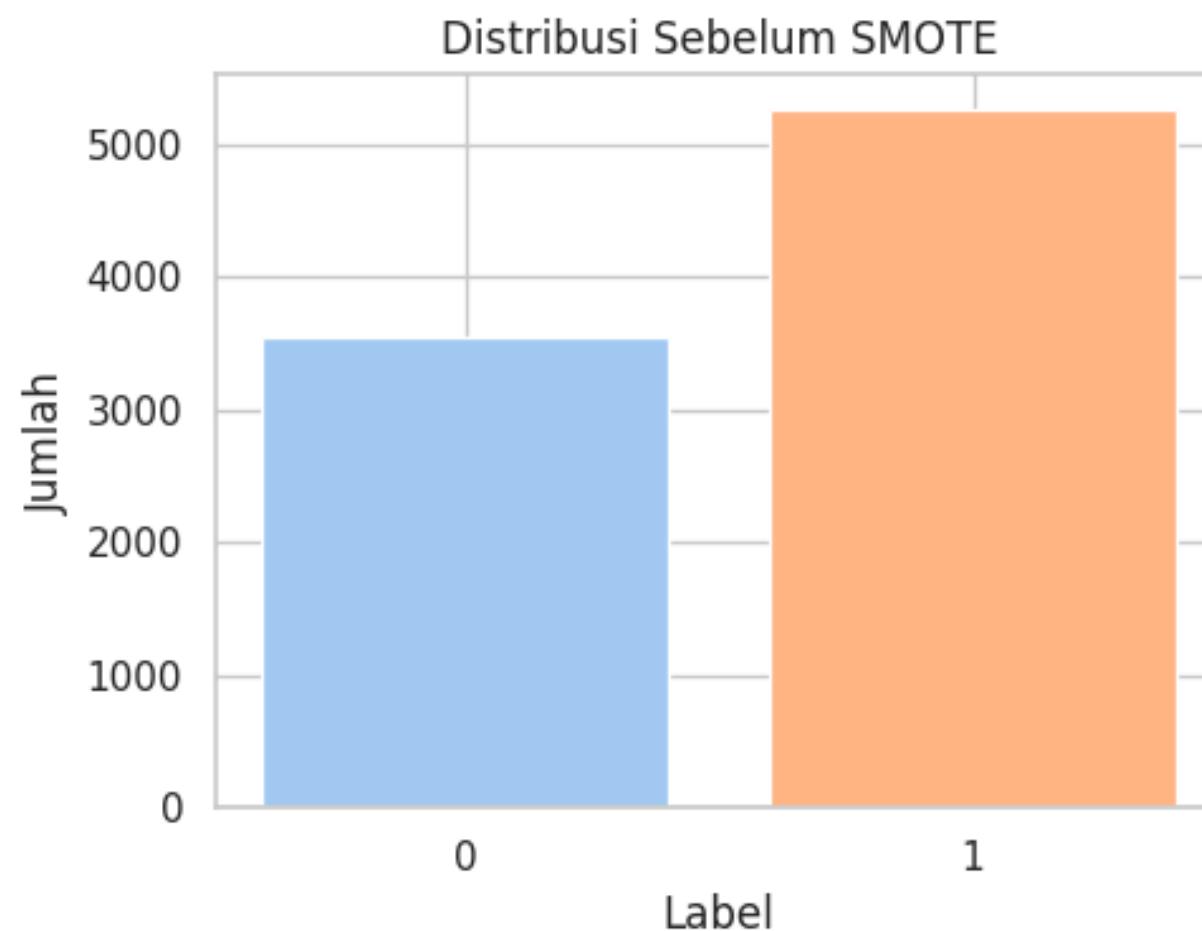
Data numerik berhasil dinormalisasi:

| | Customer_care_calls | Cost_of_the_Product | Discount_offered | Weight_in_gms |
|---|---------------------|---------------------|------------------|---------------|
| 0 | -0.047711 | -0.690666 | 1.889983 | -1.473188 |
| 1 | -0.047711 | 0.120796 | 2.815636 | -0.338634 |
| 2 | -1.799887 | -0.565826 | 2.136824 | -0.163710 |
| 3 | -0.923799 | -0.711473 | -0.208162 | -1.507438 |
| 4 | -1.799887 | -0.545019 | 2.013404 | -0.708052 |

Tabel nilai fitur numerik setelah dinormalisasi (rata-rata 0, standar deviasi 1), contoh pada 5 data pertama.

← → Q Insert your topic here

Imbalance Data Handling



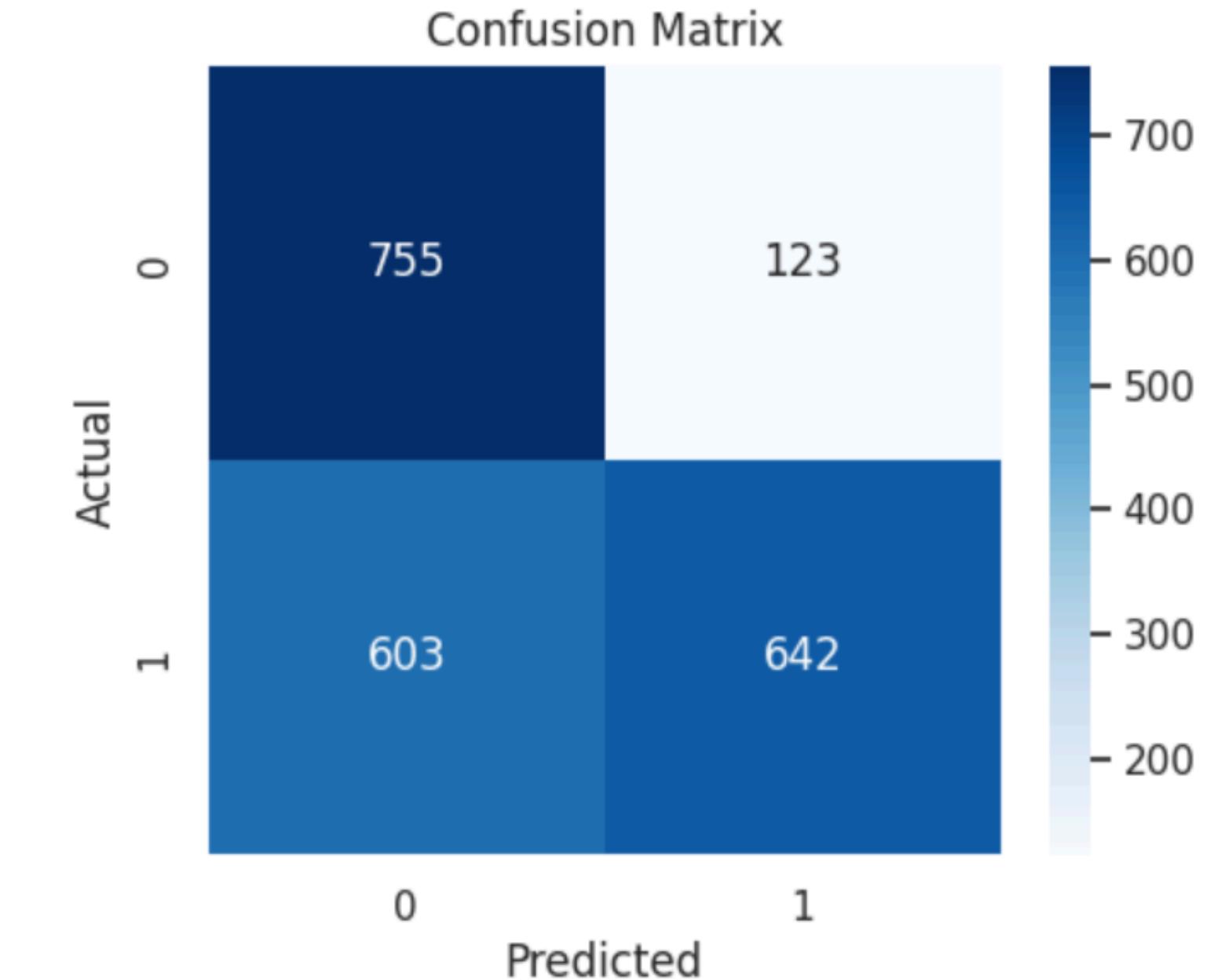
Mengatasi ketidakseimbangan jumlah data antara pengiriman “tepat waktu” dan “terlambat”.

- Label 0 = pengiriman terlambat
- Label 1 = pengiriman tepat waktu

Modelling

Best Model

| Keterangan | Nilai | Penjelasan |
|------------|-------|--|
| TN | 755 | kelas 0, dan memang benar kelasnya 0 (negatif) |
| FP | 123 | kelas 1, tapi ternyata sebenarnya 0 (salah) |
| FN | 603 | kelas 0, tapi ternyata sebenarnya 1 (salah) |
| TP | 642 | kelas 1, dan memang benar kelasnya 1 (positif) |

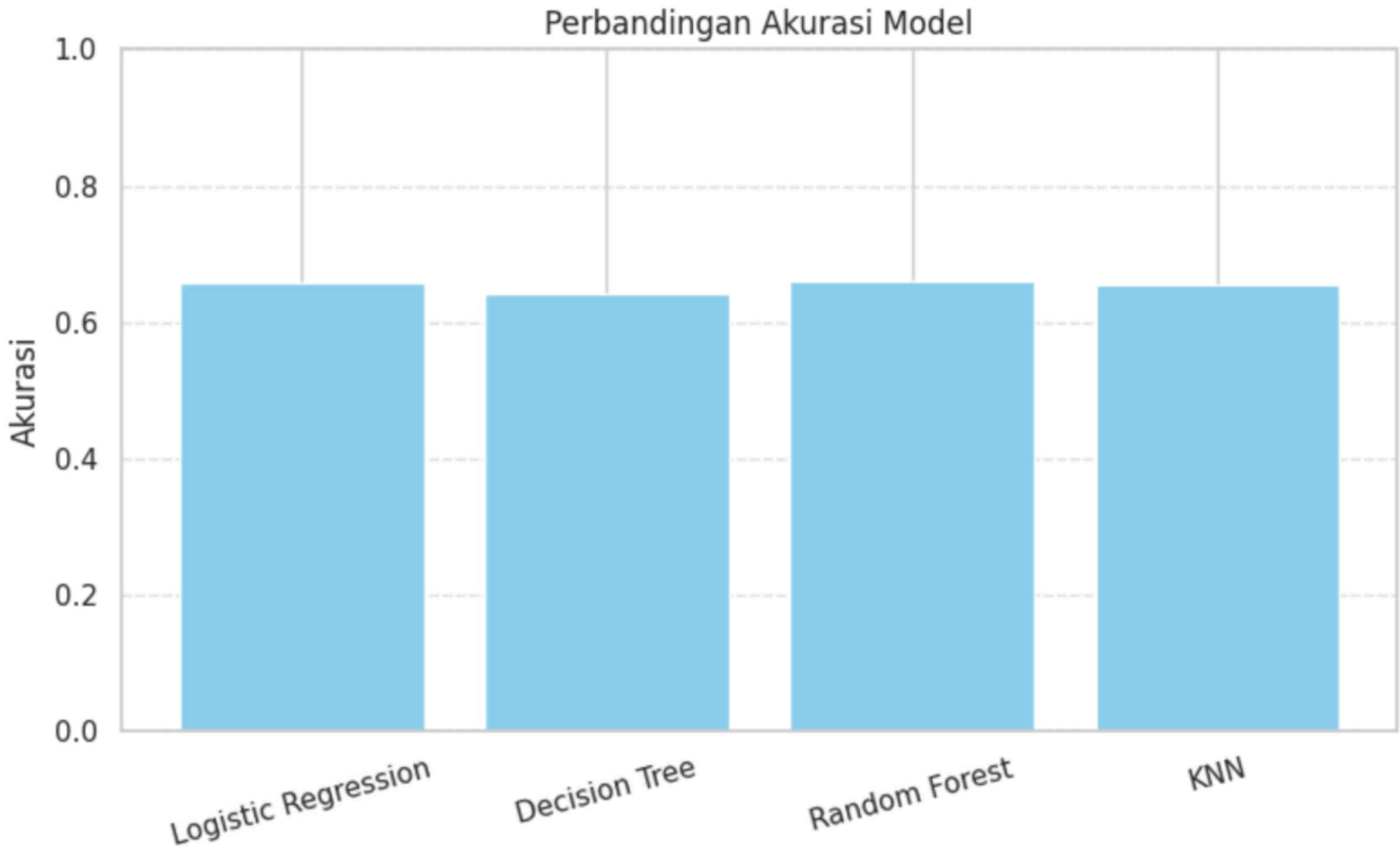


| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.56 | 0.86 | 0.68 | 878 |
| 1 | 0.84 | 0.52 | 0.64 | 1245 |
| accuracy | | | | 0.66 |
| macro avg | 0.70 | 0.69 | 0.66 | 2123 |
| weighted avg | 0.72 | 0.66 | 0.65 | 2123 |

Modelling

Best Model

Berdasarkan Perbandingan Akurasi Model didapatkan Random Forest adalah model terbaik dalam hal akurasi, sedangkan Decision Tree dan KNN memiliki akurasi lebih rendah

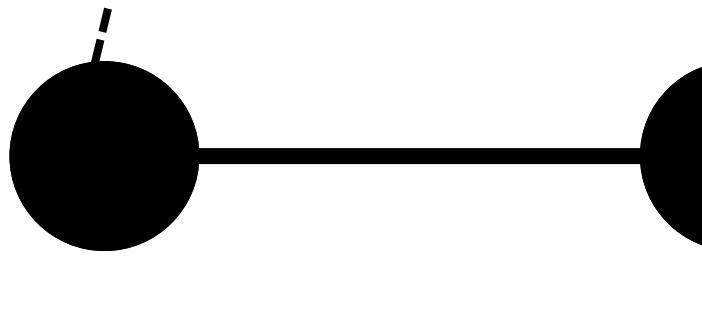


[Data Science](#)[Table of
Contents](#)[Introduction](#)[Data](#)[EDA](#)[Process](#)[Modelling](#)[Summary](#)

← → ⌂ Insert your topic here

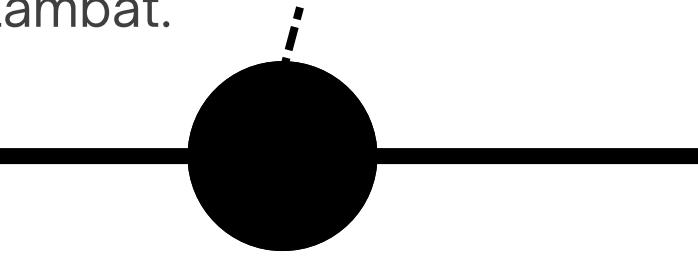
Summary

59% pengiriman
tepat waktu, sisanya
41% terlambat.



Customer care, rating, dan
harga produk tidak
berpengaruh signifikan.

Diskon tinggi: cenderung
tepat waktu
Produk berat: lebih sering
terlambat.



Ship adalah pengiriman
paling sering digunakan
dan paling tepat waktu.

Random Forest adalah
model terbaik setelah
outlier handling,
normalisasi, dan SMOTE

