

Final Project Rakamin Batch 35

# E-Commerce Shipping Data

by Hexa Avengers



<https://www.kaggle.com/datasets/prachi13/customer-analytics>

# Agenda



- **Latar Belakang**
- **Insight**
- **Modeling**
- **Rekomendasi Bisnis**

# Agenda



- **Latar Belakang**
- **Insight**
- **Modeling**
- **Rekomendasi Bisnis**

## OUR TEAM



# HEXA AVENGERS



Kevin Usmayadhy  
Wijaya  
Project Leader



Vicky Clarissa  
Jennie Damara  
Data Analyst



Nabil Abduh  
Aqil  
Data Scientist



Febiya Jomy  
Pratiwi  
Machine Learning Engineer

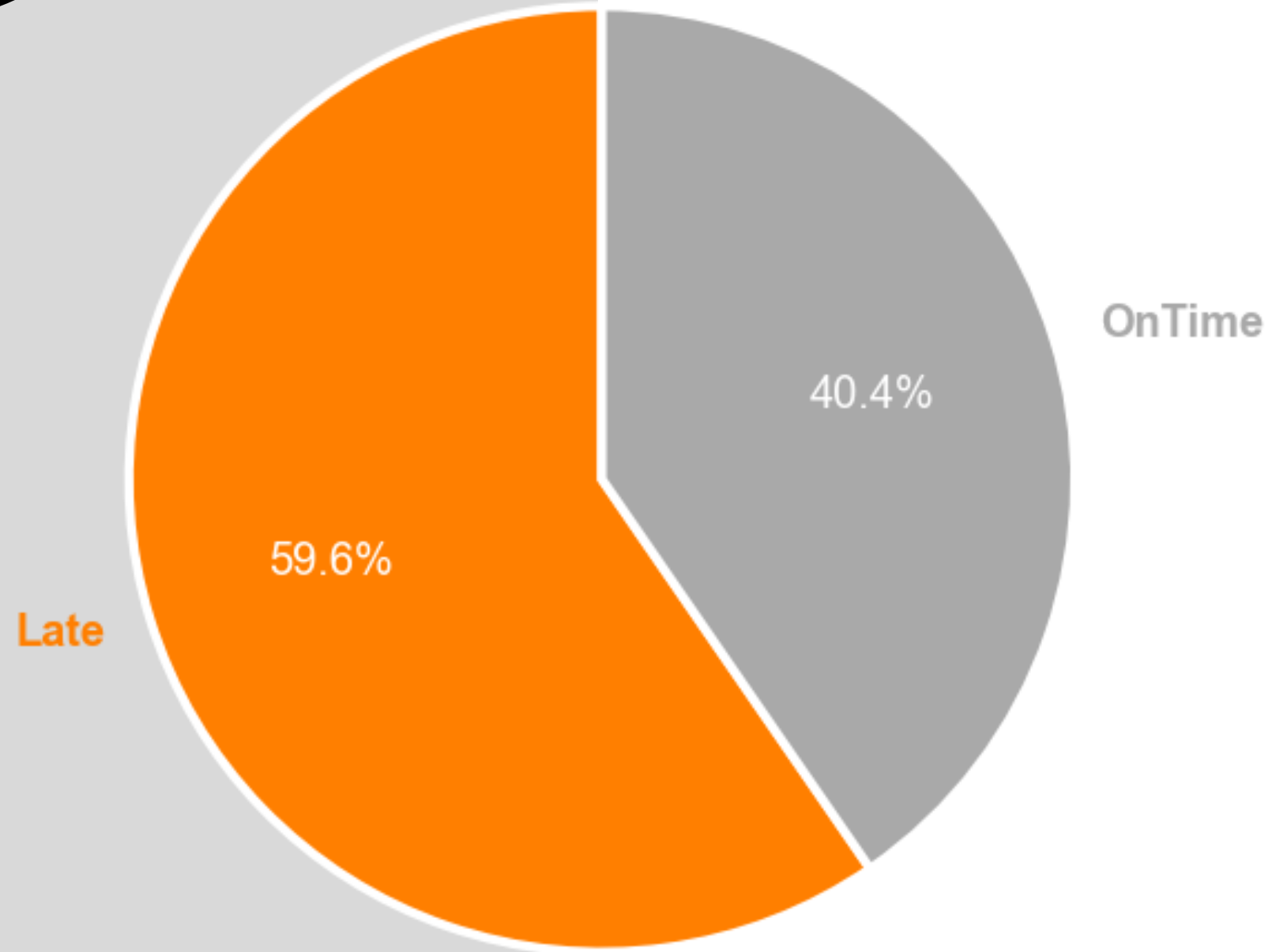


Qistina  
Muharrifa  
Business Analyst



Riel Jeremy  
Jordan Umboh  
Business Analyst

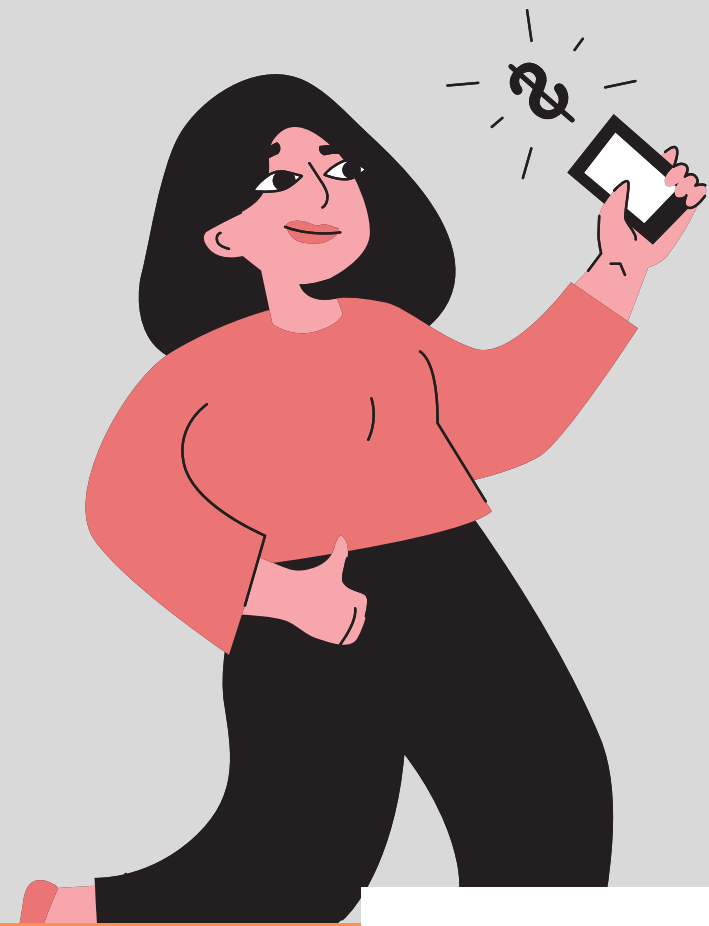
## Late Rate



## Latar Belakang

PT. Avengers merupakan perusahaan di bidang *e-commerce* yang sudah memiliki **10.999 transaksi**. Namun terdapat temuan bahwa sebanyak **6.563 (59.6%) transaksi mengalami keterlambatan** hal ini diduga akan mempengaruhi **satisfaction customer**.





# Problem

Besarnya persentase keterlambatan barang



## Goal

Menurunkan  
keterlambatan barang



## Objective

- Membuat model klasifikasi yang bisa memprediksi keterlambatan barang
- Mencari faktor-faktor yang mempengaruhi keterlambatan



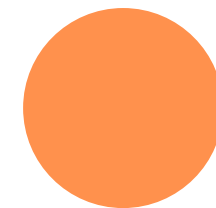
## Business Metrics

Late Rate

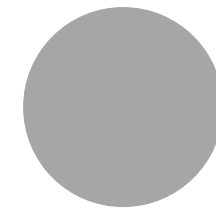
# Agenda



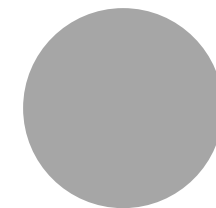
**Latar Belakang**



**Insight**



**Modeling**



**Rekomendasi Bisnis**



# Data Understanding

Dataset yang digunakan merupakan dataset **E-Commerce Shipping Data** yang terdiri dari 10.999 baris dan 12 kolom.



## Customer

- Id
- Gender
- Warehouse Block



## Layanan

- Mode of Shipment
- Customer Care Calls
- Customer Rating
- Cost of the Product
- Prior Purchases
- Product Importance
- Discount Offered
- Weight in gms

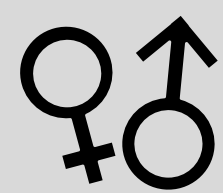


## Target

- Is Late



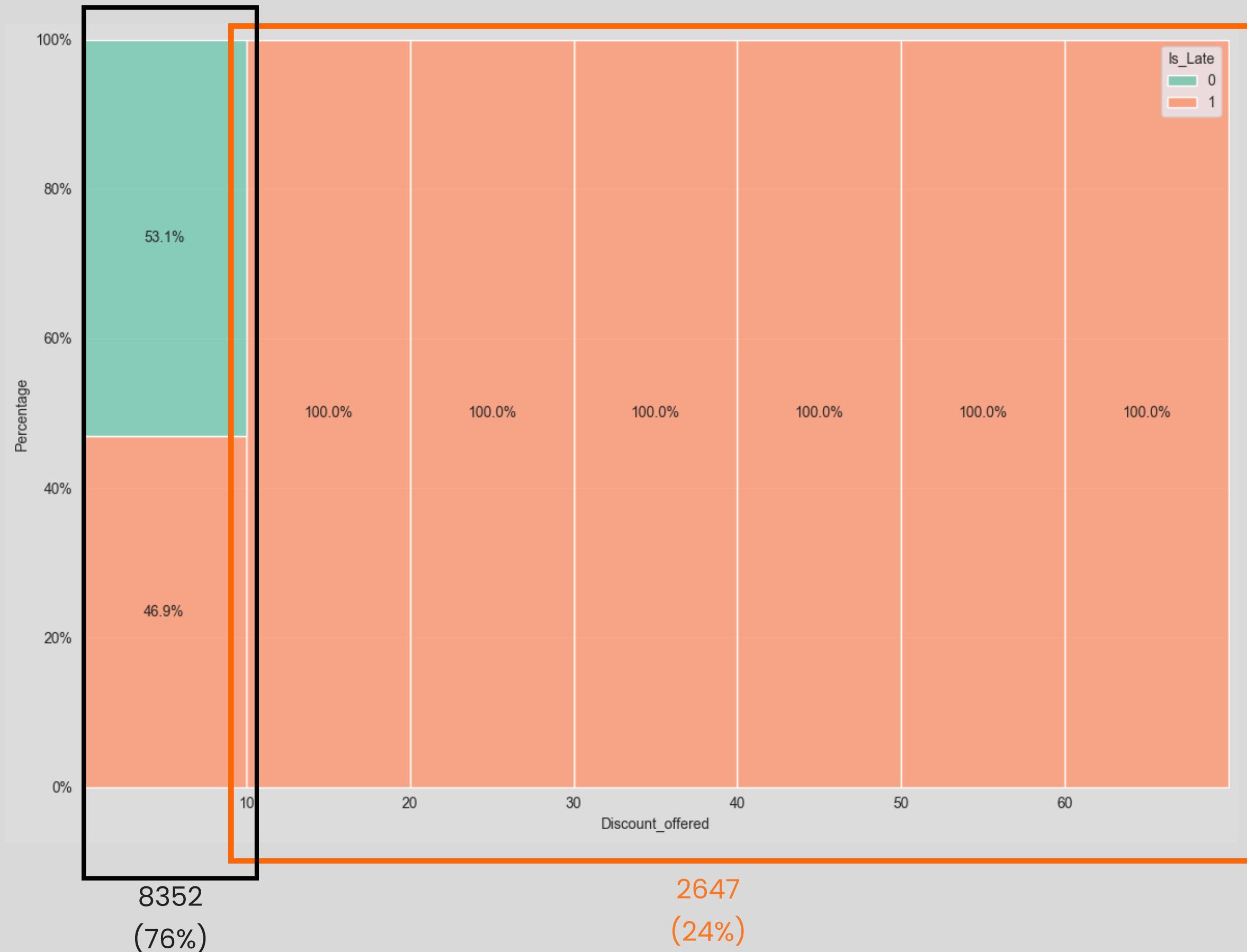
# Customer Order Profile



Gender	Warehouse Block	Mode of Shipment	Product_Importance
Male 49.6%	Block A 16.7%	Ship 67.8%	Low 48.2%
Female 50.4%	Block B 16.7%	Road 16.0%	Medium 43.2%
-	Block C 16.7%	Flight 16.2%	High 8.6%
-	Block D 16.7%	-	-
-	Block F 33.3%	-	-

# Discount Problem

Diskon **diatas 10%** semuanya **terlambat**



## Asumsi

Diskon produk tidak hanya memotong harga produk namun turut **memotong shipment cost**.

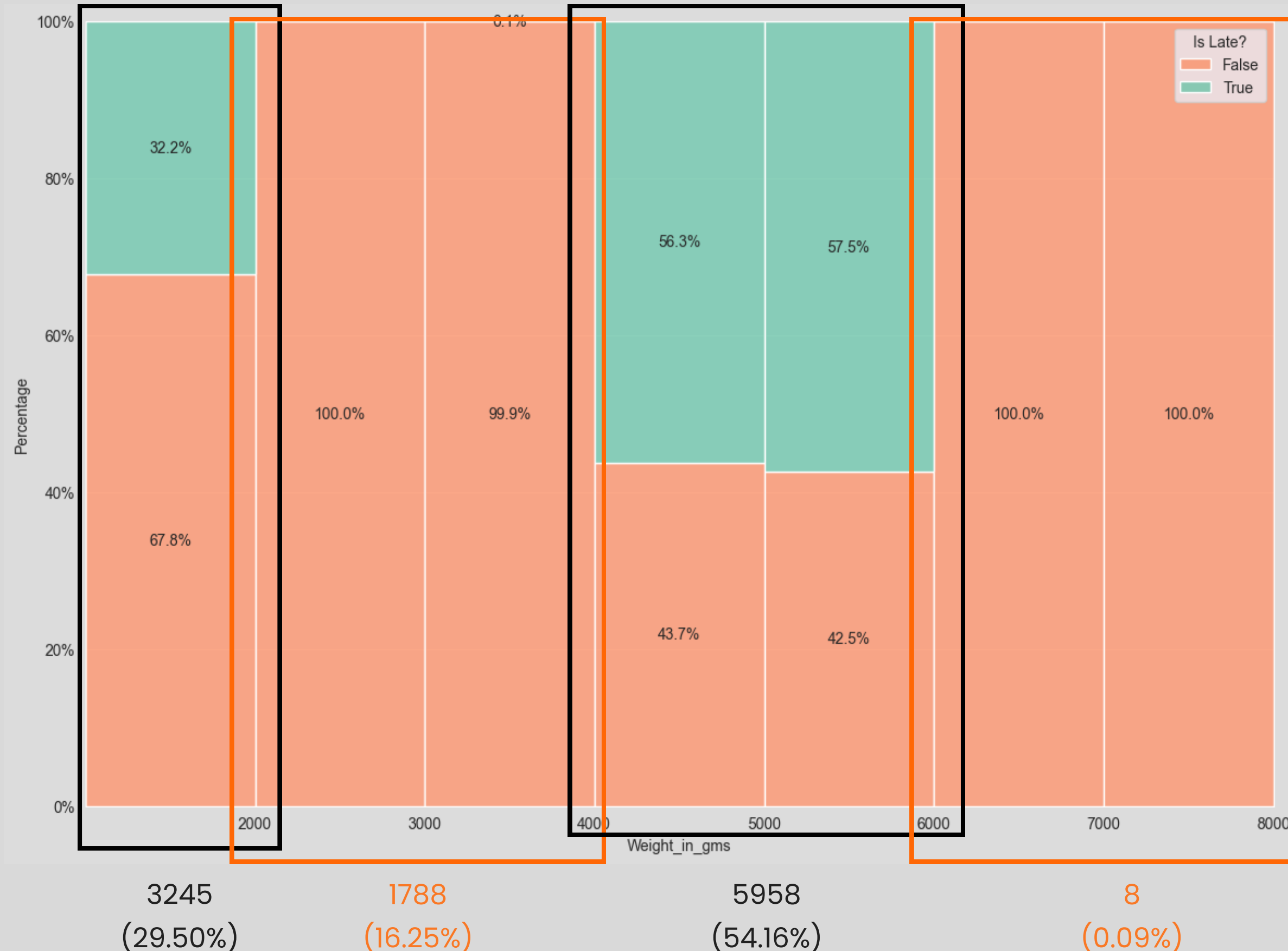
## Rekomendasi

- **Tidak memberikan diskon** apabila akan memotong shipment costnya juga.
- **Membatasi pemberian diskon** hanya sebesar maximal 10%.

\*disclaimer: semua produk memiliki diskon  $\geq 1\%$

# Weight Problem

Berat **2-4 kg** dan **>6 kg** semuanya mengalami **keterlambatan**



## Asumsi

- Berat 2-4 kg merupakan berat yang **tanggung** (memiliki berat yang tidak ringan dan tidak terlalu berat namun memiliki **shipment cost** yang sama dengan 1000-2000).
- Berat diatas 6 kg terlambat dikarenakan jumlah barang yang dapat **diantar dalam satu kali pengiriman** terbatas.

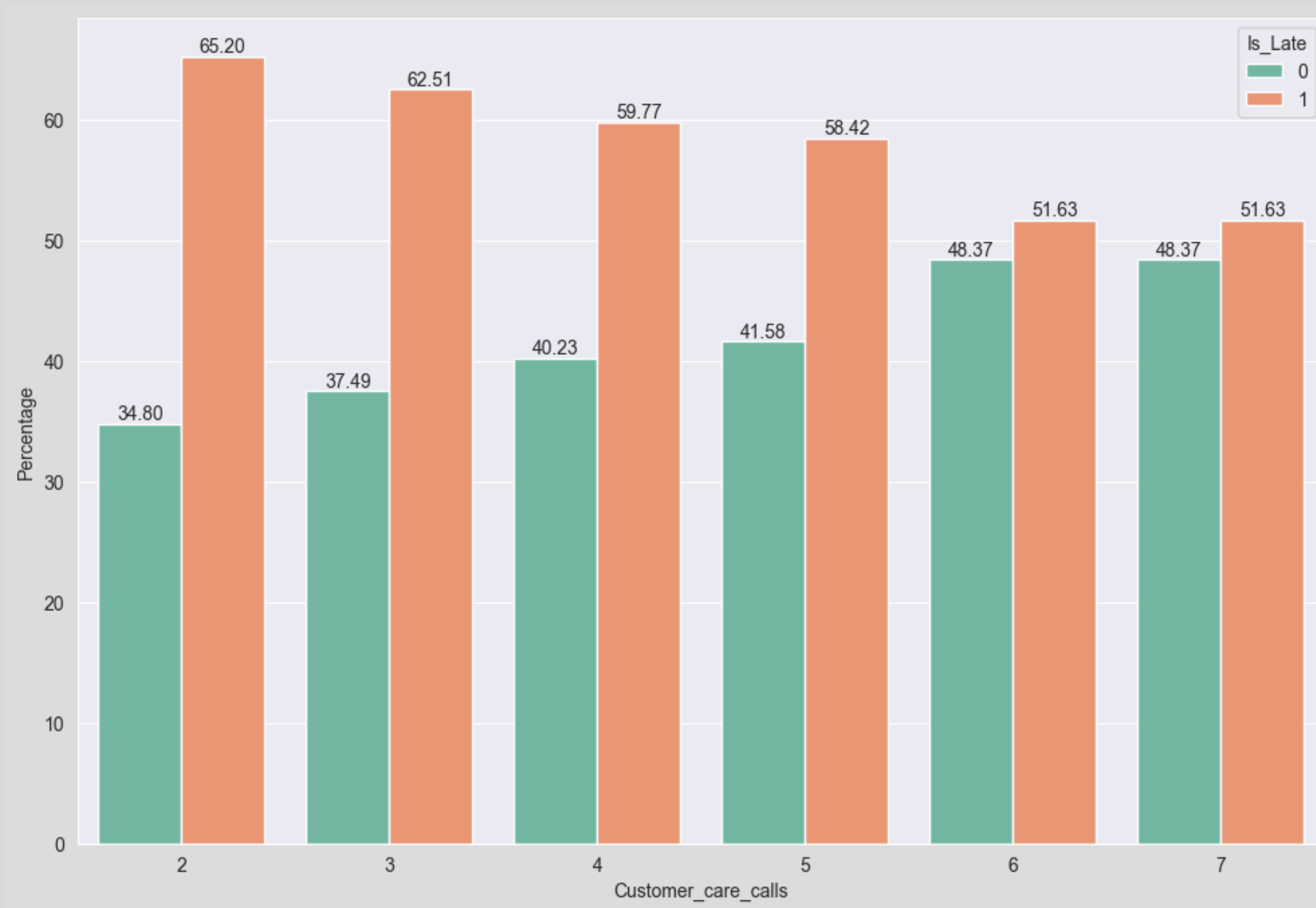
## Rekomendasi

Mengevaluasi kembali kategori shipment cost yang diberikan.

\*disclaimer: semua produk memiliki berat  $\geq 1$  kg

# Call Problem

Semakin sering customer menelpon, semakin rendah persentase keterlambatan



## Asumsi

Hanya customer yang sering melakukan panggilan yang difollow up barangnya.

## Rekomendasi

Memperbaiki sistem antrian, jangan hanya memprioritaskan customer yang sering melakukan panggilan saja.

\*disclaimer: customer sudah melakukan panggilan paling tidak 2x.

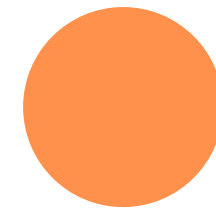
# Agenda



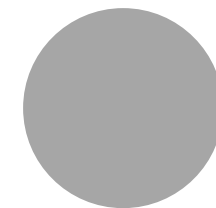
**Latar Belakang**



**Insight**



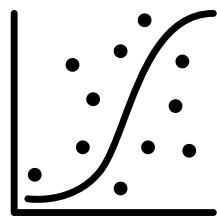
**Modeling**



**Rekomendasi Bisnis**

# Preprocessing

Membersihkan dan menyesuaikan data agar lebih efisien untuk pelatihan model



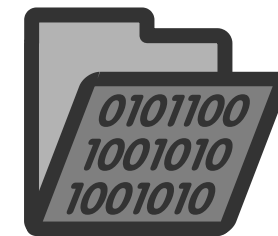
## Handling Outlier

Log Transformation untuk  
prior\_purchases dan  
discount\_offered



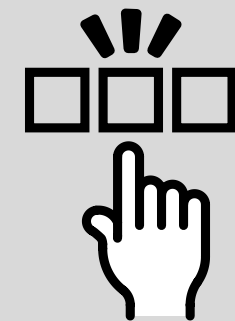
## Fitur Scaling

MinMax dan Standard Scaller,  
untuk Weight\_in\_gms dan  
Cost\_Of\_the\_Product



## Fitur Encoding

- Label Encoding: product importance dan Gender
- OHE: Warehouse\_block dan Mode\_of\_Shipment



## Fitur Selection & Splitting data

- Drop fitur warehouse block
- Split 80:20

# Modeling

Memprediksi keterlambatan barang (Is\_Late)

Model	Logistic Regression	KNN	Decision Tree	Random Forest	AdaBoost	XGBoost
Metrics						
Accuracy (Test Score)	61%	64%	66%	63%	60%	68%
Precision (Test Score)	63%	71%	71%	66%	60%	86%
Recall (Test score)	86%	68%	72%	80%	100%	55%
	Best Model					

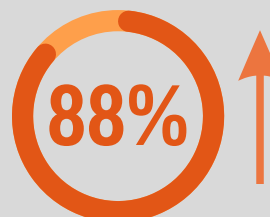
False Negative ↓  
Late Rate ↓  
Customer Satisfaction ↑



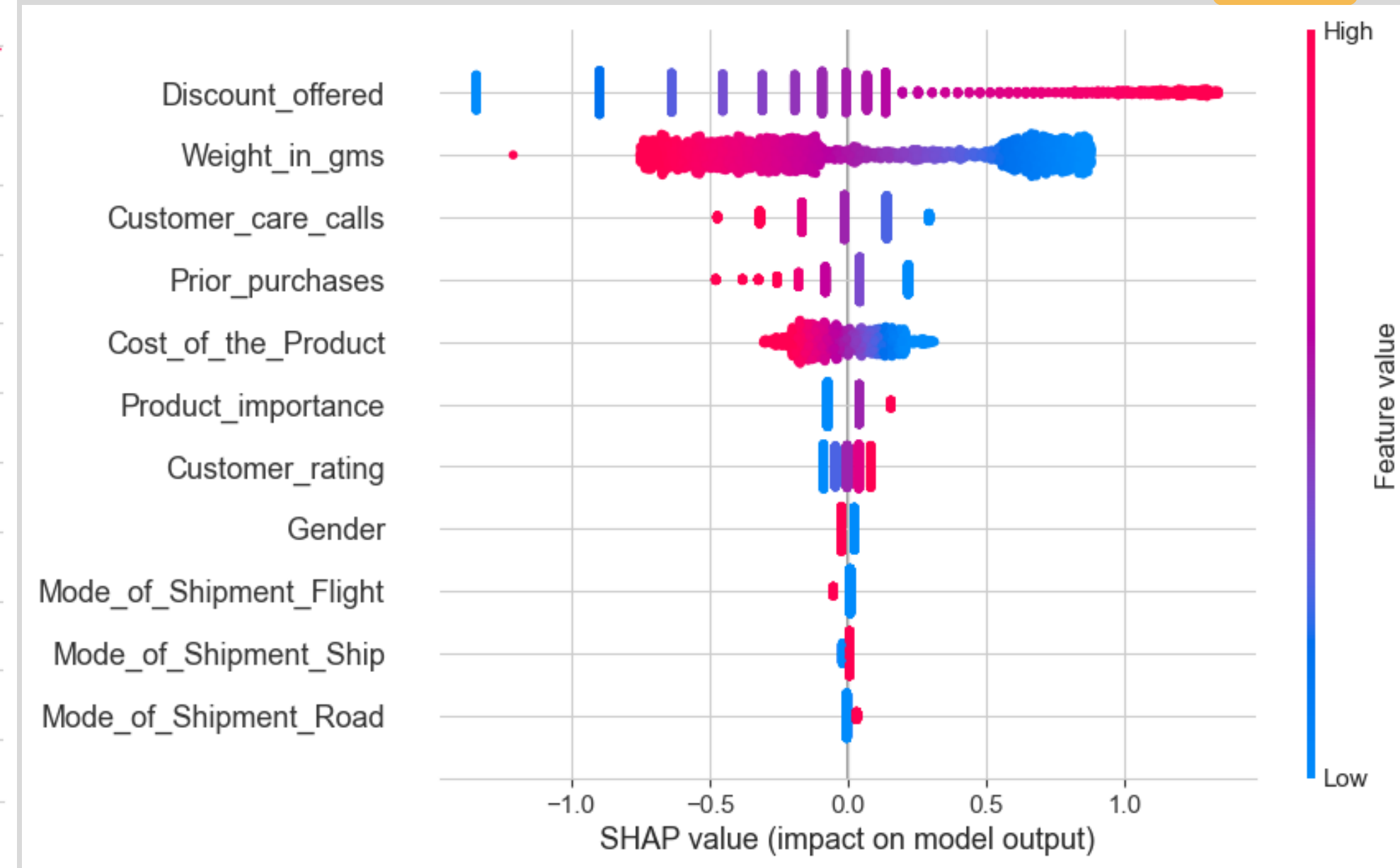
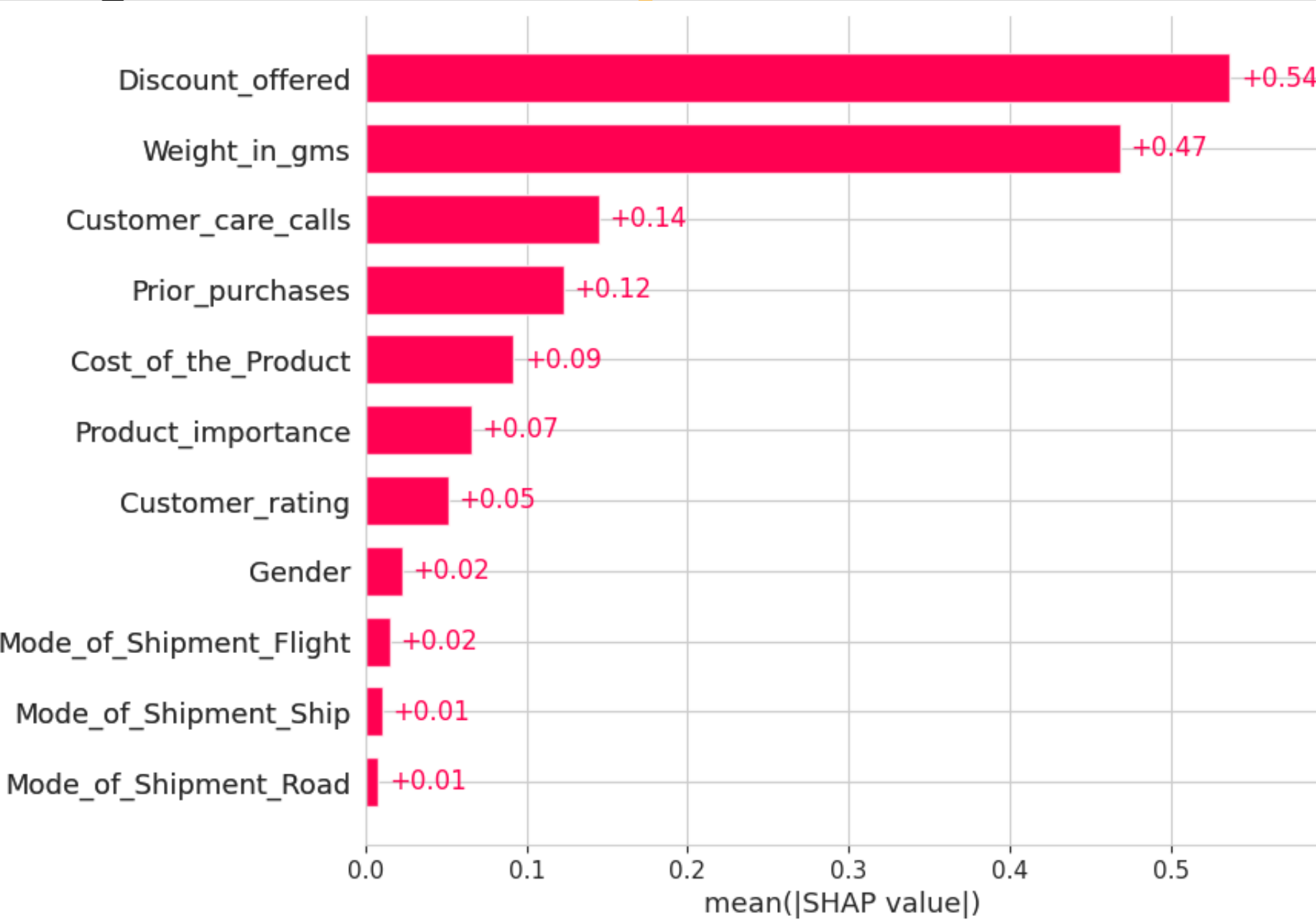
Hyperparameter  
Tuning



Recall  
(Test score)



# Feature Importance





# Agenda



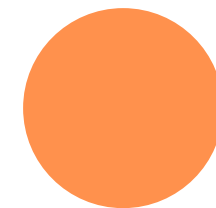
**Latar Belakang**



**Insight**



**Modeling**



**Rekomendasi Bisnis**

# Business Simulation

Bulan ini perusahaan mendapatkan 2200 transaksi dengan keterlambatan sebanyak 1319 (59.9%).

## Hasil Model Prediksi

- Customer late yang terprediksi late (TP) = 1158
- Customer late yang terprediksi on time (FN) = 161
- Customer on time terprediksi on time (TN) = 171
- Customer on tiime terprediksi late (FP) = 710

Aktual on time: 881

Aktual late: 1319

## Potential revenue loss

**Total Sales** = jumlah Cost\_of\_the\_Product = **\$458964**

**Total Diskon** = jumlah harga diskon = **\$59048**

**Total Revenue** = Total Sales - Total Diskon = **\$399916**

**Revenue per product** = Total Revenue/Jumlah product  
= **\$181.78**

**Potential Revenue Loss** = Revenue per product \* aktual late  
= **\$239767.87**

## Shipment Cost

Berdasarkan [www.freightos.com](http://www.freightos.com) harga metode shipment adalah sebagai berikut:

- Road = 2\$/kg
- Ship = 4\$/kg
- Flight = 8\$/kg

**Total Shipment Cost = \$39536 atau \$18 per product**

## Budget yang tersedia

Agar product dapat dikirim secara on time maka biaya yang perlu dikeluarkan adalah **2x biaya normal**. Perusahaan memberikan budget sebesar **\$50000** untuk mengurangi **late rate** yang terjadi.

# Business Simulation (2)

Budget Total \$50000

Jumlah Produk yang dapat diberikan tambahan

Harga per produk agar on time = shipment cost \* 2  
=  $18 * 2 = \$36$

**Produk yang dapat diberikan tambahan biaya**

= Budget Total / 36

=  $50000 / 36 = 1389$  Barang

Jumlah Barang yang terlambat

**Jumlah produk terlambat sebelumnya = 1319 (59.9%)**

**Jumlah produk terlambat setelah diberikan treatment**

= Total product - (1389 + TN)

=  $2200 - (1389 + 171)$

= **640 (29%)**

**Penurunan late rate** =  $1319 - 640 = 679$  (51.4%) ⚡

## Potential Revenue Loss After Treatment

**Potential Revenue Loss After Treatment** = Revenue per product \* jumlah barang late

=  $181.78 * 640$

= **\$116339.2**

**Penurunan Potential Revenue Loss** = Potential Revenue Loss Before Treatment - Potential Revenue Loss After Treatment

=  $\$239767.87 - \$116339.2$

= **\$123428.64 (51.4%)** ⚡

# Rekomendasi **Bisnis #1**

## Late Notification dan Tracking Location



### Team MyStore

Uh-oh! We're sorry your package may be delayed with [carrier]. We'll keep you posted. [tracking\_link]

Microsoft Dynamics

- ✓ Two-way SMS
- ✓ Automated SMS through workflow
- ✓ Send to Accounts, Contacts, Leads and Users
- ✓ Send Emojis, Images and Files
- ✓ CRM to TxtSync Contact Sync
- ✓ TxtSync to CRM SMS Sync
- ✓ Quickly Send Bulk SMS From Marketing Lists
- ✓ Campaign Tracking & Delivery Reporting
- ✓ Outside CRM Roaming
- ✓ Inbound Screen Pops
- ✓ Natural Chat Window For SMS
- ✓ SMS Engagement Statistics
- ✓ SMS Contact Preference
- ✓ Auto Set Regarding
- ✓ Unlimited Users
- ✓ Unlimited Records

Starting from:

~~£650.00~~

£450.00

Discounts available - subject to approval (Don't be afraid to ask - we don't bite)

per organisation per month

# Rekomendasi **Bisnis #2**

## Discount Optimization

- Total biaya discount awal = \$ 297.334,62
- Total biaya discount >10 = \$198.511,26 (2.647 cust, disc. >10%)
- Asumsi 2.647 cust. dengan disc awal > 10%, menjadi 10% (max discount). Beban disc = \$ 51.933,1 (turun 74%)
- Total biaya discount new = \$ 150.756,46
- Selisih = \$ 146.578,16 (digunakan untuk alokasi shipment)

# Rekomendasi **Bisnis #3**

## Increase Man Power dan Handling Time

- Penambahan man power.
- Memperbaiki sistem antrian agar memberikan prioritas yang adil.
- Untuk package dengan berat <= 2.000 dan waktu order diterima < pukul 11.00, maka waktu handling < 1 hari.
- Penetapan waktu maksimal handling package adalah 1 hari setelah order diterima.
- Same-day delivery 75 packages/hari dapat menurunkan late rate hingga 34%.

# Agenda



**Latar Belakang**



**Insight**



**Modeling**



**Rekomendasi Bisnis**

Thank you for  
listening!

