Final Project

E-Commerce Shipping Data
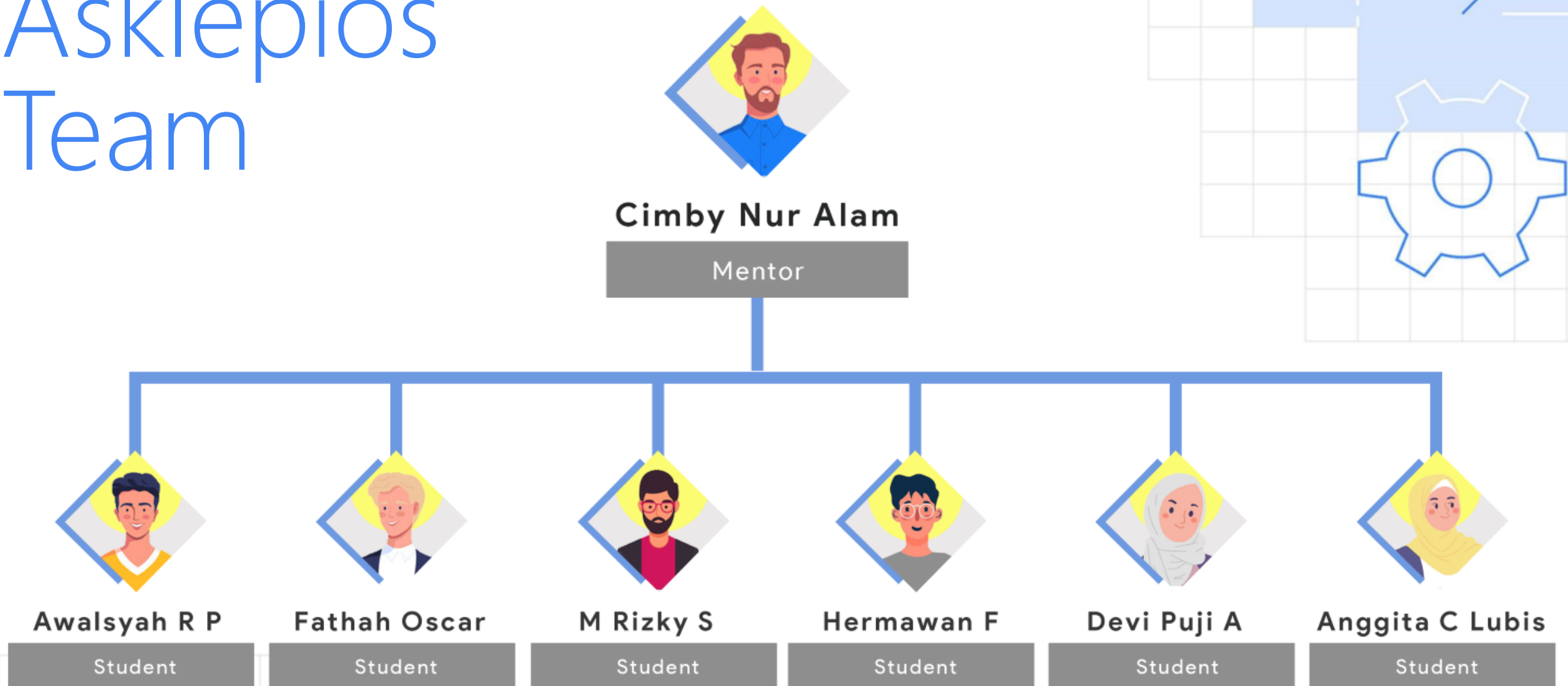
Asklepios

Data Scientist Team at Asklepios.ID

Asklepios

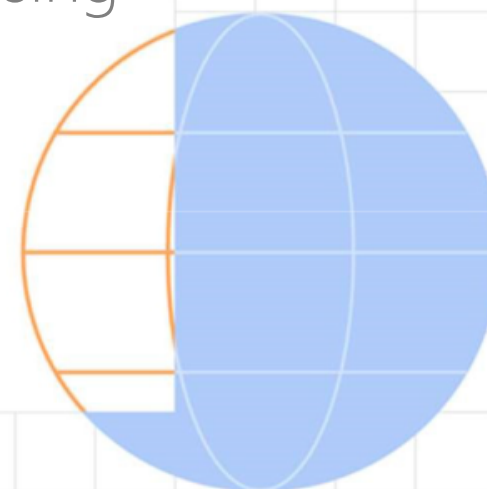# Asklepios Team

Cimby Nur Alam

Mentor

Awalsyah R P	**Fathah Oscar**	**M Rizky S**	**Hermawan F**	**Devi Puji A**	**Anggita C Lubis**
Student	Student	Student	Student	Student	Student

Asklepios

Table of Content

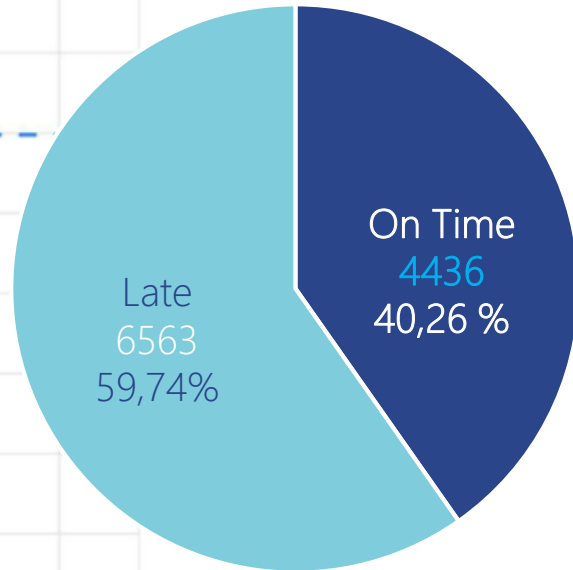Asklepios

Business Understanding

Business Background?

Project E-Commerce Shipping Data

- Membuat Model untuk Memprediksi Keterlambatan Pengiriman.
- Memberikan Business Insight dan Recommendation.

ASKLEPIOS TEAM

Data Science Consultant

Asklepios

E-Commerce Shipping Data

Problem Statement

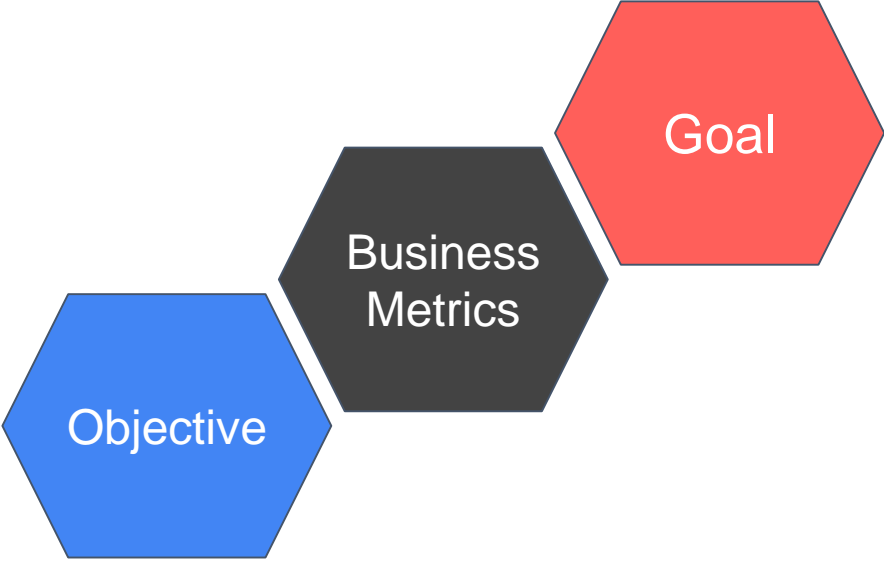Asklepios

Late Shipment Percentage

On Time
4436
40,26 %

Late
6563
59,74%

"Variabel Ketepatan Waktu Pengiriman berpengaruh signifikan terhadap kepuasan customer" ,

Berdasarkan Siburian dan Kartika (2021)

Siburian, Wahyuni dan Kartika, L.N. 2021. Pengaruh Kualitas Layanan, Persepsi Harga, dan Ketepatan Waktu Pengiriman Terhadap Kepuasan Pelanggan pada J&T Express di Perawang. *E-Jurnal Universitas Kristen Immanuel Yogyakarta*. Diakses melalui

Objective, Business Metrics, Goal.

Goal

Business Metrics

Objective

Finpro

Meningkatkan Customer Rating, Menaikan Persentase On-Time dan Menurunkan Potential Revenue Loss.

Customer Rating, Ontime Rate dan Potential Revenue Loss.

Memprediksi ketepatan waktu pengiriman barang terlambat/tidak dan memberikan rekomendasi metode pengiriman.

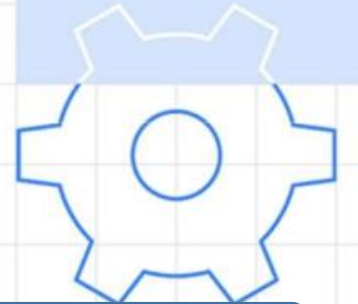# Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA)
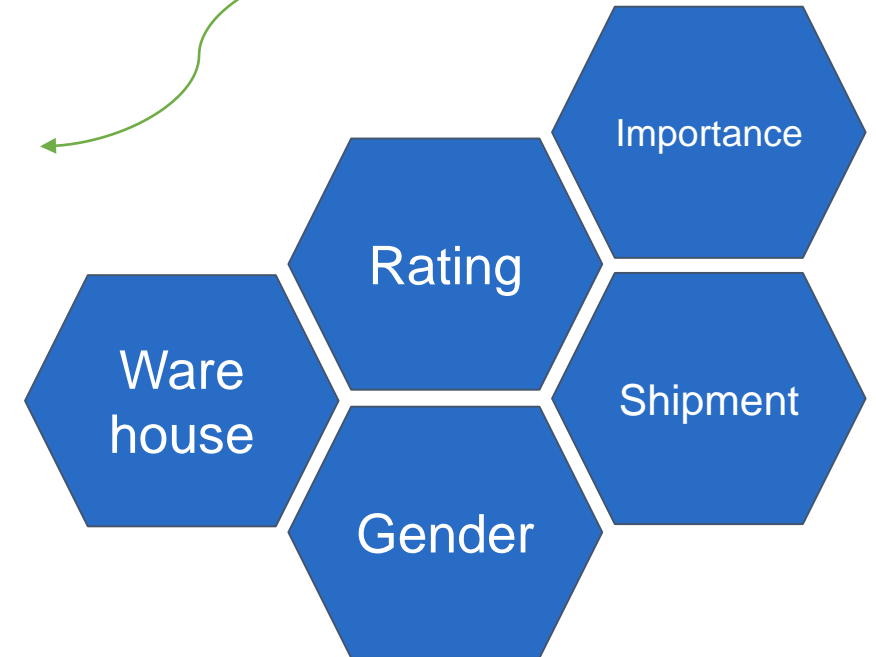
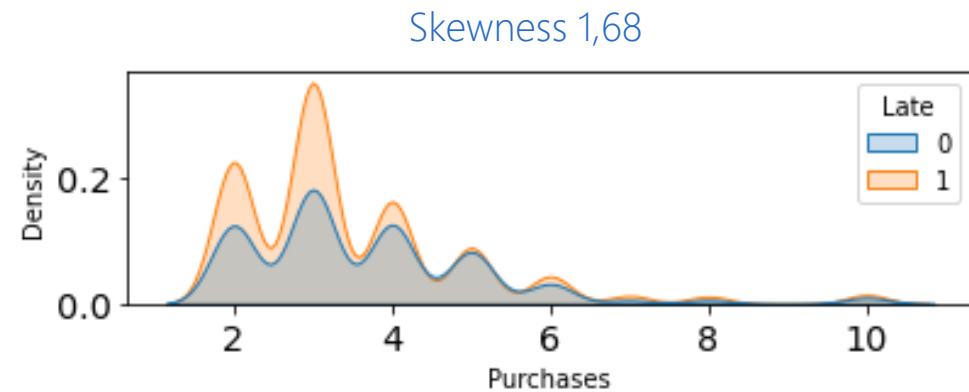| 10,999 Rows	10 Features	0 Duplicate	0 Missing Value

5 Numericals

1 Target

5 Categoricals
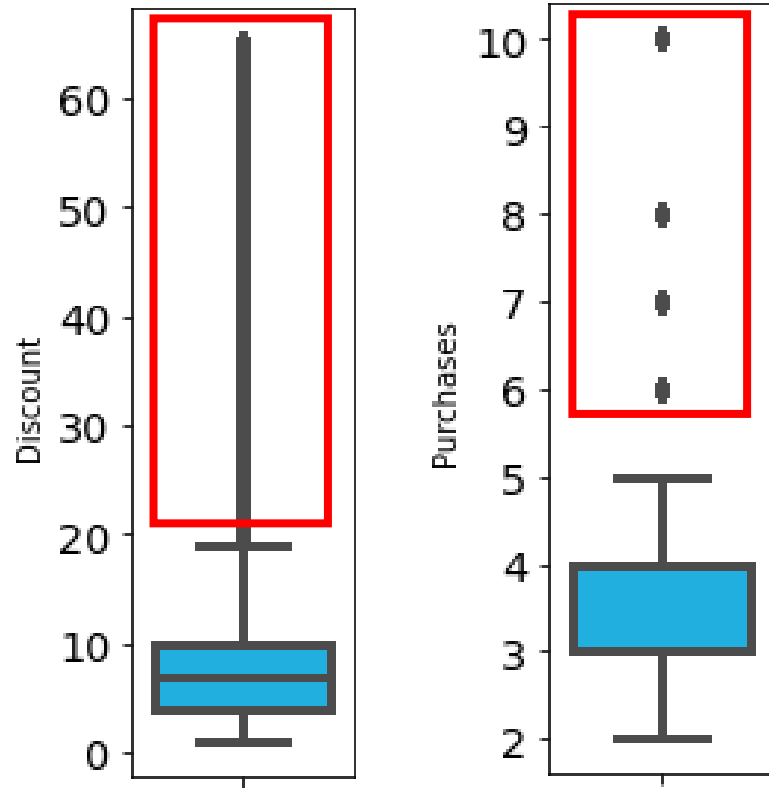
Cost

Discount

Weight

Calls

Purchases

Late

Importance

Rating

Ware house

Shipment

Gender

Asklepi

EDA - Univariate Analysis

Skewness 1,79

Skewness 1,68

Variabel Purchases dan Discount memiliki outlier dan membentuk pola positively Skewed
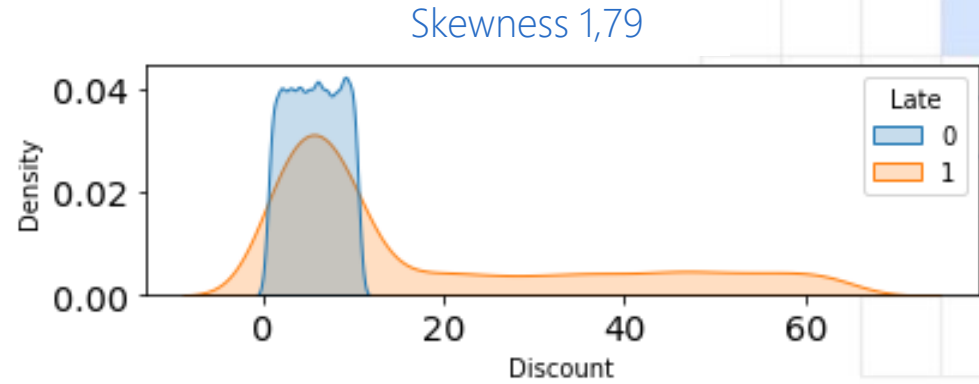
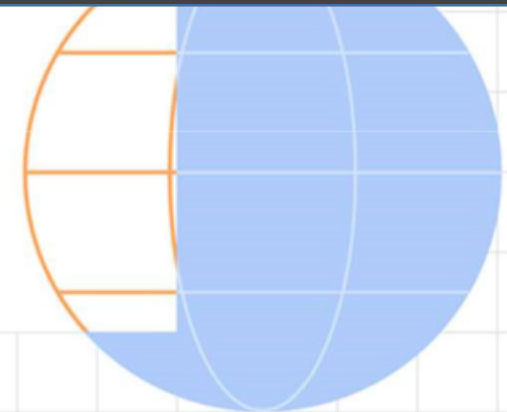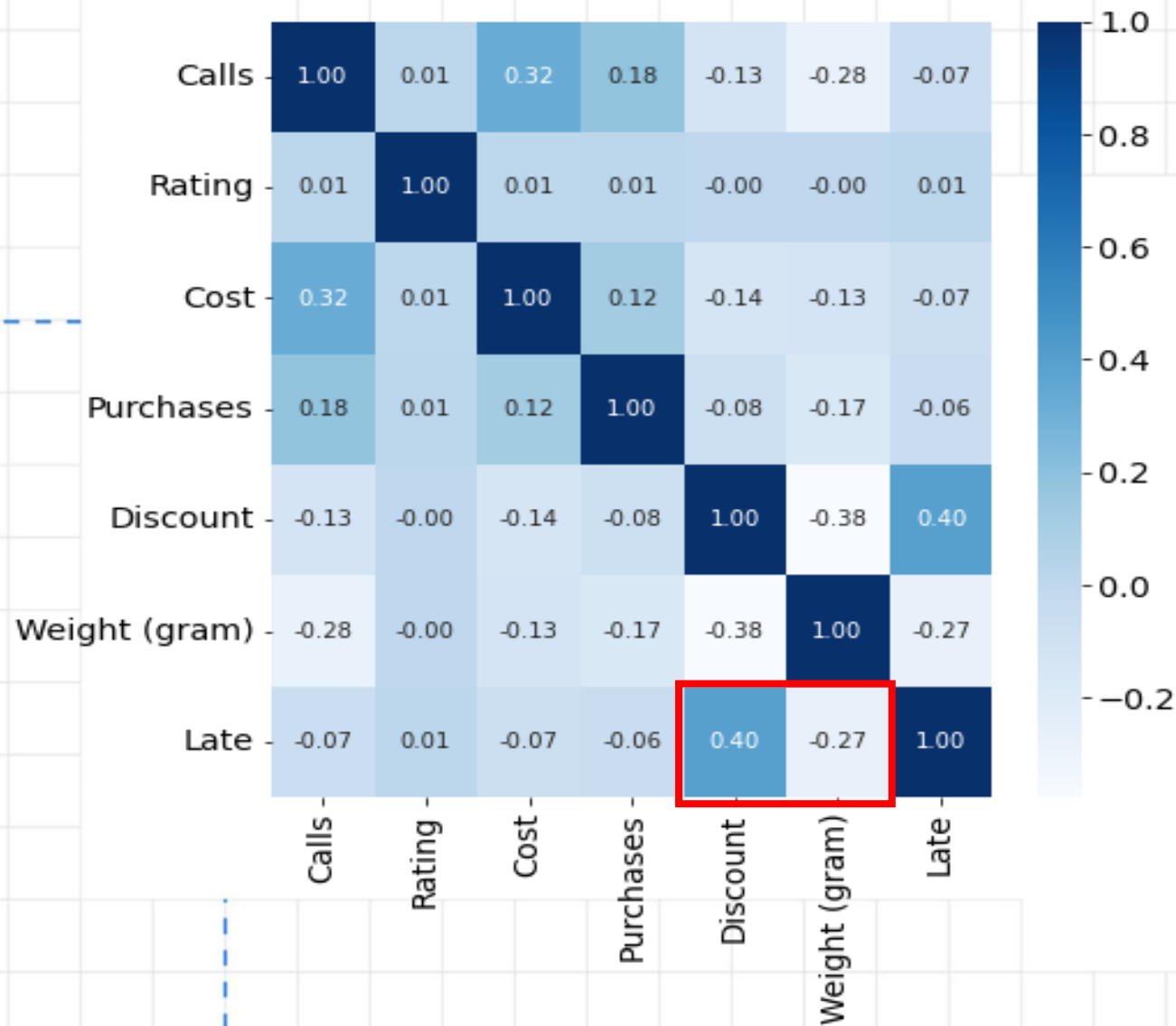# EDA – Univariate Analysis
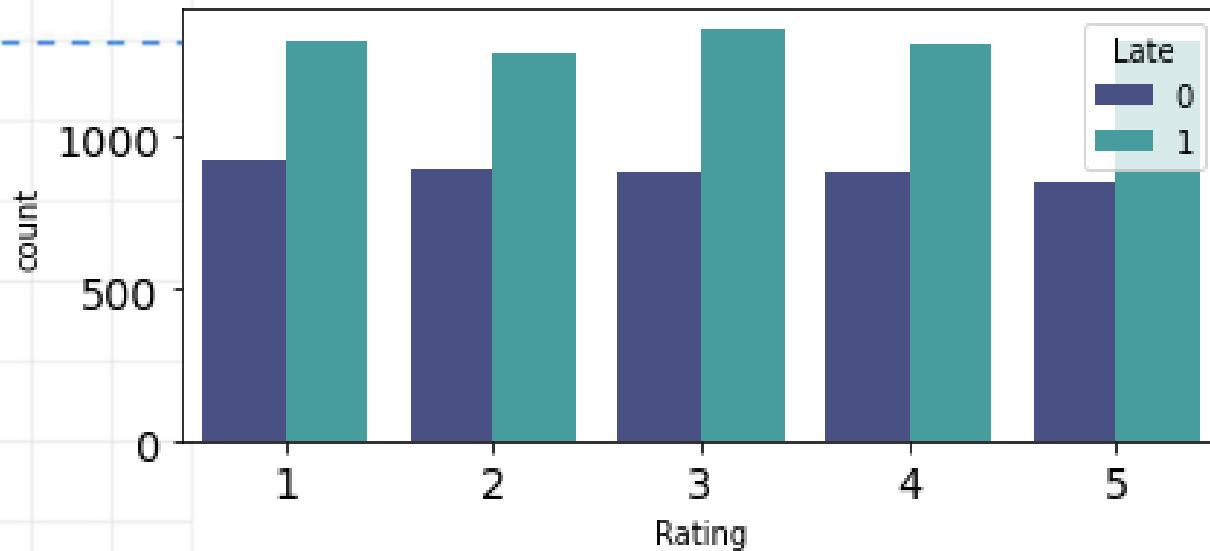
Dari countplot feature rating, customer yang memberikan rating 1-3 (6639 customer) lebih banyak daripada yang memberikan rating 4-5 (4360 customer), yang menandakan bahwa sebagian besar customer kurang puas terhadap shipment.
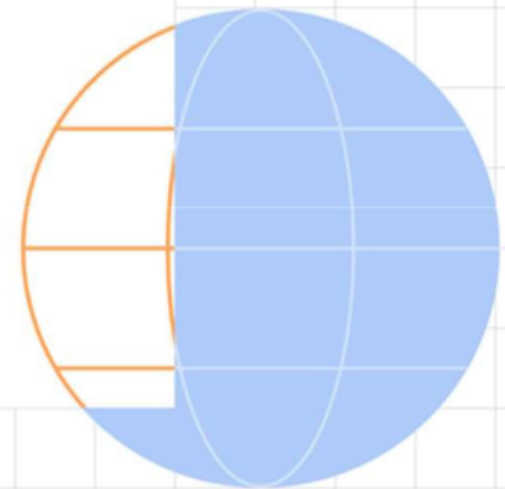
EDA – Multivariate Analysis

Feature Discount dan Weight adalah feature yang paling berkolerasi dengan feature target yaitu Late.

Pre-Processing

Feature	Dataset 1 (n = 10642)	Dataset 2 (n = 7794)
Discount Offered	Remove outlier z-score, log transformation, standardization	Remove outlier z-score + IQR, log transformation, standardization
Prior Purchase	Remove outlier z-score, log transformation, standardization	Remove outlier z-score + IQR, log transformation, standardization
Customer Care Calls	Standardization	
Cost of the Product	Standardization	
Weight (gram)	Standardization	
Product Importance	Label Encoding	
Gender	Label Encoding	
Warehouse Block	One Hot Encoding	
Shipment Mode	One Hot Encoding	
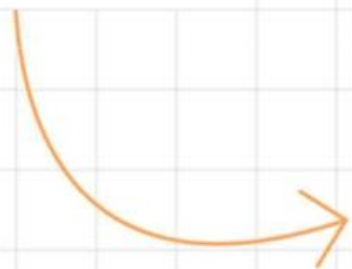
Modeling

Modeling

Asklepios

Algoritma

- Logistic Regression
- k-Nearest Neighbor
- Decision Tree
- Random Forest
- AdaBoost
- XGBoost

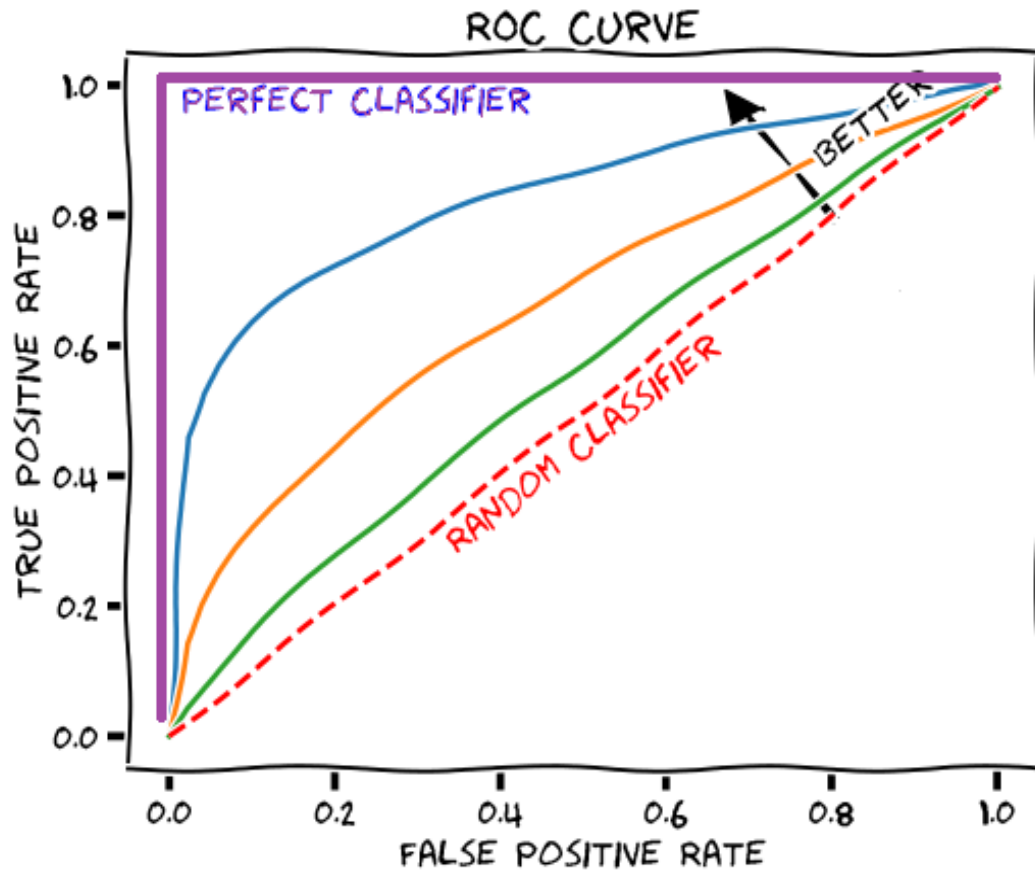## Metode Score Evaluasi

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

Modeling

ROC-AUC memiliki sifat robust terhadap dataset yang imbalance pada target, bagus untuk urgensi value false negative dan false positive.

Pada dataset terdapat 59.7% Late, 40.3% On Time.

Finpro   Asklepios

Modeling

True Class

	Positive	Negative
Predicted Class Positive	TP	FP
Negative	FN	TN

False Positive : Model memprediksi pengiriman Late, aktual On Time.

Impact :   failed to meet customer's expectation.

False Negative: Model memprediksi pengiriman On Time, aktual Late.

Impact: failed to meet customer's expectation.

Modeling

ROC-AUC					
Dataset 1 (Removing Outlier Z-Score)					
Method	Logreg	kNN	Decision Tree	Random Forest	AdaBoost	XGBoost
Train	0.71	0.76	0.79	0.75	0.76	0.79
Test	0.72	0.72	0.72	0.73	0.74	0.73

ROC-AUC					
Dataset 2 (Removing Outlier Z-Score & IQR)					
Method	Logreg	kNN	Decision Tree	Random Forest	AdaBoost	XGBoost
Train	0.58	0.58	0.61	0.62	0.63	0.88
Test	0.58	0.64	0.57	0.60	0.61	0.60
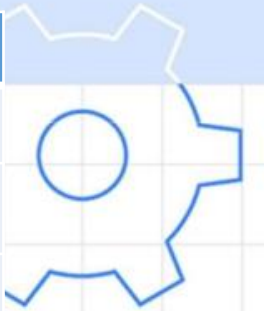
Asklepios

Modeling

Hyperparameter Terbaik (ROC-AUC Dataset 1)
AdaBoost
n_estimator = 225
learning_rate = 0.08358
algorithm = SAMMER. R
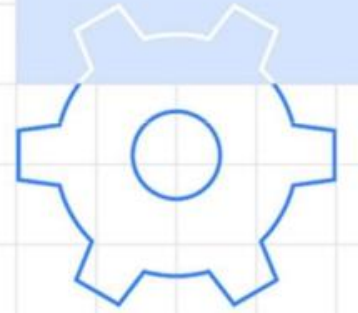
Asklepios

Adaboost Confusion Matrix
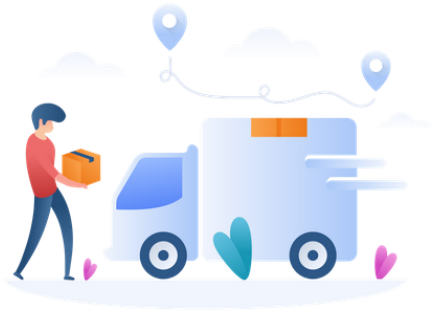
Confusion Matrix	Predicted Label	
Actual Label	True Positive 1175 (36.68 %)	False Negative 697 (21.82%)
	False Positive 363 (11.36 %)	True Negative 958 (30.00%)

Positive = Late   Negative = On Time
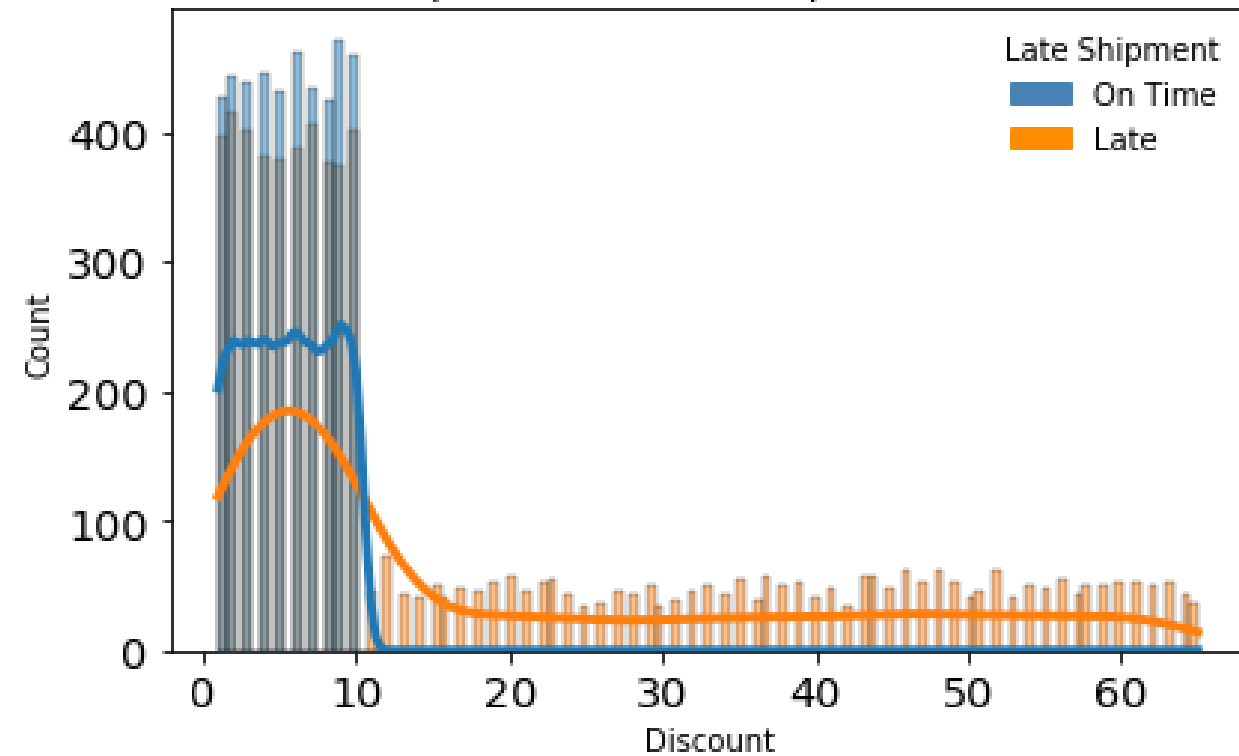
Adaboost Feature Importance
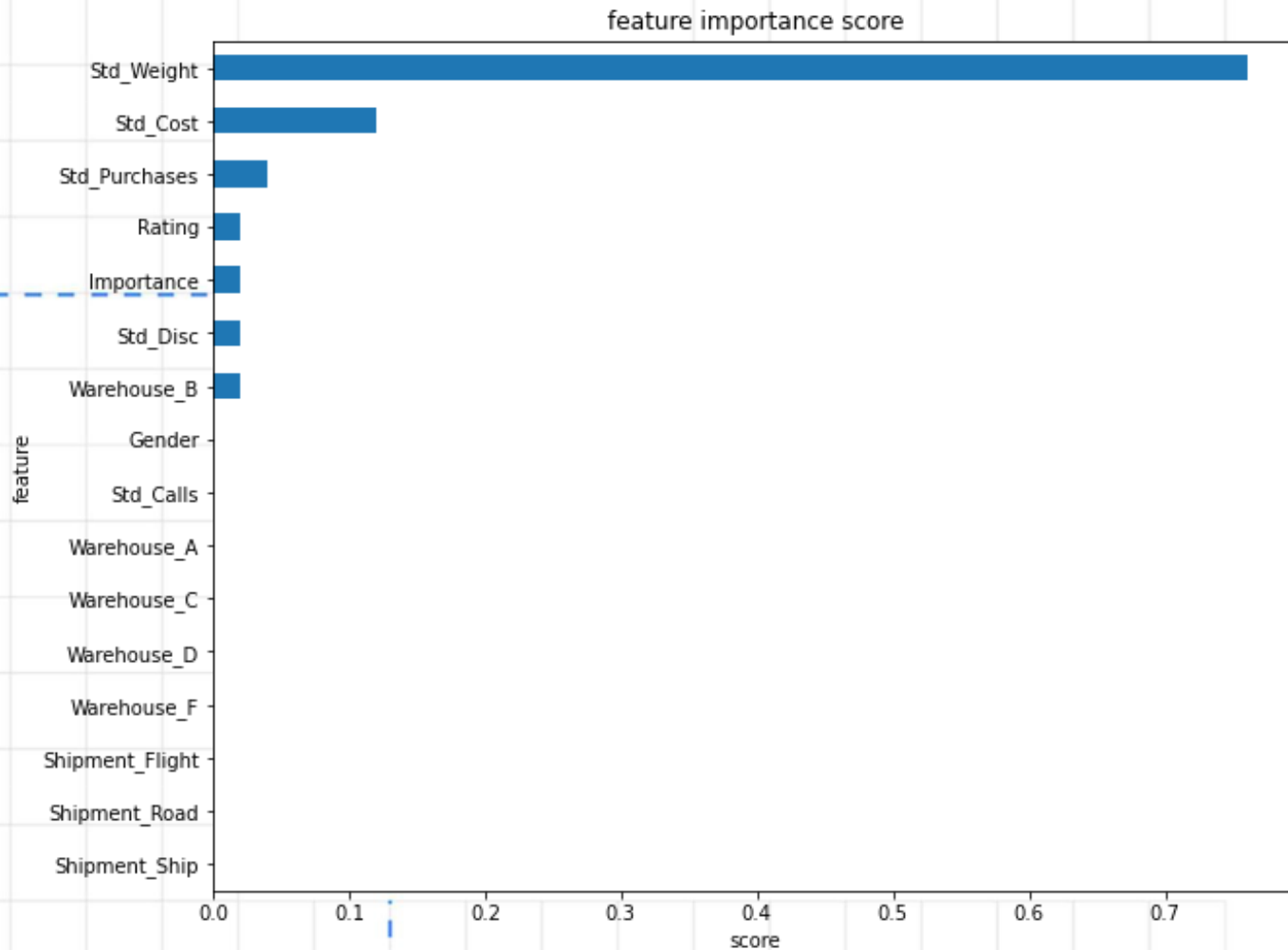
Feature Importance :
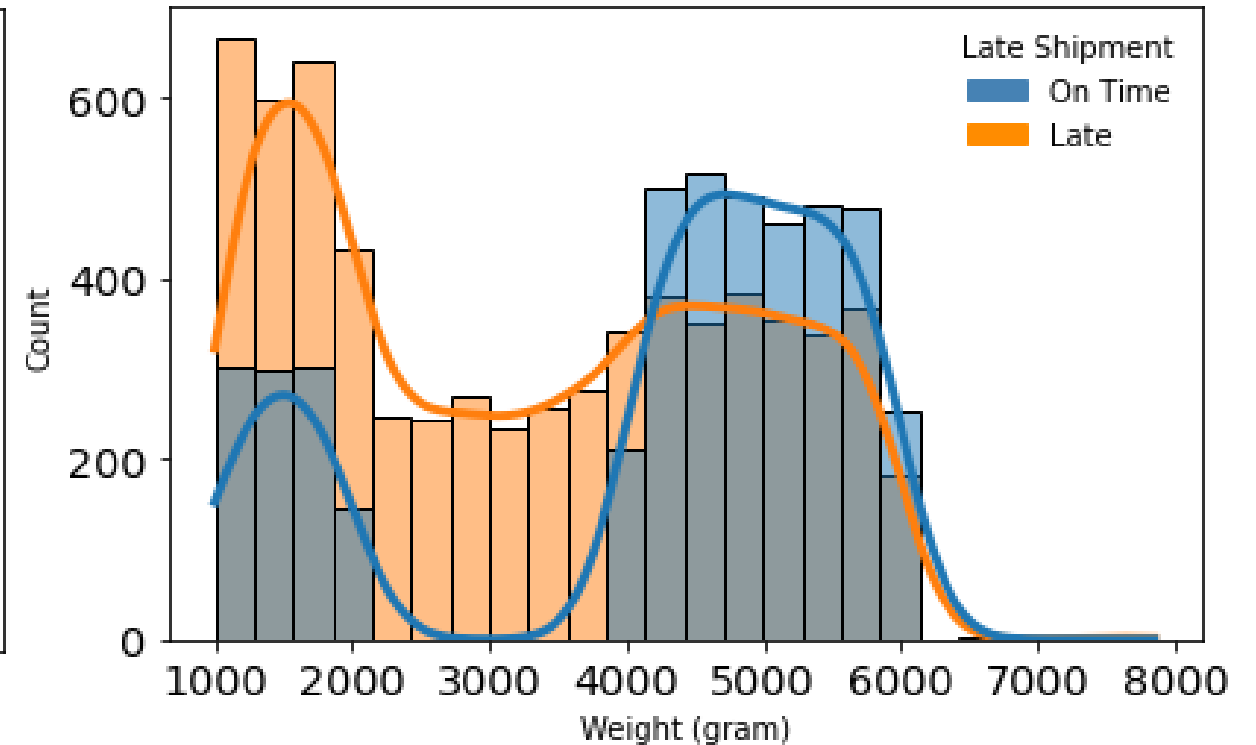
- Std_Weight

- Std_Cost

- Std_Purchases

Business Insight &
Recommendation

Business Insight
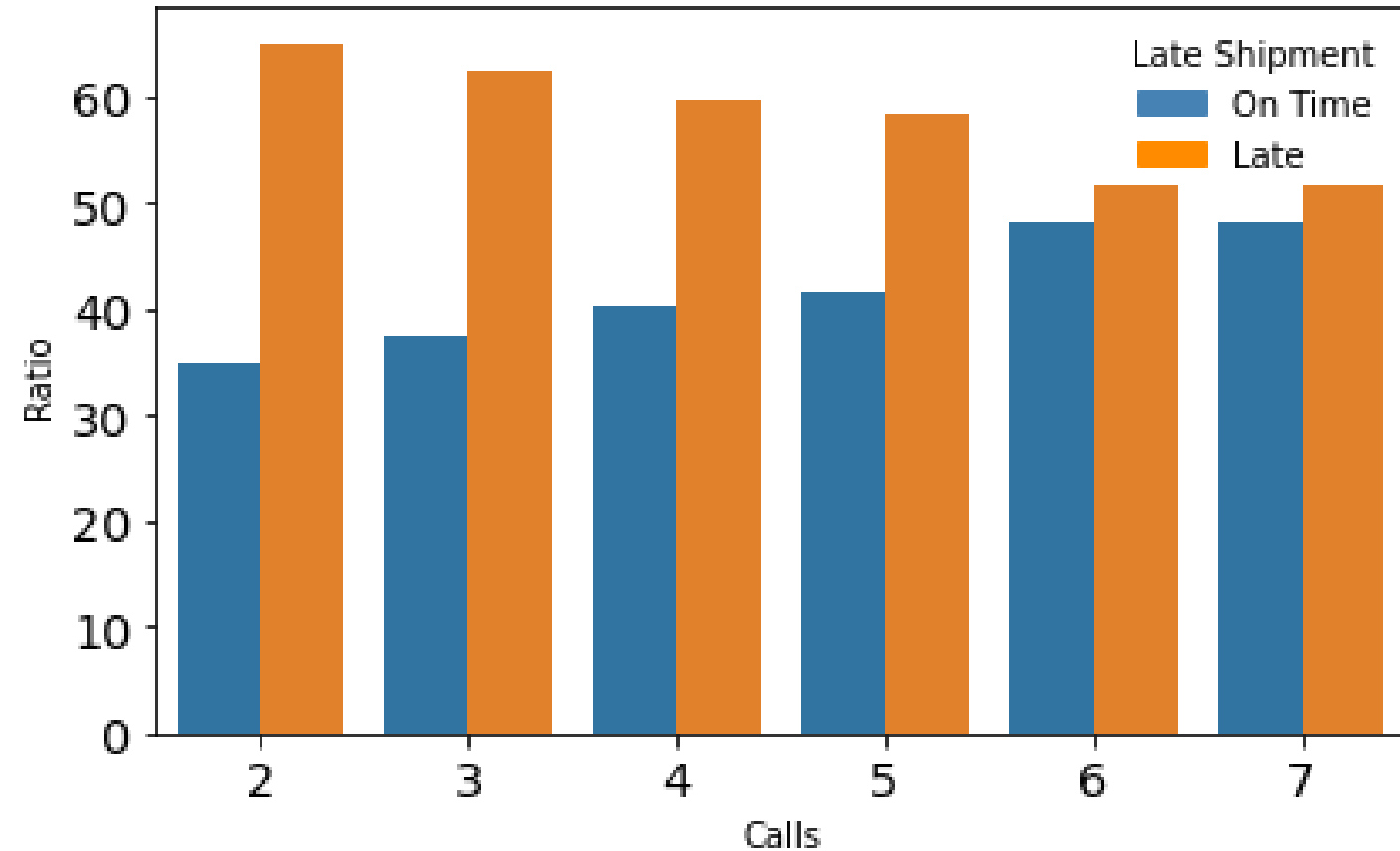
# Business Insight

Relationship Between Late Shipment and Customer Care Calls

Berdasarkan data jumlah keterlambatan pengiriman menurun dengan meningkatnya jumlah telepon yang diterima oleh customer care.

Perusahaan perlu mencari informasi mengenai isi telepon customer kepada customer care (siapa penelpon, isi telepon), sehingga bisa menentukan korelasi dengan jumlah keterlambatan pengiriman.

Recommendation
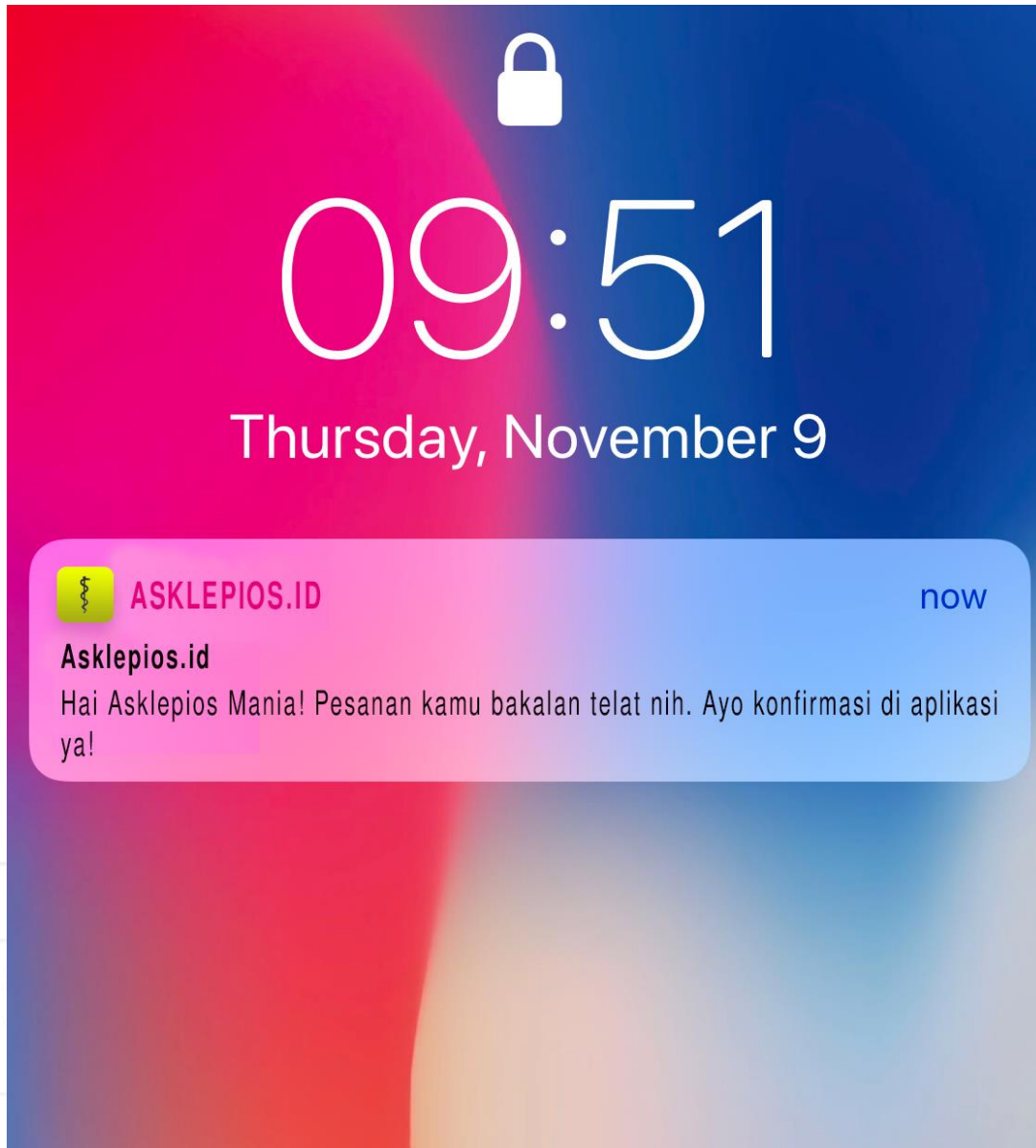
- Membuat sistem notifikasi untuk pengiriman yang terlambat agar customer rating tidak turun.

- Menambahkan estimasi waktu pengiriman pada mode pengiriman yang dipilih.
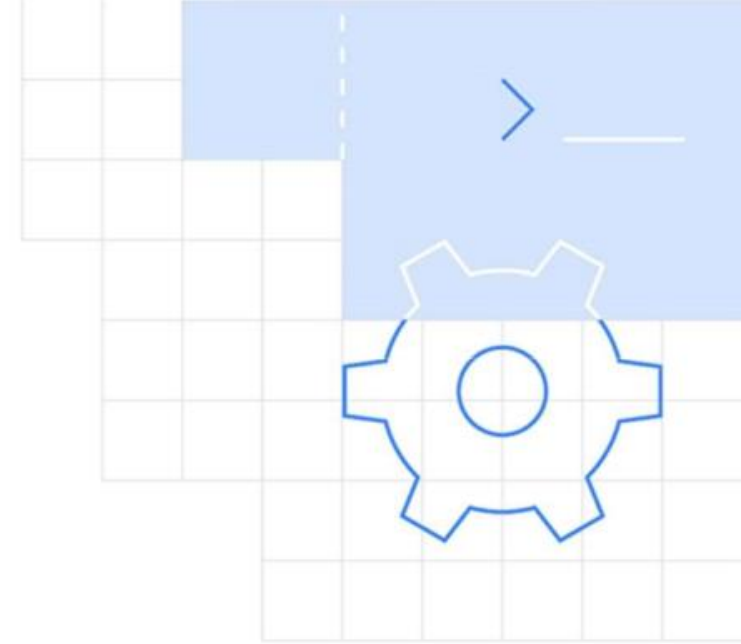
Models Impact to Business Metrics

- On Time Growth Rate
- Customer Rating
- Potential Revenue Loss

On Time Growth Rate Calculation

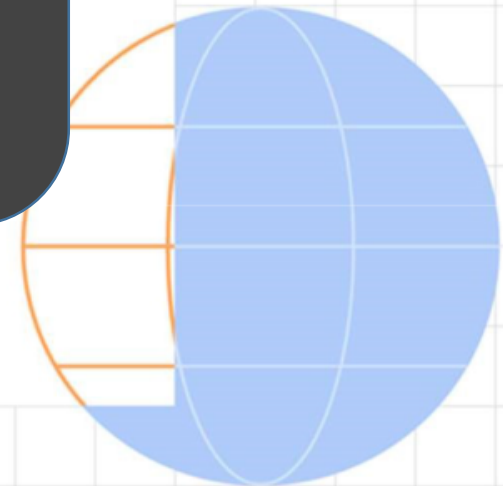| Before Modelling	Count	Percentage
Total Delivery	10999	100 %
Late Delivery	6563	59.7 %
On Time Delivery	4436	40.3 %

On Time Rate	Percentage
Before Modelling	40.3 %
After Modelling	84.48 %
Growth Rate	44.18 %

After Modelling	Count	Percentage
Total Delivery	10999	100 %
Late Delivery	6563	59.7 % of Total Delivery
Predicted Late on Late Delivery	4857	74 % of Late Delivery
Predicted On Time on Late Delivery or Late after Model Prediction	1706	26 % of Late Delivery
On Time Delivery	4436	40.3 % of Total Delivery
On Time Delivery after Model Prediction	9293	84.48 % of Total Delivery
On Time Growth Rate	-	+ 44.18 %

Customer Rating Calculation

Variabel	Jumlah
Initial total rating	32.893
Initial average rating	2,99
Total additional rating	3.882
Updated total rating	32.893 + 3.882 = 36.675
Updated average rating	3,34

Customer Rating	Rating
Before Modelling	2,99
After Modelling	3,34
Growth Rate	11, 8 %

Asumsi : Tiap customer late terprediksi late menaikan rating 1 poin.

Syarat : Customer late dengan rating 5 tidak bisa menaikkan rating lagi.

Potential Revenue Loss Calculation

Asklepios

Dalam studi kasus ini, potential revenue loss berasal dari customer yang mengalami late.

Variable	Total ($)
Total Sales	2.311.955
Total Discount	297.334,62
Total Revenue	2.014.620,38
Total Revenue Customer Late	1.200.713,75
Total Revenue Customer late Predicted late	885.828,17
Total Revenue Customer Late Predicted On Time	312.185,57

Potential Revenue Loss	Total ($)
Before Modelling	1.200.713,75
After Modelling	312.185,57
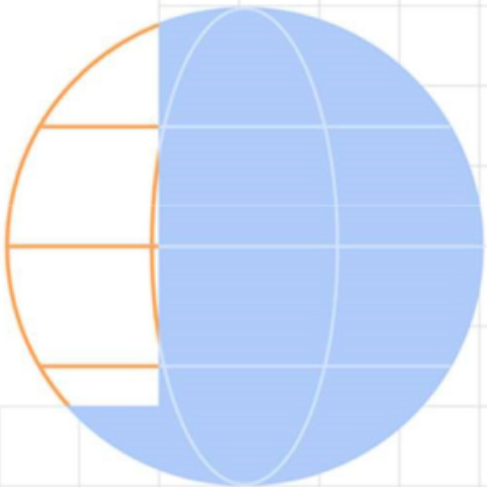
Comparison

	Metrics	Before	After	Growth Rate
	On Time Rate	40.3%	84.48%	44,18%
	Customer Rating	2.99	3.34	11,8%
	Revenue Loss	$1,200,713.75	$312,185.57	-74%

Finpro

Asklepios

Thank You.

ASKLEPIOS
TEAM