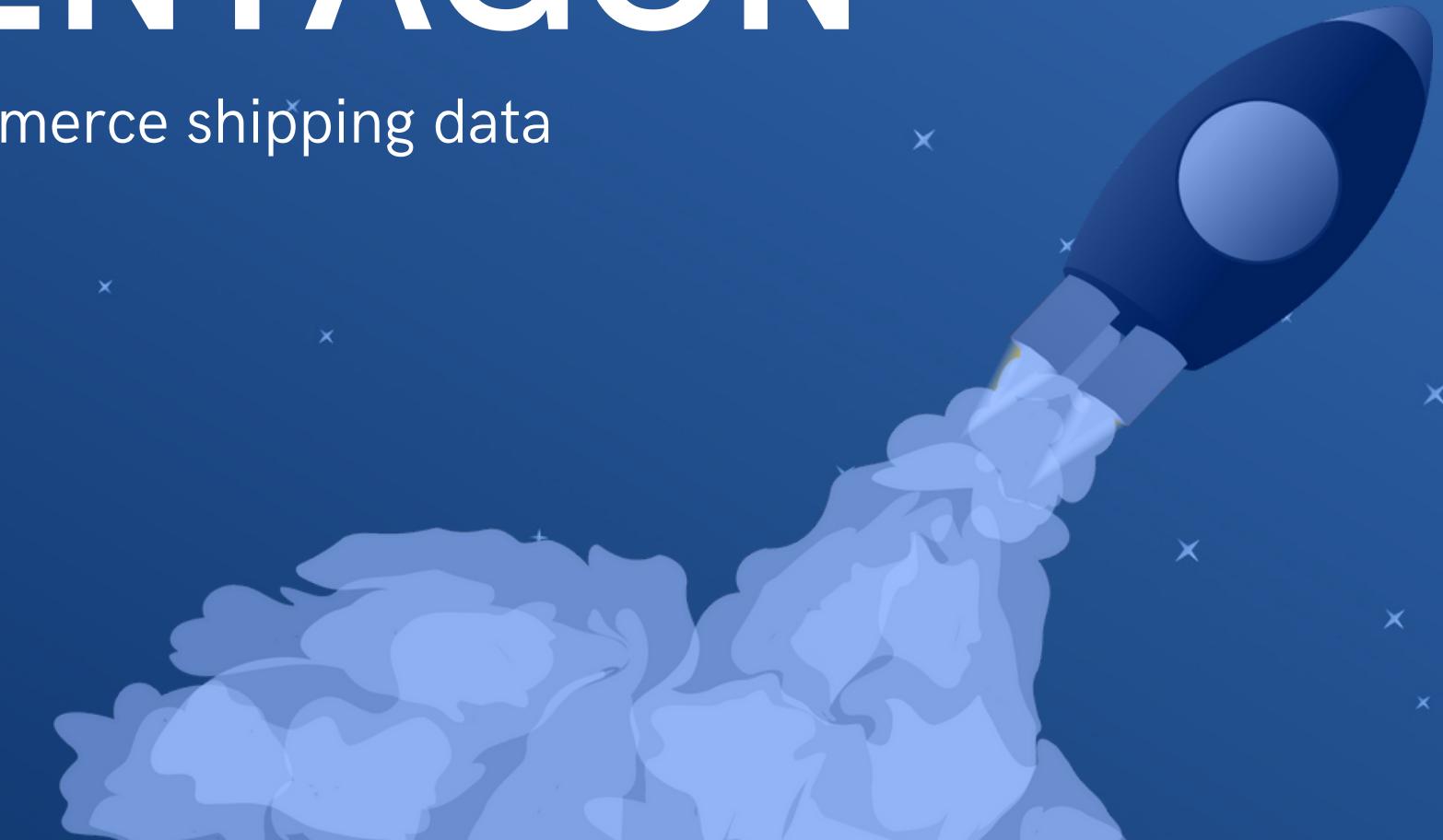




Business Presentation

# Final Project PENTAGON

e-commerce shipping data



# BUSINESS ANALYST TEAM



M. Hendrawan  
(Mentor)



Ilman  
Naafian



Freddy  
Kurniawan



Salma  
Immala



Divyah  
Laskmi



Akmalia  
Lutfia



Putri  
Qonitah

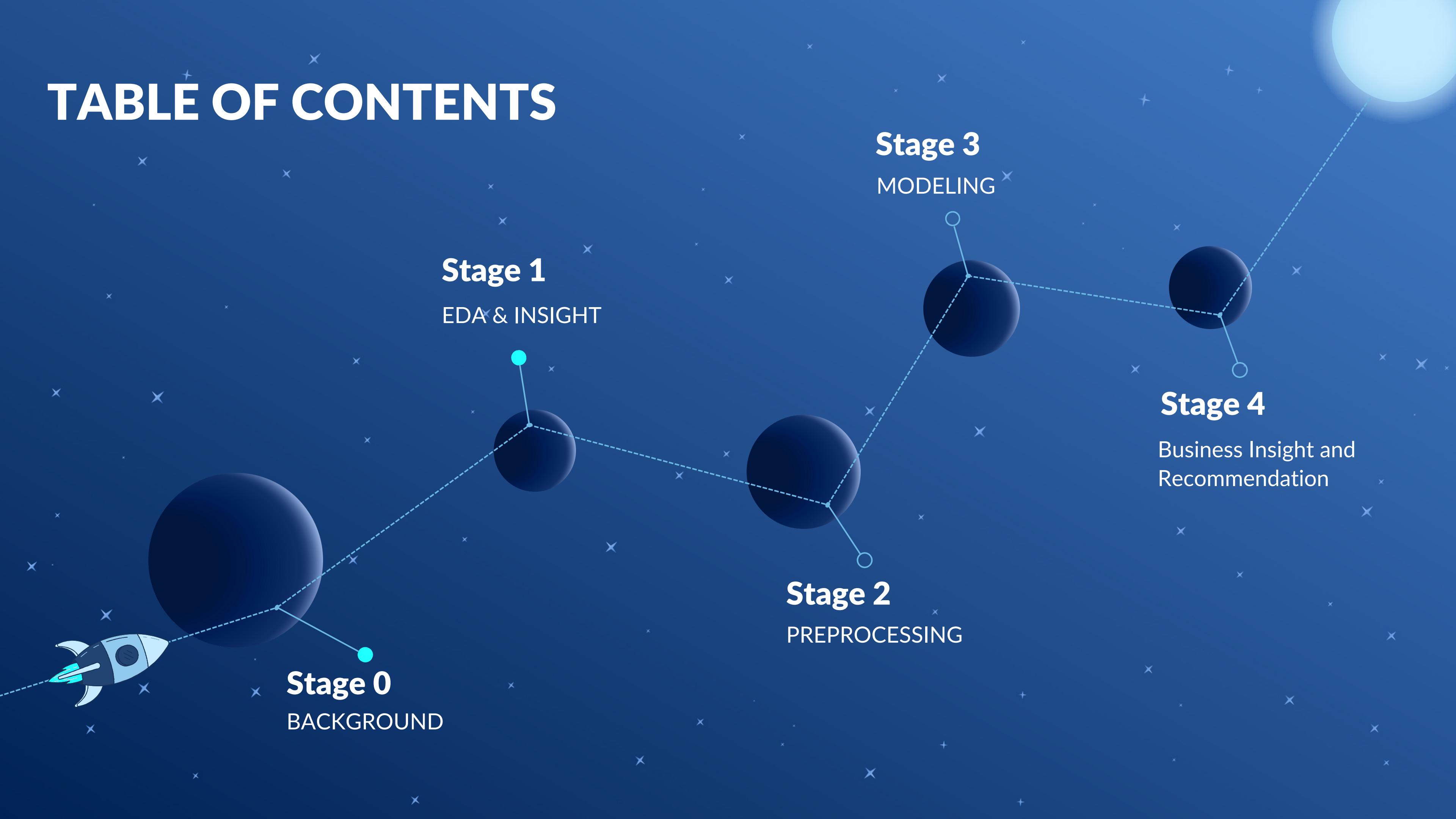


M Luthfi  
Alfatih



Adam  
Adiwangsa

# TABLE OF CONTENTS



# Background

**59.7% of shipments  
experiencing late delivery**

\*Research use 12 months shipments data

On Time  
40.3%

Late  
59.7%



# Our Objective



## Goals

Decrease late delivery rate become  
**10% in 12 months**



## Business Metrics

On time delivery rate



## Objective

Create a late delivery model prediction to provide internal and customer notification



# Dataset Overview

1 Year Historical Data

10999  
Transaction

10  
Features

0 Missing  
Value

0  
Duplicate

## Numerical Features

5 Features

- Calls
- Cost
- Prior purchases
- Discount
- Weight gram

## Categorical Features

5 Features

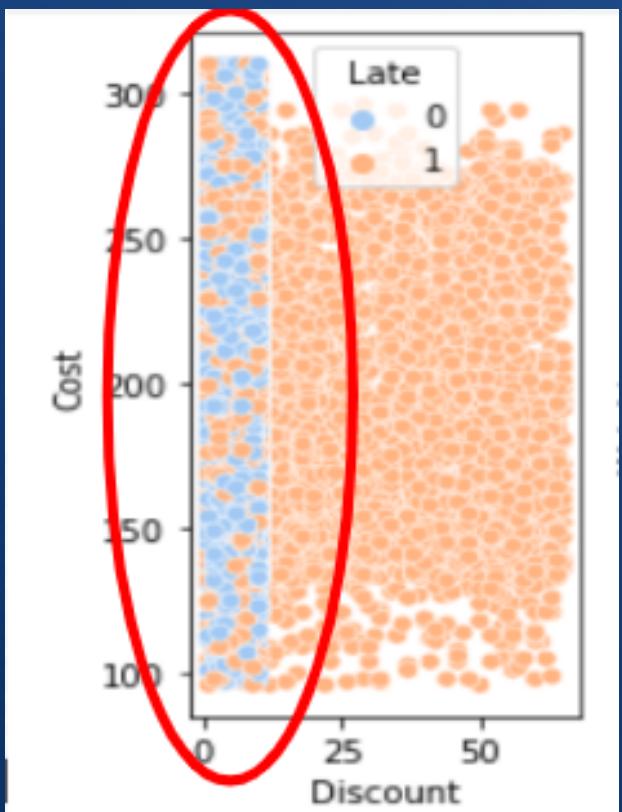
- Warehouse
- Shipment
- Rating
- Importance
- Gender

## Target

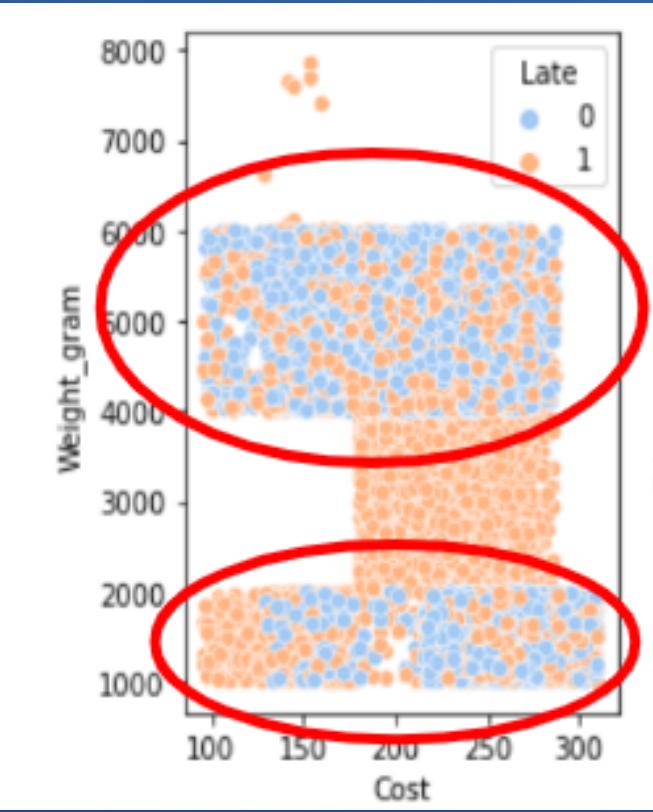
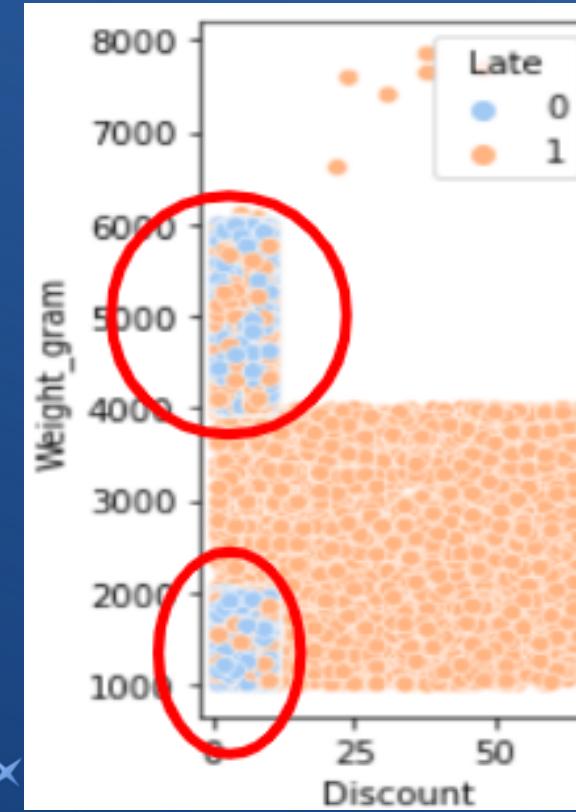
1 Target

Late

# Exploratory Data Analysis



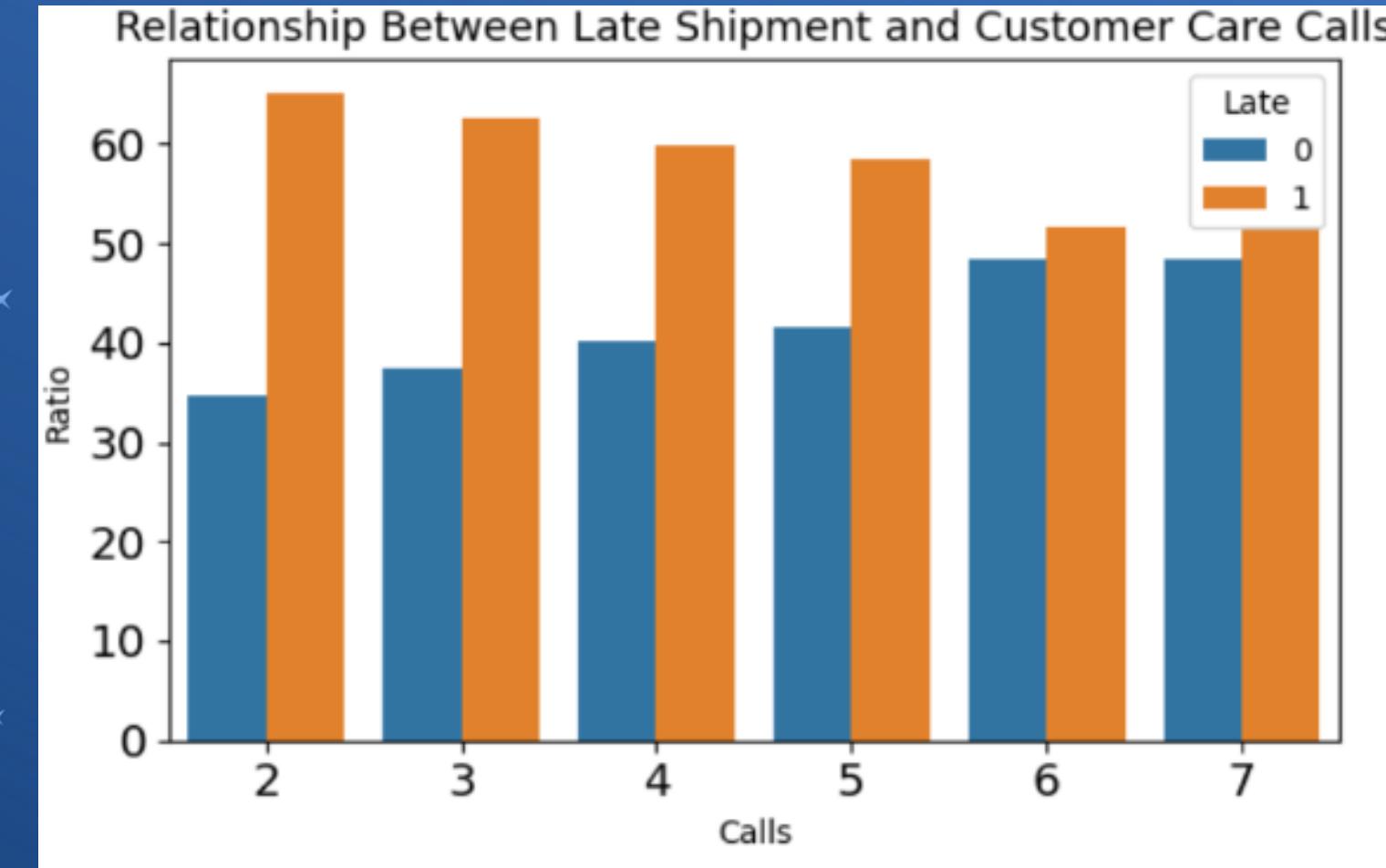
Most of on-time delivery comes from packages with discount <15%. This means that package with higher discount tends to be late.



On-time delivery comes from packages with weight between 1000 gr <= weight <= 2000 and 4000 gr <= weight <= 6000

0 = Not Late  
1 = Late

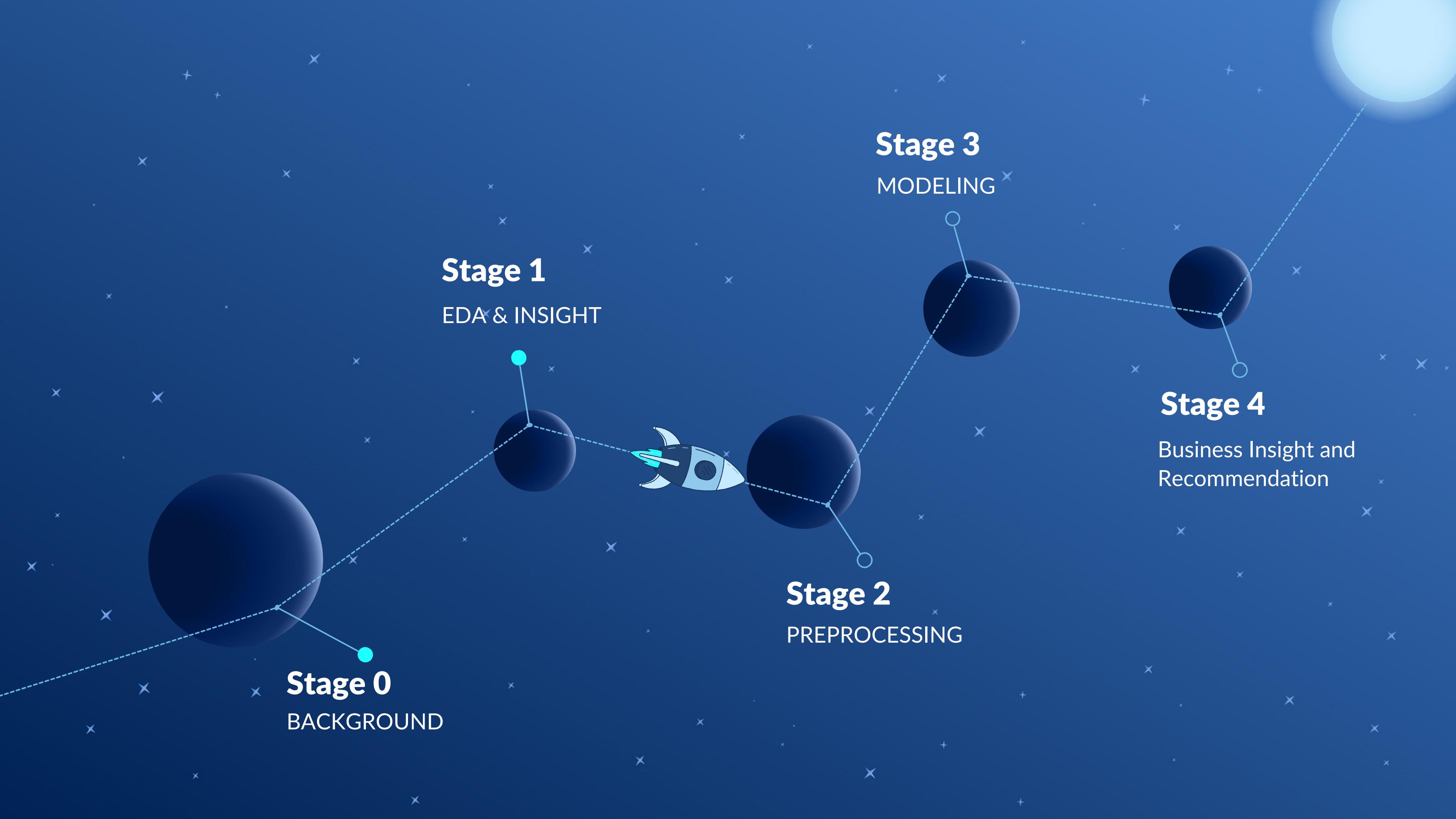
# Exploratory Data Analysis



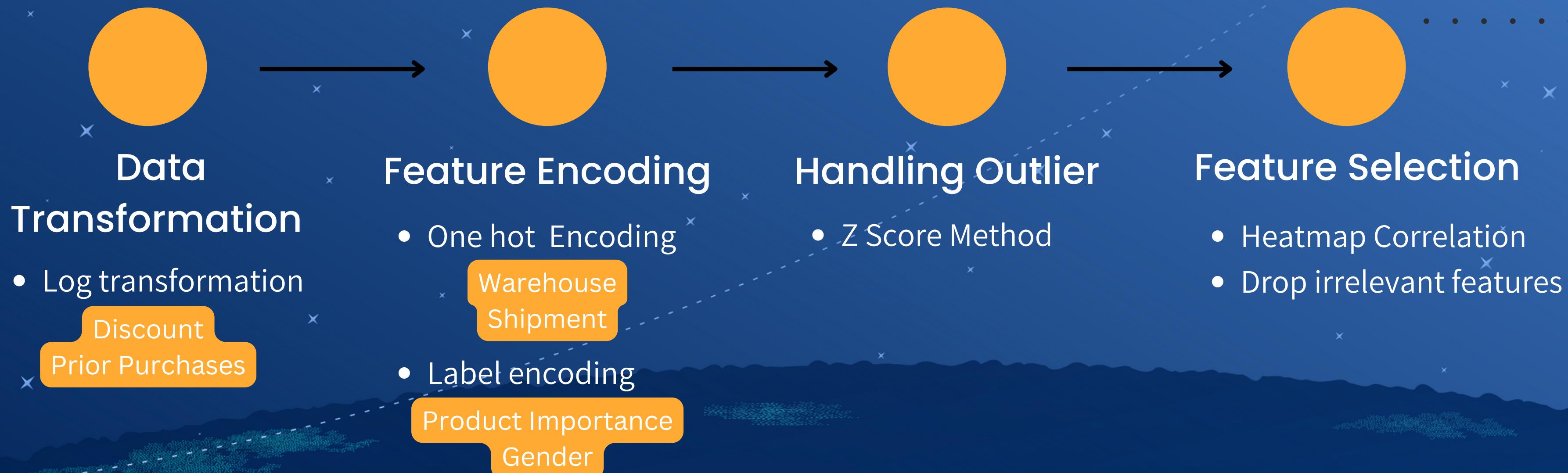
- From heatmap correlation discount (0.4) & weight (0.3) has the highest correlation with late.
- From hypothesis test (Anova) it shows that there are significant difference in Calls for late and not late. This means that calls also has a correlation with late packages
- Based on the histogram, the higher the number of the calls the more on-time the package arrived.

0 = Not Late

1 = Late



# Data Pre - Processing





# MODELLING : DATASET

Dataset 1

## Features:

1. Calls
2. Rating
3. Cost
4. Importance
5. Gender
6. Weight\_gram
7. Log\_discount
8. Log\_prior\_purchases
9. Warehouse\_A
10. Warehouse\_B
11. Warehouse\_C
12. Warehouse\_D
13. Warehouse\_F
14. Shipment\_Flight
15. Shipment\_Road
16. Shipment\_Ship

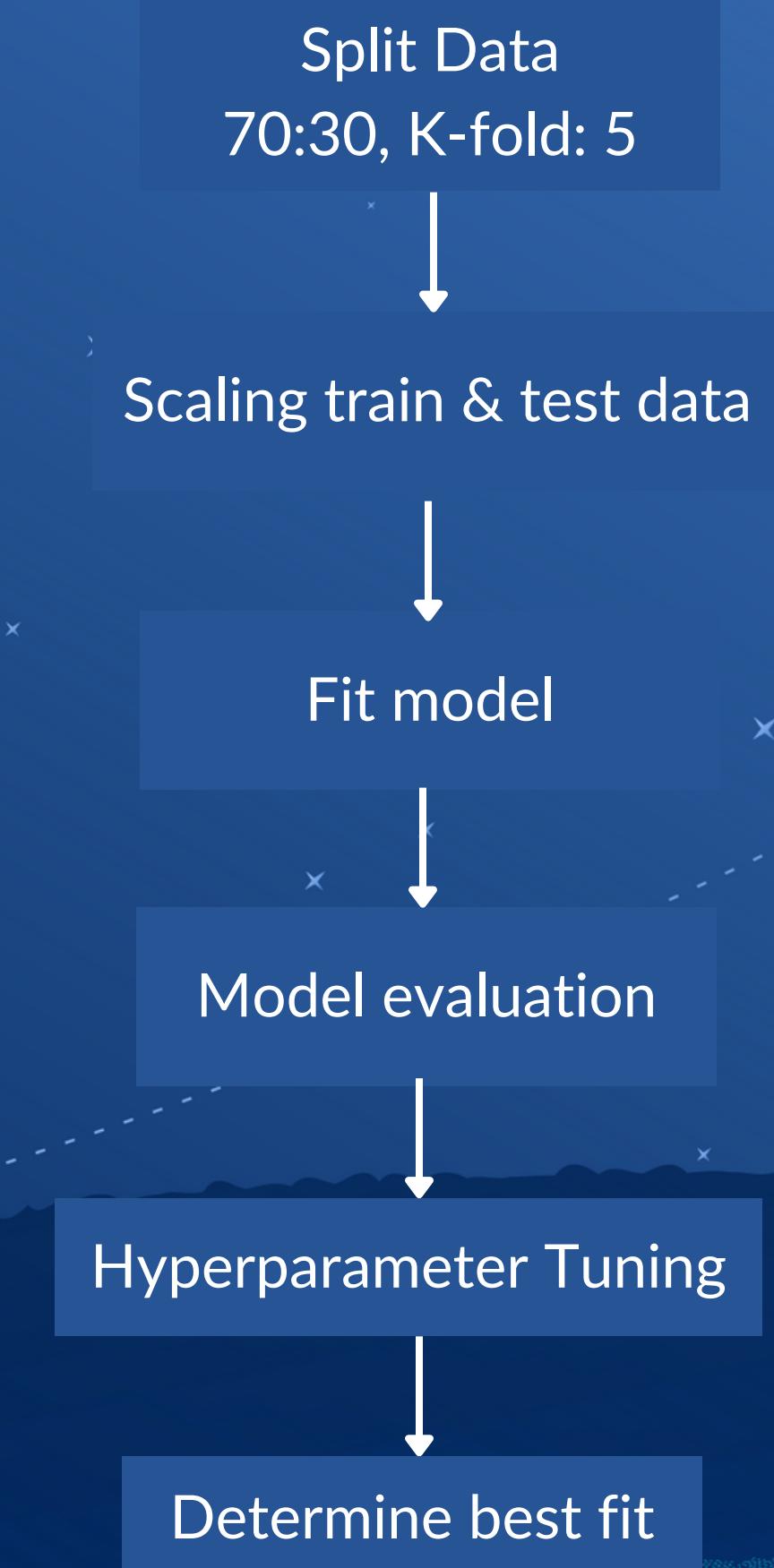
Dataset 2

(Remove the features  
with a correlation <0.05  
to the target)

## Features:

1. Calls
2. Cost
3. Weight\_gram
4. Log\_discount
5. Log\_prior\_purchases

# MODELLING: FLOWCHART



# MODELLING : EVALUATION METRICS

- ROC-AUC are chosen as evaluation metrics because the goal of the model is to predict whether incoming order have potential to become late.
- Thus, both false positive and false negative would be critical to ensure cost efficiency, if too many data had false positive error, this will results in overworking the logistics. Too high of false negative errors also cause drop in company ratings in the eye of customers as customers have negativity bias as stated by Geng Cui, et.al studies.

**EVALUATION METRIC  
ROC-AUC**

# MODELLING : SUMMARY

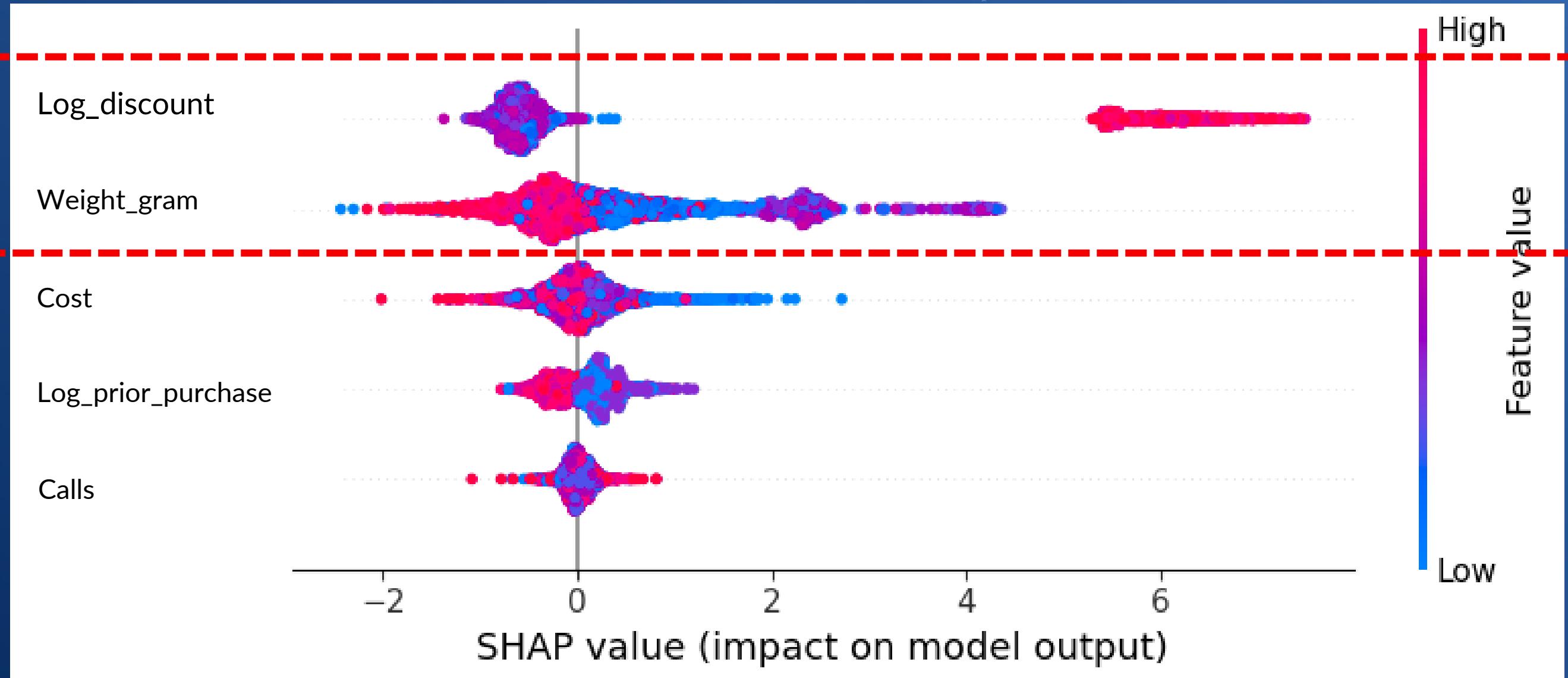
ROC-AUC							
Dataset 1 (Menggunakan semua fitur)							
Method	LogReg	Random Forest	Decision Tree	kNN	Adaboost	XGBoost	Naive Bayes
Train	0.72214	1.0	1.0	0.85295	0.76528	0.97187	0.73057
Test	0.70488	0.72748	0.61778	0.67225	0.72077	0.72921	0.70238
Tuning Train	-	0.75946	0.81438	0.77601	-	0.77837	-
Tuning Test	-	0.71815	0.72242	0.68958	-	0.72462	-

ROC-AUC							
Dataset 2 (Drop fitur dengan korelasi <0.05)							
Method	LogReg	Random Forest	Decision Tree	kNN	Adaboost	XGBoost	Naive Bayes
Train	0.72043	1.0	1.0	0.86094	0.76350	0.95499	0.72788
Test	0.70475	0.73466	0.61833	0.70688	0.71894	0.73631	0.70279
Tuning Train	-	0.76914	0.81125	0.80046	-	0.77932	-
Tuning Test	-	0.72496	0.73150	0.721312	-	0.72565	-

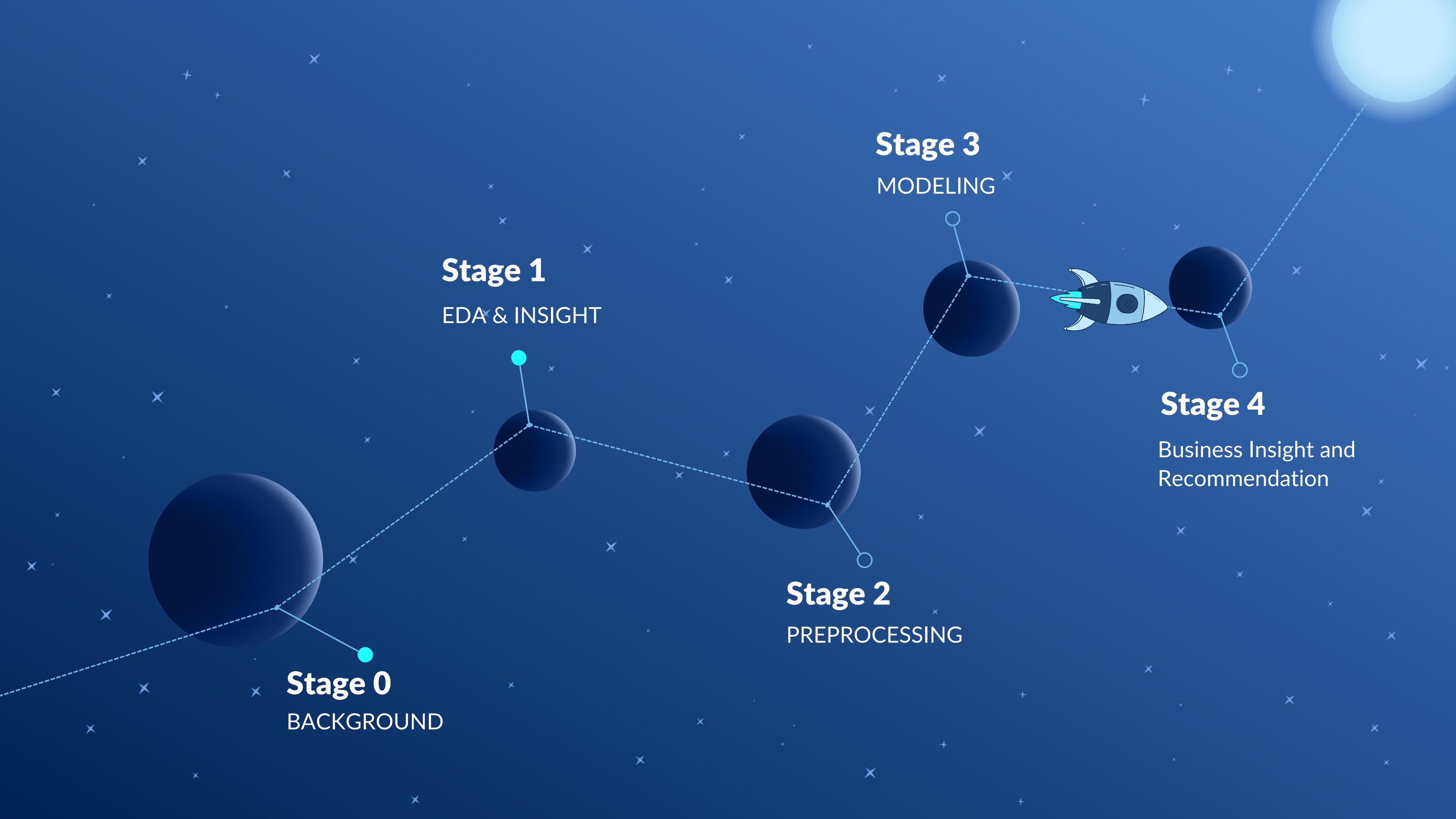
- Both dataset are fitted into several models.
- Models that had ROC-AUC gap <0.1 are not tuned as it's deemed already at best fit.
- Dataset 2 are chosen as the feature selection for the predictive model as it shows little difference in terms of ROC-AUC score and have faster processing time due to lower feature count.
- Based on ROC-AUC (crossval) score, it could be concluded that best predictive model is XGBoost, due to highest ROC-AUC score, and with small gap between train and test (<10% or <0.1).

# FEATURE IMPORTANCE

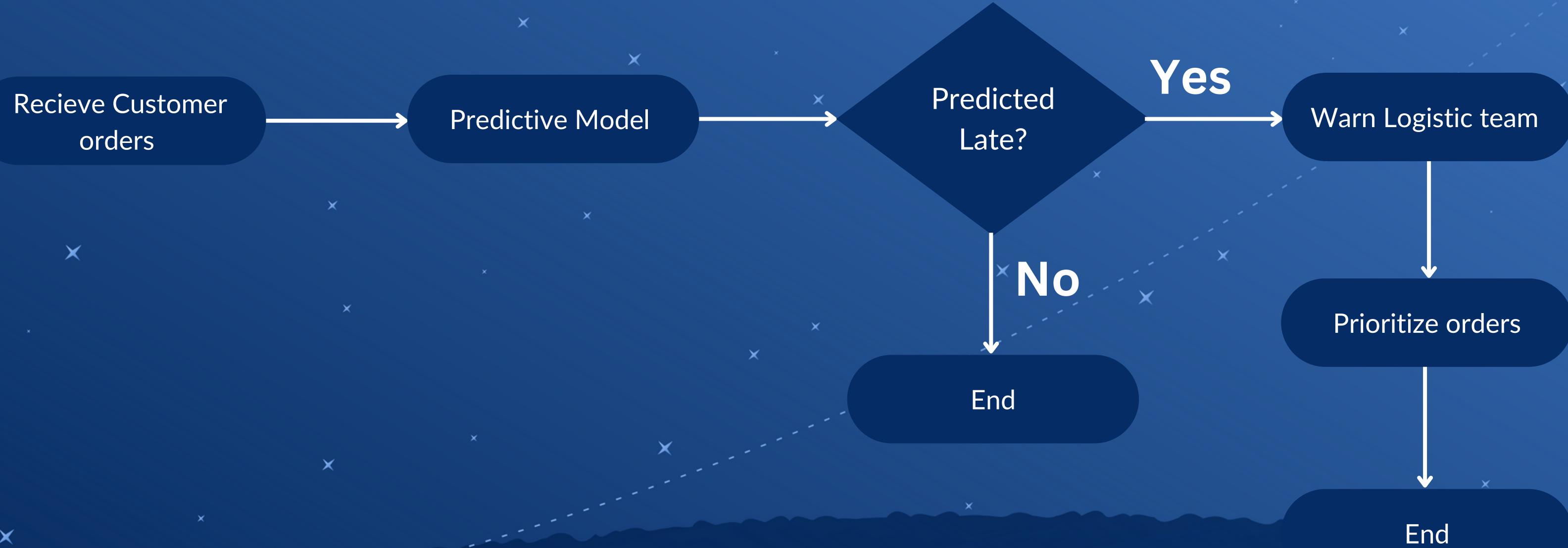


# XGBOOST CONFUSION MATRIX

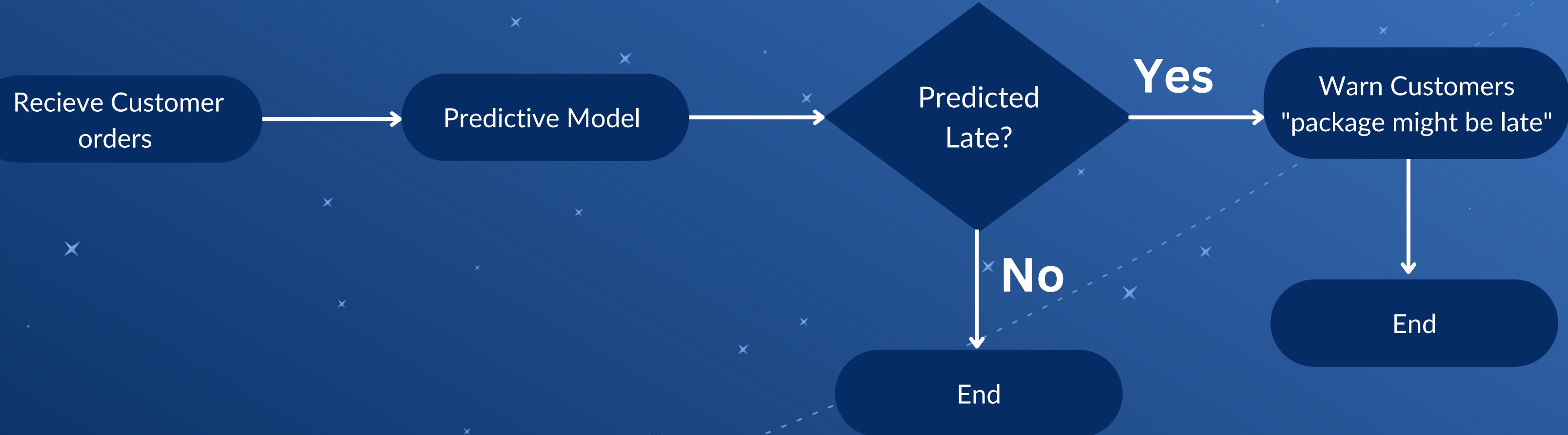
	Predicted Late	Predicted Not Late
Actually Late	<b>True Positive</b> 1205 (35.5%)	<b>False Negative</b> 107 (3.24%)
Actually Not Late	<b>False Positive</b> 933 (28.27%)	<b>True Negative</b> 1055 (31.97%)



# Warn Late Notifications Model



# Warn Late Notifications Model



# Models Impact to Business Metrics

Ideal Condition

**On Time Delivery**

before

**40.3%**

4436 packages

**On Time Delivery**

after

**96.8%**

10642 packages

**Late Delivery**

before

**59.7%**

6563 packages

**Late Delivery**

after

**3.24 %**

357 packages

Realistic Condition: considering business constraints

**40.3%**

4436 packages

**79.01%**

8690 packages

**59.7%**

6563 packages

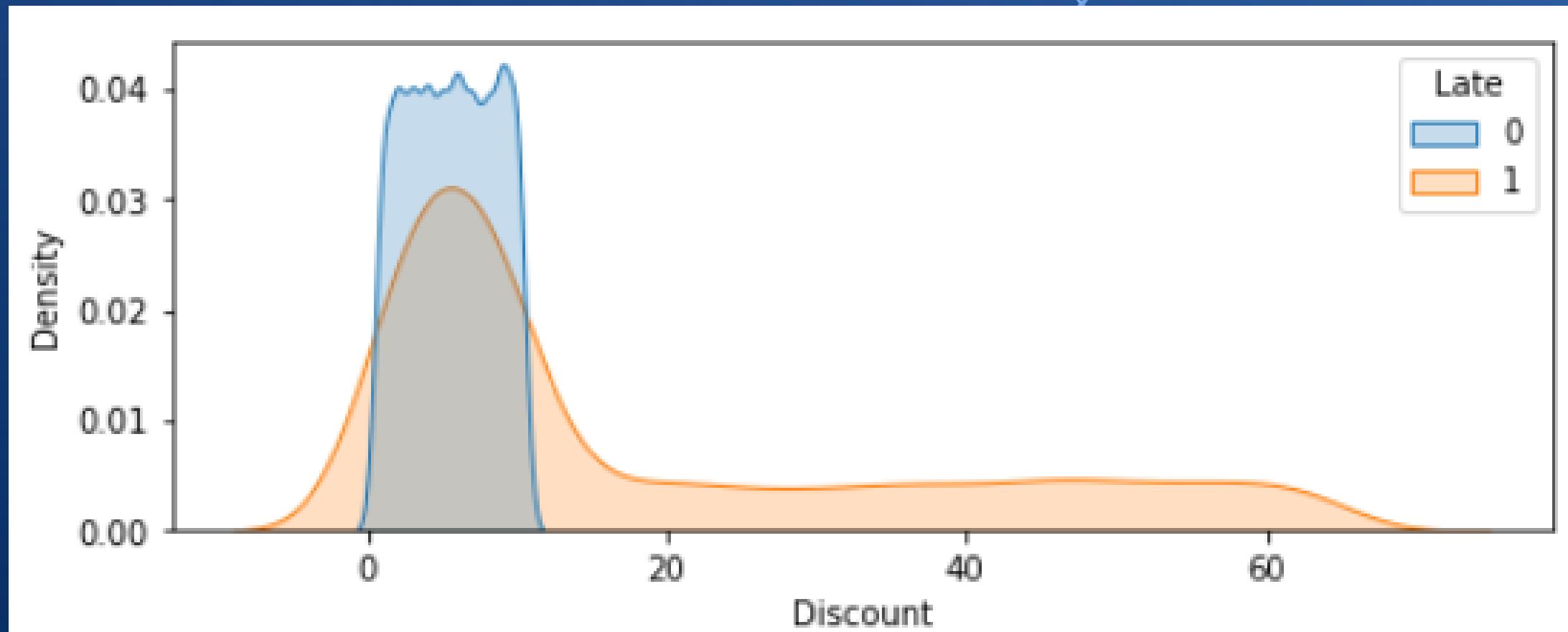
**20.9 %**

2309 packages

Assumption :

- Due to business constraint (inefficiency in logistics / warehouse), 50% of True positive results are assumed to be late even after being tagged.

# Business Insight & Recommendation



0 = Not Late

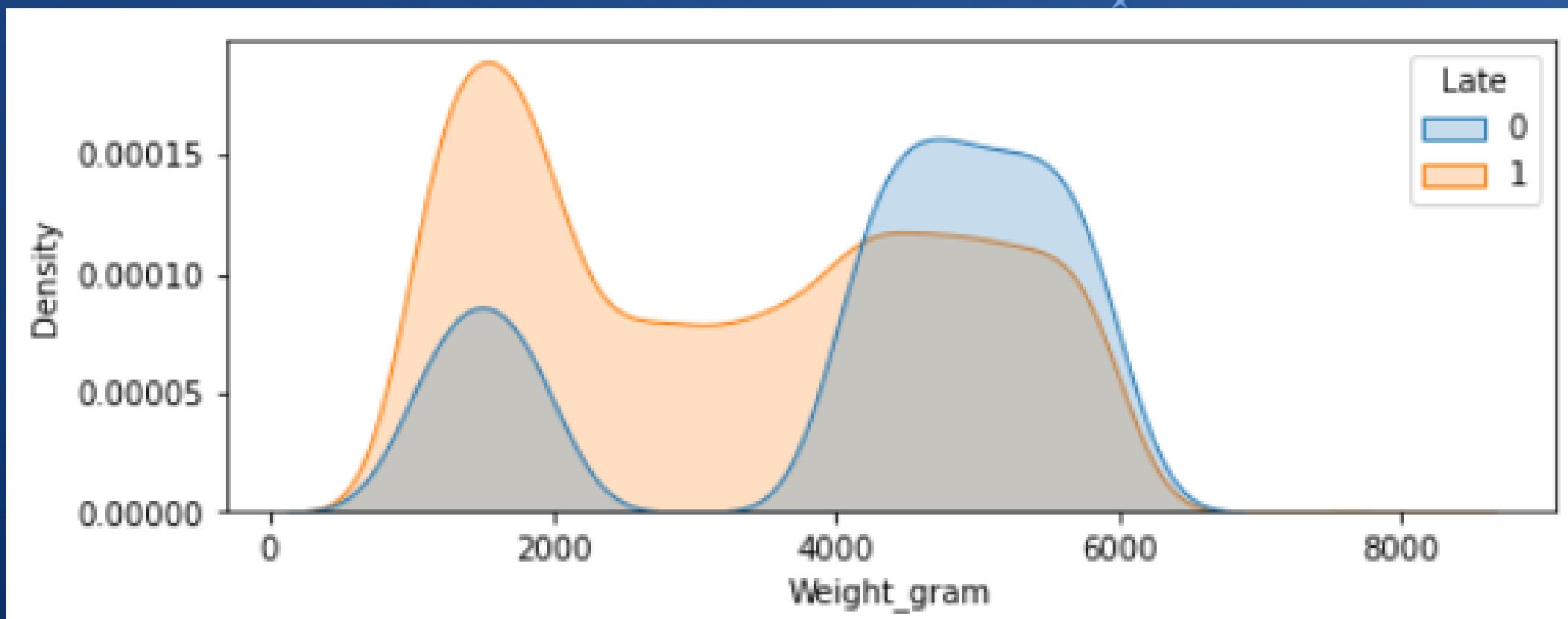
1 = Late

- Packages with high discount tends to be late. Only packages with discount <15% is on-time.

## Recommendation :

- To have a system warning for customer when there's high discount so that their expectation can be meet
- Add more resources e.g daily worker which can help to fasten the process during company campaign.

# Business Insight & Recommendation



- Packages with weight between 2 - 4 kg are always late

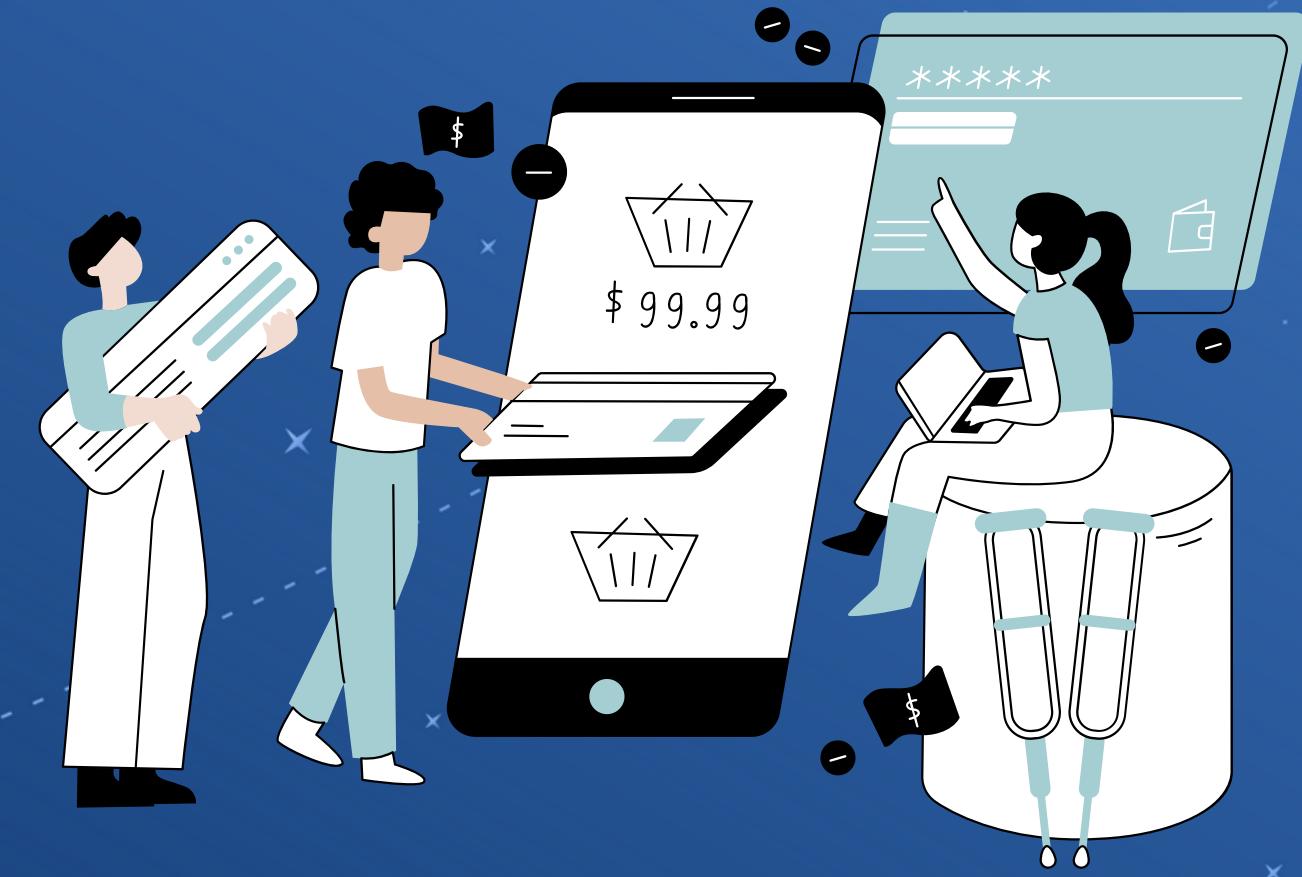
## Recommendation :

- Have a more analysis on what kind of goods are on this weight & when this type of goods are sent

0 = Not Late

1 = Late

# Thank you.



# REFERENCE

- Geng Cui, et.al. 2012 The Effect of Online Consumer Reviews on New Product Sales
- [https://deo.shopeemobile.com/shopee/cms\\_cdn\\_bucket/3668b6b1577c48a394d2d0d3ac322441\\_System%20Point%20Penalty.pdf](https://deo.shopeemobile.com/shopee/cms_cdn_bucket/3668b6b1577c48a394d2d0d3ac322441_System%20Point%20Penalty.pdf)