Data Science Team, Hunky Dory

Shipping Prediction

"Reduce the products are delivered **not on time** & customer care calls to Improve Company's Revenue"

About Us

Hunky-Dory is a Data Science Team in **Hunky,** which is a growing **e-commerce**.

Arif Romadhan

Tutor

Team

Boma Wikanthyasa I

will present about:

Problem Statement

Andriana Butar-butar

will present about:

Insight & Data Pre Processing

Reny Nur Hidayah

will present about:

Modelling & Recommendations

Contents of The Report

Problem Statement

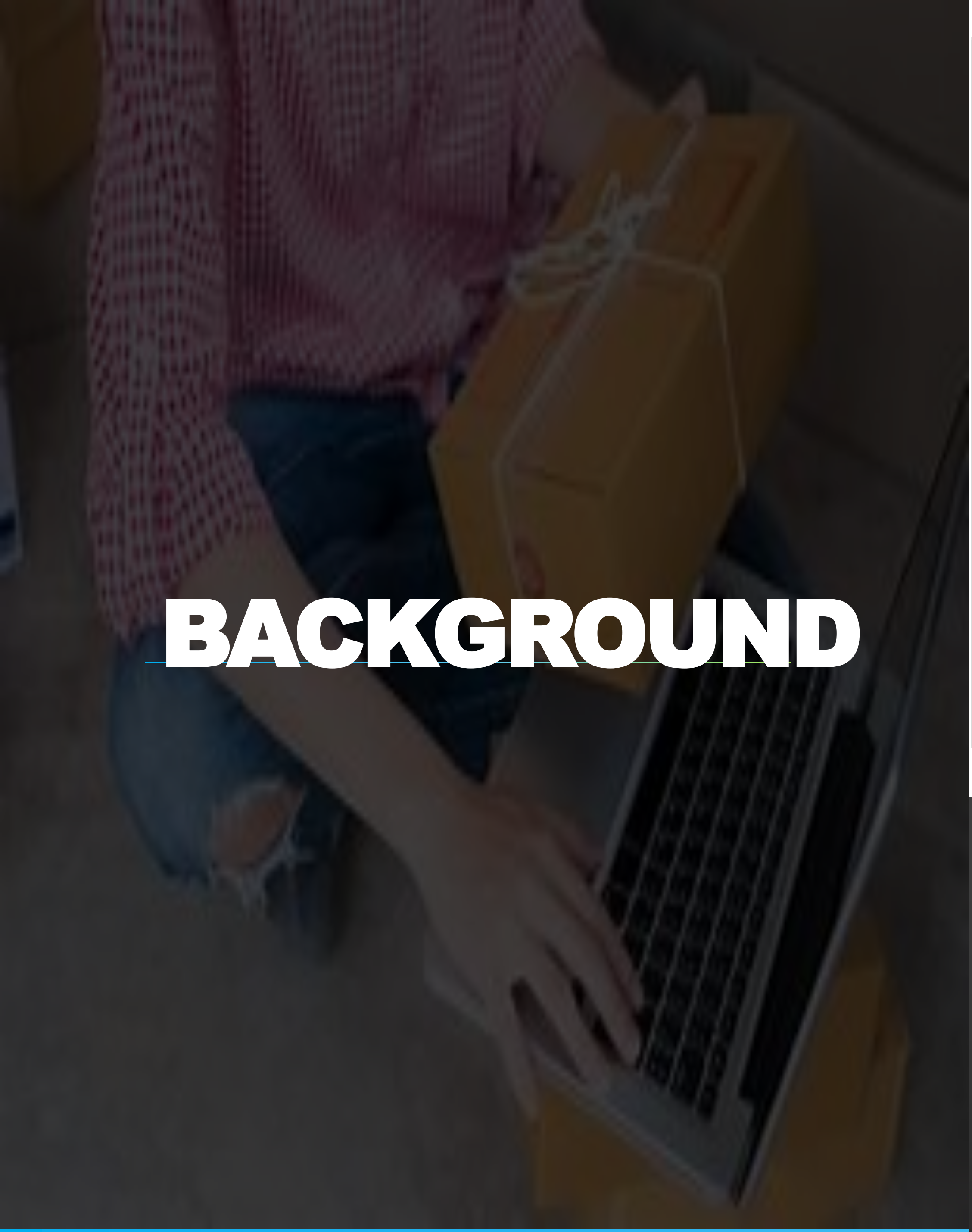
Insight

Data Pre Processing

Modelling & Recommendations

Data Science Team, Hunky Dory

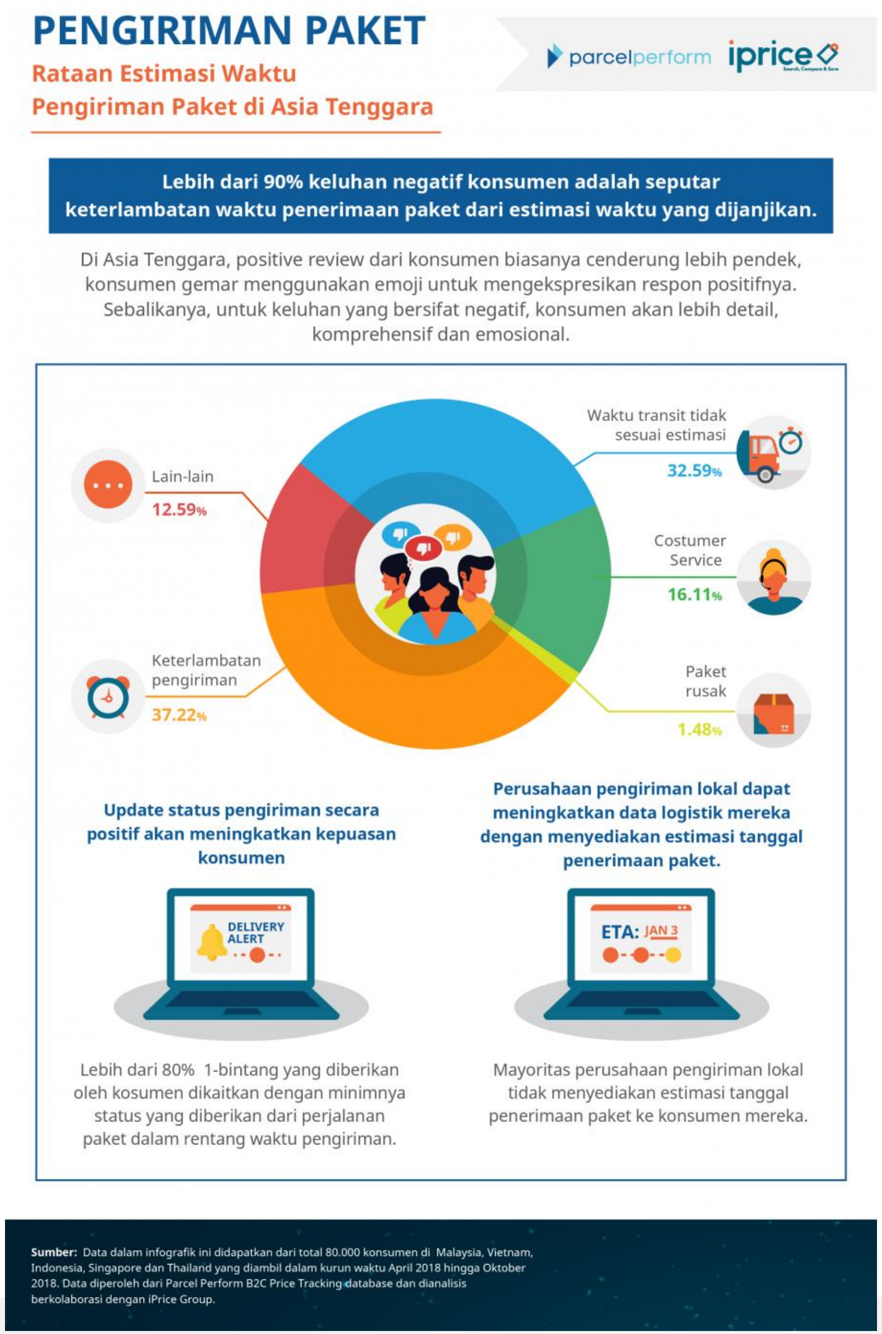
PROBLEM STATEMENT



BACKGROUND

The e-commerce business in Indonesia is increasingly promising. In the midst of a pandemic, this digital-based trading business is even projected to grow 33.2 percent from 2020, which reached IDR253 trillion to IDR337 trillion this year.

(Sumber: Katadata.co)



(Based on iprice, 2018)

Delivery service is a **biggest problem** on e-commerce

- Late shipment 37.22%
- Transit time does not match estimates 32.59%

Impact

Customer care calls

Data Science Team, Hunky Dory

Internal issues

Most of the Products are delivered **not on time**

Poor Delivery Performance

☹

Data Science Team, Hunky Dory

ShipBob

38% Customer Will not shop at the same store due to a **negative shipping experience**

https://www.shipbob.com/blog/ecommerce-shipping/

→

Potential of **4,180** customer who will not shop at our store

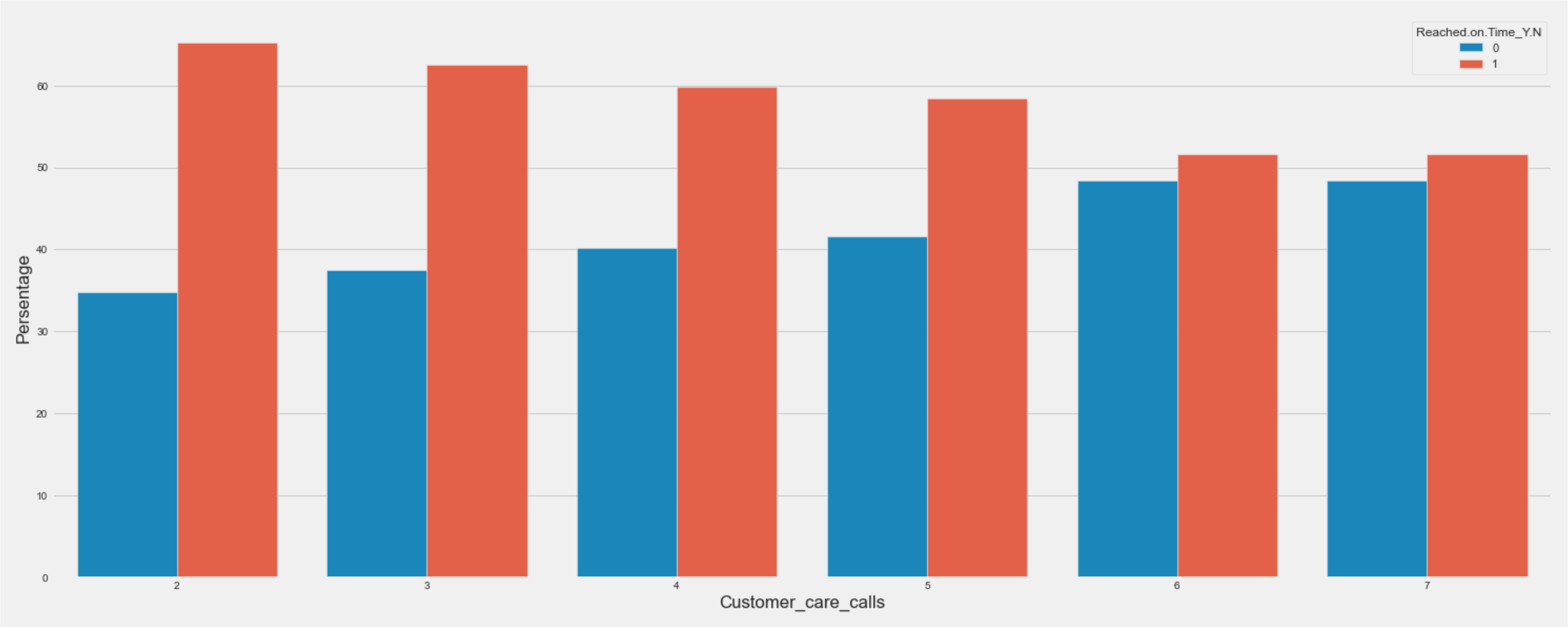
↓

We're going to prevent this

" Shipping Prediction "

Data Science Team, Hunky Dory

Internal issues



Customers who frequently make customer care calls tend to be more **on time**

Customers must often make customer care calls so that the products delivered on time

Data Science Team, Hunky Dory

Goals

Reduce the products that delivered not on time

Impact:
1. **Prevent Potential of customer who will not shop at our store**

Increase Company's revenue

Reduce customer care calls

Matrics

Late shipment rate

Customer care calls rate

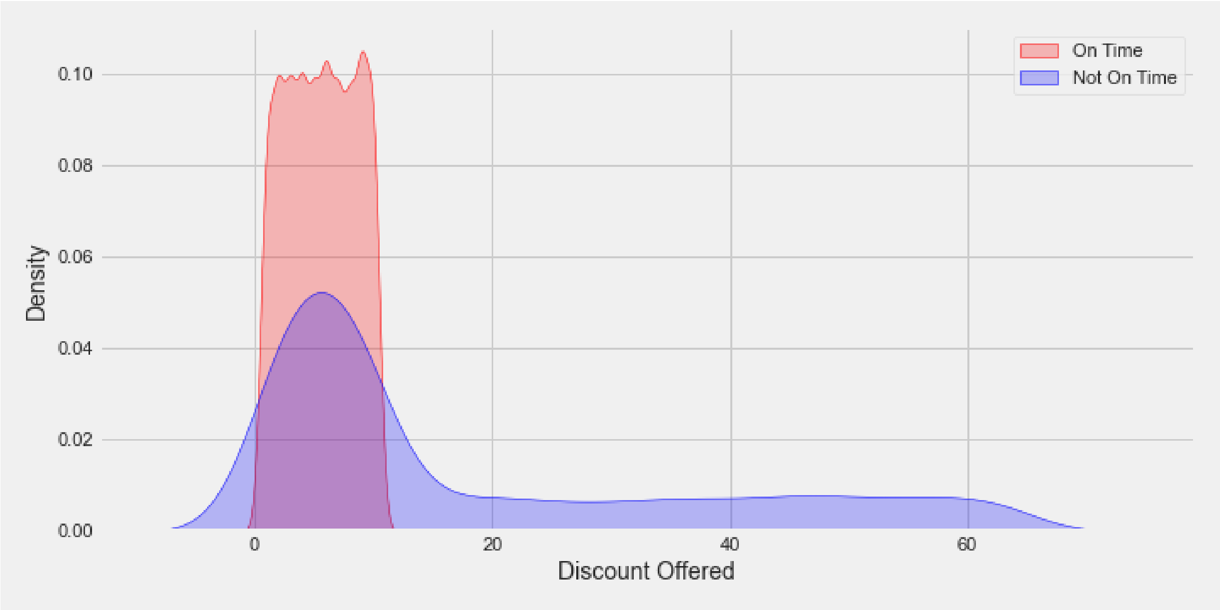
Objective

Modelling "Creating a prediction model"

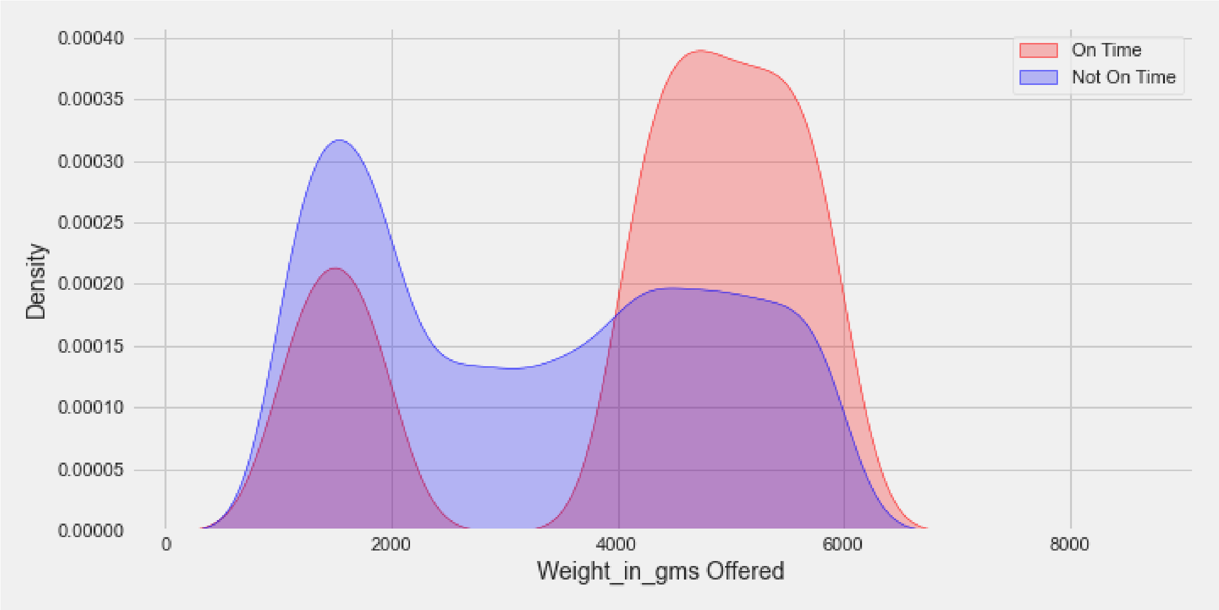
INSIGHT

INSIGHT

Discount offered



Weight



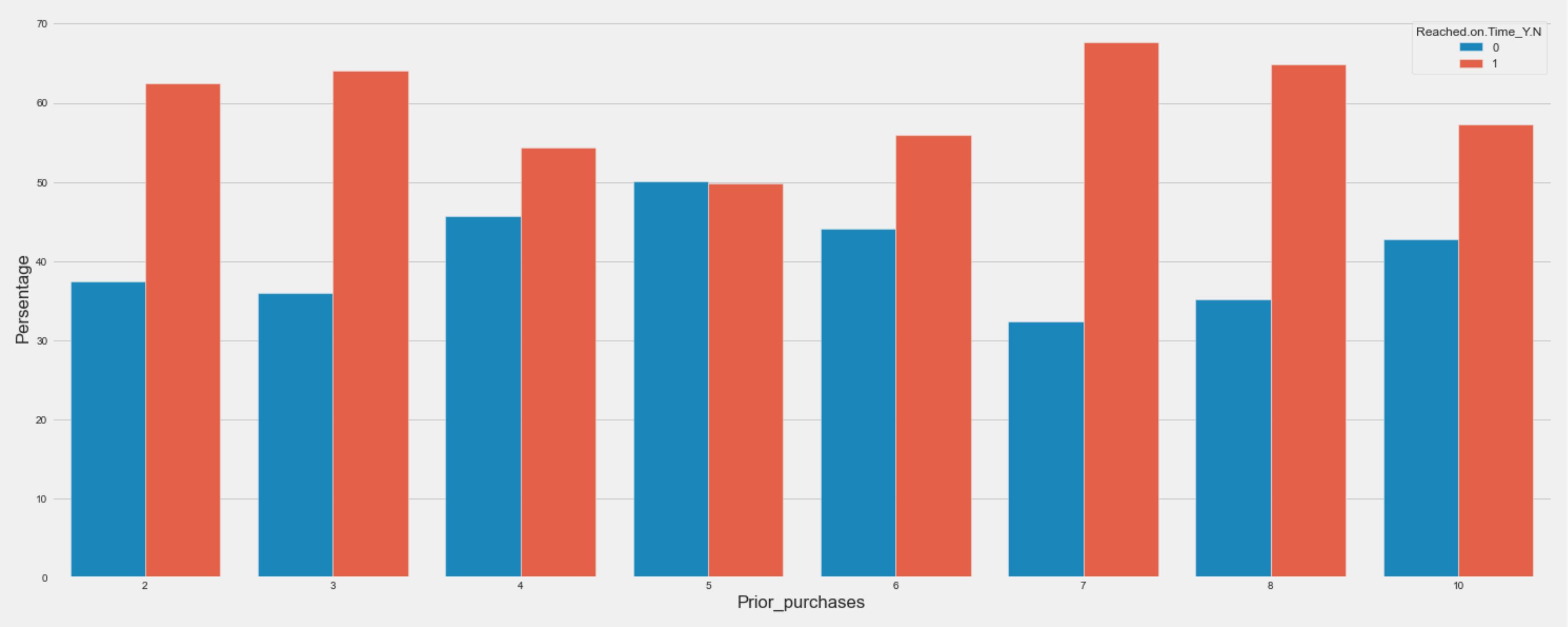
The **most On-time products have a discount in the range of 1-15 USD**, while discounts above 15 USD tend not to be on time

The **most On-time products have a weight in the range of 4-6 kg**, while **weight in the range 1-4 kg** tend not to be on time

Data Science Team, Hunky Dory

INSIGHT

Prior Purchase

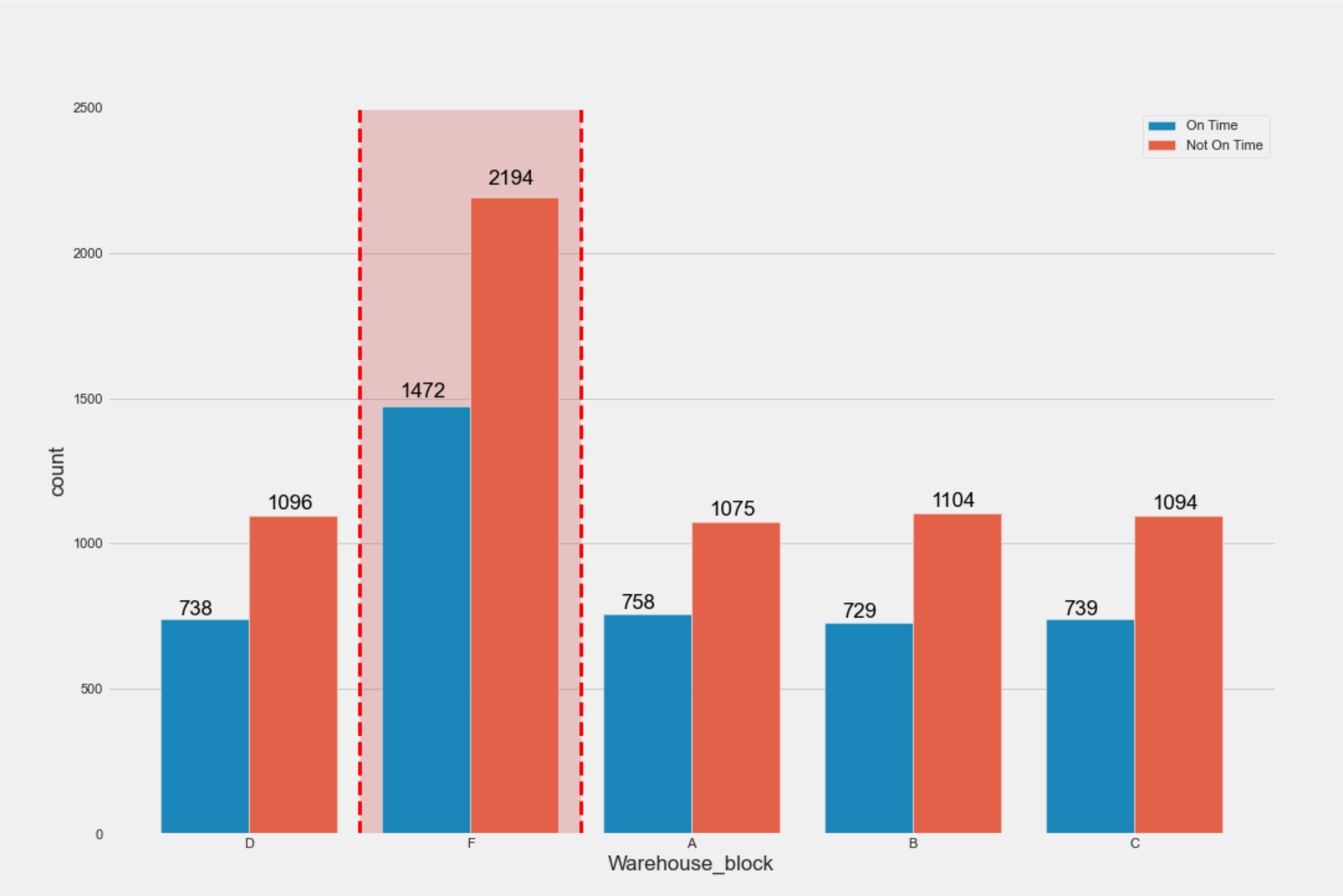


Customers having prior purchases below 5 tend not to be on time and after more than 5 products ordered are no longer on time

Data Science Team, Hunky Dory

INSIGHT

Warehouse block



About **30%** of the product ships from **Warehouse F**

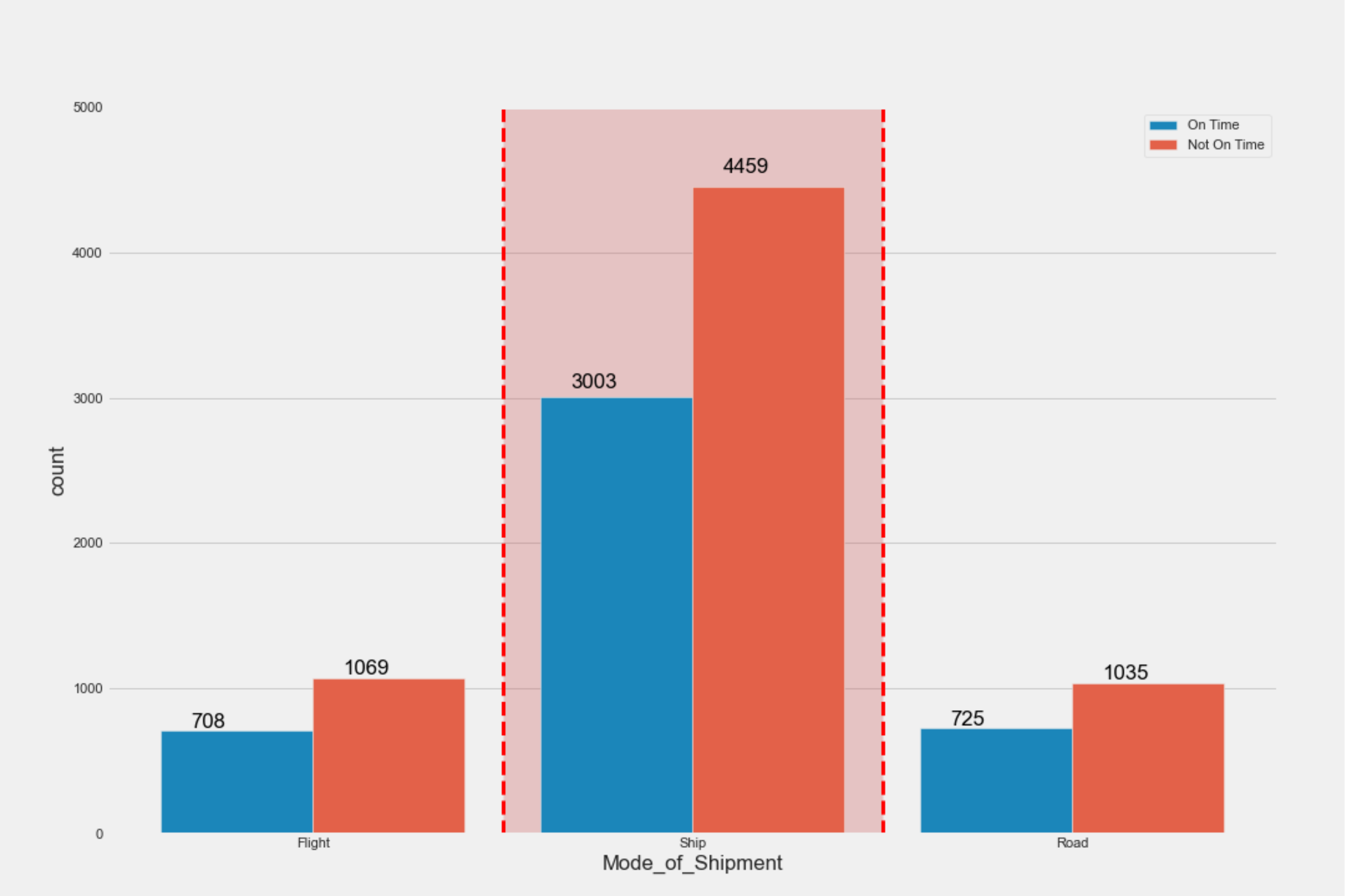
Evaluate each warehouse

1. Optimizing Warehouse Management System
2. Are we short on employees at every warehouse?

Data Science Team, Hunky Dory

INSIGHT

Mode of Sipment



Most of the products are shipped via ship

Prioritas = Ship

WHY SHIP?

1. Efficient in terms of shipping **costs**
2. There is **More Cargo Space** in Ocean Shipping
3. **Flexible** in receiving goods sent via sea shipping

Data Science Team, Hunky Dory

DATA PRE - PROCESSING

Our Dataset

10,999 Rows

12 Columns

Features

- ID
- Warehouse_block
- Mode_of_Shipment
- Customer_care_calls
- Customer_rating
- Cost_of_the_Product
- Prior_purchases
- Product_importance
- Gender
- Discount_offered
- Weight_in_gms
- Reached.on.Time_Y.N

Duplicated Data

No duplicate values found

Missing Value

No missing value found

Outliers

- Discount Offered
- Prior Purchases
- But we using data whose outliers haven’t been removed, because there is no significant difference

Features Engineering

- Discount_group
- Weight_group
- Cost_group

Delete Columns

- ID
- Customer_care_calls
- Customer_rating

Label Encoding

Gender (F:0, M:1)

One hot Endcoding

- Product_importance
- Warehouse_block
- Mode_of_shipment

MODELLING & RECOMMENDATIONS

Modelling

No	Method	Accuracy	Precision	Recall	AUC
1	Logistic Regression	0.68	0.78	0.64	0.69
2	**Decision Tree Classifier**	**0.69**	**0.88**	**0.54**	**0.72**
3	Random Forest Classifier	0.68	0.86	0.55	0.71
4	XGB Classifier	0.68	0.88	0.54	0.72
5	AdaBoost Classifier	0.678	0.82	0.59	0.70
6	LGBM Classifier	0.68	0.87	0.55	0.72

Final model decision

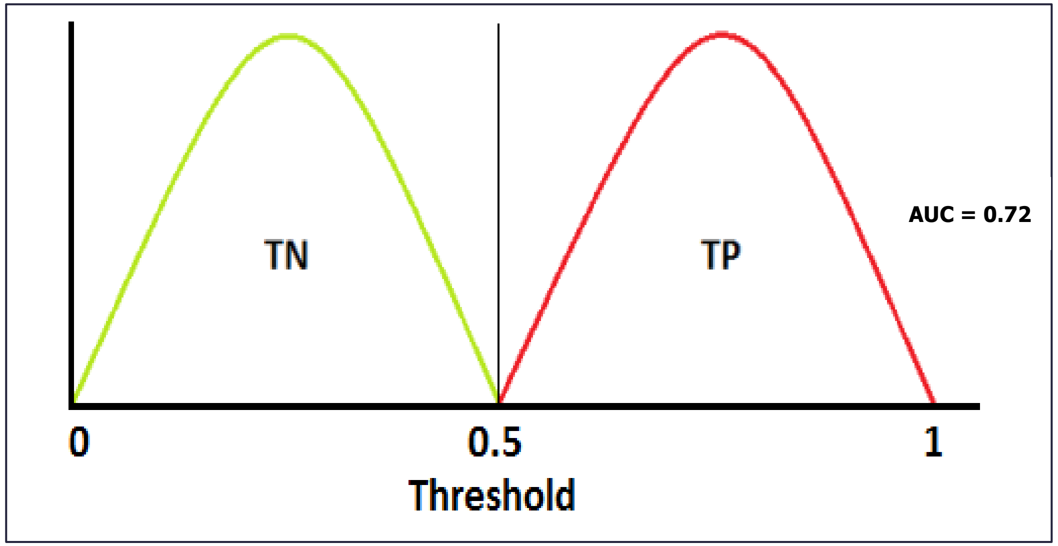
“ **Decission Tree Classifier** ”

Why?

1. AUC & Precission (Higest Score)
2. Small Data → stick to simple methods

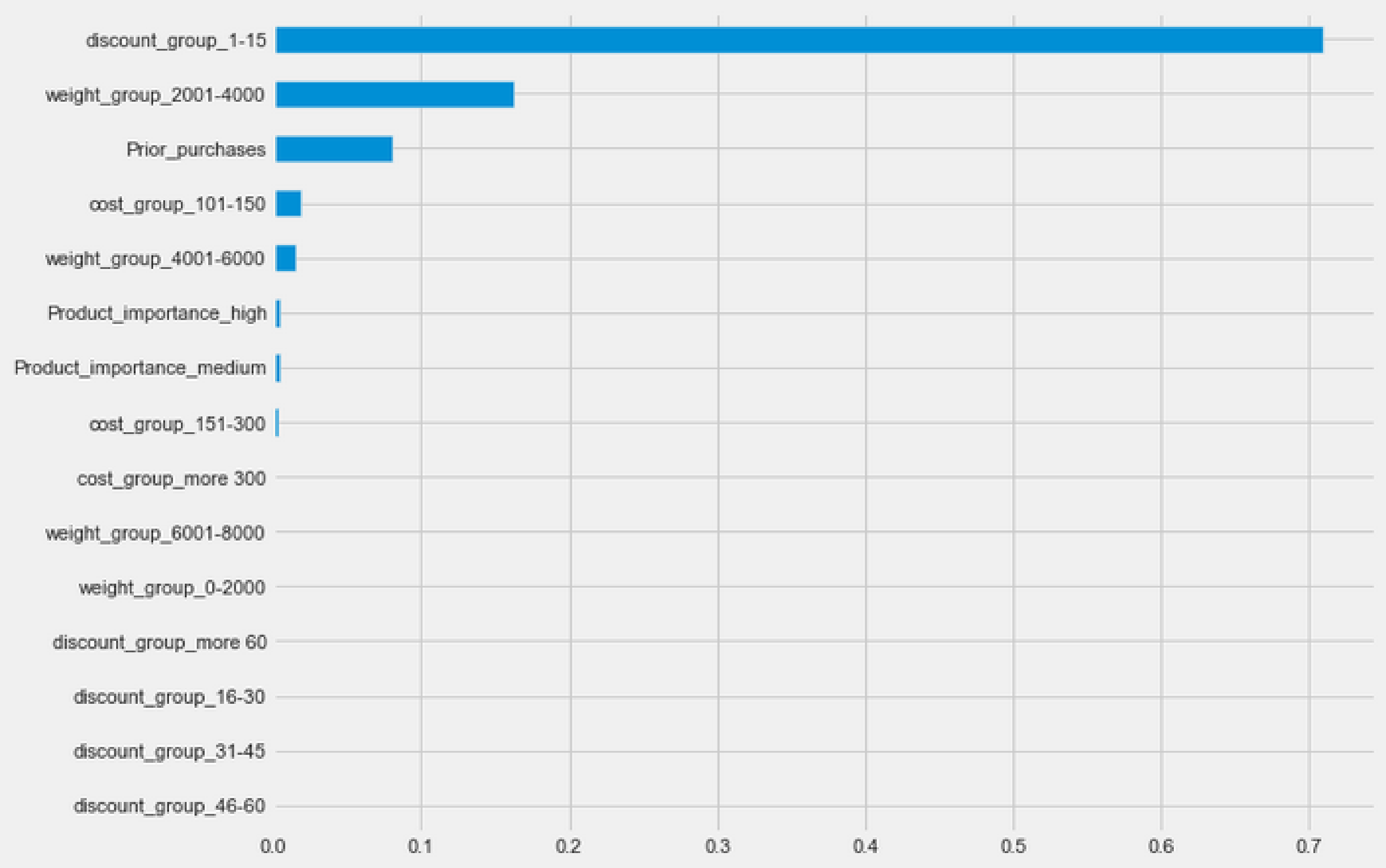
Modelling

Final model decision ⟶ **Decission Tree Classifier**



AUC Score = 0.72, which shows that the model **can predict both on-time and late shipment by 72%** (can provide a class separation correctly by 72%)

Feature Importance



Recommendations

Recommendation Actionable

1. Using the mode of shipment via Liner

- Because it has a **fixed route and schedule** so it arrives **faster** than the trump service ship
- Loading and unloading only takes 2 - 3 hours
- **Reducing** the products are delivered **not on time**

2. Notify the delivery status

- **Reducing** the number of customer care calls

Data Science Team, Hunky Dory

Recommendations

Recommendation Actionable

3. **Free Shipping** on the next transaction

- For those who are predict to be late
- So that customers **don't leave** the store because of a negative shipping experience

74% Free shipping may affect checkout

https://www.shipbob.com/blog/ecommerce-shipping/

ShipBob

4. **Evaluation** at each warehouse

- Evaluation at each warehouse in order to further optimize the warehouse management system, by giving importance to designing a data warehouse model.

Comparison before & after modelling

"Reduce products are delivered **not on time** & customer care calls"

- **Reduce the products are delivered "not on time"**

Before

- ❑ **6563** products are **not on time**

After

- ❑ **72%** (accurately predict) so we could Reduce **4726** the products are delivered **not on time**

Data Science Team, Hunky Dory

Reduce customer care calls

Before

- **Mean** of customer care calls **4**
- **37.2%** Complaint of delay (based on survey)
- 10,999 x 4 = 43,996 calls/day
- 37,2% x 43,996 = 16,367 calls (Complaint of delay calls)
- 72% x 16,367 = **11,784** calls
- 80*21 = 1,680 calls/day
- 43.996 / 1680 = **26 Customer Care**

After

- **72%** (accurately predict) so we could possibly reduce **11,784** customer care calls
- 43,996 – 11,784 = 32,212 calls/day
- 80*21 = 1,680 calls/day
- 32,212 / 1,680 = **19 Customer Care**

Reduce 7 staff customer care
1 Customer care = Rp 4.400.000 (UMR Jakarta)

Rp 30.800.000 Save/bulan

1 Customer Care = 80 calls/day
Estimating cargo = 21 days

Data Science Team, Hunky Dory

Data Science Team, Hunky Dory

Thank You

LAMPIRAN

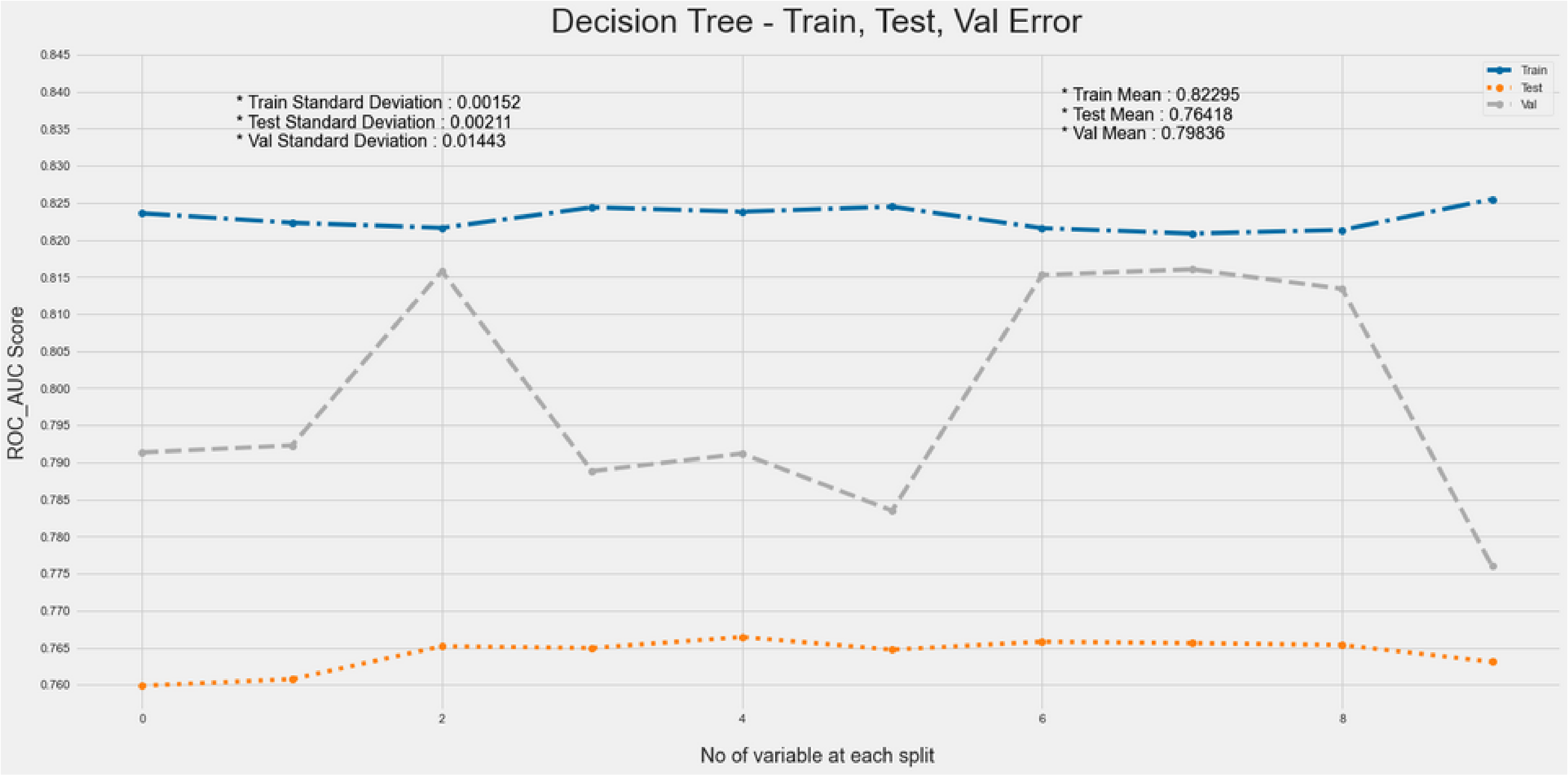
DecisionTreeClassifier Tuning

```
RandomizedSearchCV(cv=5, estimator=DecisionTreeClassifier(random_state=42),
                   param_distributions={'max_depth': [1, 4, 8, 12, 16, 19, 23,
                                                      27, 31, 34, 38, 42, 46,
                                                      49, 53, 57, 61, 64, 68,
                                                      72, 76, 79, 83, 87, 91,
                                                      94, 98, 102, 106, 110],
                                        'max_features': ['auto', 'sqrt'],
                                        'min_samples_leaf': [1, 2, 4, 10, 20,
                                                             50],
                                        'min_samples_split': [2, 5, 10, 100]},
                   random_state=42, scoring='roc_auc')
```

```
best_params = model.best_params_
best_params
```

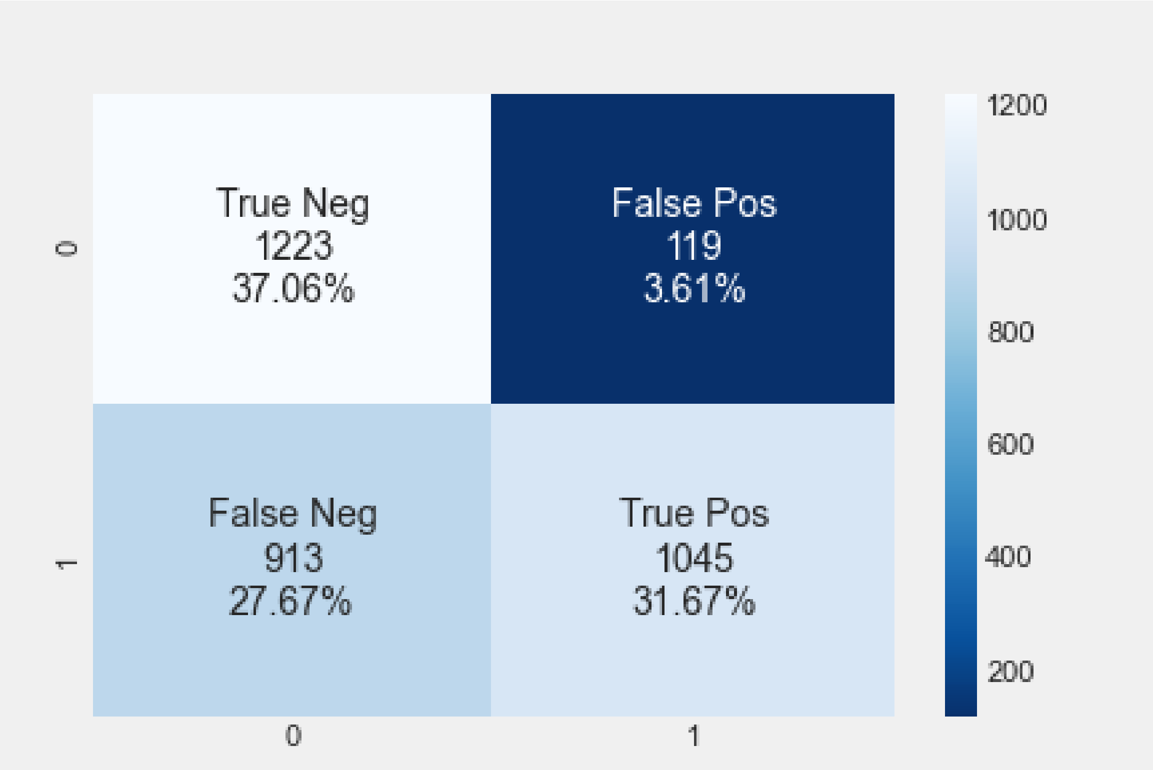
```
{'min_samples_split': 10,
 'min_samples_leaf': 20,
 'max_features': 'sqrt',
 'max_depth': 23}
```

Cross Validation



Data Science Team, Hunky Dory

Confussion Matrics



Decision Tree Classifier

AUC : 712%