

---

# Airlines Passenger Satisfaction

---

Darwin

---



# Table of contents

01

---

**Data Preprocessing**

02

---

**Exploratory Data  
Analysis**

03

---

**Modeling**

04

---

**Feature Selection  
and  
Hyperparameter  
Tuning**

05

---

**Model Evaluation  
and Interpretation**

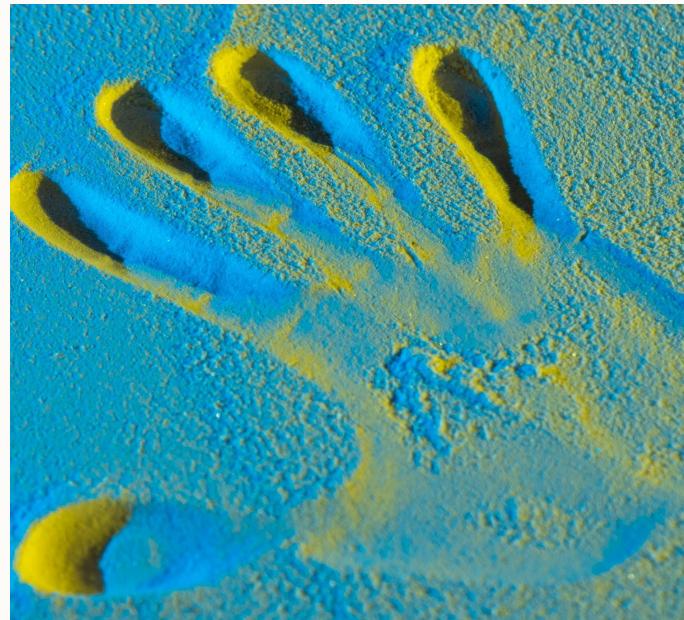
06

---

**Business Insight**

# Objective

To predict whether the airline customer is  
**Satisfied** or **Neutral/Dissatisfied**

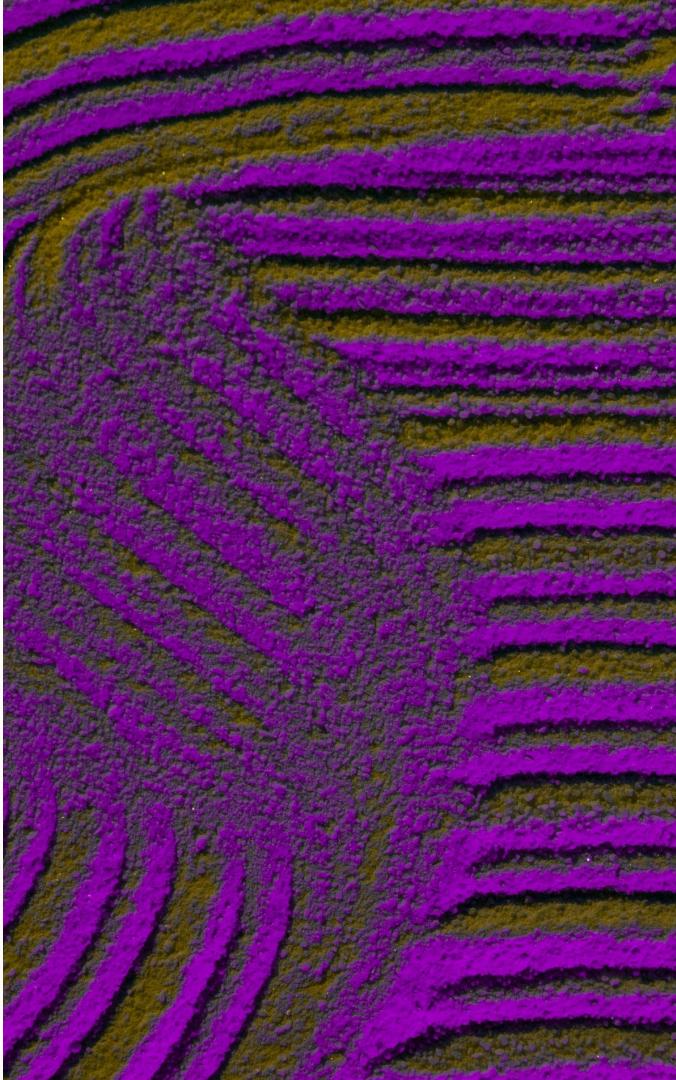


01

---

# Data Preprocessing

---



# Dataset

## Train data

25 columns and 103.000 rows

## Test data

25 columns and 25.000 rows

---

### Drop columns:

- Unnamed: 0
- id

There is no  
duplicate data

Impute missing values on  
“Arrival Delay in Minutes”  
column with median

## **3 Types of Data:**

1. Numerical : Age, Flight Distance, Departure Delay in Minutes, Arrival Delay in Minutes
  2. Categorical : Gender, Customer Type, Type of Travel, Class, Satisfaction
  3. Likert Scale (1-5) : Inflight wifi service, Ease of Online booking, Gate location, etc...
- 

## **Categorical Encoding**

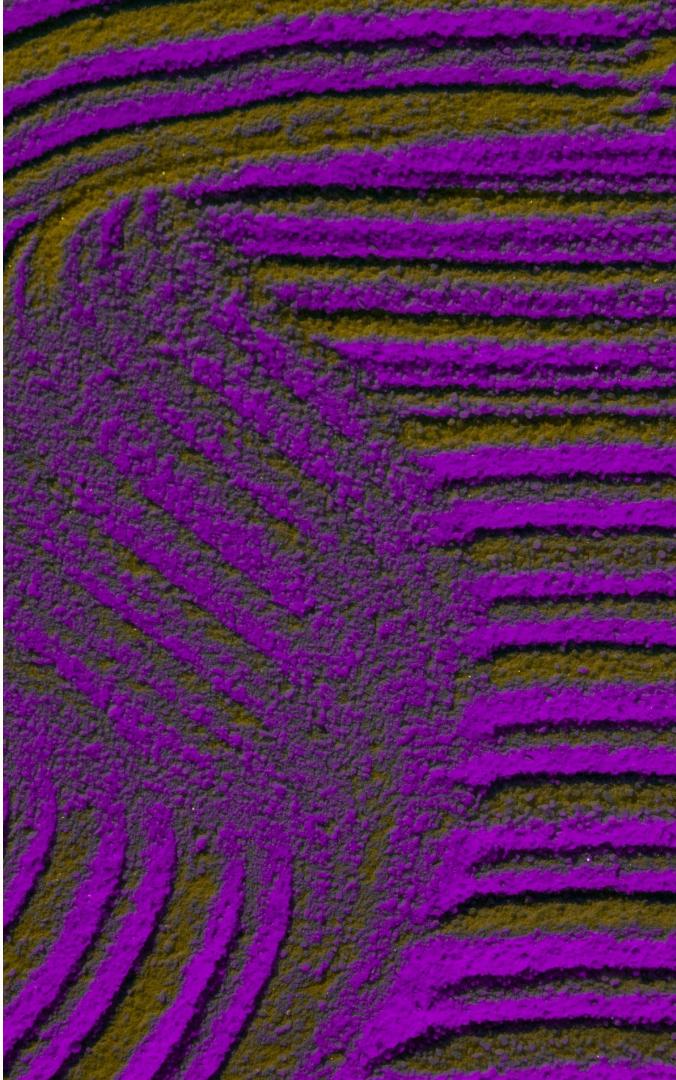
1. Label Encoding
  2. Ordinal Encoding
-

02

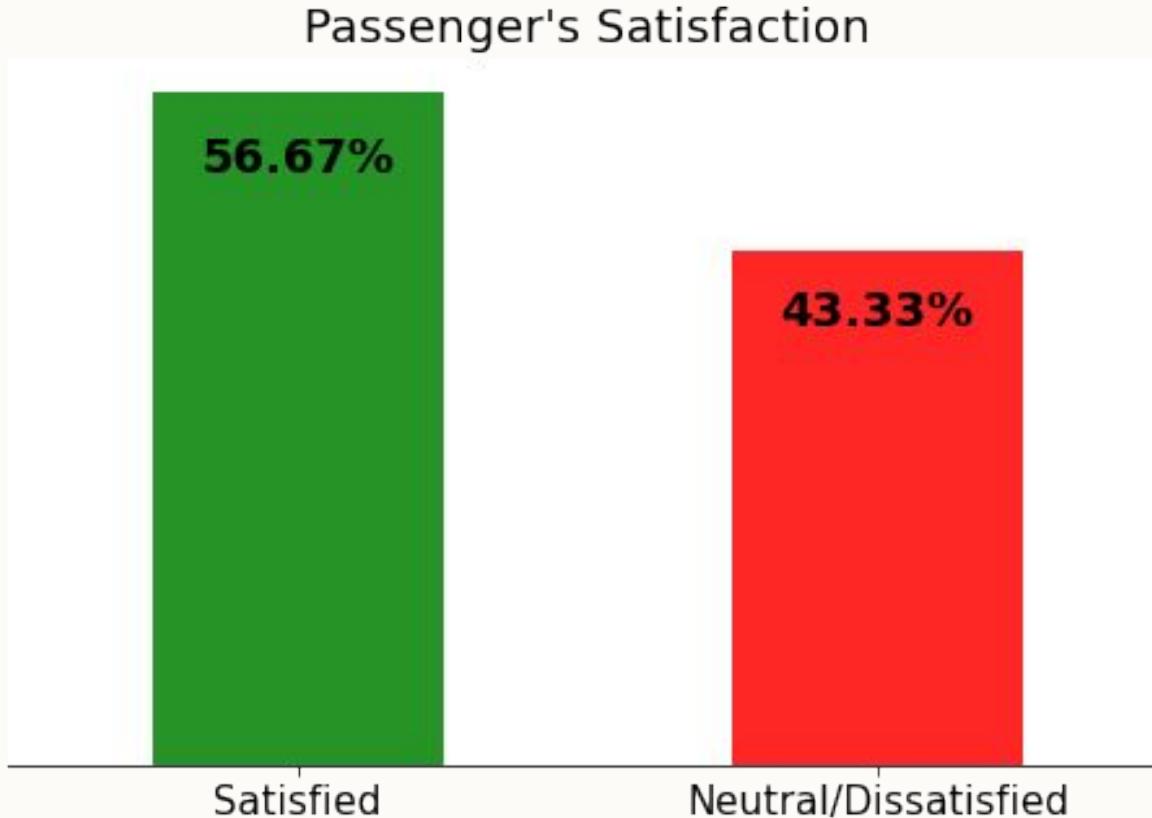
---

# Exploratory Data Analysis

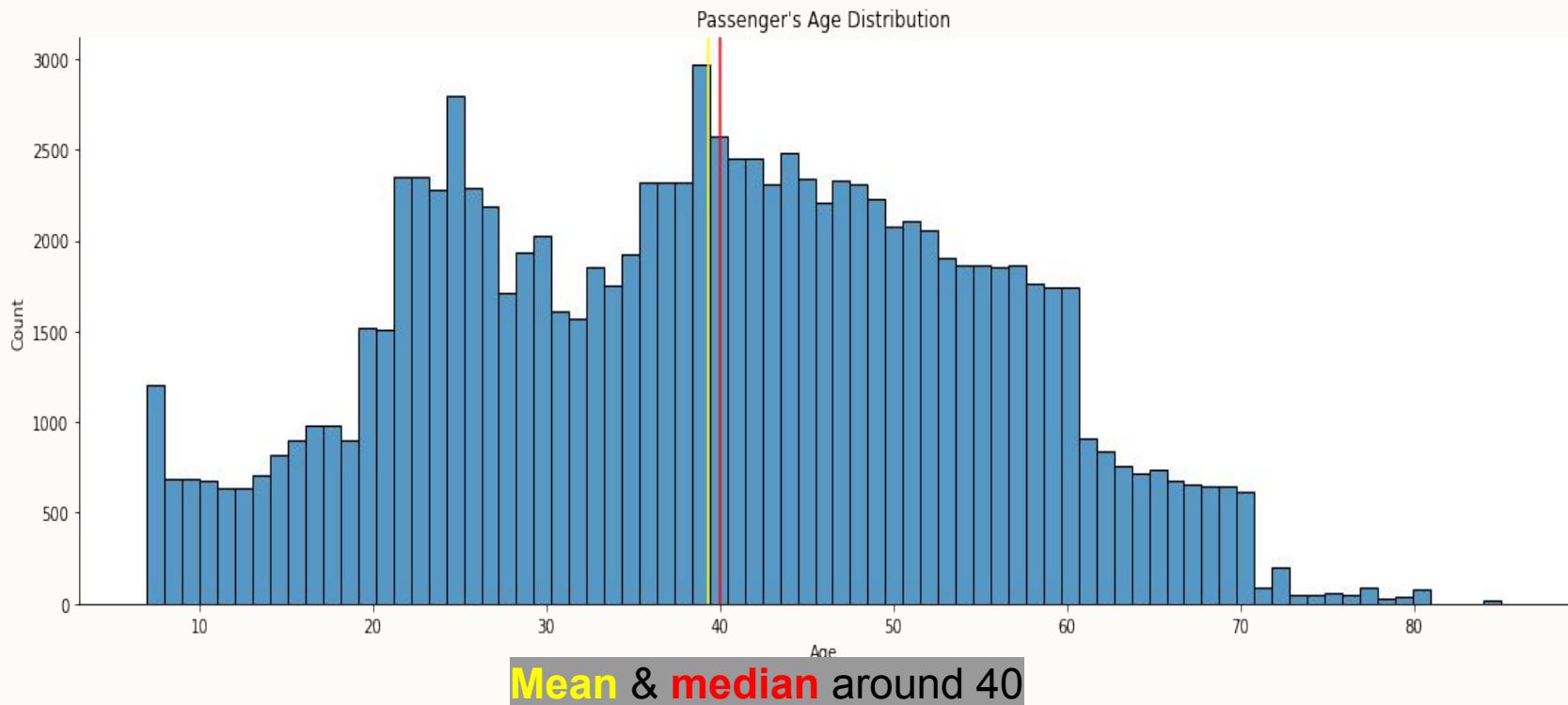
---



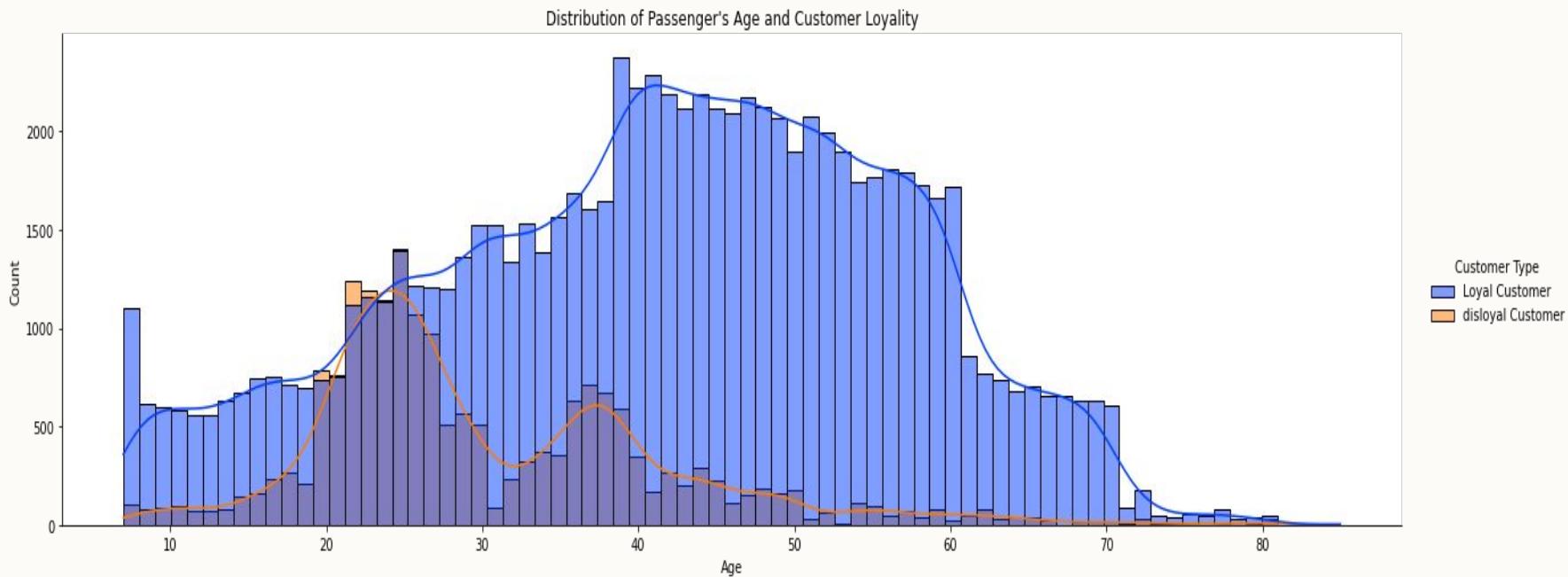
# Not balanced target...



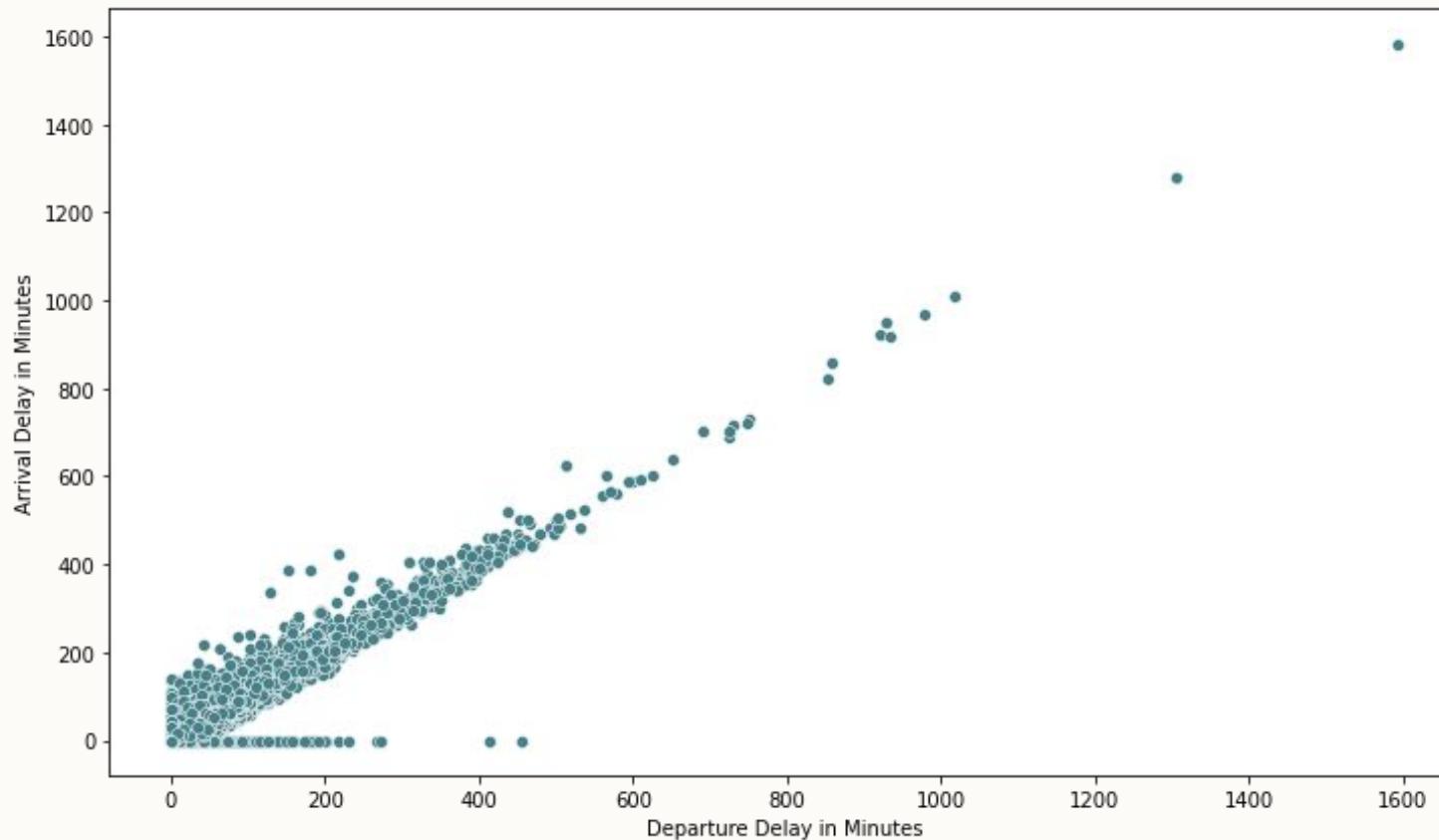
# And a decently normal distribution of Passenger's Age



# Older passengers tend to be more loyal



# Arrival Delay and Departure Delay have a high correlation

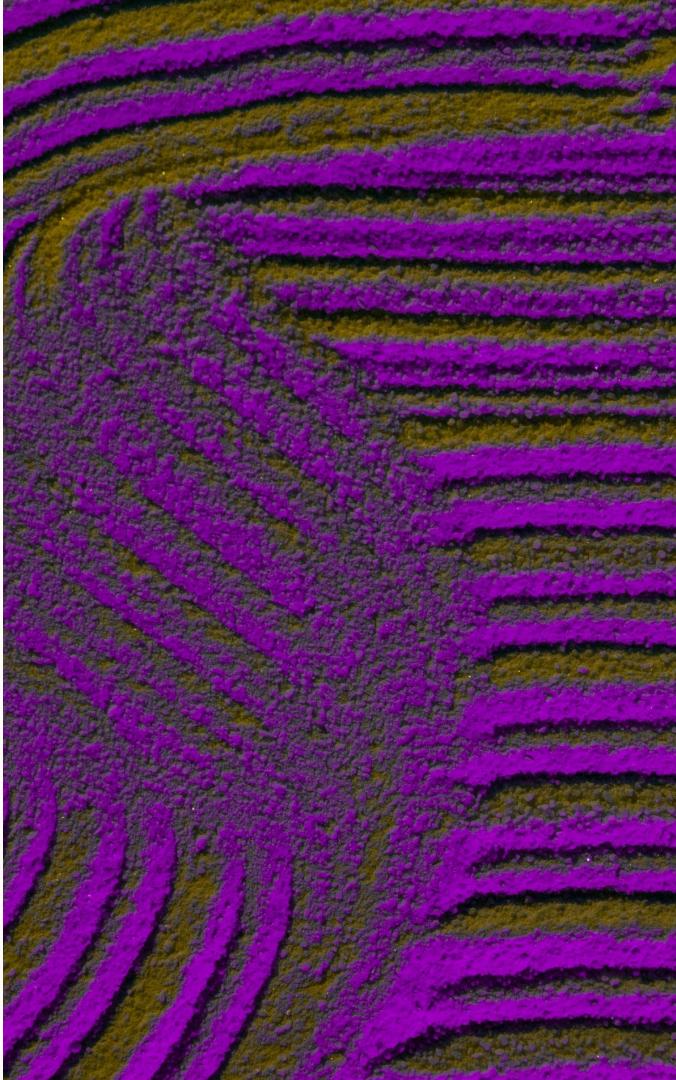


# 03

---

## Modeling

---



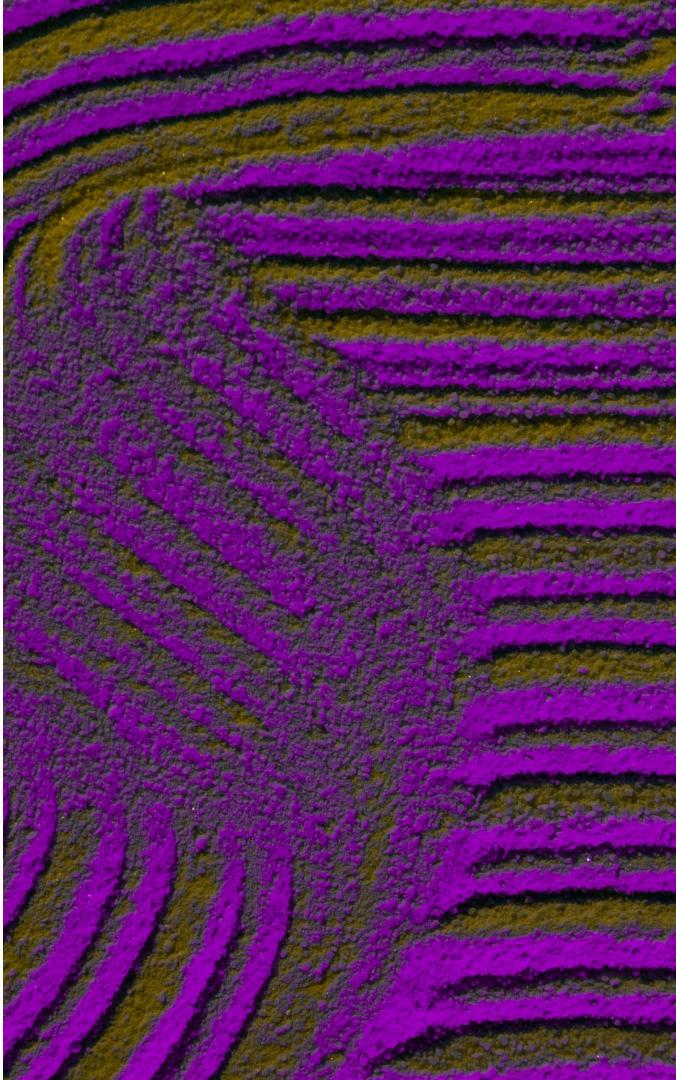
Accuracy Means	Accuracy Errors	Models
0.810	0.0145	LogisticRegression
0.645	0.0030	SVC
0.942	0.0029	DecisionTree
0.961	0.0018	RandomForest
<b>0.9629</b>	<b>0.0015</b>	<b>LightGBM</b>

# 04

---

## Feature Selection and Hyperparameter Tuning

---



# Forward Feature Selection

22

Features

0.9599 accuracy



16

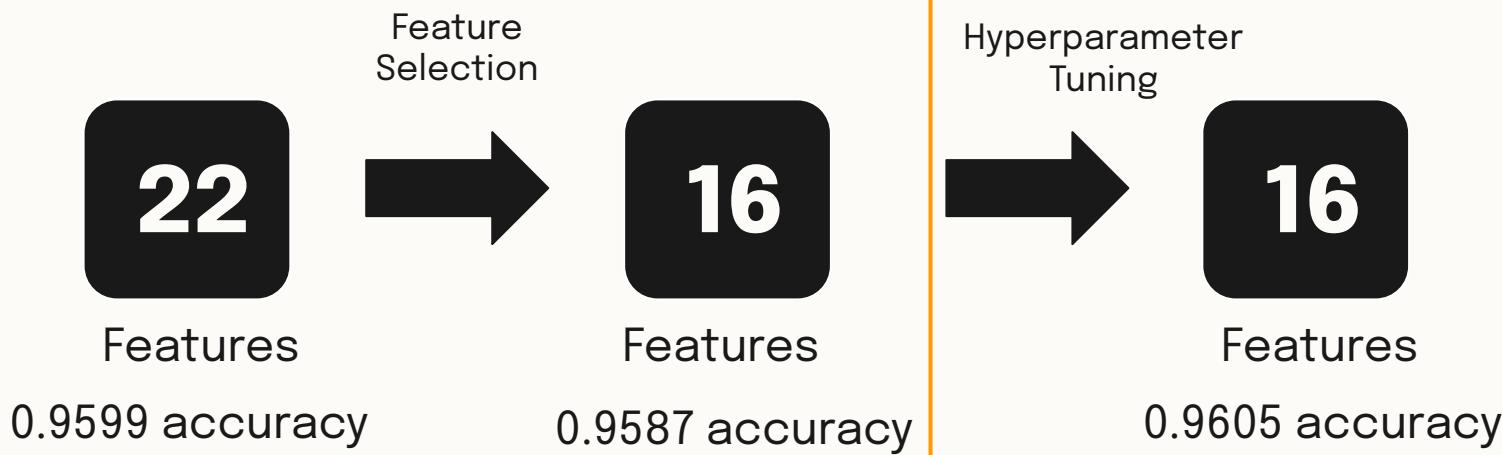
Features

0.9587 accuracy

---

## Removed Features:

- Gender
- Flight Distance
- Departure/Arrival time convenient
- Food and drink
- On-board service
- Departure Delay in Minutes

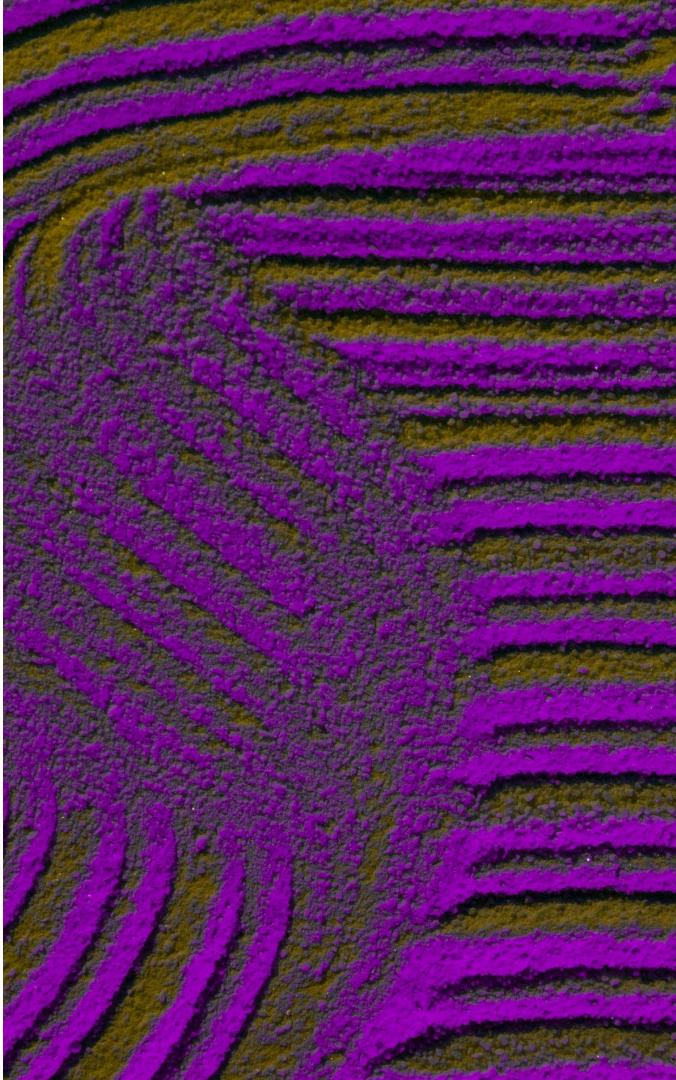


# 05

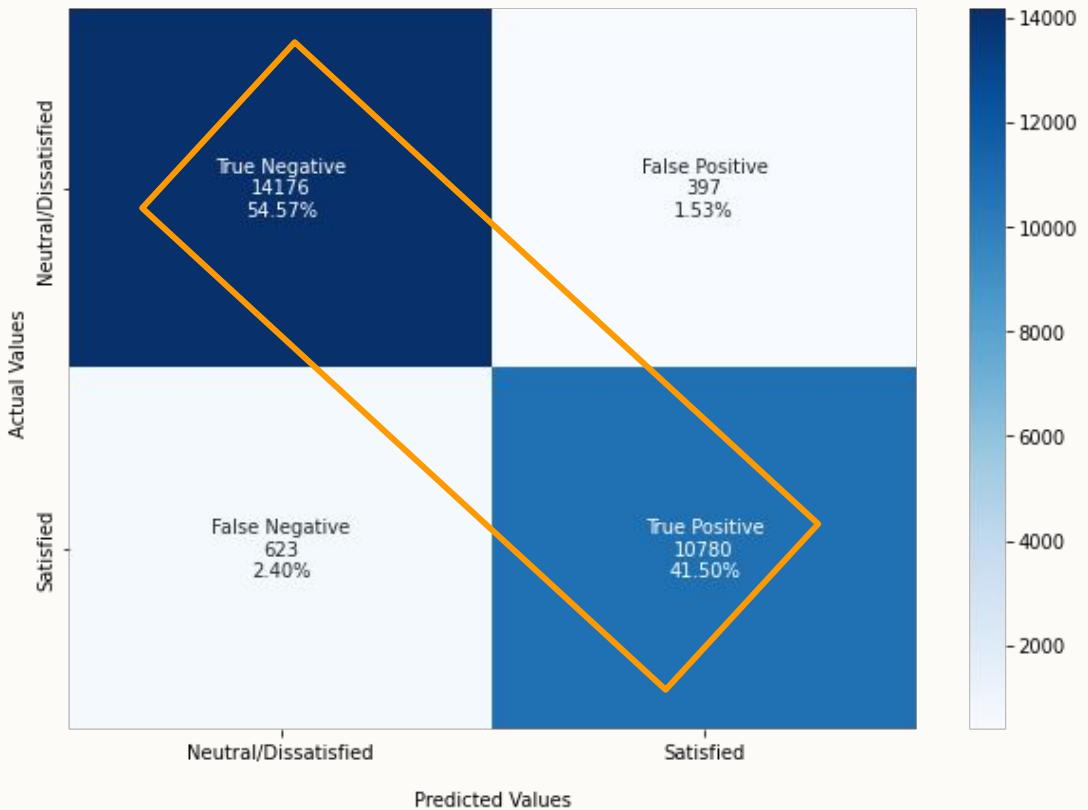
---

## Model Evaluation and Interpretation

---



Confusion Matrix with labels

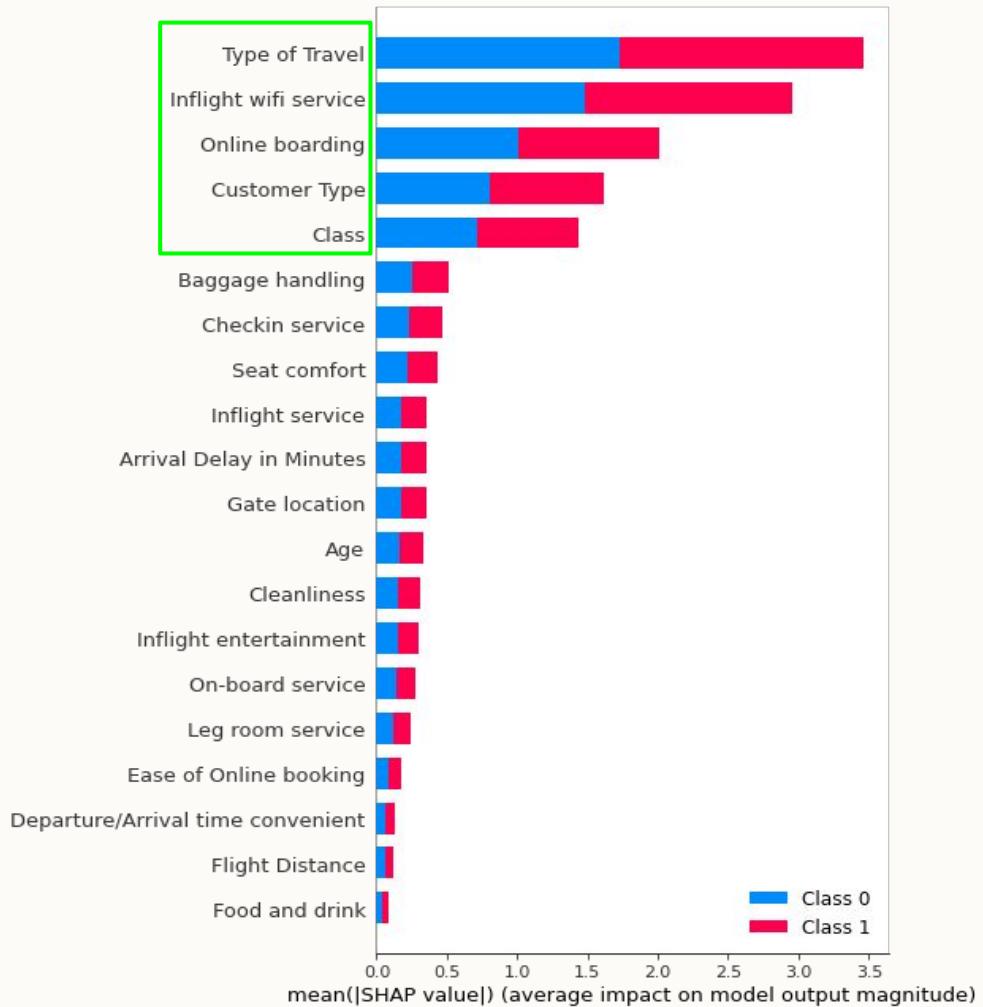


**0.9607  
accuracy**

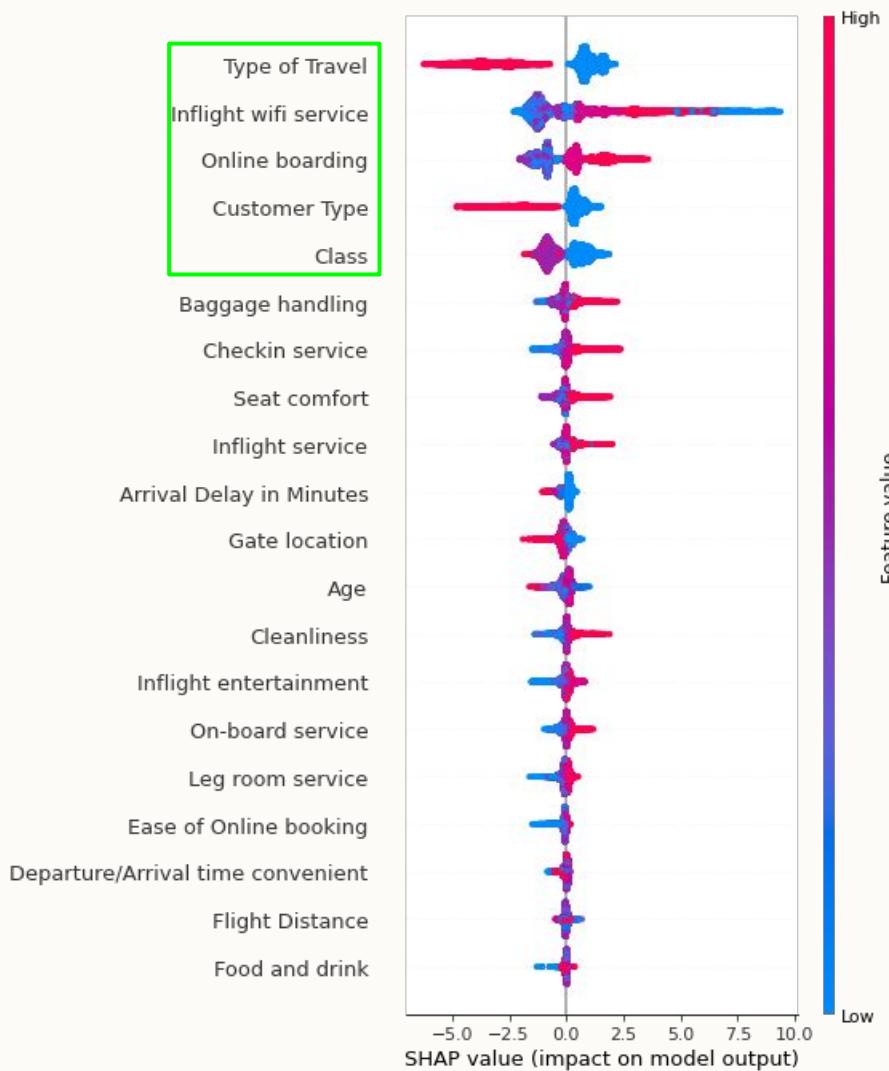


**96.07%  
True Prediction**

# SHAP values



# SHAP values

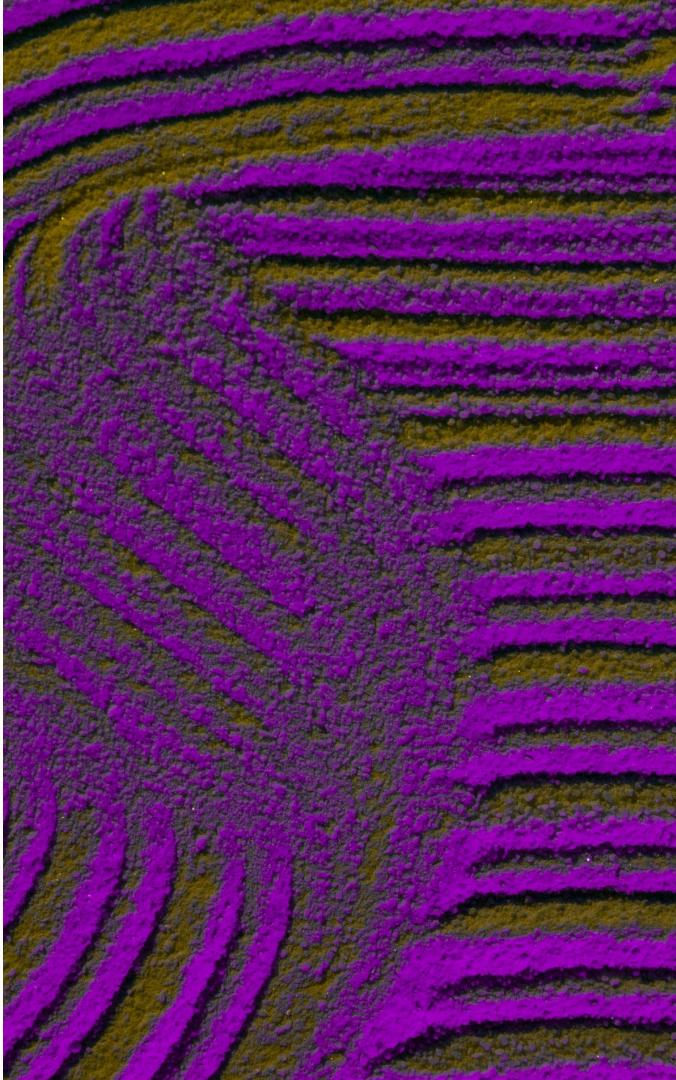


06

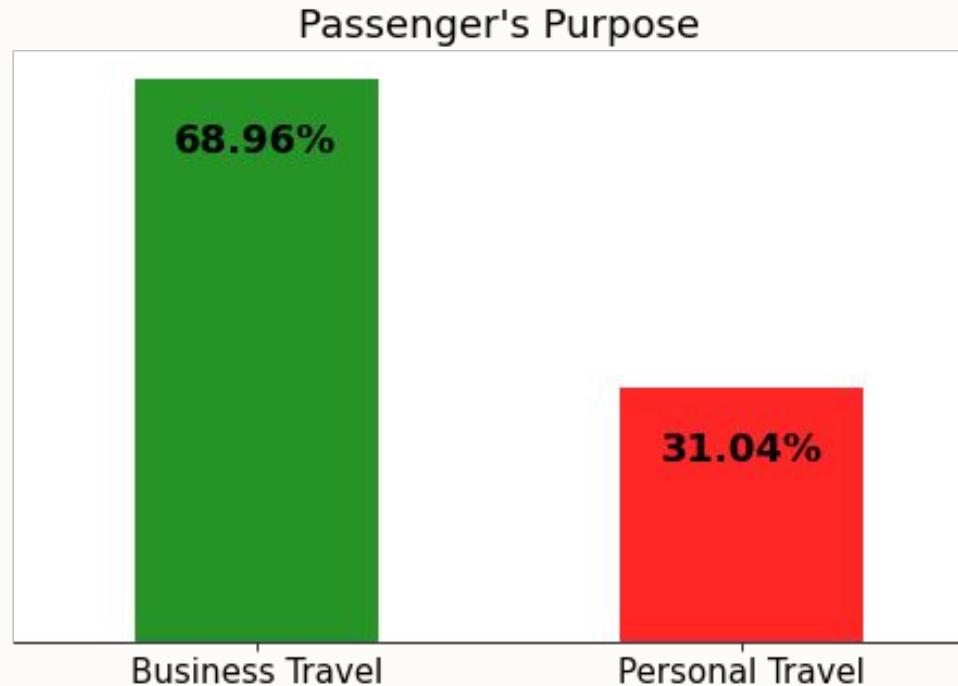
---

# Business Insight

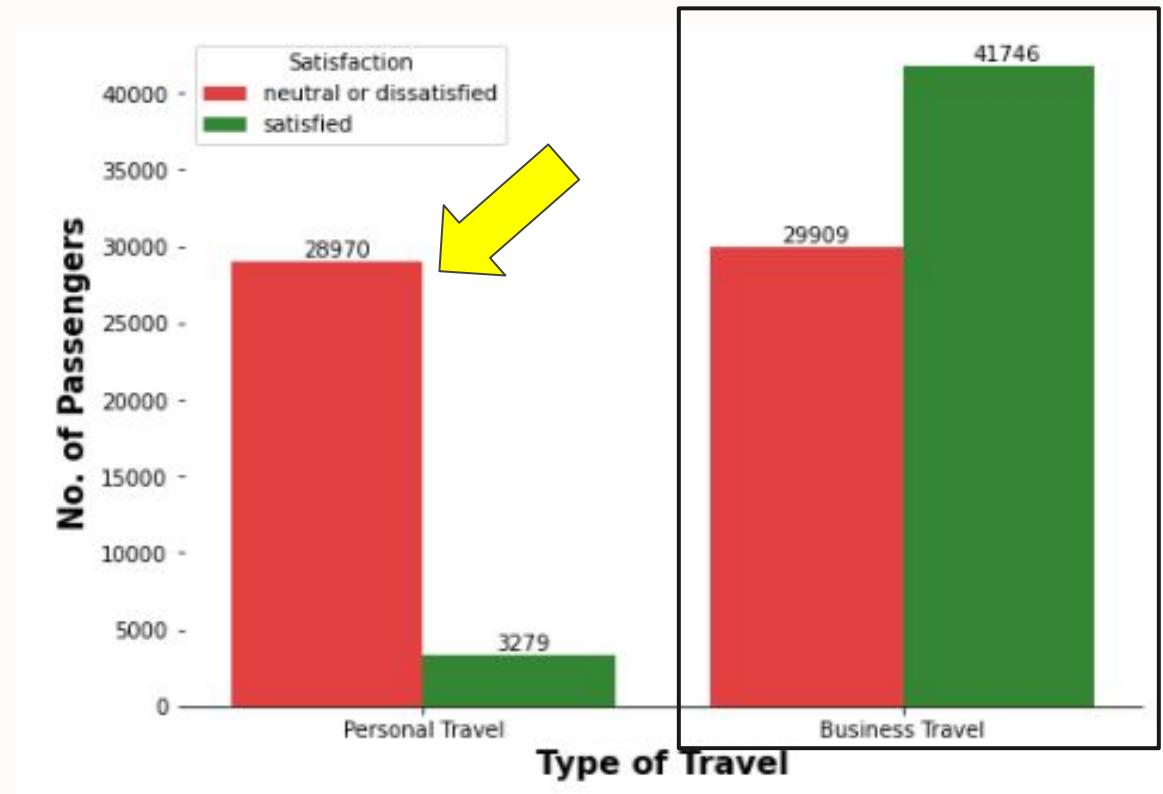
---



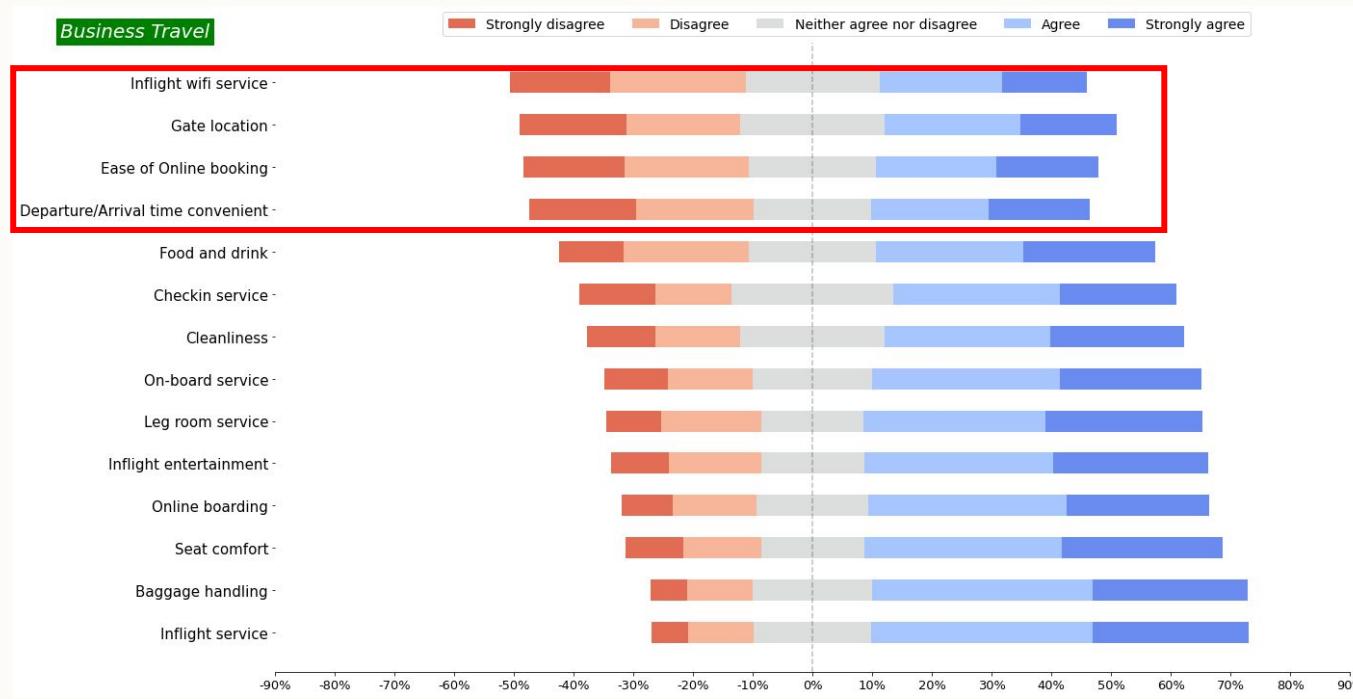
# **69% of our passenger's are doing travel for business purpose**



# Passengers who doing Business travel are more likely to become satisfied



# Business Traveler



Business traveler tend to have a good:

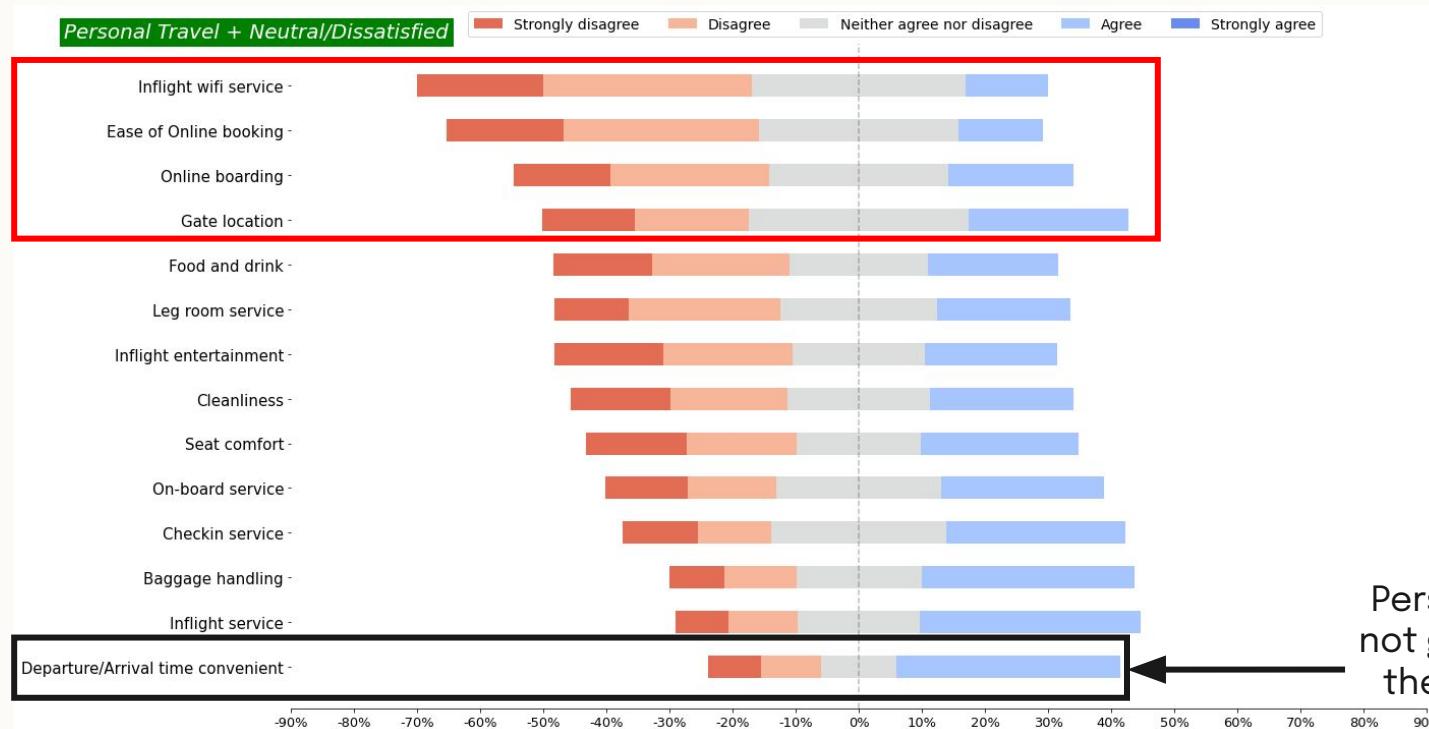
**Inflight wifi service** : for working, not wasting time

**Gate location** : saving time

**Departure/arrival time convenient** : saving time

**Ease of online booking** : saving time

# Personal Traveler + neutral/dissatisfied



Personal travelers are not giving a low score to the delay convenient

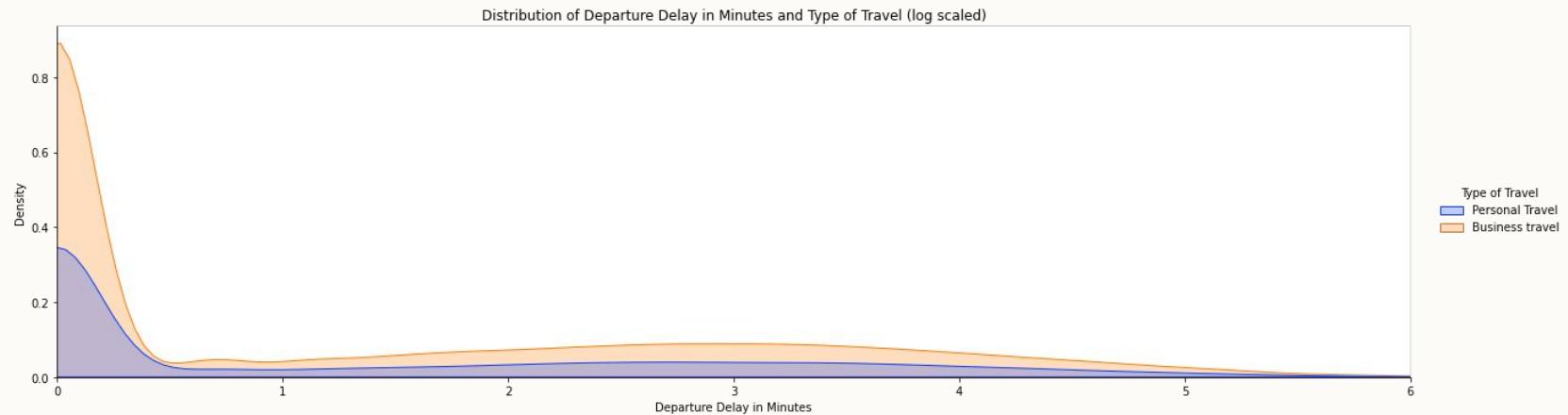
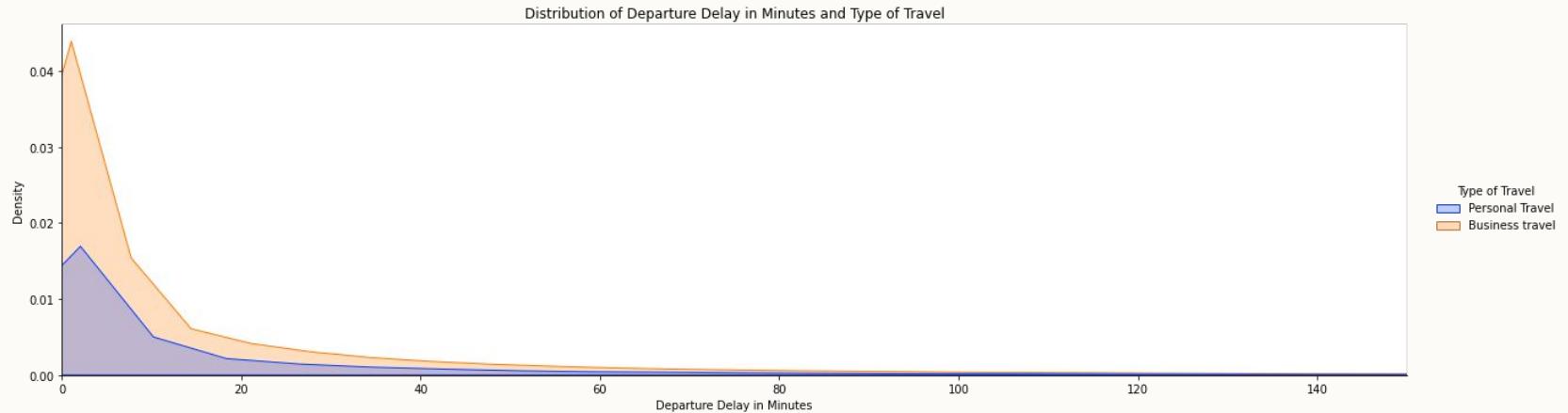
Personal traveler tend to have a good:

**Inflight wifi service** : for working, not wasting time

**Ease of online booking** : saving time

**Online boarding** : saving time

**Gate location** : saving time



	Type of Travel	Business travel	Personal Travel
Arrival Delay in Minutes	Mean	15.29	14.80
	Median	0.00	0.00
Departure Delay in Minutes	Mean	14.95	14.51
	Median	0.00	0.00

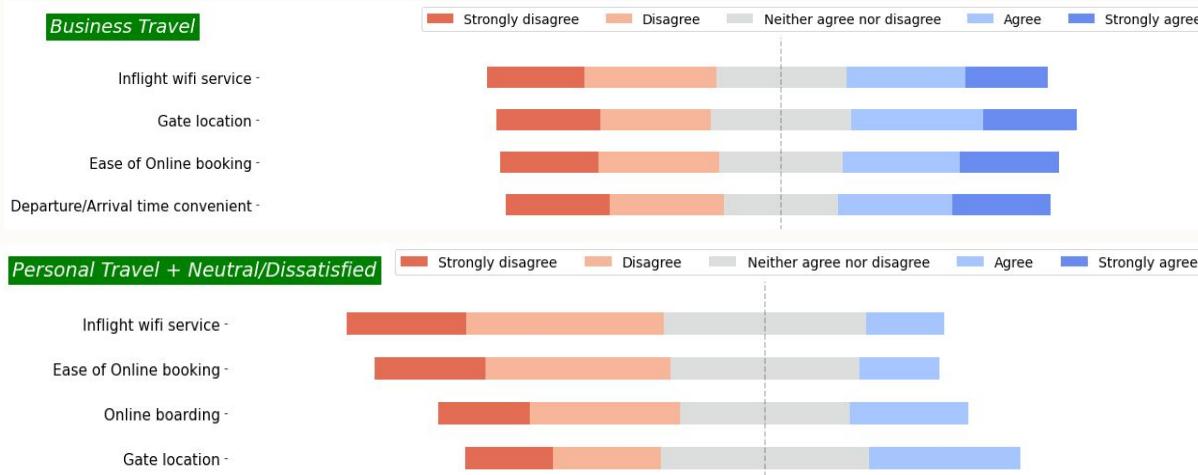
56.53%

Flights are not delay amongst **Personal Traveler**

56.44%

Flights are not delay amongst **Business Traveler**

# Business Recommendation



1. Focus on Business Traveler as the majority passengers, while keep improving the service for Personal Traveler that is Neutral/Dissatisfied
2. Improve the Inflight wifi service : Faster wifi for work or entertainment
3. Improve the online booking service : user friendly, responsive apps, livechat customer service
4. Gate location : maps, closer gate
5. Online boarding and Departure/Arrival time convenient

# Thanks!

---