SUPERVISED MACHINE LEARNING

AIRLINE PASSENGER SATISFACTION
WHAT FACTORS LEAD TO CUSTOMER SATISFACTION FOR AN AIRLINE COMPANY

KATERINA PSALLIDA –
WORK ASSOCIATED WITH THE "DATA SCIENCE BOOTCAMP" @ BIG BLUE DATA ACADEMY

AGENDA

Analysing the Dataset

Outcomes from Data Analysis

Prediction of Customer Satisfaction

Characteristics Affecting Customer Satisfactoin

THE DATASET
FROM KAGGLE

Airline Passenger Satisfaction (kaggle.com)

129880 Rows	○ 24 Features	2 Types : ✈ Descriptive ✈ Satisfaction Rating

DESCRIPTIVE:

Gender: Gender of the passengers (Female, Male)

Customer Type: The customer type (Loyal customer, disloyal customer)

Age: The actual age of the passengers

Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)

Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)

Flight distance: The flight distance of this journey

Departure Delay in Minutes: Minutes delayed when departure

Arrival Delay in Minutes: Minutes delayed when Arrival

Satisfaction: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Satisfaction Rating [0,5]

Inflight wifi service: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)

Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient

Ease of Online booking: Satisfaction level of online booking

Gate location: Satisfaction level of Gate location

Food and drink: Satisfaction level of Food and drink

Online boarding: Satisfaction level of online boarding

Seat comfort: Satisfaction level of Seat comfort

Inflight entertainment: Satisfaction level of inflight entertainment

On-board service: Satisfaction level of On-board service

Leg room service: Satisfaction level of Leg room service

Baggage handling: Satisfaction level of baggage handling

Check-in service: Satisfaction level of Check-in service

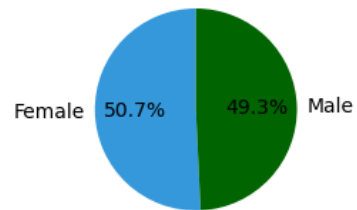**Inflight service:** Satisfaction level of inflight service

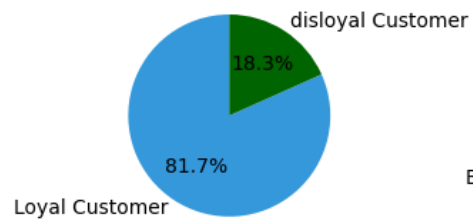**Cleanliness:** Satisfaction level of Cleanliness

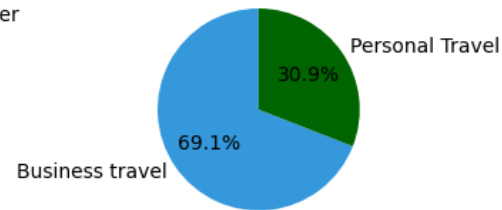# EXPLORATORY DATA ANALYSIS

DISTRIBUTION OF SOME OF THE VARIABLES

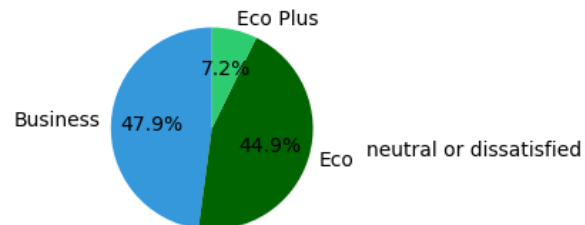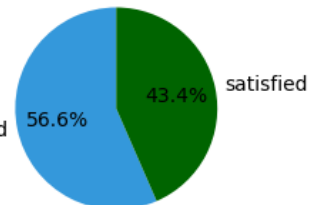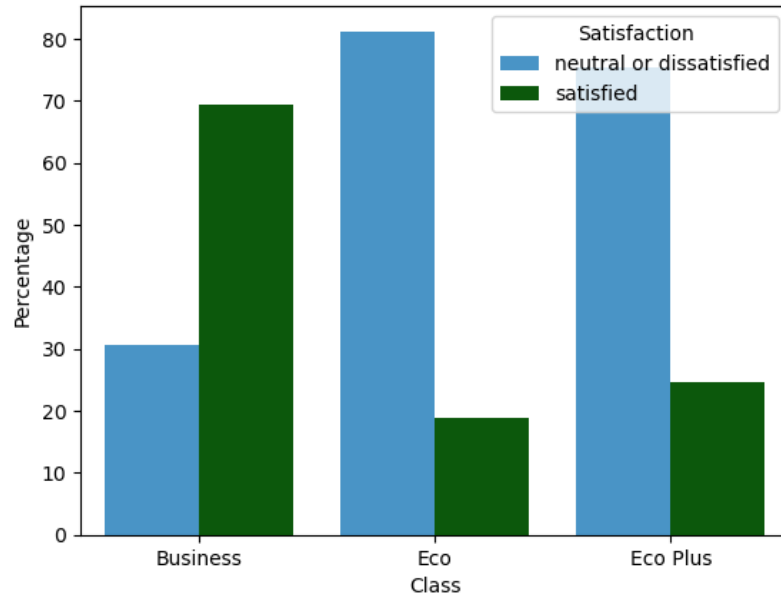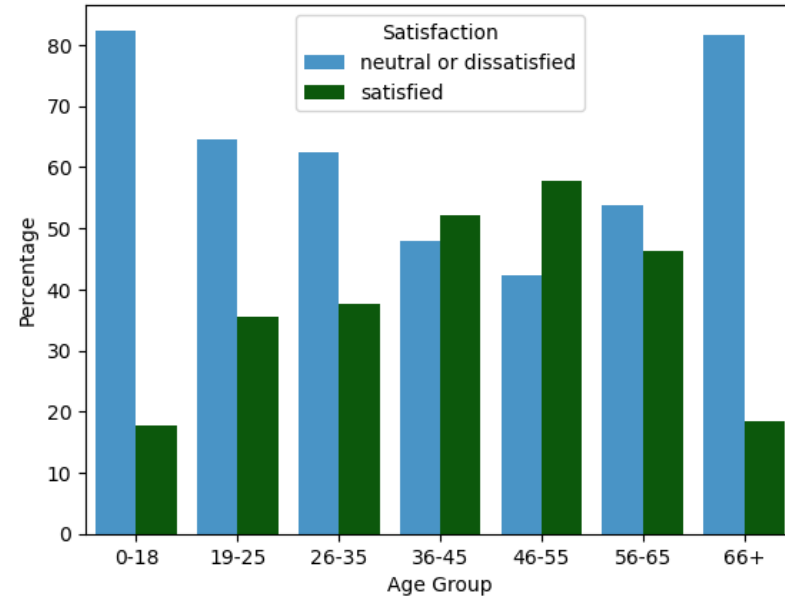- Same distribution of genders in giving feedback
- 80% of customers are loyal
- 70% of traveling is done for business
- Eco Plus is not a Class Preferred (7%).Business Class and Eco are almost in the same percentage 45%.
- The DataFrame is balanced considering the total satisfaction

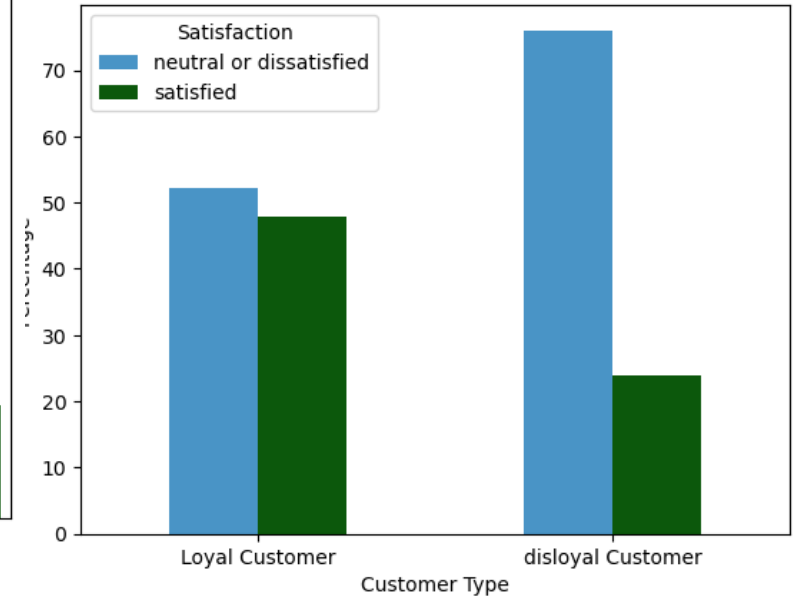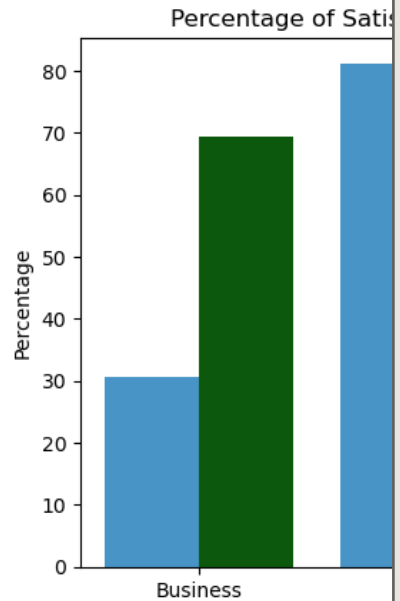# SATISFACTION FOR CLASS, AGE, CUSTOMER TYPE

SATI

Percentage of Satis...

Percentage

80
70
60
50
40
30
20
10
0

Business

Customer Type

disloyal Customer

Most satisfied customers are travelling business class
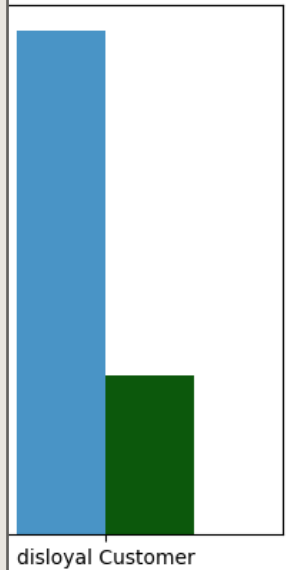Not satisfied are the ones from the less expensive classes.

Loyal Customers show the same percentace of satisfaction / non-satisfaction
Disloyal Customers are mostly dissatisfied

Mostly satisfied are the group ages 36-55.
Almost 50 -50 for the group 56-65.
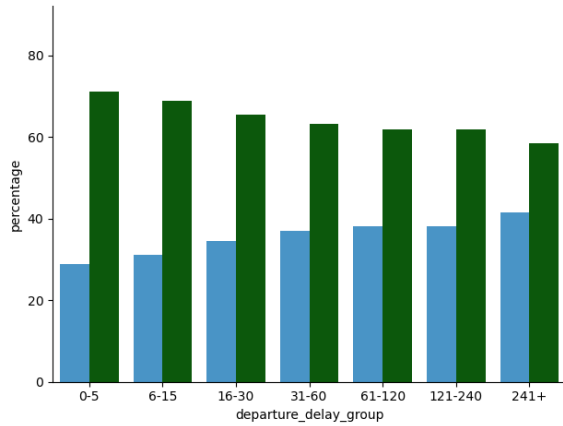Low satisfaction we have for the groups of very young 0-35 and 66+.
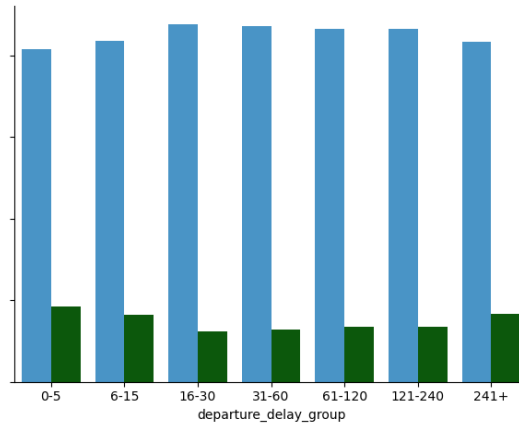
SATISFACTION DUE TO DEPARTURE DELAY
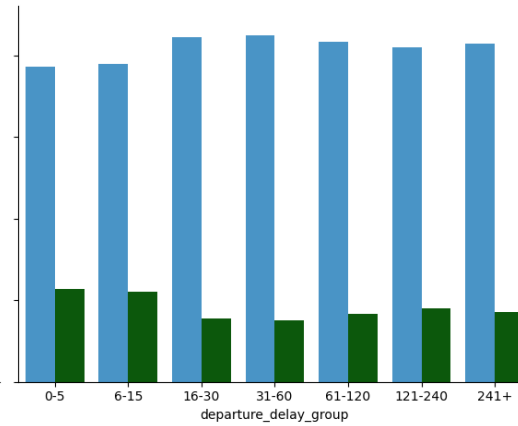
FOR CLASS, TYPE OF TRAVEL, CUSTOMER TYPE

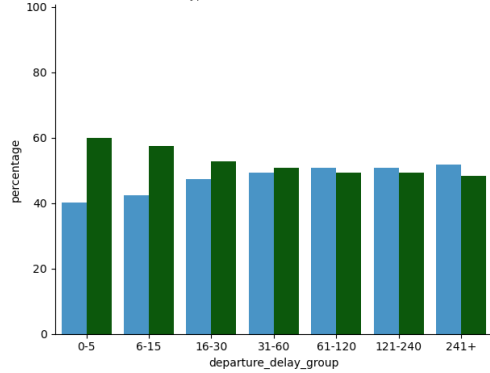Percentage of Satisfaction by Class and Departure Delay

Percentage of Satisfaction by Type of Travel and Departure Delay
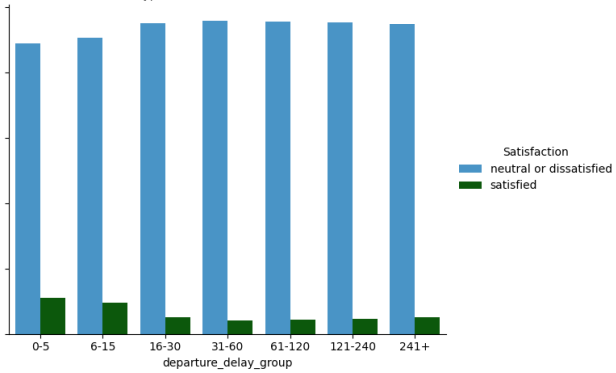
Percentage of Satisfaction by Customer Type and Departure Delay

SATISFACTION DUE TO DEPARTURE DELAY

FOR CLASS, TYPE OF TRAVEL, CUSTOMER TYPE

✈ Largest percentage of ECO and ECO Plus are dissatisfied, regardless of the flight delay
The group that is mostly satisfied is the BUSINESS class which is satisfied again regardless of the DELAY.
For larger delays there is a slight decline in satisfaction and more dissatisfaction occurring

✈ Disloyal Customers have a high percentage of being disatisfied - small percentage of them might be satisfied , again regardless of the delay
Loyal Customers have a higher percentage of dissatisfaction although this drops with increasing delays

✈ People travelling for Personal reasons sow more dissatisfaction.
Business Travellers' satisfaction and dissatisfaction is in equal proportions for large delays but for smaller delays the satisfaction is higher
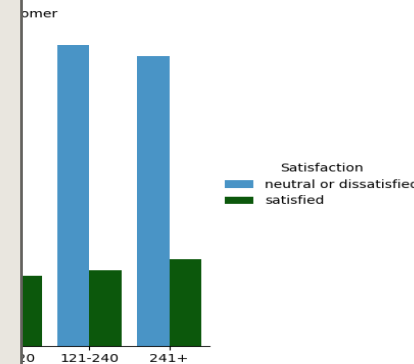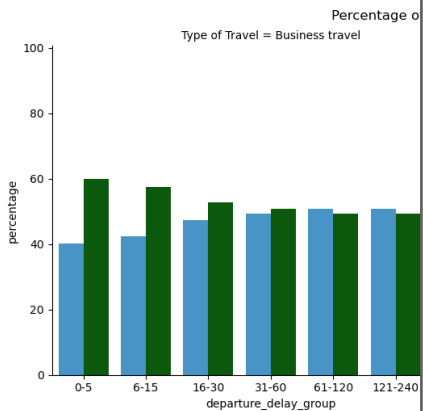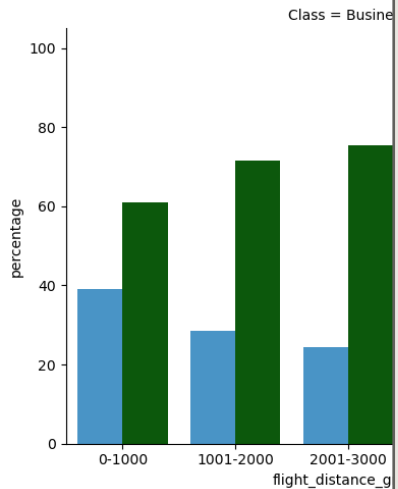
SATISFACTION DUE TO FLIGHT DURATION

FOR CLASS, TYPE OF TRAVEL, CUSTOMER TYPE
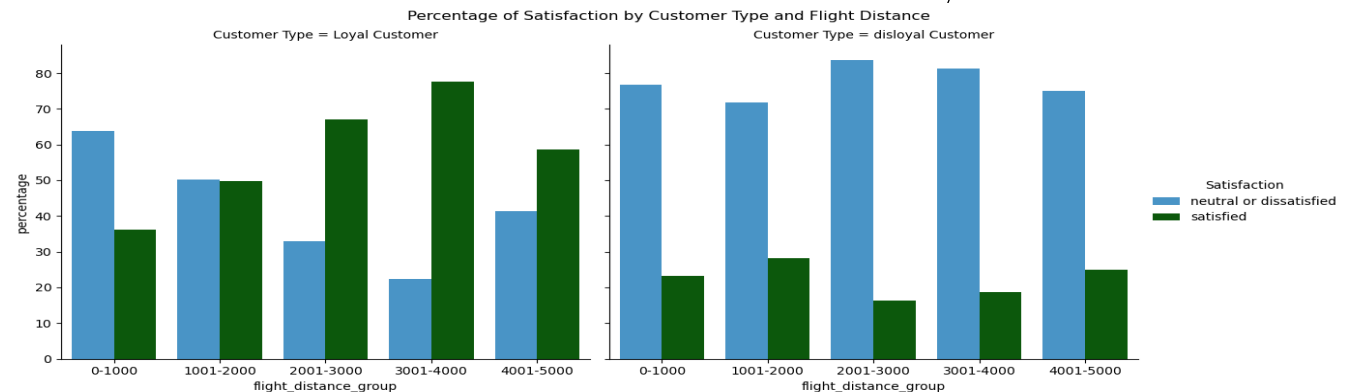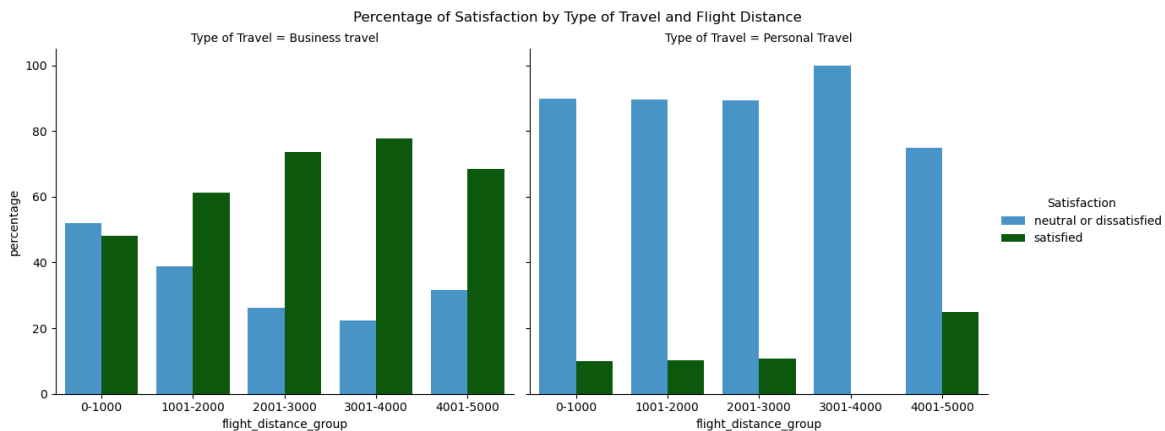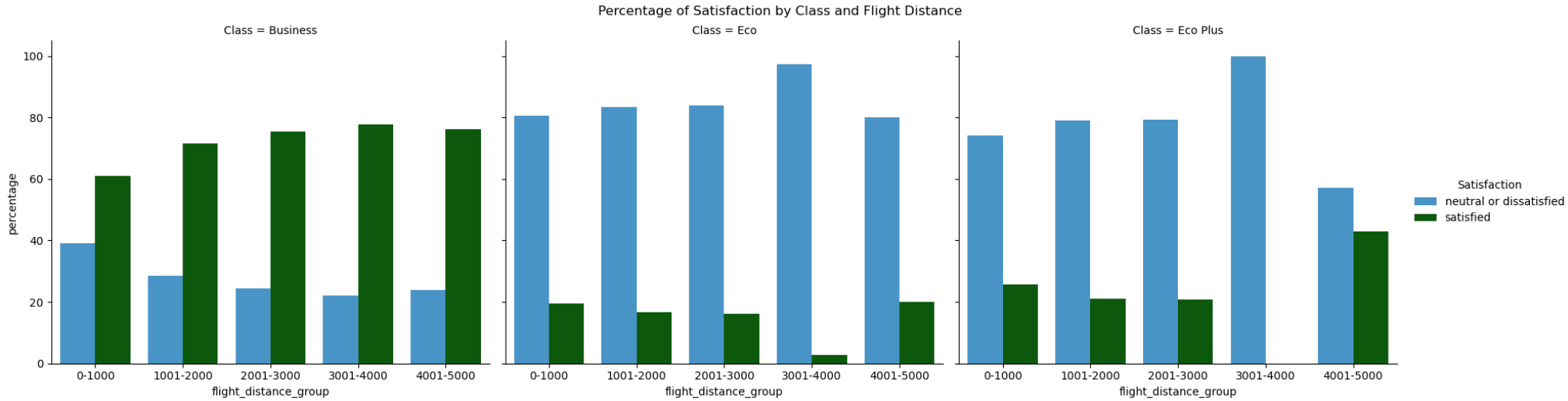
SATISFACTION DUE TO FLIGHT DURATION

FOR CLASS, TYPE OF TRAVEL, CUSTOMER TYPE

✈ ECO and ECO Plus are generally dissatisfied in very high percentages for all travel durations. For Long Flights and for the ECO Plus Group the percentage of satisfied increases a bit

✈ ECO and ECO PLUS have extremely small satisfaction percentage for the duration of 3000 - 4000 mins. Explore with Marketing

✈ Business Class is satisfied in higher percentages for all travel durations. Percentage of satisfied customers in business class increases with duration

✈ Disloyal customers have very high proportions of dissatisfaction for all travel durations

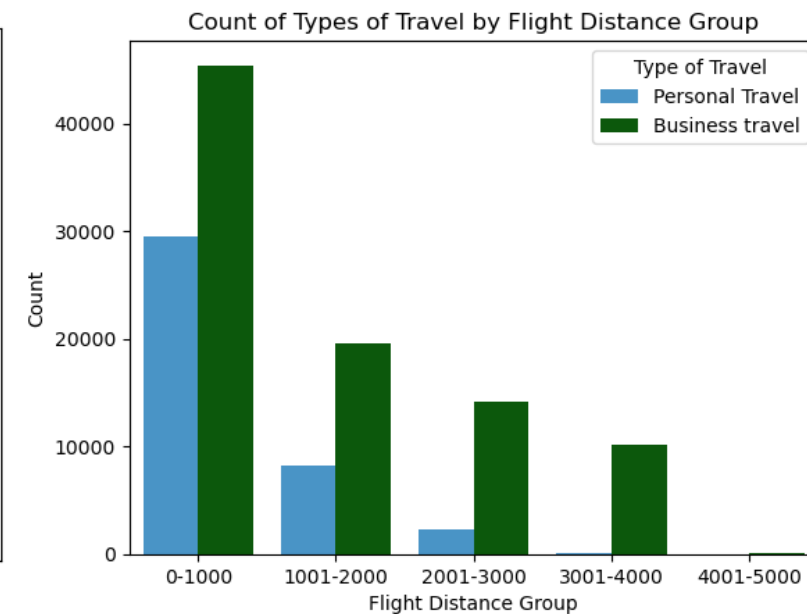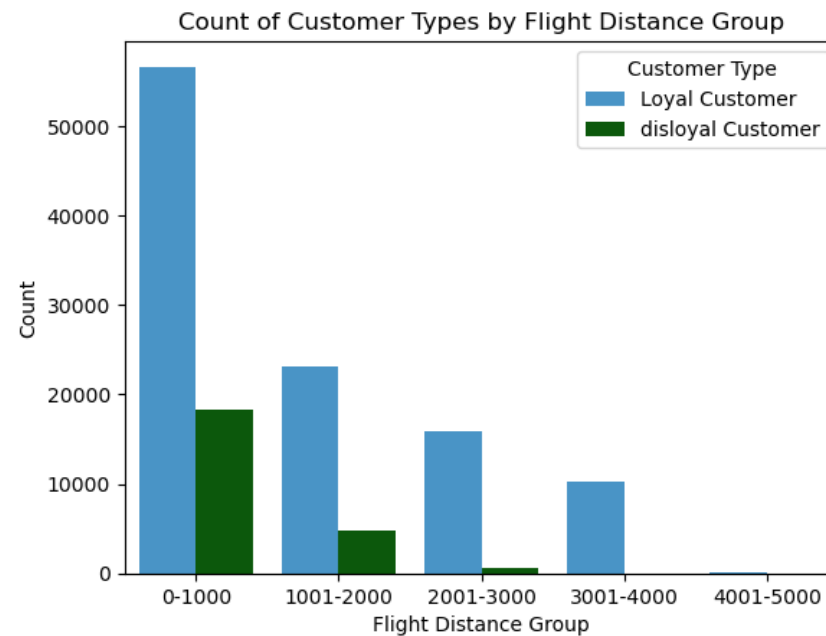✈ Loyal Customers have increasing proportions of satisfaction for larger travel durations. For small flight durations (30 - 1000 mins), there is a larger proportion of loyal customers being dissatisfied. Explore with Marketing

✈ Personal Travelers have highest proportions of being dissatisfied for all durations For 3000 - 4000 mins we have very low satisfaction rates. Investigate what happens for those flights
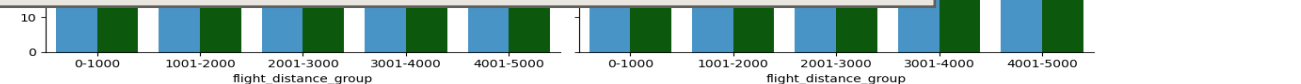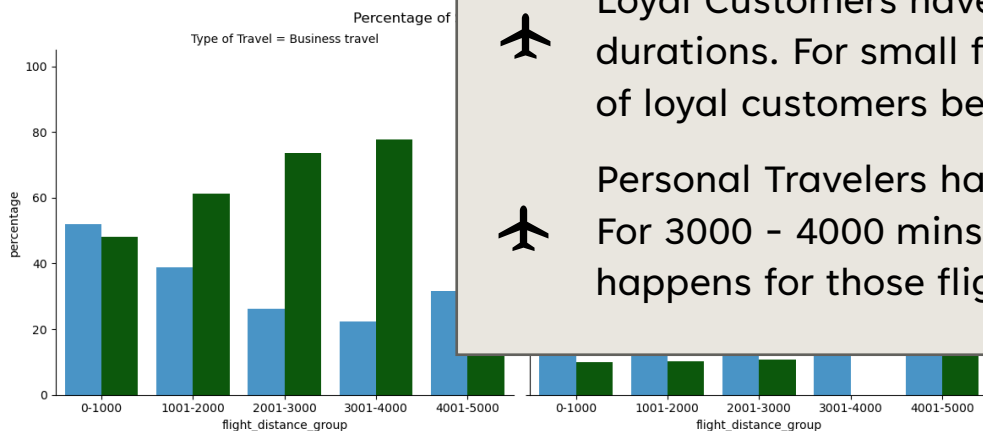
CLASS, TYPE OF TRAVEL, CUSTOMER TYPE
FOR VARIOUS FLIGHT DISTANCES

Count of Classes by Flight Distance Group

Count of Customer Types by Flight Distance Group

Count of Types of Travel by Flight Distance Group
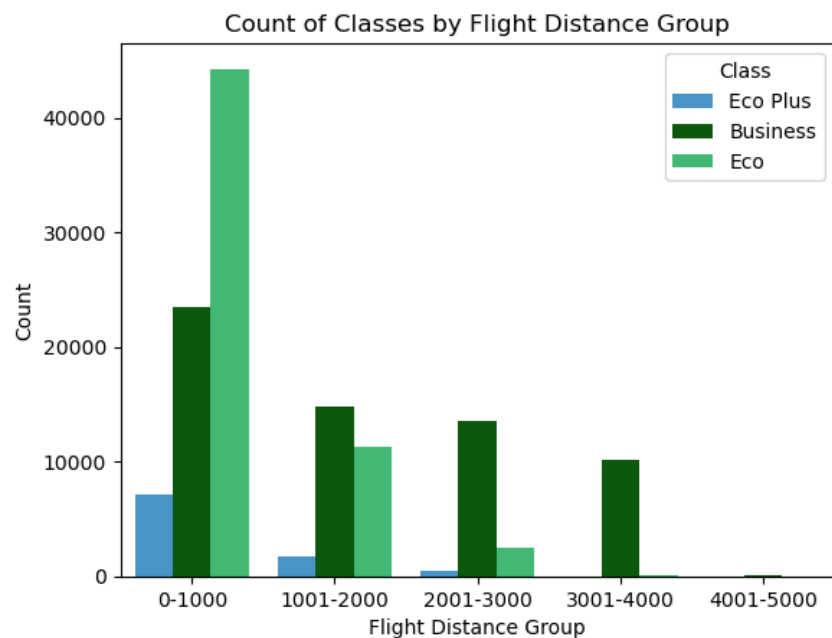
CLASS, TYPE OF TRAVEL CUSTOMER TYPE
IN VARIOU
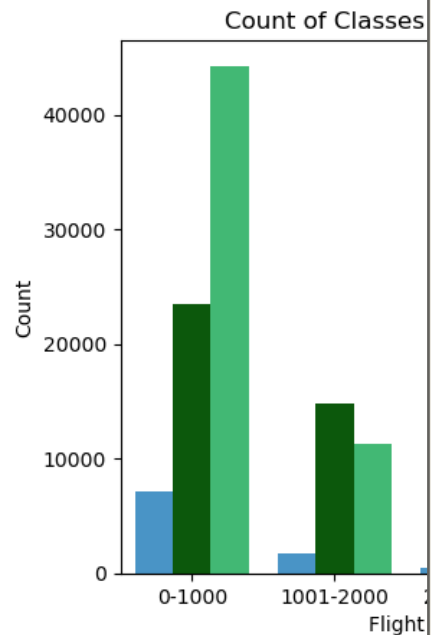

✈ Most customers are traveling in short flights 0-1000 mins, and belong to the ECO (cheapest group), followed by the business class.

✈ We know before hand that we do not have many ECO PLUS customers. This is either due to the dataset or due to a fault in how ECO PLUS is priced. Inform Pricing Unit

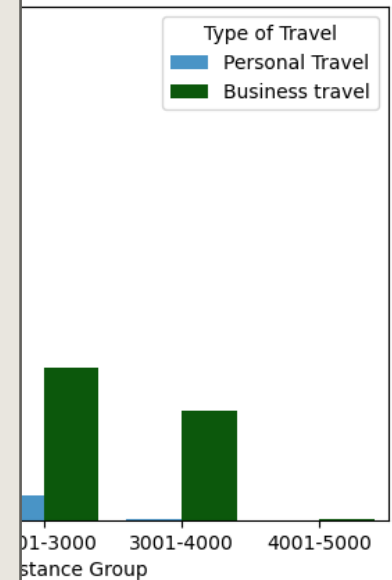✈ Most customers in all flight distances are Loyal

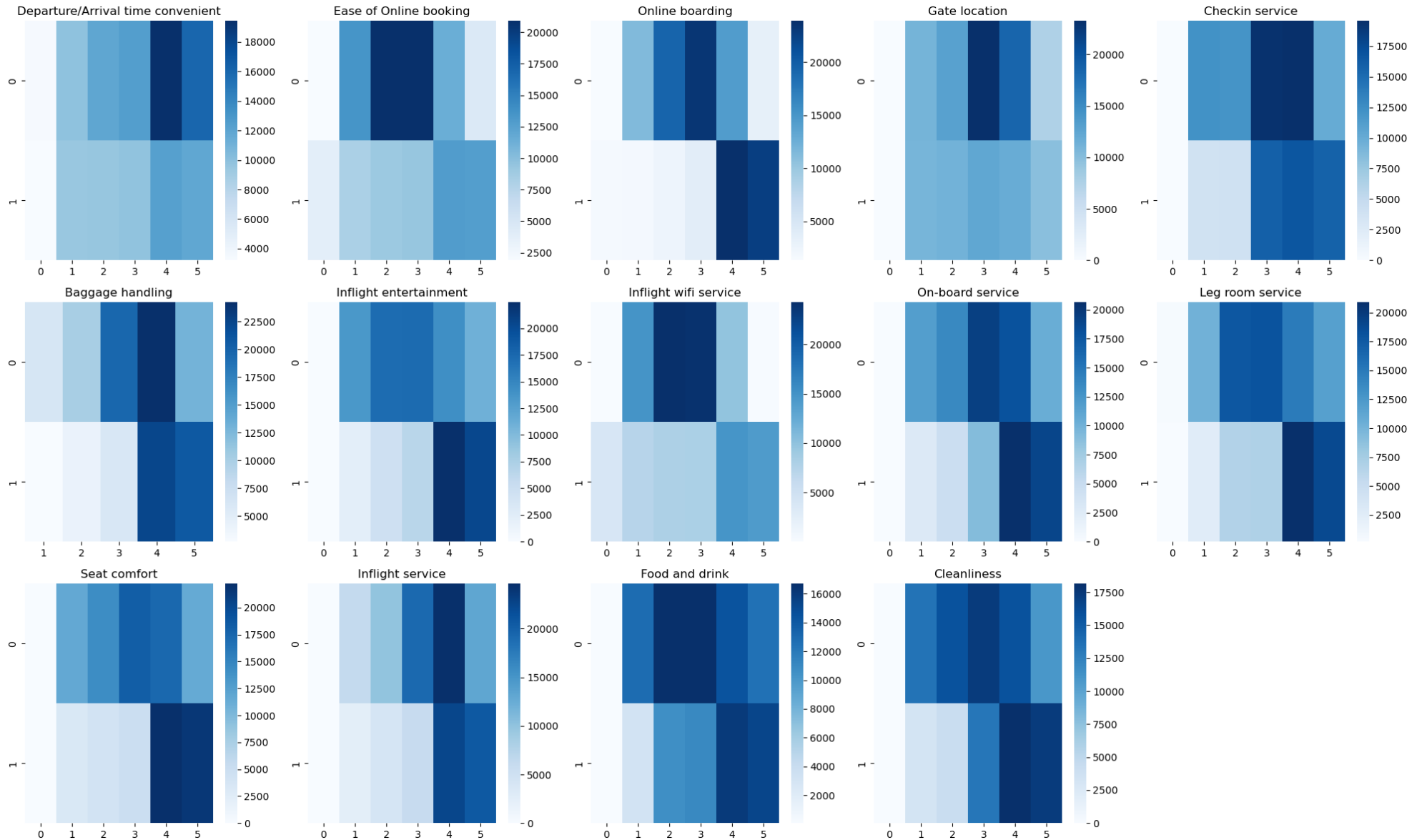✈ Most types of travel for all distances are Business Travel

✈ Personal travel is usually shorter than 2000 mins
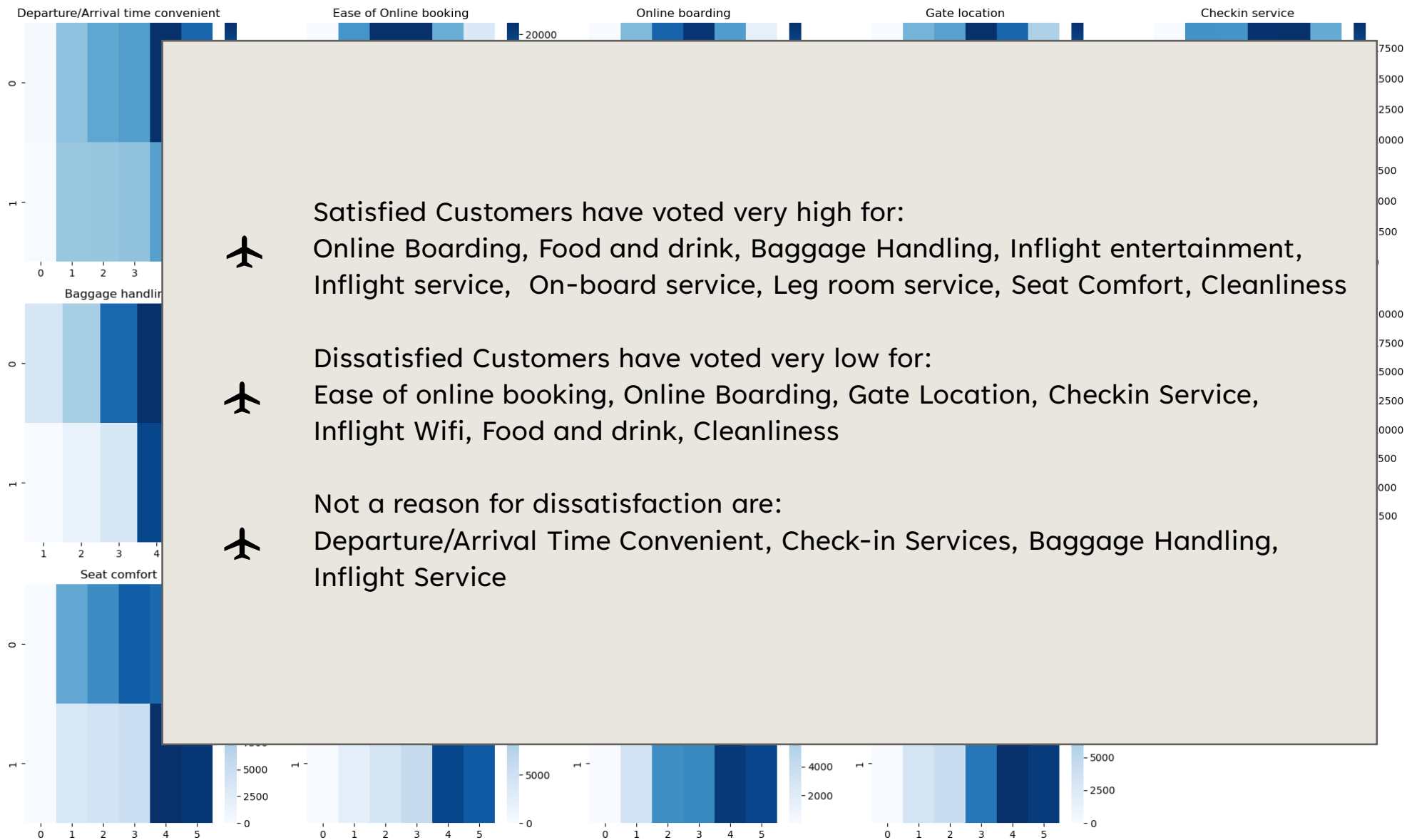Flight Durations larger than 4000 mins are not used frequently at all.

SATISFACTION RATING OF INDIVIDUAL SERVICES
CONTRIBUTING TO OVERALL SATISFACTION

SATISFACTION RATE OF SERVICES
CONTRIBUTING TO OVERALL SATISFACTION

✈ Satisfied Customers have voted very high for:
Online Boarding, Food and drink, Baggage Handling, Inflight entertainment, Inflight service, On-board service, Leg room service, Seat Comfort, Cleanliness

✈ Dissatisfied Customers have voted very low for:
Ease of online booking, Online Boarding, Gate Location, Checkin Service, Inflight Wifi, Food and drink, Cleanliness

✈ Not a reason for dissatisfaction are:
Departure/Arrival Time Convenient, Check-in Services, Baggage Handling, Inflight Service

PREDICT
A SATISFIED CUSTOMER

PREPARE THE DATA

NAN Values

Distribution of Departure Delays

Deleted 300 rows in 123K

NEW Feature creation of Delay =[0,1] due to extreme skewness of Departure delay

Correlated Variables
Did not act upfront

- Transformed categorical variables into dummies

- Ordinal Encoding of the travel Class (Eco, EcoPlus Business)
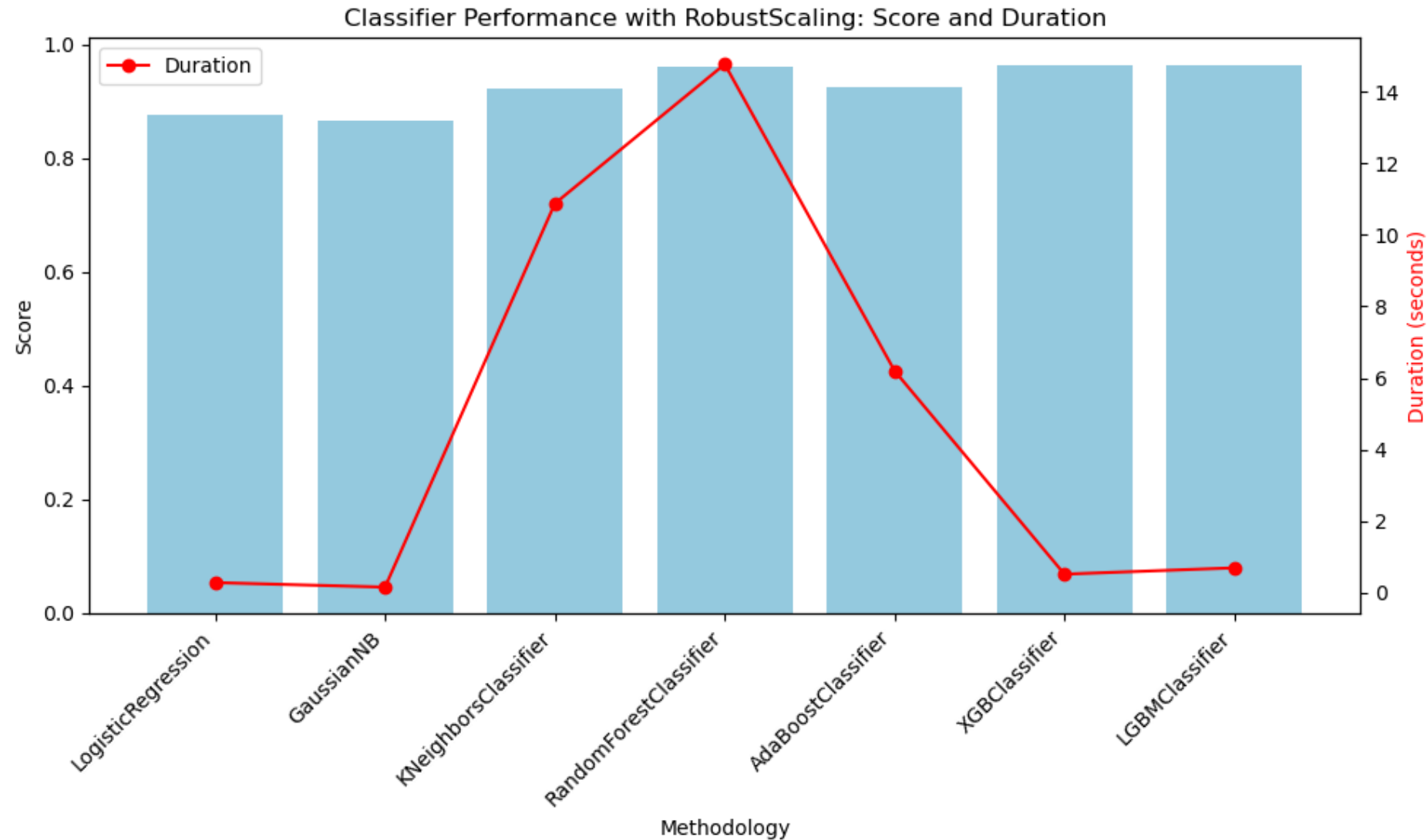
- Scaled with Robust Scaler due to Outliers
- Tested also with MinMax, Standard, and Power Scalers

- Outliers were not dropped as they seemed viable for a business case

TESTING CLASSIFIER MODELS

Classifier Performance with RobustScaling: Score and Duration

- Here we are presenting models with running duration < 15 second

- SVC run for 1799 secs

- Score selected was accuracy since the dataset was balanced

- Selected XGBClassifier for best score and best runtime
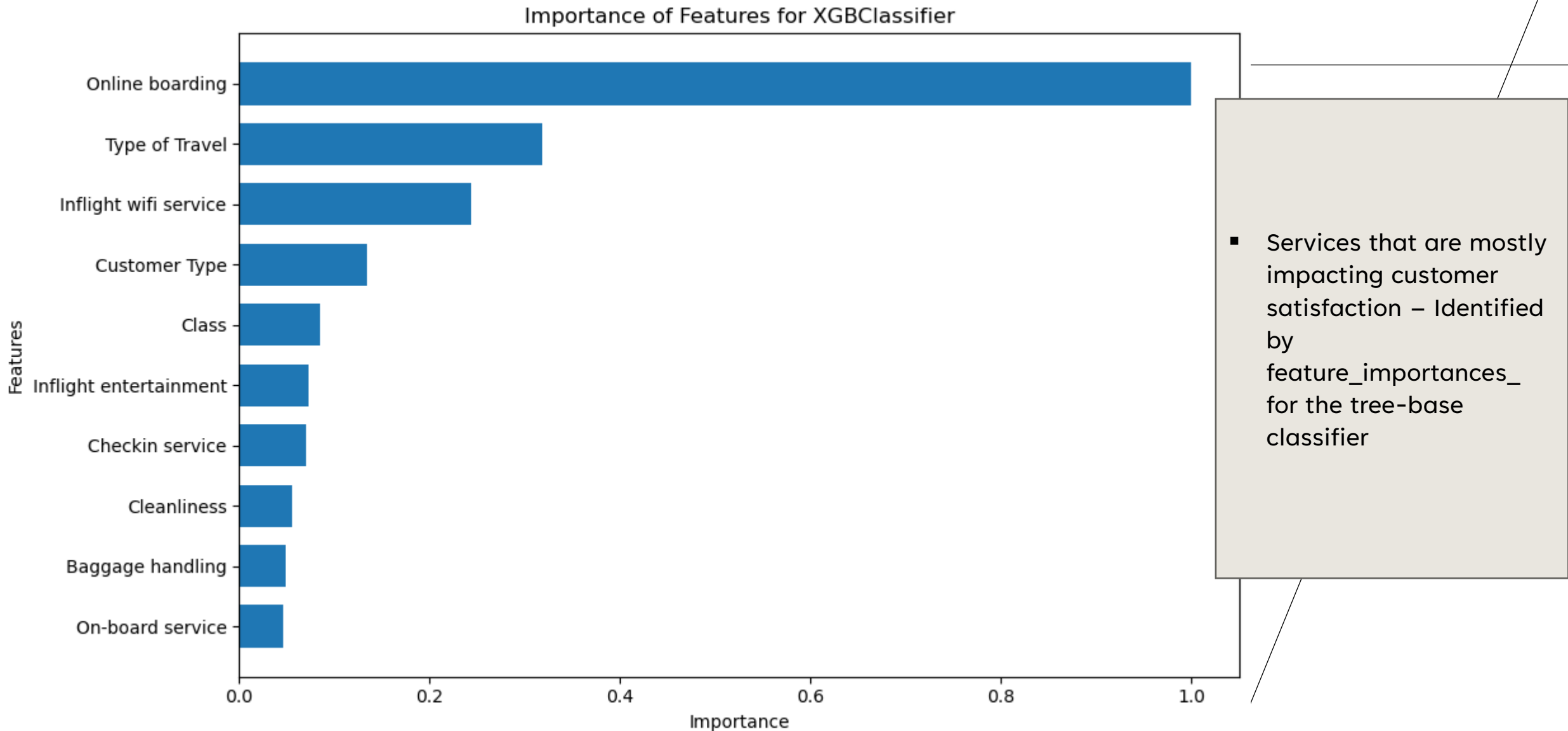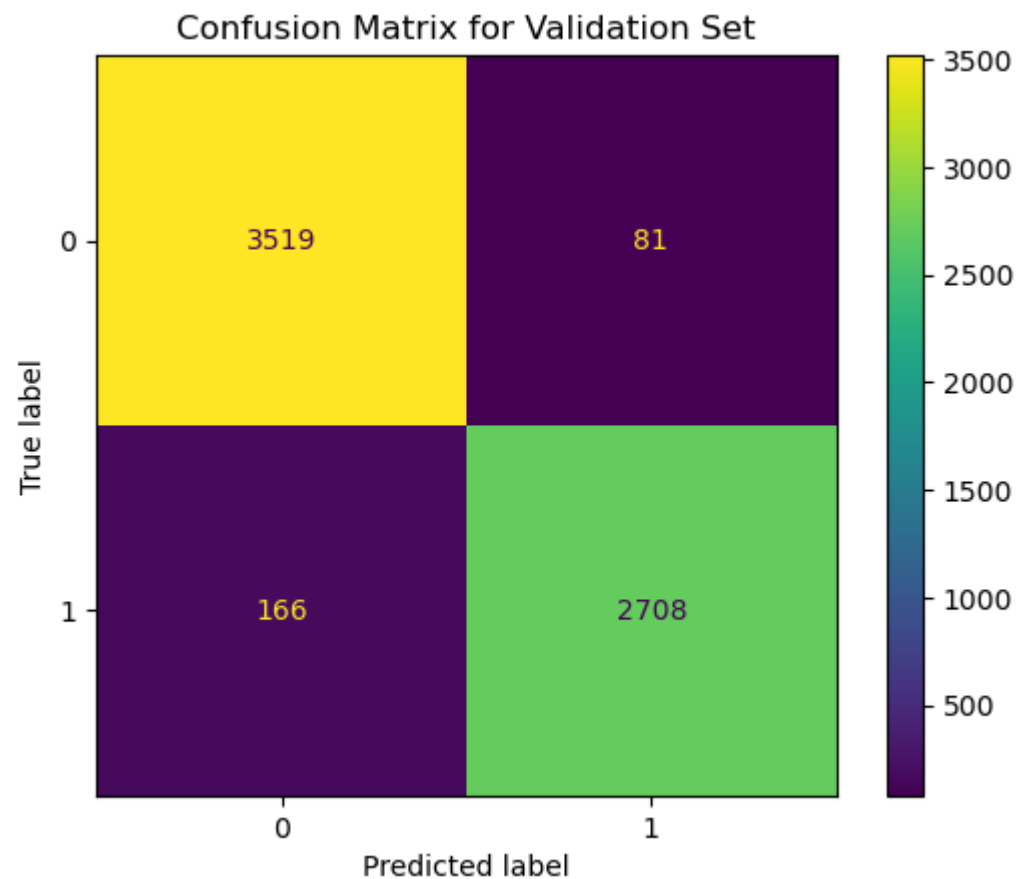- XGB accuracy score = 0.964

SELECTING XGBCLASSIFIER

ROC Curve for All Classifiers

Confusion Matrix for XGBClassifier
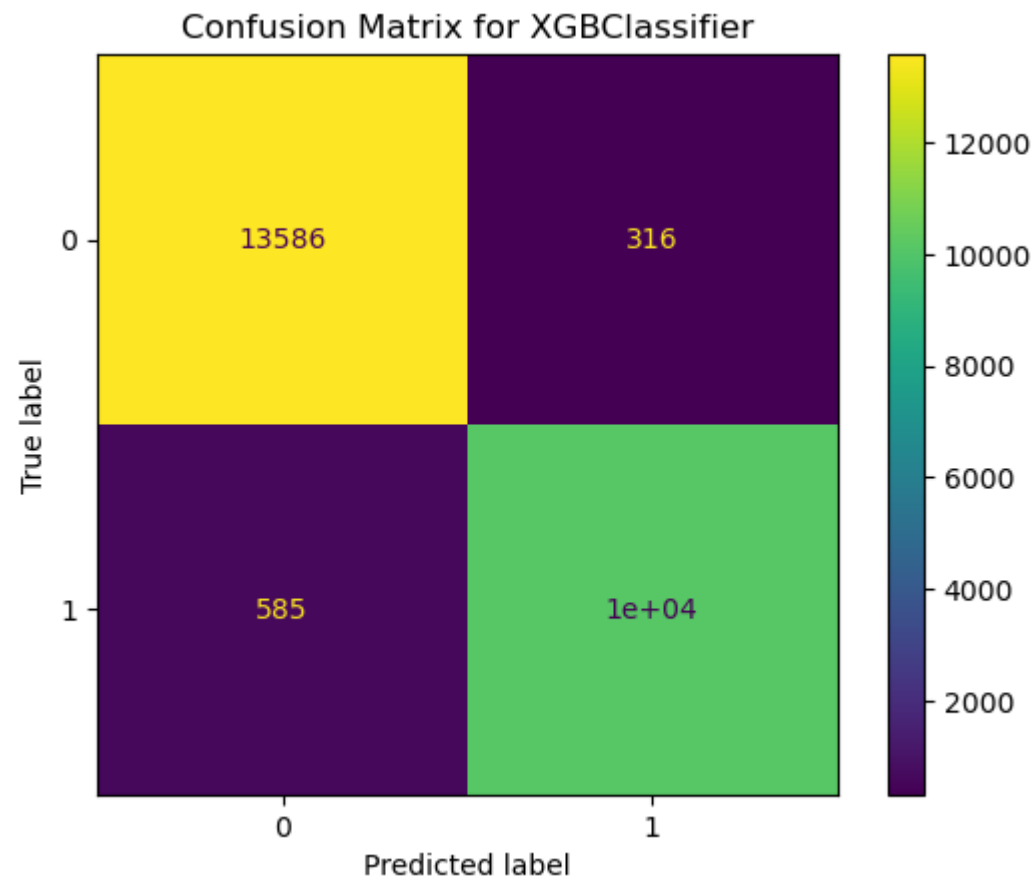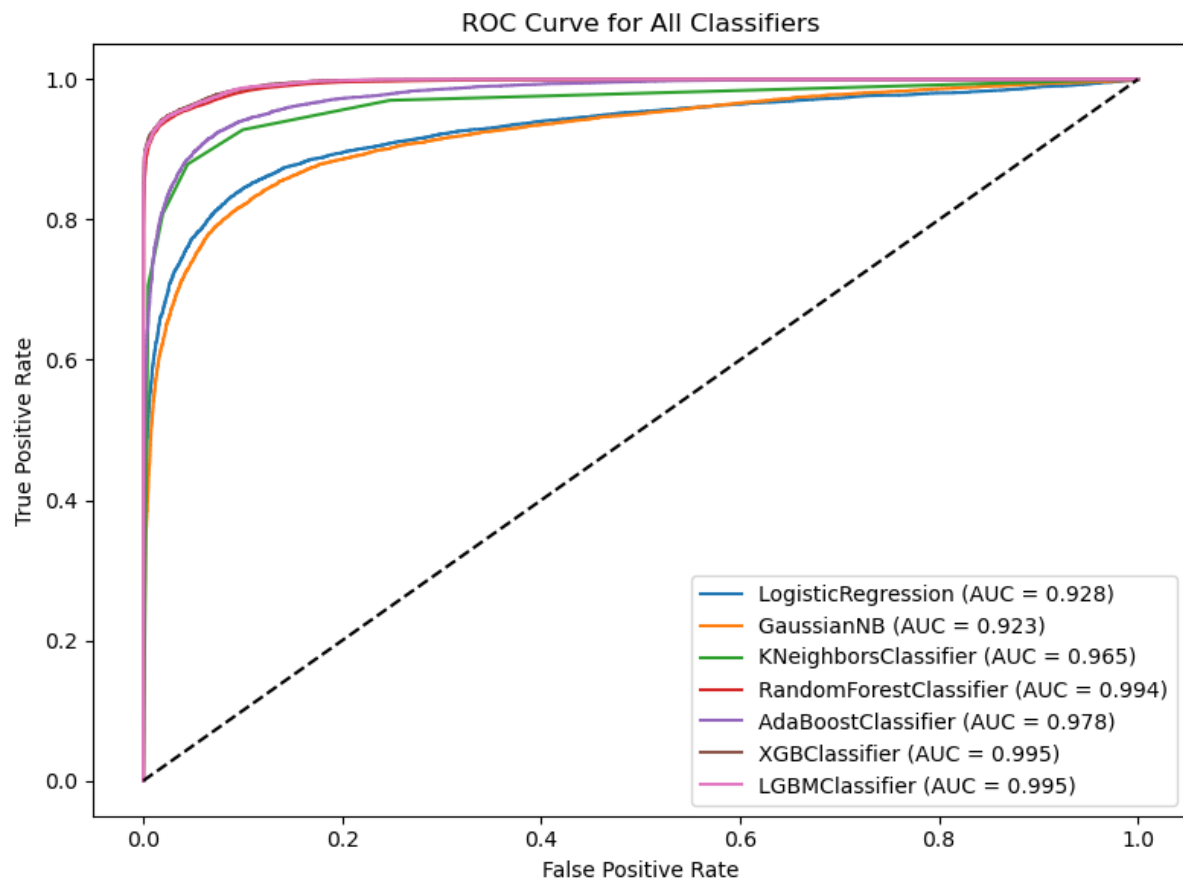
WHAT AFFECTS CUSTOMER SATISFACTION

10 MOST IMPORTANT FEATURES / SERVICES TO BE PROVIDED

Importance of Features for XGBClassifier

- Services that are mostly impacting customer satisfaction – Identified by feature_importances_ for the tree-base classifier

VALIDATION

Confusion Matrix for Validation Set

- Validated the model to a part of the dataset that was never seen by the model.
- The accuracy of the validation was 0.942 ensuring that the model has not overfitted.

THE END