



Final Project

---

# Airline Passenger Satisfaction

---

Presented by  
Ivan Reinaldo  
Margaretha Elaine Nauli

---



# Group Profile

## Ivan Reinaldo

### Work Experience

**Junior & Senior Unity Developer, FXMedia (2018–2022)**

- Developed XR (AR, VR, MR, and Desktop) applications using Unity and other tools for internal and clients
- Collaborated with other teams, such as 3D Artists, Animators, 2D Designers and Web on several projects
- Handled application publishing to app stores for both internal and clients' projects
- Mentored and planned tasks for interns

**Information Technology Intern, The Ivy League Group, LLC (2017)**

- Redesigned clients' websites using WordPress
- Wrote contents and edited videos for sites
- Assisted on wireless network configuration



reinaldoivan30@gmail.com



<https://www.linkedin.com/in/ivan-reinaldo/>

## Margaretha Elaine Nauli

### Work Experience

**Data Analyst, Sampoerna Kayoe (2020–2022)**

- Handled and restarted the performance management system (SAP Success Factor)
- Support initiative programs for improving company's performance
- Create ad hoc report for Board of Directors
- Generate monthly report for directorate usage

**Document Controller, PT AEON Indonesia (2019–2020)**

- Identified operation problem and find root cause for Standard Operation Procedure (SOP) revision
- Improved business processes related to operation division
- Generated operation related SOP



margaretha.elaine@gmail.com



<https://www.linkedin.com/in/margaretha-elaine-n-307b0347/>

# Presentation Outline



## Today's Topics

Group Profile >

Problem Formulation >

Data Understanding >

Findings and Solution >

Conclusion and Recommendation >



# Airline Passenger Satisfaction

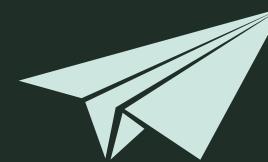
---

Understanding the factors that influence travellers in their selection of an airline is fundamental for achieving growth in this market.

Therefore, this study is to work out the determinants of customer satisfaction in the airlines industry because if passengers remain satisfied, the industry will be more competitive and this will help ensure its survival.

## Dataset information:

- Source : <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>
- The data represent customer survey done by an **US Airline in 2018**
- Survey data included basic customer's information and their rating to extensive list of airline services
- This project is dedicated for Airline companies to understand their customer and to calculate the possible income/cost that will occur in the process





Did you know?



# 79% Customers

who have a bad experience  
will tell others about it

next

85% just want to warn others  
66% want to actively try and stop others  
from using the company's services

Meanwhile  
If they have the best experince, only 55%  
are willing to recommend a company's  
service to others

# Model Valuation Metric

Confusion Matrix is a performance measurement for machine learning classification. It is a table with 4 different combinations of predicted and actual values.

		predicted	
		n	P
actual	n	TP	FN
	P	FP	TN

*Confusion matrix*

<b>False Positive</b>	<ul style="list-style-type: none"><li>Not satisfied, but predicted as one</li><li>Find potential loss (per Customer)</li></ul>
<b>True Positive</b>	<ul style="list-style-type: none"><li>Actual satisfied customers</li><li>Find saving on incentive cost</li></ul>
<b>True Negative</b>	<ul style="list-style-type: none"><li>Actual neutral or dissatisfied customers</li><li>Find campaign cost and potential income</li></ul>
<b>False Negative</b>	<ul style="list-style-type: none"><li>Not dissatisfied, but predicted as one</li><li>Find campaign cost</li></ul>

The final model result shows that this model will give the company a possible income of \$1.273.034

# Problem Formulation

## ● ○ ○ Insights

What are the insights that we can get from the data?

## ● ● ○ Factors

What are the significant factors that will affect customer's satisfaction?

## ● ● ● Improvements

Are there any improvements that we can do to increase the customer's satisfaction level?



# Data Understanding

- 103.594 rows
- 23 features
- 19 numerical features (N)
- 4 categorical features (C)

Each row represent a customer satisfaction survey for airplane services.

## Basic Info

- Gender (C)
- Customer Type (C)
- Age (N)
- Type of Travel (C)
- Class (C)
- Departure Delay in Minutes (N)
- Arrival Delay in Minutes (N)
- Satisfaction (C)

## Pre-Flight Exp.

- Ease of online booking (N)
- Online boarding (N)
- Departure/Arrival time convenient (N)

## Airport Exp.

- Checkin service (N)
- Baggage handling (N)
- Gate location (N)

## Inflight Exp.

- Cleanliness (N)
- Inflight Wifi Service (N)
- Leg Room Service (N)
- Inflight Entertainment (N)
- Seat Comfort (N)
- Food and Drink (N)
- On-board Service (N)



Score range: 1-5

**Target feature**  
Satisfaction (C)



Neutral or Dissatisfied  
Satisfied

# Feature Engineering

We will add 5 new features that will help us in getting more insight from this dataset.

<b>Total Delay</b>	Departure Delay in Minutes + Arrival Delay in Minutes
<b>Overall Satisfaction</b>	Average score of all rating
<b>Pre-Flight Experience</b>	Ease of online booking + Online boarding + Departure/Arrival time convenient
<b>Airport Experience</b>	Checkin service + Baggage handling + Gate location
<b>Inflight Experience</b>	Cleanliness + Inflight wifi service + Inflight service + Leg room service + Inflight entertainment + Seat comfort + Food and drink + On-board service

# Findings and Solution

## ● ○ ○ Basic Assumptions

Generate assumptions

## ● ● ○ EDA

Explore dataset to prove assumptions

## ● ● ● Modeling

Build model to find feasible solutions



# Basic Assumptions

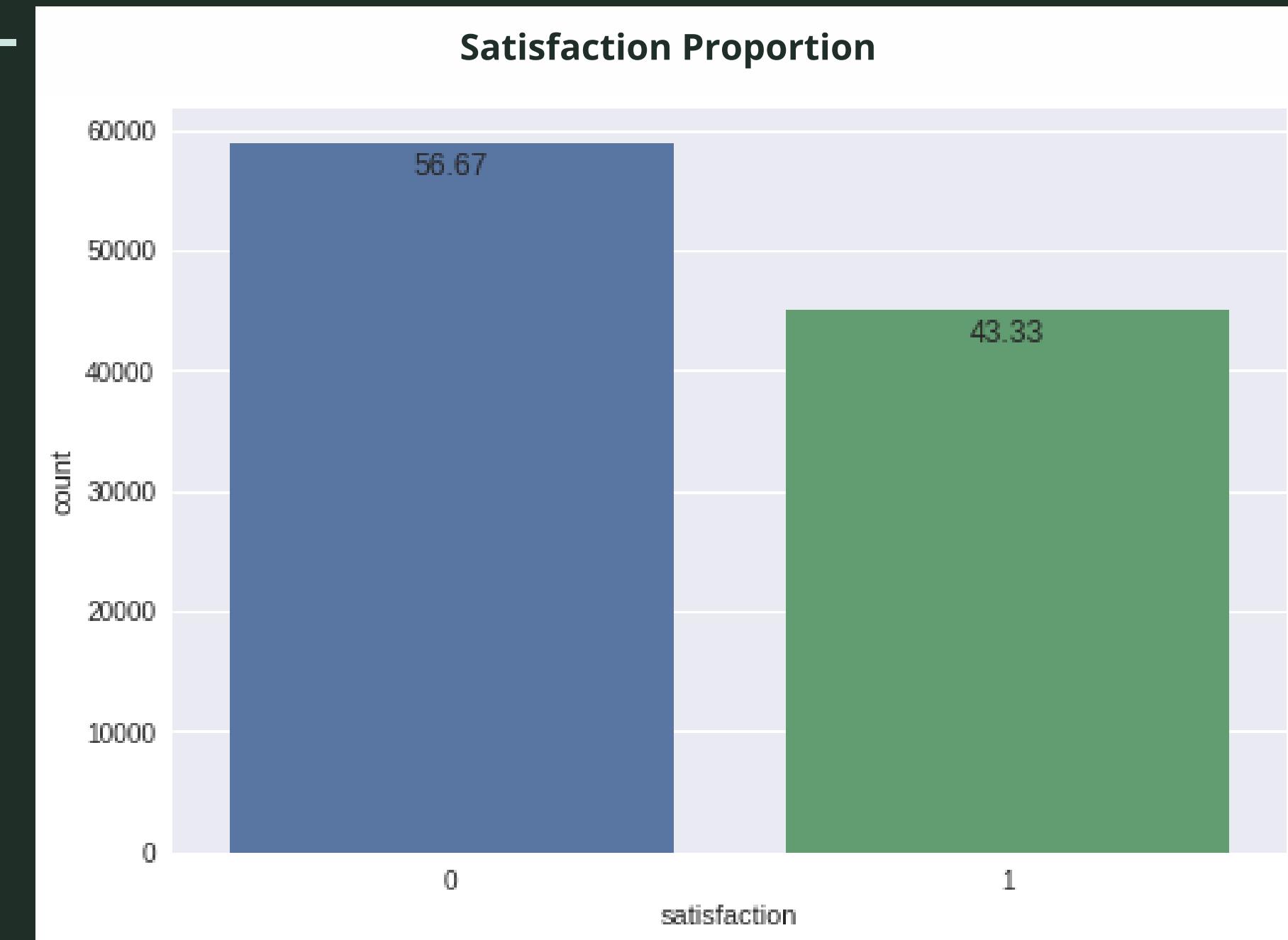
---

1. The number of satisfied customer should be more than dissatisfied
2. Business Class customer should have higher satisfaction
3. Flight distance should be more or less the same across all class
4. The number of loyal customer should be more than the disloyal
5. People who travel for Business usually take Business Class
6. Customer's loyalty should increase parallel with customer's age
7. Older customer prefer Business class

# Exploratory Data Analysis

01 .  
The number of satisfied customer  
should be more than dissatisfied

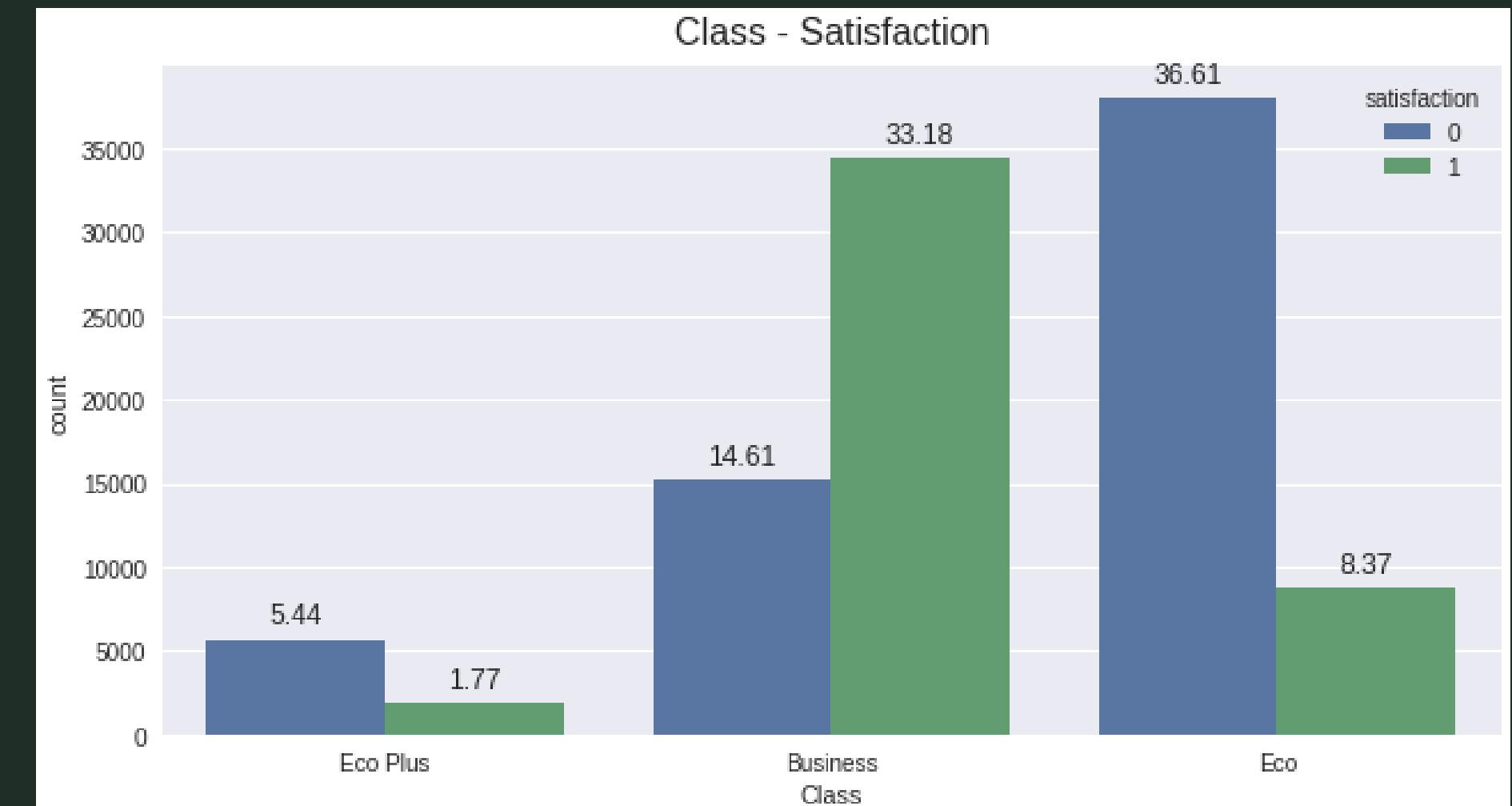
The assumption is false,  
as the number of neutral or  
dissatisfied customers (56.67%)  
is more than the satisfied ones  
(43.33%).



# Exploratory Data Analysis

02.  
Business Class customer should have  
higher satisfaction

The assumption is true,  
as the highest number of  
satisfied customers is on  
Business class (33.18%).  
However we should also pay  
attention to the number of  
neutral or dissatisfied customer  
on Eco class (36.61%).

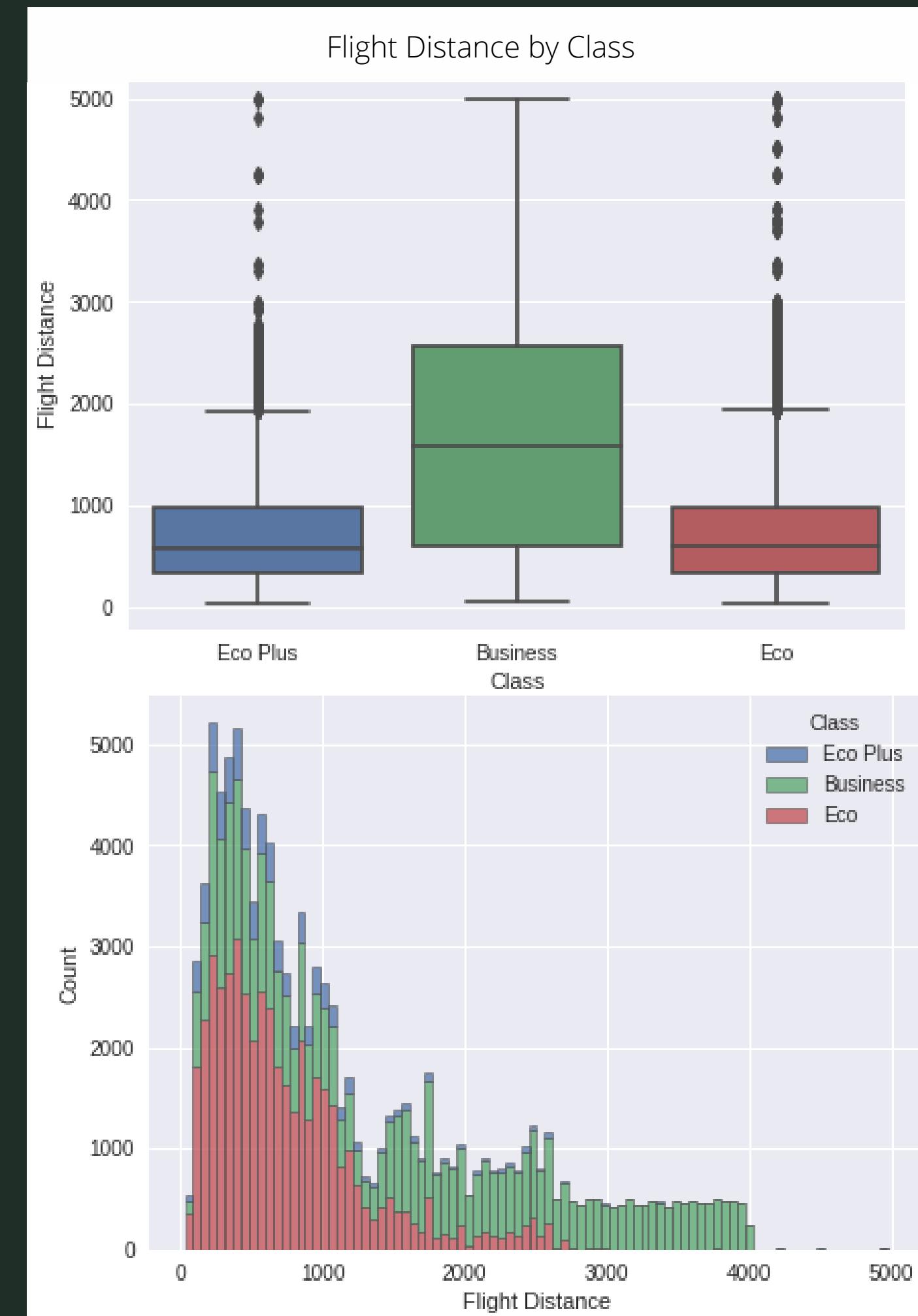


# Exploratory Data Analysis

03.  
Flight distance should be more or less the same across all class

The assumption is false,  
as more customers are using Business  
class on longer flight than the Eco  
Plus or Eco class.

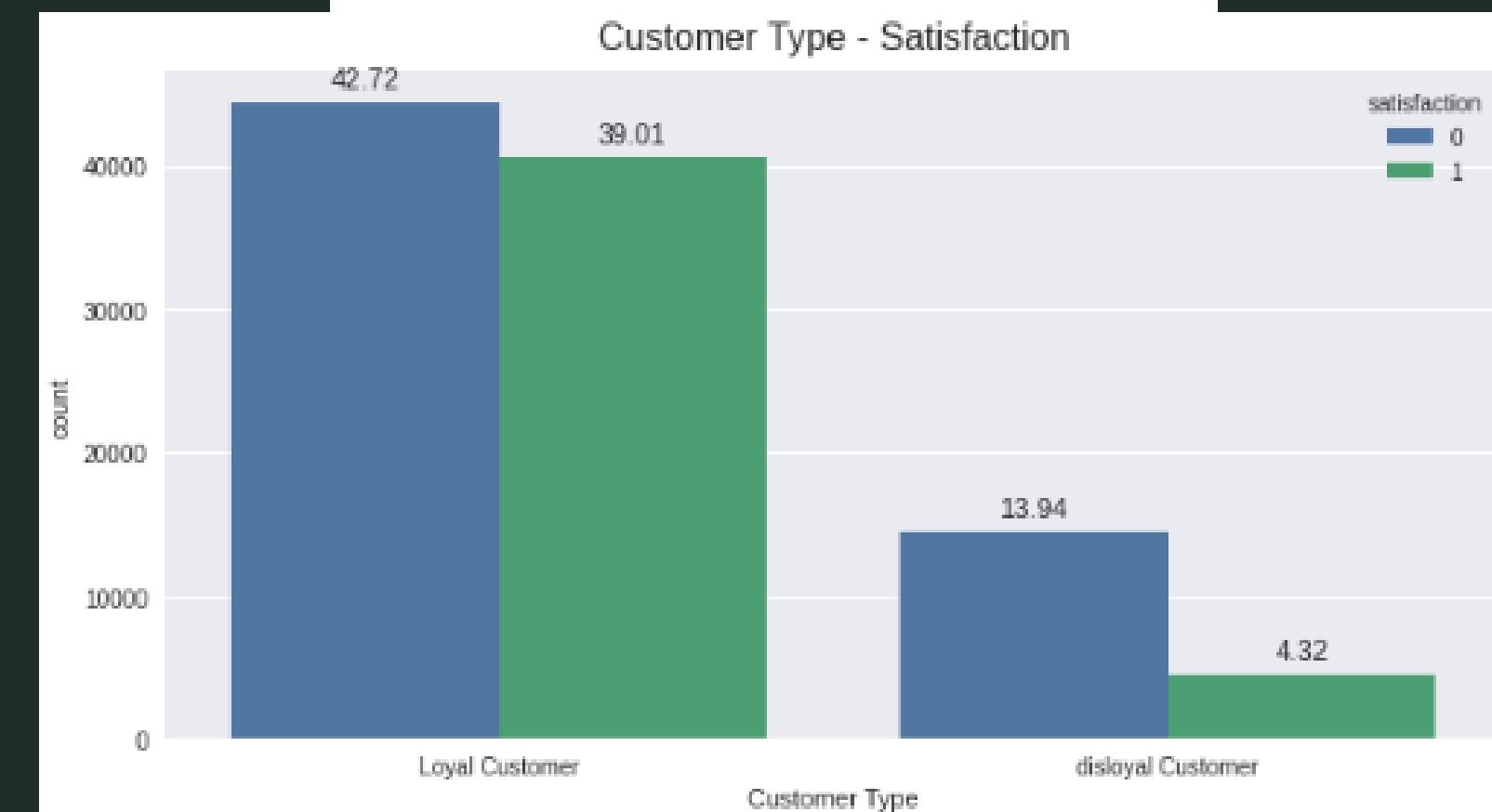
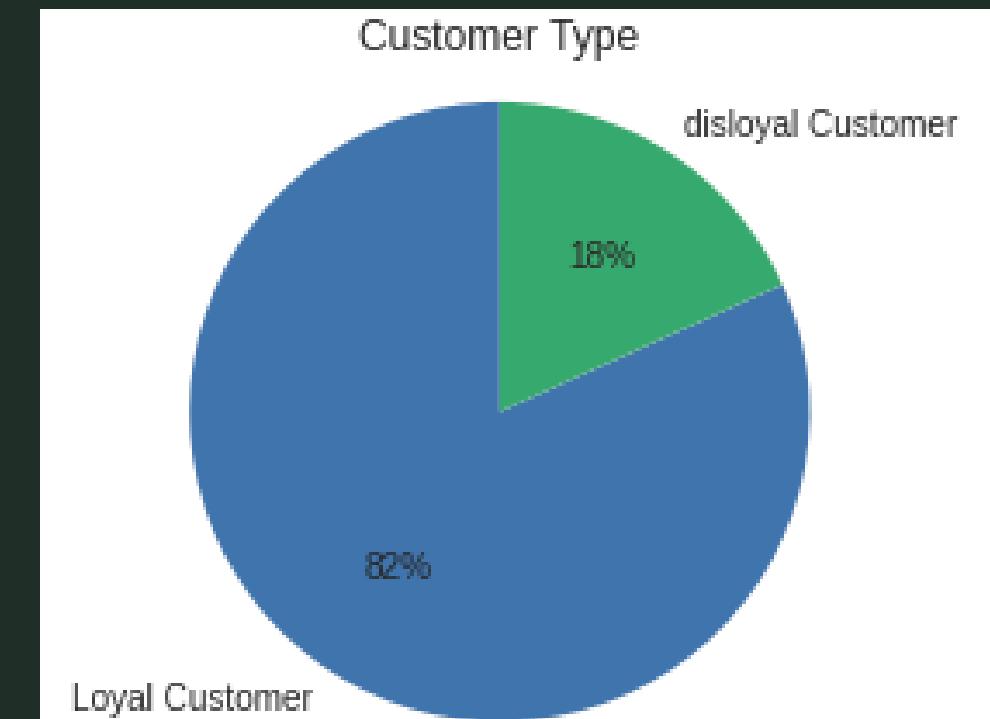
Though on the Eco Plus and Eco class  
have customer that have longer flight,  
they are outlier, which means that  
normally they wouldn't use the Eco Plus  
and Eco flight for longer flight.



# Exploratory Data Analysis

04.  
The number of loyal customer should  
be more than the disloyal

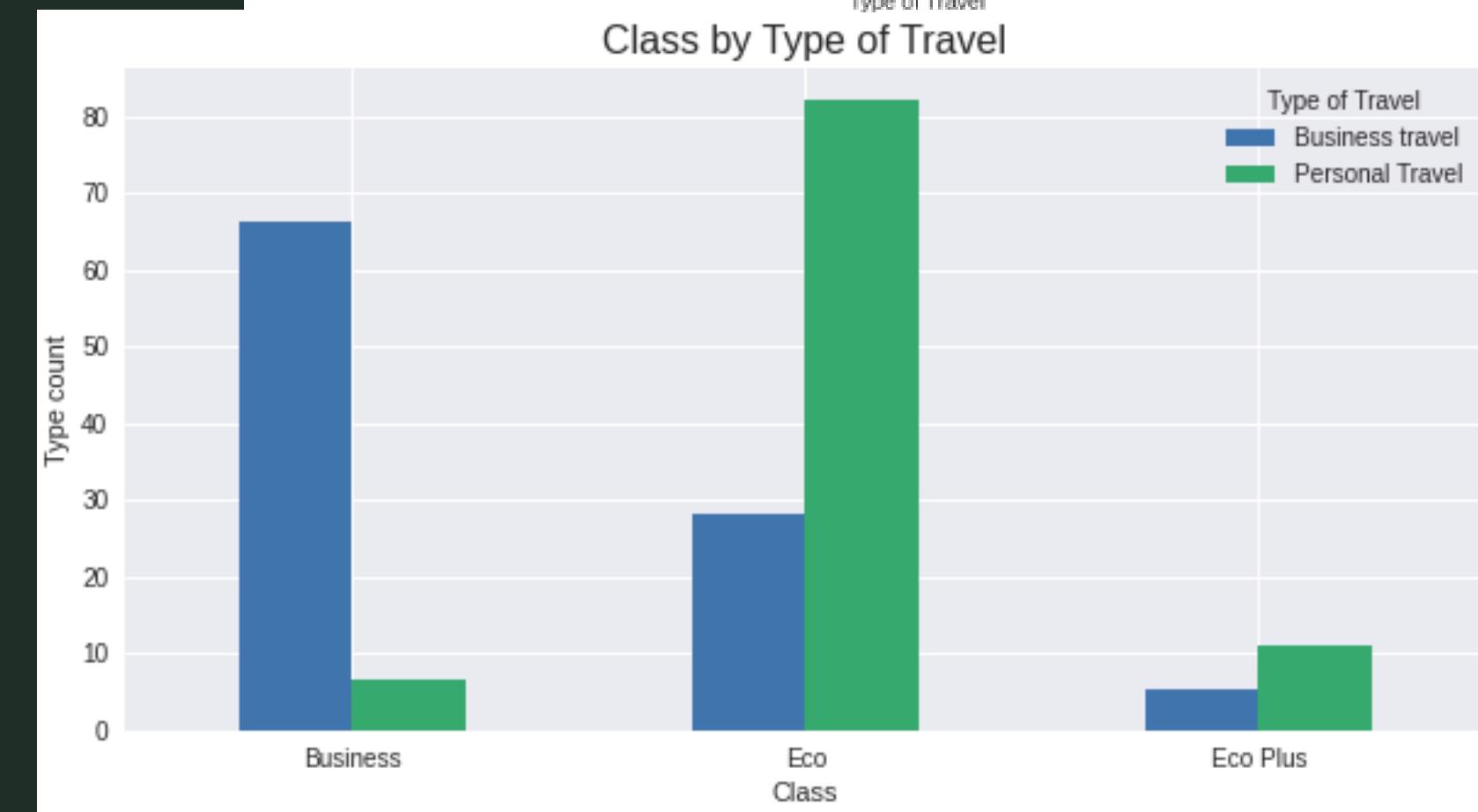
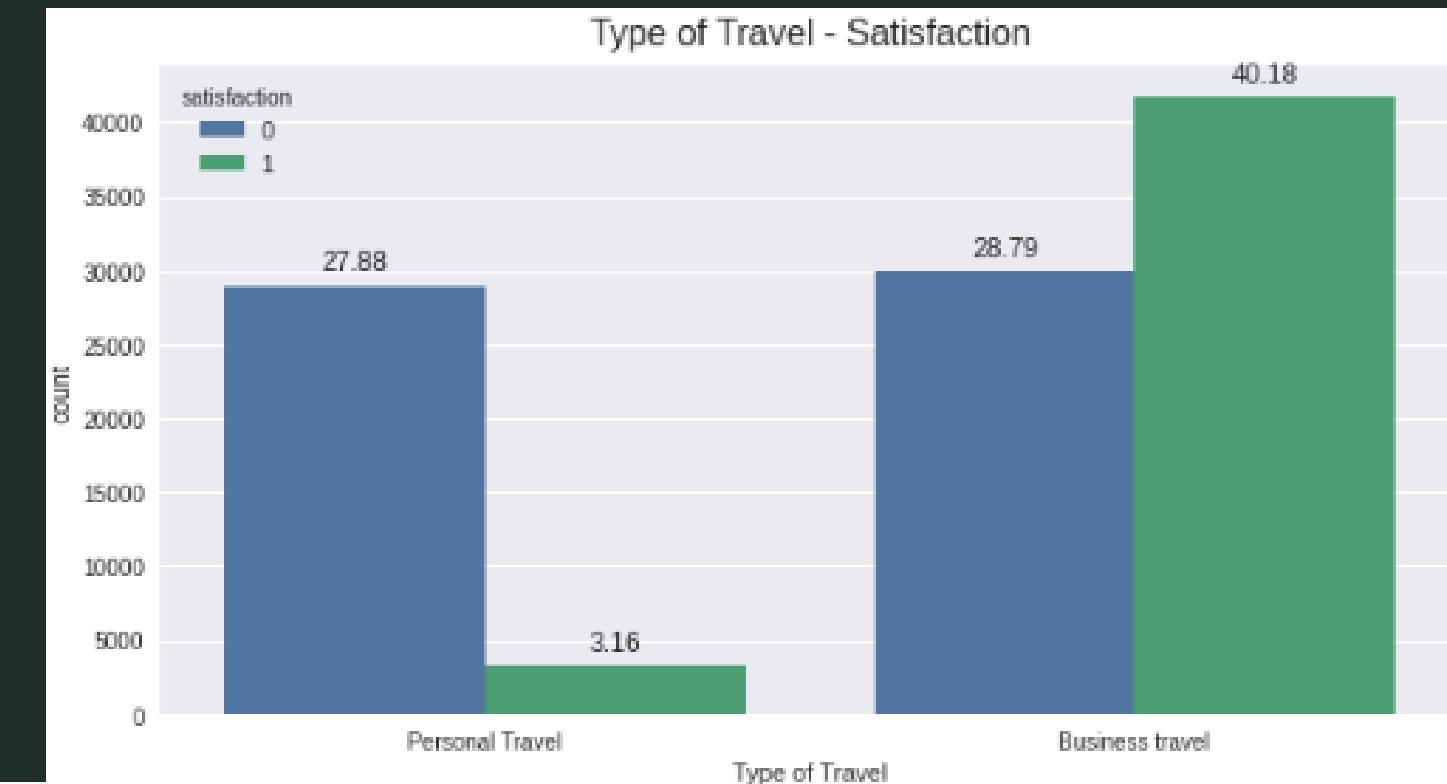
The assumption is true,  
as more customers are loyal customer and  
only 18% is disloyal.  
However if we see the satisfaction of  
both category, most of them are still  
neutral or dissatisfied with the Airline.



# Exploratory Data Analysis

05.  
People who travel for Business  
usually take Business Class

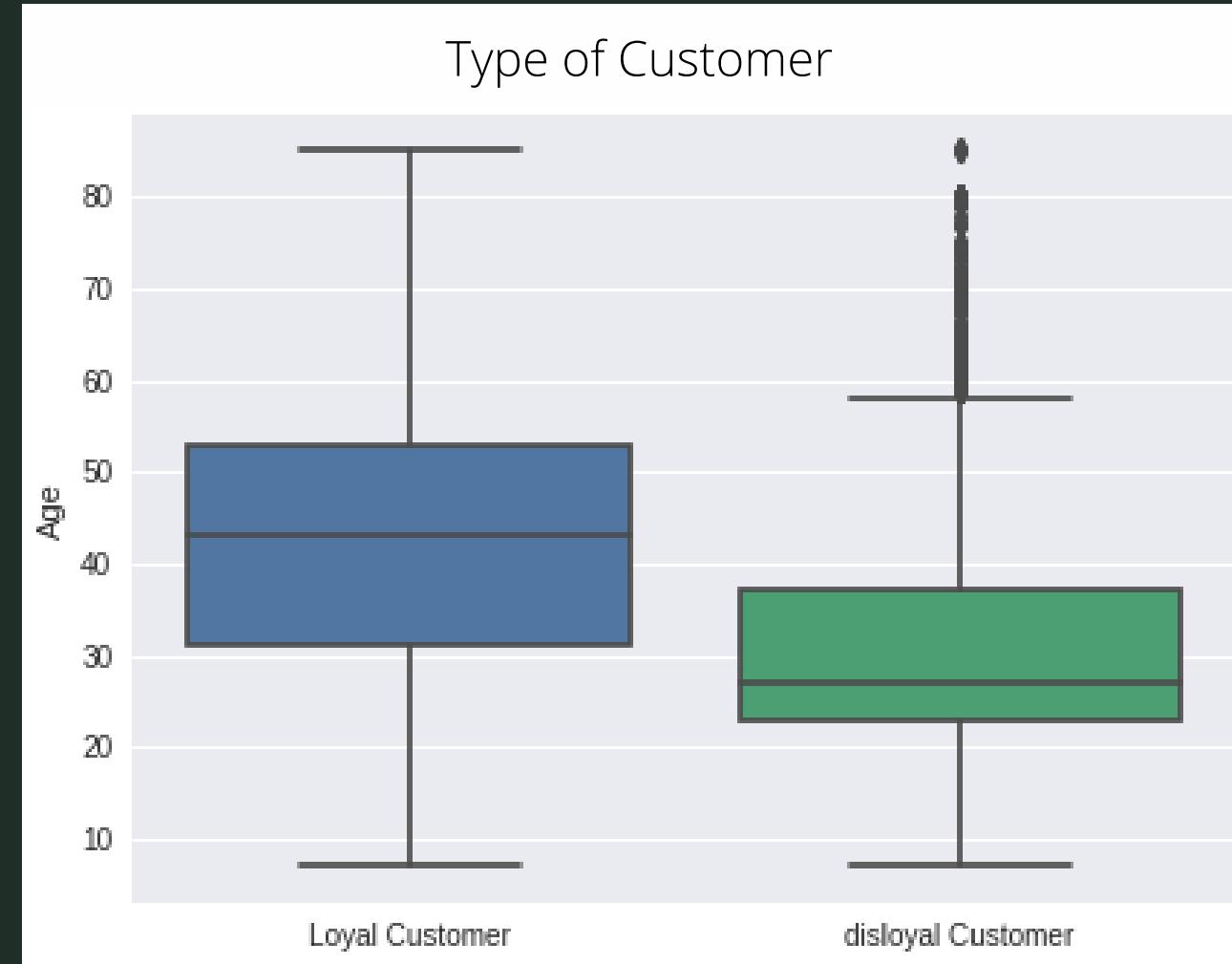
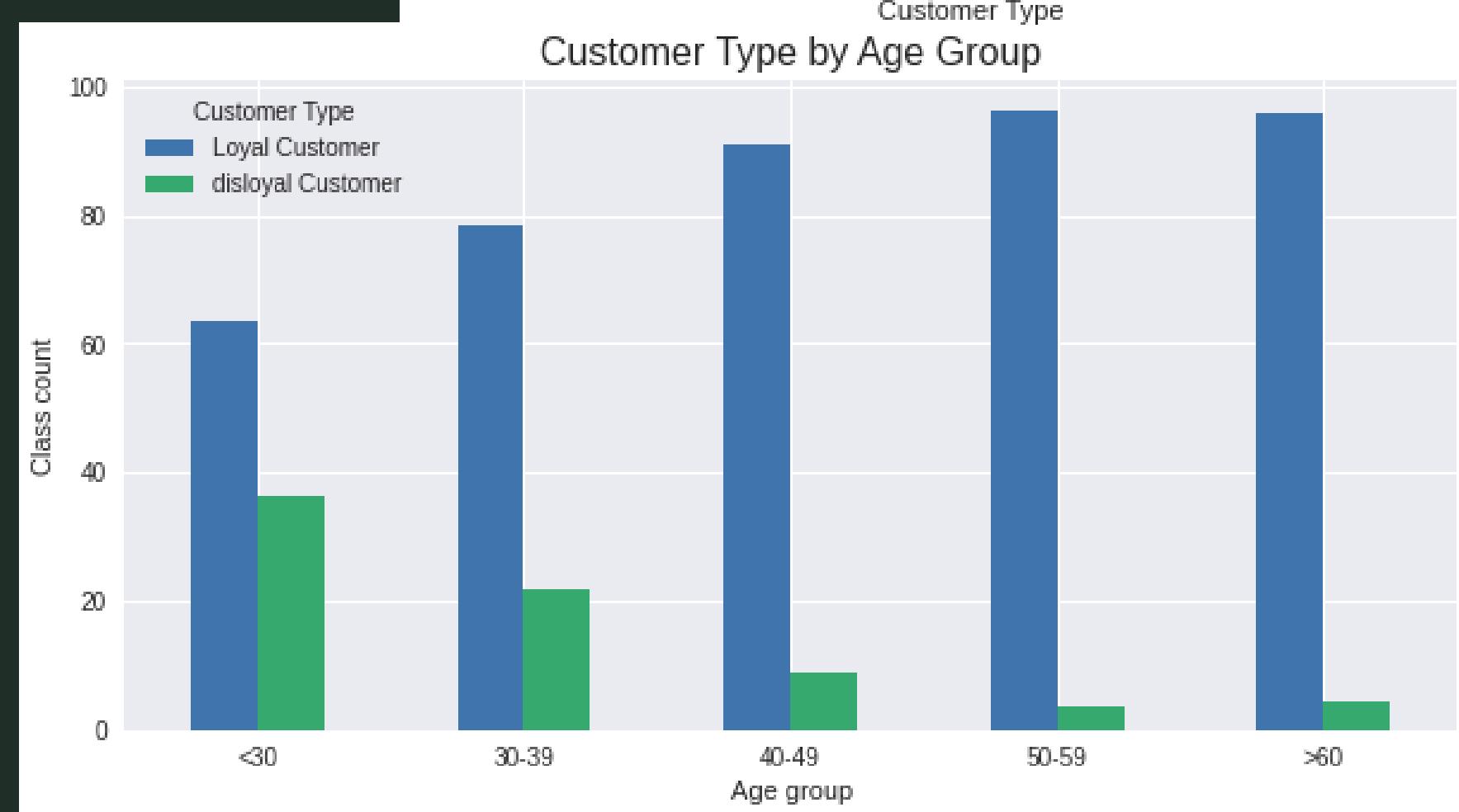
The assumption is true,  
as more customers are using the Airline  
service for Business Travel and most of  
them are using Business Class. While  
more passenger are using Eco and Eco  
Class for Personal Travel.



# Exploratory Data Analysis

06.  
Customer's loyalty should increase parallel with customer's age

The assumption is true, the older the customer they tend to be a loyal customer. The disloyal customer number also decreasing as the customer's age increases.

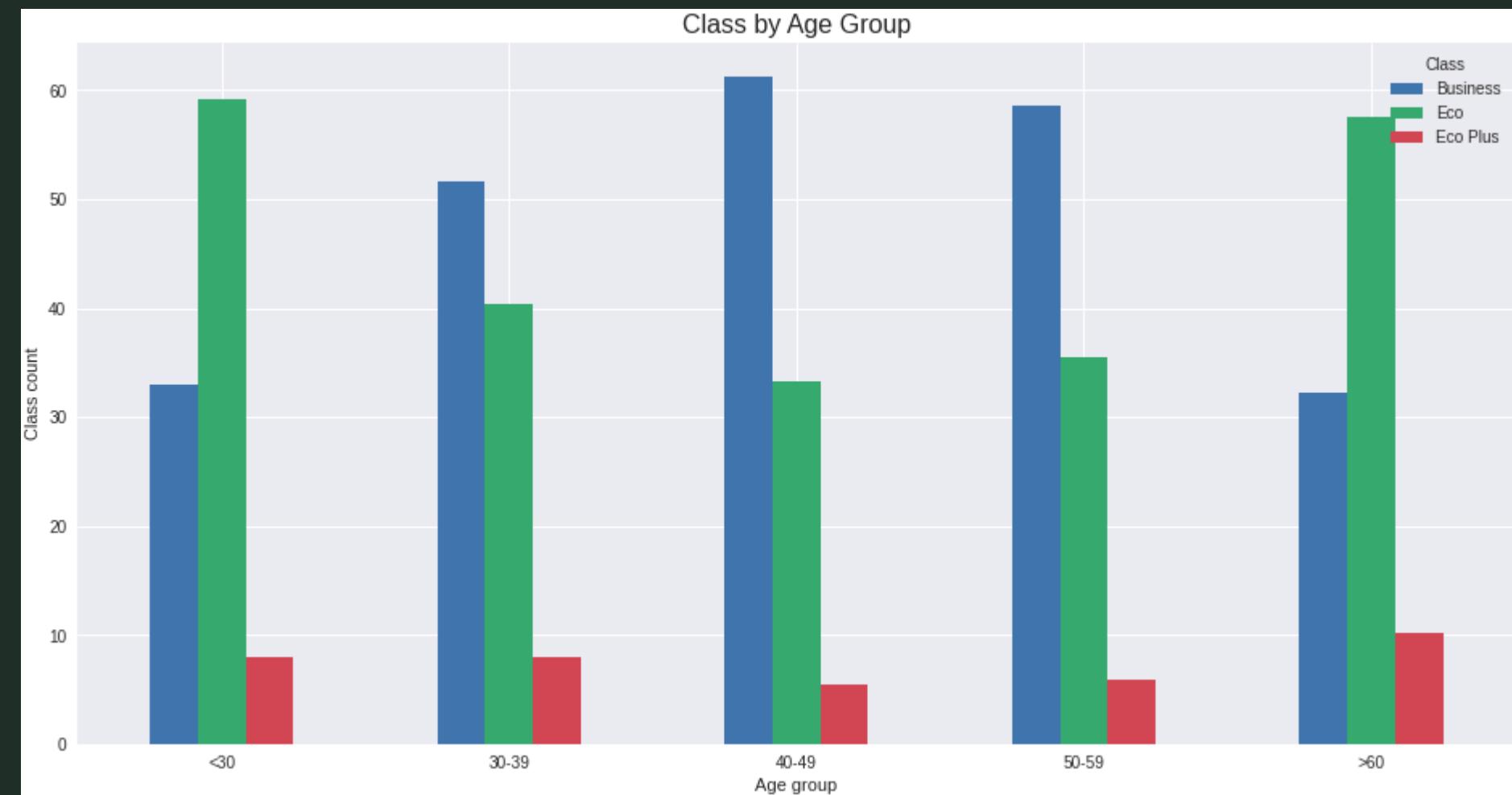


# Exploratory Data Analysis

07.

## Older customer prefer Business Class

The assumption is partially true, the older the customer, the more they use Business Class, however this only remains true until age group 50–59. On group aged >60 Eco Class placed higher than Business Class. Eco Plus constantly remains the lowest of all three classes.





# Modeling

## ● ○ ○ Preprocessing

Prepare the data

## ● ● ○ Benchmark Model

Find the best model

## ● ● ● Hyperparameter Tuning

Improve the model

```
state={  
  products: storeProducts  
}  
  
render() {  
  return (  
    <React.Fragment>  
      <div className="py-5">  
        <div className="container">  
          <Title name="our-products" />  
          <div className="row row-cols-1 row-cols-md-2 g-4">  
            <ProductCard product={products[0]} />  
            <ProductCard product={products[1]} />  
            <ProductCard product={products[2]} />  
            <ProductCard product={products[3]} />  
          </div>  
        </div>  
      </div>  
    </React.Fragment>  
  )  
}  
;
```



# Preprocessing

Data is cleaned and processed before used for modelling in order to get more accurate model.

## Drop Features

- 'Unnamed: 0' and 'id' (not relevant)
- 'Departure Delay in Minutes' and 'Arrival Delay in Minutes' (multicollinearity and combined into 'Total Delay' before dropping)

## Missing Values

'Arrival Delay in Minutes' with Iterative Imputer

## Outliers

Not dropped due to big information loss (+- 30%) and features are important

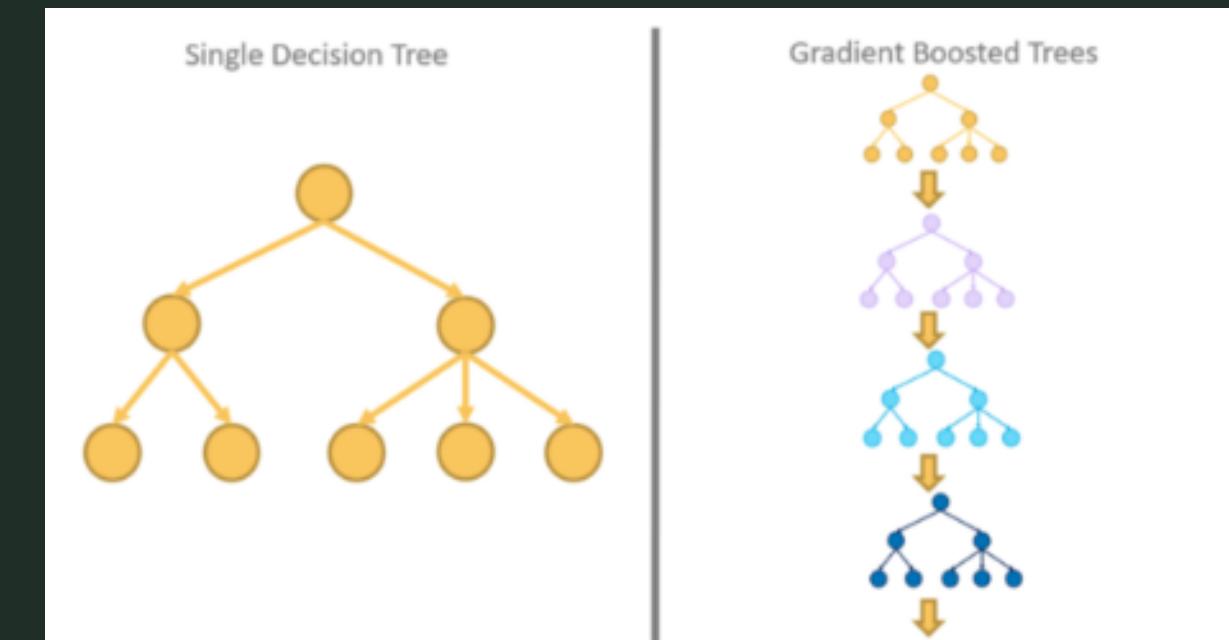
## Encoding

Categorical values are encoded using One Hot ('Gender', 'Customer Type', 'Type of Travel') and Ordinal ('Class') Encoder

# Benchmark Model

The best model to use is **CatBoostClassifier** that has the highest average of f1-score and a low average of f1-score standard deviation from cross validation.

**CatBoost is based on gradient boosted decision trees.** During training, a set of decision trees is built consecutively. Each successive tree is built with reduced loss compared to the previous trees.



model	f1-score mean	f1-score std
CatBoost	0.956894	0.001858
Random Forest	0.953560	0.001013
Decision Tree	0.935733	0.001807
XGBoost	0.928496	0.001771
AdaBoost	0.915353	0.002238
KNN	0.905829	0.002445
Logistic Regression	0.852833	0.002274

# Hyperparameter Tuning

To achieve best model, we have done hyperparameter tuning to get the best parameter with higher metric scores (f1 and ROC AUC).

**Best Parameter:**

```
{'model__learning_rate': 0.04,  
'model__l2_leaf_reg': 10,  
'model__iterations': 825,  
'model__depth': 8,  
'model__border_count': 176}
```

Metric	Before/After Tuning	Score
F1-Score Train	Before	0.956894
F1-Score Train	After	0.957916
F1-Score Test	Before	0.9579368207669351
F1-Score Test	After	0.9582581227436823
ROC AUC Test	Before	0.9952735723511094
ROC AUC Test	After	0.995314584062598

# Conclusion and Recommendation

## ● ○ ○ Conclusion

Predict possible revenue from model

## ● ● ○ Recommendation

Give insights on how to raise customer satisfaction based on model

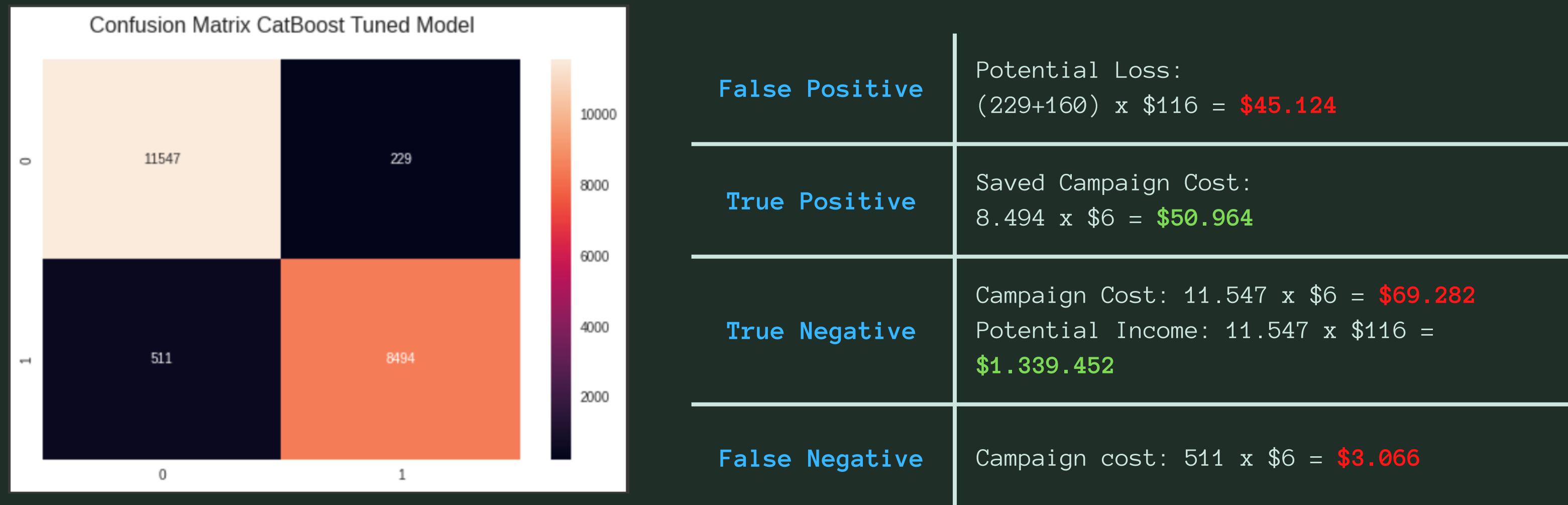
## ● ● ● Limitation

Show limitation of model



# Conclusion

Using confusion matrix, we can find the possible revenue and cost from 20.781 customers. We assume that the ticket fare per customer is \$116 and the campaign cost to improve satisfaction per customer is \$6.



The final model result shows that this model will give the company a possible income of **\$1.273.034**

# Conclusion

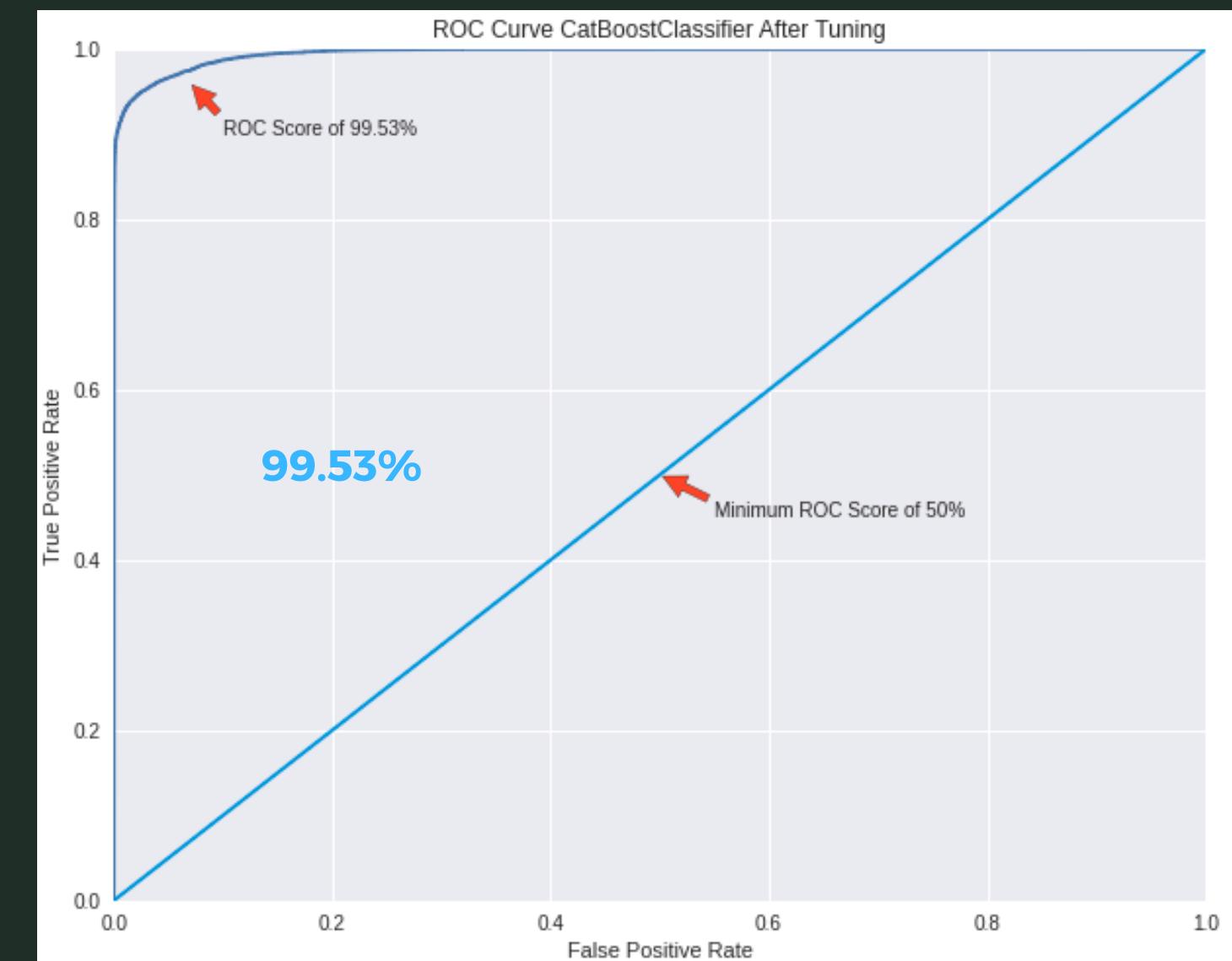
**Accuracy**

Correctly predict **96%** the satisfied and neutral or dissatisfied customer

**ROC AUC**

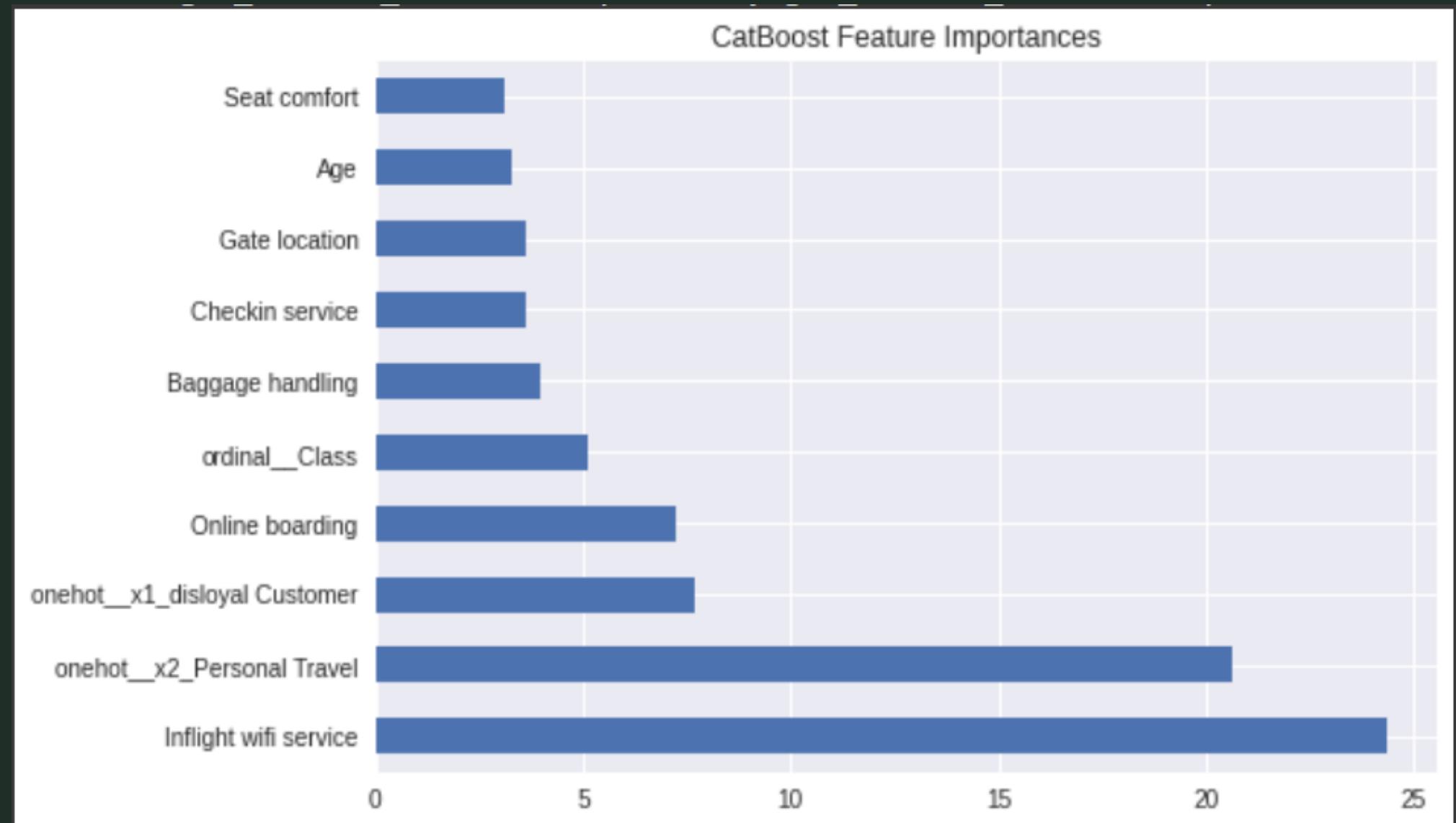
This model can almost completely (**99.53%**) distinguish between satisfied customers and those who are neutral or dissatisfied

Classification Report Tuned CatBoost :				
	precision	recall	f1-score	support
0	0.96	0.98	0.97	11776
1	0.97	0.94	0.96	9005
accuracy			0.96	20781
macro avg	0.97	0.96	0.96	20781
weighted avg	0.96	0.96	0.96	20781



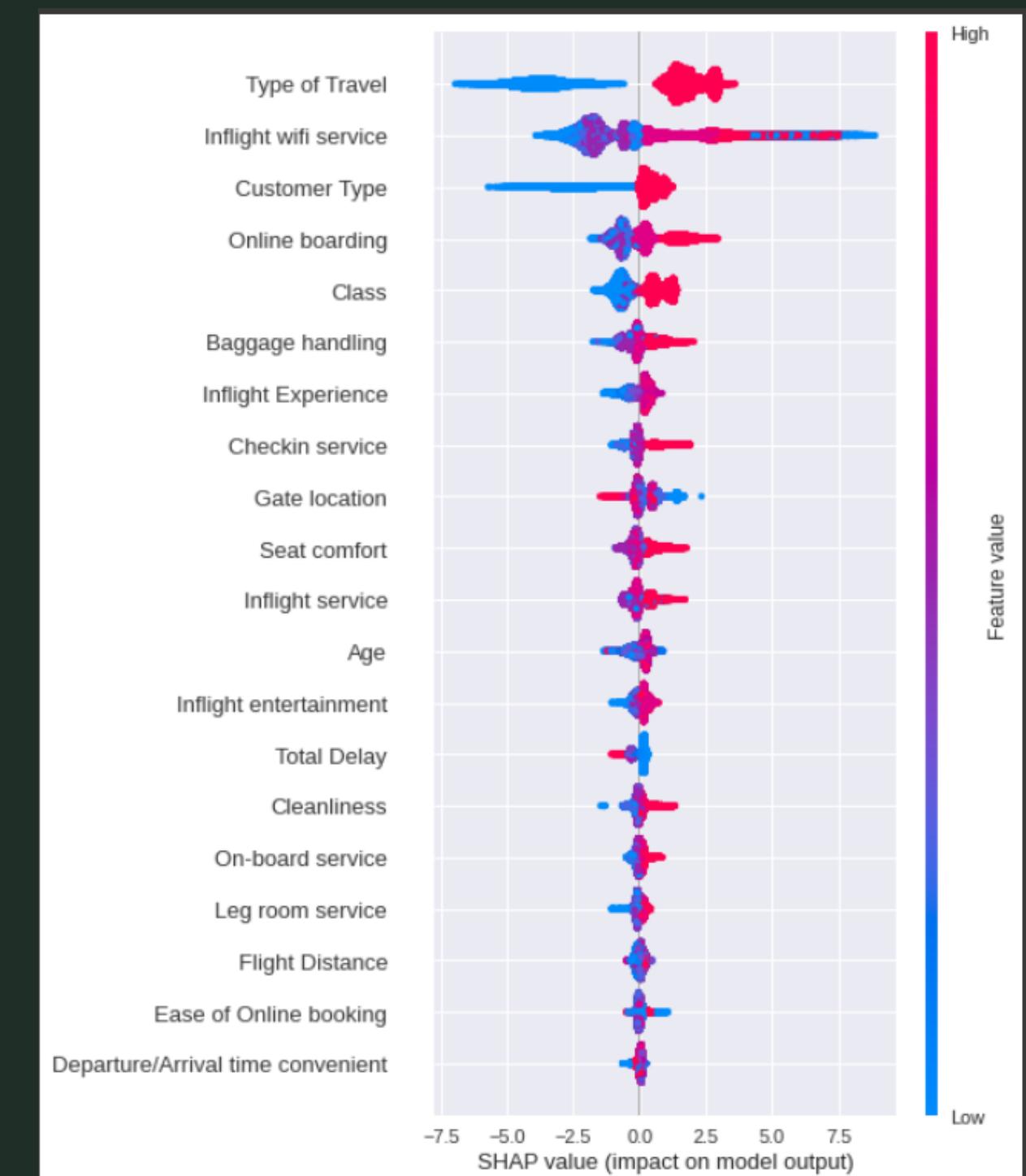
# Recommendation

From the feature importance chart we understand that the feature that have most impact to our model is **Inflight wifi service**, followed by **Type of Travel** (Personal Travel and Business Travel). To further understand this we will also look at the **SHAP values**.



# Recommendation

Type of Travel	Business Travel impact the target <b>positively</b> (1:Personal, 2:Business)
Customer Type	Loyal customers impact the target <b>positively</b> (1:disloyal, 2:loyal)
Class	Business Class impact the target <b>positively</b> (1:Eco, 2:Eco Plus, 3:Business)
Overall	Overall the <b>higher the satisfaction</b> on each features, will impact <b>positively</b> on our target. However this is not applicable on <b>Total Delay, Gate Location and Ease of Online Booking</b>



# Recommendation

---

---

01.

## Inflight wifi service

- Offer affordable inflight wifi service on online booking
  - Free wifi with limited time for Eco and Eco Plus customers that have fulfilled a certain condition
- 

---

02.

## Type of Travel

- 69% of our customers travel for business
- Develop a B2B campaign to upgrade from Eco/Eco Plus to Business Class

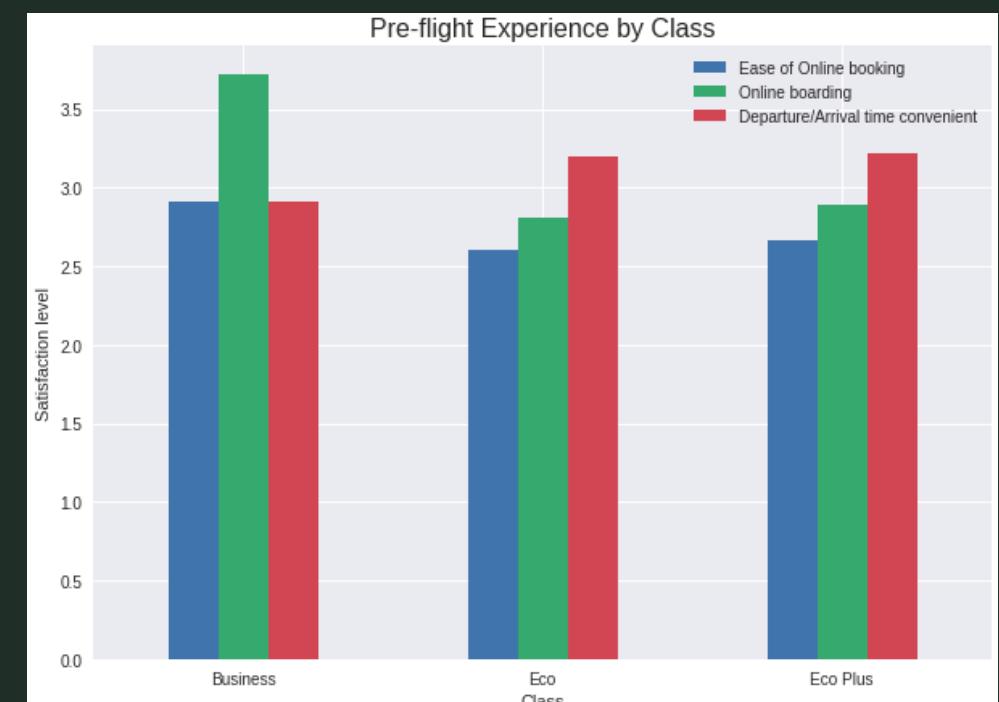
# Recommendation

## 03. Loyalty

- Loyalty placed 3rd on our feature importance chart
- Mileage point bonus can push our customer to use our service more often
- Variations in mileage point usage (hotels, airport lounge, etc.)

## 04. Online Boarding

- Improve online boarding service on Eco and Eco Plus as this feature also place among the top on feature importance



# Recommendation

## 05. Class Prioritization

- The number of customer on Business Class and Eco class are similar
- Though the number of dissatisfied customers are higher on Eco Class, we recommend to prioritize the satisfaction on Business class as it is less customer but more higher in value

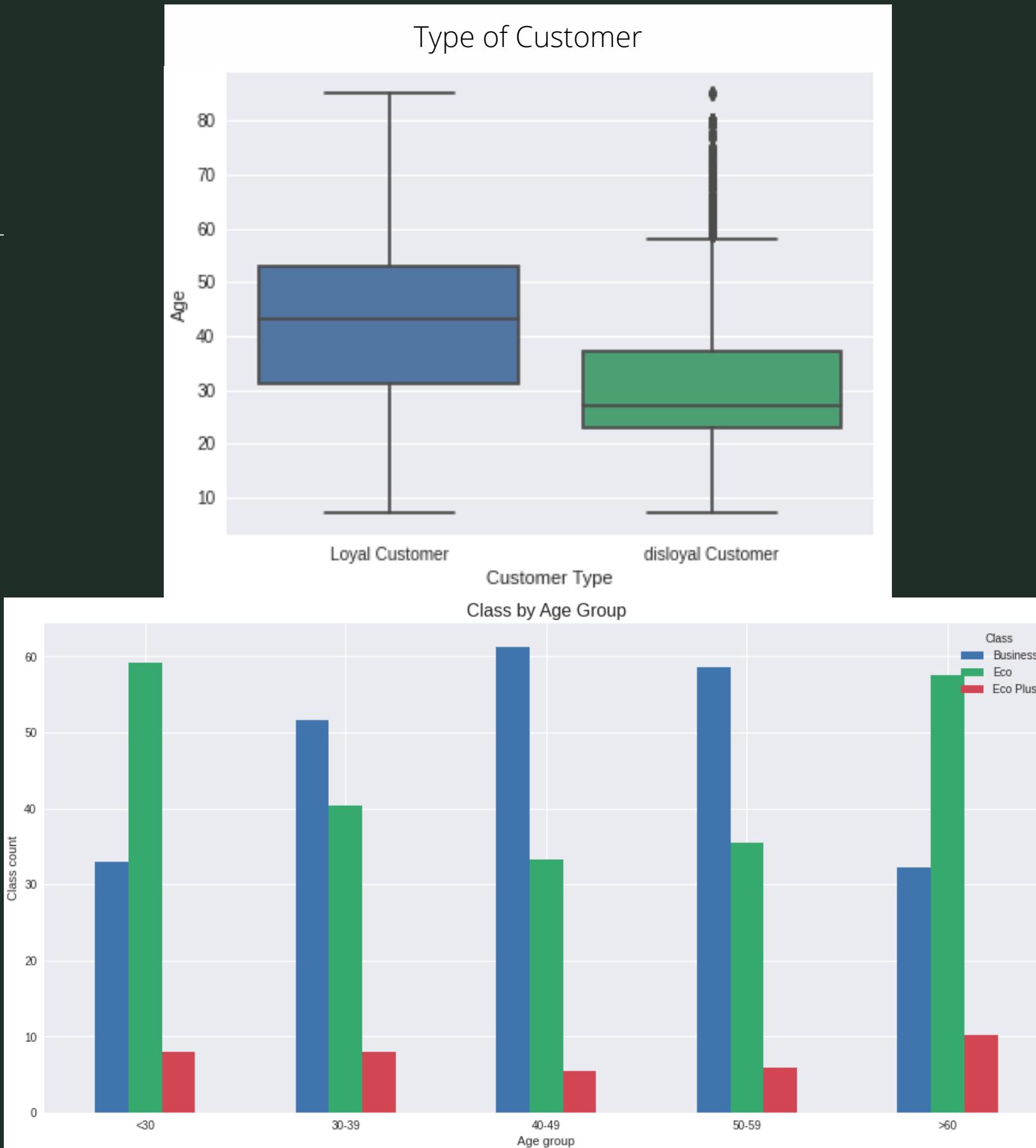


# Recommendation

06.

Age

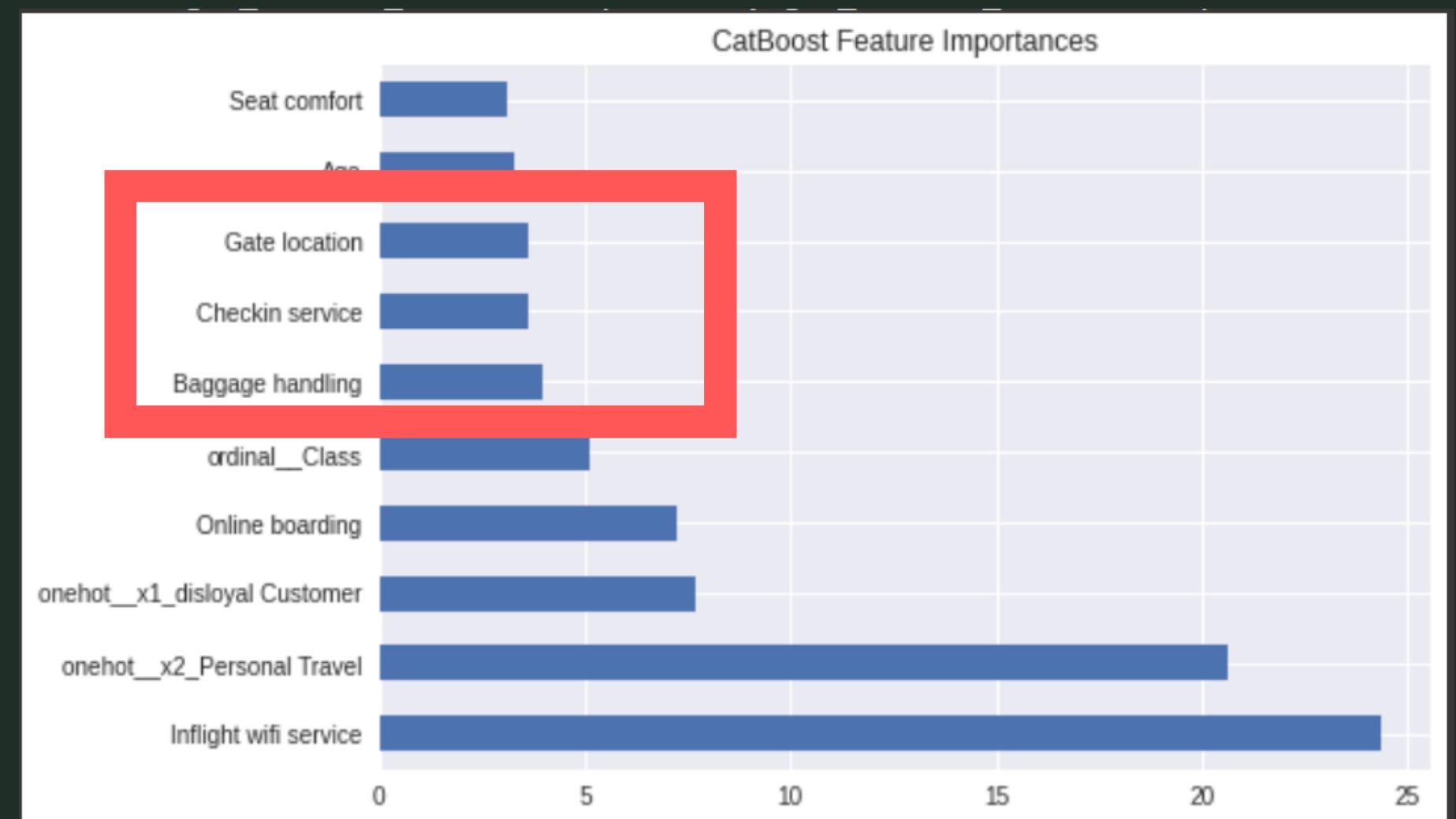
- For younger customer, focused on increasing loyalty by giving mileage point bonus
- For elder customer, focused on giving senior discount or special price on class upgrade



# Recommendation

07.

Focused on improving Airport Experience as there are 3 features that have high importance



# Recommendation

---

---

## Recommendation for project and model improvements:

- Try other Machine Learning algorithm, and add parameters on hyperparameter tuning
- Use gridsearch for hyperparameter tuning
- Commence another survey regarding features that significantly influence the model  
(Online Boarding & Inflight Wifi Service)

# Limitation

---

This model works best with the following limitations:

- Age : 7–85 years
- Flight Distance : 31–4.983 miles
- Departure/Arrival Delay : 38–1.592 minutes



---

# Thank You

---