AIRLINE PASSENGER SATISFACTION

TEAM MEMBERS

- Akshaya ES
- Nikita Ann George
- Pala Sheshu
- Ujjwal Kumar Sharma
- Venkat Paritala
- Viresh Raj Sah

MENTOR:

Mr Srikar Muppidi

GREAT LAKES
INSTITUTE OF MANAGEMENT, CHENNAI

greatlearning
Learning for Life

PROBLEM STATEMENT

As the old adage goes:" Customers are king."
This is an age-old mantra reflection the importance of customers in every business. With the recent pandemic completely disrupting the way we travel and having a huge impact on the aviation industry, one must now ponder on how we can get back to the norm

What can airlines prepare to give themselves a competitive edge when the crowd finally arrives?
We believe that to recover from the disastrous impact the pandemic had on the industry, we need to find a way to engage passengers in a better manner, by enhancing their experience and ultimately gaining their loyalty .

With this project we aim to do just that by creating a highly precise classification model for airlines to identify critical bottleneck to raise passenger satisfaction.

DATA DESCRIPTION

- 129880 observations
- 25 features

VARIABLE	DESCRIPTION
1.Unnamed:0	Index value
2. Id	Unique identifier
3. Gender	Gender of the passengers (Female, Male)
4. Customer Type	The customer type (Loyal customer, disloyal customer)
5. Age	The actual age of the passengers
6. Type of Travel	Purpose of the flight of the passengers (Personal Travel, Business Travel)
7. Class	Travel class in the plane of the passengers (Business, Eco, Eco Plus)
8. Flight distance	The flight distance of this journey
9. Inflight wifi service	Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
10. Departure/Arrival time convenient	Satisfaction level of Departure/Arrival time convenient
11. Ease of Online booking	Satisfaction level of online booking
12. Gate location	Satisfaction level of Gate location
13. Food and drink	Satisfaction level of Food and drink

VARIABLE	DESCRIPTION
14. Online boarding	Satisfaction level of online boarding
15. Seat comfort	Satisfaction level of Seat comfort
16. Inflight entertainment	Satisfaction level of inflight entertainment
17. On-board service	Satisfaction level of On-board service
18. Leg room service	Satisfaction level of Leg room service
19. Baggage handling	Satisfaction level of baggage handling
20. Check-in service	Satisfaction level of Check-in service
21. Inflight service	Satisfaction level of inflight service
22. Cleanliness	Satisfaction level of Cleanliness
23. Departure Delay in Minutes	Minutes delayed when departure
24. Arrival Delay in Minutes	Minutes delayed when Arrival
25. Satisfaction	Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

DATA PREPROCESSING

1. Treatment of Null values:

We found null values in the feature: Arrival Delay in Minutes.The percentage of null values was around 0.3%,.We decided to drop these values.

2.Outlier treatment:

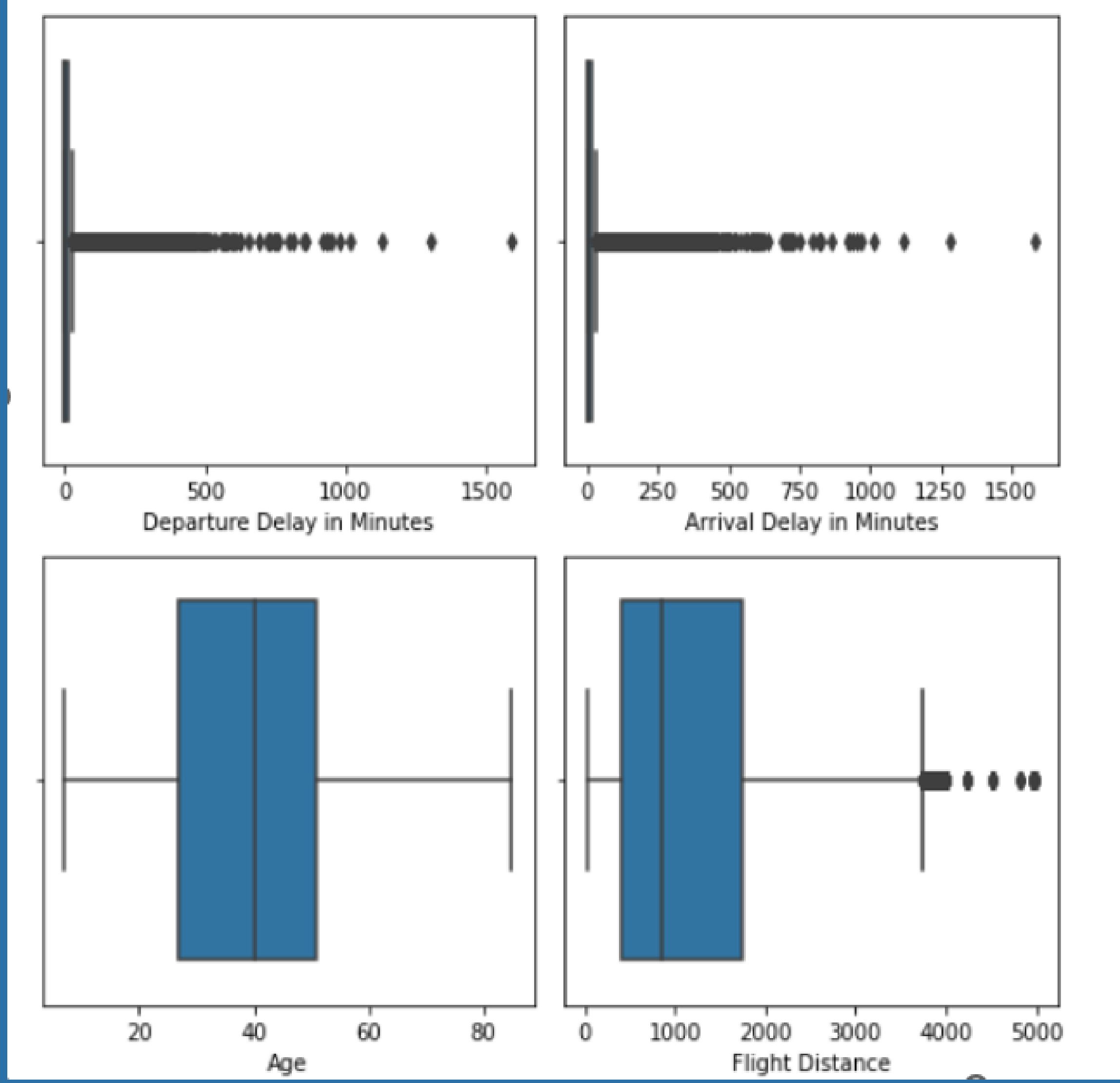
There are outliers present in the Flight Distance, Departure Delay in Minutes and Arrival Delay in Minutes column. As these variables play a role in customer satisfaction (more the delay,we can assume less satisfied the customer), we are not excluding the outliers . As we proceed with model building we can transform the extreme outliers. But for the base model we choose to leave the outliers as it is.

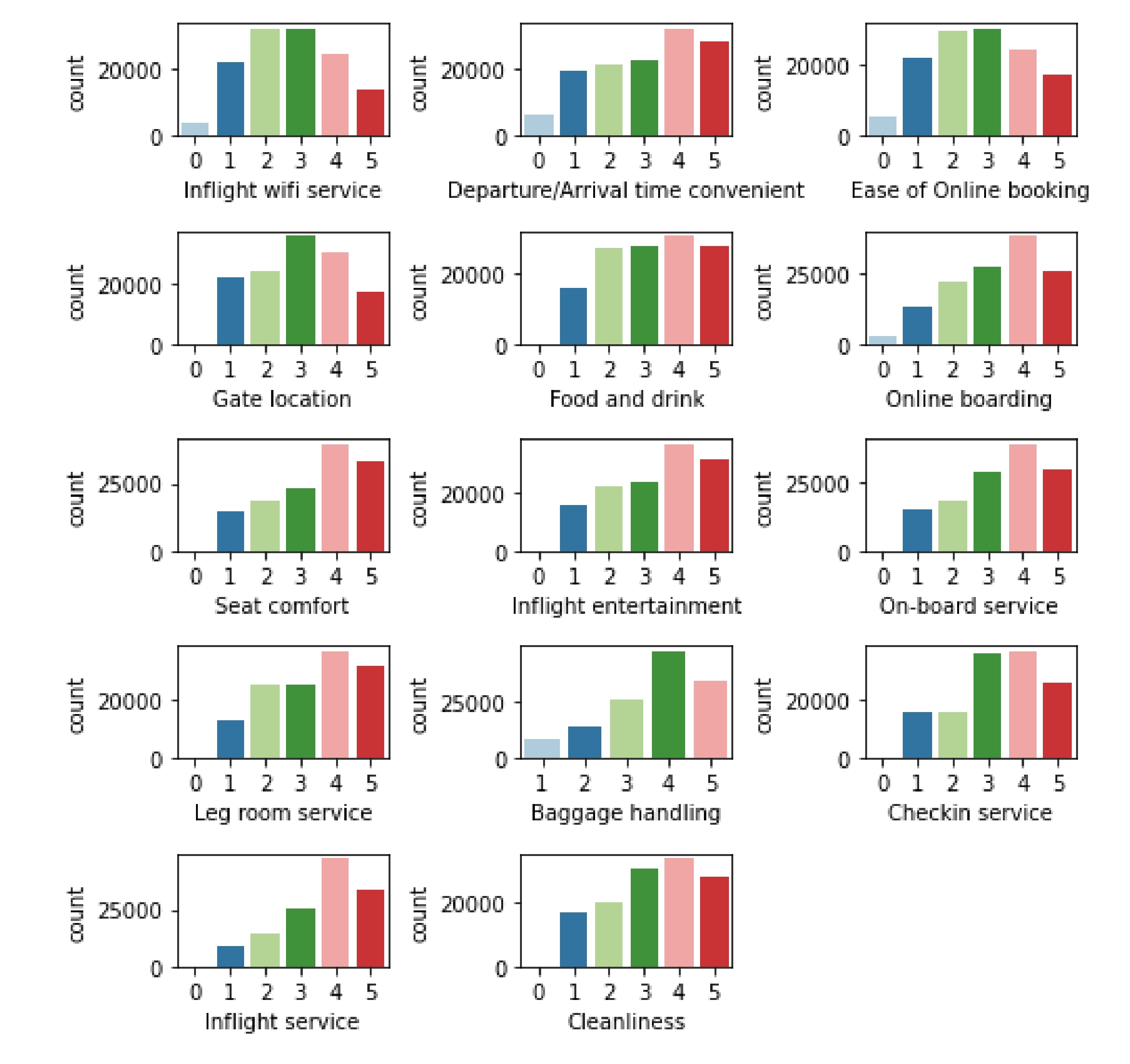
3.**Insignificant columns:**

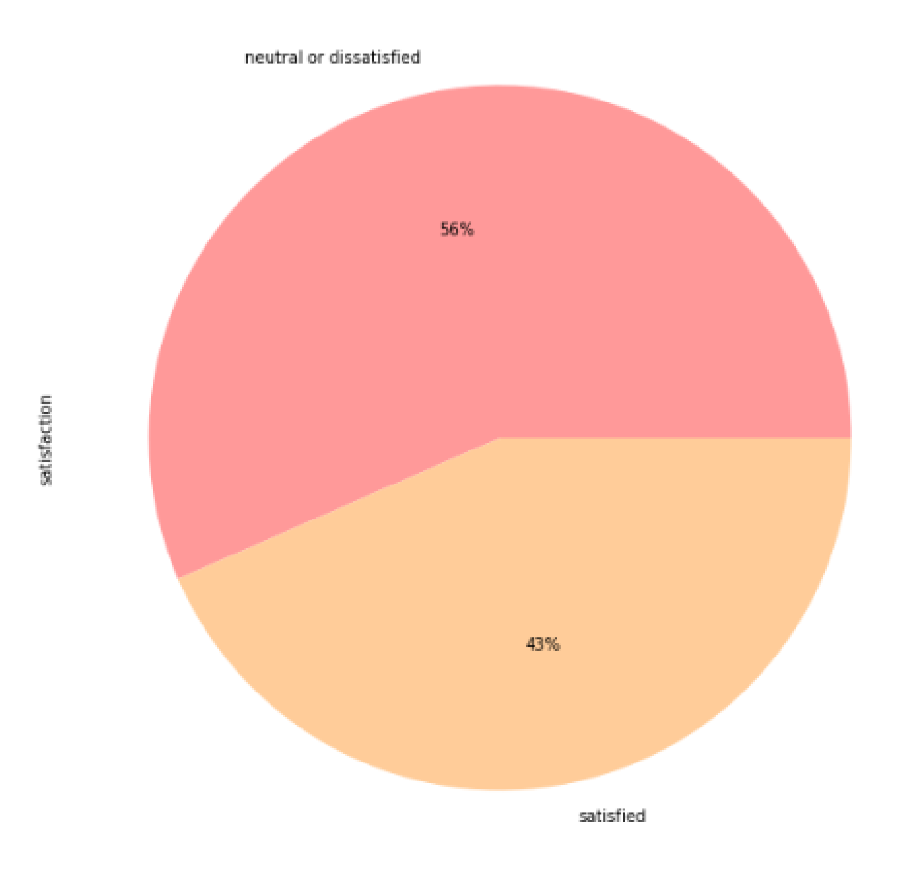
Unnamed: 0- Index value

Id- Unique identifiers of each row, has no impact on the model

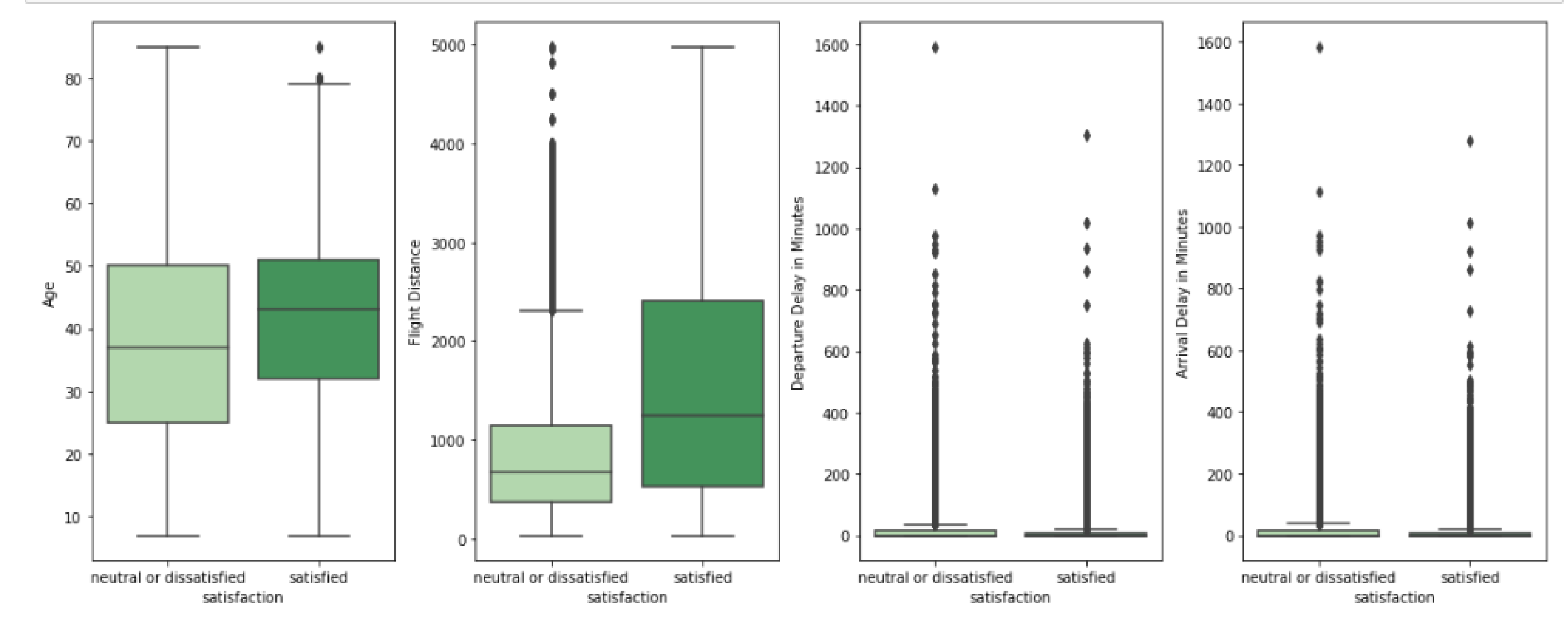
UNIVARIATE ANALYSIS



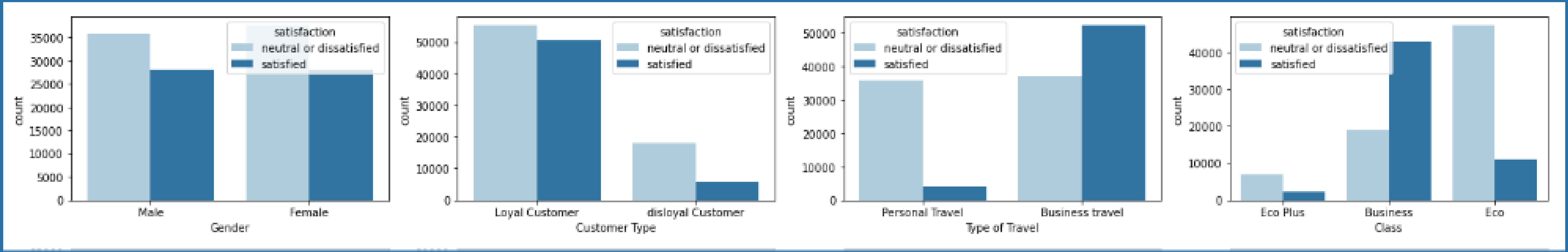




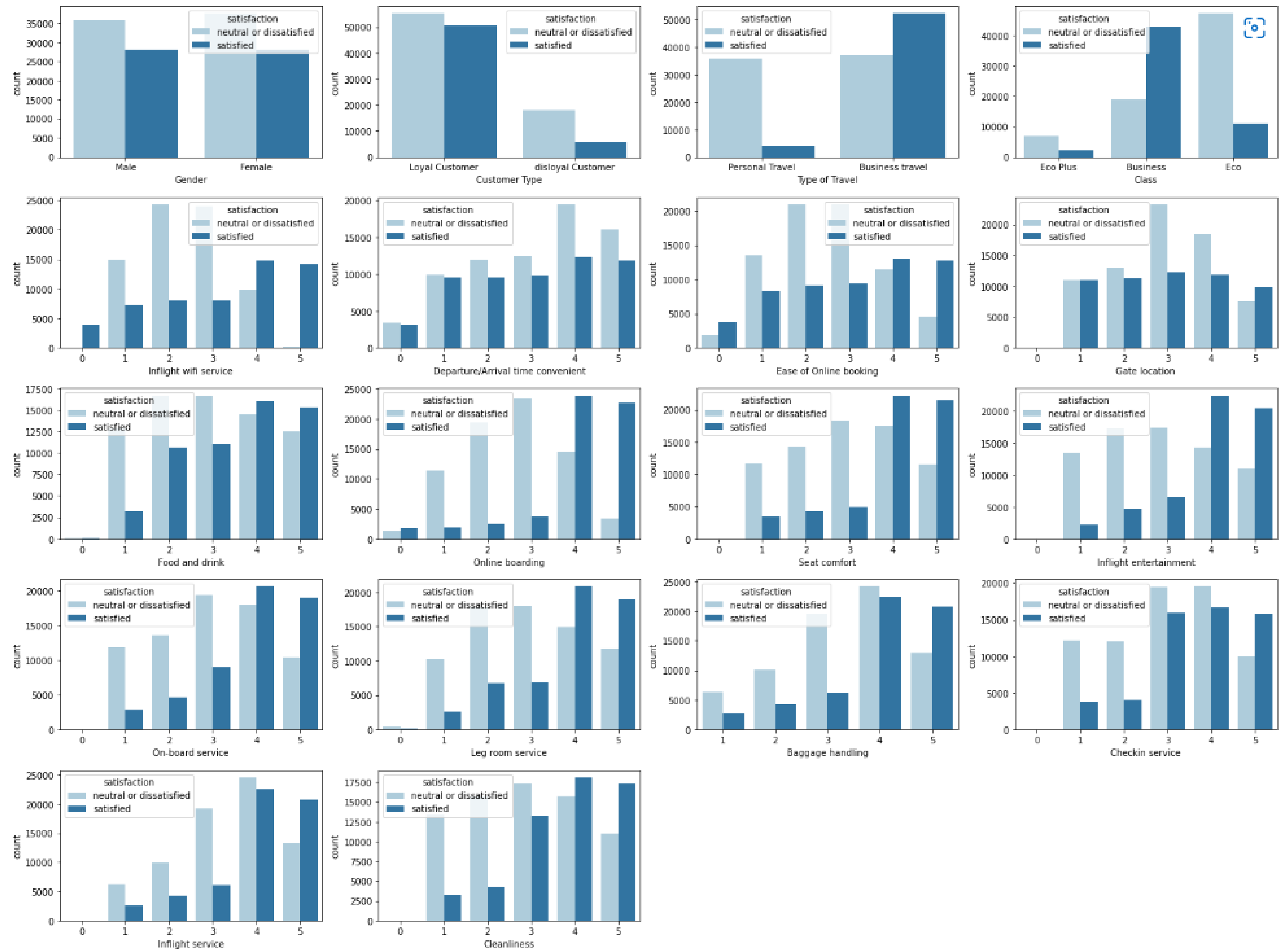
BIVARIATE ANALYSIS



- Age :The average age for satisfied and dissatisfied were found to be 37 and 41 respectively
- Flight Distance: The average distance travelled by satisfied customers was higher than that of neutral/dissatisfied



- Disloyal customers were found to be more dissatisfied than satisfied
- Business travelers' are highly satisfied
- Moreover passengers business class get high satisfied count, while passengers in the eco and eco plus class were found to be more dissatisfied


satisfaction
neutral or dissatisfied
satisfied
count
Male
Female
Gender
Loyal Customer
disloyal Customer
Customer Type
Personal Travel
Business travel
Type of Travel
Eco Plus
Business
Eco
Class
Inflight wifi service
Departure/Arrival time convenient
Ease of Online booking
Gate location
Food and drink
Online boarding
Seat comfort
Inflight entertainment
On-board service
Leg room service
Baggage handling
Checkin service
Inflight service
Cleanliness

BIVARIATE ANALYSIS

- The Main Causes of dissatisfaction are: In-Flight Wifi, Online Booking, Inflight Entertainment and Online boarding
- Less But Observable Dissatisfaction is caused by Leg-Room Services and Cleanliness
- Some Dissatisfaction is also caused by Gate Location, Food and Drinks, Seat Comfort, On Board Services
- Online Boarding Shows generally satisfactory ratings, but is one of the main cause of dissatisfaction among the dissatisfied customers

STATISTICAL TEST

Chi2_Contigency

	p_value
Gender	5.135590e-05
Customer Type	0.000000e+00
Type of Travel	0.000000e+00
Class	0.000000e+00
Inflight wifi service	0.000000e+00
Departure/Arrival time convenient	9.767302e-128
Ease of Online booking	0.000000e+00
Gate location	0.000000e+00
Food and drink	0.000000e+00
Online boarding	0.000000e+00
Seat comfort	0.000000e+00
Inflight entertainment	0.000000e+00
On-board service	0.000000e+00
Leg room service	0.000000e+00
Baggage handling	0.000000e+00
Checkin service	0.000000e+00
Inflight service	0.000000e+00
Cleanliness	0.000000e+00

Skewness

	Skewness
Age	-0.003376
Flight Distance	1.108433
Departure Delay in Minutes	6.853578
Arrival Delay in Minutes	6.670125

Kruskal Wallis test

	p_value
Age	0.000000e+00
Flight Distance	0.000000e+00
Departure Delay in Minutes	3.580163e-137
Arrival Delay in Minutes	7.087659e-287

MODELLING

- We created several models to improve upon the various metrics that help establish the most efficient model.
- From our business perspective we looked at the best way to reduce both types of false predictions and hence we used "Accuracy" and "F1-score" as our Target metrics
- We finally created a scorecard which stores all the metrics for each model , which helps us to analyze the performance of various models.

HYPER-PARAMETER TUNING

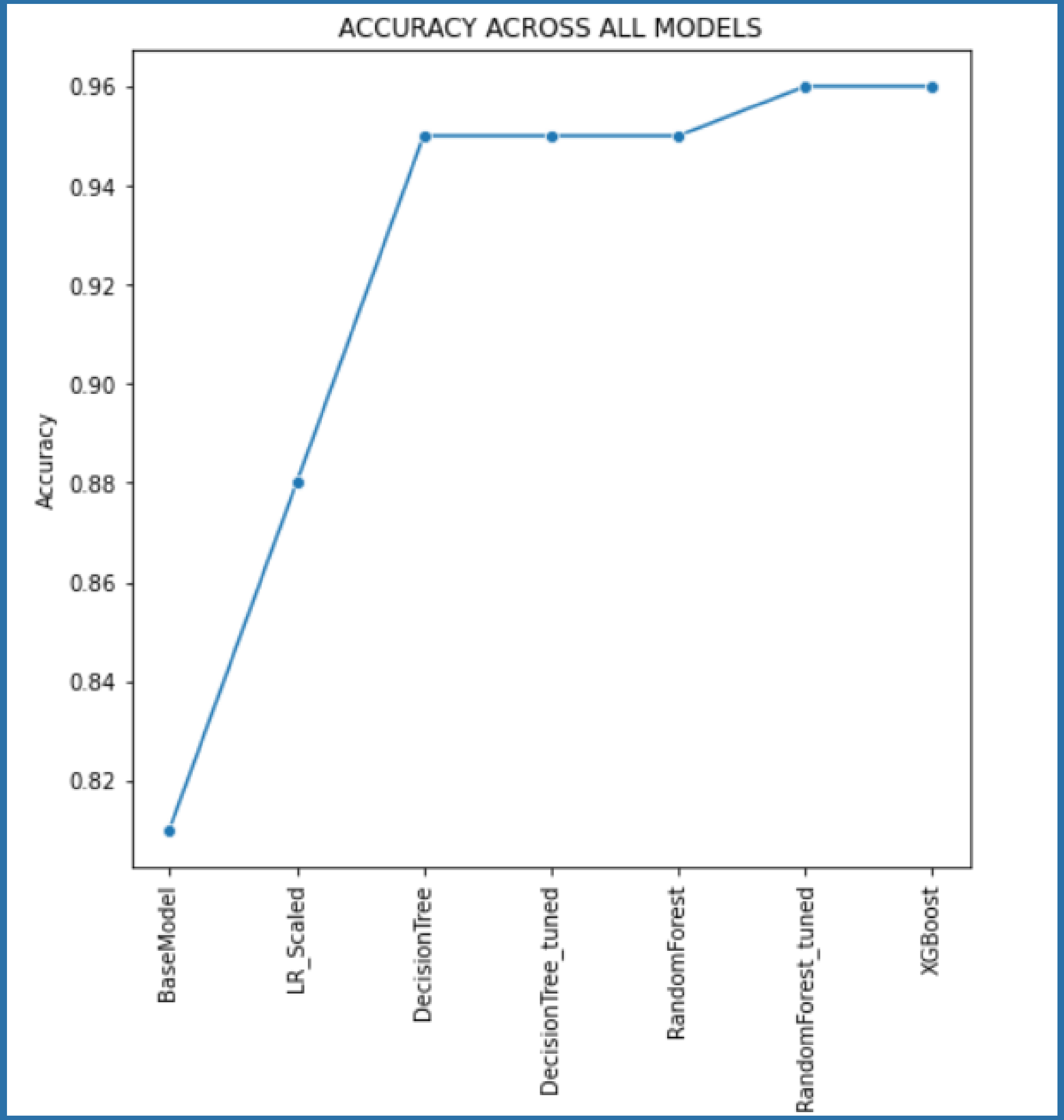
- For the Decision Tree and Random Forest models we performed hyper parameter tuning using GridSearchCV to get better results.
- We tuned the hyperparameters: criterion,max_depth, min_samples_split and n_estimators(RandomForest) for the two models.We saw a slight improvement in overall metrics on tuning the RandomForest model

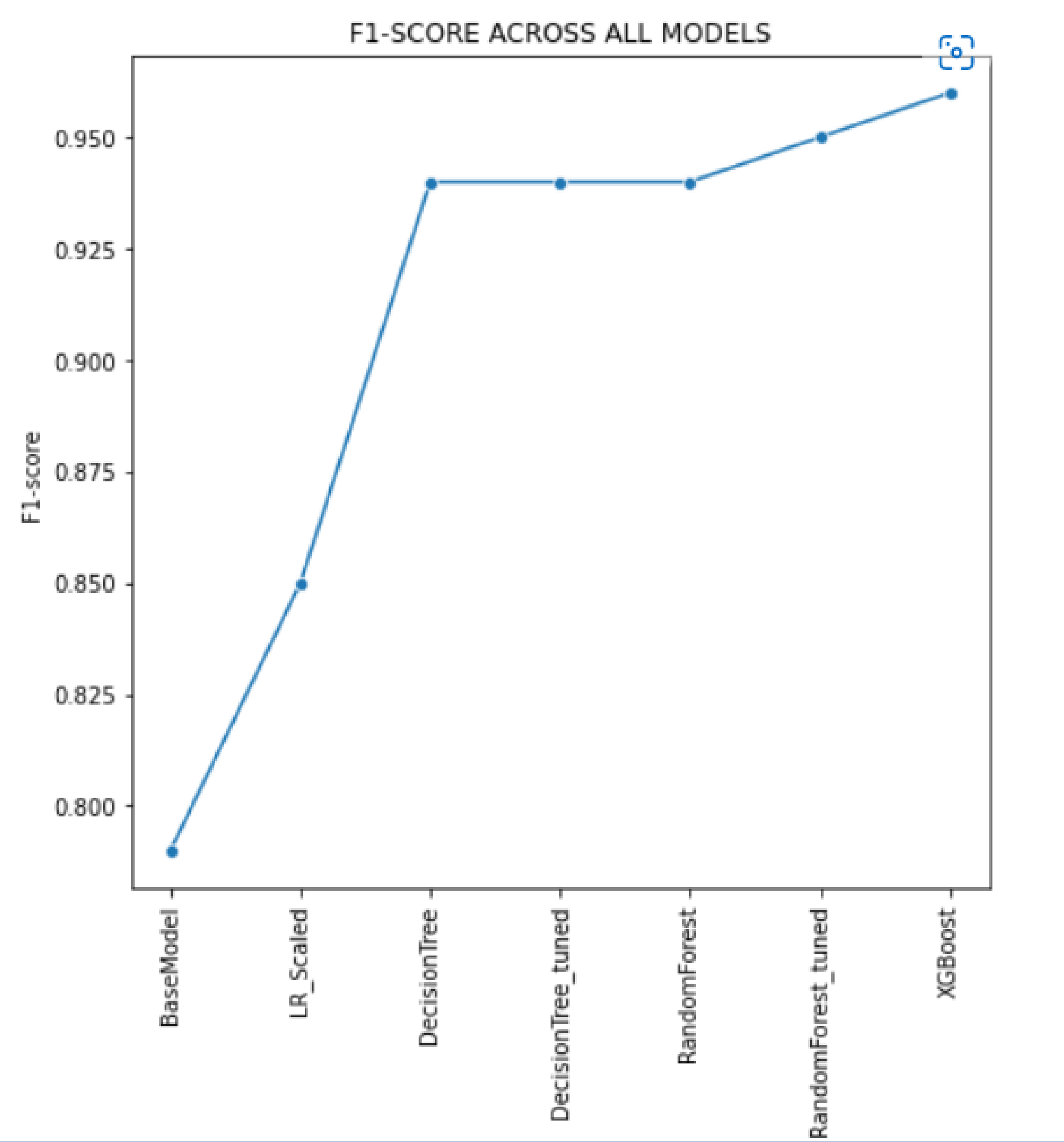
Model Comparison

	Model_Name	TN	FN	TP	FP	Accuracy	Kfold_accuracy	F1-score	ROC-AUCscore
Base_Model	BaseModel	17702	3101	13774	4270	0.81	0.81	0.79	0.81
LR_scaled	LR_Scaled	19837	2713	14162	2135	0.88	0.81	0.85	0.87
DecisionTree	DecisionTree	20953	1063	15812	1019	0.95	0.95	0.94	0.95
DecisionTreeTuned	DecisionTree_tuned	21189	1238	15637	783	0.95	0.95	0.94	0.95
RandomForest	RandomForest	21258	1149	15726	714	0.95	0.95	0.94	0.95
RandomForestTuned	RandomForest_tuned	21503	1069	15806	469	0.96	0.96	0.95	0.96
XGBoost	XGBoost	21497	993	15882	475	0.96	0.96	0.96	0.96

We found that the XGBoost model was able to give us the best results overall -96% Accuracy and 96% F1-Score

Graphical Representation of Performance





BUSINESS RECOMMENDATIONS

- We observed the following features have the most influence on customer satisfaction based on the final model: Online Boarding, Class, Inflight Wi-Fi service, Inflight Entertainment and Cleanliness

- We believe that by focussing on these aspects for starters we may be able to ultimately provide a more satisfying experience for the customers which will ultimately help in increasing the loyal customer base for the airline.

- Example :changing the internet service provider/subscribing to OTT platforms with better content/optimizing the online boarding process etc can enhance customer experience

REFERENCES

1.	https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction
2.	https://towardsdatascience.com/predicting-satisfaction-of-airline-passengers-with-classification-76f1516e1d16
3.	https://www.researchgate.net/publication/350552031_Predicting_Airline_Passenger_Satisfaction_with_Classification_Algorithms
4.	https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
5.	https://www.knowledgeisle.com/wp-content/uploads/2019/12/2-Aur%C3%A9lien-G%C3%A9ron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O%E2%80%99Reilly-Media-2019.pdf

THANK YOU