Airlines Customer Satisfaction Analysis

By : M. Haekal Akiyat

About myself

Passionate about building career in data science, with a commitment to mastering advanced techniques and adapting to the dynamic demands of the field. With a background in electrical engineering, I bring a strong analytical foundation and problem solving mindset to support my career transition from Asset Manager to Data Scientist. Currently refining my expertise in data analytics, machine learning, and statistical modeling through the Dibimbing.ID Data Science Bootcamp.

M. Haekal Akiyat

Bandung, West Java

Bachelor Degree, Electrical Engineering

Background

In today's competitive aviation industry, customer satisfaction plays a pivotal role in shaping an airline's reputation and profitability. Leveraging the power of data, this project aims to predict airline customer satisfaction by analyzing key features. By employing advanced machine learning techniques, the project seeks to provide actionable insights that empower airlines to enhance their services, foster customer loyalty, and maintain a competitive edge.

OBJECTIVE

Using a classification model to forecast customer loyalty and determine the primary determinants of airline customer satisfaction.

GOAL

- Improving the overall customer experience by focusing on primary feature affect customer satisfaction
- Providing business recommendations to optimize resource allocation in airlines service aspects having a high influence on satisfaction, implementing differentiation strategies for client segments, and lowering complaints by addressing low value service areas that cause customers not satisfied.

Table of contents

01 EDA

Descriptive Analysis, Univariate Analysis, Multivariate Analysis

02 Data Pre-processing

Missing value & outlier handling, feature engineering, Standardization

03 Modelling

Modeling using Logistic Regression, Random Forest Classifier, XGBoost, LightGBM

04 Explanation AI

Explanation AI using Permutation Feature Importance, Partial Dependence, shapley dan lime

05 Analysis and Recommendation

Actionable Insight and Business Recommendation

06 Deployment

Deployment using streamlit

Dataset Overview

Rows = 103904
Columns = 23

- **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
- **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
- **Departure Delay in Minutes**: Minutes delayed when departure
- **Arrival Delay in Minutes**: Minutes delayed when Arrival
- **Satisfaction**: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

DATASET ACCESS:
https://www.kaggle.com/datasets/sjleshrac/airlines-customer-satisfaction

01

Exploratory Data Analysis

Descriptive Analysis, Univariate Analysis,
Multivariate Analysis

Numerical Columns statistics

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	103904.0	51951.500000	29994.645522	0.0	25975.75	51951.5	77927.25	103903.0
id	103904.0	64924.210502	37463.812252	1.0	32533.75	64856.5	97368.25	129880.0
Age	103904.0	39.379706	15.114964	7.0	27.00	40.0	51.00	85.0
Flight Distance	103904.0	1189.448375	997.147281	31.0	414.00	843.0	1743.00	4983.0
Inflight wifi service	103904.0	2.729683	1.327829	0.0	2.00	3.0	4.00	5.0
Departure/Arrival time convenient	103904.0	3.060296	1.525075	0.0	2.00	3.0	4.00	5.0
Ease of Online booking	103904.0	2.756901	1.398929	0.0	2.00	3.0	4.00	5.0
Gate location	103904.0	2.976883	1.277621	0.0	2.00	3.0	4.00	5.0
Food and drink	103904.0	3.202129	1.329533	0.0	2.00	3.0	4.00	5.0
Online boarding	103904.0	3.250375	1.349509	0.0	2.00	3.0	4.00	5.0
Seat comfort	103904.0	3.439396	1.319088	0.0	2.00	4.0	5.00	5.0
Inflight entertainment	103904.0	3.358158	1.332991	0.0	2.00	4.0	4.00	5.0
On-board service	103904.0	3.382363	1.288354	0.0	2.00	4.0	4.00	5.0
Leg room service	103904.0	3.351055	1.315605	0.0	2.00	4.0	4.00	5.0
Baggage handling	103904.0	3.631833	1.180903	1.0	3.00	4.0	5.00	5.0
Checkin service	103904.0	3.304290	1.265396	0.0	3.00	3.0	4.00	5.0
Inflight service	103904.0	3.640428	1.175663	0.0	3.00	4.0	5.00	5.0
Cleanliness	103904.0	3.286351	1.312273	0.0	2.00	3.0	4.00	5.0
Departure Delay in Minutes	103904.0	14.815618	38.230901	0.0	0.00	0.0	12.00	1592.0
Arrival Delay in Minutes	103594.0	15.178678	38.698682	0.0	0.00	0.0	13.00	1584.0

Key Points:

- Most of the categorical columns are ordinal columns. From range 0 to 5.
- Remove **Unnamed:0** and **id** because those columns don't have any significance value to the data.

Categorical Columns statistics

	count	unique	top	freq
Gender	103904	2	Female	52727
Customer Type	103904	2	Loyal Customer	84923
Type of Travel	103904	2	Business travel	71655
Class	103904	3	Business	49665
satisfaction	103904	2	neutral or dissatisfied	58879

Unique counts for categorical feature have less than 4 unique value

Dataset Summary

	feature	data_type	null_num	%null	nunique
0	Gender	object	0	0.000000	2
1	Customer Type	object	0	0.000000	2
2	Age	int64	0	0.000000	75
3	Type of Travel	object	0	0.000000	2
4	Class	object	0	0.000000	3
5	Flight Distance	int64	0	0.000000	3802
6	Inflight wifi service	int64	0	0.000000	6
7	Departure/Arrival time convenient	int64	0	0.000000	6
8	Ease of Online booking	int64	0	0.000000	6
9	Gate location	int64	0	0.000000	6
10	Food and drink	int64	0	0.000000	6
11	Online boarding	int64	0	0.000000	6
12	Seat comfort	int64	0	0.000000	6
13	Inflight entertainment	int64	0	0.000000	6
14	On-board service	int64	0	0.000000	6
15	Leg room service	int64	0	0.000000	6
16	Baggage handling	int64	0	0.000000	5
17	Checkin service	int64	0	0.000000	6
18	Inflight service	int64	0	0.000000	6
19	Cleanliness	int64	0	0.000000	6
20	Departure Delay in Minutes	int64	0	0.000000	446
21	Arrival Delay in Minutes	float64	310	0.298352	455
22	satisfaction	object	0	0.000000	2

Key Points:

- Missing value on Arrival Delay in Minutes, 0,028% data missing.

Descriptive Analysis

Data Types

5 Categorical
18 numerical

Missing Value

310 missing value on
Arrival Delay in Minutes

Duplicated Rows

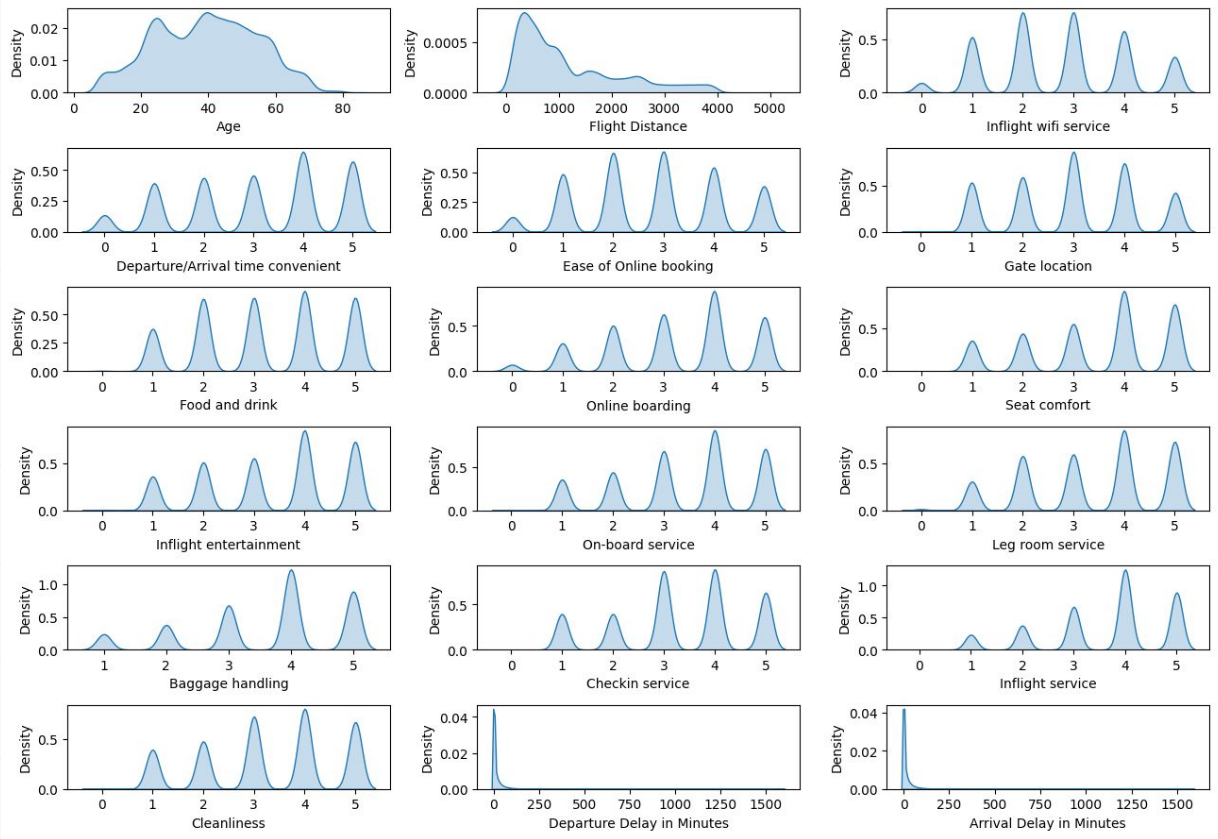
0 duplicated rows

Customer Satisfaction Proportion

Key takes:

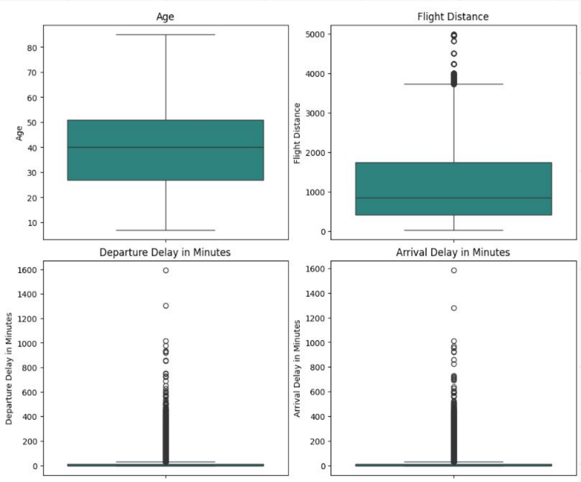
- The data quite unbalance between **neutral or dissatisfied** and **satisfied**
- **Neural or dissatisfied** are on the same level of satisfaction
- We change class **neutral or dissatisfied** class as **not satisfied.**

Distribution Plot

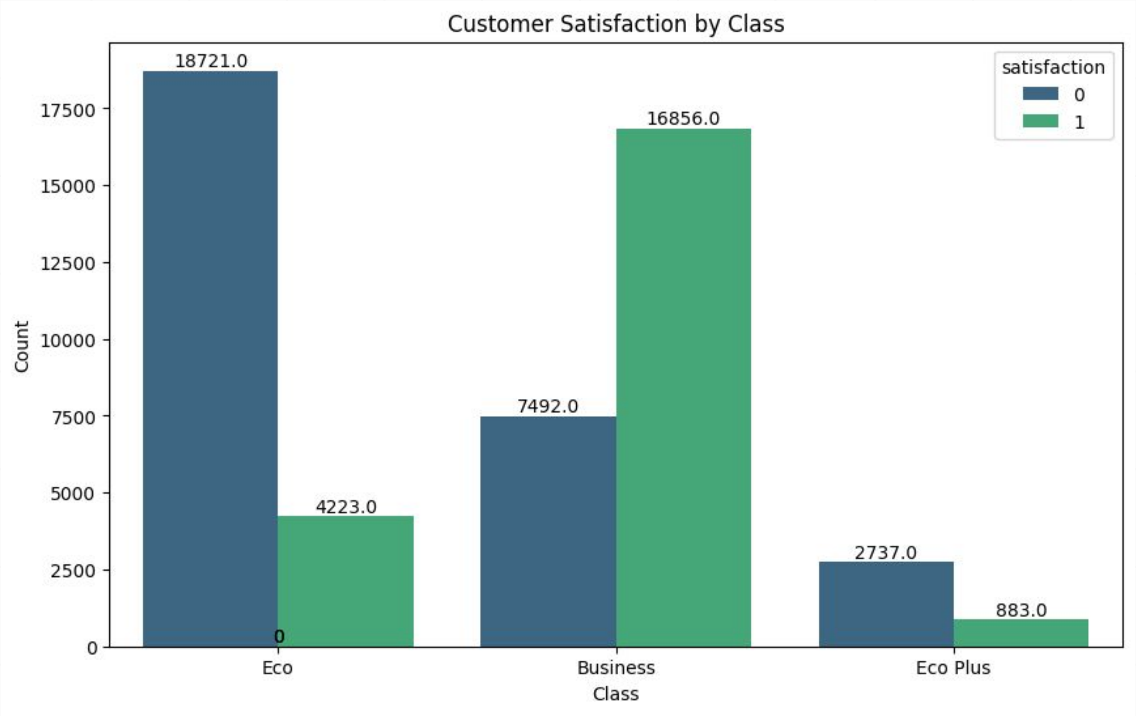


Key takes:

- **Departure delay in minutes** and **Arrival delay in minutes** both has many large outliers with skewed right distribution
- Most of the numerical columns are categorical ordinal from level 0 to 5.



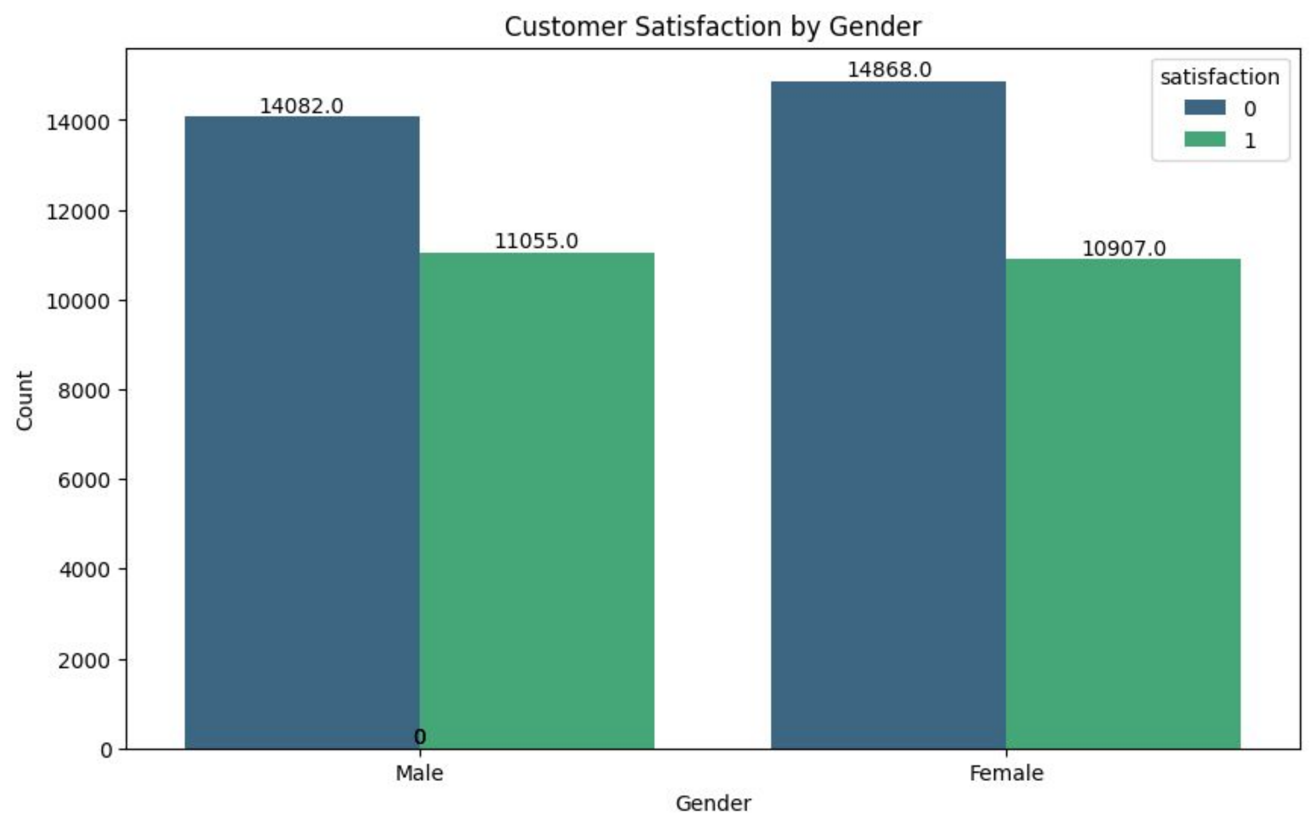
Customer Satisfaction by Class



Key takes:

- Many of the **Eco** flight class were dissatisfied with the service.
- Many of the **business** class are satisfied with flight services.
- **Eco Plus** class is the minority class
- flight **Class** has significant impact on customer satisfaction

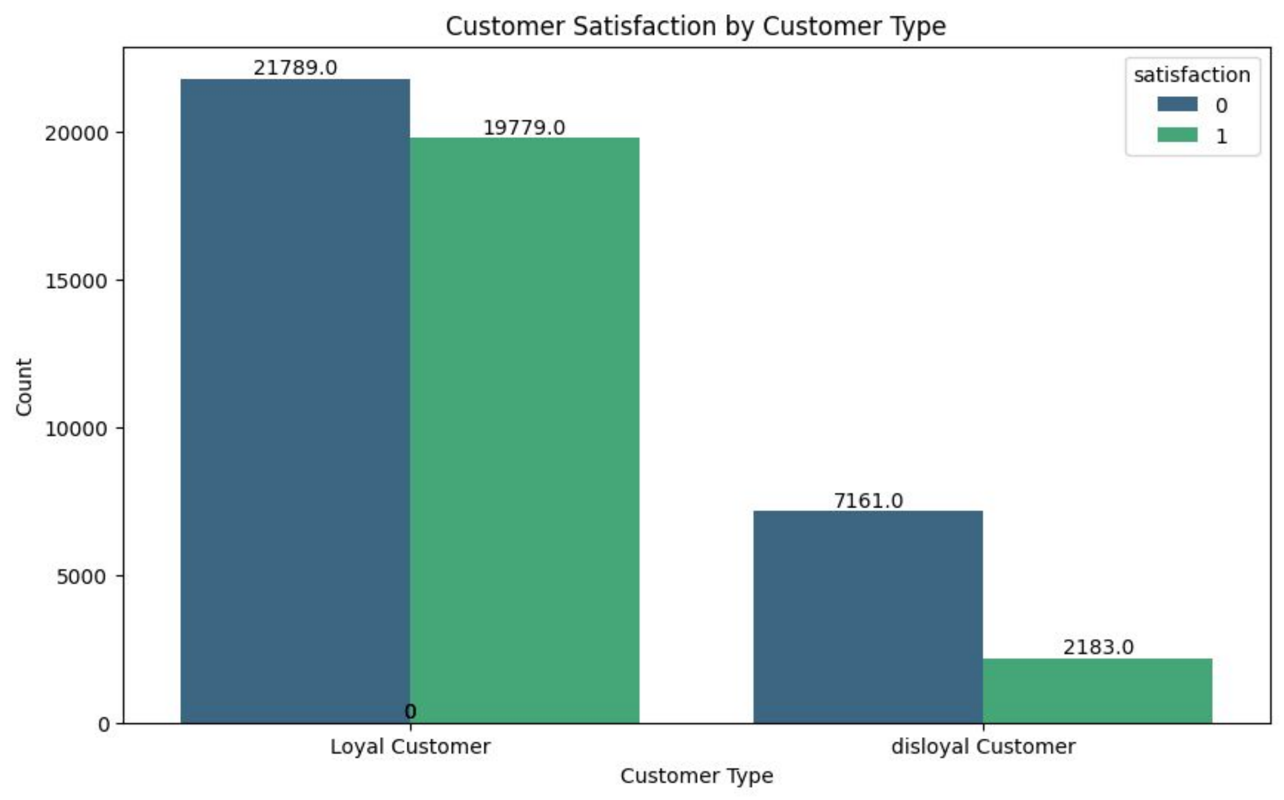
Customer Satisfaction by Gender



Key takes:

- **Gender** does not have much influence on customer satisfaction, both genders have almost the same data distribution

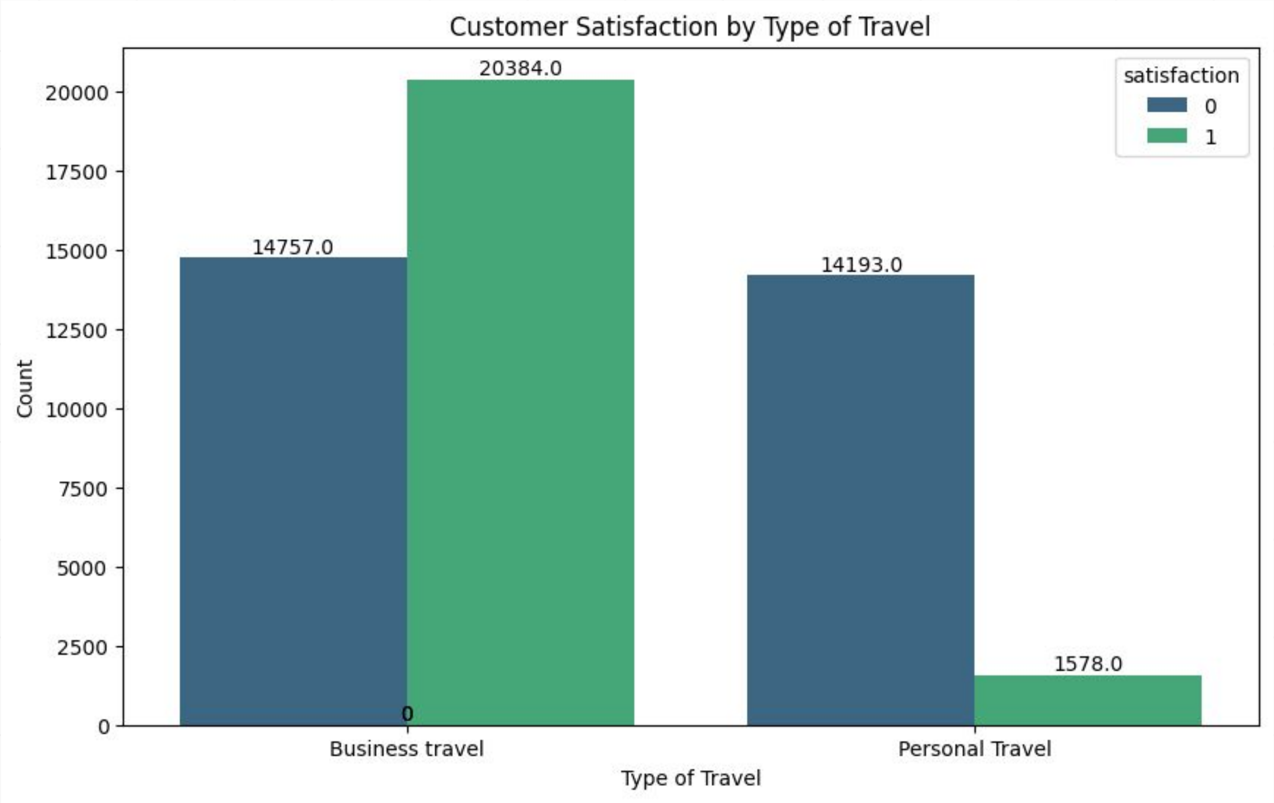
Customer Satisfaction by Customer Type



Key takes:

- Many of the customers who are **loyal**, satisfied with flight services.
- Most of the flight customer are from loyal customers
- **Customer type** has significant impact on customer satisfaction

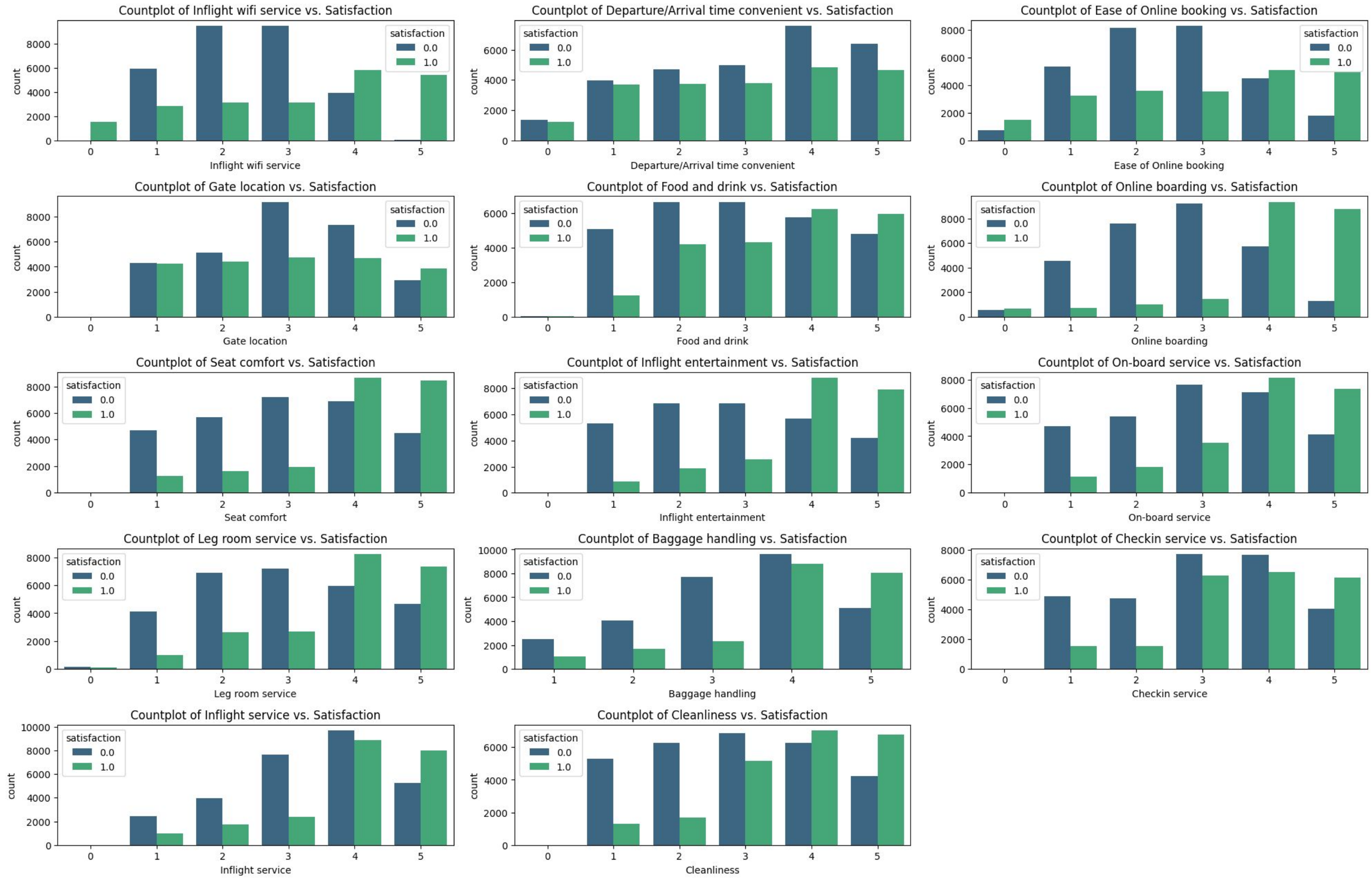
Customer Satisfaction by Type of Travel



Key takes:

- Most of the **personal travel** customer are not satisfied with the flight services
- This makes sense that personal customers use their own funds to travel while **business travel** uses company funds. They have higher expectations for flight services.
- **Type Of Travel** has significant impact on customer satisfaction

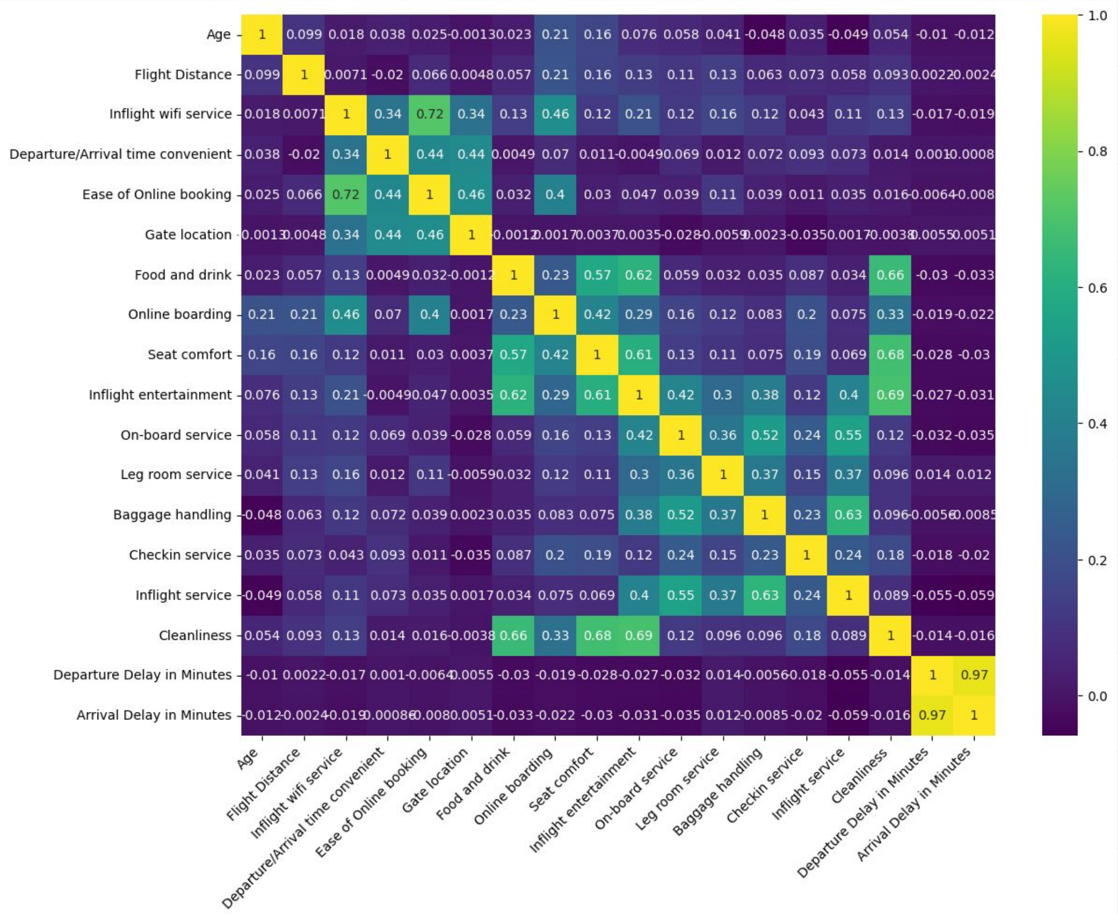
Countplot Ordinal Category



Key takes:

- **inflight wifi services, Ease of Online booking, seat comfort, inflight entertainment, On board services, baggage handling, inflight service** have significant impact on customer satisfaction

Correlation Matrix (heatmap)



Key takes:

- **Departure delay in Minutes** and **arrival Delay in Minutes** has high positive correlation, multicollinearity potential.
- **Ease of Online Booking** has high correlation with **Inflight Wi-Fi Services**
- **Cleanliness** have high positive correlation with **Food and drink, seat comfort and inflight entertainment**

02

Data Pre-processing

Missing value, feature engineering,
Standardization

Missing value and Duplicated Rows

Missing Value

310 missing value on
Arrival Delay in Minutes
filled with median().

Duplicated Rows

0 duplicated rows

Feature Engineering (redundant feature)

	feature	vif_score
18	Arrival Delay in Minutes	13.539626
17	Departure Delay in Minutes	13.529211
10	Inflight entertainment	3.729435
16	Cleanliness	2.826332
5	Ease of Online booking	2.656251
3	Inflight wifi service	2.399639
9	Seat comfort	2.353622
7	Food and drink	2.174948
15	Inflight service	2.053000
8	Online boarding	1.878693
13	Baggage handling	1.874160
11	On-board service	1.754817
6	Gate location	1.480994
4	Departure/Arrival time convenient	1.431220
12	Leg room service	1.300523
14	Checkin service	1.212913
1	Age	1.098030
2	Flight Distance	1.095129

Key takes:

- **Arrival Delay in Minutes** and **Departure Delay in Minutes** have high correlation with other features, besides with their own features
- **Remove redundant column**

We **remove Arrival Delay in Minutes**. Late departure means late arrival

	feature	vif_score
10	Inflight entertainment	3.729344
16	Cleanliness	2.826331
5	Ease of Online booking	2.656251
3	Inflight wifi service	2.399588
9	Seat comfort	2.353594
7	Food and drink	2.174782
15	Inflight service	2.052328
8	Online boarding	1.878658
13	Baggage handling	1.874120
11	On-board service	1.754614
6	Gate location	1.480992
4	Departure/Arrival time convenient	1.431219
12	Leg room service	1.300496
14	Checkin service	1.212911
1	Age	1.098007
2	Flight Distance	1.095047
17	Departure Delay in Minutes	1.007694

Feature Engineering

Encoding
Encode object column with LabelEncoder()
Note : Only use **LabelEncoder.fit_transform(train)** on train data, do not use in validation or test data. For data validation and data test use **LabelEncoder.transform(val/test)**

Gender	Encoded
Female	1
Male	0

Customer Type	Encoded
Loyal Customer	1
disloyal customer	0

Class	Encoded
Business	0
Eco	1
Eco Plus	2

Type of Travel	Encoded
Business Travel	0
Personal Travel	1

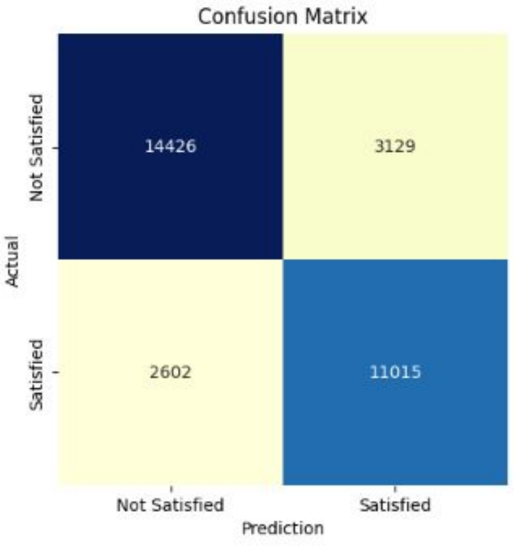
Standardize Data
Standardize our data with standard scaler
The purpose of a standard scaler is to ensure that all features are scaled appropriately. Especially for algorithms which are sensitive to scale differences in features. (Model with distance based : SVM, KNN, K Means)

03

Modelling & Evaluation

Modeling using Logistic Regression, Random Forest Classifier, XGboost, LightGBM

Confusion Matrix



Logistic Regression

Logistic Regression F1-Score (0)

83%

Classification Report

LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.85	0.82	0.83	17555
1	0.78	0.81	0.79	13617
accuracy			0.82	31172
macro avg	0.81	0.82	0.81	31172
weighted avg	0.82	0.82	0.82	31172

Classification Report

```
RANDOM_FOREST
              precision    recall  f1-score   support

           0       0.95      0.98      0.97     17555
           1       0.97      0.94      0.96     13617

    accuracy                           0.96     31172
   macro avg       0.96      0.96      0.96     31172
weighted avg       0.96      0.96      0.96     31172
```

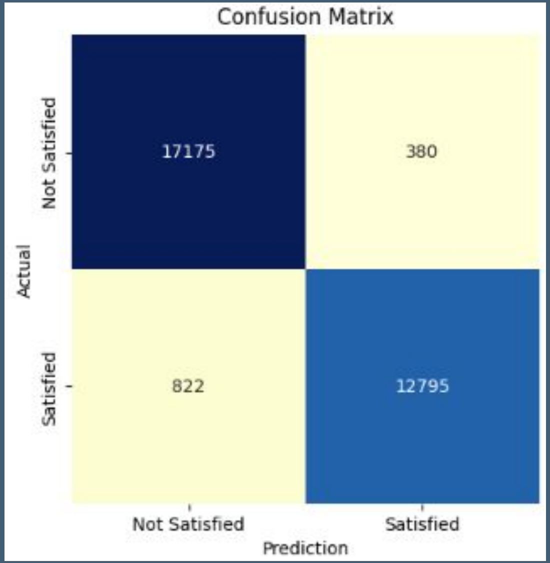
Best Parameter

Best hyperparameters for random forest:

- 'n_estimators': 150
- 'min_samples_split': 2
- 'min_samples_leaf': 1
- 'max_depth': 30

Feature Importance

	Feature	Importance
11	Online boarding	0.179435
6	Inflight wifi service	0.141845
4	Class	0.114538
3	Type of Travel	0.093106
12	Seat comfort	0.056255
13	Inflight entertainment	0.053940
8	Ease of Online booking	0.039699
15	Leg room service	0.038291
5	Flight Distance	0.037481
1	Customer Type	0.036417
2	Age	0.030579
14	On-board service	0.026910
16	Baggage handling	0.024619
19	Cleanliness	0.023264
17	Checkin service	0.023020
18	Inflight service	0.022636
7	Departure/Arrival time convenient	0.015604
9	Gate location	0.014948
20	Departure Delay in Minutes	0.011810
10	Food and drink	0.011730
0	Gender	0.003872



Random Forest Classifier

Random Forest F1-Score (0)

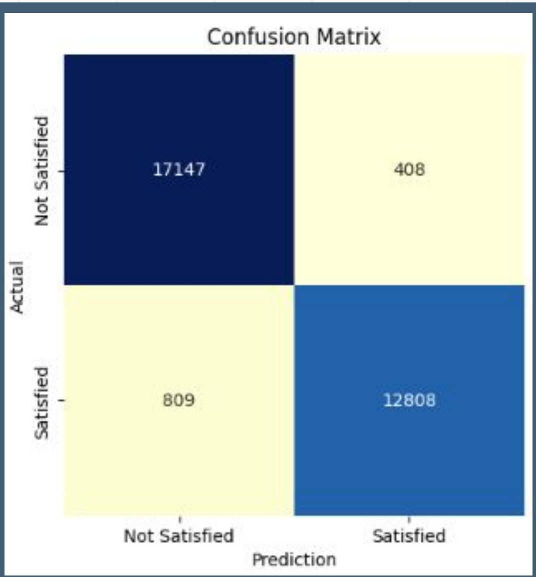
97 %

Classification Report

```
XGBOOST
              precision    recall  f1-score   support

           0       0.95      0.98      0.97     17555
           1       0.97      0.94      0.95     13617

    accuracy                           0.96     31172
   macro avg       0.96      0.96      0.96     31172
weighted avg       0.96      0.96      0.96     31172
```



Best Parameter

Best hyperparameters for XGBoost:

- 'n_estimators': 200
- 'min_child_weight': 5
- 'max_depth': 11
- 'learning_rate': 0.1

XGBOOST Classifier

XGBOOST F1-Score (0)

97 %

Feature Importance

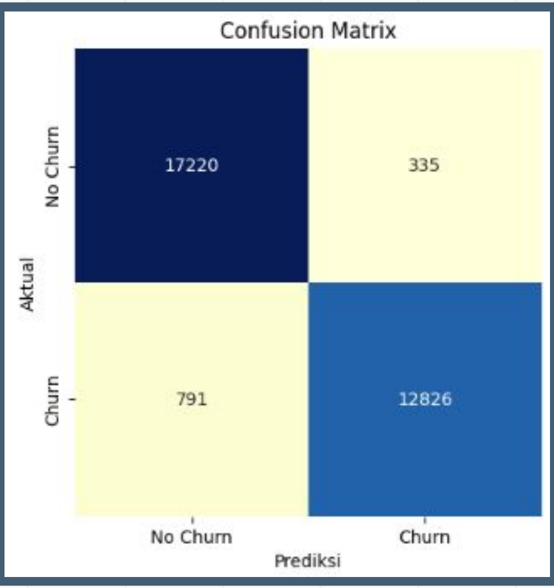
	Feature	Importance
11	Online boarding	0.403573
3	Type of Travel	0.203012
6	Inflight wifi service	0.115738
1	Customer Type	0.053760
13	Inflight entertainment	0.042273
4	Class	0.037004
17	Checkin service	0.020411
9	Gate location	0.015955
15	Leg room service	0.014664
16	Baggage handling	0.013978
19	Cleanliness	0.013144
12	Seat comfort	0.013093
18	Inflight service	0.012142
14	On-board service	0.007517
8	Ease of Online booking	0.006239
2	Age	0.006018
7	Departure/Arrival time convenient	0.005788
10	Food and drink	0.004666
20	Departure Delay in Minutes	0.004007
5	Flight Distance	0.003761
0	Gender	0.003257

Classification Report

```
LIGHT_GBM
              precision    recall  f1-score   support

           0       0.96      0.98      0.97     17555
           1       0.97      0.94      0.96     13617

    accuracy                           0.96     31172
   macro avg       0.97      0.96      0.96     31172
weighted avg       0.96      0.96      0.96     31172
```



Best Parameter

Best hyperparameters for LightGBM:

- 'num_leaves': 50
- 'n_estimators': 100
- 'min_child_samples': 20
- 'max_depth': -1
- 'learning_rate': 0.1

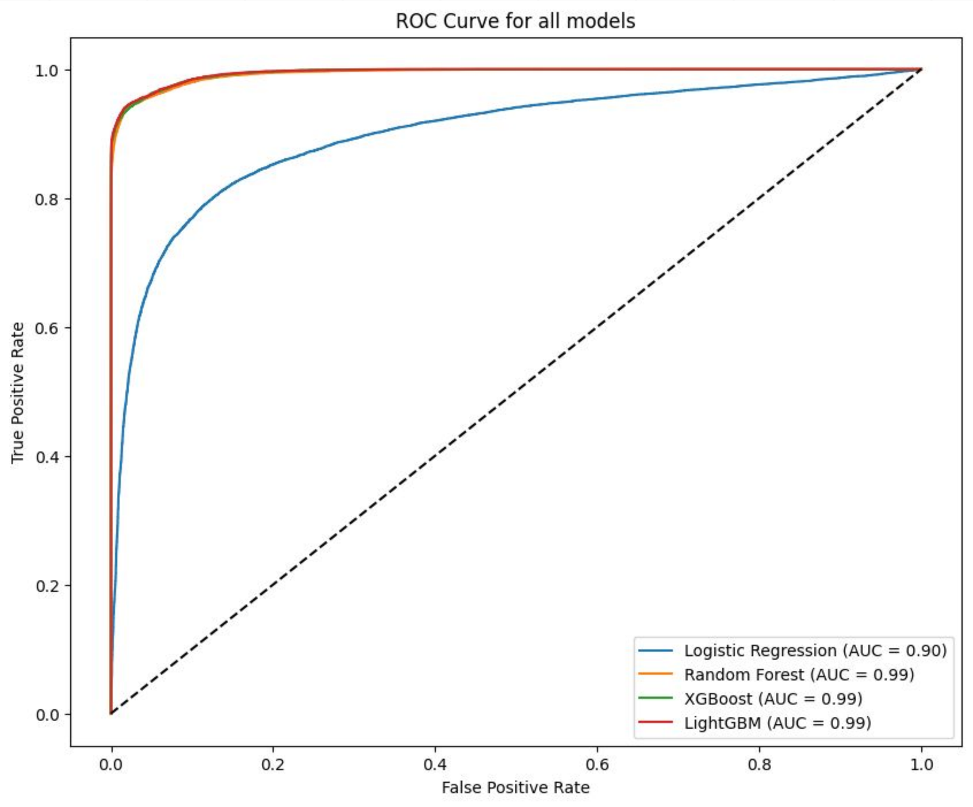
LGBM Classifier

LGBM F1-Score 97%

Feature Importance

	Feature	Importance
6	Inflight wifi service	596
2	Age	520
5	Flight Distance	442
16	Baggage handling	305
1	Customer Type	278
18	Inflight service	241
9	Gate location	240
20	Departure Delay in Minutes	224
12	Seat comfort	217
17	Checkin service	214
3	Type of Travel	211
13	Inflight entertainment	210
11	Online boarding	206
4	Class	205
15	Leg room service	172
14	On-board service	153
19	Cleanliness	134
7	Departure/Arrival time convenient	125
8	Ease of Online booking	109
10	Food and drink	78
0	Gender	20

Model Summary using AUC



Key takes:

- The Logistic Regression Model is not better than Random Forest, XGBoost, and LightGBM, as shown by the ROC curve above. This is typical since complex datasets are not a good fit for logistic regression. As our baseline model, we are employing logistic regression.

Model Performance Summary

For Class 1, Satisfied Customer

	Model	Precision (1)	Recall (1)	F1-score (1)	Support (1)	Accuracy	Macro avg F1	Weighted avg F1
1	LightGBM	97,45	94,19	95,80	13617.0	96,39	96,31	96,38
2	Random Forest	97,20	93,96	95,56	13617.0	96,18	96,11	96,17
3	XGBoost	96,91	94,06	95,46	13617.0	96,10	96,02	96,09
4	Logistic Regression	81,51	81,67	81,59	13617.0	83,90	83,64	83,90

Key takes:

- Model LGBM has better performance compared to other models. This was shown by better precision, recall, F1_score and Accuracy score for **satisfies customer (class 1)**
- For this analysis, we're focusing on **F1 Score**, since there was **unbalance** data between class target.

Interpretation LGBM:

- **Precision** = Out of 100 customer who are predicted satisfied, 97 customer are actually satisfied
- **Recall** = Out of 100 customer who are actually satisfied, our model can only detect 94 of them.
- **F1 Score** = The harmony between Precision and Recall

Model Performance Summary

For Class 0, Not Satisfied Customer

	Model	Precision (0)	Recall (0)	F1-score (0)	Support (0)	Accuracy	Macro avg F1	Weighted avg F1
1	LightGBM	95,61	98,09	96,83	13617.0	96,39	96,31	96,38
2	Random Forest	95,44	97,90	96,65	13617.0	96,18	96,11	96,17
3	XGBoost	95,49	97,68	96,57	13617.0	96,10	96,02	96,09
4	Logistic Regression	85,76	85,63	85,69	13617.0	83,90	83,64	83,90

Key takes:

- Model LGBM has better performance compared to other models. This was shown by better precision, recall, F1_score and Accuracy score for **not satisfies customer (class 0)**
- For this analysis, we're focusing on **F1 Score**, since there was **unbalance** data between class target.

Interpretation LGBM:

- **Precision** = Out of 100 customer who are predicted not satisfied, 95 customer are actually not satisfied
- **Recall** = Out of 100 customer who are actually not satisfied, our model can only detect 98 of them.
- **F1 Score** = The harmony between Precision and Recall

Should we implement model with best performance?

Keep in mind, model with better performance not always have a good reasonable feature. We'll use explainable AI to interpret our model

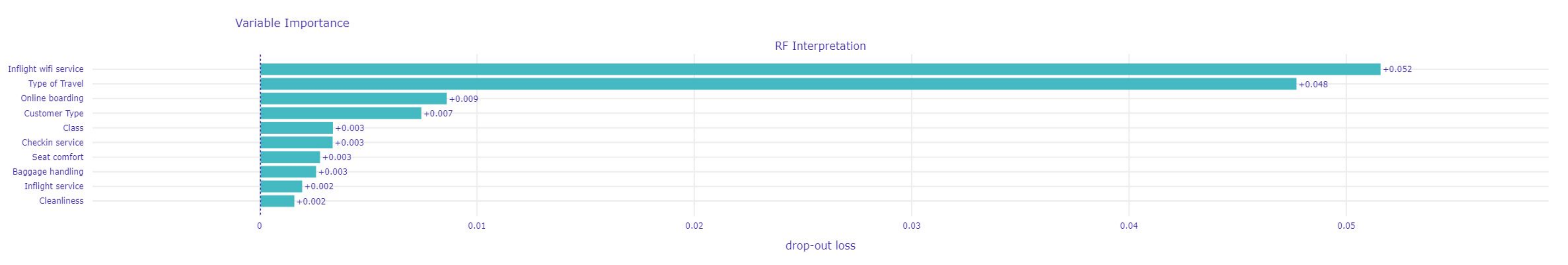
04

Explanation AI

Explanation AI using Permutation Feature Importance, Partial Dependence Plot and Lime

Permutable Feature Importance

Random Forest Classifier

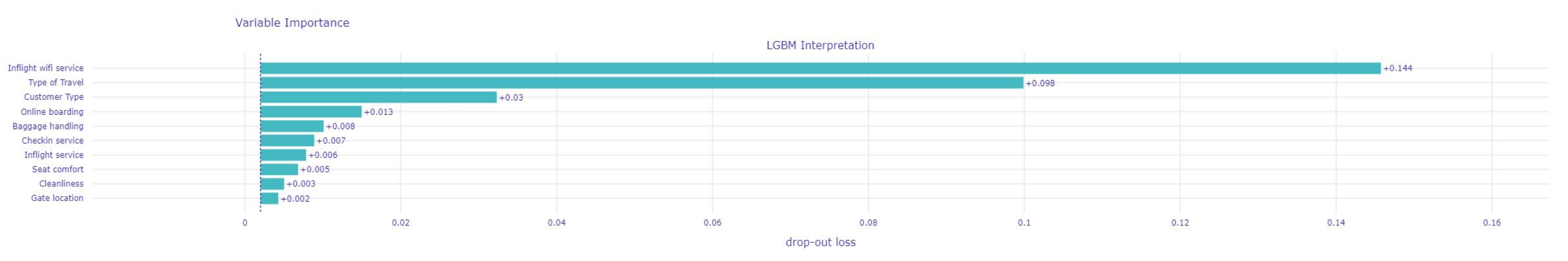


Key takes:

- Random Forest model have 10 features as an important feature with drop loss above 0.002 on each important feature.
- The **inflight wifi service** and **type of travel** features have a significant influence on model predictions

Permutable Feature Importance

LGBM Classifier

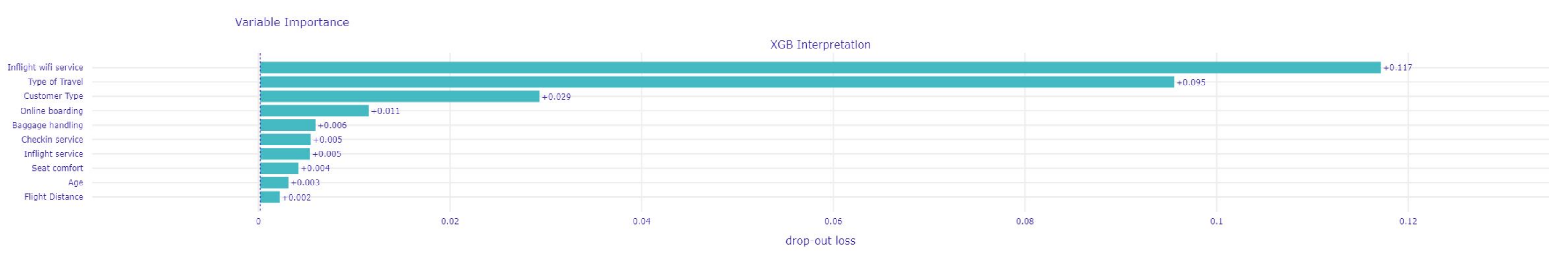


Key takes:

- LGBM model have 10 features as an important feature with drop loss above 0.002 on each important feature.
- The **inflight wifi service** and **type of travel** features have a significant influence on model predictions

Permutable Feature Importance

XGBOOST Classifier



Key takes:

- XGboost 10 model features as an important feature with drop loss above 0.002 on each important feature.
- The **inflight wifi service** and **type of travel** features have a significant influence on model predictions

Partial Dependence Plot

Random Forest Classifier

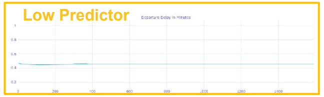
Note:
Mod Predictor : Moderate Predictor
High Predictor : High Predictor
Low Predictor : Low Predictor

Interpretation:

- Customers with `Customer Type` = 1 or disloyal customers have a 31% probability of being satisfied with flight services. While customers with `Customer Type` = 0 or Loyal customers have a 49% probability of being satisfied with flight services.
- Customers with `Type of Travel` = 1 or personal travel customers have a 22% probability of being satisfied with flight services. Meanwhile, customers with `Type of Travel` = 0 or business travel customers have a 55% probability of being satisfied with flight services.
- Customers who rate `Inflight Wifi Service` as 0 have a 71% probability of being satisfied. Customers who give a rating of 1 to 3 have a 37% probability of being satisfied. Customers who give a rating of 4 have a 55% probability of being satisfied. Customers who give a rating of 5 have a 74% probability of being satisfied.
- Customers with `Class` = 0 or business class customers have a 47.5% probability of being satisfied with the flight service. Customers with `Class` = 1 or Eco Class customers have a probability and `Class` = 2 or Eco Plus class customers have a 37.5% probability of being satisfied with the flight service
- Customers who rate `Online Onboarding` as 0 have a 43% probability of being satisfied. Customers who give a rating of 1 have a 43% probability of being satisfied. Customers who give a rating of 2 have a 41% probability of being satisfied. Customers who rate 3 have a 40% probability of being satisfied. Customers who give a rating of 4 have a 49% probability of being satisfied. Customers who give a rating of 5 have a 60% probability of being satisfied.
- Customers who give a `Seat Comfort` rating of 0 to 2 have a 44% probability of being satisfied. Customers who give a rating of 3 have a 43% probability of being satisfied. Customers who give a rating of 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 52% probability of being satisfied.
- Customers who rate `Inflight Entertainment` as 0 have a 41% probability of being satisfied. Customers who give a rating of 1 have a 42% probability of being satisfied. Customers who give a rating of 2 have a 44% probability of being satisfied. Customers who rate 3 have a 45% probability of being satisfied. Customers who give a rating of 4 have a 48% probability of being satisfied. Customers who rate 5 have a 49% probability of being satisfied.
- Customers who rate `Checkin service` as 0 to 2 have a 43% probability of being satisfied. Customers who rate 3 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 54% probability of being satisfied.
- Customers who rate `Baggage handling` at 0 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 51% probability of being satisfied.
- `Class` flight class, Business Class have 44% of being satisfied, class Eco and Eco plus have a 35% probability of being satisfied.

Partial Dependence Plot

LGBM Classifier



Note:
Mod Predictor : Moderate Predictor
High Predictor : High Predictor
Low Predictor : Low Predictor

Interpretation:

- Customers with `Customer Type` = 1 or disloyal customers have a 26% probability of being satisfied with flight services. While customers with `Customer Type` = 0 or Loyal customers have a 52% probability of being satisfied with flight services.
- Customers with `Type of Travel` = 1 or personal travel customers have a 20% probability of being satisfied with flight services. Customers with `Type of Travel` = 0 or business travel customers have a 60% probability of being satisfied with flight services.
- Customers who rate `Inflight wifi service` as 0 have a 99.7% probability of being satisfied. While customers who give a rating of 1 to 4 have a probability between 36% to 53% will be satisfied. Customers who give a rating of 5 have a 93.5% probability of being satisfied.
- Customers who rate `Online Onboarding` 1 to 4 have a 47% probability of being satisfied. Customers who give a rating of 5 have a 61% probability of being satisfied.
- Customers who rate `Seat Comfort` 1 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 56% probability of being satisfied.
- Customers who rate `On Board Service` 0 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 50% probability of being satisfied.
- Customers who rate `Baggage handling` as 1 to 4 have a 47% to 50% probability of being satisfied. Customers who give a rating of 5 have a 57% probability of being satisfied.
- Customers who rate `Checkin service` at 0 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 55% probability of being satisfied.
- Customers who rate `Inflight service` as 0 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 55% probability of being satisfied.- Customers who give a `Cleanliness` rating of 0 to 4 have a 46% probability of being satisfied. Customers who give a rating of 5 have a 53.9% probability of being satisfied.
- In each `Class` flight class, Business Class, Eco Class and Eco Plus Class have approximately the same probability of 44% to be satisfied.

Partial Dependence Plot

XGBOOST Classifier

Note:
Mod Predictor : Moderate Predictor
High Predictor : High Predictor
Low Predictor : Low Predictor

Interpretation:

- Customers with `Customer Type` = 1 or disloyal customers have a 22% probability of being satisfied with flight services. While customers with `Customer Type` = 0 or Loyal customers have a 46% probability of being satisfied with flight services.
- Customers with `Type of Travel` = 1 or personal travel customers have an 18% probability of being satisfied with flight services. Customers with `Type of Travel` = 0 or business travel customers have a 57.5% probability of being satisfied with the flight service.
- Customers who rate `Inflight wifi service` as 0 have a 95% probability of being satisfied. While customers who give a rating of 1 to 4 have a probability of between 33% to 36% will be satisfied. Customers who give a rating of 5 have a 92% probability of being satisfied.- Customers who rate `Online Onboarding` as 0 have a 43% probability of being satisfied. While customers who give a rating of 1 to 4 have a probability between 42% and 43% will be satisfied. Customers who give a rating of 5 have a 53% probability of being satisfied.
- Customers who give a `Seat Comfort` rating of 0 have a 41% probability of being satisfied. While customers who give a rating of 1 to 4 have a probability between 40% to 41% will be satisfied. Customers who give a rating of 5 have a 50% probability of being satisfied.
- Customers who rate `On Board Service` as 0 to 4 have a probability between 39% to 40% will be satisfied. Customers who give a rating of 5 have a 44% probability of being satisfied.- Customers who rate `Baggage handling` at 1 have a 43% probability of being satisfied. Customers who give a rating of 2 have a 40% probability of being satisfied. Customers who give a rating of 3 have a 33% probability of being satisfied. Customers who give a rating of 4 have a 38% probability of being satisfied. Customers who rate 5 have a 50% probability of being satisfied.
- Customers who give a `Checkin service` rating of 0 to 4 have a probability between 38% to 40% will be satisfied. Customers who give a rating of 5 have a 51% probability of being satisfied.
- Customers who rate `Inflight service` at 0 have a 43% probability of being satisfied. Customers who give a rating of 1 have a 43% probability of being satisfied. Customers who give a rating of 2 have a 41% probability of being satisfied. Customers who give a rating of 3 have a 39% probability of being satisfied. Customers who give a rating of 4 have a 40% probability of being satisfied. Customers who give a rating of 5 have a 50% probability of being satisfied.
- Customers who give a `Cleanliness` rating of 0 to 4 have a 40% probability of being satisfied. Customers who give a rating of 5 have a 46% probability of being satisfied.
- For each `Class` flight class, Business Class, Eco Class and Eco Plus Class have a 39% probability of being satisfied.

Partial Dependence Plot Summary

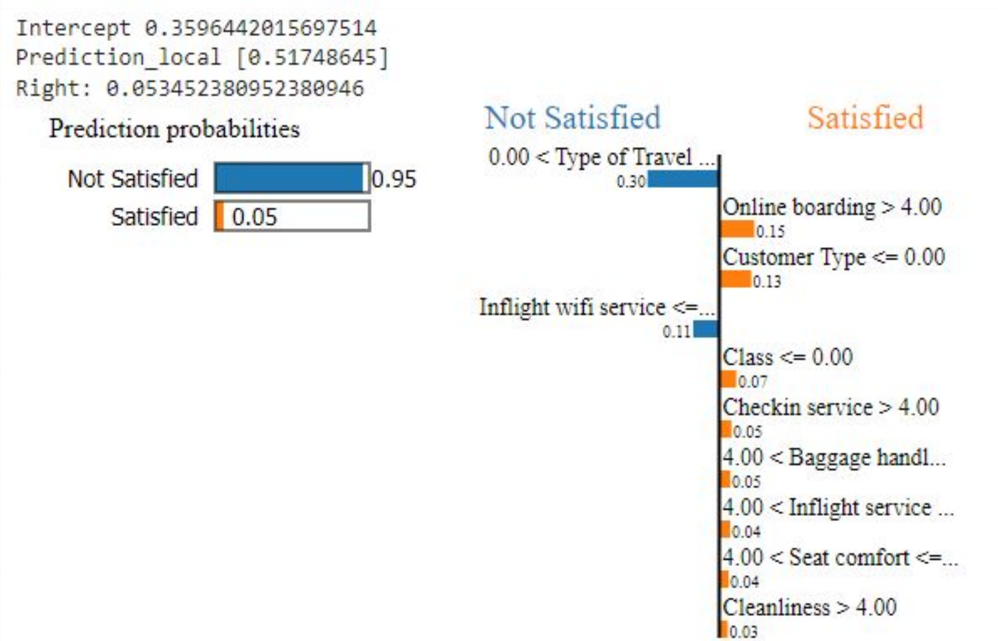
	High Predictor Count	Moderate Predictor Count	Low Predictor Count
Random Forest	5	7	9
LightGBM	3	5	13
XGBoost	3	3	15

Key takes:

- We choose a model that has more high predictors and moderate predictors
- Random Forest Model has 5 High predictor, 7 Moderate and 9 Low Predictor.

Local Interpretable Model-Agnostic Explanations

Random Forest Classifier



Type of Travel <= 0,00	29,11642
Customer Type <= 0,00	14,74793
Inflight wifi service <= 2,00	-11,33609
Online boarding <= 2,00	-8,897915
0,00 < Class <= 1,00	-6,284505
Baggage handling <= 3,00	-3,607608
Leg room service > 4,00	3,251692
Cleanliness <= 2,00	-2,710878
Inflight service <= 3,00	-2,376596
3,00 < Ease of Online booking <= 4,00	2,142828

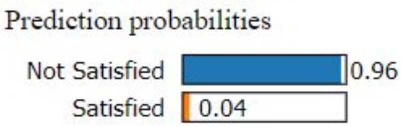
Feature	Value
Type of Travel	1.00
Online boarding	5.00
Customer Type	0.00
Inflight wifi service	1.00
Class	0.00
Checkin service	5.00
Baggage handling	5.00
Inflight service	5.00
Seat comfort	5.00
Cleanliness	5.00

- **Type of travel** has 0,30 or 30% influence on customer satisfaction. Business travel (Type of Travel = 0) will likely not satisfied with airlines services.
- **Inflight Wifi Service** has 0,11 or 11% influence on customer satisfaction. Customer who gives rating less than 2 on **Inflight Wifi Service**) will likely not satisfied with airlines services.

Local Interpretable Model-Agnostic Explanations

LGBM Classifier

```
Intercept -0.022848909262146
Prediction_local [0.51413932]
Right: 0.04224975969854429
```

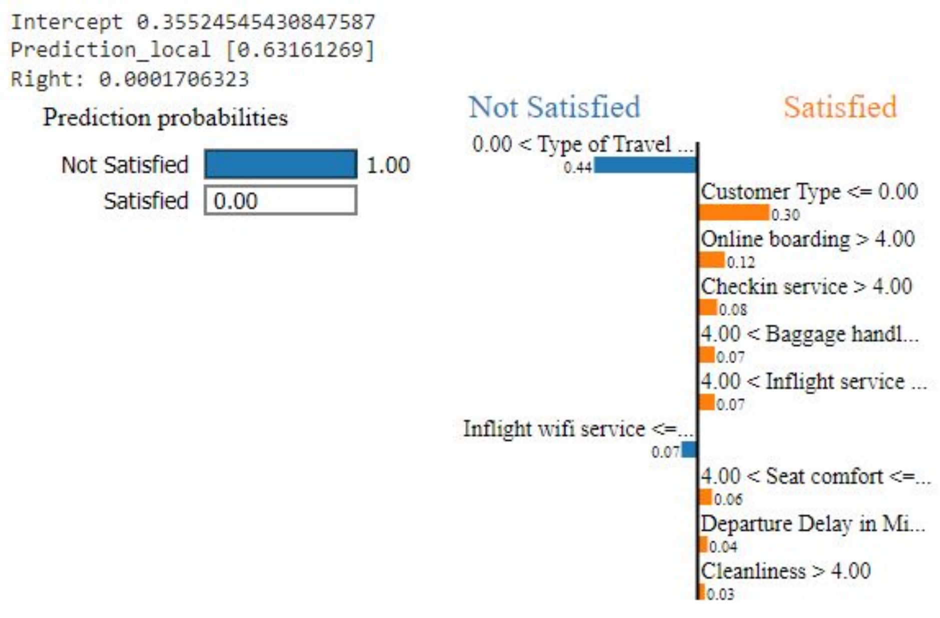


Feature	Value
Type of Travel	0.00
Customer Type	0.00
Inflight wifi service	2.00
Baggage handling	3.00
Online boarding	2.00
Cleanliness	2.00
Departure Delay in Minutes	0.00
Inflight service	3.00
Gate location	2.00
Inflight entertainment	2.00

- **Type of travel** has 0,50 or 50% influence on customer satisfaction. Business travel (Type of Travel = 1) will likely satisfied with airlines services.
- **Inflight Wifi Service** has 0,19 or 19% influence on customer satisfaction. Customer who gives rating less than 2 on **Inflight Wifi Service**) will likely not satisfied with airlines services.

Local Interpretable Model-Agnostic Explanations

XGBOOST Classifier



Type of Travel <= 0,00	42,54
Customer Type <= 0,00	32,44
Inflight entertainment <= 2,00	-4,72
Inflight wifi service <= 2,00	-4,19
Online boarding <= 2,00	-3,96
Leg room service > 4,00	3,73
Cleanliness <= 2,00	-3,70
Departure Delay in Minutes <= 0,00	3,23
Gate location <= 2,00	3,21
Baggage handling <= 3,00	-2,14

Feature	Value
Type of Travel	1.00
Customer Type	0.00
Online boarding	5.00
Checkin service	5.00
Baggage handling	5.00
Inflight service	5.00
Inflight wifi service	1.00
Seat comfort	5.00
Departure Delay in Minutes	0.00
Cleanliness	5.00

- **Type of travel** has 0,42 or 42% influence on customer satisfaction. Business travel (Type of Travel = 0) will likely satisfied with airlines services.
- **Inflight Wifi Service** has 0,0419 or 4,1% influence on customer satisfaction. Customer who gives rating less than 2 on **Inflight Wifi Service**) will likely not satisfied with airlines services.

LIME XAI Model Summary

Random Forest Classifier

Rules	% impact
Type of Travel <= 0,00	29,11642
Customer Type <= 0,00	14,74793
Inflight wifi service <= 2,00	-11,33609
Online boarding <= 2,00	-8,897915
0,00 < Class <= 1,00	-6,284505
Baggage handling <= 3,00	-3,607608
Leg room service > 4,00	3,251692
Cleanliness <= 2,00	-2,710878
Inflight service <= 3,00	-2,376596
3,00 < Ease of Online booking <= 4,00	2,142828

LGBM Classifier

Rules	% impact
Type of Travel <= 0,00	50,79
Customer Type <= 0,00	38,70
Inflight wifi service <= 2,00	-19,08
Online boarding <= 2,00	-5,22
Cleanliness <= 2,00	-4,31
Baggage handling <= 3,00	-3,70
Inflight service <= 3,00	-3,13
Leg room service > 4,00	2,40
837,00 < Flight Distance <= 1742,00	2,03
Departure/Arrival time convenient <= 2,00	-1,941

XGBOOST Classifier

Rules	% impact
Type of Travel <= 0,00	42,54
Customer Type <= 0,00	32,44
Inflight entertainment <= 2,00	-4,72
Inflight wifi service <= 2,00	-4,19
Online boarding <= 2,00	-3,96
Leg room service > 4,00	3,73
Cleanliness <= 2,00	-3,70
Departure Delay in Minutes <= 0,00	3,23
Gate location <= 2,00	3,21
Baggage handling <= 3,00	-2,14

Notes:

- **Positive** on % impact, means direction to **Satisfied** class
- **Negative** on % impact means direction to **Not Satisfied** class

05

Analysis and Business Recommendation

Model review and actionable insight

Model Review

- **XAI Interpretation Results using Permutable Feature importance**
  - Random Forest model has 10 features as feature importance with drop loss above 0.002 on each important feature. `**inflight wifi service**` and `**type of travel**` have significant influence on customer satisfaction.
  - LGBM model 10 features as feature importance with drop loss above 0.002 on each important feature. `**inflight wifi service**` and `**type of travel**` have a significant influence on customer satisfaction, the drop loss of these two features is 2 times greater than random forest.
  - XGBOOST model 10 features as important features with drop loss above 0.002 on each important feature. `**inflight wifi service**` and `**type of travel**` have a significant influence on customer satisfaction, the drop loss of these two features is 2 times greater than random forest.
- **XAI Interpretation Results using partial dependency**
  The features taken into account are almost similar for all models. However, in the RANDOM FOREST model, each increase in the value of the feature has a significant effect on the chances of not satisfied customers and satisfied.PDP in random forest has more features that have a high effect on customer satisfaction compared to LGBM and XGBOOST.
- **XAI Interpretation Results using LIME**
  Features that affect the not satisfied and not satisfied classes are **more reasonable** in the **LIGHTGBM model**. Customer type 0 or loyal customer tends to be satisfied, type of travel = 0 or business travel customers, tends to be satisfied and has a considerable significance value in both features. At each categorical or rating given to each aspect, the LIGHTGBM model provides a reasonable limit for customer satisfaction, giving a rating below 3 or 2 will cause the customer to be not satisfied.
- Based on the XAI results above, I **recommend** using the **LIGHTGBM** model to predict customer satisfaction because it considers more features that cause customers to be dissatisfied and has a more logical reason than random forest and XGBOOST.

Recommendation

1. **Customer Segmentation by Type and Loyalty Enhancement Strategy**

   Loyal customers have a high chance of being satisfied, while disloyal customers tend to be dissatisfied.

   Recommendation:

   - Loyalty Programs: Increase efforts to retain loyal customers through loyalty programs, such as special offers, priority access, or exclusive discounts.
   - New Customer Retention: Identify disloyal customers and implement better onboarding programs, such as first-trip offers

2. **Focus on Business Travel Customers**

   Business travel customers tend to be more satisfied than personal travel customers.

   Recommendation:

   - Premium Services: Improve premium services for business customers, such as lounge facilities, quick rescheduling services, or high-speed internet access on board.
   - Personalization for Personal Travelers: Identify the unique needs of personal travel customers to enhance the experience, such as with additional entertainment, destination promotions, or family packages.

3. **Key Service Improvements Based on Aspects Affecting Dissatisfaction**

   Low ratings on aspects such as inflight wifi, online boarding, leg room, cleanliness, baggage handling, inflight service, and gate location contribute greatly to unsatisfied customers.

   Recommendation:

   - Inflight Wifi Service: Improve the speed and stability of inflight wifi service by investing in technology infrastructure.
   - Online Boarding: Optimize the online boarding experience, for example using user friendly application UI UX and clearer integration of flight information.
   - Leg Room: Add seat options with more legroom for customers in competitive price.
   - Cleanliness: Improve aircraft or waiting room cleanliness with more frequent inspections.
   - Baggage Handling: Reduce complaints related to baggage handling by speeding up the process and improving communication of baggage status through the app.

Recommendation

4. Positive Feedback Improvement Program

Customers give low ratings for various services if they are unsatisfied.

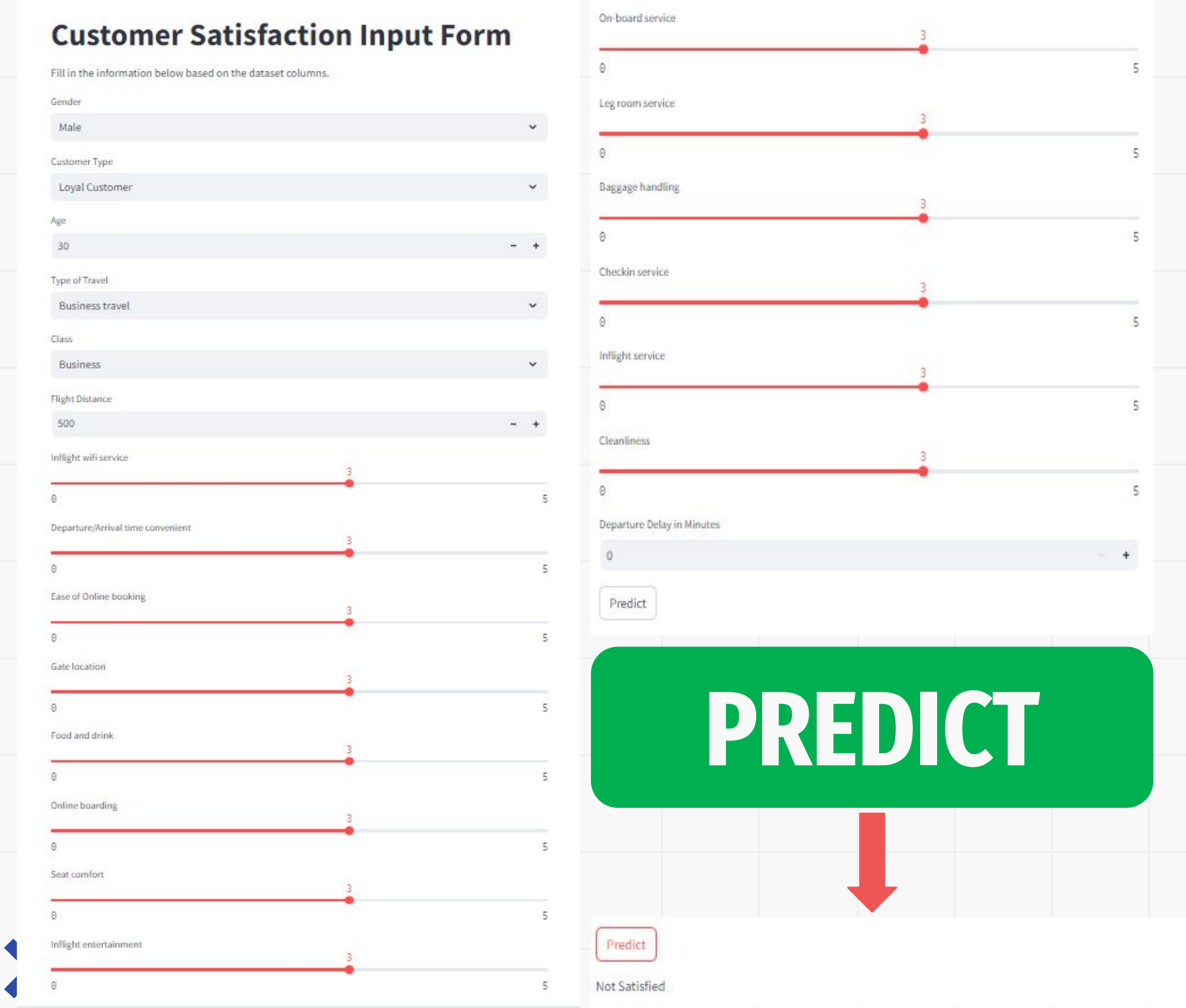
Recommendation:

- Post-flight survey: Improve the process of collecting post flight feedback with reward points for participation.
- Quick Wins: Focus on improving areas that received low ratings in previous surveys to make an immediate impact on customer satisfaction.

06

Deployment

Deploy Using Streamlit



Streamlit

- Here is a preview of the application built to classify airlines customer satisfaction
- Perform **imputation** on several features, then predict what the results will be.
- This application makes predictions by using the **streamlit** app with **ngrock** tunnel. Can be accessed at the following link

https://afa6-34-30-3-199.ngrok-free.app

THANK YOU