Airplane Satisfiction prediction

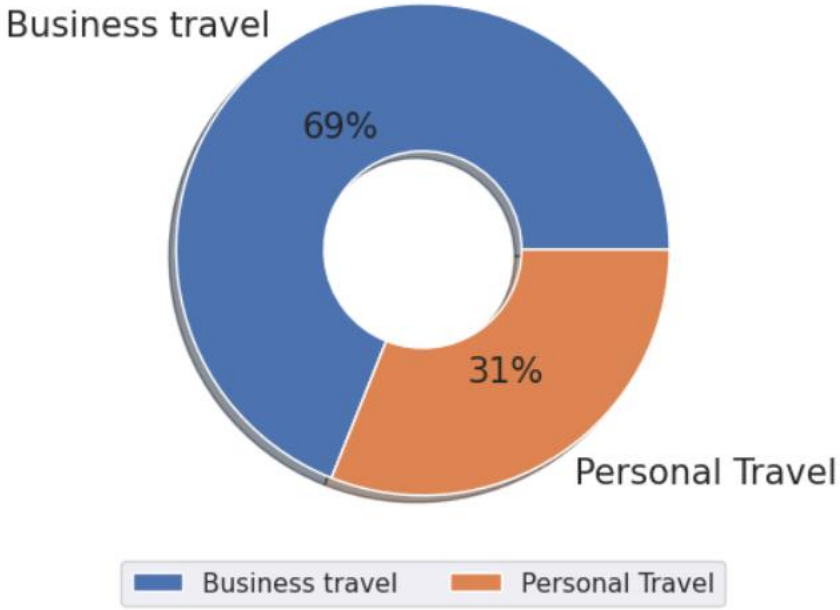
MACHIN LEARNING PROJECT BY: DIAA YOUSEF

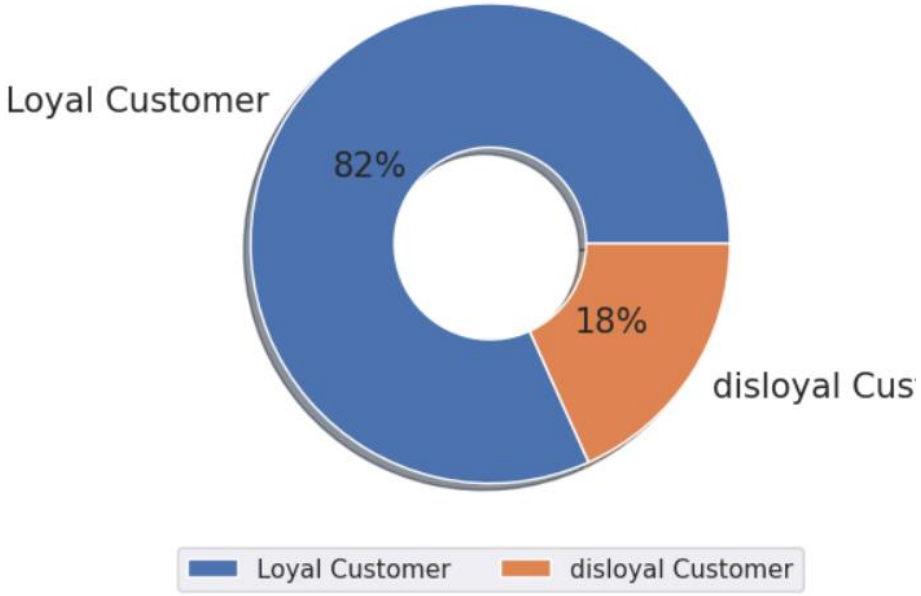
About Dataset

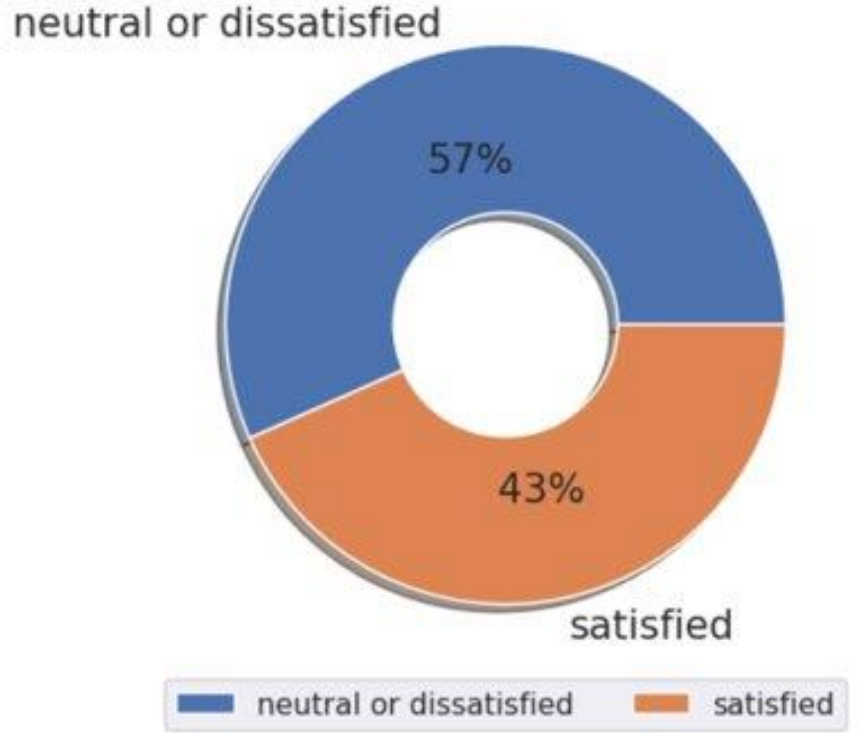
103904 ,24

Context This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger? Can you predict passenger satisfaction

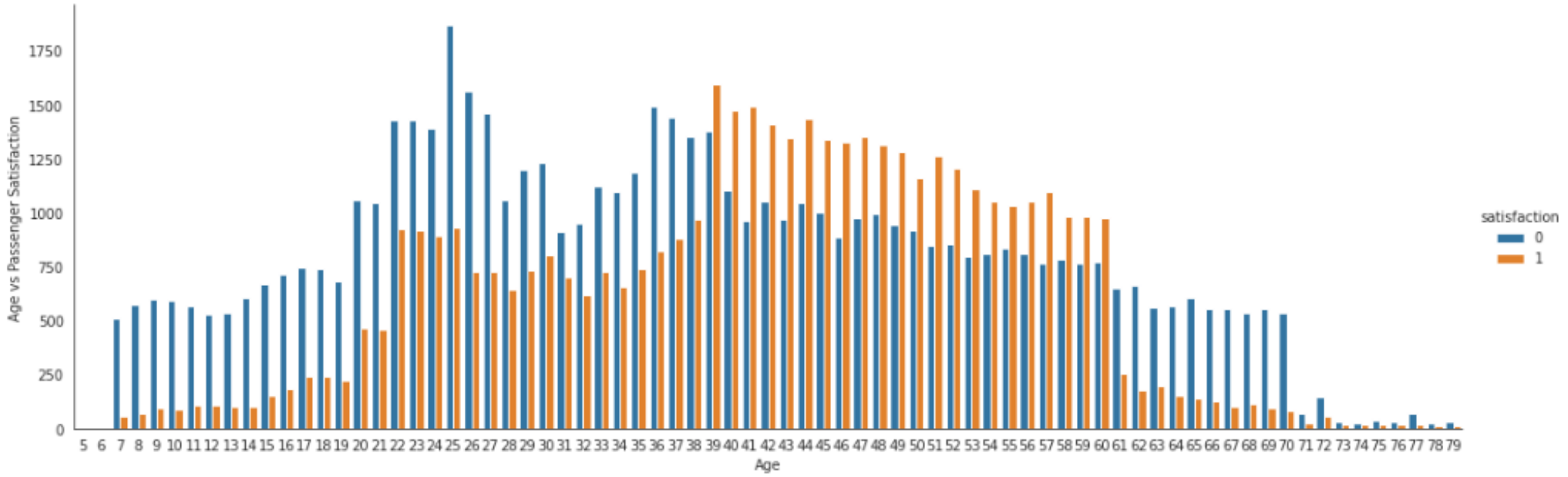
Exploratory Data Analysis



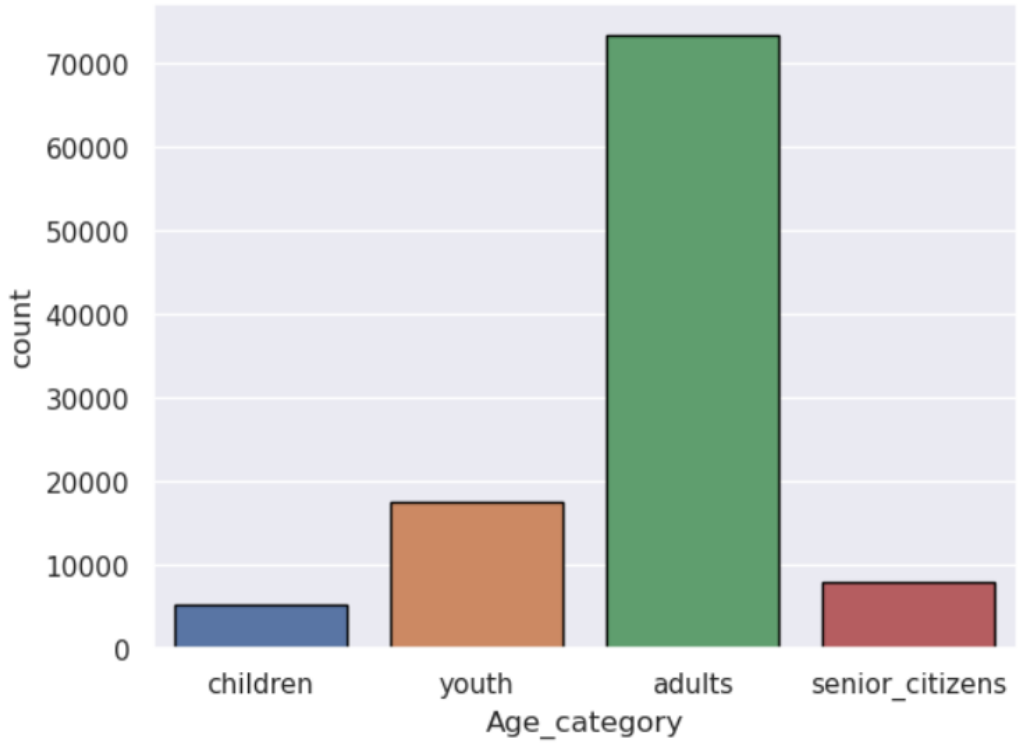




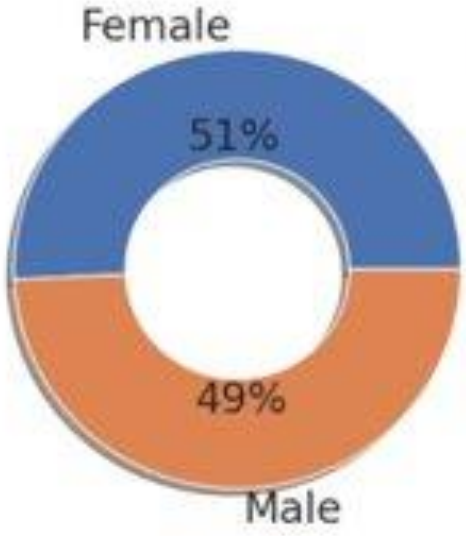
Age vs Satisfaction

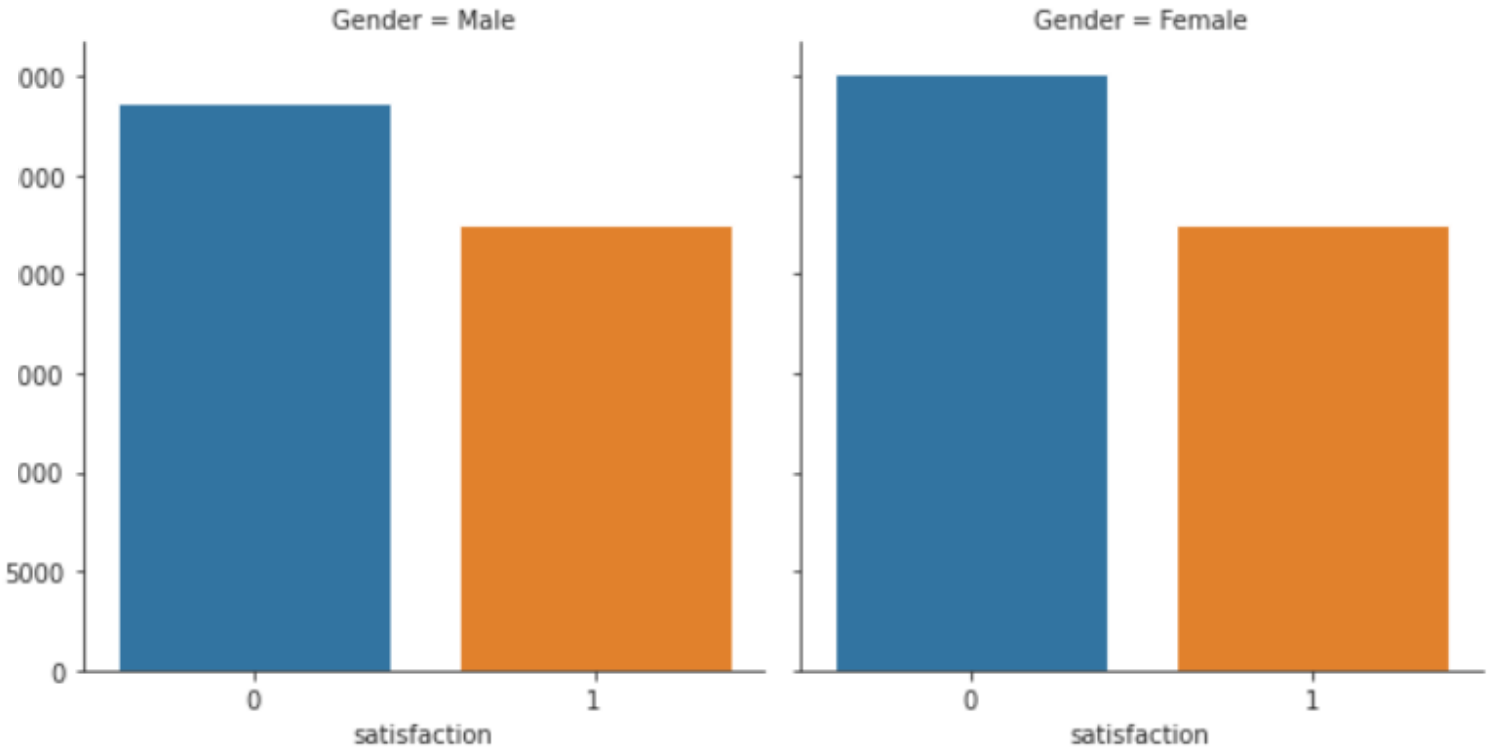


- Adult were the most Travellers which is between 25 to 40

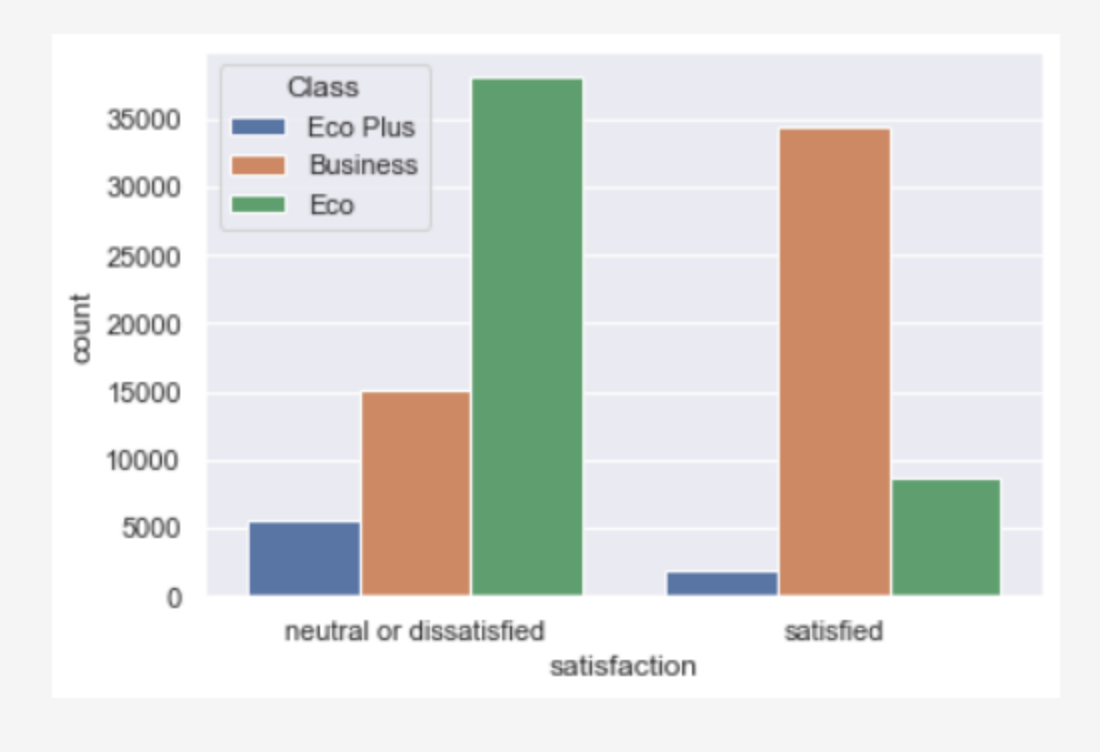


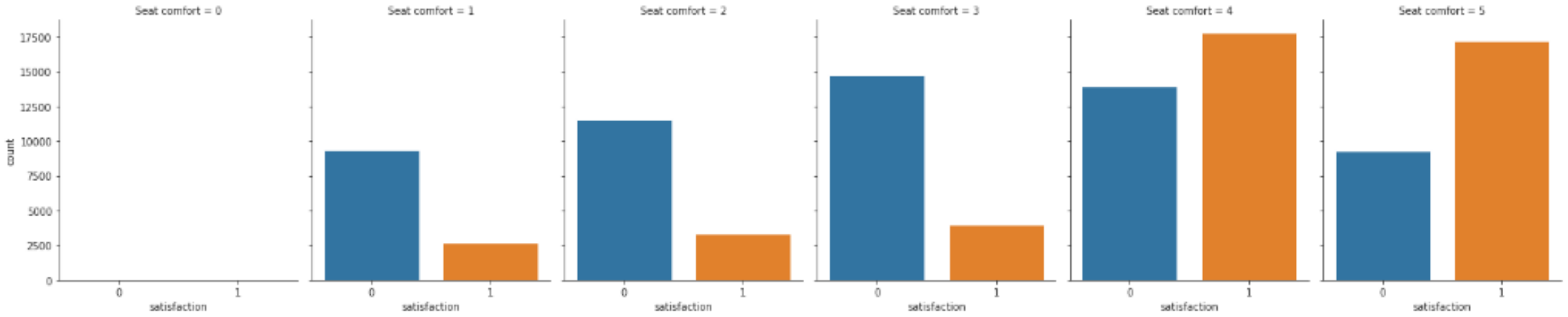
Gender vs Satisfaction



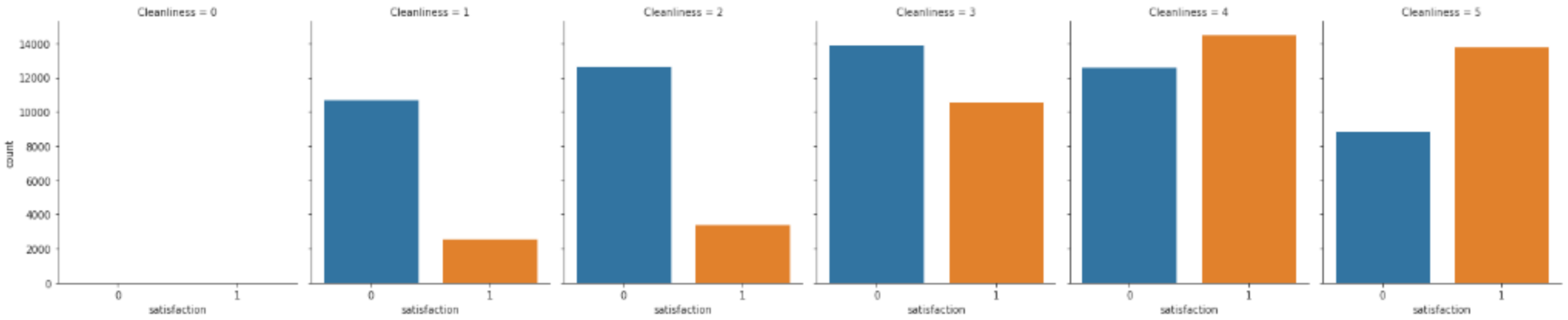


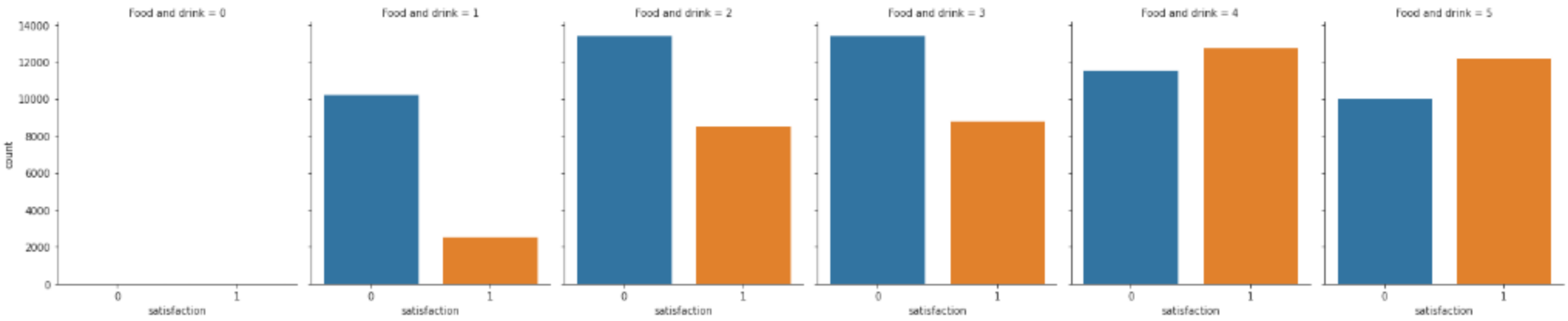
Class vs Satisfaction





Some services rating to Satisfiction


Cleanliness = 0
Cleanliness = 1
Cleanliness = 2
Cleanliness = 3
Cleanliness = 4
Cleanliness = 5
count
0
2000
4000
6000
8000
10000
12000
14000
0
1
satisfaction


Food and drink = 0
Food and drink = 1
Food and drink = 2
Food and drink = 3
Food and drink = 4
Food and drink = 5
count
14000
12000
10000
8000
6000
4000
2000
0
0
1
satisfaction

Insights

Female and male were equal travellers, And Loyal Customers were more

Mostly Business travels were done

Most of the people use Business and Eco class

Mostly Customers are neutal and Dissatisfied with the service

The Average rating of 3 and 4 were the most given by both males and females

Arrival delay and Departure delay are highly correlated

Adult were the most Travellers which is between 25 to 40

Preprocessing

Filling missing values with median (numerical data)

Handle outliers by Datasist

Split data into input and output

Dealing with categorical data (pandas get dummies)

Split into train and test

Feature selection

Feature scaling

Handle imbalance

Modeling

Model selection

	Train acc	Test acc
RandomForest	1.00	0.96
LogisticRegression	0.87	0.88
knn	0.95	0.92
naive_bayes	0.86	0.86

Cross validation by Kfold k=5

Deployment

Using streamlit

Heroku server