

Airways Passengers Satisfaction Analysis

BADOUR BIN SALAMAH

Overview



Introduction

- Data analytics and machine learning are playing very essential role in setting up organization's strategies and directions.
- Moreover, such analyzing datasets can be very helpful and supportive for many different organizations such as airways, airports and loyalty programs providers.

Dataset description

- The dataset that will be used in this work is **Passenger Satisfaction** found in **kaggle**.
- Contains 24 columns and 129880 rows

Data Cleaning

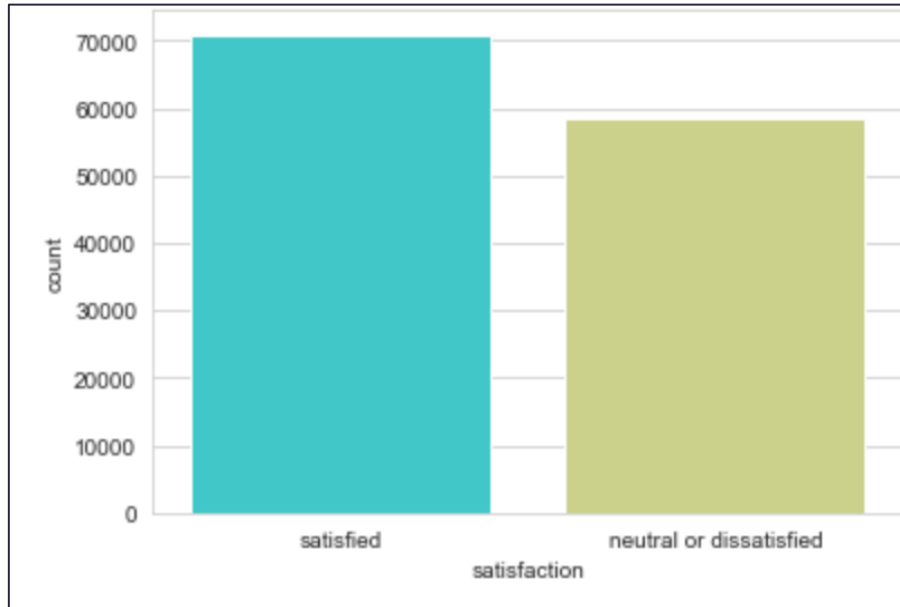
Handling missing data

- Drop rows contain null values.

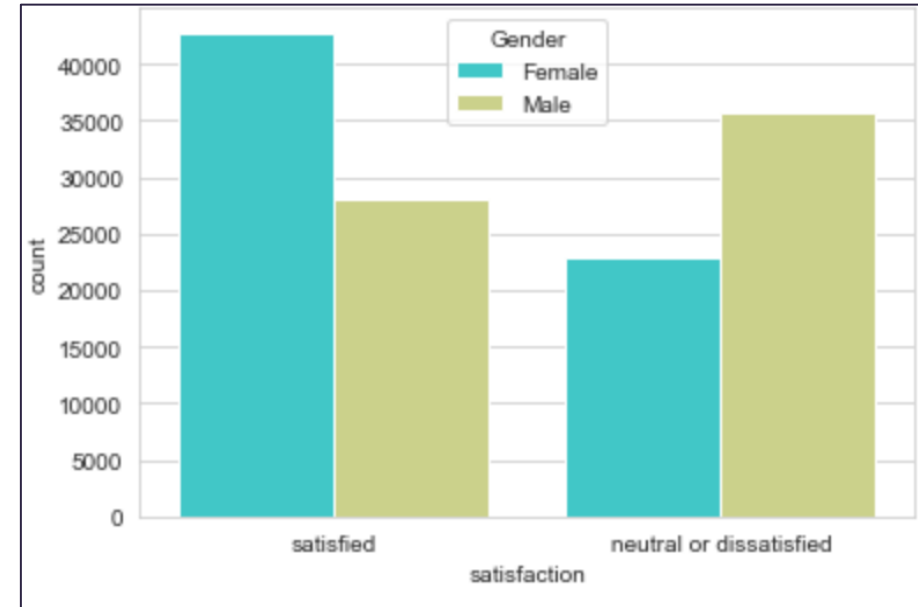
Filtering out data outliers

- Check data validity

Exploratory Data Analysis (EDA)

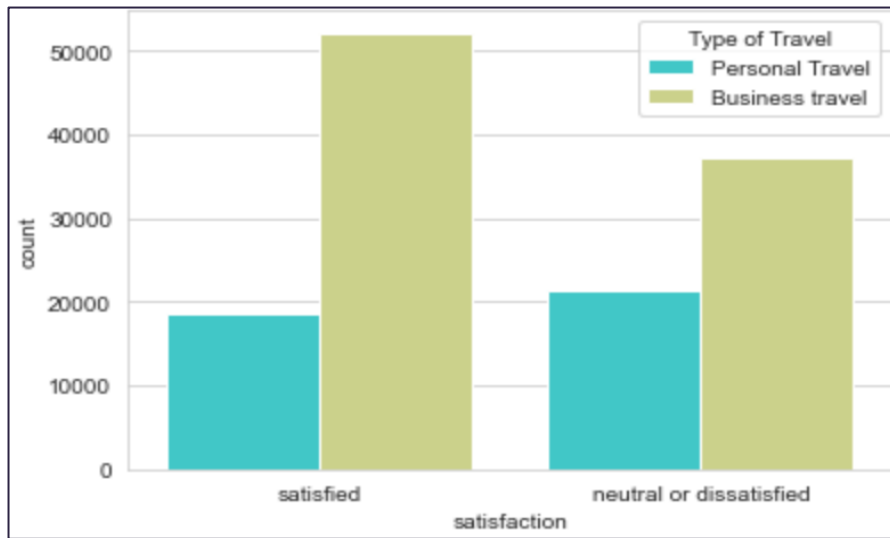


Count of
satisfied / dissatisfied passengers

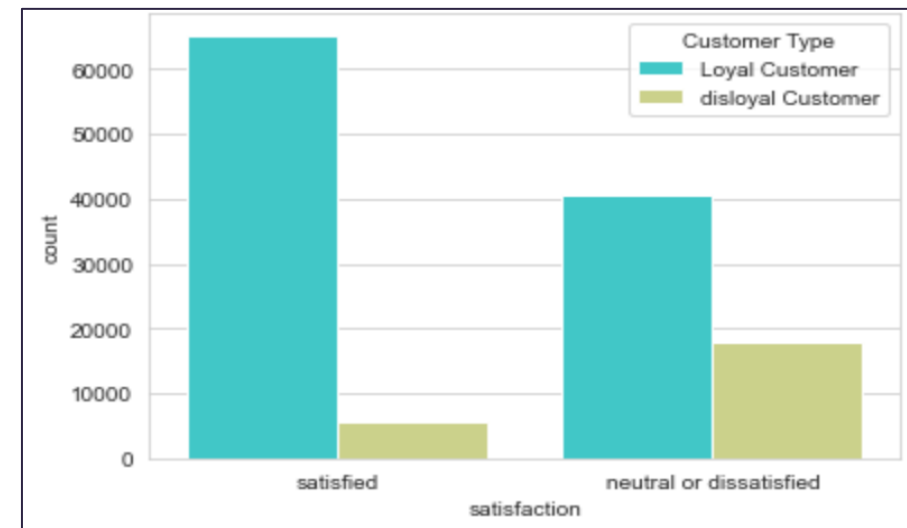


Correlation between
Satisfaction / Gender

Exploratory Data Analysis (EDA)

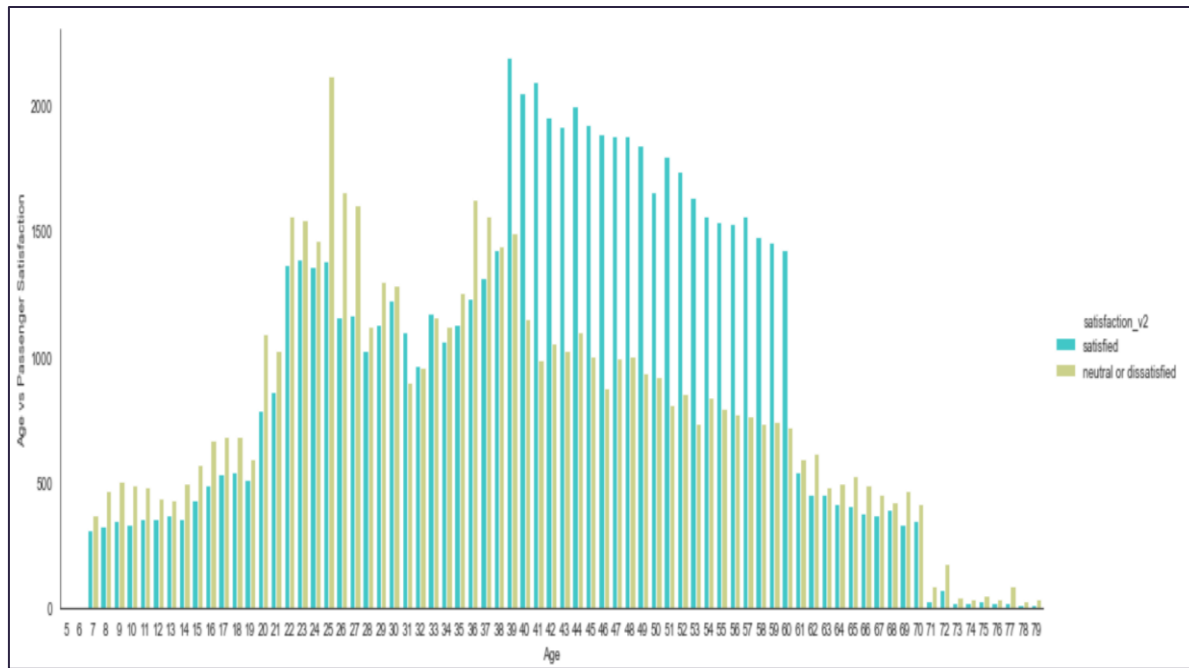


Correlation between
Satisfaction / Type of travel

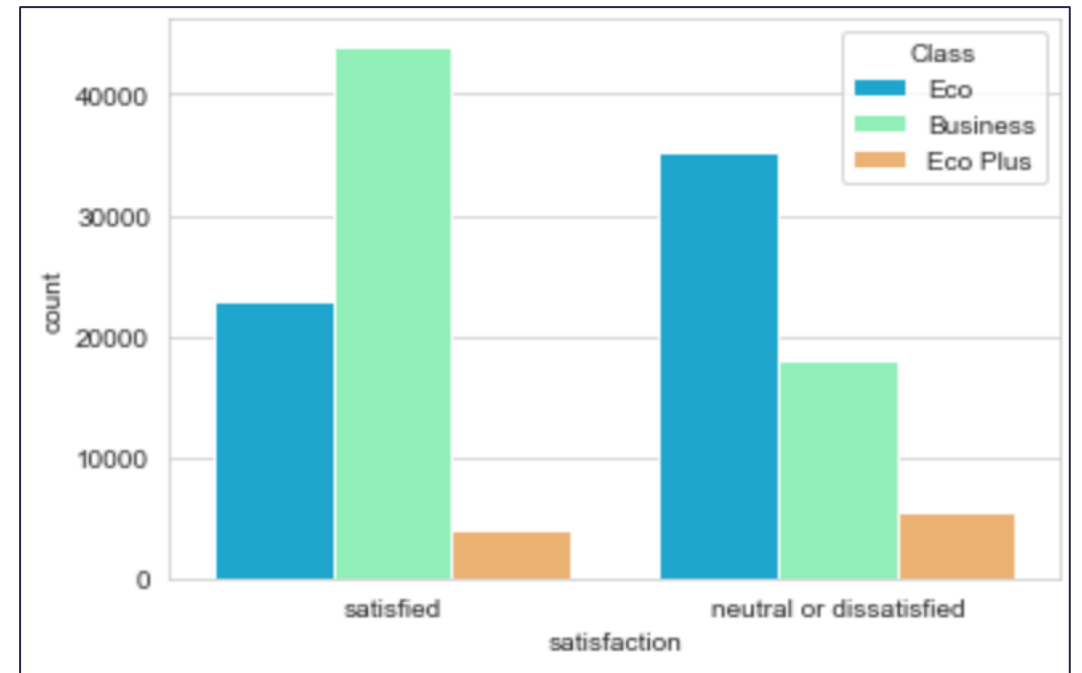


Correlation between
Satisfaction / Customer Type

Exploratory Data Analysis (EDA)



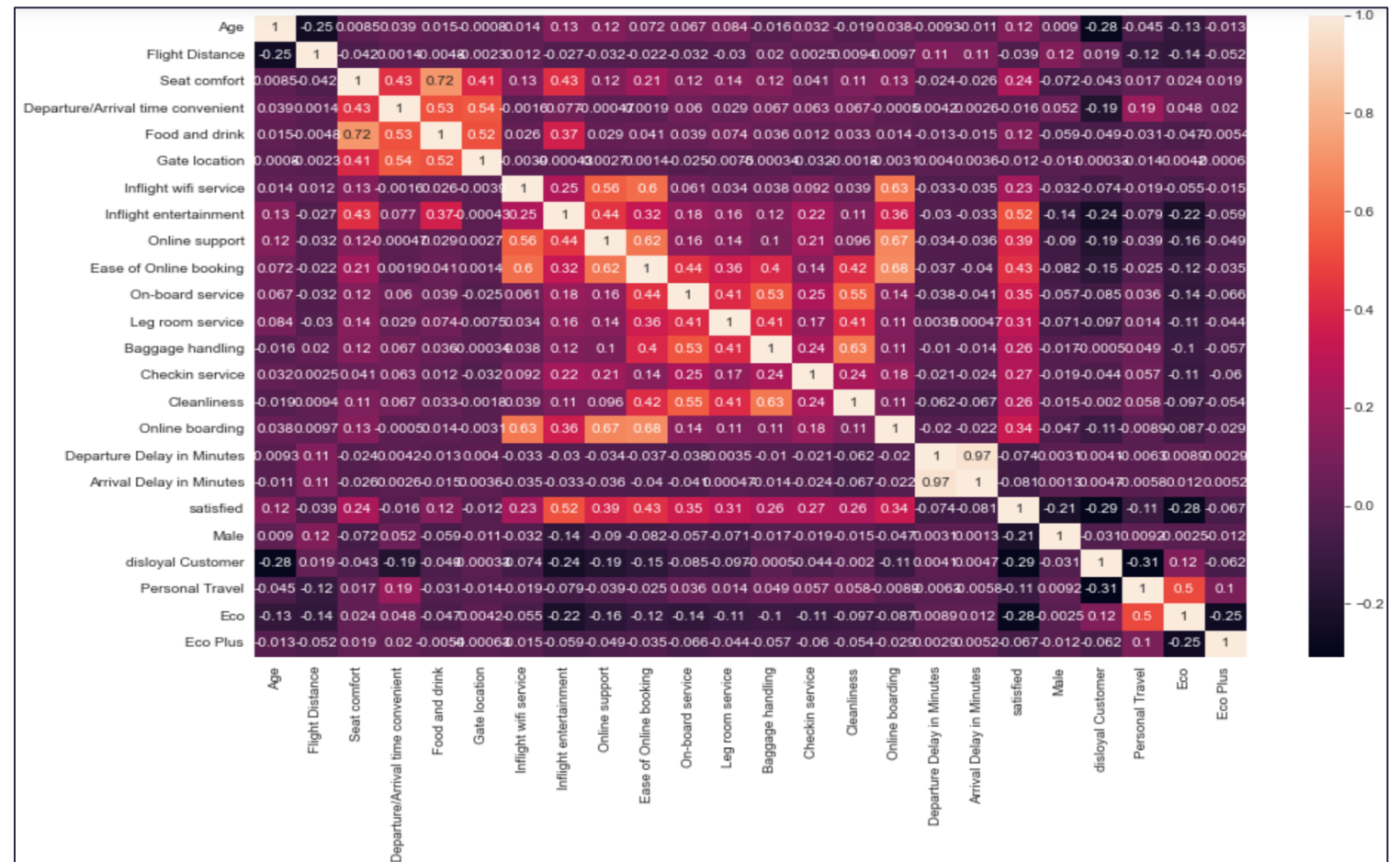
Correlation between
Satisfaction / Age



Correlation between
Satisfaction / Class

Exploratory Data Analysis (EDA)

Correlations between features



Modeling

Started with

Logistic Regression Model

- A logistic regression (LR) model was used and the findings were as following:

Test accuracy	0.84
Train accuracy	0.83
RMSE	0.39
Precision	0.84
Recall	0.84
F1	0.84

KFold

- By applying Kfold using 5 splits the accuracy was 0.83

Then

Random Forest Model

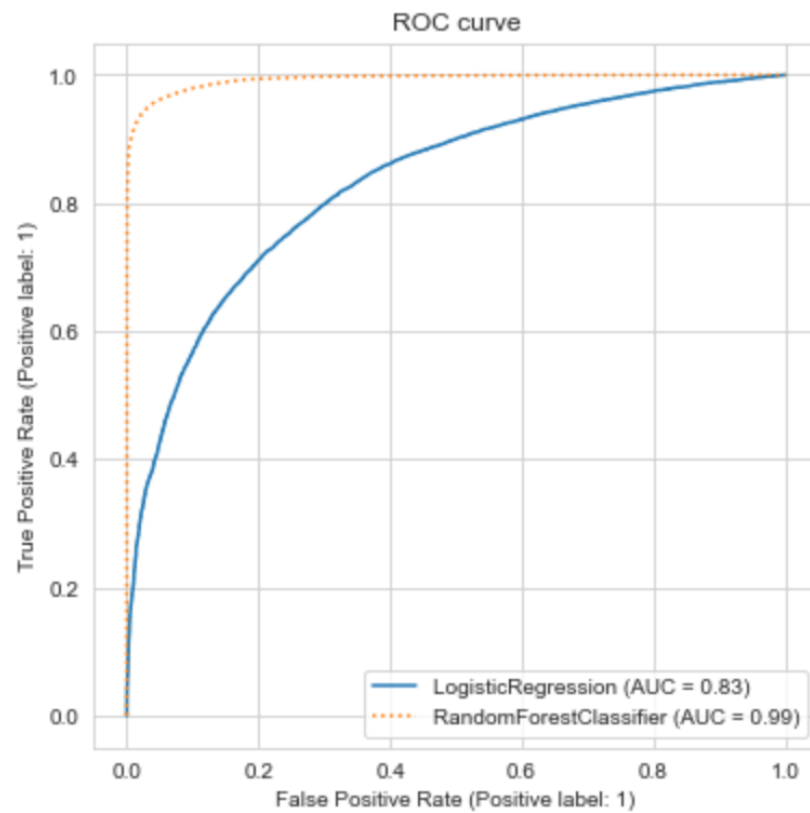
- A Random Forest (RF) model was used and the findings were as following:

Test accuracy	0.95
Train accuracy	0.99
RMSE	0.21
Precision	0.95
Recall	0.95
F1	0.95

KFold

- By applying Kfold using 5 splits the accuracy was 0.95

Conclusion



Thank you