AIRLINE PASSENGER SATISFACTION

Business Understanding

Business Overview

A major key to the success of any business is the satisfaction of its customers. For an airline, having passengers unsatisfied with its services means decreased revenue, as the passengers are unlikely to return to the airline. Additionally, having bad reviews hurts the reputation of an airline, deterring new customers from employing its services. Customer satisfaction is therefore essential in the airline industry. It is affected by how the airline's overall services are, right from booking a ticket with them to the various conditions during the flight, factors also highlighted in this article by Perficient. Being able to determine passenger satisfaction levels, as well as the top factors that influence this, will aid airlines in improving their services.

Business Objective

The main objective of this report is to build a model that predicts a passenger's satisfaction with an airline.

Business Success Criteria

To do a thorough analysis and build a model to predicts whether a passenger's satisfied, dissatisfied or neutral with the services offered by the airline.

Assessing the Situation

1. **Resource Inventory**

a. Datasets:
   - https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction?select=test.csv

b. Software used (Github, R-Studio, Google Docs, IBM® SPSS® Modeler)

c. Resources available for reference is found (here)

2. **Assumptions**

- The data provided is correct and up to date.

- All my data is relevant for analysis.
- There are no existing typos in my dataset.

3. Constraints

There are no constraints

Data Mining Goals

Our data mining goals for this project are as follows:

- Determine the relationships between overall level of satisfaction and various predictor variables (such as gender, travel class, inflight wifi service, etc).
- Determine the correlation between overall level of satisfaction and the predictor variables.
- Determine the top factors affecting satisfaction with an airline.
- Build a model that can predict a passenger's overall satisfaction with an airline.

Data Mining Success Criteria

Our success criteria will be measured by the following criteria;

- Create a clear picture of how close the company is to its business goals and strategies to ensure overall business success and growth.
- Easy application of the analytics results to gain substantial and sustainable benefits for the company.
- Customer satisfaction, brand awareness or customer engagement to help evaluate the efficiency of the use and consumption of the product(Airlines).

Data Understanding

Data Understanding Overview

For this project, we are using the availed dataset. Data was extracted from Kaggle, where the availability information was available in real-time. The information of the scheduled departures is highlighted below.

<u>Different Feature Names</u>

1. **Index** column
2. **Id**
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
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 Description

We have one dataset available for this project. A detailed description of the datasets is provided as follows:

- This dataset consists of 25976 rows instances and 25 columns. The data contains the following fields:
    1. **Index**
    2. **Id**
    3. **Gender**
    4. **Customer Type**
    5. **Age**
    6. **Type of Travel**
    7. **Class**
    8. **Flight distance**
    9. **Inflight wifi service**
    10. **Departure/Arrival time convenient**

11. **Ease of Online booking**
12. **Gate location**
13. **Food and drink**
14. **Online boarding**
15. **Seat comfort**
16. **Inflight entertainment**
17. **On-board service**
18. **Leg room service**
19. **Baggage handling**
20. **Check-in service**
21. **Inflight service**
22. **Cleanliness**
23. **Departure Delay in Minutes**
24. **Arrival Delay in Minutes**
25. **Satisfaction**

Verifying Data Quality

There were 83 missing values in the arrival delay in the minutes column. Given that the dataset has 25976 rows, the rows with missing values in this column will be dropped.

Our dataset had no duplicates as well.

Data Preparation

These are the steps followed in preparing the data

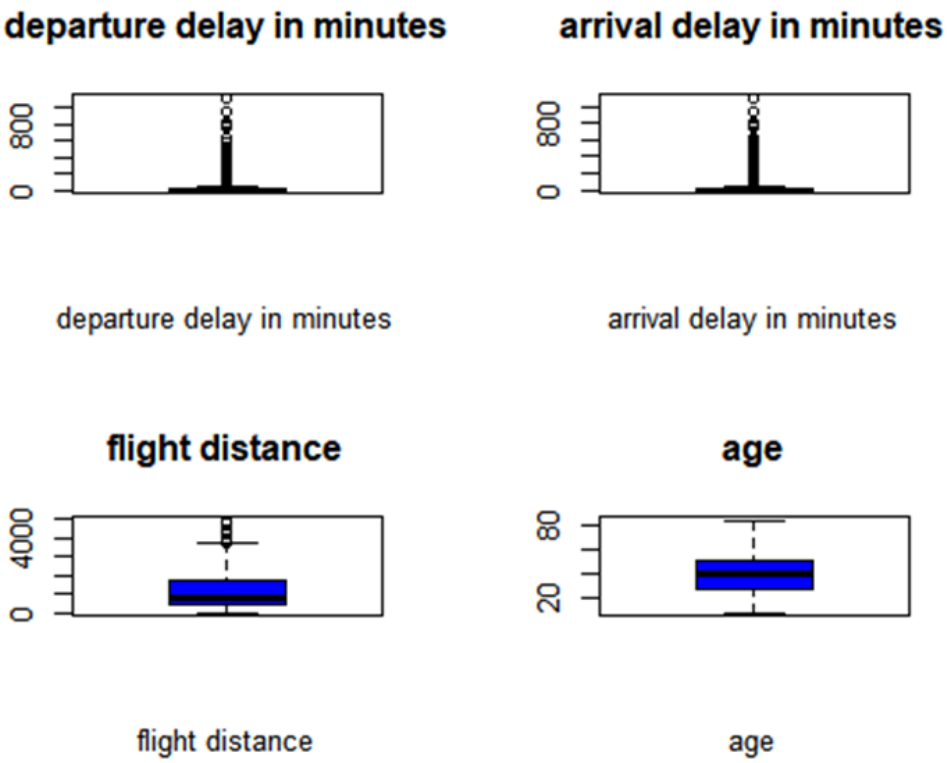
Loading Data

Loading the dataset from the CSV file and then creating a dataframe to be used.

Cleaning Data

- We converted column names to lowercase from uppercase
- We dropped column X because it is an index column. Similarly, the id column just consist of unique identifications for each entry so it will be dropped too
- There were 83 missing values in the arrival delay in minutes column. Given that the dataset has 25976 rows, the rows with missing values in this column will be dropped.

- There were outliers in the departure delay in minutes,arrival delay in minutes, and flight distance columns. They will not be dropped because it is possible for the outlier delays and distances recorded to occur.



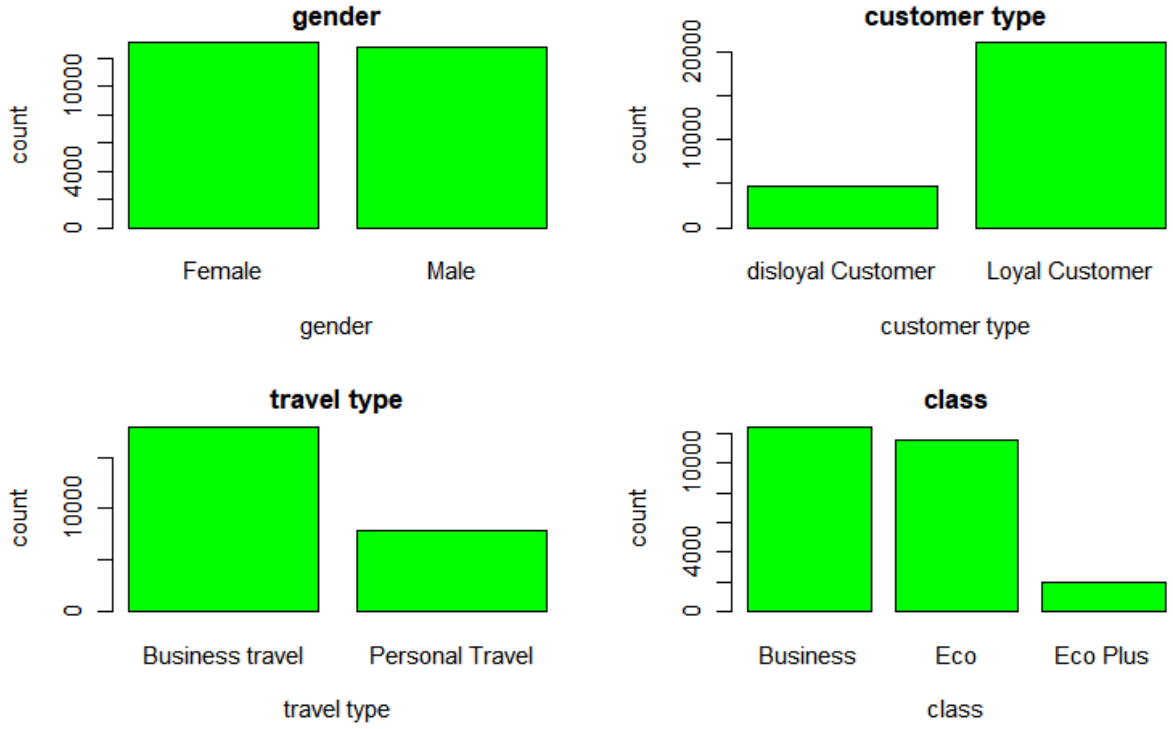
- No anomalous values observed in the data

Analysis

During our analysis, the following was covered;

1. Univariate
2. Bivariate

<u>Univariate</u>

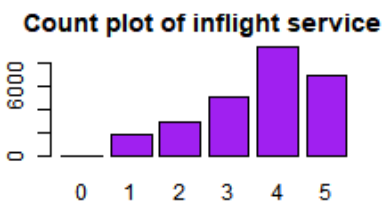


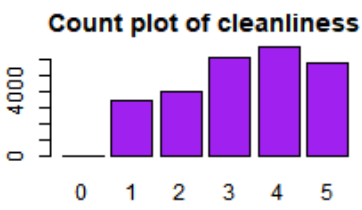
Most travels were for business purposes

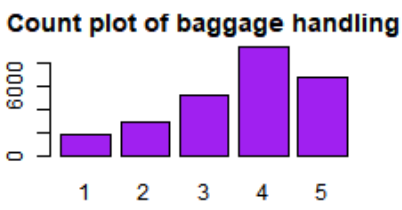
Business class most frequent, closely followed by economy

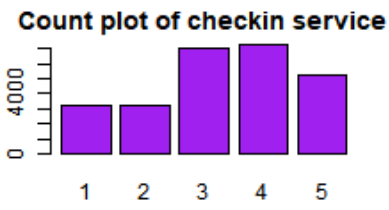
Slightly more female than male passengers

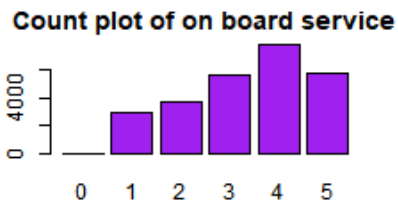
Most passengers were loyal customer type

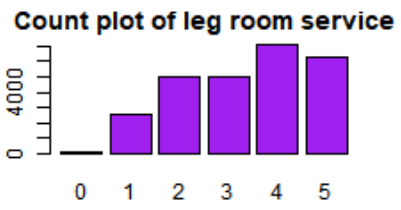


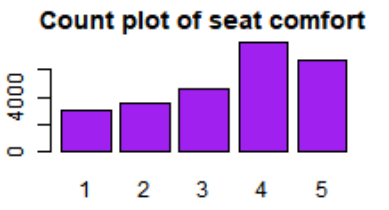


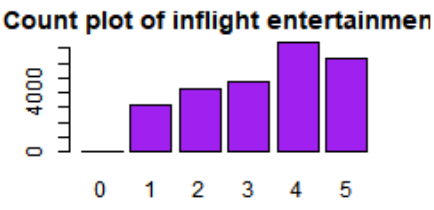






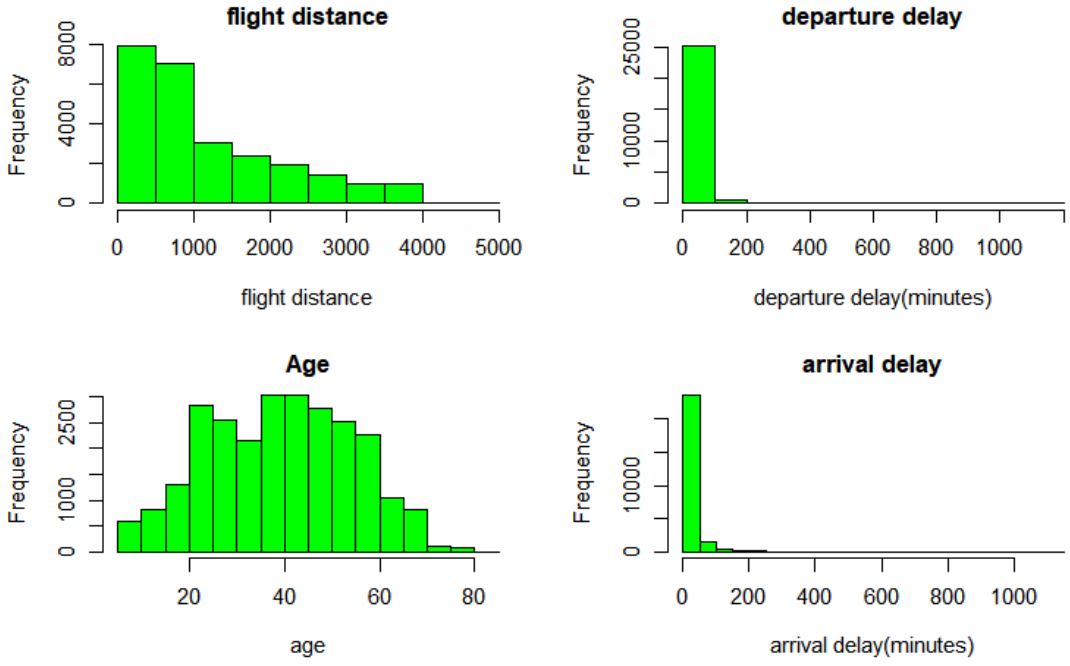






For each of the features above (inflight service, cleanliness, baggage handling, checkin service, on board service, leg room service, seat comfort, and in flight entertainment), the most frequent rating was 4.

There were more neutral/dissatisfied customers than satisfied passengers.

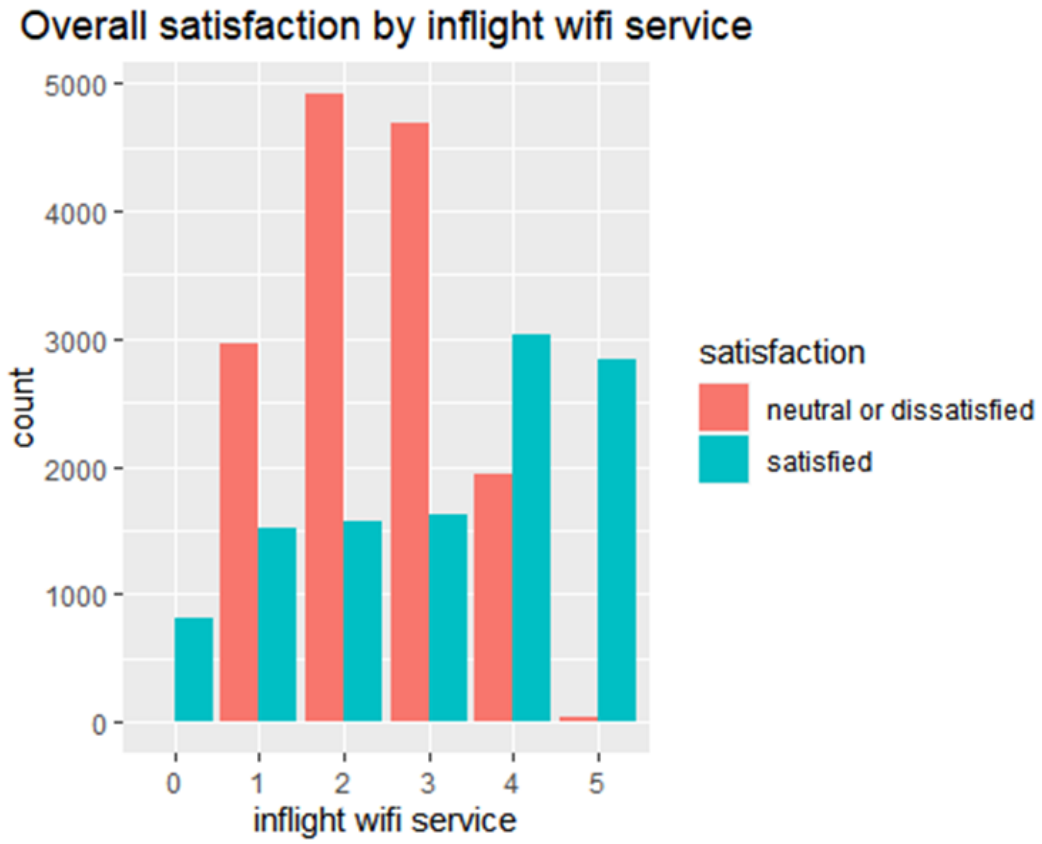


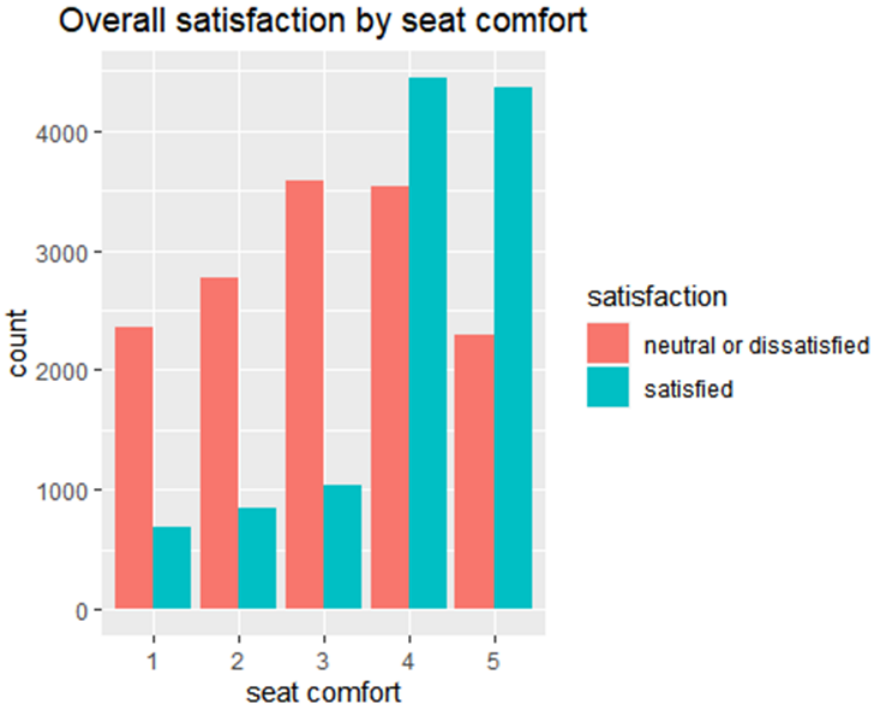
Most delays were less than 100 minutes

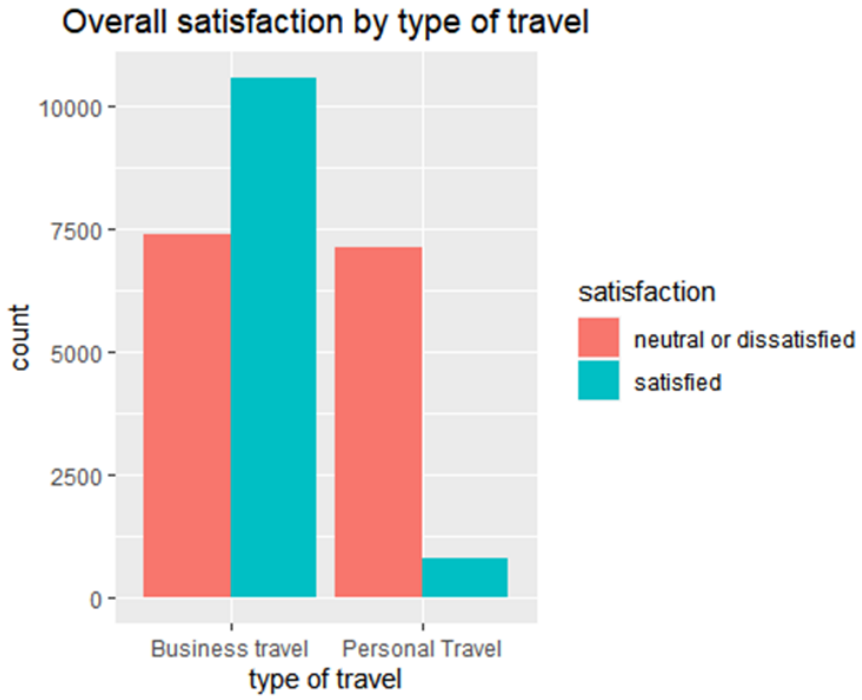
Most flight distances fell in the 0 to 500 bin.

Most passengers were between 20 and 60 years old

Bivariate


Overall satisfaction by inflight wifi service
5000
4000
3000
2000
1000
0
count
0
1
2
3
4
5
inflight wifi service
satisfaction
neutral or dissatisfied
satisfied

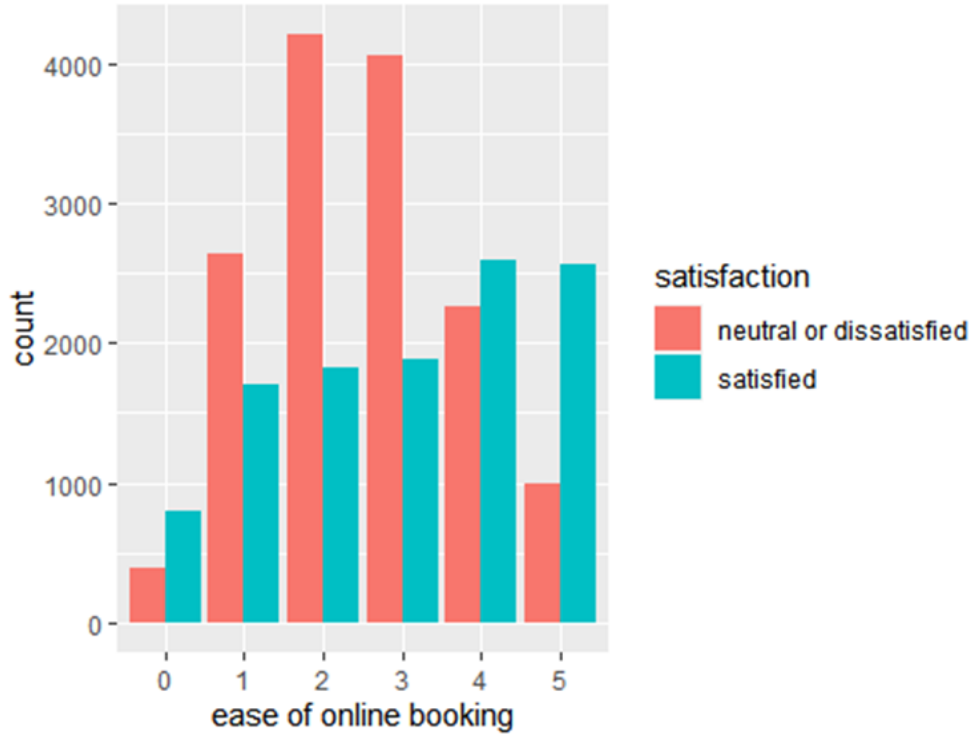

Overall satisfaction by seat comfort
4000
3000
2000
1000
0
count
1
2
3
4
5
seat comfort
satisfaction
neutral or dissatisfied
satisfied



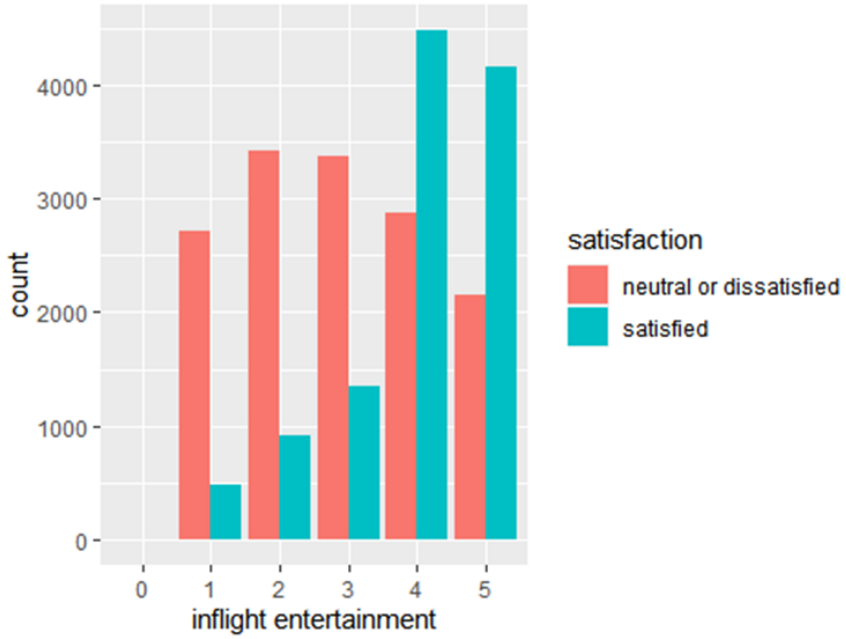
The proportion of those whose overall satisfaction level was "satisfied" was higher than that of neutral/dissatisfied among those who gave online boarding, inflight wifi service, and seat comfort ratings of 4 or 5.

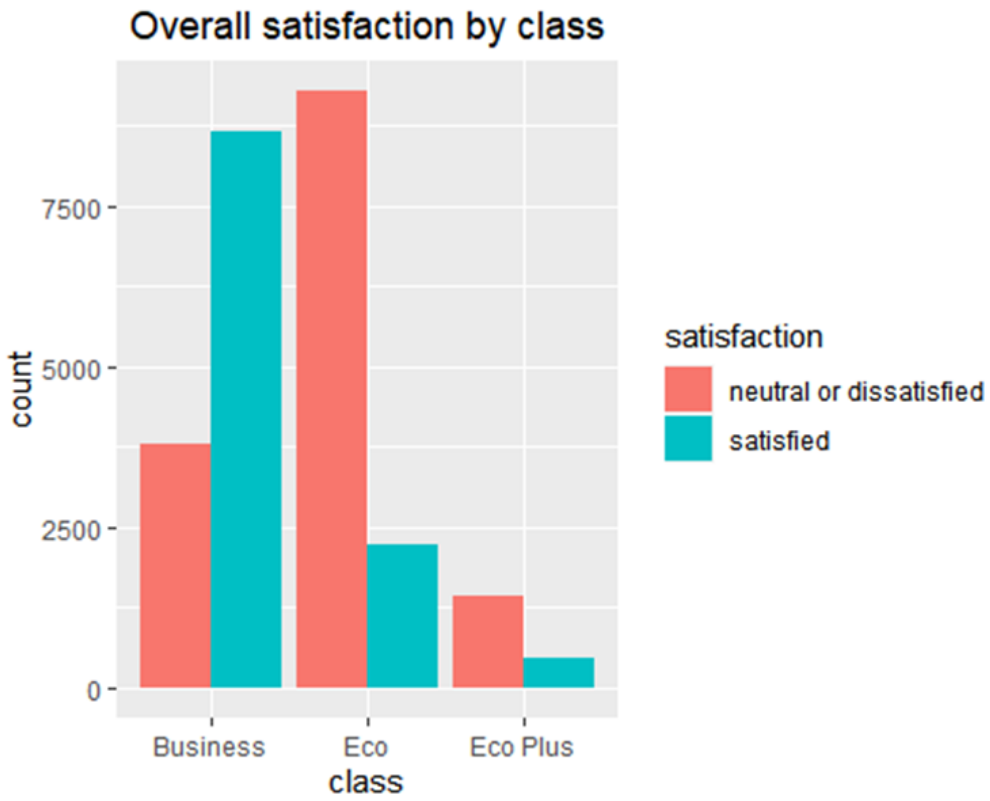
Most people travelling for business purposes were overall satisfied with the airline while most personal travellers were dissatisfied or neutral.

Overall satisfaction by ease of online booking



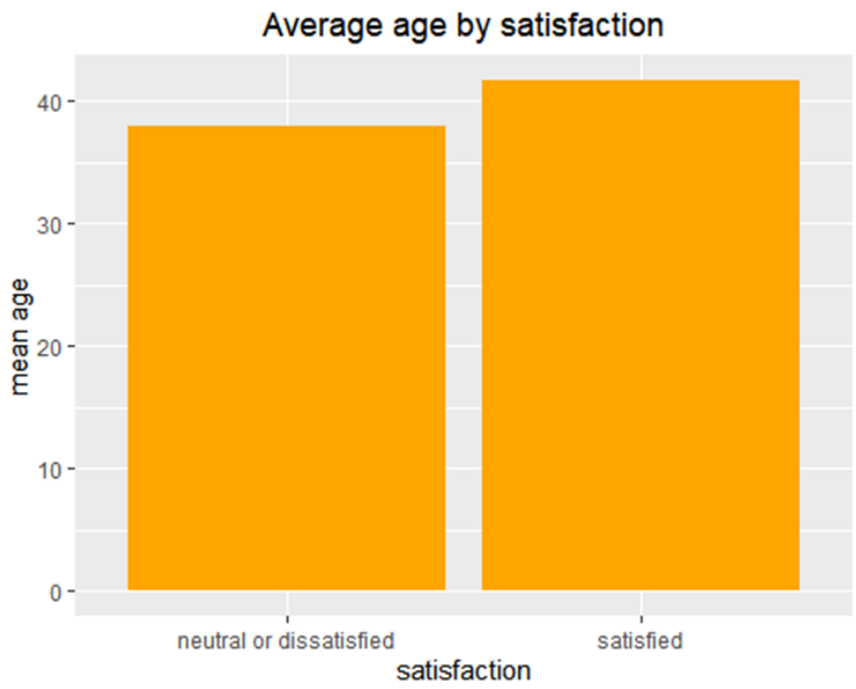
Overall satisfaction by inflight entertainment

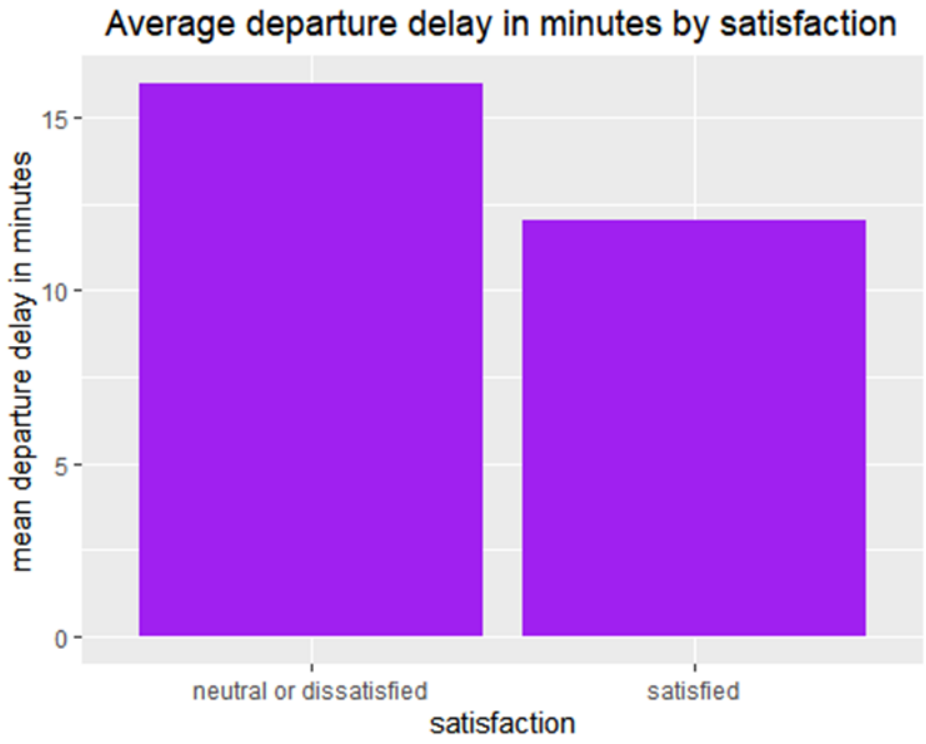




The proportion of those whose overall satisfaction level was "satisfied" was higher than that of neutral/dissatisfied among those who gave inflight entertainment and ease of online booking ratings of 4 or 5.

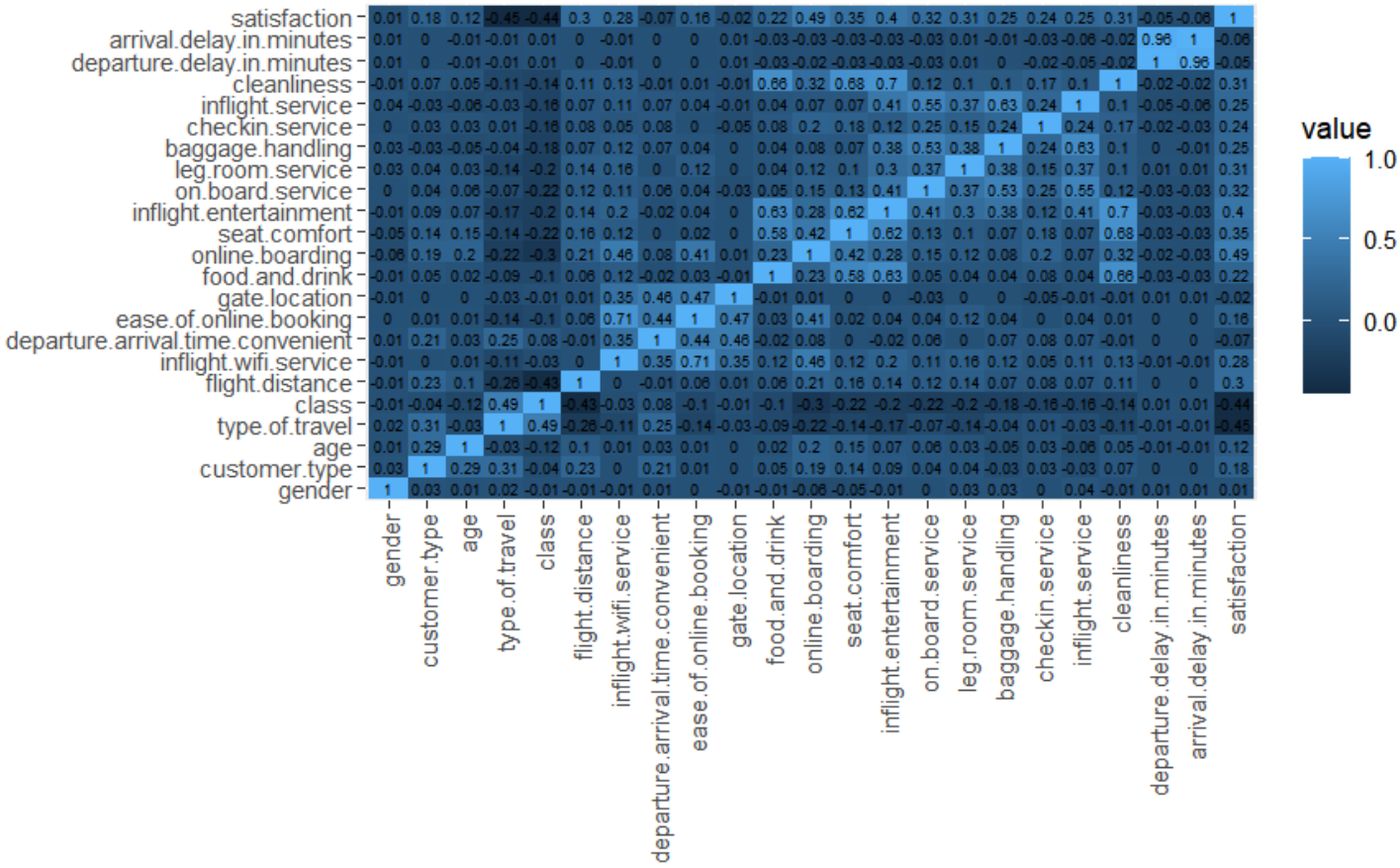
Most in business class were satisfied with the airline.





Those satisfied was higher. The mean departure delay in minutes experienced was higher among neutral/dissatisfied customers than satisfied customers.

Correlation heatmap



According to the matrix, satisfaction appears most correlated to online boarding, type of travel, and class. Modelling will reveal more on the relationships between the predictors and level of passenger satisfaction.

Modeling

The models tested were built applying 4 types of classifiers:

- Decision trees (rpart)
- Random forests (ranger)
- Support vector machines
- K-Nearest-Neighbors

The main metric that was being optimized was the f1-score, because a good balance between precision and recall was being sought, that is, minimising both false positives

and negatives. Feature scaling was done before using KNN and SVM. Train-test split of 70-30.

Decision trees:

1. The first model gave an F1 score of 0.8831
2. Tuning the cp (complexity parameter), f1 improved to 0.9139

Random forests:

1. The first model - mtry and splitrule were tuned. F1 of 0.9529 was obtained
2. The two above and min node size additionally were tuned, giving F1 of 0.9536-benchmark

K-nearest neighbours:

1. Initially, an arbitrary odd value (because we have 2 classes) of k as 5 was used. F1 was 0.9052.
2. Using k as the square root of training sample size, f1 was 0.8774
3. Tuning k further resulted in identification of k as 9. That model gave F1 of 0.9059

SVM

1. Linear kernel - Tuning the c parameter. Model gave F1 of 0.8583
2. Polynomial kernel - tuning degree of model (3). Model gave F1-0.9402
3. Rbf kernel - tuning c, sigma constant. F1-0.9362

Best model - random forest with mtry =12, split rule =extra trees, min.node.size =10: F1 0.9536

Conclusion

The main objectives of the project were achieved:

a. A model that can predict a passenger's level of satisfaction with high accuracy was built (Best model - a random forest model giving F1 - 0.9536, as well as high values for other metrics (precision : 0.9641, recall: 0.9432,accuracy: 0.9596"))
b. The top factors affecting a passenger's overall satisfaction level with an airline were identified: (highlighting top 10 respectively) inflight wifi service, online boarding, type of travel, class, customer type, inflight entertainment, seat comfort, ease of online booking, baggage handling, and checkin service.
c. The relationships between overall level of satisfaction and various predictor variables were determined. Some insights:

- The proportion of those whose overall satisfaction level was "satisfied" was higher than that of neutral/dissatisfied among those who gave online boarding,

inflight wifi service, seat comfort, inflight entertainment and ease of online booking ratings of 4 or 5.
- Most people travelling for business purposes were overall satisfied with the airline while most personal travellers were dissatisfied or neutral.
- Most in business class were satisfied with the airline as opposed to economy and economy plus where most were dissatified/neutral.

Recommendations

Recommendations based on some of the top features:

- Airlines should improve the quality and availability of their in flight wifi service to improve overall passenger satisfaction with their airline.
- Customer type. Advertising and encouraging passengers to join loyalty programs and experience perks such as gaining redeemable miles, deals with travel destination accommodation etc, will be beneficial because these programs are likely to increase passenger satisfaction with the airline overall. Customer retention also boosts revenue.
- Online boarding. Checking in online allows passengers to check in in advance, choose their seat, and print their boarding pass. They can therefore avoid the long queues in the airport. A smooth process in doing this improves the passenger experience.
- Airlines should diversify the available selections on their inflight entertainment platforms, such as movies, music, etc to increase the likelihood of a passenger finding something enjoyable.
- Airlines should upgrade the quality of their seats when necessary (for example those that are hard, worn and torn) to maximize on improving comfort for the passengers.
- Ease of online booking was also determined to be an important factor. Airlines should ensure that their online platforms are user-friendly, straightforward and up-to-date when it comes to booking.
- Baggage handling is also a top 10 factor in determining passenger satisfaction. Airlines should ensure that their systems are well-coordinated to avoid cases of delays/loss/damage to baggage.
- The staff at the physical airline check in stations in airports should have good customer service skills, as the impression they make affects overall passenger satisfaction with the airline.