PROJECT REPORT

Analysis of Airline Passenger Satisfaction using R

Setting:

This dataset encompasses information regarding airline passenger satisfaction, capturing diverse aspects of the travel experience. It includes data from various airlines, covering factors such as flight details, customer demographics, service quality perceptions, and overall satisfaction ratings. The dataset consists of structured information obtained through surveys, with attributes like flight distance, departure/arrival delays, seat comfort, onboard services, and customer feedback. Additionally, it incorporates categorical variables reflecting passenger characteristics such as age, gender, travel class, and frequent flyer status. This comprehensive dataset offers valuable insights into the factors influencing passenger satisfaction within the aviation industry, aiding in the analysis of trends, identification of key drivers, and formulation of strategies for enhancing the travel experience.

Problem Statement:

The challenge faced by airlines is to effectively analyze the vast amount of data generated from various touchpoints in the passenger journey to gain actionable insights that drive business decisions. Specifically, airlines aim to understand the factors that contribute most significantly to passenger satisfaction and dissatisfaction. By identifying these factors, airlines can prioritize resources and initiatives to address pain points and enhance the overall travel experience.

Business Problems:

1. **Identifying Key Drivers of Passenger Satisfaction Drivers:** Analyzing the dataset to identify the key factors influencing passenger satisfaction, including flight details, service quality, and customer demographics. This involves exploring correlations between different variables and overall satisfaction ratings to pinpoint areas where improvements are needed.

2. **Predicting Passenger Satisfaction Levels:** Developing machine learning models to predict passenger satisfaction ratings based on relevant features such as flight distance, delays, seat comfort, and onboard services.

3. **Formulating Enhancement Strategies:** Based on the insights gained from data analysis and machine learning models, formulating targeted strategies to improve satisfaction levels, address pain points, and enhance the overall travel experience.

The questions we are trying to answer are:

1. What factors contribute most to passenger satisfaction?

Identify the key features (e.g., seat comfort, in-flight entertainment, customer service) that significantly influence passenger satisfaction ratings.

2. How do flight-related factors impact satisfaction?

Analyze the effects of flight duration, departure/arrival delays, and flight class on passenger satisfaction levels.

3. How does inflight service impact passenger satisfaction?

Analyze the effects of Inflight Wifi Service, Food and Drink,Seat Comfort, Inflight entertainment, Leg room service, Cleanliness.

Data Description:

The dataset used for this analysis was obtained from Kaggle (https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data).
This dataset encompasses information regarding airline passenger satisfaction, capturing diverse aspects of the travel experience. It includes data from various airlines, covering factors such as flight details, customer demographics, service quality perceptions, and overall satisfaction ratings. The dataset consists of structured information obtained through surveys, with attributes like flight distance, departure/arrival delays, seat comfort, onboard services, and customer feedback. Additionally, it incorporates categorical variables reflecting passenger characteristics such as age, gender, travel class, and frequent flyer status. This comprehensive dataset offers valuable insights into the factors influencing passenger satisfaction within the aviation industry, aiding in the analysis of trends, identification of key drivers, and formulation of strategies for enhancing the travel experience. Below is a brief description of the dataset:

No	Attribute	Description
1	ID	It is the unique number given to each response
2	Gender	Gender of the passengers (Female, Male)
3	Customer Type	Type of the customer. (Loyal, Disloyal)
4	Age	The actual age of the passengers
5	Type of Travel	Purpose of the flight of the passengers (Personal, Business)
6	Class	Travel class of the passenger (Economy, Business)
7	Flight Distance	The flight distance of this journey
8	Inflight WiFi Service	Satisfaction level of the inflight Wi-Fi service (0:Not Applicable;1- 5)
9	Departure/Arrival time convenient	Satisfaction level of Departure/Arrival time convenient
10	Ease of Online Booking	Satisfaction level of online booking
11	Gate Location	Satisfaction level of Gate location
12	Food and Drink	Satisfaction level of Food and drink
13	Online Boarding	Satisfaction level of online boarding
14	Seat Comfort	Satisfaction level of Seat comfort
15	Inflight entertainment	Satisfaction level of inflight entertainment
16	On-Board Service	Satisfaction level of On-board service
17	Leg room service	Satisfaction level of Leg room service
18	Baggage Handling	Satisfaction of baggage handling
19	Checkin Service	Satisfaction of checkin service
20	Inflight Service	Satisfaction of inflight entertainment
21	Cleanliness	Satisfaction of cleanliness
22	Departure Delay in Minutes	Minutes delayed when departure
23	Arrival Delay in Minutes	Minutes delayed when Arrival
24	Satisfaction	Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

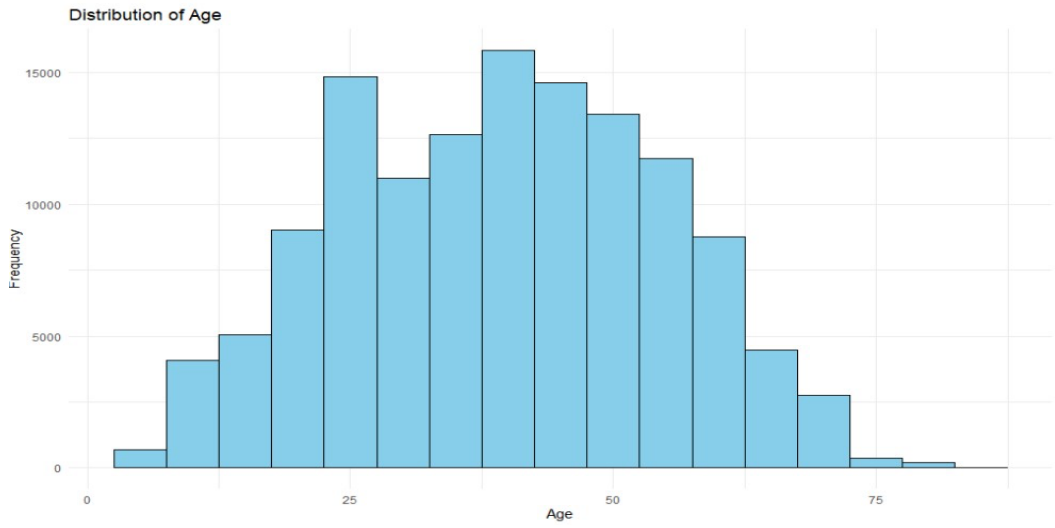
Challenges Encountered:

1. Understanding feature importance: Determining which features have the most significant impact on passenger satisfaction and how they interact with each other required thorough exploration and analysis of the dataset.
2. Model interpretation: Interpreting the results of machine learning models to derive actionable insights for improving passenger satisfaction posed a challenge, as it involved understanding the underlying relationships between features and satisfaction levels.
3. Dealing with missing values is one of the most common challenges. You may need to decide whether to impute missing values, remove rows or columns with missing values, or leave them as is.We have remove the coloumns with null values.

Analysis and Discussion:

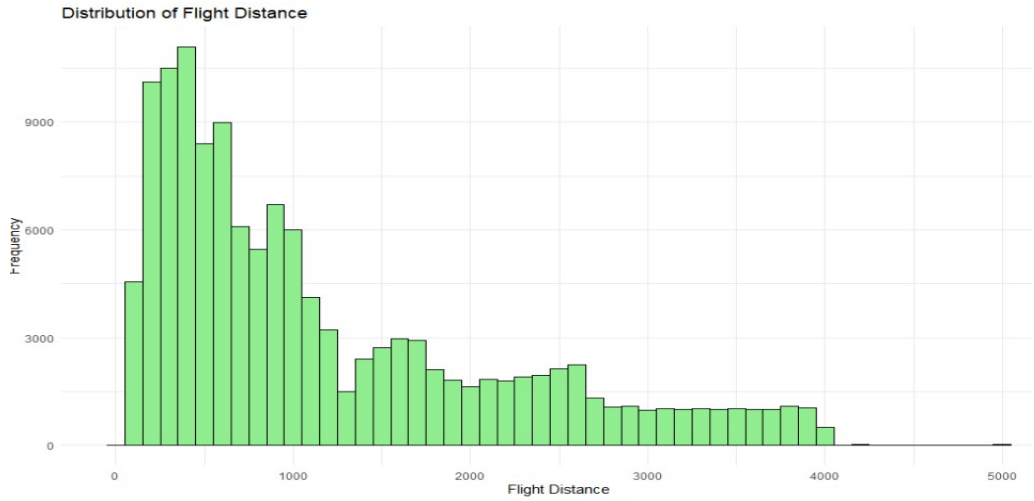
Exploratory data analysis using Different Visualization Techniques

Histogram for Age:



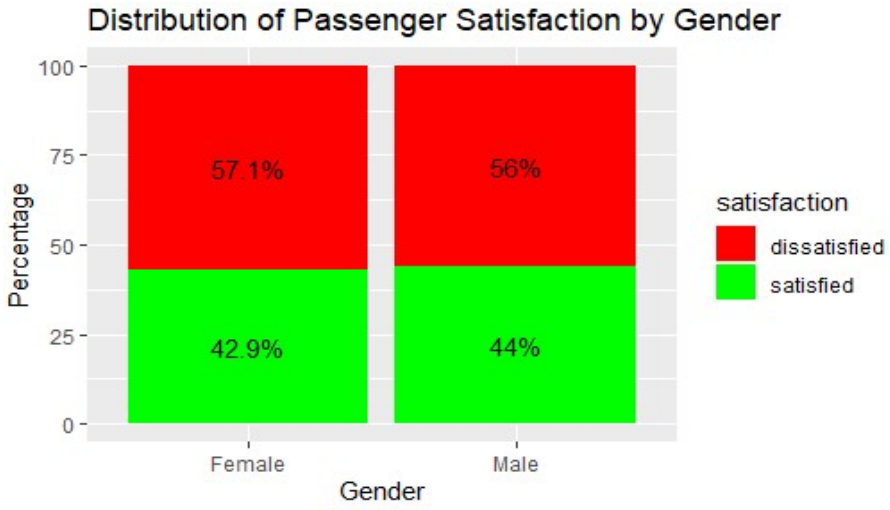
It can be inferred from the graph that the passengers between the age groups (23-27) and (42-47) are frequent flyers.

Histogram for Flight Distance:



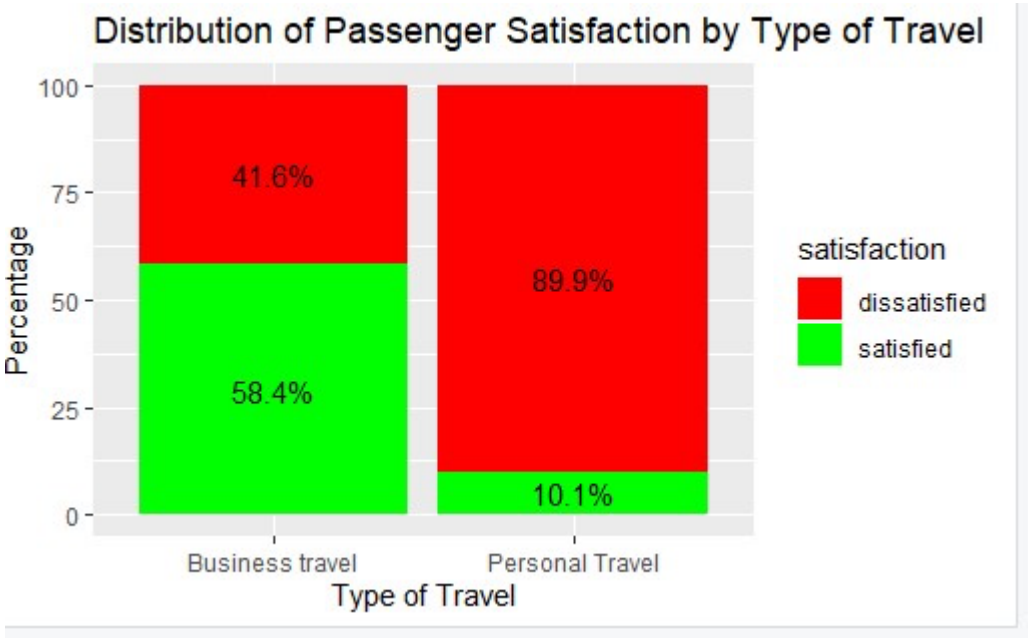
From the graph it can be inferred that there are more passenger trips who travel less flight distances which are less than 1500 miles. Therefore most of the trips are domestic.

Bar chart for Gender with percentages:



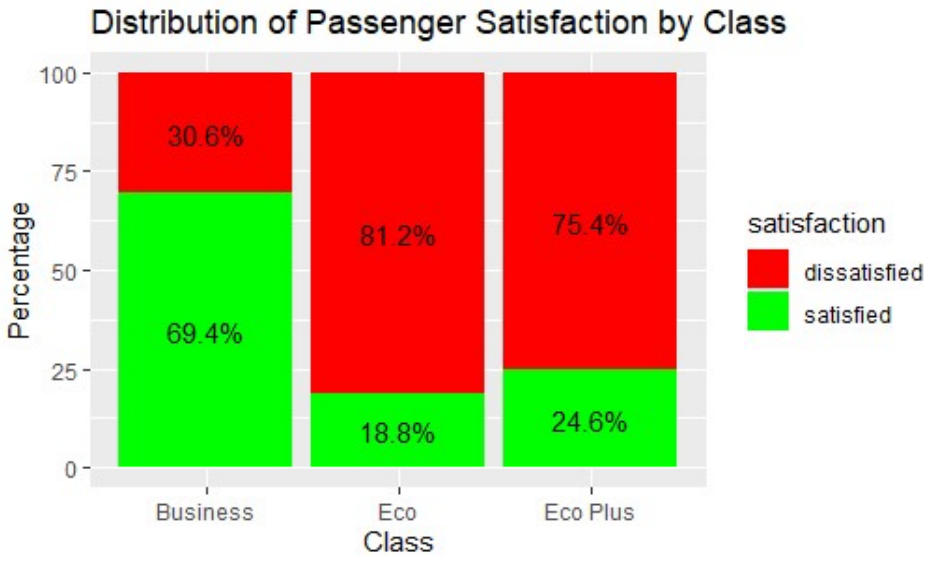
From the Bar chart it can be inferred that in both male and female passenger satisfaction the percentage is low which is less than 50% but in male passenger satisfaction it is slightly higher than female.

Bar chart for Type of Travel with percentages:



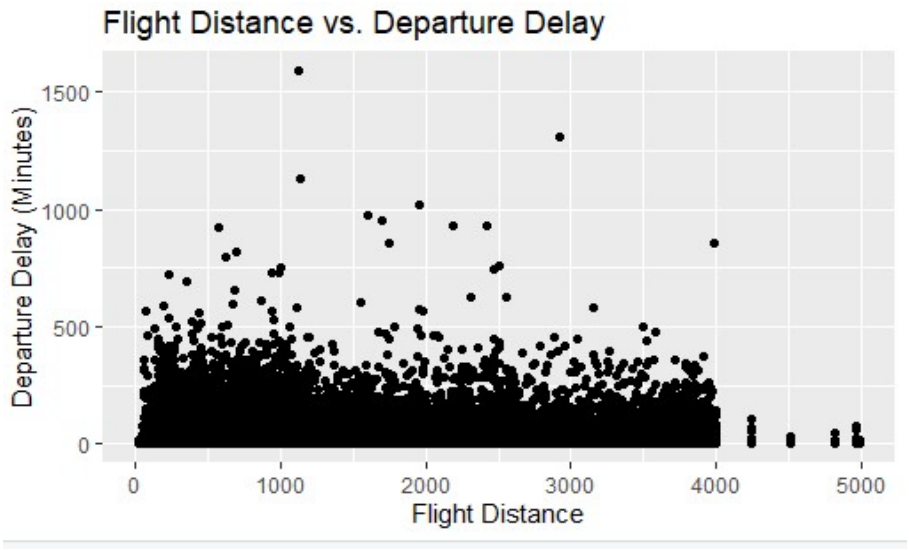
From the bar chart it can be inferred that passenger satisfaction is very low for personal travel(10.1%) compared to Business travel(58.4%). The airline should improve the services for personal travel.

Bar chart for Class with percentages:



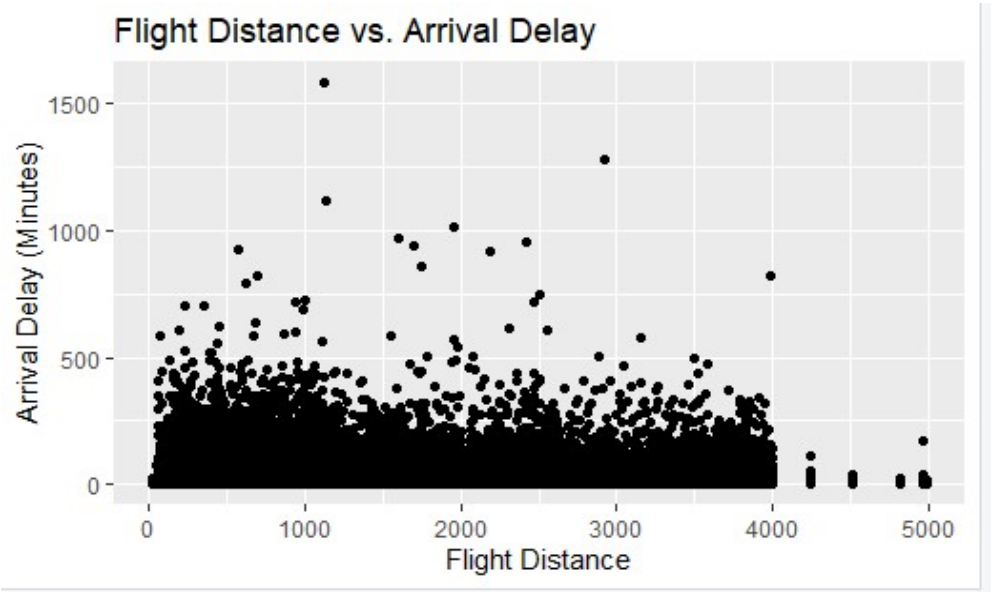
From the bar chart, it can be inferred that Passenger satisfaction for Eco and Eco plus is very low as compared to the business class. Thus the airline has to improve the services for Eco and Eco plus.

Scatter plot for Flight Distance vs. Departure Delay:



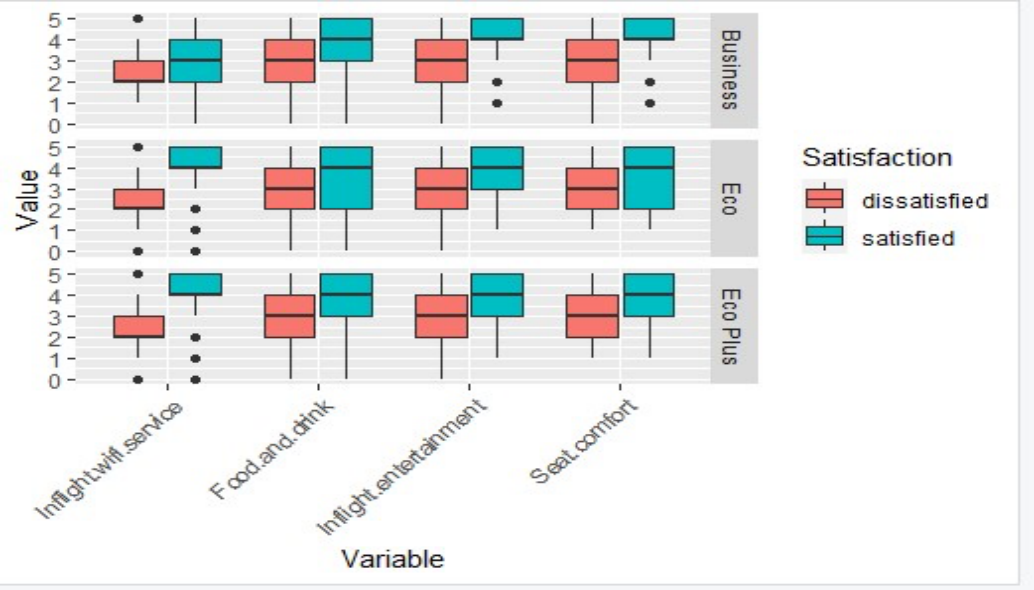
From the above scatter plot it can be inferred that departure delay is more for short flight distances. Therefore airlines should take necessary measures to reduce departure delay for short distances.

Scatter plot for Flight Distance vs. Arrival Delay:



From the above scatter plot it can be inferred that arrival delay is more for short flight distances. Therefore, airlines should take necessary measures to reduce arrival delay for short distances.

Box plots to visualize the distribution of numerical variables across different levels of categorical variables such as Class.

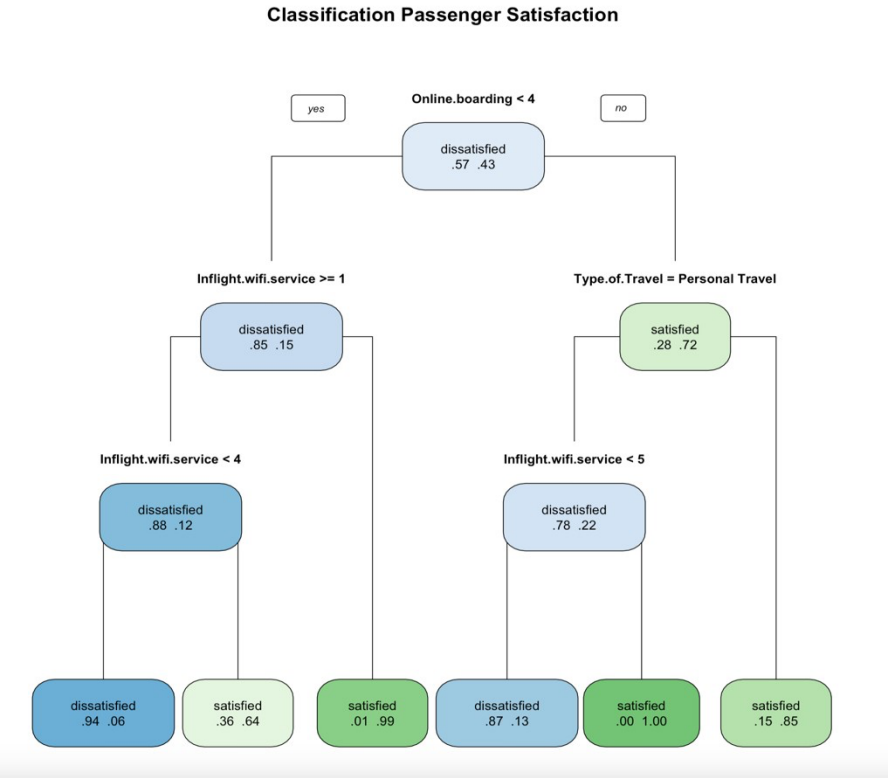


According to the box plot, for “Inflight WiFI Services”, the mean for the satisfied passenger is highest for the “Eco Class”. Whereas, for the “Food & Drink”, the mean for the satisfied customers is highest for the “Economy” class passengers. For “Inflight Entertainment” the mean of satisfied passengers is highest for “Business Class” and that for the “Seat Comfort”, it is observed to be highest for the “Business Class” as well.

Goal of the Project:

The primary goal of this project is to analyze and understand the factors influencing airline passenger satisfaction. By leveraging the dataset provided, we aim to uncover patterns, trends, and correlations among various attributes related to the travel experience. Through comprehensive analysis, the project seeks to identify key drivers of passenger satisfaction and dissatisfaction, thereby enabling airlines to improve their services, enhance customer experience, and ultimately increase customer loyalty and retention.

Goal 1: Factors contributing most to passenger satisfaction and identifying the features (e.g., seat comfort, in-flight entertainment, customer service) that significantly influence passenger satisfaction ratings.



On the Training model, the accuracy comes out to be 88.46% and that for the testing data is 88.38% which is approximately equal.

The root node of Decision tree is the “Online boarding” factor , meaning it has the highest information gain following “Inflight wifi service” and “Type of travel” which further drives satisfaction of passengers.

Managerial Insights:

- Optimize Online Boarding: Streamline online check-in processes, offer clear instructions, and provide personalized assistance to improve the online boarding experience.
- Enhance Inflight WiFi: Invest in technology to improve the quality and reliability of onboard WiFi, and consider offering complimentary or affordable access to enhance passenger satisfaction.
- Tailor Services by Travel Type: Customize services and amenities based on the needs of different types of travelers (e.g., leisure, business) to cater to their specific preferences and enhance satisfaction levels.

Goal 2: How do flight-related factors impact satisfaction?

Analyze the effects of flight duration, departure/arrival delays, and flight class on passenger satisfaction levels.

To analyze the effects of flight-related factors on the customer satisfaction, two different analyses were performed on the dataset; namely, logistic regression and Naive Bayes Classifier.

The flight related factors considered are: Online boarding service, Departure Delay in Minutes, Arrival Delay in Minutes, Departure Arrival TIme Convenient, Flight Distance.

```
Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -2.934e+00  4.213e-02 -69.655  < 2e-16 ***
Online.boarding                  8.581e-01  8.200e-03 104.645  < 2e-16 ***
Checkin.service                  2.633e-01  7.366e-03  35.741  < 2e-16 ***
Departure.Delay.in.Minutes       2.942e-03  9.150e-04   3.215   0.0013 **
Arrival.Delay.in.Minutes        -7.102e-03  8.993e-04  -7.897 2.86e-15 ***
Departure.Arrival.time.convenient -1.298e-01  6.086e-03 -21.322  < 2e-16 ***
Flight.Distance                  1.608e-04  9.989e-06  16.096  < 2e-16 ***
ClassEco                        -1.810e+00  2.065e-02 -87.653  < 2e-16 ***
ClassEco Plus                   -1.451e+00  3.601e-02 -40.307  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 118149  on 86323  degrees of freedom
Residual deviance:  77222  on 86315  degrees of freedom
AIC: 77240

Number of Fisher Scoring iterations: 5
```

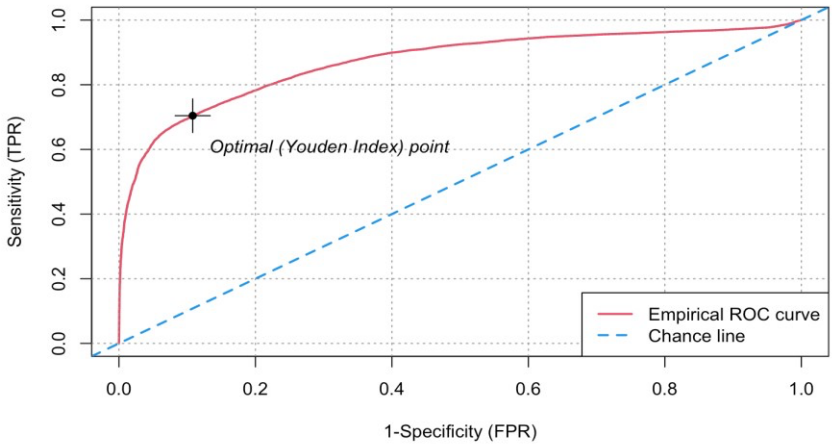
The above output shows that p-value for all the variables considered is less than 0.05 except for “Departure Delay in Minutes”. This helps in understanding that all the

flightrelated factors affect passenger satisfaction other than “Departure Delay in Minutes”.

The Confusion Matrix built is:

predict	actual dissatisfied	satisfied
0	20406	4737
1	3936	14084

The Receiver Operating Characteristic (ROC) Curve from the Logistic Regression is as follows:



The optimal point/ best cut-off point (Youden Index point) plotted is 0.567. It means that at sensitivity being 0.567, the regression best classifies the satisfied and dissatisfied airline passengers.

Below is the output of Naive Bayes Classifier

```
Confusion Matrix and Statistics

              Reference
Prediction     dissatisfied satisfied
  dissatisfied        17347      2400
  satisfied            6995     16421

               Accuracy : 0.7823
                 95% CI : (0.7784, 0.7862)
    No Information Rate : 0.564
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5694

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8725
            Specificity : 0.7126
         Pos Pred Value : 0.7013
         Neg Pred Value : 0.8785
             Prevalence : 0.4360
         Detection Rate : 0.3804
   Detection Prevalence : 0.5425
      Balanced Accuracy : 0.7926

       'Positive' Class : satisfied
```

The accuracy of Logistic Regression is observed to be 79.9%, while that of Naive Bayes Classifier is observed as 78.23%. Hence, Logistic Regression is the better fit for this dataset.

Managerial Insights :

- Focus on Arrival Delays and Online Boarding: Our analysis shows arrival delays and a smooth online boarding experience are key factors influencing satisfaction. Let's prioritize initiatives to minimize arrival delays and enhance the online boarding system for a more positive passenger experience.

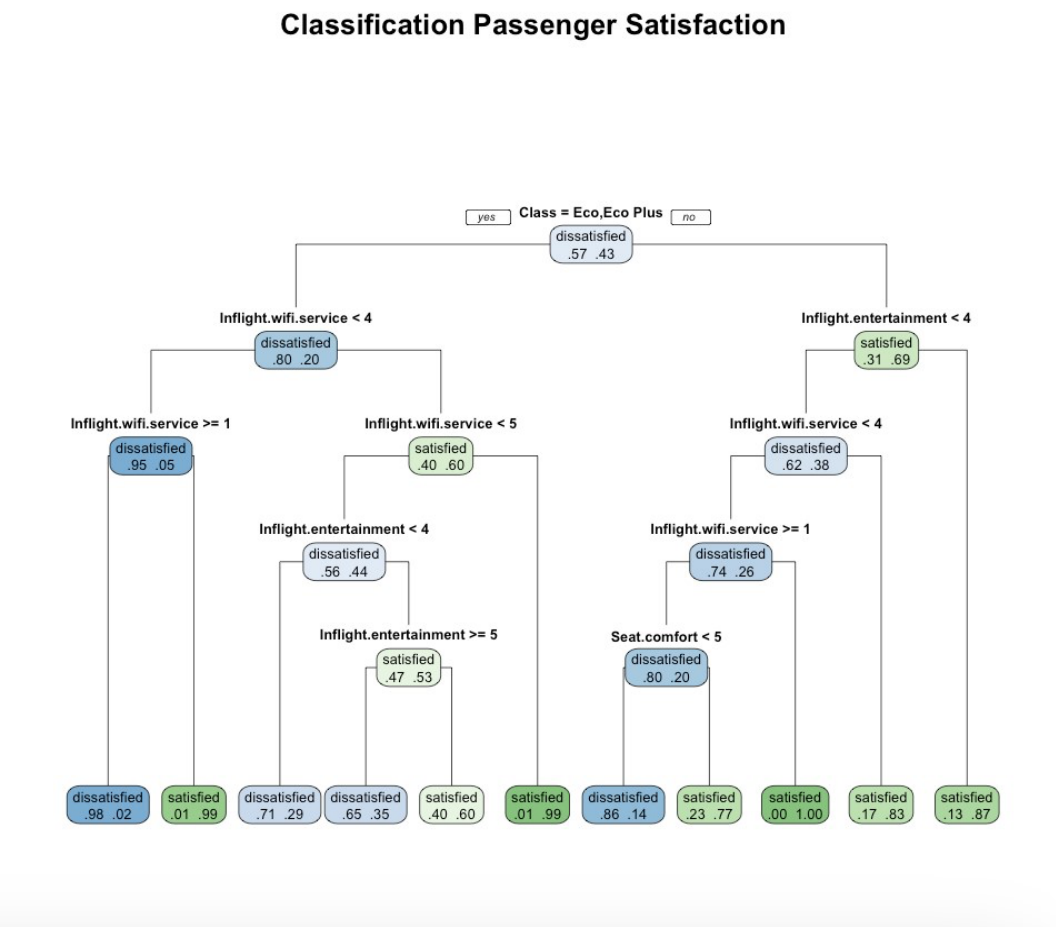
- Departure Delays Might Need Rethinking: Interestingly, departure delays seem less impactful on satisfaction compared to other factors. We should investigate further through passenger surveys to understand why. This could reveal opportunities to streamline pre-departure processes and improve communication during delays.

- Passenger Satisfaction Prediction Model: We developed a Logistic Regression model with 79.9% accuracy in predicting passenger satisfaction based on flightrelated factors. This model can be a powerful tool. We can use it to proactively identify flights with a higher risk of dissatisfaction and take targeted actions to address potential concerns before passengers even board.

Goal 3: How does inflight service impact passenger satisfaction?
Analyze the effects of Inflight Wifi Service, Food and Drink,Seat Comfort, Inflight entertainment, Leg room service, Cleanliness.

To analyze what inflight factors contribute the most in passenger satisfaction, Decision Tree Algorithm is used. The inflight factors taken into consideration are: Class, Inflight WiFi, Food and Drink, Seat Comfort, Inflight Entertainment, Leg Room Service, Cleanliness. The method to split nodes is “Gini” i.e. this method focuses on minimizing the Gini index in order to get maximum information.

Decision Tree Algorithm Output obtained is:



The Decision Tree above gives the inference that feature "Class" is the most important inflight feature in affecting the passenger satisfaction as compared to other features into consideration. This is because it is the root node in the Decision Tree. Further, "Inflight WiFi" followed by "Inflight Entertainment" followed by "Seat Comfort" are the factors that affect the satisfaction levels of the customers.

The Confusion Matrix is as follows:

	satisfaction.actual	
satisfaction.pred	dissatisfied	satisfied
dissatisfied	43155	3425
satisfied	5728	34016

Accuracy of the above model is 89.39%

Managerial Insights:

- Prioritize In-flight Experience by Class (with Examples):
  - ➢ Economy Class: Focus on providing clean and well-maintained cabins. Offer a selection of complimentary beverages and snacks. Ensure seat cushions are comfortable and legroom is adequate within industry standards.
  - ➢ Business Class: Invest in wider, reclining seats with adjustable headrests and footrests. Offer a wider variety of complimentary beverages and healthy meal options. Provide access to priority boarding and baggage claim.
  - ➢ First Class: Elevate the experience with lie-flat seats, high-quality linens, and personalized service. Offer gourmet meals with a wider selection and the ability to preorder. Include premium entertainment options and noise-canceling headphones.
- Focus on WiFi, Entertainment, and Seat Comfort (with Examples):
  - ➢ In-flight WiFi: Invest in reliable and high-speed internet connectivity to allow passengers to stay connected, work, or stream entertainment during the flight. Offer tiered Wi-Fi packages catering to different usage needs (browsing, streaming).
  - ➢ In-flight Entertainment: Provide a diverse range of movies, TV shows, music, and games catering to various tastes and ages. Partner with streaming services to offer the latest content. Implement user-friendly interfaces for easy access and control.
  - ➢ Seat Comfort: Invest in ergonomic seat designs with adjustable lumbar support for better posture. Offer different seat cushion firmness options based on passenger preferences. Consider wider seats with increased legroom, especially on longer flights.
- High-Accuracy Predictive Model with Proactive Actions (with Examples):
  - ➢ Utilize the Decision Tree model to identify flights with a higher risk of passenger dissatisfaction based on in-flight service factors (e.g., limited entertainment options on a long flight).
  - ➢ Pre-flight Communication: Inform passengers about potential limitations (e.g., limited entertainment selection) and offer upgrade options for in-flight entertainment packages.
  - ➢ Onboard Adjustments: Offer complimentary amenity kits or entertainment vouchers to passengers on flights predicted to have lower satisfaction. Prioritize attentive service and address passenger concerns promptly.