Data and Prediction Analysis of Airline Passenger Satisfaction

Project Team

RUTURAJ JOSHI - A20497857
VIKAS PATHAK - A20460927

Overview

- Transportation services have become core community necessities for both daily activities and travel

- The airline industry is currently the biggest industry of transportation in the world

- Global air transport increased by 5% per year, or doubling every 15 years

- Customer behavior analysis is crucial in today's market

- Understanding the variables that affect airline passengers' pleasure is essential to marketing techniques like promotional vouchers, target campaigns, etc

Problem Statement

- Identify trends from data that will help drive business decisions

- Identify critical factors that ensure customer satisfaction

- Identify areas in which airline operator needs improvement

- Identify customer satisfaction levels by building a efficient classifier
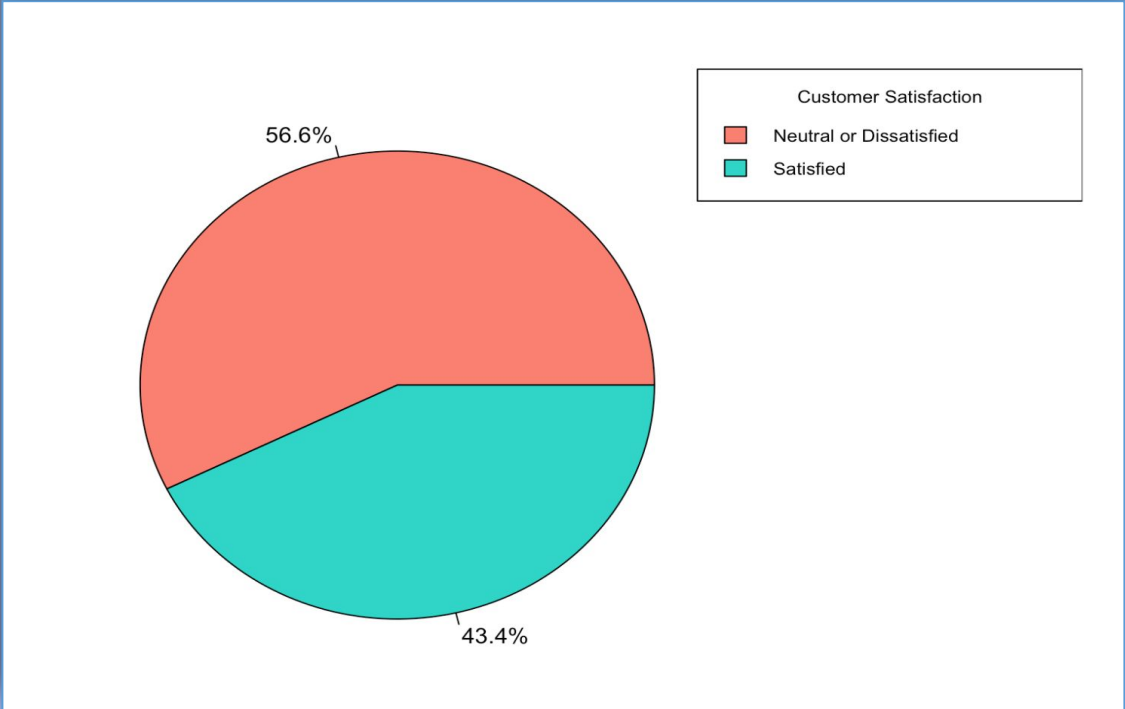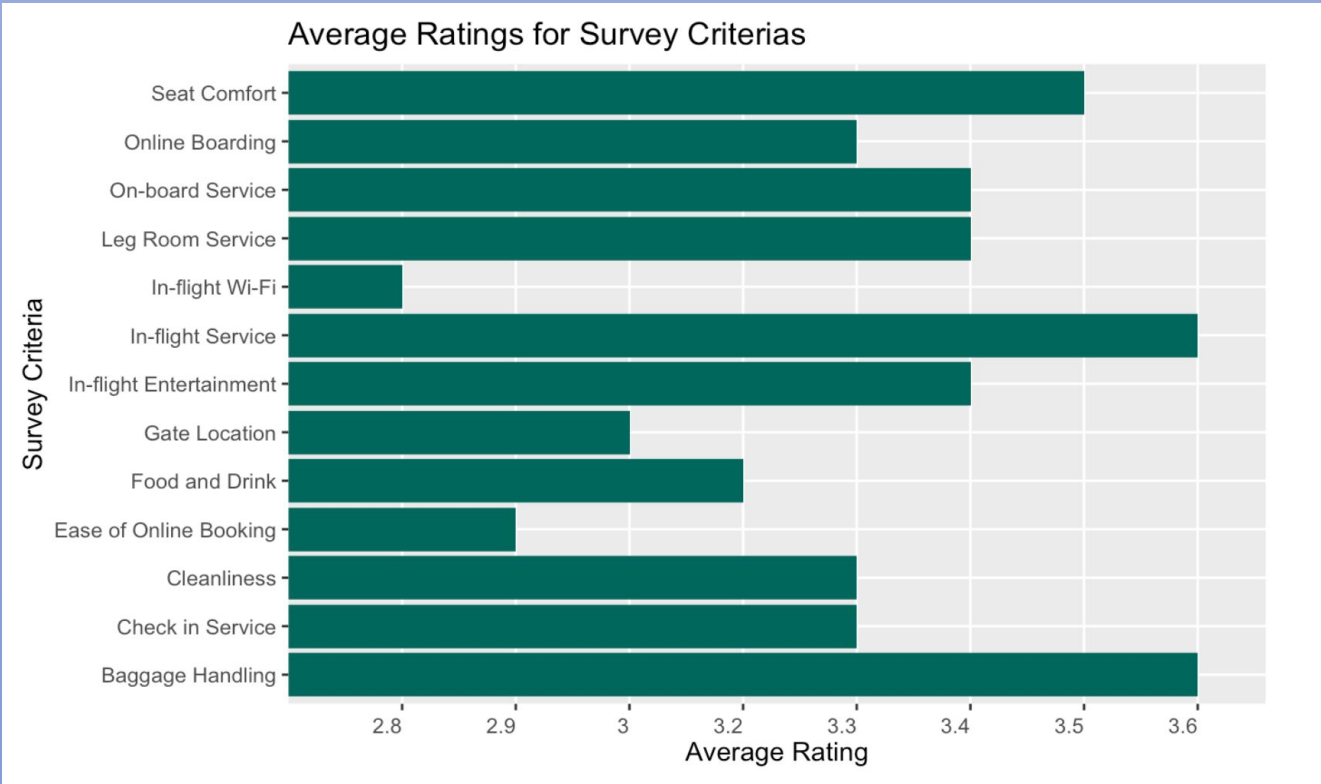
Dataset

Target

	Satisfaction	Gender	Customer Type	Age	Type Of Travel	Class	Flight Distance	Inflight Wifi Service	Departure/ Arrival Time Convenience	Ease Of Online Booking	Gate Location	Food And Drink	Online Boarding	Seat Comfort	Inflight Entertain ment	On-board Service	Leg Room	Baggage Handling	Checkin Service	Inflight Service	Cleanliness	Total Delay
0	Neutral/ Dissatisfied	Male	Returning Customer	13	Personal Travel	Economy	460	3	4	3	1	5	3	5	5	4	3	4	4	5	5	43.0
1	Neutral/ Dissatisfied	Male	First-time Customer	25	Business travel	Business	235	3	2	3	3	1	3	1	1	1	5	3	1	4	1	7.0
2	Neutral/ Dissatisfied	Female	Returning Customer	25	Business travel	Business	562	2	5	5	5	2	2	2	2	2	5	3	1	4	2	20.0
3	Satisfied	Male	Returning Customer	61	Business travel	Business	214	3	3	3	3	4	5	5	3	3	4	4	3	3	3	0.0
4	Neutral/ Dissatisfied	Female	Returning Customer	26	Personal Travel	Economy	1180	3	4	2	1	1	2	1	1	3	4	4	4	4	1	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
69061	Neutral/ Dissatisfied	Female	First-time Customer	36	Business travel	Economy	432	1	5	1	3	4	1	4	4	5	2	5	2	3	4	0.0
69062	Neutral/ Dissatisfied	Male	First-time Customer	34	Business travel	Business	526	3	3	3	1	4	3	4	4	3	2	4	4	5	4	0.0
69063	Neutral/ Dissatisfied	Female	Returning Customer	17	Personal Travel	Economy	828	2	5	1	5	2	1	2	2	4	3	4	5	4	2	0.0
69064	Satisfied	Male	Returning Customer	14	Business travel	Business	1127	3	3	3	3	4	4	4	4	3	2	5	4	5	4	0.0
69065	Neutral/ Dissatisfied	Female	Returning Customer	42	Personal Travel	Economy	264	2	5	2	5	4	2	2	1	1	2	1	1	1	1	0.0

Data Preparation

- Typecast string values into factor i.e. categorical type

- Identified & removed rows that has *NA* value for at least one predictor variable

- Assumption: Survey criteria that has 0 as rating might be skipped by passenger and not the other way around. Such rows should be removed as it could affect the predictions

Data Visualization

Obtained dataset is balanced dataset since it has enough representation of each class label.

Average Ratings for Survey Criterias

Customer Satisfaction
- Neutral or Dissatisfied
- Satisfied

56.6%

43.4%

We can identify services where airline operator is doing well and also the services that needs improvements.

Gender Vs Satisfaction

In our dataset, fewer female passengers are satisfied with airline service when compared to male passengers. On the other hand, fewer male passengers are dissatisfied with airline service when compared to female passengers.
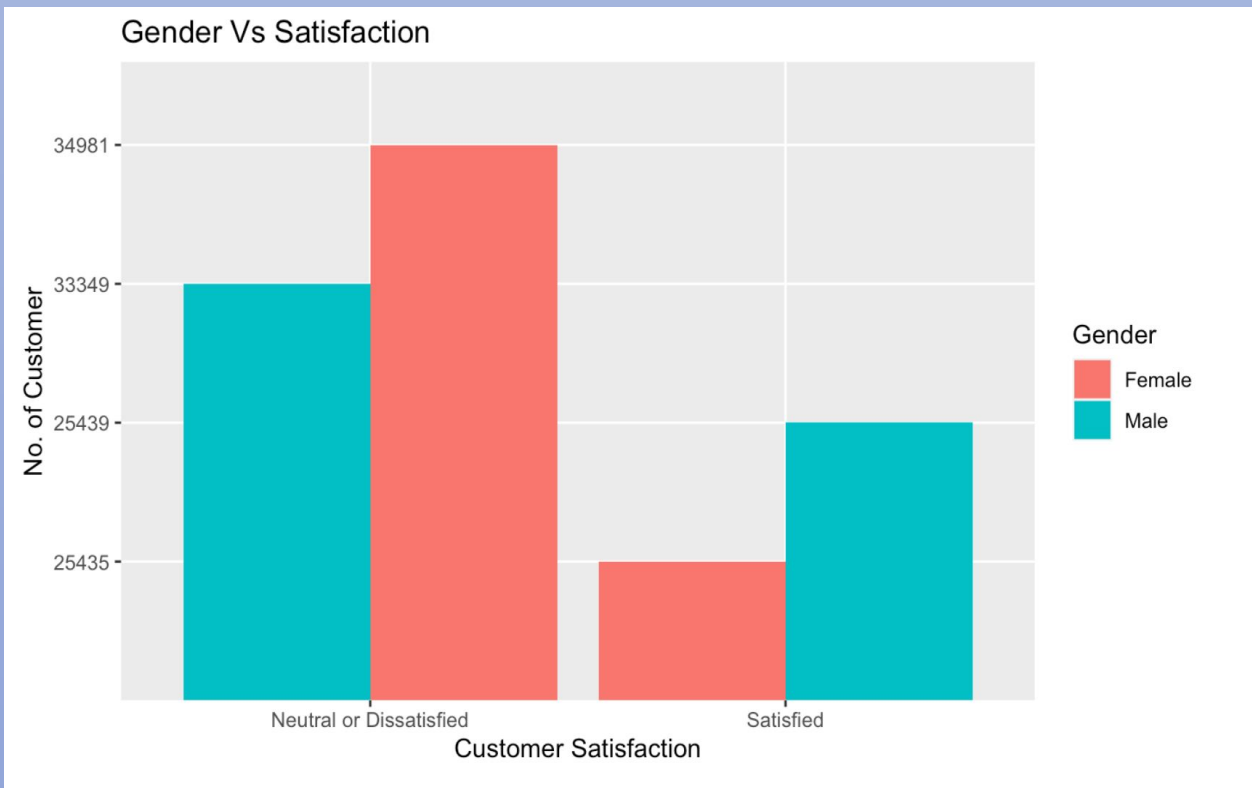
First-time passengers are more likely to be satisfied with airline service. On the contrary, returning passengers are more likely to to be dissatisfied with the service.

Customer Type Vs Satisfaction
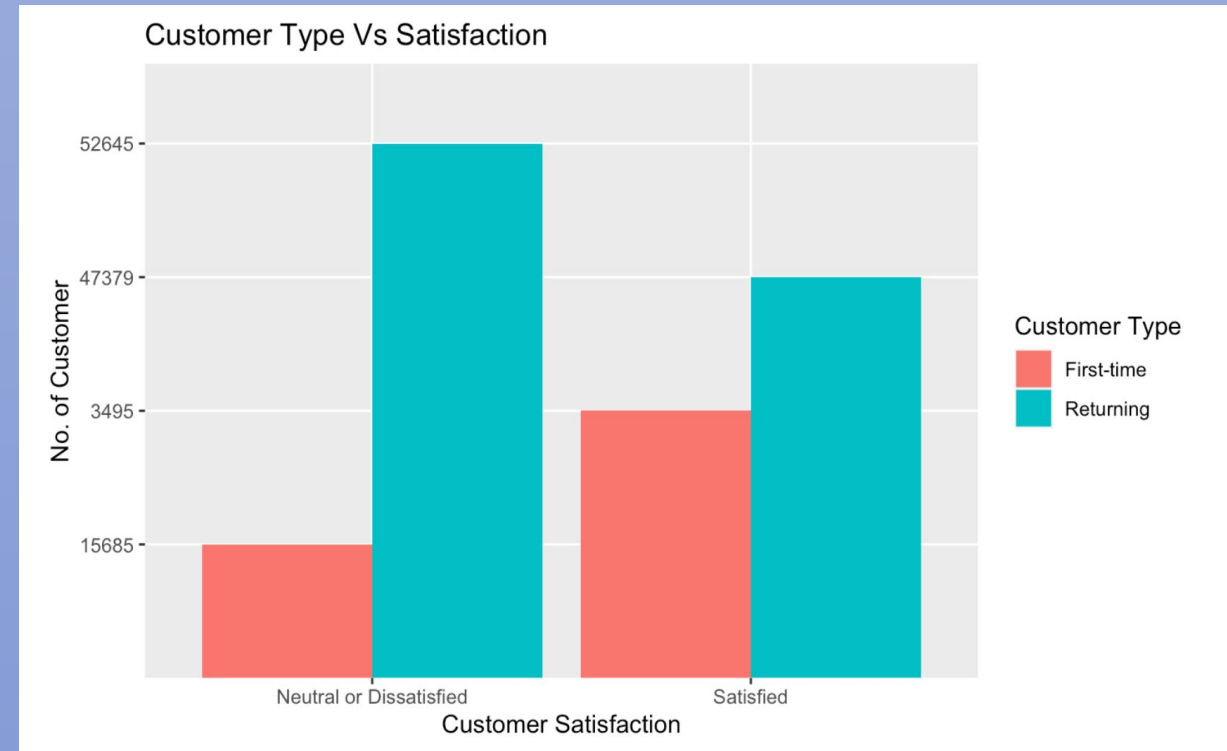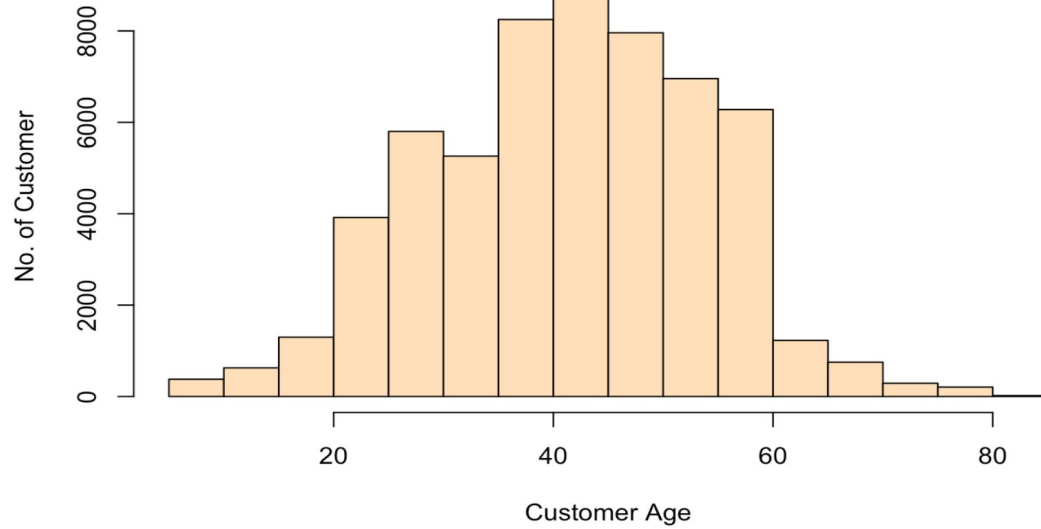
Travel Class Vs Satisfaction
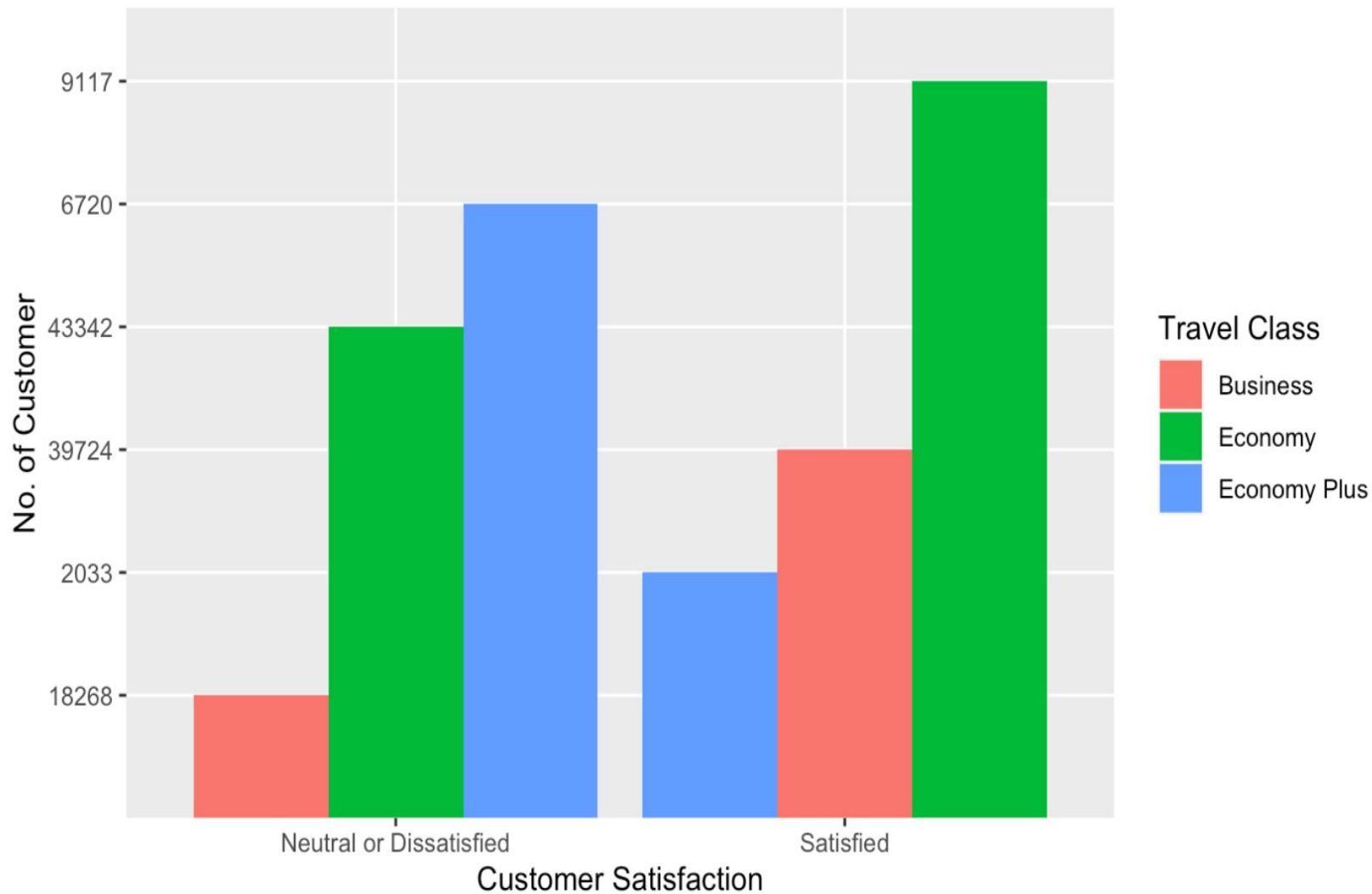
Passengers travelling with economy class tends to be more satisfied with airline service when compared to passengers travelling with business class and economy plus class.
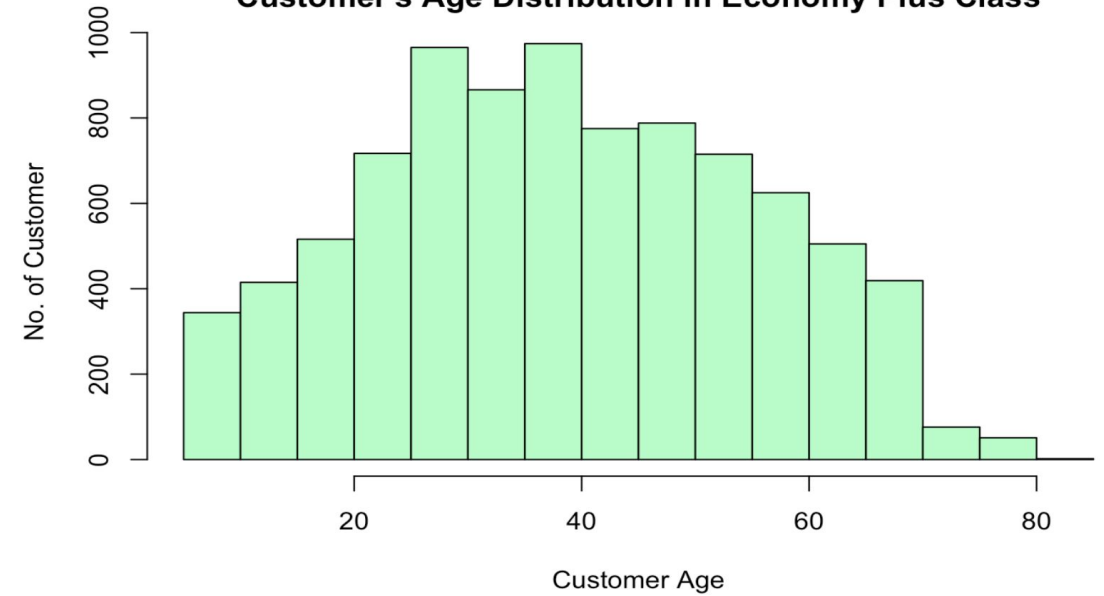
Passengers travelling with economy plus class tends to be more dissatisfied when compared with passengers travelling with economy class and business class.
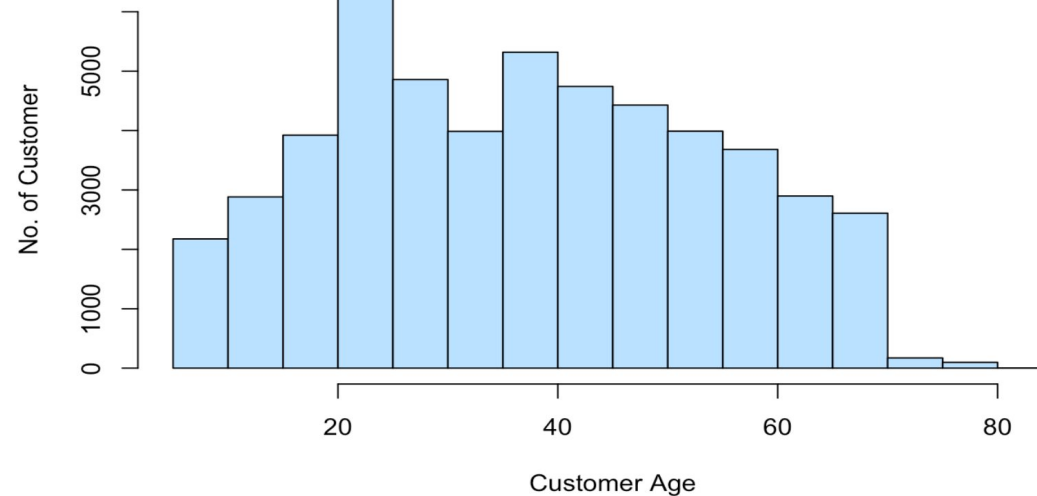
Passengers age distribution across all classes

Data Partitions

- We are using 80-20% split for training and testing dataset

- No of rows in Training dataset: 95364

- No of rows in Testing dataset: 23840

Classification Models

We have built below classifiers:

- Decision tree

- Random forest

- Naive Bayes

- Logistic regression

Comparative Analysis

Models	Accuracy	Balanced Accuracy	Specificity	Sensitivity	Precision	Error
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Logistic Model	0.872	0.863	0.798	0.927	0.859	0.128
Decision Tree	0.88	0.882	0.898	0.866	0.919	0.12
Random Forest	0.946	0.943	0.922	0.964	0.942	0.054
Naive Bayes	0.866	0.866	0.866	0.867	0.896	0.134

4 rows

Conclusion

- We were able to identify anomalies in the underlying data set and counter them with strategically sound decisions

- We were also able to produce some insightful visualization that depicts trends in the dataset

- With visualization, we could point out areas the airline operator is doing well and also the areas that need improvements to achieve a better customer experience

- We built several classification models to determine customer satisfaction and from the experimentation results, we conclude that random forest is the best suitable for our application

Future Work

- Create data pipeline to utilize data from several sources

- Explore other classification models like KNN, Support vector machine, Artificial neural network etc

- Create interactive and user friendly dashboard for data visualization

Thank You!