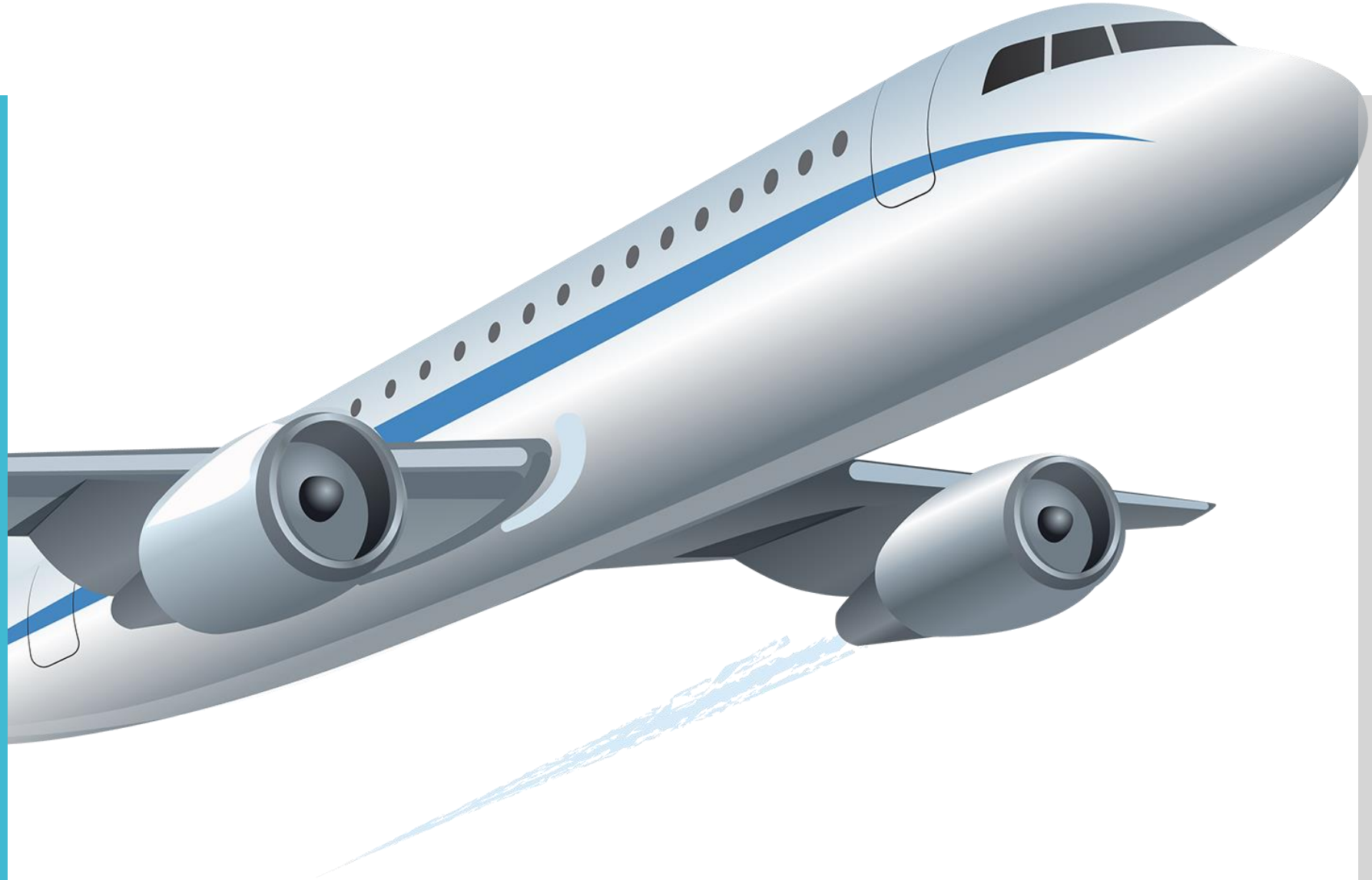
Airline Passenger Satisfaction

Group 3

Naina Grover, Ronald Morris, Sarah Saltzman, Scott Slutsky, Mischa Subhan

Business Understanding

Analyze what factors contribute to customer satisfaction



Team

Naina Grover

Sarah Saltzman

Scott Slutsky

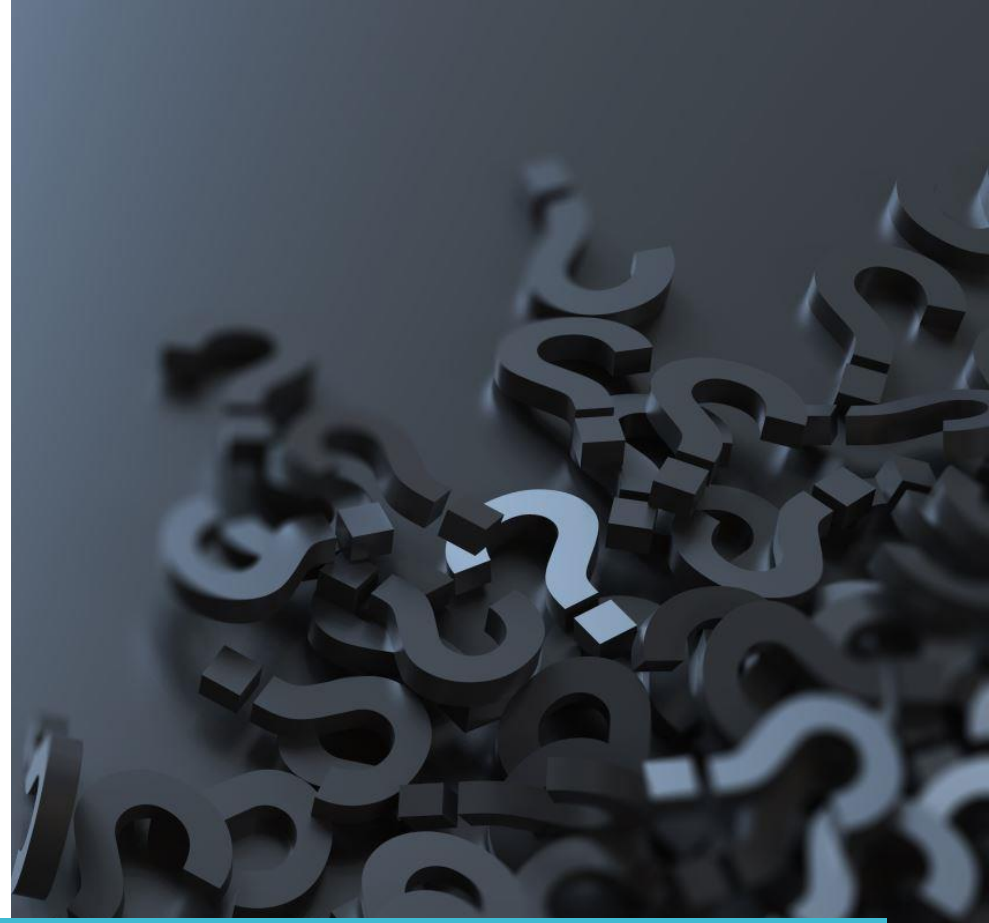
Mischa Subhan

Ronald Morris

Data Instance and Variables

- Customer loyalty
- Gender
- Customer type
- Age
- Type of travel
- Class
- Flight distance
- Flight wifi service
- Departure/Arrival time convenience

Customer Satisfaction



Data science solutions address business problems.

Review

Male	LoyalCustomer	BusinessTravel	Satisfied	EcoClass
1	1	0	0	0
1	0	1	0	0
0	1	1	1	0
0	1	1	0	0
1	1	1	1	0

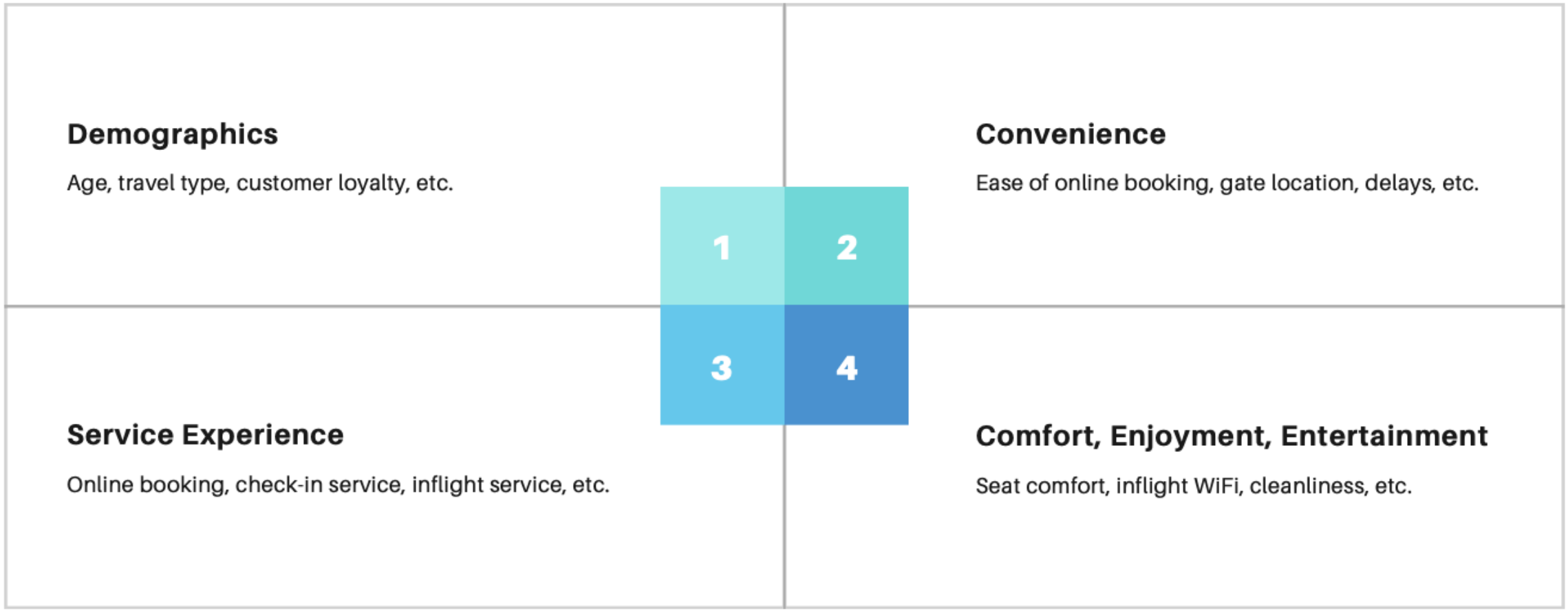
Inflight.entertainment	On.board.service	Leg.room.service	Baggage.handling	Checkin.service	Inflight.service	Cleanliness	Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes
5	4	3	4	4	5	5	25	18
1	1	5	3	1	4	1	1	6
5	4	3	4	4	4	5	0	0
2	2	5	3	1	4	2	11	9
3	3	4	4	3	3	3	0	0

row names	Age	Flight.Distance	Inflight.wifi.service	Departure.Arrival.time.convenient	Ease.of.Online.booking	Gate.location	Food.and.drink	Online.boarding	Seat.comfort
1	[illegible]	460	3	4	3	1	5	3	5
2	25	235	3	2	3	3	1	3	1
3	26	1142	2	2	2	2	5	5	5
4	25	562	2	5	5	5	2	2	2
5	61	214	3	3	3	3	4	5	5

ABOUT THE DATA

Airline Passenger Satisfaction (Source: Kaggle)

n = 119,204 observations



49% Male (48-52 years old)

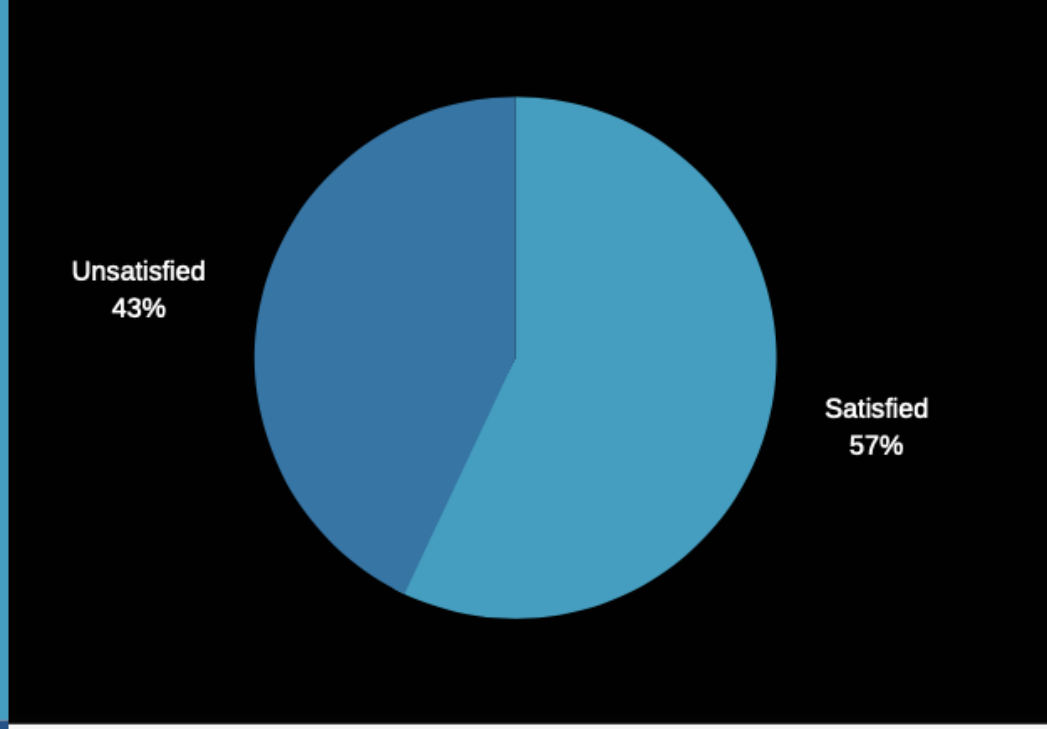
60% Business Class (48-52 years old)

76% Business Travel (48-52 years old)

95% Loyal (48-52 years old)

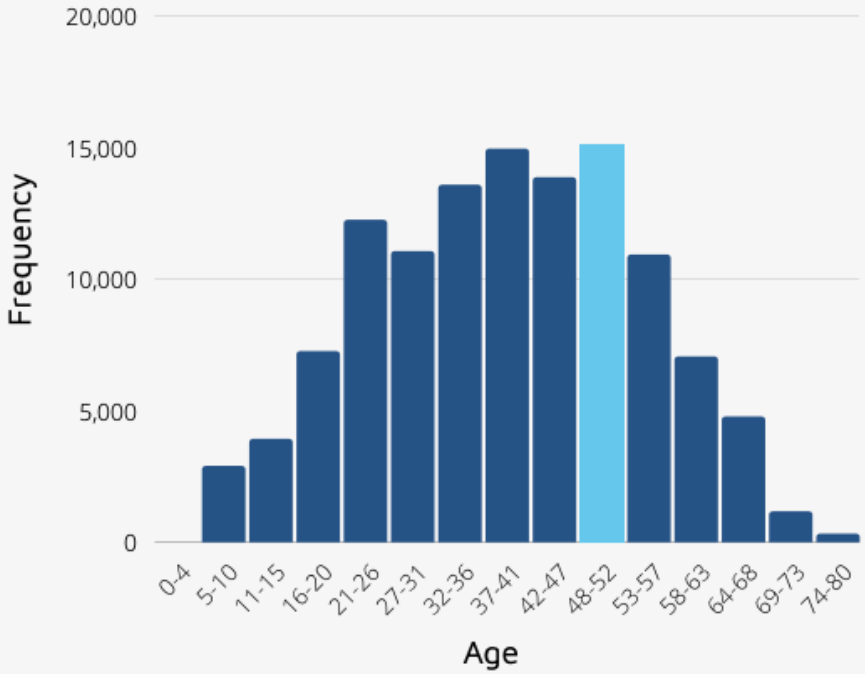
PASSENGER SATISFACTION

Pie chart featuring passenger satisfaction distribution for all age groups

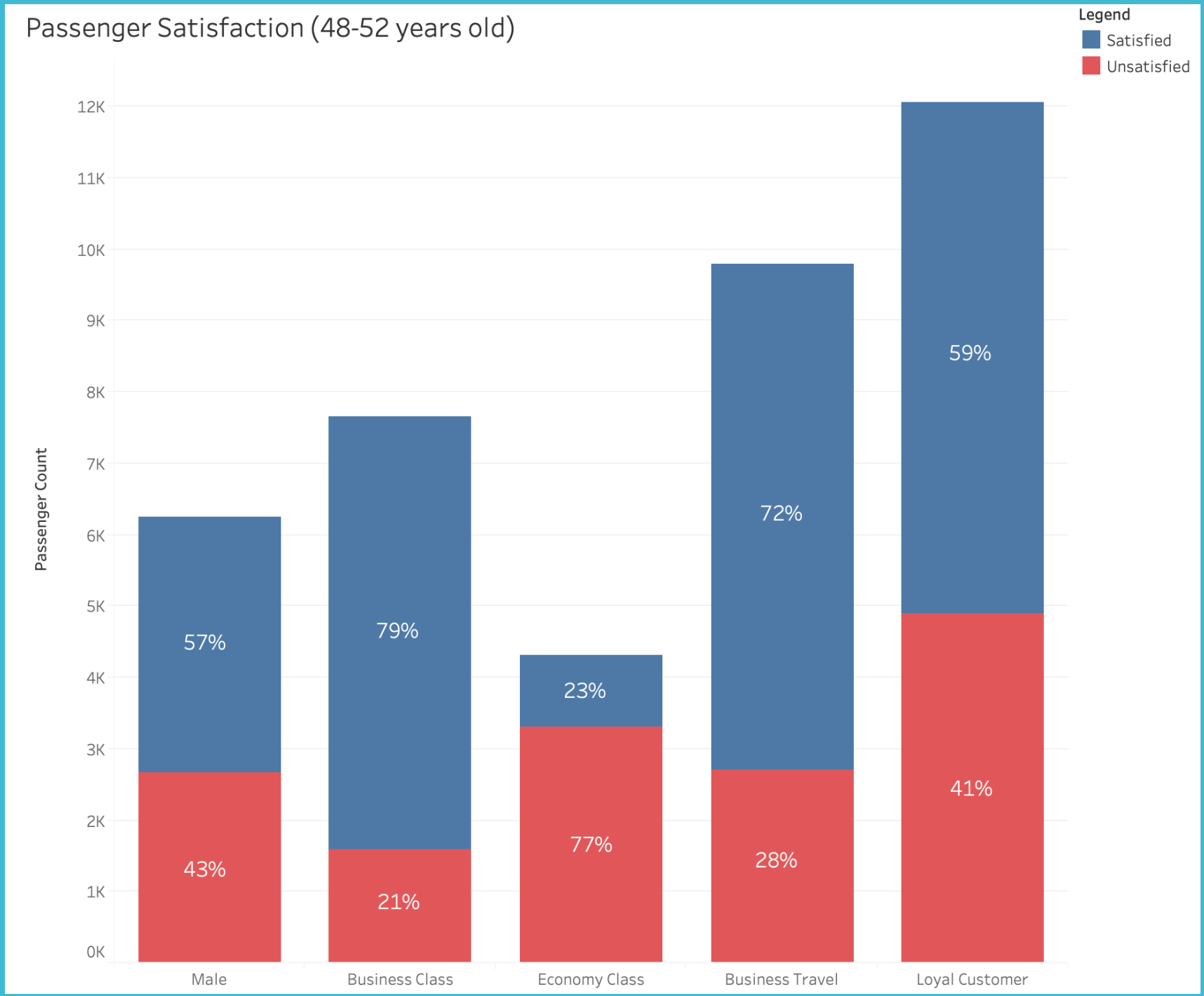


PASSENGER AGE DISTRIBUTION

Histogram featuring passenger age distribution



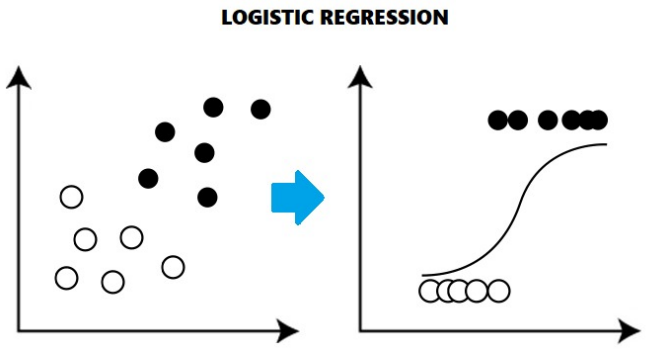
Passenger Satisfaction by Demographic

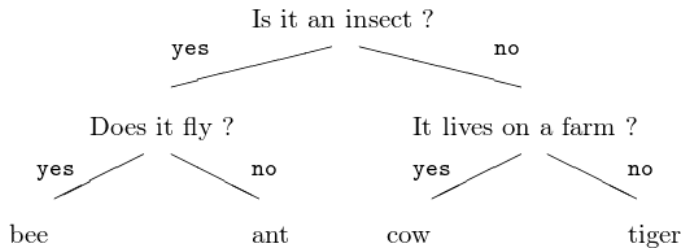


Modeling

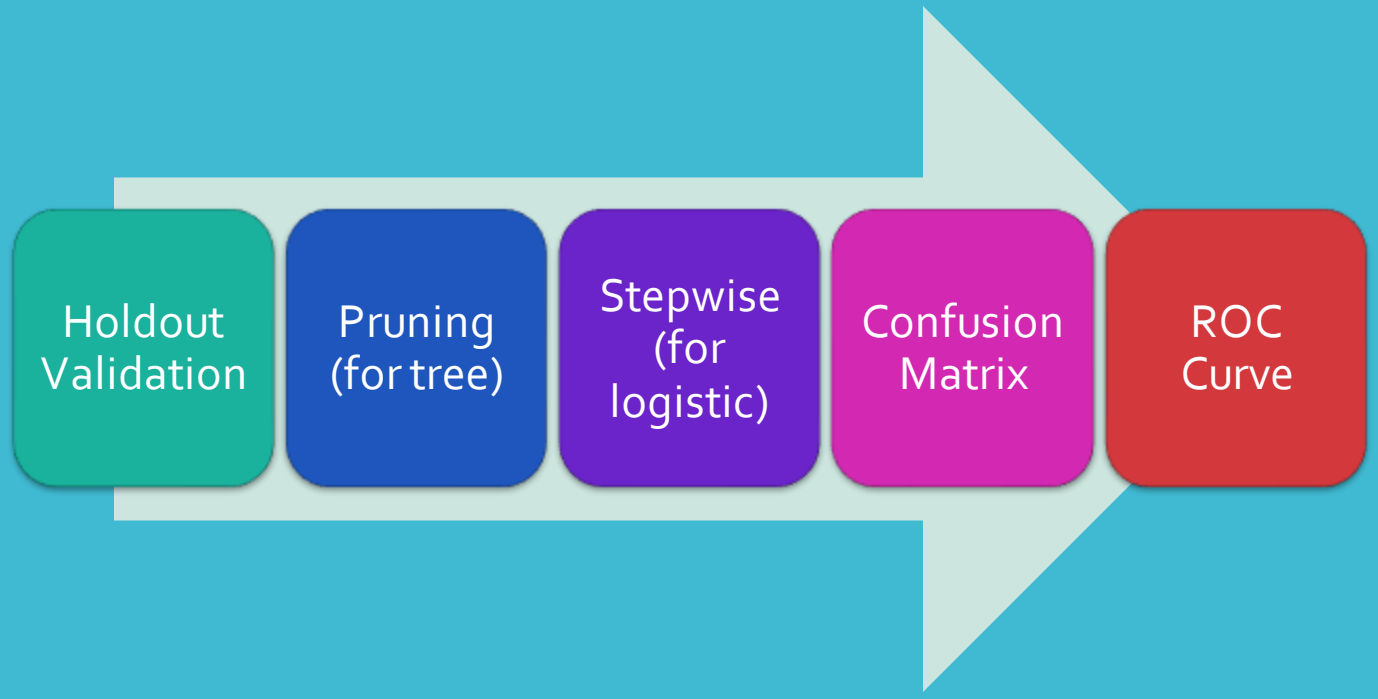
Models Considered

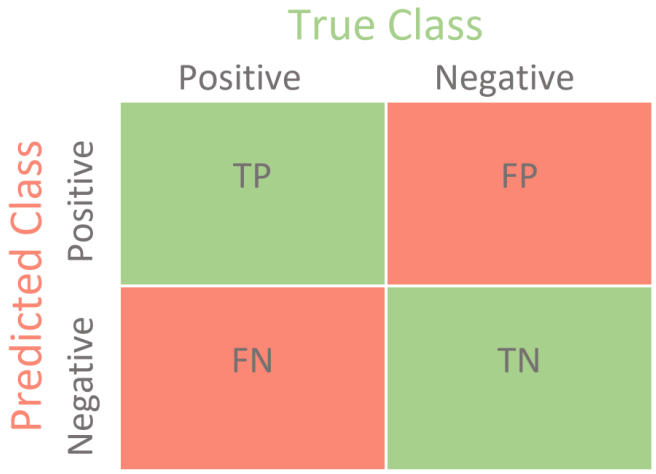
Logistic Regression	SVM	Classification Trees	Random Forest

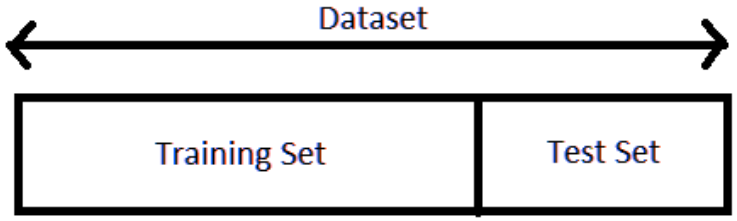




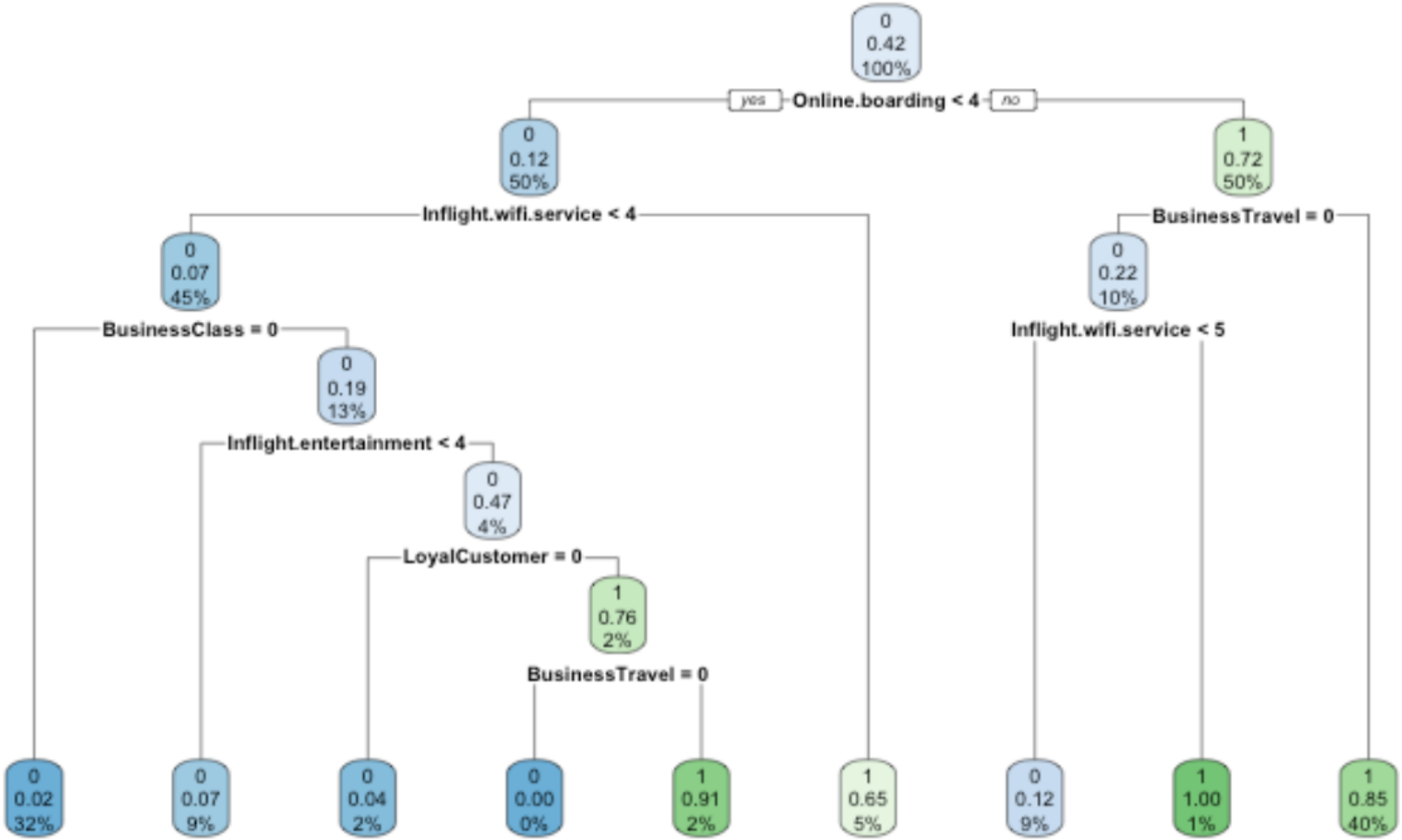
Modeling


Holdout Validation
Pruning (for tree)
Stepwise (for logistic)
Confusion Matrix
ROC Curve


True Class
Positive
Negative
Predicted Class
Positive
Negative
TP
FP
FN
TN


Dataset
Training Set
Test Set

Classification Tree



Logistic Regression

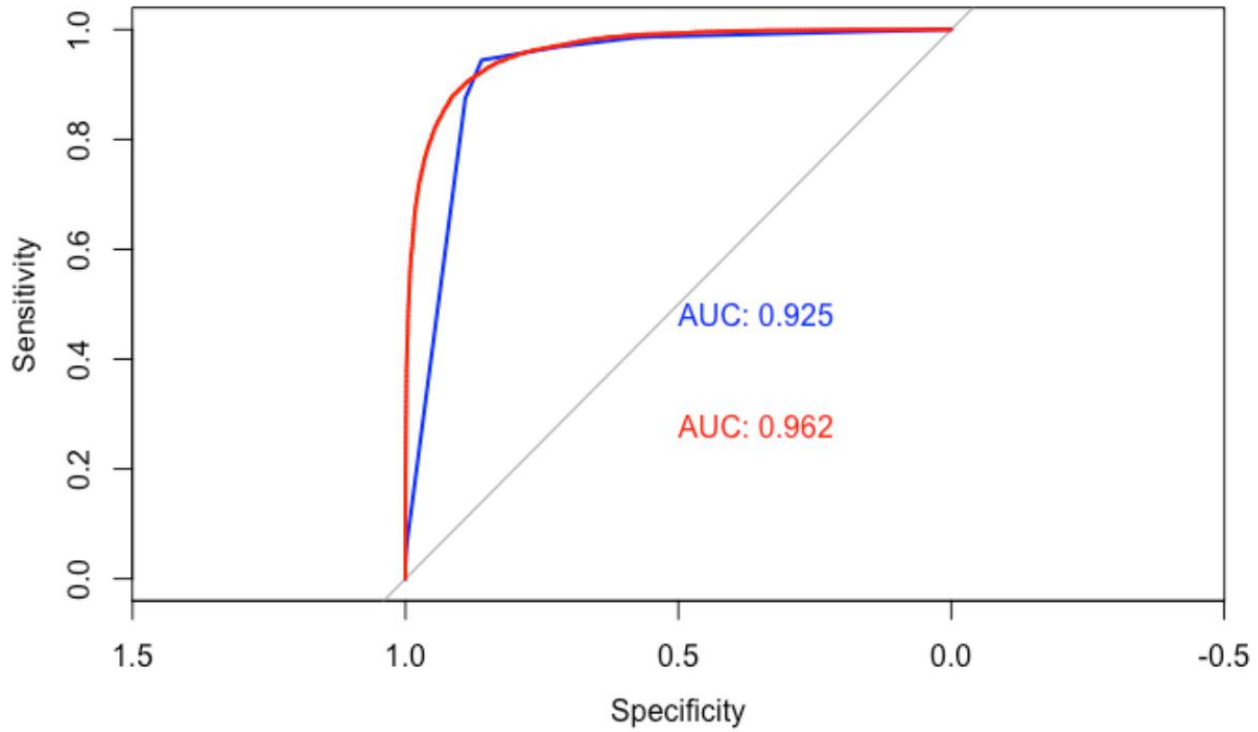
```
Coefficients:
                                  Estimate Std. Error  z value Pr(>|z|)
(Intercept)                     -1.663e+01  1.325e-01 -125.522  < 2e-16 ***
Flight.Distance                  3.245e-05  1.332e-05    2.436 0.014867 *
Inflight.wifi.service            7.915e-01  1.433e-02   55.223  < 2e-16 ***
Departure.Arrival.time.convenient -2.826e-01  1.073e-02  -26.332  < 2e-16 ***
Ease.of.Online.booking           3.692e-01  1.493e-02   24.729  < 2e-16 ***
Gate.location                    -2.673e-01  1.256e-02  -21.278  < 2e-16 ***
Food.and.drink                   -4.567e-02  1.261e-02   -3.622 0.000292 ***
Online.boarding                  9.212e-01  1.348e-02   68.345  < 2e-16 ***
Seat.comfort                     1.939e-02  1.348e-02    1.438 0.150400
Inflight.entertainment           1.003e-01  1.722e-02    5.824 5.73e-09 ***
On.board.service                 3.481e-01  1.234e-02   28.202  < 2e-16 ***
Leg.room.service                 3.119e-01  1.060e-02   29.417  < 2e-16 ***
Baggage.handling                 1.599e-01  1.380e-02   11.591  < 2e-16 ***
Checkin.service                  3.471e-01  1.021e-02   33.992  < 2e-16 ***
Inflight.service                 1.625e-01  1.441e-02   11.281  < 2e-16 ***
Cleanliness                      2.451e-01  1.405e-02   17.450  < 2e-16 ***
Departure.Delay.in.Minutes       -3.833e-03  3.045e-04  -12.586  < 2e-16 ***
Male                             2.947e-02  2.342e-02    1.258 0.208251
LoyalCustomer                    2.790e+00  3.804e-02   73.349  < 2e-16 ***
BusinessTravel                   3.387e+00  3.987e-02   84.952  < 2e-16 ***
EcoClass                         2.446e-01  4.912e-02    4.980 6.37e-07 ***
BusinessClass                    1.109e+00  5.088e-02   21.787  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

	actual	
predicted	0	1
0	13636	694
1	540	9852

	actual	
predicted	0	1
0	12196	586
1	1980	9960

Confusion Matrices

ROC Curve


Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.5
1.0
0.5
0.0
-0.5
Specificity
AUC: 0.925
AUC: 0.962

Business Case

Deployment

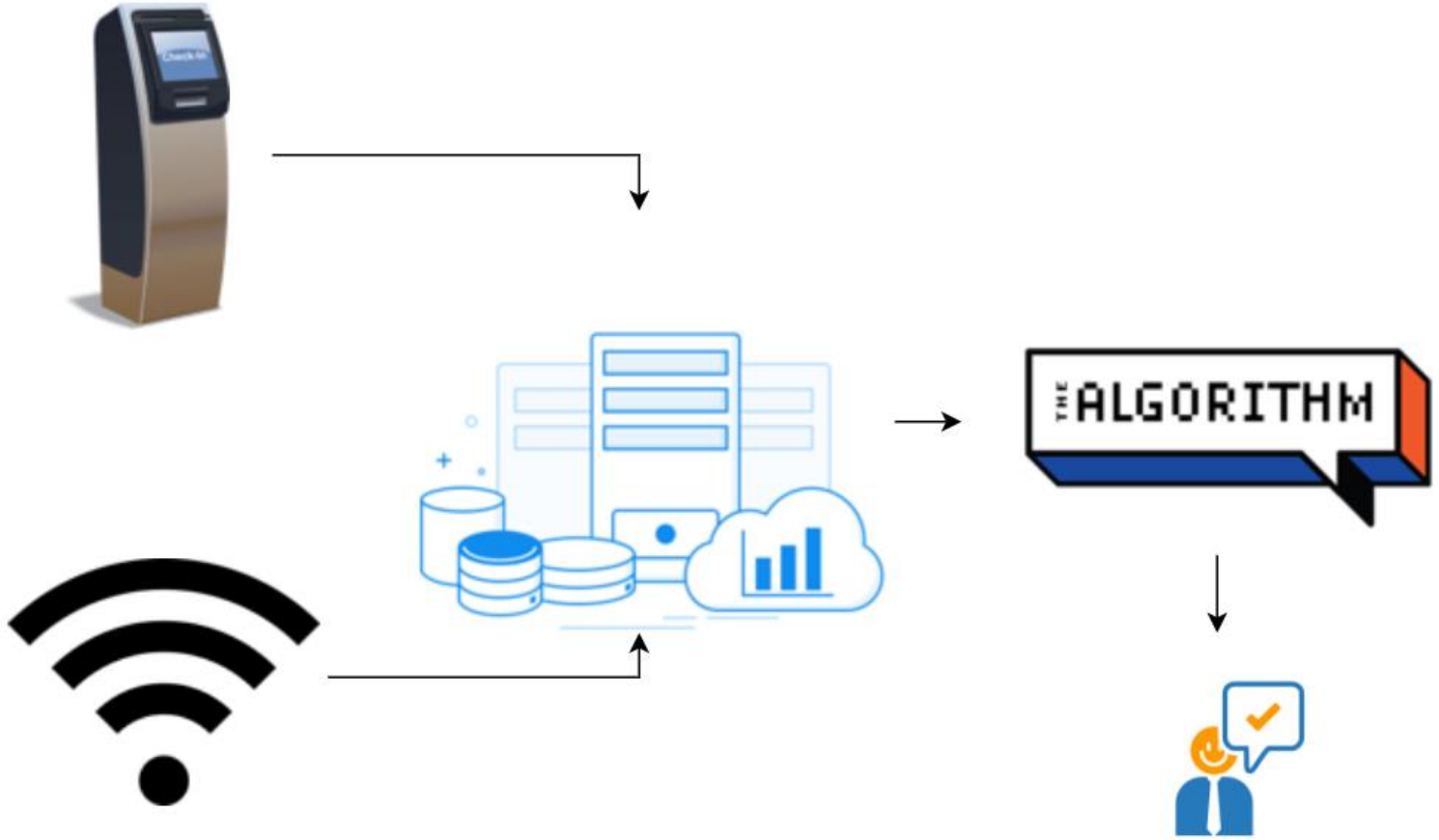
Review the online boarding process on the company's website for improvement

Reach out to either R&D or seat supplier to increase seat comfort

Increase inflight movie library

Decide between increasing WiFi capabilities and reducing the price

Increasing value of business class



Deployment cont.

Ethics and Risks

Ethics: The airline would need to be careful not to increase its value so much that it creates a class divide to the point that economy class satisfaction drops.

Risk: The process of collecting and analyzing the data should have a human component to it to make sure everything is working as planned.

Risk: Customer satisfaction is a never-ending struggle as customers will always find something for improvement. Therefore, the airline needs to set a goal at which point the company will be happy with customer satisfaction.

References

Equiskill. (2018, July 5). Understanding Logistic Regression [Graph]. https://www.equiskill.com/understanding-logistic-regression/

Klein, TJ. (2020, June). *Airline Passenger Satisfaction* (Version 1) [Data set]. Kaggle. https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction/activity

(n.d.). [Diagram of Confusion Matrix]. https://miro.medium.com/max/2102/1*fxiTNIgOyvAombPJx5KGeA.png

(n.d.). [Diagram of Training and Testing Data]. https://data-flair.training/blogs/wp-content/uploads/sites/2/2018/08/1-16.png

University of Illinois at Chicago. (n.d.). Fig Animal Tree [Graph]. http://homepages.math.uic.edu/~jan/mcs275/mcs275notes/_images/figanimaltree.png

Thank You!

Group 3

Naina Grover, Ronald Morris, Sarah Saltzman, Scott Slutsky, Mischa Subhan

Appendix

Descriptive Statistics of Relevant Variables

	Age	Flight Distance	Inflight Wifi Service	Flight Time Convenience	Ease of Online Booking
Minimum	7	31	1	1	1
1st Quartile	28	440	2	2	2
Median	40	867	3	3	3
Mean	39.86	1224	2.818	3.207	2.879
3rd Quartile	51	1775	4	4	4
Maximum	85	4983	5	5	5

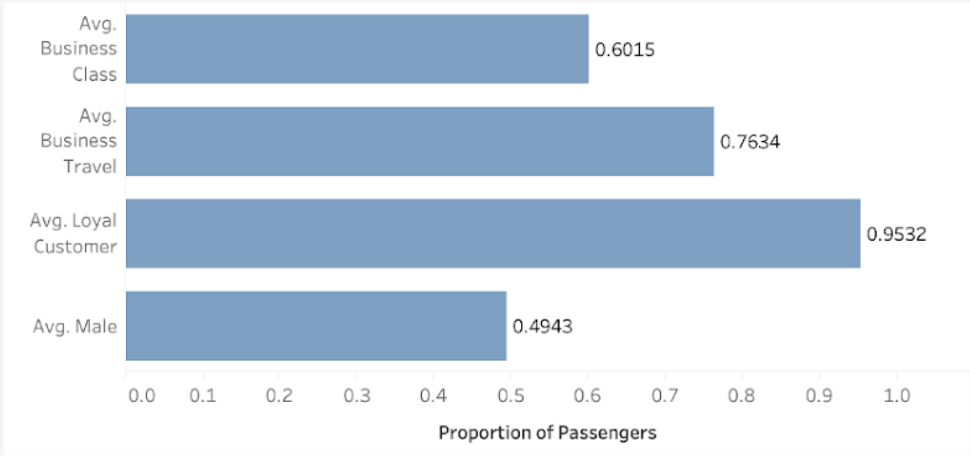
	Gate Location	Food and Drink	Online Boarding	Seat Comfort	Inflight Entertainment
Minimum	1	1	1	1	1
1st Quartile	2	2	2	2	2
Median	3	3	4	4	4
Mean	2.987	3.214	3.331	3.457	3.381
3rd Quartile	4	4	4	5	4
Maximum	5	5	5	5	5

	On Board Service	Leg Room Service	Baggage Handling	Check-in Service	Inflight Service
Minimum	1	1	1	1	1
1st Quartile	2	2	3	2	3
Median	4	4	4	3	4
Mean	3.386	3.381	3.637	3.295	3.647
3rd Quartile	4	4	5	4	5
Maximum	5	5	5	5	5

	Cleanliness	Departure Delay (minutes)	Arrival Delay (minutes)	Male	Loyal Customer
Minimum	1	0	0	0	0
1st Quartile	2	0	0	0	1
Median	3	0	0	0	1
Mean	3.294	14.84	15.28	0.4932	0.8391
3rd Quartile	4	13	13	1	1
Maximum	5	1592	1584	1	1

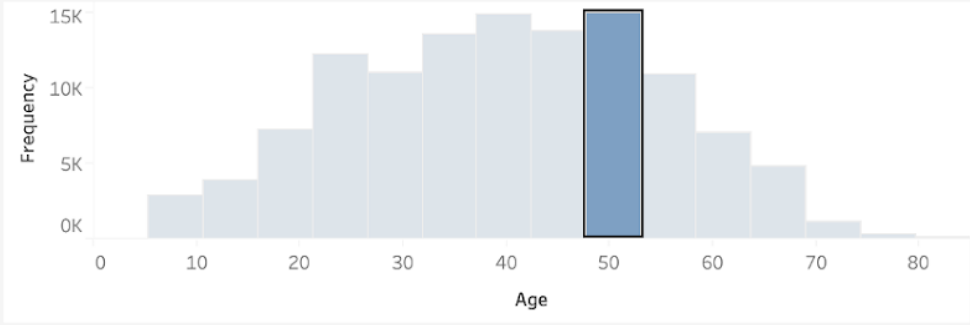
	Business Travel	Satisfied	Economy Class	Business Class
Minimum	0	0	0	0
1st Quartile	0	0	0	0
Median	1	0	0	0
Mean	0.6916	0.4268	0.4401	0.4865
3rd Quartile	1	1	1	1
Maximum	1	1	1	1

Demographics of Passengers



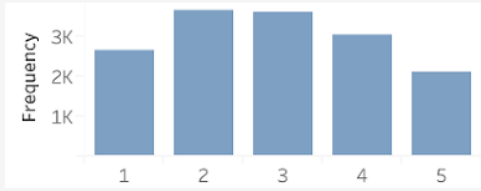
This figure presents the proportions of passengers who flew on business class, were on business travel, were loyal customers, and were male. These variables are composed solely of binary values, so computing the average of these values will give us the proportion of passengers who fit the descriptions in this figure.

Age Distribution of Passengers



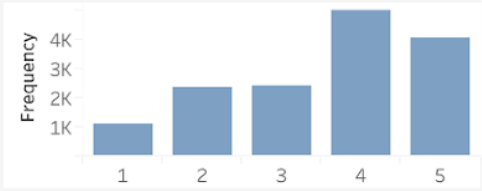
This figure presents the age distribution of passengers, which ranged from 7 to 85 years old.

Inflight WiFi Service Ratings



This figure presents the frequency of each inflight WiFi service rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Legroom Service Ratings

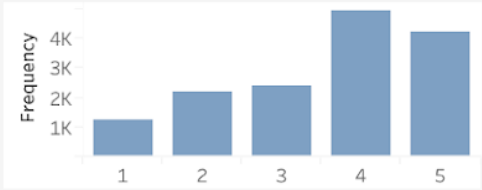


This figure presents the frequency of each legroom service rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Cleanliness Ratings

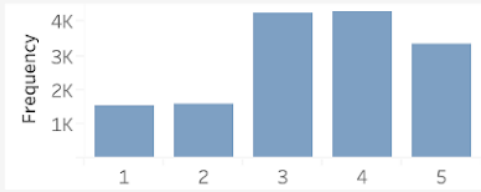
This figure presents the frequency of each cleanliness rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Inflight Entertainment Ratings



This figure presents the frequency of each inflight entertainment rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Check-in Service Ratings

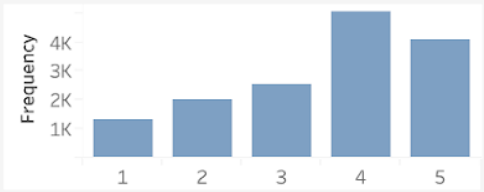


This figure presents the frequency of each checkin-in service rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Food and Drink Ratings

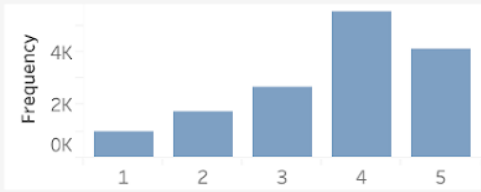
This figure presents the frequency of each food and drink rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Onboard Service Ratings



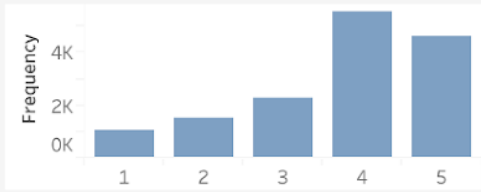
This figure presents the frequency of each onbard service rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Online Boarding Ratings



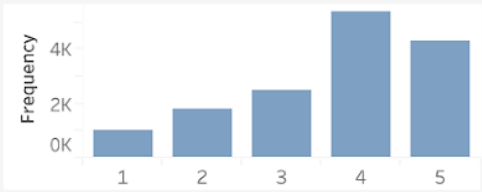
This figure presents the frequency of each online boarding rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Seat Comfort Ratings



This figure presents the frequency of each seat comfort rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).

Inflight Service Ratings



This figure presents the frequency of each inflight service rating. Ratings are discrete values that range from 1 (poor) to 5 (excellent).