ASSIGNMENT 1

"Invistico Airline"

Patricia Jolly Gariando
991 446 494
PROG 25211 - 1245_66383

Table of Contents

1. Ask a Question

Kaggle Dataset:

https://www.kaggle.com/datasets/sjleshrac/airlines-customer-satisfaction/data#

The Question:

Will I have a satisfying flight with Invistico Airlines given the following –

Gender: Female

Customer Type: Disloyal Customer

Age: 26

Type of Travel: Personal

Seat Comfort: 3

Food and Drink: 2

Inflight wifi service : 0

Inflight entertainment: 4

On-board service: 4

Leg room service: 3

Baggage handling: 5

Checkin service: 4

Cleanliness: 2

Online boarding: 3

Departure Delay in Minutes: 0

2. Clean Your Data

2.1 What changes did you make to the table and why?

2.1.1 Dropped

- Class – there are too many class types to convert though the majority fell between economy and business.
- Flight Distance – people care more about the flight time rather than how far the destination is.
- Departure/Arrival time convenient – airlines are usually on time unless there is bad weather that must delay or ground a flight.
- Gate Location – serves no importance in customer's satisfaction level. Gate location is either controlled by the airport or the airline itself.
- Online support and Ease of Online booking – wanted the algorithm to focus more on human interaction than online service.
- Arrival Delay in Minutes – if the flight departure was 3 hours late, then the arrival would also be 3 hours. Late departures have a bigger impact on a customer's satisfaction than late arrivals.

2.1.2 Altered

The following was altered so that the algorithm can process the data.

- Gender – 'Male' = 0 and 'Female' = 1.
- Satisfied – 'satisfied' = 0 and 'dissatisfied' = 1.
- Customer Type – 'Loyal Customer' = 0 and 'disloyal Customer' = 1.

3. Graph Your Data

3.1 Graph 1 – Line Plot

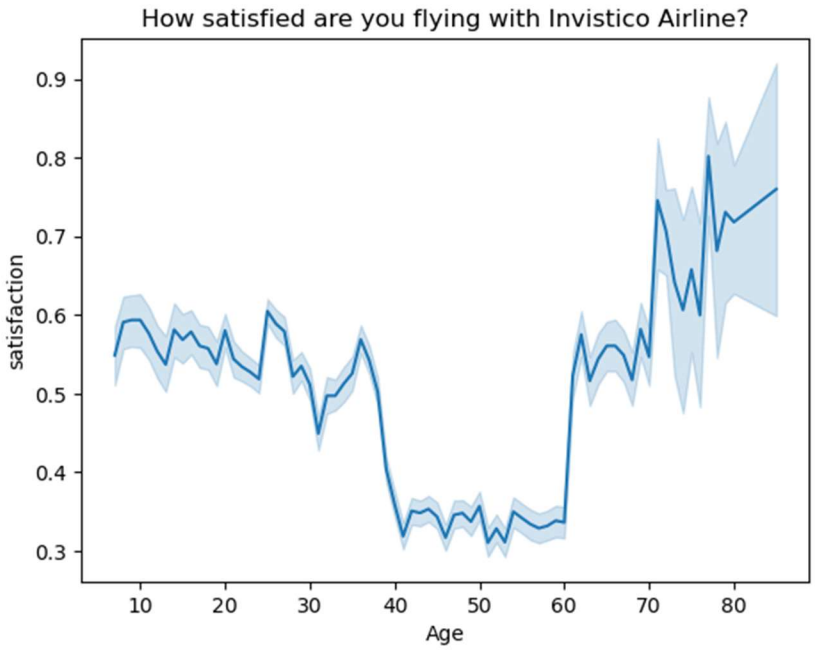


Figure 1: Passengers between mid 30 to 60 are not satisfied with the airlines.

3.2 Graph 2 – Violin Plot

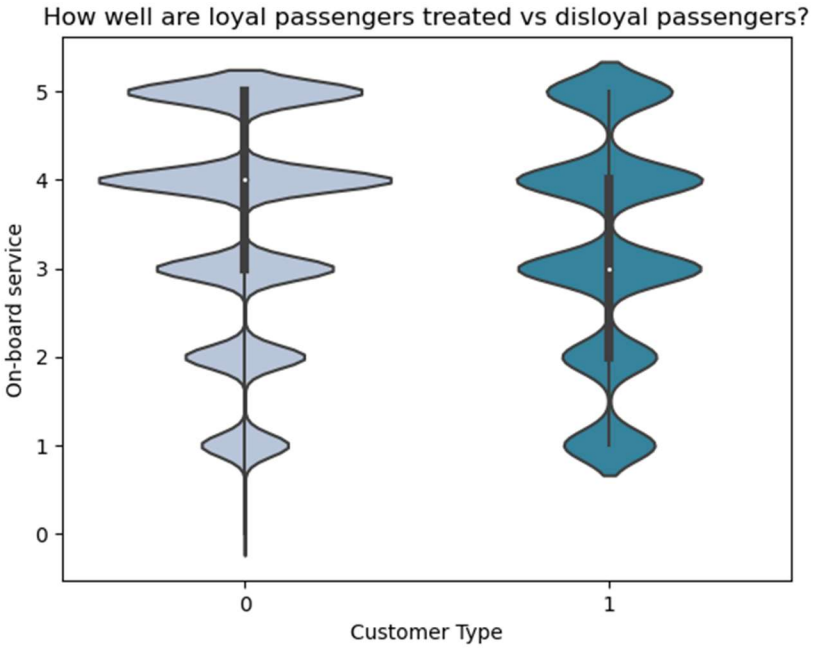


Figure 2: Loyal (0) passengers are more satisfied with the service .

3.3 Graph 3 – Bar Plot

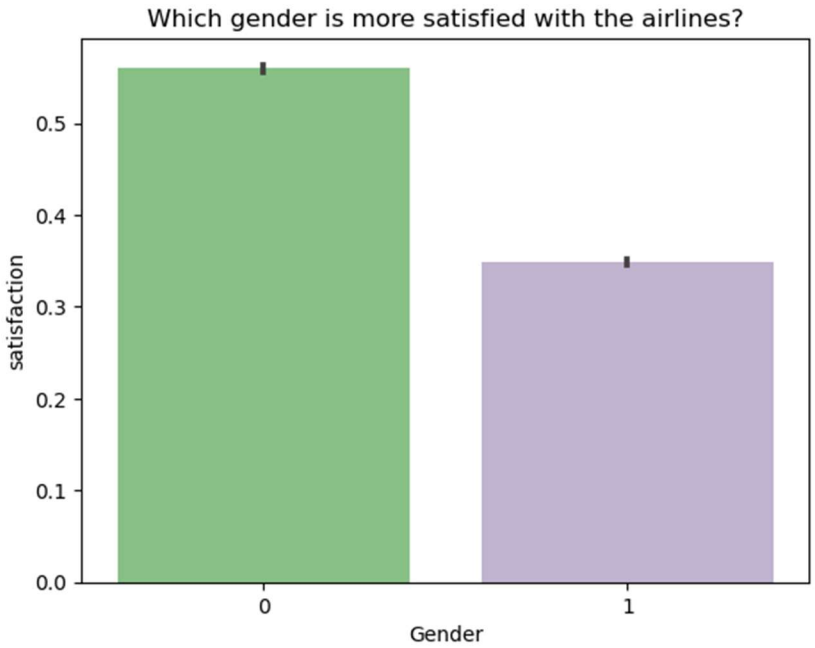


Figure 3: Males are more satisfied with the airlines than females.

3.4 Graph 4 – Box Plot

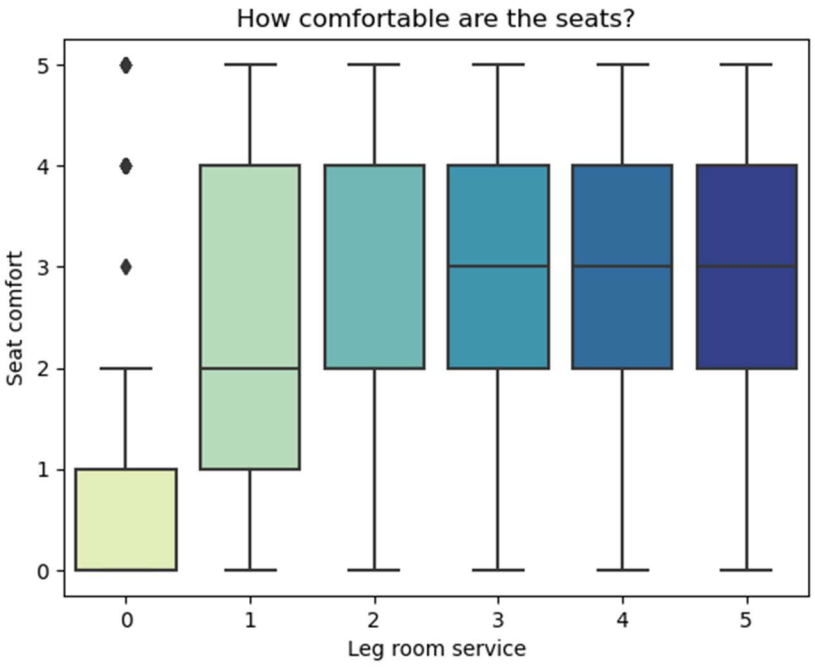


Figure 4: Passengers find the seats average.

3.5 Graph 5 – Boxen Plot

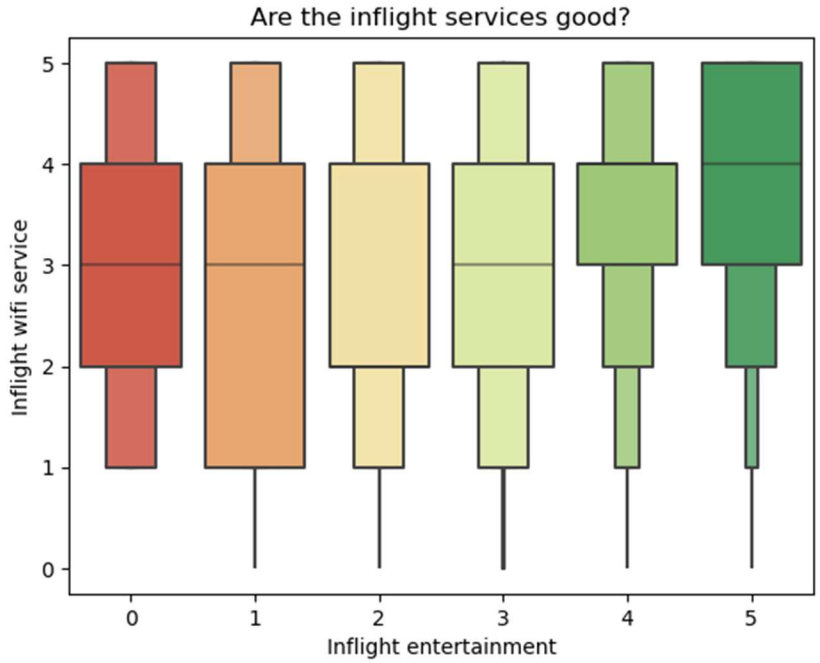


Figure 5: Majority of the passengers find the inflight services to be adequate.

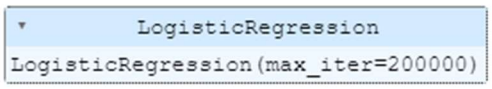
4. Train and Test Your Algorithm

4.1 Code

```
# We need to split our data into x and y. We use X to predict y.
X, y = airline.drop('satisfaction', axis=1), airline['satisfaction'].values


# Split the data into a training set and a testing set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=87)


# Train our algorithm
from sklearn.linear_model import LogisticRegression
logmodel = LogisticRegression(max_iter=200000)
logmodel.fit(X_train, y_train)
```



5. Evaluate Your Model

5.1 How accurate is the model?

The model is 0.83% accurate.

To make my model more accurate, I would have included more columns that could have changed the outcome of the results. In part 2, columns such as ‘Class’ and ‘Flight Distance’ were dropped. The type of ‘Class’ that passengers are in can change the satisfaction rate. Customers would want to be comfortable during a lengthy flight but not everyone gets to experience that luxury. In addition, how passengers are treated on a flight that is 5000km+ compared to something that’s 100km can influence the ‘On-board service’ rate. Are they treated well throughout the duration of the flight, are they treated well during the first half then badly the next or vice-versa.

My model may have not been as accurate as I would like it to be because there could have been duplicate entries in the .csv file that I am not aware of. There are more than 100,000 entries so there is a possibility. Rather than having a variety of information, we are dealt with biased and repetitive results. Moreover, if the ratings were scaled from 1 to 10 inside of 1 to 5, there could have been more information to analyze.

6. Conclusion

6.1 Answer the Question

To answer the question from part 1, I will NOT have a satisfying flight with Invistico Airlines.

6.2 Closing Statement

In conclusion, the algorithm is capable of making predictions with an 83% accuracy which means that there is room for improvement. We can adjust our input and see how changing the ratings for each column affects our results and possibly factor in the columns that were dropped during the cleaning stage. All in all, the following algorithm can provide insight for many airlines, using their own data, on ways to improve their services so that passengers will leave happy.