

AIRLINE PASSENGER SATISFACTION



TEAM MEMBERS

- Akshaya ES
- Nikita Ann George
- Pala Sheshu
- Ujjwal Kumar Sharma
- Venkat Paritala
- Viresh Raj Sah

MENTOR:

Mr Srikar Muppidi

PROBLEM STATEMENT

As the old adage goes: "Customers are king."

This is an age-old mantra reflecting the importance of customers in every business. With the recent pandemic completely disrupting the way we travel and having a huge impact on the aviation industry, one must now ponder on how we can get back to the norm.

What can airlines prepare to give themselves a competitive edge when the crowd finally arrives?

We believe that to recover from the disastrous impact the pandemic had on the industry, we need to find a way to engage passengers in a better manner, by enhancing their experience and ultimately gaining their loyalty.

With this project we aim to do just that by creating a highly precise classification model for airlines to identify critical bottlenecks to raise passenger satisfaction.

DATA DESCRIPTION

- 129880 observations
- 25 features

VARIABLE	DESCRIPTION
1.Unnamed:0	Index value
2. Id	Unique identifier
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

VARIABLE	DESCRIPTION
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 PREPROCESSING

1. Treatment of Null values:

We found null values in the feature: Arrival Delay in Minutes. The percentage of null values was around 0.3%. We decided to drop these values.

2. Outlier treatment:

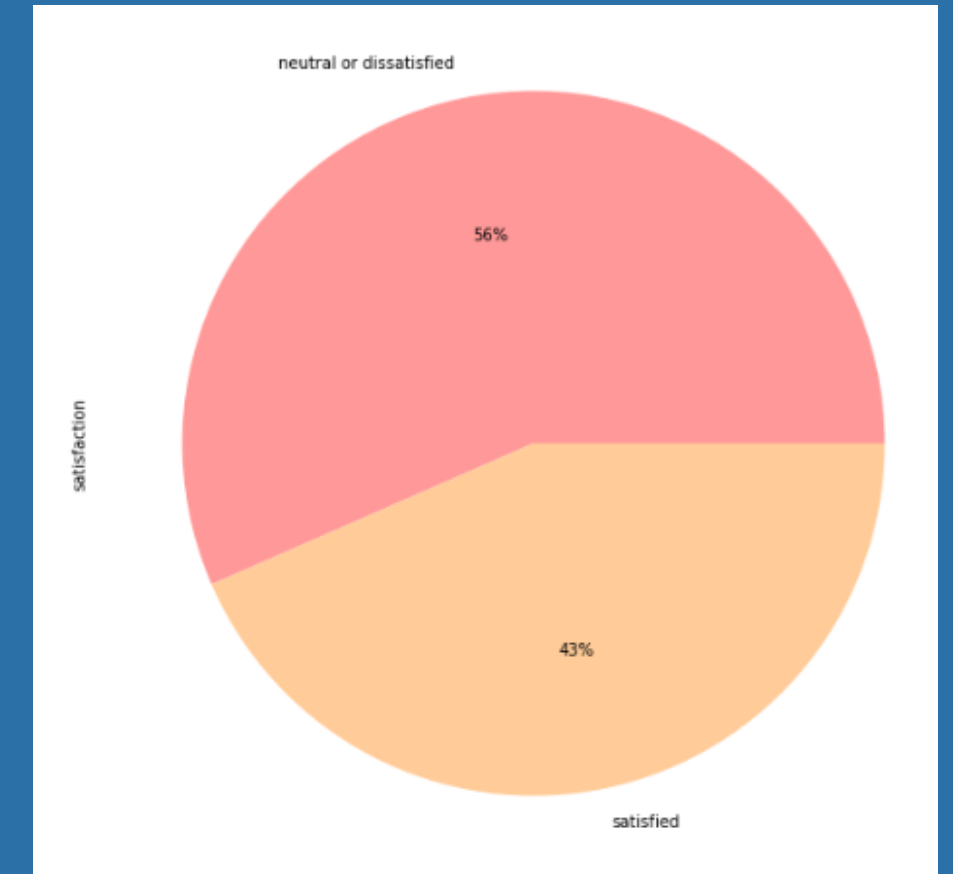
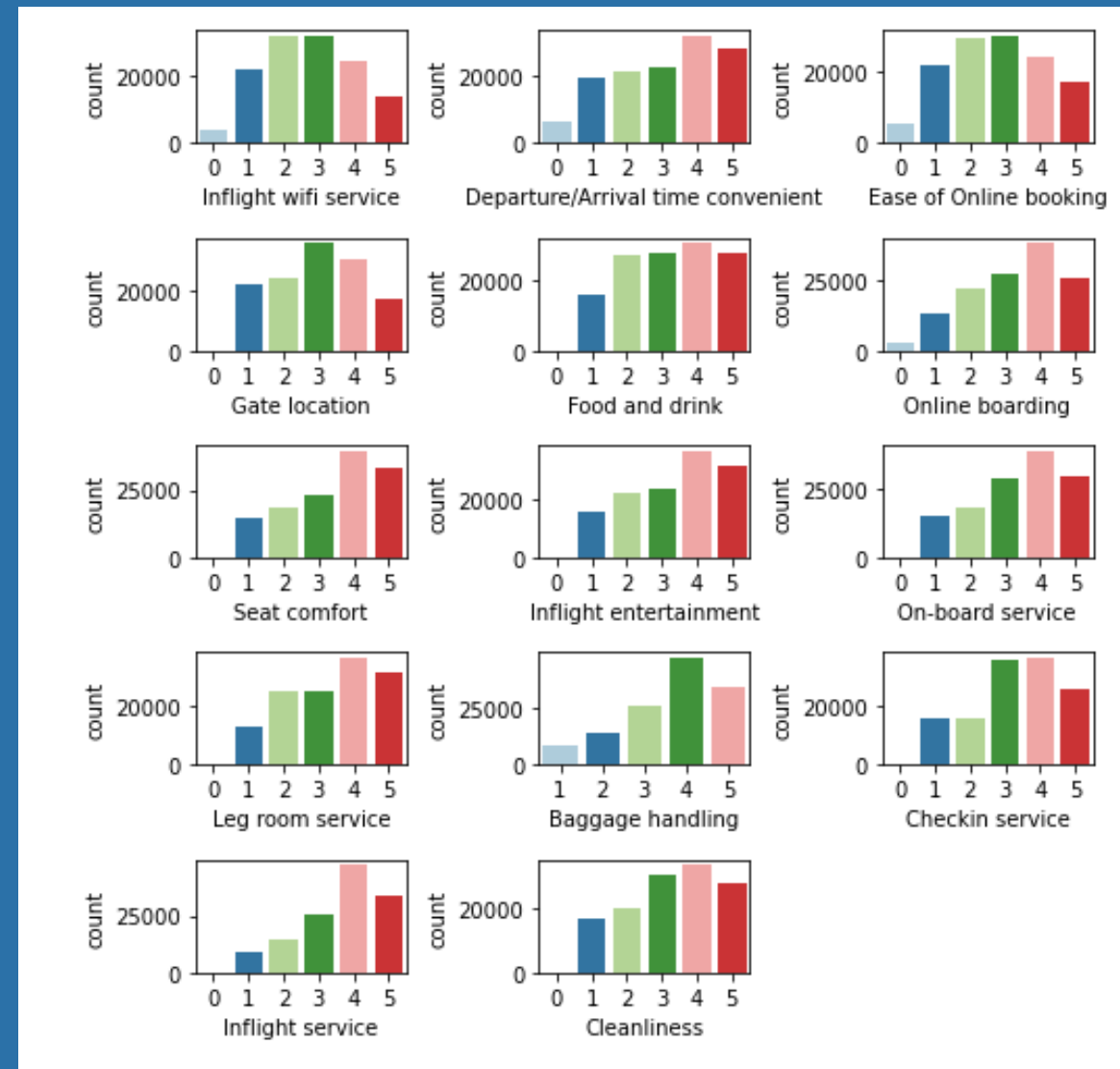
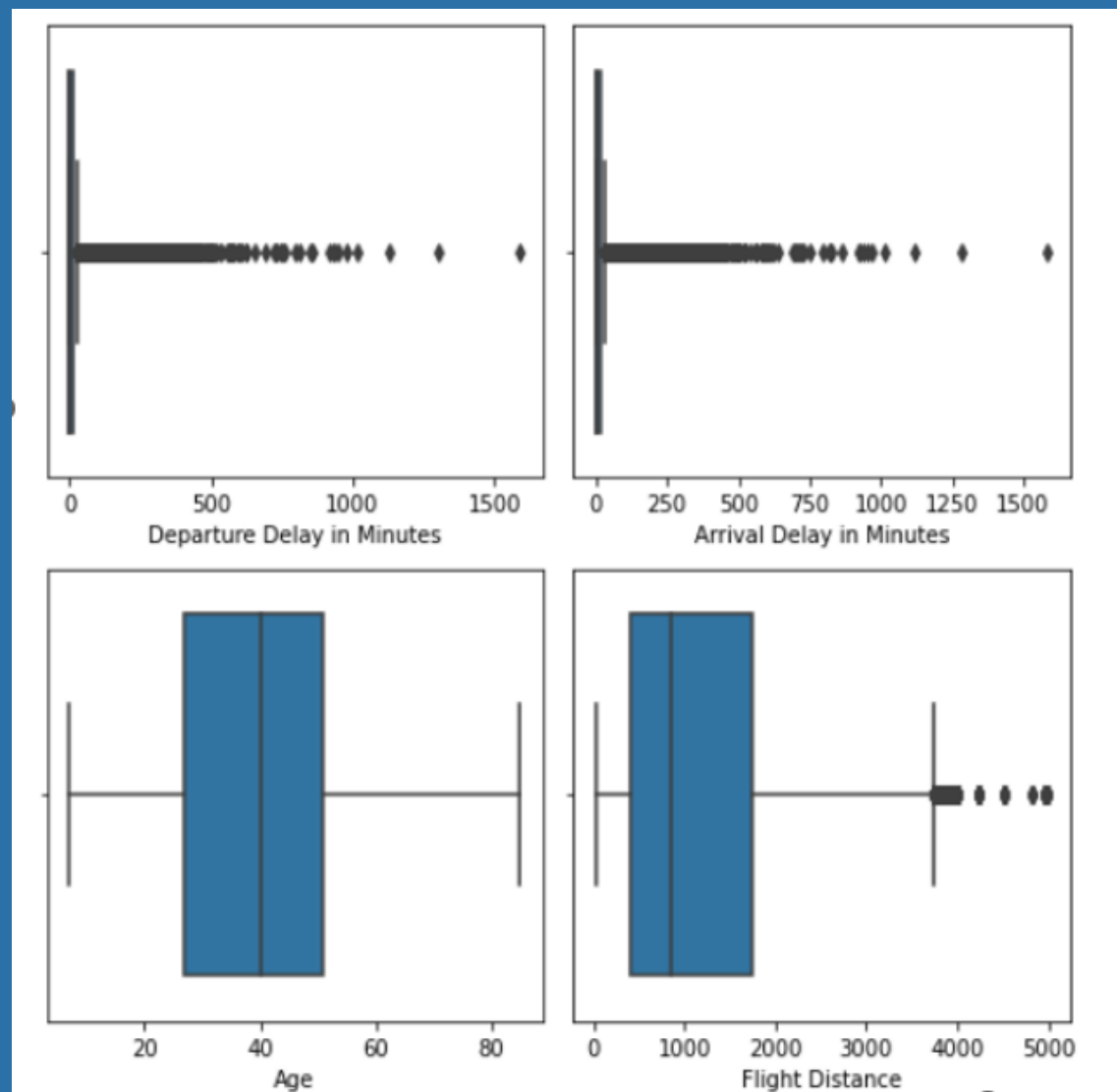
There are outliers present in the Flight Distance, Departure Delay in Minutes and Arrival Delay in Minutes column. As these variables play a role in customer satisfaction (more the delay, we can assume less satisfied the customer), we are not excluding the outliers. As we proceed with model building we can transform the extreme outliers. But for the base model we choose to leave the outliers as it is.

3. Insignificant columns:

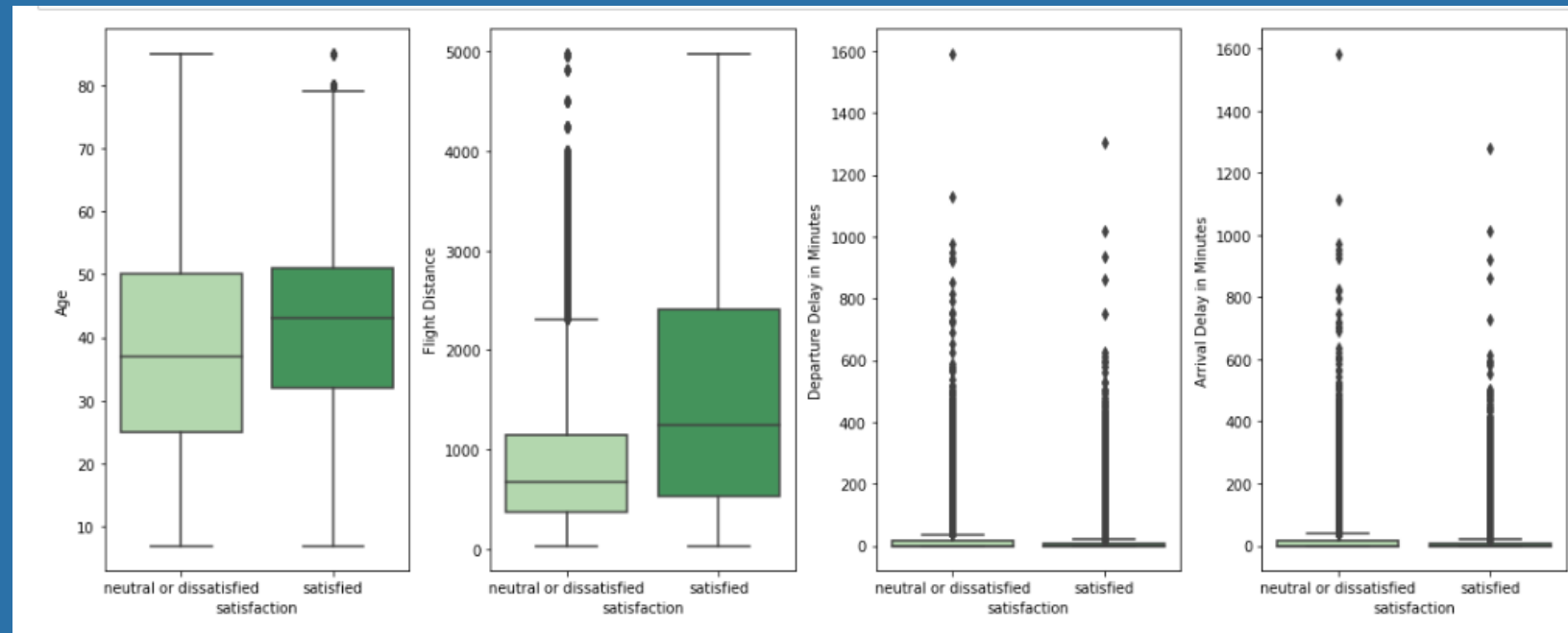
Unnamed: 0 – Index value

Id – Unique identifiers of each row, has no impact on the model

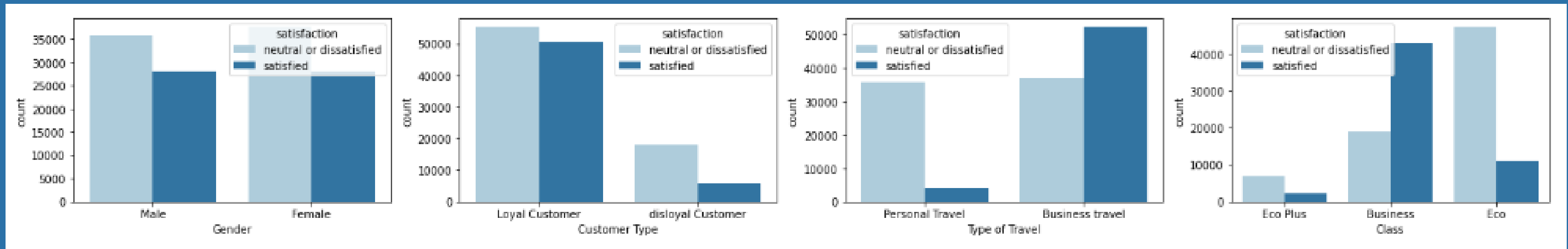
UNIVARIATE ANALYSIS



BIVARIATE ANALYSIS



- **Age :**The average age for satisfied and dissatisfied were found to be 37 and 41 respectively
- **Flight Distance:** The average distance travelled by satisfied customers was higher than that of neutral/dissatisfied



- Disloyal customers were found to be more dissatisfied than satisfied
- Business travelers' are highly satisfied
- Moreover passengers business class get high satisfied count, while passengers in the eco and eco plus class were found to be more dissatisfied



BIVARIATE ANALYSIS

- The Main Causes of dissatisfaction are: In-Flight Wifi, Online Booking and Online boarding
- Less But Observable Dissatisfaction is caused by Leg-Room Services and Cleanliness
- Some Dissatisfaction is also caused by Gate Location, Food and Drinks, Seat Comfort, Inflight Entertainment, On Board Services
- Online Boarding Shows generally satisfactory ratings, but is one of the main cause of dissatisfaction among the dissatisfied customers

STATISTICAL TEST

Chi2_Contingency

	p_value
Gender	5.135590e-05
Customer Type	0.000000e+00
Type of Travel	0.000000e+00
Class	0.000000e+00
Inflight wifi service	0.000000e+00
Departure/Arrival time convenient	9.767302e-128
Ease of Online booking	0.000000e+00
Gate location	0.000000e+00
Food and drink	0.000000e+00
Online boarding	0.000000e+00
Seat comfort	0.000000e+00
Inflight entertainment	0.000000e+00
On-board service	0.000000e+00
Leg room service	0.000000e+00
Baggage handling	0.000000e+00
Checkin service	0.000000e+00
Inflight service	0.000000e+00
Cleanliness	0.000000e+00

Skewness

	Skewness
Age	-0.003376
Flight Distance	1.108433
Departure Delay in Minutes	6.853578
Arrival Delay in Minutes	6.670125

Kruskal Wallis test

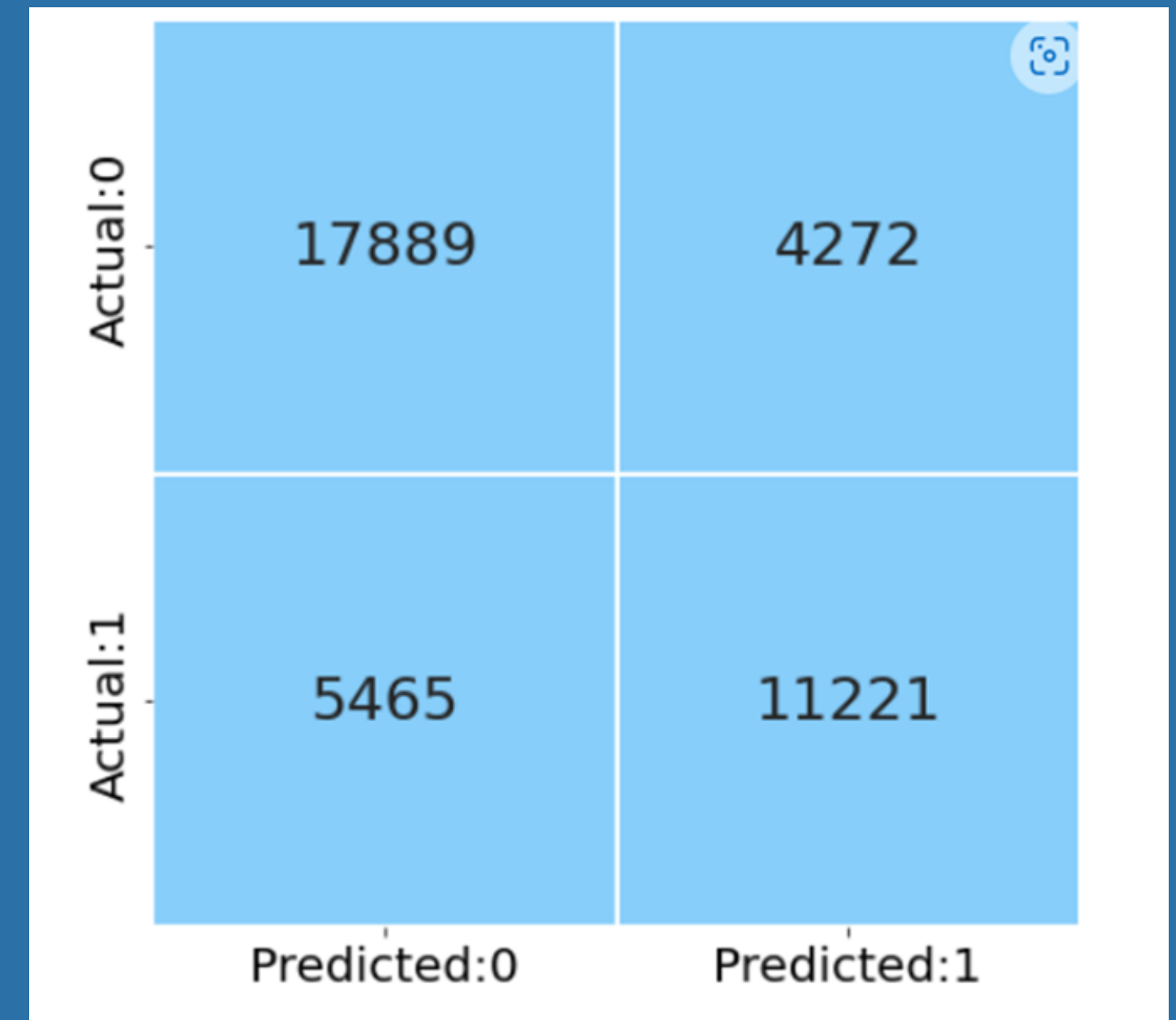
	p_value
Age	0.000000e+00
Flight Distance	0.000000e+00
Departure Delay in Minutes	3.580163e-137
Arrival Delay in Minutes	7.087659e-287

BASE MODEL

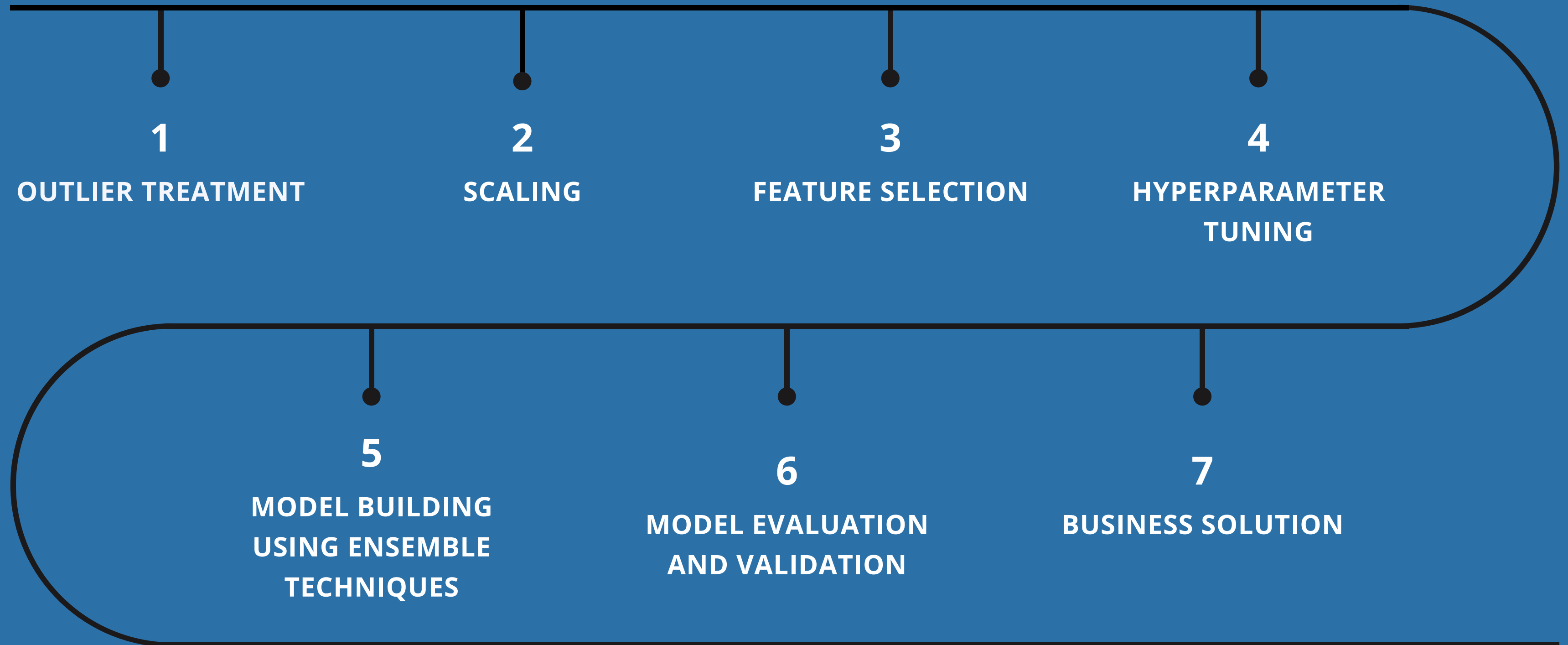
LOGISTIC REGRESSION

```
from sklearn.linear_model import LogisticRegression
LR=LogisticRegression()
model1=LR.fit(xtrain,ytrain)
ypred=model1.predict(xtest)
from sklearn.metrics import classification_report
print(classification_report(ytest,ypred))
```

	precision	recall	f1-score	support
0	0.86	0.84	0.85	22161
1	0.79	0.82	0.81	16686
accuracy			0.83	38847
macro avg	0.83	0.83	0.83	38847
weighted avg	0.83	0.83	0.83	38847



FUTURE WORK



THANK YOU