Flight Ticket Price Predictor

Final Project
DA - Ironhack
Clara Balcells

Intro to the dataset

Dataset contains information about flight booking options from the website Easemytrip for flight travel between India's top 6 cities.

1) Airline: The name of the airline company is stored in the airline column. It is a categorical feature having 6 different airlines.

2) Flight: Flight stores information regarding the plane's flight code. It is a categorical feature.

3) Source City: City from which the flight takes off. It is a categorical feature having 6 unique cities.

4) Departure Time: This is a derived categorical feature obtained created by grouping time periods into bins.

5) Stops: A categorical feature with 3 distinct values that stores the number of stops between the source and destination cities.

6) Arrival Time: This is a derived categorical feature created by grouping time intervals into bins. It has six distinct time labels and keeps information about the arrival time.

7) Destination City: City where the flight will land. It is a categorical feature having 6 unique cities.

8) Class: A categorical feature that contains information on seat class; it has two distinct values: Business and Economy.

9) Duration: A continuous feature that displays the overall amount of time it takes to travel between cities in hours.

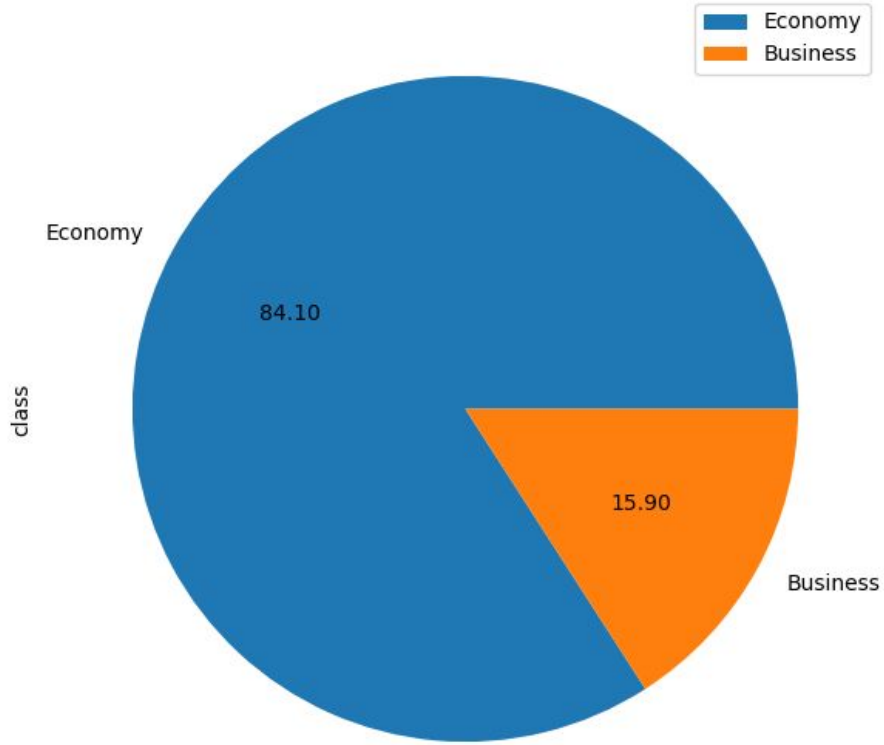
10)Days Left: This is a derived characteristic that is calculated by subtracting the trip date by the booking date.

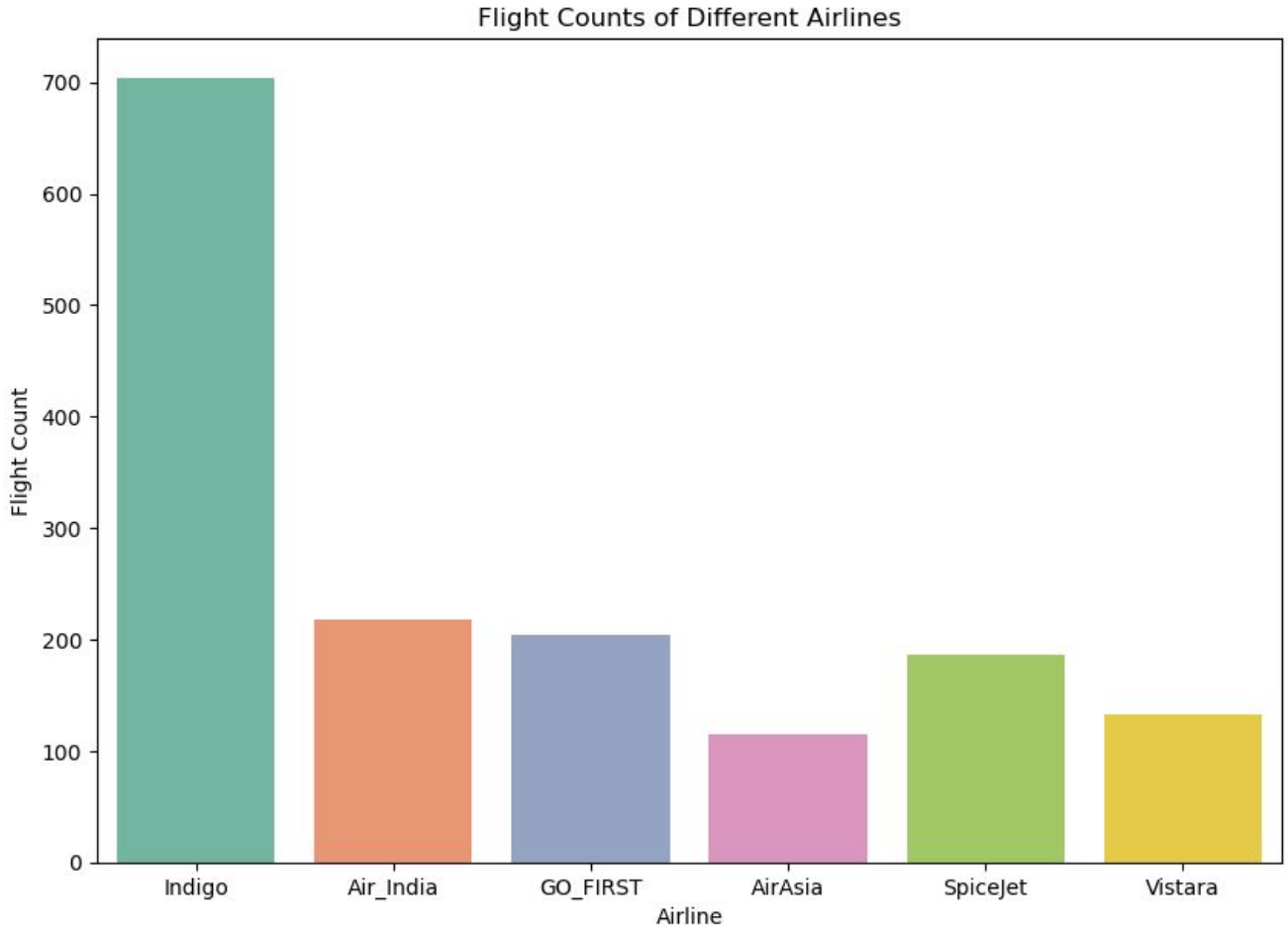
11) Price: Target variable stores information of the ticket price.

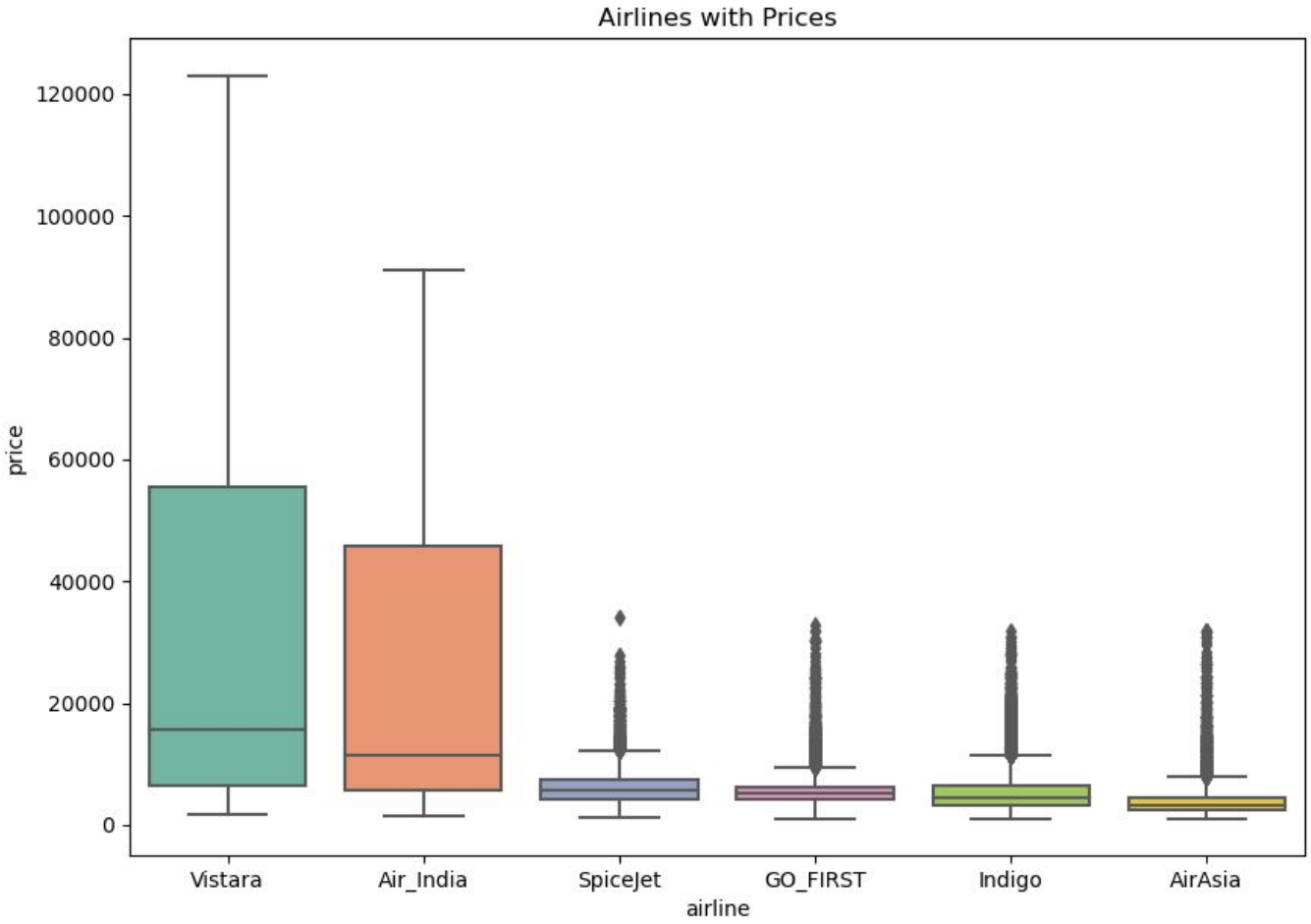
Business Questions

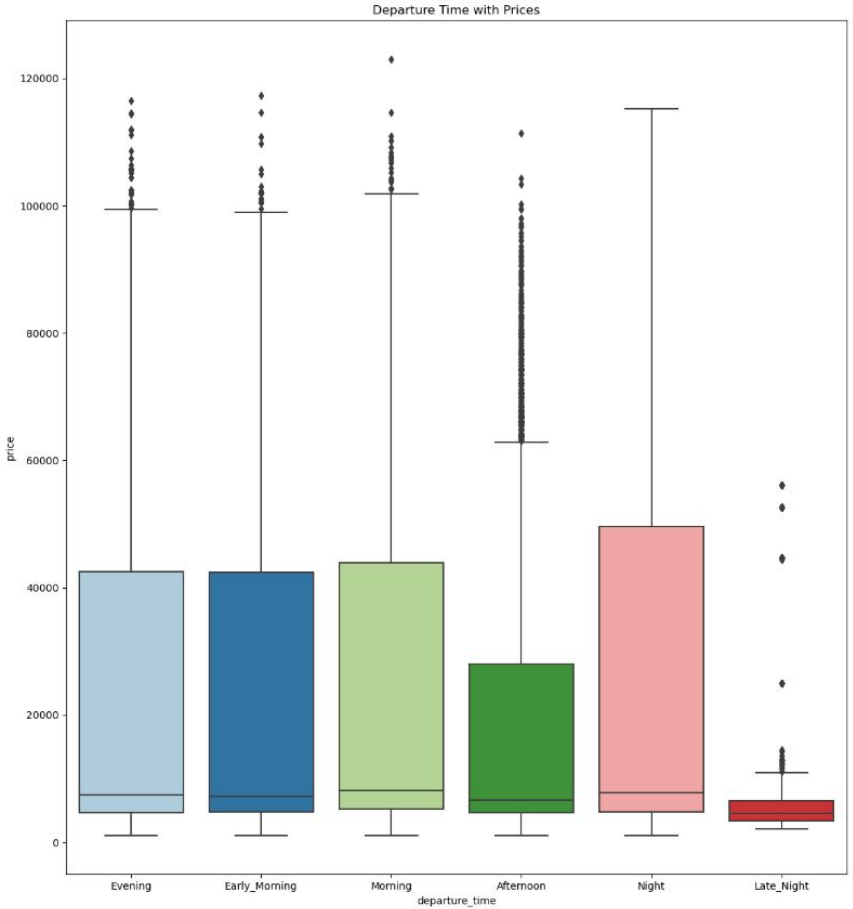
- How does the ticket price vary between Economy and Business class?
- Does price vary within Airlines?
- How the price changes when tickets are bought before departure?
- Does ticket price change based on the departure time and arrival time?
- How the price changes based on Source and Destination?
- Does the number of stops influences the price?

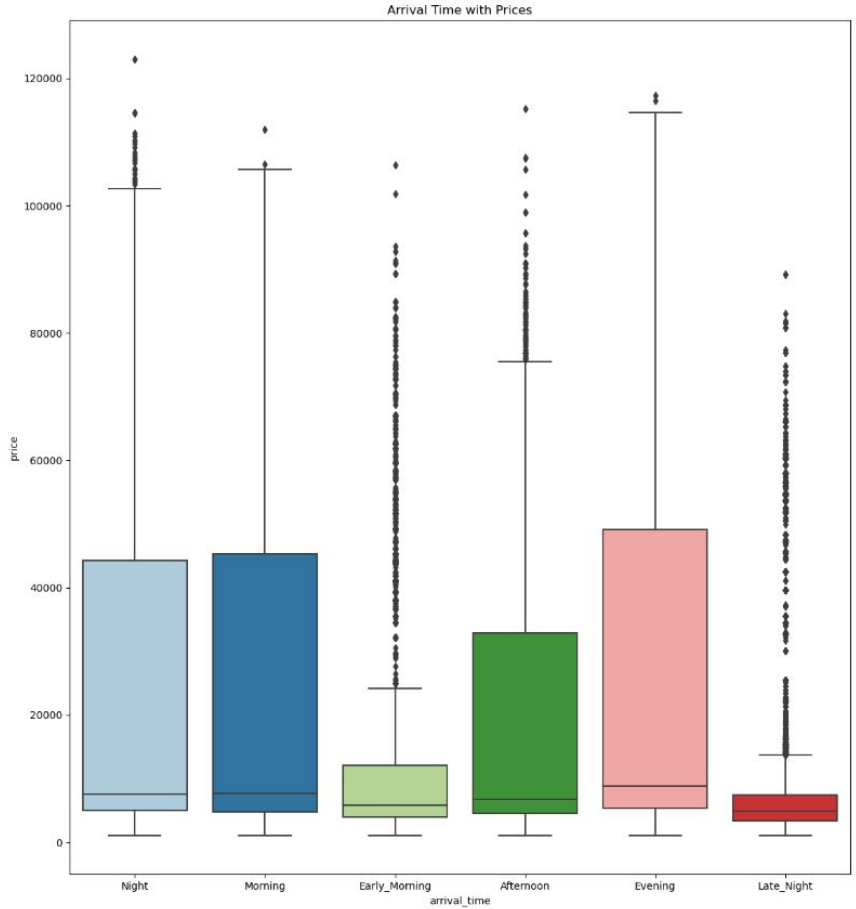
Initial EDA

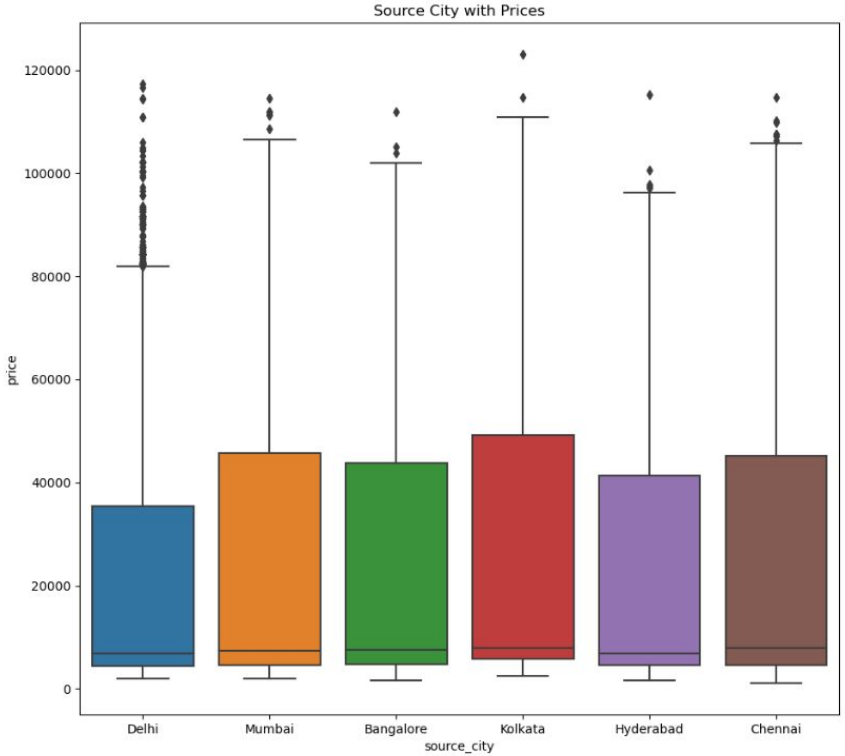


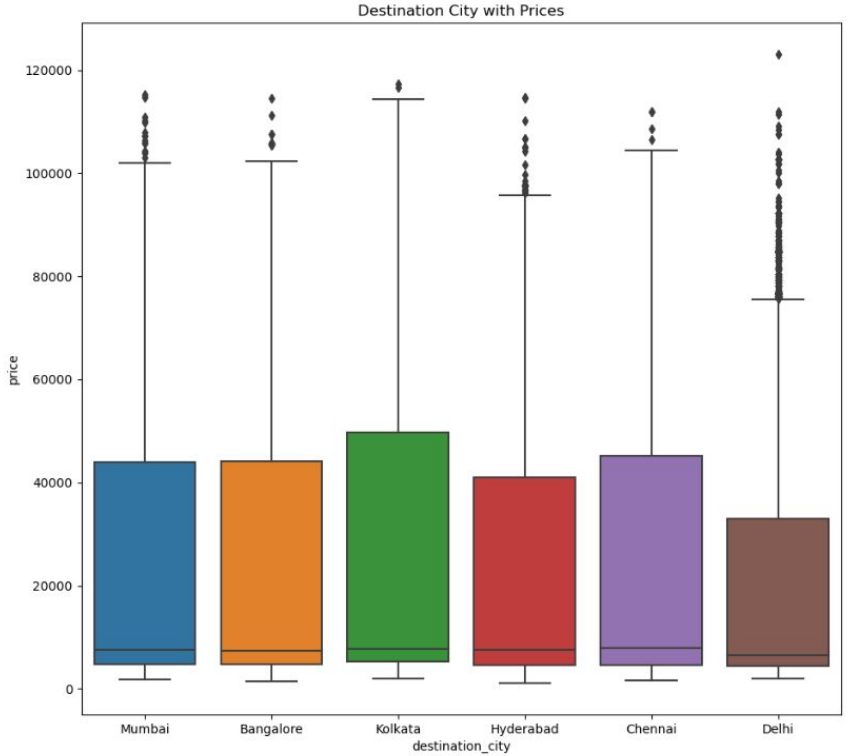

Flight Counts of Different Airlines
Flight Count
0
100
200
300
400
500
600
700
Indigo
Air_India
GO_FIRST
AirAsia
SpiceJet
Vistara
Airline

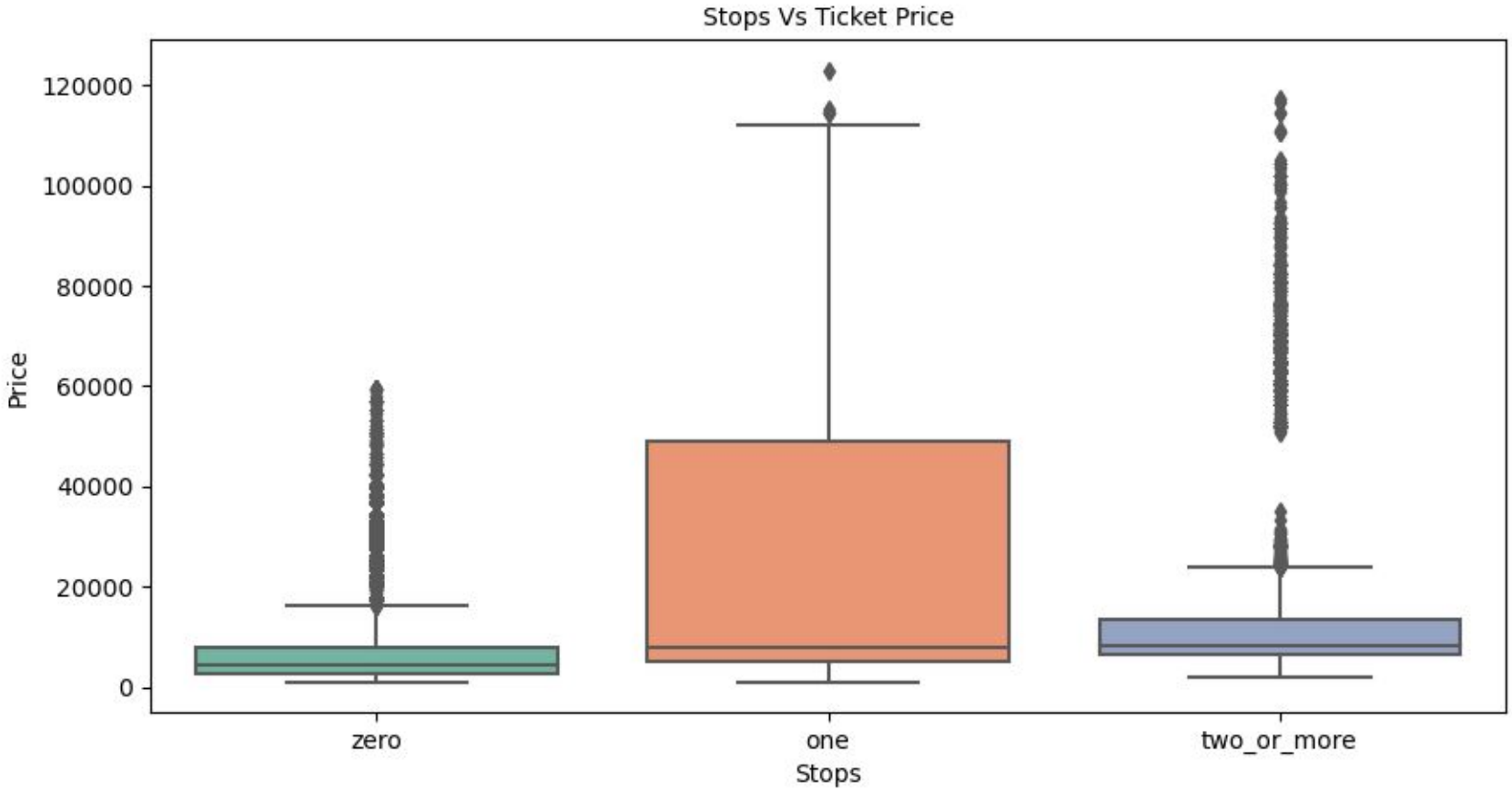

Airlines with Prices
120000
100000
80000
60000
40000
20000
0
price
Vistara
Air_India
SpiceJet
GO_FIRST
Indigo
AirAsia
airline

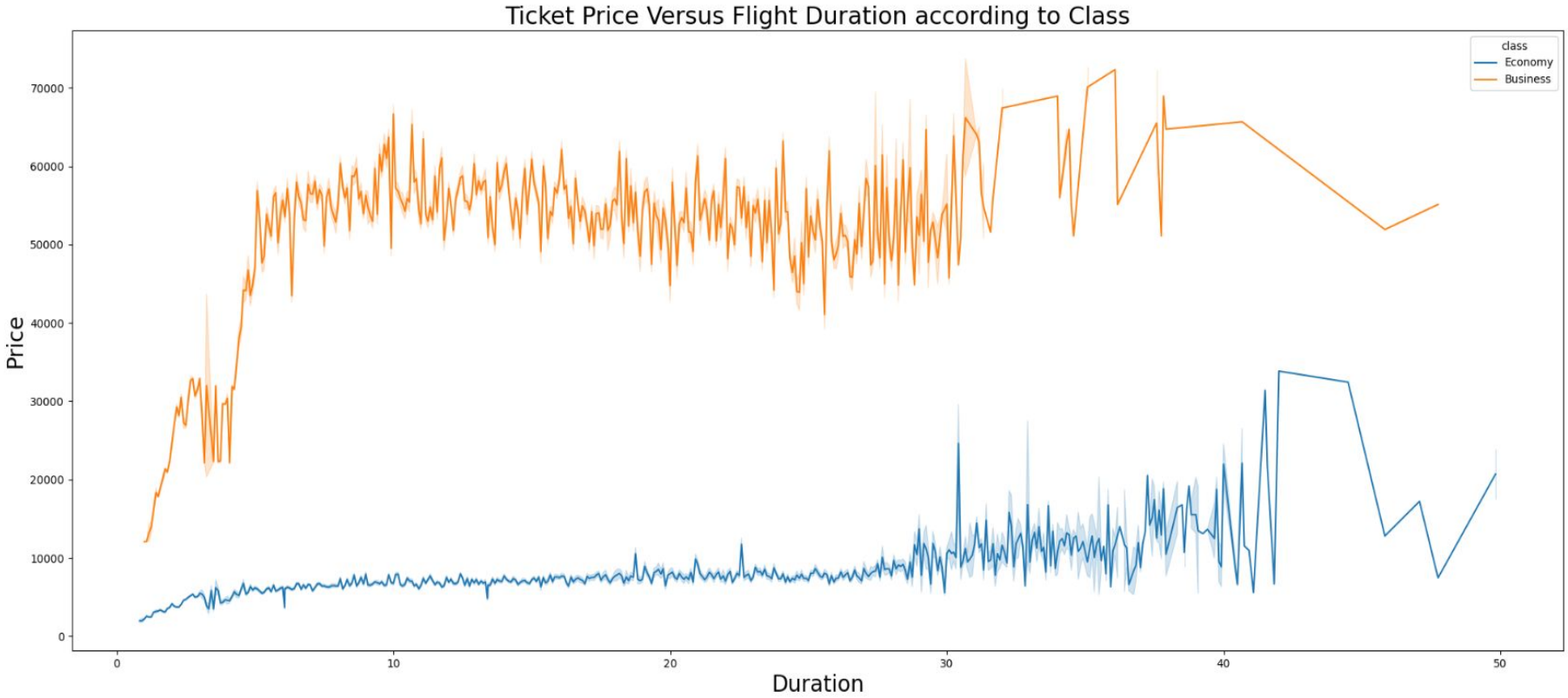

Departure Time with Prices
price
120000
100000
80000
60000
40000
20000
0
Evening
Early_Morning
Morning
Afternoon
Night
Late_Night
departure_time

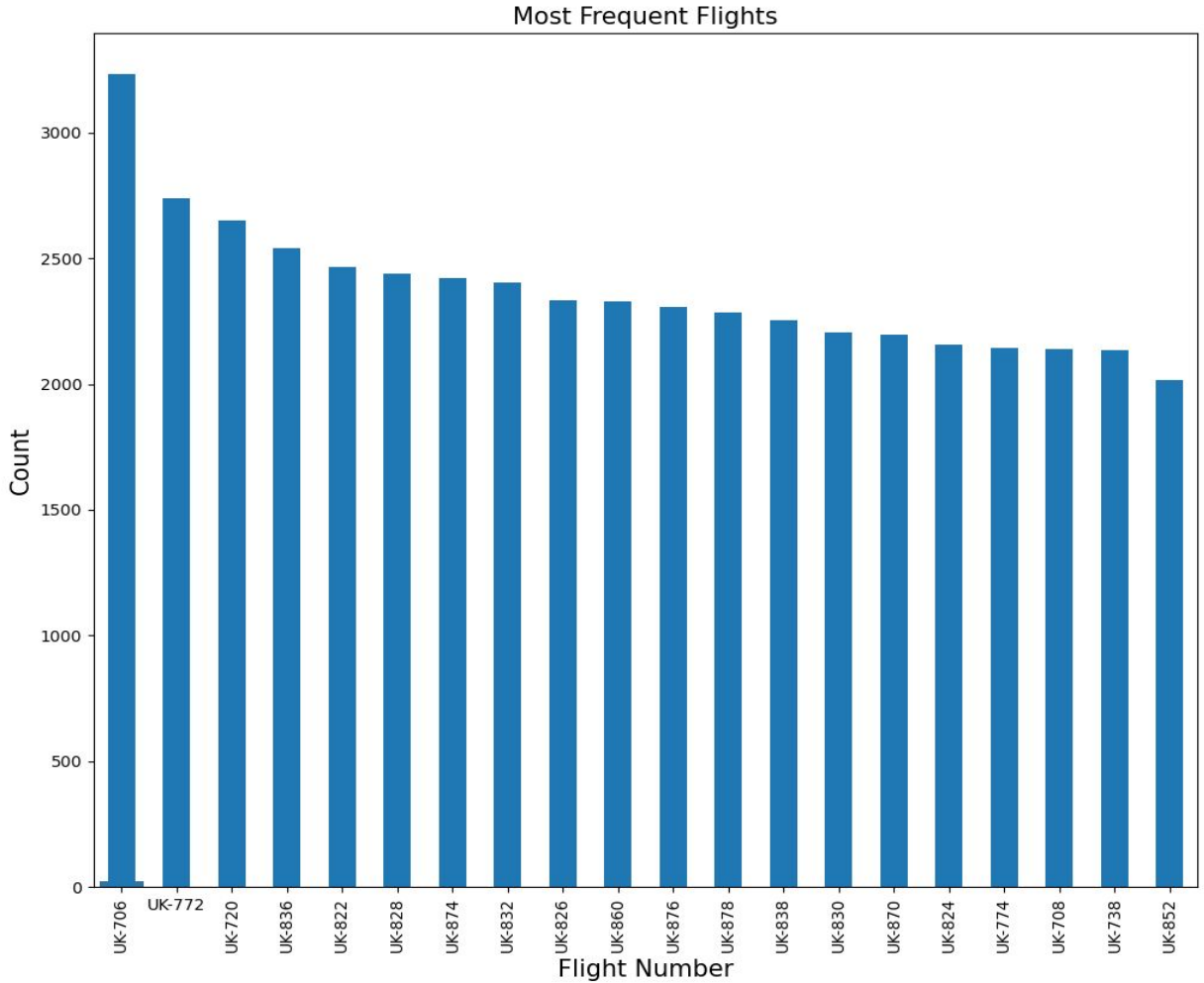

Arrival Time with Prices
price
120000
100000
80000
60000
40000
20000
0
Night
Morning
Early_Morning
Afternoon
Evening
Late_Night
arrival_time

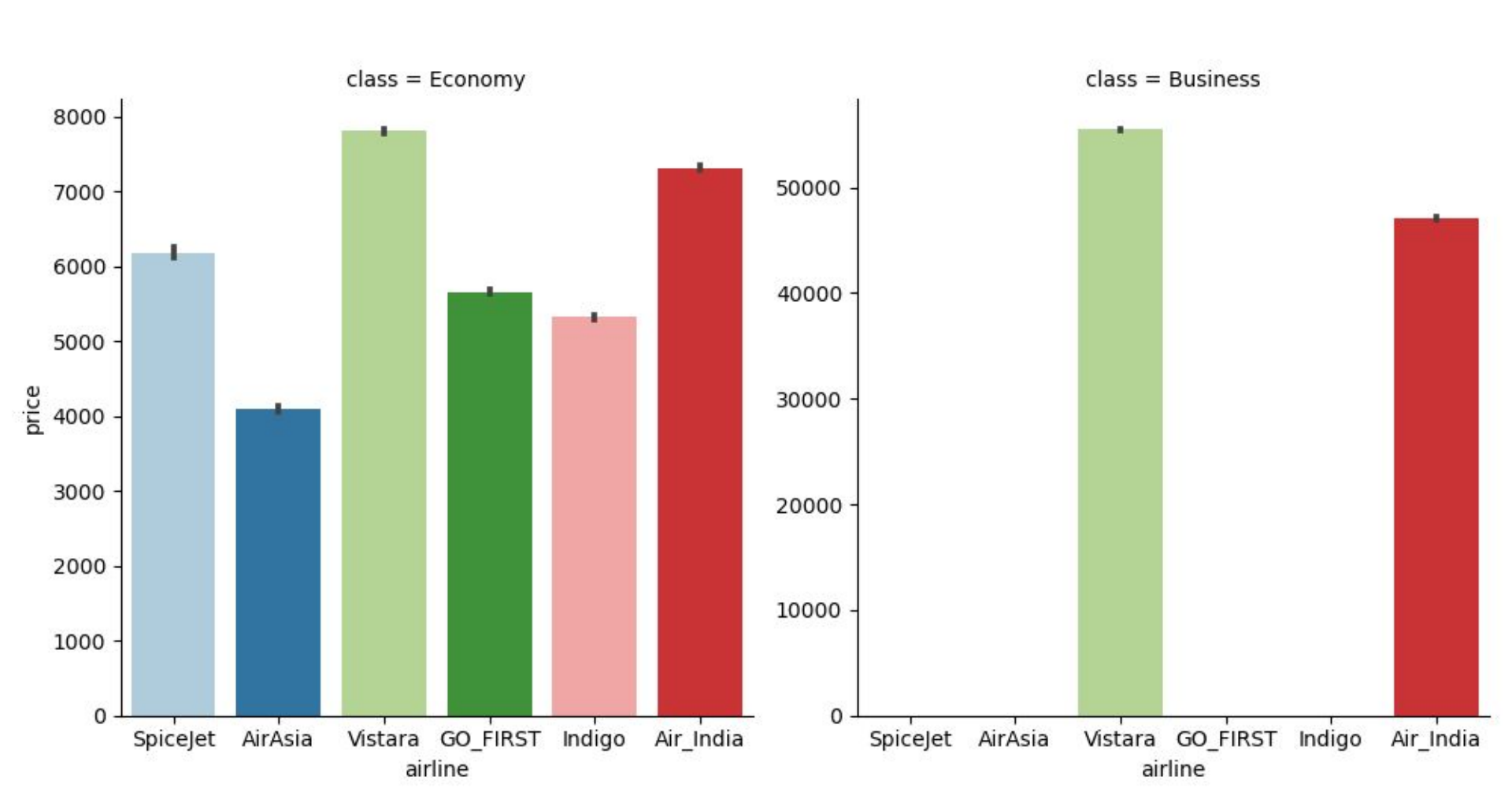

Source City with Prices
price
120000
100000
80000
60000
40000
20000
0
Delhi
Mumbai
Bangalore
Kolkata
Hyderabad
Chennai
source_city

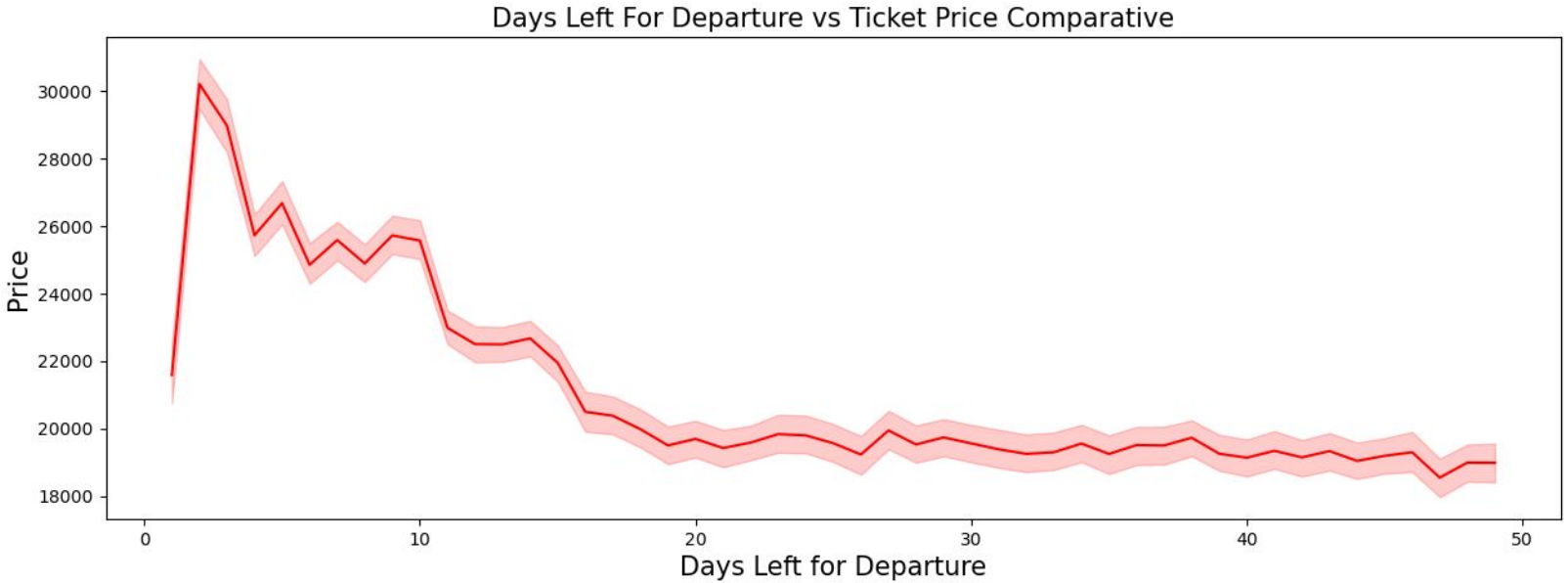

Destination City with Prices
price
120000
100000
80000
60000
40000
20000
0
Mumbai
Bangalore
Kolkata
Hyderabad
Chennai
Delhi
destination_city


Stops Vs Ticket Price
120000
100000
80000
60000
40000
20000
0
Price
zero
one
two_or_more
Stops


Ticket Price Versus Flight Duration according to Class
class
Economy
Business
Price
Duration
0
10000
20000
30000
40000
50000
60000
70000
0
10
20
30
40
50


Most Frequent Flights
Count
0
500
1000
1500
2000
2500
3000
UK-706
UK-772
UK-720
UK-836
UK-822
UK-828
UK-874
UK-832
UK-826
UK-860
UK-876
UK-878
UK-838
UK-830
UK-870
UK-824
UK-774
UK-708
UK-738
UK-852
Flight Number


class = Economy
class = Business
price
airline
SpiceJet
AirAsia
Vistara
GO_FIRST
Indigo
Air_India
0
1000
2000
3000
4000
5000
6000
7000
8000
10000
20000
30000
40000
50000


Days Left For Departure vs Ticket Price Comparative
Price
30000
28000
26000
24000
22000
20000
18000
0
10
20
30
40
50
Days Left for Departure

EDA Conclusions

- There is a big gap between flight tickets in business and economy. Business tickets are 6 times more expensive than economy tickets.
- The most expensive airlines appear to be Vistara and AirIndia, while AirAsia is the least expensive. There are just two airlines that provide business class tickets, Vistara and AirIndia, and Vistara is significantly more expensive.
- Prices generally increase gradually until 20 days prior to the flight, at which point they dramatically increase. However, there are frequently open seats that haven't been filled a day before takeoff. As a result, tickets can be found for three times less than the previous day.
- It appears that early morning and late night departures are less expensive, whereas evening departures are more expensive. It appears that early departure, late departure, and late departure are less expensive than night time departure.
- Flights to Chenai are slightly more expensive than those to other cities, but flights from Delhi are the least expensive overall.
- Generally, the cost of an airline ticket increases with the number of stops.

Models used:

- Linear Regression Model
- Decision Tree Model
- Gradient Boosting Model
- Random Forest Regression Model
- KNeighbors Regression
- Bagging
- Extra Trees Regression

Best model performance:

Random Forest Regression Model

- R2 Score: 0.9894574805166351
- Mean Squared Error is 5398550.098449273
- Mean Absolute Error is 888.2363460939549
- Root Mean Squared Error is 2323.478017638487

Random Forest Model Prediction