Flight Ticket Prices Prediction

Submitted by

RAVINDER

Submitted to

iNeuron

Project Title	Flight Ticket Prices Prediction
Technology	Machine Learning
Domain	Aviation
Project Difficulty	Advanced
Programming Language Used	Python
Tool Used	Jupyter Notebook

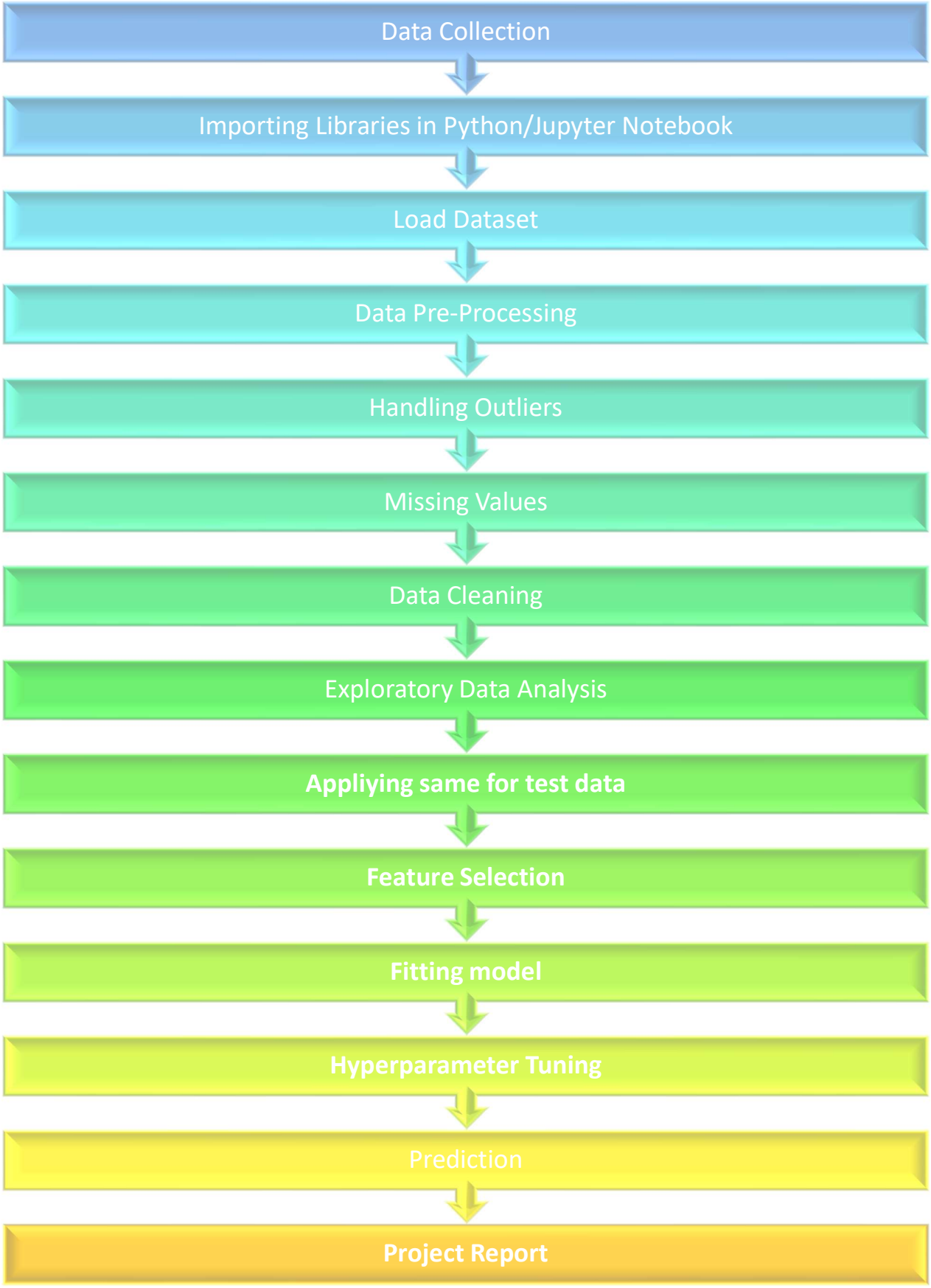
Problem Statement:

For purchasing an airplane ticket, the traditional purchase approach is to buy a ticket far in advance of the flight's departure date to avoid the risk that the price may increase quickly before the date of departure. However, this is not always the case; if airline corporations wish to increase sales, they can lower prices. Airlines employ a variety of factors to decide flight ticket rates, including whether the trip is around the holidays, the quantity of available seats on the plane, and even the month. Some of the variables can be seen, while others are hidden. In this context, customers are attempting to discover the best day to purchase a ticket, while airline firms, on the other hand, are attempting to maximize overall revenue.

The main objective here is –

- Airline companies have the freedom to change the flight ticket prices at any moment. Travelers can save money if they choose to buy a ticket when its price is the lowest.
- The problem is how to determine when is the best time to buy flight ticket for the desired destination and period.
- Airline companies use many different variables to determine the flight ticket prices: indicator whether the travel is during the holidays, the number of free seats in the plane etc. Some of the variables are observed, but some of them are hidden
- In other word, when given the historical price and the current price of a flight for a specific departure date, algorithms need to determine whether it is suitable to buy or wait.

ARCHITECTURE



Imports

1. numpy
2. pandas
3. matplotlib.pyplot
4. seaborn
5. sklearn.ensemble
6. sklearn.model_selection
7. RandomizedSearchCV
8. pickle

Dataset

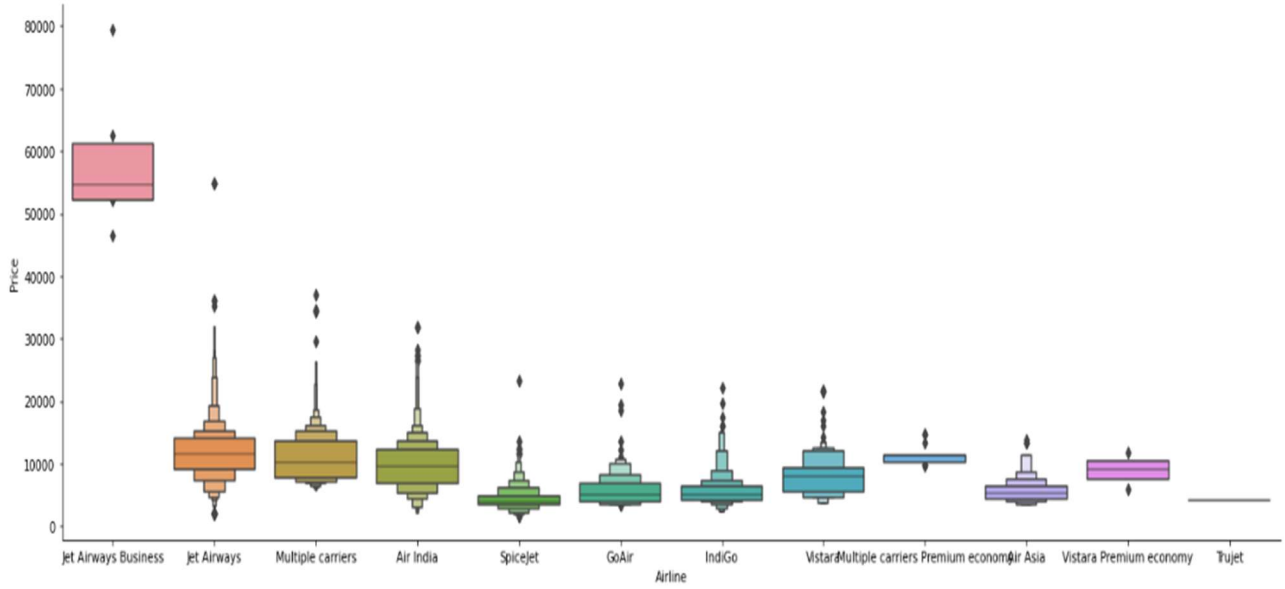
1. Airline: Name of Airline
2. Date_of_Journey: Date of Journey
3. Source: Place from which Journey started
4. Destination: Place which passenger reached
5. Route: Flight Route
6. Dep_Time: Departure Time from source
7. Arrival_Time: Arrival Time at destination
8. Duration: Duration of journey
9. Total_Stops: total no of stops between source and destination
10. Additional_Info: Any additional information
11. Price: Price of journey

Exploratory Data Analysis

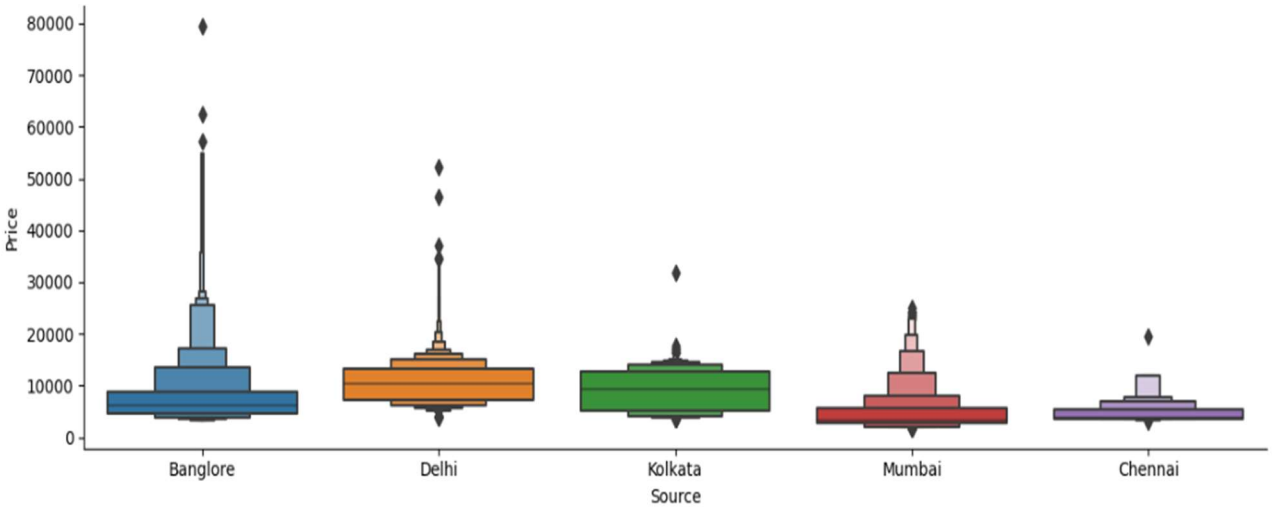
- Date_of_Journey is an object data type; Therefore, we have to convert this datatype into timestamp so as to use this column properly for prediction
- Similar to Date_of_Journey we can extract values from Arrival_Time
- Extracting Hours, Minutes

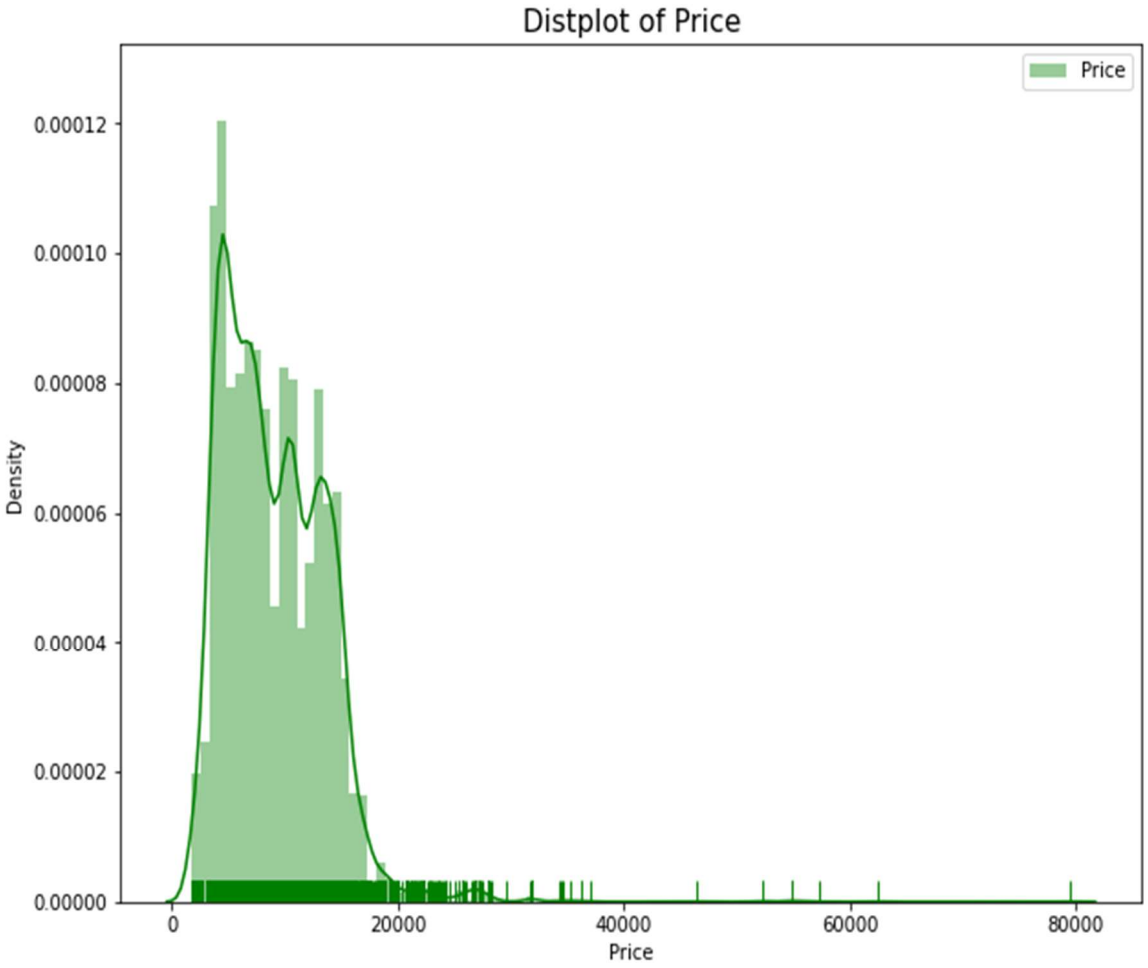
Insights

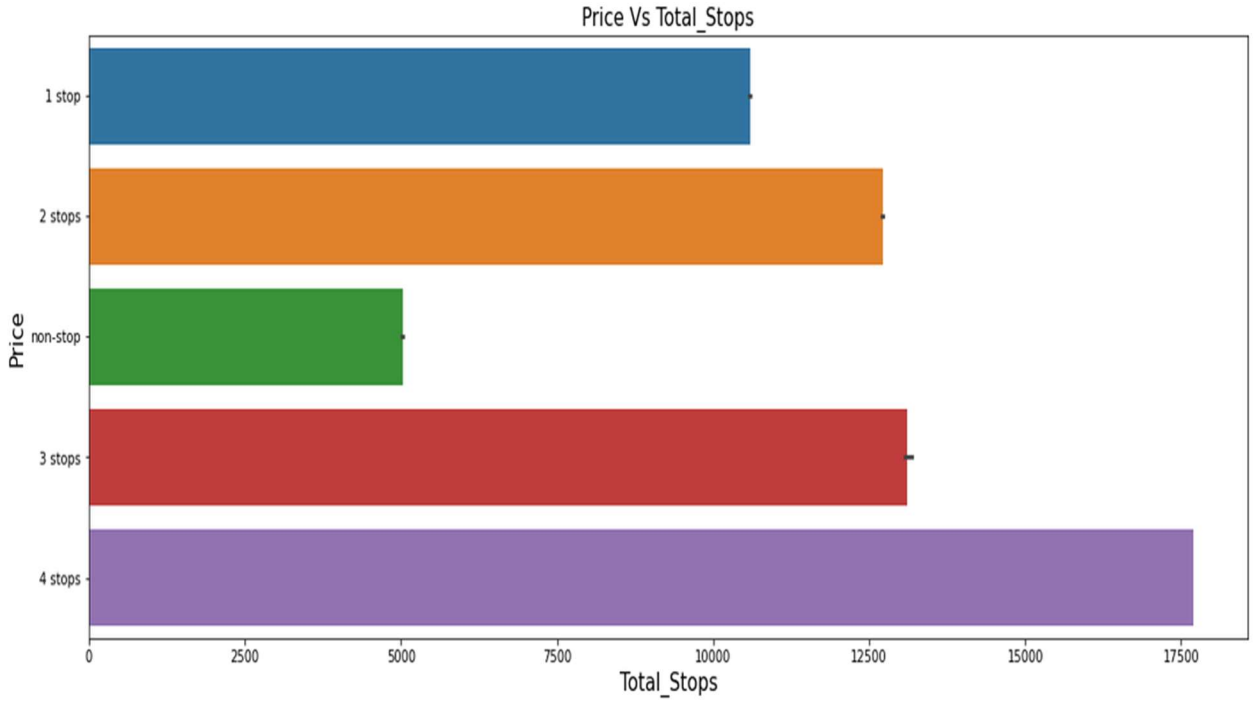
Airline vs Price



Source vs Price

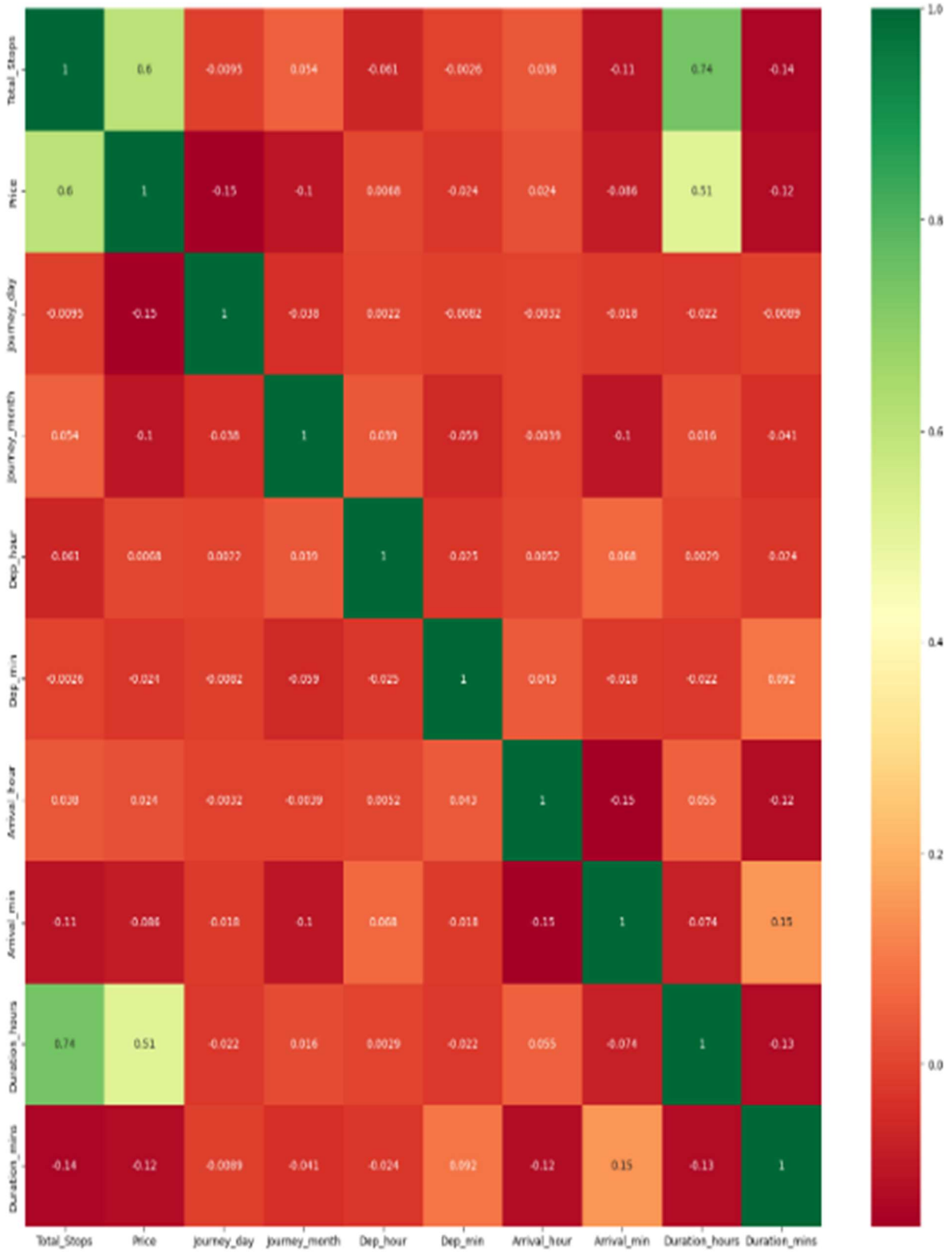



Distplot of Price
Price
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Density
0
20000
40000
60000
80000
Price

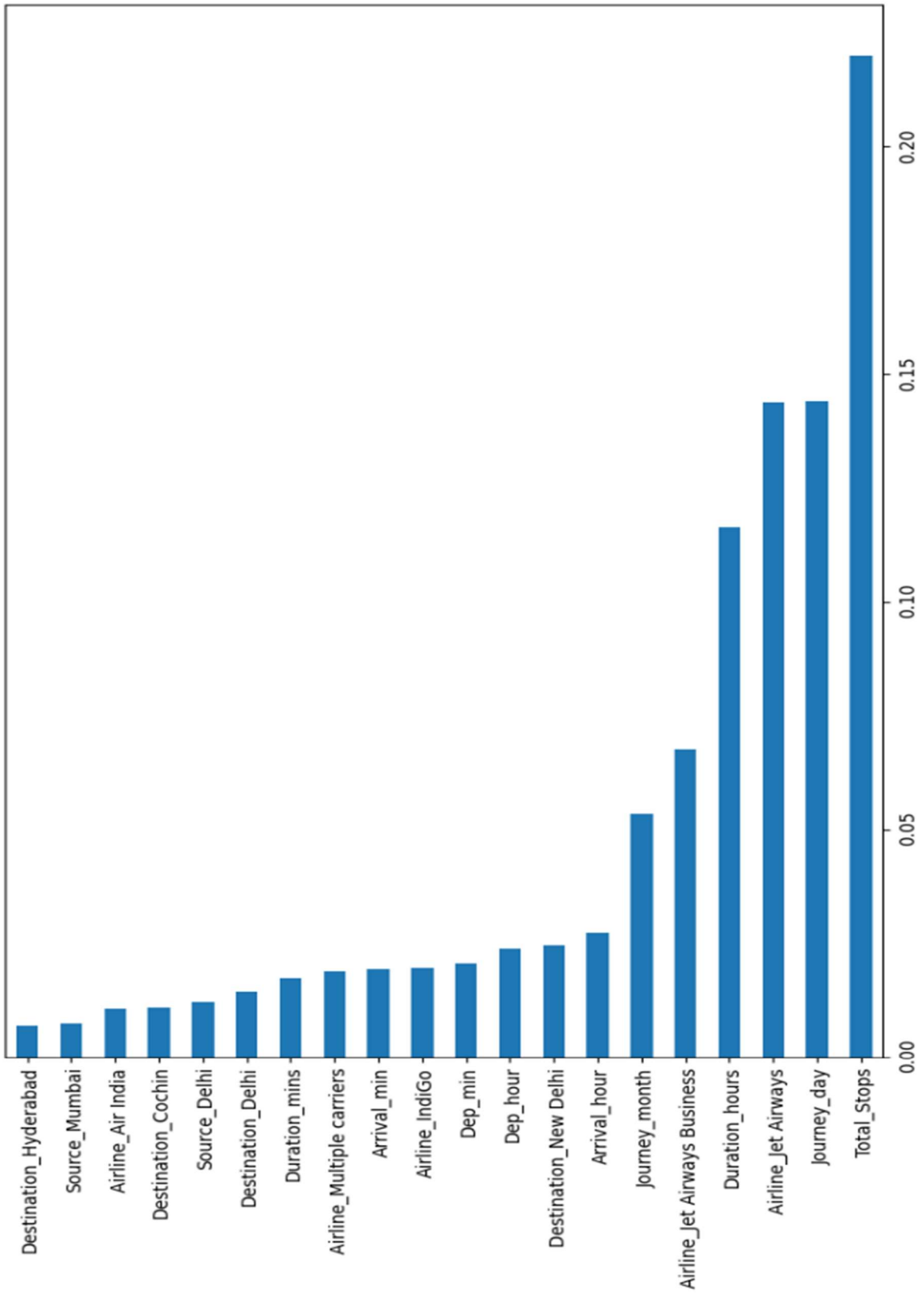

Price Vs Total_Stops
1 stop
2 stops
non-stop
3 stops
4 stops
Price
0
2500
5000
7500
10000
12500
15000
17500
Total_Stops

Correlation between Independent and Dependent Attributes

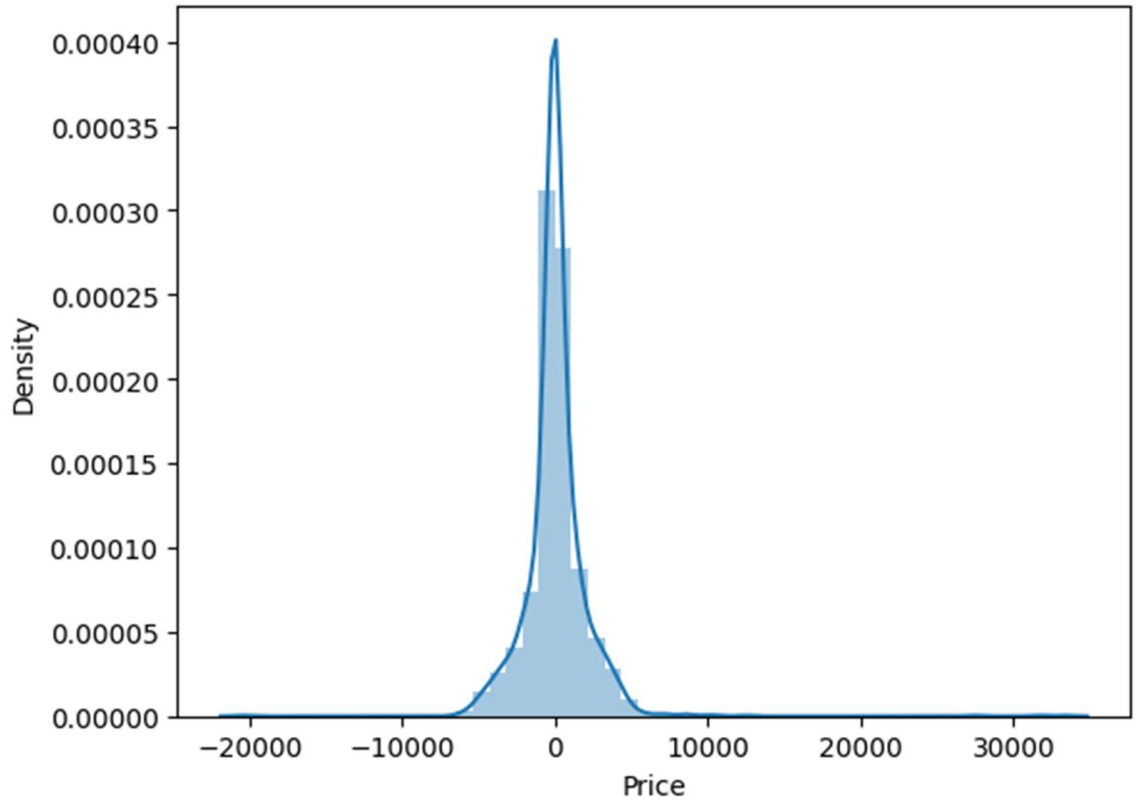
Heatmap

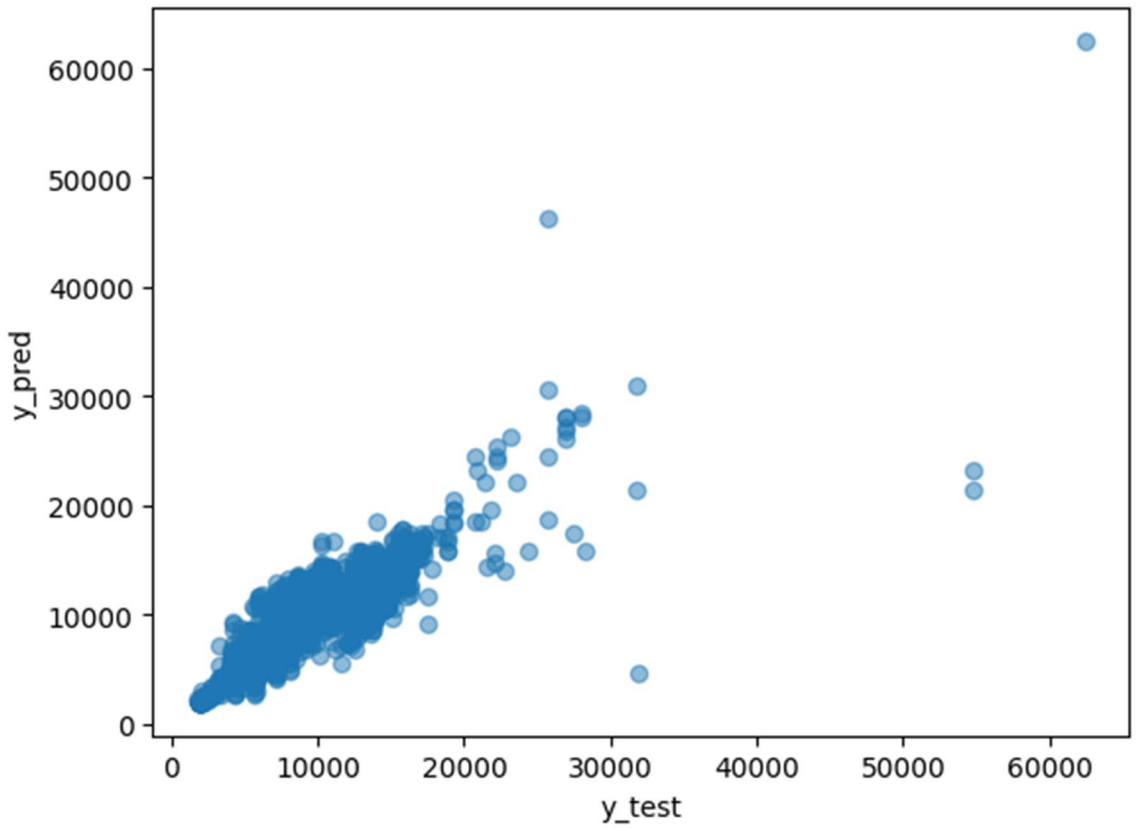


Feature importance for better visualization



Distplot


0.00040
0.00035
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
Density
−20000
−10000
0
10000
20000
30000
Price


60000
50000
40000
30000
20000
10000
0
y_pred
0
10000
20000
30000
40000
50000
60000
y_test