

Flight Booking Price Prediction

Table Content

**Importing the Libraries Loading the Data**

**02**

**01**

**Data Visualization One Hot Encoding**

**04**

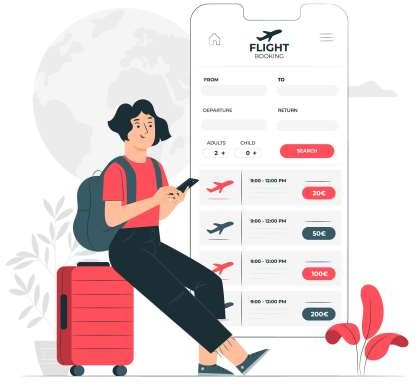
**03**

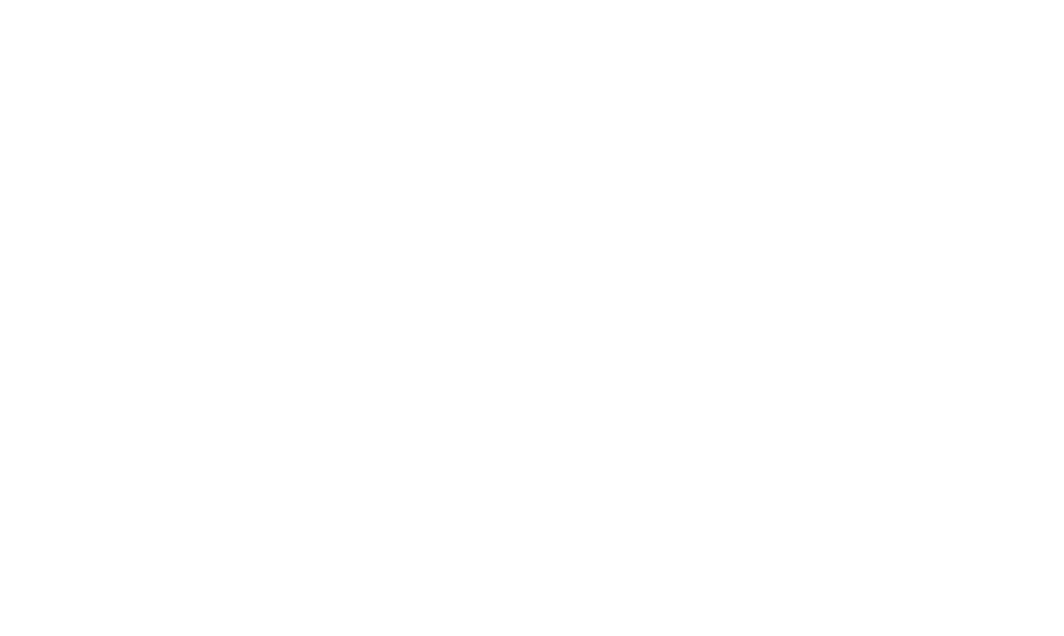
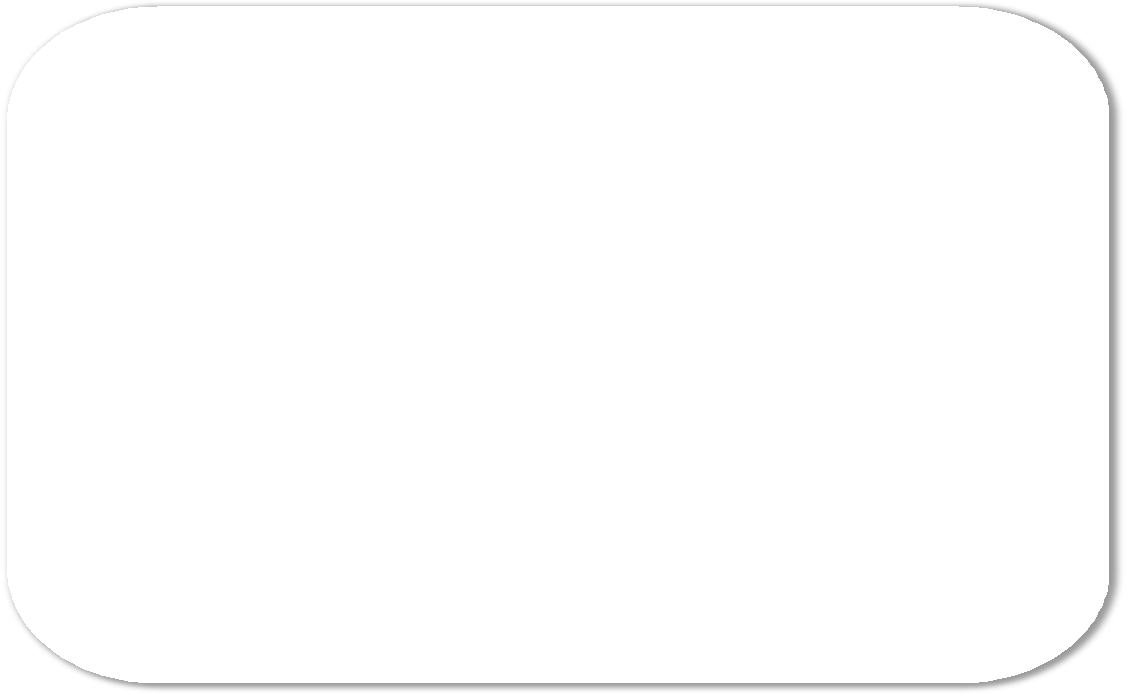
**Feature Selection Implementing ML Algorithms**

**06**

**05**

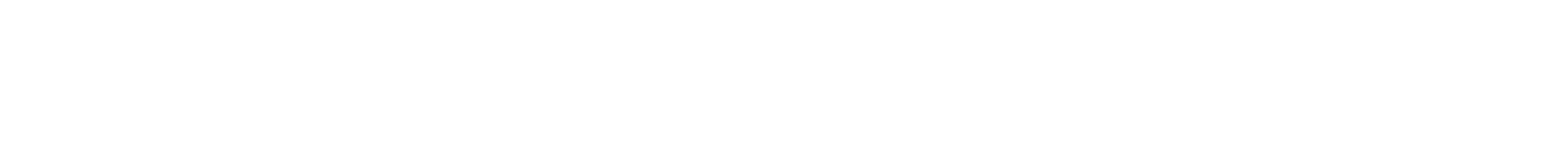
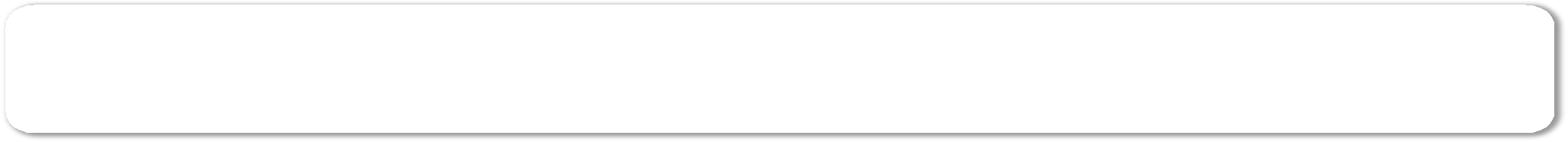
Problem Statement





The objective is to analyze the flight booking dataset obtained from a platform which is used to book flight tickets. A thorough study of the data will aid in the discovery of valuable insights that will be of enormous value to passengers. Applied EDA, statistical methods and Machine learning algorithms in order to get meaningful information from it.

Dataset Information



Flight booking price prediction dataset contains around 3 lacs records

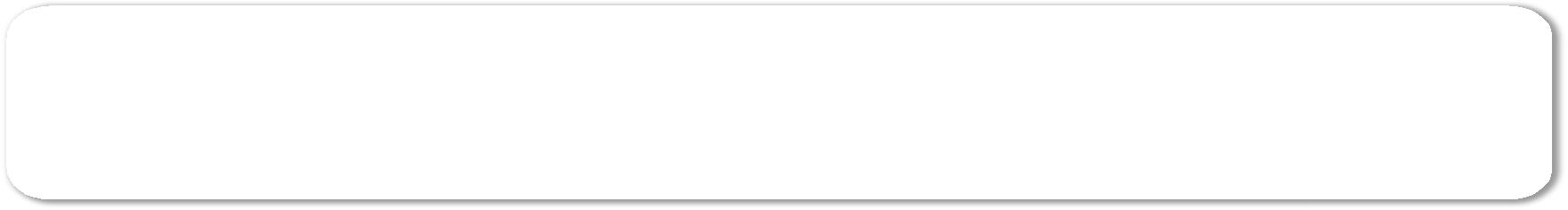
with 11 attributes .



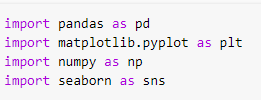
Dataset Information

|  |  |
| --- | --- |
| **Attributes** | **Description** |
| Airline | Name of the airline company |
| Flight | Plane's flight code |
| Source City | City from which the flight takes off |
| Departure Time | Time of Departure |
| Stops | Number of stops between the source and  destination cities |
| Arrival Time | Time of Arrival |
| Destination City | City where the flight will land |
| Class | Contains information on seat class |
| Duration | Overall amount of time taken to travel  between cities in hours. |
| Days left | Subtracting the trip date by the booking  date. |
| Price | Ticket price |

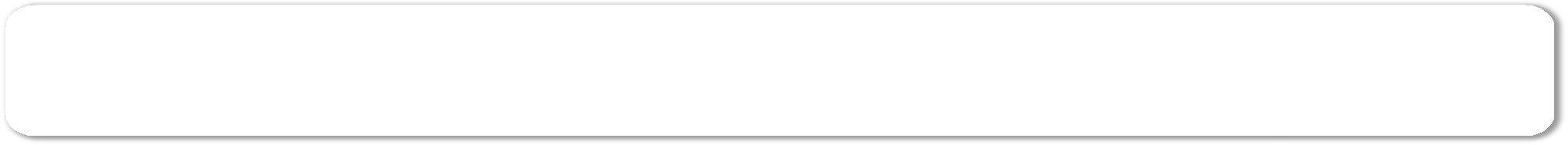
Importing the Libraries



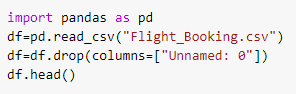
We start off this project by importing all the necessary libraries that will be required for the process.

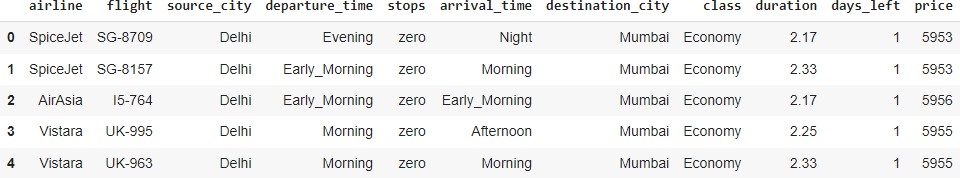


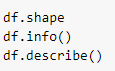
Loading the Data

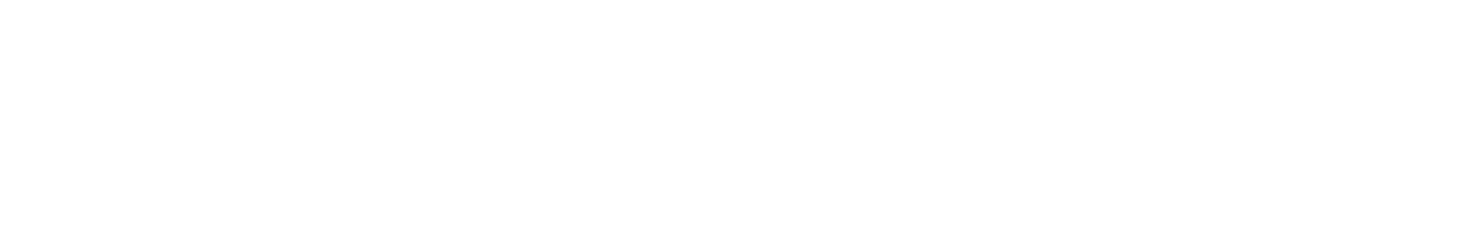
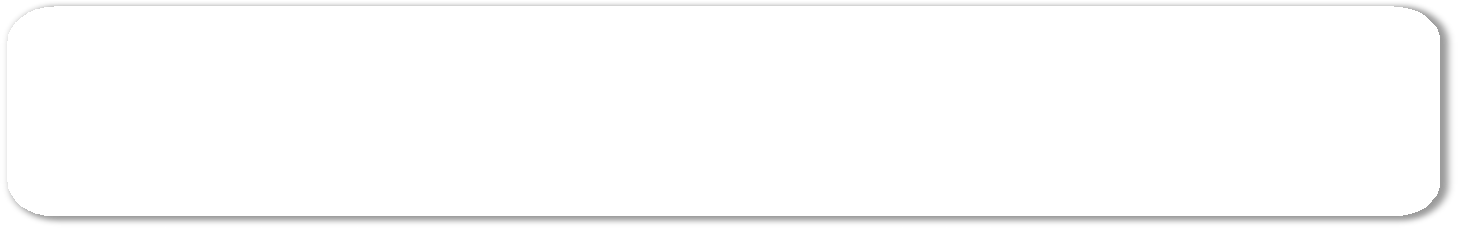


Loading the data and removing unnecessary column from the dataframe

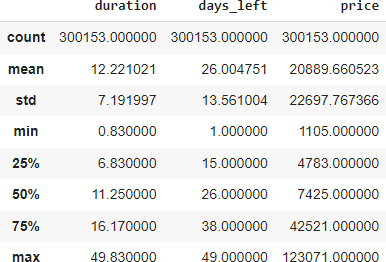
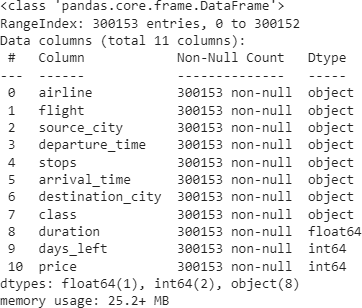




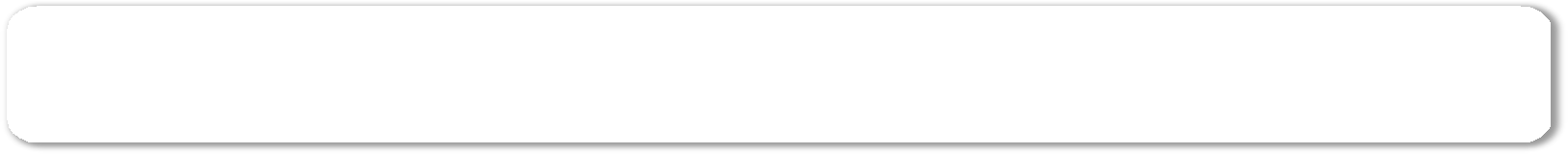
C:\Users\HP\Downloads\5.pngLoading the Data



Checking the shape of a dataframe and datatypes of all columns along with calculating the statistical data.

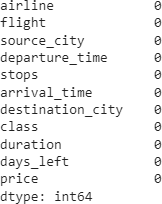


Missing Values

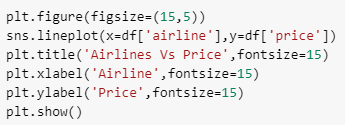


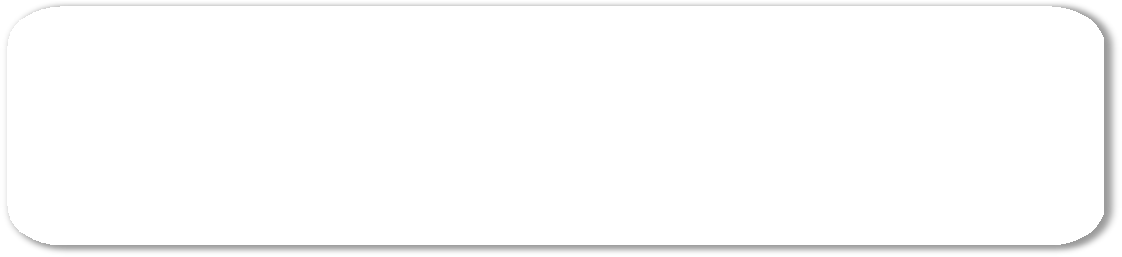
Checking out the missing values in a dataframe

C:\Users\HP\Downloads\8.png

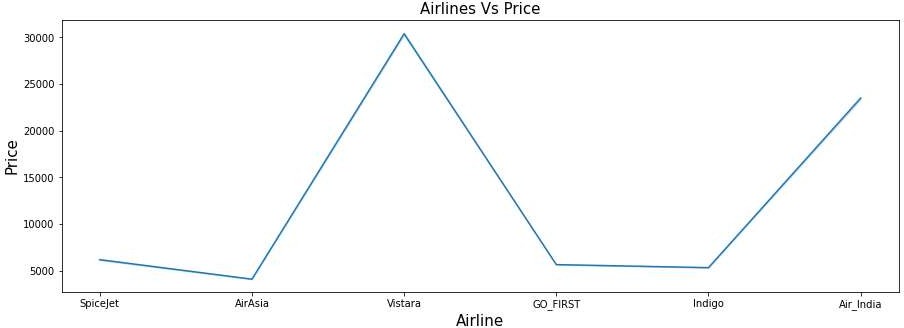


Data Visualization

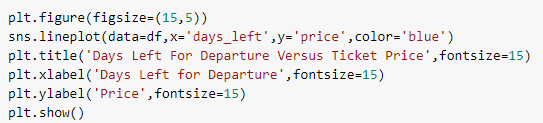


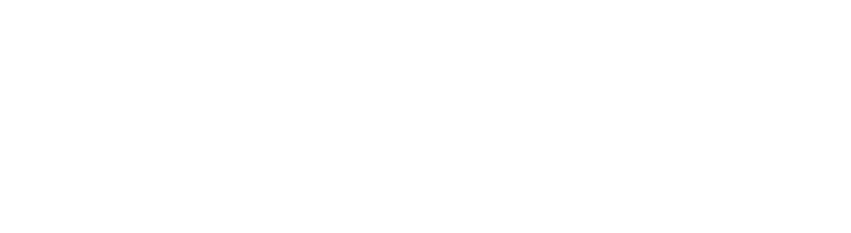
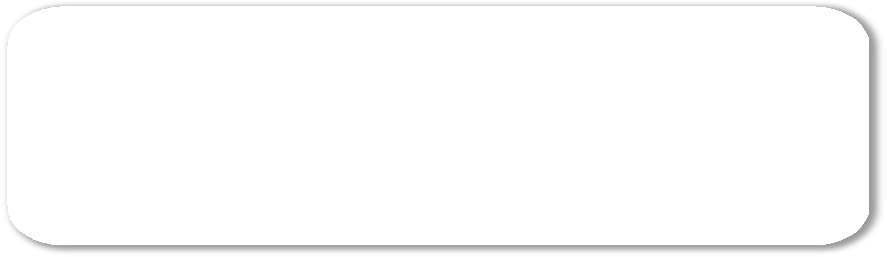


There is a variation in price with different airlines

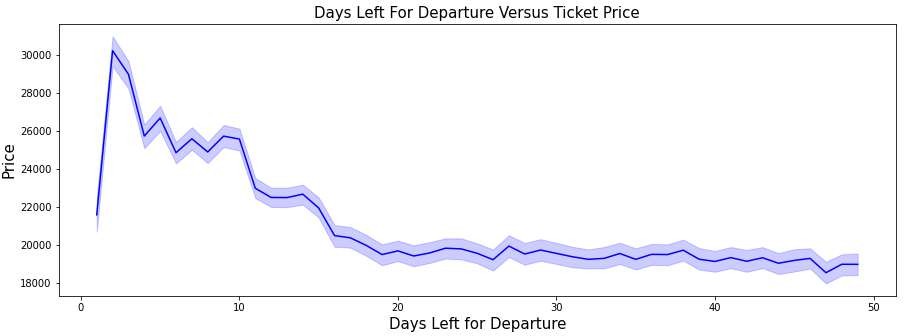


Data Visualization



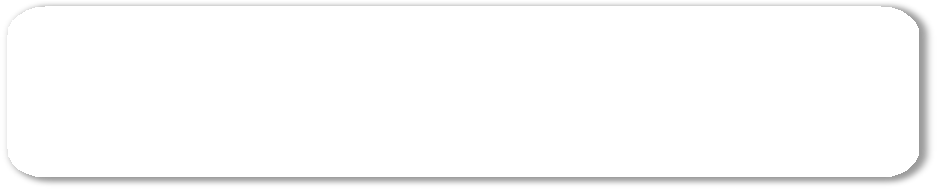


The price of the ticket increases as the days left for departure decreases

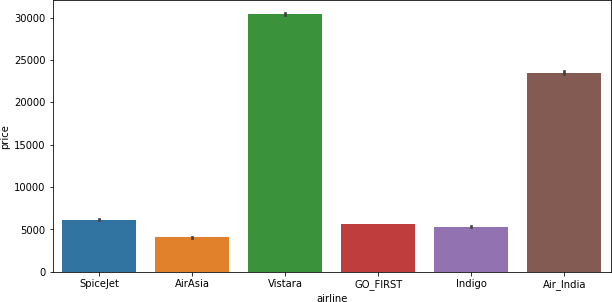


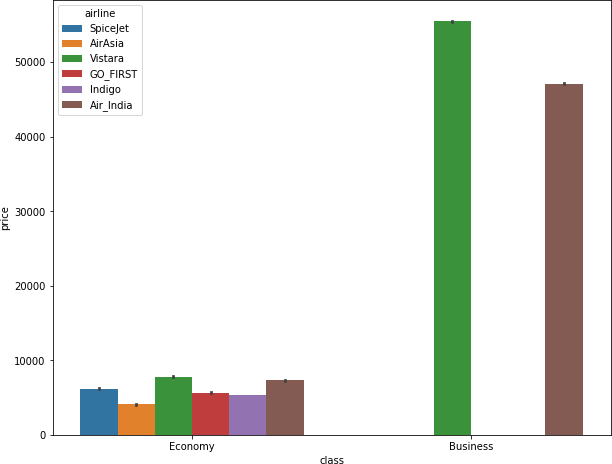
Data Visualization

C:\Users\HP\Downloads\11.png

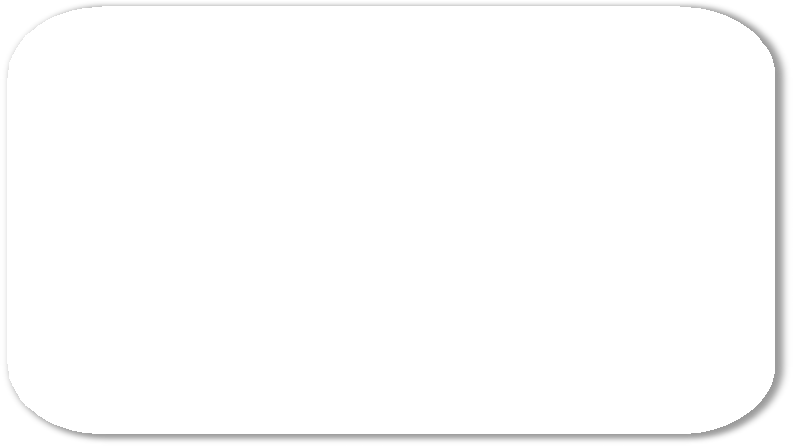


Price range of all the flights



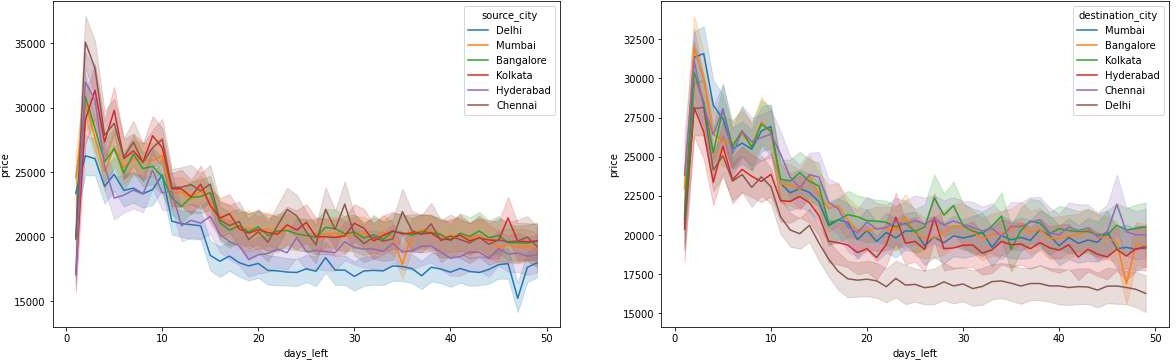
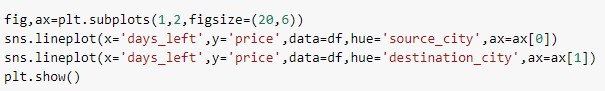
Data Visualization

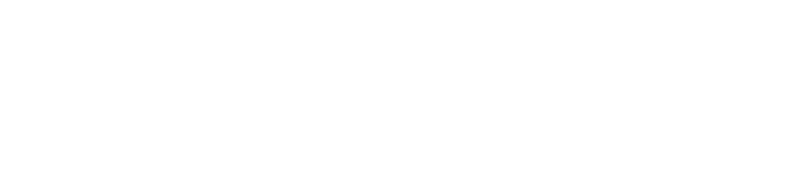
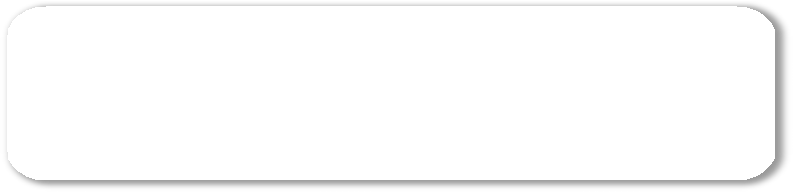
C:\Users\HP\Downloads\12.png



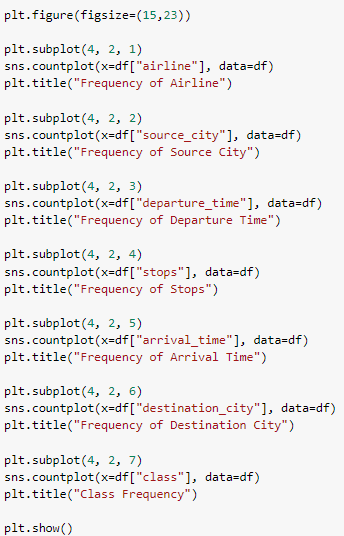
Range of price of all the flights of Economy and Business class

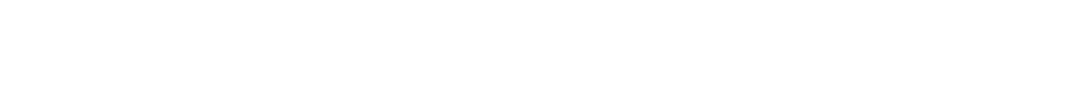
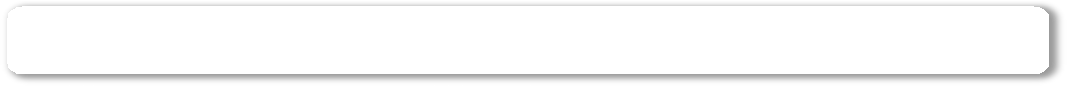
Data Visualization



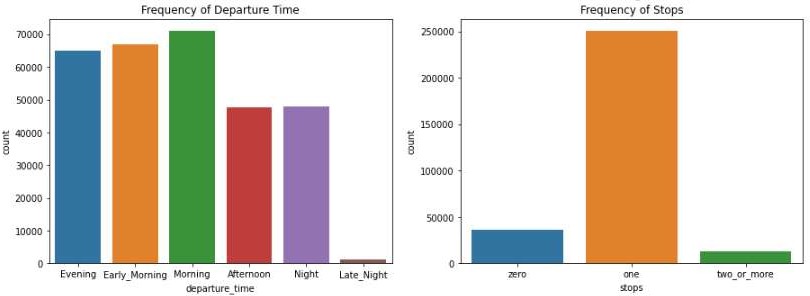
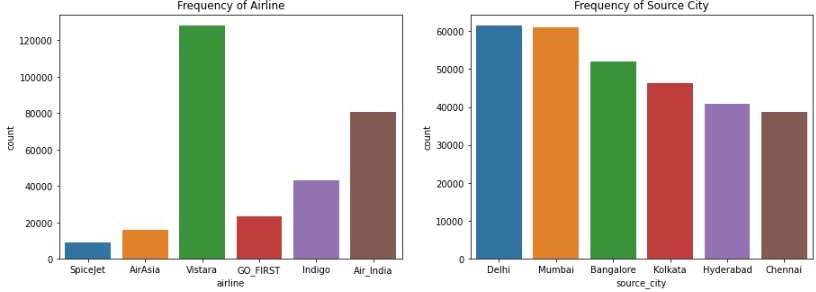


Range of price of flights with source and destination city according to the days left

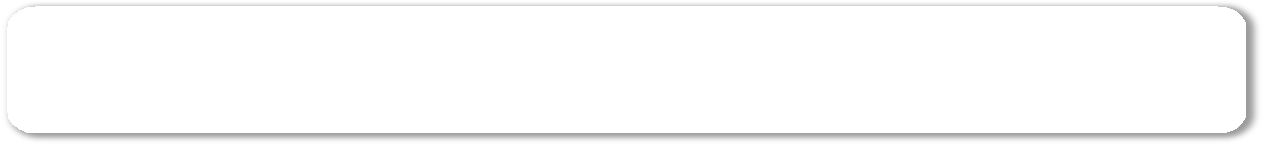
Data Visualization



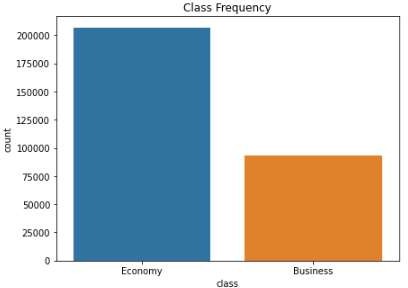
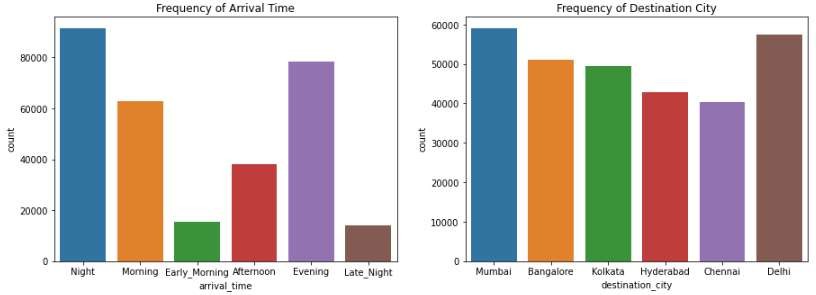
Visualization of categorical features with countplot



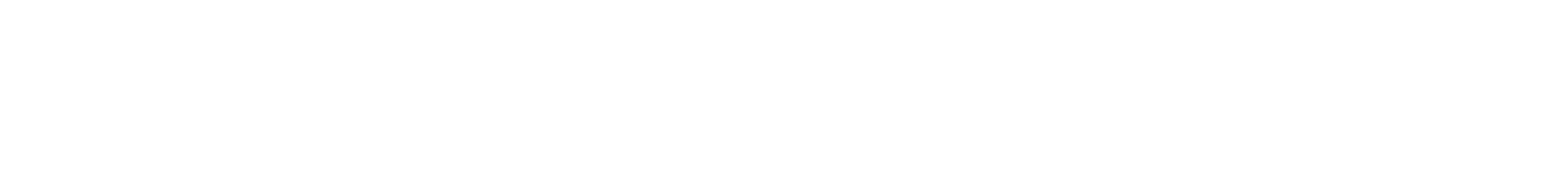
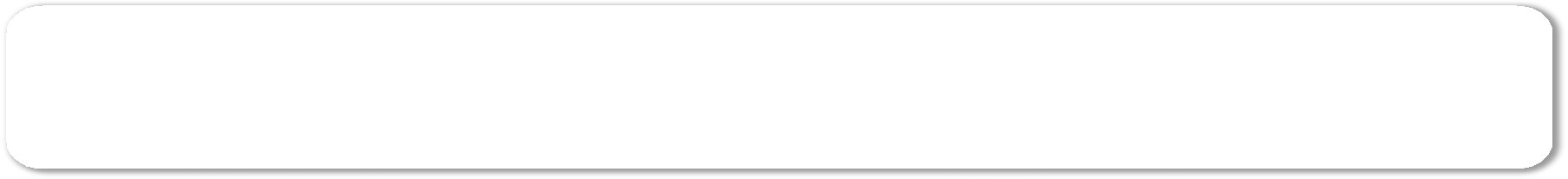
Data Visualization



Visualization of categorical features with countplot

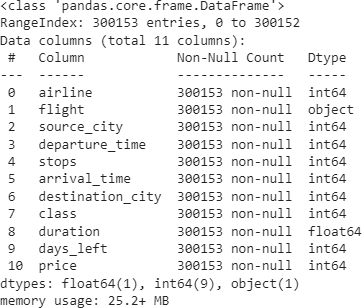
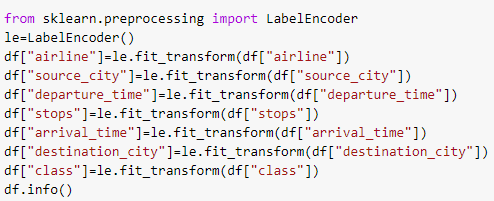


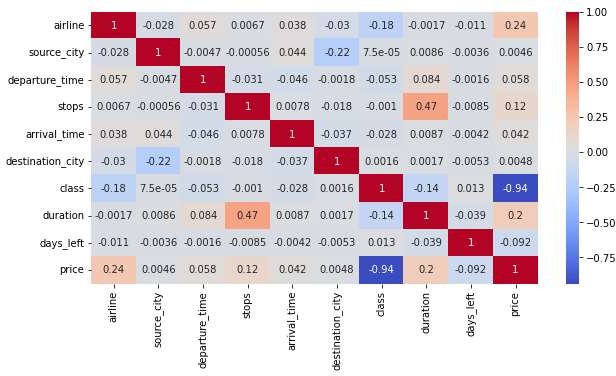
Label Encoding

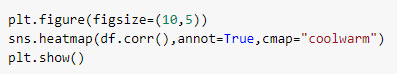


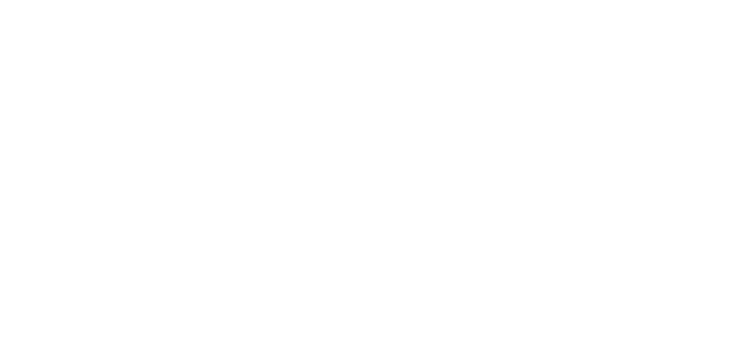
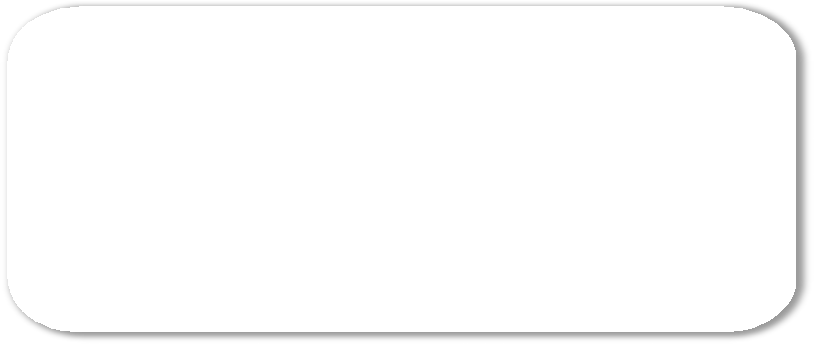
Performing One Hot Encoding for categorical features of a

dataframe



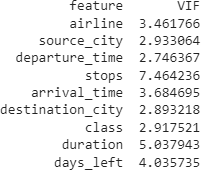
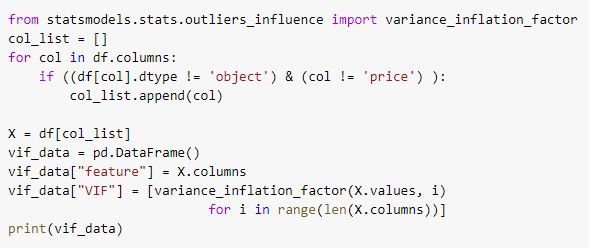
Feature Selection

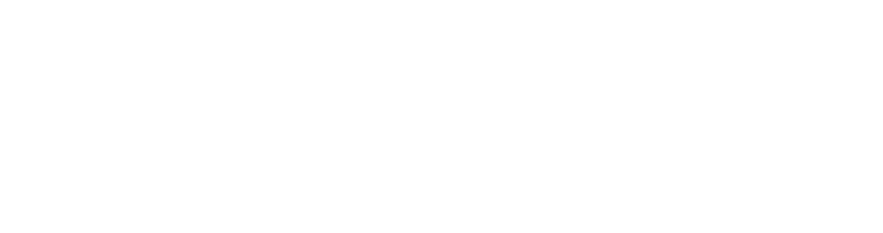
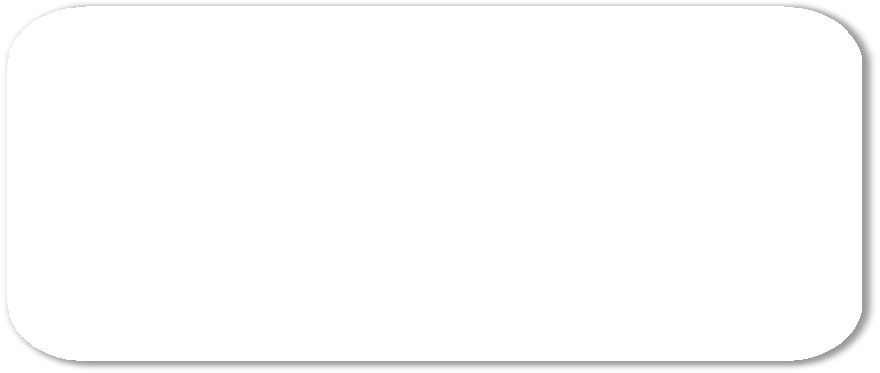




Plotting the correlation graph to see the correlation between features and dependent variable.

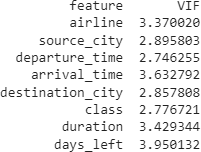
Feature Selection

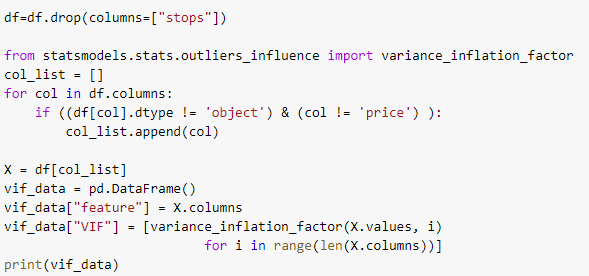


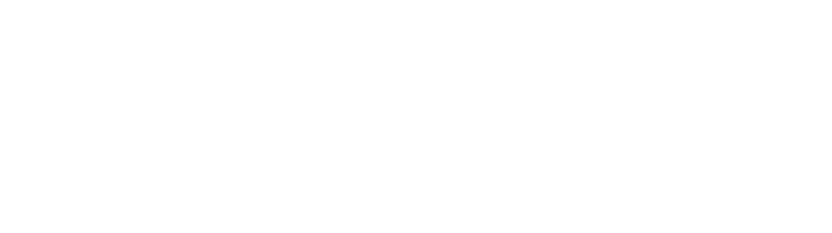
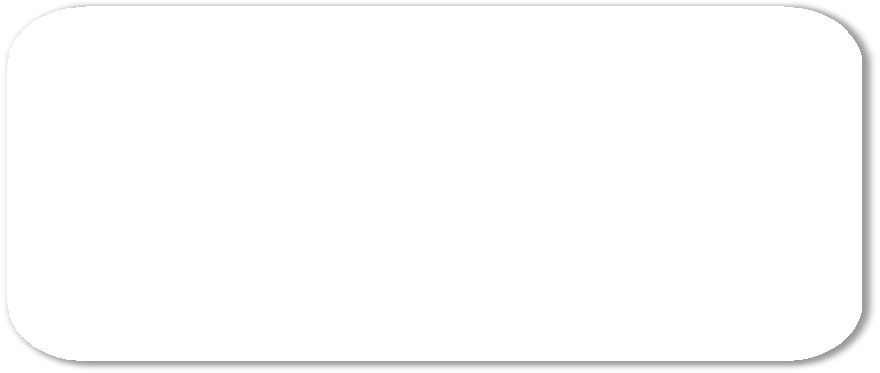


Selecting the features using VIF. VIF should be less than 5.

So drop the stops feature.

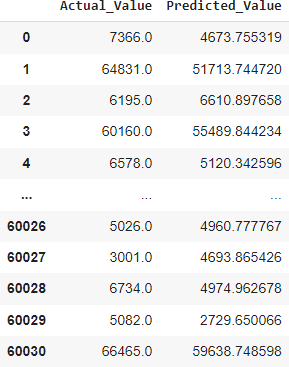
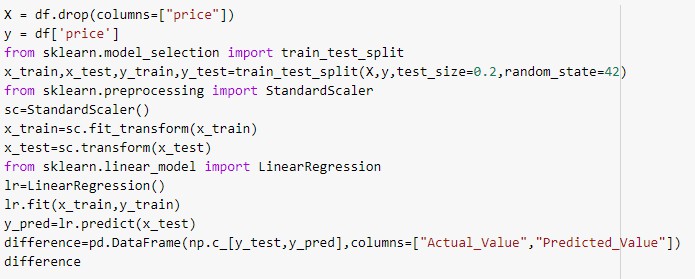
Feature Selection

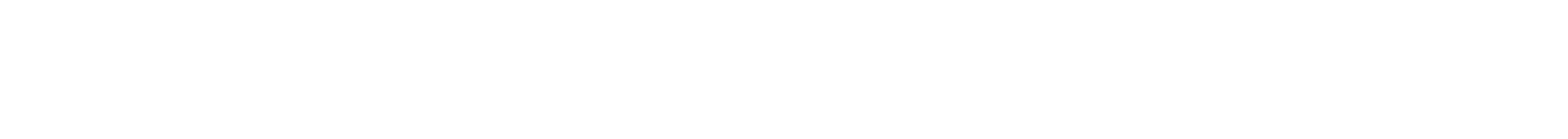
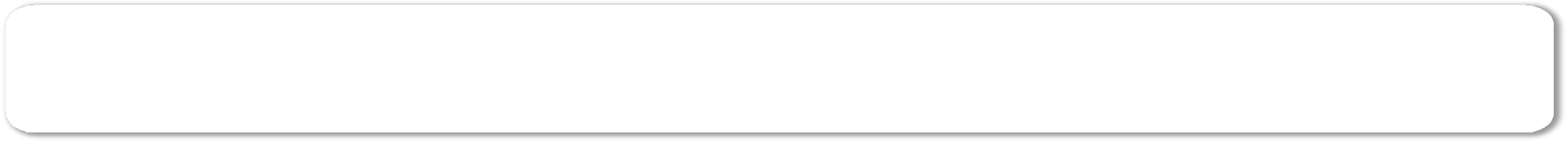




Dropping the stops column. All features are having VIF less than 5.

Linear Regression

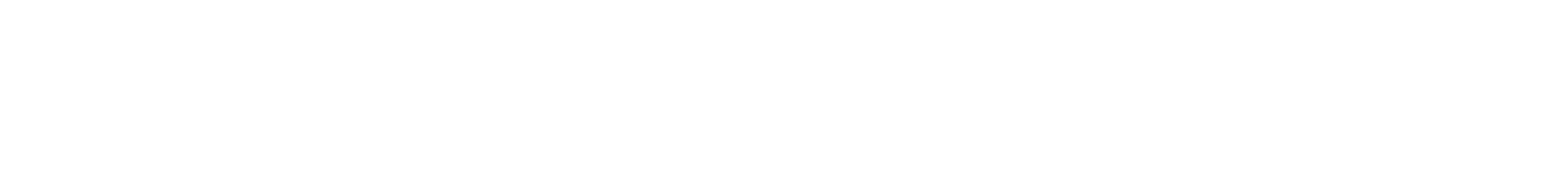
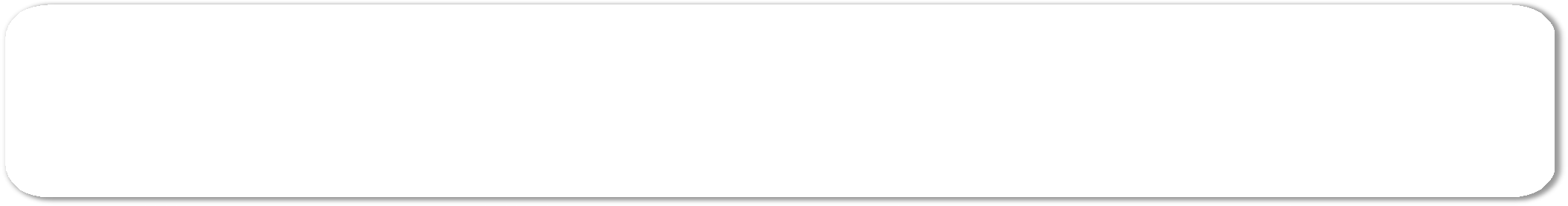




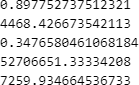
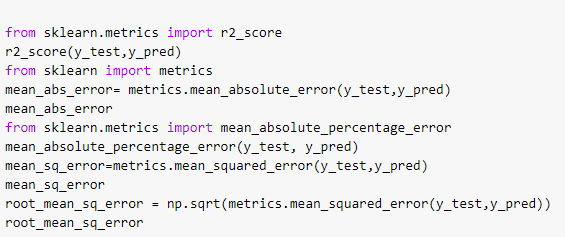
Applying standardization and implementing Linear Regression Model to

predict the price of a flight.

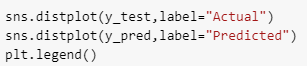
Linear Regression

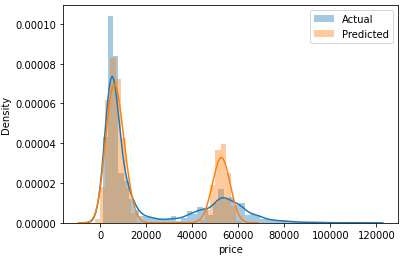


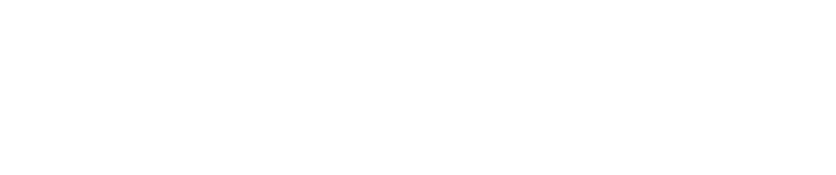
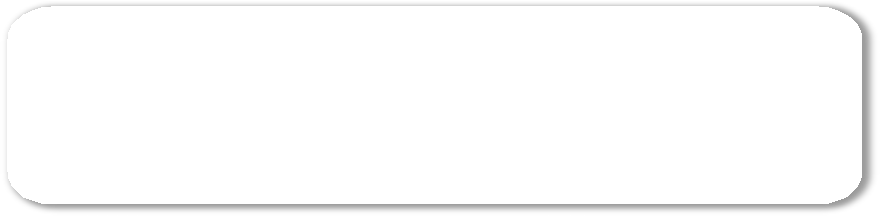
Calculating r2 score,MAE, MAPE, MSE, RMSE. Root Mean square error(RMSE) of the Linear regression model is 7259.93 and Mean absolute percentage error(MAPE) is 34 percent. Lower the RMSE and MAPE better the model.



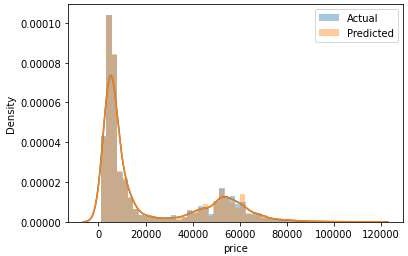
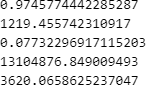
Linear Regression

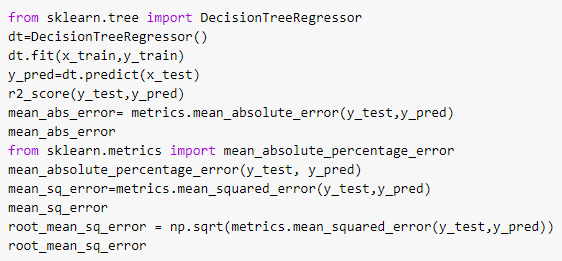


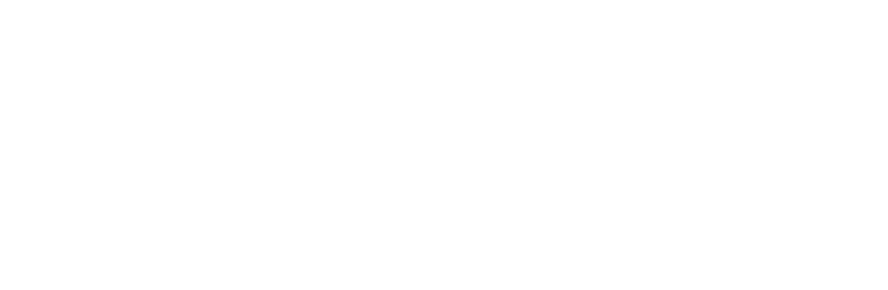
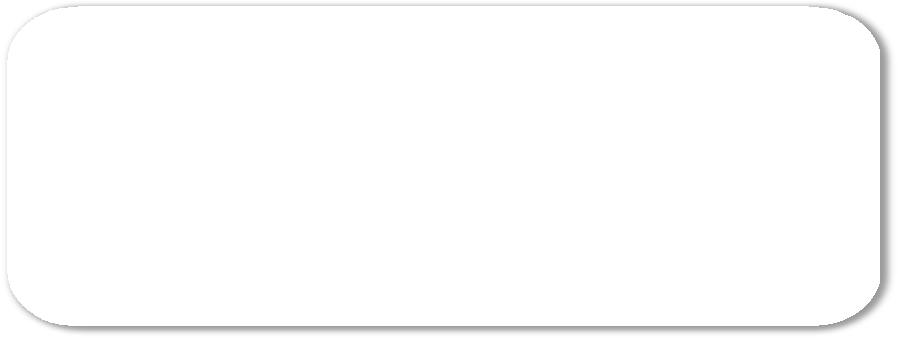




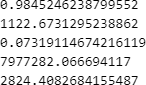
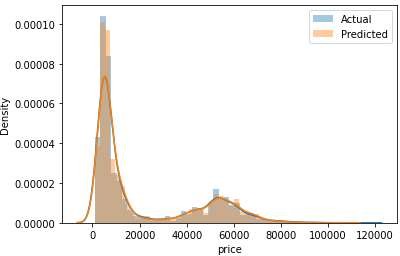
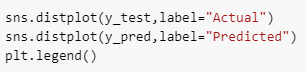
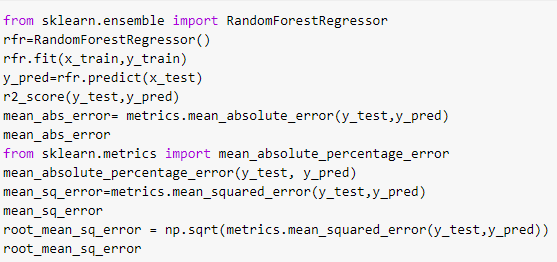
Plotting the graph of actual and predicted price of flight

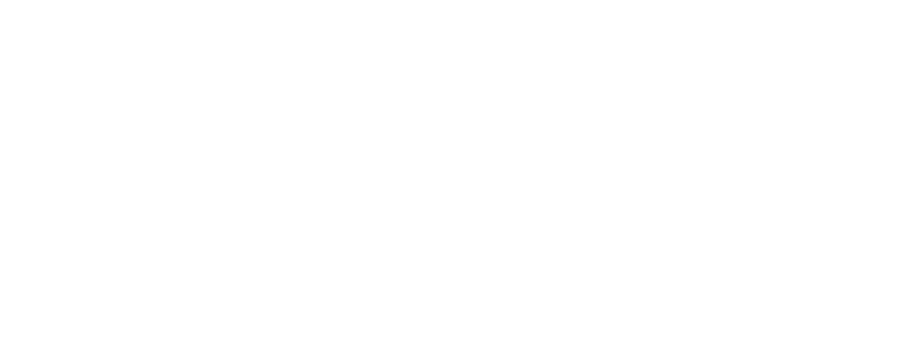
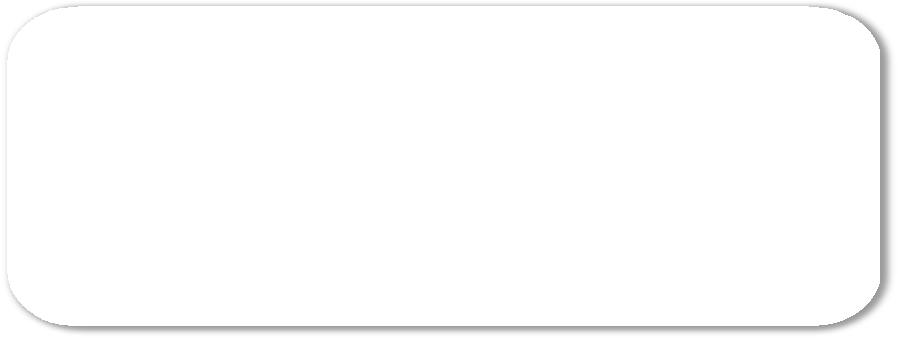
Decision Tree Regressor





Mean absolute percentage error is 7.7 percent and RMSE is 3620 which is less than the linear regression model

Random Forest Regressor



Mean absolute percentage error is 7.3 percent and RMSE is 2824 which is less than the linear regression and decision tree model