

Problem Statement

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, and it will be a different story.

To solve this problem, we have been provided with prices of flight tickets for various airlines between the months of March and June of 2019 and between various cities, using which we aim to build a model which predicts the prices of the flights using various input features.

We have 2 datasets here — training set and test set.

The training set contains the features, along with the prices of the flights. It contains 10683 records, 10 input features and 1 output column — ‘Price’.

The test set contains 2671 records and 10 input features. The output ‘Price’ column needs to be predicted in this set. We will use Regression techniques here, since the predicted output will be a continuous value.

Following is the description of features available in the dataset —

1. **Airline:** The name of the airline.
2. **Date_of_Journey:** The date of the journey

3. **Source:** The source from which the service begins.
4. **Destination:** The destination where the service ends.
5. **Route:** The route taken by the flight to reach the destination.
6. **Dep_Time:** The time when the journey starts from the source.
7. **Arrival_Time:** Time of arrival at the destination.
8. **Duration:** Total duration of the flight.
9. **Total_Stops:** Total stops between the source and destination.
10. **Additional_Info:** Additional information about the flight
11. **Price:** The price of the ticket