Flight Price Prediction

# Objective:

The objective of this project is to predict the price of a flight when a set of relevant flight details are provided. This kind of model can be used to predict the fair price of a flight. The flight details has information related to flight timings, source and destination details, route details, stops in between etc.

The dataset consists of about four months of information about various flights and their respective prices. You are to analyze the data and provide the below

1. Identify the Independent variables which are significant

2. Establish the relationship between time of journey and flight prices

3. Develop and test the hypothesis: (a) Flight Prices on Weekdays are cheaper than flight prices on weekends.

(b) Flight Prices during peak hours (9 AM till 9 PM ) are costlier than flights at other times.

4. Build predictive models to predict flight prices. Split the data into training and test sets. Build supervised models on training data and test it on the test.

Use RMSE as the metric for model evaluation.

5. Use the best model to predict flight prices for which we do not have the flight prices. This is the second dataset attached. (It has 2500 observations for which flight prices have to be predicted)

# About the dataset:

|  |
| --- |
| Airline |
| Date\_of\_Journey |
| Source |
| Destination |
| Route |
| Dep\_Time |
| Arrival\_Time |
| Duration |
| Total\_Stops |
| Additional\_Info |
| Price |
|  |

|  |  |
| --- | --- |
| **1) Introduction** |  |
| a) Defining problem statement |  |
| b) Need of the study/project |
| c) Understanding business/social opportunity |
|  | |
| **2)Data Report** |  |
| a) Understanding how data was collected in terms of time, frequency and methodology |  |
| b) Visual inspection of data (rows, columns, descriptive details) |
| c) Understanding of attributes (variable info, renaming if required) |
|  | |
| **3) Initial Exploratory data analysis** |  |
| a) Univariate analysis (distribution and spread for every continuous attribute, distribution of data in categories for categorical ones) |  |
| b) Bivariate analysis (relationship between different variables , correlations) |