

Flight Fares Prediction Project Report

Submitted by

Naresh Danthala

202195637

Problem statement: Flight Prices are dynamic in nature i.e they change from time to time . so in this project we are building two models that will predict the Flight Fares and we are going to compare which Model is going to give accurate results.

Learning Outcomes from the Project:

- 1.Exploratory Data Analysis
2. Understanding, Processing, and visualizing the data
- 3.Making the machine learning models
4. Model comparison and choosing the best model for accurate results.

Input: we have two datasets one for training and other for testing the results in this project we use Lasso and Decision tree regression models and find out which is the best model to predict the flight prices for the given data.

About the dataset:

The categories that our dataset contains are Airline, Date_of_Journey, Source, Destination, Route, Arrival_Time, Duration, Total_Stops, Additional_Info, Price

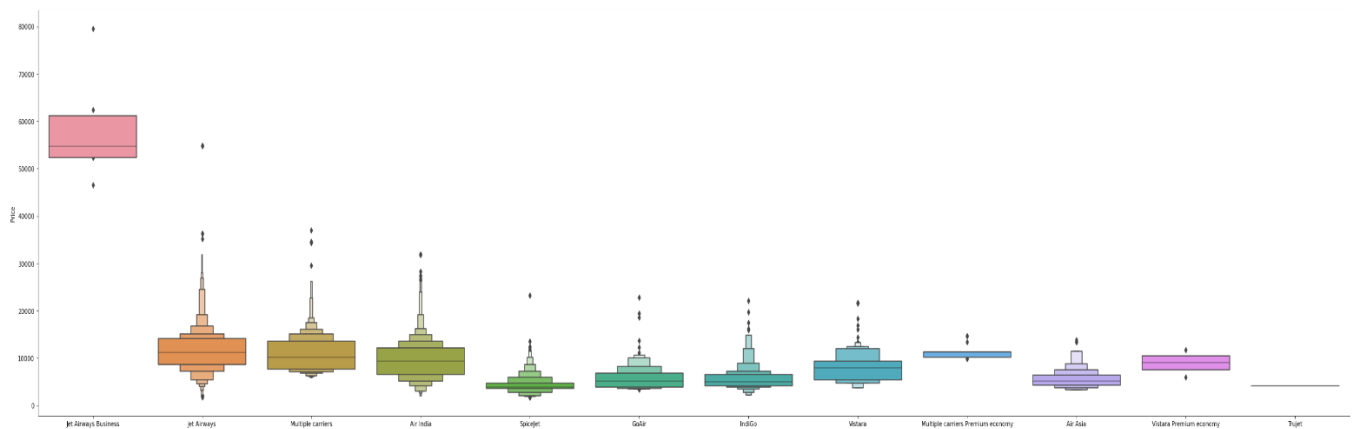
Operations Performed:

we have loaded the given data into data frames with the help of pandas and have performed many operations on them such as Data pre-processing, Data Categorization, Encoding

Once the data is processed then we have used that data for modelling we have used Lasso and Decision tree regression models as part of my project in the end we can conclude which model gives us the more accurate results.

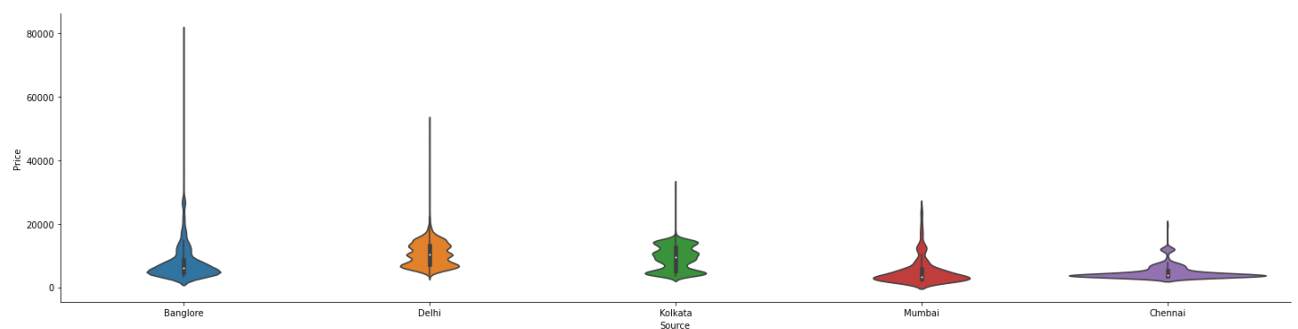
Analysis and Graphs:

we are plotting the Airline versus its price with the help of a cat plot from the below figure we come to know that Jet airways has the most outliers with respect to price

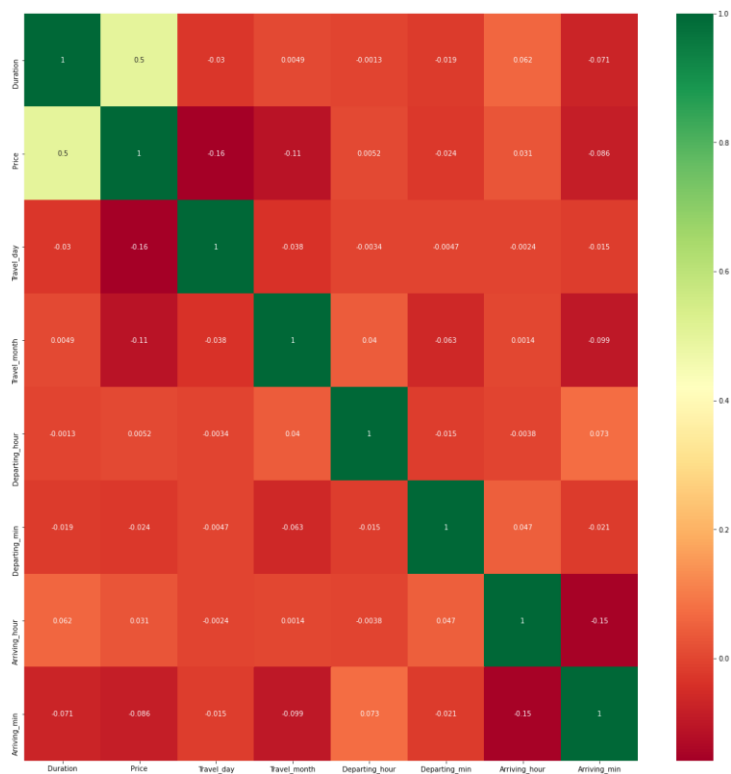
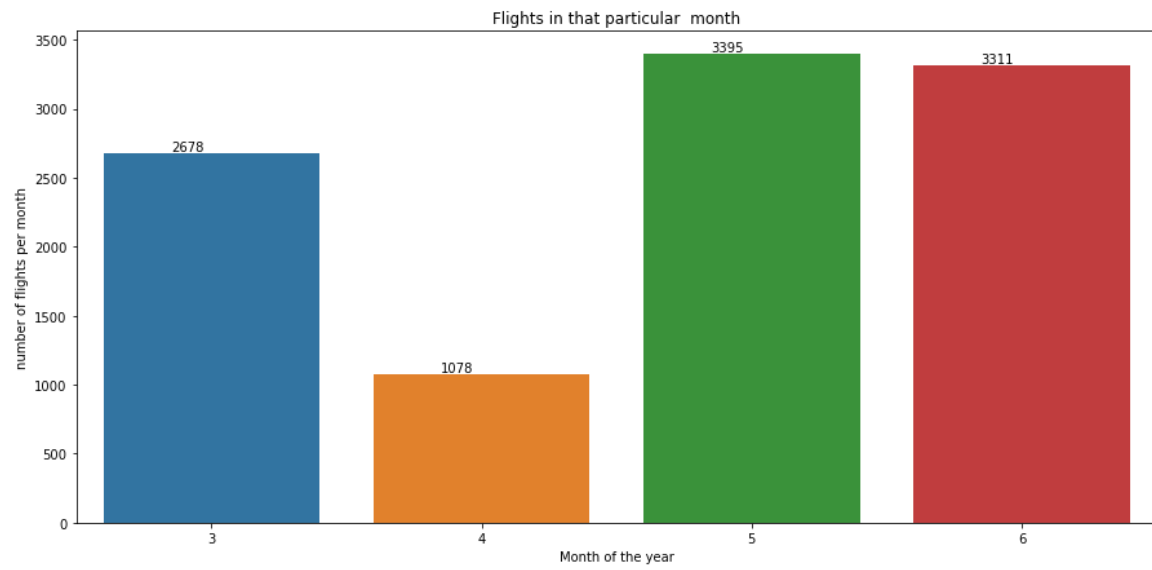


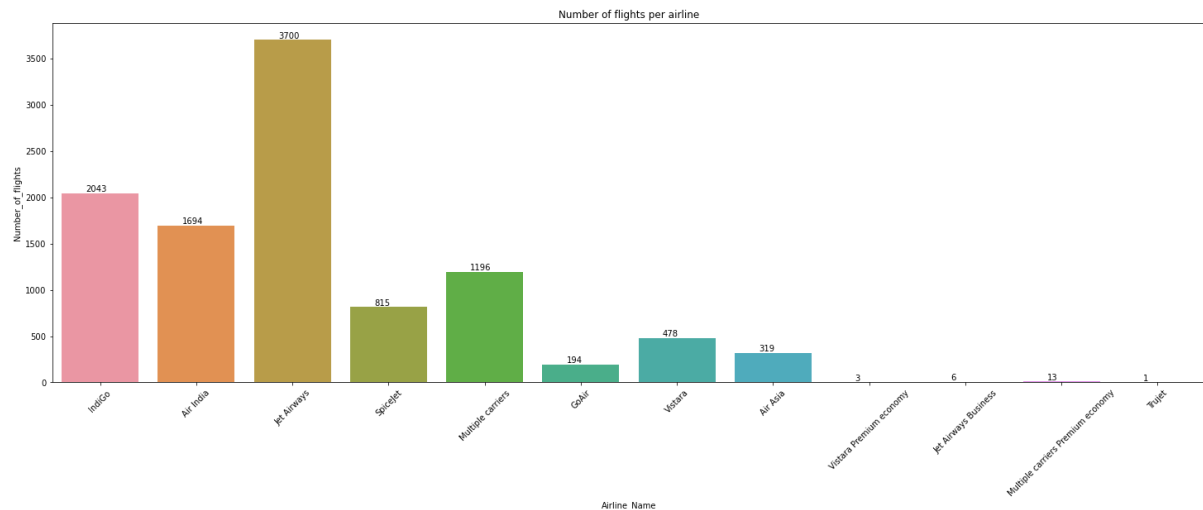
we are plotting the source versus its price with the help of a Violin plot from the below figure we come to know that passengers who are travelling from Bangalore has the most outliers with respect to price

and Chennai has the lease outliers.



Below are few more graphs as a part of visualizing the data.





Once this data is processed then we use this data to train our models and will use these models on the test data and find out which is going to give us accurate results

Conclusion:

We trained both the Lasso regressor and Decision tree regression models and we came to the conclusion that Decision tree regression model has achieved a r^2 -score of 82% whereas the lasso regression model has only got 42% so with the help of Decision tree regression we can predict the flight fares that were actually close to the original prices.

```

*****
      Name_of_the_Model  model_train_Score  Model_Test_Score
0      Lasso Regression      -252062.50      -248119.29
1  Decision Tree Regressor       96.63         81.52
*****

```

Operating system : Windows

Softwares used: Jupyter Notebook, Visual Studio code, Python

Libraries used: Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn,
DecisionTreeRegressor, Lasso, r2_score