

FLIGHT PRICE PREDICTION

Submitted by:

NITIN GUPTA

**ACKNOWLEDGMENT**

I wish to express our gratitude to our mentor .Sapna Verma for providing me an opportunity to do my research project work on "CAR PRICE PREDICTION". Under her guidance I have completed my project report and tried my best to implement what I had learnt till now .She also helped me by updating me about the information of what to do and not to do during my research project and help me with all

**ABSTRACT**

This dataset consists information about used car listed on easemytrip.com. It has 9 columns each columns consists information about specific features like Flight\_Name gives information about Airlines info .Source indication where flight start to fly. Destination indicate where to go. Arrival Time is start time when flight start to fly and Reached time is when flight reached to Destination. Total time is taken the total time taken in journey. Date of Journey is the Journey date when journey start. Flight Price is the total price for particular journey on particular date.

**INTRODUCTION**

* Business Problem Framing

With Anyone who has booked a flight ticket knows how unexpectedly the prices vary. The cheapest available ticket on a given flight gets more and less expensive over time. This usually happens as an attempt to maximize revenue based on -

1. Time of purchase patterns (making sure last-minute purchases are expensive)

2. Keeping the flight as full as they want it (raising prices on a flight which is filling up in order to reduce sales and hold back inventory for those expensive last-minute expensive purchases)

So, you have to work on a project where you collect data of flight fares with other features and work to make a model to predict fares of flights.

**Data Collection Phase:**

I have collected data from easemytrip.com.

I have collected data from easemytrip.com using Data Scraping using Selenium technique.

**Analytical Problem Framing**

* Data Sources and their formats

Data scrap from easemytrip.comusing selenium.

* Data Preprocessing Done

In Pre-processing step I changed feature data as ML friendly to removing commas, Make data columns Flight\_Price in Float type. Checking counts as per countplots for object data type data distribution among variables.

* Data Inputs- Logic- Output Relationships

We have split data into features and target columns

* Hardware and Software Requirements and Tools Used

I have used Jupyter NoteBook to run the code for both Data Scraping as well as model Building.

**Model/s Development and Evaluation**

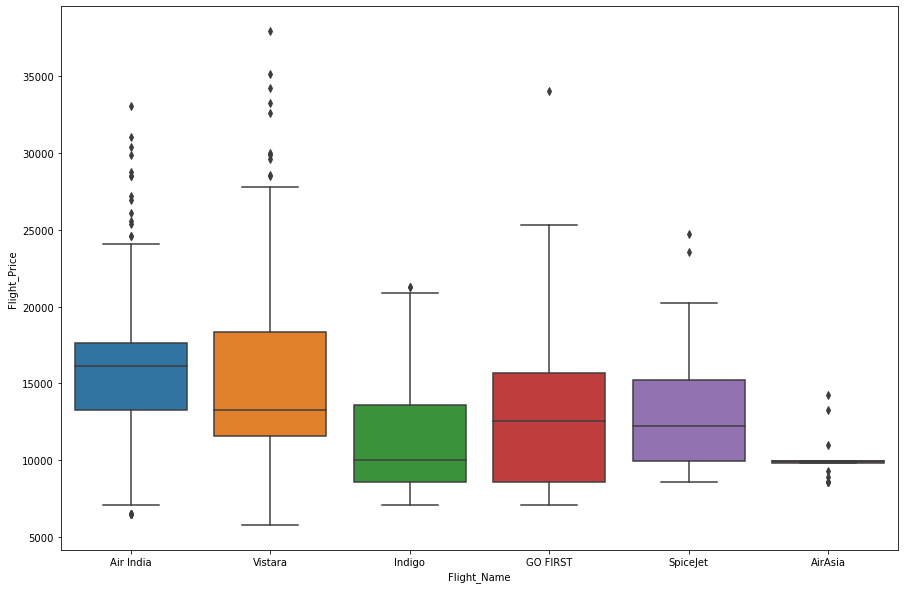
* Identification of possible problem-solving approaches (methods)

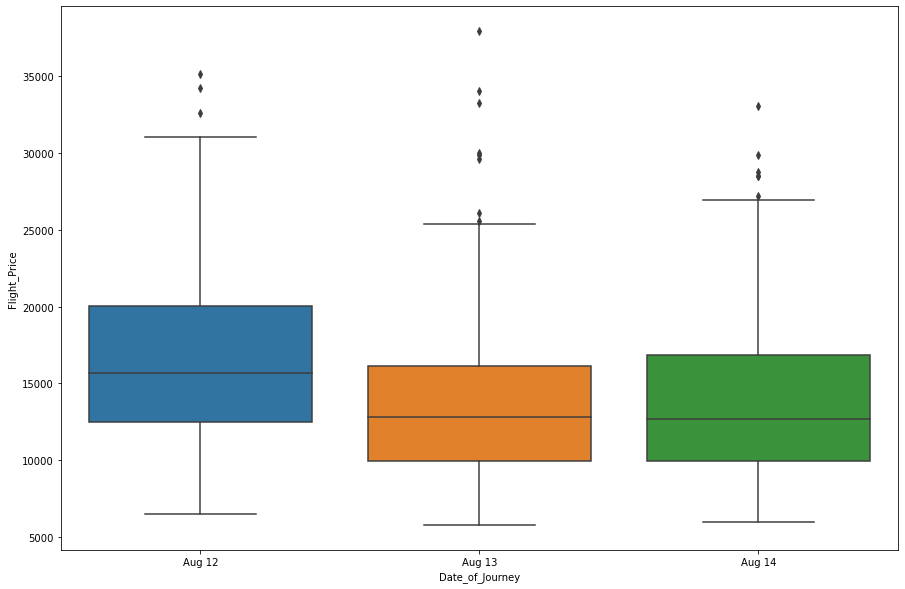
I have imported 5 models K-Neighbors Regressor, Decision Tree Regressor, Random Forest Regressor, Adda Boost Regressor and Ridge Regressor for if overfitting exist. While I checked R2 score on different random states for different models I found Random forest regressor gives the best result on Random\_state 67

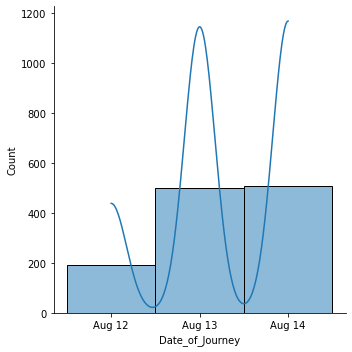
* Run and Evaluate selected models

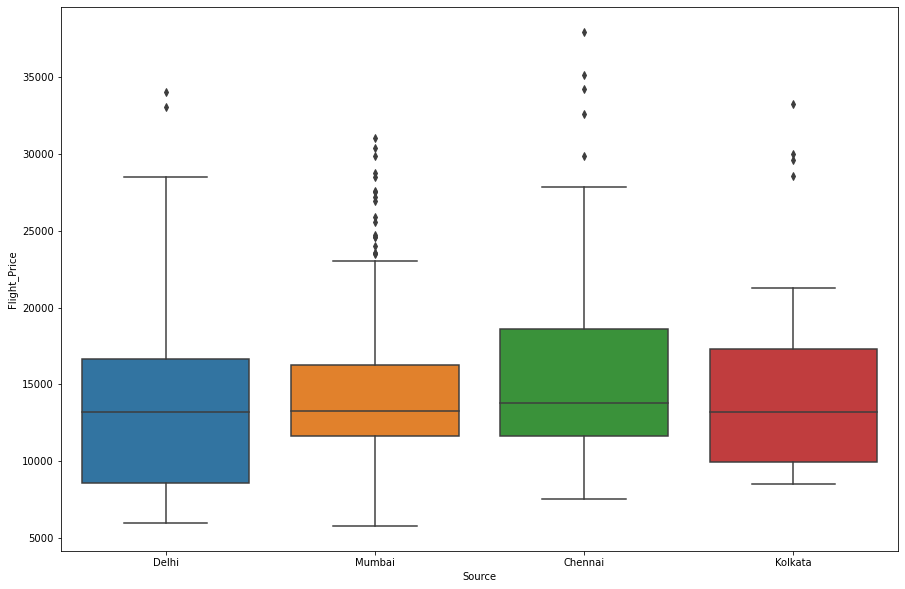
So I have selected Random Forest model is best model as per random state=67.

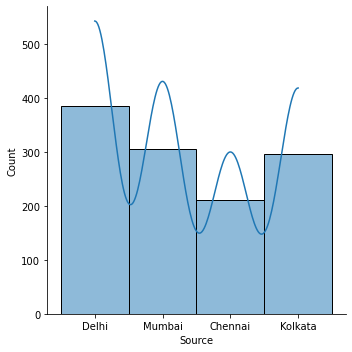
* Visualizations

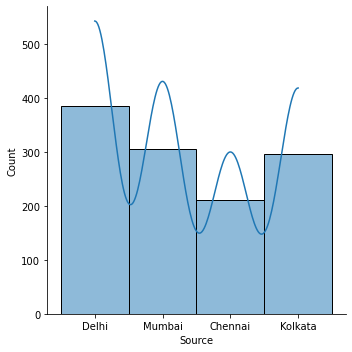


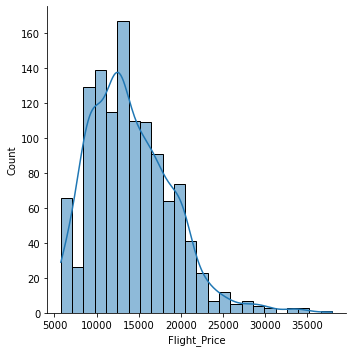


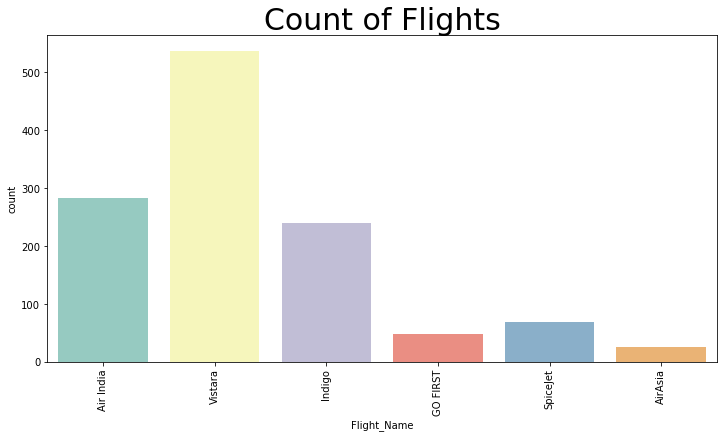


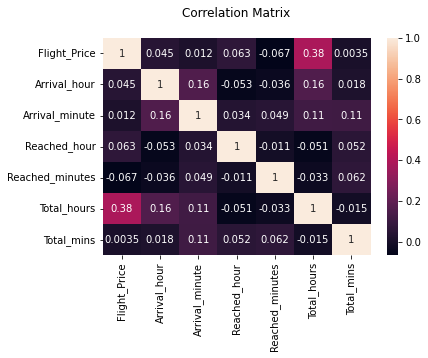


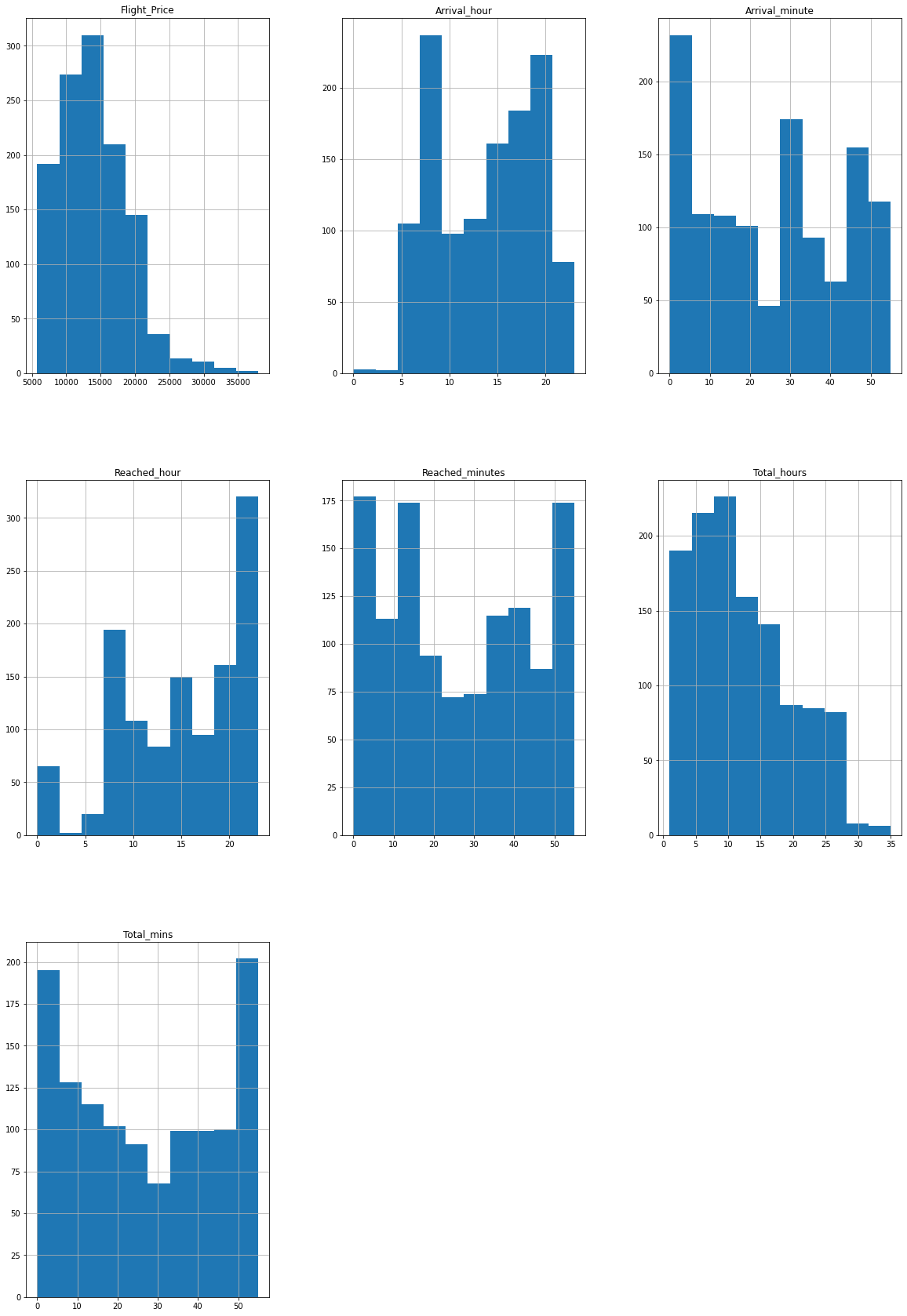






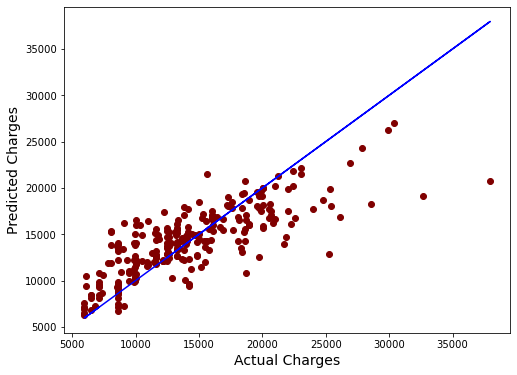






* Interpretation of the Results

I have selected the Random forest is best model now using gris search cv Hyperparameter tunning done.



We have found the best parameters to fit under model and gives the best result.

**CONCLUSION**

* Key Findings and Conclusions of the Study

We have found that the model is running good and not biased and it gives the r2\_score below.

* Learning Outcomes of the Study in respect of Data Science

The conclusion is the model is running best.