

Flight Price Prediction

Submitted by:

Manoj Kumar Saxena

**ACKNOWLEDGMENT**

I would like to express my gratitude towards FlipRobo Technologies for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons and my mentor (Ms. Swati Mahaseth) for giving me such attention and time as and whenever required.

**INTRODUCTION**

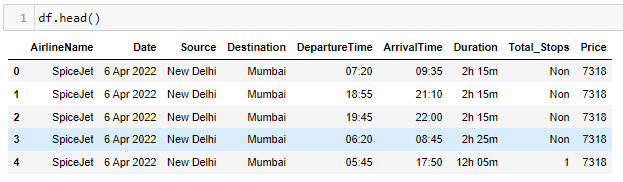
## Business Problem Statement

* 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 -
* Time of purchase patterns (making sure last-minute purchases are expensive)
* 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, we 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.

## Conceptual Background of the Domain Problem

* The airline industry is considered as one of the most sophisticated industries in using complex pricing strategies. Nowadays, ticket prices can vary dynamically and significantly for the same flight, even for nearby
* Customers are seeking to get the lowest price for their ticket, while airline companies are trying to keep their overall revenue as high as possible and maximize their profit.
* **Analytical Problem Framing**

Dataset Representation:



As Clearly, we can see that the target variable i.e., Price is of integer type. So, it’s a regression-based problem.

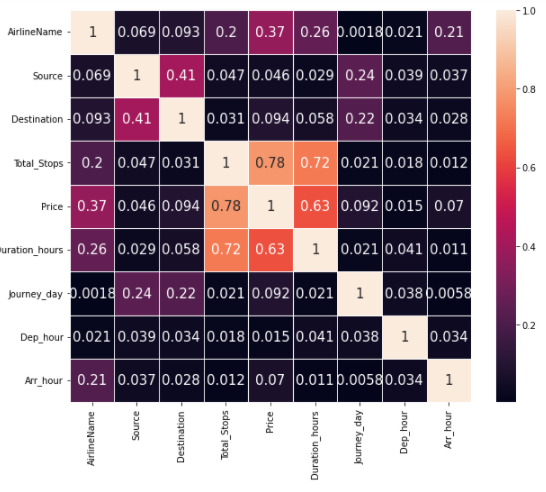
## **Data Sources and their formats & inferences**

* **Airline Name:** The name of the airline.
* **Date**: The date of the journey
* **Source**: The source from which the service begins.
* **Destination**: The destination where the service ends.
* **DepartureTime**: The time when the journey starts from the source.
* **ArrivalTime**: Time of arrival at the destination.
* **Duration**: Total duration of the flight.
* **Total\_Stops**: Total stops between the source and destination.
* **Price**: The price of the ticket

Observation:

* 8 Independent variables with Price as target variables.
* From the dataset we can infer that it is clearly a regression problem.
* The dataset consists of 1892 rows and 10 columns including one unnamed column which will be dropped later on.
* Data Inputs- Logic- Output Relationships

Corelation:



#### Observation:

Total no of stops and duration taken is highly correlated with price.

## Assumption for the problem:

From the dataset we can infer that it is clearly a regression problem.

* Hardware and Software Requirements and Tools Used

Software Used:

* + Jupyter Notebook
  + Ms-Paint
  + MS-PowerPoint
  + MS-Word

Hardware used:

* + Laptop
  + Good internet connectivity

# Model/s Development and Evaluation

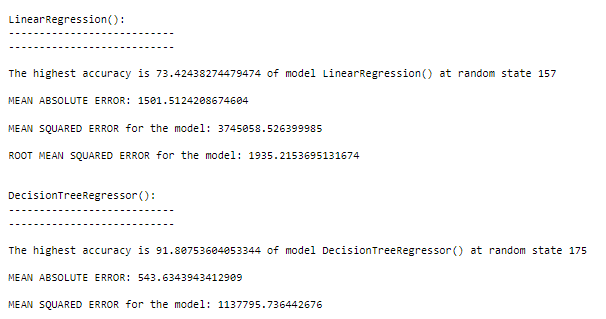
## Testing of Identified Approaches (Algorithms)

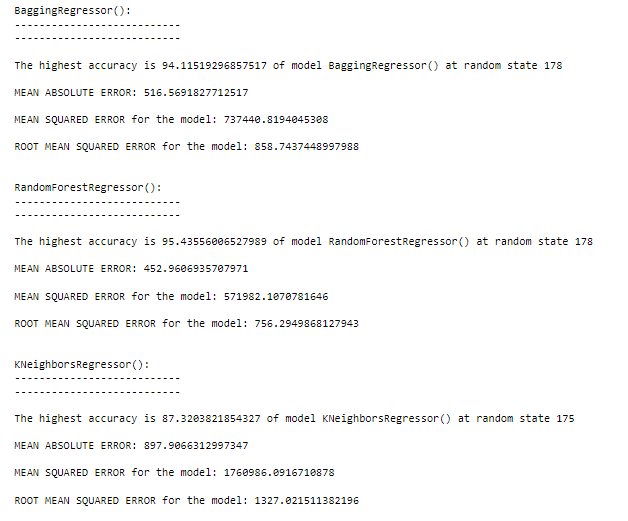
* Linear Regression
* Decision Tree Regressor
* K-Neighbours Regressor
* Bagging Regressor
* Random Forest Regressor
* AdaBoost Regressor
* Run and Evaluate selected models

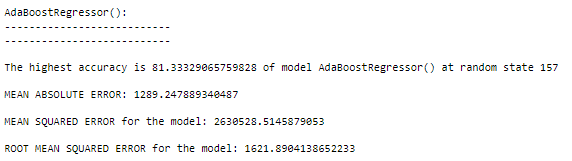
Code:

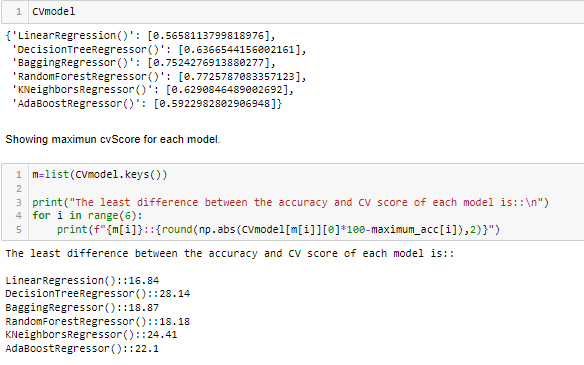


Output:









Here from the above Code and Outputs we finalized model on the basis of first we see the least difference between CV Score and r2 score of each model that we got for Linear regressor but the linear Regressor has high residual. So, we taken second least difference of cv Score and r2 score i.e., Random Forest model.

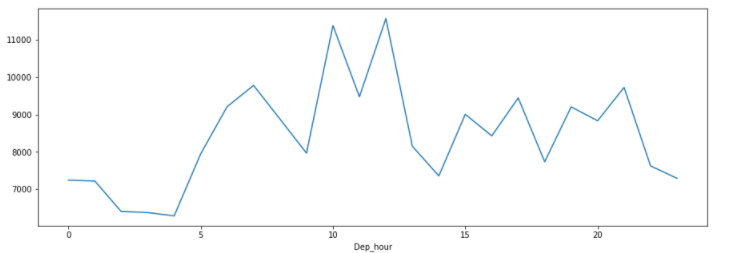
* Key Metrics for success in solving problem under consideration

## Hyper Tuning the Models Random Forest Regressor:



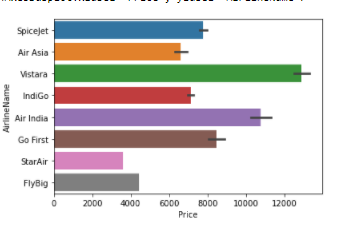
### OBSERVATION:

* Finally, we hyper tuned the finalized model but here I got very minute difference in the r2 score.
* Visualizations



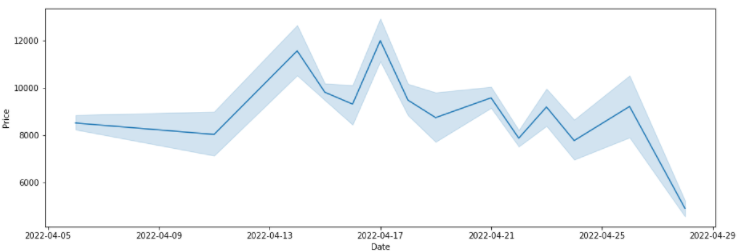
Observations:

Morning flights are less expensive in comparison to others flights.



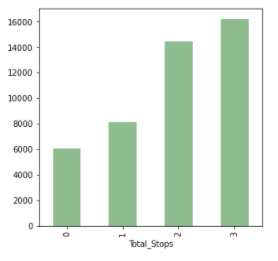
Observations:

* Indigo flights are less expensive in comparison to Spicejet.
* Vistara flights are most expensive.



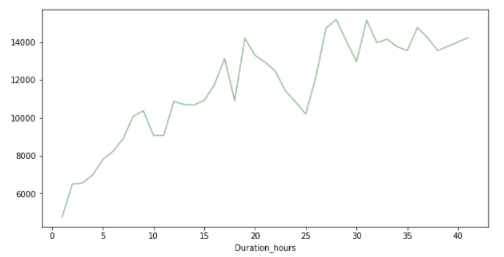
Observations:

* Flights fares are increasing as the departure date of the flights is getting closed. The fares increment over time is with some up and downs.



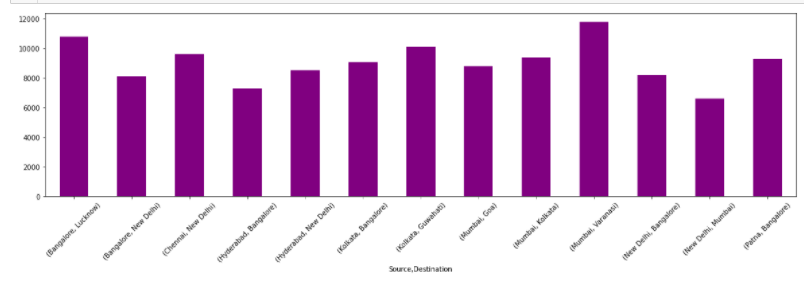
Observations:

* As No. of stops increments for the consumer to reach its destination with cost him expensive.



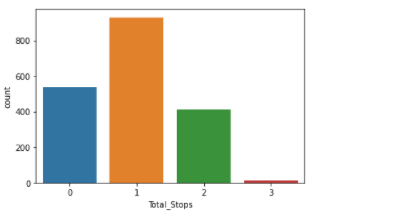
Observations:

* The Customer spends more time in flights cost him expensive to reach the destination.



Observations:

* Mumbai To Varanasi flights are most Expensive.



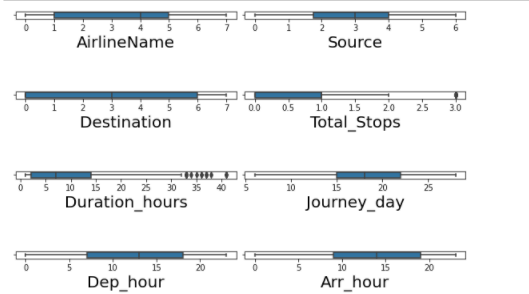
Observations:

Flights with 3 stoppages to reach the destination are very rare.

Conclusion:

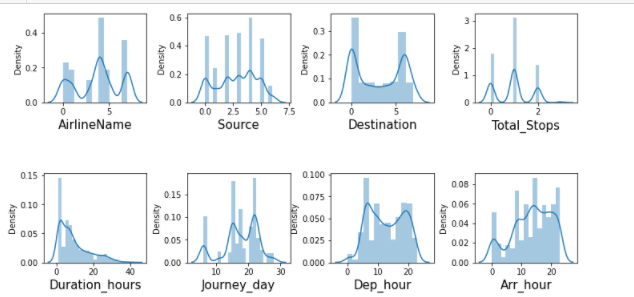
* Person should take early morning flights to reach the destination with less no of stoppages.
* Flights should be pre booked before 12 to 14 days from departure. So, that the person should not get the expensive tickets to buy.

Outlier detection:

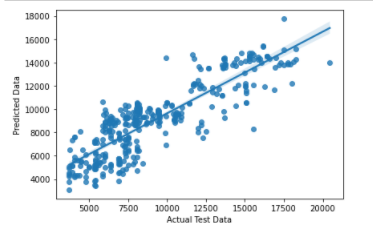


* Seems like Duration hours & Total Stops has outliers init.

Distribution Plot:

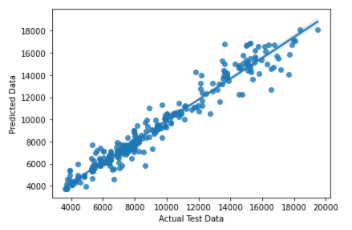


Journey day & Duration hour are not normally distributed.



Linear Regressor

In this graph it is clear that data points are very scatters to the best fit line.



Random Forest Regressor

In this graph it is clear that data points are very closer to the best fit line.

So, from above two graph we get to know the Random Forest is should be our Finalized Model.

**Interpretation of the Results**

* Hence we can go with normal Random Forest Regressor model with is also giving good accuracy.
* This model may help people solve real life problems, before predicting and comparing flight prices and make

a wise decision.

# CONCLUSION

Finally, we have saved the Hyper tuned Random Forest Regressor Model.

## **Learning Outcomes of the Study in respect of Data Science**

* Hence, we can go with normal Random Forest Regressor model which is also giving good accuracy.
* Finally, we have saved the Random Forest Regressor Model.

