

Flight Price Prediction

Tinu Shiby

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 -

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

Data was collected from the website of yatra.com

About the dataset

Dataset have 1630 entries with 10 variables, which is split into train data and test data.

Approach towards the problem

Considered Price as target variable and rest everything as feature variables.

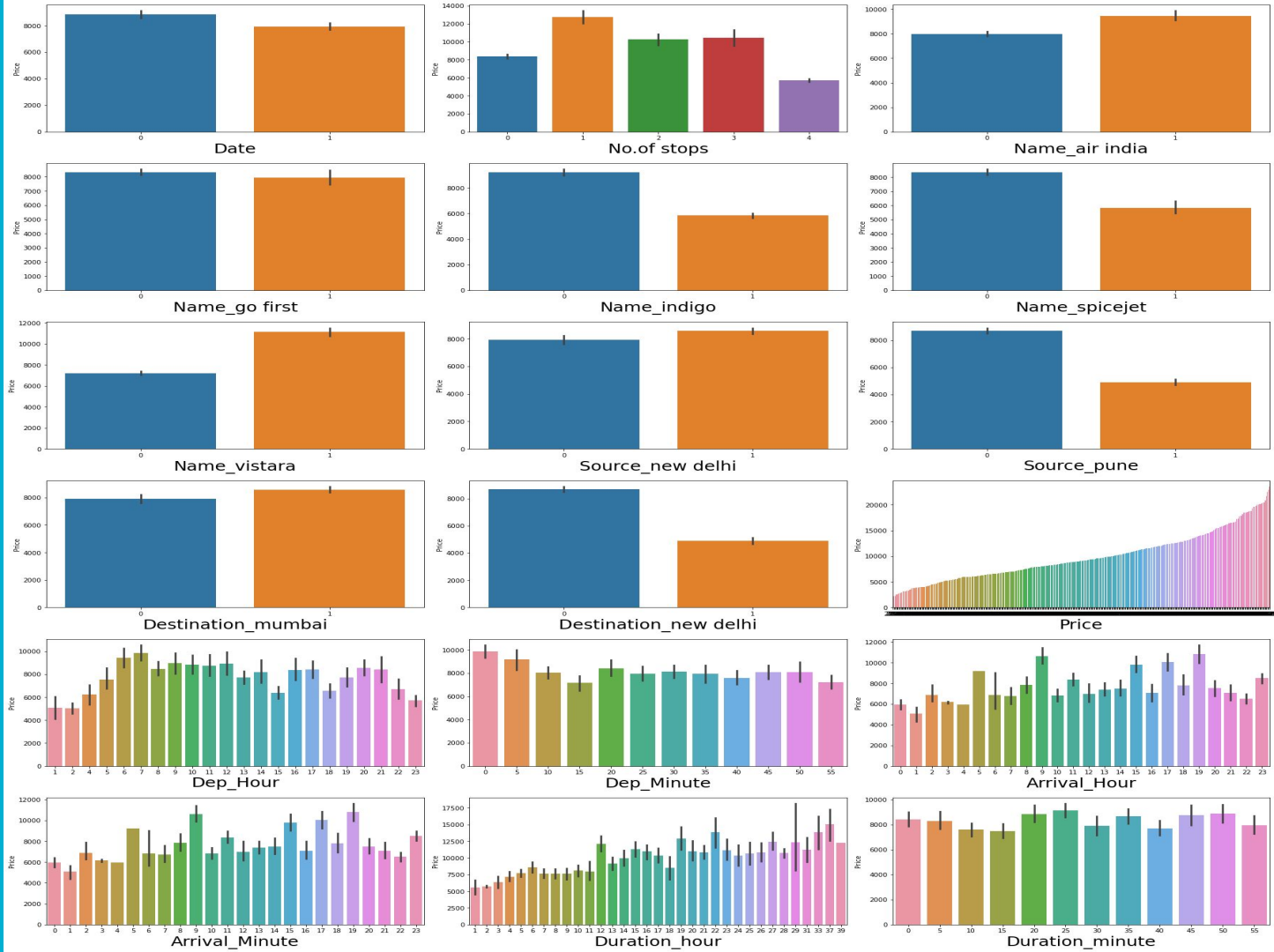
By bivariate analysis tried to figure out which all factors have significance on price prediction

Exploratory Data Analysis

After careful data analysis it was found out that following factors have high correlation with the target variable:

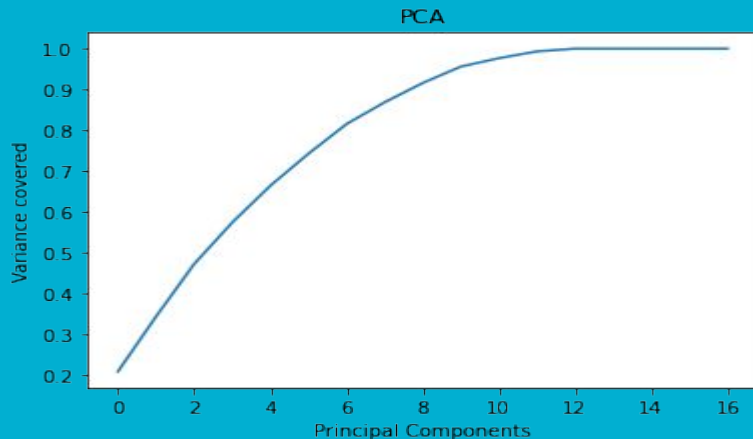
- ❖ Date
- ❖ Number of Stops
- ❖ Airline Name

Graph



Data Preprocessing Steps

- ❖ Removed outliers by z-score method
- ❖ Removed the skewness using power transform
- ❖ Dropped the columns Unnamed: 0 as they have no contribution towards price
- ❖ Applied StandardScalar to standardize the data
- ❖ Applied PCA



Model Training

Developed the following models:

- ❑ Linear Regression
- ❑ Decision Tree
- ❑ Random Forest
- ❑ SVR
- ❑ XGBoost

Best Model

Random Forest Regressor turns out to be the best model as it has the least difference between R2 score and cross-validation score

Hyper Parameter Tuning

Using GridSearchCV hyper parameter tuning is applied on XGBoost model with the following parameters:

- Max_depth : 10
- Min_samples_leaf : 2
- Min_samples_split : 12
- Random_state : 1

After Hyper Parameter tuning the accuracy increased to 50.67%

Conclusion

In this project, I have tried to uncover the underlying trends in flight prices. Nowadays the flight price varies significantly even for the same flight. So customers are seeking to get flight prices at the lowest rate and there comes the scope of this project. For this project I have collected the data from the site Yatra.com and did the feature engineering, exploratory data analysis and data preprocessing steps. Tried with different models and chose Random Forest Regressor as the best model. Random Forest Regressor is working with an accuracy of 51.67%.

Thank You