FLIP ROBO

FLIGHT PRICE PREDICTION

Submitted by:

Rutuja Patil

ACKNOWLEDGMENT

As domestic flight travel in India is becoming increasingly popular with different flight ticket booking channels coming online these days, passengers are trying to understand how these airline companies make decisions over time about ticket prices. Therefore, many methods are ready to provide the proper time to do so. The customer who buys an air ticket by estimating the price of the airfare is recently proposed. The majority of these strategies make use of sophisticated Computational Intelligence Prediction Models an area of science known as Machine Learning (ML). This project highlights the parameters and also includes the guidelines that are important for project work to be developed that is indicated above.

INTRODUCTION

Business Problem Framing

Any individual who has booked a flight ticket previously knows how dynamically costs change. Aircraft uses advanced strategies called Revenue Management to execute a distinctive valuing strategy. The least expensive accessible ticket changes over a period the cost of a ticket might be high or low. This valuing method naturally modifies the toll as per the time like morning, afternoon or night. Cost may likewise change with the seasons like winter, summer and celebration seasons. The extreme goal of the carrier is to build its income yet on the opposite side purchaser is searching at the least expensive cost. Purchasers generally endeavor to purchase the ticket in advance to the takeoff day. Since they trust that airfare will be most likely high when the date of buying a ticket is closer to the takeoff date, yet it is not generally true. Purchaser may finish up with the paying more than they ought to for a similar seat.

Conceptual Background of the Domain Problem

The airline implements dynamic pricing for the flight ticket. According to the survey, flight ticket prices change during the morning and evening time of the day. Also, it changes with the holidays or festival season. There are several different factors on which the price of the flight ticket depends. The seller has information about all the factors, but buyers are able to access limited information only which is not enough to predict the airfare prices. Considering the features such as departure time, the number of days left for departure and time of the day it will give the best time to buy the ticket. The purpose of the paper is to study the factors which influence the fluctuations in the airfare prices and how they are related to the change in the prices. Then using this information, build a system that can help buyers whether to buy a ticket or not.

Review of Literature

It is very difficult for the customer to purchase a flight ticket at the minimum price. For this several techniques are used to obtain the day at which the price of air ticket will be minimum. Most of these techniques are using sophisticated artificial intelligence(AI) research is known as Machine Learning.

Motivation for the Problem Undertaken

Optimal timing for airline ticket purchasing from the consumer’s perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. In this project we simulate various models for computing expected future prices and classifying whether this is the best time to buy the ticket.

Analytical Problem Framing

- ## Mathematical/ Analytical Modeling of the Problem

To develop the model for the flight price prediction, many conventional machine learning algorithms are evaluated. They are as follows: Linear regression, Decision tree, Random Forest Algorithm, K-Nearest neighbors, Multilayer Perceptron, Support Vector Machine (SVM) and Gradient Boosting. All these models are implemented in the scikit learn. To evaluate the performance of this model, certain parameters are considered. They are as follows: R-squared value, Mean Absolute Error (MAE) and Mean Squared Error (MSE).

- ## Data Sources and their formats

The script extracts the information from the website and creates a CSV file as output. This file contains the information with features and its details. Now an important aspect is to select the features that might be needed for the flight prediction algorithm. Output collected from the website contains numerous variable for each flight but not all are required, so only the following feature is considered.

a. Airline
b. flight
c. source_city
d. departure_time
e. stops
f. arrival_time
g. destination_city
h. class
i. duration
j. days_left
k. price

- ## Data Preprocessing Done

All the collected da ta needed a lot of work so after the collection of data, it is needed to be clean and prepare according to the model requirements. All the unnecessary data is removed like duplicates and null values. In all machine learning this technology, this is the most important and time consuming step. Various statistical techniques and logic builtin python are used to clean and prepare the data.

- ## Hardware and Software Requirements and Tools Used

a. Jupyter Notebook

b. Libraries –

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier,
ExtraTreesClassifier, GradientBoostingClassifier
```

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis,
QuadraticDiscriminantAnalysis
from sklearn.metrics import accuracy_score, log_loss
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler, RobustScaler
from sklearn.naive_bayes import GaussianNB
import statsmodels.api as sm
from scipy.stats import probplot
from sklearn import metrics
from sklearn import preprocessing
from sklearn import utils
from sklearn.metrics import mean_absolute_error
from statistics import mean, stdev
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score, roc_curve, auc, confusion_matrix,
plot_confusion_matrix
from sklearn.model_selection import KFold, cross_val_score, cross_validate,
StratifiedKFold
from sklearn.svm import SVC
from sklearn.feature_selection import chi2
from sklearn.feature_selection import SelectKBest
```

Model/s Development and Evaluation

• Identification of possible problem-solving approaches (methods)

To develop the model for the flight price prediction, many conventional machine learning algorithms are evaluated. They are as follows: Linear regression, Decision tree[8], Random Forest Algorithm. All these models are implemented in the scikit learn. To evaluate the performance of this model, certain parameters are considered. They are as follows: R-squared value, Mean Absolute Error (MAE) and Mean Squared Error (MSE).

• Testing of Identified Approaches (Algorithms)

- Linear Regression

  Regression is a method of modeling a target value based on predictors that are independent. It is mostly based on the number of independent variables and the relationship between independent and dependent variables. linear regression is a type of analysis where the number of independent variables is one and the relationship between the dependent and independent variables vary linearly. The important concept to understand linear regressions are cost function and Gradient decent.

- Decision tree

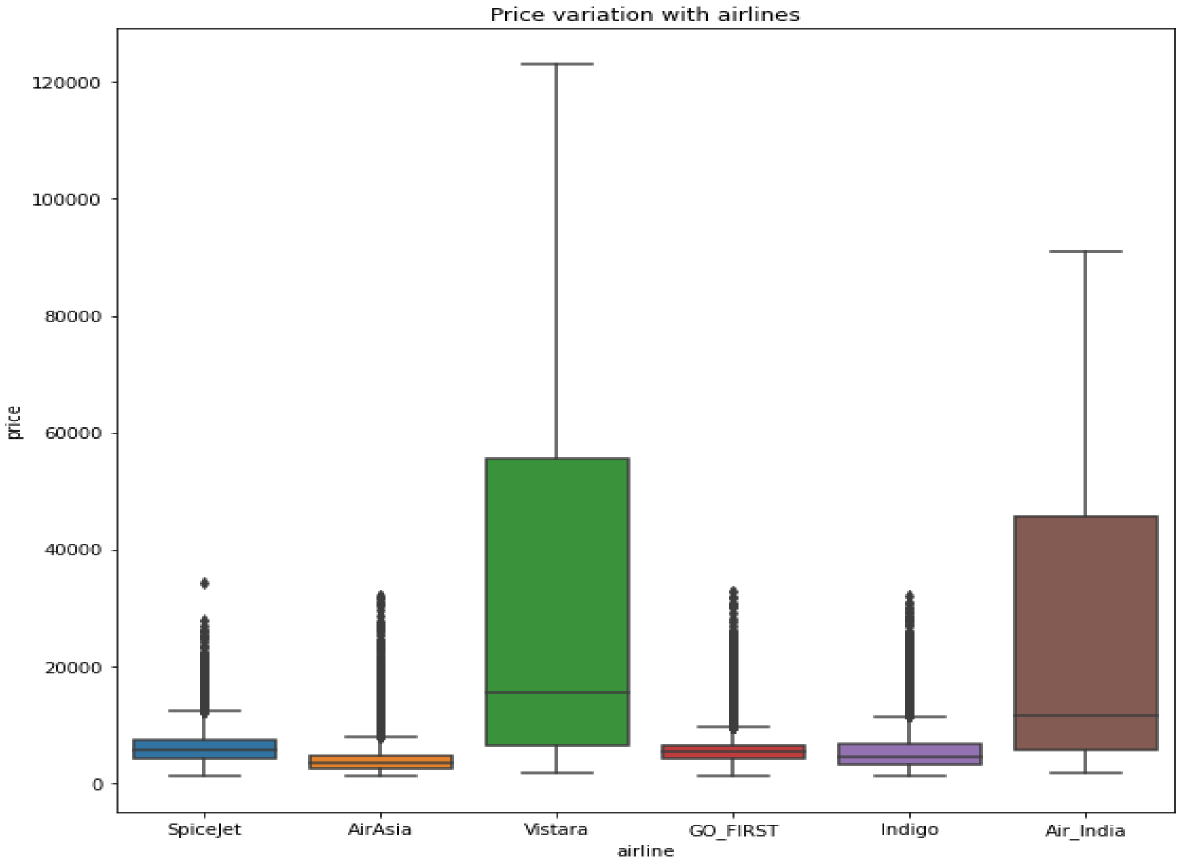
  The Decision tree calculation separates the informational collection into small subsets, at a similar same time it creates gradually. The last outcomes are the tree with the decision nodes, whats more, the leaf nodes. A decision hub may have at least two branches. In the beginning, consider the entire informational collection as root. Highlight esteems are wanted to be downright. On the off chance that the qualities are constant then they are discretized before structure the model. Based on characteristic qualities records are dispersed recursively. There are two primary characteristics in the decision tree calculation. One is Information Gain and another is the Gini index. Information Gain is the proportion of Change in entropy. Higher the entropy more the instructive substance, where the entropy is a proportion of vulnerability of arbitrary variable .

- Random Forest

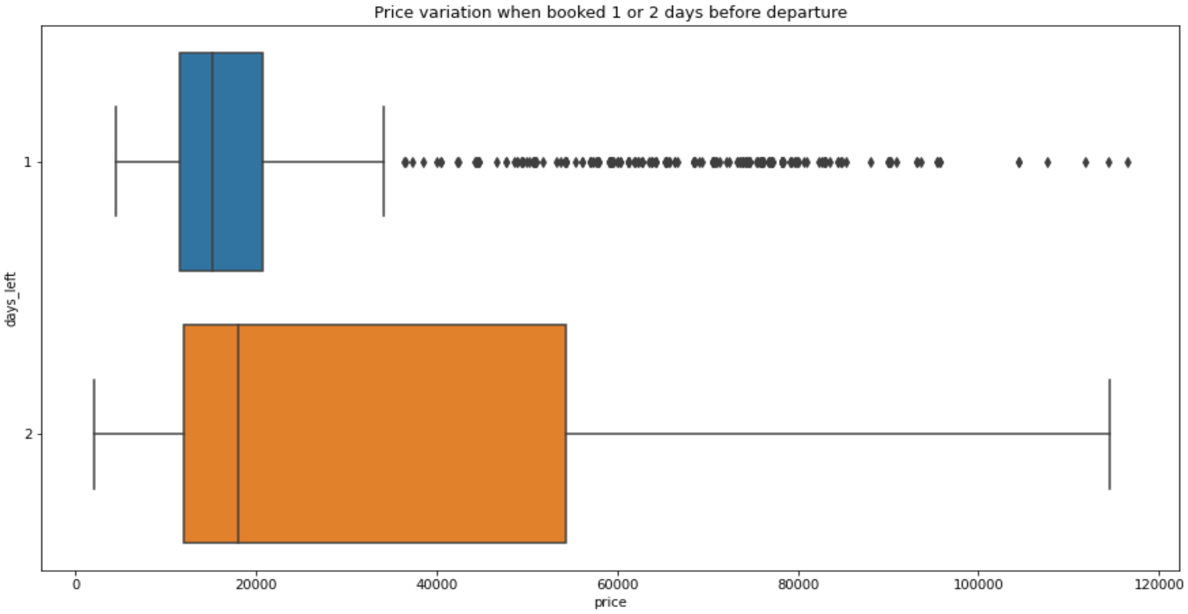
  It is a supervised learning algorithm. The benefit of the random forest is, it very well may be utilized for both characterization and relapse issue which structure most of current machine learning framework. Random forest forms numerous decision trees, whats more, adds them together to get an increasingly exact and stable expectation. Random Forest has nearly the equivalent parameters as a decision tree or a stowing classifier model. It is very simple to discover the significance of each element on the expectation when contrasted with others in this calculation.

  .

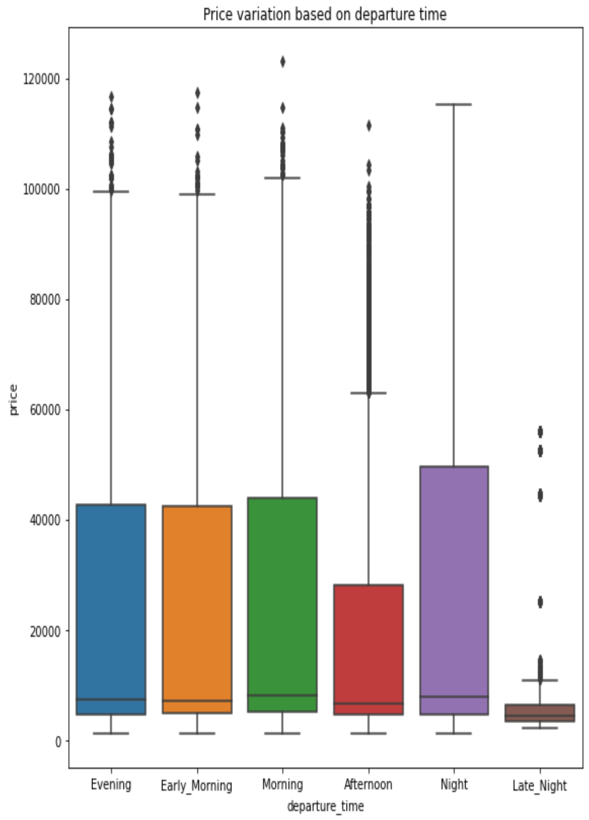
- Interpretation of the Results
- **Does prices vary with airlines?**

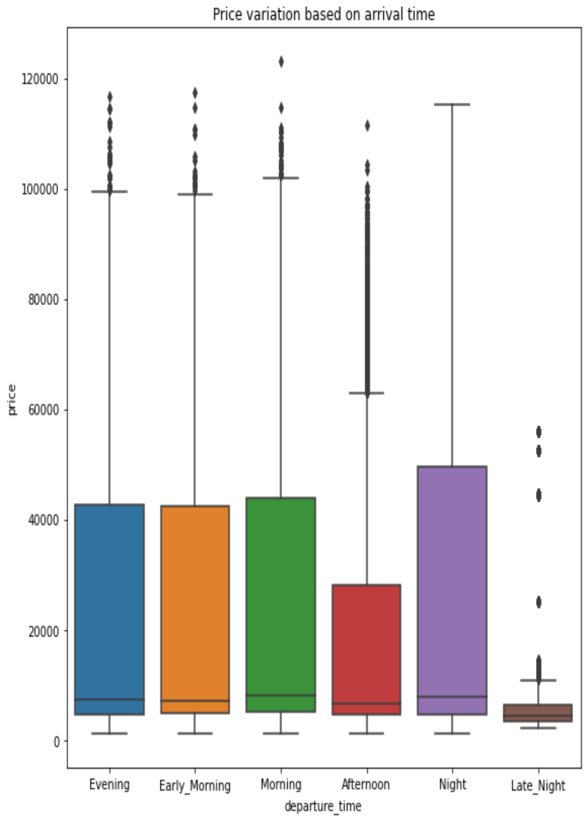


- **How is the price affected when tickets are bought in just 1 or 2 days before departure?**

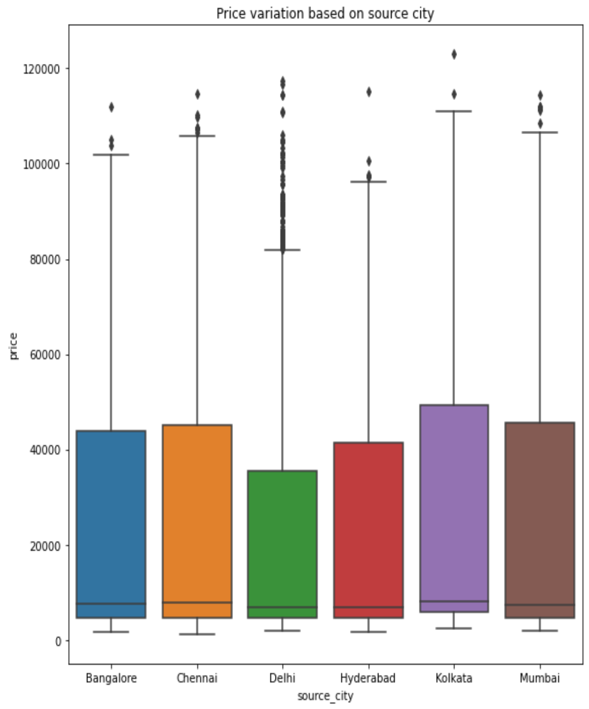


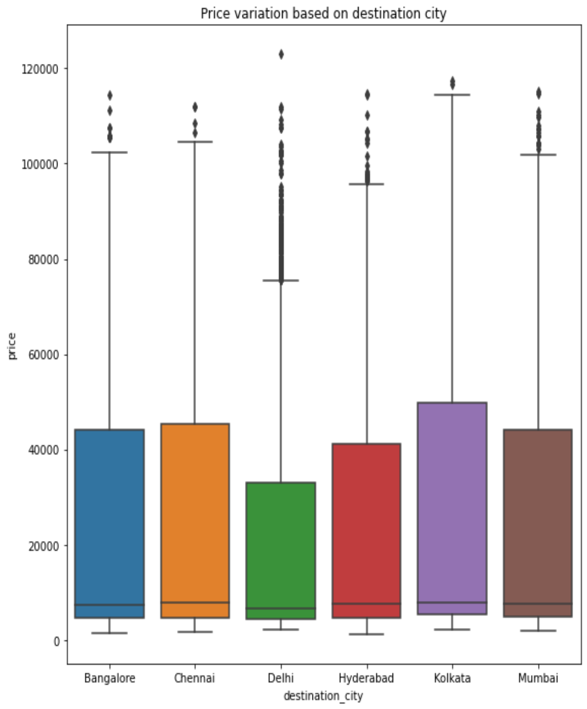
- **Does ticket price change based on the departure time and arrival time?**



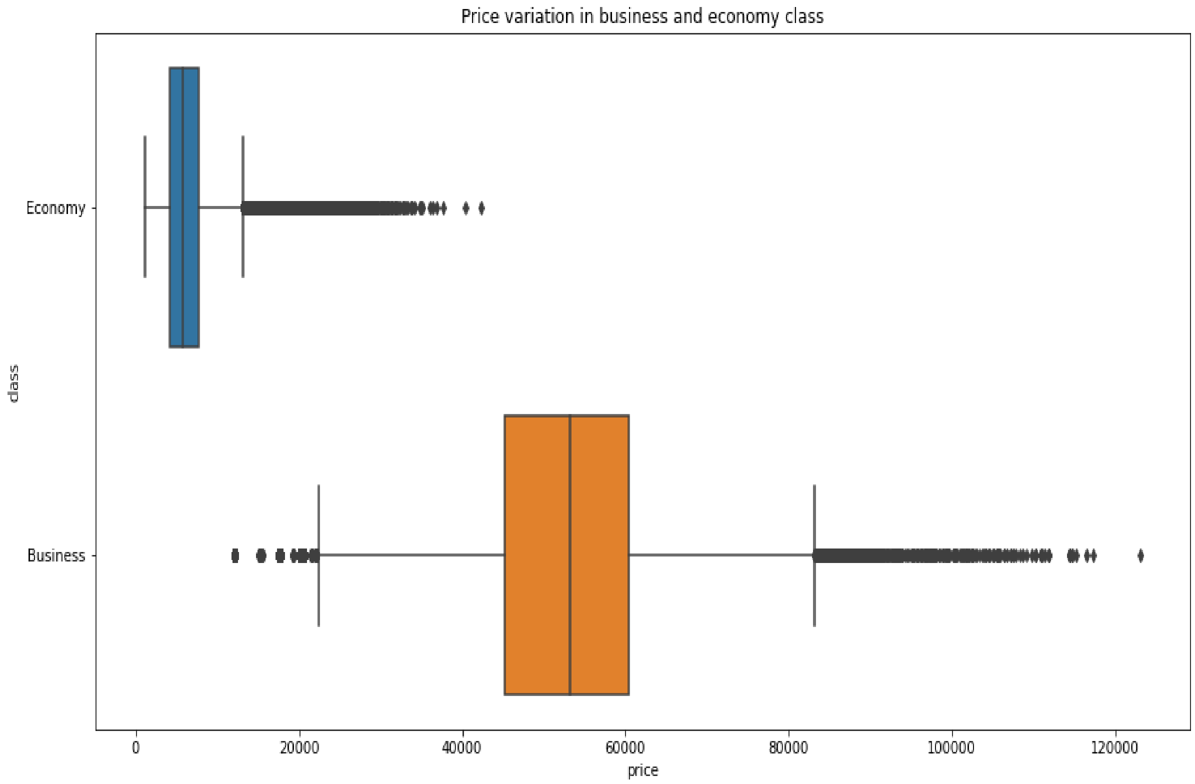


- **How the price changes with change in Source and Destination?**





- **How does the ticket price vary between Economy and Business class?**



CONCLUSION

From the data collected and through exploratory data analysis, we can determine the following:

- The trend of flight prices vary over various months and across the holiday
- The airfare varies depending on the time of departure, making timeslot used in analysis an important parameter.
- The airfare increases during a holiday season. In our time period, during Diwali the fare remained high for all the values of days to departure. We haven’t considered holiday season as a parameter now, since we are looking at data for a few months.
- Airfare varies according to the day of the week of travel. It is higher for weekends and Monday and slightly lower for the other days.
- There are a few times when an offer is run by an airline because of which the prices drop suddenly. These are difficult to incorporate in our mathematical models, and hence lead to error.
- Along the Mumbai-Delhi route, we find that the price of flights increases or remains constant as the days to departure decreases. This is because of the high frequency of the flights, high demand and also could be due to heavy competition.