# FLIGHT PRICING MODEL

Problem Statement

Airline companies use complex algorithms to calculate flight prices given various conditions present at that particular time. These methods take financial, marketing, and various social factors into account to predict flight prices.

Nowadays, the number of people using flights has increased significantly. It is difficult for airlines to maintain prices since prices change dynamically due to different conditions. That's why we will try to use machine learning to solve this problem. This can help airlines by predicting what prices they can maintain. It can also help customers to predict future flight prices and plan their journey accordingly. Anyone who has booked a flight ticket knows how unexpectedly the prices vary.

Theoretical Background

- . 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.
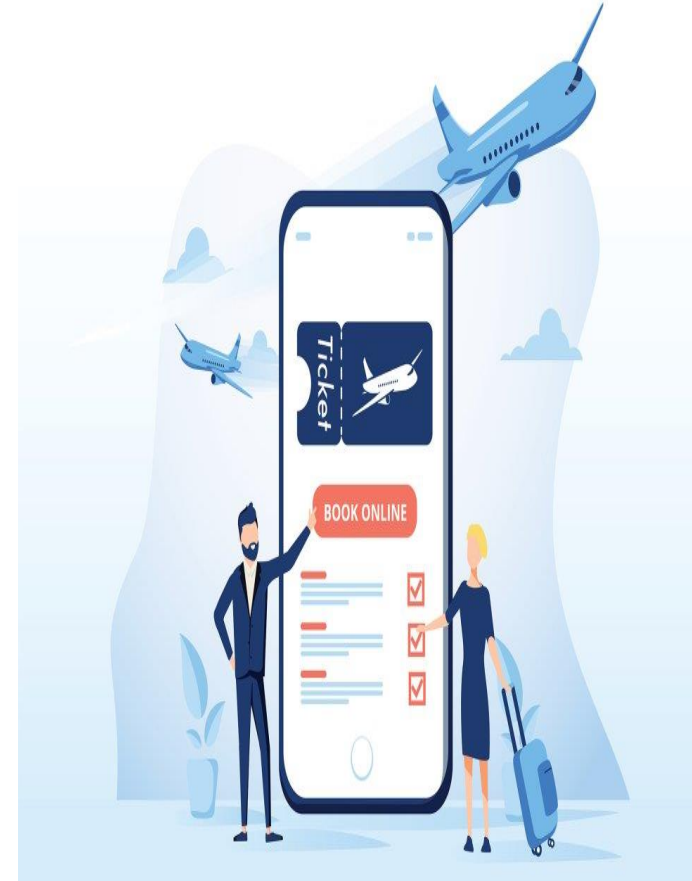
EDA

```
ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1230 entries, 0 to 1229
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Airlines        1230 non-null   object
 1   Date_of_Journey 1230 non-null   object
 2   Source          1230 non-null   object
 3   Destination     1230 non-null   object
 4   Dep_times       1230 non-null   object
 5   Arrival_time    1230 non-null   object
 6   Duration        1230 non-null   object
 7   Stops           1230 non-null   object
 8   Price           1230 non-null   int64
dtypes: int64(1), object(8)
memory usage: 86.6+ KB
```
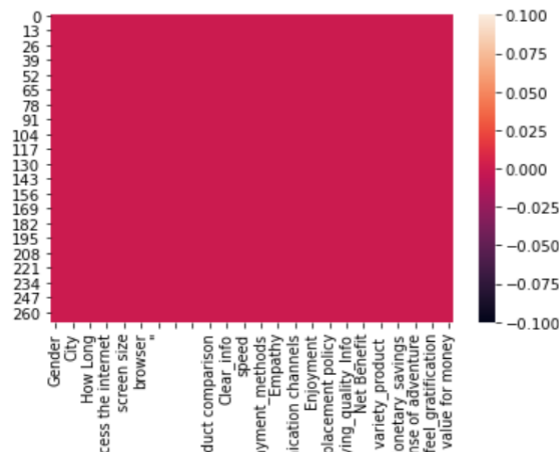
here are 1203 rows and 9 columns.All are object type varibale. Target varibale is continous and interger type.

The dataset contains 1203 rows and 9 columns in the dataset.

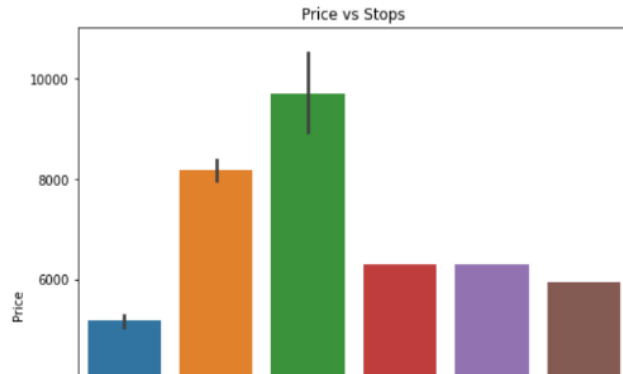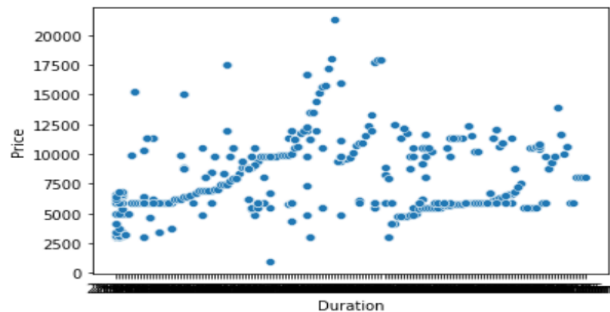**The heat map shows there is no missing values in the data set.**

Preparing the data for Analysis and Model Building

```
plt.figure(figsize=(8,8))
sns.barplot(x='Stops',y='Price',data=ds)
plt.title("Price vs Stops")
plt.xticks(rotation=45)
plt.show()
```


Price vs Stops

```
plt.figure(figsize=(8,8))
sns.barplot(x='Airlines',y='Price',data=ds)
plt.title("Price vs Airlines")
plt.xticks(rotation=45)
plt.show()
```


Price vs Airlines

```
#duration v/s AveragePrice
sns.scatterplot(data=ds, x='Duration', y='Price')
```

```
<AxesSubplot:xlabel='Duration', ylabel='Price'>
```
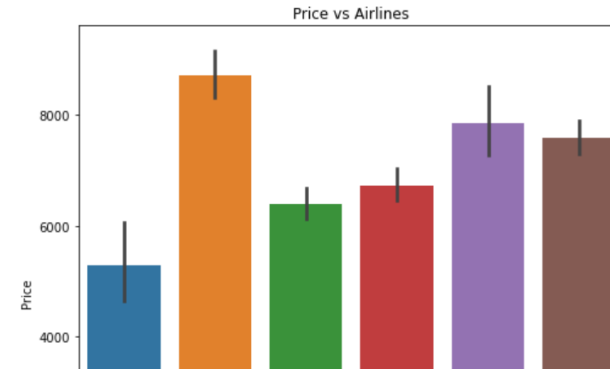

We know that duration is important and plays a major role in affecting air ticket prices but we see no such pattern here, there must be other significant factors affecting air fare like type of airline, destination of flight, date of journey of flight.

• The flights with destination Mumbai has the highest flight price then Hyderbad ,New delhi respectively.

• As the number of stops increases the price of flight also increases.

Model Building

We built LinearRegression, Decision Tree Regessor, and Random Forest as machine learning. We built base models of LinearRegression, XGBoost, and RandomForest so there is not much to show about these models but we can see the model summary and how they converge with deep learning models that we built.

- Majority of the customer agree and strongly agree that monetary savings and empathy toward the customers helps in purchasing the product.

Model Evaluation

Based on the results of above models, and capomaring the R2 score and other evalution matrics result of MAE,MSE and RMSE.
We can find the Random Forest Regression model is best model to predict the price of used cars Score

Score	R2 Score	MAE	MSE	RMSE 1	
LinearRegression	0.47	0.16	110	1800	134
DecisionTreeRegression	0.97	0.71	469	113	1392
KNeighbors Regression	0.69	0.53	70	10229	1795
Random Forest Regression	0.95	0.84	522	498	1017

Since the Random Forest Regression has the hishest score(1) and R2 score(0.99) and lowest values of MAE, MSE,RMSE , it is the best model among the above four models

It can be observed that the "Random Forest Regression " algorithim has the almost some positive accuracy score after the cross validation. So the best model with highet accuracy score and best on evalution with other matrix is "DecisionTree" is the best model for predicting the Price.

Conclusion

❑ In this project, we tried predicting the flight price using the various parameters that were provided in the data about the flight. We build machine learning models to predict prices and saw that machine learning-based models performed well at this data.

❑ By performing different ML models, we aim to get a better result or less error with max accuracy. Our purpose was to predict the price of the used cars having 9 predictors and 1203 data entries. Initially, data cleaning is performed to remove the null values and outliers from the dataset then ML models are implemented to predict the price of flight. Next, with the help of data visualization features were explored deeply. The relation between the features is examined.

❑ From the below table, it can be concluded that Random Forest Regressor is the best model for the prediction of flight prices. The regression model gave the best MSLE and RMSLE values

Thank You

Submitted by
Dinesh Sharma