FLIGHT PRICE PREDICTION

Submitted by:

Ganesh kumbhar

INTRODUCTION

- ## Business Problem Framing

  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)

- ## Conceptual Background of the Domain Problem

  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

- Review of Literature

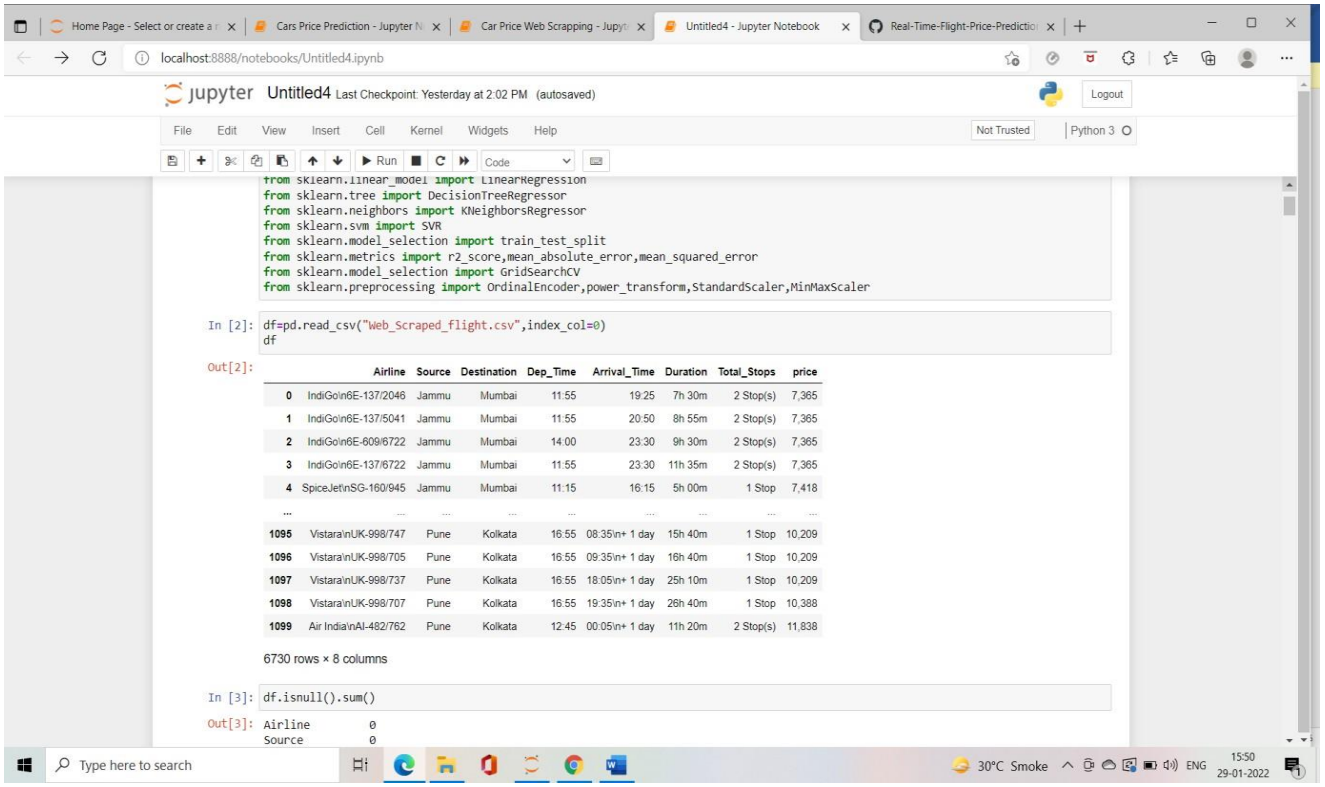
  As per the requirement of client, I have scrapped the data from online sites and based on that data I have did analysis like for based on which feature of my data prices are changing. and checked the relationship of flight price with all the feature like what flight he should choose.

- Motivation for the Problem Undertaken

  I have worked on this on the bases of client requirements and followed all the steps till model deployment.

Analytical Problem Framing

After scrapping my data using selenium I have loaded my data into python with the help of pandas.



The size of the data is 6730*8

Data Pre-processing

Checking null values

```
data.isna().sum()
```

```
Unnamed: 0      0
Unnamed: 0.1    0
Airline         0
Source          0
Destination     0
Dep_Time        0
Arrival_Time    0
Duration        0
Total_Stops     0
Price           0
dtype: int64
```

There are no missing values

as there is no null values so I can move forward

- Data Sources and their formats

  I have collected data from web scrapping and I have converted it into csv format

- Data Preprocessing Done

  Doing pre-processing where I am dropping some columns and filling missing values in total stops

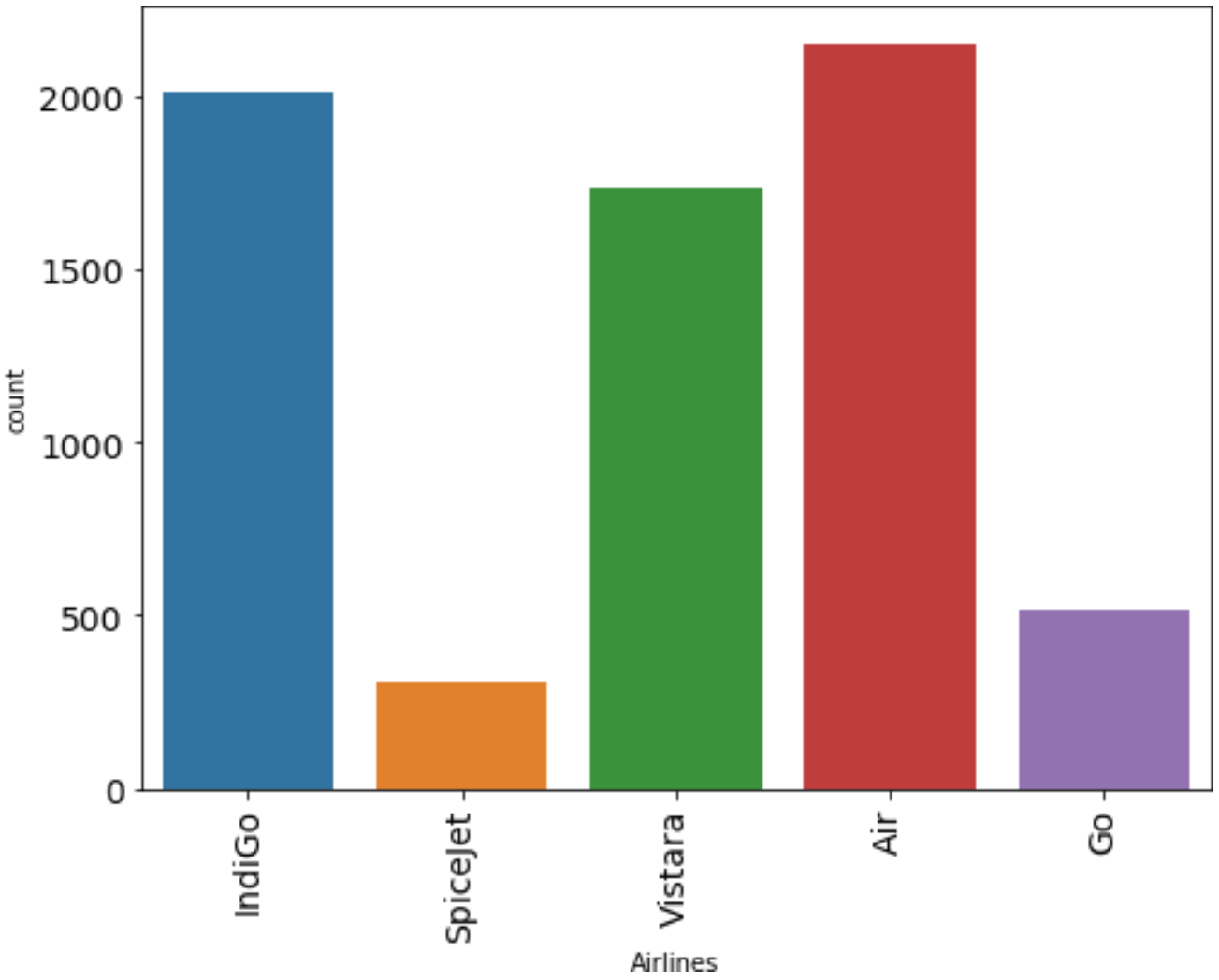
```
def preprocess1(df):
  df['Total_Stops']=df['Total_Stops'].fillna(df['Total_Stops'].mode()[0])
  df=df.drop(['Duration'],axis=1)
  return df
```

```python
def preprocess2(df):
  df['Dep_hour']=pd.to_datetime(df['Dep_Time']).dt.hour
  df['Dep_minute']=pd.to_datetime(df['Dep_Time']).dt.minute
  df=df.drop(['Dep_Time'],axis=1)
  df['arrival_hour']=pd.to_datetime(df['Arrival_Time']).dt.hour
  df['arrival_minute']=pd.to_datetime(df['Arrival_Time']).dt.minute
  df=df.drop(['Arrival_Time'],axis=1)
  return df
```

Here I am converting time into hour and minute and also dropping some columns that are not useful for my model.

- Data Inputs- Logic- Output Relationships

I have did EDA to understand the feature relationship.



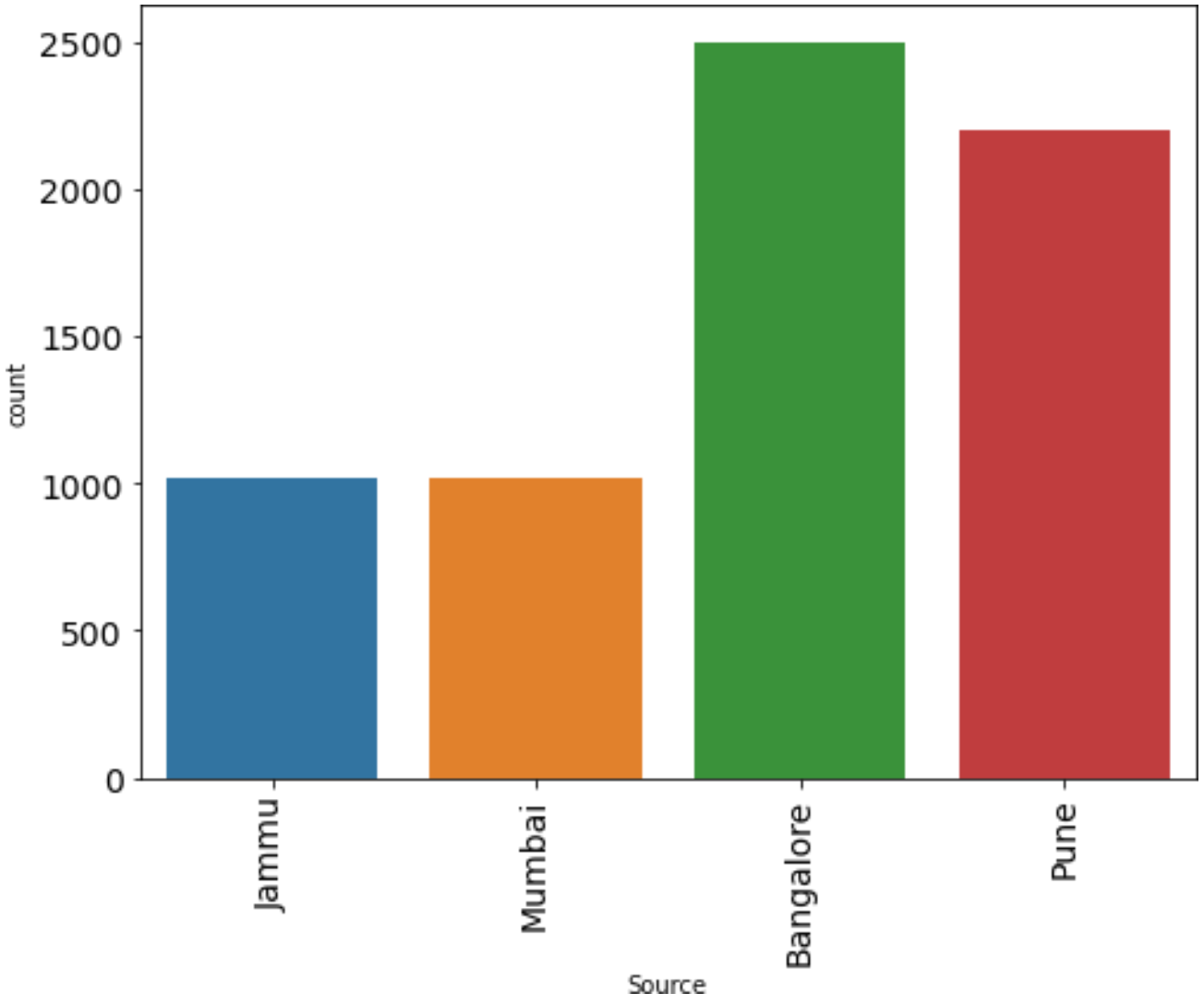
Obseravtion

1- Mostly people use to travel with Air India

2- After air india people use to travel with IndiGo

3- ANd spicejet has the least count

Countplot of Source



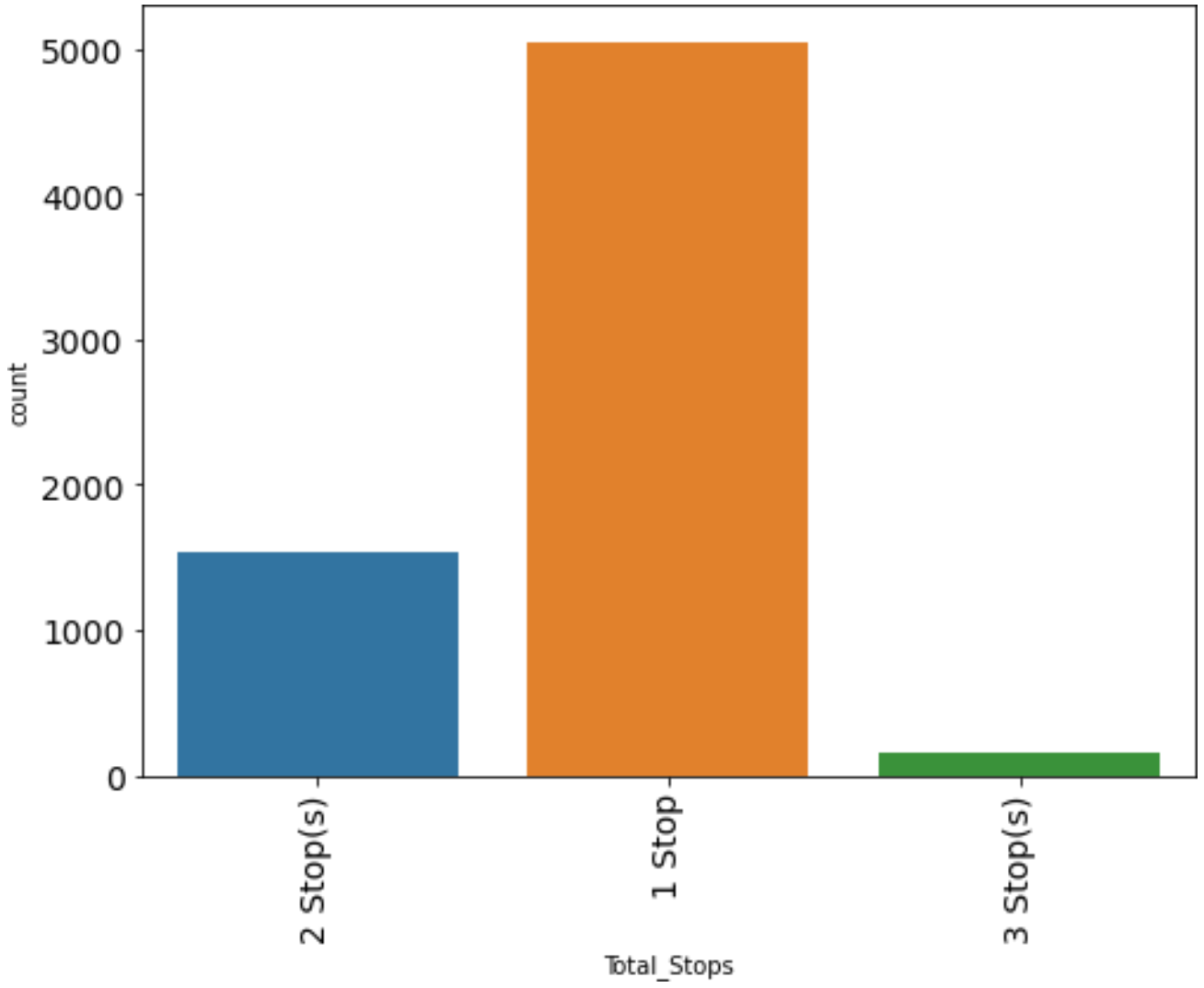
Obseravtion

1-Mostly Source has bangalore as high count

2-after bangalore pune has 2nd high count

3-and at least jammu and mumbai equal count

Countplot of Total Stops



Obseravtion

1-Mostly people use to take a fligh who has only one stop

3-ANd only approx 1500 people use to take 2 stop flights

4-ANd there are very less people who used to take 3 stops or 4 stops fligh

- State the set of assumptions (if any) related to the problem under consideration

  Here, you can describe any presumptions taken by you.

- Hardware and Software Requirements and Tools Used

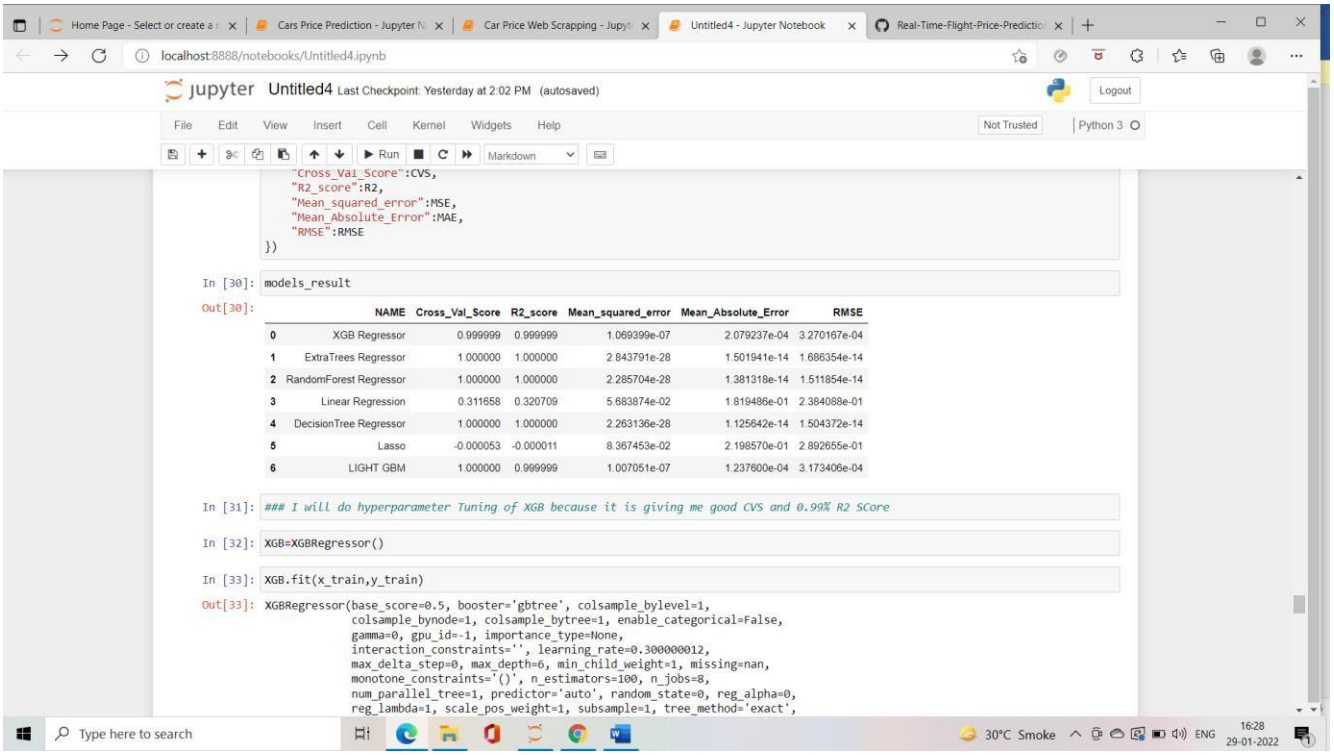
Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

I have did Analysis on this data to understand the value of each feature and the contribution of each feature for model creating and effect of all the feature on the prices. And considering all the point I have built a model that can predict the prices.

Testing of Identified Approaches (Algorithms)

I have trained many model and even evaluate them using all the performance metrics of regression. Here is the screenshot
I have make a dataframe of all the model and metrics so here we can see the performance of every model. I have selected LIGHT



GBM as a final model because of its accuracy and performance metrics.

- Key Metrics for success in solving problem under consideration

  I have make a dataframe of all the model and metrics so here we can see the performance of every model. I have selected LIGHT GBM as a final model because of its accuracy and performance metrics.

- Interpretation of the Results

  From the above eda we can easily understand the relationship between features and and we can even see which things are effecting the price of flights.

CONCLUSION

- Key Findings and Conclusions of the Study

  Describe the key findings, inferences, observations from the whole problem.

- Learning Outcomes of the Study in respect of Data Science

  The above research will help our client to study the latest flight price market and with the help of the model built he can easily predict the price ranges of the flight, and also will helps him to uLimitatinderstand Based on what factors the fight price is decided.

ons of this work and Scope for Future Work

The limitation of the study is that in the volatile changing market we have taken the data, to be more precise we have taken the data at the time of pandemic and recent data, so when the pandemic ends the market correction might happen slowly. So based on that again the deciding factors of the might change and we have shortlisted and taken these data from the important cities across india, if the customer is from the different city our model might fail to predict the accuracy prize of that flight.

THANK YOU !!!