NAME OF THE PROJECT

Flight Price Prediction

Submitted by:

Nomaan Sayed

ACKNOWLEDGMENT

Different websites like Yatra, Makemytrip etc where we got to know about the frequency of flights and tariffs.

INTRODUCTION

- Business Problem Framing
  Many flights are remaining empty at the time of departure, So to avoid this many companies are using the last moment high rates strategy.
- Conceptual Background of the Domain Problem
  If someone books the flight 4-5 months prior to its departure then he will get maximum discount and other who is booking close to the departure he has to pay more prices, this is done to fill-up the entire flight.

Analytical Problem Framing

- Data Sources and their formats
  Data sources is from the website of Yatra.com.
- Data Preprocessing Done
- Data Inputs- Logic- Output Relationships

1- Selenium was used to scrape the data
2- Checked the Data type
3- Checked the null values
4- Checked the outliers
5- Data visualization using seaborn

- State the set of assumptions (if any) related to the problem under consideration
  No such assumptions.
- Hardware and Software Requirements and Tools Used

Hardware- core i5, 8 gb ram

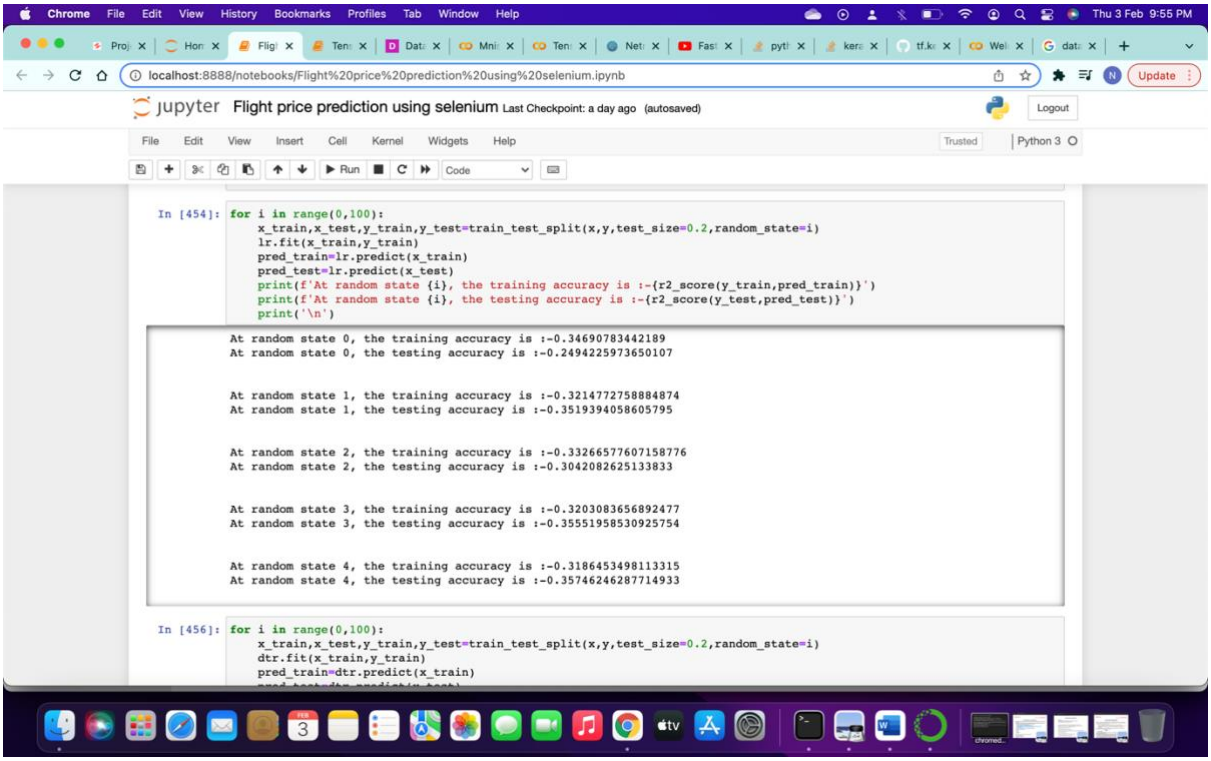
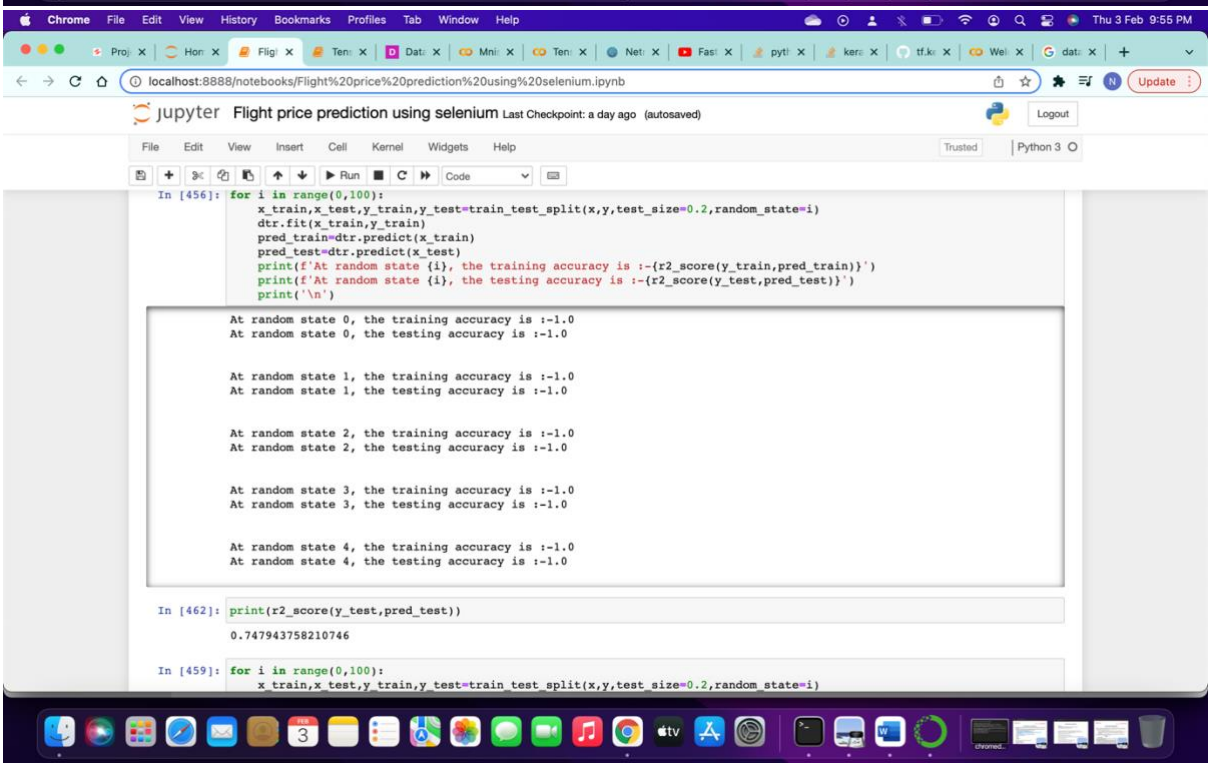
Software- Jupyter notebook- Python

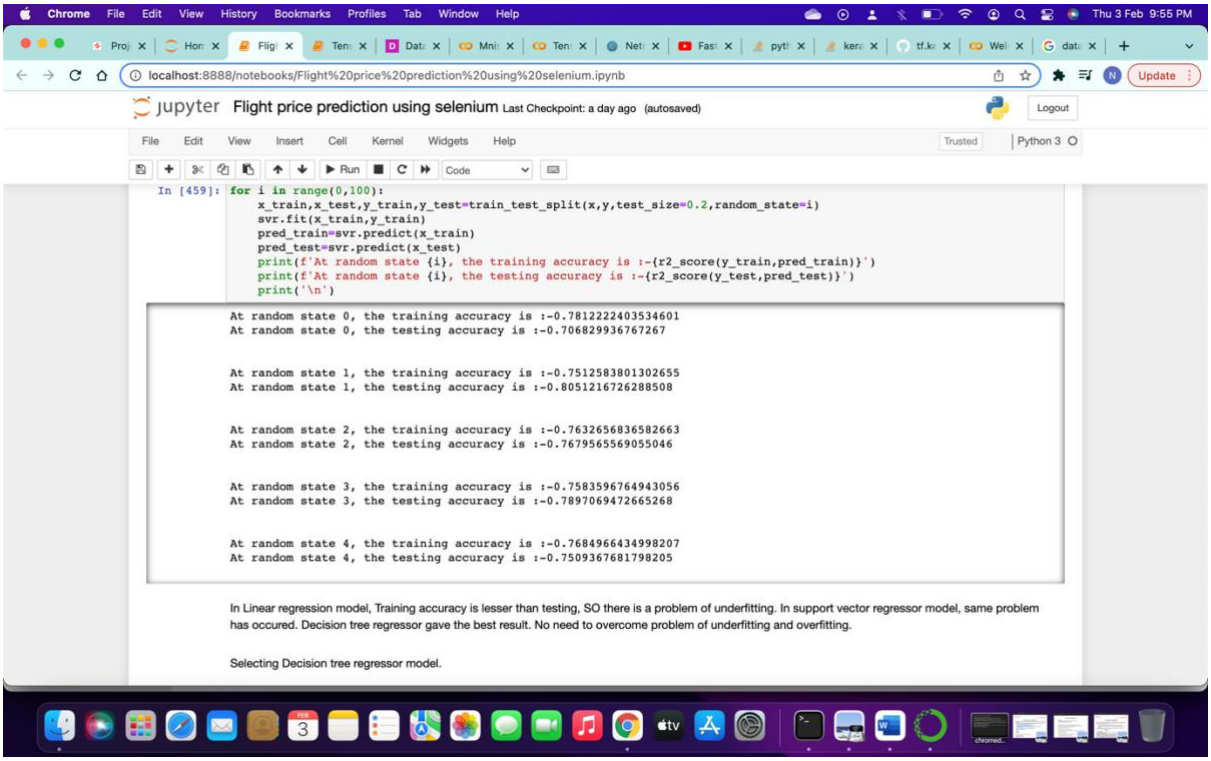
Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)
  1- Data scraped from different websites
  2- Made the DataFrame using Pandas
  3- Checked the outliers and null values.
  4- Encoding of DataFrame
  5- Data Visualization
  6- Describing the dataset
  7- Correlation of the dataset
  8- Creating heatmap
  9- Separating x and y
  10- Used power_transform

11- Scaling the data
12- Model building

- Testing of Identified Approaches (Algorithms)
  1- Linear regression.
  2- Decision tree regressor
  3- Support vector regressor
- Run and Evaluate selected models


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Thu 3 Feb 9:55 PM
localhost:8888/notebooks/Flight%20price%20prediction%20using%20selenium.ipynb
Update
jupyter Flight price prediction using selenium Last Checkpoint: a day ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [454]:
for i in range(0,100):
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=i)
lr.fit(x_train,y_train)
pred_train=lr.predict(x_train)
pred_test=lr.predict(x_test)
print(f'At random state {i}, the training accuracy is :-{r2_score(y_train,pred_train)}')
print(f'At random state {i}, the testing accuracy is :-{r2_score(y_test,pred_test)}')
print('\n')
At random state 0, the training accuracy is :-0.34690783442189
At random state 0, the testing accuracy is :-0.2494225973650107
At random state 1, the training accuracy is :-0.3214772758884874
At random state 1, the testing accuracy is :-0.3519394058605795
At random state 2, the training accuracy is :-0.33266577607158776
At random state 2, the testing accuracy is :-0.3042082625133833
At random state 3, the training accuracy is :-0.3203083656892477
At random state 3, the testing accuracy is :-0.35551958530925754
At random state 4, the training accuracy is :-0.3186453498113315
At random state 4, the testing accuracy is :-0.35746246287714933
In [456]:
for i in range(0,100):
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=i)
dtr.fit(x_train,y_train)
pred_train=dtr.predict(x_train)

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Thu 3 Feb 9:55 PM
localhost:8888/notebooks/Flight%20price%20prediction%20using%20selenium.ipynb
Update
jupyter Flight price prediction using selenium Last Checkpoint: a day ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [456]:
for i in range(0,100):
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=i)
dtr.fit(x_train,y_train)
pred_train=dtr.predict(x_train)
pred_test=dtr.predict(x_test)
print(f'At random state {i}, the training accuracy is :-{r2_score(y_train,pred_train)}')
print(f'At random state {i}, the testing accuracy is :-{r2_score(y_test,pred_test)}')
print('\n')
At random state 0, the training accuracy is :-1.0
At random state 0, the testing accuracy is :-1.0
At random state 1, the training accuracy is :-1.0
At random state 1, the testing accuracy is :-1.0
At random state 2, the training accuracy is :-1.0
At random state 2, the testing accuracy is :-1.0
At random state 3, the training accuracy is :-1.0
At random state 3, the testing accuracy is :-1.0
At random state 4, the training accuracy is :-1.0
At random state 4, the testing accuracy is :-1.0
In [462]:
print(r2_score(y_test,pred_test))
0.747943758210746
In [459]:
for i in range(0,100):
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=i)



- Key Metrics for success in solving problem under consideration

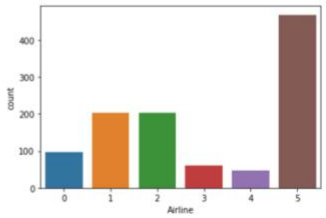
  R2_score metrics is used.
- Visualizations

Visualization of DataSet

In [419]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

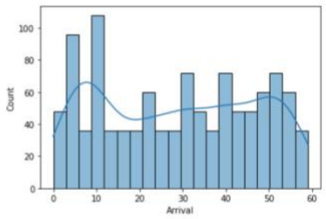
In [420]:
```
sns.countplot(x='Airline',data=df)
print(df['Airline'].value_counts())
```

```
5    468
2    204
1    204
0     96
3     60
4     48
Name: Airline, dtype: int64
```



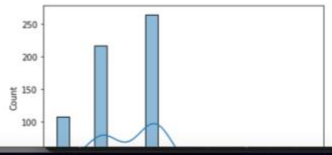
In [422]:
```
sns.histplot(x='Arrival',data=df,bins=20,kde=True)
```

Out[422]:
```
<AxesSubplot:xlabel='Arrival', ylabel='Count'>
```



In [423]:
```
sns.histplot(x='Departure',data=df,bins=20,kde=True)
```

Out[423]:
```
<AxesSubplot:xlabel='Departure', ylabel='Count'>
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help Thu 3 Feb 9:58 PM

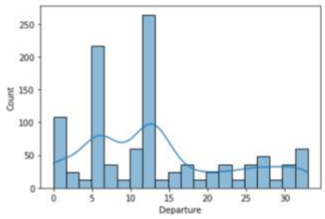
localhost:8888/notebooks/Flight%20price%20prediction%20using%20selenium.ipynb

jupyter **Flight price prediction using selenium** Last Checkpoint: a day ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

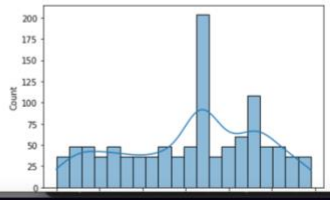
Run Code

Out[423]: <AxesSubplot:xlabel='Departure', ylabel='Count'>



In [425]:
```
sns.histplot(x='Duration',data=df,bins=20,kde=True)
```

Out[425]: <AxesSubplot:xlabel='Duration', ylabel='Count'>



Chrome File Edit View History Bookmarks Profiles Tab Window Help Thu 3 Feb 9:58 PM

localhost:8888/notebooks/Flight%20price%20prediction%20using%20selenium.ipynb

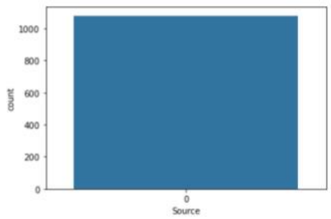
jupyter **Flight price prediction using selenium** Last Checkpoint: a day ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

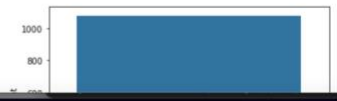
In [428]:
```
sns.countplot(x='Source',data=df)
print(df['Source'].value_counts())
```

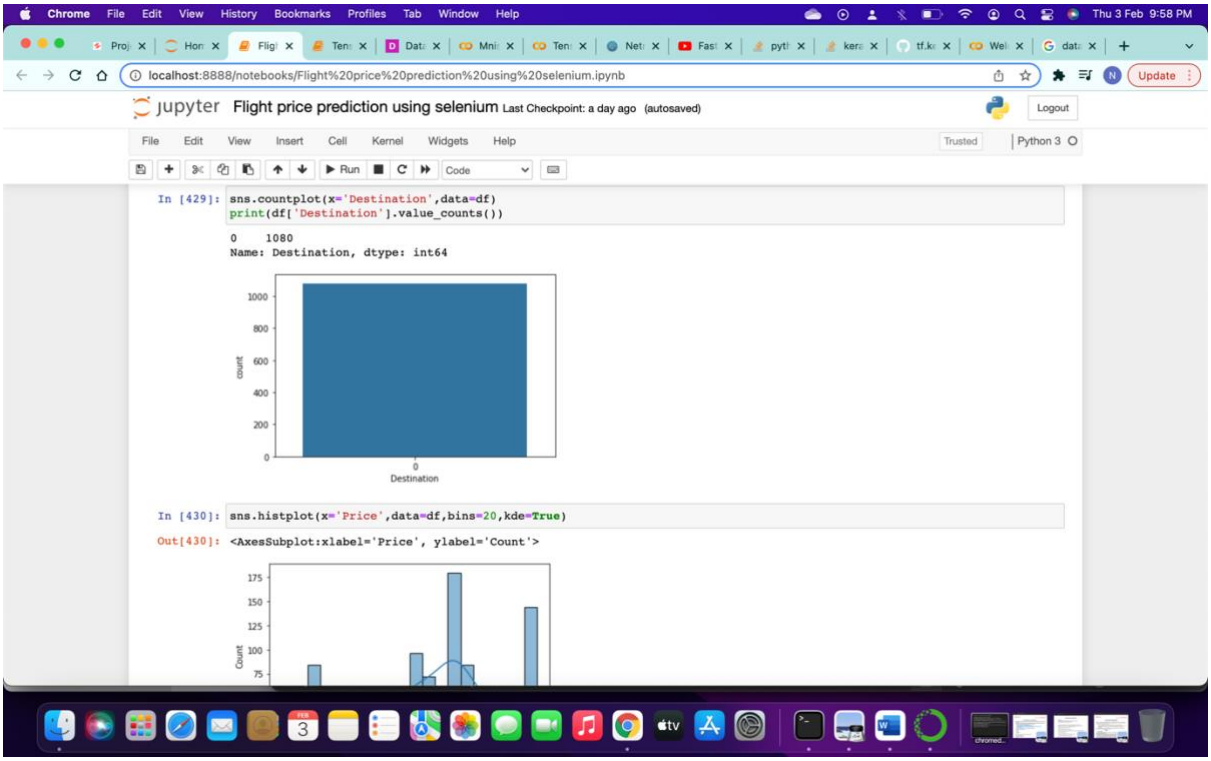
```
0    1080
Name: Source, dtype: int64
```

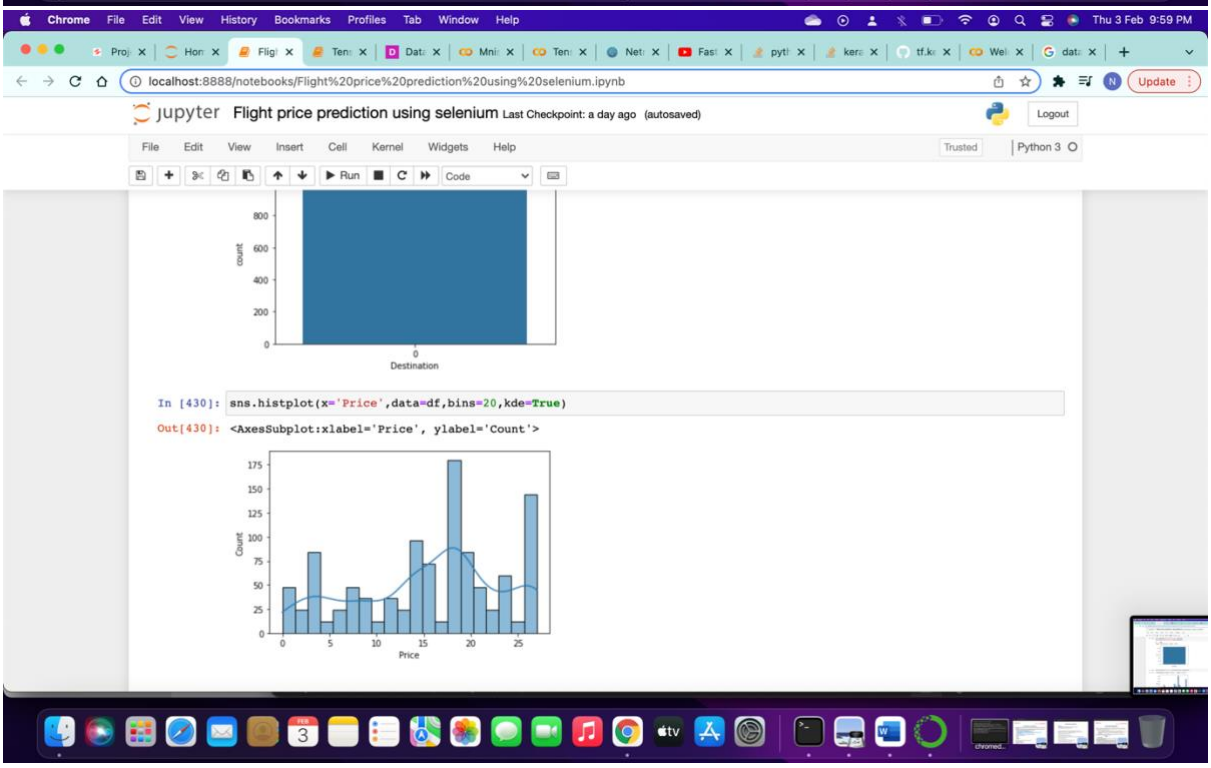


In [429]:
```
sns.countplot(x='Destination',data=df)
print(df['Destination'].value_counts())
```

```
0    1080
Name: Destination, dtype: int64
```







CONCLUSION

- Key Findings and Conclusions of the Study

Duration of the flights increases, customer decreases like for an example if a passenger wants to move from Mumbai to Delhi and if it will take 15-16 hours then very few passenger would like to travel from that flight.

- Learning Outcomes of the Study in respect of Data Science

  Scrapping of data is the challenge faced. Decision tree regressor works best.

- Limitations of this work and Scope for Future Work

  Increase more number of columns and rows to get better accuracy.