

Desktop/project/ x app.py - Jupyter Text Editor x flight\_price - Jupyter Notebook x Flight Price Prediction x +

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [6]: train\_data.head()

Out[6]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302

In [7]: train\_data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 11 columns):
 #   column              non-null count  dtype
---  --
 0   Airline             10683 non-null object
 1   Date_of_Journey     10683 non-null object
 2   Source              10683 non-null object
 3   Destination         10683 non-null object
 4   Route              10682 non-null object
 5   Dep_Time            10683 non-null object
 6   Arrival_Time        10683 non-null object
 7   Duration            10683 non-null object
 8   Total_Stops         10682 non-null object
 9   Additional_Info     10683 non-null object
10   Price              10683 non-null int64
dtypes: int64(1), object(10)
memory usage: 918.2+ KB
```

In [8]: train\_data["Duration"].value\_counts()

27°C Mostly cloudy

Desktop/project/ x app.py - Jupyter Text Editor x flight\_price - Jupyter Notebook x Flight Price Prediction x +

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Out[8]:

Duration	count
2h 50m	550
1h 30m	386
2h 45m	337
2h 55m	337
2h 35m	329
...	
31h 30m	1
30h 25m	1
42h 5m	1
4h 10m	1
47h 40m	1

Name: Duration, Length: 368, dtype: int64

In [9]: train\_data.dropna(inplace = True)

In [10]: train\_data.isnull().sum()

Out[10]:

column	count
Airline	0
Date_of_Journey	0
Source	0
Destination	0
Route	0
Dep_Time	0
Arrival_Time	0
Duration	0
Total_Stops	0
Additional_Info	0
Price	0

dtype: int64

In [11]: train\_data["Journey\_day"] = pd.to\_datetime(train\_data.Date\_of\_Journey, format="%d/%m/%Y").dt.day

27°C Mostly cloudy

Desktop/project/ x app.py - Jupyter Text Editor x flight\_price - Jupyter Notebook x Flight Price Prediction x +

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [16]: train\_data.head()

Out[16]:

	Airline	Source	Destination	Route	Arrival_Time	Duration	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min
0	IndiGo	Banglore	New Delhi	BLR → DEL	01:10 22 Mar	2h 50m	non-stop	No info	3897	24	3	22	20
1	Air India	Kolkata	Banglore	CCU → IXR → BBI → BLR	13:15	7h 25m	2 stops	No info	7662	1	5	5	50
2	Jet Airways	Delhi	Cochin	DEL → LKO → BOM → COK	04:25 10 Jun	19h	2 stops	No info	13882	9	6	9	25
3	IndiGo	Kolkata	Banglore	CCU → NAG → BLR	23:30	5h 25m	1 stop	No info	6218	12	5	18	5
4	IndiGo	Banglore	New Delhi	BLR → NAG → DEL	21:35	4h 45m	1 stop	No info	13302	1	3	16	50

In [17]: train\_data["Arrival\_hour"] = pd.to\_datetime(train\_data.Arrival\_Time).dt.hour

```
# Extracting Minutes
train_data["Arrival_min"] = pd.to_datetime(train_data.Arrival_Time).dt.minute
# Now we can drop Arrival_Time as it is of no use
train_data.drop(["Arrival_Time"], axis = 1, inplace = True)
```

In [18]: train\_data.head()

Out[18]:

	Airline	Source	Destination	Route	Duration	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min
0	IndiGo	Banglore	New Delhi	BLR → DEL	2h 50m	non-stop	No info	3897	24	3	22	20	1	10

27°C Mostly cloudy

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (pykernel)

```
In [13]: train_data.head()
```

```
Out[13]:
```

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price	Journey_day	Journey_m
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897	24	
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662	1	
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882	9	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218	12	
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302	1	

```
In [14]: train_data.drop(["Date_of_Journey"], axis = 1, inplace = True)
```

```
In [15]: train_data["Dep_hour"] = pd.to_datetime(train_data["Dep_Time"]).dt.hour
```

27°C Mostly cloudy

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (pykernel)

```
In [20]: train_data["Duration_hours"] = duration_hours
train_data["Duration_mins"] = duration_mins
train_data.drop(["Duration"], axis = 1, inplace = True)
train_data.head()
```

```
Out[20]:
```

	Airline	Source	Destination	Route	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min
0	IndiGo	Banglore	New Delhi	BLR → DEL	non-stop	No info	3897	24	3	22	20	1	10
1	Air India	Kolkata	Banglore	CCU → IXR → BBI → BLR	2 stops	No info	7662	1	5	5	50	13	15
2	Jet Airways	Delhi	Cochin	DEL → LKO → BOM → COK	2 stops	No info	13882	9	6	9	25	4	25
3	IndiGo	Kolkata	Banglore	CCU → NAG → BLR	1 stop	No info	6218	12	5	18	5	23	30
4	IndiGo	Banglore	New Delhi	BLR → NAG → DEL	1 stop	No info	13302	1	3	16	50	21	35

```
In [21]: train_data["Airline"].value_counts()
```

27°C Mostly cloudy

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (pykernel)

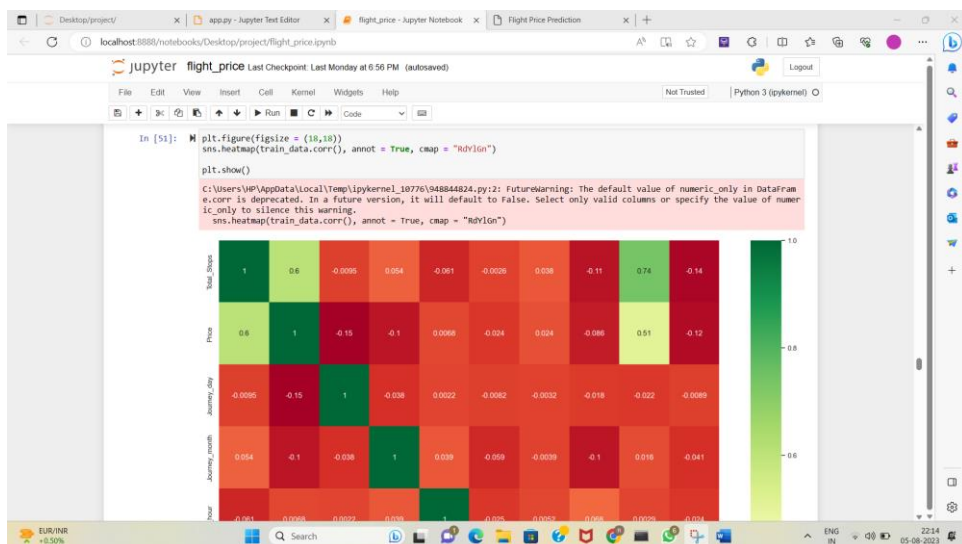
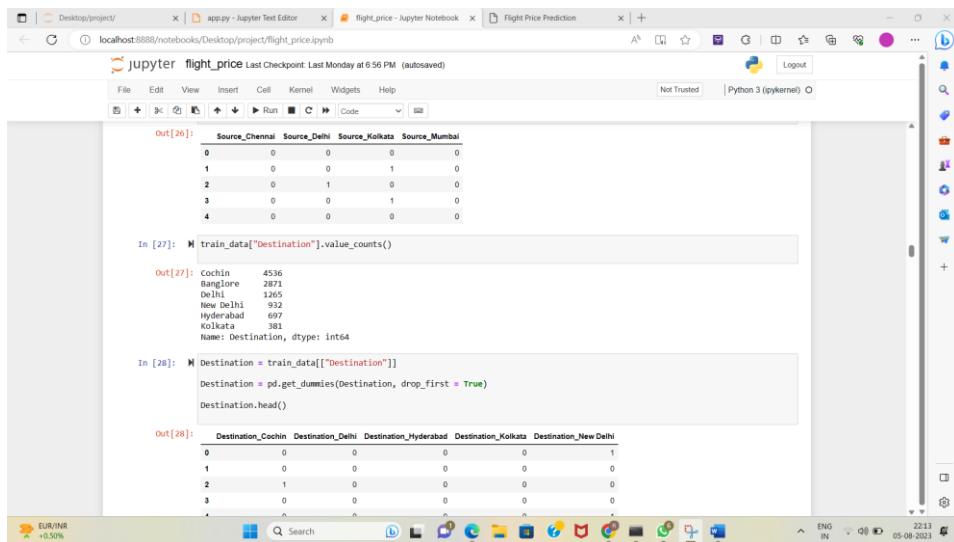
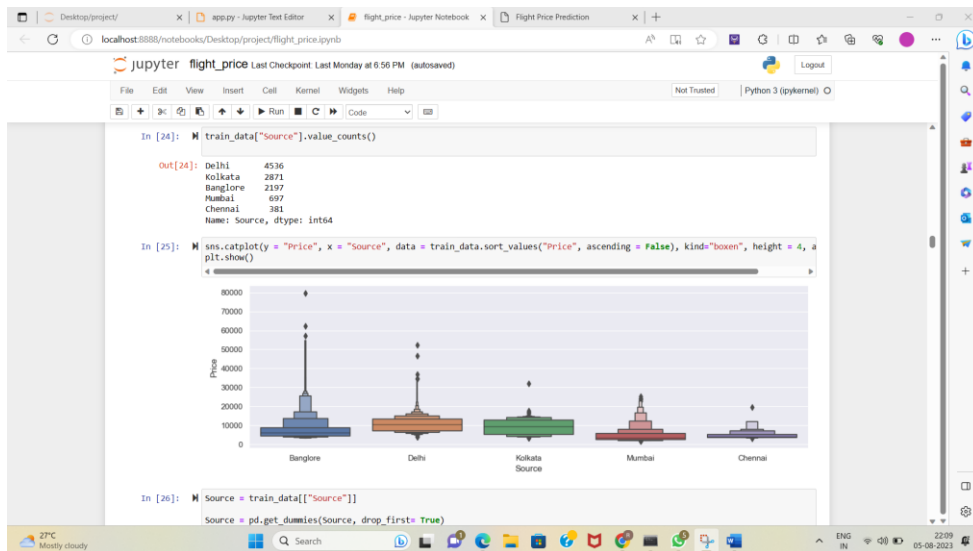
```
In [22]: sns.catplot(y = "Price", x = "Airline", data = train_data.sort_values("Price", ascending = False), kind="boxen", height = 6, plt.show())
```

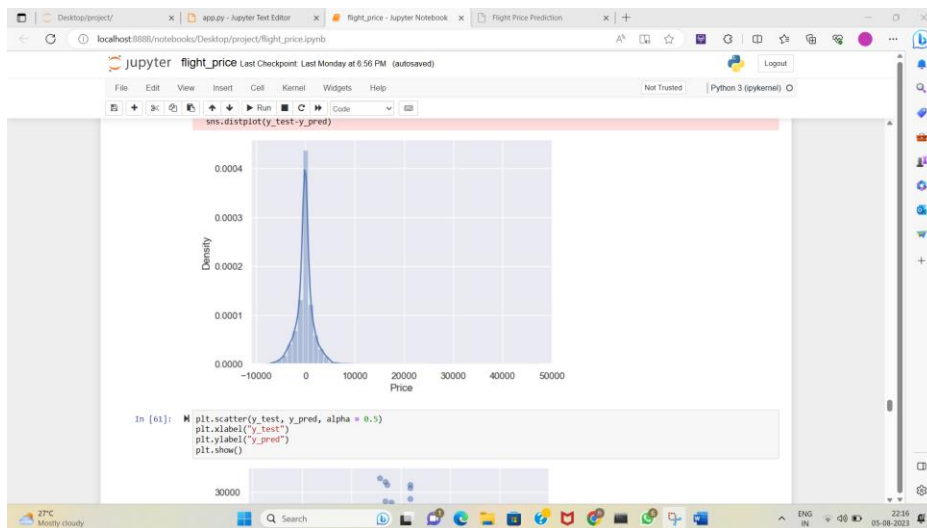
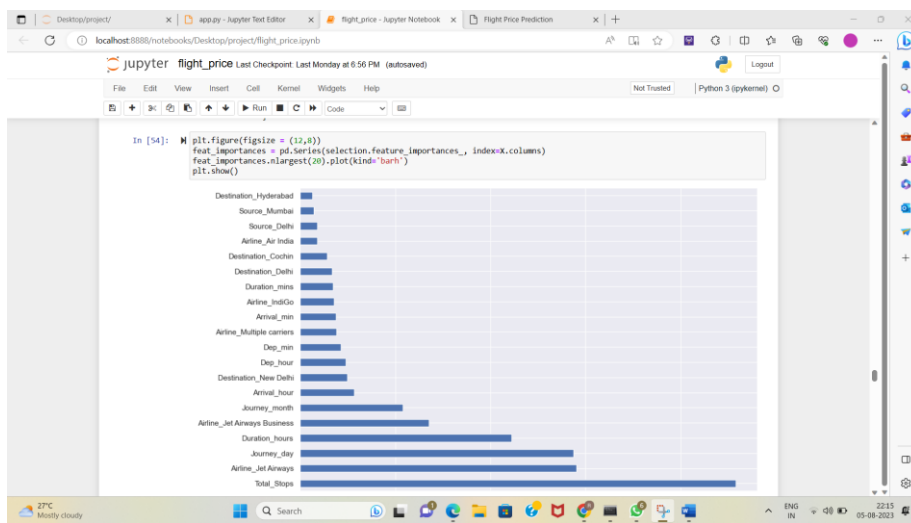
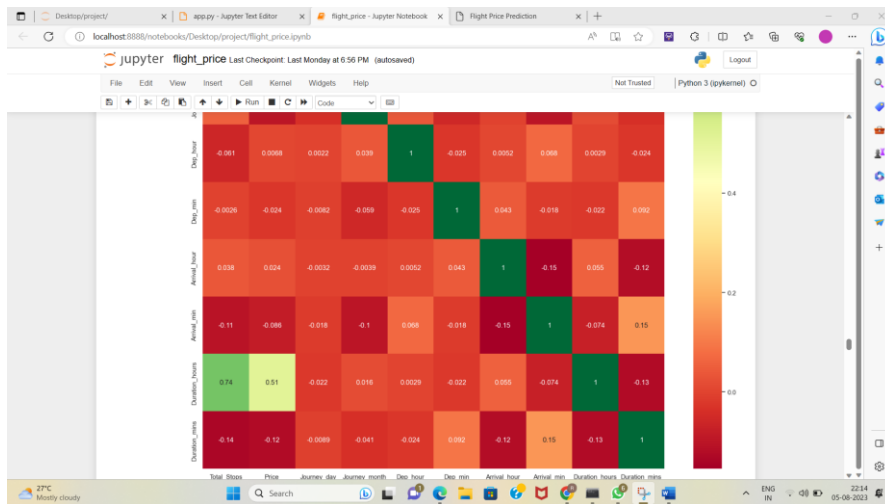
```
In [23]: Airline = train_data["Airline"]
Airline = pd.get_dummies(Airline, drop_first = True)
Airline.head()
```

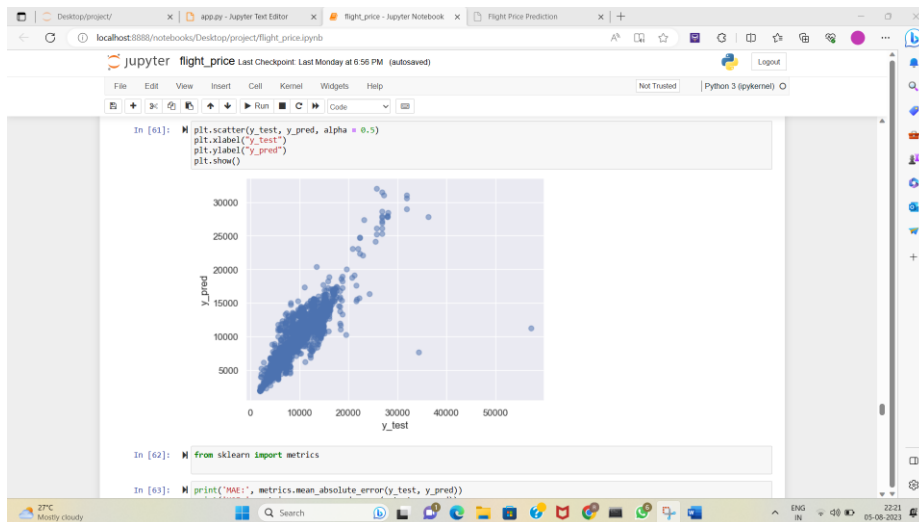
```
Out[23]:
```

	Airline_Air India	Airline_GoAir	Airline_IndiGo	Airline_Jet Airways	Airline_Jet Airways Business	Airline_Multiple carriers	Airline_Multiple carriers Premium economy	Airline_SpiceJet	Airline_Trujet	Airline_Vistara	Airline_Unknown
0	0	0	1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0

27°C Mostly cloudy







Desktop/project/ x app.py - Jupyter Text Editor x flight\_price - Jupyter Notebook x Flight Price Prediction x +

localhost:8888/notebooks/Desktop/project/flight\_price.ipynb

jupyter flight\_price Last Checkpoint: Last Monday at 6:56 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [76]: print('MAE:', metrics.mean_absolute_error(y_test, prediction))
print('MSE:', metrics.mean_squared_error(y_test, prediction))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, prediction)))

MAE: 1164.5571751510895
MSE: 4857291.2196157915
RMSE: 2204.2718832411358

In [86]: import pickle
# open a file, where you want to store the data
file = open('flight_rf.pkl', 'wb')
pickle.dump(reg_rf, file)

In [87]: model = open('flight_rf.pkl', 'rb')
forest = pickle.load(model)

In [88]: y_prediction = forest.predict(x_test)

In [89]: metrics.r2_score(y_test, y_prediction)

Out[89]: 0.7981365267969486

In [ ]:

In [ ]:
```

27°C Mostly cloudy 22:22 05-08-2023