flight-price-prediction

March 21, 2025

Importing Libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
from sklearn.model_selection import train_test_split
```

Loading the data

[5]:
```
df=pd.read_excel(r"C:\Users\DELL\Downloads\archive (5)\Data_Train.xlsx")
```

[6]:
```
df.head()
```

[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

clean data

[7]:
```
df.isnull().sum()
```

[7]:
```
Airline            0
Date_of_Journey    0
Source             0
Destination        0
Route              1
```

```
Dep_Time           0
Arrival_Time       0
Duration           0
Total_Stops        1
Additional_Info    0
Price              0
dtype: int64
```

data preprocessing

[21]:
```
#encoding Airline
a_l=LabelEncoder()
df["Airline encode"]=a_l.fit_transform(df["Airline"])
df.head()
```

[21]:

	Airline	Date_of_Journey	Source	Destination	Route	\
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	

	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price	\
0	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897	
1	05:50	13:15	7h 25m	2 stops	No info	7662	
2	09:25	04:25 10 Jun	19h	2 stops	No info	13882	
3	18:05	23:30	5h 25m	1 stop	No info	6218	
4	16:50	21:35	4h 45m	1 stop	No info	13302	

	Airline encode	Source encode	Destination encode
0	3	0	5
1	1	3	0
2	4	2	1
3	3	3	0
4	3	0	5

[22]:
```
#encoding Source
sou=LabelEncoder()
df["Source encode"]=sou.fit_transform(df["Source"])
df.head()
```

[22]:

	Airline	Date_of_Journey	Source	Destination	Route	\
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	

```
  Dep_Time  Arrival_Time Duration Total_Stops Additional_Info  Price  \
0    22:20  01:10 22 Mar   2h 50m    non-stop         No info   3897
1    05:50         13:15   7h 25m     2 stops         No info   7662
2    09:25  04:25 10 Jun      19h     2 stops         No info  13882
3    18:05         23:30   5h 25m      1 stop         No info   6218
4    16:50         21:35   4h 45m      1 stop         No info  13302

   Airline encode  Source encode  Destination encode
0               3              0                   5
1               1              3                   0
2               4              2                   1
3               3              3                   0
4               3              0                   5
```

[23]:
```
#encoding Destination
des=LabelEncoder()
df["Destination encode"]=des.fit_transform(df["Destination"])
df.head()
```

[23]:
```
       Airline Date_of_Journey    Source Destination                  Route  \
0       IndiGo      24/03/2019  Banglore   New Delhi              BLR → DEL
1    Air India       1/05/2019   Kolkata    Banglore  CCU → IXR → BBI → BLR
2  Jet Airways       9/06/2019     Delhi      Cochin  DEL → LKO → BOM → COK
3       IndiGo      12/05/2019   Kolkata    Banglore        CCU → NAG → BLR
4       IndiGo      01/03/2019  Banglore   New Delhi        BLR → NAG → DEL

  Dep_Time  Arrival_Time Duration Total_Stops Additional_Info  Price  \
0    22:20  01:10 22 Mar   2h 50m    non-stop         No info   3897
1    05:50         13:15   7h 25m     2 stops         No info   7662
2    09:25  04:25 10 Jun      19h     2 stops         No info  13882
3    18:05         23:30   5h 25m      1 stop         No info   6218
4    16:50         21:35   4h 45m      1 stop         No info  13302

   Airline encode  Source encode  Destination encode
0               3              0                   5
1               1              3                   0
2               4              2                   1
3               3              3                   0
4               3              0                   5
```

split-ind,dep

[24]:
```
X=df[["Airline encode","Source encode","Destination encode"]]
Y=df["Price"]
```

split-train and test

[25]:
```
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.
↪2,random_state=42)
```

create and train

[26]:
```
price_model=LinearRegression()
price_model.fit(X_train,Y_train)
LinearRegression()
```

[26]:
```
LinearRegression()
```

Test

[27]:
```
Air=input("Enter the Airline")
Source=input("Enter the Source")
Destination=input("Enter the Destination")
Price=float(input("Enter the Price"))
```

```
Enter the Airline Air India
Enter the Source Kolkata
Enter the Destination Banglore
Enter the Price 7662
```

[28]:
```
Air_enc=a_l.transform([Air])[0]
Sou_enc=sou.transform([Source])[0]
Des_enc=des.transform([Destination])[0]
print(Air_enc,Sou_enc,Des_enc)
```

```
1 3 0
```

[29]:
```
result=price_model.predict([[Air_enc,Sou_enc,Des_enc]])
print("The predicted price is : ",result[0])
```

```
The predicted price is :  9635.136999310234
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:439: UserWarning: X
does not have valid feature names, but LinearRegression was fitted with feature
names
  warnings.warn(
```

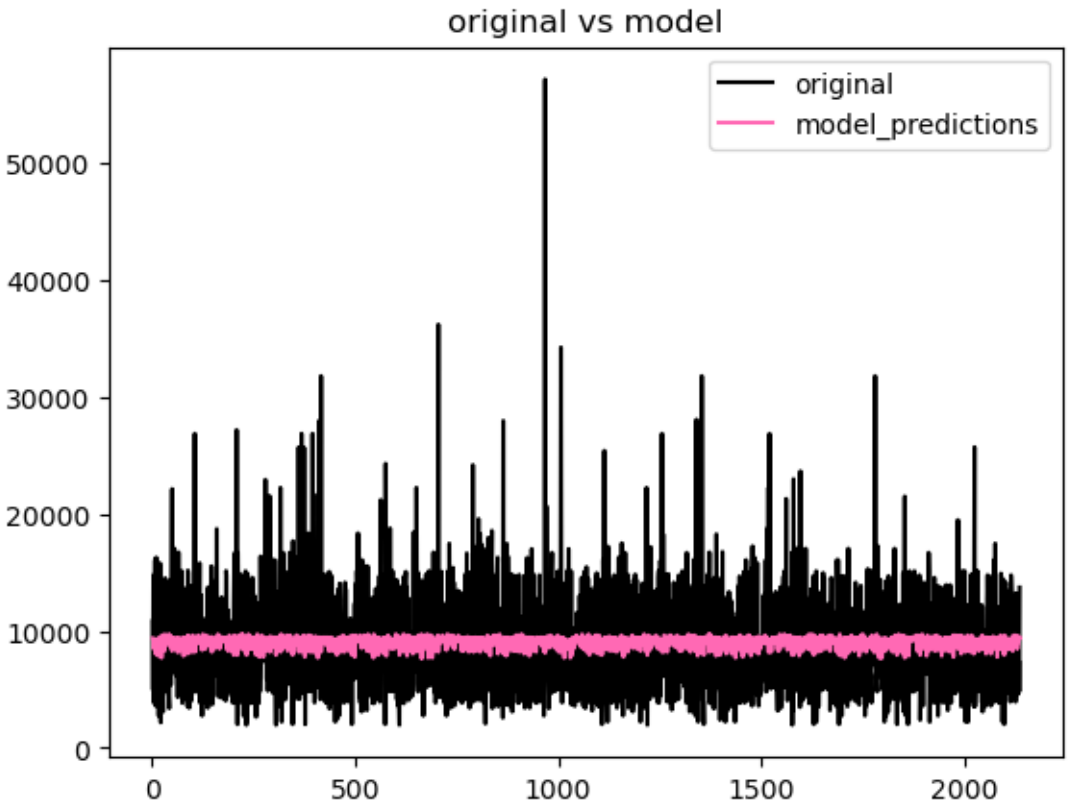
Evaluation:

1.predict test values 2.visualize 3.metrics

[30]:
```
model_predictions=price_model.predict(X_test)
len(Y_test)
```

[30]:
```
2137
```

[32]:
```
#plotting original values
plt.plot(np.arange(1,2138),Y_test,color="k",label="original")
plt.plot(np.
  ↪arange(1,2138),model_predictions,color="hotpink",label="model_predictions")
plt.title("original vs model")
plt.legend()
plt.show()
```



[33]:
```
r2score=r2_score(Y_test,model_predictions)
print(r2score)
```

```
-8.215607105532818e-05
```

MSE

[ ]:
```
mse=mean_squared_error(Y_test,model_predictions)
p
```