Flight Price Dataset

```
import pandas as pd
import numpy as np

df = pd.read_excel(r"C:\Users\DELL\Downloads\flight_price.xlsx")
df
```

```
           Airline Date_of_Journey    Source Destination  \
0           IndiGo      24/03/2019  Banglore   New Delhi
1        Air India       1/05/2019   Kolkata    Banglore
2      Jet Airways       9/06/2019     Delhi      Cochin
3           IndiGo      12/05/2019   Kolkata    Banglore
4           IndiGo      01/03/2019  Banglore   New Delhi
...            ...             ...       ...         ...
10678     Air Asia       9/04/2019   Kolkata    Banglore
10679    Air India      27/04/2019   Kolkata    Banglore
10680  Jet Airways      27/04/2019  Banglore       Delhi
10681      Vistara      01/03/2019  Banglore   New Delhi
10682    Air India       9/05/2019     Delhi      Cochin

                        Route Dep_Time  Arrival_Time Duration
Total_Stops  \
0                   BLR → DEL    22:20  01:10 22 Mar   2h 50m    non-
stop
1      CCU → IXR → BBI → BLR    05:50         13:15   7h 25m     2
stops
2      DEL → LKO → BOM → COK    09:25  04:25 10 Jun      19h     2
stops
3            CCU → NAG → BLR    18:05         23:30   5h 25m      1
stop
4            BLR → NAG → DEL    16:50         21:35   4h 45m      1
stop
...                      ...      ...           ...      ...         .
..
10678             CCU → BLR    19:55         22:25   2h 30m    non-
stop
10679             CCU → BLR    20:45         23:20   2h 35m    non-
stop
10680             BLR → DEL    08:20         11:20       3h    non-
stop
10681             BLR → DEL    11:30         14:10   2h 40m    non-
stop
10682  DEL → GOI → BOM → COK    10:55         19:15   8h 20m     2
stops

      Additional_Info  Price
0             No info   3897
1             No info   7662
```

```
2              No info  13882
3              No info   6218
4              No info  13302
...                ...    ...
10678          No info   4107
10679          No info   4145
10680          No info   7229
10681          No info  12648
10682          No info  11753

[10683 rows x 11 columns]
```

df.tail()

```
           Airline Date_of_Journey    Source Destination  \
10678     Air Asia       9/04/2019   Kolkata    Banglore
10679    Air India      27/04/2019   Kolkata    Banglore
10680  Jet Airways      27/04/2019  Banglore       Delhi
10681      Vistara      01/03/2019  Banglore   New Delhi
10682    Air India       9/05/2019     Delhi      Cochin

                        Route Dep_Time Arrival_Time Duration
Total_Stops  \
10678               CCU → BLR    19:55        22:25   2h 30m    non-
stop
10679               CCU → BLR    20:45        23:20   2h 35m    non-
stop
10680               BLR → DEL    08:20        11:20       3h    non-
stop
10681               BLR → DEL    11:30        14:10   2h 40m    non-
stop
10682  DEL → GOI → BOM → COK    10:55        19:15   8h 20m     2
stops

      Additional_Info  Price
10678          No info   4107
10679          No info   4145
10680          No info   7229
10681          No info  12648
10682          No info  11753
```

df.head()

```
       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

df.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

df.describe()

              Price
count  10683.000000
mean    9087.064121
std     4611.359167
min     1759.000000
25%     5277.000000
50%     8372.000000
75%    12373.000000
max    79512.000000

df.head(3)

       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

  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
```

```
## Feature Engineering
df['Date']=df['Date_of_Journey'].str.split('/').str[0]
df['Month']=df['Date_of_Journey'].str.split('/').str[1]
df['Year']=df['Date_of_Journey'].str.split('/').str[2]

df
```

```
           Airline Date_of_Journey    Source Destination  \
0           IndiGo      24/03/2019  Banglore   New Delhi
1        Air India       1/05/2019   Kolkata    Banglore
2      Jet Airways       9/06/2019     Delhi      Cochin
3           IndiGo      12/05/2019   Kolkata    Banglore
4           IndiGo      01/03/2019  Banglore   New Delhi
...            ...             ...       ...         ...
10678     Air Asia       9/04/2019   Kolkata    Banglore
10679    Air India      27/04/2019   Kolkata    Banglore
10680  Jet Airways      27/04/2019  Banglore       Delhi
10681      Vistara      01/03/2019  Banglore   New Delhi
10682    Air India       9/05/2019     Delhi      Cochin

                        Route Dep_Time  Arrival_Time Duration
Total_Stops  \
0                   BLR → DEL    22:20  01:10 22 Mar   2h 50m    non-
stop
1      CCU → IXR → BBI → BLR    05:50         13:15   7h 25m     2
stops
2      DEL → LKO → BOM → COK    09:25  04:25 10 Jun      19h     2
stops
3            CCU → NAG → BLR    18:05         23:30   5h 25m      1
stop
4            BLR → NAG → DEL    16:50         21:35   4h 45m      1
stop
...                      ...      ...           ...      ...         .
..
10678             CCU → BLR    19:55         22:25   2h 30m    non-
stop
10679             CCU → BLR    20:45         23:20   2h 35m    non-
stop
10680             BLR → DEL    08:20         11:20       3h    non-
```

```
stop
10681              BLR → DEL    11:30          14:10   2h 40m    non-
stop
10682  DEL → GOI → BOM → COK    10:55          19:15   8h 20m     2
stops

      Additional_Info  Price Date Month  Year
0             No info   3897   24    03  2019
1             No info   7662    1    05  2019
2             No info  13882    9    06  2019
3             No info   6218   12    05  2019
4             No info  13302   01    03  2019
...               ...    ...  ...   ...   ...
10678         No info   4107    9    04  2019
10679         No info   4145   27    04  2019
10680         No info   7229   27    04  2019
10681         No info  12648   01    03  2019
10682         No info  11753    9    05  2019

[10683 rows x 14 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 14 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 
 11  Date             10683 non-null  object
 12  Month            10683 non-null  object
 13  Year             10683 non-null  object
dtypes: int64(1), object(13)
memory usage: 1.1+ MB
```

```
df['Date'] = df['Date'].astype(int)
df['Month'] = df['Month'].astype(int)
df['Year'] = df['Year'].astype(int)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 14 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 
 11  Date             10683 non-null  int32 
 12  Month            10683 non-null  int32 
 13  Year             10683 non-null  int32 
dtypes: int32(3), int64(1), object(10)
memory usage: 1.0+ MB
```

```
df.drop("Date_of_Journey",axis=1,inplace=True)
```

```
df.head()
```

```
       Airline    Source Destination                  Route  \
Dep_Time  \
0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20

1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50

2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25

3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05

4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50


   Arrival_Time Duration Total_Stops Additional_Info  Price  Date  
Month  Year
0  01:10 22 Mar   2h 50m    non-stop         No info   3897    24  
3  2019
1         13:15   7h 25m     2 stops         No info   7662     1  
5  2019
2  04:25 10 Jun      19h     2 stops         No info  13882     9  
6  2019
3         23:30   5h 25m      1 stop         No info   6218    12  
5  2019
4         21:35   4h 45m      1 stop         No info  13302     1  
3  2019
```

```
df['departure_hour']=df['Dep_Time'].str.split(':').str[0]
df['departure_min']=df['Dep_Time'].str.split(':').str[1]

df.head()
```

```
       Airline    Source Destination                  Route
Dep_Time  \
0       IndiGo  Banglore   New Delhi              BLR → DEL    22:20

1    Air India   Kolkata    Banglore  CCU → IXR → BBI → BLR    05:50

2  Jet Airways     Delhi      Cochin  DEL → LKO → BOM → COK    09:25

3       IndiGo   Kolkata    Banglore        CCU → NAG → BLR    18:05

4       IndiGo  Banglore   New Delhi        BLR → NAG → DEL    16:50


   Arrival_Time Duration Total_Stops Additional_Info  Price  Date
Month  \
0  01:10 22 Mar   2h 50m    non-stop         No info   3897    24
3
1         13:15   7h 25m     2 stops         No info   7662     1
5
2  04:25 10 Jun      19h     2 stops         No info  13882     9
6
3         23:30   5h 25m      1 stop         No info   6218    12
5
4         21:35   4h 45m      1 stop         No info  13302     1
3

   Year departure_hour departure_min
0  2019             22            20
1  2019             05            50
2  2019             09            25
3  2019             18            05
4  2019             16            50
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Airline          10683 non-null  object
 1   Source           10683 non-null  object
 2   Destination      10683 non-null  object
 3   Route            10682 non-null  object
 4   Dep_Time         10683 non-null  object
 5   Arrival_Time     10683 non-null  object
```

```
 6   Duration         10683 non-null  object
 7   Total_Stops      10682 non-null  object
 8   Additional_Info  10683 non-null  object
 9   Price            10683 non-null  int64
 10  Date             10683 non-null  int32
 11  Month            10683 non-null  int32
 12  Year             10683 non-null  int32
 13  departure_hour   10683 non-null  object
 14  departure_min    10683 non-null  object
dtypes: int32(3), int64(1), object(11)
memory usage: 1.1+ MB
```

```
df['departure_hour']=df['departure_hour'].astype(int)
df['departure_min']=df['departure_min'].astype(int)

df.drop("Dep_Time",axis=1,inplace=True)

df.drop('Route',axis=1,inplace=True)

df.head(2)
```

```
     Airline    Source Destination  Arrival_Time Duration Total_Stops
\
0     IndiGo  Banglore   New Delhi  01:10 22 Mar   2h 50m    non-stop

1  Air India   Kolkata    Banglore         13:15   7h 25m     2 stops


  Additional_Info  Price  Date  Month  Year  departure_hour
departure_min
0         No info   3897    24      3  2019              22
20
1         No info   7662     1      5  2019               5
50
```

```
df["Arrival_Time"] = df["Arrival_Time"].apply(lambda x:x.split(" ")
[0])

df['arrival_hour']=df['Arrival_Time'].str.split(':').str[0]
df['arrival_min']=df['Arrival_Time'].str.split(':').str[1]

df['arrival_hour']=df['arrival_hour'].astype(int)
df['arrival_min']=df['arrival_min'].astype(int)

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Airline          10683 non-null  object
```

```
 1   Source           10683 non-null  object
 2   Destination      10683 non-null  object
 3   Arrival_Time     10683 non-null  object
 4   Duration         10683 non-null  object
 5   Total_Stops      10682 non-null  object
 6   Additional_Info  10683 non-null  object
 7   Price            10683 non-null  int64
 8   Date             10683 non-null  int32
 9   Month            10683 non-null  int32
 10  Year             10683 non-null  int32
 11  departure_hour   10683 non-null  int32
 12  departure_min    10683 non-null  int32
 13  arrival_hour     10683 non-null  int32
 14  arrival_min      10683 non-null  int32
dtypes: int32(7), int64(1), object(7)
memory usage: 959.9+ KB
```

```
df.drop("Arrival_Time",axis=1,inplace=True)
```

```
df.head(3)
```

```
       Airline    Source Destination Duration Total_Stops
Additional_Info  \
0       IndiGo  Banglore   New Delhi   2h 50m    non-stop         No
info
1    Air India   Kolkata    Banglore   7h 25m     2 stops         No
info
2  Jet Airways     Delhi      Cochin      19h     2 stops         No
info

   Price  Date  Month  Year  departure_hour  departure_min
arrival_hour  \
0   3897    24      3  2019              22             20
1
1   7662     1      5  2019               5             50
13
2  13882     9      6  2019               9             25
4

   arrival_min
0           10
1           15
2           25
```

```
df[df["Total_Stops"].isnull()]
```

```
        Airline Source Destination Duration Total_Stops
Additional_Info  \
9039  Air India  Delhi      Cochin  23h 40m         NaN         No
info
```

```
      Price  Date  Month  Year  departure_hour  departure_min
arrival_hour  \
9039   7480     6      5  2019               9             45
9

      arrival_min
9039           25
```

```
df["Total_Stops"].mode()
```

```
0    1 stop
Name: Total_Stops, dtype: object
```

```
df["Total_Stops"] =df["Total_Stops"].map({'non-stop':0,'1 stop':1,'2
stops':2,'3 stops':3,'4 stops':4,np.nan:1})
```

```
df.head()
```

```
       Airline    Source Destination Duration  Total_Stops
Additional_Info  \
0       IndiGo  Banglore   New Delhi   2h 50m            0         No
info
1    Air India   Kolkata    Banglore   7h 25m            2         No
info
2  Jet Airways     Delhi      Cochin      19h            2         No
info
3       IndiGo   Kolkata    Banglore   5h 25m            1         No
info
4       IndiGo  Banglore   New Delhi   4h 45m            1         No
info

   Price  Date  Month  Year  departure_hour  departure_min
arrival_hour  \
0   3897    24      3  2019              22             20
1
1   7662     1      5  2019               5             50
13
2  13882     9      6  2019               9             25
4
3   6218    12      5  2019              18              5
23
4  13302     1      3  2019              16             50
21

   arrival_min
0           10
1           15
2           25
3           30
4           35
```

```
df['duration_hour']=df['Duration'].str.split('
').str[0].str.split("h").str[0]
df['duration_min']=df['Duration'].str.split('
').str[1].str.split("m").str[1]

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Airline          10683 non-null  object
 1   Source           10683 non-null  object
 2   Destination      10683 non-null  object
 3   Duration         10683 non-null  object
 4   Total_Stops      10683 non-null  int64 
 5   Additional_Info  10683 non-null  object
 6   Price            10683 non-null  int64 
 7   Date             10683 non-null  int32 
 8   Month            10683 non-null  int32 
 9   Year             10683 non-null  int32 
 10  departure_hour   10683 non-null  int32 
 11  departure_min    10683 non-null  int32 
 12  arrival_hour     10683 non-null  int32 
 13  arrival_min      10683 non-null  int32 
 14  duration_hour    10683 non-null  object
 15  duration_min     9651 non-null   object
dtypes: int32(7), int64(2), object(7)
memory usage: 1.0+ MB
```

```
df[~df['duration_hour'].str.isnumeric()]
```

```
        Airline  Source Destination Duration  Total_Stops
Additional_Info  \
6474  Air India  Mumbai   Hyderabad       5m            2         No
info

      Price  Date  Month  Year  departure_hour  departure_min
arrival_hour  \
6474  17327     6      3  2019              16             50
16

      arrival_min duration_hour duration_min
6474           55            5m          NaN
```

```
df['duration_min'] = pd.to_numeric(df['duration_min'],
errors='coerce').fillna(0).astype(int)
```

```
df['duration_hour'] = pd.to_numeric(df['duration_hour'],
errors='coerce').fillna(0).astype(int)

df['duration_hour']=df['duration_hour'].astype(int)
df['duration_min']=df['duration_min'].astype(int)

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Airline          10683 non-null  object
 1   Source           10683 non-null  object
 2   Destination      10683 non-null  object
 3   Duration         10683 non-null  object
 4   Total_Stops      10683 non-null  int64
 5   Additional_Info  10683 non-null  object
 6   Price            10683 non-null  int64
 7   Date             10683 non-null  int32
 8   Month            10683 non-null  int32
 9   Year             10683 non-null  int32
 10  departure_hour   10683 non-null  int32
 11  departure_min    10683 non-null  int32
 12  arrival_hour     10683 non-null  int32
 13  arrival_min      10683 non-null  int32
 14  duration_hour    10683 non-null  int32
 15  duration_min     10683 non-null  int32
dtypes: int32(9), int64(2), object(5)
memory usage: 959.9+ KB
```

```
# Check column names
print(df.columns)
# Strip leading/trailing spaces from column names
df.columns = df.columns.str.strip()



# Drop the column
df.drop("Duration", axis=1, inplace=True)
```

```
Index(['Airline', 'Source', 'Destination', 'Duration', 'Total_Stops',
       'Additional_Info', 'Price', 'Date', 'Month', 'Year',
'departure_hour',
       'departure_min', 'arrival_hour', 'arrival_min',
'duration_hour',
       'duration_min'],
      dtype='object')
```

```
df.head(3)
```

```
       Airline    Source Destination  Total_Stops Additional_Info
Price  \
0       IndiGo  Banglore   New Delhi            0         No info
3897
1    Air India   Kolkata    Banglore            2         No info
7662
2  Jet Airways     Delhi      Cochin            2         No info
13882

   Date  Month  Year  departure_hour  departure_min  arrival_hour  \
0    24      3  2019              22             20             1
1     1      5  2019               5             50            13
2     9      6  2019               9             25             4

   arrival_min  duration_hour  duration_min
0           10              2             0
1           15              7             0
2           25             19             0
```

```
df['Airline'].unique()
```

```
array(['IndiGo', 'Air India', 'Jet Airways', 'SpiceJet',
       'Multiple carriers', 'GoAir', 'Vistara', 'Air Asia',
       'Vistara Premium economy', 'Jet Airways Business',
       'Multiple carriers Premium economy', 'Trujet'], dtype=object)
```

```
df["Additional_Info"].unique()
```

```
array(['No info', 'In-flight meal not included',
       'No check-in baggage included', '1 Short layover', 'No Info',
       '1 Long layover', 'Change airports', 'Business class',
       'Red-eye flight', '2 Long layover'], dtype=object)
```

```
df["Source"].unique()
```

```
array(['Banglore', 'Kolkata', 'Delhi', 'Chennai', 'Mumbai'],
dtype=object)
```

```
df["Destination"].unique()
```

```
array(['New Delhi', 'Banglore', 'Cochin', 'Kolkata', 'Delhi',
'Hyderabad'],
      dtype=object)
```

```
from sklearn.preprocessing import OneHotEncoder
```

```
encoder = OneHotEncoder()
```

```
encoder.fit_transform(df[["Airline","Additional_Info","Source","Destin
ation"]]).toarray()
```

```
array([[0., 0., 0., ..., 0., 0., 1.],
       [0., 1., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 1.],
       [0., 1., 0., ..., 0., 0., 0.]])
```

```
encoded_df=pd.DataFrame(encoder.fit_transform(df[['Airline',"Additiona
l_Info",'Source','Destination']]).toarray(),columns=encoder.get_featur
e_names_out())
```

```
df_dropped = df.drop(columns=['Airline', 'Additional_Info', 'Source',
'Destination'])
```

```
# Concatenate the one-hot encoded DataFrame with the remaining columns
df= pd.concat([df_dropped, encoded_df], axis=1)
```

```
df
```

```
       Total_Stops  Price  Date  Month  Year  departure_hour
departure_min  \
0                0   3897    24      3  2019              22
20
1                2   7662     1      5  2019               5
50
2                2  13882     9      6  2019               9
25
3                1   6218    12      5  2019              18
5
4                1  13302     1      3  2019              16
50
...            ...    ...   ...    ...   ...             ...
...
10678            0   4107     9      4  2019              19
55
10679            0   4145    27      4  2019              20
45
10680            0   7229    27      4  2019               8
20
10681            0  12648     1      3  2019              11
30
10682            2  11753     9      5  2019              10
55

       arrival_hour  arrival_min  duration_hour  ...
Source_Chennai  \
0                 1           10              2  ...             0.0

1                13           15              7  ...             0.0
```

```
2                  4           25             19  ...                0.0
3                 23           30              5  ...                0.0
4                 21           35              4  ...                0.0
...              ...          ...            ...  ...                ...
10678             22           25              2  ...                0.0
10679             23           20              2  ...                0.0
10680             11           20              3  ...                0.0
10681             14           10              2  ...                0.0
10682             19           15              8  ...                0.0


       Source_Delhi  Source_Kolkata  Source_Mumbai
Destination_Banglore  \
0                0.0             0.0            0.0
0.0
1                0.0             1.0            0.0
1.0
2                1.0             0.0            0.0
0.0
3                0.0             1.0            0.0
1.0
4                0.0             0.0            0.0
0.0
...              ...             ...            ...                    .
..
10678            0.0             1.0            0.0
1.0
10679            0.0             1.0            0.0
1.0
10680            0.0             0.0            0.0
0.0
10681            0.0             0.0            0.0
0.0
10682            1.0             0.0            0.0
0.0

       Destination_Cochin  Destination_Delhi  Destination_Hyderabad  \
0                     0.0                0.0                    0.0
1                     0.0                0.0                    0.0
2                     1.0                0.0                    0.0
3                     0.0                0.0                    0.0
4                     0.0                0.0                    0.0
```

```
...                   ...                 ...                   ...
10678                 0.0                 0.0                   0.0
10679                 0.0                 0.0                   0.0
10680                 0.0                 1.0                   0.0
10681                 0.0                 0.0                   0.0
10682                 1.0                 0.0                   0.0
```

```
       Destination_Kolkata  Destination_New Delhi
0                      0.0                   1.0
1                      0.0                   0.0
2                      0.0                   0.0
3                      0.0                   0.0
4                      0.0                   1.0
...                    ...                   ...
10678                  0.0                   0.0
10679                  0.0                   0.0
10680                  0.0                   0.0
10681                  0.0                   1.0
10682                  0.0                   0.0

[10683 rows x 44 columns]
```

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error,
r2_score

# Features (X) and Target (y)
X = df.drop(columns=['Price'])  # Assuming 'Price' is the target
column
y = df['Price']

# Split data into training and testing sets (80-20 split)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

print("Training set shape:", X_train.shape)
print("Testing set shape:", X_test.shape)
```

```
Training set shape: (8546, 43)
Testing set shape: (2137, 43)
```

```
# Initialize the model (adjust hyperparameters as needed)
model = RandomForestRegressor(
    n_estimators=100,  # Number of trees
    max_depth=10,      # Maximum depth of trees
    random_state=42
)

# Train the model
model.fit(X_train, y_train)
```

```python
# Predict on the test set
y_pred = model.predict(X_test)

y_pred
```

```
array([10997.33811834,  6924.69671977, 14019.57310912, ...,
        6502.90227531,  4421.01390585, 14057.11320598])
```

```python
# Calculate evaluation metrics
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Absolute Error (MAE):", mae)
print("Mean Squared Error (MSE):", mse)
print("R-squared (R²):", r2)
```

```
Mean Absolute Error (MAE): 920.7899443340879
Mean Squared Error (MSE): 3319622.319251574
R-squared (R²): 0.8431474369918184
```

```python
# Get feature importances
feature_importances = pd.DataFrame({
    'Feature': X.columns,
    'Importance': model.feature_importances_
}).sort_values(by='Importance', ascending=False)

print("Top 5 Important Features:")
print(feature_importances.head(5))
```

```
Top 5 Important Features:
                                          Feature  Importance
8                                  duration_hour    0.456275
1                                           Date    0.086840
27  Additional_Info_In-flight meal not included    0.082442
15                Airline_Jet Airways Business    0.069707
14                         Airline_Jet Airways    0.068296
```

```python
import pickle
```

```python
import pickle
pickle_out = open("model.pkl", "wb")
pickle.dump(model, pickle_out)
pickle_out.close()
```