Untitled2

May 17, 2024

[1]:
```
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
#loading dataset
df_1= pd.read_csv('Bangalore Flight.csv')
```

[3]:
```
df_1.head()
```

[3]:

	Airline	Flight No.	Source	Departure	No. of stops	Arrival	Destination	Ticket Class	Flight Duration (hrs)	Days left	Price
0	AirAsia	I5-996	Bangalore	Evening	0	Night	Delhi	Economy	2.83	1	7489
1	Vistara	UK-820	Bangalore	Evening	0	Night	Delhi	Economy	2.67	1	7489
2	Vistara	UK-802	Bangalore	Evening	0	Night	Delhi	Economy	2.75	1	7489
3	Indigo	6E-6139	Bangalore	Afternoon	0	Evening	Delhi	Economy	2.75	1	7488
4	Indigo	6E-2514	Bangalore	Evening	0	Evening	Delhi	Economy	2.83	1	7488

[6]:
```
df_1.tail()
```

[6]:

	Airline	Flight No.	Source	Departure	No. of stops	Arrival	Destination	Ticket Class	Flight Duration (hrs)	Days left	Price
52056	Air_India	AI-503	Bangalore	Evening	1	Afternoon	Chennai	Business	18.83	49	60396
52057	Air_India	AI-501	Bangalore	Afternoon	1	Morning	Chennai	Business	19.58	49	60396
52058	Air_India	AI-505	Bangalore	Morning	1	Morning					
52059	Air_India	AI-501	Bangalore	Afternoon	1	Afternoon					
52060	Air_India	AI-505	Bangalore	Morning	1	Afternoon					

```
52058     Chennai     Business                 22.50          49  60396
52059     Chennai     Business                 23.33          49  60396
52060     Chennai     Business                 26.25          49  60396
```

[4]:
```
df_1.shape
```

[4]:
```
(52061, 11)
```

[5]:
```
df_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52061 entries, 0 to 52060
Data columns (total 11 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Airline                 52061 non-null  object
 1   Flight No.              52061 non-null  object
 2   Source                  52061 non-null  object
 3   Departure               52061 non-null  object
 4   No. of stops            52061 non-null  object
 5   Arrival                 52061 non-null  object
 6   Destination             52061 non-null  object
 7   Ticket Class            52061 non-null  object
 8   Flight Duration (hrs)   52061 non-null  float64
 9   Days left               52061 non-null  int64
 10  Price                   52061 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 4.4+ MB
```

[7]:
```
#dropping Null values
df_1.isnull().sum()
```

[7]:
```
Airline                 0
Flight No.              0
Source                  0
Departure               0
No. of stops            0
Arrival                 0
Destination             0
Ticket Class            0
Flight Duration (hrs)   0
Days left               0
Price                   0
dtype: int64
```

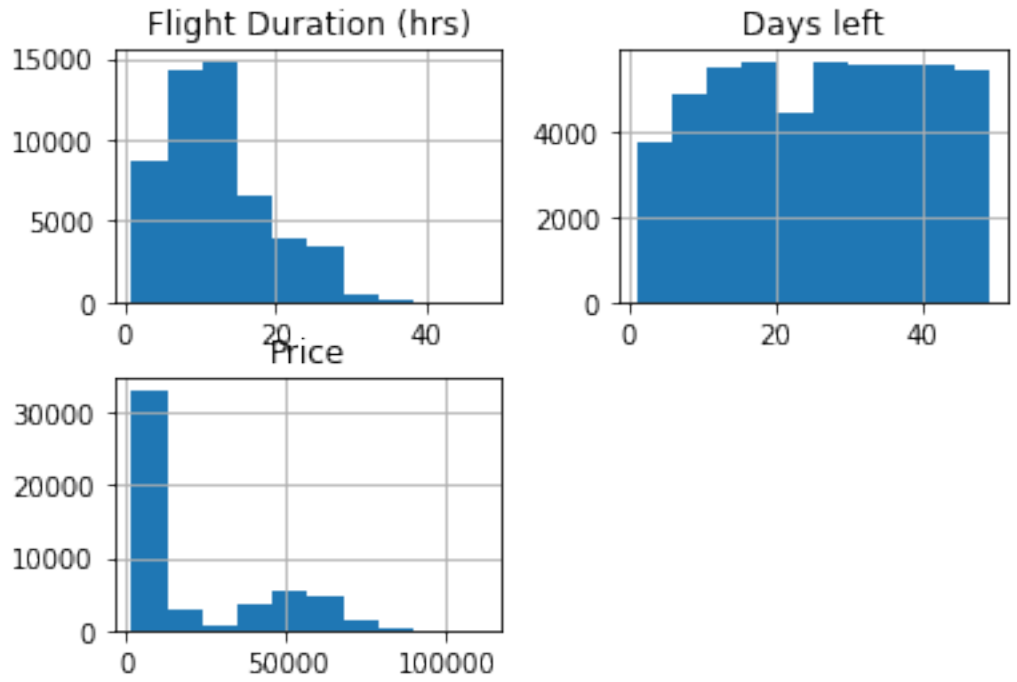
[25]:
```
#describing the dataset
df_1.describe()
```

[25]:

	Flight Duration (hrs)	Days left	Price
count	52061.000000	52061.000000	52061.000000
mean	12.029203	25.955398	21469.460575
std	6.966650	13.627462	23171.386433
min	0.830000	1.000000	1603.000000
25%	7.000000	14.000000	4687.000000
50%	11.420000	26.000000	7488.000000
75%	15.330000	38.000000	43865.000000
max	47.750000	49.000000	111883.000000

[11]:
```
#identifying the trend of the dataset with a histogram
df_1.hist()
```

[11]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[12]:
```
#loading dataset
df_2= pd.read_excel('Delhi 1 Flights.xlsx')
```

[13]:
```
df_2.head()
```

[13]:
```
   Airline Flight No. Source      Departure No. of stops     Arrival  \
0  AirAsia     I5-548  Delhi  Early_Morning            1   Afternoon
1  AirAsia     I5-559  Delhi      Afternoon    2_or_more       Night
2  AirAsia     I5-787  Delhi  Early_Morning            1   Afternoon
3  AirAsia     I5-713  Delhi        Evening    2_or_more  Late_Night
4  AirAsia     I5-744  Delhi        Morning            1     Evening

  Destination Ticket Class  Flight Duration (hrs)  Days left  Price
0   Bangalore      Economy                   7.58         41   3090
1   Bangalore      Economy                   8.00         41   3090
2   Bangalore      Economy                   8.17         41   3090
3   Bangalore      Economy                   8.75         41   3090
4   Bangalore      Economy                   9.92         41   3090
```

[16]:
```
df_2.tail()
```

[16]:
```
         Airline Flight No. Source      Departure No. of stops  Arrival  \
43135    Vistara     UK-955  Delhi        Evening            1  Morning
43136  Air_India     AI-883  Delhi          Night    2_or_more  Morning
43137    Vistara     UK-829  Delhi  Early_Morning    2_or_more  Evening
43138    Vistara     UK-819  Delhi      Afternoon    2_or_more    Night
43139    Vistara     UK-705  Delhi  Early_Morning    2_or_more  Evening

      Destination Ticket Class  Flight Duration (hrs)  Days left  Price
43135     Chennai     Business                  15.08         49  68184
43136     Chennai     Business                  12.92         49  70323
43137     Chennai     Business                  12.83         49  71376
43138     Chennai     Business                   8.83         49  76684
43139     Chennai     Business                  12.42         49  82553
```

[15]:
```
df_2.shape
```

[15]:
```
(43140, 11)
```

[17]:
```
df_2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43140 entries, 0 to 43139
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Airline               43140 non-null  object
 1   Flight No.            43140 non-null  object
 2   Source                43140 non-null  object
 3   Departure             43140 non-null  object
 4   No. of stops          43140 non-null  object
 5   Arrival               43140 non-null  object
```

```
 6   Destination             43140 non-null  object
 7   Ticket Class            43140 non-null  object
 8   Flight Duration (hrs)   43140 non-null  float64
 9   Days left               43140 non-null  int64
 10  Price                   43140 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 3.6+ MB
```

[18]:
```
#dropping Null values
df_2.isnull().sum()
```

[18]:
```
Airline                 0
Flight No.              0
Source                  0
Departure               0
No. of stops            0
Arrival                 0
Destination             0
Ticket Class            0
Flight Duration (hrs)   0
Days left               0
Price                   0
dtype: int64
```

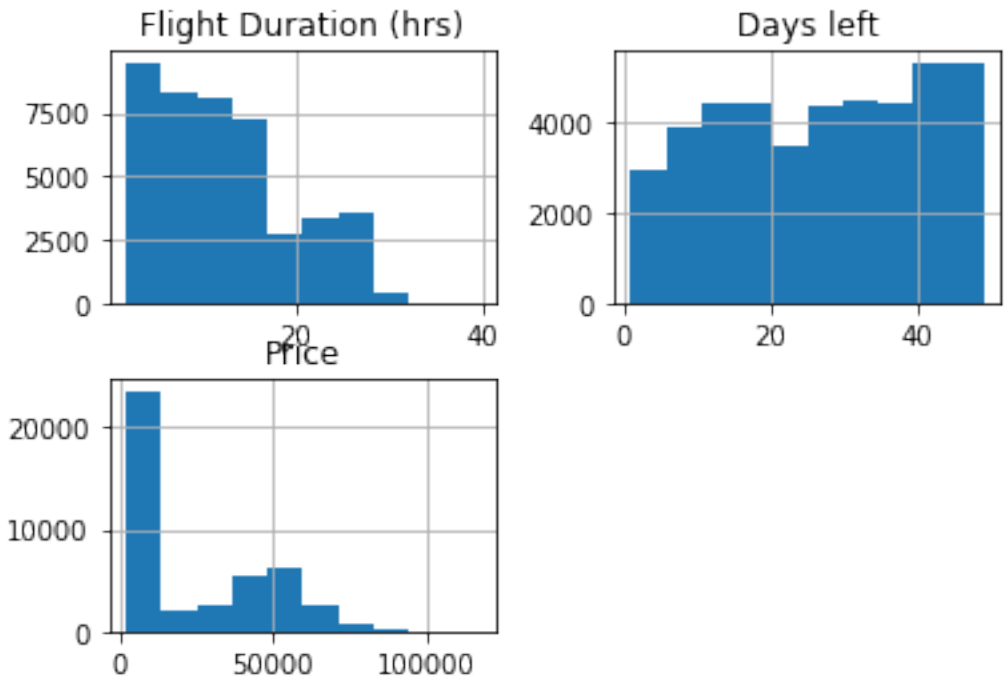
[19]:
```
df_2.describe()
```

[19]:

	Flight Duration (hrs)	Days left	Price
count	43140.000000	43140.000000	43140.000000
mean	12.248589	26.808507	24293.741145
std	7.431686	13.906185	22807.587741
min	1.920000	1.000000	1998.000000
25%	6.330000	15.000000	5097.000000
50%	11.330000	27.000000	9840.000000
75%	16.420000	39.000000	45257.000000
max	39.670000	49.000000	117307.000000

[20]:
```
#identifying the trend of the dataset with a histogram
df_2.hist()
```

[20]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[22]:
```
#loading dataset
df_3= pd.read_excel('Delhi 2 Flights .xlsx')
```

[23]:
```
df_3.head()
```

[23]:

	ï»¿Airline	Flight No.	Source	Departure	No. of stops	Arrival	Destination	Ticket Class	Flight Duration (hrs)	Days left	Price
0	SpiceJet	SG-8709	Delhi	Evening	0	Night	Mumbai	Economy	2.17	1	5953
1	SpiceJet	SG-8157	Delhi	Early_Morning	0	Morning	Mumbai	Economy	2.33	1	5953
2	AirAsia	I5-764	Delhi	Early_Morning	0	Early_Morning	Mumbai	Economy	2.17	1	5956
3	Vistara	UK-995	Delhi	Morning	0	Afternoon	Mumbai	Economy	2.25	1	5955
4	Vistara	UK-963	Delhi	Morning	0	Morning	Mumbai	Economy	2.33	1	5955

[24]:
```
df_3.tail()
```

[24]:

	ï»¿Airline	Flight No.	Source	Departure	No. of stops	Arrival
18198	AirAsia	I5-768	Delhi	Afternoon	1	Night
18199	AirAsia	I5-713	Delhi	Evening	2_or_more	Night

```
18200    AirAsia     I5-711  Delhi  Early_Morning            1   Afternoon
18201    AirAsia     I5-710  Delhi        Morning            1     Evening
18202    AirAsia     I5-550  Delhi        Evening            1  Late_Night

      Destination Ticket Class  Flight Duration (hrs)  Days left  Price
18198   Bangalore      Economy                   5.42         41   3090
18199   Bangalore      Economy                   7.00         41   3090
18200   Bangalore      Economy                   7.42         41   3090
18201   Bangalore      Economy                   7.42         41   3090
18202   Bangalore      Economy                   7.50         41   3090
```

[26]:
```
df_3.shape
```

[26]:
```
(18203, 11)
```

[27]:
```
df_3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18203 entries, 0 to 18202
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   ï»¿Airline            18203 non-null  object
 1   Flight No.             18203 non-null  object
 2   Source                 18203 non-null  object
 3   Departure              18203 non-null  object
 4   No. of stops           18203 non-null  object
 5   Arrival                18203 non-null  object
 6   Destination            18203 non-null  object
 7   Ticket Class           18203 non-null  object
 8   Flight Duration (hrs)  18203 non-null  float64
 9   Days left              18203 non-null  int64
 10  Price                  18203 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 1.5+ MB
```

[29]:
```
#dropping the null values
df_3.isnull().sum()
```

[29]:
```
ï»¿Airline            0
Flight No.             0
Source                 0
Departure              0
No. of stops           0
Arrival                0
Destination            0
Ticket Class           0
```

```
Flight Duration (hrs)    0
Days left                0
Price                    0
dtype: int64
```

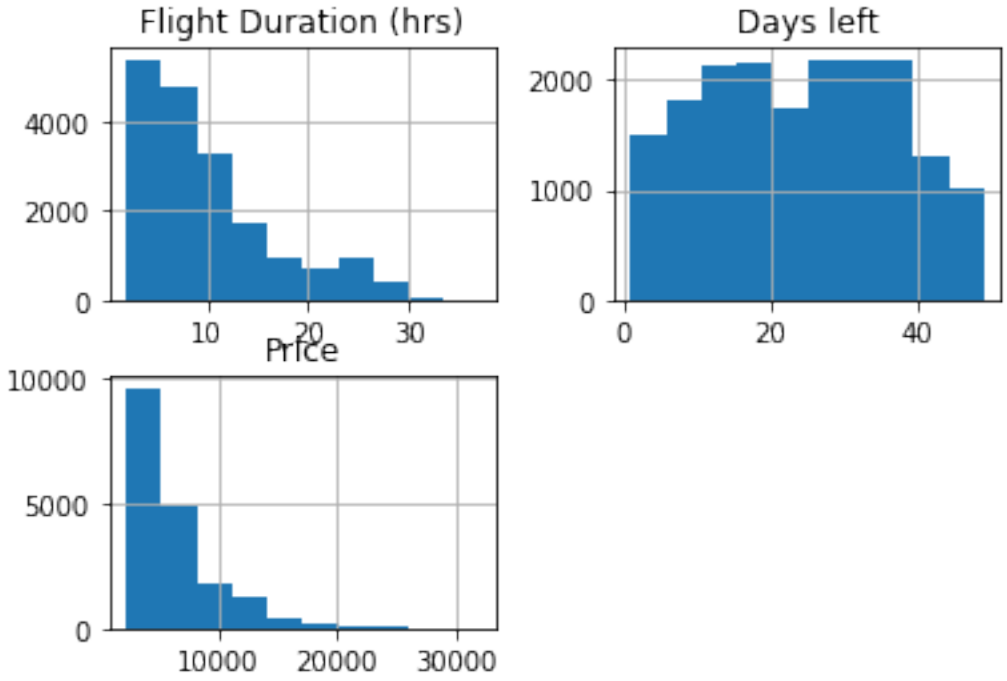
[30]:
```
#describing the dataset
df_3.describe()
```

[30]:

	Flight Duration (hrs)	Days left	Price
count	18203.000000	18203.000000	18203.000000
mean	9.778119	23.983354	6290.130034
std	6.796409	12.745930	3761.429737
min	2.000000	1.000000	2281.000000
25%	4.830000	13.000000	3855.000000
50%	8.170000	24.000000	5110.000000
75%	12.750000	34.000000	7425.000000
max	36.920000	49.000000	31917.000000

[31]:
```
#identifying the trend of the datast
df_3.hist()
```

[31]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[32]:
```
df_4= pd.read_csv('Hyderabad 1 flights.csv')
```

[33]:
```
df_4.head()
```

[33]:
```
    Airline Flight No.     Source  Departure No. of stops     Arrival  \
0   AirAsia    I5-1543  Hyderabad  Afternoon            1       Night
1   AirAsia    I5-1543  Hyderabad  Afternoon            1  Late_Night
2   Vistara     UK-880  Hyderabad  Afternoon            0   Afternoon
3   AirAsia    I5-1517  Hyderabad    Evening            1  Late_Night
4  SpiceJet    SG-3283  Hyderabad    Evening            1     Evening

  Destination Ticket Class  Flight Duration (hrs)  Days left    Price
0       Delhi      Economy                   7.92        1.0   8393.0
1       Delhi      Economy                  11.00        1.0   8393.0
2       Delhi      Economy                   2.25        1.0  11698.0
3       Delhi      Economy                   9.58        1.0  11488.0
4       Delhi      Economy                  26.92        1.0  11591.0
```

[36]:
```
df_4.tail()
```

[36]:
```
      Airline Flight No. Source Departure No. of stops Arrival Destination  \
27929     NaN        NaN    NaN       NaN          NaN     NaN         NaN
27930     NaN        NaN    NaN       NaN          NaN     NaN         NaN
27931     NaN        NaN    NaN       NaN          NaN     NaN         NaN
27932     NaN        NaN    NaN       NaN          NaN     NaN         NaN
27933     NaN        NaN    NaN       NaN          NaN     NaN         NaN

      Ticket Class  Flight Duration (hrs)  Days left  Price
27929          NaN                    NaN        NaN    NaN
27930          NaN                    NaN        NaN    NaN
27931          NaN                    NaN        NaN    NaN
27932          NaN                    NaN        NaN    NaN
27933          NaN                    NaN        NaN    NaN
```

[35]:
```
df_4.shape
```

[35]:
```
(27934, 11)
```

[37]:
```
df_4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27934 entries, 0 to 27933
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Airline                22948 non-null  object
```

```
 1   Flight No.              22948 non-null  object
 2   Source                  22948 non-null  object
 3   Departure               22948 non-null  object
 4   No. of stops            22948 non-null  object
 5   Arrival                 22948 non-null  object
 6   Destination             22948 non-null  object
 7   Ticket Class            22948 non-null  object
 8   Flight Duration (hrs)   22948 non-null  float64
 9   Days left               22948 non-null  float64
 10  Price                   22948 non-null  float64
dtypes: float64(3), object(8)
memory usage: 2.3+ MB
```

[38]:
```
df_4.isnull().sum()
```

[38]:
```
Airline                 4986
Flight No.              4986
Source                  4986
Departure               4986
No. of stops            4986
Arrival                 4986
Destination             4986
Ticket Class            4986
Flight Duration (hrs)   4986
Days left               4986
Price                   4986
dtype: int64
```

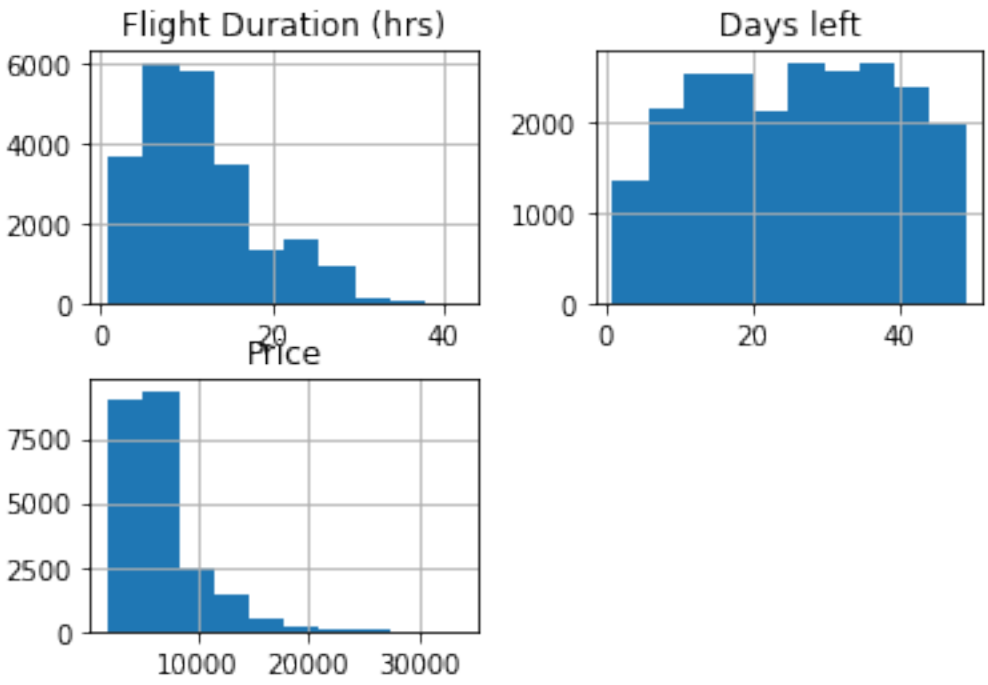
[39]:
```
df_4.describe()
```

[39]:

	Flight Duration (hrs)	Days left	Price
count	22948.000000	22948.000000	22948.000000
mean	11.349399	25.823383	6299.803251
std	6.911996	13.064234	3477.532128
min	0.920000	1.000000	1755.000000
25%	6.250000	15.000000	4363.000000
50%	10.250000	26.000000	5527.000000
75%	15.000000	37.000000	7496.000000
max	42.000000	49.000000	33853.000000

[40]:
```
df_4.hist()
```

[40]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[ ]:
```
df_5= pd.read_json('Hyderabad 2 flight.json')
```

[43]:
```
df_6= pd.read_json('Mumbai Flight.json')
```

[44]:
```
df_6.head()
```

[44]:

	Airline	Flight No.	Source	Departure	No. of stops	Arrival	Destination	Ticket Class	Flight Duration (hrs)	Days left	Price
0	Vistara	UK-994	Mumbai	Morning	0	Afternoon	Delhi	Economy	2.25	1	5943
1	Vistara	UK-910	Mumbai	Evening	0	Evening	Delhi	Economy	2.17	1	6048
2	GO_FIRST	G8-339	Mumbai	Evening	0	Night	Delhi	Economy	2.17	1	5942
3	Indigo	6E-6004	Mumbai	Evening	0	Night	Delhi	Economy	2.08	1	5943
4	Indigo	6E-2077	Mumbai	Afternoon	0	Afternoon	Delhi	Economy	2.17	1	5943

[45]:
```
df_6.tail()
```

[45]:

	Airline	Flight No.	Source	Departure	No. of stops	Arrival
60891	Vistara	UK-996	Mumbai	Evening	1	Morning

```
60892  Vistara     UK-940  Mumbai    Evening            1    Night
60893  Vistara     UK-996  Mumbai    Evening            1    Night
60894  Vistara     UK-613  Mumbai  Afternoon    2_or_more    Night
60895  Vistara     UK-877  Mumbai    Morning    2_or_more    Night

      Destination Ticket Class  Flight Duration (hrs)  Days left  Price
60891     Chennai     Business                  15.42         49  64173
60892     Chennai     Business                  24.50         49  64173
60893     Chennai     Business                  25.75         49  64173
60894     Chennai     Business                  10.50         49  64879
60895     Chennai     Business                   9.50         49  68743
```

[46]:
```
df_6.shape
```

[46]:
```
(60896, 11)
```

[47]:
```
df_6.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60896 entries, 0 to 60895
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Airline                60896 non-null  object
 1   Flight No.             60896 non-null  object
 2   Source                 60896 non-null  object
 3   Departure              60896 non-null  object
 4   No. of stops           60896 non-null  object
 5   Arrival                60896 non-null  object
 6   Destination            60896 non-null  object
 7   Ticket Class           60896 non-null  object
 8   Flight Duration (hrs)  60896 non-null  float64
 9   Days left              60896 non-null  int64
 10  Price                  60896 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 5.1+ MB
```

[48]:
```
df_6.isnull().sum()
```

[48]:
```
Airline                0
Flight No.             0
Source                 0
Departure              0
No. of stops           0
Arrival                0
Destination            0
Ticket Class           0
```

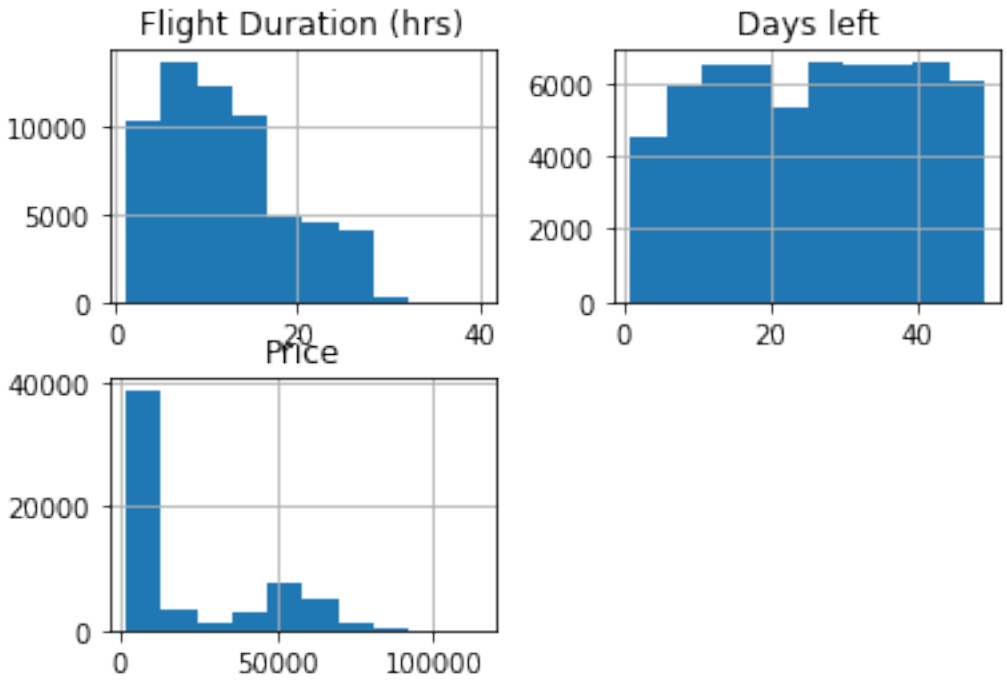
```
Flight Duration (hrs)    0
Days left                0
Price                    0
dtype: int64
```

[49]:
```
df_6.describe()
```

[49]:
```
       Flight Duration (hrs)     Days left          Price
count           60896.000000  60896.000000   60896.000000
mean               11.888448     25.769755   21483.818839
std                 7.121188     13.619620   23394.131784
min                 1.170000      1.000000    1890.000000
25%                 6.420000     14.000000    4658.000000
50%                11.080000     26.000000    7413.000000
75%                16.170000     38.000000   45693.000000
max                40.000000     49.000000  114523.000000
```

[50]:
```
df_6.hist()
```

[50]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[51]:
```
df_7= pd.read_csv('Kolkata Flight.csv')
```

[52]:
```
df_7.head()
```

[52]:
```
    Airline Flight No.   Source Departure No. of stops     Arrival  \
0  SpiceJet    SG-8264  Kolkata     Night            0  Late_Night
1   AirAsia     I5-582  Kolkata   Morning            1     Evening
2   AirAsia    I5-2473  Kolkata   Morning            1       Night
3   AirAsia    I5-1563  Kolkata   Evening    2_or_more   Afternoon
4    Indigo    6E-2009  Kolkata     Night            0  Late_Night

  Destination Ticket Class  Flight Duration (hrs)  Days left  Price
0       Delhi      Economy                   2.50          1   6488
1       Delhi      Economy                   9.25          1   6353
2       Delhi      Economy                  12.42          1   6353
3       Delhi      Economy                  18.33          1   6353
4       Delhi      Economy                   2.50          1   6489
```

[53]:
```
df_7.tail()
```

[53]:
```
         Airline Flight No.   Source      Departure No. of stops  \
46342  Air_India     AI-402  Kolkata        Morning            1
46343  Air_India     AI-763  Kolkata  Early_Morning            1
46344  Air_India     AI-402  Kolkata        Morning            1
46345  Air_India     AI-732  Kolkata      Afternoon    2_or_more
46346  Air_India     AI-768  Kolkata      Afternoon            1

             Arrival Destination Ticket Class  Flight Duration (hrs)  \
46342        Morning     Chennai     Business                  22.50
46343        Morning     Chennai     Business                  26.00
46344      Afternoon     Chennai     Business                  26.25
46345          Night     Chennai     Business                  10.58
46346  Early_Morning     Chennai     Business                  15.75

       Days left  Price
46342         49  55983
46343         49  55983
46344         49  55983
46345         49  60355
46346         49  73377
```

[54]:
```
df_7.shape
```

[54]:
```
(46347, 11)
```

[55]:
```
df_7.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46347 entries, 0 to 46346
Data columns (total 11 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Airline                 46347 non-null  object
 1   Flight No.              46347 non-null  object
 2   Source                  46347 non-null  object
 3   Departure               46347 non-null  object
 4   No. of stops            46347 non-null  object
 5   Arrival                 46347 non-null  object
 6   Destination             46347 non-null  object
 7   Ticket Class            46347 non-null  object
 8   Flight Duration (hrs)   46347 non-null  float64
 9   Days left               46347 non-null  int64
 10  Price                   46347 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 3.9+ MB
```

[56]:
```
df_7.isnull().sum()
```

[56]:
```
Airline                 0
Flight No.              0
Source                  0
Departure               0
No. of stops            0
Arrival                 0
Destination             0
Ticket Class            0
Flight Duration (hrs)   0
Days left               0
Price                   0
dtype: int64
```

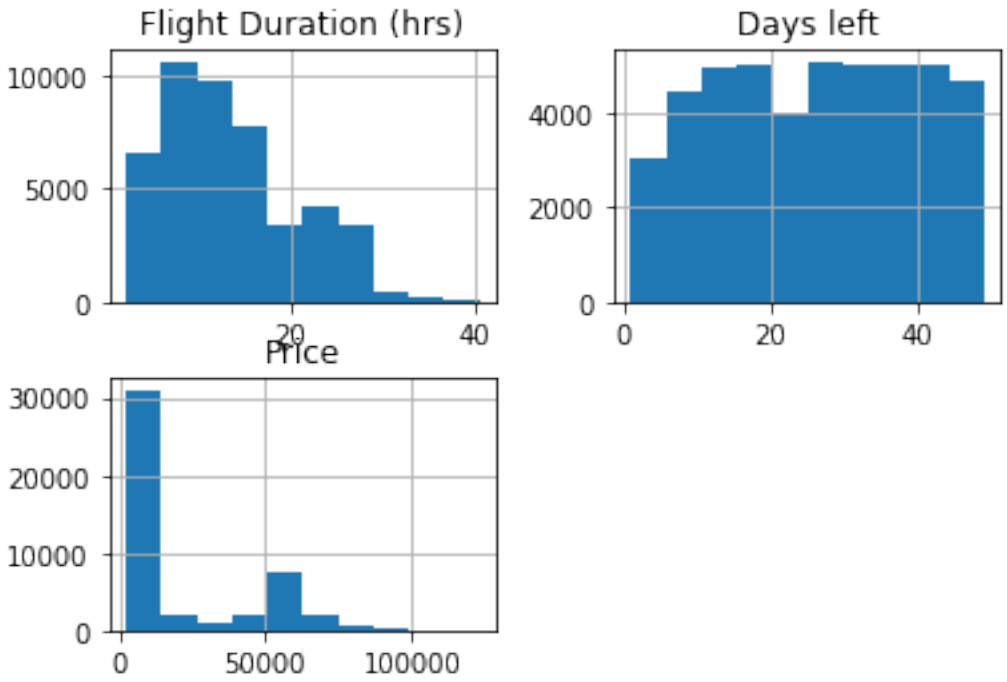
[57]:
```
df_7.describe()
```

[57]:

	Flight Duration (hrs)	Days left	Price
count	46347.000000	46347.000000	46347.000000
mean	13.249898	26.013162	21746.235679
std	7.223163	13.511999	23439.972854
min	2.080000	1.000000	2436.000000
25%	7.670000	15.000000	5853.000000
50%	12.250000	26.000000	7958.000000
75%	17.500000	38.000000	49207.000000
max	40.500000	49.000000	123071.000000

[58]:
```
df_7.hist()
```

[58]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[59]:
```
df_8= pd.read_json('Chennai Flight.json')
```

[60]:
```
df_8.head()
```

[60]:

	Airline	Flight No.	Source	Departure	No. of stops	Arrival	Destination	Ticket Class	Flight Duration (hrs)	Days left	Price
0	Indigo	6E-2752	Chennai	Evening	0	Night	Delhi	Economy	2.75	1	15752
1	Indigo	6E-613	Chennai	Evening	0	Night	Delhi	Economy	2.75	1	18902
2	Air_India	AI-539	Chennai	Evening	0	Evening	Delhi	Economy	2.83	1	20268
3	SpiceJet	SG-612	Chennai	Night	1	Morning	Delhi	Economy	11.75	2	7352
4	SpiceJet	SG-612	Chennai	Night	1	Afternoon	Delhi	Economy	14.83	2	7352

[61]:
```
df_8.tail()
```

[61]:
```
       Airline Flight No.   Source      Departure No. of stops  Arrival  \
38695  Vistara     UK-822  Chennai        Morning            1  Evening
38696  Vistara     UK-826  Chennai      Afternoon            1    Night
38697  Vistara     UK-832  Chennai  Early_Morning            1    Night
38698  Vistara     UK-828  Chennai  Early_Morning            1  Evening
38699  Vistara     UK-822  Chennai        Morning            1  Evening

      Destination Ticket Class  Flight Duration (hrs)  Days left  Price
38695   Hyderabad     Business                  10.08         49  69265
38696   Hyderabad     Business                  10.42         49  77105
38697   Hyderabad     Business                  13.83         49  79099
38698   Hyderabad     Business                  10.00         49  81585
38699   Hyderabad     Business                  10.08         49  81585
```

[62]:
```
df_8.shape
```

[62]:
```
(38700, 11)
```

[63]:
```
df_8.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38700 entries, 0 to 38699
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Airline                38700 non-null  object
 1   Flight No.             38700 non-null  object
 2   Source                 38700 non-null  object
 3   Departure              38700 non-null  object
 4   No. of stops           38700 non-null  object
 5   Arrival                38700 non-null  object
 6   Destination            38700 non-null  object
 7   Ticket Class           38700 non-null  object
 8   Flight Duration (hrs)  38700 non-null  float64
 9   Days left              38700 non-null  int64
 10  Price                  38700 non-null  int64
dtypes: float64(1), int64(2), object(8)
memory usage: 3.2+ MB
```

[64]:
```
df_8.isnull().sum()
```

[64]:
```
Airline                 0
Flight No.              0
Source                  0
Departure               0
No. of stops            0
Arrival                 0
```

```
Destination              0
Ticket Class             0
Flight Duration (hrs)    0
Days left                0
Price                    0
dtype: int64
```

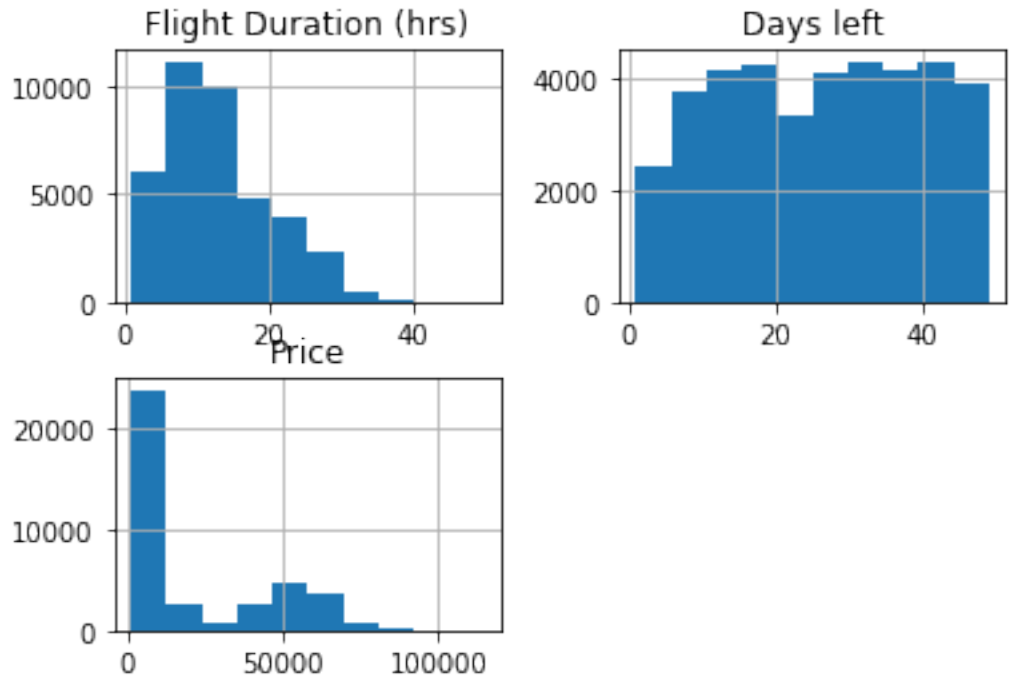
[65]:
```
df_8.describe()
```

[65]:

	Flight Duration (hrs)	Days left	Price
count	38700.000000	38700.000000	38700.000000
mean	12.838901	26.095736	21995.339871
std	7.353063	13.460905	23527.227014
min	0.830000	1.000000	1105.000000
25%	7.500000	15.000000	4672.000000
50%	11.580000	26.000000	7846.000000
75%	16.750000	38.000000	45185.000000
max	49.830000	49.000000	114704.000000

[66]:
```
df_8.hist()
```

[66]:
```
array([[<AxesSubplot: title={'center': 'Flight Duration (hrs)'}>,
        <AxesSubplot: title={'center': 'Days left'}>],
       [<AxesSubplot: title={'center': 'Price'}>, <AxesSubplot: >]],
      dtype=object)
```



[ ]:
```
combined_df= pd.concat(['df_1', 'df_2', 'df_3', 'df_4', 'df_5', 'df_6', 'df_7',␣
↪'df_8' ]),
print(combined_df.shape)
```