In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:
```
df = pd.read_csv('Downloads/Data_Train.csv')
```

In [3]:
```
df.head()
```

Out[3]:

	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 [4]:
```
df.tail()
```

Out[4]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
10678	Air Asia	9/04/2019	Kolkata	Banglore	CCU → BLR	19:55	22:25	2h 30m	non-stop	No info	4107
10679	Air India	27/04/2019	Kolkata	Banglore	CCU → BLR	20:45	23:20	2h 35m	non-stop	No info	4145
10680	Jet Airways	27/04/2019	Banglore	Delhi	BLR → DEL	08:20	11:20	3h	non-stop	No info	7229
10681	Vistara	01/03/2019	Banglore	New Delhi	BLR → DEL	11:30	14:10	2h 40m	non-stop	No info	12648
10682	Air India	9/05/2019	Delhi	Cochin	DEL → GOI → BOM → COK	10:55	19:15	8h 20m	2 stops	No info	11753

In [5]:
```
df.shape
```

Out[5]:
```
(10683, 11)
```

In [6]:
```
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
```

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

Out[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
```

In [8]:
```
df.dropna(inplace=True)
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
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 [10]:
```
df.describe()
```

Out[10]:

	Price
count	10682.000000
mean	9087.214567
std	4611.359167
min	1759.000000
25%	5277.000000
50%	8372.000000
75%	12373.000000
max	79512.000000

In [11]:
```
df.describe(include='object')
```

Out[11]:

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info
count	10682	10682	10682	10682	10682	10682	10682	10682	10682	10682
unique	12	44	5	6	128	222	1343	368	5	10
top	Jet Airways	18/05/2019	Delhi	Cochin	DEL → BOM → COK	17:30	19:00	2h 50m	1 stop	No info
freq	3849	504	4536	4536	2376	233	151	550	5625	8344

In [12]:
```
df['Airline'].value_counts()
```

Out[12]:
```
Jet Airways                          3849
IndiGo                               2053
Air India                            1751
Multiple carriers                    1196
SpiceJet                              818
Vistara                               479
Air Asia                              319
GoAir                                 194
Multiple carriers Premium economy      13
Jet Airways Business                    6
Vistara Premium economy                 3
Trujet                                  1
Name: Airline, dtype: int64
```

In [13]:
```
df['Source'].value_counts()
```

Out[13]:
```
Delhi       4536
Kolkata     2871
Banglore    2197
Mumbai       697
Chennai      381
Name: Source, dtype: int64
```

In [14]:
```
df['Destination'].value_counts()
```

Out[14]:
```
Cochin       4536
Banglore     2871
Delhi        1265
New Delhi     932
Hyderabad     697
Kolkata       381
Name: Destination, dtype: int64
```

In [15]:
```
df['Route'].value_counts()
```

Out[15]:
```
DEL → BOM → COK             2376
BLR → DEL                   1552
CCU → BOM → BLR              979
CCU → BLR                    724
BOM → HYD                    621
                            ... 
CCU → VTZ → BLR                1
CCU → IDR → JDH → DEL → BLR    1
BOM → DED → DEL → HYD          1
BOM → UDR → DEL → HYD          1
BOM → VNS → DEL → HYD          1
Name: Route, Length: 128, dtype: int64
```

In [16]:
```
df['Total_Stops'].value_counts()
```

Out[16]:
```
1 stop      5625
non-stop    3491
2 stops     1520
3 stops       43
4 stops        3
Name: Total_Stops, dtype: int64
```

In [17]:
```
df['Additional_Info'].value_counts()
```

Out[17]:
```
No info                         8344
In-flight meal not included     1982
No check-in baggage included     320
1 Long layover                    19
Change airports                    7
Business class                     4
No Info                            3
1 Short layover                    1
Red-eye flight                     1
2 Long layover                     1
Name: Additional_Info, dtype: int64
```

In [18]:
```
df['Date_of_Journey'].value_counts()
```

Out[18]:
```
18/05/2019    504
21/05/2019    480
24/05/2019    475
1/06/2019     465
9/05/2019     465
             ... 
6/06/2019     165
3/06/2019     162
27/06/2019    158
3/03/2019     155
1/03/2019     155
Name: Date_of_Journey, Length: 44, dtype: int64
```

In [19]:
```
df['Dep_Time'].value_counts()
```

Out[19]:
```
17:30    233
10:00    230
18:00    228
18:55    222
08:00    220
        ... 
05:25      1
05:40      1
05:05      1
05:30      1
05:35      1
Name: Dep_Time, Length: 222, dtype: int64
```

In [20]:
```
df['Arrival_Time'].value_counts()
```

Out[20]:
```
19:00          151
10:00          150
19:00 19 Jun    149
21:00          144
16:10          140
              ... 
00:30 10 Jun     1
00:30 12 Jun     1
00:30 13 Jun     1
00:30 15 Jun     1
00:30 22 Jun     1
Name: Arrival_Time, Length: 1343, dtype: int64
```

In [21]:
```
df['Duration'].value_counts()
```

Out[21]:
```
2h 50m     550
1h 30m     386
2h 55m     337
2h 45m     329
2h 40m     320
          ... 
42h 5m       1
42h 30m      1
42h 40m      1
43h 10m      1
47h 40m      1
Name: Duration, Length: 368, dtype: int64
```

In [22]:
```
df['Price'].value_counts()
```

Out[22]:
```
10262    258
10844    238
10703    233
7229     228
4804     208
        ... 
3714       1
3708       1
3704       1
3694       1
3687       1
Name: Price, Length: 1870, dtype: int64
```

In [23]:
```
df['Airline'].unique()
```

Out[23]:
```
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)
```

In [24]:
```
df['Source'].unique()
```

Out[24]:
```
array(['Banglore', 'Kolkata', 'Delhi', 'Chennai', 'Mumbai'],
      dtype=object)
```

In [25]:
```
df['Destination'].unique()
```

Out[25]:
```
array(['New Delhi', 'Banglore', 'Cochin', 'Kolkata', 'Delhi',
       'Hyderabad'], dtype=object)
```

In [26]:
```
df['Route'].unique()
```

Out[26]:
```
array(['BLR → DEL', 'CCU → IXR → BBI → BLR', 'DEL → LKO → BOM → COK',
       'CCU → NAG → BLR', 'BLR → NAG → DEL', 'CCU → BLR',
       'BLR → BOM → DEL', 'DEL → BOM → COK', 'DEL → BLR → COK',
       'MAA → CCU', 'CCU → BOM → BLR', 'DEL → BOM → COK',
       'DEL → JAI → BOM → COK', 'DEL → COK', 'CCU → DEL → BLR',
       'BLR → BOM → DEL', 'BLR → MAA → DEL', 'BLR → HYD → DEL',
       'BLR → HYD → DEL', 'CCU → HYD → BLR', 'BLR → DEL',
       'CCU → GAU → BLR', 'DEL → HYD → COK', 'CCU → DEL → BLR',
       'BLR → BOM → DEL', 'BOM → HYD', 'DEL → BOM → COK',
       'DEL → BOM → COK', 'BOM → DEL → HYD', 'DEL → BOM → COK',
       'DEL → BOM → COK', 'DEL → BOM → COK
```