

```
In [77]: # Import Lib
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [78]: # Load the dataset
df = pd.read_csv('planes.csv')
```

```
In [79]: # Understand its structure and contents of the dataset
print("Columns :\n",df.columns)
print("Shape :\n",df.shape)
```

```
Columns :
Index(['Airline', 'Date_of_Journey', 'Source', 'Destination', 'Route',
      'Dep_Time', 'Arrival_Time', 'Duration', 'Total_Stops',
      'Additional_Info', 'Price'],
      dtype='object')
Shape :
(10660, 11)
```

```
In [80]: # How many unique airlines are represented in the dataset?
airlines = df['Airline'].unique()
print(airlines)
```

```
['Jet Airways' 'IndiGo' 'SpiceJet' 'Multiple carriers' 'Air India' 'GoAir'
 'Vistara' nan 'Air Asia']
```

```
In [81]: # What is the range of flight prices present in the dataset?
min = df['Price'].min()
max = df['Price'].max()
print("Range of flight price is from '",min,'" to '",max,'" )
```

```
Range of flight price is from ' 1759.0 ' to ' 54826.0 '
```

```
In [82]: # Are there any duplicate records? If so, how many?
print("Duplicated Records :",df.index.duplicated().sum())
```

```
Duplicated Records : 0
```

```
In [83]: # What is the frequency distribution of flights among different sources and destinations?
city_count_s = df['Source'].value_counts()
city_count_d = df['Destination'].value_counts()
print("Frequency Distribution of Source Flights :\n",city_count_s)
print("Frequency Distribution of Destination Flights :\n",city_count_d)
```

```
Frequency Distribution of Source Flights :
```

```
Delhi      4451
Kolkata    2814
Banglore   2149
Mumbai     685
Chennai    374
```

```
Name: Source, dtype: int64
```

```
Frequency Distribution of Destination Flights :
```

```
Cochin     4391
Banglore    2773
Delhi       1219
New Delhi   888
Hyderabad   673
Kolkata     369
```

```
Name: Destination, dtype: int64
```

```
In [84]: airlines = df['Airline'].value_counts()
print(airlines)
```

```
Jet Airways      3685
IndiGo           1981
Air India        1686
Multiple carriers 1148
SpiceJet         787
Vistara          455
Air Asia         309
GoAir            182
Name: Airline, dtype: int64
```

```
In [85]: # Are there any missing values in crucial columns like Price, Duration, or Total_Stops
print("Missing Values in column Price ",df['Price'].isnull().sum())
print("Missing Values in column Duration ",df['Duration'].isnull().sum())
print("Missing Values in column Total_Stops ",df['Total_Stops'].isnull().sum())
```

```
Missing Values in column Price  616
Missing Values in column Duration  214
Missing Values in column Total_Stops  212
```

```
In [86]: #Validate the format of Date_of_Journey, Dep_Time, and Arrival_Time. Are they consistent?
df.dtypes
```

```
Out[86]: Airline      object
Date_of_Journey  object
Source          object
Destination     object
Route           object
Dep_Time        object
Arrival_Time    object
Duration        object
Total_Stops     object
Additional_Info  object
Price           float64
dtype: object
```

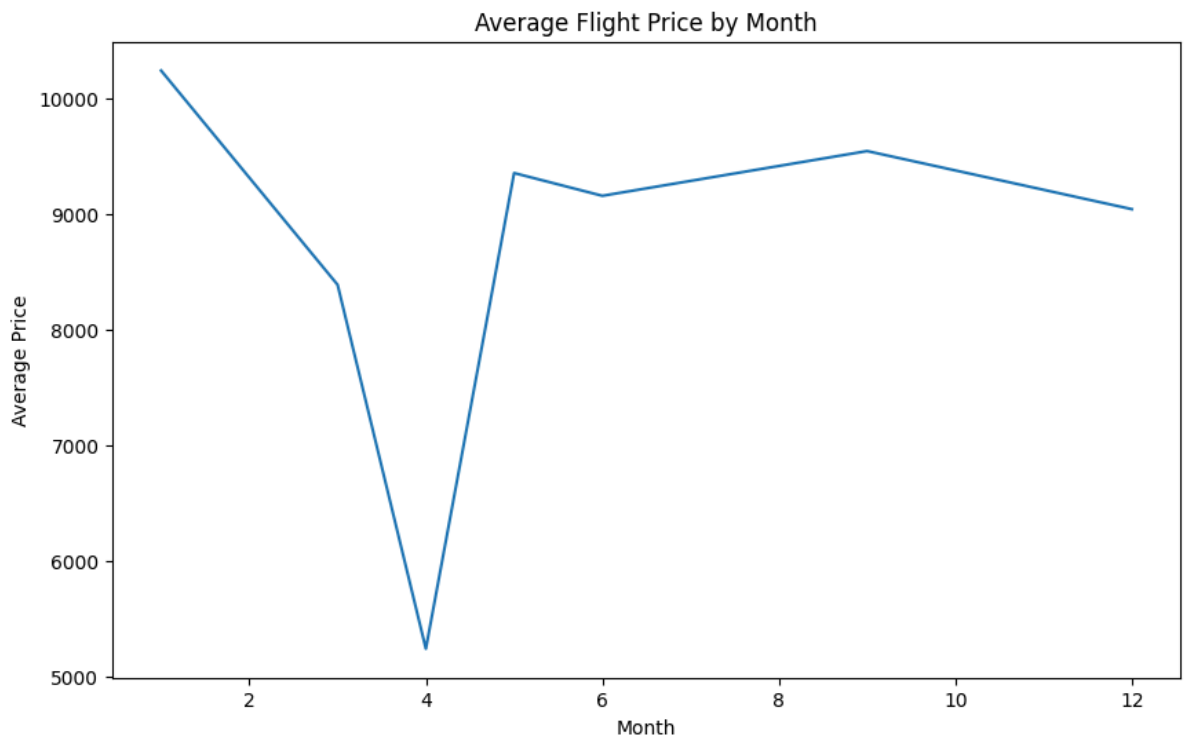
```
In [88]: #How does the price of flights vary with the date of journey?
df['Journey_day'] = pd.to_datetime(df['Date_of_Journey']).dt.day
df['Journey_month'] = pd.to_datetime(df['Date_of_Journey']).dt.month
df.drop('Date_of_Journey', axis=1, inplace=True)
df_price_by_month = df.groupby('Journey_month')['Price'].mean().reset_index()
df_price_by_day = df.groupby('Journey_day')['Price'].mean().reset_index()
plt.figure(figsize=(10, 6))
sns.lineplot(data=df_price_by_month, x='Journey_month', y='Price')
plt.xlabel('Month')
plt.ylabel('Average Price')
plt.title('Average Flight Price by Month')
plt.show()
```

<ipython-input-88-436b1ee0dead>:2: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a format to ensure consistent parsing.

```
df['Journey_day'] = pd.to_datetime(df['Date_of_Journey']).dt.day
```

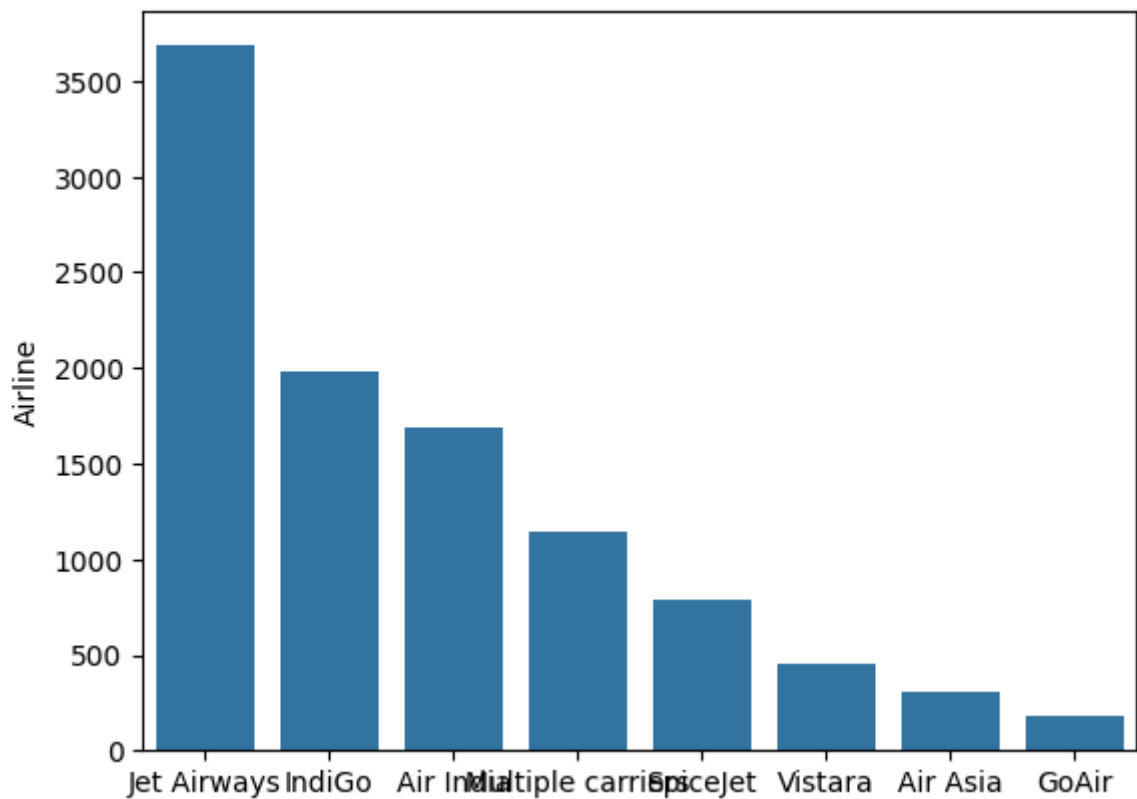
<ipython-input-88-436b1ee0dead>:3: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a format to ensure consistent parsing.

```
df['Journey_month'] = pd.to_datetime(df['Date_of_Journey']).dt.month
```



```
In [97]: #Visualize the number of flights per airline using a bar chart.
airlines_count = df['Airline'].value_counts()
sns.barplot(airlines_count)
```

```
Out[97]: <Axes: ylabel='Airline'>
```



```
In [89]: #Is there a noticeable trend in flight prices over months?
```

```
In [87]: # How does the flight duration vary across different airlines?
df1 = df[df['Airline'] == "Jet Airways"]
df2 = df[df['Airline'] == "IndiGo"]
df3 = df[df['Airline'] == "Air India"]
df4 = df[df['Airline'] == "Multiple carriers"]
```

```
df5 = df[df['Airline'] == "SpiceJet"]
df6 = df[df['Airline'] == "Vistara"]
df7 = df[df['Airline'] == "Air Asia"]
df8 = df[df['Airline'] == "GoAir"]

print("Jet Airways has Max flight duration :",df1['Duration'].index.max())
```

Jet Airways has Max flight duration : 10657

In [90]: *#Which airline has the highest average flight price? What about the lowest?*

In [91]: *#Explore the relationship between flight duration and price. Is there a correlation?*

In [93]: *#Use a boxplot to compare the flight prices across different airlines. What insights can you derive?*

In [94]: