```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("/content/Clean_Dataset.csv")
```

```
df.head()
```

	Unnamed: 0	airline	flight	source_city	departure_time	stops	arrival_time	destination_city
0	0	SpiceJet	SG-8709	Delhi	Evening	zero	Night	Mumbai
1	1	SpiceJet	SG-8157	Delhi	Early_Morning	zero	Morning	Mumbai
2	2	AirAsia	I5-764	Delhi	Early_Morning	zero	Early_Morning	Mumbai
3	3	Vistara	UK-995	Delhi	Morning	zero	Afternoon	Mumbai
4	4	Vistara	UK-963	Delhi	Morning	zero	Morning	Mumbai

```
df.tail()
```

	Unnamed: 0	airline	flight	source_city	departure_time	stops	arrival_time	destination_
300148	300148	Vistara	UK-822	Chennai	Morning	one	Evening	Hy
300149	300149	Vistara	UK-826	Chennai	Afternoon	one	Night	Hy
300150	300150	Vistara	UK-832	Chennai	Early_Morning	one	Night	Hy
300151	300151	Vistara	UK-828	Chennai	Early_Morning	one	Evening	Hy
300152	300152	Vistara	UK-822	Chennai	Morning	one	Evening	Hy

```
df.isnull().sum()
df.dropna(inplace=True)
```

```
df.describe()
```

	Unnamed: 0	duration	days_left	price
count	300153.000000	300153.000000	300153.000000	300153.000000
mean	150076.000000	12.221021	26.004751	20889.660523
std	86646.852011	7.191997	13.561004	22697.767366
min	0.000000	0.830000	1.000000	1105.000000
25%	75038.000000	6.830000	15.000000	4783.000000
50%	150076.000000	11.250000	26.000000	7425.000000
75%	225114.000000	16.170000	38.000000	42521.000000
max	300152.000000	49.830000	49.000000	123071.000000

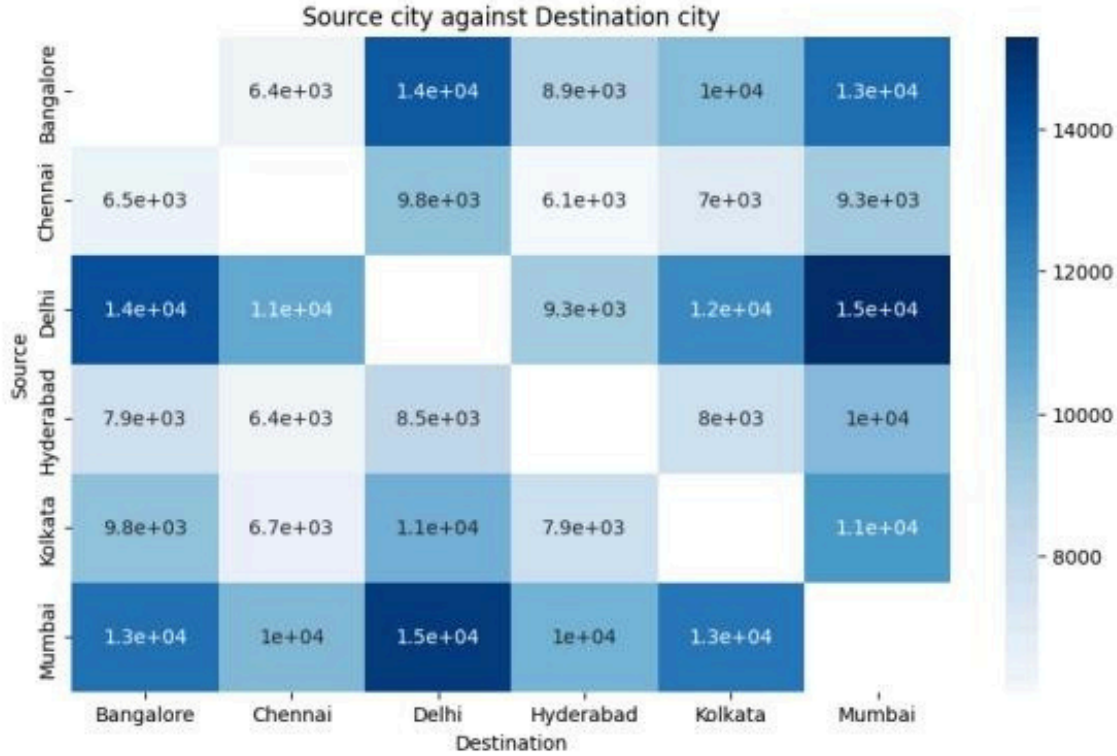
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300153 entries, 0 to 300152
Data columns (total 12 columns):
 #   Column            Non-Null Count   Dtype  
---  ------            --------------   -----  
 0   Unnamed: 0        300153 non-null  int64  
 1   airline           300153 non-null  object 
 2   flight            300153 non-null  object 
 3   source_city       300153 non-null  object 
 4   departure_time    300153 non-null  object 
 5   stops             300153 non-null  object 
 6   arrival_time      300153 non-null  object 
 7   destination_city  300153 non-null  object 
 8   class             300153 non-null  object 
 9   duration          300153 non-null  float64
 10  days_left         300153 non-null  int64  
 11  price             300153 non-null  int64  
dtypes: float64(1), int64(3), object(8)
memory usage: 27.5+ MB
```

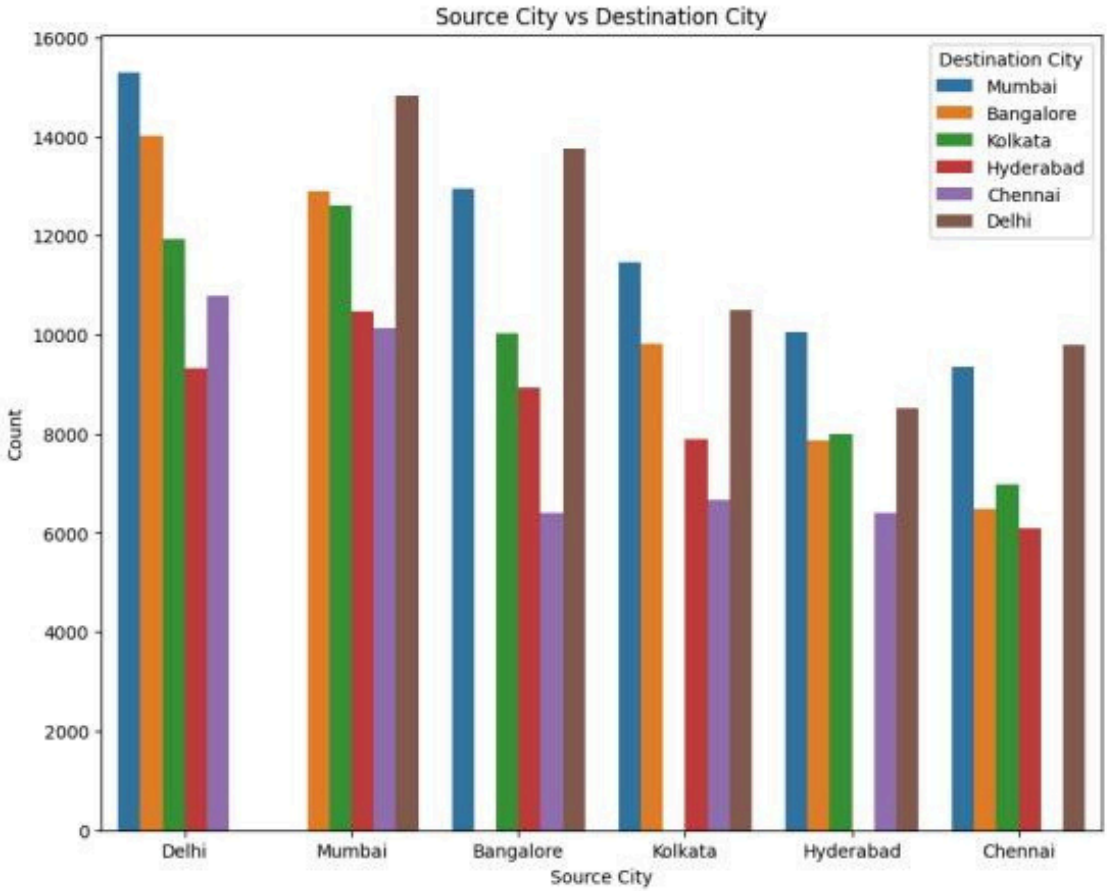
```
airline_freq = df['airline'].value_counts()
print(airline_freq)
```

```
airline
Vistara      127859
Air_India     80892
Indigo        43120
GO_FIRST      23173
AirAsia       16098
SpiceJet       9011
Name: count, dtype: int64
```

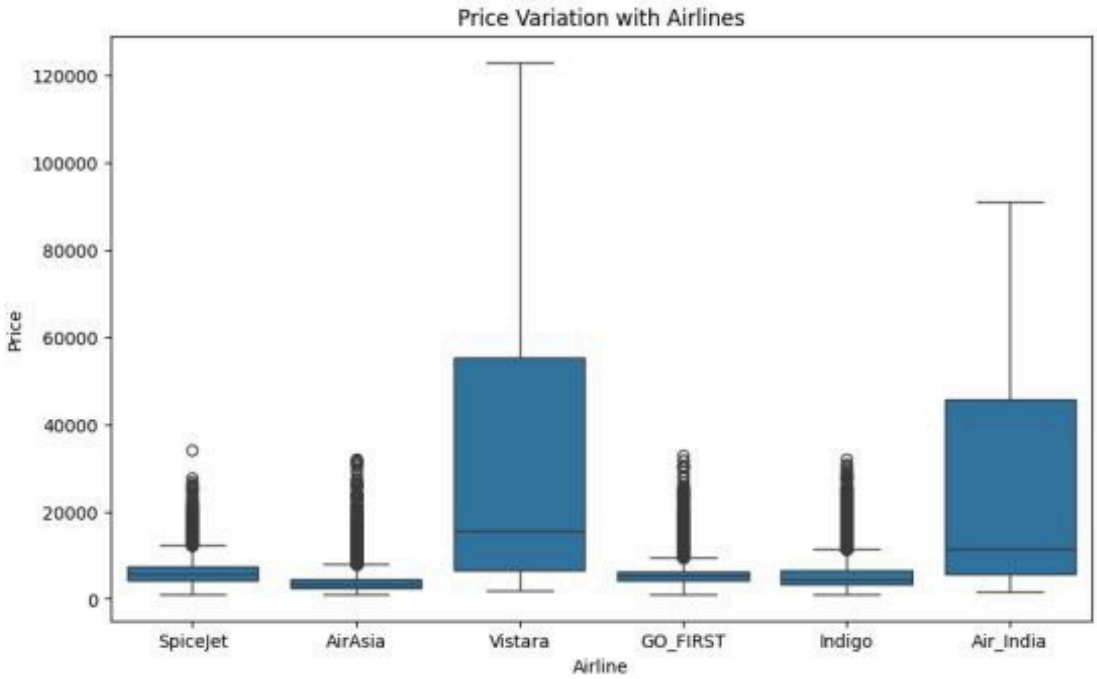
```
source_dest = df.groupby('source_city')['destination_city'].value_counts().unstack()
plt.figure(figsize=(10,6))
sns.heatmap(source_dest, annot=True, cmap='Blues')
plt.title('Source city against Destination city')
plt.xlabel('Destination')
plt.ylabel('Source')
plt.show()
```



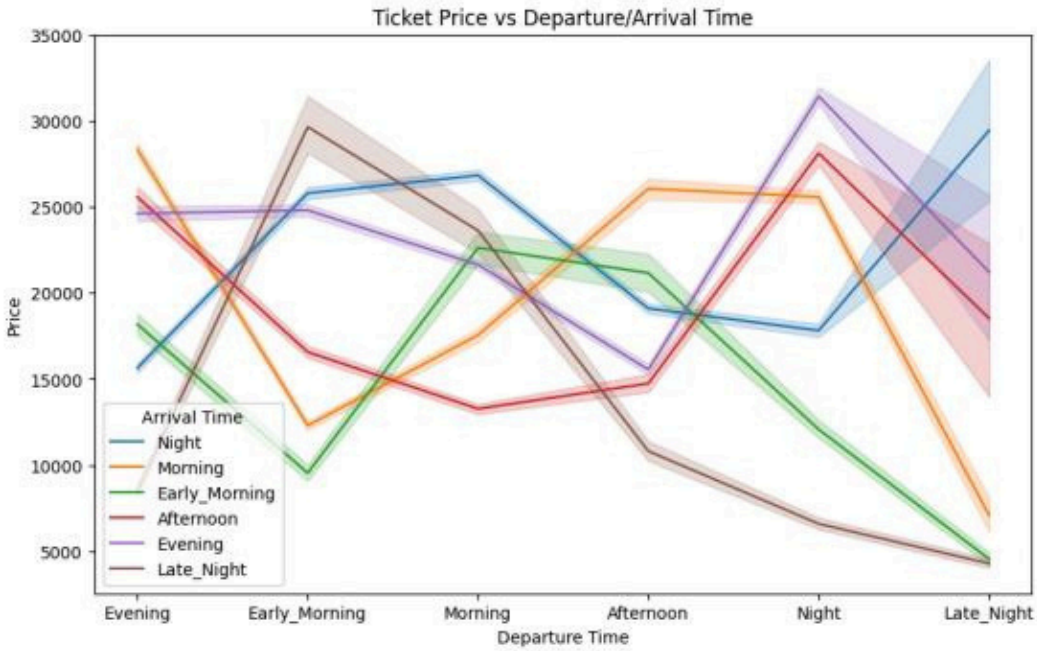

```
plt.figure(figsize=(10, 8))
sns.countplot(x='source_city', hue='destination_city', data=df)
plt.title('Source City vs Destination City')
plt.xlabel('Source City')
plt.ylabel('Count')
plt.legend(title='Destination City')
plt.show()
```



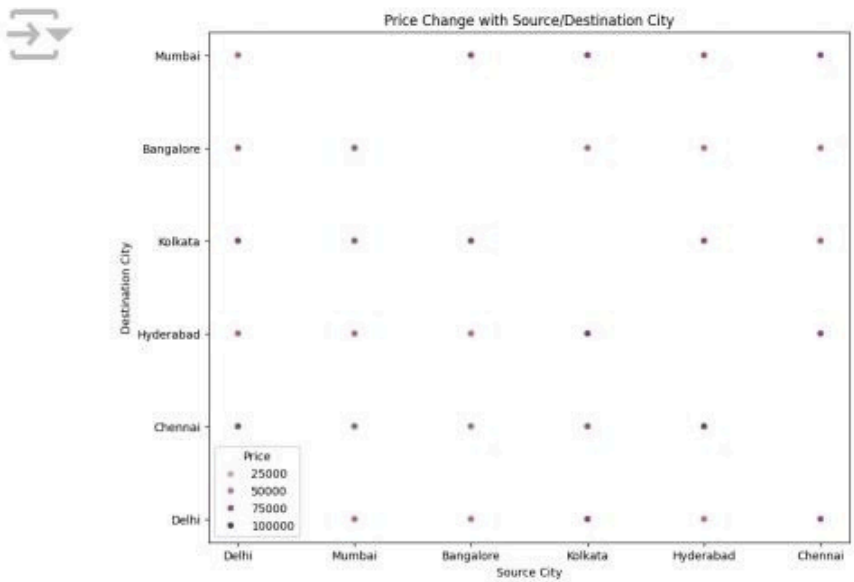

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='airline', y='price', data=df)
plt.title('Price Variation with Airlines')
plt.xlabel('Airline')
plt.ylabel('Price')
plt.show()
```



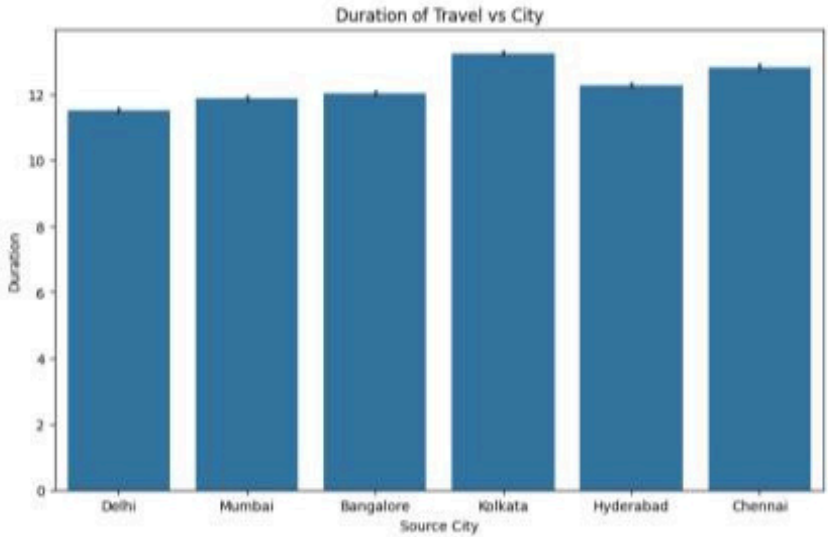

```
plt.figure(figsize=(10, 6))
sns.lineplot(x='departure_time', y='price', hue='arrival_time', data=df)
plt.title('Ticket Price vs Departure/Arrival Time')
plt.xlabel('Departure Time')
plt.ylabel('Price')
plt.legend(title='Arrival Time')
plt.show()
```


Ticket Price vs Departure/Arrival Time
35000
30000
25000
20000
15000
10000
5000
Price
Arrival Time
Night
Morning
Early_Morning
Afternoon
Evening
Late_Night
Evening
Early_Morning
Morning
Afternoon
Night
Late_Night
Departure Time

```
plt.figure(figsize=(10, 8))
sns.scatterplot(x='source_city', y='
plt.title('Price Change with Source/
plt.xlabel('Source City')
plt.ylabel('Destination City')
plt.legend(title='Price')
plt.show()
```




```
plt.figure(figsize=(10, 6))
sns.barplot(x='source_city', y='dura
plt.title('Duration of Travel vs Cit
plt.xlabel('Source City')
plt.ylabel('Duration')
plt.show()
```




```
high_price_flights = df[df['price']
plt.figure(figsize=(10,6))
```

```
<Figure size 1000x600 with 0
Axes>
<Figure size 1000x600 with 0
Axes>
```