In [1]:
```
import os
```

In [2]:
```
os.getcwd()
```

Out[2]:
```
'C:\\Users\\Rahul'
```

In [3]:
```
os.chdir("D:\\revolve assignment\\data")
```

In [4]:
```
os.getcwd()
```

Out[4]:
```
'D:\\revolve assignment\\data'
```

In [5]:
```
import pandas as pd
```

In [6]:
```
df=pd.read_csv('flights.csv')
```

how many total number of days does the flights table cover?

In [8]:
```
Total = df['day'].count()
```

In [9]:
```
print(Total)
```

```
336776
```

how many departure cities (not airports) does the flights database cover?

In [23]:
```
DepartureCity=df["origin"].value_counts()
```

In [24]:
```
print(DepartureCity)
```

```
EWR    120835
JFK    111279
LGA    104662
Name: origin, dtype: int64
```

In [13]:
```
df1= pd.read_csv('planes.csv')
```

what is the relationship between flights and planes tables?

tailnum helps to relate flghts and planes tables

In [14]:
```
df2 = pd.merge(df, df1)
```

In [20]:
```
df2
```

Out[20]:

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	...	distance	hour	minute	type	manufacturer	model	engines	seats	speed	engine
0	2013	1	18	1846.0	36.0	2156.0	36.0	UA	N37465	1292	...	1065	18.0	46.0	Fixed wing multi engine	BOEING	737-924ER	2	191	NaN	Turbo-fan
1	2013	10	3	1257.0	0.0	1544.0	-18.0	UA	N37465	1158	...	1085	12.0	57.0	Fixed wing multi engine	BOEING	737-924ER	2	191	NaN	Turbo-fan
2	2013	10	3	2058.0	-1.0	2323.0	-35.0	UA	N37465	1416	...	937	20.0	58.0	Fixed wing multi engine	BOEING	737-924ER	2	191	NaN	Turbo-fan
3	2013	10	4	1003.0	3.0	1300.0	-6.0	UA	N37465	1735	...	2454	10.0	3.0	Fixed wing multi engine	BOEING	737-924ER	2	191	NaN	Turbo-fan
4	2013	10	7	1926.0	81.0	2123.0	55.0	UA	N37465	1139	...	1605	19.0	26.0	Fixed wing multi engine	BOEING	737-924ER	2	191	NaN	Turbo-fan
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4625	2013	4	22	624.0	-1.0	758.0	13.0	WN	N8611F	3493	...	725	6.0	24.0	Fixed wing multi engine	BOEING	737-8H4	2	140	NaN	Turbo-fan
4626	2013	4	28	1023.0	43.0	1144.0	29.0	WN	N8611F	367	...	738	10.0	23.0	Fixed wing multi engine	BOEING	737-8H4	2	140	NaN	Turbo-fan
4627	2013	5	17	624.0	-1.0	736.0	-9.0	WN	N8611F	3493	...	725	6.0	24.0	Fixed wing multi engine	BOEING	737-8H4	2	140	NaN	Turbo-fan
4628	2013	5	21	644.0	19.0	752.0	7.0	WN	N8611F	3493	...	725	6.0	44.0	Fixed wing multi engine	BOEING	737-8H4	2	140	NaN	Turbo-fan
4629	2013	7	24	611.0	1.0	722.0	-3.0	WN	N8611F	273	...	725	6.0	11.0	Fixed wing multi engine	BOEING	737-8H4	2	140	NaN	Turbo-fan

4630 rows × 23 columns

which are the two most connected cities?

In [24]:
```
df.groupby(['origin','dest']).size().idxmax()
```

Out[24]:
```
('JFK', 'LAX')
```

which airplane manufacturer incurred the most delays in the analysis period?

In [21]:
```
df2.groupby('manufacturer')['dep_delay'].sum().nlargest(1)
```

Out[21]:
```
manufacturer
BOEING    19505.0
Name: dep_delay, dtype: float64
```

In [ ]:
```

```