

```
In [1]: import os

In [4]: os.getcwd()

Out[4]: 'C:\\Users\\Rahul'

In [6]: os.chdir("D:\\revolve assignment\\data")

In [7]: os.getcwd()

Out[7]: 'D:\\revolve assignment\\data'

In [8]: import pandas as pd

In [9]: df=pd.read_csv('flights.csv')

In [12]: print(df)
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	\
0	2013	1	1	517.0	2.0	830.0	11.0	UA	
1	2013	1	1	533.0	4.0	850.0	20.0	UA	
2	2013	1	1	542.0	2.0	923.0	33.0	AA	
3	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	
4	2013	1	1	554.0	-6.0	812.0	-25.0	DL	
...	
336771	2013	9	30	NaN	NaN	NaN	NaN	9E	
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	
336773	2013	9	30	NaN	NaN	NaN	NaN	MQ	
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	
	tailnum	flight	origin	dest	air_time	distance	hour	minute	
0	N14228	1545	EWR	IAH	227.0	1400	5.0	17.0	
1	N24211	1714	LGA	IAH	227.0	1416	5.0	33.0	
2	N619AA	1141	JFK	MIA	160.0	1089	5.0	42.0	
3	N804JB	725	JFK	BQN	183.0	1576	5.0	44.0	
4	N668DN	461	LGA	ATL	116.0	762	5.0	54.0	
...	
336771	NaN	3393	JFK	DCA	NaN	213	NaN	NaN	
336772	NaN	3525	LGA	SYR	NaN	198	NaN	NaN	
336773	N535MQ	3461	LGA	BNA	NaN	764	NaN	NaN	
336774	N511MQ	3572	LGA	CLE	NaN	419	NaN	NaN	
336775	N839MQ	3531	LGA	RDU	NaN	431	NaN	NaN	
[336776 rows x 16 columns]									

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

```
In [14]: Total = df['day'].sum()

In [15]: print(Total)

5291016
```

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

```
In [16]: DepartureCity=df["origin"].value_counts()

In [26]: print(DepartureCity)

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

In [18]: df1= pd.read_csv('planes.csv')
```

what is the relationship between flights and planes tables?

```
In [19]: 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 x 23 columns

```
In [22]: maximum = df2['dep_delay'].max()

print(maximum)

411.0
```

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

```
In [23]: df2[df2['dep_delay']==df2['dep_delay'].max()]

Out[23]:
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	...	distance	hour	minute		type	manufacturer	model	engines	seats	speed	engine
336	2013	8	3	1521.0	411.0	1752.0	403.0	B6	N355JB	227	...	937	15.0	21.0	Fixed wing multi engine	EMBRAER	ERJ 190-100 IGW	2	20	NaN	Turbo-fan	

1 rows x 23 columns

which are the two most connected cities?

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

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

In [ ]:
```