

```
In [76]: import pandas as pd
```

```
In [77]: df1 = pd.read_csv("dataRevolve/flights.csv")
```

```
In [78]: df1
```

```
Out[78]:
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
0	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR
1	2013	1	1	533.0	4.0	850.0	20.0	UA	N24211	1714	LGA
2	2013	1	1	542.0	2.0	923.0	33.0	AA	N619AA	1141	JFK
3	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	N804JB	725	JFK
4	2013	1	1	554.0	-6.0	812.0	-25.0	DL	N668DN	461	LGA
...
336771	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3393	JFK
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3525	LGA
336773	2013	9	30	NaN	NaN	NaN	NaN	MQ	N535MQ	3461	LGA
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	N511MQ	3572	LGA
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	N839MQ	3531	LGA

336776 rows × 16 columns



```
In [79]: df1["date"] = pd.to_datetime((df1.year*10000+df1.month*100+df1.day), format = "%Y%m%d")
```

```
In [80]: df1
```

```
Out[80]:
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
0	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR
1	2013	1	1	533.0	4.0	850.0	20.0	UA	N24211	1714	LGA
2	2013	1	1	542.0	2.0	923.0	33.0	AA	N619AA	1141	JFK
3	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	N804JB	725	JFK
4	2013	1	1	554.0	-6.0	812.0	-25.0	DL	N668DN	461	LGA
...
336771	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3393	JFK

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3525	LGA
336773	2013	9	30	NaN	NaN	NaN	NaN	MQ	N535MQ	3461	LGA
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	N511MQ	3572	LGA
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	N839MQ	3531	LGA

336776 rows × 17 columns



```
In [91]: print(df1.date.value_counts().count(), 'is the total number of days covered in the data')
365 is the total number of days covered in the dataset
```

365 is the total number of days covered in the dataset

```
In [ ]:
```

```
In [ ]:
```

```
In [82]: df2 = pd.read_csv("dataRevolve/airports.csv")
```

```
In [83]: df2
```

Out[83]:	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
0	ABE	Lehigh Valley International Airport	Allentown	PA	USA	40.65236	-75.44040
1	ABI	Abilene Regional Airport	Abilene	TX	USA	32.41132	-99.68190
2	ABQ	Albuquerque International Sunport	Albuquerque	NM	USA	35.04022	-106.60919
3	ABR	Aberdeen Regional Airport	Aberdeen	SD	USA	45.44906	-98.42183
4	ABY	Southwest Georgia Regional Airport	Albany	GA	USA	31.53552	-84.19447
...
317	WRG	Wrangell Airport	Wrangell	AK	USA	56.48433	-132.36982

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
318	WYS	Westerly State Airport	West Yellowstone	MT	USA	44.68840	-111.11764
319	XNA	Northwest Arkansas Regional Airport	Fayetteville/Springdale/Rogers	AR	USA	36.28187	-94.30681
320	YAK	Yakutat Airport	Yakutat	AK	USA	59.50336	-139.66023
321	YUM	Yuma International Airport	Yuma	AZ	USA	32.65658	-114.60597

322 rows × 7 columns

```
In [92]: print(df2.CITY.value_counts().count(), 'is the total number of city')
308 is the total number of city
```

308 is the total number of city

```
In [ ]:
```

```
In [ ]:
```

```
In [85]: df3 = pd.read_csv('dataRevolve/planes.csv')
```

```
In [86]: df3
```

Out[86]:

	tailnum	year	type	manufacturer	model	engines	seats	speed	engine
0	N10156	2004.0	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NaN	Turbo-fan
1	N102UW	1998.0	Fixed wing multi engine	AIRBUS INDUSTRIE	A320-214	2	182	NaN	Turbo-fan
2	N103US	1999.0	Fixed wing multi engine	AIRBUS INDUSTRIE	A320-214	2	182	NaN	Turbo-fan
3	N104UW	1999.0	Fixed wing multi engine	AIRBUS INDUSTRIE	A320-214	2	182	NaN	Turbo-fan
4	N10575	2002.0	Fixed wing multi engine	EMBRAER	EMB-145LR	2	55	NaN	Turbo-fan
...
3317	N997AT	2002.0	Fixed wing multi engine	BOEING	717-200	2	100	NaN	Turbo-fan
3318	N997DL	1992.0	Fixed wing multi engine	MCDONNELL DOUGLAS AIRCRAFT CO	MD-88	2	142	NaN	Turbo-fan

	tailnum	year	type	manufacturer	model	engines	seats	speed	engine
3319	N998AT	2002.0	Fixed wing multi engine	BOEING	717-200	2	100	NaN	Turbo-fan
3320	N998DL	1992.0	Fixed wing multi engine	MCDONNELL DOUGLAS CORPORATION	MD-88	2	142	NaN	Turbo-jet
3321	N999DN	1992.0	Fixed wing multi engine	MCDONNELL DOUGLAS CORPORATION	MD-88	2	142	NaN	Turbo-jet

3322 rows × 9 columns

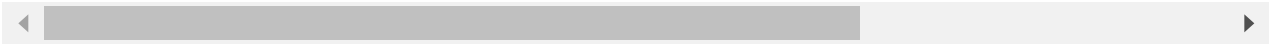
In [87]:

df1

Out[87]:

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
0	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR
1	2013	1	1	533.0	4.0	850.0	20.0	UA	N24211	1714	LGA
2	2013	1	1	542.0	2.0	923.0	33.0	AA	N619AA	1141	JFK
3	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	N804JB	725	JFK
4	2013	1	1	554.0	-6.0	812.0	-25.0	DL	N668DN	461	LGA
...
336771	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3393	JFK
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3525	LGA
336773	2013	9	30	NaN	NaN	NaN	NaN	MQ	N535MQ	3461	LGA
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	N511MQ	3572	LGA
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	N839MQ	3531	LGA

336776 rows × 17 columns



we can see that the both columns contain a same column, tailnum it is the number of the plane's tail number and the flights datasets indicates which plane is going to which destination at what time and date

we can merge the data on that column to see the details of the flying plane with the tailnum

```
In [103... df1[df1['dep_delay']>0].tailnum.value_counts()
```

```
Out[103... N258JB    186
N228JB    165
N15980    158
N190JB    157
N725MQ    152
...
N363SW      1
N7ASAA      1
N646UA      1
N59630      1
N512SW      1
Name: tailnum, Length: 3885, dtype: int64
```

here we can see that the N258JB tailnum has the most delay now we have to find its manufacturer

```
In [ ]:
```

```
In [ ]:
```

```
In [109... df3[df3['tailnum']=='N258JB'].manufacturer
```

```
Out[109... 543    EMBRAER
Name: manufacturer, dtype: object
```

EMBRAER manufacturer is the most delays in the analysis period

```
In [ ]:
```

```
In [ ]:
```

```
In [110... df1.dest.value_counts()
```

```
Out[110... ORD    17283
ATL    17215
LAX    16174
BOS    15508
MCO    14082
...
MTJ      15
SBN      10
ANC       8
LGA       1
LEX       1
Name: dest, Length: 105, dtype: int64
```

```
In [111... df1
```

```
Out[111...   year  month  day  dep_time  dep_delay  arr_time  arr_delay  carrier  tailnum  flight  origin
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
0	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR
1	2013	1	1	533.0	4.0	850.0	20.0	UA	N24211	1714	LGA
2	2013	1	1	542.0	2.0	923.0	33.0	AA	N619AA	1141	JFK
3	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	N804JB	725	JFK
4	2013	1	1	554.0	-6.0	812.0	-25.0	DL	N668DN	461	LGA
...
336771	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3393	JFK
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3525	LGA
336773	2013	9	30	NaN	NaN	NaN	NaN	MQ	N535MQ	3461	LGA
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	N511MQ	3572	LGA
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	N839MQ	3531	LGA

336776 rows × 17 columns



In [130...

df2[df2['IATA_CODE']=='EWR']

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
106	EWR	Newark Liberty International Airport	Newark	NJ	USA	40.6925	-74.16866

In [131...

df1[df1['origin']=='EWR']

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
0	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR
5	2013	1	1	554.0	-4.0	740.0	12.0	UA	N39463	1696	EWR
6	2013	1	1	555.0	-5.0	913.0	19.0	B6	N516JB	507	EWR
13	2013	1	1	558.0	-2.0	923.0	-14.0	UA	N53441	1124	EWR

3/16/22, 11:07 PM

REVOLVE_ASSIGNMENT

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin
16	2013	1	1	559.0	-1.0	854.0	-8.0	UA	N76515	1187	EWR
...
336752	2013	9	30	2142.0	13.0	2250.0	11.0	EV	N12957	4509	EWR
336755	2013	9	30	2149.0	-7.0	2245.0	-23.0	UA	N813UA	523	EWR
336756	2013	9	30	2150.0	-9.0	2250.0	-16.0	EV	N10575	3842	EWR
336760	2013	9	30	2211.0	72.0	2339.0	57.0	EV	N12145	4672	EWR
336762	2013	9	30	2233.0	80.0	112.0	42.0	UA	N578UA	471	EWR

120835 rows × 17 columns



In [119...

df2

Out[119...

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
0	ABE	Lehigh Valley International Airport	Allentown	PA	USA	40.65236	-75.44040
1	ABI	Abilene Regional Airport	Abilene	TX	USA	32.41132	-99.68190
2	ABQ	Albuquerque International Sunport	Albuquerque	NM	USA	35.04022	-106.60919
3	ABR	Aberdeen Regional Airport	Aberdeen	SD	USA	45.44906	-98.42183
4	ABY	Southwest Georgia Regional Airport	Albany	GA	USA	31.53552	-84.19447
...
317	WRG	Wrangell Airport	Wrangell	AK	USA	56.48433	-132.36982
318	WYS	Westerly State Airport	West Yellowstone	MT	USA	44.68840	-111.11764

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
319	XNA	Northwest Arkansas Regional Airport	Fayetteville/Springdale/Rogers	AR	USA	36.28187	-94.30681
320	YAK	Yakutat Airport	Yakutat	AK	USA	59.50336	-139.66023
321	YUM	Yuma International Airport	Yuma	AZ	USA	32.65658	-114.60597

322 rows × 7 columns

```
In [117... df2.IATA_CODE.value_counts()
```

```
Out[117... ISN      1
MKE      1
ISP      1
CLD      1
LAR      1
...
PPG      1
FAT      1
SHV      1
BNA      1
BTV      1
Name: IATA_CODE, Length: 322, dtype: int64
```

```
In [128... df1.value_counts(df1['dest'])
```

```
Out[128... dest
ORD      17283
ATL      17215
LAX      16174
BOS      15508
MCO      14082
...
HDN       15
SBN       10
ANC        8
LGA        1
LEX        1
Length: 105, dtype: int64
```

```
In [129... df1.value_counts(df1['origin'])
```

```
Out[129... origin
EWR      120835
JFK      111279
LGA      104662
dtype: int64
```

```
In [138... df1.groupby(['dest', 'origin']).size().sort_values()
```

```
Out[138... dest origin
MEM  JFK      1
LGA  EWR      1
BHM  JFK      1
STL  JFK      1
```



```
LEX  LGA      1
...
CLT  LGA     6168
SFO  JFK     8204
ORD  LGA     8857
ATL  LGA    10263
LAX  JFK    11262
Length: 224, dtype: int64
```

Here we can see LAX and JFK has the most number of flights which indicates there are the most connected

```
In [140... df2[df2['IATA_CODE']=='LAX']
```

Out[140...

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
176	LAX	Los Angeles International Airport	Los Angeles	CA	USA	33.94254	-118.40807

```
In [141... df2[df2['IATA_CODE']=='JFK']
```

Out[141...

	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
166	JFK	John F. Kennedy International Airport (New Yor...	New York	NY	USA	40.63975	-73.77893

NEW YORK and LOS ANGELES are the two most connected cities

```
In [ ]:
```