PANDAS

DATA ANALYSIS ASSIGNMENT 2

FLIGHTS FROM NEW YORK CITY AIRPORTS IN 2013

PART ONE: QUESTIONS 1 - 10

LINK: [NEW YORK CITY AIRPORTS] (https://www.ny.com/transportation/airports/)

[Link: Pandas Documentation]

(https://pandas.pydata.org/docs/)

In [1]:
```
# set up notebook to display multiple output in one cell

from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

print('The notebook is set up to display multiple output in one cell.')
```

```
The notebook is set up to display multiple output in one cell.
```

In [4]:
```
import pandas as pd
import numpy as np
```

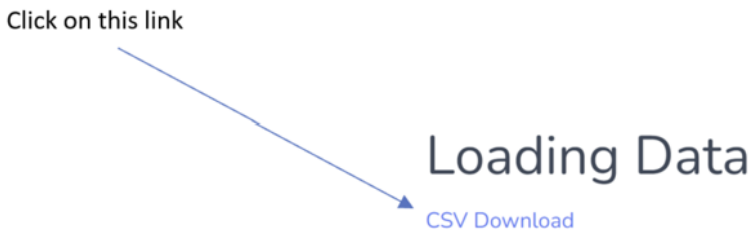
Files needed for this assignment:

For this assignment, read in the **nycflights.csv** dataset.

Data Source: **Wickham H. 2014. nycflights13: Data about flights departing NYC in 2013. R package version 0.1.**

Hadley Wickham Bio: **Hadley Wickham**

Link for Dataset: **nycflights.csv**



Variables

- **year**: Year.
- **month**: Month.
- **day**: Day.
- **dep_time**: Departure time, in Eastern time zone.
- **dep_delay**: Departure delay, in minutes.
- **arr_time**: Arrival time, in the local time zone.
- **arr_delay**: Arrival delay, in minutes.
- **carrier**: Carrier, abbreviated.
- **tailnum**: Tail number of the airplane.
- **flight**: Flight number.
- **origin**: Flight origin, airport code.
- **dest**: Flight destination, airport code.
- **air_time**: Time in the air, in minutes.
- **distance**: Distance between the departure and arrival airports, in miles.
- **hour**: Scheduled departure hour.
- **minute**: Scheduled departure minute.

In [10]:
```
flights = pd.read_csv("nycflights.csv")
flights.info()
flights.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32735 entries, 0 to 32734
Data columns (total 16 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   year       32735 non-null  int64
 1   month      32735 non-null  int64
 2   day        32735 non-null  int64
 3   dep_time   32735 non-null  int64
 4   dep_delay  32735 non-null  int64
 5   arr_time   32735 non-null  int64
 6   arr_delay  32735 non-null  int64
 7   carrier    32735 non-null  object
 8   tailnum    32735 non-null  object
 9   flight     32735 non-null  int64
 10  origin     32735 non-null  object
 11  dest       32735 non-null  object
 12  air_time   32735 non-null  int64
 13  distance   32735 non-null  int64
 14  hour       32735 non-null  int64
 15  minute     32735 non-null  int64
dtypes: int64(12), object(4)
memory usage: 4.0+ MB
```

Out[10]:

```
year         0
month        0
day          0
dep_time     0
dep_delay    0
arr_time     0
arr_delay    0
carrier      0
tailnum      0
flight       0
origin       0
dest         0
air_time     0
distance     0
hour         0
minute       0
dtype: int64
```

Note: Before answering the questions below, use appropriate attributes and methods to inspect the data.

Question 1: How many flights were there from NYC airports to Miami in 2013?

In [22]:

```
flights[flights.dest =="MIA"]
```

Out[22]:

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin	d
8	2013	9	26	725	-10	1027	-8	AA	N3FSAA	2279	LGA	
87	2013	9	28	1652	32	2013	33	AA	N3DYAA	1410	LGA	
92	2013	7	4	1125	0	1411	-29	AA	N3BNAA	2099	LGA	
104	2013	6	11	556	-9	911	1	AA	N3EXAA	1837	LGA	
139	2013	11	24	1920	-5	2218	-27	AA	N3FHAA	2437	LGA	
...	...	...	...	...	...	...	...	...	...	...	...	...
32627	2013	1	20	1554	-6	1910	-23	DL	N947DL	161	JFK	
32673	2013	4	19	1805	216	2116	213	DL	N318US	2175	LGA	
32674	2013	4	3	1004	9	1301	-9	AA	N3GEAA	1871	LGA	
32679	2013	8	21	812	-8	1122	-13	AA	N3KJAA	2267	LGA	
32700	2013	6	3	1420	22	1804	49	DL	N366NW	1331	LGA	

1220 rows × 16 columns

Question 2: What was the most frequent destination for flights from NYC airports in 2013?

In [26]:
```
flights.dest.value_counts()
```

Out[26]:
```
ATL    1653
ORD    1588
LAX    1583
BOS    1470
CLT    1383
       ...
JAC       2
PSP       2
CHO       1
EYW       1
ANC       1
Name: dest, Length: 102, dtype: int64
```

Question 3: Which New York City airport had the most flight departures in 2013?

In [27]:
```
flights.origin.value_counts()
```

Out[27]:
```
EWR    11771
JFK    10897
LGA    10067
Name: origin, dtype: int64
```

Question 4: How many airlines flew from NYC to LAX in 2013?

In [63]:
```
#Carrier is the airline while tailnumber is the specific aircraft(also gives origin of
flights[flights.dest == 'LAX'].carrier.value_counts()
```

Out[63]:
```
UA    553
AA    374
DL    262
VX    235
B6    159
Name: carrier, dtype: int64
```

Question 5: How many unique air planes flew from NYC to Atlanta in 2013?

In [41]:
```
flights[flights.dest == 'ATL'].carrier.value_counts()
```

Out[41]:
```
DL    1049
FL     212
MQ     201
EV     167
UA      14
WN       5
9E       5
Name: carrier, dtype: int64
```

Question 6: What was the average arrival delay for flights from NC to Chicago O'Hare International Airport (ORD) in 2013?

In [43]:
```
flights[flights.dest == 'ORD'].arr_delay.mean()
```

Out[43]:
```
6.663098236775818
```

Question 7: What proportion of flights to Chicago O'Hare International Airport (ORD) in 2013 come from each NYC airport?

In [44]:
```
flights[flights.dest == 'ORD'].origin.max()
```

Out[44]:
```
'LGA'
```

Question 8: What was the longest departure delay for a flight from a New York City airport in 2013? On which day did that delay occur?

In [56]:
```
flights.dep_delay.max()
flights[flights.dep_delay == flights.dep_delay.max()]
print("The delay happened on January 9th.")
```

Out[56]:
```
1301
```

Out[56]:

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin	d
30381	2013	1	9	641	1301	1242	1272	HA	N384HA	51	JFK	H

```
The delay happened on January 9th.
```

Question 9: What was the longest arrival delay for a flight from a New York City airport in 2013? What airport was that flight to?

In [59]:
```
flights.arr_delay.max()
flights[flights.arr_delay == flights.arr_delay.max()].dest
print("Airport in Hawaii")
```

Out[59]:
```
1272
```

Out[59]:
```
30381    HNL
Name: dest, dtype: object
Airport in Hawaii
```

Question 10: Which carrier had the most flights from a New York City airport to Milwaukee's General Mitchell International Airport (MKE) in 2013?

In [60]:
```
flights[flights.dest == 'MKE'].carrier.max()
```

Out[60]:
```
'WN'
```

In [ ]: