PANDAS

DATA ANALYSIS ASSIGNMENT 2

FLIGHTS FROM NEW YORK CITY AIRPORTS IN 2013

PART Two: QUESTIONS 11 - 20

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

[Link: Pandas Documentation]

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

In [91]:
```
# 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 [92]:
```
import pandas as pd
import numpy as np
```

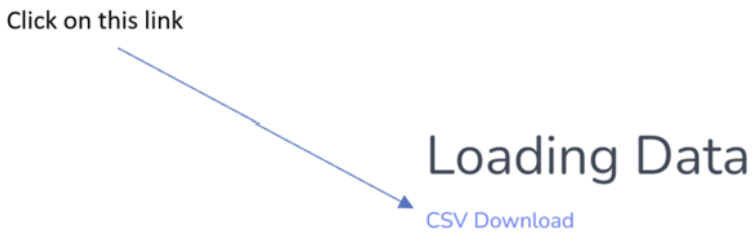
Files needed for this assignment:

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 [93]:
```
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[93]:

```
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 11: What is the most common departure time from NYC airports to MKE?

In [94]:

```
flights[flights.dest=='MKE'].dep_time.value_counts()
```

Out[94]:
```
1252    4
940     4
945     4
655     4
1255    3
       ..
1619    1
614     1
1301    1
1054    1
604     1
Name: dep_time, Length: 217, dtype: int64
```

Question 12: Which carrier has the longest delay in arrival time?

In [95]:
```
flights[flights.arr_delay==flights.arr_delay.max()].carrier
```

Out[95]:
```
30381    HA
Name: carrier, dtype: object
```

Question 13: Which carrier has the longest delay in departure time?

In [96]:
```
flights[flights.dep_delay==flights.dep_delay.max()].carrier
```

Out[96]:
```
30381    HA
Name: carrier, dtype: object
```

Question 14: Which airplane (tail number) has the greatest delay?

In [97]:
```
flights[flights.dep_delay==flights.dep_delay.max()].tailnum
```

Out[97]:
```
30381    N384HA
Name: tailnum, dtype: object
```

Question 15: Does a specific carrier specialize in longer flights?

In [98]:
```
#flights.distance.describe(percentiles=[.1, .25,.50,.75, .95])
flights[flights.distance >= 2475].carrier.value_counts()
```

Out[98]:
```
UA    909
AA    476
DL    467
VX    386
B6    336
HA     34
Name: carrier, dtype: int64
```

Question 16: Which airline carrier is the fastest? (**Use the distance and airtime variables.**)

In [99]:
```
flights['DistancePerMin'] = flights.distance / flights.air_time
flights.sort_values('DistancePerMin').max().carrier
```

Out[99]:
```
'YV'
```

Question 17: What are the top 5 busiest days of the year?

In [100]:
```
datedf = pd.DataFrame({'year': 2013,
                       'month': flights.month,
                       'day': flights.day})
flights['Date'] = pd.to_datetime(datedf)
flights.Date.value_counts().head()
```

Out[100]:
```
2013-05-13    121
2013-08-21    118
2013-11-14    116
2013-10-23    115
2013-05-09    115
Name: Date, dtype: int64
```

Question 18: What is the busiest day of the week with respect to trafic in and out of New York City airports?

In [101]:
```
DayOfWeekMap = {
    0 : 'Monday',
    1 : 'Tuesday',
    2 : 'Wednesday',
    3 : 'Thursday',
    4 : 'Friday',
    5 : 'Saturday',
    6 : 'Sunday'
}
flights['Date'].dt.dayofweek.map(DayOfWeekMap).value_counts()
```

Out[101]:
```
Monday       4962
Tuesday      4943
Friday       4855
Wednesday    4830
Thursday     4794
Sunday       4632
Saturday     3719
Name: Date, dtype: int64
```

Question 19: What is the most popular destination for each month? (Seasonality)

In [102]:
```
months = {
        1:'January',
        2:'February',
        3:'March',
        4:'April',
        5:'May',
        6:'June',
        7:'July',
        8:'August',
        9:'September',
        10:'October',
        11:'November',
```

```
        12:'December'
}
flights['month'] = flights['month'].map(months)
```

In [122]:
```
grouped = flights.groupby('month')
for month in months.values():
    print(f"{month}:  {grouped.get_group(month).dest.value_counts().head(1).to_string()
```

```
January:  ATL    137
February:  ATL    125
March:  BOS    133
April:  LAX    148
May:  ORD    151
June:  ATL    160
July:  ORD    148
August:  ATL    159
September:  LAX    159
October:  ORD    152
November:  ORD    156
December:  ATL    144
```

Question 20: What is the average departure delay time for each airport?

In [132]:
```
grouped = flights.groupby('origin')
for group in grouped.groups:
    print(f"{group}: {flights[flights.origin == group].dep_delay.mean():.2f}")
```

```
EWR: 15.31
JFK: 12.27
LGA: 10.13
```