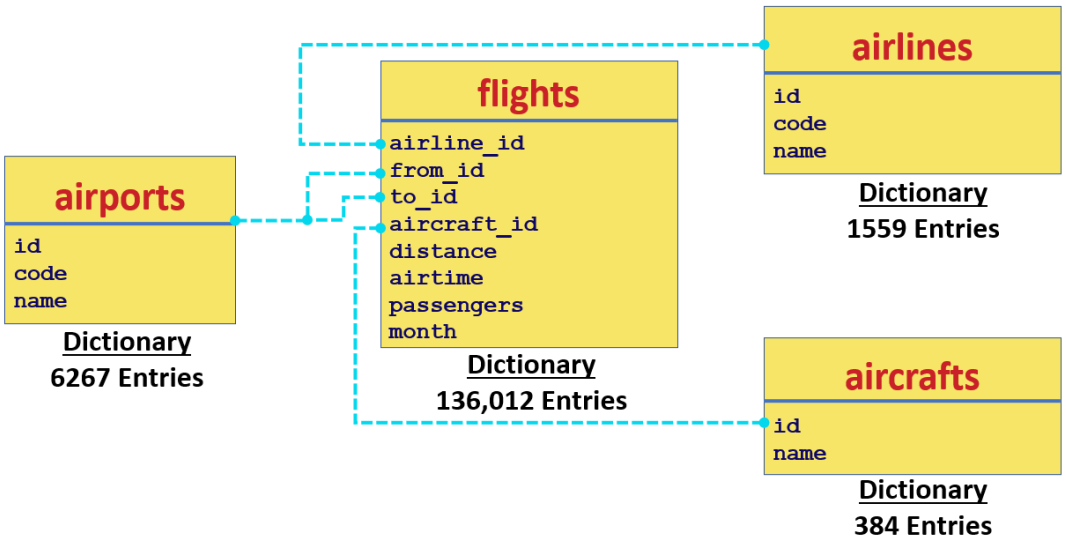
Assignment # 01

Introduction

The subject of your project is US domestic air traffic. The data required for this assignment is available in four csv files. Specifically, it has a portion of the US domestic traffic data from January 2013 to May 2013. This data is made available by the US Bureau of Transportation Statistics.

The Database Parts:

The four csv files represent four datasets, and the information connected between the four datasets is shown below and explained in next section. There is a starting code provided in next section to load the csv files which will be loaded as a dictionary. Notice that there are over 136,000 flights, 300 aircraft types, 1500 airlines, and 6000 airports in the database.



Loading the Database

There are four csv files, and you should first load and view the structure of data inside those. For that, just put the csv files inside the working folder and run the following code to view size and one item of each of the four files. Important point to note is that all data values will be loaded as string though they represent some numeric value.

```
import csv
with open('aircrafts.csv','r') as f:
    r=csv.DictReader(f)
    airCrafts=list(r)
print(len(airCrafts))
print(airCrafts[0])
```

```
with open('airlines.csv','r') as f:
    r=csv.DictReader(f)
    airLines=list(r)
print(len(airLines))
print(airLines[0])

with open('airports.csv','r') as f:
    r=csv.DictReader(f)
    airPorts=list(r)
print(len(airPorts))
print(airPorts[0])

with open('flights.csv','r') as f:
    r=csv.DictReader(f)
    flights=list(r)
print(len(flights))
print(flights[0])
```

Each csv file data is explained here:

The data of **flights.cvs** file when converted to a list of dictionaries, contains one **dictionary** for each flight segment recorded in the database. A flight segment represents monthly air travel between two points (i.e., from one airport to another airport) by a unique carrier (i.e., an airline) using one type of aircraft. **The data for a flight segment can represent one or more actual flights.** For example, if American Airlines flies from Nashville to Miami, one flight segment in **flights** will represent all flights ***from*** Nashville ***to*** Miami ***by*** American Airlines ***using*** a particular aircraft type ***during*** a particular month.

You can view the detail of any flight and one example is shown here:

```
print(flights[58344])
```

The output will be:

```
{'airline_id': '138', 'from_id': '699', 'to_id': '3285', 'aircraft_id':
'228', 'distance': '806', 'airtime': '3316', 'passengers': '3825',
'month': '3'}
```

As you can see, each flight segment has an **airline_id**, **from_id**, and **to_id**. Also, the distance between airports (in miles), the total airtime of all associated flights (in minutes), the number of passengers transported, and the month of reporting is given.

For the example above, the distance between the two airports is 806 miles. The number of passengers transported during the month of March was 3,825 passengers. The total airtime of all corresponding flights was 3,316 minutes.

The **airline_id**, **from_id**, **to_id**, and **aircraft_id** keys of the dictionaries in the **flights** list are important: they are IDs of the given airline, airports (both origin and destination airports), and

aircraft type, respectively. These numbers are used to identify the airline, airports, and aircraft type in the other three csv files (i.e., **aircraft**, **airlines**, and **airports**) of the database.

For the flight segment above, the associated airline is "American Airlines Inc." This airline is also known with the airline code "AA." And the detail of the airline can be viewed through the ID **138** which is same as the index of list of airlines.

```
print(airLines[138])
```

The output will be:

```
{'id': '138', 'code': 'AA', 'name': 'American Airlines Inc.'}
```

Also, the flight segment originated from "Nashville, TN: Nashville International." This airport is also known with the airport code "BNA". Again, the detail of the airport can be viewed through the ID **699** which is same as the index of list of airports.

```
print(airPorts[699])
```

The output will be:

```
{'id': '699', 'code': 'BNA', 'name': 'Nashville, TN: Nashville International'}
```

And, the flight segment flew to "Miami, FL: Miami International," also known as "MIA" that can be viewed through the ID **3285** which is same as the index of list of airports.

```
print(airPorts[3285])
```

The output will be:

```
{'Code': 'MIA', 'Name': 'Miami, FL: Miami International'}
```

Finally, the aircraft that was used was a "Boeing 737-800." The detail of the aircraft can be viewed through the ID **228** which is same as the index of list of aircraft.

```
print(airCrafts[228])
```

```
{'id': '228', 'name': 'Boeing 737-800'}
```

An Example (Helper) Function

To illustrate how to use the information in the database, consider the following function:

```
def findAirportByName(name_check):
    for a in airPorts:
        if a['name']==name_check:
```

```
            return a['id']
    return -1
```

This function finds the airport in the database with the matching name. If there is no matching airport, the function returns -1. For example, if you want to find the record for "Memphis, TN: Memphis International," you should just type:

```
print(findAirportByName('Memphis, TN: Memphis International'))
3226
```

Now you know that the ID for the Memphis International Airport is **3226** and to make sure you can simply type:

```
print(airPorts[3226])
```

```
{'id': '3226', 'code': 'MEM', 'name': 'Memphis, TN: Memphis International'}
```

You are free, and in fact, encouraged, to create similar functions of your own and call them from your code in order to make your code better organized and cleaner. Just remember that you must submit just **ONE PYTHON** file for all the problems assigned to you.

You should create a function for each of problem assigned to you with name **`problemN`** (where **`N`** is the problem number e.g., 1, 2, 15, etc.). And then use that in main program to get the desired result. You might observe that some of the problems can better be solved without creating a function, but to be consistent, in that case too, you should create a function, though without any input argument.

Problem List

1. How many flight segments originated from a particular airport? The function should consider that the input can be code or name of the airport. Use the function for *Nashville International Airport*. Show output both for (BNA) and (Nashville, TN: Nashville International) provided as input.
2. How many flight segments used a particular aircraft? The function should consider that the input can be code or name of the aircraft. Use the function for *Boeing 737-800*? Show output both for (228) and (Boeing 737-800) provided as input.
3. How many passengers were transported by the flight segment carrying the largest number of passengers?
4. How many miles was the longest flight segment?
5. Which airline had the most flight segments and how many?
6. How many total passengers were transported each month from January 2013 to May 2013? Provide a list with a total for each month.
7. Create a list of unique airlines that served a particular airport as a destination. Provide a list of IDs of the airline. The function should consider that the input can be code or name of the airport. Use the function for *Orlando International Airport.* Show output both for (MCO) and (Orlando, FL: Orlando International)
8. Create a list of unique aircraft types that were used to fly out of a particular airport. Provide a list of IDs of the aircrafts. The function should consider that the input can be code or name of the

airport. Use the function for *Seattle/Tacoma International Airport*. Show output both for (SEA) and (Seattle, WA: Seattle/Tacoma International)

9. Create a list of airlines that flew to two particular airports. Provide a list of IDs of the airline. The function should consider that the input can be code or name of the airports. Use the function for *New York, NY: John F. Kennedy International* and *Los Angeles International Airport*. Show three output; one with both airports specified by ID, one with both airports specified by name and third with one airport specified as ID and one as name.

10. Create a list of airports that had more than *1000* flight segments taking into account both segments coming in or going out of the airport. Provide a list of IDs of the airports.

11. How many flight segments did aircraft manufactured by *Airbus* and *Boeing* operate as respective groups? Provide a list containing two values; one total for *Airbus* and one total for *Boeing*.

12. Create a list of airlines that operated both the *Boeing 737-800* and *Boeing 737-900*. Provide a list of names of the airlines.

13. Create a list of airlines that are both in the top five in terms of passengers transported and total flight segments. Provide a list of names of the airlines.

14. Which airline had the most diverse fleet (i.e., the most number of different aircraft types). What aircraft types did the airline operate? Provide a dictionary with key as name of the airline (there can be more than one airline with most diverse fleet) and value as a list of IDs of the aircrafts.

15. Create a list of airlines that had a positive increase of passengers transported between each month from January 2013 to May 2013. Provide a list of IDs of the airlines.

16. Which airport was served by the most number of airlines? What airlines served this airport? Provide a dictionary with key as name of the airport (there can be more than one airports with most number of airlines) and value as a list of IDs of the airlines.

17. Which aircraft type logged the most airtime and which airline was responsible for the most airtime for this aircraft type? Provide the name of the aircraft, a list of names of the airlines and the name of one airline responsible for the most airtime for this aircraft type.

18. Which two airports had the most passengers transported between them and how many? Provide the names of the two airports.

19. For the first two months (i.e., January – February), which route (i.e., from one airport to another) that was served by "Delta Air Lines Inc." was operated by the most types of different aircraft? Provide a nested list where each inner list represents one month and the contains the IDs of the airports (to and from) and the total number of aircraft types.

20.

i) Which airport had the most direct flights to or from other airports and how many? Provide the name of the airport.

ii) In addition, which airport had the most flight segments with the airport determined from the previous question? Provide the name of the airport.