**Exercise 4 – Lists & Functions**

**Assignment Specification**

**The Dataset:**

This assignment is using data from the NYC public bike-sharing program, Citi Bike (http://www.citibikenyc.com).

On the System Data page in the website, there are links to datasets related to bikes usage (https://ride.citibikenyc.com/system-data).

A portion of the data to be analyzed in this exercise is from November 2021, a portion from November 2016.

Each row has data related to a ride. The attributes of the ride (columns/values of the files) are: Trip Duration, Start Time, Stop Time, Start Station ID, Start Station Name, Start Station Latitude, Start Station Longitude, End Station ID, End Station Name, End Station Latitude, End Station Longitude, Bike ID, User Type, Birth Year, Gender

**Description:** Write a program to read in the Citi bike data files and store it as a list of lists, and do the processing specified below. Compare - as specified in the Procedure - the average daily trip duration (from ended\_at - started\_at), the most popular starting and ending stations, the number of members vs casual users. Merge the 2 files in a pandas structure and print the result. Write a report with your findings.

**Input**: No user input. The input are the files JC-202111-citibike-tripdata.csv and JC-201611-citibike-tripdata.csv

**Output**: Print a blank line, then print the output as specified by the steps in the procedure.

**Procedure**:

1. Read the 2 files and create 2 lists of lists.

Note that using one index for data refers to the whole line of data for that date, while using two indices refers to a specific cell on a specific date. For example (not from the files above):

* data [2] = ['6/3/13', '8599', '81135', '31547.41', '323206.16', '29611', '1160', '1160', '221']

which is the list of data items for June 3 (the third line - index starts at 0)

* data [4] [0] = 6/5/13 (the date of the 5th entry)

1. Create a function “print\_details” taking the list(s) generated above as input and doing the following:
   1. Loop through each list of your list of lists to collect the daily trip duration (from ended\_at - started\_at), the number of times starting and ending stations have been used, the number of members and casual users
   2. Print a blank line
   3. Print the average daily trip duration, the 5 most popular starting and ending stations, the number of members and casual users
2. Compare the results from the 2 files. Comparison should create the elements to evaluate the ridership change over the 2 years.
3. Merge the files into a pandas structure and print it (just “raw” print). Any form of “merging” is acceptable. See details at the following URL: <https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html>
4. Print “This is the end of the files processing”
5. Write a 2 pages report describing - based on the data from the above analysis - how ridership change.