**Exercise 4 – Lists & Functions**

**Assignment Specification**

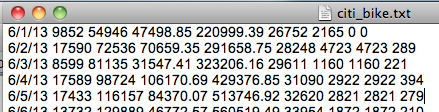
**The Dataset:**

NYC has a public bike-sharing program called Citi Bike, which began May 27, 2013: http://www.citibikenyc.com/

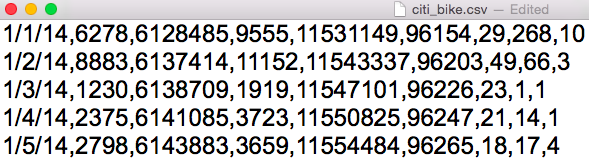
If you go to the System Data page in the website, there are links to datasets related to bikes usage.

A portion of the data, for June through September 2013, is contained in the file citi\_bike.txt and another portion of the data for early 2014 is contained in the file citi\_bike.csv

The text file data look like this:



The csv file data look like this:



The columns/values represent the following:

1. Date
2. Trips over the past 24-hours (5 pm to 5 pm)
3. Cumulative trips (since launch)
4. Miles traveled today (5 pm to 5 pm)
5. Miles traveled to date
6. Total Annual Members
7. Annual Member Sign-Ups (5pm - 5pm)
8. 24-Hour Passes Purchased (5 pm - 5 pm)
9. 7-Day Passes Purchased (5 pm - 5 pm)

**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 at 2.a - the average daily miles traveled (item 4) and the 24-­hour pass purchased (item 8) for the months of June (summer) and January (winter). Merge the 2 files in a pandas structure and print the result.

**Input**: No user input. The input are the files citi\_bike.txt and citi\_bike.csv

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

**Procedure**:

1. Read in the citi\_bike.txt file and create a list of lists.

To do that, follow the instructions below:

* file = open (“citi\_bike.txt”, “r”) # open citi bike text file for reading
* data = [] # create empty list
* for line in file: # read each line
* parts = line.strip().split() # strip the white space and create list of items
* data.append(parts) # append that line’s items list to the data list

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:

* 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 generated above as input and doing the following:
   1. Loop through each list of your list of lists to calculate the average daily miles traveled and total24-­hour pass purchases for that month
   2. Print a blank line
   3. Print the first and last day in your data using list indexes. Output should look like:

"The following data is from 1/1/14 to 1/31/14:”

* 1. Print a blank line
  2. Print the average miles
  3. Print the total number of pass purchased
  4. Print top 5 days with higher number of trips (item 2).

1. Read in the citi\_bike.csv file, create a list of lists and print the same information as for the txt file, using the function “print\_details”
2. Merge the csv and txt file into a pandas structure and print it (just “raw” print)
3. Print a blank line, then “This is the end of the files processing”.