**Exercise 3**

**Assignment Specification**

**Description**: This assignment has three parts: processing a csv file related to January-February period, processing a second csv file related to April-May period and comparing the two.

Part 1:

**Input**: No input from the user required. The input file name to use in the script is *CitiBike- NYC-CitiBike-Jan\_Feb2016.csv*

Please note:

* file name has to be without its complete file path. For example: c:/stevens/em624/NYC-CitiBike- NYC-CitiBike-Jan\_Feb2016.csv is NOT right. If you change platform or move the file to another directory, the script wouldn’t work. Be sure your script and your file are in the same directory.
* No data structure management library (like pandas) should be used for this assignment.

**Output**: Skip a line/print a blank line, then print:

* + These are the last five lines in the file:
    - [5 lines of file]
    - [Skip a line]
  + The file has n1 lines. *x* of them have "Customer" as *usertype*, *y* have "Subscriber" as usertype. "Customer" are *z*% of the total

**Procedure**:

1. Open the file
2. For each line in the file
   1. Keep a count of the number of lines (this is *n0*) - Note: skip the header
   2. Keep a count of the number of lines with "Customer" as *usertype* (this is *n1*)
   3. Keep a count of the number of lines with "Subscriber" as usertype (this is *n2*)
   4. Print the last 5 lines
3. After processing the file:
   1. Calculate the z1 % of Customer
   2. Print: *The file has n0 lines. Of those n1 have usertype as Customer, n2 have usertype as Subscriber. "Customer" are z1% of the total*.

Part 2:

**Input**: No input from the user required. The input file name to use in the script is *NYC-CitiBike-Apr\_May2016.csv*

See above for file naming in your script.

**Output**: Similarly to Part 1: skip a line, then print:

* + These are the first five lines in the file:
    - [5 lines of file] (Note: skip the header):
    - [Skip a line]
  + The first file is (smaller or larger) than the second one. The percentage of Customer in the first file is (smaller or larger) than in the second one

**Procedure**:

1. Open the file
2. For each line in the file
   * Keep a count of the number of lines (this is *n3*) - Note: skip the header
   * Keep a count of the number of lines with "Customer" as *usertype* (this is *n4*)
   * Keep a count of the number of lines with "Subscriber" as usertype (this is *n5*)
   * Print the first 5 lines
3. After processing the file:
   * Calculate the z2 % of Customer
   * Print: *The file has n3 lines, of which n4 have usertype as Customer, n5 have usertype as Subscriber.*

Part 3:

**Input**: No input from the user required. Use the data from Part 1 and Part 2

**Output**: Comparison of the data in the 2 files. Considering the files are related to different periods, we want to compare them to better serve the customers in the different times of the year.

**Procedure**:

1. Check IF n0>n3:
   * IF n0>n3, print: *The Winter riders are more than the Spring*
   * else, print: *The Spring riders are more than the Winter*
2. Check IF z1>z2:
   * IF z1>z2, print: *During the Winter there are more Customers/non-Subscribers than in the Spring*
   * else, print: *During the Spring there are more Customers/non-Subscribers than in the Winter*
3. Write a 1 page interpretation. The interpretation would be a narrative describing/explaining in plain English the results of your Python script.

Submit the 3 parts as a **single** .py file via Canvas and the interpretation in a separate doc/pdf file.