Homework 4
(*Individual Homework)

Description

Given the *input_file1, input_file2, n_element* and *app_name* in "user_definition.py", complete the python code "hw4.py" to:
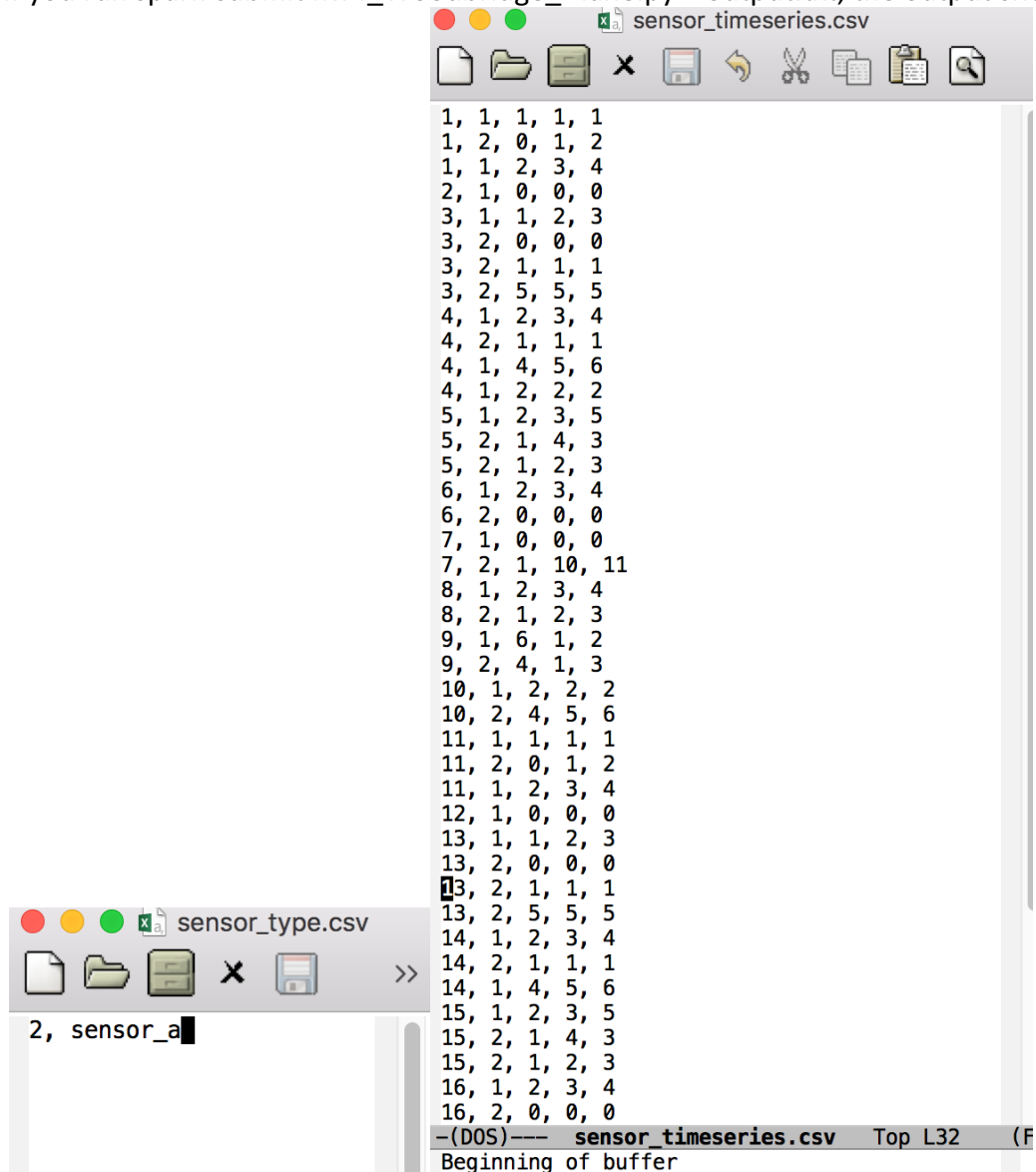
1. Preprocess the input data.
    -   *Input_file1* includes timeseries sensor data and has 5 fields with a format of "**timestamp, sensor_id, x_axis reading, y_axis reading, z_axis reading**".
    -   *Input_file2* includes sensor information and has 2 fields with a format of **"sensor_id, sensor_name"**
    -   Input_file1 samples data every time there are any changes in the x, y and/or z axis and could have multiple readings with the same timestamp.
        -   If there are multiple readings with the same timestamp, the average values should be used.
2. Print **sensor_id** and its name <u>**ordered by** *sensor_id*</u> in **input_file1**.
    -   If corresponding sensor information is not given in **input_file2,** the sensor_name is **None**.
    -   Each line should be formatted as **sensor_id : sensor_name**.
3. For each sensor data, print **first and last n_element** <u>**ordered by** *timestamp***.**</u>
    -   For example, if n_element is 10, then print the first 10 and last 10 sensor readings (ordered by the respective timestamp.
    -   Each line should be formatted as **[timestamp, [x_axis average, y_axis average, z_axis average]],** where the average is necessary if there is multiple data points with the same timestamp.
    -   Between first n_element and last n_element, print "**…**" (example below).

Submit the hw4.py file (**ONLY**) - the name of your file should be **<u>hw4_LastName_Firstname.py</u>** on Canvas. Make sure it runs in **Python 2.7**.

This assignment counts **6%** of your final grade.

We provide two example input file(input_1/ sensor_timeseries.csv, sensor_type.csv, input_2/ sensor_timeseries.csv, sensor_type.csv) and corresponding output.txt.
If you run spark-submit hw4_Woodbridge_Diane.py > output.txt, the output should be:

sensor_timeseries.csv

```
1, 1, 1, 1, 1
1, 2, 0, 1, 2
1, 1, 2, 3, 4
2, 1, 0, 0, 0
3, 1, 1, 2, 3
3, 2, 0, 0, 0
3, 2, 1, 1, 1
3, 2, 5, 5, 5
4, 1, 2, 3, 4
4, 2, 1, 1, 1
4, 1, 4, 5, 6
4, 1, 2, 2, 2
5, 1, 2, 3, 5
5, 2, 1, 4, 3
5, 2, 1, 2, 3
6, 1, 2, 3, 4
6, 2, 0, 0, 0
7, 1, 0, 0, 0
7, 2, 1, 10, 11
8, 1, 2, 3, 4
8, 2, 1, 2, 3
9, 1, 6, 1, 2
9, 2, 4, 1, 3
10, 1, 2, 2, 2
10, 2, 4, 5, 6
11, 1, 1, 1, 1
11, 2, 0, 1, 2
11, 1, 2, 3, 4
12, 1, 0, 0, 0
13, 1, 1, 2, 3
13, 2, 0, 0, 0
13, 2, 1, 1, 1
13, 2, 5, 5, 5
14, 1, 2, 3, 4
14, 2, 1, 1, 1
14, 1, 4, 5, 6
15, 1, 2, 3, 5
15, 2, 1, 4, 3
15, 2, 1, 2, 3
16, 1, 2, 3, 4
16, 2, 0, 0, 0
```
-(DOS)--- sensor_timeseries.csv Top L32 (F
Beginning of buffer

sensor_type.csv

```
2, sensor_a
```

```
1 : None
[1.0, [1.5, 2.0, 2.5]]
[2.0, [0.0, 0.0, 0.0]]
[3.0, [1.0, 2.0, 3.0]]
[4.0, [2.6666666666666665, 3.3333333333333335,
4.0]]
[5.0, [2.0, 3.0, 5.0]]
[6.0, [2.0, 3.0, 4.0]]
[7.0, [0.0, 0.0, 0.0]]
[8.0, [2.0, 3.0, 4.0]]
[9.0, [6.0, 1.0, 2.0]]
[10.0, [2.0, 2.0, 2.0]]
...
[11.0, [1.5, 2.0, 2.5]]
[12.0, [0.0, 0.0, 0.0]]
[13.0, [1.0, 2.0, 3.0]]
[14.0, [3.0, 4.0, 5.0]]
[15.0, [2.0, 3.0, 5.0]]
[16.0, [2.0, 3.0, 4.0]]
[17.0, [0.0, 0.0, 0.0]]
[18.0, [2.0, 3.0, 4.0]]
[19.0, [6.0, 1.0, 2.0]]
[21.0, [2.0, 2.0, 2.0]]
2 :   sensor_a
[1.0, [0.0, 1.0, 2.0]]
[3.0, [2.0, 2.0, 2.0]]
[4.0, [1.0, 1.0, 1.0]]
[5.0, [1.0, 3.0, 3.0]]
[6.0, [0.0, 0.0, 0.0]]
[7.0, [1.0, 10.0, 11.0]]
[8.0, [1.0, 2.0, 3.0]]
[9.0, [4.0, 1.0, 3.0]]
[10.0, [4.0, 5.0, 6.0]]
[11.0, [0.0, 1.0, 2.0]]
...
[13.0, [2.0, 2.0, 2.0]]
[14.0, [1.0, 1.0, 1.0]]
[15.0, [1.0, 3.0, 3.0]]
[16.0, [0.0, 0.0, 0.0]]
[17.0, [1.0, 10.0, 11.0]]
[18.0, [1.0, 2.0, 3.0]]
[19.0, [4.0, 1.0, 3.0]]
[21.0, [4.0, 5.0, 6.0]]
[23.0, [5.0, 6.0, 7.0]]
[24.0, [5.0, 8.0, 9.0]]
```