Homework-9

**Out Date:** 11/10/2019 (Sunday)

**Due Date:** 11/23/2018 (Saturday) 11:59PM

Team#: \_\_\_

Team Member-1:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) \_\_

Team Member-2:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Member’s Contribution (in %) \_\_

**Submission**

1. Work on the homework requirements.
2. Prepare your Python file (e.g., HW9\_P1\_Team#.py).
3. Upload the file to blackboard.

**Problem Statement:** Three datasets on an adult literacy program are available with this document. The dataset **Tutor.xlsx** stores tutor information (tutor ID, certificate date, and tutor status). Before the tutors teach any student, they are required to complete a certification class offered by the agency. Their certificate completion dates are stored under the **CertDate** column. The column **TutorStatus** reports each tutor’s current status as Active (currently serving), Temp Stop (temporarily not working), or Dropped (dropped out of the program).

The dataset **Student.xlsx** stores student information (student ID, student group, and read score). Students are required to complete an assessment interview before they attend any tutoring session. The assessment results are recorded under the **ReadScore** column and are shared with the tutors. Each student belongs to a student group (**StudentGroup**).

When matched with a student, a tutor meets with the student on one-on-one basis. The dataset **Match\_History.xlsx** stores the match data (match ID, tutor ID, student ID, start of tutoring, and end of tutoring). Students change tutors if their learning style does not match the tutor’s tutoring style. Tutors may have more than one student. The **StartDate** column shows the beginning of tutoring for an individual student. The **EndDate** column shows the end of tutoring. An empty cell of the **EndDate** column represents ongoing tutoring.

Write a python script to perform the following queries **[10 points each]**:

1. Which tutors have a Dropped status and have achieved their certification after 4/01/2018?
2. What is the average length of time a student stayed (or has stayed) in the program? You may use the current date for ongoing tutoring.
3. Identify all students who have been matched in 2018 with a tutor whose status is Temp Stop.
4. List the Read scores of students who were ever taught by tutors whose status is Dropped.
5. List the tutors who taught two or more students.
6. Display a list of all students, their read score, their tutors, and tutors status. Store this information in a file names **Student\_Tutor.xlsx**.
7. For each student group, list the number of tutors who have been matched with that group.
8. List all active students who started in May and June.
9. Find students who have not been tutored yet.
10. Find tutors who did not tutor any students.

Use pandas, and numpy for data wrangling. Submit the Python script on blackboard. The script should be self-contained, meaning it should read the data files (Tutor.xlsx, Student.xlsx and Match\_History.xlsx) from the folder where the script resides and produce required outputs. The output includes display of query results on the console and creation of an output file for query-6.

Your program will be generic enough to handle any datasets of the same structures. Your programs may be tested with different test sets.