Python Screening Test
Oxbridge Health
July 2023

Directions: Please provide one python script per problem. The code should be in working condition. Data for problem 3 is provided. Please generate some basic test data to illustrate the functionality of code for problem 1 and problem 2.

- **Problem 1.** Write a python function (does not need to be a decorator) which calls pandas merge() and adds the following functionality:
    - A parameter which can optionally enforce that
        - The left dataframe contains a superset of observations available in the right dataframe (defined in terms of the merge keys)
        - The right dataframe contains a superset of observations available in the left dataframe (defined in terms of the merge keys)
        - The two dataframes contain the same sets of observations (defined in terms of the merge keys)
    - When specifying different sets of columns for *left_on* and *right_on*, have the resulting merged dataset contain a single set of key columns. [e.g. if left_on='person_id' and right_on='id', we only would like to have 'person_id' (tip: pay attention to the situation when neither dataframe is a superset of the other)
    - Check merge columns across the left and right dataframe for consistency of their types and raise an error in they are not compatible (e.g. one is an integer, the other one is a string)
    - Add an option to raise an error if there are no matched observations from the two input datasets in the resulting dataset
    - Make -validate- in pd.merge() a required argument

- **Problem 2.** Assume that you have two events whose beginning and end dates are captured by four columns in a dataframe:
    Event 1: [event_1_start_date, event_1_end_date]
    Event 2: [event_2_start_date, event_2_end_date]
- Write a function to calculate the number of days of overlap between the two events. Create a test dataset and provide the output from the calculations.

- **Problem 3.** You are given three tables (please see corresponding tabs in -all_test_data.xlsx-):
    - A medical event level table for patients, indicating the start and end dates of the events.

- An enrollment table, specifying start and end months of a person's enrollment in health insurance plans. In cases of gaps in enrollment, a person will have multiple records in the enrollment table. The data may not be clean, though - it is possible for enrollment records to overlap (i.e. the presence of multiple records does not necessarily imply gaps). Enrollment status can change monthly.
- A person-level file indicating death dates

Determine the set of medical events for which the person was continuously enrolled in a health plan while alive (i.e. the patient must be continuously enrolled through the month corresponding to the death date).