DAP Assignment: Individual Assignment

This component has 50% weightage in the course.

Deliverables:

1. The Python (.ipynb) file used for solving the assignment. Please number your answers properly.

Note: The Assignment submission form should be submitted as well. Your submission will not be considered without the Assignment Submission form being submitted.

Questions for the assignment:

1. Write a function in Python that takes three integers as arguments, and that prints either “Yes” or “No”, depending on whether you can or cannot form a triangle from the given lengths. **[10 Marks]**
   Hint: If any of the three lengths is greater than the sum of the other two, then you cannot form a triangle. Otherwise, you can.

2. Using Pandas, do the following **[15 Marks]**:
   a. Create 3 different dataframe based on the following raw data. Name the dataframes as df1, df2, df3

```
raw_data_1 = {
    'subject_id': ['1', '2', '3', '4', '5'],
    'first_name': ['Alexa', 'Amaya', 'Alminas', 'Alicia', 'Ayoub'],
    'last_name': ['Atherton', 'Sujaya', 'Dolas', 'Young', 'Khan']}


raw_data_2 = {
    'subject_id': ['4', '5', '6', '7', '8'],
    'first_name': ['Bijoy', 'Brijesh', 'Bran', 'Bharat', 'Betsy'],
    'last_name': ['Bondi', 'Bhardwaj', 'Bruce', 'Bisht', 'Binny']}


raw_data_3 = {
    'subject_id': ['1', '2', '3', '4', '5', '7', '8', '9', '10', '11'],
    'test_id': [51, 15, 15, 61, 16, 14, 15, 1, 61, 16]}
```

b. Stack dataframes df1 and df2 by rows (**Hint:** You don't have to merge but concatenate). Call new dataframe as df1_df2_rows
c. Stack dataframes df1 and df2 by columns. Call new dataframe as df1_df2_cols
d. Merge df1_df2_rows with df3 based on subject_id.
e. Perform a full join on df1 and df2 based on subject_id.

3. Using Pandas, do the following **[25 Marks]**:
    a. Import **restaurant.tsv** file (it is in a tab separated format) and check if the data has been loaded properly.
    b. Provide summary statistics on the columns in the dataset and also identify the shape (number of rows, columns) and type of columns in the dataset.
    c. Identify the item that has been ordered the most in the dataset and how many quantities of that item have been ordered?
    d. What were the sales of the restaurant during the entire period of the dataset?
    e. On average, how much does the restaurant make on any order?

4. Analyze Earthquakes **[25 Marks] (earthquakes.csv, Earthquake variable description.xlsx)**

> *The National Earthquake Information Center (NEIC) determines the location and size of all significant earthquakes that occur worldwide and disseminates this information immediately to national and international agencies, scientists, critical facilities, and the general public. The NEIC compiles and provides to scientists and to the public an extensive seismic database that serves as a foundation for scientific research through the operation of modern digital national and global seismograph networks and cooperative international agreements. The NEIC is the national data centre and archive for earthquake information.*
>
> *This dataset includes a record of the date, time, location, depth, magnitude, and source of every earthquake with a reported magnitude 5.5 or higher since 1965.*

Conduct data analysis on the dataset and try to uncover meaningful and/or interesting insights from the dataset.

5. The dataset contains data on Indian cities (**Indian_cities.csv**) (sourced from Govt of India website) and has information about various characteristics. **[25 Marks]**

   Columns for the dataset are as follows:

   - 'name_of_city' : Name of the City
   - 'state_code' : State Code of the City
   - 'state_name' : State Name of the City
   - 'dist_code' : District code where the city belongs (99 means multiple districts)
   - 'population_total' : Total Population
   - 'population_male' : Male Population
   - 'population_female': Female Population
   - '0---6_population_total' : 0---6 Age Total Population
   - '0---6_population_male': 0---6 Age Male Population
   - '0---6_population_female': 0---6 Age Female Population
   - 'literates_total' : Total Literates
   - 'literates_male' : Male Literates
   - 'literates_female' : Female Literates
   - 'sex_ratio': Sex Ratio
   - 'child_sex_ratio' : Sex ratio in 0-6
   - 'effective_literacy_rate_total' : Literacy rate over Age 7
   - 'effective_literacy_rate_male' : Male Literacy rate over Age 7
   - 'effective_literacy_rate_female': Female Literacy rate over Age 7
   - 'location' : Lat,Lng
   - 'total_graduates' : Total Number of Graduates
   - 'male_graduates' : Male Graduates
   - 'female_graduates': Female Graduates

Conduct data analysis on the dataset and try to uncover meaningful and/or interesting insights from the dataset. Before conducting the analyses state the questions that you would want to address from the dataset, and then conduct analyses accordingly.

General Instructions:

1. This is an individual assignment.

2. Do NOT submit .zip files otherwise the submission will not be considered.
3. Late submission is applicable as per the course outline.
4. Please name your files properly and make sure your name and PGID is added to every submission made for this assignment.
5. The honour code for this submission is **2N-b**, please look through the honor code restrictions carefully before attempting the assignment as they will be strong consequences for breaking them.
6. There are a total of 5 problems. Attempt all problems to achieve maximum marks.

Deadline: 29th January, 11:55 pm IST