TPT 1201
GROUP ASSIGNMENT 2

INSTRUCTIONS TO STUDENTS:

1. The total mark for this assignment is **40%,** with Question 1 (20%), Question 2 (10%), and Presentation (10%).

2. This assignment has TWO (2) parts. The first part is Research Proposal Writing, the second part is Statistical Data Analysis using Python. The assignment should to be completed in a group of **minimum** 3 **members** and **maximum** 4 **members**.

3. There will be a presentation for Assignment 2. Each group will be allocated maximum **30 minutes** for the presentation; each member must present. Presentation marks may vary from one member to another.

4. For Question 1, identify **ONE** research domain and find at least 15 research papers related to it. The papers must be recent than year 2017.The output of this assignment is a PDF report. The maximum number of pages is 5, including the references. All reference and citation must follow APA formatting. Use the LaTeX template given.

5. If plagiarism is detected, the assignment will be given 0% for all members of the group.

6. Deadline for submission is on 24/7/2022 at 11.59pm. Submission is to be made via Google Classroom. Timestamp will be logged as proof of submission. Late submission will be deducted marks; 5 marks each day.

7. Only the project leader should submit one **ZIP** file consisting of one report in PDF, research papers used as references, and a Jupyter Lab file.

QUESTION 1 (Research Proposal) – 20%

Based on your findings from Assignment 1, you are requested to write a *research proposal*. In this proposal, you should only focus on a specific research problem and subsequently present appropriate method(s) to solve the problem. You research proposal should consist of the following sections:

(i) Abstract of Research Proposal (max 300 words)
   a. What is the problem you are solving?
   b. What is the state-of-the-art research work related to the problem?
   c. What expected outcome from your proposal? Better performance?

(ii) Introduction
   a. Provide brief introduction to the problem you are solving. The terminology and the state-of-the-art research work.
   b. Write the research objectives.

(iii) Motivation of the Research
   a. State clearly why the problem you want to solve worth solving? In other words, why is it so important to be tackled.

(iv) Research Objectives
   a. List down at least 2-3 objectives

(v) Literature Review
   a. Provide a brief discussion on existing research works that have also highlighted that the problem you are solving is an important issue.
   b. Hint: you may consider extending your findings from Assignment 1. Use table(s) or graph(s) to summarize the findings.

(vi) Research Method
   a. Discuss the methods you are proposing? Experiments? User Survey…etc?
   b. How are you going to evaluate the findings from your experiments?
   c. Use a flowchart or similar diagram to present the high-level discussion of method.

(vii) Expected Outcomes
   a. What do you expect from your proposed method?
   b. How could the expected outcomes contribute to the research community?

(viii) References
   a. List down 15 most relevant references to your work in APA format.

Continued...

QUESTION 2 (Statistical Data Analysis) – 10%

Scenario : People-to-PPV assignment

You are given two datasets, namely, `people.csv` and `ppv.csv`. The first dataset consists of 10000 vaccinees' locations while the second dataset represents 100 vaccination centers' locations. All the locations are given by the latitudes and longitudes. Your task is to assign vaccinees to vaccination centers. The assignment criterion is based on *shortest* distances.

Task 1: Is there any significant difference between the execution time for **2** computers?

Write a Python program for the scenario above and compare its execution time using 2 different computers. You can use Google Colab or JetBrains Datalore (https://datalore.jetbrains.com/) or Python in your local machined. You need to run the program 50 times in each computer. You must provide the specification of RAM, hard disk type, and CPU of the computers. You need to use *shaded density plot* to show the distribution difference. Make sure you provide discussion of the experiment setting.

Table 1. Execution Time Comparison using 2 Computers

No	Computer 1	Computer 2
1		
2		
...		
n		

Task 2: Is there any significant difference between the execution time for **3** different Python programs?

Write 3 different methods/ways to solve the *people-to-ppv* assignment problem as stated in the scenario. You need to run the program 50 times for each method. You need to use *shaded density plot* to show the distribution difference.

Table 1. Execution Time Comparison for 3 Different Methods

No	Method 1	Method 2	Method 3
1			
2			
...			
n			

Continued...

Take note that the differences must be substantial. Some examples, but not limited to, are shown below:

Acceptable	Not Acceptable
Changing from for-loop to while-loop ***and*** Changing from list to numpy array	***Only*** changing from for-loop to while-loop
Changing Python dictionary to JSON ***and*** Changing for-loop to using SQL	***Only*** changing from Python dictionary to JSON
Using different Python packages	Introduce new variables or change variable names
…	…

* *Refer to your **tutor** if you are not sure.*

End of Pages.