TDS 3301
GROUP ASSIGNMENT

INSTRUCTIONS TO STUDENTS:

1. This assignment has two questions, with Question 1 10% and Question 2 40%. The total marks will be normalized to **20%.**

2. The assignment is to be completed in a group of a **maximum 4 members**.

3. For Question 1, the maximum number of pages is 5, including the front cover and the references. All reference and citation formats must follow APA. Report with more than 5 pages will resort to 0%.

4. For Question 2, make sure you submit the JupyterLab file containing your solution.

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

6. Deadline for submission is on 5/12/2022, 12pm. Submission is made via Google Classroom. Timestamp will be logged as proof of submission. Late submission will be deducted marks.

7. Report must be prepared using the template given.

8. The project leader only should submit THREE items in a ZIP file: (i) all the research papers you referred to, (ii) a report for Question 1 in PDF, and (iii) a JupyterLab file. Name your zip file **`<Student ID>_<Project Leader Name>.ZIP`**

QUESTION 1: Applications of Data Mining [10 marks]

Select any __ONE__ domain. Examples of domain including, but not limited to, the following:
- (i) Location Analytics
- (ii) Healthcare
- (iii) Retail
- (iv) Natural Disaster
- (v) Banking or Insurance
- (vi) Airlines

Discuss the techniques proposed and implemented by researchers to tackle challenges in the selected domain. Your findings must be supported by at least **20** journal papers recent than 2017. The literature must be related to data mining. Your report should also include the conclusion, limitations of current findings, future work, and recommendation. Your report must have both the Table 1 and Table 2. Provide a discussion to each table.

Table 1: Domain problems tackled by researchers

Author	Domain Problem 1	Domain Problem 2	Domain Problem 3	Domain Problem *n*
Ali(2018)		x	x	x
Chong & Ting (2019)				x
James et al.(2010)	x			
...	...	...	...	...

Example of discussion:

Table 1 shows the challenges and problems solved by different researchers when dealing with COVID-19. From the table, most of the reports have been centering around vaccine supply and delivery. This can be showed by the number of counts where 10 out of 20 papers focused on vaccine supply optimization. The second most challenging work as reported by researchers are predicting the outbreak of an area in next 14 days....

[Note: you should provide critical analysis about the findings, and not just provide the statistical information about your findings]

Table 2: Data mining techniques used by researchers

Author	Data Mining Technique 1	Data Mining Technique 2	Data Mining Technique 3	Data Mining Technique *n*
Ali(2018)	x			x
Chong & Ting (2019)			x	
James et al.(2010)	x	x		
…	…	…	…	…

Example of discussion:

Table 2 depicts the information about different data mining techniques used in work related to COVID-19. From the above table, decision-tree has been widely accepted in most of the reported research work. Based on the 20 papers, 10 out of them employed decision-tree. The focus areas are vaccine supply chain optimization and forecasting. More recent work has also reported on using geography particularly geospatial analytics to predict the outbreak….

[Note: you should provide critical analysis about the findings, and not just provide the statistical information about your findings]

Continued…

QUESTION 2: Python Programming [40 marks]

Refer to the attached JupyterLab file to complete this question. Marks are given for each question.

You are requested to provide information about the members in the JupyterLab file. Include the Student ID, Student Name, and Contact Number.

End of Pages.