Project 20%
Big Data Analytics (Python)
Dataset: Refer to list of datasets for groups

Your group work is to conduct investigation and determine hidden pattern based on dataset provided to your team. As a data analyst, you are responsible to conduct data analytics and **present your findings** and **prepare a short report (less than 5-10 pages)** to explain your findings based on the **CRISP DM framework**. Your short report content as below: (Arial 12, 1.5 spacing).

Table of content

1. Background of the study – follow CRISP-DM
    a. Explain the dataset (dataset nature understanding)
        i. Data coming from which field, data source, nature of the data
    b. Objective of the analytics study (3 objectives) – purpose of analytics
    c. Elaborate how the CRISP Methodology implemented in the project

2. Dataset analysis and visualization
    a. Analyze and describe the dataset
        i. describe properties, size, dimension, datatype, distribution
        ii. explore issues exist in the dataset
    b. Use descriptive analytics to describe the issues exist in the dataset
    c. Solve the issue stated above.
        i. Explain the methods you use to solve the issue exist in the dataset?
        ii. Use exploratory data analysis to describe the dataset.
        iii. Describe hidden insight revealed from the created data visualization. What can you tell from this dataset description?
    d. Use one of the following classifiers to train and test your data, such as: DecisionTreeClassifier, LogisticRegression, KNeighborsClassifier, RandomForestClassifier, Support Vector Machine, GaussianProcessClassifier
    e. Describe and provide a comprehensive insight on your output in step 2(d).

3. Summary - Prepare a paragraph of summary on how you process the data into final output.

Datasets:
Dataset and data description is available at: Dataset for BDA (Jan 2022)
DEADLINE: 9/3/2022, by 1PM

Presentation (10 minutes)
Submission Date: TBA

i) Hardcopy-Presentation session
ii) Softcopy -Ulearn

Rules and Regulations

i) This is a group project.
ii) Zero mark will be awarded if plagiarism detected.
iii) Any late submission without prior approval will not be entertained.