WQD7004 Programming for Data Science (Group Project – 20%)

Your company has received several new projects this month. All these projects are equally important, with a certain level of technical difficulties and challenges. You have been instructed by the CIO to lead a team of **5 members**, to handle one of the projects. You have been given **10 weeks** to complete the project and prepare a presentation. Wishing all of you best of luck.

Details

1. Choose a specific title for your project and get a dataset related to your project.

2. Based on the dataset selected, perform the following steps:
    a. Determine the details about the dataset (e.g. title, year, purpose of dataset, dimension, content, structure, and the summary) by exploring the raw data.
    b. Data cleaning: Putting it all together as a new cleaned/processed dataset. Which section of the data do you need to tidy? You can use any cleaning packages in R
    c. Data analysis: Identify the questions, what is the objective/goal of processing this dataset? What answers are you interested to find from this dataset? You must have at least two questions. **It is best to have different types of problems**, i.e. one regression, and one classification.

Requirements

1. Each group must have 5 members.

2. Report and presentation submission in week 13. Late submission will be subjected to mark deductions.
    a. Report: Submission will be an **R Markdown published at RPubs**, and the link is to be submitted in spectrum or any other medium as specified. The R Markdown may include the following:
       • Short introduction and objectives of the project
       • Explanation of all the processes involved in the project.
       • Results, which may include visualization, prediction, evaluation of models and discussion of output.
       • Conclusion
    b. Presentation: prepare the presentation (video recording, maximum 12 minutes)
    c. Only the group leader is required to submit the report which include the R Markdown link and the presentation link.

3. Marks will be given based on your presentation and report.

4. **Plagiarism is strictly PROHIBITED.** You will be penalized if any of your work is found online.