

**Assignment 1 (25%)**  
**STQD6324 Data Management**  
**SEMESTER 2 2023/2024**

As you aspire to become a proficient Data Scientist in the near future, it is crucial to hone your DATA MANAGEMENT skills, which are integral to your training. For this assignment, you are tasked with selecting a topic relevant to the industry you aim to join post-graduation from the MASTER OF SCIENCE (DATA SCIENCE AND ANALYTICS) program. You may refer to sources like the "[Top Big Data Analytics Companies in Malaysia \(2024\)](#)" or related media to guide your choice of industry.

Once you have identified your preferred industry, proceed to platforms like [kaggle.com](#) or any online database where you can access relevant RAW DATA SETS to demonstrate your DATA SCIENCE and DATA MANAGEMENT abilities. To prepare for future job interviews and showcase your skills set effectively, **you are required to utilize the GitHub platform to host your analysis.** This should include sections covering Data Cleaning, Data Visualizations, Insights and Explanations, Recommendations, Conclusion and more. Feel free to host and organize your GitHub presentation in any creative manner you prefer. You can refer to this [Git Tutorial](#) for training purpose if you are now familiar with the GitHub platform.

It is important to note that this assignment is designed to help you exhibit your DATA MANAGEMENT skills. Therefore, emphasis will be placed on how well you present your case study in alignment with the **ten elements of the DATA LIFE CYCLE** (please refer to Lecture Note Week 1).

You may choose to utilize either R (Rmarkdown) or Python (Google Colab) for this assignment. The submission deadline is **2024-04-30**. Kindly add my email, [bernardlkb@ukm.edu.my](mailto:bernardlkb@ukm.edu.my), as your GitHub co-host.