

## WQD 7005 Data Mining 2022/2023 Semester 2

### Group Assignment (at most 4 members in a group)

Select one of data set from the following web site

**UC Irvine Machine Learning Repository** (<http://archive.ics.uci.edu/ml/>)

but not limited to the above.

Write a report consisting of the first 2 characters of SAS **SEMMA**. The emphasis is on the use of data mining methodology in real-world applications.

Group Presentation – 5 marks

Group Report writing – 15 marks

Hints:

Your report topics will be covered the following items but not limited to the below:

- a) Accessing and Assaying Prepared Data
- b) Analysis goal
- c) Analysis data
- d) Show the column metadata
- e) Exploring the data source (creating a histogram, changing the graph properties for a histogram, adding a "Missing" Bin to a histogram, creating a pie chart, exploring variable associations)

Due: Week 8