Jamhuriya University of Science & Technology
Faculty of Computer & Information Technology

Guidance to the final course project of **Data Science**
The final project should implement as many data analysis and exploration steps as appropriate as possible including, but not limited, the following:

Each Group or student should come up with / produce their Unique Dataset from online repositories.

1. Reading external dataset

2. Producing various descriptive statistics of the analytic dataset. Provide verbal comments on the obtained descriptive statistics and justifications for your choices.

3. Building various possible visualizations of dataset variables including distributions of numerical variables with histograms. Provide comments on the choice of your generated graphs and descriptions of the data represented in your graphs and any insights it shows.

4. Checking any records in the dataset with missing values and handling them appropriately (e.g., missing values, duplicates, etc). Provide comments on the result of your checks and justify the methods selected for handling missing values.

5. Investigate the unique values of categorical variables and comment on your findings.

6. Build a linear regression model for predicting appropriate dependent variables and interpret the model parameters and results

7. Report summarizing all steps you have already done and interpreting the results.

Project Deadline information
Submission deadline: January 2024

Data, TBD by Lecturer.

The final project should include:

1. A Jupypter notebook including implementation of the entire project pipeline from beginning to end.
2. A PDF report version of the project implementation generated from the notebook
   You will be required to present your data analysis project in January 2024.