CS3121 Introduction to Data Science

Data Analysis Lab Exercise

Used the pre-processed dataset from **Data Pre-processing Lab** for this exercise.

For this lab you can use any Python library required for data analysis, e.g., Pandas, Scikit-learn, Numpy, Matplotlib, Seaborn, etc.

What to do

Create one page report answering the following questions.

- Provide the five number summary for `Insulin`, `BMI` and `Age` variable.
- Generate a bar chart for `Outcome` variable.
- Generate a pairwise relationship scatterplot for predictor variables (8x8 subplots) using Seaborn Python library.
- Calculate the covariance between each predictor variable and the target variable (Outcome).
- Name two most influential factors (variables) that increase the risk of diabetics. Justify your answer.

Submission

Your report should be in PDF format and renamed as <<Your_Student_Id.pdf>. Submit the report to Moodle by **7th May 2023 11.59 p.m.**

If you have any question email to *sandarekaw@cse.mrt.ac.lk*.