PROPOSAL OF PROJECT 2
PROBABILITY AND STATISTICAL DATA ANALYSIS
SECI / SCSI 2143

Source of dataset (URL)	:	https://www.kaggle.com/volodymyrgavrysh/heart-disease
Dataset description (collected by who and for what purpose?)	:	Collected by VolodymyrGavrysh The goal refers to the presence of heart disease in the patient. The value of patient that having heart disease is 1 according to the data in column 'target'. This dataset provides information on the risk factors for heart disease. Experiments with the Cleveland database have concentrated on attempting to distinguish presence (value 1) or absence (value 0) of heart disease in the patient.
Variables (name of variables and type – categorical / ordinal / interval / ratio)	:	• Age (Ratio) • Gender (Nominal) • Chest Pain (Ordinal) • Resting Blood Pressure (Interval) • Cholesterol (Interval) • Exercise induced angina (Nominal) • Max Heart Rate (Interval)
Description of purpose of study	:	To describe the relationship between the age and the risk factor for the heart disease. It will be illustrated based on the gender and all the possible cause (the level of resting blood pressure, levels of cholesterol and the maximum heart rate) individually.
Specification of target population	:	Patient that have heart disease which age within 34 to 76 years old.
Selection of variables (potential variables that will be selected for analysis)	:	• Age • Gender • Resting Blood Pressure • Cholesterol • Max Heart Rate
Proposed analysis (potential statistical test analysis related to the variables chosen)	:	There will be 5 statistical test analysis will be carried out which are 1 sample test, correlation, regression, chi square test of independence and ANOVA. Selected information will be test according to a suitable condition.
Expected outcome for analysis	:	The outcome will be which factor that most cause the patient having heart disease. Hence it will be easy for them to keep track their health.

Note: Submit this proposal within 2 weeks after briefing done by the lecturer.