PROJECT 2 PROPOSAL
PROBABILITY AND STATISTICAL DATA ANALYSIS
SECI 2143
Submission date: 30th May 2020

Dataset description (collected by who and for what purpose?) :	This dataset is about the death statistics in Malaysia, throughout the whole year of 2019. The whole description is mainly on the principal causes of death in this country. The data was collected by the Department of Statistics Malaysia to know some of the main causes of death in this country and also possibly finding ways to prevent in from happening again in the following years to come.
Variables (name of variables and type – categorical / ordinal / interval / ratio) :	• **Ratio** = The number of deaths each disease causes. • **Ordinal** = The knowledge level of the people. • **Nominal** = Types of deaths. • **Interval** = The age group of the deaths.
Description of purpose of study :	This study will aim to determine the relationship between the number of deaths that each disease causes and the knowledge level of the people.
Specification of target population :	The target population are all of the citizens of Malaysia, while also taking count of their age.
Selection of variables (potential variables that will be selected for analysis) :	The main objective of this study is to determine the most critical diseases in this country that are causing more and more deaths every year, especially in the year 2019. This study hopefully can show and educate people on the grave danger of this whole issue, which is the types of dangerous diseases, the number of deaths each disease causes and the age group of the most frequent casual

Proposed analysis (potential statistical test analysis related to the variables chosen)	:	**Hypothesis :** The main hypothesis in between two samples, which are the number of deaths each disease causes and the knowledge people have about these deaths. **Correlation :** Relationship between the number of deaths each disease causes and the knowledge people have about these deaths. **Regression :** Relationship between the knowledge of people about their lifestyle and the ways people get these diseases.
Expected outcome for analysis	:	**Hypothesis :** The knowledge of people about these deaths is clearly related to the number of deaths each disease causes. **Correlation :** The more knowledge people have about these deaths, the lower the number of deaths that each disease can cause. **Regression :** A proper lifestyle can definitely prevent people from suddenly getting these diseases.
Source of dataset (URL)	:	• https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=RUxlSDNkcnRVazJnakNCNVN2VGgrdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09 • https://www.thestar.com.my/news/nation/2019/10/31/heart-attack-leading-cause-of-death