ETC1010 Project

The project is designed to give you experience collecting or finding your own dataset, determining the appropriate questions to answer about the data, and planning how to execute analysis of the data. The project involves several parts. The project represents 15% of your final grade for ETC1010.

1. Locate a suitable data source and determine appropriate questions that could be answered using this data. It cannot be data set from kaggle. It needs to be from an original source. If it is is csv format, there need to be more than one file or multiple sheets. Challenge yourself to work with data addressing a problem in today's world.

2. Cleaning of your data, in order to answer your questions. This is the important part to illustrate in your project, because we are expecting you to be able to demonstrate your ability to take a messy data set and organise it for later analysis.

3. Simple analysis using methods covered in class; exploratory data analysis, numerical and visual summaries of the data, and the application of basic modeling strategies. The focus is on trying to answer some of the questions you posed. You are not expected to answer all, if you have a long lots of questions.

4. Describe your cleaning procedures and analytics in web story board, which can be done using any of these R packages: **learnr**, **flexdashboard**, or a simple **shiny** app. You should include why you chose the data and what learned about the problem by completing this project. We will upload these to the departmental shiny server for everyone to see, and so that you can show it off to future employers or your family members.

5. Present your data analysis in class, 10 minute oral presentation.

This project will be conducted collaboratively, with team of your choices, and with i**a maximum team size of 4.** To ensure correct marks are awarded, please carefully document, in detail, your individual contributions to the project. Each team member is expected to participate substantially in all aspects of the work, including the writing and oral presentation. The important deadlines are as follows:

Due Date	Turn in	Points
Apr 11, end of class	Prospective team members and topics	5
Apr 18 end of class	team members and team name, and paragraph describing possible data sets, with links to the data sources.	5
May 9, end of class	Electronic copy of your data, and a page of data description, and cleaning done, or needing to be done.	10
May 23, end of class	Final version of story board uploaded	40
Wed, Fri Oct 17, 19	Project presentations during class periods. All students are expected to attend, and points will be deducted for non-attendance.	30 (peer evaluation) 5 points will be deducted from your presentation score if you do not attend for the entire class, and 5 points if you skip the class where you did not present.

No late turn-ins accepted.