The old-school model of passively learning facts and reciting them out of context is no longer sufficient to prepare students to survive in today's world. Solving highly complex problems requires that students have both fundamental skills (reading, writing, and math) and 21st century skills (teamwork, problem solving, research gathering, time management, information synthesizing, utilizing high tech tools). With this combination of skills, students gain all-round knowledge of the subject and are ready to efficiently use it in their respective real-worlds applications.

> **Project Requirements**

Your project should clearly state how the problem you have chosen is essential. You are expected to:

1. **Consider a significant problem in the current scenario**
2. **Choose a target population**
3. **Data collection:**
   - Select the sampling method to be used to collect relevant data, or
   - Select an existing dataset(s) and describe why this data fits to the project goal
4. **Provide Descriptive Statistical Analysis**
5. **Provide Inferential Statistical Results.**

Optional elements of the project:
- ✓ Correlation analysis of multiple factors
- ✓ Use data fitting and provide conclusions

> **Deliverables**

1. **Project Report (PDF Format):**
   - All groups should provide a project report including all of the above-mentioned items with detailed calculations and graphs. The project report guidelines are provided at the end of this file.
   - The report should be less than 10 pages.
   - Students are highly recommended (but not required) to use RMarkdown to generate the project report.
2. **PowerPoint presentation (PPT Format):**
   - All groups will present their projects on *Presentation Day* determined by the course schedule.
3. **Original Source code: A original R or RMarkdown file of the project**
   - Please zip all files together and specify the submission files(s) name as:

        "IE6200 Project - <Section No.> - <Group No.>",

   - ✓ Please submit your project via course platform (i.e. Canvas).
   - ✓ One submission per group is sufficient.

> **Grading:**
   - This project is intended for you to apply the concepts learned in class in an open-ended setting. You will be evaluated on how well you "sell" your proposed solution, as well as the justification you use. You may pull in any publicly available data (please cite your sources) to help justify your answer and list any assumptions you make during the course of the project. Part of selling is the look and flow of your deliverables, so please try and make the report and presentation (if applies) as professional as possible.

> **Honor Code:**
   - Students are expected to hand in their own work on the projects. Students are requested to mention the percentage of effort contributed by each student in the group in their final project report.

## Project Report guidelines

This section provides a typical report structure and requirements.

### Data Collection

In this section you should describe your approach towards collecting the data for your project. For example:

- Did you choose a readily available dataset from web?
- Did you select two different datasets and combined them to form your dataset?
- Did you create your own dataset using information from any relevant articles from web?
- Did you collect the data by yourself by conducting a survey?

Describe the variables (columns) in your dataset. Briefly explain about what information is conveyed by them.

- Ex: If you have a "Temp" column in your dataset, you can say that it represents the average Temperature on a specific day.

### Data Visualization:

In this section, you should draw plots that give information about any existing patterns in your dataset. For this, you can use any R package (ggplot2, lattice, highcharter, etc).

- Some examples of plots include.
- Distribution plots to determine what distribution is followed by a specific variable in your data.
- Correlation plot.
- Scatter plot to see how the data points are distributed overall.
- Box plots to determine median, range, interquartile range of a specific variable.
- QQ plots, PP plots for goodness of fit (recommended).
- Heatmaps for joint probability distributions.

### Statistical Analysis:

In this section you should use inferential statistical methods described in class to perform statistical analysis on your data. Some examples include:

- Formulate a question and perform hypothesis tests (left, right or two-tailed tests) to solve it.
- Find confidence interval of means, confidence of difference of means, confidence interval of proportions.
- Find probability using sampling distribution of means, variance etc.
- ✓ **Note:** Run at least 3 different statistical tests on your data.

### Advanced Analytics:

Perform regression analysis which is an excellent tool to make predictions or forecast future trends. This can be done using simple linear regression or logistic regression in R.

- ✓ **Note:** All the figures and tables that you plot should be numbered and captioned. Use figure numbers and table numbers while analyzing your results. Example: From figure 1, we can see that.