ASDS5303 Final Project

Presentation on Dec.2nd; Report due on Dec. 6th

In this final project, students will have the opportunity to showcase their skills and knowledge acquired throughout the course by conducting in-depth analyses of real-world dataset(s). The project will involve exploratory data analysis, application of methods/tools, and the presentation of findings in a final presentation. In this project, students will work as a team consisting of 2 students. Each team will give a presentation which will be 10 minutes long talk. Each team member must speak during the presentation. There will be 5 minutes (approx.) questions and answers session. Note that some questions may pop up while a team is presenting, in which case the presentation time will be adjusted accordingly. **All students are required to be present for the entire duration of the final project** and they are encouraged to ask questions at the end of each presentation.

Final Project Grading:

The final project includes four parts **a report (30%), R-code and output (30%), a presentation (30%) and attendance (10%).** The final grade will be on a team basis not individually. All students in the same team will receive the same grade for the final project.

Requirements and Guidelines:

Report:

a) You need to submit the following three files for the final project by Friday 12/06 midnight. The deadline is the same for all students no matter when the presentation is.
    1. A PowerPoint presentation
    2. A report document in PDF or Word.
    3. The original code file that can regenerate the output in your report doc.

    There will be an assignment on Canvas **group submission**. All groups must submit the three files.
b) The report should be the results of the analysis of real data set(s) and how the analysis is carried out. The analysis part should cover **one of the three** major topics taught in class, **regression, classification or pca**.
c) The dataset(s) that you are going to analyze should **not** be the one(s) that has/have already been used in lectures or in lab projects or existing R-datasets. You are expected to look for suitable dataset(s) from a research paper, database, or any other source. If you do not know how to find an appropriate dataset, please follow the suggestions below (**these are NOT requirements**) to find a data set.
    1. Google 'R dataset for regression/classification'.

2. Find datasets from the two following websites.
    i. https://www.kaggle.com
    ii. https://archive.ics.uci.edu/datasets
3. The size of the dataset is smaller than 1000 rows.
4. The missing values are less than 50%.
5. The number of columns less than 30.

d) Your report **must** include the following, among others:
    i. A suitable title
    ii. Description of the dataset(s)
    iii. Main objective(s) or research questions you want to answer.
    iv. Methods and statistical tools and techniques used to meet research objectives(s) or answer questions of interest.
    v. Detailed analysis and interpretation of results
    vi. Appropriate conclusions wherever necessary
    vii. R codes and corresponding outputs, including figures.
    viii. References such as data source links, and any resources you reach for help (papers, websites, or others).

Presentation:

The presentation will take 10-15 mins for one team, which includes a 7-10 mins talk (All students in the team must speak) and a 5 mins Q&A (Q&A is not required). You must prepare a PPT and properly time your presentation.

The final presentation schedule is as follows.

Date	**Teams**
December 2nd	**1-7 (the class may extend 20 more mins)**

Attendance:

Please be on time to get the full credit and to avoid disrupting the presentations. Please don't be late. If you are late for the final presentation, you will not get the 10% attendance.