Project One

STAT4011

Introduction

Students are required to do a group project. The project will require you to do some computer programming based on `R/Python`. At the end of the semester, each team is required to submit a report on your project as well as the corresponding `R/Python` code file. A guideline for its length is at most **5 pages** (excluding references, tables and figures). The total number of figures and tables is **at most 5**. All team members are expected to contribute to all aspects of the project work (statistical analysis, coding, writing, presentation). Students will be asked to provide confidential feedback about the work contribution of his/her team members through peer evaluations.

You are free to use any data set if you like.

Project Requirement

The project is related to the analysis of data by statistical learning approaches. The students are required to conduct the project under the following requirements.

- The students are encouraged to find the data by themselves. However, they can also use the following website to search for the dataset.
  - Kaggle https://www.kaggle.com/datasets
  - UCI https://archive.ics.uci.edu/ml/datasets.php
- The main methodology for analyzing the data should be mentioned in the lecture note or in the book (neural network is excluded)
  - Friedman, J. H. (2017). The elements of statistical learning: Data mining, inference, and prediction. springer open https://link.springer.com/book/10.1007/978-0-387-84858-7.

Sections for Project

The final project should include the following sections.

1. **Introduction/Background**
   - Why you are interested in this problem?
   - Assume that your audience is not an expert in statistics, what do people need to know to understand?
   - From where did the data come? Is this an experiment or observation study? Who collected the data? Why are the data collected?
   - What are the questions of interest that you hope to answer?

2. **Methods**
    - Summarize and explore the data.
    - What analyses are most appropriate to answer the question of interest?
    - Describe the analyses used.
3. **Result**
    - Present relevant graphics.
    - Interpret result of the analysis.
4. **Conclusion/Summary**
    - What are your conclusions?

Code

You are required to upload the `RPython` code together with your report. Please write clean and readability code. Document your code with comments so that the grader can understand the link between your code and your report.

Submission

You need to submit a report and a `RPython` code in one `.zip` file on Canvas. The file name of your zip file should be `Group#YOUR_GROUP_NUMBER.zip`, where you need to replace YOUR_GROUP_NUMBER with your group number. Only one team member needs to submit the zip file. The file should contain the following:

1. Your report in `.pdf` format. The report should be **11 point Times New Roman, and double spaced**. The report should be **no more than 5 pages** (excluding references, tables and figures). The total number of both figures and tables is **at most 5**.
2. Your `R` code needs to be ready to run on any computer with `R` and `R` Studio installed, not just yours. For instance, all dependent data files must be included (and in the folder that is referenced in your code); Your code needs to check if the required libraries are installed (and install them if needed); There should be absolutely no run-time error.

Grading of the report

The grading of the team project report depends on the following.

1. Whether you follow the instruction for different sections of the report. (65%)
2. English writing of the report (typos and grammers). (15%)
3. Well documented code. (20%)

Presentation

The major components should be similar as the components mentioned in Section 2.1. For example, it should contain the proposed problem, data, methodology, outcomes, and your conclusion. Here is something important about the presentation you should know.

- The presentation will hold in the 6th week (OCT 10 – OCT 16). The exact time will be determined later.
- The presentation of each group cannot exceed 20 minutes and each group member must join the presentation.
- Here are some Dos and Don'ts

1. Use as few words as possible on each slide. Insightful tables and figures are recommended instead of lengthy text.
2. Talk to your audience and have eye contact with them occasionally. Simply reading the presentation contents on your cell phones or tablets to audience should be AVOIDED. You can prepare some notecards in case you are blank out during the presentation.
3. Make your presentation coherent. (I understand that each of you is responsible for only a part of the project. But all of you need to make sure that the presentation is a whole story of your project rather than a combination of several individual works.)

Grading of the presentation

The grading of the team project presentation depends on the following.

1. Whether you follow the instruction of the presentation. (40%)
2. Whether your presentation is clear, is easy to understand, and deliver useful messages of your project to the audience. (60%)