BSE2610: Project (100 marks in total)

Students are required to do a group project. The project will require you to do some computer programming based on `Python`. At the end of the project, each team is required to make a **presentation** to describe the main content of their outcomes. Besides that, each team also needs to submit a report on your project as well as the corresponding `Python` code file. A guideline for its length is at most **7 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). Students will be individually asked to provide a confidential evaluation of the work contribution of his/her team members as a supplement to the report.

1 Project Requirement

The project is related to the data analysis. The students are required to conduct the project under the following datasets and use at least one method (e.g. Linear regression, ANN, Trees) we mentioned in this course. The students need to choose one dataset and find the question by themselves.

- Commercial Building Energy Consumption Survey (CBECS) 2018 microdata.
  - This dataset can be found in

    https://www.eia.gov/consumption/commercial/data/2018/.
  - We have used a similar dataset (CBECS 2012 microdata) as an example to show the performance of Lasso, Ridge and Linear regression. (The recording is provided on blackboard)
- Your valid dataset. The students are encouraged to use other dataset satisfying the requirements:
  - The data is real data and must be related to the buildings. It is publicly available, and cannot be CBECS dataset.
  - The number of data should be larger than 200 and the number of features should be larger than 10.
  - Bonus: using the valida dataset will receive 5 marks as bonus (total score will not exceed 100).

1.1 Evaluation

- Report: (60 marks)
  - Main content (48 marks)
  - Contribution Evaluation (12 marks)
- Presentation: (40 marks)
- Bonus (5 marks)
- Overall grade will not exceed 100.

2 Report

2.1 Sections for Report

The report should include the following sections.

1. **Introduction/Background**
   - What is the question of interest that you hope to answer?
   - Why you are interested in this question?
   - Assume that your audience is not an expert in data science, what do people need to know to understand?
   - From where did the data come? (In case you use other dataset.)
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?

2.2 Code

You are required to upload the `Python` 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.

2.3 Submission

2.3.1 Submission for main content

Your team need to submit a report and a `Python` code in one `.zip` file on Blackboard before Apr. 10, 23:59. 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 7 pages** (excluding references, tables and figures). The total number of both figures and tables is **at most 5**.
2. Your `Python` code needs to be ready to run on any computer with `Python` 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/packages are installed (and install them if needed); There should be absolutely no run-time error.

2.3.2 Submission for contribution evaluation

You need to submit additional individual file to evaluate your and your teammate's contribution on Blackboard before Apr. 17, 23:59. The submission will be opened at Apr. 11, 14:00. Specficially, you should write the following sentences:

1. I am a member of Group XX. My contribution is XXXXXXX. If enough contribution deserves 12 marks, I will give myself XX marks.
2. In terms of my teammates' contribution, I will give XX marks to my teammate XXX. I will give XX marks to my teammate XXX. I will give XX marks to my teammate XXX. I will give XX marks to my teammate XXX.

2.4 Grading for reports

The grading of the report depends on the following.

1. Whether you follow the instruction for different sections of the report. (40%)
2. English writing of the report (typos and grammers). (15%)
3. Well documented code. (15%)
4. Reference. (10%)
5. Contribution Evaluation. (20%) [This part is confidential and solely depends on the evaluation from the teammates.]

3 Presentation

There is no special guideline about the presentation, but 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 room **Z406 on Apr. 10 from 8:30 to 11:30**.
- The presentation of each group cannot exceed 15 minutes. The schedule for each groups is as follows. The presenters should come to the their sessions on time. Other group members are encouraged to join the presentation.
  - 8:30-9:30 (Groups 1 - 4)
  - 9:45-10:45 (Groups 5 - 8)
  - 11:00-11:30 (Groups 9 - 10)