STA 4164 - Statistical Method III

Project 2 – Guidelines

Summer 2022

Please READ this document CAREFULLY before starting your project

GENERAL GUIDELINES:

1. This assignment is a group assignment. You may use notes from our class and online resources, such as your preferred programming language documentation and help files. You may also discuss your approach towards the problem with your fellow students but make sure to prepare and write your report exclusively by your teammate.

2. The purpose of this study is to do a comprehensive and complete regression analysis from what you have learned in this class. Therefore, the report should be detailed and concise at the same time. You should also consider your audience as an educated but non-technical person and do your best to explain your perspective as more uncomplicated as possible.

3. Your report's body ***MUST*** be free of code. Only if there is a novelty in your code never represented in the literature, it can be included it in the body. You should, however, append your code to the report as supplementary material or appendix.

4. You may use the software's output to describe what you are discussing in the text. The result (plot, table, etc.) ***MUST*** be clearly explained by its caption and referred to in the body of your report. Making your own tables to summarize the software's output is highly recommended.

5. Page limit for your report is 10 pages at most (excluding reference and code in the appendix).

6. My intention in this assignment is to evaluate your coding and writing skills and how you can convince the audience that you followed a professional methodology to answer a potential question. You will also be assessed on your ability to simply and efficiently express your finding to the general audience.

7. Be aware of plagiarism and academic integrity while writing your report.

8. Try to be innovative and push **your limits** as far as you can. This is what matters most.

9. Please put all your report into a single pdf file.

STA 4164 - Statistical Method III

Project 2– Guidelines

Summer 2022

REPORTS OUTLINE:

1. Abstract (You should write it at the beginning and summarize your research in a paragraph)

2. Objectives (Express your Main question(s) doing this study)

3. Source of your data (Where did you find your data, how is it collected, dimension, etc.)

4. Environment of your study (OS, programming language, software, version, …)

5. Data exploration
    a. Frequency exploration (mean, range, etc. of your variables)
    b. General graphs (primal explanatory scatter plots, histograms, bar charts, etc.)
    c. Exploring general assumptions (are all regression assumptions satisfied)
    d. Collinearity evaluation (Possible colinear variables and your remedial actions)

6. Data preparation
    a. Dealing with missing values (size of missed values, any imputation, discarding, …)
    b. Treating Outliers (distinguished observations, how did you deal with them, …)
    c. Possible transformations (Reasoning behind and type of transformation)
    d. Possible definition of dummy variables (merging, dummy definition, etc.)

7. Modeling:
    a. Univariate model exploration (Brief result of univariate analysis)
    b. Interaction and confounding exploration (discovered interactions, reasoning, etc)
    c. Comprehensive model (Result of your comprehensive model exploration)
    d. Model selection (Approaches you followed and findings)
    e. Diagnostics (Model diagnostic and redefining the model)
    f. Final best model

8. Discussion:
    a. Final model interpretation (What does your model mean? Clear and simple.)
    b. Model performance (accuracy) (if you pursued testing your model)

9. Conclusion (Clearly answer your main question from part 2 above)

10. References

11. Appendix (code, etc.)