JSC370 Final Project 2023

Final Project

- Recorded presentation (5 minutes, 10% of final grade): April 28th, 2023
- Written report (single spaced, length up to you but please do not show code dumps or unformatted output dumps! 90% of final grade): April 28th, 2023

Learning Objective: To apply the skills learned in JSC370 by analyzing and interpreting a dataset of your choice.

Narrative: Through this project you will begin to develop a portfolio of data science projects that will be showcased on your personal github website.

Using the dataset from your midterm, make sure you have formulated a clear and concise question to answer. You will apply the skills learned throughout the semester to answer this question.

Deliverables: 1) A 5-minute presentation where you walk through your website and main findings. Upload to Quercus or link on your website; 2) A written report with embedded tables and figures that is submitted as a PDF to a final project-specific github repository *and* as a downloadable link on your website. Please see the checklist below for additional details.

The report should have the following sections (elaborate from what was written in the midterm): **Introduction** (provide background on your dataset and formulated question), **Methods** (include how and where the data were acquired, how you cleaned and wrangled the data, what tools you used for data exploration), description of modeling, **Results** (provide final, publication ready tables and figures from your analysis, refer to your website if needed), and a brief **Conclusions and Summary** where you describe your findings.

In your report, please do not include code (so make sure `echo = FALSE`), unformatted output, or dataset summaries (e.g. output from head(), str(), etc.). Summarize model output in tables (e.g. R^2, RMSE, coefficients, cross validation results, etc.) and figures (e.g. variable importance, scatterplot of test obs vs pred, etc.).

Checklist for Final Project

1. **Create a website** (HTML document and all the required files, including figures). It should feature:
    1. A brief description of the project,
    2. Interactive visualizations, also with a description so that people can understand what they are looking at, and
    3. A link to the PDF version of the actual report (i.e. a link to "Download the report.").
    4. Your home page should feature no more than five interactive tables and figures. If you want to include more to showcase your skills, we request you do so by adding extra pages to the website.

       The actual analysis should be included in the PDF report. **The PDF report can refer to visualizations included in the website**.
2. **Upload**, source code, website files, and PDF report, to your GitHub repository.
3. Make sure that the **website**, which is to be **hosted in GitHub pages**, actually **works**, i.e., figures and interactive visualizations are properly rendered when visiting the website.

4. **Have a README.md file** in the upper level of the repository. This file provides general information about the project, like title, brief description, etc.
5. The **README.md file links to the website**, e.g. https://username.github.io/JSC370-finalproject, .
6. Have a **"data" folder** with either the data or instructions about how to acquire it. You should provide the instructions in a README.md within that folder.
7. **(optional)** Your document is **fully reproducible**, meaning you don't have paths to files not shared in the repository. For example, if you are loading a dataset with "fread," it should be something like "fread("https://data.com/dataset.csv")," so you are directly using data online, or "fread("data/dataset.csv")," so the data needed has been shared on GitHub and is contained in the path "data/dataset.csv."