Stat 6021: Project 2

In this project, each group will use the statistical techniques covered in this course to answer a data analysis question. The subject of the project will not be assigned; instead it will be chosen collaboratively by the members of each group. Each group will complete a written report and oral presentation. Each project should feature the following:

- Clear central analytic goal(s) and/or question(s) to answer – the more interesting, the better.
- Identifiable data that can be obtained in support of the project.
- Use of methods learned in this course.

There are several parts of the project as detailed below.

Part 1: Proposal

Each group will prepare a project proposal for review.

- The proposal should be no more than six pages.
- The associated data set for the project should be provided, in one of the following ways: 1) providing the names of the R package and the R dataframe; 2) a link to the data set; 3) a file containing the data, as an Excel spreadsheet, a .csv file, or a .txt file.
- It should provide information about the project that the group plans to conduct, including project objectives/goals, the data that the group plans to use, expected data cleaning, and expected methods of analysis. Your project should involve both linear regression and logistic regression, so clearly state the response variables involved.
- Some data visualizations and related commentary related to the project objectives/goals. At least one visualization related to linear regression, and at least one visualization related to logistic regression.
- The more interesting the project objectives/goals, the better.

- Please note that regression methods assume the observations are independent. One way of assessing whether your observations are independent is to ask if the order of the rows in your dataframe matters. If you can scramble the order of the rows without affecting any structure in your data, then your observations are likely to be independent. If scrambling the rows upends the structure of the dataframe, then your observations are not independent.

- Also note that you have about 3 to 4 weeks to work on this project, so be realistic in your proposal.

- For the categorical response variable, be sure it is binary. We have not covered enough material to tackle categorical response variables with more than 2 classes.

Feedback on the proposal will be provided that will be based on the instructor's evaluation of the challenges that the project presents. Some suggestions may also be provided. Groups will have the option of submitting one revised proposal (but this should be avoided).

Please submit your group's proposal via Assignments (1 upload per group).

Part 2: Group Expectations Agreement

On a single document, put your names and list the expectations you agree as a group to adopt for this project. You can deal with any or all aspects outlined below: preparation for and attendance at meetings, contributing to the project, raising and discussing questions, communicating frankly and respectfully when disagreements arise, etc. Each group member should sign the sheet, indicating acceptance of the expectations and intention to fulfill them.

The expectations are for your use and benefit; they will not be graded or commented on unless you specifically ask for comments. Note that the list should be fairly thorough without being unrealistic. For example, "We will solve every problem completely" or "We will get a perfect score for the project" or "We will never miss a meeting" are unrealistic, but "We will be sure to understand our tasks prior to our next meeting" and "We will try to be as flexible as possible in scheduling meetings" are realistic.

Group work is not always easy; group members sometimes cannot prepare for or attend all meetings due to other responsibilities, and conflicts may arise from different skill levels and work ethics. When groups work and communicate well, the benefits more than compensate for the difficulties. One way to improve the chances a group works well is to agree beforehand on what everyone on the group expects from everyone else. Reaching this understanding is the goal of the Group Expectations Agreement.

I have listed some possible items to list in your Group Expectations Agreement:

- Designate an organizer. The organizer is tasked to keep the group on track and to guide the meeting back on track if the discussion goes off tangent. Your group may choose to rotate who is the designated organizer.

- Designate a recorder for each meeting. The recorder is tasked with documenting the group's questions, discussions, as well as assigned tasks for each group member during

a meeting. Using a google doc that is shared among all group members may be a good idea. Your group may choose to rotate who is the recorder.

- Designate a scheduler who arranges a meeting time for the week. Group members are expected to give the scheduler a list of times they are available during the week. Your group may choose to rotate who is the designated scheduler.
- Inform other group members if you are unable to attend a meeting with good reason.
- Make a good faith effort to share your thoughts and ideas for the project.
- Listen to the ideas and questions of other group members respectfully.
- If you are unsure about a task, to raise your doubts to the group.
- If there are disagreements about how to proceed with the project, to seek direction from the instructor
- If a group member is not contributing or putting in a good faith effort to the project, the other group members will respectfully inform this group member, and potentially, the instructor as well.
- Complete the peer reviews honestly and respectfully.

Please submit your Group Expectations Agreement via Assignments. Every student will submit a document. Documents within a group should be identical. Submission indicates agreement with the document. Failure to do so will result in getting a 0 for Project 2.

Part 3: Report

Your group is to type up a report for this project. One member of your group is to upload the report and the R script via Assignments on Collab. The report is to include the following:

1. An executive summary that describes the high-level results of the analysis. This executive summary should be written in a way that can be understood by a wide variety of readers, including readers with no background in statistics. A way to think about this is how newspaper articles report results from various studies, so avoid technical jargon. This section should be no more than 2 pages.
2. A detailed description for a professional audience. The audience for this section is another classmate your client may hire to review your report. This section should include:
    - Detailed description of the data including challenges, data cleaning before, during, or after analysis
    - Exploratory data analysis and what you learned
    - Clear reasons given for model(s) considered

- Appropriate model diagnostics provided and checked
- Attempts to improve the model, as well as reasons for decisions made on how to improve the model.
- Relevant R output and graphical summaries

3. Discussion of analysis should always be done contextually.

Report Guidelines

- Include the names of the group members and group number in the heading of your report.
- Have sections that are clearly labeled.
- Aim for no more than 30 pages. If you go over this limit a bit, that is fine.
- Do not use appendices as a way to work around the page limit. Anything that belongs in the main body of the report should be in the main body and not be tucked away in an appendix. I will not read anything in the appendix.
- The report should contain correct grammar, clear explanations, and professional presentation.
- I should be able to repeat your analysis without looking at your R code.
- Your report does not need to include any R code. Relevant output from R (e.g. graphs, results from hypothesis tests, etc) should be included if the output is referenced to in the report.
- The text in your document should be readable after printing out. Difficult to read documents will be penalized.

Please submit your group's report via Assignments (1 upload per group).

Part 4: Presentation

Each group will give a presentation that is no more than 15 minutes long. This presentation should be designed to be understandable by anyone familiar with the course material, but who has not read the project report. Each group is free to organize who talks about which topics during the presentation. Not everyone needs to talk, but all group members should be active contributors to the presentation materials. The presentation is to include the following:

- What are the goals/questions of this project?
- What are the results of the analysis? (include any recommendations)

- What is the nature of the data used?
- Does your project answer interesting and/or important questions? Explain.
- Use of graphical summaries to explore the data.
- A description of the model building process. This part will be much more technical. The previously listed items should be aimed for a non technical audience.

Presentation Guidelines

- The presentation should use PowerPoint or something similar as a visual.
- Each slide should be clear and easy to read.
- The presenter should be clear, with good pace and logical flow.
- R output should be clearly labeled.

Please submit your group's presentation via Assignments (1 upload per group).

Part 5: Feedback on Classmates' Presentation

After you have completed your project, you will need to review another presentation in Collab:

- View the presentation you have been assigned to review.
- Provide anonymous, constructive, and concise feedback.
- Your feedback will be evaluated on the helpfulness and conciseness of the feedback you provide.
- Half the points will be awarded for turning the feedback in on time.

Each student will submit the feedback via Assignments.

Part 6: Self- and Peer-Evaluation of Group Participation in Project 2

- You will anonymously evaluate each group member's contributions to the project.
- Complete this by giving an honest of your own performance and that of your group members in project 2.
- Half the points will be awarded for turning the evaluation in on time.

Each student will submit the group participation evaluation via Test & Quizzes.