

## Data Science Project Details

The data science project will consist of 5 separate graded components, described below:

**1. Project Proposals (5%):** You will be required to come up with multiple project ideas and do some preliminary investigating to determine how feasible the projects will be to complete and what kinds of questions can be answered. More details on this component are given in a separate file.

**2. Project Updates (5%):** Once projects and groups are assigned, you are expected to begin work, meeting regularly with your group members to tackle the project. Part way through the semester – usually about 70% of the way through the course – you will provide a short ( $\sim 5$  minute) oral presentation to the rest of the class outlining what you’ve done so far and what you plan to do going forward. Your presentation should consist of the following components:

- Overview of problem and data ( $\sim 1$  minute)
- Overview of methods and results thus far ( $\sim 1$  minute)
- Plans for future analysis ( $\sim 1$  minute)
- Observations/findings thus far & other potential directions ( $\sim 1$ -2 minutes)

These updates do not need to be overly formal, but they should efficiently convey the problem and what progress has been made thus far. You are also free to discuss potential challenges and ask for advice or opinions from the rest of the class in terms of how to best move forward. Each group member should contribute roughly the same amount to these presentations.

**3. Oral Presentation (15%):** The course will conclude with each group giving a formal presentation of their projects designed to last approximately 15-20 minutes. Each group member should plan to speak for approximately the same amount of time. Each presentation should consist of two primary components:

### Part 1: Technical Details (10-12 minutes)

- An overview of the key problem(s) of interest
- The dataset(s) used to investigate these problems as well as any issues related to data collection, missing data etc.
- The statistical and data science techniques applied to the data
- The results of these techniques

- A summary of key results and findings, as well as an explanation of any “contradictory” findings

## **Part 2: General Findings/Suggestions/Recommendations (3-5 minutes)**

- A conclusion that goes beyond simply the model results. That is, based on what you’ve observed with the different techniques and approaches, what do you really believe can safely be concluded? Here you need to convey the results to people with little to no statistics background. What would you recommend if someone wanted to make practical use of your results.

There should be something of a brief “reset” between the two parts of your oral presentation. The basic idea is that in part one, you should communicate to the rest of the class as you typically would and assume that the audience knows all of the technical terminology that you do. In part two, **you must clearly communicate your results, state your findings, and make recommendations to an audience without relying on jargon or specific pieces of model output.** Note that the oral presentation should not be overly detailed, but is designed to present a concise summary of the questions, data, methods, and results.

**Grading Criteria:** You will receive both a group grade and an individual grade. The group grade will be based on:

- **Overall Length:** The entire presentation should fit comfortably within the 15-20 minute timeframe.
- **Overall Organization and Message:** The presentation should follow the general outline above. The purpose of the study, techniques employed, and results should be clearly displayed and expressed.
- **Non-technical Communication of Results:** This is primarily in relation to part 2 of the presentation. The presentation should clearly communicate results in a non-technical manner in such a way that an audience unfamiliar with statistics has a clear sense of what was found and how those results could potentially be acted upon.

The individual grade will be based on:

- **Slide Quality:** The slides should be easy to follow and display the appropriate information. You’ll need to think about striking a balance between providing enough details but not so many that the overall point is lost. You (personally) should create all of the slides from which you plan to present.
- **Individual Length:** Each group member should speak for roughly the same amount of time.
- **Presentation and Clarity:** You should speak at an appropriate volume and pace and the topics you discuss should be clear and easy to follow.

**4. Written Report (20%):** Each student is required to put together a final 8-10 page project report. This is where you should go into more details on the specifics of the project. Each written report must include the following sections:

1. Introduction: An overview of the problem of interest and details of the specific dataset as well as a clear description of the problems of interest. ( $\sim 1$  page)
2. Methods Overview: A summary of the methods explored and the various approaches that were considered. ( $\sim 0.5 - 1$  page)
3. Method Details: This is where you should provide details on the methods that you personally were involved in fitting and how those results influenced other models that were attempted. You may include plots here, but they should not overwhelm the document. If you need (or would like) to provide many plots and/or code, these can be put into an appendix. ( $\sim 3-5$  pages)
4. Summary of Results: This should provide an overview of all of the results obtained by everyone in your group. Comment on overall trends, contradictions between models, etc. You can include a table here if it helps summarize the findings. Note that this should only be a fairly concise summary (i.e. the “What”; the “Why” should come in the next section). ( $\sim 1$  page)
5. Conclusions and Takeaway: This is where you should expand upon the results observed in the previous section. Describe what you feel can safely be concluded. Which models might you prefer? Why would some models agree/disagree? What results might be most relevant and how might you report them to someone without a strong statistics background? You may provide extra plots/code to investigate anything further that seems relevant after looking at the results from the previous section. As in Section 3, continue these in an appendix if necessary. ( $\sim 2-3$  pages)

Sections 1, 2, and 4 can be written as a group (i.e. each group member can and should turn in the same thing for these sections). Sections 3 and 5 must be done independently. Section 3 should provide the details of your individual contributions to the project and Section 5 should summarize your takeaways of the results. A separate grading rubric will be provided.

**In addition to the above sections,** you should also include (as part of the appendix) a section clearly labeled as “Peer Evaluation”. At a minimum, you should include the name of each of your group members along with a rough “percent effort” for that person. For example, if there are four people in your group and you feel that everyone contributed equally, you would say:

- Your Name: 25%
- Name Two: 25%
- Name Three: 25%
- Name Four: 25%

You may also include a paragraph with specific details if you feel it appropriate. Please understand that students in this course have a wide variety of backgrounds and experience levels.

Please do not downgrade group members for struggling with particular aspects of the project if there was clear commitment and effort on their part. On the other hand, if one member was not responsive and contributed very little, this is something you should bring to my attention.

**5. Peer Evaluations (5%):** For each project (other than your own), you will be required to provide an evaluation of the work presented. Standard rubrics will be provided. You should evaluate the work on the following :

**Problem Statement:** Is it clear what the overall purpose of the project is and what specific questions are trying to be answered?

**Data Description:** Is the data described clearly? This may include things like the size, source, quality, and missingness of the data, as well as the types of predictor variables and response and how they're measured.

**Methods Description:** For the modeling methods that were attempted, are they described appropriately? Is enough information given to understand how the process was carried out without being tedious?

**Methods Execution:** For the modeling methods that were attempted, are they performed correctly?

**Thoroughness:** Were all relevant modeling methods attempted and discussed? Are there obvious things that could have been attempted that were not?

**Overall Clarity & Organization:** Was the presentation well-organized? Did the order in which things were discussed make sense? Were the methods and results clearly stated?

**Slide Quality:** Were the slides easy to follow? Was enough information provided without being overly busy or difficult to read?

**Appropriateness of Summary & Conclusions:** Are the stated conclusions appropriate given the models and results provided?

**Non-technical Explanation:** Was the non-technical summary provided appropriate? Did it cover the important takeaways in a clear manner without resorting to technical jargon?

**Overall Ranking:** How did this presentation rank according to what you would expect and relative to the others presented? 1 = well below average, 2 = below average, 3 = average, 4 = above average, 5 = well above average.

In addition to these numerical rankings, you'll also have the ability to adjust scores in the direction you feel appropriate. For example, you might chose to subtract some points if you felt the group took too many shortcuts in cleaning the data prior to the analysis. On the other hand, you might add a few points if, for example, you think the group dealt with a particularly tough problem or was particularly thorough in their analysis. If you choose to adjust a group's

score, you must provide a justification for your decision.

Finally, you'll have the opportunity to add written comments for each group. You should focus in particular on any shortcomings you see and any ways in which you feel the project could have been improved.