Final Project EDA Report Rubric

HTML report, due Monday, April 28th at 11:59 PM ET

This report (produced from Quarto) will act as a public-facing document of your work. Though due on April 28th, this report will be made public for the Statistics & Data Science department to see on April 30th. After that date, we will leave the reports publicly available unless any member of the team asks for the reports to be taken down after finals week.

HTML reports should start with a short paragraph and/or bullet-point list describing the data, followed by a short paragraph and/or bullet-point list describing the main research questions of the project (**you are required one research question per group member, so that groups of size three have 3 research questions while groups of size four have 4 research questions**). After that, the middle part of your report should consist of your graphs (at least six or eight depending on your group size), bookend-ed by statements motivating how the graphs fit into the overarching narrative of the project as well as their interpretation and main takeaways. For example, an HTML report could take the following format:

> We wanted to learn about [INSERT RESEARCH QUESTION 1], which suggests we should examine [NSERT RELEVANT VARIABLES].
>
> [INSERT YOUR AMAZING GRAPH HERE]
>
> The above graph suggests that [INSERT MAIN TAKEAWAYS, INTERPRETATIONS, AND CONCLUSIONS].
>
> We also wanted to better understand [INSERT RESEARCH QUESTION 2]. To do this, we plotted [INSERT RELEVANT VARIABLES] using [INSERT GRAPH CHOICE EXCEPT PIE CHARTS HERE]:
>
> [INSERT YOUR NEXT AMAZING GRAPH HERE]

And so on. (Please do not just copy-and-paste the above format; it's supposed to be a broad idea of how you should structure your document, rather than literal instruction.) Additionally, **you need to include at least one formal statistical analysis to complement a graph in your report**, e.g., t-test, KS test, regression analysis, etc. At the end of your report, you should include a short paragraph (or paragraphs) discussing the main conclusions of your project and potential directions for future work.

Grading Rubric (100 points)

We will be grading final reports based on *basic items* and *detailed items*. The *basic items* are in bold below, and will generally be graded in an all-or-nothing fashion. For example, the first *basic item*, "Does the project provide a description of the dataset?", is worth 5pts. If the answer to that question is **Yes**, then the team is awarded 5pts; if the answer is **No**, the team is deducted 5pts. The basic items are meant to set a minimum grade that every team should receive based on very minimum effort; the basic items add up to 60pts. The remaining 40pts are divided into *detailed items* that we will rate on a 0-to-5 scale. Each *detailed item* is under a *basic item* and is worth 5pts each. Thus, you'll find 8 *detailed items* below, adding up to the remaining 40pts.

Basic item (5pts): Does the report provide a description of the dataset?

- Detailed item (rated 0 to 5): What is the quality of this description? Does it clearly communicate what the rows and columns (i.e., subjects and variables) are in the dataset in a way that is understandable to a CMU undergraduate?

Basic item (9pts): Does the report have the required number of clearly stated research questions? You need one research question for each member in the group, so that groups of size three have 3 research questions while groups of size four have 4 research questions. (Minus 3pts for groups of size three or 2.25pts for groups of size four, for each question that is missing or not clearly stated.)

- Detailed item (rated 0 to 5): What is the quality of the research questions? Are they well-motivated by real-world/scientific interests? Or are they shallow? For example, a shallow question would be in the form, *What does the distribution of this variable look like?* A more interesting question would motivate why we would like to inspect particular distributions with real-world context.

Basic item (16 pts): Does the report have the minimum number of required graphs (4 points) that also follow the three report requirements listed in the 'Final Project Requirements' file on Canvas? (each worth 4pts)

- Detailed item (0 to 5): What is the quality of these graphs? Are these graphs easily readable, interpretable, and properly labeled?

- Detailed item (0 to 5): Are the graphs well-motivated, given the research questions of the project? In other words, do the graphs address/answer the research questions of the project, or do they only provide tangential (or even irrelevant) details?

Basic item (10 pts): Does the report provide further descriptions/interpretations for each graph? (minus 2.5pts for each description that is missing, up to 10pts)

- Detailed item (0 to 5): What is the quality of these descriptions? After reading these descriptions, is it crystal clear what is being displayed in the graph and what the main takeaways are?

Basic item (5 pts): Does the report include any kind of statistical analyses and/or summary statistics/tables to complement their graphs?

- Detailed item (0 to 5): Are the statistical analyses appropriate given the type of data and research questions of interest? Are the statistical analyses interpreted correctly?

Basic item (10 pts): Does the report provide some clear conclusions that can be made from their graphs and analyses?

- Detailed item (0 to 5): What is the accuracy of the claims being made? Are the claims well-supported by the graphs and analyses presented? Are the conclusions well-aligned with the research questions of the project?

Basic item (5 pts): Does the report discuss questions that have not been answered by the project, but could be answered with future work?

- Detailed item (0 to 5): Did the team provide adequate reasons as to why these questions were left as future work (e.g., they need more data, need more nuanced statistical techniques they haven't learned, etc.)? Are these future-work questions well-motivated given what the team has completed for this project?