

# Group Final Report Instructions

## Assignment Submission Instructions:

- This is a group assignment. Only one team member needs to submit this report.
- Clearly **indicate your project group number** at the top of your report.
- Use the **correct submission page based on your group number**, otherwise there will be a 10% penalty.
- Create an electronic report with a maximum of 2000 words (excluding citations) using Jupyter Notebook.
- **Due date:** Dec 6, 2025
- Do not try to include all tasks completed by each group member in their individual assignments. Reach a consensus among all team members on the final content of the report.
- **The final report is an extended version of assignment 1 and 2.** Thus, even if the selected content may have been developed by only one student, everyone in the group should be able to **contribute with some additional content** and completion of the final report. Consult your TA if you need further clarification.
- You must submit **two files**:
  - the source Jupyter Notebook (`.ipynb` file)
  - the rendered final document (`.html` file)

## What Sections to Include?

**(1) Title Page:** title, group number, group members, course and date

**(2) Introduction**

Start with relevant background information on the topic to prepare those unfamiliar for the rest of your proposal. Motivate the question you are about to add.

Formulate one broad question for investigation that would require you to incorporate fitting several models in your analysis, using the same response variable and 2 or more covariates. Indicate if the primary goal is inference or prediction. Make sure that the question(s) can be answered with the data available. For example, you won't be able to quantify the "effect of X on Y" or "the influence of X on Y" or "how X affects Y" using the methods learned in class and observational data.

If available, align your question/objectives with the existing literature on the topic. You can add a reference to a scientific publication if available and listed in the References section (not mandatory).

### (3) Methods and Results

In this section, you will include the following subsections:

#### a) Data

- read the data into R using reproducible code (i.e., from an open source **and not a local directory in your server or computer**)
- include a citation of its source
- include any information you have about data collection (e.g., observational vs experimental)
- describe the variables as done in your Stage 1 Report.
- if (absolutely) needed, indicate which variables will be pre-selected (or dropped) and provide a clear justification of your selection.
  - If your goal is prediction, you should keep all variables in the analysis and perform variable selection based on model performance.

#### b) Exploratory Data Analysis (EDA)

- Clean and wrangle your data into a tidy format (review [Tidyverse's style guide](https://style.tidyverse.org/syntax.html) (<https://style.tidyverse.org/syntax.html>) if needed)
- Include 2 effective and creative visualizations
  - explore the association of some potential explanatory variables with the response (use colours, point types, point size and/or faceting to include more variables)
  - highlight potential problems (e.g., multicollinearity or outliers)
  - You may utilize sub-plots as you did in Stage 1 Report.
  - Use easily readable main/axis/legend titles, appropriately sized and without any underscores.
- Transform some variables if needed and include a clear explanation (e.g. log-transformation may be useful when outliers are present)
- Any summary tables that are relevant to your analysis (e.g., summarize number of observation in groups, indicate if NAs exist)
- Be sure not to print output that takes up a lot of screen space!

Your EDA must be comprehensive with high quality plots.

#### c) Methods: Plan

- Describe in written English the methods/models you used to perform your analysis from beginning to end.
- Provide a detailed justification of the method(s) used. The analysis must be based on *methods learned in class*.
  - Make sure that the analysis responded the question posed and that the proposed method is appropriate for the characteristics of the data.
- If a variable selection method is used, you need to describe and justify the method. Furthermore, explain what data will be used, and how final model will be chosen.

- Include a careful model assessment plan relevant to your goal (i.e. diagnostics and/or evaluation, however appropriate), with justifications.

#### d) Code and Results

- all the analysis code, from reading the data to visualizing results, must be based on clean, reproducible (e.g. read from an open source **and not a local directory in your server or computer**), and well-commented code.
- Include no more than 3 visualizations and/or tables to summarize and highlight your results. Ensure your tables and/or figures are labelled with a figure/table number and readable fonts.
  - You may utilize sub-plots as you did in Stage 1 Report.
  - Use easily readable main/axis/legend titles, appropriately sized and without any underscores.
- Make sure to interpret/explain the results you obtain. It's not enough to just say, "I fitted a linear model with these covariates, and my R-square is 0.87".
  - If inference is the aim of your project, a detailed interpretation of your fitted models will be required, as well as a discussion of relevant quantities.
    - For example, which coefficient(s) is(are) statistically significant? What are some hypothesis tests of interest? Interpretation of coefficients, how does the model fit the data? among other points.
    - Also explain briefly the key differences between your fitted models.
  - If prediction is the aim, you must highlight the key outcomes from your model fitting/selection/prediction in written English.

#### (4) Discussion

In this section, you'll interpret and reflect on the results you obtained in the previous section with respect to the main question/goal of your project.

- Summarize what you found and the implications/impact of your findings
- If relevant, discuss whether your results were what you expected to find
- Discuss how your model could be improved
- Discuss future questions/research this study could lead to

#### (5) References

Include any citation of literature relevant to the project. The citation format is your choice – just be consistent. Make sure to cite the source of your data as well.

**A rubric will be added soon on assignment submission page. Check before submitting your report!**