

Final Report Instructions - Big Data in Social Sciences 2025/2026

Group list: [Google Doc](#)

Instructor: Prof. Nicola Barban

Deadline: December 16, 2025

Submission to: n.barban@unibo.it and riccardo.omenti2@unibo.it

A. Essential Information:

1. **Title:** Make an informative title
2. **Author:** Name and student ID for individual work; or group members' names and student IDs for group work.
3. **Keywords:** Include at least three keywords at submission.
4. **Group Size:** The maximum number of participants is 3.
5. **Data Choice:** You may use data from *Programming Lab 2* or choose a completely different dataset. Note that *Programming Lab* focuses on **programming/coding**, whereas this report emphasizes *using quantitative methods to answer a research question*.
6. **Data Source:** Describe the data source, its main characteristics such the population of interest, the sample size, and the sample composition. Mention how these data were found, how to access it (web address), and download instructions.
7. **Group Work Attribution:** If it's a group project, specify who did which part. If work was equally distributed, state that.
8. **File Submission:** If the document is created in Quarto (recommended), produce a PDF file and include the `.qmd` file as an attachment. Otherwise, submit a PDF file and an R script as attachments.

B. Analysis:

1. Provide a brief overview of the phenomenon you're analyzing.
2. State your research question.
3. (Optional) If applicable, specify the hypotheses you aim to test.
4. Describe the data source.
5. Describe the type of analysis conducted, e.g., descriptive analysis, association between variables, data reduction, classification, regression model.
6. Detail the variables used, their type (categorical, numerical, etc.), missing values, and any additional notes.
7. Comment on the results, highlighting key findings related to your research question.
8. Conclude with the contribution of this analysis.
9. (Optional) List of scientific articles referred to in the report.

C. Report Limitations:

1. Report length: Minimum 6 pages, maximum 12 pages.
2. Include a minimum of 4 and a maximum of 8 medium-size graphical elements (i.e., figures and tables).

3. D. Possible Analyses:

- Any statistical models or techniques already known can be used. Examples include exploratory analyses, frequency tables, correlations, linear models, logistic regression, survival analysis, time series, etc.

E. Suggestions:

- Aim the report at an informed but non-specialist audience.
- If categorical data has too many levels, consider deriving a categorical variable from a numeric one.
- If you have temporal data, specify if you're considering a single time point or a time range.
- Tip: You can repeat some numerical or categorical analyses on a subset of data instead of the entire dataset.

F. Some ideas on data sources

- [Google dataset search](#).
- <https://libraryguides.missouri.edu/datasets/public-use>

- <https://www.icpsr.umich.edu/web/pages/>