

Final Assignment

Instructions for the fourth assignment

DEADLINE: 23/01/2024 at 23:59

In the last assignment, you **merge** your results of assignments and the respective revisions you did. You want to analyze your data using descriptive statistics, visualizations and any other method you deem appropriate, to answer your research question. You create a project presentation. Please upload the final version on <https://www.dropbox.com/request/gT56kMNS3fsEHk99w8I9> until **February 23rd**. You should also include your **commented** Python code/notebook and final processed data.

1 Final presentation

- The presentation is short (only **15-20 minutes**). The emphasis is on the content and findings of your research and the hurdles you had to overcome.
- *Structure*: Spend up to 4 minutes to motivate your topic, about 4 minutes about the data collection strategy and description, and 7 minutes on your analysis. Conclude by summarizing your findings.
- *Format*: All graphs and tables should be numbered and have a title (above or below), and a legend and source information.

Questions You are required ask a question to another group during the final presentations.

2 Structure of the presentation

2.1 Title page (including title of work and names of participants in group)

2.2 Introduction ($\approx 15\%$)

- Provide a motivation for your topic. Why is it important?
- Provide your research question and explain why it is interesting.
- Explain how you proceed (which data you collect, which method to analyse the data)
- Summarize your findings

2.3 Background ($\approx 10\%$)

Here you give supplemental information about anything around your case study that you do not analyse yourself. You have the following options:

- Summarize the findings of the literature you cited in your first report.

- Describe the institutional setting (e.g. legislation)
- Describe the market, its size and its evolution (e.g. from fact sheets / other statistics)
- Describe the platform where you gather the data (if your topic is specifically linked to the topic).

2.4 Data collection strategy (≈5%)

This should be an updated response compared to your first assignment, taking our response into account. You should also assess what has not worked as you had envisioned it in the beginning.

- Data source, purpose of this data, access to the data (API, scraping, both, etc.)
- Technical implementation (Python packages, overcoming obstacles like getting blocked). What solved your problem?
- When and how long did you collect? What did you collect / what is one observational unit in the data? How did you generate the sample (searching on the website, filtering, etc.)
- How **complete** and how **representative** is your sample? Does your number of observational units and the composition of the data coincide with other sources? If not, give reasons why.
- Enumerate any relevant steps in data preparation and cleaning.

2.5 Data description (≈10%)

Provide a short description of the data set: How many units of observation do you have? What is one of these units? What are the variables/features you have. Provide a table with descriptive statistics like average, median, standard deviation, and describe them shortly. Consider only the variables that you will make use in your analysis!

2.6 Analysis (≈50%)

- For descriptive tables, charts and map should follow the guidelines discussed in class. Don't have anything in the main body that you think is not interesting. You can still move them in the Annex.
- Every table, chart or map you include here must be described in the text. First describe what can be seen, and then give an interpretation.

2.7 Conclusion (≈5%)

Describe here quickly your main findings, potential shortcomings, and avenues for future research if you would have had more time do this project.