Preparation

- Select one group leader; mark their name with * on the title page.
- Group leader submits all files on Moodle for the group.
- Assign responsibilities for:
  - Data collection
  - Data preprocessing
  - Statistical/data analysis
  - LaTeX report writing
  - Notebook development
  - Final packaging

Data Collection (`Group#_Data.csv`)

- For each dataset, record in a notebook (Markdown cell) and mention in the report:
  - Source URL
  - Date accessed
  - License
  - One-line justification for use
- Optionally save this information in `metadata.txt`.
- Use CSV or a single Excel workbook with clearly named sheets (e.g., `customers`, `transactions`, `metadata`). Combine multiple datasets into a single file.
- Add a README cell in the notebook or in `metadata.txt` that describes column meanings, units, and missing-value codes.

Implementation (`Group#_Code.ipynb`)

- Submit one runnable, well-commented Jupyter notebook that reproduces main results.
- Top markdown cells must include:
  - "How to run" instructions
  - "Environment and dependencies" with pinned versions and fixed seed

```
**Project title:**
**Group#:**
## Overview
One-paragraph summary of the project objective, dataset(s) used, approach,
```

```
## How to run
1. Create a virtual environment: `python -m venv .......`.
2. Install dependencies: `pip install -r requirements.txt`
3. Open the notebook and run `Group#_Code.ipynb` .
## Contact
For questions about reproducibility or data access contact: NAME, EMAIL.
```

- Example requirements:

```
pandas==1.5.2
numpy==1.23.5
matplotlib==3.6.2
```

- Use relative paths for data files. If your data file is larger than 8 MB, host it publicly (e.g., GitHub) and load via:

```
url = "https://raw.githubusercontent.com/<user>/<repo>/main/datasets/yourf
df = pd.read_csv(url)
```

Report Writing (`Group#_Topic.pdf`)

- Use the provided LaTeX report template unchanged.
- Cite all external resources appropriately.
- Plagiarism check: Reports above 15% similarity will be penalized.
- Paraphrase and correctly cite external information, including DOIs or URLs in references.
- File name should be `<Group#_Topic>`, e.g., `Group5_MidCap`.

Submission

- Upload **only** these files to Moodle:

    - **<Group#_Topic.pdf>**
    - **<Group#_Code.ipynb>**
    - **<Group#_Data.csv>**

- Late submissions are not accepted; a zero grade will be assigned for late upload.