# 2025 Mini-Project Instruction

This mini-project is meant to be an evaluation of your understanding coming into the course and should be **completed without collaboration with other groups.** You may ask for logistical help concerning formatting of submission.

Any extra tools in the internet is allowed, including using generative AI for helping in the report writing. However, the result will be read and interprete by human. Therefore human readability and flow of the report is the most important, **the AI-generated content that are poorly organized or edited will perform badly in the evaluation process**; beware of this.

**Instructions (4-5 persons per group)**

In this project, there are datasets provided in LEB2 in the theme of COVID-19 data (curating from open dataset **with some number adjustment from real cases**). The description of each dataset has been provided in the separate files. According to these dataset, please create the comprehensive report using Python/Excel/SQL EDA techniques and/or Looker Dashboard to answer several analysis questions below.

Note that you must register your group member with TA. We allow only 16 groups and only one group that can have 4 members in, so **first-come-first-serve**.

**Analysis Questions:**

Each group will address the following core questions while incorporating their own unique elements by select a specific continent or region of 5-8 countries for in-depth analysis of every questions (please state it clearly in the report):

**1. Regional Vaccination Analysis**

- Analyze vaccination patterns and effectiveness within your chosen region
- Compare vaccination rates, strategies, and outcomes with global averages
- **Personalization Element of each group**: Develop a unique "Vaccination Effectiveness Index" for your region that combines multiple metrics (each group must create their own methodology)

**2. Variant Spread and Impact Assessment**

- Track the spread of COVID-19 variants in your chosen region
- Analyze the relationship between variant prevalence and case severity (using deaths/recoveries as proxies).

**3. Vaccine Manufacturer Comparison**

- Compare the distribution and performance of different vaccine manufacturers
- Analyze if certain vaccines were more prevalent or effective in specific contexts
- **Personalization Element of each group**: Develop a methodology to assess vaccine manufacturer performance in your chosen region

**4. Data Quality and Reporting Analysis**

- Assess data completeness and quality across countries and develop a methodology to handle missing or inconsistent data.
- Create a data pipeline that standardizes and cleans the various datasets for consistent analysis.

**Documentation Requirements:**

**Project Journal in .Docx format:**

- Maintain a detailed journal with regular entries (at least 2-3 per week)
- Document all key decisions, methods, code development, and analysis steps
- Include screenshots of works-in-progress, preliminary visualizations, and team discussions

**Analysis Narratives:**

For each analysis question, provide:

- Clear methodology explanation
- Rationale for analytical approaches
- Interpretation of findings
- Limitations and potential biases
- Recommendations based on findings
- Final submission must include all code, visualizations, journals, and presentations

**Evaluation Criteria**

- Technical implementation quality (35%)
- Analytical depth and insight (30%)
- Documentation thoroughness (20%)
- Creativity and originality of personalized elements (15%)

**Note:** Clear documentation of your unique analytical approach is essential. The project's emphasis is not just on results but on demonstrating your thought process and analytical skills.

**Submission**

Include the following files in your submission:

- Project journal and analysis report have to submit in **.{pdf or docx} Be sure to indicate both of your names and student ID in this file**
- All code or Excel file for calculation of each analysis question. **Be sure to indicate both of your names and student ID in the filename (StudentID_FirstName)**
- The instruction file in **.{pdf or docx}** of how to load the dataset in your code, Excel or Looker. The final data organization is required in this file to explain why and how you treat the dataset.
- Deadline June 2nd, 10.00