
# 2025-HX05-COS30045-Data Visualisation (HCMC)-H1

# Project Stand Up 1

**Instructor:** Mr.Tristan Nguyen

**Student name:** Anh Khoa Nguyen

Arlene Phuong Brown

Thien Thao Vy Tran

**Submitted date:**29/06/2025

**MEETING SCHEDULE**

Regular Meeting Slot: Monday and Thursday at 5:00 PM

Platform: Discord

Frequency: Twice weekly

**WHAT WE HAVE DONE SO FAR**

Our team so far has been focused on establishing a strong foundation for the data visualization project. We have been planning out the entire project process from start to finish, ensuring we have a clear roadmap for successful completion. A significant portion of our time has been dedicated to studying and analyzing the provided dataset, examining its structure, identifying key variables and understanding the data quality and potential insights it contains.

We have conducted several productive team meetings on Discord where we discussed our individual findings about the dataset, shared insights we discovered during our analysis and brainstormed potential approaches to the project itself. These discussions have helped us understand what each team member brings to the project and allowed us to identify our respective strengths for task allocation.

**WHAT WE PLAN TO COMPLETE BEFORE NEXT STANDUP**

Our immediate focus will be completing the planning phase and transitioning into active data preparation. We plan to finalize our project approach and move forward with data cleaning and preprocessing activities. This includes handling missing values, standardizing data formats, and ensuring data quality meets our visualization requirements.

Additionally, we will be permanently assigning specific project tasks to each team member based on our individual strengths and interests identified during our planning discussions. This will ensure clear ownership of deliverables and efficient workflow management moving forward.

**OBSTACLES WE ARE FACING**

Currently, we have encountered minimal obstacles in our project progression. The primary challenge we are navigating is determining the most effective data preprocessing approach for our specific dataset characteristics. We are working through decisions about which data cleaning techniques will best preserve data integrity while optimizing it for visualization purposes.

**SUMMARY OF TO-DOs BEFORE NEXT MEETING**

- Complete final project planning and approach finalization
- Begin comprehensive data cleaning and preprocessing
- Permanently assign individual tasks and responsibilities to team members
- Establish data quality standards and validation criteria
- Create initial data exploration visualizations to guide further analysis

**PROJECT SCHEDULE ESTIMATE**

Week 1-2: Project planning, dataset analysis, and team coordination *(Current Phase)*

Week 3: Data cleaning, preprocessing, and task assignment

Week 4-5: Individual visualization development and implementation

Week 6: Integration, testing, and refinement of visualizations

Week 7: Final presentation preparation and documentation

Week 8: Project submission and presentation delivery

**TECHNICAL DECISIONS AND TOOL DECISION**

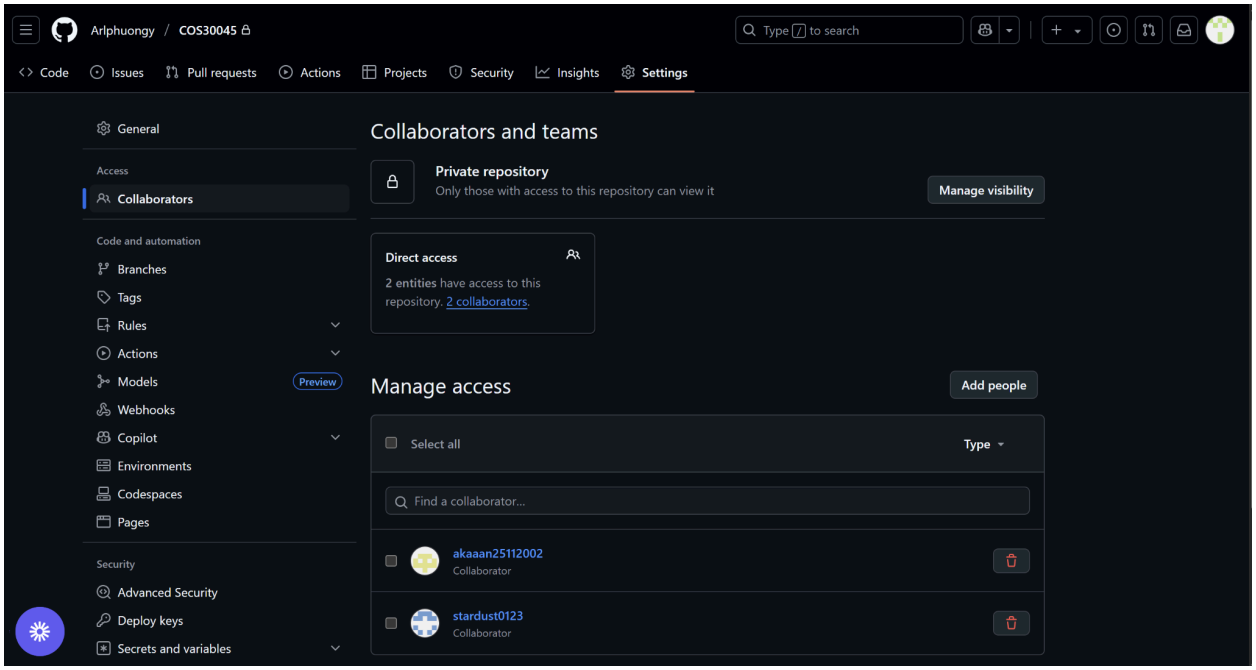
Our team has been actively discussing the technical stack and tools for our data visualization project. Key decisions we are evaluating include:

- **Programming Language**: We have decided to use **Python** as our primary development language, leveraging its robust ecosystem for data analysis, visualization, and application development.
- **Data Processing**: We are comparing **pandas** for its comprehensive capabilities in data cleaning, transformation, and feature engineering. Its integration with other Python libraries makes it a strong candidate for efficient data handling.
- **Visualization Libraries**: We are evaluating **Seaborn** and **Matplotlib (plt)** for creating visualizations. While Matplotlib offers fine-grained control and flexibility, Seaborn provides aesthetically pleasing statistical plots with less code. The choice may depend on the complexity and type of data representations required.
- **Deployment Strategy**: Several deployment options are under consideration. These include:
    - **Power BI**, known for creating interactive dashboards with ease, especially when integrating with enterprise environments.
    - **Streamlit**, a Python-based open-source framework that allows us to quickly build and deploy custom interactive web applications. It provides seamless integration

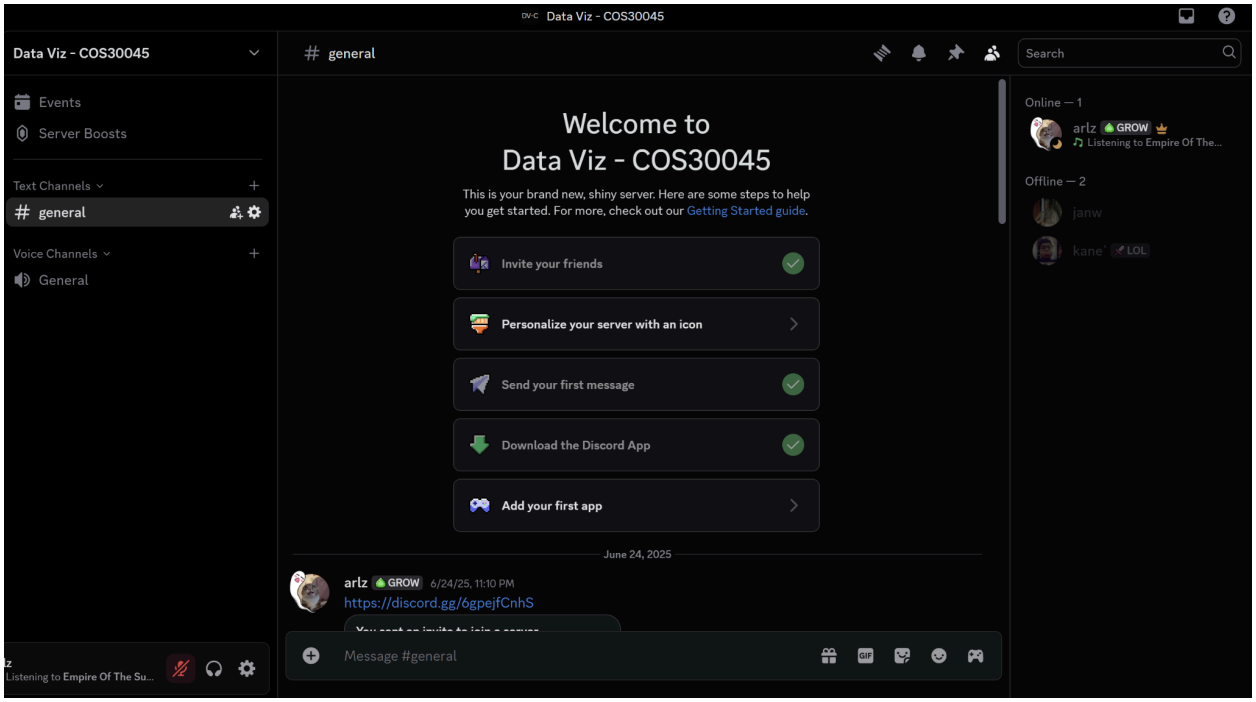
with our Python code and is ideal for showcasing data science and machine learning workflows in a user-friendly web interface.

## SCREENSHOTS

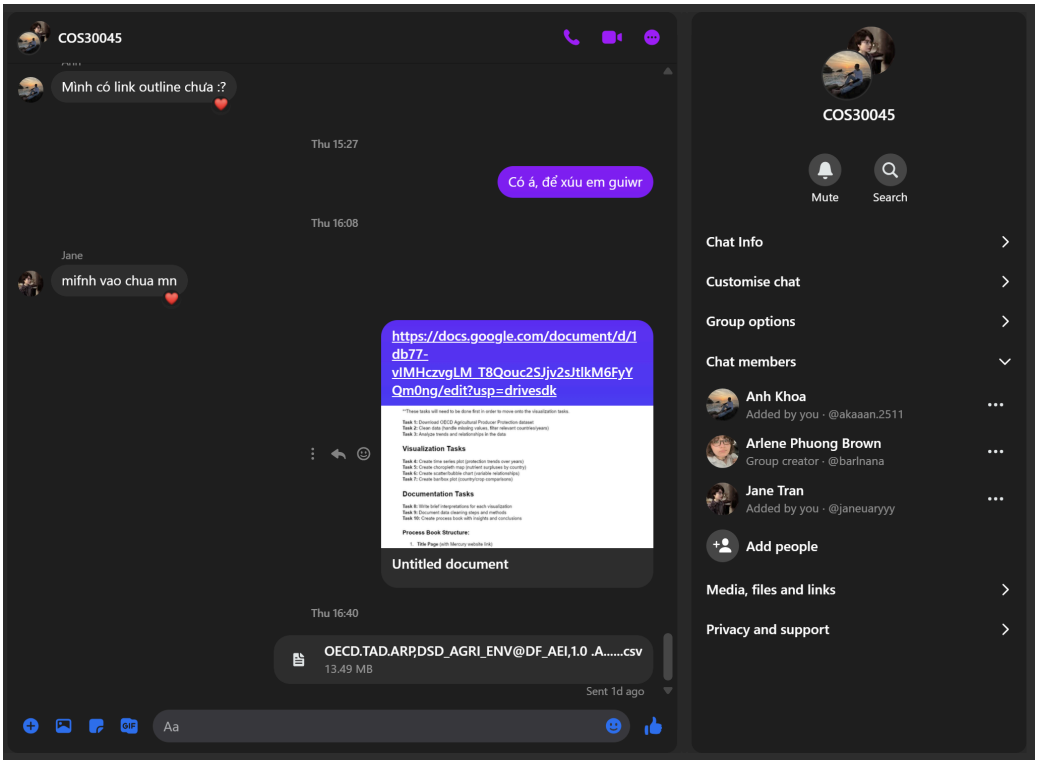
Our shared Github repository



Our Discord communication channel




Our Messenger communication channel (what we use to communicate with each other when we are not in meetings)



Our shared drive