Looking back at your proposal, **in 2 pages max**, answer these questions.

- What exactly has been done? Provide a detailed description of the work completed according to the proposal. (3 points)
- What results do you have so far? Include the results you have so far. (2 points)
- What challenges have you faced so far? Provide a detailed description of any challenges encountered and the solutions you've implemented, if applicable. (2 points)
- What is left to do? Provide a detailed description of what still needs to be implemented, along with a timeline for completing the remaining parts of the project. (3 points)

Completed Tasks

So far, I've made good progress on the setup and first phase of the project. I created a GitHub repository (https://github.com/abstractafua/co2nnections.git) to store all scripts, data, and documentation. This was one of the final deliverables, so I wanted to get it done early and use it as a way to track my progress and organize files as I go.

For data collection, I wrote a Python script to scrape information from the Celebrity Private Jet Tracker website (the main data source mentioned in my proposal). The script collects data on celebrities' CO_2 emissions, flight counts, and distances flown. I also started looking into other possible data sources used in other research papers like FlightRadar24 in case I need more information for the later research questions (especially 2–4).

For Research Question 1 ("What is the average degree of the private jet emission network?"), I cleaned the scraped data, created a node list, and used R with the igraph package to plot the initial emission similarity network.

Results

For Research Question 1 ("What is the average degree of the private jet emission network?", the average degree of the network came out to:

```
> avg_degree
[1] 10.3125
```

This means that on average each celebrity in the network is connected to around 10 others based on emission similarities. It shows there's a fair amount of overlap in flight behavior and that certain celebrities have similar flying and emission patterns. This is an okay foundation to continue my analysis on. However, I'm wondering if the emissions network could display more connectedness.

Challenges

One challenge has been data quality and cleaning. The celebrity tracker data isn't always formatted consistently, so extracting celebrity names from text has taken a lot of time.

Another issue is that right now my similarity scale is too broad, which is creating lots of disconnected components in the network. I'll need to tweak how I calculate similarity so that the network reflects more connections.

Lastly, a big future challenge will be figuring out how to get flight route data (origins and destinations). It's not directly available in the current dataset, so I may have to infer it through the TAIL_REGISTRATION number attribute and data from other sources like OWID or FlightRadar24.

Remaining Deliverables & Timeline

Analysis II → (Research Questions 2 & 3)
Week 4 & 5 (November 10th, 2025 - November 23rd, 2025)

- Implement Robust Data Cleaning script to improve celebrity name recognition
- Adjust similarity scale to produce a more balanced network connectivity
- Finalize and compute centrality measures (betweenness, closeness, degree) to identify the most central "flyers."
- Conduct analysis for Research Question II and III

Analysis III → (Research Question 4 & Conclusion/Key Findings)
Week 6 (November 24th, 2025 - November 30th, 2025)

- Investigate potential shared flight paths.
- Integrate other CO_2 emission datasets to contextualize findings.
- Conduct final analysis of results for Research Question IIII.
- Start integrating github results and findings into the final report template

Review Phase → Week 7 (December 1st, 2025 - December 7th, 2025)

- Peer Review and feedback
- Finalize Github documentation and files.