

# Week 2: Team Log

This week we finalized the project proposal and got feedback from UDL on the proposal. After reviewing the comments and editing the proposal, we submitted the final copy on Friday. The data from the influxDB database was exported and added to the google drive to be able to use in our initial code. Another client meeting with UDL was held to clarify the project objectives and the project overview flowchart was changed to reflect this. Jiachen (from UDL) also held an additional Exploratory Data Analysis (EDA) meeting with us to help us better understand the data and where to look for the metadata. A GitHub workflow plan was also created in order to be prepared for code sharing effectively. Lots of initial research was also done to look into the machine learning aspect of the project, including comparing different clustering methods and looking into what distant measurement is most appropriate. Connor was able to get sklearn running in Databricks and we all played around with clustering in Databricks using the data we pulled from the database. In terms of meetings we were able to stick to our schedule and every meeting ran smoothly and on time. One thing we agreed on as a team was that we need to make our tasks more granular (more specific) in order for them to be completed in our weekly sprints. This is because we noticed the tasks were too general and weren't getting completed, idling in the *In-Progress* section on our Kanban board. Overall, all of our organizational tasks for the project have been completed and we are starting to dive into the data and research more of the specifics of the project.

## Tasks for the Next Week

1. Identify Relevant Features
2. Research Data Flow Connections
3. Transform Data for ML Tasks
4. Develop a tool to identify missing information
5. Develop a tool to populate missing information
6. Identify NRCAN's classifications