Introduction: This can be copied from the proposal.

This paper aims to estimate building energy efficiency using publicly available data (street & aerial view images, building footprints, and satellite-derived Land Surface Temperature). For ground truth labels, the study used the EU's EPC registry data. If results are reliable, such a model could be used to identify buildings or even general geographical areas where buildings are energy-inefficient. This would be immensely useful for sustainable building initiatives and decarbonization plans—with such a tool, one could quickly and cheaply identify potential high-yield retrofits compared with traditional methods like on-site assessments. We chose this paper not only because of its potential societal benefits, but also because it would allow us to apply many concepts we've covered in class. The paper combines many different modalities of data as well as multiple types of machine learning models, giving us the opportunity to work with all of them.

Challenges: What has been the hardest part of the project you've encountered so far?

The most difficult part of the project has been processing the data—we've experienced significant slowdowns we weren't expecting due to the amount of time it took to first gather all the data, clean it, and rewrite and compile it in the form of numpy arrays (probably over 30 hours total). However, after compiling the three cleaned, reformatted datasets, training the model really only took less than an hour to get good results.

Insights: Are there any concrete results you can show at this point?

- How is your model performing compared with expectations?
    - Our model performs within a percentage of the original study's results for precision, recall and F1 for the MLP head. However, the results for our linear head differed more significantly, performing about 5% worse for precision and F1 but 5% better for recall.

Plan: Are you on track with your project?

- What do you need to dedicate more time to?
    - We need now to conduct the ablation studies in order to be completely done. We also would like to do more experiments, possibly by building and training the other types of models to replicate the baseline comparisons they made in their study.
- What are you thinking of changing, if anything?
    - We would like to try and get higher metrics for our model, if possible, by possibly tuning hyperparameters a bit more.