What dataset you are planning to work with

We will be working on the San Francisco Crime Classification Dataset (https://www.kaggle.com/c/sf-crime)

Who your partner is

Group Members: Jee Hyun Kim, Taejin Kim

For each group member: what are you hoping to learn from this project? These goals could be specific data manipulation skills, software engineering skills, or communication skills.

TJ

From this project, I am hoping to get familiar with using datasets and improving my model to get a higher score on kaggle than I did in my warmup project. This process will include cleaning data more effectively, altering variables to align with the machine learning process, and write clean efficient code that is easy to comprehend. I also want to continue to apply matplotlib and seaborn to visualize my findings effectively.

JH

1. learn about efficient method for cleaning data and apply them2.
2. learn how to deal with bigger dataset
3. think about how the concept of space (coordinates, address, police district) and time are interrelated to the recorded crimes (caught cases)
4. become a better pair programmer
5. understand the workings behind sckit-learn
6. apply various visualization methods (especially seaborn)

How do you foresee this project fulfilling the learning goals identified above. Be specific. If there are adjustments that could be made to bring things into better alignment, let me know (I won't guarantee that I will allow it, but it's worth a discussion).

TJ

The project in its current state will offer plenty of opportunities for me to aim for my learning goals. The project seems very similar to the warm up project, just with a lot less guidance. Therefore, I will be able to explore data very visually, clean and manipulate data, and apply machine learning to attempt to get a higher score than 75%.

JH

1 & 2: I would like to learn not just about writing the shortest code but how to process the data such that it is the most efficient way. I think this won't just apply to this project but it will be a learning process throughout the entire class.

3: As we work on data exploration, I guess I will be exploring these concepts

4: We will be doing pair programming so both of us will understand/know everything that is going on in the code, and the experience will help us become better pair programmer

5 & 6: As we work on the model iteration, we will be understanding (more!) and using seaborn