Meeting with Risako on 4/11 3:30pm-3:45pm:

High Level overview of project: We want to scrape food related text-based only tweets (i.e. the food was very spicy) and determine a corresponding food-related emoji that would fit that tweet based on manual inspection and creating a model and comparing the result between the two.

Current status with questions:

* Compiled a list of 75 keywords to scrape emojis with, about 40 are related to taste and 35 related to texture
  + Are all 75 keywords audacious (in accordance to time constraint)?
    - Yes, 40 should be sufficient.
* Used snscrape to scrape about 700 tweets using ~40 taste adjectives to describe
  + Accomplished by using keyword adjective + “food” (i.e. acidic food, bitter food) for each word
    - How can we make this dataset more broad?
      * Risako recommended including more searches rather than just food, use “snack” or “meal”
  + Is the date of tweets important?
    - Yes, should use a year round sample and show in report of certain emojis possible appearing more often based off the time of year (for example, an egg emoji being more popular around Easter
* Compiled a list of 60 food/drink emojis to use as a subset of data the model can respond with
  + Is 60 too much?
    - Yes, honestly something around 5-10 could even be sufficient. Remove unpopular emojis such as cucumber.

Next Steps:

* limit emoji dataset to less than 20 emojis
* scrape more tweets using broader keywords “snack” “meal”
  + also need to take note of adjectives such as “Bittersweet” which may not have as many tweets as “spicy” and record that in report
* manually label data with predicted emoji
* incorporate our dataset and ask chatgpt to respond with a food-related emoji
  + Will most likely need to do a little more secondary research as we are just using chatgpt to ask for a response. Need to come up with a more efficient method of looping through our tweets and using the provided emoji dataset
* Create poster
* Create report