I. The Problem

- **What is the problem you want to solve?**
- I want to find a way to execute discrimination between negative, neutral, and positive reviews of restaurants more efficiently and accurately.

II. Stakeholders and Significance

- **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**
- Being able to discriminate between negative, neutral, and positive reviews can enable restaurants to better understand their customers and help to inform customer acquisition and maintenance efforts.

III. The Data

- **What data are you using? How will you acquire the data?**
- The dataset, which contains over 6.5 million yelp reviews, is available here: https://www.yelp.com/dataset.

IV. My Approach

- **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**
- I will approach this as a supervised classification problem, where the labels will be derived from number of stars. The goal is to divide reviews into the categories 'negative', 'neutral', or 'positive'. I will divide the Yelp dataset into training and test datasets. As a baseline, I will first construct at least two simple models (e.g. logistic regression and multinomial naive bayes), and then subsequently consider constructing other models to improve upon the baseline.

V. Deliverables

- **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**
- My deliverables, as required, will include all jupyter notebooks I develop, a final report, and a presentation slide deck.