

# Final Project Requirements

*Charley Ferrari*

*July 10, 2016*

I will be using the Yelp Challenge Dataset to create a recommender for the restaurants and users included.

Unlike my project 4 dataset, this is a very straight forward recommendation set. It includes users and restaurants in a sparse matrix, with numerical ratings users give to restaurants. In this way, the degree to which a user likes a restaurant can be inferred, and the matrix can be filled in.

This is also a highly contextual database. Restaurants have metadata, including location and type of food, that can be used in a content based scheme, which can be used to create a content-based recommendation system. Latitude and longitude can be used as actual distance measures, while other features can be used to calculate item to item similarity.

This is a large dataset, and I plan on bringing the data into spark. Depending on how much data munging is needed I may turn to Python, or do everything in the sparkR environment.