CS 109a Final Project: Recommendations

Group #20:
Maciej Holubiec, Paul Von Chamier, (potential members to be added)

Have you met with group members?
Yes

Have you met with your TF?
Yes

Project statement

We want to predict the ratings for restaurants for different users. In order to do that, we want to check what affects the ratings of different types of restaurants by different types of user. We want to account for different cuisines, price levels, locations. We might narrow down the restaurants to a given area if it turns out that the dataset is too large or the relationships not significant enough.

Preliminary EDA

- Database is big so we will have to sample
    - Sample the data so that we have same users across three data sets
- Show how different latent factors affect ratings
    - Spiciness of food, type of cuisine, price, etc
- Check collinearity between different latent factors and with the respose variable
- Try to narrow down the list of latent factors for the models using plots and graphs (hopefully to the point where PCA is not needed so we can still interpret the model)
- Try to determine whether to use regression with polynomial terms or interaction terms for the predictors we chose

Timeline:
- November 20th
    - Review references – Paul and Maciej
    - Completed baseline models
        - Regression baseline model – Paul
        - Averages baseline model – Maciej
    - Completed EDA – Maciej
- November 28th
    - EDA summary and milestone 3 submission - Paul
- December 7th
    - Project due