Springboard Data Science Career Track | Nikki Seegars
Capstone Project 2 Project Proposal

Topic: Yelp Restaurant Recommendations Based on Similar User Ratings
Data Source: Yelp.com
Link to Data: https://www.yelp.com/dataset
Updated: 2018
Number of Features: 52
Number of Records: 6,685,900
Problem Type: Recommender System
Customer: Restaurant Review Website

Yelp is an online business review company allowing users to rate local businesses and share reviews of their experiences. On the Yelp website, there is a large subset of data provided by the company with over 6 million reviews for10 metropolitan areas. The data is split into six json files and contain 52 features including business name, user id, review, star rating, and list of user's friends.

The problem to be solved in this project is recommending restaurants to individual users based on their geographical location, friend's star ratings and reviews. Potential clients include restaurant review websites, such as Yelp or Google reviews, and their users. The client will have an improved recommendation system for its users. The users will be able to quickly discover restaurants they have yet to visit with confidence that the experience will be enjoyable.

For this project, only the restaurant text data will be used for analysis. The network connectivity of the users and friends will be considered along with other similarities among the users. Collaborative filtering with user and item similarities will be implemented. A recommendation engine will be built by using matrix factorization. Several metrics including root mean squared error and mean reciprocal rank will be calculated to evaluate results from the recommendation engine. Deliverables for this project include a python code and a slide deck.