

Approach

1. Group businesses by city, since users most likely won't want to travel far for a business (so for a given user, it is practical to only consider those businesses which are nearby).
2. Group users into those city-groups based on the number of reviews they have written for businesses in a group, since that indicates that they probably live near to that city. Additionally, if the user's current location is provided, then that can be used to directly associate them with a city.
3. (Perhaps) pick one city to analyze & generate recommendations for, to decrease the size of the dataset if it is difficult to be managed at full size (TBD).

Using metadata to improve the recommendation

1. weighted by location - consider closer businesses to be more important
2. weighted by review time - consider more recent reviews to be more important
3. weighted by business category - consider the desired category of business by the user
4. only recommend businesses that are currently open

CSV columns to extract from the JSON raw data

1. User
 - a. gPlusUserId
 - b. currentPlace (just the city, the first element in the list)
2. Business
 - a. Name
 - b. Address
 - c. gPlusPlaceId
 - d. Latitude (from gps)
 - e. Longitude (from gps)
3. Review
 - a. Rating
 - b. Categories (one hot encoding?)
 - c. gPlusPlaceId
 - d. unixReviewTime
 - e. gPlusUserId