The method we will use is the Pearson correlation. This determines a relationship between the user we are trying to recommend to (we will call this the user) and other raters that have ranked similar things to that user. The approximate steps of applying this in the most naïve way possible are: 1) Determine the relationship between each rater and the user. 2) Select all users above a certain relatedness threshold. 3) Collect the average rating that similar raters gave to businesses the user hasn’t rated yet. The recommendations will be saying “users with similar tastes to you liked this business n percent of the time” where n is the average calculated in step 3. As described, the algorithm is inefficient, especially for a large number of users. One of our goals in the project will be to make this algorithm run quickly so that new recommendations can be calculated in a reasonable amount of time. We will look into steps like only selecting a subset of the data that might still give us good results.

There are also modifications that we will need to make to get the most accurate results. The threshold in step 2 will have a great effect on the accuracy of the results. If it is too high, we might not get a good sample size very often. If it is too low, we might not use users that are related enough to the querying user. We also need to decide what do to with users that have a high negative relationship to the querying user. Will these users still be reliable for making predictions? Will they be reliable at the same threshold as positive relationship users will?