Problem discussion:

Assume a client is looking to open up a pizza business in Toronto. Which neighborhood should they choose?

Data:

I will use Foursquare API to collect information about local venues in each Toronto neighborhood. I will find characteristics of neighborhoods with pre-existing pizza joints (e.g. the other types of venues within that neighborhood), creating a venue one-shot matrix.

With this, I can find the neighborhoods without pizza places that are most similar to the neighborhoods with pizza places (lowest dissimilarity) to make a recommendation. Furthermore, I can utilize information about the population of each neighborhood alongside the dissimilarity values to allow the client to make an informed decision.

Alternatively, use logistic regression (classification) on the venue matrix, using most of the non-pizza neighborhoods as false labels and testing the rest. Logistic regression could reveal which of the non-pizza neighborhoods would most likely be classified as one with a pizza business – which neighborhoods ‘should’ have a pizza store. This approach will find appropriate weights for population and each venue matrix entry.