Data Description

2.1 Data Sources

Toronto neighborhood data and geospatial data (latitude and longitude) can be found in the website https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and http://cocl.us/Geospatial_data . Venue listings are queried from FourSquare.

2.2 Data Preprocessing

First, the data from the above sources are merged on postal code to obtain the latitude and longitude for each neighborhood and then the Toronto neighborhood data was extracted.

Second, the neighborhood venue listings were queried from FourSquare and the Toronto neighborhood venue listing were extracted. After checking FourSquare's neighborhood and postal code neighborhood, I found that FourSquare's neighborhoods are not the same as the postal code neighborhood boundaries, so some venues will show up more than once. I assign each venue to the postal code neighborhood it is closest to.

Third, I checked how many different kinds of venues we have in each neighborhood. Some of these are not restaurants, so I drop them. In addition, since this list is long, I drop any types that have 5 or fewer venues; it may not be feasible to open another one of these restaurants because they are not very popular.