1. Introduction

Mocha Joe is looking for a new location in the Toronto area. We are expanding into the area for the first time and we need to explore the neighborhoods within the city of Toronto. Ideally we want to find the neighborhood with the least amount of competition from other coffee shops, so we will be looking for a neighborhood without many preexisting coffee shops.

2. Data Acquisition and Data Cleaning

The data I will be using using to discover the neighborhoods of Toronto was scraped from Wikipedia using Python's beautifulsoup package. The geospatial coordinates were provided by Coursera in the form of a csv file. Venues in the area were discovered by using the Foursquare API to gather information about the most common types of venue within each neighborhood. The wikipedia article included neighborhoods that were outside of the area we are interested in, so I cleaned the dataset to include only that have a borough with the word Toronto in it.

3. Methodology

To explore this data in order to discover this neighborhood in need of a new coffee shop I used a couple of common machine learning techniques. First I prepared the dataset for K-means clustering by one-hot encoding all of the venues in every neighborhood. Then I use the clusters to narrow down which neighborhood has the least amount of coffee shops. I mapped these clusters using folium for a visual aid in the selection process.