

Comparing Neighborhoods of Toronto And NYC to Predict Restaurant Success

Keegan McCleary-Sharpe

December 3, 2020

Introduction -

This project showcases the ability to use data science toolkits on real-life problems. Geographical data can be scraped from the web, segmented, and subsequently applied to a relevant business problem. The analysis for this particular problem will be completed within Python, using primarily the Pandas, Geocoder, and Folium libraries, as well as Scikit-learn for K Means clustering.

Business Problem:

A potential client has established the following question:

“We have a series of restaurants that perform very well throughout several neighborhoods of New York City. We are looking to expand our operations into Canada, particularly the Toronto area, however, we are unsure which neighborhood we should target in order to access a similar clientele to our New York chain.”

Problem Analysis:

Based on the question, it stands to reason that we can address this business problem using geospatial data and clustering the neighborhoods of New York City and Toronto. We can look first at the neighborhoods of New York City that contain existing locations of our client's restaurant, and we can characterize those neighborhoods based on surrounding venues.

Once characterized, we can do the same for the neighborhoods of Toronto, and then perform unsupervised clustering on the neighborhoods of both cities. If two neighborhoods cluster together, they will have similar venues throughout the neighborhoods, and for this we can infer similar clientele. We can then highlight the neighborhoods of Toronto that cluster with New York City neighborhoods with our client's successful locations.

Data -

To perform the above analysis, we will need a dataset that outlines the five boroughs of New York City, as well as the neighborhoods within them. We have obtained that dataset in the form of a JSON file from IBM's server. We will also need a similar dataset for the Toronto area, which we can scrape from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

After we have loaded, cleaned, and began exploratory analysis on our neighborhood data for each city of interest, we will also need to access Foursquare's API. This API will allow us to identify the most frequent type of venue within a neighborhood and identify the characteristics of a neighborhood that allow our client's locations to be successful.