Applied Data Science Capstone Project – The Battle of Neighbourhoods

Baltej Toor

INTRODUCTION

Background

Given its growing population and corresponding increased development and investment into the city, Toronto provides an exceptional opportunity to grow a new business. Regardless of the numerous opportunities in the GTA, there are important factors associated with starting any new venture such as determining a location suited to success. When deciding on a location, essential aspects of the area include security/safety, marketing and promotion opportunities, and overall business concerns such as organized capital improvements and planning. The aforementioned considerations need to be made in order to set a small business up to thrive in a new environment. However, there are resources available to small business owners to establish a setting in which businesses can operate in a safe and competitive climate. A primary example of such a resource is Toronto's BIA program. A BIA (Business Improvement Area) is an association of commercial property owners and tenants that work together in a partnership, defined in a given area, with the city of Toronto to create an atmosphere in which business can be conducted safely and competitive in such a way that attracts customers and new businesses. BIAs in many ways are the catalysts for the improvements that create the kinds of strong communities that allow new businesses to integrate into and contribute to Toronto's economy.

Business Problem

Despite the opportunities and resources afforded to new businesses, there are difficulties that face entrepreneurs and business owners looking to grow their operation in one of Toronto's neighbourhoods. Depending on the nature of the business it may be challenging to secure enough capital towards efforts to gather extensive information on competitors and BIA initiatives. Entrepreneurs typically enter a market with limited resources, the majority of which needing to go towards inventory, leases/rental costs, and labour. This project aims to provide a tool in which new business owners can evaluate the level of competition in each Toronto neighbourhood and cross-reference these locations against established BIAs. Businesses will be able to evaluate locations on the basis of competition density and enhancement via the Business Improvement Areas.

DATA

In order to determine the viability of a particular area a combination of Foursquare location data and the Toronto BIA open dataset will be utilized. The two sources will be used in conjunction to determine the density of competition on the neighbourhood level as well as in regards to a common specified radius.

Foursquare Data

Data Setup

In order to take advantage of the Foursquare API's venue exploration feature we'll need neighbourhood geolocation data. Geospatial coordinates data will be read from an external dataset (via http://cocl.us/Geospatial_data). This data will be cross-referenced with postal code, borough and neighbourhood data web scraped from a listing of Toronto neighbourhoods (through https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). After modifications to concatenate neighbourhood values and ignore incomplete records, the foundational dataset with which the Foursquare API call is made will look similar to the following:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494
...	...	...	...	...	...

Figure 1 *Constructed dataframe to be used for Foursquare API call including fields PostalCode, Borough, Neighbourhood, Latitude & Longitude*

Area Competition Density

The Foursquare API will be used to retrieve information on nearby venues, namely pertaining to the category/type of venue as an indication whether a business is competition. An API call utilizes the latitude, longitude coordinates as well as a predefined RADIUS (among other parameters if necessary) to return a list of venues within the area. The initial result would resemble data similar to the following frame:

	Neighborhood	Venue	Venue Category
0	Parkwoods	Brookbanks Park	Park
1	Parkwoods	GTA Restoration	Fireworks Store
2	Parkwoods	Variety Store	Food & Drink Shop
3	Victoria Village	Victoria Village Arena	Hockey Arena
4	Victoria Village	Tim Hortons	Coffee Shop
5	Victoria Village	Portugril	Portuguese Restaurant

Figure 2 *Formatted dataframe containing venue category data for analysis on neighbourhood competition values*

This data would then be analyzed for the most ***n*** frequent venue categories present in the area with variable ***n*** chosen to maximize relevance of the categories to the result while ignoring unnecessary noise in the data (minimally occurring venue types complicating the analysis). Based on the venue category, in comparison to the target business category (e.g. Bakery), a count of the competitors in the area will serve as one component of the overall evaluation metric of whether a business should consider the location.

Toronto BIA Data

The second component of the metric would primarily incorporate the location data provided for each Toronto BIA. The base dataset provides latitude, longitude geometries of the BIAs with central coordinates and identifies (e.g. AREA_NAME). The distance to each BIA's central point will be determined for each neighbourhood and the BIA will be considered based on the RADIUS as specified during the initial Foursquare venue exploration. The metric would reflect the difference between two areas with similar competition factors based on availability of a BIA within the range of the neighbourhood. To illustrate, if the BIA shown below fell within the range of one of two prospective neighbourhoods (based on LONGITUDE, LATITUDE as shown) with Bakery-related business competition counts of 10 and 10, the neighbourhood with the BIA in range would rank higher than that of the neighbourhood out of range. Contrarily if both neighbourhoods were within range of the BIA, it would be the competition value that would hold higher weight in regards to their relative ranks.

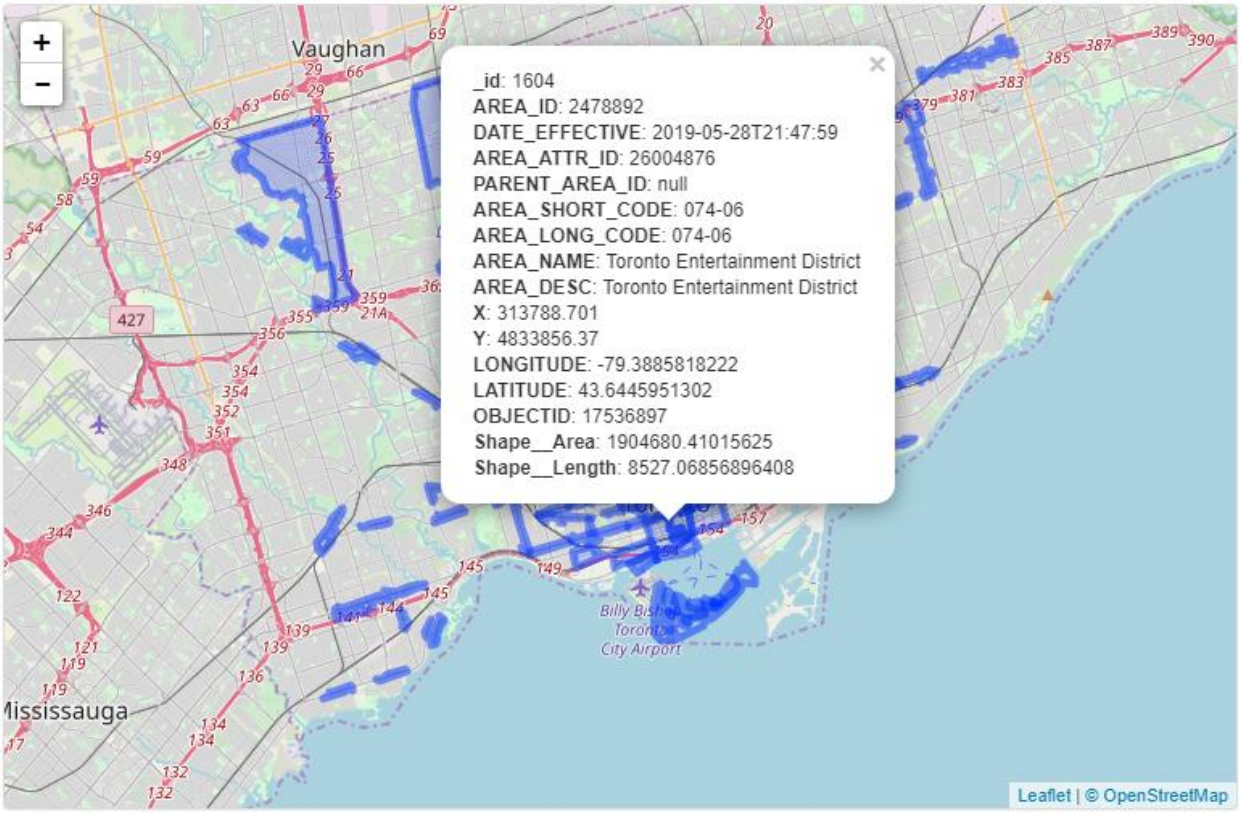

Figure 3 *Data preview of BIA geometry and data features*