

# The Battle of Neighbourhoods: Bar Venues in Toronto

Kristian Koschany

21 June 2019

## Introduction

An entertainment company from Vancouver is looking to expand to the Toronto market by opening some bar properties.

They have approached our consultancy to figure out which bar-containing neighbourhoods in Toronto proper contain the least bars. Their intention is to open establishments in the neighbourhoods they view as having the least competition, and therefore the most room to grow for their new establishments.

## Data

To produce a recommendation for our client, our consultancy will require the following data:

- List of neighbourhoods in Toronto
  - Data source: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
  - Description: A listing of neighbourhoods in Toronto containing postal code, borough, and neighbourhood name. This dataset can be easily scraped with the wikipedia Python library
- Latitude and longitude of each of Toronto's neighbourhoods
  - Data source: [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)
  - Description: A .csv file containing the latitude and longitude of each of Toronto's postal codes. This can be combined with the dataset scraped from Wikipedia to provide coordinates for each of Toronto's neighbourhoods
- List of bars in each neighbourhood in Toronto
  - Data source: Foursquare API
  - Description: Using Foursquare's API we can generate a dataset with this information, as stored within Foursquare's own large database. We can filter out venue entries so that we only have a dataset of bars, and we can combine it with our other data in order to attach a specific neighbourhood to each bar

## Methodology

Initially, a list of neighbourhoods in Toronto and their corresponding postal code and borough was collected by scraping the [List of postal codes of Canada: M](#) Wikipedia page. As multiple neighbourhoods can be contained within the same postal code, we combined these into one neighbourhood for statistical and ease-of-use purposes. For the purposes of this report, each postal code with its combined neighbourhoods will now be referred to as a *neighbourhood*. We then added a latitude and longitude to each neighbourhood, obtained from the [Geospatial\\_data.csv](#) file mentioned in the Data section.

Next, we stripped out the data we did not need from this dataset: neighbourhoods not contained in Toronto proper. To be specific, we only kept neighbourhoods contained within the *East Toronto*, *Central Toronto*, *Downtown Toronto*, or *West Toronto* boroughs, as per our client's specifications.

This complete, we required a metric for determining how many bar venues were within each neighbourhood, once that data was pulled from Foursquare's API. One solution would have been to simply count the number of bars contained exclusively within each neighbourhood; not only would that have been more difficult to code, but it also ignored the fact that people living near the boundary of their own neighbourhood likely frequent bars outside their neighbourhood, and not just within it.

To that end, a neighbourhood's boundaries were redefined as a certain radius centred on the latitude and longitude coordinates previously added to our dataset. As neighbourhoods vary greatly in size depending on location, we clustered them into four different clusters using the *kmeans* algorithm from the sklearn Python library, with the intention of using a different radius for neighbourhoods within each of the four clusters. This done, the data was exported to Excel, where a rough calculation was performed to determine an appropriate radius size for each cluster.<sup>1</sup> This led to the following radii for the following neighbourhoods (see next page):

---

<sup>1</sup> The calculation is as follows: the median distance of each neighbourhood's centre from the median centre of the cluster, divided by 2, rounded to the closest 50m increment

1000m	Christie
	Dovercourt Village, Dufferin
	Little Portugal, Trinity
	Brockton, Exhibition Place, Parkdale Village
	High Park, The Junction South
	Parkdale, Roncesvalles
	Runnymede, Swansea
600m	Rosedale
	Cabbagetown, St. James Town
	Church and Wellesley
	Harbourfront, Regent Park
	Ryerson, Garden District
	St. James Town
	Berczy Park
	Central Bay Street
	Adelaide, King, Richmond
	Harbourfront East, Toronto Islands, Union Station
	Design Exchange, Toronto Dominion Centre
	Commerce Court, Victoria Hotel
	The Annex, North Midtown, Yorkville
	Harbord, University of Toronto
	Chinatown, Grange Park, Kensington Market
	CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara
	Stn A PO Boxes 25 The Esplanade
	First Canadian Place, Underground city
900m	Lawrence Park
	Davisville North
	North Toronto West
	Davisville
	Moore Park, Summerhill East
	Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
	Roselawn
	Forest Hill North, Forest Hill West
950m	The Beaches
	The Danforth West, Riverdale
	The Beaches West, India Bazaar
	Studio District
	Business Reply Mail Processing Centre 969 Eastern

This complete, a function *get\_venues* was defined, which pulled the top 100 venues from each neighbourhood, using the specified radii from each neighbourhood's latitude and longitude as the neighbourhood boundaries.<sup>2</sup> Next, all venues that were not classified as some sort of bar venue were excised from this new dataset.

At this point we now had a list of bar venues for each neighbourhood, which was our goal. We were able to count the number of venues per neighbourhood using *pandas* as well as create maps displaying the data using *folium*, which will be displayed and discussed in the next sections.

---

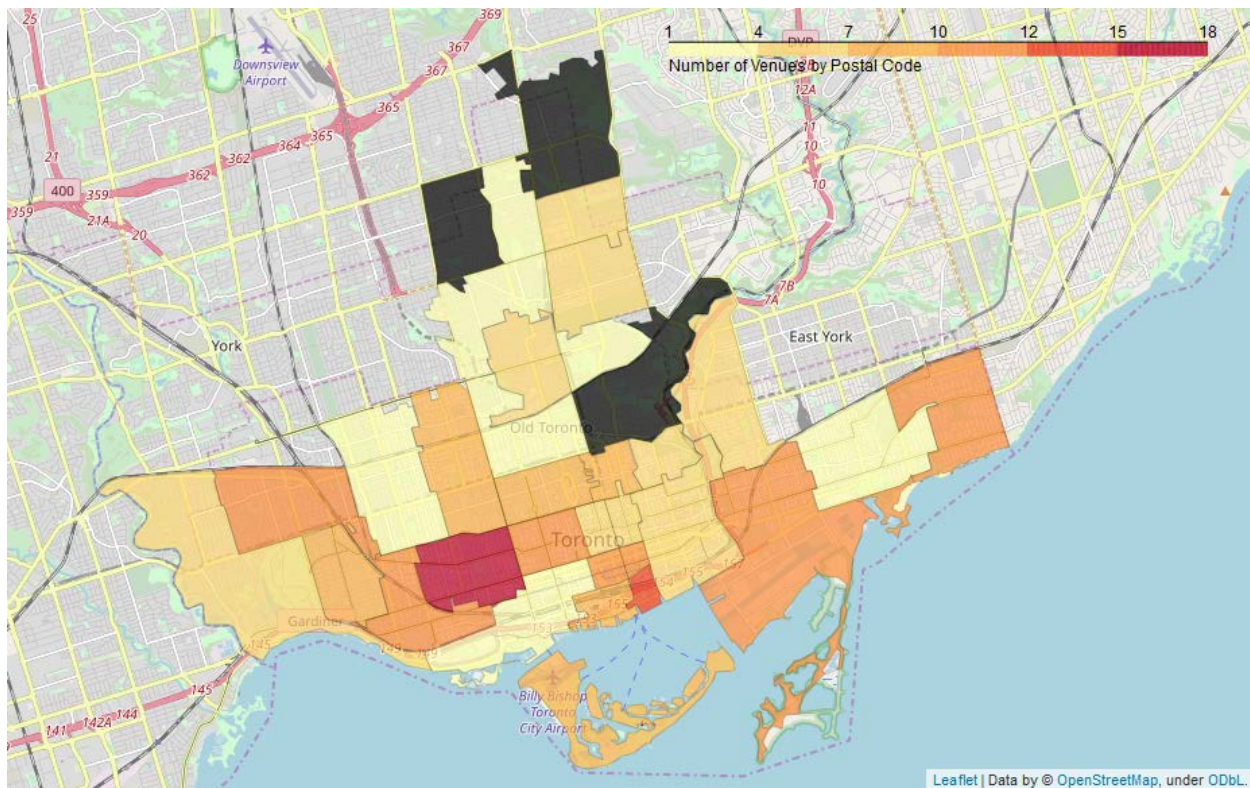
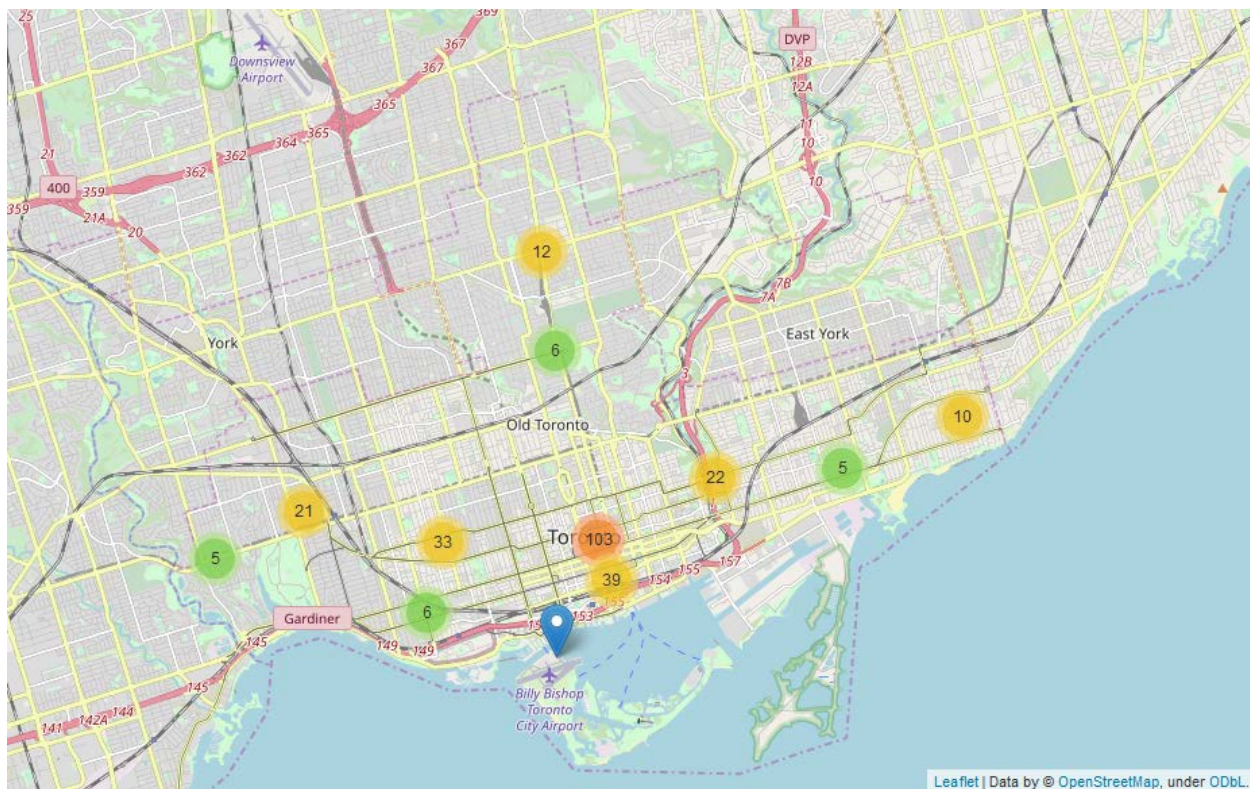
<sup>2</sup> Only being able to pull the top 100 is an arbitrary limit set by Foursquare on those accessing their API from a free account

## Results

Below are the results gathered. Figure 1 displays the number of bars in each Toronto neighbourhood, and figures 2 and 3 display maps of our processed data. The following section, *Discussion*, will explain and interpret these results.

Neighbourhood	Number of Bars
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	1
Forest Hill North, Forest Hill West	1
Moore Park, Summerhill East	1
Business Reply Mail Processing Centre 969 Eastern	2
North Toronto West	2
The Beaches West, India Bazaar	2
Dovercourt Village, Dufferin	3
The Annex, North Midtown, Yorkville	3
Central Bay Street	4
Davisville North	4
Harbourfront, Regent Park	4
Cabbagetown, St. James Town	5
Davisville	5
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	5
Runnymede, Swansea	5
The Danforth West, Riverdale	5
Ryerson, Garden District	6
Harbord, University of Toronto	8
Christie	9
Church and Wellesley	9
Harbourfront East, Toronto Islands, Union Station	9
Parkdale, Roncesvalles	9
Adelaide, King, Richmond	10
Chinatown, Grange Park, Kensington Market	10
St. James Town	10
The Beaches	10
Brockton, Exhibition Place, Parkdale Village	11
High Park, The Junction South	11
Studio District	12
Commerce Court, Victoria Hotel	13
Stn A PO Boxes 25 The Esplanade	13
Berczy Park	14
Design Exchange, Toronto Dominion Centre	14
First Canadian Place, Underground city	14
Little Portugal, Trinity	18

Figure 1: The number of bar venues in each neighbourhood of Toronto





## Discussion

As seen in Figure 1, the top three neighbourhoods that best meet our client's requirements are as follows:

- CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara
- Forest Hill North, Forest Hill West
- Moore Park, Summerhill East

Each of these neighbourhoods has only one bar contained within, making them good prospects.

However, by looking at figure 3, we can see that *CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara* is directly next to *Little Portugal, Trinity*, which contains the most bars at 18. Although our clustering and radius methodology accounts for bar overlap, we should still take our maps into account from an analytical perspective. Therefore, the two best prospects for our client are *Forest Hill North, Forest Hill West* and *Moore Park, Summerhill East*.

There are three neighbourhoods that contain no bar venues – they appear on our choropleth map in black, but do not appear on our list since they contain no data. These are not neighbourhoods we are recommending to our clients, as if there are no bars in them at all it likely means they are not neighbourhoods suitable for retail business.

## Conclusion

In this study, we analysed the neighbourhoods of Toronto proper for the number of bar venues that they contain. We created a list containing our data, as well as plotting two maps displaying the data in representative formats. From this data and our analysis and knowledge of Toronto geography, we were able to create a recommendation for our client on which neighbourhoods are their best prospect for opening a bar in the city: *Forest Hill North, Forest Hill West* and *Moore Park, Summerhill East*.