

Market Analysis of Small Businesses in Toronto

Byron Seppala

April 12, 2020

1. Introduction

Small business startups are a risky venture. Before small business owners can receive a capital infusion, they need to demonstrate the viability of the venture. The investor, whether a bank or a venture capitalist, will need some measure of confidence that the investment will yield returns. The small business sector is notoriously volatile, with many businesses failing in the first few years. Is there a way in which to measure the likelihood that a business will succeed? This is the question that this report will address.

1.1 Business Problem

A key factor in the success of a small business is location. One type of business may fail in one area and succeed in another. This success hinges partly on whether the product being offered matches the consumption patterns of the residents in the area. This report tackles this problem by analyzing consumption patterns in different neighborhoods around Toronto and identifying opportunities to fill market gaps.

1.2 Interest

This report is targeted at parties who are interested in the viability of small businesses in the Toronto area. A few possible recipients of this report are loan officers at lending institutions, analysis working for venture capital, city planners, and consultants.

2. Data

2.1 Data Sources

The venue data is imported from the Foursquare database. The Toronto neighborhood data is acquired by scraping the Wikipedia page:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The geographical coordinates of these neighborhoods are used to compile a list of venues in each neighborhood. These lists are used to perform cluster analysis to analyze consumption patterns.

2.2 Data Description

The project data for each neighborhood will consist of a Toronto postal codes, borough name, neighborhood name, latitude coordinates, and longitude coordinates. The coordinates will be used to populate a map of Toronto neighborhoods and gather venue data from the Foursquare database. The venue data will then be collected to build a consumption profile of each neighborhood. The consumption profile can be understood as the distribution of venue categories in the neighborhood.

2.3 Data Cleaning

There are several problems with the data that need to be sorted out. Many of the Toronto postal codes are unassigned. Since these postal codes do not represent actual neighborhoods, they are useless to this analysis. Thus, the non-assigned postal codes are dropped. Many postal codes contain more than one neighborhood. For each postal code, the neighborhoods are organized into a comma separated list. The latitude and longitude values for each postal code are loaded into a separate dataset. The datasets with borough and neighborhood and the dataset with the coordinates are merged by an inner join keyed on postal codes. This merged dataset is then used to complete the analysis.

3. Exploratory Analysis

3.1 Neighborhoods

The neighborhoods of Toronto can be displayed in an interactive map seen below.



Figure 1. Map of Toronto Neighborhoods

These are the neighborhoods that this report will analyze.

3.2 Venues by Neighborhood

As an initial exploration of the consumption profiles of each neighborhood, one can analyze the venue makeup of each neighborhood. A couple of examples are given below.

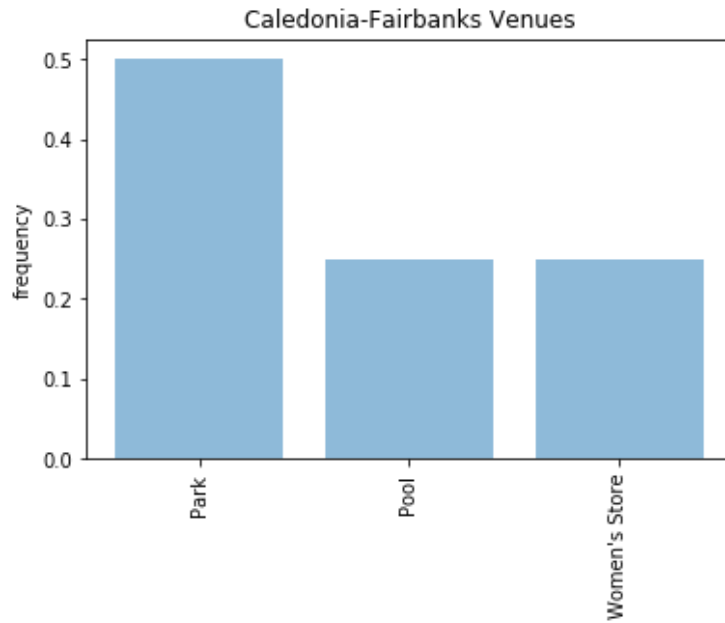


Figure 2. Venues of Caledonia-Fairbanks

Some neighborhoods have few venues, such as Caledonia-Fairbanks shown here.

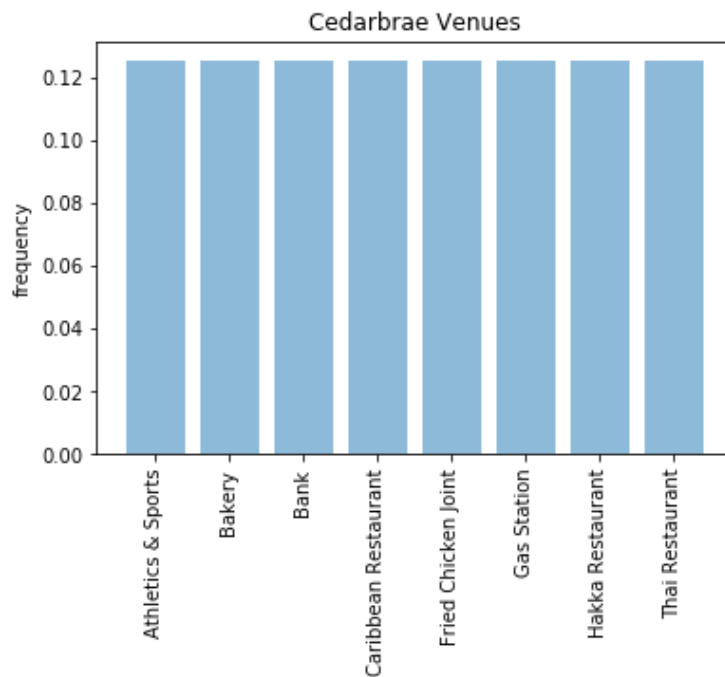


Figure 3. Venues of Cedarbrae

The venues of Cedarbrae, shown here, are uniform across the categories.

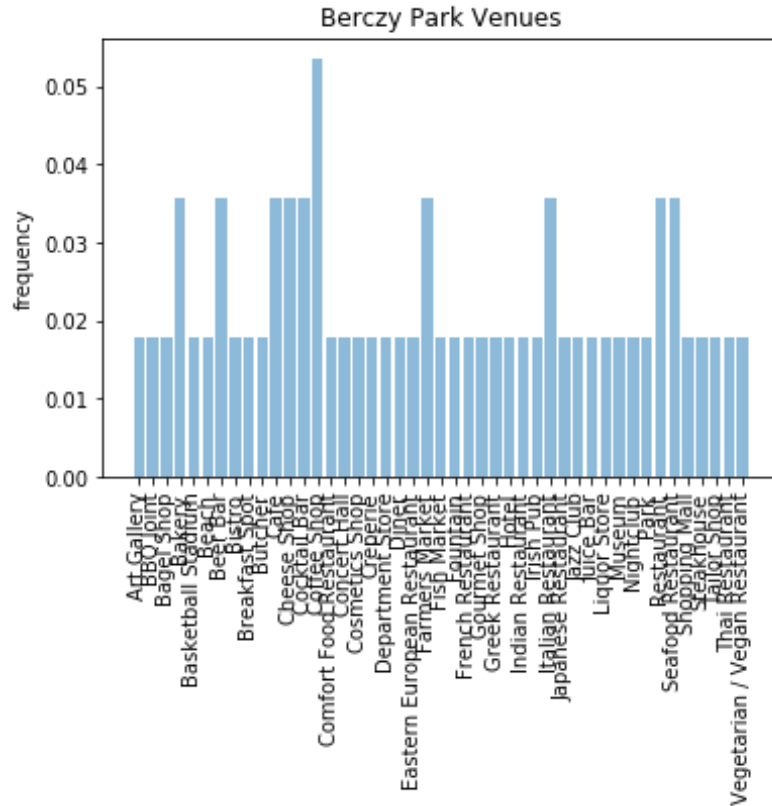


Figure 4. Venues of Berczy Park

Berczy Park has a large amount and a variety of venues, with coffee shop being the most common category.

4. Methodology

4.1 K-Means Clustering

The basis of the analysis in this report is k-means clustering. This is an unsupervised machine learning algorithm that classifies datapoints based on their position in higher dimensional space. The main hyperparameter to be tuned (or chosen) is the number of clusters. In this report, one venue category is chosen for analysis. This category is removed from the consumption profiles of all neighborhoods. A k-means cluster model is fit to the remaining data.

An example of k-means clustering is shown below.

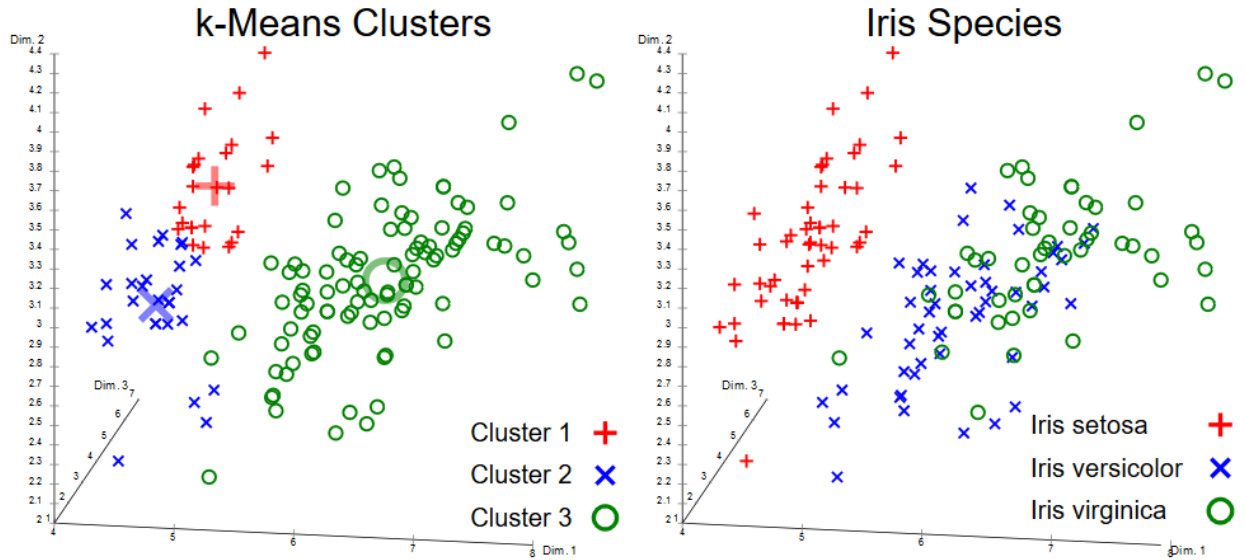


Figure 5. Visual example of k-means clustering. Source: https://en.wikipedia.org/wiki/K-means_clustering#/media/File:Iris_Flowers_Clustering_kMeans.svg

4.2 Dual Ranking

For each cluster, all neighborhoods are then ranked according to their distance to the cluster mean. The neighborhoods are also ranked according to their level of saturation of the category of interest. These two ranks are added together to get a composite rank. I.e. for each cluster, the neighborhood that has the least saturation of the category of interest and is most similar to the cluster mean is given the highest rank. For example, if there is a neighborhood that has a very similar consumption profile (low distance to the cluster mean) but has a low saturation of the target category (few venues of that type), then this neighborhood would receive a high rank. The neighborhood with the maximum composite rank for each cluster is chosen to be a suggested location for the target venue type. The number of suggestions can be adjusted by changing the number of clusters in the k-means cluster model.

5. Results

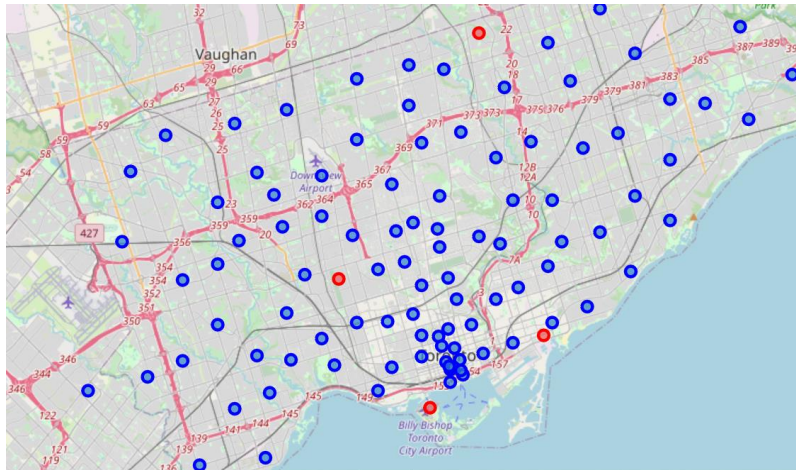
This report analyzes a number of venue types and gives suggested neighborhood locations for each one. These suggested neighborhoods represent the most “typical” neighborhood in their cluster (closest to the mean) which lack venues of the target category.

5.1 Coffee Shops

The first target category under analysis is coffee shops. According to this model, coffee shops in the following neighborhoods have the highest chance of success.

- Caledonia-Fairbanks
- CN Tower
- King and Spadina
- Railway Lands
- Harbourfront West
- Bathurst Quay
- South Niagara
- Island airport
- Business reply mail Processing Centre
- Hillcrest Village

A map of these areas is shown below. Note that multiple neighborhoods are collected into single nodes. The red nodes represent the best locations.

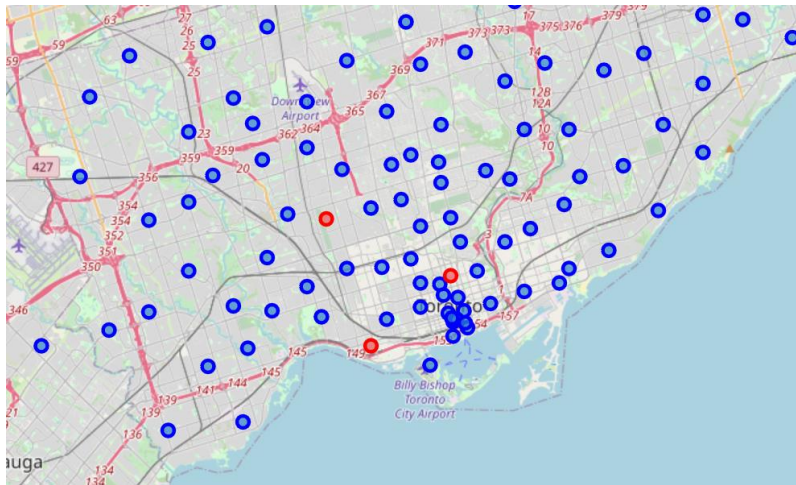


5.2 Bagel Shops

Next, this report analyzes locations for a bagel shop. The ideal neighborhoods are as follows.

- Church and Wellesley
- Brockton
- Parkdale Village
- Exhibition Place
- Caledonia-Fairbanks

The corresponding map is shown below.

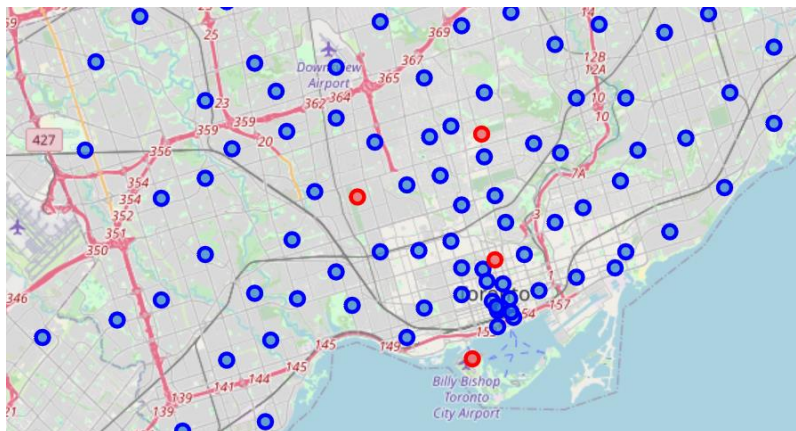


5.3 Nightclubs

The following neighborhoods are ideal for nightclubs.

- Church and Wellesley
- Davisville North
- CN Tower
- King and Spadina
- Railway Lands
- Harbourfront West
- Bathurst Quay
- South Niagara
- Island airport
- Caledonia-Fairbanks

The corresponding map is below.

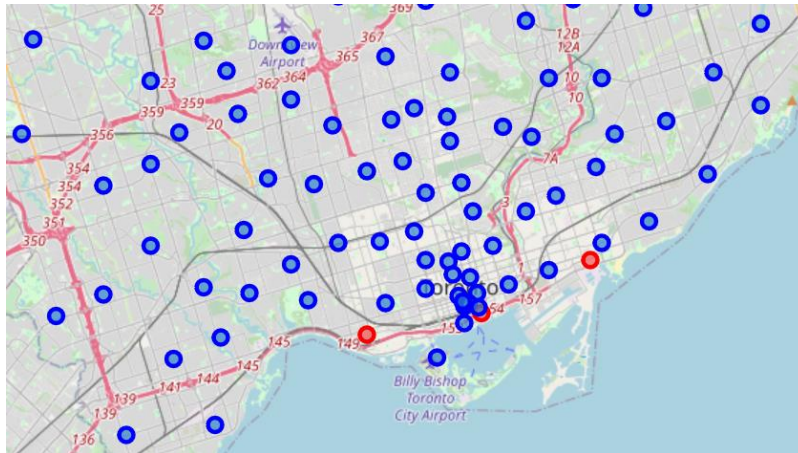


5.4 Women's Stores

The following neighborhoods are ideal for Women's Stores.

- Berczy Park
- Brockton
- Parkdale Village
- Exhibition Place
- Business reply mail Processing Centre

The corresponding map is below.

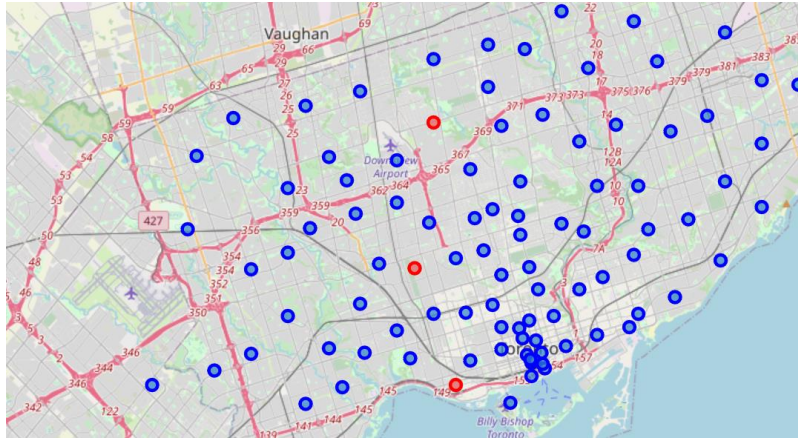


5.5 Thai Restaurants

The following neighborhoods are ideal for Thai Restaurants

- Bathurst Manor
- Wilson Heights
- Downsview North
- Brockton
- Parkdale Village
- Exhibition Place
- Caledonia-Fairbanks

The corresponding map is below.



6. Discussion

One interesting result from this report is that the ideal location for coffee shops is outside the downtown core. This suggests that there are neighborhoods outside of downtown Toronto that are underserved with respect to coffee shop venues. Another interesting result is the overlap between category types. For example, Caledonia-Fairbanks comes up a few times as an underserved neighborhood, most notably with coffee and bagels, suggesting a real opportunity to capture a market need.

Another notable result is that the suggested locations for women's stores are along the shoreline, which also includes the neighborhoods of Brockton, Parkdale Village, and Exhibition Place. These neighborhoods come up a few times as an underserved area and again suggest opportunity to invest in small businesses in this area.

7. Conclusion

This report highlights the ideal locations for small business startups of various types. Although only a handful of venue types are discussed here, this analysis can be extended to cover any venue category included in the Foursquare dataset. The analysis presented here attempts to identify underserved neighborhoods in the Toronto metro and highlight the types of businesses that may succeed in such areas. This analysis relies on the fact that residents tend to patronize businesses in the areas in which they conduct their daily lives, whether that is where they work or where they sleep. By identifying the types of businesses that are already successful, and the ones that are missing, this report puts forward compelling evidence with respect to small business success in Toronto.