***The Battle of Neighborhoods***

**Capstone Project: Doug Marcum**

**Contents**

|  |  |  |
| --- | --- | --- |
| 1. [**Introduction/Business Problem**](#Intro) | 1. [**Data**](#Data) | 1. [**Methodology**](#Methodology) |
| 1. [**Results**](#Results) | 1. [**Discussion**](#Discussion) | 1. [**Conclusion**](#Conclusion) |

**1)** **Introduction/Business Problem**

The basis of this study is to help a small group of investors planning to open their first U.S. based brewery / restaurant expansion in Toronto. Being that Toronto is the most populated city in Canada, and continually ranks as an important global city based on a high quality of living, the choice to expand into the neighbor of the north market was an easy selection for the investing group. However, with limited knowledge of the Toronto market, the group of investors have selected us to assist in the selection of which areas of Toronto will facilitate a launch of their brewery / restaurant expansion.

They are interested in building in an area that meets the following criteria:

* *A neighborhood with an average to above average median total population*
* *Above average populations of 25-40-year-old male and female professionals*
* *A high concentration of the population having secondary education*
* *Average to above average median net household incomes*

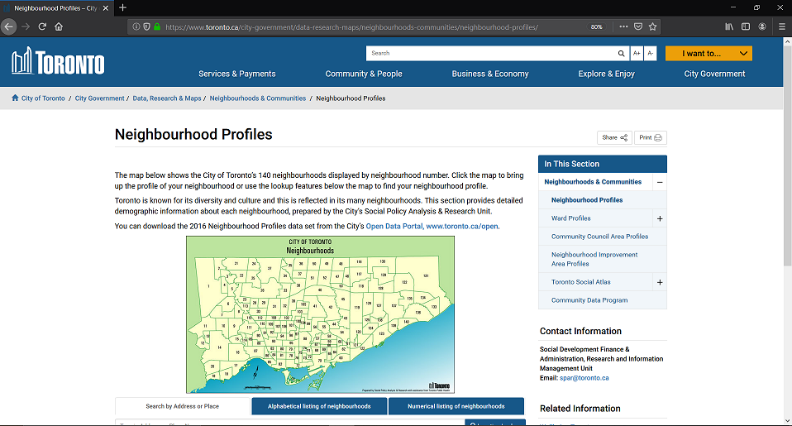
With these criteria given by the investing group, based on previous success in other markets, the objective is to locate and recommend to the investors, the target audience, which neighborhood(s) of Toronto will be the best choice to start their international growth plan. The information gained will assist in choosing the right location by providing data about the population of each neighborhood, in addition to other established venues present in these areas.

Additionally, this information could be of interest to other potential investors looking to open a new restaurant or entertainment venue in Toronto.

**2)** **Data**

The necessary information needed by the investing group will come from the following sources:

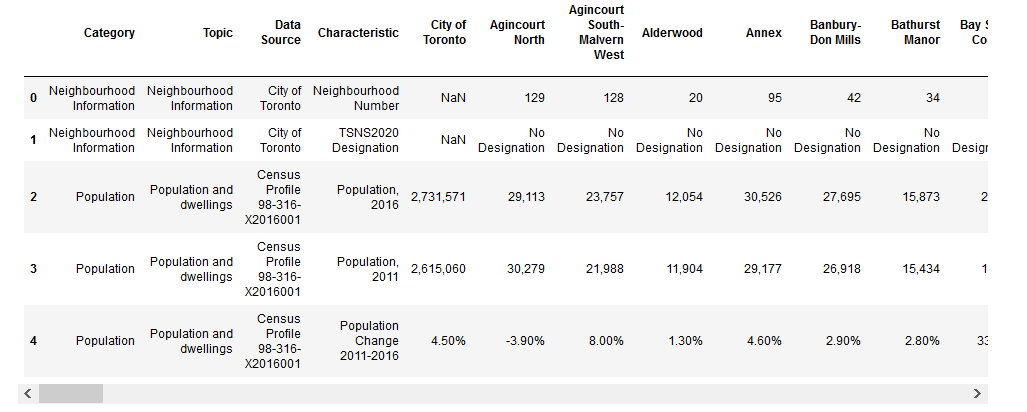
* [**City of Toronto Neighborhood Profiles**](https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/) for providing an overview of the neighborhoods in Toronto

[](https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/)

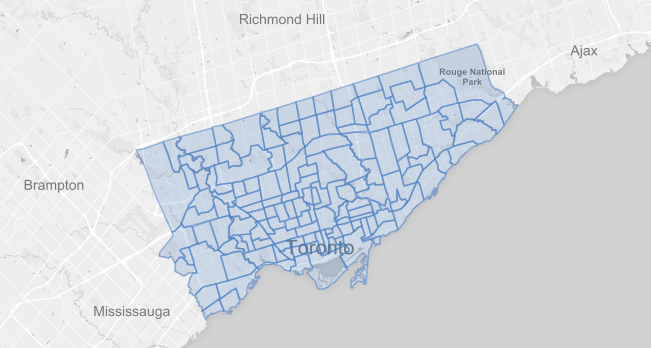
* [**City of Toronto Open Data Catalogue**](https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a) : *The Census of Population is held across Canada every five years (the last being in 2016), and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labor. City of Toronto Neighborhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighborhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighborhood.*

*In these profiles of the City of Toronto's 140 social planning neighborhoods. These social planning neighborhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighborhoods are consistent over time, allowing for comparison between Census years. Neighborhood level data from a variety of other sources are also available through the City's Wellbeing Toronto mapping application and here on the Open Data portal.*

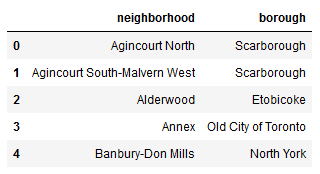
*Each data point in this file is presented for the City's 140 neighborhoods, as well as for the City of Toronto as a whole. The data is sourced from several Census tables released by Statistics Canada. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information.* [**CSV File**](https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv)



* [**City of Toronto Neighborhood Shapes**](https://portal0.cf.opendata.inter.sandbox-toronto.ca/dataset/neighbourhoods/) for mapping : [**GeoJSON File**](https://ckan0.cf.opendata.inter.sandbox-toronto.ca/download_resource/1d02b0f0-d735-4469-8f71-ea6d96b319e4?format=geojson&projection=4326)



* [**Wikipedia for Toronto Neighborhood Borough Designation**](https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto) : *Each of the 140 social planning neighborhoods of Toronto reside within a defined borough. While the City of Toronto is a singular municipality, the 140 neighborhoods are still grouped into six distinct boroughs.*



* [**Foursquare API**](https://developer.foursquare.com/) to collect information on other venues/competitors in the neighborhoods of Toronto



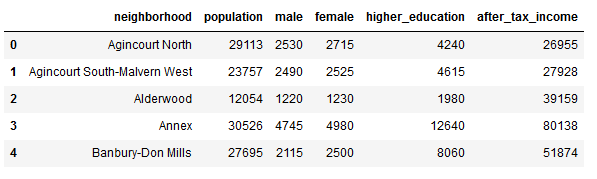
**3)** **Methodology**

In order to establish the targeted neighborhood(s), we will explore the demographics of the neighborhoods in the city of Toronto by segmenting the data and conducting descriptive analysis using Panda. Additional data will be gleaned by web scraping and API will be used to generate data.

**Data Group 1**

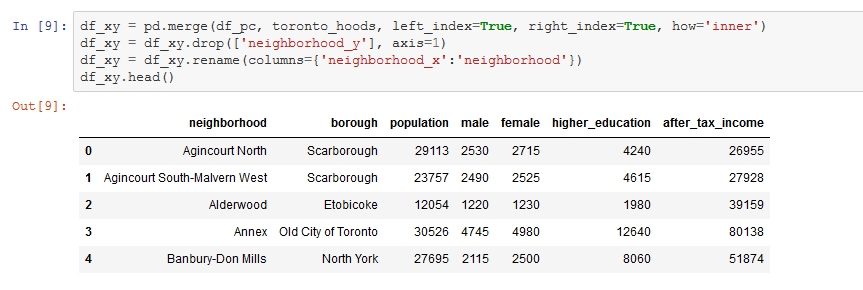
Stage A – Census Data

* 1. Data was pulled into from the City of Toronto Neighborhoods Profile Census CSV File to create a dataframe.
  2. This dataframe contains all the census data (2016) of the neighborhoods of Toronto that will be filtered.
  3. Data is filtered into columns based on neighborhood population, male and female age groups, education level, and after-tax income.



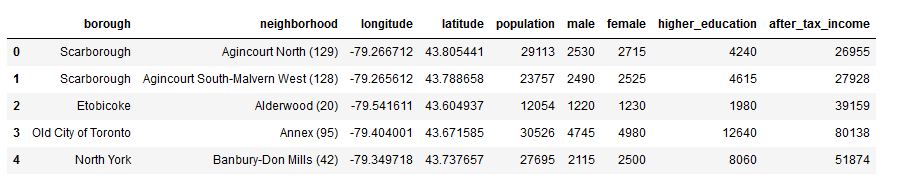
Stage B – Web scraping to align neighborhoods with boroughs

1. [Wikipedia page](https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto) for Toronto Neighborhood Borough Designations is scraped using BeautifulSoup.
2. Scraped data is transformed to dataframe.
3. Merge this dataframe with Census Data dataframe.



Stage C – Pull Toronto shape file

1. Get the [shape file](https://ckan0.cf.opendata.inter.sandbox-toronto.ca/download_resource/1d02b0f0-d735-4469-8f71-ea6d96b319e4?format=csv&projection=4326).
2. Remove unnecessary data and merge to previous dataframe.



**Data Group 2**

Stage A – Establish Medians and scoring system

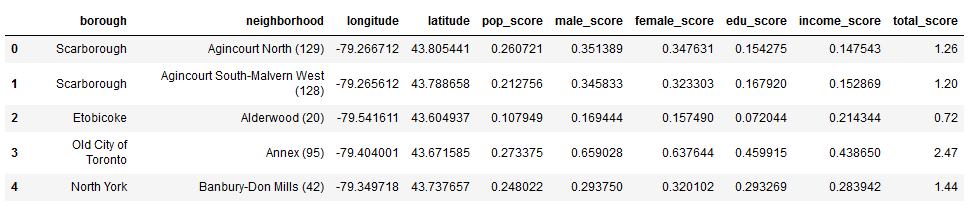
1. Calculate medians of the demographic columns across the 140 neighborhoods.

Median Population: 16749.5 Median Higher Education: 4122.5

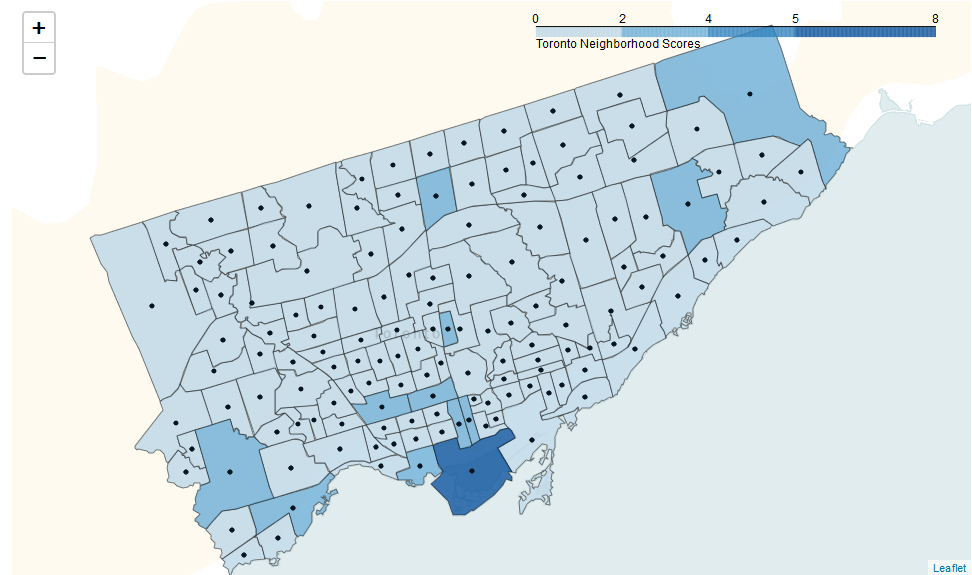
Median Male: 1800.0 Median After Tax Income: 36538.5

Median Female: 1952.5

1. From the criteria delivered by the investor group, each category was given a standardized score based the category being divided by its median score and then multiplied by a factor of given importance. The columns are the summed to create a total score for each neighborhood. The dataframe is then merged to create a dataframe with all necessary data.



1. From here, utilizing a choropleth folium map, a clearer picture of the neighborhoods of Toronto becomes apparent.



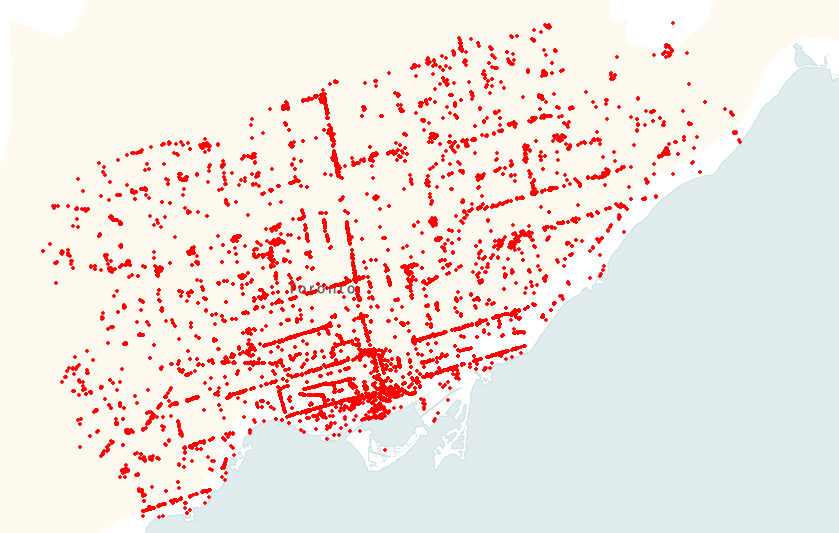
**Data Group 3**

Stage A – Toronto geographical data is utilized as the input into the Foursquare API, that returned venues from all of Toronto

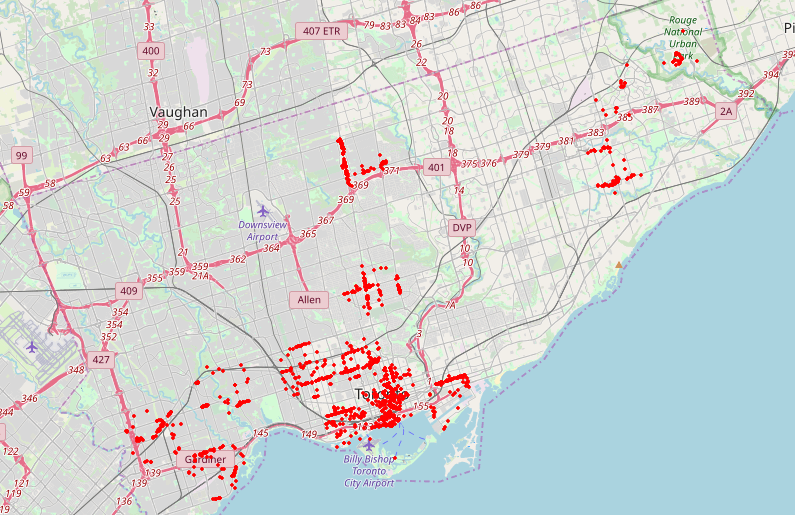
* 1. Using the geographical coordinates of each neighborhood in Toronto, calls are made to the Foursquare API to return the top 100 venues in a radius of 1610 meters, approximately a one-mile radius.



* 1. The data is then visualized via a folium map.



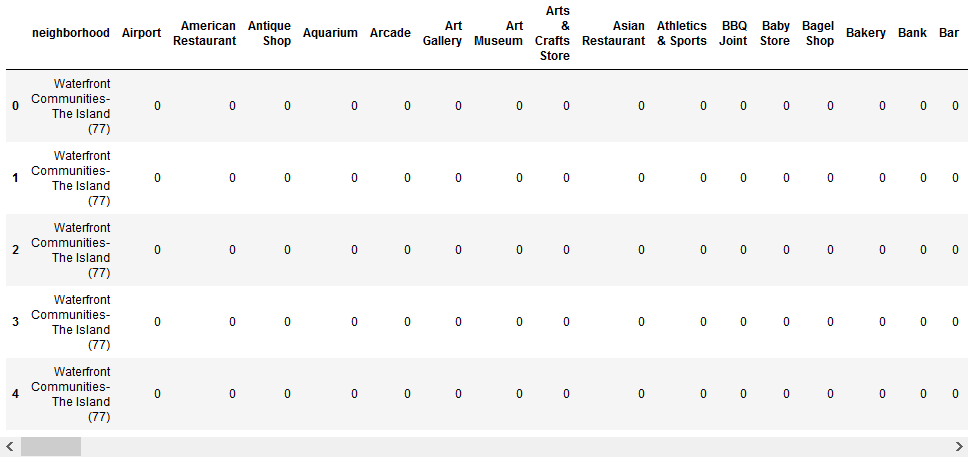
* 1. Foursquare is called on again to narrow the list to the neighborhoods within the top fifteen (15) total score (with the same parameters as the previous call), and this data is mapped as well.



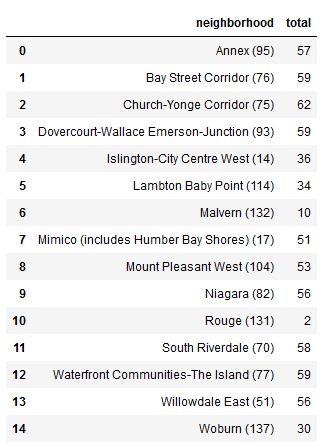
**Data Group 4**

Stage A – Analyze the Top 15 neighborhoods in Toronto

* 1. Utilize one hot encoding to transform our list of established venues in the top 15 neighborhoods to return a shape of (1198, 208).



### Group by neighborhoods. Filter out venues related to retail and personal care categories to focus on activity centered venues (i.e. bars, nightclubs, restaurants, attractions, etc...). Create a small dataframe to display totals for each neighborhood.

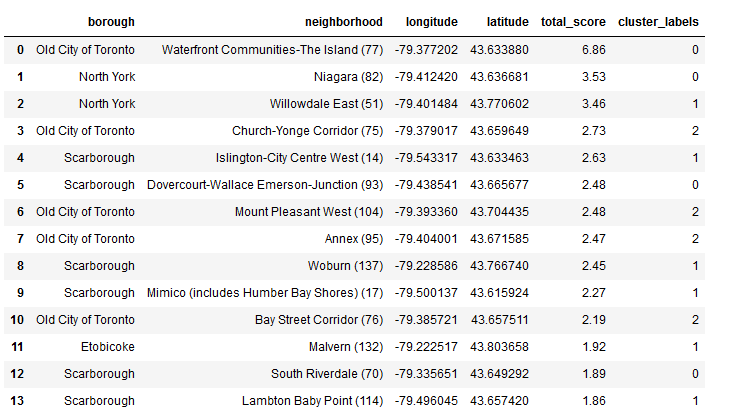


* 1. Drop neighborhood Rogue due to extreme limitations on venues after final filter.

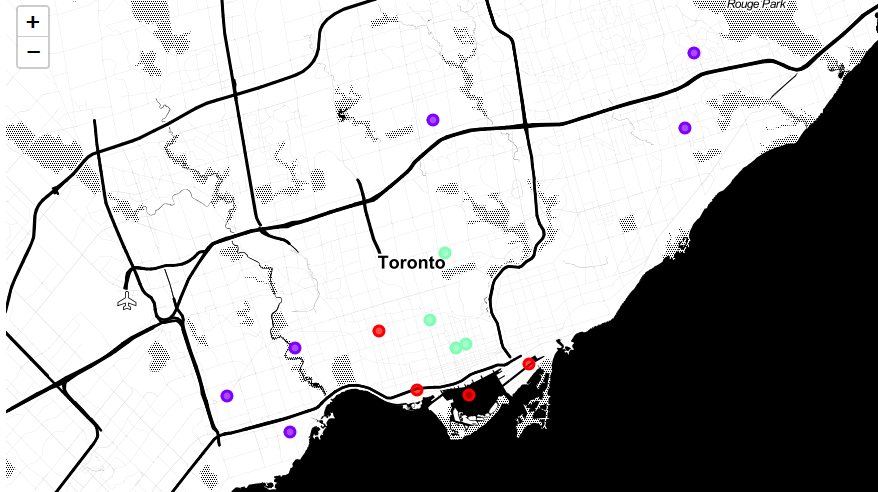
**4)** **Results**

With the data now ready, we run k-means to cluster the neighborhoods into three (3) clusters. The cluster number was established after multiple samplings and iterations. With our clusters established, this dataframe is merged with the total scores data to provide us with our final pieces of criteria in selecting the appropriate neighborhood(s).

1. The final dataframe:



1. The clusters are visualized via a stamen toner folium map:



**5)** **Discussion**

From the results discovered and presented, the following observations and recommendations can be made:

1. Based on the criteria given by the investor group and the cluster data, the main neighborhood recommendation would be for the neighborhood of Waterfront Communities – The Island. With a total score of 6.86 and being in the cluster 0 (a mean of 58 venues/neighborhood), this area is a prime location for a flagship brewery / restaurant.
2. A secondary recommendation is made for the neighborhood of Islington-City Centre West. This neighborhood a ranking in the top five total score assessment, as well as having less competition (only 36 venues within a one-mile radius).
3. Additionally, from the clustering, it becomes clear that city of Toronto radiates from that point in a circular pattern outward in terms of entertaining attractions. While it was not included in the research and analysis, commercial square footage rental rates tend to follow this trend as well.

**6)** **Conclusion**

In conclusion, the scope of this of the analysis is somewhat limited. The hospitality industry is ever changing, and the information afforded us may be dated due to relying on user information via Foursquare. Overall though, the model created can easily be replicated again and again with monitored data via the Foursquare API and the data from the forthcoming census in 2021. With the data analyzed and scoring system established by the investor group, we stand by the recommendations made.