# Predicting the Best Location to Start a Business in Toronto, Canada

**Shivam Maria**

**May 2, 2020**

## 1. Introduction

### 1.1 Background.

Prospective investors and entrepreneurs consider a wide variety of factors when considering when and where to establish a business venture, like a new factory or restaurant. One such critical aspect of an exercise of this nature, is selecting a location, that is, often what type of community, city, state or even country to locate their business in. This capstone project hopes to make this process a little less burdensome and scientific, by providing a robust and grounded analytical framework through which one can better assess the costs and benefits of a location relative to each other.

### 1.2 Problem

The main premise of our research paper will center on investigating and determining the ideal locale in Toronto, Canada, for the establishment of a new dining experience, a restaurant targeting people whose taste palette is unconventional and outside the box. In this study, I endeavor to identify the right location by finding the right cluster of people and amenities, where income, age, education, household size combined with the right competitive mixture to create an enabling environment to support business and those who depend on it.

Toronto is Canada's business and financial capital, a growing financial hub in North America, and a top ten global financial centre. The Toronto region's GDP accounts for 18 per cent of Canada's GDP and Toronto's economy grew by 2.4 per cent in 2014. It is home to Canada's five major banks, most foreign banks operating in Canada, and the Toronto Stock Exchange (TSX) – the world's principal exchange for mining, oil and gas and a leader in cleantech listings.

### 1.3 Interest

Toronto is competitive in almost every other major business sector from technology and life sciences to green energy; from fashion and design to food and beverage; from film and television production to music and digital media. Toronto's rich industrial diversity drives growth, innovation and cross-sectoral synergies and knowledge spillovers have spawned new leading-edge hybrid sectors including med-tech, green-tech and food-tech.

Toronto is Canada's business and financial capital. The city is the second largest financial services centre in North America and has one of the highest concentrations of financial services company headquarters in the Americas. With its reputation for safety, soundness and stability, Toronto is fast becoming a global location destination for financial services. (Toronto.ca/business)

Consequently, this paper will focus on countries, more specifically, cities and communities located in North America, mainly the United States of America, USA, and Canada. The USA and Canada are two of the worlds most developed and advance economies, exhibiting growing populations, rising incomes, underpinned by a dynamic legal system, stable political and economic environments. These are all attractive and necessary factors that allow sound business ideas to germinate and become profitable. While there are excellent locations in both countries, Toronto, Canada was chosen as the place of focus. It is not known to a lot of Americans and others as an incredibly diverse and economically rich city.

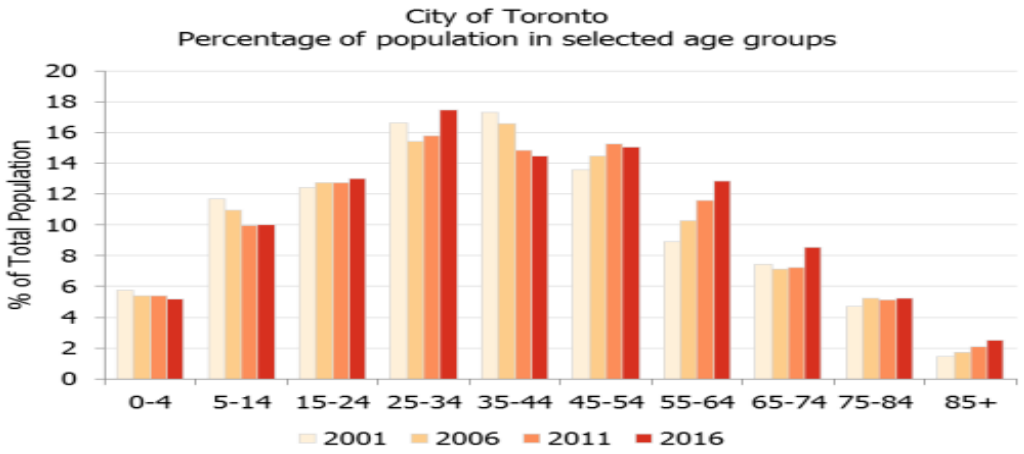
## 2. Data Sourcing, Cleaning and Wrangling

Projects of this nature always hinge on the quality and manner of data that can be had to inform the analysis, and in this case, it was no different. Data is messy and never comes ready and prepped to use, and so, much tweaking, messing and munging took place with the data used in the report. The census data and the data extracted from web site previously, was a challenge.

Initially, I obtained demographic and household data on Toronto and its environments from the parsing the 2016 Canadian census. In Toronto's 2016 Census that contains Population, Average income per Neighborhood with Toronto's Neighborhoods shapefile

and Foursquare API to collect competitors on the same neighborhoods.

Figure 4: Percentage of population in selected Age Groups, city of Toronto (2001-2016)




The figure above shows that the productive age group of 16-65, is about 65 percent, giving business a large pool of talent. This data however, had to be cleaned and transformed.

In order to ensure high quality social data, the neighborhoods were defined based on Statistics Canada Census Tract boundaries. Census Tracts include several city blocks and have on average about 4,000 people. Neighborhoods are comprised of from 2 to 5 Census Tracts.

Like Census Tracts, most service agencies and their programs have service areas that are defined by main streets, former municipal boundaries, or natural boundaries such as rivers. These service areas include several census tracts. It is not uncommon for service areas of community agencies to overlap. Choices about neighborhood boundaries were made to make the data in the profiles useful to as many users as possible and are not intended to be statements or judgments about where a neighborhood starts or ends.

Along with relying on Python for statistical and data analysis, we leveraged the Foursquare API to obtain longitudinal and latitudinal values for location and venues around Toronto, which ultimately informed our analysis. As previously noted, fair degree of cleaning and formatting was necessary on transform the data in a digestible and usable format.

| Toronto's Census data is publicly available at this website: |
| --- |
| https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a |
| |
| https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, |

### 3. Methodology

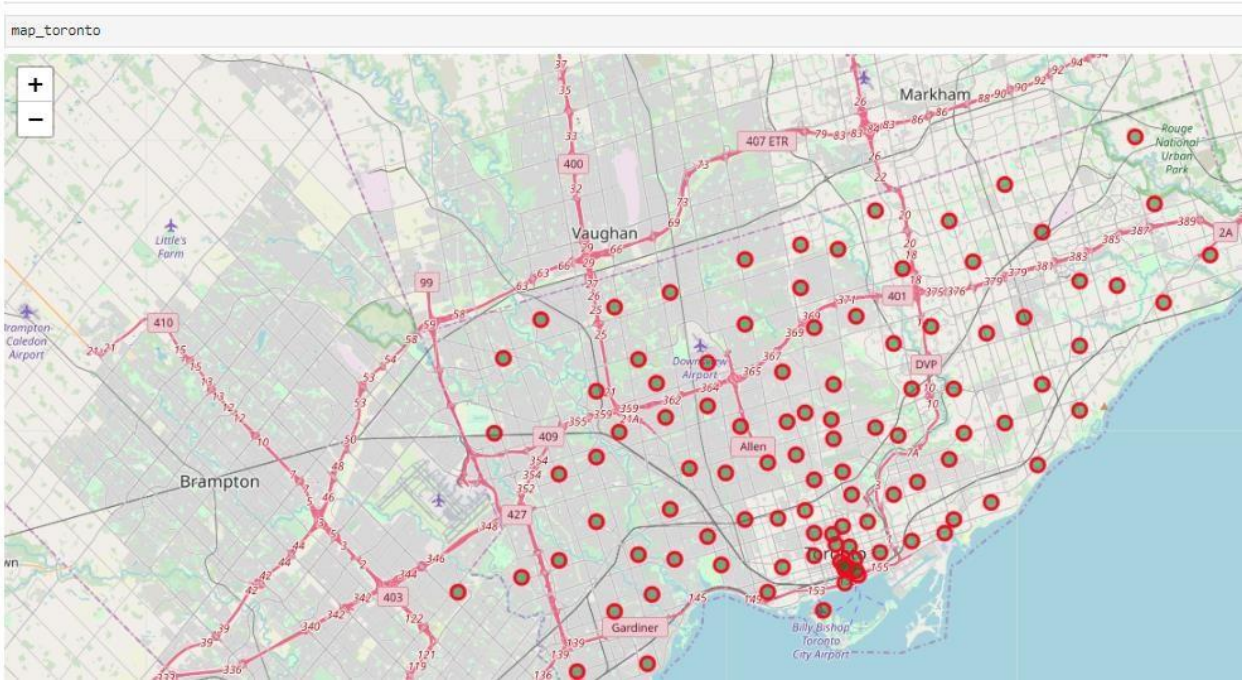
| | Borough | Neighbourhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | East Toronto | The Beaches | 43.676357 | -79.293031 |
| 1 | East Toronto | Riverdale, The Danforth West | 43.679557 | -79.352188 |
| 2 | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| 3 | East Toronto | Studio District | 43.659526 | -79.340923 |
| 4 | Central Toronto | Lawrence Park | 43.728020 | -79.388790 |
| 5 | Central Toronto | Davisville North | 43.712751 | -79.390197 |
| 6 | Central Toronto | North Toronto West | 43.715383 | -79.405678 |
| 7 | Central Toronto | Davisville | 43.704324 | -79.388790 |
| 8 | Central Toronto | Moore Park, Summerhill East | 43.689574 | -79.383160 |
| 9 | Central Toronto | Deer Park, Forest Hill SE, Rathnelly, South Hi... | 43.686412 | -79.400049 |

For this report I used a few different maps that could help a new investor to decide the best neighborhood to open a restaurant in Toronto based on its income, population and available competitors. In order to do that I've used the 2016 Census information combined with maps, tables and graphs to visually display the wealthier and more populational neighborhoods and Foursquare data to display the current restaurants in each region.

Toronto Neighborhoods' shapefile is publicly available at this :
website: https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#a45bd45a-ede8-730e-1abc-93105b2c439f

This problem is best suited for a classification methodology as opposed to a regression. Therefore, I employ the popular, somewhat simpler relative to other ML models available, but efficient and power, K-means Clustering in Python. Many clustering algorithms are available in Scikit-Learn and elsewhere, but perhaps the simplest to understand is an algorithm known as *k-means clustering*, which is implemented in sklearn. cluster. KMeans.

K-means clustering is a clustering algorithm that aims to partition nn observations into kk clusters. Primarily, there are 3 key steps to executing a k-means method, that is: Initialization – K initial "means" (centroids) are generated at random; next Assignment – K clusters are created by associating each observation with the nearest centroid, and finally Update – The centroid of the clusters becomes the new mean Assignment. This process continues, and Update are repeated iteratively until convergence, doesn't always happen however.



The result is that the sum of squared errors is minimized between points and their respective centroids. Clustering algorithms seek to learn, from the properties of the data, an optimal division or discrete labeling of groups of points.

Essentially, I looked at the various locations around Toronto, ranked them based on their characteristic metric and assigned to a scored group 1-10 to each. Then they were later grouped into clusters of 5. Ultimately, the location or venue with the most ranked scores will be presumed to be the most advantageous location in which to locate the new business entity.

Following initial preprocessing and exploratory analysis, I specified k=5, that is, clusters, 0-5. Thereafter, relied on the statistical analysis score given to the various neighborhoods, before the model can be deployed to be used a predictive tool. The diagram below shows a pictorial view of the model, as we can model parameters and the methods being deployed. And that is exactly what we observed when we take a look at the output results generated by the model.

```
# Run k-means to cluster the neighborhood into 6 clusters
# set number of clusters
kclusters = 10

toronto_grouped_clustering = toronto_grouped.drop('Neighbourhood', 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=1).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
print(kmeans.labels_[0:20])
print(len(kmeans.labels_))
```

```
[0 8 0 7 7 7 0 7 7 7 0 7 9 7 0 7 0 5 0 7]
38
```

We can clearly see that of the 10 neighborhoods, which we established some begin to emerge as leaders on the metric score.

### 4) Results

Comparing the maps, we can notice many of the restaurants grouped on main streets and on the south of the city, although some of the wealthiest neighborhoods are up to the north. Also, the areas with a dense population don't reflect on the number of restaurants.

```
toronto_onehot.shape
```

```
(836, 193)
```

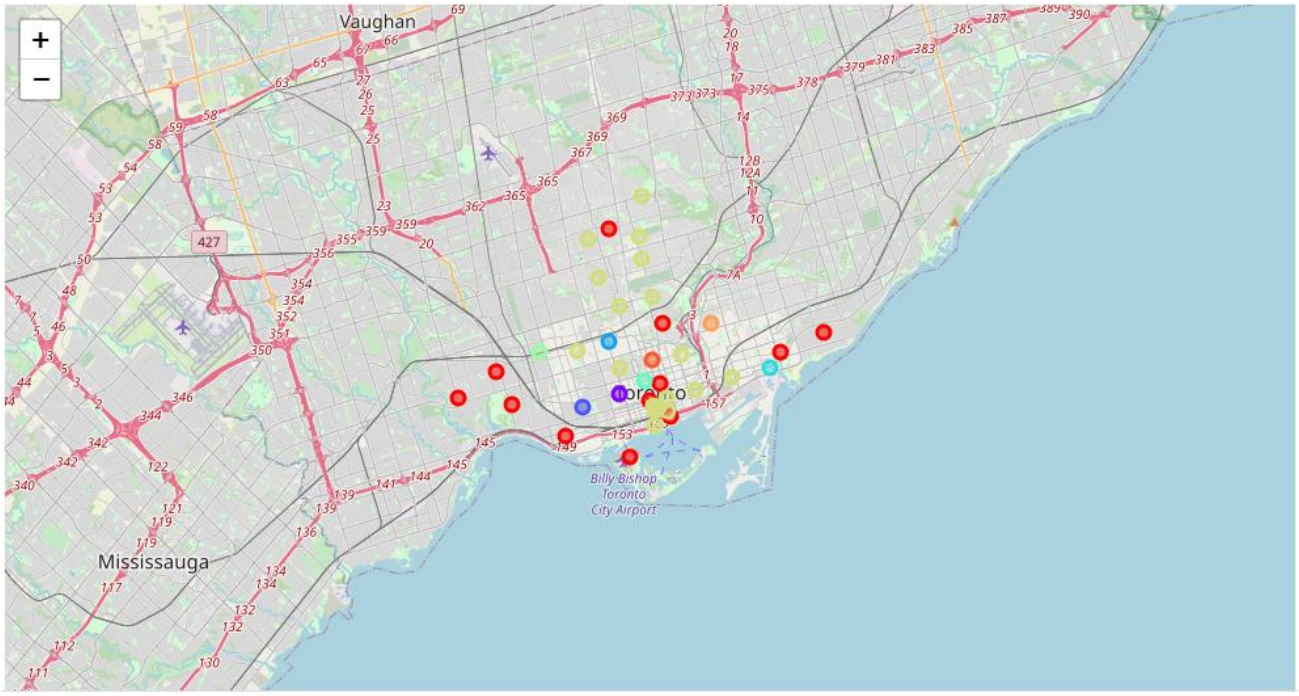
```
toronto_onehot.describe()
```

| | Airport | Airport Food Court | Airport Gate | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Antique Shop | Aquarium | Art Gallery | Art Museum | Arts Cra Sto |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836.000000 | 836 |
| mean | 0.001196 | 0.001196 | 0.001196 | 0.002392 | 0.003589 | 0.002392 | 0.009569 | 0.001196 | 0.001196 | 0.009569 | 0.001196 | 0.00 |
| std | 0.034586 | 0.034586 | 0.034586 | 0.048882 | 0.059832 | 0.048882 | 0.097412 | 0.034586 | 0.034586 | 0.097412 | 0.034586 | 0.05 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| 25% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| 50% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| 75% | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| max | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.00 |

```
toronto_neighborhoods.head()
```

| | Borough | Neighbourhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | East Toronto | The Beaches | 43.676357 | -79.293031 |
| 1 | East Toronto | Riverdale, The Danforth West | 43.679557 | -79.352188 |
| 2 | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| 3 | East Toronto | Studio District | 43.659526 | -79.340923 |
| 4 | Central Toronto | Lawrence Park | 43.728020 | -79.388790 |

The illustration below shows the final neighborhood cluster, East Toronto, in yellow as the best location.




## 5) Discussion

This report may be helpful for someone planning on opening a restaurant in Toronto, as any similar such type of business. They can be confident that by comparing the current offers and neighborhoods profiles, however it may not cover all variables such as access to public transportation or even the restaurants profiles, so it shall not be used as a single decision-making tool. Reports of this nature have a shelf life and one should always appreciate and contextualize their process going forward. However, I have no doubt that businesses will be aware of all these factors.

## 6) Conclusion

After examining the above 5 clusters, I believe and would recommend that cluster label (7), East Toronto is the best Neighborhood to set up shop for a new business venture. While there are great locations like downtown, Lawrence Park and Riverdale, East Toronto presents the best opportunity to be a first mover as well an early entrant into a community that is poised to experience phenomenal growth. East Toronto is the most suitable locale as it has an expanding population, colleges, a dependable public transport system, supporting entities for the productive population a business will need and attract.