## **Predicting a good location to start a Coffee/Take Away Shop in Toronto**

## **Introduction**

### **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 Statement**

The main premise of our research will center on investigating and determining the ideal locale in Toronto, Canada, for the establishment of a new coffee/take away shop, a shop targeting people whose are on the go. In this study, I endeavor to identify the right location by finding the right cluster of people and amenities, and lack of quick eats. We must investigate the issue in a scientific method using Data Science as there is an opportunity for wasting money and effort by investing at a non-optimal location. Data Science provides best possible location based on available data.

## **Data Sourcing, Cleaning and Wrangling**

The neighborhoods in by the City of Toronto were developed to help government and community organizations with their local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighborhoods do not change over time, allowing researchers to examine changes over time. 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.

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. Fair degree of cleaning and formatting was necessary on transform the data in a digestible and usable format.

Postal codes are found using Wikipedia Link : <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

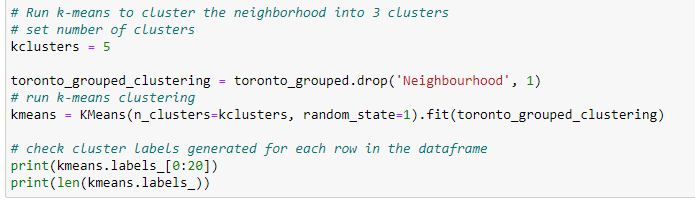
## **Methodology**

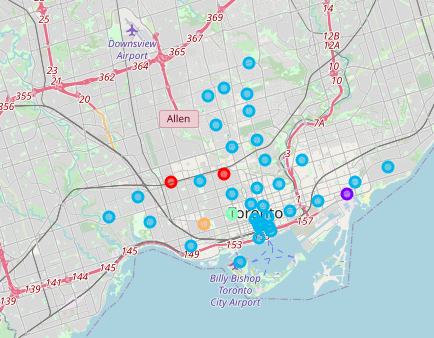
For this report we use map based of foursqure API that could help a new investor to decide the best neighborhood to open a Coffee shop in Toronto based on available competitors.

This problem is best suited for a classification methodology as opposed to a regression. Therefore, we 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 but 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.

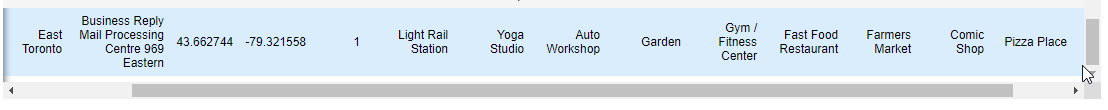
Essentially, we looked at the various locations around Toronto, ranked them based on their characteristic metric and assigned to a scored group 1-10 to each. Then they were later grouped into clusters of 5. Ultimately, the location or venue with the most ranked scores will be presumed to be the most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





1. **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. You can also find that towards the east of the CBD, especially at Mail processing center has least amount of café shop listed.



One could imagine, considering there is light rail . there is will be a lot of foot traffic. It is my recommendation from the results we have seen that this a good place to invest on a coffee shop. For a new investor.

1. **Discussion:**

This report may be helpful for someone planning on opening a coffee shop in Toronto, as any similar such type of business. They can be confident that by comparing the current offers and neighborhoods profiles. 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.

1. **Conclusion:**

After examining the above clusters, I believe and would recommend that cluster label (1), East Toronto, business mail processing center is the best Neighborhood to set up shop for a new business venture. While there are great locations like downtown this location 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.