# Introduction

### Description and Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Toronto is the fastest growing city in North America. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. The city proper has since expanded past its original borders through both annexation and amalgamation to its current area of 630.2 km2 (243.3 sq mi).

### Problem

As we can see there is lot of opportunities for business in this city, I would like to create a report to understand which business should be setup in which part of the city so that their business should florish.  
For example if you open a retaurant in a place which is more popular among the people because of local GYM or any fitness centre then our restaurant may not be that much of success then the place where we already have different restaurants in the area. Bcause people would try to look for new and better option if we are going to open an restaurant in an area where one restaurant is already present. Or you can say people may want to try out new GYM eqipments provided by new gym centre and their services rather than if place a GYM in an area which is popular among foody kind of people looking for new food items.

# Data

### Data Collection

Data for neighborhoods present in Toronto is collected from below Wikipedia website:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Scrape the data from Wikipedia to dataframe to analyse the data.

After that for each neighborhood in Toronto make a map.

Using Foursquare API get all the nearby venues of the neighborhood and list most common visited venues for all the neighborhoods.

### Data Cleansing

Now let’s group the most common visited venues as a single entity.

1. Thus the venues containing words like: "Restaurant", "Food", "Bar", "Pizza", "Ice Cream", "Chips", "Cheese", "Cafe", "Breakfast", "Coffee" are grouped as a single entity “Food/Restaurant/Bar”.
2. The venues containing words like: "Gym", "Fitness", "Yoga", "Martial" are grouped as single entity: 'Gym/Fitness/Yoga'
3. The venues containing word: Airport is grouped as a single entity: “Airport”

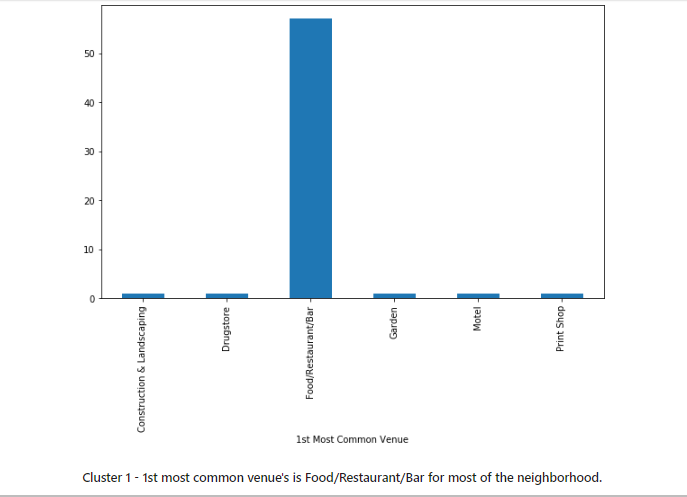
# Methodology

Using one hot encoding and k-means clustering methodology cluster the different neighborhood groups consisting of areas which includes venues on basis of different categories:

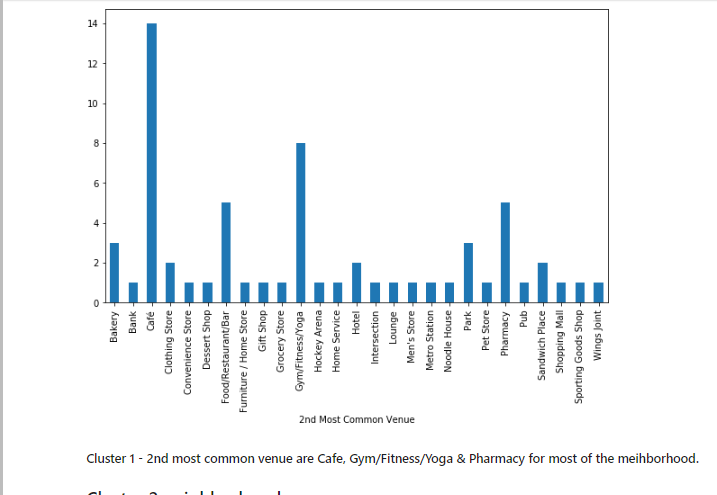
1. Food/Restauraunt/Bar, Cafe
2. Gym/Fitness/Yoga
3. Park
4. Playgrounds
5. Baseball Fields, etc

Below are the different bar graphs showing no. of neighborhoods vs 1st most common visited venue of the neighborhood in different clusters.

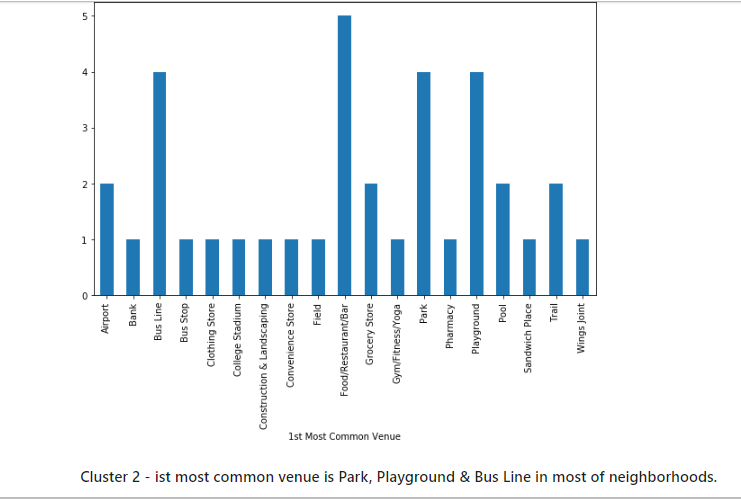
1. Cluster 1: No. of neighborhoods vs 1st most common visited venue.



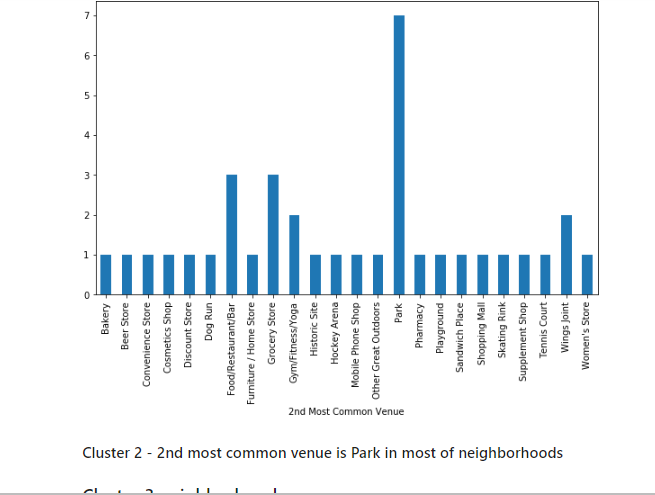
1. Cluster 1: No. of neighborhoods vs 2nd most common visited venue.



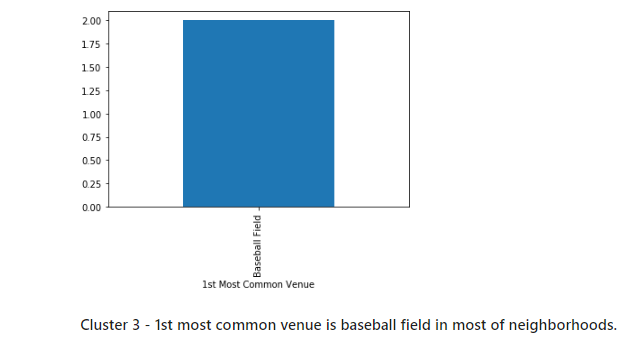
1. Cluster 2: No. of neighborhoods vs 1st most common visited venue.



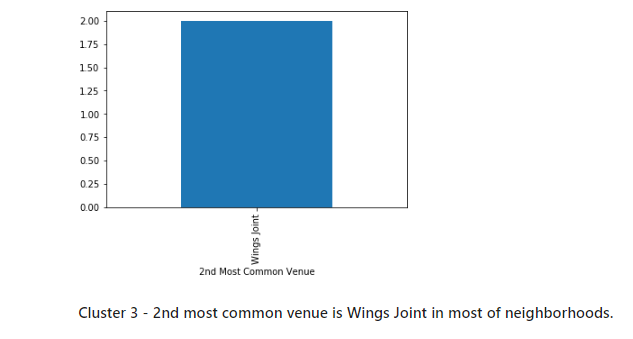
1. Cluster 2: No. of neighborhoods vs 2nd most common visited venue.



1. Cluster 3: No. of neighborhoods vs 1st most common visited venue.



1. Cluster 3: No. of neighborhoods vs 2nd most common visited venue.



# Result

In the 1st cluster we could see clearly throgh bar graph that Food/restaurant/Bar, Cafe & Gym/Fitness/Yoga are the popular venues. In 2nd cluster bar graph shows that Playground & Park are the popular venues. And similarly in 3rd cluster Baseball field and Wings joint are the most popular venues visited.

# Discussion

As per by my analysis report following recommendations can be made:

1. Neighborhoods present in cluster 1 is apt choice for the business ventures looking for opening a Restaurant/Bar. Also in few places Gym/Fitness center can be a profitable as it appeared to be the 2nd most common venue in this cluster's neighborhood.
2. Neighborhoods present in cluster 2 can be profitable for the business ventures that deals with park/playground maintenance stores. Because this neighborhood should maximum amount of visits due to Park or playground.
3. Neighborhoods present in cluster 3 can be good for stores containing items related to baseball. As it is the most common venue being visited in that area.