1. **Introduction**
   1. **Background**

According to Innovation, Science & Economic Development Canada statistics, thousands of businesses exit the marketplace every year in Canada. Business failure statistics show that about 96 percent of small businesses (1–99 employees) that enter the marketplace survive for one full year, 85 percent survive for three years and 70 percent survive for five years (Key Small Business Statistics). Approximately 7,000 businesses go bankrupt every year in Canada

The main reason for (business) failure is inexperienced management and planning. Therefore, for a good business plan it is essential that your decisions are based on the data.

* 1. **Business Problem**

With the help of data science I will analyse all the neighbourhoods in Toronto. These insights would be beneficial to someone who is looking to open a new restaurant or trying to expand his chain in this region. It will help those investors to decide where to open the restaurant and which type of cuisines would be beneficial for their business.

1. **Data acquisition and cleaning**
   1. **Data Sourcing**

[This](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) is a Wikipedia page of Canadian neighbourhood. The table format was different when I started the project but now they changed. So I went to the history section of Wikipedia page and picked the old table format from [this link](https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969). Then the data has been cleaned and filtered out all the neighbourhoods other than Toronto’s. Also, used foursquare API to get all the venues in each neighbourhood.

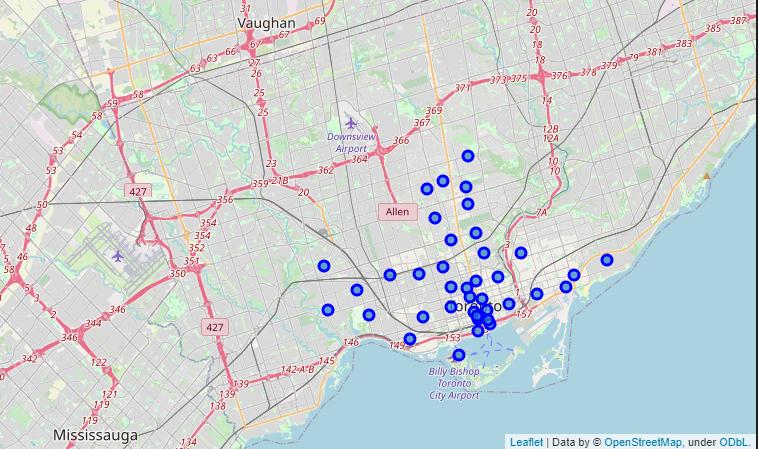
* 1. **Data Cleaning**

There were three data frames but the first one was relevant so the other two were ignored. Then column names were corrected by removing spaces in between. Then, all the boroughs and neighbourhoods named “not assigned” were removed because that was basically a null value. Then, there is another dataset which was provided by courser course. It provides the latitude and longitude values according to postal codes. This data frame was merged with the first one.

Then, only Toronto boroughs has been kept and all other rows were discarded. At this step, data has been cleaned.

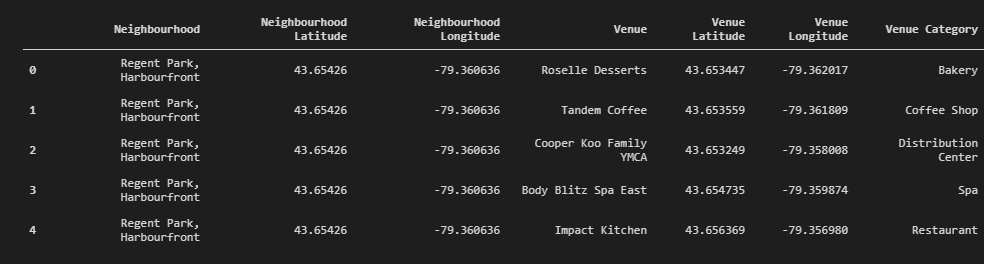
1. **Exploratory Data Analysis**
   1. **Creating Toronto Map**

Folium library has been used to create a map of Toronto with boroughs superimposed on top of it.

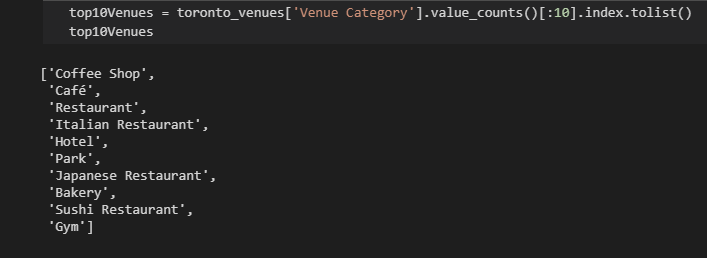


* 1. **Exploring neighbourhoods**

First, foursquare credentials and version information has been defined. Then exploration of first neighbourhood ‘Regent Park, Harbourfront’ has been started by generating foursquare API’s URL. Then, we got the results in json format. We catered it by creating a function which extracts the category of the venue. Through this function we can clean the json and structure it into pandas data frame. Now, same process has been repeated for all the neighbourhoods and hence we get a data frame in which we have venue, venue type, its location, and the neighbourhood of that venue.



After that, we found that there were 237 unique categories of venues but following were the top ten venue types which occurred frequently compared to the others.



* 1. **Finding least competitive Neighbourhood**

First, one hot encoding has been done on our main dataset and the mean has been taken to get the frequency of occurrence of any venue category in a neighbourhood.