**Introduction**

When you go to a certain it is sometimes hard find a place to visit apart from famous areas. It is obvious that every city has some places that are not widely known but considered to be one of the most visited places. Apply the data from www.foursquare.com we can get the data which shows the degree of popularity of a particular place. As a result, we can identify the places where it would beneficial to open some shop or restaurant and where not to open.

The aim of this project is to be able to explore the possibilities of where a restaurant or touristic places could be setup and be embraced in Toronto and hopefully thrive. A major benefit of pursuing this type of project is that if it is successful, it could help future restaurateurs and tourism firms to determine where next to focus their efforts.

The target audience of this project is citizens, visitors, businessmen in Toronto, Canada.

### **Data collection**

This project will require the exploration and analysis of following data:

1) Toronto Neighborhood Data - '<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada'which> which possesses a list of neighbourhoods in the Toronto area and use of a geo locator to determine the latitude and longitude of each area.

2) Foursquare Location Data, TOP tips, Favorites, User Experience, etc. will be used to cluster, segment, target, and position to craft recommendations for the new Muslim visitors.

Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, tips relating to these venues, menus and even photos. As such, the Foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

Firstly, Toronto borough neighborhood addresses are converted into their equivalent latitude and longitude using the values retrieved from Toronto geodata sheet using Pandas library in Python. After that, the Foursquare API is used to explore neighborhoods in Toronto borough. Using the explore function, the most common venued within a radius of 500 meters is extracted for each Toronto neighborhood and the number of returned venues by Foursquare API is limited to 100.

Now, the venue categories feature is used to group the neighborhoods into clusters. The k-means clustering algorithm is used to perform this task. Finally, the Folium library is employed to visualize the neighborhoods in Toronto borough and their emerging clusters.

**Approach**

The degree of popularity was identified by the frequency of visiting. Neighbourhood was divided into boroughs and 5 most visited places were taken

**Methodology**

1. Data acquisition and cleansing
2. Data preparation
3. Feature selection
4. Clustering

**Data acquisition and cleansing**

Data acquisition was a 2-step process:

1. Obtaining the postcodes for neighbourhoods in Toronto
2. Obtaining venues within these neighbourhoods

The information obtained per venue as as follows:

* Neighbourhoood
* Neighbourhood Latitude
* Neighbourhood Longitude
* Venue
* Venue Latitude
* Venue Longitude
* Venue Category

**Data preparation**

Having successfully retrieved the data, it was necessary to filter out everything except for the restaurants/food vendors. In order to achieve this, the ‘Icon Prefix’ was used as it consistently showed food information regardless of the venue category.

**Feature selection**

Data was grouped by neighbourhood and by taking the mean of the frequency of occurrence of each category. And for each neighbourhood top 5 most common venues was printed selected.

**Clustering**

Using the selected features, K-mean clustering was performed. The number of clusters chosen was 5