**Capstone Project**[**¶**](https://render.githubusercontent.com/view/ipynb?commit=8bc7a9443efd1868f5a515374a43970b718c8967&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f7365616e626f756c696e616d65656c2f436f7572736572615f43617073746f6e652f386263376139343433656664313836386635613531353337346134333937306237313863383936372f427573696e6573732532305265706f72742e6970796e62&nwo=seanboulinameel%2FCoursera_Capstone&path=Business+Report.ipynb&repository_id=183585308&repository_type=Repository#Capstone-Project)

**The Battle of the Neighborhoods**

**Introduction:**

Toronto is a vibrant, big-time city abuzz with activity. Some of the world's finest restaurants are found here, alongside happening bars and clubs and eclectic festivals. A rich, first world city known for its low crime rates and unique nightlife, Toronto stands out as a prime candidate as a place to open a new bar. However, a key piece of criteria when making a decision about where to open a new bar is getting the location right. Get it right and you could make a fortune, get it wrong and it could be a financial disaster. This project aims to help businesspeople to choose an area that is not saturated with bars and in a small way help them find a gap in the market.

**Business Problem:**

What are the best Neighborhoods in downtown Toronto for setting up a new bar? A question any budding entrepreneur worth his or her salt would be dying to know! Armed with data, this project uses Machine Learn and Data Science methods to provide answers to this business problem. Q. What are the best Neighborhoods in downtown Toronto for setting up a new bar?

**Data Acquisition and Cleaning:**

Data was pulled from the following areas -

* Canadian Postal Code Wikipedia page for lists of data about Postal Codes and their relevant Borough and Neighbourhood.
* Geocoder Package to get coordinates of Neighbourhoods in Downtown Toronto.
* Data about nearby bars from Foursquare crowdsourcing data.

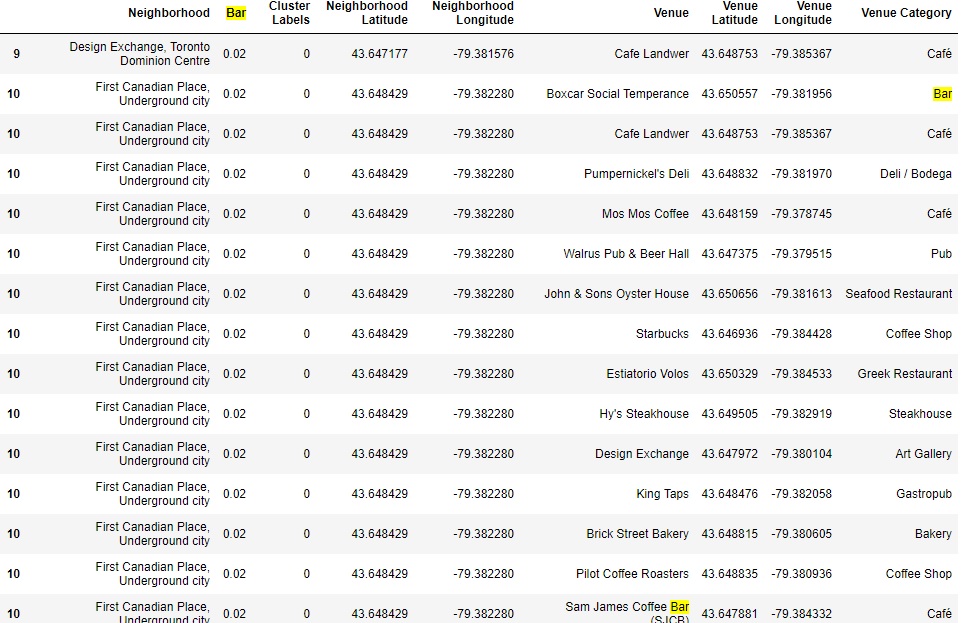
Data was scraped from the Canadian Postal Code Wikipedia page using the Beautiful Soup Library. The column needed to be re-named from 'Neighbourhood' to 'Neighborhood' so that processing issues later down the line could be avoided because 'Neighborhood' came up in multiple instances.

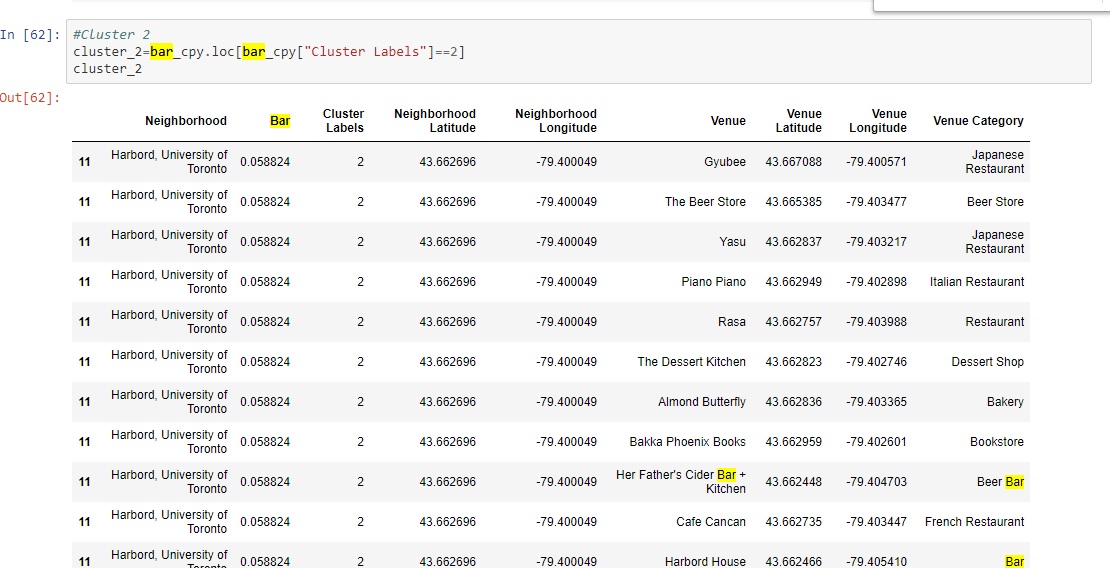
Cells with a borough were processed, cells without were ignored. I also combined postal codes which are listed twice and cells which had a borough but a Not assigned neighborhood by replacing 'Not assigned' with an entry which is same as the borough.

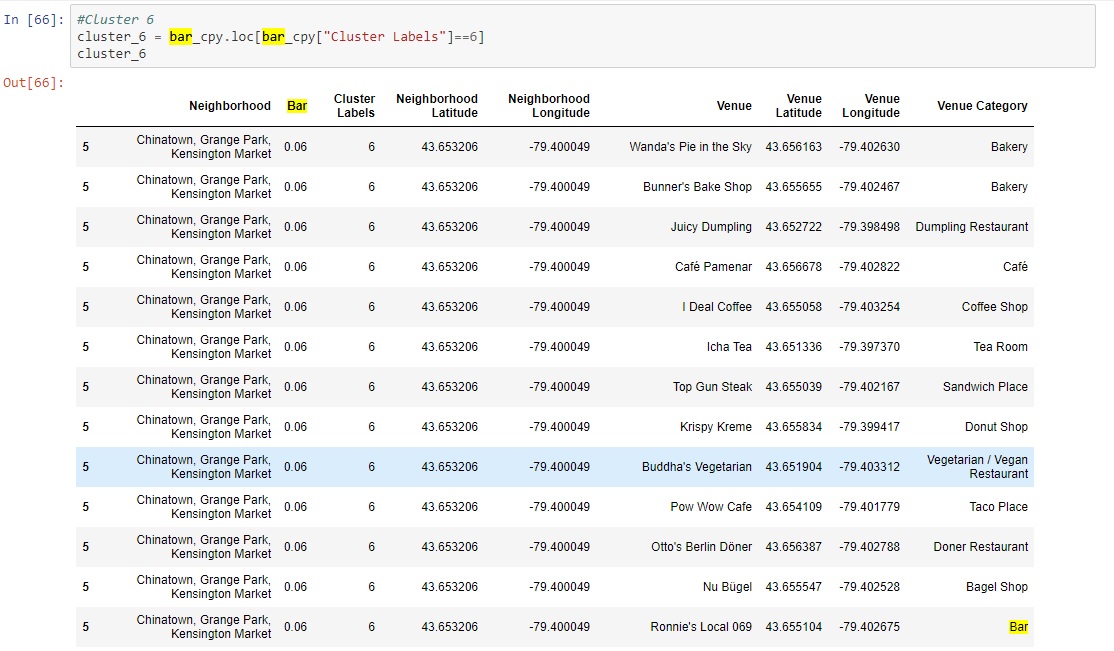
I used the CSV file to the GPS coordinates and Postal Codes of Toronto boroughs, before merging that data with data about the Named boroughs and neighborhoods - and create the final dataframe.

**Methodology:**

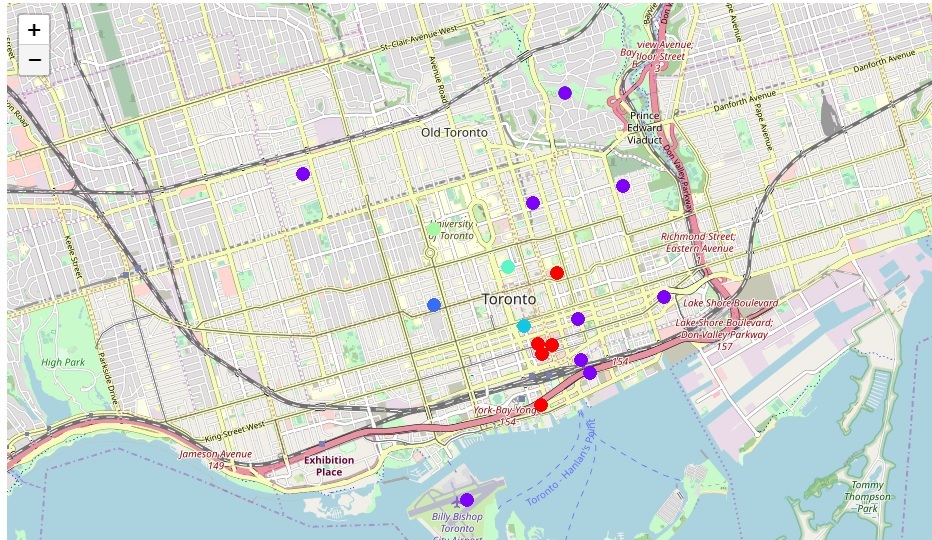
 The model we use to solve this problem is K-means clustering. We use the K-means clustering algorithm and label all the neighborhoods in downtown Toronto into 7 clusters. Here are some sample data from some of the clusters generated by our model:







We use folium to visualise the clusters and its distribution on the map.



**Results:**

There are 7 clusters:

**Cluster (0)** – A decent variety of neighbourhoods with a moderate concentration of bars.

**Cluster (1) :** The most variety of neighbourhoods with a moderate concentration of bars.

**Cluster (2) :** Harbord and University of Toronto areas, contains a high amount of bars.

**Cluster (3)** : Covers 3 neighbourhoods (Adelaide, King, Richmond) and has a moderate level of bars.

**Cluster (4) :** Two neighbourhood (Ryerson, Garden District) with a low level of bars.

**Cluster (5)** : Single neighbourhood (Central Bay Street) with a moderate concentration of bars.

**Cluster (6)** : This cluster contains a high concentration of bars, which contains the Chinatown, Grange Park and Kensington.

**Observations/Conclusion:**

It appears that the highest concentrations of bars are in Cluster 2 and 6. Cluster 2 covers the University of Toronto neighbourhood, this would make a lot of sense because it is a student neighbourhood. Cluster 6 has the highest level of bars so would be a very poor choice. Cluster 4, the Ryerson and Garden District areas have the lowest concentration of bars so would be a great place to consider investing in opening a new bar.