

# DATA SCIENCE CAPSTONE PROJECT

---

## *Battle of the Neighborhoods*

---

### **Introduction**

Data Science can be used to make business decisions. The decisions will be based on the analysis of large data sets. The use of Machine Learning Algorithms makes the analysis possible as these large data sets can be analyzed.

### **Business Problem**

For this project, Manhattan in New York City and Toronto City will be explored using Foursquare Location Data. Foursquare Location data gives venues and ratings of these venues based on Foursquare users. This data gives a fair picture of different venues found in a particular area. A comparison of Manhattan and Toronto City will be done based on the type of venues and frequency of each type.

## **Target Audience**

When starting a business, an investor needs to do SWOT analysis. This project will help potential investors who are looking at either starting a small business in either Manhattan in New York City or Toronto City. An investor needs to know who is the competition and where are they located.

When deciding on starting a certain type of business, an investor can find out which is the most common business and what is their rating according to customers. This will help in making an informed decision on location and type of business they can start.

## **Data**

For this project, data on Manhattan and Toronto Boroughs and Neighborhoods will be used. The data shows the latitude and longitude of the different boroughs and neighborhoods. This data will then be used in the Foursquare API to locate venues within the boroughs and neighborhoods. The venues will then be classified into different categories.

## newyork data. json from IBM Skills Network Labs

```
[10]: neighborhoods.head()
```

```
[10]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

## Toronto City data

Beautiful Soup is used to scrap web data from the following url: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

This data is used to identify the different boroughs and neighborhoods in Canada. The latitudes and longitudes of these places is used in the Foursquare API.

## **METHODOLOGY**

The Python GeoPy function is used to get the latitudes and longitude of New York City and Toronto City. Boroughs and neighborhoods of each city are then obtained. Using Four Square API, venues in these two cities are identified and classified since Four Square API has a venue's category and rating data. It also has the latitudes and longitudes of these venues. Venues within a given location that have been rated by Four Square API users can be identified.

Exploratory data analysis was done to identify the most common venue in each city. This information can be used by an investor to identify the most common venue and their location in the city. The Python library, Folium was used to show the locations of the different venues around the city.

Cluster analysis can be used to identify clusters by Venue categories.

# Results

## Most Common Venue in Manhattan

```
print(manhattan_venues.shape)
manhattan_venues.head()
manhattan_venues.groupby("Venue Category").count().sort_values("Neighborhood", ascending=False).head(10)
```

(3259, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Venue Category						
Coffee Shop	133	133	133	133	133	133
Italian Restaurant	130	130	130	130	130	130
Café	82	82	82	82	82	82
American Restaurant	78	78	78	78	78	78
Pizza Place	77	77	77	77	77	77
Bakery	74	74	74	74	74	74
Park	70	70	70	70	70	70
Hotel	68	68	68	68	68	68
Bar	62	62	62	62	62	62
Mexican Restaurant	55	55	55	55	55	55

Exploratory data analysis reveals that the most common venue in Manhattan is a Coffee Shop followed by Italian Restaurant.

## The most common venue in Toronto is also the Coffee Shop

```
print(toronto_venues.shape)
toronto_venues.head()
toronto_venues.groupby("Venue Category").count().sort_values("Neighborhood", ascending=False).head(10)
```

(2124, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Venue Category						
Coffee Shop	178	178	178	178	178	178
Café	99	99	99	99	99	99
Restaurant	70	70	70	70	70	70
Park	52	52	52	52	52	52
Bakery	52	52	52	52	52	52
Pizza Place	44	44	44	44	44	44
Hotel	43	43	43	43	43	43
Italian Restaurant	42	42	42	42	42	42
Japanese Restaurant	41	41	41	41	41	41
Sandwich Place	39	39	39	39	39	39

Exploratory Data Analysis is used to investigate which is the most common venue .

In Toronto, most common Coffee shop is Starbucks, with Tim Hortons coming second. The other coffee shops have fewer outlets.

```
filtered1= toronto_venues[toronto_venues['Venue Category']=='Coffee Shop']
filtered1.groupby("Venue").count().sort_values("Neighborhood",ascending=False).head(20)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category
Venue						
Starbucks	59	59	59	59	59	59
Tim Hortons	40	40	40	40	40	40
Pilot Coffee Roasters	9	9	9	9	9	9
Dineen @CommerceCourt	5	5	5	5	5	5
Aroma Espresso Bar	4	4	4	4	4	4
Balzac's Coffee	4	4	4	4	4	4
M Square Coffee Co	4	4	4	4	4	4
Bulldog On The Block	3	3	3	3	3	3
Second Cup	3	3	3	3	3	3
Everyday Gourmet (Teas & Coffees)	3	3	3	3	3	3
Mos Mos	2	2	2	2	2	2

In Manhattan, the most common Coffee shop is Starbucks followed by Blue Bottle Coffee.

```
#manhattan_venues.loc['Venue Category' == 'Coffee Shop' ]
#manhattan_venues.head()
filtered= manhattan_venues[manhattan_venues['Venue Category']=='Coffee Shop']
filtered.groupby("Venue").count().sort_values("Venue Category",ascending=False).head(10)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category
Venue						
Starbucks	15	15	15	15	15	15
Blue Bottle Coffee	9	9	9	9	9	9
Oslo Coffee Roasters	4	4	4	4	4	4
Joe Coffee Company	4	4	4	4	4	4
La Colombe Torrefaction	4	4	4	4	4	4
787 Coffee	3	3	3	3	3	3
Matto Espresso	3	3	3	3	3	3
Laughing Man Coffee & Tea	3	3	3	3	3	3
The Brown Cup	2	2	2	2	2	2
For Five Coffee Roasters	2	2	2	2	2	2

Cluster analysis of the different neighborhoods in both places was carried out.

## **Recommendations**

In both Manhattan and Toronto, the Coffee Shop business is dominated by big brands. A new investor can either set up next to these big brands or look for places where these big brands do not have outlets.

## **Conclusion**

The approach used in this report can be used by an investor to identify competition when doing a SWOT analysis. The limitation is that the report only identifies the most common Venues using Four Square API. Those venues not rated by Four Square users are not included in this analysis. Further analysis of the data will be carried on location data of the venues as an improvement of the project.