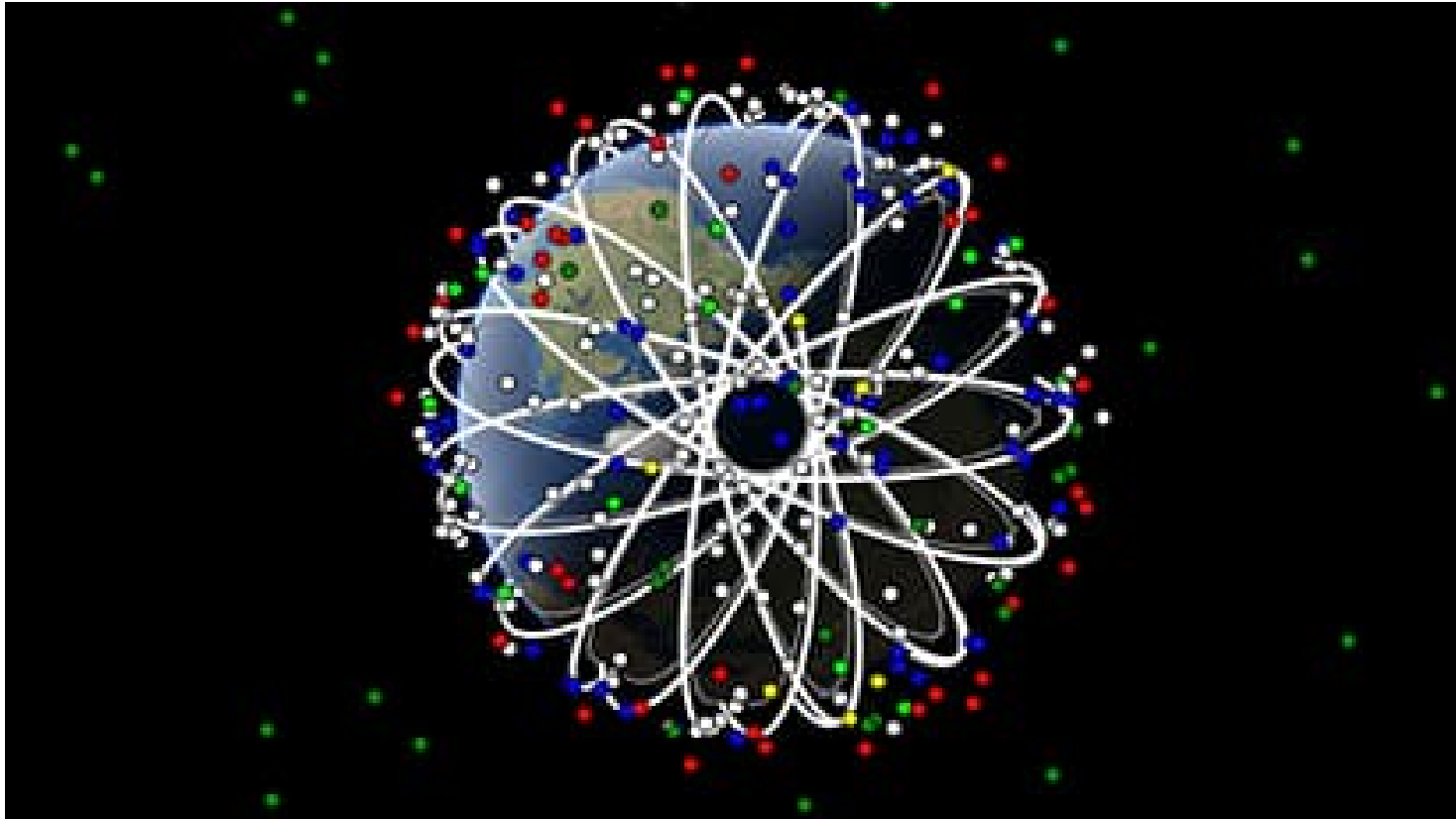


# Capstone Project On Neighborhoods of New York and Toronto



## **Business Problem :**

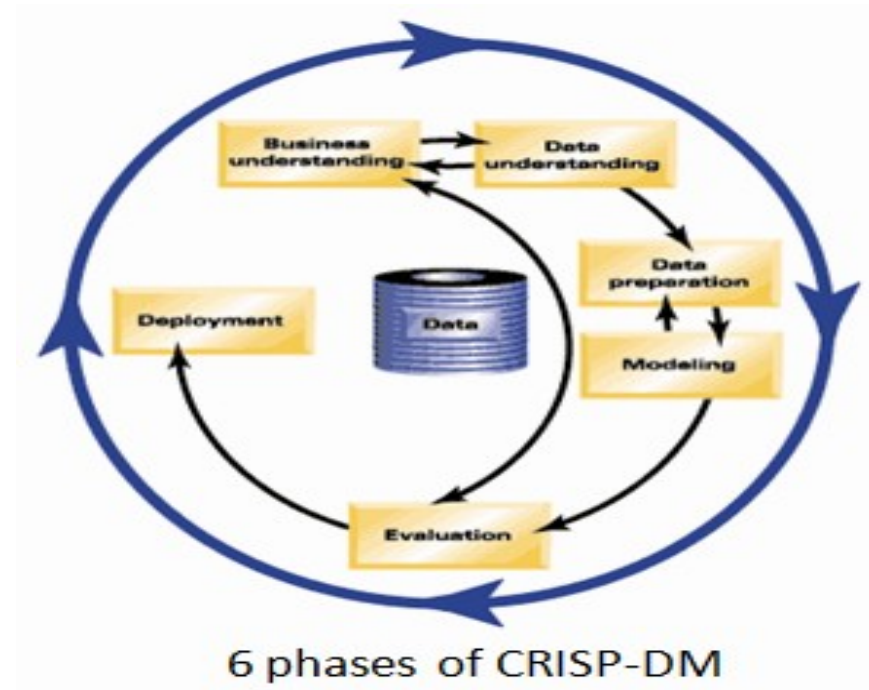
New York and Toronto are the cities of Canada and These two cities having many neighborhoods and stores. The stores are like Restaurants, Coffee shops, Super Markets etc. From these we should find out best stores and neighborhoods wherever these stores are required and need in the neighborhoods. We should find out best neighborhood and store from these two cities using with Foursquare map.

## **Requirements**

- > Need an account in FourSquare for Locations and Stores
- > Need Neighborhoods with Latitude and Longitude
- > Knowledge in Data Analysis, Data Visualization, Data Processing and Machine learning algorithms with any Language Python or R.
- > Need Data for Two Cities of New York and Toronto

## We need to follow these steps

- >> Business Problem
- >> Data Understanding
- >> Data Preparation
- >> Modeling
- >> Evaluation
- >> Deployment



Using with above steps, we can find out the best solution

## Required these Python Libraries for this Project :

```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API lab
import folium # map rendering library
```

## Required FourSquare Parameters :

Client Id, Client Secrete and Version of the URL.

```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(  
    CLIENT_ID,  
    CLIENT_SECRET,  
    VERSION,  
    neighborhood_latitude,  
    neighborhood_longitude,  
    radius,  
    LIMIT)  
url = url.replace(" ", "%20")
```

From above URL, you have to mention Latitude and Longitude for specific city (New York or Toronto).

Go through this URL

>> <https://developer.foursquare.com/>

# FourSquare Venues with ratings:


**FOURSQUARE**

Museums

Current Map View

Suggestions for **Museums** near **Toronto**

Filters: Specials Haven't Been Following Price Open Now Saved Liked




**1. Royal Ontario Museum**  
Museum • \$\$\$  
100 Queen's Pk (at Bloor St. W.), Toronto  
[Save](#)

9.2


↑

Ratings




Tobias N. • May 21, 2016

Great museum. Especially loved the exhibitions about Canada and the dinos. Come early in the morning to avoid huge crowds.




**2. Hockey Hall of Fame**  
Museum  
30 Yonge St. (at Brookfield Pl.), Toronto  
[Save](#)

9.0




Sergiy G. • June 11, 2017

Lots of hockey artefacts and The Cup! The staff is very friendly, so besides making selfie, you can even touch the cup! P.S. Ukrainians will enjoy seeing equipment of famous Terry Sawchuk (pictured)



**3. Aga Khan Museum**  
History Museum  
77 Wynford Dr (at Don Valley Pkwy), Toronto  
[Save](#)

9.0

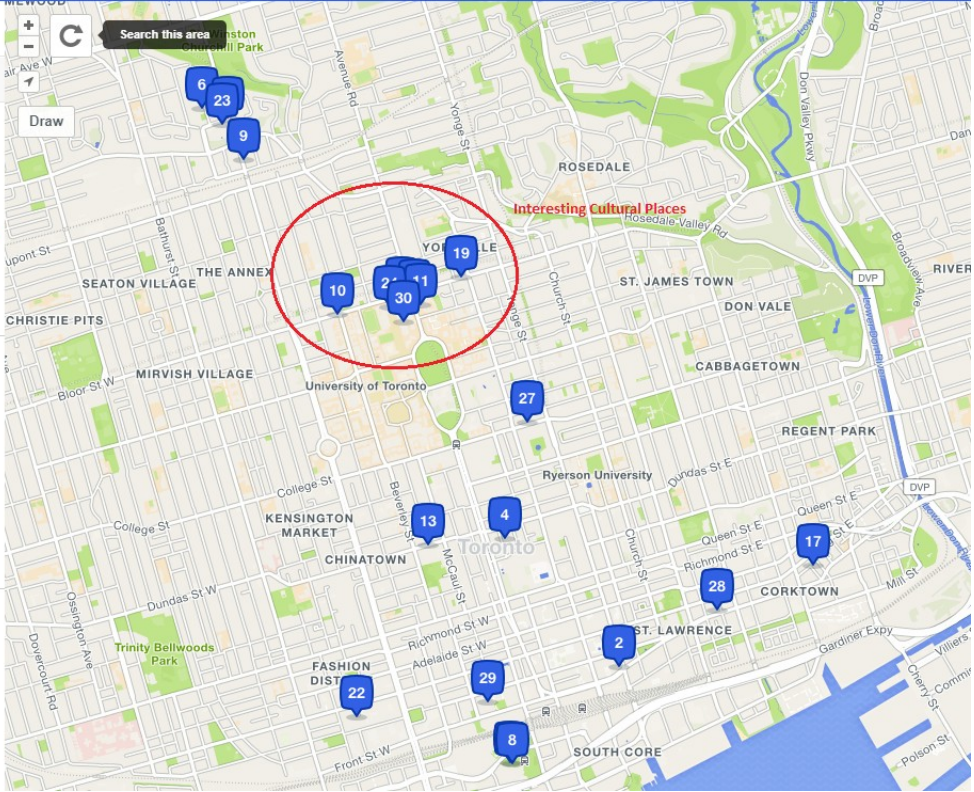


Oliver N. • February 16, 2016

Amazing museum with a lovely building.

Search this area

Draw



Map of Toronto showing museum locations with ratings. A red circle highlights the area around the University of Toronto and the Royal Ontario Museum.

## New York Data with Latitude and Longitude :

	Borough	Neighbourhood	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 Data with Latitude and Longitude :

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

## **Data Understanding :**

After Business problem statement, we need to understand the data like

- is this Data sufficient for the project or not ?
- What are the parameters needed ?
- We should understand whether this problem is Supervised problem or Unsupervised problem (Depending on the problem statement)
- What type of data, we are dealing with like Text, Table format, Image, audio and Video type of data ?
- what type of python libraries needed for this project ?



## **Data PreProcessing :**

We should acquire the data using with Pandas Library like

```
>> import pandas as pd  
>> pd.read_csv("path")
```

We can read the different type of the data like .CSV, .xls, .html, .text etc.

## **Clean the data :**

- \* Missing values
- \* Wrong data
- \* Duplicate values
- \* Normalize the data
- \* Data Conversion

# Data Modeling :

This Project is belongs to unsupervised mechanism. In this project we used K- Means clustering algorithm.

```
# import k-means from clustering stage
from sklearn.cluster import KMeans

# set number of clusters
kclusters = 5

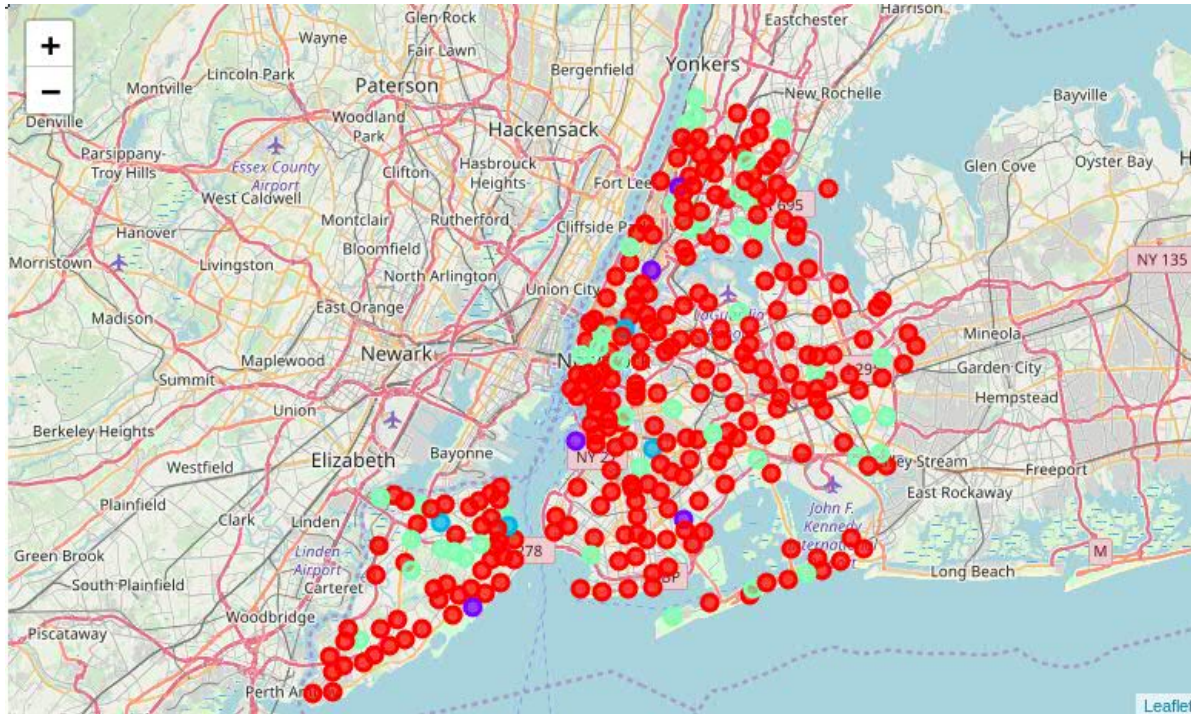
newyork_grouped_clustering = newyork_grouped.drop('Neighbourhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(newyork_grouped_clustering)

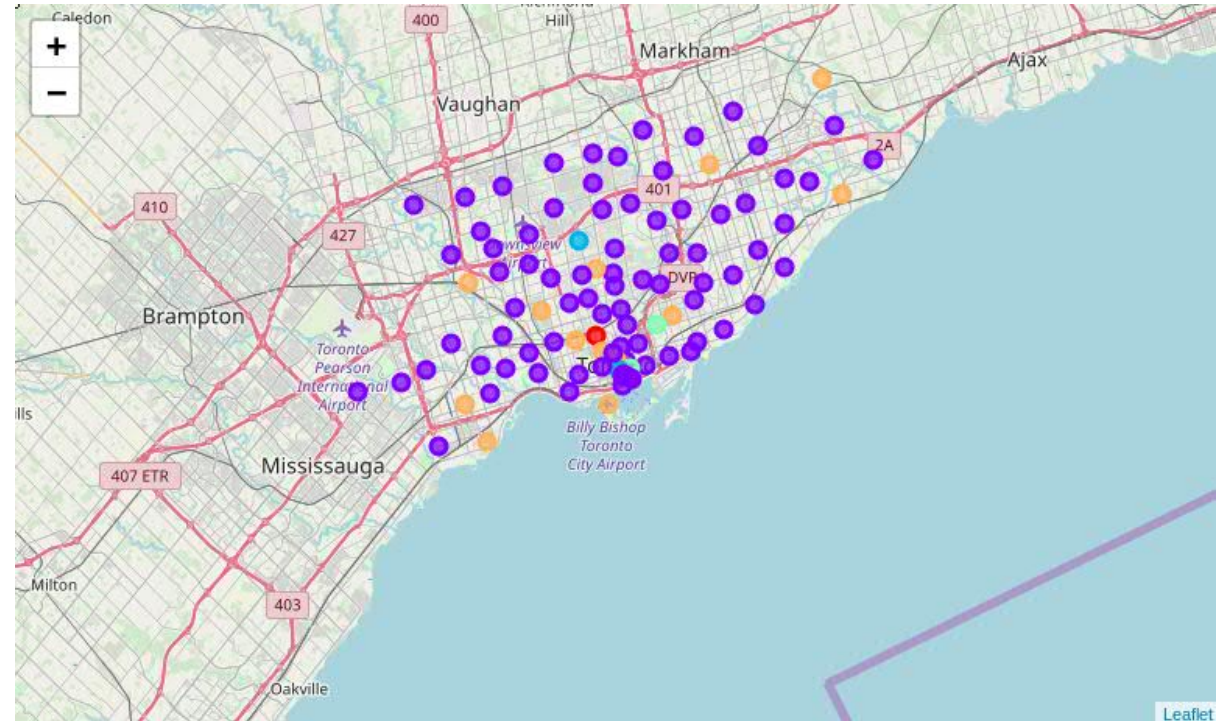
# check cluster labels generated for each row in the dataframe
kmeans.labels [0:10]
```

# Data Visualization :

## New York



## Toronto



## Data Evaluation :

We have different types of Evaluation models

>> Train\_test\_split

>> Crossvalidation\_Score

>> Metrics (Accuracy Score)

Example:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=0)
kmeans.fit(X_train, y_train)

yhat = kmeans.predict(X_test)
```

## Deploy the Project :

Deploying Machine Learning into production is hard. You need to:

- >> Build and use the right cloud infrastructure on the right cloud provider
- >> Design and implement public and internal APIs for model usage
- >> Orchestrate a load of containers
- >> Implement a load balancer to ensure you can scale to meet inference needs
- >> Integrate with data pipelines and consistently update models

## Result of New York :

	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most different Venue	2nd Most different Venue	3rd Most different Venue	4th Most different Venue	5th different Venue
0	Bronx	Wakefield	40.894705	-73.847201	0	Accessories Store	Peruvian Restaurant	Persian Restaurant	Performing Arts Venue	Pedestrian Plaza
1	Bronx	Co-op City	40.874294	-73.829939	0	Jazz Club	Peruvian Restaurant	Persian Restaurant	Performing Arts Venue	Pedestrian Plaza
2	Bronx	Eastchester	40.887556	-73.827806	3	Accessories Store	Pet Café	Peruvian Restaurant	Persian Restaurant	Performing Arts Venue
3	Bronx	Fieldston	40.895437	-73.905643	3	Accessories Store	Peruvian Restaurant	Persian Restaurant	Performing Arts Venue	Pedestrian Plaza
4	Bronx	Riverdale	40.890834	-73.912585	3	Accessories Store	Pet Café	Peruvian Restaurant	Persian Restaurant	Performing Arts Venue

## Result of Toronto :

	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most different Venue	2nd Most different Venue	3rd Most different Venue
0	Scarborough	Malvern	43.806686	-79.194353	1	Accessories Store	Mediterranean Restaurant	Men's Store
1	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	1	Accessories Store	Medical Center	Mediterranean Restaurant
2	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	4	Accessories Store	Martial Arts Dojo	Massage Studio
3	Scarborough	Woburn	43.770992	-79.216917	1	Accessories Store	Medical Center	Mediterranean Restaurant
4	Scarborough	Cedarbrae	43.773136	-79.239476	1	Accessories Store	Medical Center	Mediterranean Restaurant



# Discussion :

By comparing above two Cities

Best recommend is in Toronto city and New York, Choices to open a store "Accessories Store" we have different type of restaurants in that best recommend restaurants are

**Peruvian,Persian,Pakistani,paella Restaurants (To open in the place "Bronx" of New York city)**

**Mediterranean,Mexican,Middle Estern,Modern European Restaurants (To open in the place "Scarborough" of Toronto city)**

We can able to choose many stores like this....

## Conclusion :

- The analysis of this data is rather limited because we were to use Foursquare. The basis for the ranking positions is based on this platform and are given by users directly. It's possible to have different results if the data set were other than Foursquare, for example, official data provided by some Government division.
- The Clustering function can be changed and will yield different clusters. We assumed 5, but stakeholders can change this at will.
- Was interesting to note that Museums, are popular but not as much as one might think.
- Given the time allowed and the limited data, this was an interesting project that could very well benefit the Tourism agencies of Toronto!



**THANK YOU...**