Capstone Project On Neighborhoods of Toronto

Business Problem :

Toronto has many stores, restaurants, Coffee shops etc. The objective is to recommend the owner where to open the stores/restaurant. We shall try to find out the best neighborhood and store in Toronto using with Foursquare map.

Requirements:

1. Foursquare developer account for API calls.
2. Need Neighborhoods with Latitude and Longitude
3. Knowledge of Data Analysis, Data Visualization,Data Processing and Machine learning algorithms.
4. Need Data for Toronto City.

Methodology:

CRISP-DM – **Cr**oss**I**ndustry**S**tandard**P**rocess for **D**ata**M**ining

The typical steps involved in CRISP-DM includes :

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

Data Requirements :

- We would need the Foursquare location data to solve the problem.
- We would also need Geospatial_data for Toronto location.

Required Python Libraries for this Project :

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

import pandas as pd # library for data analsysis
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.

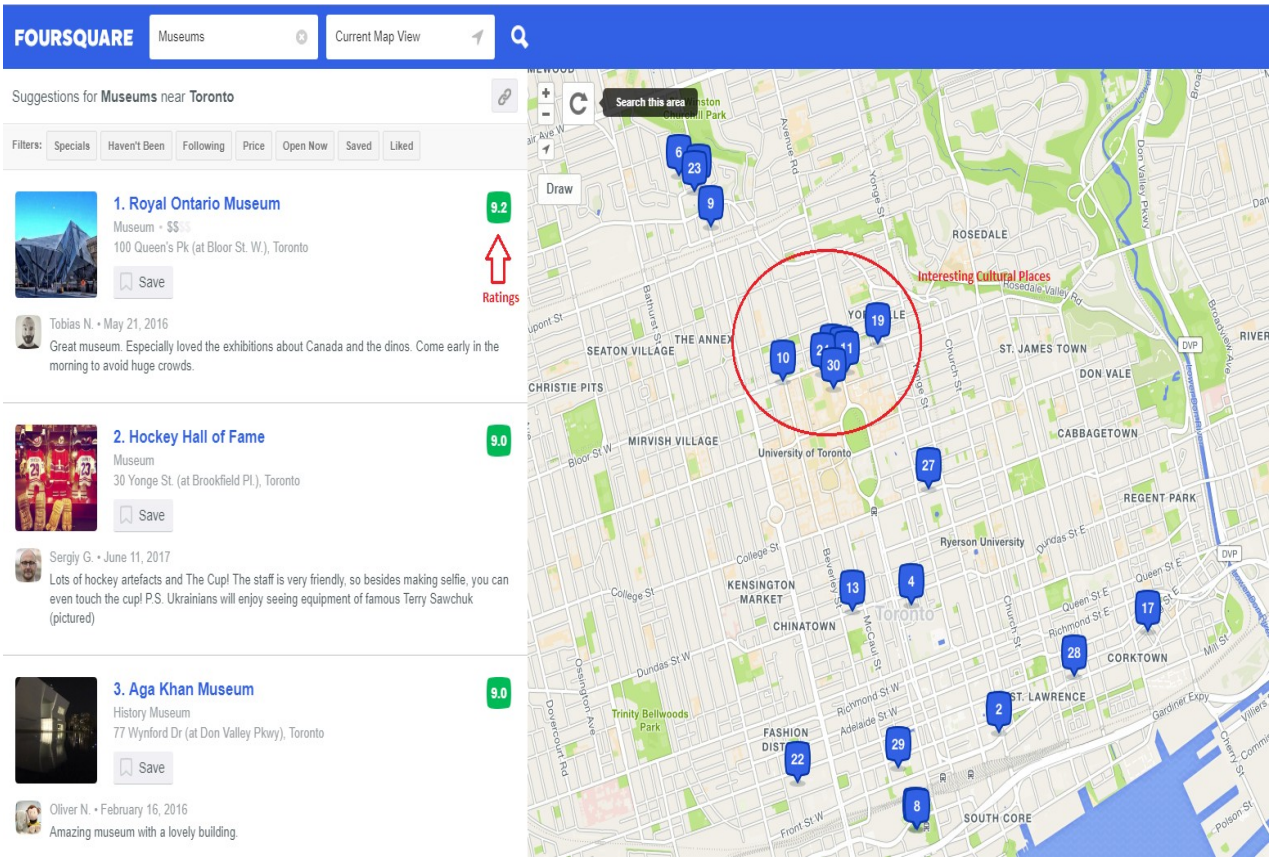
```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
```

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

Go through this URL

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

FourSquare Venues with ratings:



Data Understanding :

Things that matter :

– Is this Data sufficient ?

– What are the parameters required ?

– Is the problem Supervised or Unsupervised (Depending on the problem statement)

– Type of data we are dealing with i.e.,Text, Table format, Image, audio or Video

– What libraries are required for this project ?

Data PreProcessing :

Acquire the data using Pandas;

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

We can read the different formats of 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.

```
from sklearn.cluster import KMeans
```

```
# set number of clusters
kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighbourhood', 1)

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

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

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)
```

Discussion

0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	Accessories Store	Massage Studio	Medical Center	Mediterranean Restaurant	Men's Store	Metro Station	Mexican Restaurant	Mi Eas Restau
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Accessories Store	Massage Studio	Medical Center	Mediterranean Restaurant	Men's Store	Metro Station	Mexican Restaurant	Mi Eas Restau
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	Accessories Store	Malay Restaurant	Market	Martial Arts Dojo	Massage Studio	Mediterranean Restaurant	Men's Store	M Sta
3	M1G	Scarborough	Woburn	43.770992	-79.216917	Accessories Store	Massage Studio	Medical Center	Mediterranean Restaurant	Men's Store	Metro Station	Mexican Restaurant	Mi Eas Restau
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Accessories Store	Martial Arts Dojo	Massage Studio	Medical Center	Mediterranean Restaurant	Men's Store	Metro Station	Mex Restau

Mediterranean Restaurant, Mexican Restaurant, Middle Eastern Restaurant are recommended to open in Scaborough of Toronto.

Conclusion :

The analysis of this data is rather limited because we used only Foursquare. The ranking positions users' ratings.

The Clustering function can be changed and will yield different clusters. We assumed 5, but stakeholders can change this at will.

Given the time allowed and the limited data, this was an interesting project that could very well benefit the Tourism agencies of Toronto!