**Project Description**

One of an intriguing topic used in Data science is to compare neighborhoods of two cities or districts are similar or disimilar. People who are planning to do a bussiness, restaurants or want to construct a house would be benefitted from the comparison of these. The main purpose of the project is to ana;yse geographical locations and to help people understand the geographical advantages when starting or planning to do any sorts of investments.

**Data used in this project**

To make good use of the foursquare location data, the labs done for Toronto segmented and clustered neighborhoods gave me good examples for new ideas. I would need to leverage the foursquare location data to solve or execute. Data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that I explicitly describe the audience and why they would care about my problem. From the results on New York, the lab especially tells 10 most common venues in each district. The data set is about Toronto Venues. After transforming the data to Json files, it is easy to use Pandas to transform them into DataFrame. Then select the required data columns to appear and get the information we want. Geopy library is used to get the latitude and longitude values. With the locations the maps can be created and we will have a direct understanding of the neighborhoods and the purposed areas. The Folium library is also used to show maps on different requirements

**Code**

*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*

*get\_ipython().system("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*

*get\_ipython().system("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*

*print('Libraries imported.')*

*# In[2]:*

*address = 'Toronto, Canada'*

*geolocator = Nominatim(user\_agent='capstone\_project')*

*location = geolocator.geocode(address)*

*latitude = location.latitude*

*longitude = location.longitude*

*print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude))*

*# In[3]:*

*from bs4 import BeautifulSoup*

*import requests*

*import bs4*

*html = "https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M"*

*fsa=requests.get(html)*

*fsa.raise\_for\_status()*

*fsa.encoding=fsa.apparent\_encoding*

*fsa.text*

*x=fsa.text*

*soup = BeautifulSoup(x)*

*content = soup.findAll('td')*

*tables=[]*

*for data in content:*

*tables.append(data.text.strip())*

*new\_tables=[tables[i:i+3] for i in range(0,len(tables),3)]*

*new\_tables*

*# In[4]:*

*import pandas as pd*

*column\_names = ['Postcode','Borough', 'Neighborhood']*

*neighborhoods =pd.DataFrame(columns=column\_names)*

*# In[5]:*

*neighborhoods*

*# In[18]:*

*for x in range(0,289):*

*neighborhoods = neighborhoods.append(*

*{'Postcode': new\_tables[x][0],*

*'Borough': new\_tables[x][1],*

*'Neighborhood': new\_tables[x][2]*

*},ignore\_index=True)*

*# In[19]:*

*neighborhoods*

*# In[20]:*

*df1=neighborhoods.copy()*

*df1=df1[~df1['Borough'].isin(['Not assigned'])]*

*df1['Neighborhood']= df1['Neighborhood'].replace('Not assigned', df1['Borough'])*

*df1*

*# In[21]:*

*d2=df1.groupby(['Postcode','Borough'], as\_index=False)*

*df1=d2.agg(lambda x: ', '.join(x))*

*df1.head(20)*

*# In[22]:*

*df\_cor = pd.read\_csv('http://cocl.us/Geospatial\_data')*

*df\_cor.head(5)*

*# In[23]:*

*df\_with\_cor = pd.merge(df1, df\_cor, left\_on='Postcode', right\_on='Postal Code', suffixes=('', '')).drop('Postal Code', 1)*

*df\_with\_cor.head(10)*

*# In[29]:*

*map\_toronto = folium.Map(location=[latitude, longitude], zoom\_start=10)*

*for lat, lng, borough, neighborhood in zip(df\_with\_cor['Latitude'], df\_with\_cor['Longitude'], df\_with\_cor['Borough'], df\_with\_cor['Neighborhood']):*

*label = '{}, {}'.format(neighborhood, borough)*

*label = folium.Popup(label, parse\_html=True)*

*folium.CircleMarker(*

*[lat, lng],*

*radius=5,*

*popup=label,*

*color='blue',*

*fill\_color='#3186cc',*

*fill\_opacity=0.7,*

*).add\_to(map\_toronto)*

*map\_toronto*

*# In[30]:*

*def getNearbyVenues(names, latitudes, longitudes, radius=500):*

*venues\_list=[]*

*for name, lat, lng in zip(names, latitudes, longitudes):*

*# create the API request URL*

*url = 'https://api.foursquare.com/v2/venues/explore?&client\_id={}&client\_secret={}&v={}&ll={},{}&radius={}'.format(*

*CLIENT\_ID,*

*CLIENT\_SECRET,*

*VERSION,*

*lat,*

*lng,*

*radius)*

*# make the GET request*

*results = requests.get(url).json()["response"]['groups'][0]['items']*

*# return only relevant information for each nearby venue*

*venues\_list.append([(*

*name,*

*lat,*

*lng,*

*v['venue']['name'],*

*v['venue']['location']['lat'],*

*v['venue']['location']['lng'],*

*v['venue']['categories'][0]['name']) for v in results])*

*nearby\_venues = pd.DataFrame([item for venue\_list in venues\_list for item in venue\_list])*

*nearby\_venues.columns = ['Neighborhood',*

*'Postcode Latitude',*

*'Postcode Longitude',*

*'Venue',*

*'Venue Latitude',*

*'Venue Longitude',*

*'Venue Category']*

*return(nearby\_venues)*

*# In[31]:*

*boroughs = set(df\_with\_cor['Borough'])*

*toronto\_boroughs = {borough for borough in boroughs if 'Toronto' in borough}*

*toronto\_data = df\_with\_cor[df\_with\_cor['Borough'].isin(toronto\_boroughs)].reset\_index(drop=True)*

*toronto\_data.head(5)*

*# In[38]:*

*CLIENT\_ID = 'PXXFRZXNDT3HAAGE2PTLXXZTSNVD0PKRE41KMGDP3JOSRDN5' # your Foursquare ID*

*CLIENT\_SECRET = 'UO2TGIOWALSIT1QK4XLNIZYT2NJFNN4KC5HDGSJVRRCFYAY1' # your Foursquare Secret*

*VERSION = '20190101' # Foursquare API version*

*print('Your credentails:')*

*print('CLIENT\_ID: ' + CLIENT\_ID)*

*print('CLIENT\_SECRET:' + CLIENT\_SECRET)*

**Methodology**

This was only an exploratory analysis.First of all, the Foursquare API was utilized because basic geographical location information and particular venues can be explored through Foursquare API. With a specific credentials client ID and secret, the required foursquare data can be accessed.

Second, Folium is a great visualization library, the map can directly show us the overall shape of the place we want to know. The detailed information on different venues such as restaurants can be seen in a direct way. The investors can have a better understanding of the city of Toronto to choose his properties. People feel free to zoom into the above map, and click on each circle mark to reveal the name of the neighborhood and its respective borough.

**Results**

In Toronto, venues types are various and investors can choose their preferable neighborhoods to open the restaurants and the companies. In my opinion, their should be more crowds to make the neighborhood more developed. Data shown in my notebook told us comprehensive kinds of venues in different areas of Toronto

**Discussion**

From data of top 10 common venues in Toronto, the best place for investors to open a restaurant or start a company should not be one, but also several choices. The data only showed us the current numbers of each kind of venues but no other factors such as cost of the place or environment.

**Conclusion**

The investors should consider the competitive restaurants in the district. Try not to compete with those nearby with more similar kinds of restaurants as yours, that may do bad for the kind of investment. But for start a company, it is good to choose the place with more kinds of restaurants so that you will attract more employees to be hired.