Capstone Report

Finding the Right Neighborhoods for a Franchise in a New City

Introduction

In this project, I assume Mitsuwa Marketplace, the famous Japanese supermarket chain, is considering opening their first store in Tonronto, but where to open their store in a new city is a challenge. As Mitsuwa Marketplace has a store in New Jersey, USA, we can compare surrounding area of their New Jersey store and Toronto neighborhoods with the help of data analytics, and suggest best matched Toronto neighborhoods to open their second store in Texas.

Data

In order to perform this project, following data sets are required from different resources.

- Mitsuwa New Jersy store's latitude and longitude values are needed which should be queried from Geocoder Python package or other resources.
- Name, latitude and longitude values of Toronto neighborhood that can be pulled from Geocoder Python package or scarping from websites.
- Foursquare location data of Mitsuwa New Jersey store and Toronto neighborhoods to dig top 10 categories of places for each neighborhood and Mitsuwa New Jersey store with a radius of 1000 meters.

Finally, a clustering analysis will be performed with above data sets to discover most common neighborhood(s) in Toronto to Mitsuwa New Jersey store, and provide news store location recommendation to stakeholders.

Concretely, in order to analysis common places for each neighborhood and Mitsuwa market, we need to know their coordination and get data from Foursquare database.

First, we get Mitsuwa Coordination data from geolocator package.

```
address = 'Mitsuwa, New Jersey'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Mitsuwa Marketplace, New Jersey is {}, {}.'.format(latitude, longitude))
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main__.py:3: DeprecationWarning: Using Nominatim with the defa
ult "geopy/1.18.1" `user_agent` is strongly discouraged, as it violates Nominatim's ToS https://operations.osmfoundation.org/po
licies/nominatim/ and may possibly cause 403 and 429 HTTP errors. Please specify a custom `user_agent` with `Nominatim(user_age
nt="my-application")` or by overriding the default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-applicatio
n"`. In geopy 2.0 this will become an exception.
  app.launch_new_instance()
```

```
The geograpical coordinate of Mitsuwa Marketplace, New Jersey is 40.815989099999996, -73.97997001693285.
```

Second, we get Toronto neighborhoods' names, zip codes, coordination from webpage.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494
2	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Thirdly, we merge Mitsuwa New Jersey and Toronto neighborhoods' data into one data frame and perform clustering which will be discussed in following sections.

Adding Mitsuwa Marketplace data into the dataframe for cluestring ¶

```
df_mit ={'Borough':'New Jersey','Neighbourhood':'Mitsuwa','Latitude':location.latitude,'Longitude':location.longitude}
df_t=df_t.append(df_mit,ignore_index=True)
```

Methodology

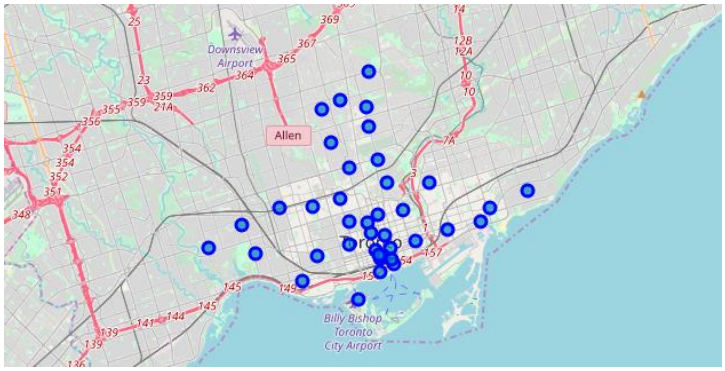
After getting geographic data sets, data visualization techniques were performed to display each neighborhood and even Mitsuwa New Jersey on the same map.

Furthermore, I call Foursquare API Data to get venues information and categories of each neighborhood with a radius of 1,000 meters. Then exploratory data analysis is performed to study each neighborhood's top 5 popular venue categories with frequency.

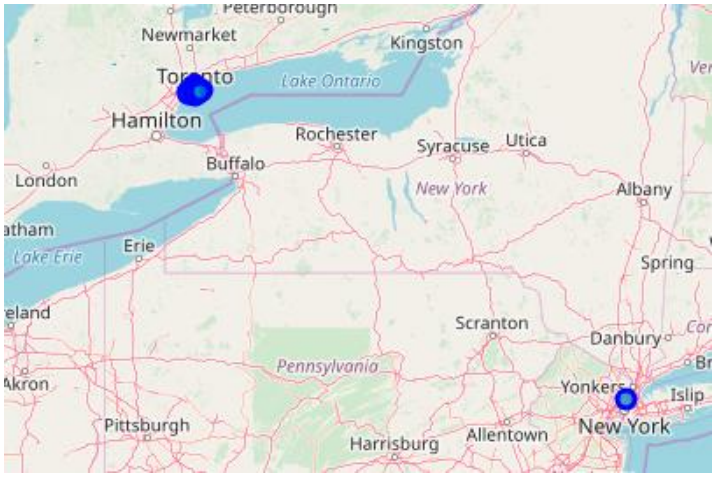
Finally, in order to find best matched Toronto neighborhoods to Mitsuwa New Jersey, clustering algorithm will be applied. I used k-nearest neighbohoodrs algorithm to cluster Toronto's neighborhood plus Mitsuwa New Jersey into 12 clusters by suing their venue categories data.

Detailed steps are illustrated as following:

Apply folium package to display Toronto Map and denote neighborhoods on the map.



And after zooming out, we actually can find a point in New Jersey area which the Mitsuwa Marketplace locates.



Next, we call Foursquare API to explore venues within 1000 meters of each neighborhood. We limit total number of venues from each neighborhood to 1000, in case we get too many venues in a certain neighborhood.

```
LIMIT = 100

radius = 1000 # define radius

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

For exploratory data analysis, we study the venues categories distribution for each neighborhood and calculate 5 most popular venue categories in each neighborhood. Partial result which includes Mitsuwa and other two Toronto neighborhoods are shown as following.

```
----Mitsuwa----
                venue  freq
0  Japanese Restaurant  0.12
1                  Gym  0.06
2         Dessert Shop  0.04
3          Pizza Place  0.04
4       Cosmetics Shop  0.04


----Moore Park,Summerhill East----
         venue  freq
0   Restaurant  0.33
1   Playground  0.33
2        Trail  0.33
3      Museum  0.00
4      Market  0.00


----North Toronto West----
                venue  freq
0        Clothing Store  0.19
1  Sporting Goods Shop  0.10
2          Coffee Shop  0.10
3  Gym / Fitness Center  0.05
4                 Café  0.05
```

Finally, KNN algorithm is applied to the data sets which include all Toronto neighborhoods and Mitsuwa New Jersey information. I set number of cluster to 12, and run k-means clustering as following:

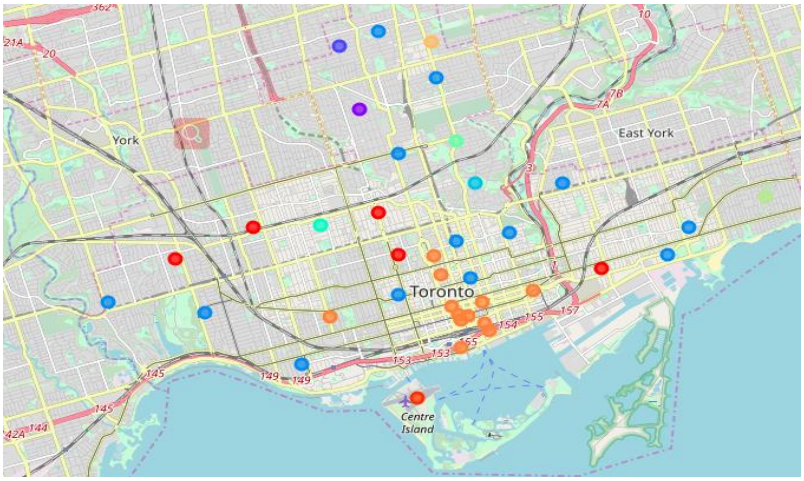
```
kclusters = 12

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

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

Results

I generated a table to display each Toronto neighborhood and Mitsuwa New Jersey with their corresponding cluster labels. The partial table with Mitsuwa and some other Toronto neighborhood are listed as following.

32	M5V	Downtown Toronto	CN Tower,Bathurst Quay,Island airport,Harbourf...	43.628947	-79.394420	11
33	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	4
34	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	10
35	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675	3
36	M5X	Downtown Toronto	First Canadian Place,Underground city	43.648429	-79.382280	10
37	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	3
38	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	3
39	NaN	New Jersey	Mitsuwa	40.815989	-73.979970	3

Visualize cluster results for Toronto neighborhood:



The all Toronto neighborhoods that match Mitsuwa New Jersey's surrounding environment are listed as following:

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels
2	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937	3
12	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	3
14	M6K	West Toronto	Brockton,Exhibition Place,Parkdale Village	43.636847	-79.428191	3
15	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	3
23	M4R	Central Toronto	North Toronto West	43.715383	-79.405678	3
25	M6R	West Toronto	Parkdale,Roncesvalles	43.648960	-79.456325	3
26	M4S	Central Toronto	Davisville	43.704324	-79.388790	3
28	M6S	West Toronto	Runnymede,Swansea	43.651571	-79.484450	3
30	M5T	Downtown Toronto	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049	3
31	M4V	Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	43.686412	-79.400049	3
35	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675	3
37	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	3
38	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	3
39	NaN	New Jersey	Mitsuwa	40.815989	-73.979970	3

Discussion

From above results, we can recommend Mitsuwa's stakeholders that there are 13 neighborhoods in Toronto have similar surrounding environment with Mitsuwa New Jearsey. If they are considering opening their first market in Toronto, those areas shall be on their top list.

Conclusion

In this project, we study the similarities of Toronto neighborhoods and Mitsuwa New Jersey surrounding area in order to find right place to open Mitsuwa market in Toronto. We pull Foursquare data to explore nearby venues for each neighborhood and perform KNN algorithm to cluster neighborhood and Mitsuwa New Jersey into 12 clusters. Finally, we locate the neighborhoods which contains same cluster label with Mitsuwa and provide location recommendations for first Mitsuwa in Toronto.