Battle of the Neighbourhoods

Toronto

Introduction

- **We are a Consultancy in Toronto which main activity is the Strategic support for new business into the City.**
- **We are supporting a new Spanish customer to select the best tye of business and the Best Venue to place it.**
- **To be able to give a proper advise to our customer, we will do an analysis of of the density of businesses in Toronto depending on the venue.**

Data Used

- **Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) to get Postal code, Borough and Neighbourhood in Toronto.**
- **Geospatial data from Toronto(http://cocl.us/Geospatial_data) to get the geographical coordinated of each postal code.**
- **Foursquare API to obtain more information about venues.**
- **Random user data, with a random number of preferences to check, how our system works.**

Methodology

1. **We will retrieve Geospatial data from for Toronto Neighbourhood from the sources.**
2. **Using Foursquare API we retrieve the data about “Popular Spots” venues**
3. **We cluster unsing Kmeans algorithm acc. Location and number of venues.**
4. **We detect areas with low density of Popular venues where giving a hint about where to place a nes business.**

1.Retrieving Geospatial Data

From **Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) to get Postal code, Borough and Neighbourhood in Toronto.**

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

1.Retrieving Geospatial Data

From Geospatial data from Toronto(http://cocl.us/Geospatial_data) to get the geographical coordinated of each postal code.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

]:

	PostCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

1.Retrieving Geospatial Data

From Geospatial data from Toronto(http://cocl.us/Geospatial_data) to get the geographical coordinated of each postal code.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

]:

	PostCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

Retrieving venues data
From APIS FOURSQUARE + postprocessing

```
'https://api.foursquare.com/v2/venues/explore?&client_id=GEHGRDNDQ0CH0GJG3DJFG4B4QLLALBZVIVHR3MXRLGYGGGPG&client_secret=5BZJ0LMQSCST0P3YHL5NPQ4P0SYC11T5TQEXS0UNPAKLZK2S&v=20200728&ll=43.806686299999996,-79.19435340000001&radius=5000&limit=1000'
```

```
114]: print(toronto_venues.shape)
      toronto_venues.head()
```

```
(2152, 8)
```

ut[114]:

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Summary
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant	This spot is popular
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	RIGHT WAY TO GOLF	43.785177	-79.161108	Golf Course	This spot is popular
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar	This spot is popular
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank	This spot is popular
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store	This spot is popular

Adding number of venues per Neighbourhood

venues per neighbourghood

In [185]:
```
toronto_venues_grouped = toronto_venues.groupby(['Neighbourhood', 'Neighborhood Latitude', 'Neighborhood Longitude'] ).count().reset_index()
toronto_venues_grouped.head()
```

Out[185]:

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Summary
0	Agincourt	43.794200	-79.262029	4	4	4	4	4
1	Alderwood, Long Branch	43.602414	-79.543484	9	9	9	9	9
2	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	22	22	22	22	22
3	Bayview Village	43.786947	-79.385975	4	4	4	4	4
4	Bedford Park, Lawrence Manor East	43.733283	-79.419750	27	27	27	27	27

In [186]:
```
toronto_venues_grouped.shape
```

Out[186]: (100, 8)

100 Elements

Sorting to appraise optimal number of clusters

Out[140]:

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Summary
62	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	1	1	1	1	1
43	Humberlea, Emery	43.724766	-79.532242	1	1	1	1	1
53	Malvern, Rouge	43.806686	-79.194353	1	1	1	1	1
92	Weston	43.706876	-79.518188	1	1	1	1	1
90	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	1	1	1	1	1
74	Scarborough Village	43.744734	-79.239476	1	1	1	1	1
54	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	2	2	2	2	2
17	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476	2	2	2	2	2
70	Roselawn	43.711695	-79.416936	2	2	2	2	2
99	York Mills West	43.752758	-79.400049	2	2	2	2	2
73	Runnymede, The Junction North	43.673185	-79.487262	2	2	2	2	2
85	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	2	2	2	2	2

Aprox 6 elements

100/6 = aprox 20 clusters

Clustering (Kmeans) and labeling

Comentario en la lógica del Kmeans

[156]:
```
# import k-means from clustering stage
from sklearn.cluster import KMeans

# set number of clusters
kclusters = 20

toronto_venues_grouped_clustering = toronto_venues_grouped.drop(['Neighbourhood','Venue Latitude','Venue Longitude','Venue Category','Venue Summary'], axis = 1)

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

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

Out[156]:
```
array([ 0,  7, 16,  0,  6, 13,  0, 16, 18, 18], dtype=int32)
```

Clustering and labeling

Cluster 11 contains the least number of venues (1 or 2)

Insertamos indice Kmeans

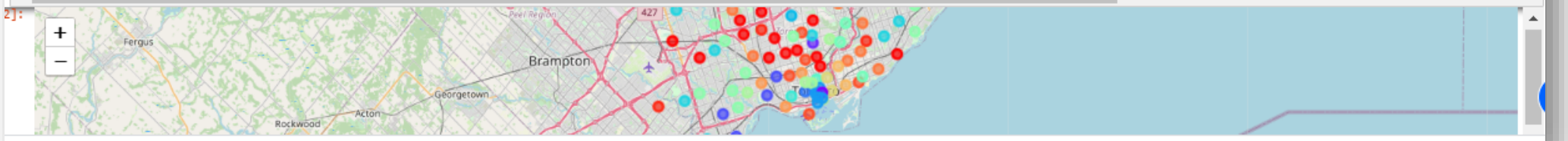
```
In [163]: toronto_venues_grouped_merged = toronto_venues_grouped

In [ ]: toronto_venues_grouped_merged.insert(0, 'Cluster Labels', kmeans.labels_)

In [179]: toronto_venues_grouped_merged_sorted = toronto_venues_grouped_merged.sort_values('Venue')

In [180]: toronto_venues_grouped_merged_sorted.head(100)
```

Out[180]:

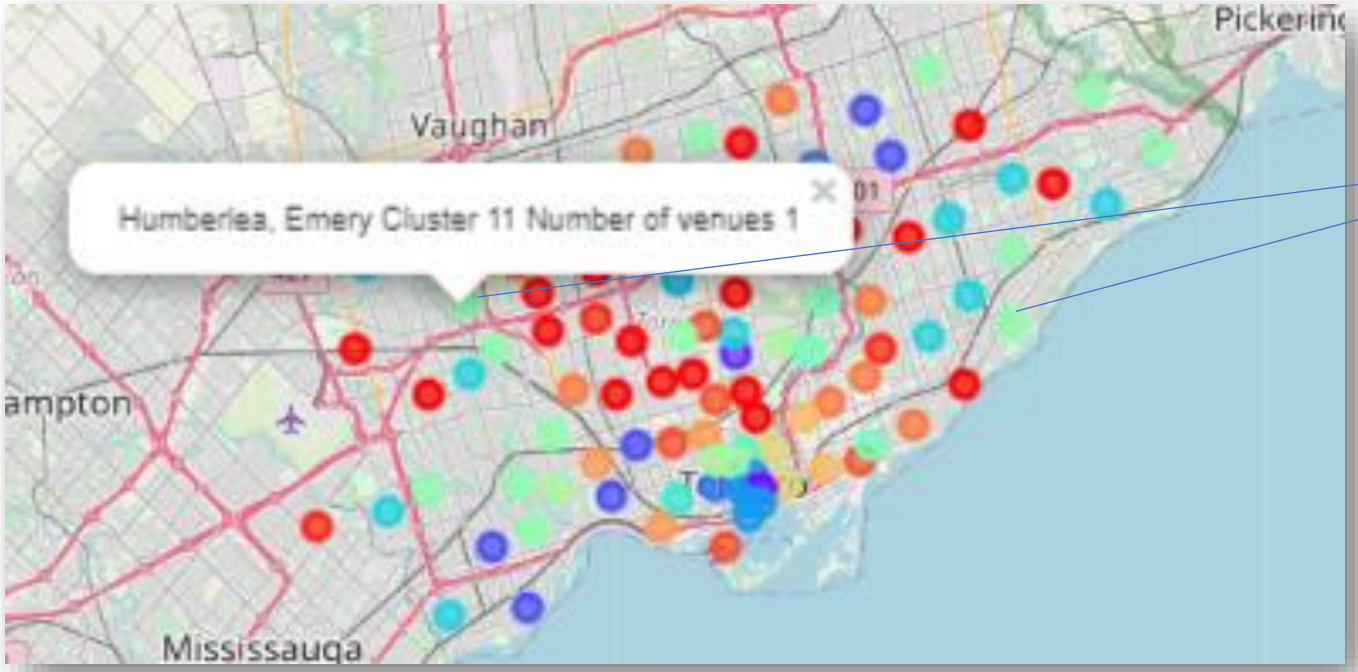
	Cluster Labels	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Summary
62	11	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	1	1	1	1	1
43	11	Humberlea, Emery	43.724766	-79.532242	1	1	1	1	1
53	11	Malvern, Rouge	43.806686	-79.194353	1	1	1	1	1
92	11	Weston	43.706876	-79.518188	1	1	1	1	1
90	11	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	1	1	1	1	1
74	11	Scarborough Village	43.744734	-79.239476	1	1	1	1	1
54	11	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	2	2	2	2	2
17	11	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476	2	2	2	2	2
70	11	Roselawn	43.711695	-79.416936	2	2	2	2	2
99	11	York Mills West	43.752758	-79.400049	2	2	2	2	2
73	11	Runnymede, The Junction North	43.673185	-79.487262	2	2	2	2	2
85	11	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	2	2	2	2	2

Mapping

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

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

# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster, venue in zip(toronto_venues_grouped_merged['Neighborhood Latitude'], toronto_venues_grouped_merged['Neighborhood Longitude'], toronto_venues_grouped_merged['Neighbourhood'], toronto_ve
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster) + '  Number of venues  ' + str(venue), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```



Conclusions
Let’s explore the light green dotted Neighbourhoods(Cluster 11) for the new business



Cluster 11

References

Jupiter notebook in Watson Studio (you see the map here)

https://eu-de.dataplatform.cloud.ibm.com/analytics/notebooks/v2/ffd5cb12-47f0-46d4-bd3b-e57bb4e91698/view?access_token=8a2fe9d14ee052fb2c1c0df5c5fb2b547d0c73fcffbf7fff8019612fc7c7c39d

GIT hub repository

https://github.com/pertres/capstone