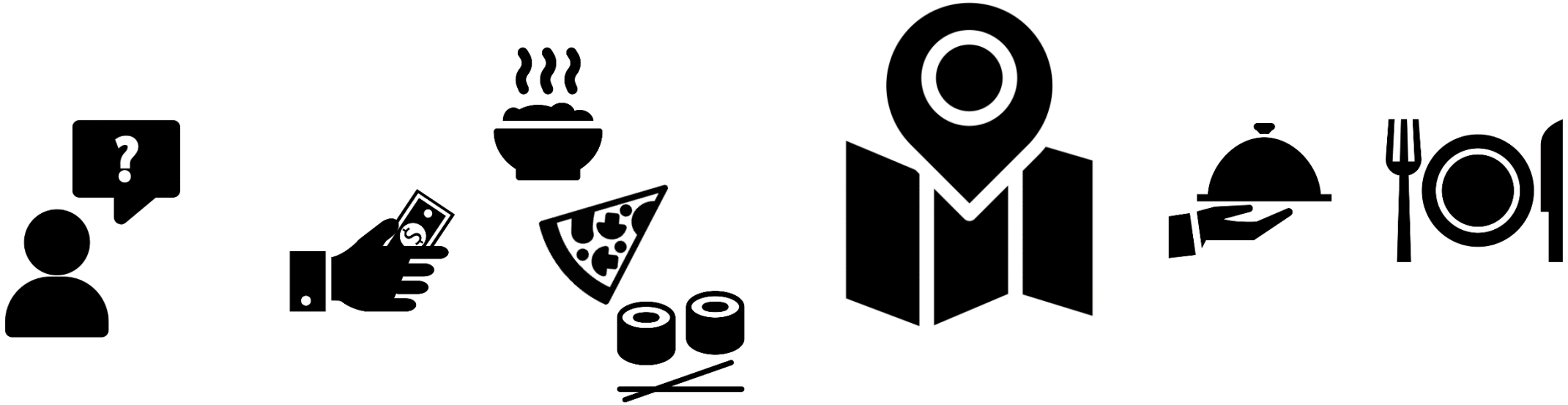
Clustering Toronto boroughs and neighborhoods by restaurants' cuisines

Coursera capstone project by Eve Belyaeva

The goal of the project

- The goal of the project is to combine information about restaurant count in neighborhoods containing different cuisines
- It is important if the customers wish to find a place with kitchen they prefer and also to have a number of these places



Gather dataset
Toronto boroughs and neighborhoods

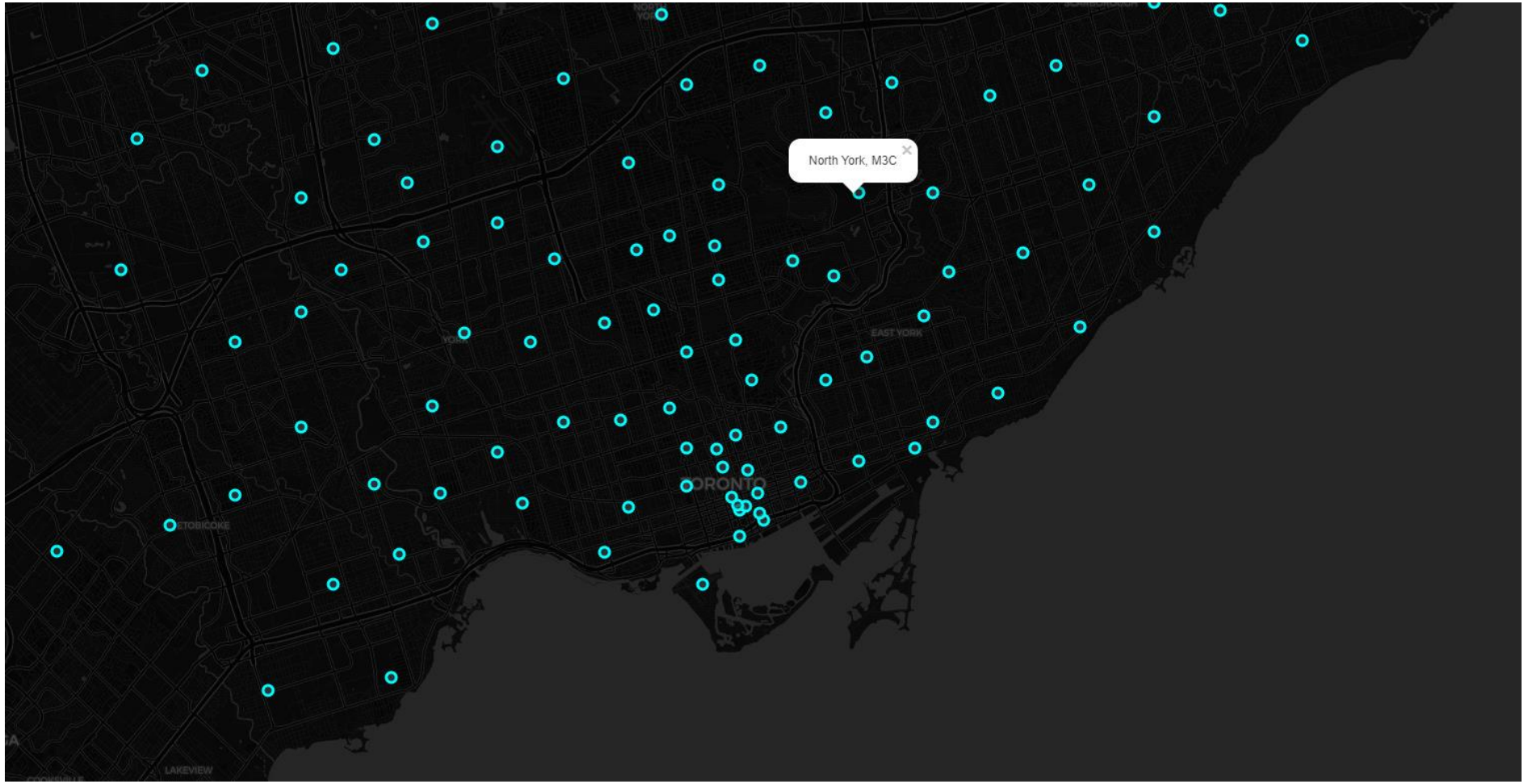
```
df = pd.read_html('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M')[0]
df.drop(df[df['Borough']=='Not assigned'].index,inplace=True)
df=df.groupby(['Postcode','Borough'])['Neighbourhood'].apply(','.join).reset_index()
df['Neighbourhood']=df['Neighbourhood'].replace('Not assigned',df['Borough'])
url='http://cocl.us/Geospatial_data'
gd=pd.read_csv(url)
gd.rename(columns={'Postal Code':'Postcode'}, inplace=True)
df.set_index('Postcode', inplace=True)
gd.set_index('Postcode', inplace=True)
mergedDf = df.merge(gd, left_index=True, right_index=True)
mergedDf=mergedDf.reset_index()
mergedDf.head()
```

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,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

Code snippet for the map

```
latitude=43.653226
longitude=-79.383184

#'OpenStreetMap', 'cartodbpositron', 'cartodbdark_matter'

toronto_all=mergedDf.copy()
print(toronto_all.shape)

map_dots = folium.Map(location=[latitude, longitude], zoom_start=12,tiles='cartodbdark_matter')
# add markers to map
for lat, lng, borough, neighborhood, fsa in zip(toronto_all['Latitude'],
                                                toronto_all['Longitude'], toronto_all['Borough'],
                                                toronto_all['Neighbourhood'], toronto_all['Postcode']):
    label = '{}, {}'.format(borough, fsa)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='#01ffff',
        fill=True,
        fill_color='#3d3c42',
        fill_opacity=0.7,
        line_opacity=0.2,
        parse_html=False).add_to(map_dots)

map_dots
```

Toronto boroughs on map



Gathering restaurants data from FourSquare and preparing the data for choropleth map

	FSA	Borough	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka
3	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Terry's Restaurant & Bar	43.774780	-79.241043	Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	terry's restaurant	43.774969	-79.240872	Italian
5	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	Thai One On	43.774468	-79.241268	Thai
6	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304	Karaikudi Chettinad South Indian Restaurant	43.756042	-79.276276	Indian

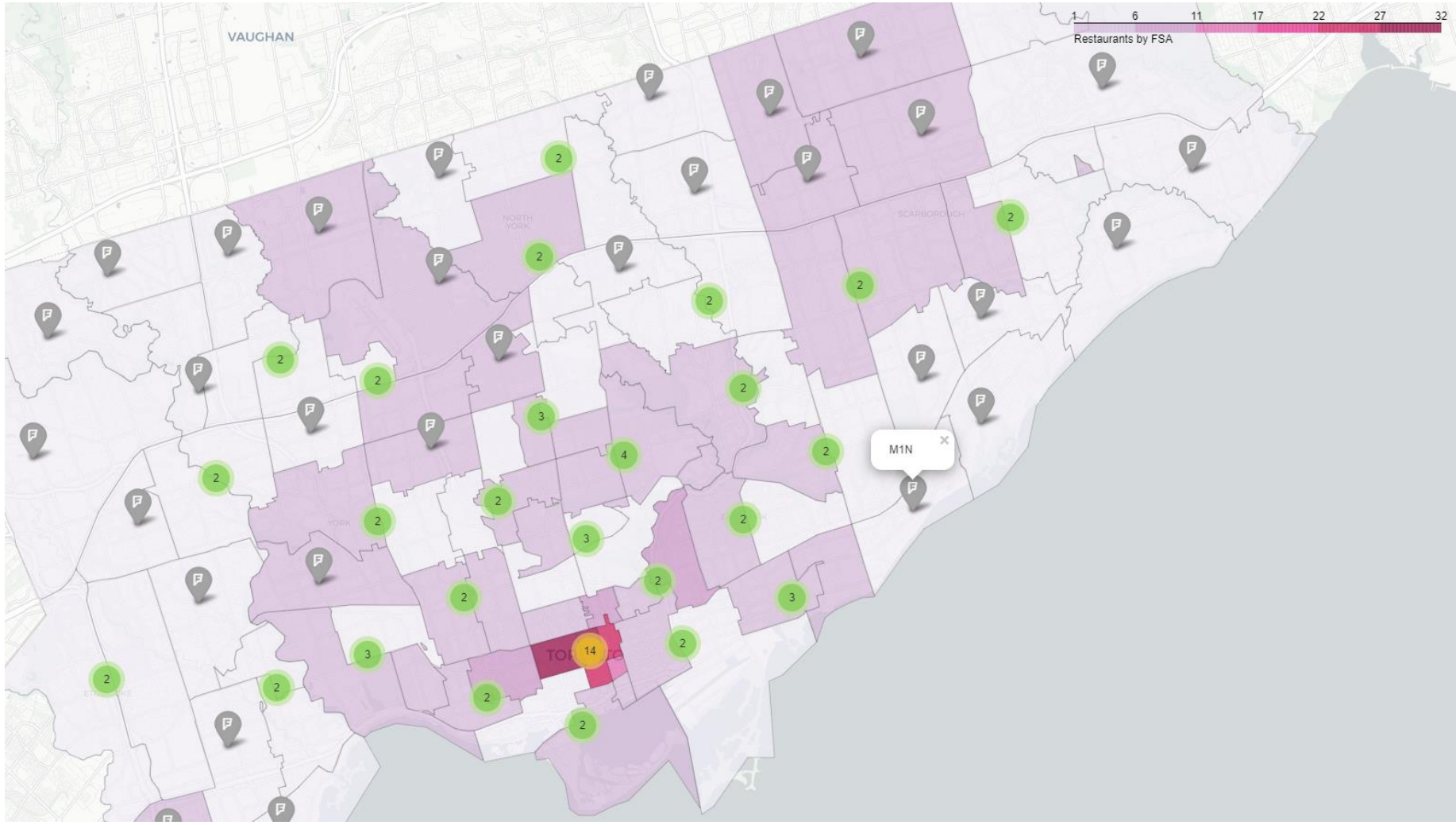
	FSA	Count
0	M1H	3
1	M1P	2
2	M1R	1
3	M1S	2
4	M1T	3

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

Code snippet for the map

```
latitude=43.653226
longitude=-79.383184

#'OpenStreetMap', 'cartodbpositron', 'cartodbdark_matter'

toronto_map = folium.Map(location=[latitude, longitude], zoom_start=10,tiles='cartodbpositron')

toronto_map.choropleth(geo_data=t_js,
    data = bn_grouped,
    columns=['FSA','Count'],
    key_on='feature.properties.CFSAUID',
    fill_color='PuRd',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Restaurants by FSA')


# instantiate a mark cluster object for the incidents in the dataframe
rests = plugins.MarkerCluster().add_to(toronto_map)

# loop through the dataframe and add each data point to the mark cluster
for lat, lng, label, in zip(fsa_geo['Latitude'], fsa_geo['Longitude'], fsa_geo['Postcode']):
    folium.Marker(
        location=[lat, lng],
        icon=folium.Icon(color='lightgray', icon_color='white',icon='foursquare',prefix='fa'),
        popup=label,
    ).add_to(rests)

toronto_map
```

Choropleth map based on restaurants count in Toronto boroughs



Preparing the data to search for different cuisines distribution in boroughs

```
topvenues=bn_restaurants.copy()
topvenues=topvenues.groupby(['Venue Category']).count().reset_index()

topvenues=topvenues.loc[:,['Venue Category','FSA']]
topvenues.rename(columns={'FSA':'Count'},inplace=True)

topvenues.sort_values(by='Count', ascending=True, inplace=True)
topvenues=topvenues[topvenues['Venue Category'].isin(cuisines)]
topvenues.set_index('Venue Category', inplace=True)
cu_top10 = topvenues['Count'].tail(10)
cu_top10.head()
```

```
Venue Category
New American    10
Vietnamese      10
Thai            10
Caribbean       12
Indian          19
Name: Count, dtype: int64
```

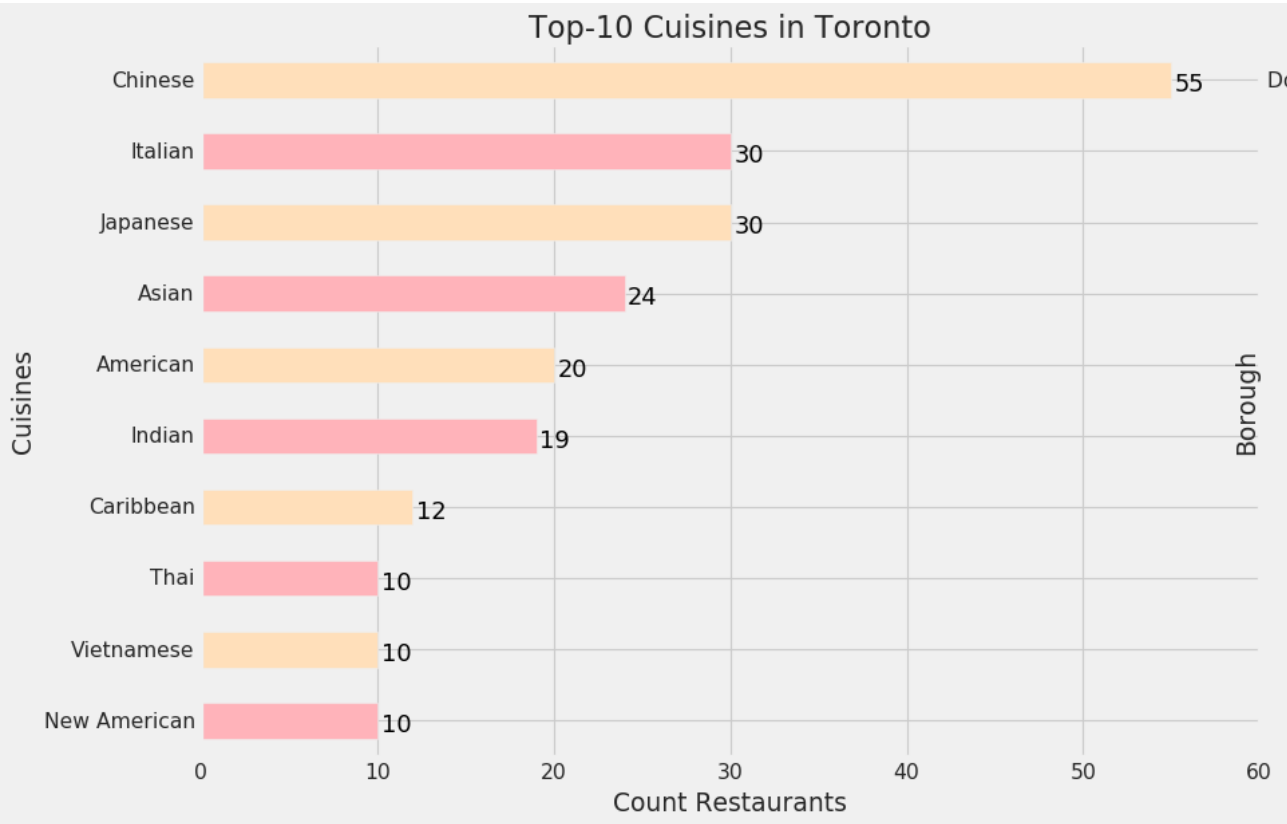
```
topboroughs=bn_restaurants.copy()
#delete Restaurants without Cuisine
topboroughs=topboroughs[topboroughs['Venue Category'].isin(cuisines)]
topboroughs=topboroughs.groupby(['Borough']).count().reset_index()
topboroughs=topboroughs.loc[:,['Borough','FSA']]
topboroughs.rename(columns={'FSA':'Count'},inplace=True)
topboroughs.sort_values(by='Count', ascending=True, inplace=True)
topboroughs.set_index('Borough', inplace=True)
topboroughs = topboroughs['Count'].tail(10)
topboroughs.head()
```

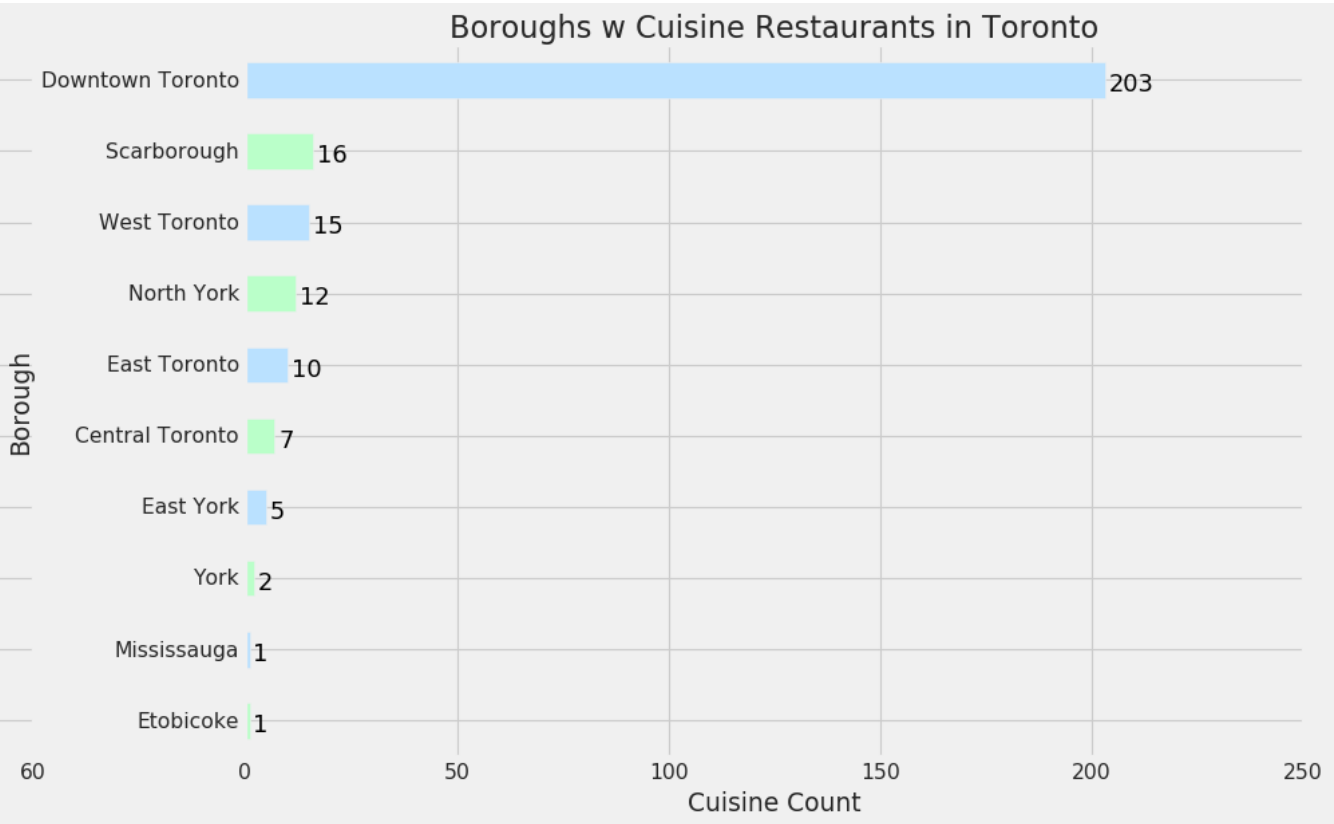
```
Borough
Etobicoke          1
Mississauga        1
York               2
East York          5
Central Toronto    7
Name: Count, dtype: int64
```

Code snippet for the barh plots

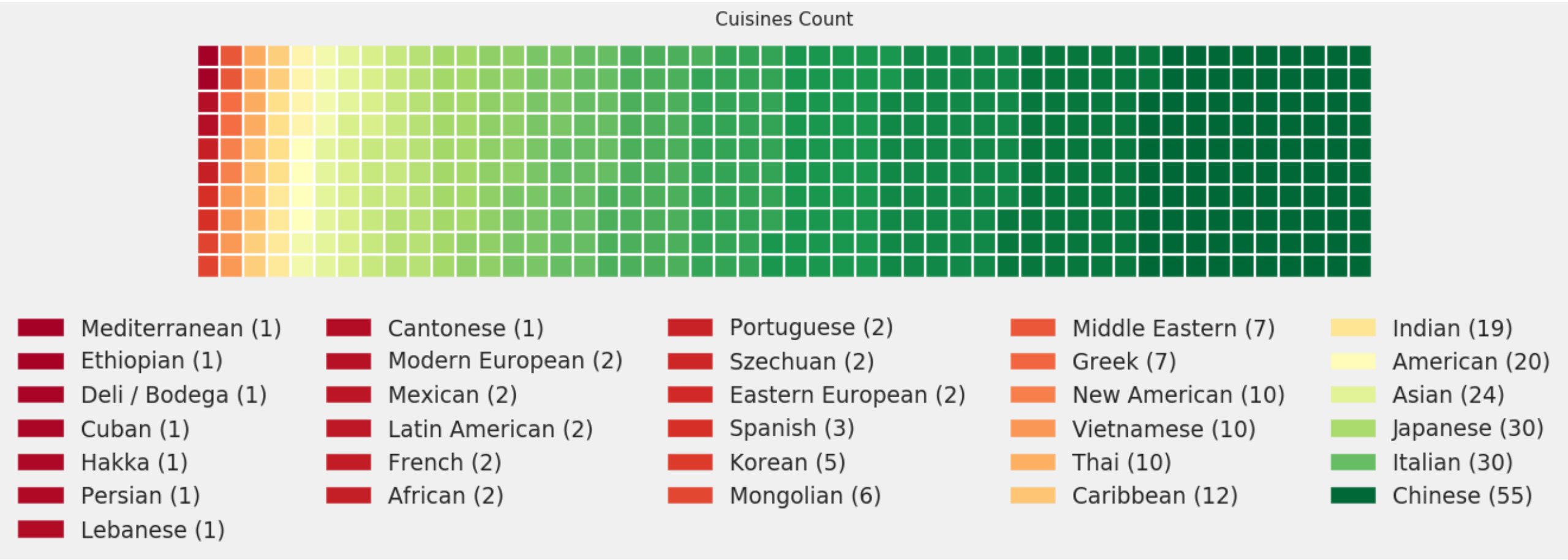
```
plt.style.use('fivethirtyeight')
colors1=['#ffb3ba','#ffdfba']
colors2=['#baffc9','#bae1ff']
colors=['#ffb3ba','#ffdfba','#ffffba','#baffc9','#bae1ff']

fig = plt.figure() # create figure

ax0 = fig.add_subplot(1, 2, 1) # add subplot 1 (1 row, 2 columns, first plot)
ax1 = fig.add_subplot(1, 2, 2) # add subplot 2 (1 row, 2 columns, second plot). See tip below**

# Subplot 1: Box plot
cu_top10.plot(kind='barh', figsize=(25,8), color=colors1, ax=ax0) # add to subplot 1
for a in ax0.patches:
    #print(a,' ',str(a.get_y()))
    ax0.annotate(str(a.get_width()),(a.get_width()+0.2,a.get_y()+0.1), color='black',fontsize=16)
ax0.set_xlabel('Count Restaurants')
ax0.set_ylabel('Cuisines')
ax0.set_title('Top-10 Cuisines in Toronto')

# Subplot 2: Line plot
topboroughs.plot(kind='barh', figsize=(25,8), color=colors2, ax=ax1) # add to subplot 2
for a in ax1.patches:
    #print(a,' ',str(a.get_y()))
    ax1.annotate(str(a.get_width()),(a.get_width()+0.9,a.get_y()+0.1), color='black',fontsize=16)
ax1.set_xlabel('Cuisine Count')
ax1.set_ylabel('Borough')
ax1.set_title('Boroughs w Cuisine Restaurants in Toronto')
#ax1.legend([])
#ax1.get_legend().remove()

plt.show()
```

Visualization of the top cuisine data





Visualization of the all cuisine data



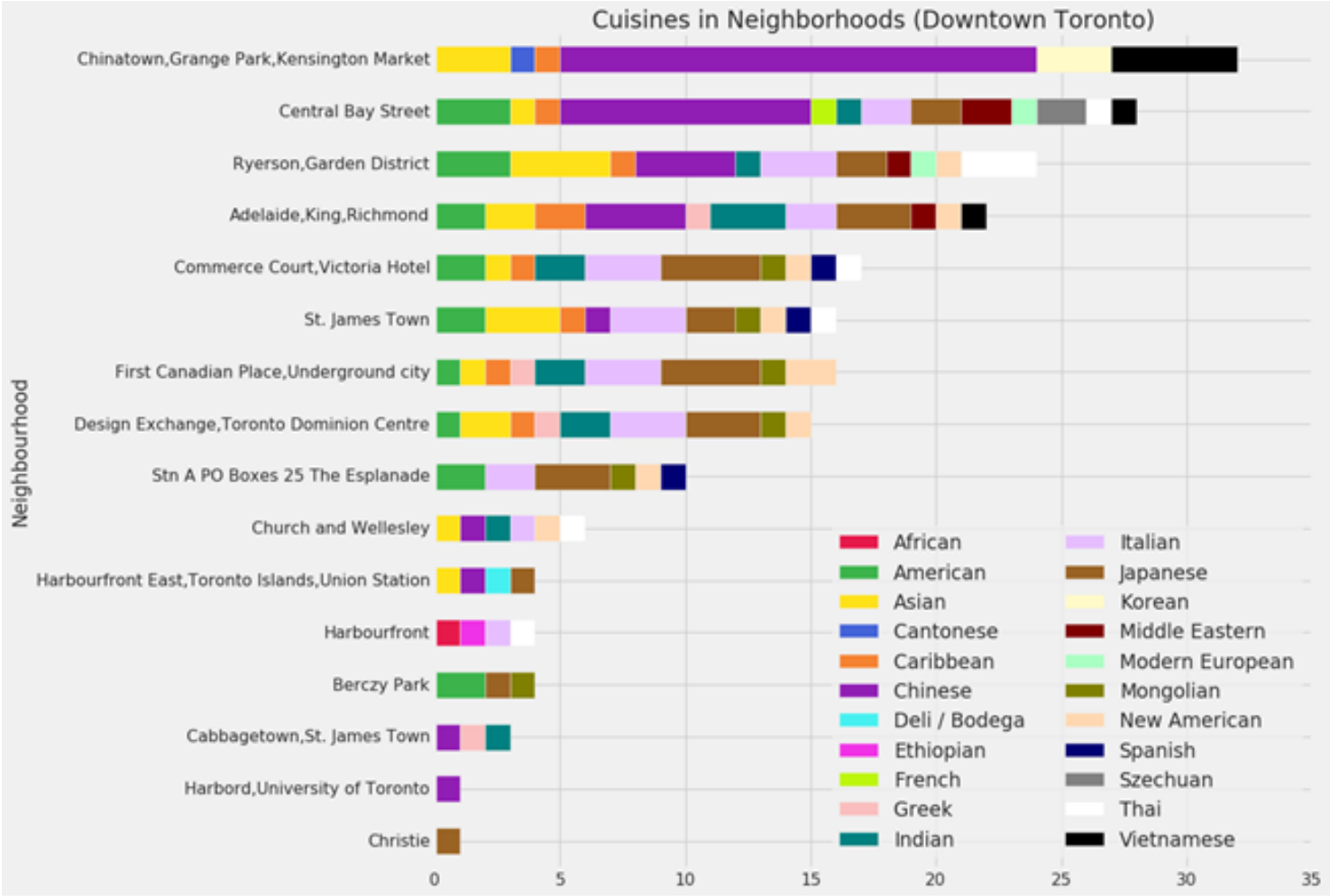
Cuisines in Downtown Toronto borough: neighborhoods

	Neighbourhood	African	American	Asian	Cantonese	Caribbean	Chinese	Deli / Bodega	Ethiopian	French	Greek	Indian	Italian	Japanese	Korean	Middle Eastern	Modern European	Mongolian	New American	Spanish	Szechuan	Thai	Vietnamese
0	Cabbagetown,St. James Town	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Cabbagetown,St. James Town	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	Cabbagetown,St. James Town	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	Church and Wellesley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
4	Church and Wellesley	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Venue Category	African	American	Asian	Cantonese	Caribbean	Chinese	Deli / Bodega	Ethiopian	French	Greek	Indian	Italian	Japanese	Korean	Middle Eastern	Modern European	Mongolian	New American	Spanish	Szechuan	Thai	Vietnamese
Neighbourhood																						
Adelaide,King,Richmond	0	2	2	0	2	4	0	0	0	1	3	2	3	0	1	0	0	1	0	0	0	1
Berczy Park	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Cabbagetown,St. James Town	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Central Bay Street	0	3	1	0	1	10	0	0	1	0	1	2	2	0	2	1	0	0	0	2	1	1
Chinatown,Grange Park,Kensington Market	0	0	3	1	1	19	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	5

	Chinese	Japanese	Italian	Asian	American	Indian	Caribbean	New American	Thai	Vietnamese	Mongolian	Greek	Middle Eastern	Korean	Spanish	Modern European	Szechuan	French	Ethiopian	Deli / Bodega	Cantonese	African
Neighbourhood																						
Commerce Court,Victoria Hotel	0	4	3	1	2	2	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0
Adelaide,King,Richmond	4	3	2	2	2	3	2	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0
Ryerson,Garden District	4	2	3	4	3	1	1	1	3	0	0	0	1	0	0	1	0	0	0	0	0	0
Central Bay Street	10	2	2	1	3	1	1	0	1	1	0	0	2	0	0	1	2	1	0	0	0	0
Chinatown,Grange Park,Kensington Market	19	0	0	3	0	0	1	0	0	5	0	0	0	3	0	0	0	0	0	0	1	0

Cuisines in Downtown Toronto in barh plot



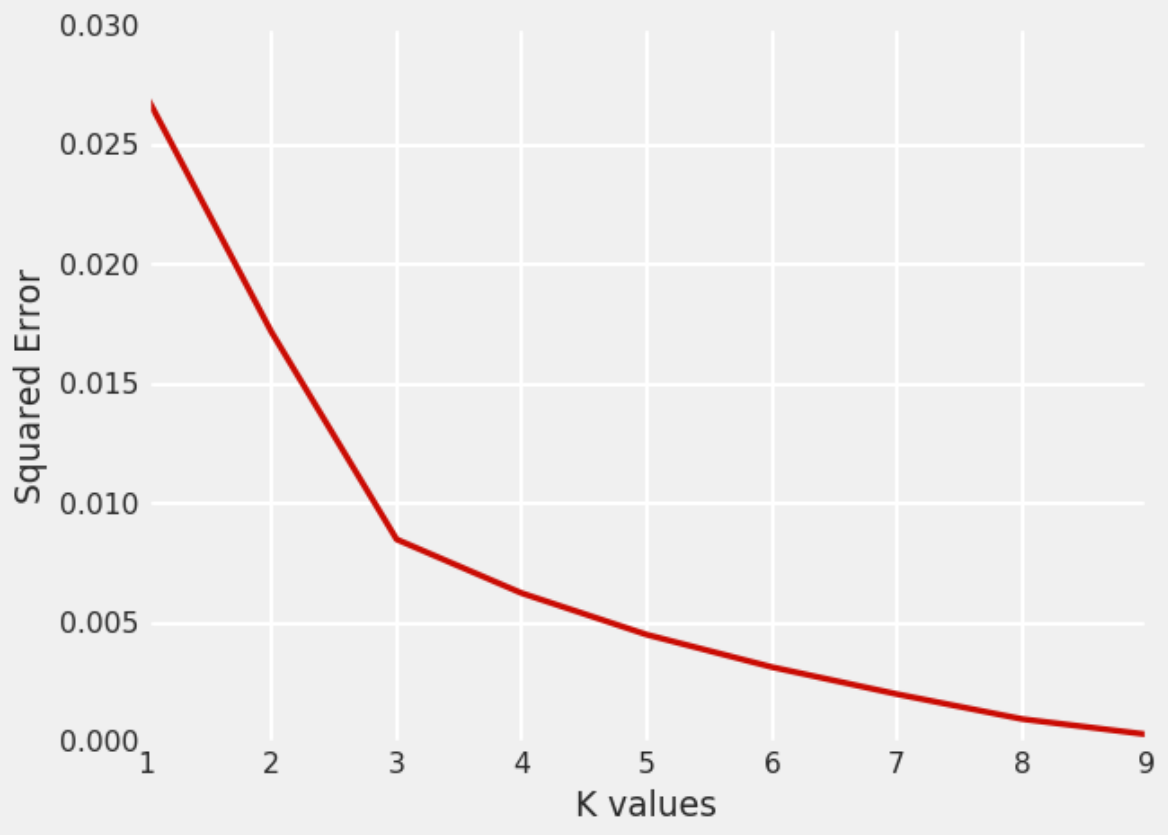
Preparing data for clusterization

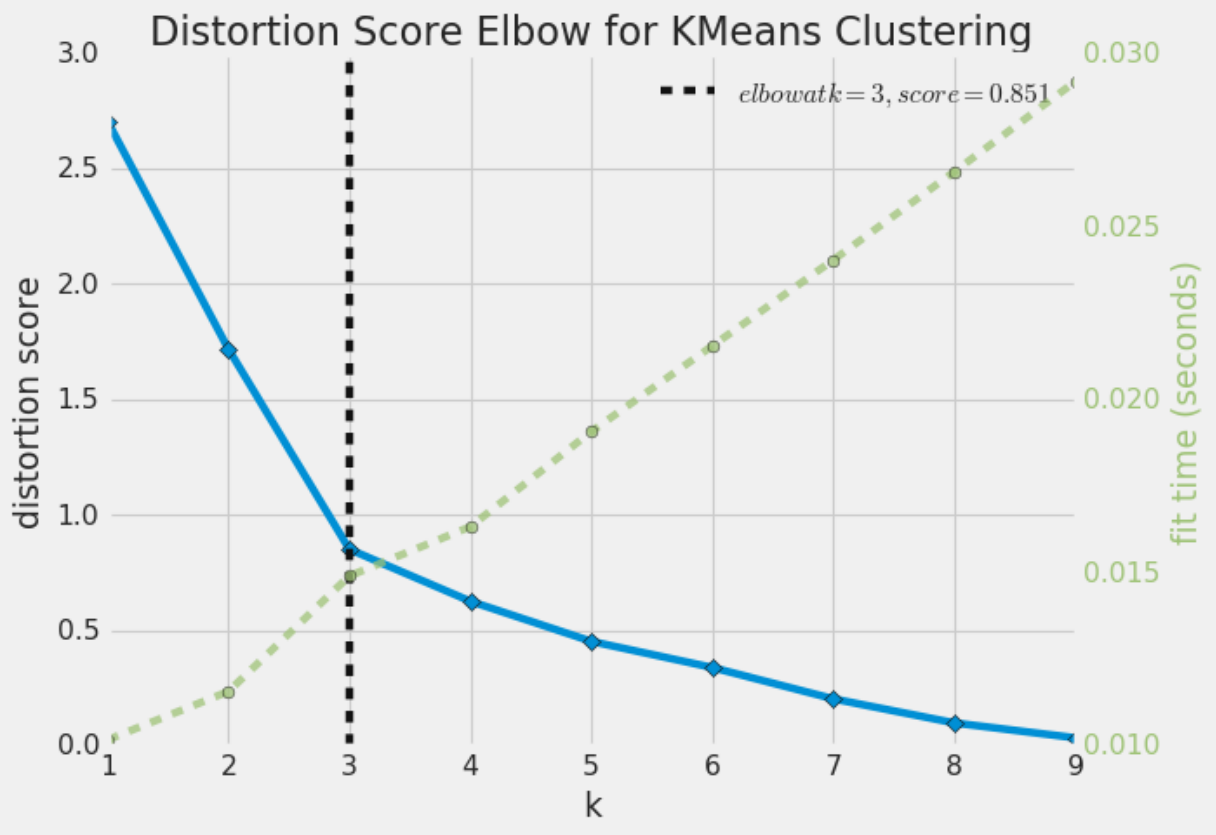
	Borough	African	American	Asian	Cantonese	Caribbean	Chinese	Cuban	Deli / Bodega	Eastern European	Ethiopian	French	Greek	Hakka	Indian	Italian	Japanese	Korean	Latin American	Lebanese	Mediterranean	Mexican	Middle Eastern	Modern European	Mongolian	New American	Persian	Port
0	Scarborough	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Scarborough	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
3	Scarborough	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Scarborough	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	

	Borough	African	American	Asian	Cantonese	Caribbean	Chinese	Cuban	Deli / Bodega	Eastern European	Ethiopian	French	Greek	Hakka	Indian	Italian	Japanese	Korean	Latin American	Lebanese	Mediterranean	Mexican	Middle Eastern	Modern European	Mongolian	Am
0	Central Toronto	0.000000	0.000000	0.142857	0.000000	0.000000	0.142857	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.0000	0.142857	0.285714	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
1	Downtown Toronto	0.004926	0.088670	0.093596	0.004926	0.044335	0.206897	0.000000	0.004926	0.000000	0.004926	0.004926	0.019704	0.0000	0.064039	0.113300	0.128079	0.014778	0.000000	0.0	0.0	0.000000	0.019704	0.009852	0.029557	0.0
2	East Toronto	0.000000	0.000000	0.100000	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.300000	0.0000	0.100000	0.100000	0.000000	0.000000	0.000000	0.1	0.1	0.000000	0.000000	0.000000	0.000000	0.0
3	East York	0.200000	0.200000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.200000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.200000	0.000000	0.000000	0.0
4	Etobicoke	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	1.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
5	Mississauga	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0
6	North York	0.000000	0.000000	0.166667	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333	0.000000	0.000000	0.000000	0.0000	0.083333	0.166667	0.250000	0.000000	0.000000	0.0	0.0	0.000000	0.083333	0.000000	0.000000	0.0
7	Scarborough	0.000000	0.000000	0.000000	0.000000	0.000000	0.562500	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0625	0.125000	0.062500	0.000000	0.062500	0.000000	0.0	0.0	0.000000	0.062500	0.000000	0.000000	0.0
8	West Toronto	0.000000	0.066667																							0.0
9	York	0.000000	0.000000																							0.0

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central Toronto	Italian	Asian	Chinese	Eastern European	Indian
1	Downtown Toronto	Chinese	Japanese	Italian	Asian	American
2	East Toronto	Greek	Indian	Mediterranean	Italian	Asian
3	East York	African	American	Caribbean	Middle Eastern	Indian
4	Etobicoke	Korean	Vietnamese	Italian	American	Asian

Choosing k-value based on squared error value



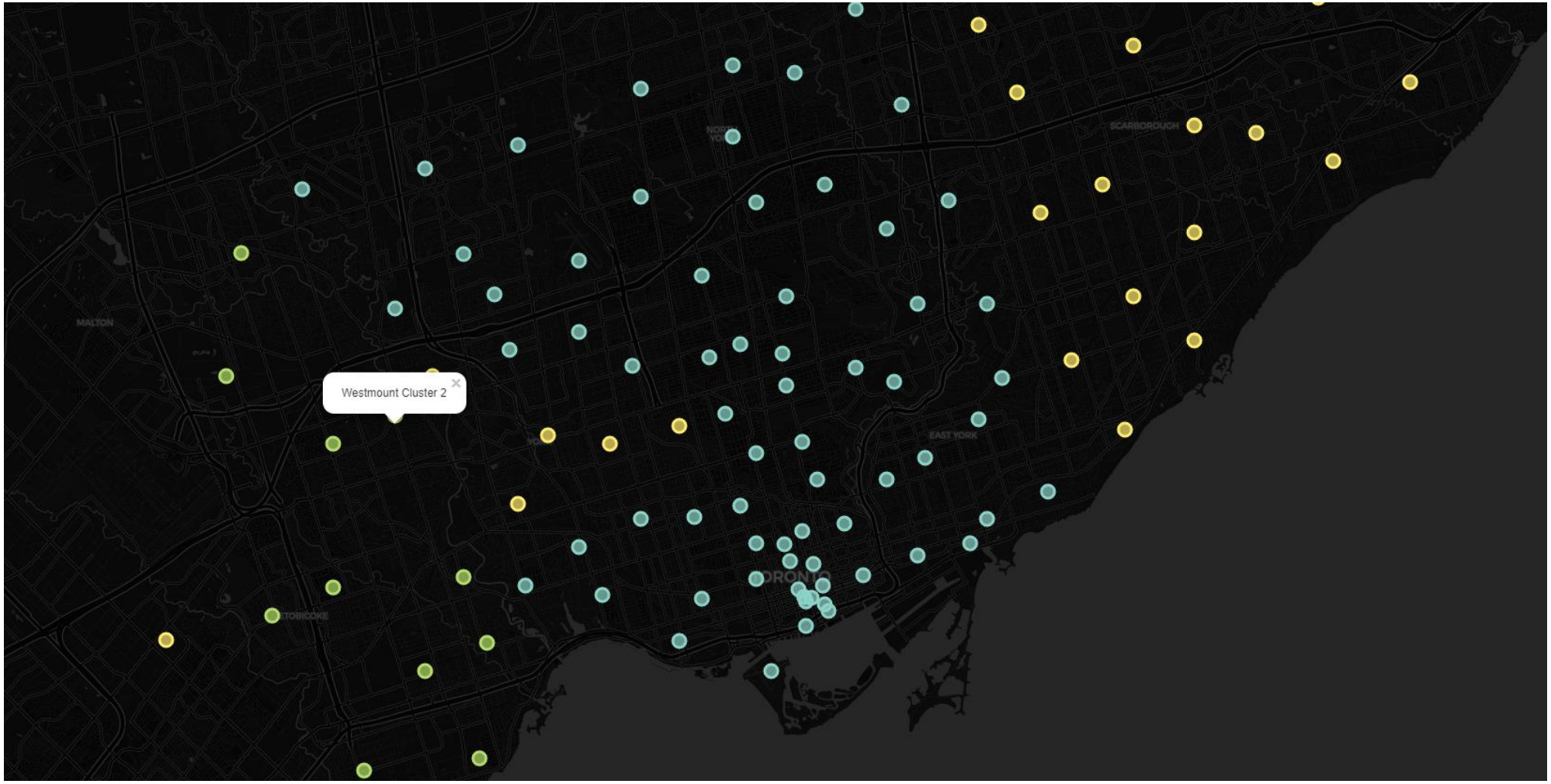


Showing the clusters on a map

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	0	Chinese	Indian	Italian	Thai	Korean
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0	Chinese	Indian	Italian	Thai	Korean
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	0	Chinese	Indian	Italian	Thai	Korean
3	M1G	Scarborough								Thai	Korean
4	M1H	Scarborough								Thai	Korean

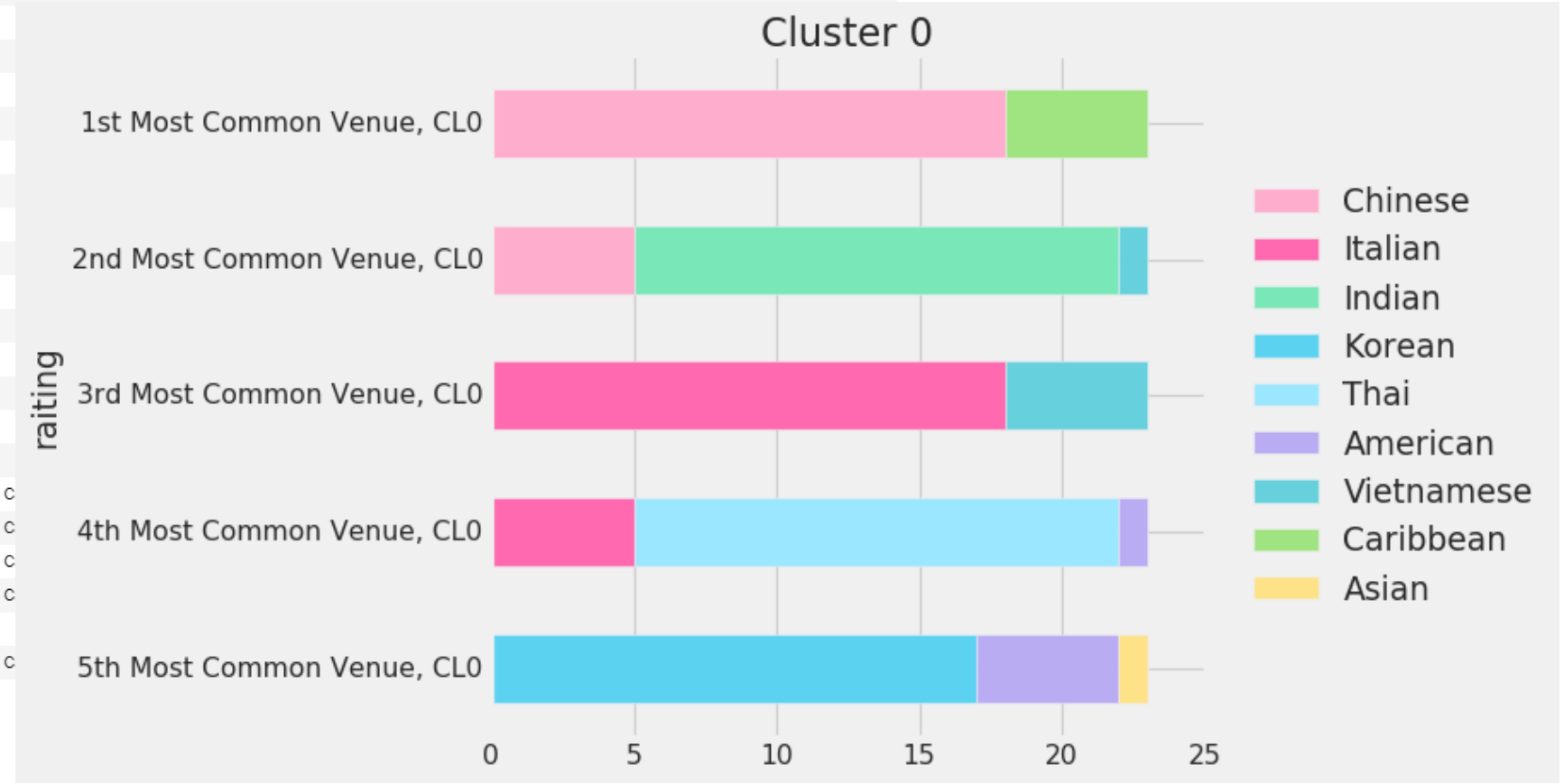
```
#'OpenStreetMap', 'cartodbpositron', 'cartodbdark_matter'

# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11,tiles='cartodbdark_matter')
# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.Set3(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 in zip(topneighs_merged['Latitude'], topneighs_merged['Longitude'],
                                  topneighs_merged['Neighbourhood'], topneighs_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=7,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Three Toronto boroughs clusters



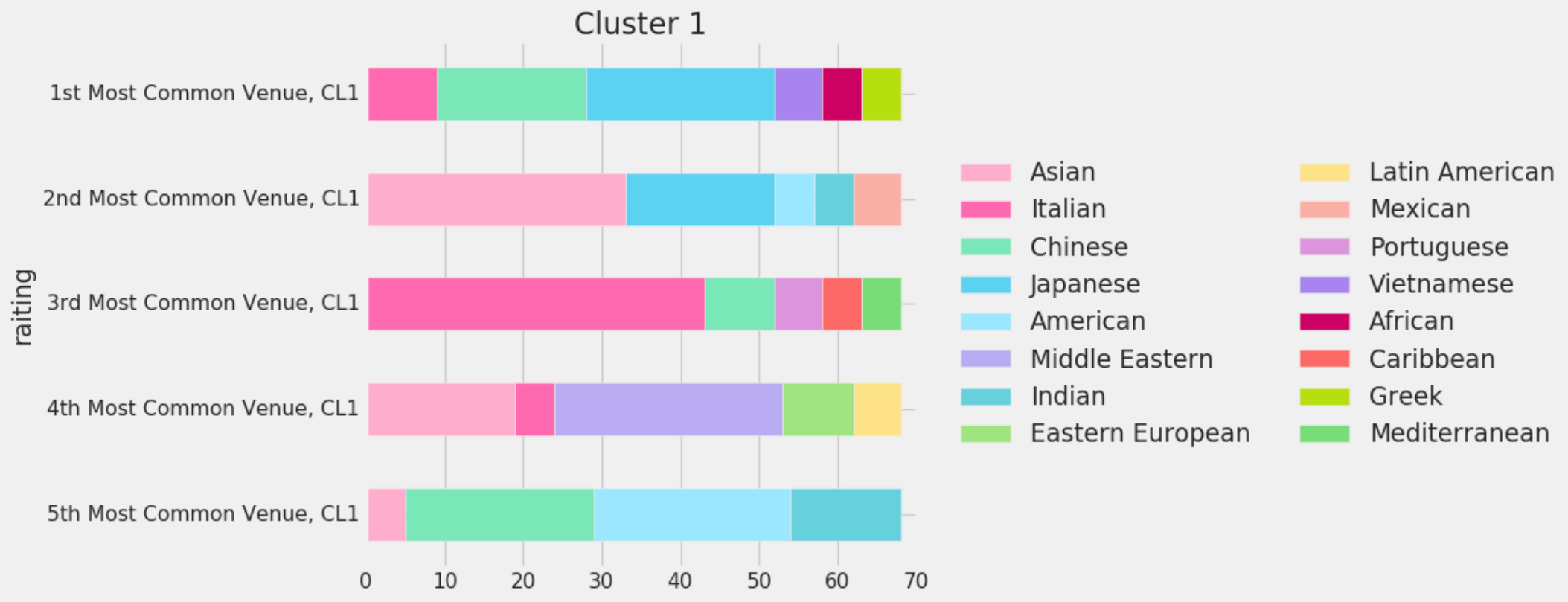
Cluster0 cuisines

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Rouge,Malvern	0	Chinese	Indian	Italian	Thai	Korean
1	Highland Creek,Rouge Hill,Port Union	0	Chinese	Indian	Italian	Thai	Korean
2	Guildwood,Morningside,West Hill	0	Chinese	Indian	Italian	Thai	Korean
3	Woburn	0					
4	Cedarbrae	0					
5	Scarborough Village	0					
6	East Birchmount Park,Ionview,Kennedy Park	0					
7	Clairlea,Golden Mile,Oakridge	0					
8	Cliffcrest,Cliffside,Scarborough Village West	0					
9	Birch Cliff,Cliffside West	0					
10	Dorset Park,Scarborough Town Centre,Wexford He...	0					
11	Maryvale,Wexford	0					
12	Agincourt	0					
13	Clarks Corners,Sullivan,Tam O'Shanter	0					
14	Agincourt North,L'Amoreaux East,Milliken,Steel...	0					
15	L'Amoreaux West	0					
16	Upper Rouge	0					
73	Humewood-Cedarvale	0	C				
74	Caledonia-Fairbanks	0	C				
80	Del Ray,Keelesdale,Mount Dennis,Silverthorn	0	C				
81	The Junction North,Runnymede	0	C				
86	Canada Post Gateway Processing Centre	0					
98	Weston	0	C				



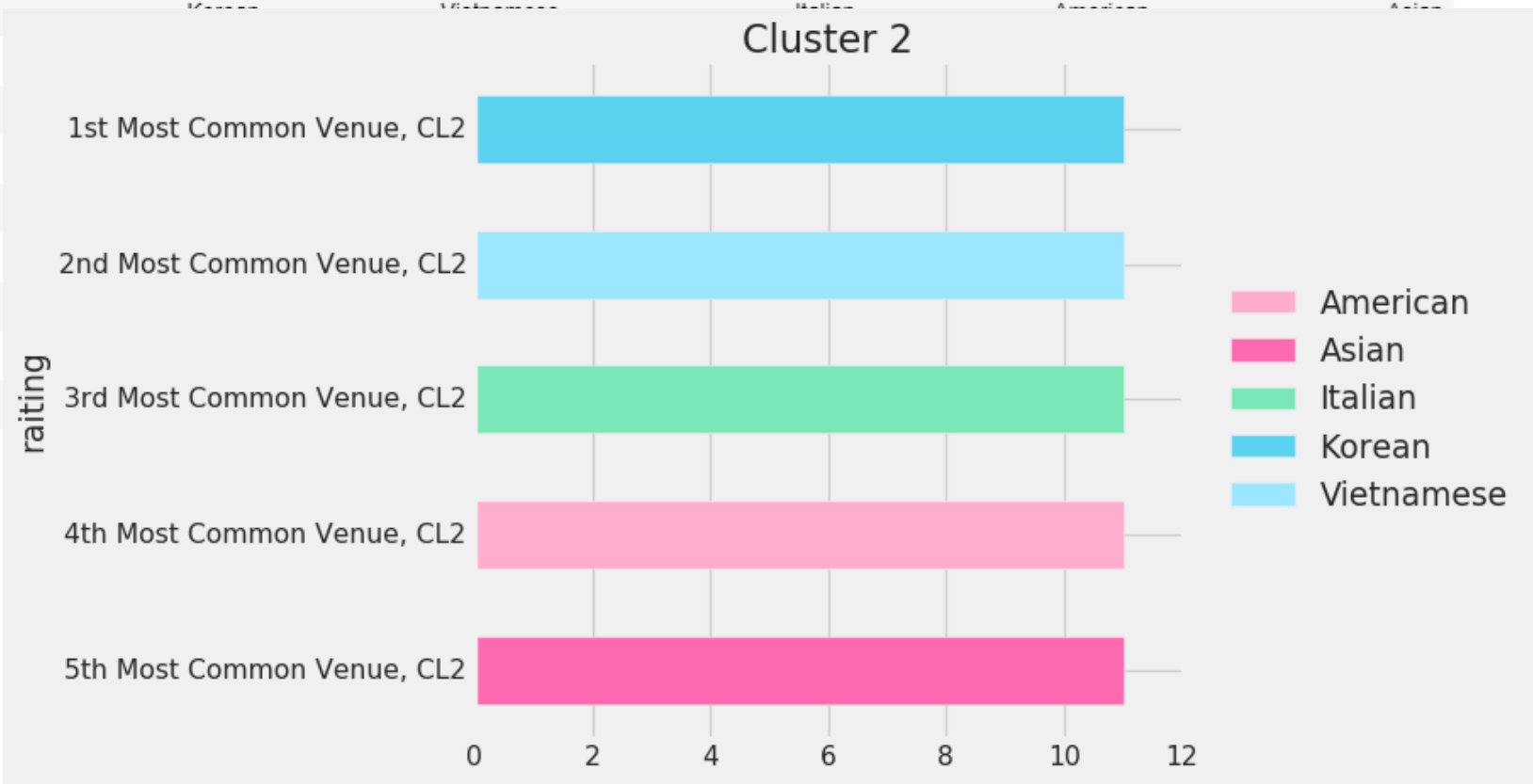
Cluster1 cuisines

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	Hillcrest Village	1	Japanese	Asian	Italian	Middle Eastern	Chinese
18	Fairview,Henry Farm,Oriole	1	Japanese	Asian	Italian	Middle Eastern	Chinese
19	Bayview Village	1	Japanese	Asian	Italian	Middle Eastern	Chinese
20	Silver Hills,York Mills	1	Japanese	Asian	Italian	Middle Eastern	Chinese
21	Newtonbrook,Willowdale	1	Japanese	Asian	Italian	Middle Eastern	Chinese
22	Willowdale South	1	Japanese	Asian	Italian	Middle Eastern	Chinese
23	York Mills West	1	Japanese	Asian	Italian	Middle Eastern	Chinese
24	Willowdale West						
25	Parkwoods						
26	Don Mills North						
27	Flemingdon Park,Don Mills South						
28	Bathurst Manor,Downsview North,Wilson Heights						
29	Northwood Park,York University						
30	CFB Toronto,Downsview East						
31	Downsview West						
32	Downsview Central						
33	Downsview Northwest						
34	Victoria Village						
35	Woodbine Gardens,Parkview Hill						
36	Woodbine Heights						
37	The Beaches						
38	Leaside						
39	Thorncliffe Park						
40	East Toronto						
41	The Danforth West,Riverdale						
42	The Beaches West,India Bazaar						
43	Studio District						
44	Lawrence Park						
45	Davisville North						
46	North Toronto West						

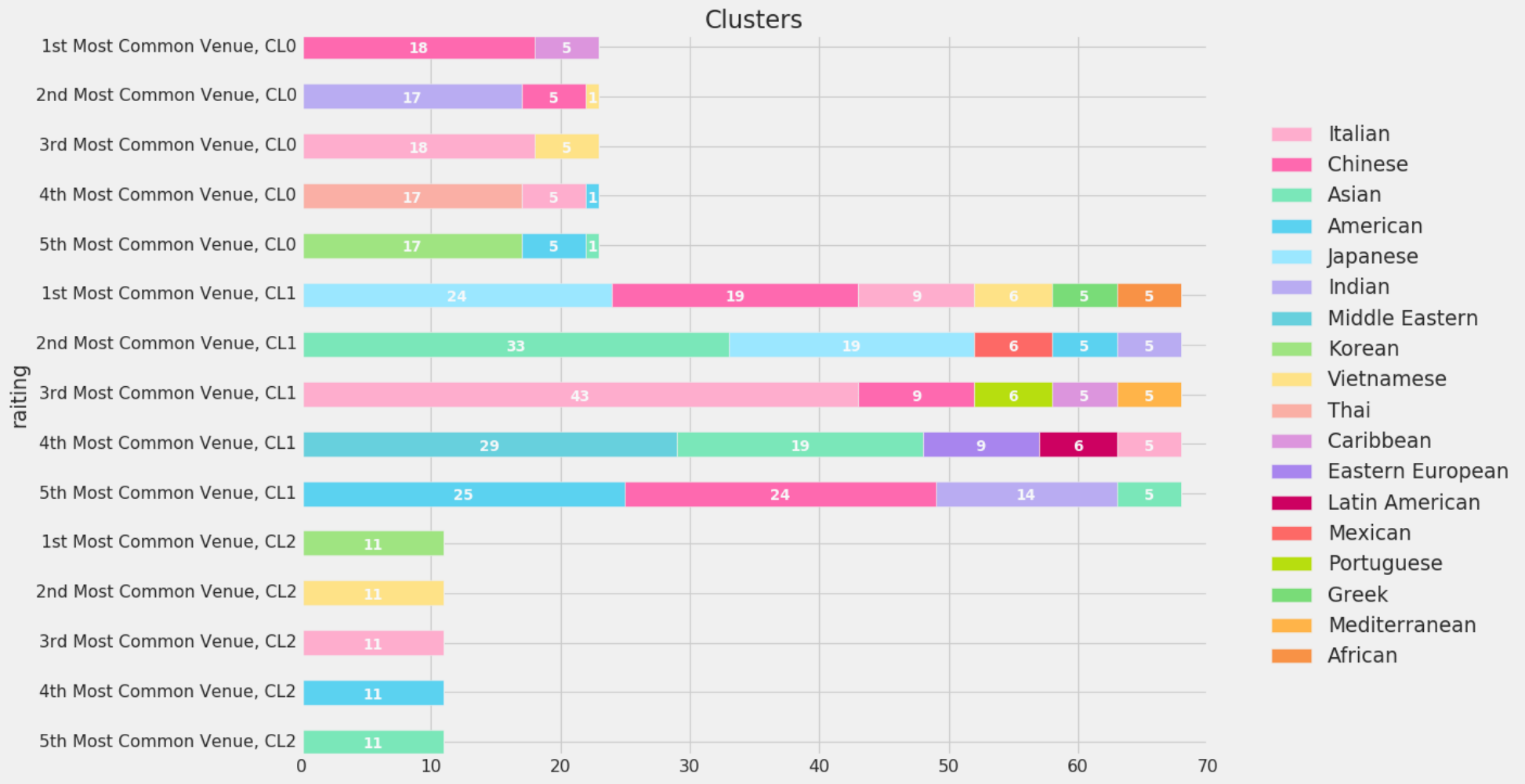


Cluster2 cuisines

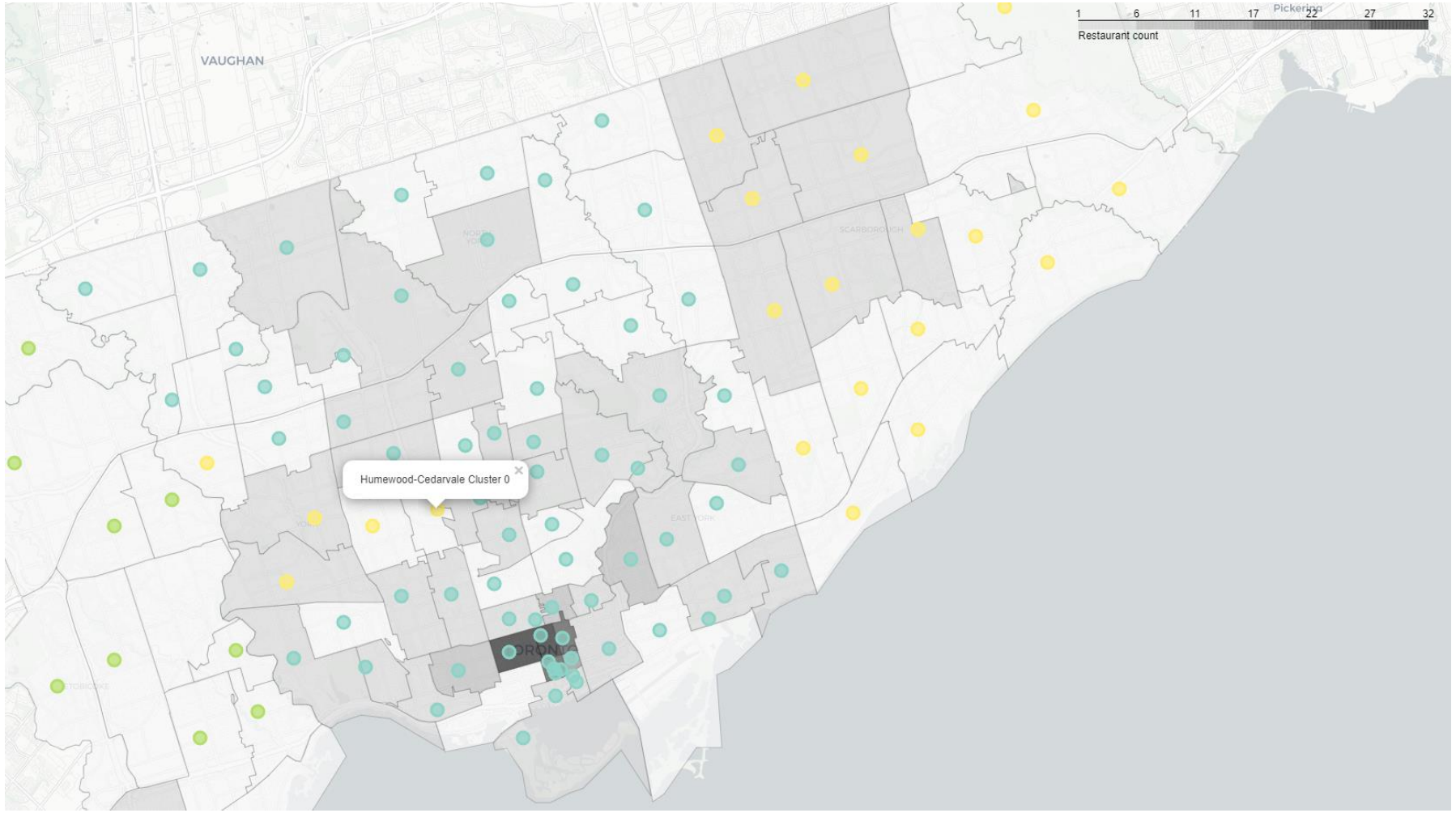
	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
88	Humber Bay Shores,Mimico South,New Toronto	2	Korean	Vietnamese	Italian	American	Asian
89	Alderwood,Long Branch	2	Korean	Vietnamese	Italian	American	Asian
90	The Kingsway,Montgomery Road,Old Mill North	2					
91	Humber Bay,King's Mill Park,Kingsway Park Sout...	2					
92	Kingsway Park South West,Mimico NW,The Queensw...	2					
94	Cloverdale,Islington,Martin Grove,Princess Gar...	2					
95	Bloordale Gardens,Eringate,Markland Wood,Old B...	2					
99	Westmount	2					
100	Kingsview Village,Martin Grove Gardens,Richvie...	2					
101	Albion Gardens,Beaumond Heights,Humbergate,Jam...	2					
102	Northwest	2					



All clusters common venues



Combined map of restaurant count and cuisine clusters



Conclusion

- Made useful map for choosing meal location.
- Many statistics based on cuisines.
- Ideas to add further:
  - ratings for restaurant;
  - tips for chosen restaurants;
  - user profile to help with choice and predict tastes.