FINAL REPORT

THE BATTLE OF NEIGHBORHOODS

INTRODUCTION

New York and Toronto are both cities with global recognition and attract a diverse set of people from different parts of the world. Being ethnically diverse cities requires them to cater to a vast variety of people hailing from different countries. New York having an immigrant population of 3.1 million as of 2018 and Toronto having a population of roughly 1.5 million as of 2017, both cities host people from a multitude of cultures and ethnicities.

With a diverse culture comes diversity in cuisine. This project looks into the availability of restaurants of varying cuisines in both cities. Both have restaurant types ranging from Italian, Chinese, Indian to American, French and German. Through this project, the magnitude and availability of these different types of restaurants is presented which can be useful for any person looking to move to either of the cities or wanting to open a new restaurant in either city.

Problem:

This project aims to achieve the following:

1. Compare the top 10 cuisines in both New York and Toronto
2. Provide a view of similar neighborhoods in terms of cuisine in both cities.

DATA

Demographics of both New York and Toronto show that they are highly diverse cities.

Thus, both are host to a variety of restaurants for everyone.

For this project, the following data has been used:

1. New York City Data containing information on boroughs, neighborhoods and their geographical coordinates

   Source: https://cocl.us/new_york_dataset

   Description: The data is used in conjugation with Foursquare API to zero in on each venue of New York City and identify the same on a map.

2. Toronto City Data containing information on boroughs, neighborhoods and their geographical coordinates

   Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for neighborhoods and boroughs.
   https://cocl.us/Geospatial_data for geospatial coordinates

Description: The data is used in conjugation with Foursquare API to zero in on each venue of Toronto City and identify the same on a map.

METHODOLOGY

We begin by extracting data for New York City using the following link:

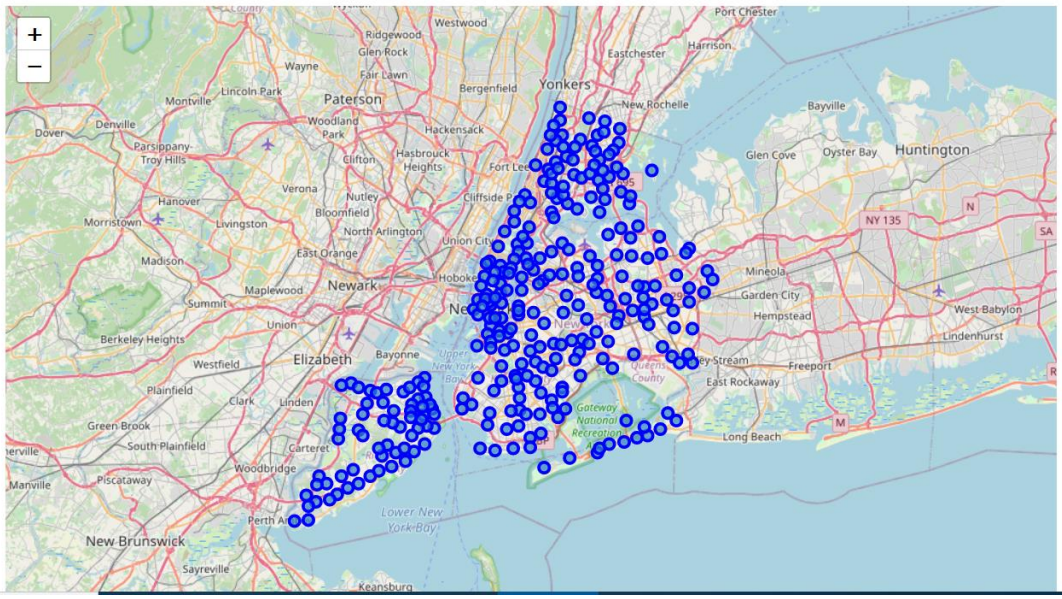
https://cocl.us/new_york_dataset

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Using the geopy library, we obtain latitude and longitude values of New York City and print a map of New York using folium

```
map_newyork
```

Out[22]:



After that, the venues of each neighborhood are extracted using Foursquare API.

Let's check the size of the resulting dataframe

In [32]:
```
print(manhattan_venues.shape)
manhattan_venues.head()
```

```
(3093, 7)
```

Out[32]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Following this, a one-hot representation of different venues in New York is created. This is grouped by mean of occurrence and narrowed down to various types of restaurants in New York.

Out[43]:

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cambodian Restaurant	Cantonese Restaurant	Caribbe Restaur
0	Battery Park City	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
1	Carnegie Hill	0.00	0.000000	0.011765	0.000000	0.011765	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
2	Central Harlem	0.00	0.068182	0.045455	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.022
3	Chelsea	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
4	Chinatown	0.00	0.000000	0.030000	0.000000	0.000000	0.020000	0.000000	0.01	0.00	0.000000	0.01	0.000(
5	Civic Center	0.00	0.000000	0.053763	0.000000	0.000000	0.010753	0.010753	0.00	0.00	0.000000	0.00	0.000(
6	Clinton	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.00	0.01	0.000000	0.00	0.000(
7	East Harlem	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
8	East Village	0.00	0.000000	0.010000	0.010000	0.010000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.010(
9	Financial District	0.00	0.000000	0.060000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
10	Flatiron	0.00	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
11	Gramercy	0.00	0.000000	0.035714	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
12	Greenwich Village	0.00	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.020(
13	Hamilton Heights	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.035(
14	Hudson Yards	0.00	0.000000	0.053571	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
15	Inwood	0.00	0.000000	0.034483	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.017
16	Lenox Hill	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(
17	Lincoln Square	0.00	0.000000	0.032258	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000(

This facilitates printing each neighborhood along with 5 most common cuisines in that area.

```
----Battery Park City----
                    venue  freq
0        Mexican Restaurant  0.03
1  Mediterranean Restaurant  0.01
2        Chinese Restaurant  0.01
3         Afghan Restaurant  0.00
4       Moroccan Restaurant  0.00


----Carnegie Hill----
                    venue  freq
0    Japanese Restaurant  0.04
1  Vietnamese Restaurant  0.02
2      French Restaurant  0.02
3       Sushi Restaurant  0.02
4     Italian Restaurant  0.02


----Central Harlem----
                    venue  freq
```

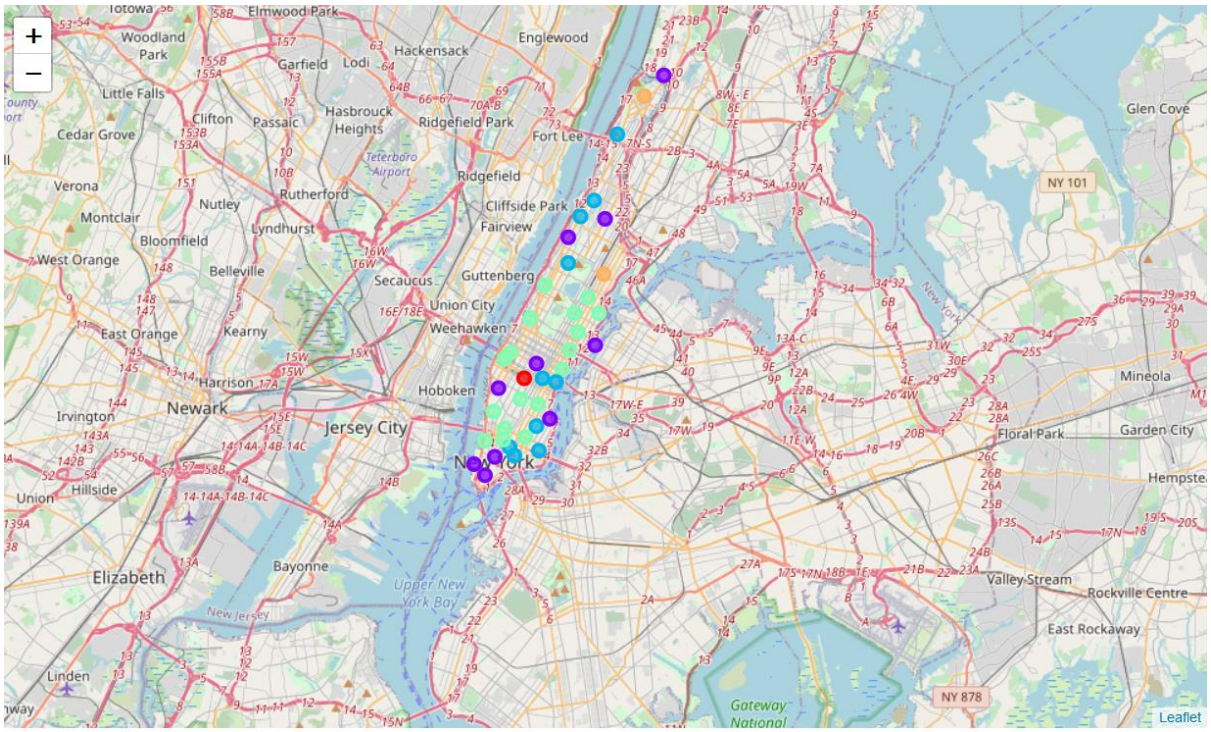
33c532c284dee935cbd6f3f2eeff1f3831d70786

Top 10 types of restaurants in each neighborhood are displayed after sorting the 'Total' column.

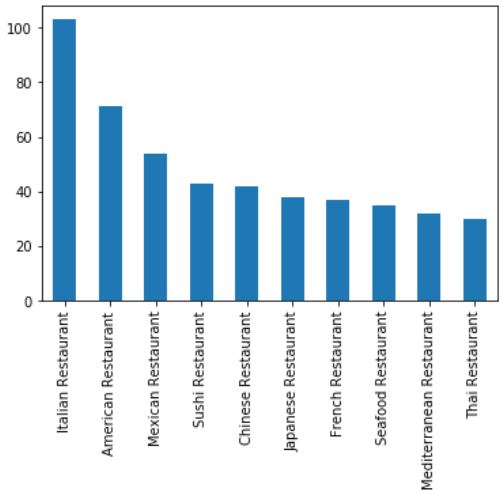
Out[46]:

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Battery Park City	Mexican Restaurant	Chinese Restaurant	Mediterranean Restaurant	French Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant
1	Carnegie Hill	Japanese Restaurant	Sushi Restaurant	French Restaurant	Italian Restaurant	Vietnamese Restaurant	New American Restaurant	American Restaurant	Mexican Restaurant	Argentinian Restaurant	Ramen Restaurant
2	Central Harlem	African Restaurant	French Restaurant	American Restaurant	Seafood Restaurant	Chinese Restaurant	Southern / Soul Food Restaurant	Ethiopian Restaurant	Caribbean Restaurant	English Restaurant	Falafel Restaurant
3	Chelsea	American Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Ramen Restaurant	Restaurant	Sushi Restaurant	Middle Eastern Restaurant
4	Chinatown	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Asian Restaurant	Malay Restaurant	Dim Sum Restaurant	Shanghai Restaurant	Mexican Restaurant	Greek Restaurant	Korean Restaurant

Using K-means clustering, the neighborhoods are sorted into clusters and similar neighborhoods are displayed on a map using folium.



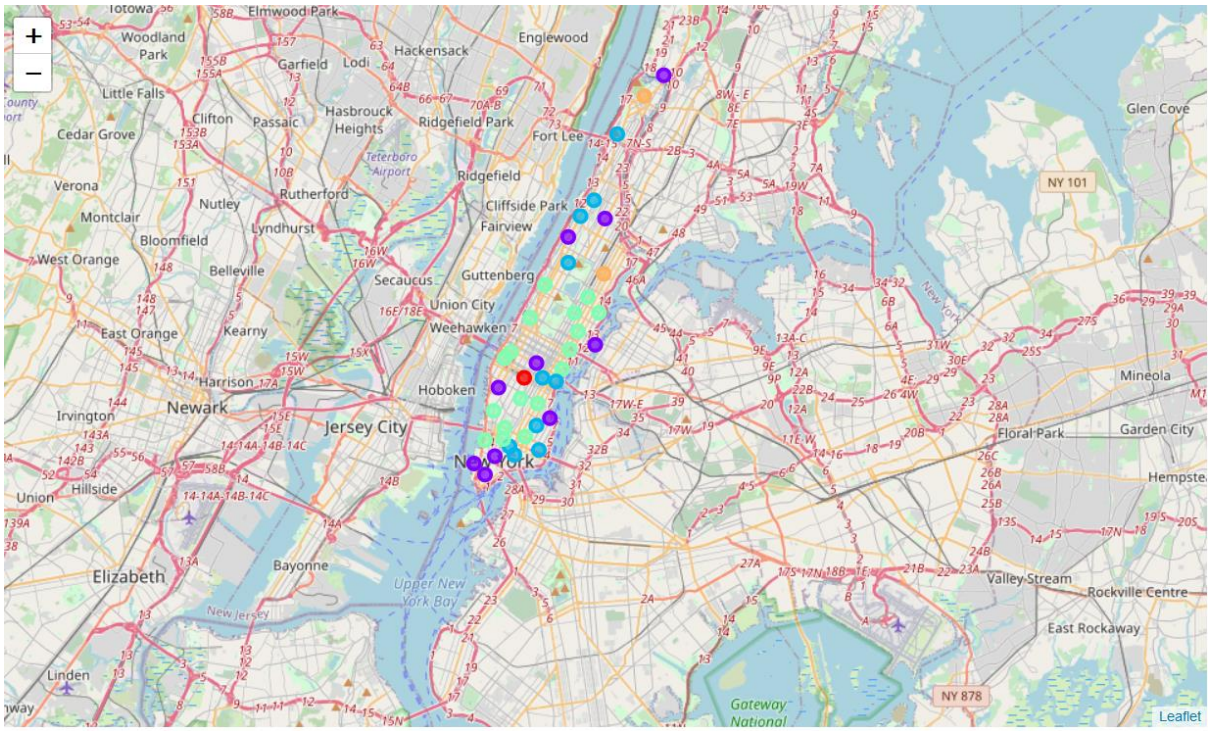
Using the one-hot representation of the neighborhood venues dataframe, the 10 most common cuisines in New York are displayed using a matplotlib bar chart.



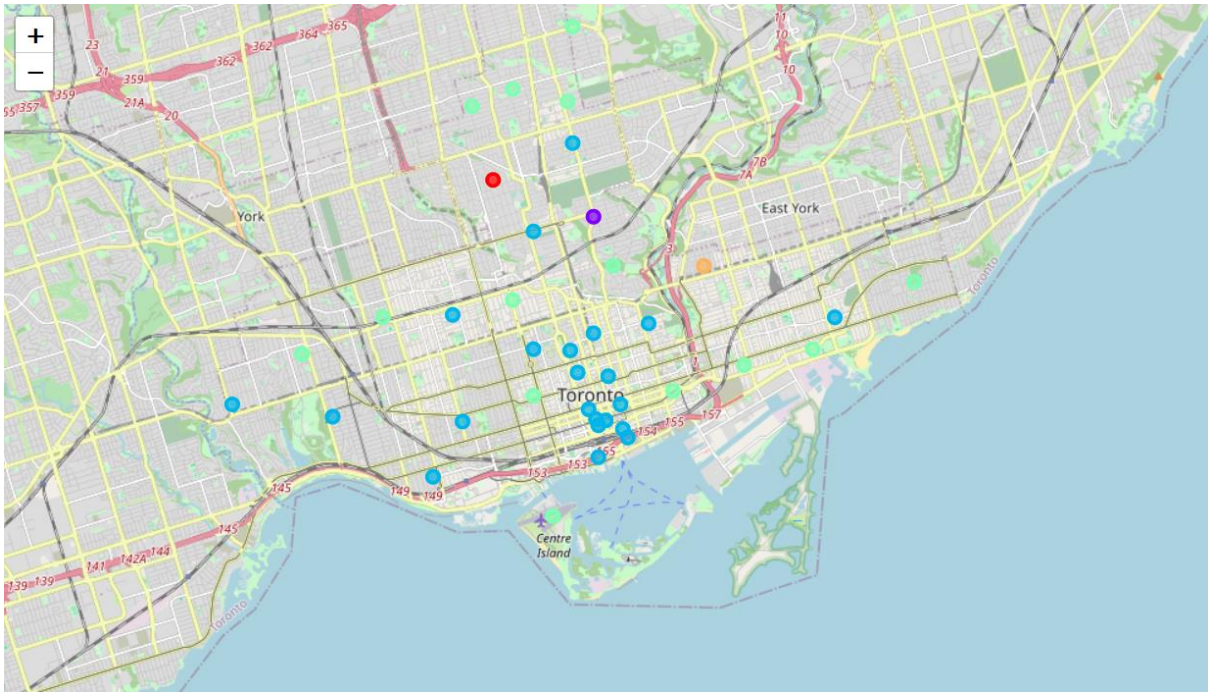
The same procedure is repeated for Toronto.

RESULTS

Clusters of neighborhoods in New York with similar cuisines of restaurants:

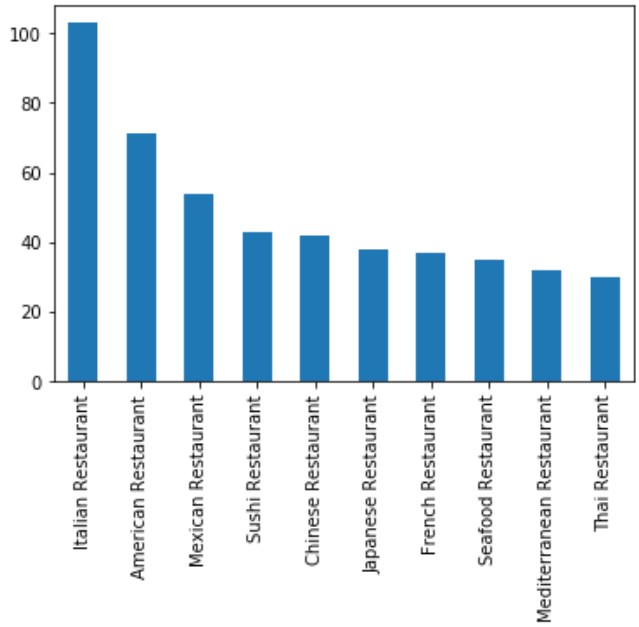


Clusters of neighborhoods in Toronto with similar cuisines of restaurants:

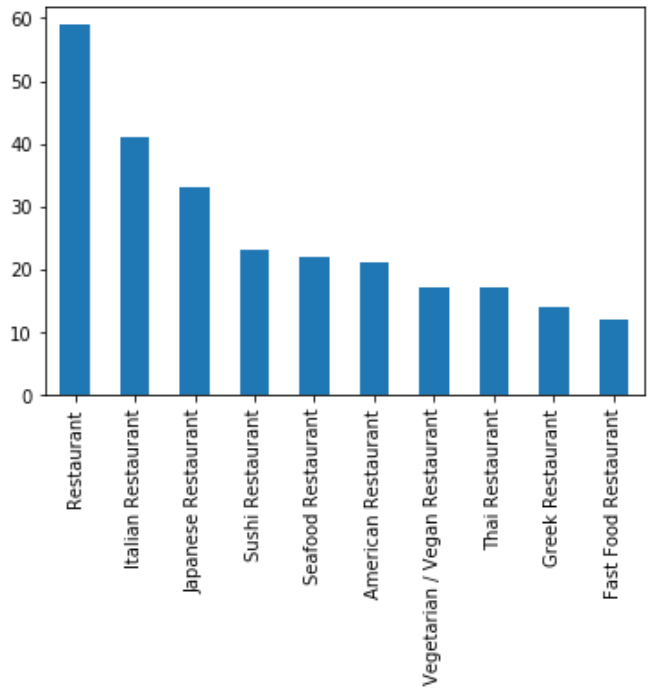


Comparative analysis of top cuisines in New York and Toronto:

New York



Toronto



- New York predominantly has Italian restaurants while Toronto has a majority of general restaurants.
- Both cities do have a wide variety of cuisines to offer and many neighborhoods have similar abundances of restaurant types

CONCLUSION

With the available data, the above report has been generated but there is always some scope of improvement in the same. Most common cuisines have been identified and clusters of similar cuisines have been portrayed.