

Capstone Project - Battle of The Neighborhoods

Comparing Restaurants within Toronto & New York City

Megan W.

Introduction: Project Description

Problem

Problem: "How many Halal or Kosher Restaurants can be found within the Downtown Toronto area, and how do the results compare to offering within New York City?" The analysis conducted within this project is to compare the downtown neighborhoods of Toronto and New York City and determine how similar or dissimilar they are.

Cities of Interest - Toronto & New York City

Project Background

Problem Background: To ensure the project for the capstone is completed, we will limit our work to specify Halal and Kosher. For individual that have a protein intolerance or allergy to Pork, Halal and Kosher Restaurants can ensure the patron is safe, because Halal and Kosher restaurant will not bring pork on the premise. With this in mind, the project will be focused on answering the following question:

"How many Halal or Kosher Restaurants can be found within the Downtown Toronto area?" "How do the results from Toronto Compare to New York City's offering?"

When someone has a protein intolerance or allergy to Pork it can be very hard to find a Restaurant that accommodate their needs; especially when cross contamination of food prep areas is a concern. There is no cure for food allergies. Strict avoidance of food allergens—and early recognition and management of allergic reactions to food—are important measures to prevent serious health consequences.

In order to help individuals with pork allergies to find safe dining options while visiting Toronto and New York City, this capstone project will develop an

algorithm that employs a methodology/approach that is tailored towards discovering Halal and Kosher Restaurants.

The results from the Downtown Toronto will be compared to the results from New York City. This comparison will then expose the venue types that offer Halal or Kosher food. Lastly the analysis will compare the neighborhoods of Toronto and New York City and determine how similar or dissimilar they are.

Target Audience - People with Pork Allergies or someone who is looking for Halal, Kosher, and/or Jewish Restaurants.

Data Use Overview

For this project, I intend to use data gathered from foursquare about halal and kosher restaurants and type of cuisine within downtown area for Toronto and New York City neighborhoods through use of API. I will use a supervised learning algorithms classify data points into categorical type of cuisine using best classification model.

Data Collection

Data Collection Steps:

The following steps will be used to collect the data:

1. I will need gather geo-locational information about that Downtown Toronto borough and the neighborhoods in that borough. We specifically and technically mean the latitude and longitude numbers of that borough. With this in mind, I will pull data from all the boroughs within the Downtown area by using the the Postal Codes that fall into that borough (Downtown).
2. I will need data about different venues in different neighborhoods of that specific borough. In order to gain that information, I will use "Foursquare" locational information. By locational information for each venue, I pull basic and advanced information about that venue. This means that I obtain the venue's precise latitude and longitude and also its distance from the center of the neighborhood, the category of that venue, whether this venue serves halal or kosher food its category, and the type of food served at this venue.
3. This process will be repeated for New York City information.

Data Sources

Following data sources will be needed to extract/generate the required information:

1. Corresponding Postal Codes from Wikipedia for Toronto

2. NYU New York City Neighborhood information
3. Centers of candidate areas will be collected from Wikipedia or will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
4. Number of restaurants for each cuisine type in neighborhood i.e approximate radius of 500 meters will be obtained using Foursquare API

Data Types - CSV, JSON, APIs

File Locations -

1. <https://foursquare.com/>
2. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
3. https://cocl.us/new_york_dataset

Data Analysis Methodology

General Plan:

I plan to use variables pertaining to venue type, location, food type, etc., obtained from foursquare in relation to downtown Toronto and New York City to see what categories of Halal and Kosher food are available within the neighborhood. I will conduct exploratory analysis on data pulled from foursquare to see what patterns can be gleamed from the data. I will then compare results from Toronto with the New York City results to see what area has the best restaurant options for someone with a pork allergy.

I will also use k-means to cluster the finding from downtown Toronto and New York City to support the comparison on the findings to see what else can be gleamed. This means that I will examine all the neighborhoods within downtown Toronto and New York City/Manhattan areas.

This analysis will be conducted looking at venue information. I will use k-means to help map out the results on interactive map to help visualize where the restaurant is located.

Data Issues and Bias:

Based on definition of our problem, there is a possibility that the following factors will influence my results:

1. Number of existing Halal and Kosher restaurants in the neighborhood
2. The limit number of types Halal and Kosher
3. Personal bias or data bias
4. Incomplete data or lack of category information

Data Visualization:

This information will be displayed in maps, tables and other visualizations.

Data Results

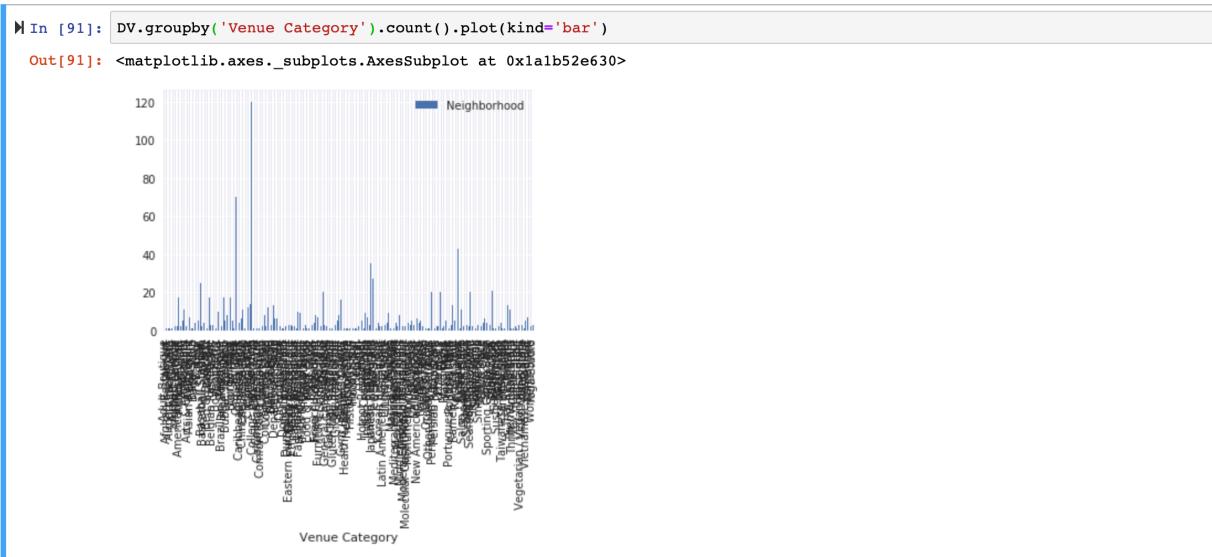
Overall Finding: Manhattan has more safe eating options and had 3 Restaurants listed as Jewish or Kosher within the Category information pulled from Foursquare API.

Kosher Findings Within Toronto:

Question: How many Halal or Kosher Restaurants can be found within the Downtown Toronto area?

Answer: No restaurant within in Toronto area was labeled Halal, Kosher, or Jewish within the Category information. However, a total number of possible safe restaurants found in the analysis is 22 within Downtown Toronto. These restaurants typically do not carry pork products: Afghan Restaurant 1, Ethiopian Restaurant 2, Falafel Restaurant 1, Persian Restaurant 1, Vegetarian / Vegan Restaurant 17.

I tried to map out the venue category count information, the plot is too hard to read.



Total number of possible safe restaurants 22 within Downtown Toronto

These restaurants typically do not carry pork products. The total breakdown is: Afghan Restaurant 1, Ethiopian Restaurant 2, Falafel Restaurant 1, Persian Restaurant 1, Vegetarian / Vegan Restaurant 17

Kosher Findings Within Manhattan:

Out[3]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights
7	M6A	North York	Lawrence Manor
8	M7A	Queen's Park	Not assigned
9	M8A	Not assigned	Not assigned

Question:
How many Halal or Kosher Restaurants can be found within the Manhattan area?

Answer:

In [92]

Italian Restaurant	123	123	123	123	123	123
Japanese Curry Restaurant	3	3	3	3	3	3
Japanese Restaurant	45	45	45	45	45	45
Jazz Club	8	8	8	8	8	8
Jewelry Store	6	6	6	6	6	6
Jewish Restaurant	3	3	3	3	3	3
Juice Bar	23	23	23	23	23	23
Karaoke Bar	3	3	3	3	3	3
Kebab Restaurant	1	1	1	1	1	1
Kids Store	3	3	3	3	3	3
Korean Restaurant	21	21	21	21	21	21
Kosher Restaurant	1	1	1	1	1	1
Latin American Restaurant	12	12	12	12	12	12

Total of 3 Restaurants are labeled as Jewish or Kosher.

Total number of possible safe restaurants 89 within Manhattan.

These restaurants typically do not carry pork products: Afghan Restaurant 1, Ethiopian Restaurant 4, Falafel Restaurant 8, Indian Restaurant 32, Israeli Restaurant 2, Jewish Restaurant 3, Kebab Restaurant 1, Kosher Restaurant 1, Lebanese Restaurant 1, Middle Eastern Restaurant 6, Moroccan Restaurant 2, North Indian Restaurant 1, Pakistani Restaurant 1, Persian Restaurant 1, South Indian Restaurant 1, and Vegetarian / Vegan Restaurant 24.

Total of 3 Restaurants in Manhattan area were labeled as Jewish or Kosher within the Category information. In addition, it was discovered that there are total number of possible safe restaurants 89 within Manhattan. These restaurants typically do not carry pork products: Afghan Restaurant 1, Ethiopian Restaurant 4, Falafel Restaurant 8, Indian Restaurant 32, Israeli Restaurant 2, Jewish Restaurant 3, Kebab Restaurant 1, Kosher Restaurant 1, Lebanese Restaurant 1, Middle Eastern Restaurant 6, Moroccan Restaurant 2, North Indian Restaurant 1, Pakistani Restaurant 1, Persian Restaurant 1, South Indian Restaurant 1, and Vegetarian / Vegan Restaurant 24.

Toronto City Analysis

1. Get Postal Data for Toronto - built the code to scrape the following Wikipedia page, [https://en.wikipedia.org/wiki/
List_of_postal_codes_of_Canada:_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), in order to obtain the data that is in the table of postal codes and to transform the data into a pandas dataframe
2. Create Dataframe from the Postal Data - The dataframe consists of three columns: PostalCode, Borough, and Neighborhood

Out[3]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights
7	M6A	North York	Lawrence Manor
8	M7A	Queen's Park	Not assigned
9	M8A	Not assigned	Not assigned

3. Clean Toronto Data -

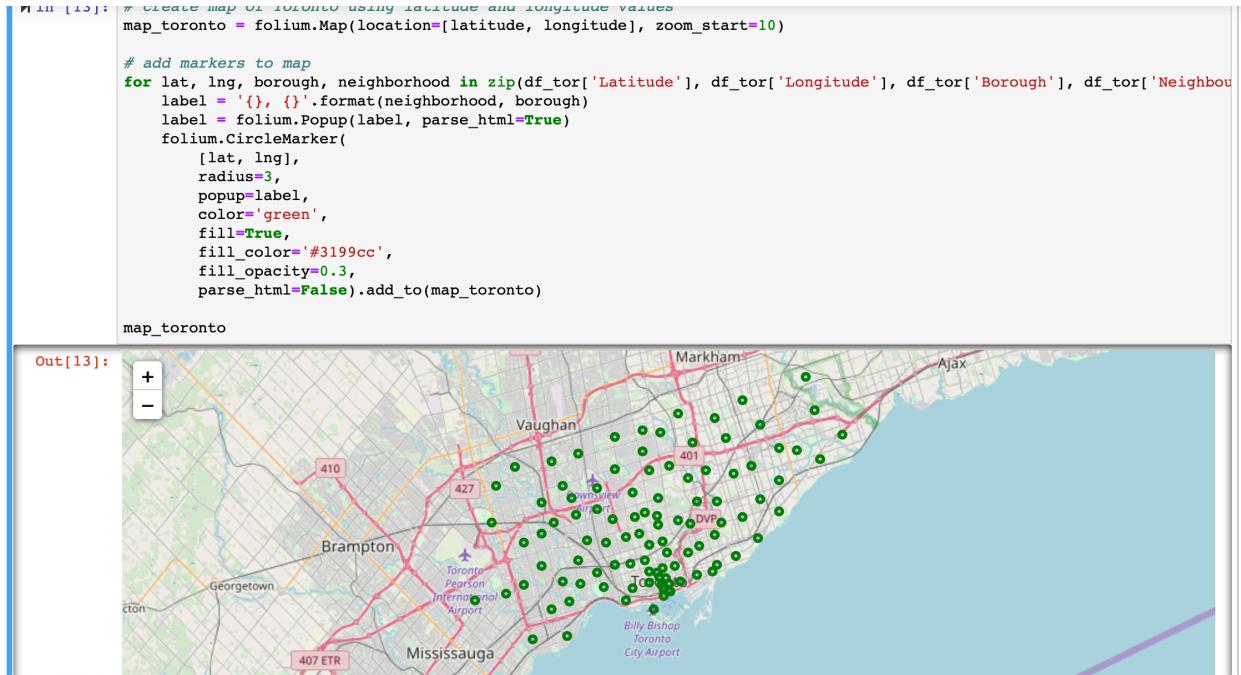
- Combine Neighborhoods with same postal code area - More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.
- Fixing Queen's Park - If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough. So for the 9th cell in the table on the Wikipedia page, the value of the Borough and the Neighborhood columns will be Queen's Park.

4. Adding Lat and Long Dataframe & Getting Lat and Long for neighborhoods -

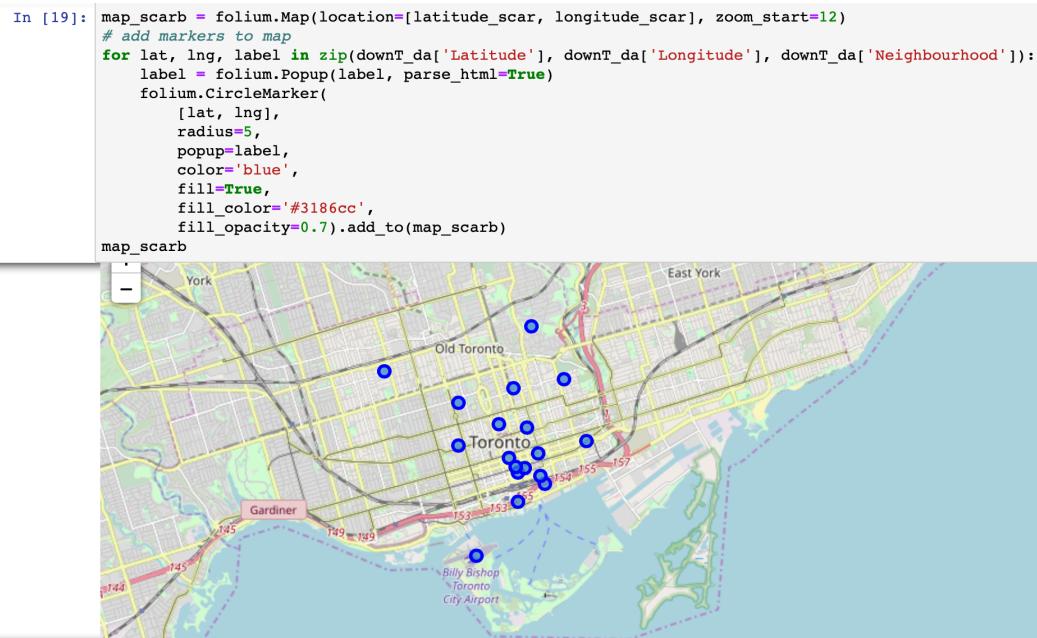
in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.

5. Explore and cluster the neighborhoods in Toronto

- Used geolocator to get the geographical coordinate of the City of Toronto: 43.653963, -79.387207.
- Built Map to see neighborhoods



- Built subset of map for downtown Toronto



- Pulled Foursquare data for neighborhoods

```
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues = nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head(10)
```

Out[25]:

	name	categories	lat	lng
0	Otto's Bierhalle	Beer Bar	43.643457	-79.422279
1	El Almacen Yerba Mate Cafe + Gallery	Café	43.643682	-79.421627
2	The Theatre Centre	Theater	43.643241	-79.423264
3	apt 200	Bar	43.644026	-79.420063
4	Bar Fancy	Bar	43.643734	-79.421326
5	The Drake Hotel	Hotel	43.643053	-79.424589
6	Cafe Neon	Café	43.644130	-79.419670
7	Union Restaurant	New American Restaurant	43.645860	-79.419532
8	Marmakas Taverna	Greek Restaurant	43.646042	-79.419679
9	OddSeoul	Korean Restaurant	43.646192	-79.419601

In [26]: `print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))`

83 venues were returned by Foursquare.

- Viewed category information within dataframe

In [30]: `downtown_venues.head(10)`

Out[30]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.679563	-79.377529	Mooredale House	43.678631	-79.380091	Building
1	Rosedale	43.679563	-79.377529	Rosedale Park	43.682328	-79.378934	Playground
2	Rosedale	43.679563	-79.377529	Whitney Park	43.682036	-79.373788	Park
3	Rosedale	43.679563	-79.377529	Alex Murray Parkette	43.678300	-79.382773	Park
4	Rosedale	43.679563	-79.377529	Milkman's Lane	43.676352	-79.373842	Trail
5	Cabbagetown, St. James Town	43.667967	-79.367675	Cranberries	43.667843	-79.369407	Diner
6	Cabbagetown, St. James Town	43.667967	-79.367675	F'Amelia	43.667536	-79.368613	Italian Restaurant
7	Cabbagetown, St. James Town	43.667967	-79.367675	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
8	Cabbagetown, St. James Town	43.667967	-79.367675	Kingyo Toronto	43.665895	-79.368415	Japanese Restaurant
9	Cabbagetown, St. James Town	43.667967	-79.367675	Murgatroid	43.667381	-79.369311	Restaurant

- Create the new dataframe and display the top 10 venues for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Thai Restaurant	Bar	American Restaurant	Hotel	Restaurant	Burger Joint	Cosmetics Shop
1	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Italian Restaurant	Bakery	Steakhouse	Beer Bar	Cheese Shop	Café	Farmers Market
2	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Lounge	Airport Terminal	Airport Service	Harbor / Marina	Boutique	Plane	Coffee Shop	Bar	Airport Gate	Airport Food Court
3	Cabbagetown, St. James Town	Coffee Shop	Restaurant	Park	Pub	Café	Bakery	Italian Restaurant	Pizza Place	Breakfast Spot	Butcher
4	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Middle Eastern Restaurant	Sandwich Place	Burger Joint	Japanese Restaurant	Bubble Tea Shop	Bar	Bakery
5	Chinatown, Grange Park, Kensington Market	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Mexican Restaurant	Dumpling Restaurant	Bar	Bakery	Chinese Restaurant	Vietnamese Restaurant	Gaming Cafe
6	Christie	Grocery Store	Café	Park	Baby Store	Italian Restaurant	Diner	Nightclub	Convenience Store	Restaurant	Coffee Shop
7	Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Café	Pub	Hotel	Fast Food Restaurant	Gym
8	Commerce Court, Victoria Hotel	Coffee Shop	Hotel	Café	Restaurant	Italian Restaurant	Seafood Restaurant	Gastropub	Deli / Bodega	American Restaurant	Steakhouse
9	Design Exchange, Toronto Dominion Centre	Coffee Shop	Café	Hotel	Restaurant	Bakery	Italian Restaurant	Steakhouse	Gastropub	Bar	Deli / Bodega
10	First Canadian Place, Underground city	Coffee Shop	Café	Restaurant	Steakhouse	Hotel	Seafood Restaurant	Asian Restaurant	Bakery	Deli / Bodega	American Restaurant
11	Harbord, University of Toronto	Café	Bookstore	Restaurant	Japanese Restaurant	Bar	Italian Restaurant	Bakery	Nightclub	Chinese Restaurant	Beer Store

- Cluster Neighborhoods - Run k-means to cluster the neighborhood into 5 clusters and build a map

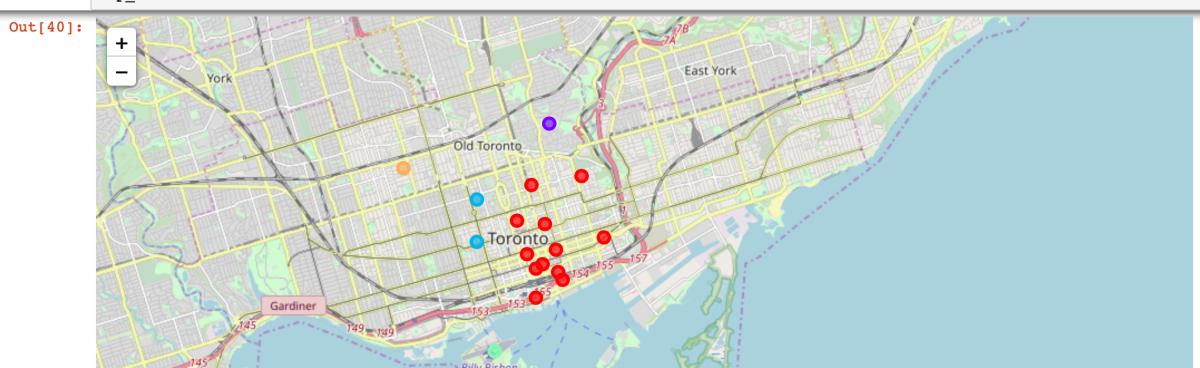
```

colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

for lat, lon, poi, cluster in zip(scarb_merged['Latitude'], scarb_merged['Longitude'], scarb_merged['Neighbourhood'], s):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), 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

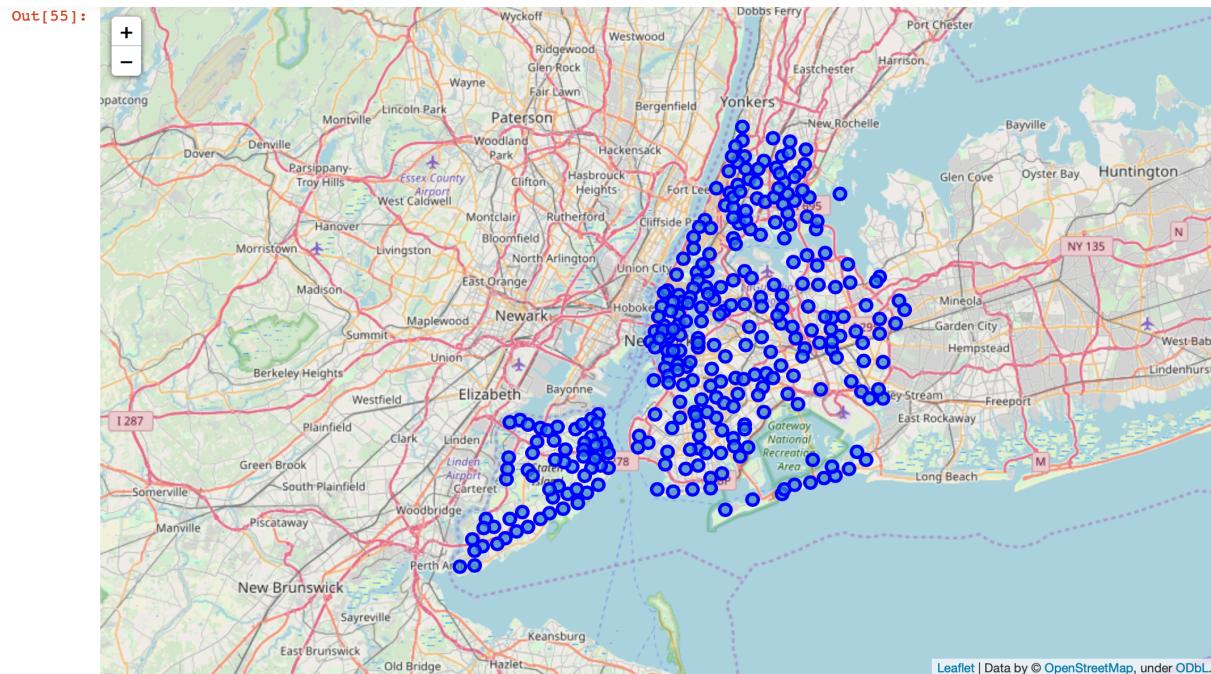
```



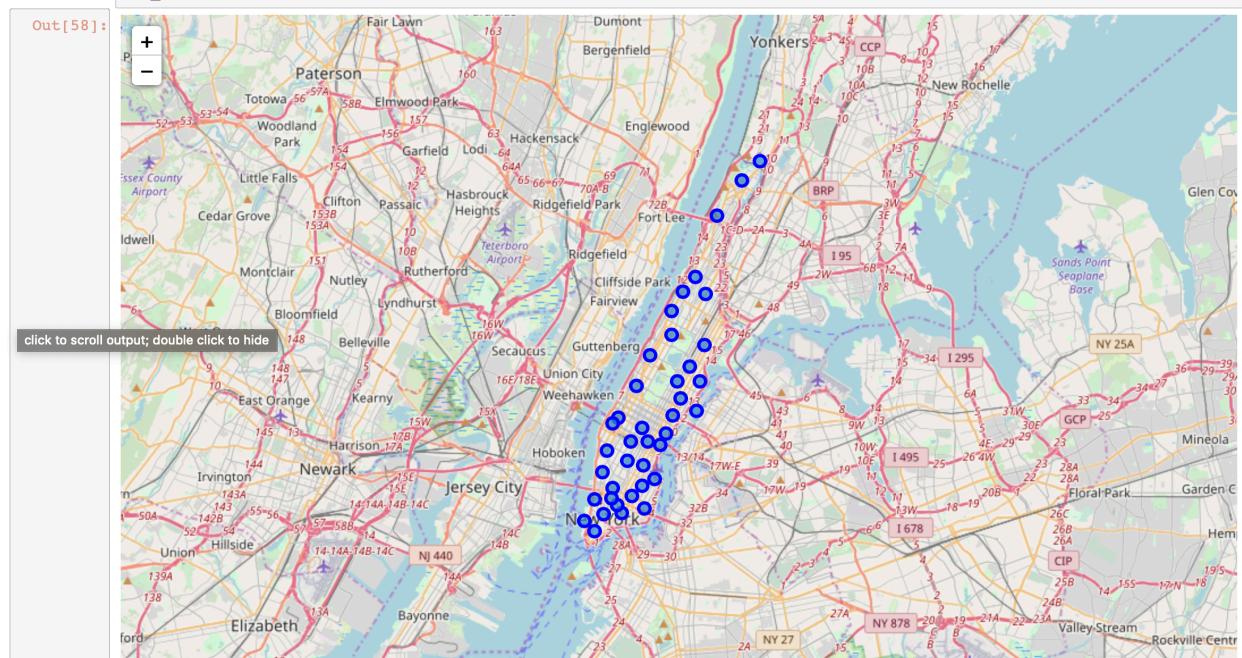
- Analysis each of the cluster results

New York Analysis

1. Get Postal Data for Toronto - built the code to scrape the following page, 'https://cocl.us/new_york_dataset' in order to obtain the data that is in the table of new york city neighborhoods and to transform the data into a pandas dataframe
2. Create Dataframe from the NYU JSON file
3. Clean NYC Data - Made smaller data set with just the manhattan neighborhood data.
4. Adding Lat and Long Dataframe & Getting Lat and Long for neighborhoods - in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.
5. Explore and cluster the neighborhoods in NYC
 - Gathered Lat and Long info
 - used geolocator to fine the geographical coordinate of New York City are 40.7127281, -74.0060152.
 - built NYC map



- built Manhattan map

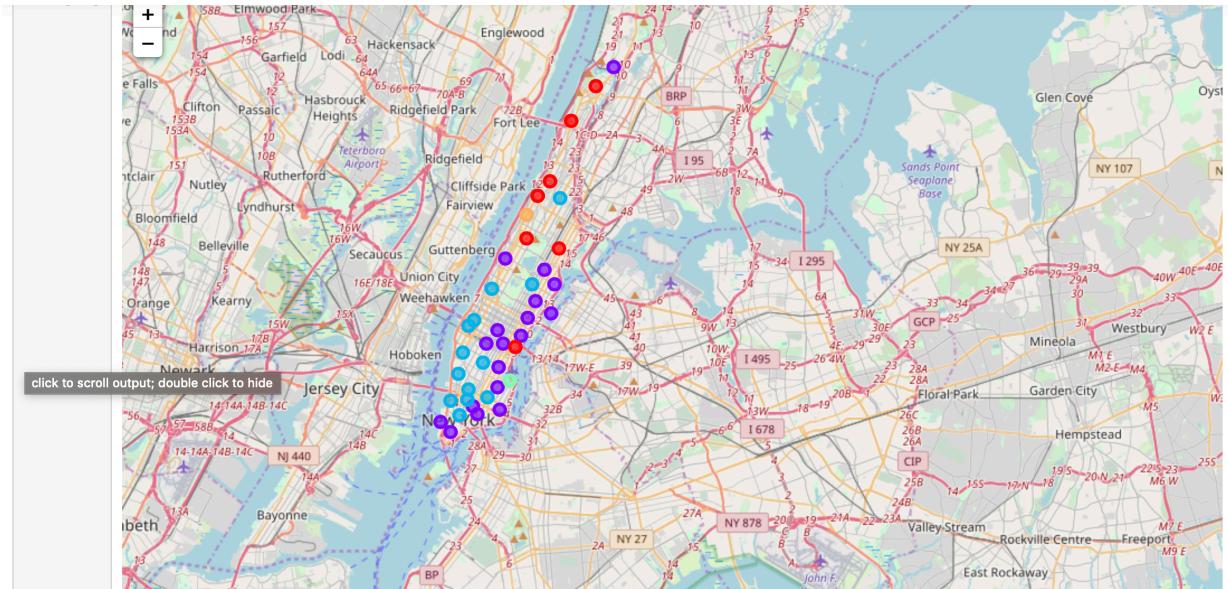


- Pulled Foursquare data for neighborhoods
- Viewed category information within dataframe
- Create the new dataframe and display the top 10 venues for each neighborhood.

Out[77]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Wine Shop	Clothing Store	Italian Restaurant	Department Store	Women's Store
1	Caregie Hill	Coffee Shop	Pizza Place	Café	Yoga Studio	Bookstore	Cosmetics Shop	French Restaurant	Bar	Japanese Restaurant	Spa
2	Central Harlem	African Restaurant	Art Gallery	Seafood Restaurant	American Restaurant	Gym / Fitness Center	French Restaurant	Cosmetics Shop	Chinese Restaurant	Public Art	Grocery Store
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Nightclub	Bakery	Seafood Restaurant	American Restaurant	Theater	Art Gallery	Hotel
4	Chinatown	Chinese Restaurant	Cocktail Bar	American Restaurant	Dumpling Restaurant	Spa	Ice Cream Shop	Bakery	Vietnamese Restaurant	Bubble Tea Shop	Dim Sum Restaurant

- Cluster Neighborhoods - Run k-means to cluster the neighborhood into 5 clusters and build a map



- Analysis each of the cluster results

Discussion:

Based on findings, I am disappointed in the fact that my question and its analysis did not provide enough data on Halal, kosher, and jewish restaurants. I had to compare Manhattan and Downtown Toronto in the same way that class assignment to ensure that I was able to meet all the requirements for the assignment. I tried a number of different way to pull the data and most these attempt came with issues- including searching on the category ids for halal and kosher restaurants.

I am glad that I was able to gather results within NYC that directly stated kosher and jewish with radius of 500, even knowing there are many kosher and halal restaurants with NYC.

I do think the category information for the venues is not easy to gather from the API and the data is dependent on whether it was entered correctly by the user.

Conclusion:

While the concept is a good one and there are number of application providing this service, I wanted to see if I could gather and analyze the information using data science techniques.

1. If the algorithm is working on to query the category ID over a larger radius area, you might be able to use the information in an application.
2. You might have to gather information across the U.S. to able to sub-categorize the types of halal, kosher, and jewish restaurants are available.