Capstone Project - The Battle of Neighborhoods (Week 2)

Introduction:

Queens is a New York City borough on Long Island across the East River from Manhattan. Flushing Meadows Corona Park, with the Unisphere, a 12-story 1964 World's Fair globe sculpture, hosts the annual U.S. Open tennis tournament. The park's Queens Museum is known for the "Panorama," a building-for-building model of New York City. Nearby Citi Field is the stadium of pro baseball team, the Mets. As being a tourist, she/he would want to visit Queens, before the tourist coming, it is always nice to know what the region would be mostly like, what kind of style of that region is. This project is going to explore the shops in Queens and provide a clear insight for tourist who are wondering how Queens is.

Data - 2014 New York City Neighborhood Names

- This New York City Neighborhood Names point file was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale.

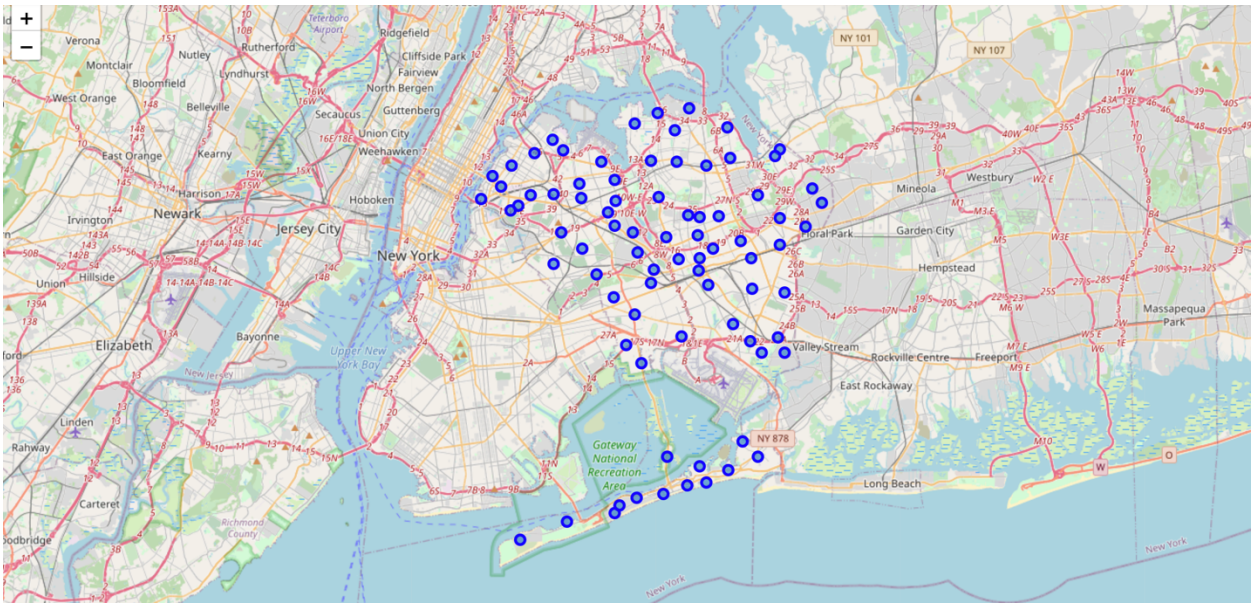
Methodology

- Foursquare API **-** the assignment will use the **explore** function to get the most common venue categories in the neighborhood of Queens, and then use this feature to group the neighborhoods into clusters.
- One-hot encoding
- Clustering - the *k*-means clustering algorithm will be applied to complete this task.

- Map - finally, the Folium library to visualize the neighborhoods in Queens and their emerging clusters will be merged.

Results

Map of Queens



Grouped Venues

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arverne	19	19	19	19	19	19
Astoria	100	100	100	100	100	100
Astoria Heights	13	13	13	13	13	13
Auburndale	20	20	20	20	20	20
Bay Terrace	36	36	36	36	36	36
Bayside	75	75	75	75	75	75
Bayswater	2	2	2	2	2	2
Beechhurst	14	14	14	14	14	14
Bellaire	11	11	11	11	11	11
Belle Harbor	17	17	17	17	17	17
Bellerose	18	18	18	18	18	18
Blissville	15	15	15	15	15	15
Breezy Point	5	5	5	5	5	5
Briarwood	11	11	11	11	11	11
Broad Channel	5	5	5	5	5	5
Brookville	1	1	1	1	1	1
Cambria Heights	12	12	12	12	12	12

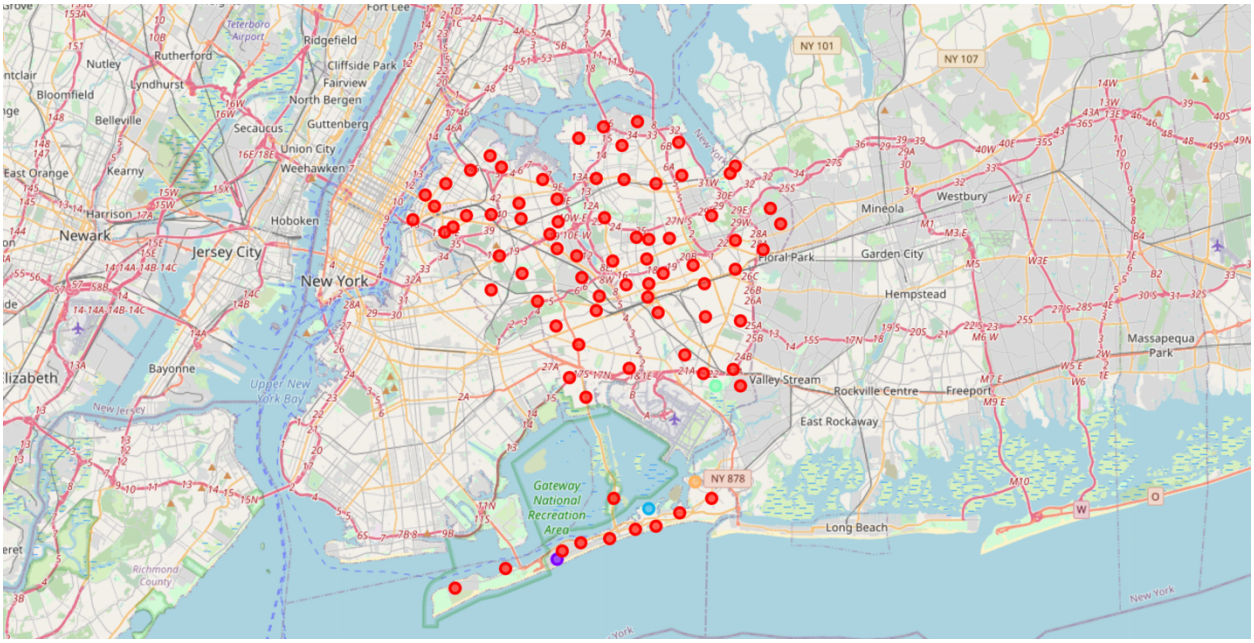
One-hot encoded results – First five rows

	Yoga Studio	Accessories Store	Afghan Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Ba
0	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	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Clustering Neighborhood

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Queens	Astoria	40.768509	-73.915654	0	Bar	Middle Eastern Restaurant	Hookah Bar	Greek Restaurant	Seafood Restaurant	Bakery	Pizza Place	Mediterranean Restaurant	Indian Restaurant	Food Truck
1	Queens	Woodside	40.746349	-73.901842	0	Grocery Store	Thai Restaurant	Latin American Restaurant	Filipino Restaurant	Bakery	Pub	Bar	Donut Shop	American Restaurant	Ice Cream Shop
2	Queens	Jackson Heights	40.751981	-73.882821	0	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mexican Restaurant	Thai Restaurant	Grocery Store	Mobile Phone Shop	Clothing Store	Supermarket
3	Queens	Elmhurst	40.744049	-73.881656	0	Thai Restaurant	Mexican Restaurant	Vietnamese Restaurant	Chinese Restaurant	Bubble Tea Shop	Park	Colombian Restaurant	Salon / Barbershop	Food Court	Snack Place
4	Queens	Howard Beach	40.654225	-73.838138	0	Italian Restaurant	Bagel Shop	Sandwich Place	Fast Food Restaurant	Pharmacy	Concert Hall	Deli / Bodega	Shipping Store	Seafood Restaurant	Salon / Barbershop

Clustering Map in Queens



Discussion

The type of shops in Queens is fewer than that in Manhattan as practiced in the Lab. The most common venues in Queens have been clustered and explored and summarized by this project. It can provide an instant glance for people who have never been to Queens and help them to realize which neighborhood fits their interest.

Conclusion

1. The characteristic of Queens in NY has been explored.
2. Foursquare API were fully integrated into the project and analysis.
3. Machine Learning algorithm such as k-means clustering and one-hot encoding have been implemented into the project.
4. Results show that the project can help tourist save time to explore venues with a data-based analysis.