1 Introduction

1.1 Background

Websites like Trulia and TripAdvisor are constantly looking for content to push to consumers. These are usually educational infographics that talk about a city or country that their consumers might be interested in traveling to.

1.2 Idea

My idea was to find out which neighborhoods in the five boroughs of New York City had the most culturally diverse collection of restaurants, to be used as one such piece of consumer-facing content for a site such as Trulia or TripAdvisor.

1.3 Target Audience

The audience would be users of services such as Trulia and TripAdvisor, but more specifically it would be anyone who is planning to make a trip to New York and is looking for guidance/advice on what to do or eat once here.

2 Data

2.1 Data Used

The New York geographical data used was provided in a previous lesson; the json file was downloaded from the cloud server to be used for this project. Restaurant information was pulled from the Foursquare API. The coordinates of New York City were found online.

3 Methodology

3.1 Pulling data from New York json

New York city geographical data, including neighborhood name, borough name and neighborhood coordinates, were pulled from the json file used in one of the earlier lessons.

	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

3.2 Exploring Foursquare restaurant data

An exploration of the way that restaurants are categorized provided an early obstacle towards the project. Some restaurants were merely labeled as “coffee shop” or “bagel place”, with some ethnic restaurants being very specifically classified (e.g. “Cha Chan Teng” to refer to a Cantonese diner). A decision was made to parse the data to only include restaurants where ‘Restaurant’ existed in the string, as this is how a majority of the ethnic restaurants were classified, even though this means that categories such as ‘Fast Food restaurant’ or ‘Seafood restaurant’ would also count towards a neighborhood’s culinary diversity.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
15	Co-op City	40.874294	-73.829939	Arby's	40.870411	-73.828606	Fast Food Restaurant
16	Co-op City	40.874294	-73.829939	Townhouse Restaurant	40.876086	-73.828868	Restaurant
20	Co-op City	40.874294	-73.829939	Kennedy's	40.876807	-73.829627	Fast Food Restaurant
26	Eastchester	40.887556	-73.827806	Fish & Ting	40.885656	-73.829197	Caribbean Restaurant
28	Eastchester	40.887556	-73.827806	Dyre Fish Market	40.889318	-73.831453	Seafood Restaurant

3.3 Processing Data

After the venue categories were filtered to contain ‘Restaurant’ in the strings, the data was then grouped by neighborhood to see how many categories existed in each. This column was then retitled ‘Restaurant Number’.

	Neighborhood	Venue Category
0	Allerton	4
1	Annadale	1
2	Arrochar	5
3	Arverne	2
4	Astoria	36

3.4 K-means clustering

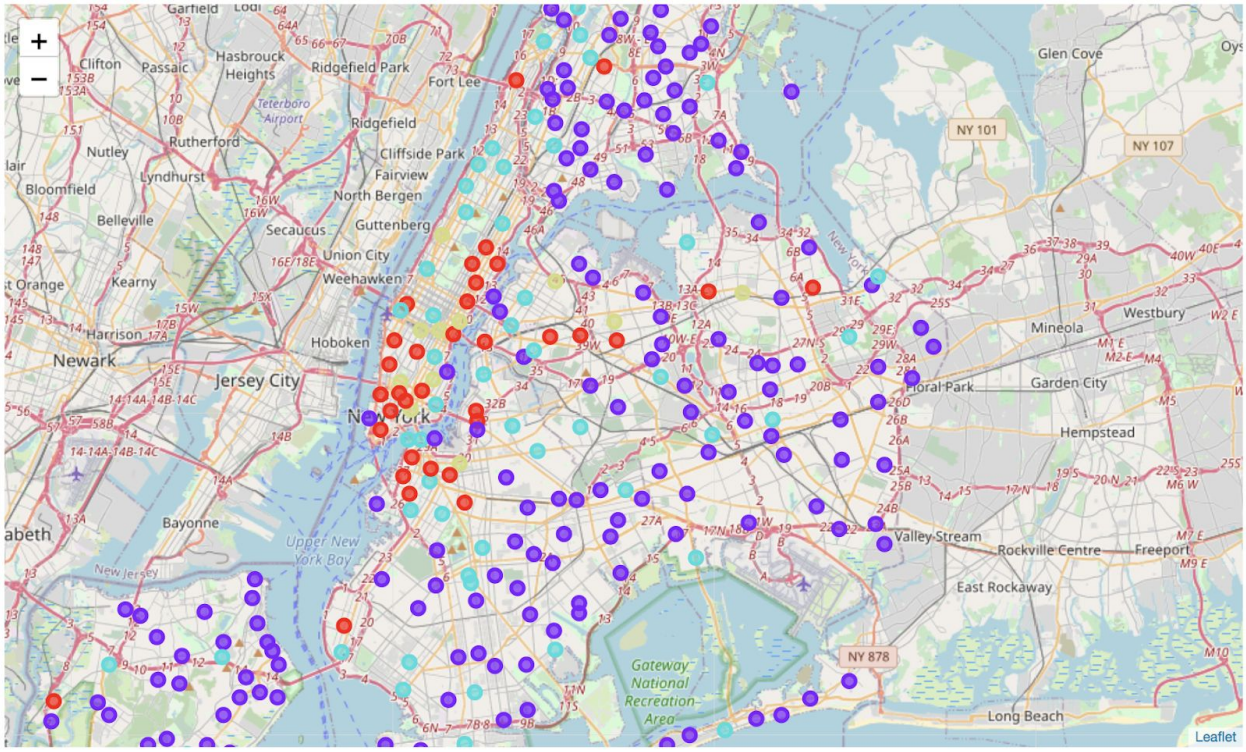
This data was then run through the k-means clustering algorithm, which was set to split the data into 4 clusters. This data was then appended to a neighborhood dataframe created earlier as shown below.

	Neighborhood	Number of Restaurants	Cluster Labels	Borough	Latitude	Longitude
0	Allerton	4	1	Bronx	40.865788	-73.859319
1	Annadale	1	1	Staten Island	40.538114	-74.178549
2	Arrochar	5	1	Staten Island	40.596313	-74.067124
3	Arverne	2	1	Queens	40.589144	-73.791992
4	Astoria	36	3	Queens	40.768509	-73.915654

4 Results

4.1 Visualization

This data was then mapped using Folium to show the location of each neighborhood, with the colors indicating the cluster it was split into. The color coding is as follows: Red = cluster 0, Purple = cluster 1, Blue = cluster 2 and Yellow = cluster 3.



4.2 Cluster Classification

Attached below are tables to describe each cluster label:

	Neighborhood	Number of Restaurants	Cluster Labels	Borough	Latitude	Longitude
4	Astoria	36	3	Queens	40.768509	-73.915654
39	Chinatown	40	3	Manhattan	40.715618	-73.994279
47	Clinton Hill	36	3	Brooklyn	40.693229	-73.967843
67	East Village	36	3	Manhattan	40.727847	-73.982226
102	Greenwich Village	43	3	Manhattan	40.726933	-73.999914
117	Jackson Heights	36	3	Queens	40.751981	-73.882821

	Neighborhood	Number of Restaurants	Cluster Labels	Borough	Latitude	Longitude
7	Bath Beach	17	2	Brooklyn	40.599519	-73.998752
20	Bensonhurst	12	2	Brooklyn	40.611009	-73.995180
22	Boerum Hill	16	2	Brooklyn	40.685683	-73.983748
25	Brighton Beach	13	2	Brooklyn	40.576825	-73.965094
30	Bulls Head	9	2	Staten Island	40.609592	-74.159409
31	Bushwick	18	2	Brooklyn	40.698116	-73.925258

	Neighborhood	Number of Restaurants	Cluster Labels	Borough	Latitude	Longitude
0	Allerton	4	1	Bronx	40.865788	-73.859319
1	Annadale	1	1	Staten Island	40.538114	-74.178549
2	Arrochar	5	1	Staten Island	40.596313	-74.067124
3	Arverne	2	1	Queens	40.589144	-73.791992
5	Astoria Heights	3	1	Queens	40.770317	-73.894680
6	Auburndale	5	1	Queens	40.761730	-73.791762
8	Battery Park City	6	1	Manhattan	40.711932	-74.016869

	Neighborhood	Number of Restaurants	Cluster Labels	Borough	Latitude	Longitude
9	Bay Ridge	25	0	Brooklyn	40.625801	-74.030621
12	Bayside	26	0	Queens	40.766041	-73.774274
19	Belmont	30	0	Bronx	40.857277	-73.888452
28	Brooklyn Heights	23	0	Brooklyn	40.695864	-73.993782
34	Carnegie Hill	20	0	Manhattan	40.782683	-73.953256
35	Carroll Gardens	23	0	Brooklyn	40.680540	-73.994654
38	Chelsea	23	0	Manhattan	40.744035	-74.003116
38	Chelsea	23	0	Staten Island	40.594726	-74.189560
42	Civic Center	24	0	Manhattan	40.715229	-74.005415
46	Clinton	24	0	Manhattan	40.759101	-73.996119

From the above we can determine that restaurants in cluster label 3 (yellow) are the most diverse, followed by 0 (red), 2 (blue) and 1 (purple), which has the lowest restaurant diversity.

5 Discussion

As one might assume from general knowledge, Manhattan is the borough with the most diverse neighborhoods in terms of restaurants, although there are pockets of neighborhoods throughout the other four boroughs that also contain high restaurant diversity. This information could be used by travelers who might hope to explore neighborhoods other than Manhattan, and want to travel to places with a large variety of cuisines to dine on as they do so.

As mentioned previously, this classification is not the best reflection of cultural or ethnic diversity because of the way the restaurants are categorized by the Foursquare API; for this to reflect true cultural diversity, further data cleaning would have been required. However, this could be used as a good general guide for which neighborhoods are the most gastronomically diverse.

6 Conclusion

This exploration can be seen as a preliminary exploration into the gastronomic diversity in New York City, and although the data could be cleaner we can already draw interesting conclusions. For example, travel advisory sites might want to introduce people to neighborhoods in Queens and Brooklyn as alternatives to those in Manhattan, if they were to want to explore neighborhoods rich in culinary diversity.