Tastes of New York
IBM Data Science Professional Capstone Project

Table of Contents

1. Introduction
New York City (NYC), also known as the City of New York or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. With an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial,] and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to the headquarters of the United Nations, New York is an important centre for international diplomacy. This makes it definitely an interesting place to open a restaurant. Yet, depending on the sources there are already 20 to 36.000 restaurants in NYC. So it's crucial to get it right.

Some of the most important decisions when opening a restaurant are about What (Which cuisine?) and Where?

To choose a Location and Lease a Commercial Space the following features are among the most important: Visibility and accessibility. Select a spot that can be seen by those driving or walking by. You should also look for an area that gets plenty of passers-by on foot or in cars. In addition, consider if there is parking and ease of access by foot or car. The demographics. Ensure the target market of your restaurant matches the demographics of the area. And competition is actually great as this proves there is a place where a certain demographic like to go and eat a certain type of food.

On the other hand, on of the most compelling reasons not to choose a restaurant in a certain neighbourhood is safety. Crime rates in New York City spiked in the 1980s and early 1990s as the crack epidemic hit,[1][2] but they have been dropping since 1991,[3] and, as of 2017, they are among the lowest of major cities in the United States. And though there is no objective reason to avoid NYC and its offers based on safety, there is a good rationale that people and potential clients might chose to ignore and avoid areas with higher crime areas than others.

Let's explore the diversity and business opportunities which come with opening a restaurant in a diverse city as NYC:

As quoted in an article - What Food Tells Us About Culture
"Traditional cuisine is passed down from one generation to the next. It also operates as an expression of cultural identity. Immigrants bring the food of their countries with them wherever they go and cooking traditional food is a way of preserving their culture when they move to new places."

2. Data

To examine the above said, following data sources will be used:

1. **New York City Dataset**
    - Link: https://geo.nyu.edu/catalog/nyu_2451_34572
    - Description: 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. This dataset will provide the addresses of neighborhood of NYC in json format. An extract of the json is as follows:
        - `{'type': 'Feature',`
        - `'id': 'nyu_2451_34572.306',`
        - `'geometry': {'type': 'Point',`
        - `'coordinates': [-74.08173992211962, 40.61731079252983 ]},`
        - `'geometry_name': 'geom',`
        - `'properties': {'name': 'Fox Hills',`
        - `'stacked': 2,`
        - `'annoline1': 'Fox',`
        - `'annoline2': 'Hills',`
        - `'annoline3': None,`
        - `'annoangle': 0.0,`
        - `'borough': 'Staten Island',`
        - `'bbox': [-74.08173992211962,`
        - `40.61731079252983,`
        - `-74.08173992211962, 40.61731079252983]}}`
2. **Foursquare API:**
    - Link: https://developer.foursquare.com/docs
    - Description: Foursquare API, a location data provider, will be used to make RESTful API calls to retrieve data about venues in different neighbourhoods. This is the link to Foursquare Venue Category Hierarchy. Venues retrieved from all the neighbourhoods are categorized broadly into "Arts & Entertainment", "College & University", "Event", "Food", "Nightlife Spot", "Outdoors & Recreation", etc. An extract of an API call is as follows: ``` 'categories': [{'id': '4bf58dd8d48988d110941735', 'name': 'Italian Restaurant', 'pluralName': 'Italian Restaurants', 'shortName': 'Italian', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/italian_', 'suffix': '.png'}, 'primary': True}], 'verified': False, 'stats': {'tipCount': 17}, 'url': 'http://eccorestaurantny.com', 'price': {'tier': 4, 'message': 'Very Expensive', 'currency'
3. **New York City Crime Database:** ```https://www1.nyc.gov/site/nypd/stats/crime-statistics/crime-statistics-landing.page
    - Description: New York City provides and maintains a detailed database on all their government actions, also relating to policing crime. This includes the location, purpose, GPS coordinates and type or crime, misdemeanour, felony tried or committed. For the purpose of this analysis we will focus on the number of committed felonies per neighbourhood

Download and Explore New York City Dataset

In order to segment the neighbourhoods of New York City, a dataset is required that contains the 5 boroughs and the neighbourhoods, that exist in each borough, with respective latitude and longitude coordinates.

This dataset exists for free on the web. Here is the link

- To the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572, and
- To its downloadable json format file: https://cocl.us/new_york_dataset/newyork_data.json
- `newyork_data key and value's length are:`
- `type 17`
- `totalFeatures 3`
- `features 124743`
- `crs 70`
- `bbox 74`

All the relevant data is in the *features* key, which is basically a list of the neighbourhoods. So, let's define a new variable that includes this data.

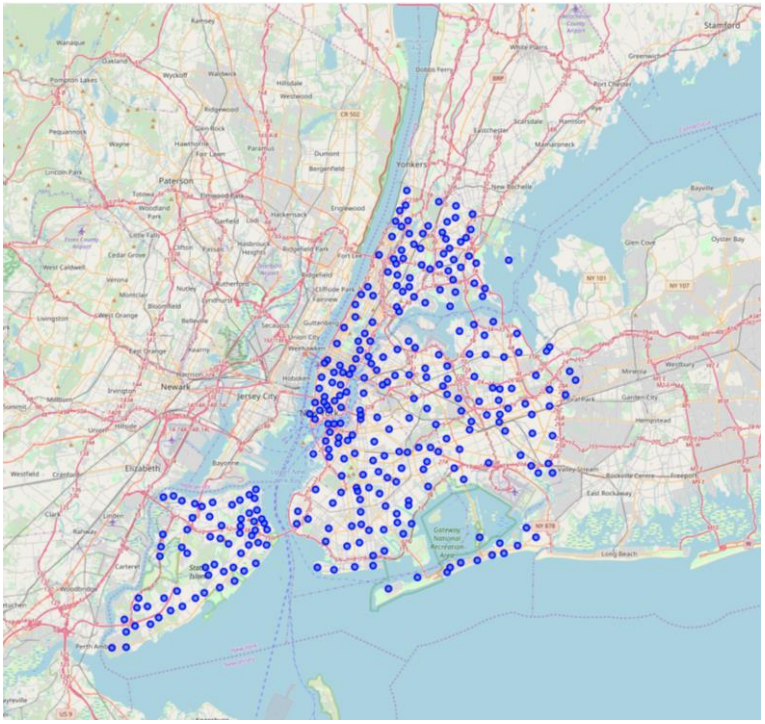
```
{'geometry': {'coordinates': [-74.08173992211962, 40.61731079252983],
              'type': 'Point'},
 'geometry_name': 'geom',
 'id': 'nyu_2451_34572.306',
 'properties': {'annoangle': 0.0,
                'annoline1': 'Fox',
                'annoline2': 'Hills',
                'annoline3': None,
                'bbox': [-74.08173992211962,
                         40.61731079252983,
                         -74.08173992211962,
                         40.61731079252983],
                'borough': 'Staten Island',
                'name': 'Fox Hills',
                'stacked': 2},
 'type': 'Feature'}
```

The resulting dataframe has 5 boroughs and 306 neighbourhoods:

	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

No we used Use geopy library to get the latitude and longitude values of New York City.
The geograpical coordinate of New York City are 40.7127281, -74.0060152.

Then we create a map of New York with neighbourhoods superimposed on top as per below:



Map of Now York City and its Neighbourhoods

We proceed to utilizing the Foursquare API to explore the neighbourhoods, its restaurants with different cuisines and follow up with segmenting them.

Food **& It's Sub-Categories**

We create a function to return a dictionary with Ids' & Names of *Food* & It's Sub-Categories. Now, we have all the categories in *Food* with their id's.

Explore the first neighbourhood to understand the results of GET Request


```
[{'id': '4c783cef3badb1f7e4244b54',
 'name': 'Carvel Ice Cream',
 'location': {'address': '1006 E 233rd St',
  'lat': 40.890486685759605,
  'lng': -73.84856772568665,
  'labeledLatLngs': [{'label': 'display',
    'lat': 40.890486685759605,
    'lng': -73.84856772568665}],
  'distance': 483,
  'postalCode': '10466',
  'cc': 'US',
  'city': 'Bronx',
  'state': 'NY',
  'country': 'United States',
  'formattedAddress': ['1006 E 233rd St',
   'Bronx, NY 10466',
   'United States']},
 'categories': [{'id': '4bf58dd8d48988d1c9941735',
   'name': 'Ice Cream Shop',
   'pluralName': 'Ice Cream Shops',
   'shortName': 'Ice Cream',
   'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/icecream_',
    'suffix': '.png'},
   'primary': True}],
 'referralId': 'v-1578825612',
 'hasPerk': False}]
```

As, our aim is to segment the neighbourhoods of NYC with respect to the *Food* in its vicinity. We need to proceed further to fetch this data from all the 306 neighbourhoods' venues.

Let's create a function to repeat the following process to all the neighbourhoods in NYC:

- Loop through neighbourhoods
  - Create the API request URL with radius=1000, LIMIT=50
  - Make the GET request
  - For each neighbourhood, return only relevant information for each nearby venue
  - Append all nearby venues to a list
- Unfold the list & append it to dataframe being returned

The *categoryId* parameter in the API request URL can be a comma separated string. So, lets create a comma separated string from *category___dict*.

3. Analysis & Machine Learning

Let's find out how many unique categories can be curated from all the returned venues

We find there are 176 unique categories:

Coffee Shop	1250	Vegetarian / Vegan Restaurant	30	Austrian Restaurant	6
Fast Food Restaurant	1092	Beer Garden	29	Soup Place	6
Donut Shop	1064	Snack Place	29	Farmers Market	5
Pizza Place	920	Bookstore	29	Pakistani Restaurant	5
Deli / Bodega	802	Greek Restaurant	29	Pet Café	5
Italian Restaurant	726	Sports Bar	28	Bowling Alley	5
Bakery	620	Convenience Store	27	Indie Movie Theater	5
American Restaurant	548	Udon Restaurant	27	Portuguese Restaurant	5
Bagel Shop	441	Theme Restaurant	25	Moroccan Restaurant	4
Diner	415	Supermarket	24	Market	4
Chinese Restaurant	408	Cuban Restaurant	23	Lebanese Restaurant	4
Café	378	Buffet	23	Cha Chaan Teng	4
Mexican Restaurant	303	Eastern European Restaurant	22	South American Restaurant	4
Fried Chicken Joint	284	Hotel	21	Food Service	4
Sandwich Place	265	Peruvian Restaurant	21	Pie Shop	3
Burger Joint	246	Wine Bar	20	Burmese Restaurant	3
Restaurant	245	Noodle House	20	Nightclub	3
Caribbean Restaurant	236	Cajun / Creole Restaurant	20	Swiss Restaurant	3
Food Court	215	Food & Drink Shop	19	Park	3
Seafood Restaurant	182	Flea Market	19	Australian Restaurant	3
Bar	173	Event Space	19	Hawaiian Restaurant	3
Ice Cream Shop	172	Tea Room	18	Music Venue	3
Sushi Restaurant	136	Taiwanese Restaurant	18	Cupcake Shop	3
Breakfast Spot	136	Gym / Fitness Center	18	Ethiopian Restaurant	3
Spanish Restaurant	136	Tapas Restaurant	17	Tibetan Restaurant	2
BBQ Joint	133	Burrito Place	16	Airport Food Court	2
New American Restaurant	131	Bistro	16	Modern Greek Restaurant	2
Dessert Shop	123	Falafel Restaurant	16	English Restaurant	2
Korean Restaurant	122	Comfort Food Restaurant	16	Gaming Cafe	2
Japanese Restaurant	115	Hot Dog Joint	16	Dosa Place	2
Asian Restaurant	115	Beer Bar	15	Hotel Bar	2
Taco Place	108	Ukrainian Restaurant	15	Shabu-Shabu Restaurant	2
Cocktail Bar	100	Shanghai Restaurant	15	Church	1

Grocery Store	95	Club House	14	Harbor / Marina	1
Thai Restaurant	94	Other Nightlife	13	Gift Shop	1
Bubble Tea Shop	91	Frozen Yogurt Shop	13	Gluten-free Restaurant	1
Food Truck	90	Health Food Store	13	Himalayan Restaurant	1
Ramen Restaurant	86	Lounge	13	Varenyky restaurant	1
Steakhouse	83	Poke Place	13	Fish Market	1
Food	82	Sri Lankan Restaurant	12	Warehouse Store	1
Wings Joint	81	Residential Building (Apartment / Condo)	12	Beach	1
Gastropub	81	Halal Restaurant	11	Dutch Restaurant	1
Latin American Restaurant	74	Gas Station	11	Speakeasy	1
Vietnamese Restaurant	70	Modern European Restaurant	11	Souvlaki Shop	1
Juice Bar	68	Colombian Restaurant	11	Casino	1
Southern / Soul Food Restaurant	67	Malay Restaurant	11	Hookah Bar	1
Dim Sum Restaurant	67	Skating Rink	10	Brazilian Restaurant	1
French Restaurant	64	Russian Restaurant	10	Soba Restaurant	1
Pub	60	African Restaurant	9	Pool	1
Gourmet Shop	50	Szechuan Restaurant	9	Dive Bar	1
Mediterranean Restaurant	50	Empanada Restaurant	8	Israeli Restaurant	1
Middle Eastern Restaurant	45	Tex-Mex Restaurant	8	Butcher	1
Salad Place	44	Jewish Restaurant	8	Kebab Restaurant	1
Indian Restaurant	42	Fish & Chips Shop	8	Churrascaria	1
Filipino Restaurant	38	Polish Restaurant	8	Flower Shop	1
Dumpling Restaurant	35	Beach Bar	8		
Hotpot Restaurant	33	Kosher Restaurant	8		
German Restaurant	32	Cafeteria	7		
Turkish Restaurant	32	Arepa Restaurant	7		
Cantonese Restaurant	31	Arcade	7		

As we are interested in exploring the diversity of the neighbourhood. Let's remove the generalized categories, like Coffee Shop, Cafe, etc.

```
Dim Sum Restaurant, American Restaurant, Souvlaki Shop, Spanish Restaurant, Shangha
i Restaurant, Speakeasy, Food Service, Austrian Restaurant, Udon Restaurant, Vietna
mese Restaurant, Malay Restaurant, Modern Greek Restaurant, Mediterranean Restauran
t, Pool, Moroccan Restaurant, South American Restaurant, English Restaurant, Tibeta
n Restaurant, Italian Restaurant, Tex-Mex Restaurant, Himalayan Restaurant, German
Restaurant, Thai Restaurant, Arepa Restaurant, Soba Restaurant, BBQ Joint, Fried Ch
icken Joint, Turkish Restaurant, Falafel Restaurant, Shabu-Shabu Restaurant, Harbor
/ Marina, Kebab Restaurant, Fast Food Restaurant, Southern / Soul Food Restaurant,
Burrito Place, Pakistani Restaurant, Australian Restaurant, Cha Chaan Teng, Casino,
Seafood Restaurant, Halal Restaurant, Mexican Restaurant, New American Restaurant,
Ramen Restaurant, Burmese Restaurant, Warehouse Store, Airport Food Court, Szechuan
Restaurant, Eastern European Restaurant, Kosher Restaurant, African Restaurant, Col
ombian Restaurant, Filipino Restaurant, Swiss Restaurant, Lebanese Restaurant, Chur
ch, Empanada Restaurant, Farmers Market, Chinese Restaurant, Varenyky restaurant, E
thiopian Restaurant, Greek Restaurant, Polish Restaurant, Hawaiian Restaurant, Cuba
n Restaurant, Jewish Restaurant, Cajun / Creole Restaurant, Brazilian Restaurant, I
sraeli Restaurant, Steakhouse, Taco Place, Deli / Bodega, Latin American Restaurant
, Gaming Cafe, Dosa Place, Peruvian Restaurant, Dumpling Restaurant, Sushi Restaura
nt, French Restaurant, Portuguese Restaurant, Tapas Restaurant, Taiwanese Restauran
t, Indian Restaurant, Middle Eastern Restaurant, Ukrainian Restaurant, Russian Rest
aurant, Caribbean Restaurant, Noodle House, Cantonese Restaurant, Churrascaria, Dut
ch Restaurant, Pizza Place, Asian Restaurant, Sri Lankan Restaurant, Korean Restaur
ant, Japanese Restaurant, Modern European Restaurant, Vegetarian / Vegan Restaurant
, Irish Pub
```

We combine these data with the Neighbourhood data and run a test on Wakefield and see to verity:

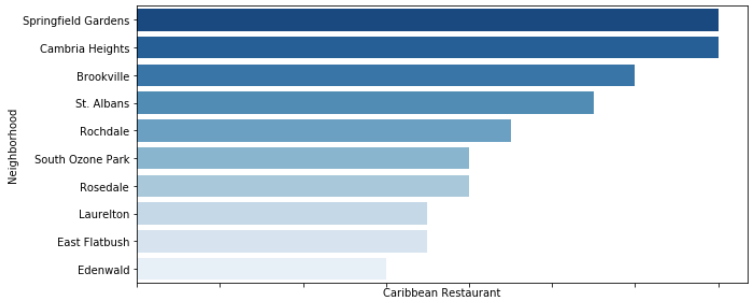
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
Wakefield	40.894705	-73.847201	Burger King	40.876911	-73.845978	Fast Food Restaurant
Wakefield	40.894705	-73.847201	Domino's Pizza	40.876564	-73.846898	Pizza Place
Wakefield	40.894705	-73.847201	Dunkin'	40.894054	-73.824016	Donut Shop
Wakefield	40.894705	-73.847201	Fairway Market	40.891762	-73.820376	Grocery Store

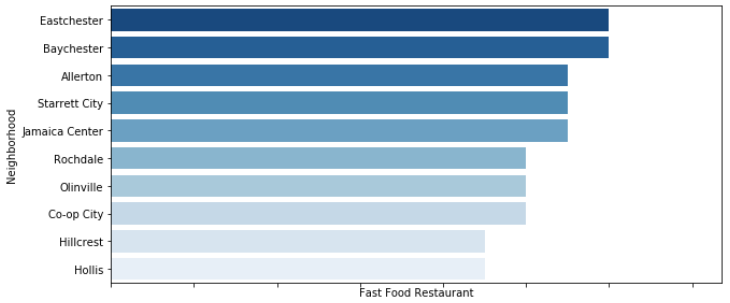
We can find that there are 99 unique FOOD categories within 1810 unique venues. We will proceed with top to bottom analysis now using these filtered Foursquare in c

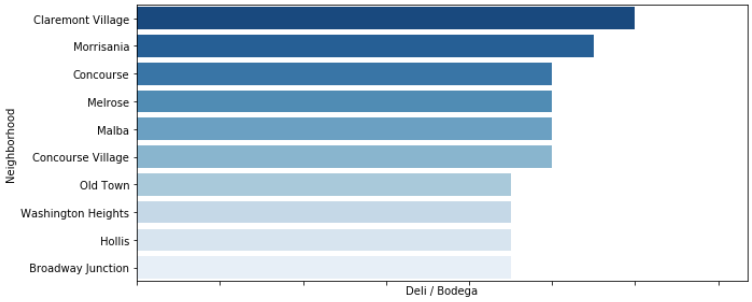
Analyse Each Neighbourhood

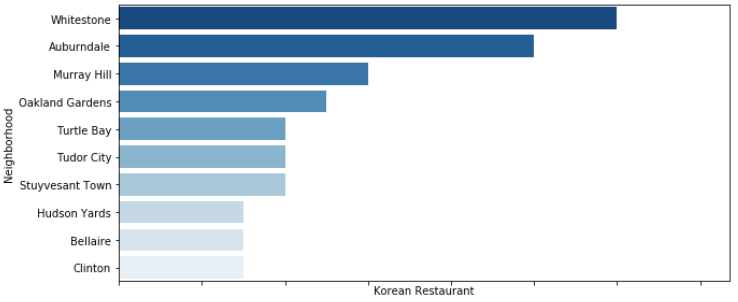
Let's find out the top 10 food categories in NYC. Not surpassingly the most wide-spread categories are Fast Food, Pizza and Deli/Bodega. Interestingly the main culturally driven categori es are Italian, American followed by Chinese. As the mean representation shows a huge deviatio n from the maximum per neighbourhood it will be interesting to follow up with a clustering and geographical representation to see if the are local hotspots.

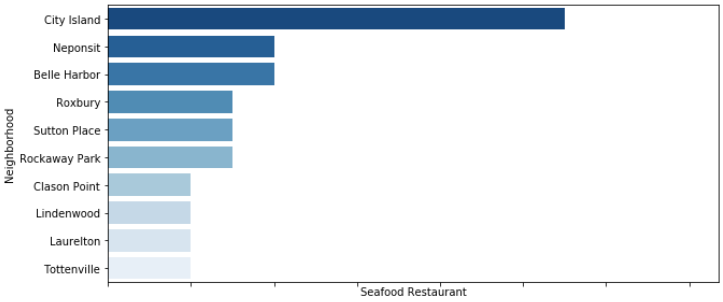
	mean	std	max
Fast Food Restaurant	13.2%	9.3%	40.0%
Pizza Place	11.6%	6.7%	33.3%
Deli / Bodega	9.4%	8.5%	38.7%
Italian Restaurant	9.1%	8.2%	37.0%
American Restaurant	7.7%	7.8%	66.7%
Chinese Restaurant	5.1%	5.4%	27.3%
Caribbean Restaurant	2.6%	6.2%	38.9%
Seafood Restaurant	2.4%	3.8%	40.7%
Sushi Restaurant	1.6%	3.3%	22.7%
Korean Restaurant	1.6%	4.7%	40.0%

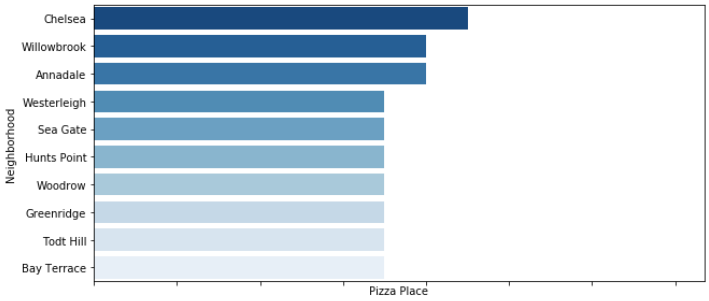


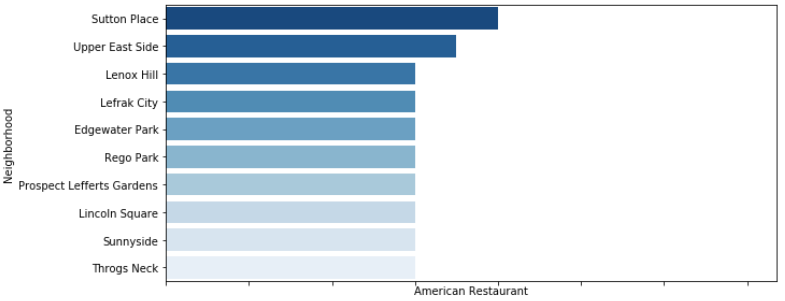


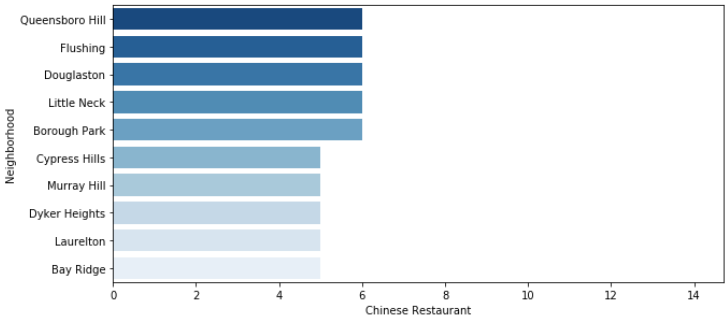


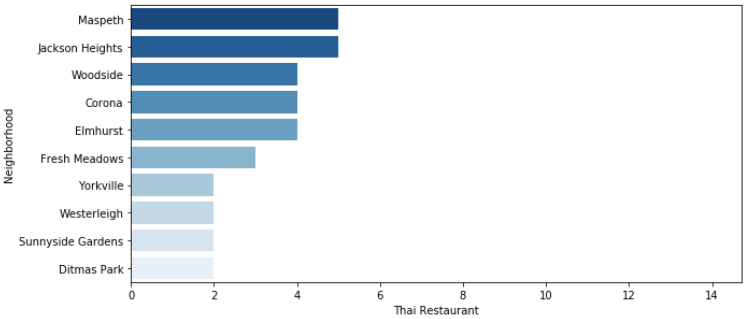




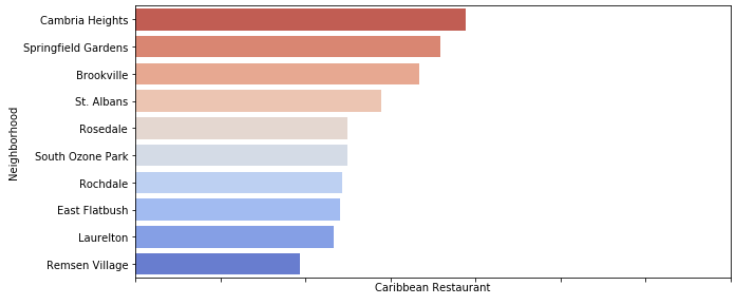


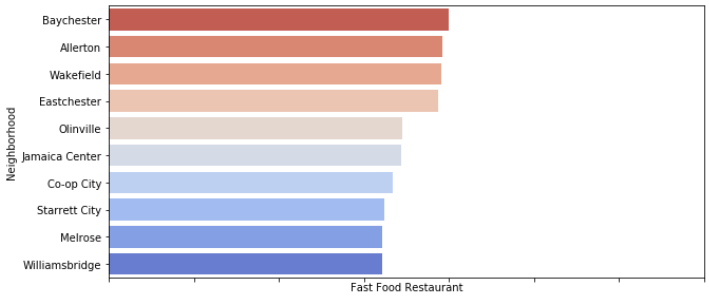


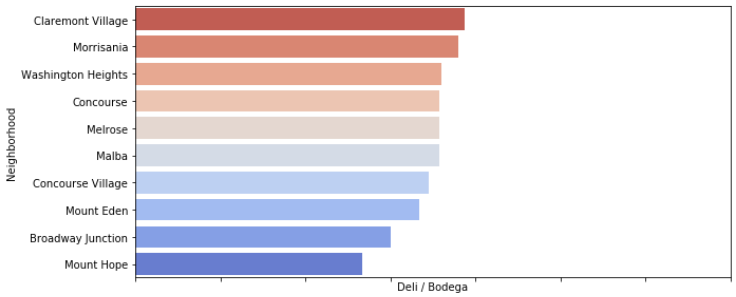


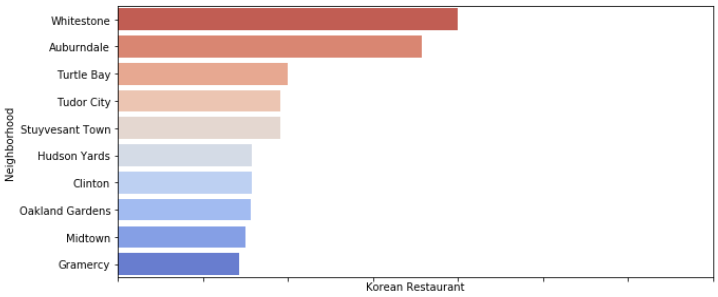


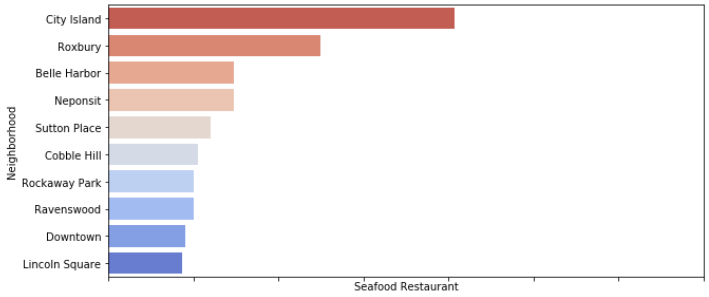
The comparison of numbers of restaurants per top category and how it's represented by Neighbourhood is already giving interesting details and could help deciding on the best location to choose. Yet, as there are different overall numbers of restaurant in general per location/neighbourhood, indexing and expressing the amount as percentage will help us further to refine the choice as per below graphs (see next page):

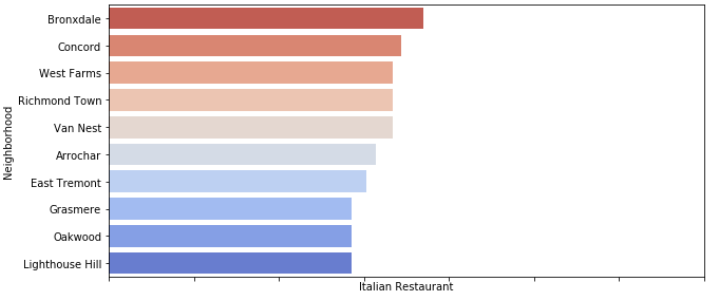


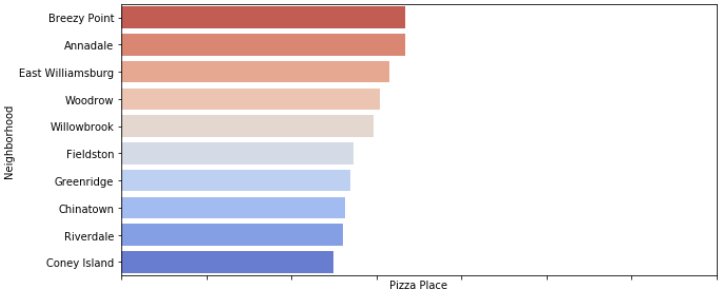


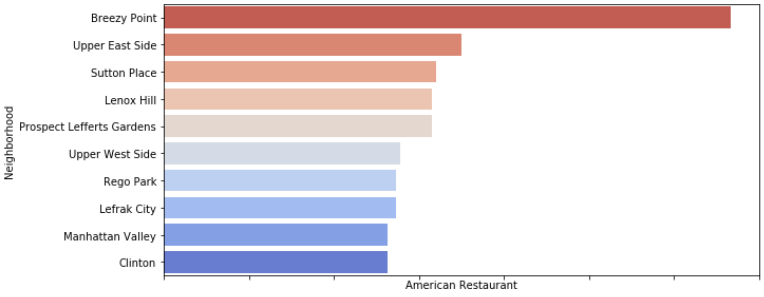


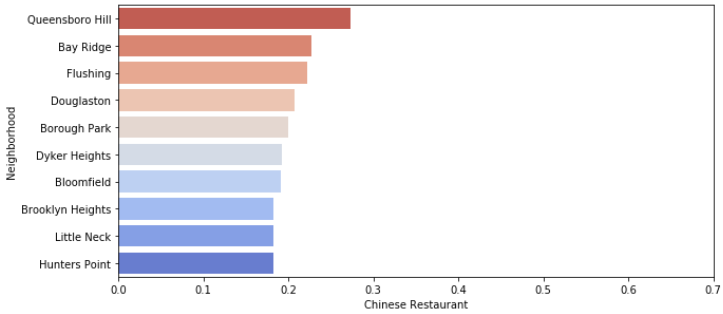


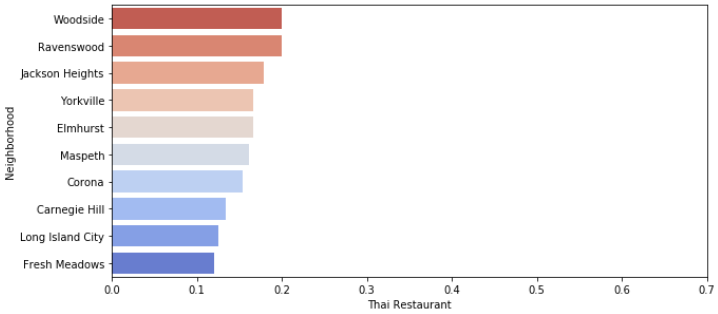












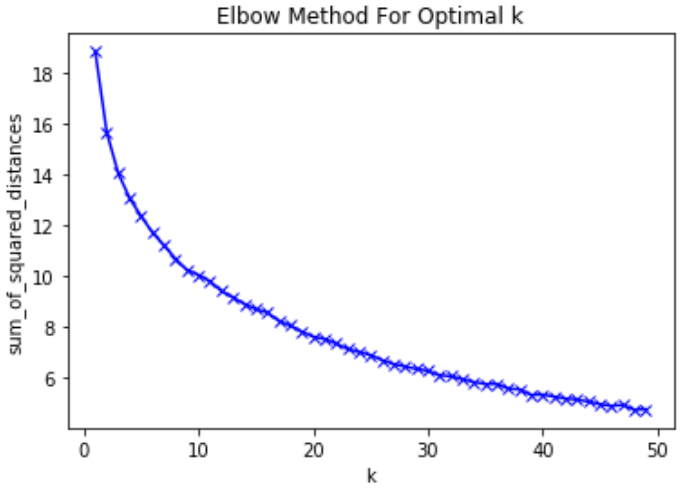
We can see that the differences ger clearer and more pronounced.

Clustering Restaurants in Neighbourhoods

We run *k*-means clustering to count Neighbourhoods for each cluster label for variable cluster size. This has been run on one-hot-encoded data for the restaurant categories, which have been represented as Euclidian distance in this case.

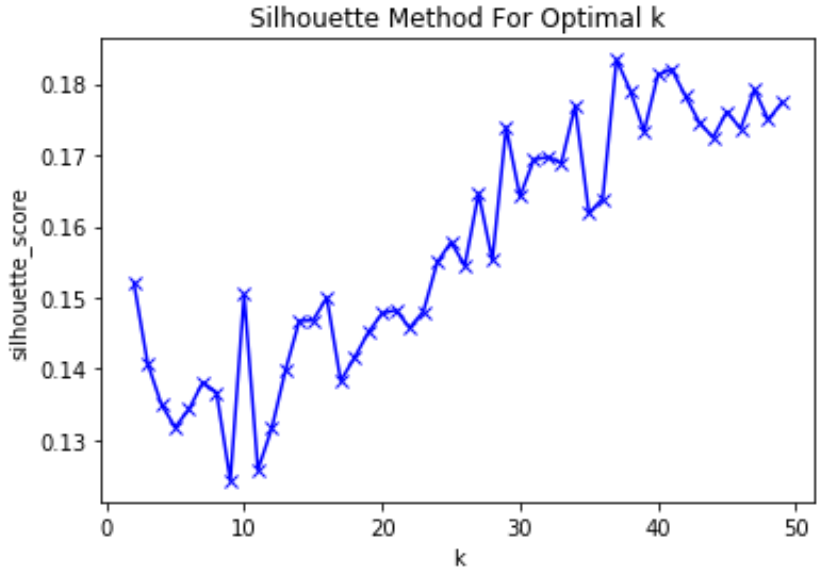
We proceed to firstly determine the optimal and most meaningful number of clusters for k-means clustering. Two methods were employed: Elbow and Silhouette:

The Elbow Method - calculate the sum of squared distances of samples to their closest cluster centre for different values of k. The value of k after which there is no significant decrease in sum of squared distances is chosen.



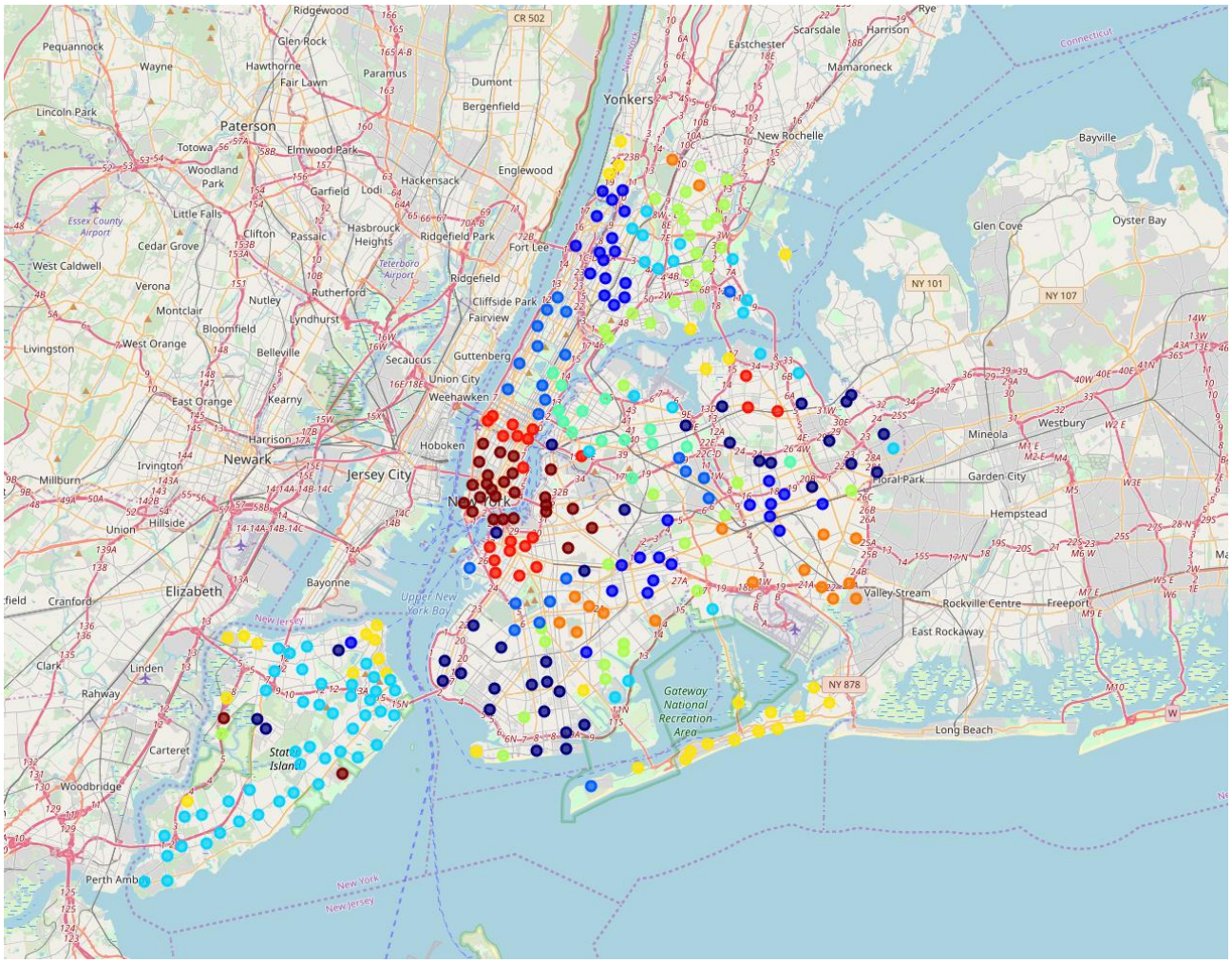
Elbow method does not seem to help us to determine the optimal number of clusters in this case.

The Silhouette Method - The silhouette value measures how similar a point is to its own cluster (cohesion) compared to other clusters (separation).



There is a local peak each at k = 2, k = 10. Two clusters will give a very broad classification of the venues. And we proceeded to chose k=10 as the next higher value of was just reached at 28 which doesn't seem to add enough data reduction in our opinion.

Below is a geographical representation of the different clusters. It reveals that most of the clusters also have a pronounced local density which would allow us to see a clear link between local culture, probably population/clients and the special cuisines offered most often:



Folium Map with different clusters marked in different colours

Detailed Description of the cluster

In the following a more detailed description and analysis of the clusters shall be conducted:

Cluster 1

Most Common Venue		2nd Most Common Venue		Boroughs	
American Restaurant	20	Fast Food Restaurant	5	Manhattan	11
Fast Food Restaurant	2	American Restaurant	5	Queens	8
Italian Restaurant	2	Italian Restaurant	5	Brooklyn	5
Chinese Restaurant	1	Pizza Place	4	Bronx	1
		Seafood Restaurant	2		
		Chinese Restaurant	1		
		Southern / Soul Food Restaurant	1		
		Steakhouse	1		
		Mexican Restaurant	1		

Cluster 2

Most Common Venue		2nd Most Common Venue		Boroughs	
Pizza Place	22	Italian Restaurant	13	Staten Island	18
American Restaurant	5	Pizza Place	9	Queens	9
Fast Food Restaurant	5	Fast Food Restaurant	5	Bronx	5
Italian Restaurant	4	Sushi Restaurant	3	Brooklyn	5
Sushi Restaurant	1	American Restaurant	3	Manhattan	1
Chinese Restaurant	1	Deli / Bodega	2		
		Chinese Restaurant	1		
		Dumpling Restaurant	1		
		Indian Restaurant	1		

Cluster 3

Most Common Venue		2nd Most Common Venue		Boroughs	
Deli / Bodega	21	Fast Food Restaurant	18	Bronx	13
Fast Food Restaurant	11	Deli / Bodega	10	Brooklyn	8
		Pizza Place	3	Queens	7
		Italian Restaurant	1	Manhattan	3
				Staten Island	1

Cluster 4

Most Common Venue		2nd Most Common Venue		Boroughs	
Fast Food Restaurant	39	Pizza Place	19	Bronx	21
Pizza Place	1	Deli / Bodega	5	Queens	10
		American Restaurant	5	Brooklyn	8
		Italian Restaurant	4	Staten Island	1
		Caribbean Restaurant	3		
		Fried Chicken Joint	2		
		Fast Food Restaurant	1		
		Mexican Restaurant	1		

Cluster 5

Most Common Venue		2nd Most Common Venue		Boroughs	
American Restaurant	5	American Restaurant	4	Brooklyn	10
Ramen Restaurant	3	Pizza Place	3	Manhattan	3
Southern / Soul Food Restaurant	2	Ramen Restaurant	2	Queens	2
Italian Restaurant	2	Chinese Restaurant	1		
Chinese Restaurant	1	Seafood Restaurant	1		
Mexican Restaurant	1	Mexican Restaurant	1		
BBQ Joint	1	BBQ Joint	1		
		Deli / Bodega	1		
		Thai Restaurant	1		

Cluster 6

Most Common Venue		2nd Most Common Venue		Boroughs	
Korean Restaurant	8	Korean Restaurant	5	Manhattan	10
American Restaurant	3	American Restaurant	5	Queens	3
Pizza Place	1	Fast Food Restaurant	1		
French Restaurant	1	Deli / Bodega	1		
		New American Restaurant	1		

Cluster 7

Most Common Venue		2nd Most Common Venue		Boroughs	
Deli / Bodega	17	Deli / Bodega	12	Queens	15
Pizza Place	9	Pizza Place	9	Staten Island	15
Italian Restaurant	4	Fast Food Restaurant	5	Bronx	3
Fast Food Restaurant	2	Chinese Restaurant	3	Brooklyn	3
Seafood Restaurant	2	Seafood Restaurant	3		
Spanish Restaurant	1	Italian Restaurant	3		
Caribbean Restaurant	1	American Restaurant	1		

Cluster 8

Most Common Venue		2nd Most Common Venue		Boroughs	
Fast Food Restaurant	9	Fast Food Restaurant	11	Brooklyn	18
Chinese Restaurant	8	Pizza Place	4	Queens	15
Thai Restaurant	4	Sushi Restaurant	3		
Turkish Restaurant	3	Chinese Restaurant	2		
Pizza Place	2	Dumpling Restaurant	2		
Sushi Restaurant	1	American Restaurant	2		
Korean Restaurant	1	Dim Sum Restaurant	2		
Seafood Restaurant	1	Thai Restaurant	2		
Middle Eastern Restaurant	1	Mediterranean Restaurant	1		
Fried Chicken Joint	1	Vietnamese Restaurant	1		
Italian Restaurant	1	Fried Chicken Joint	1		
Deli / Bodega	1	Italian Restaurant	1		
		Korean Restaurant	1		

Cluster 9

Most Common Venue		2nd Most Common Venue		Boroughs	
Caribbean Restaurant	7	Caribbean Restaurant	6	Queens	8
Fast Food Restaurant	7	Deli / Bodega	4	Brooklyn	5
Italian Restaurant	1	Fast Food Restaurant	4	Bronx	2
		Chinese Restaurant	1		

Cluster 10

Most Common Venue		2nd Most Common Venue		Boroughs	
Pizza Place	11	New American Restaurant	4	Manhattan	12
Dim Sum Restaurant	3	Pizza Place	4	Brooklyn	7
American Restaurant	2	American Restaurant	3		
Italian Restaurant	2	French Restaurant	2		
French Restaurant	1	Italian Restaurant	2		
		Dim Sum Restaurant	2		
		Chinese Restaurant	1		
		Taiwanese Restaurant	1		

Crime in Clusters and Neighbourhoods

At this point we want to come back to show how as of this point additional data can be used to further ease the decision which kind of restaurant to open at which location.

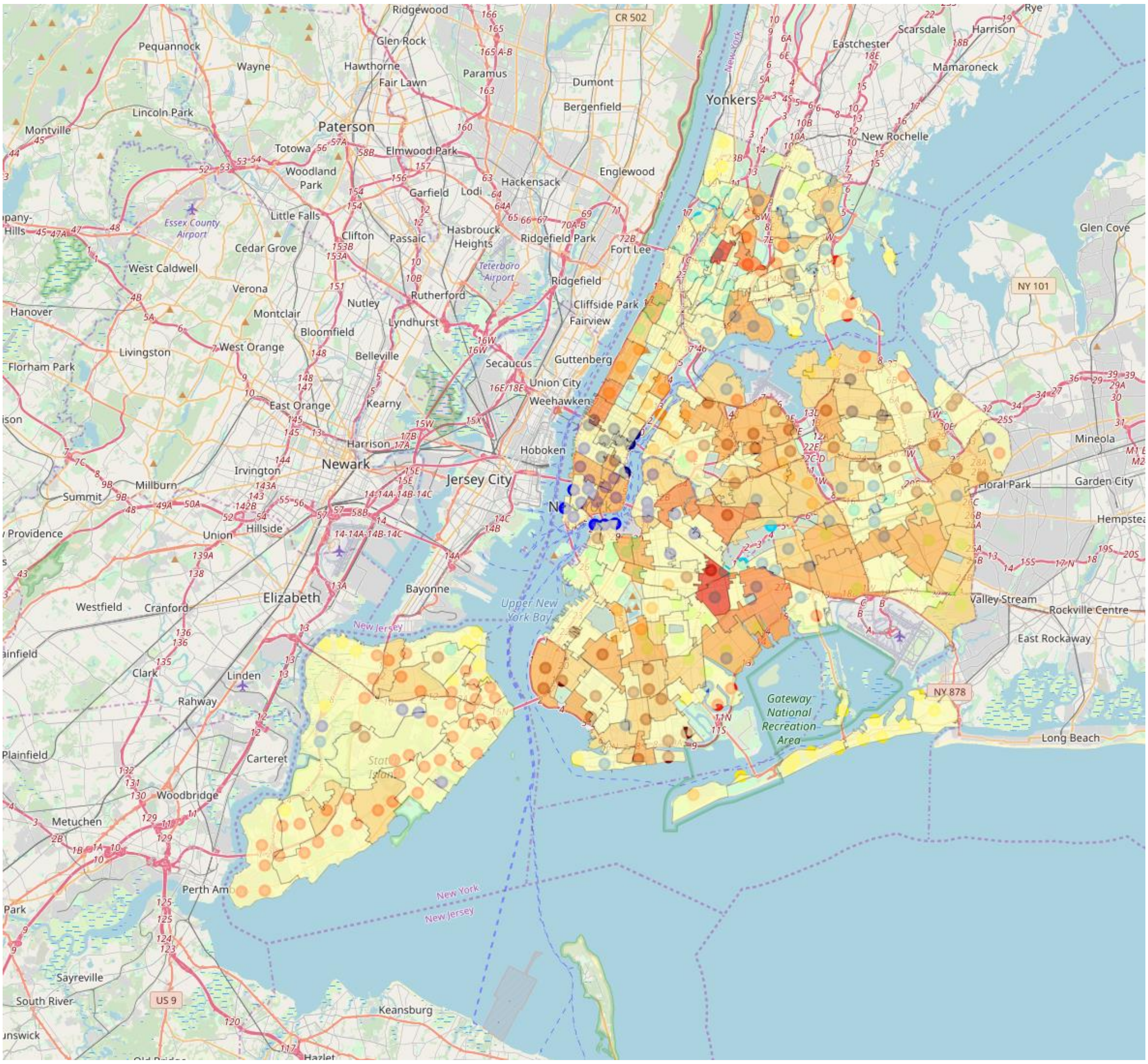
The New Work City Crime database was analysed for this purpose. This was filtered with a condition of committed felonies. Via nearest neighbour algorithms the locations (latitude, longitude) as given in the database where each assigned to each a cluster as per analysis as well as to the neighbourhoods to allow for a two-way interpretation and visual special representation of the data.

Below are the summarized committed felonies in 2018 per cluster showing a 20-fold difference between cluster 8 and cluster 4.

Cluster	Crime*
8	104
2	438
3	528
9	568
0	693
1	729
5	764
7	913
6	924
4	2158

*New York City Crime Database, Summarized Number of Felonies

Below is a geographical representation of the felonies per neighbourhood in 2018 via a coloured /choropleth map



Ridgewood
Glen Rock
Hawthorne
Fair Lawn
Paramus
Dumont
Bergenfield
Pequannock
Wayne
Lincoln Park
Montville
Paterson
Totowa
Woodland Park
Elmwood Park
Garfield
Lodi
Hackensack
Englewood
Little Falls
Clifton
Passaic
Hasbrouck Heights
Ridgefield Park
Fort Lee
Essex County Airport
Cedar Grove
West Caldwell
Verona
Teterboro Airport
Ridgefield
Cliffside Park
Fairview
Nutley
Rutherford
Lyndhurst
Montclair
Bloomfield
Hanover
Livingston
West Orange
Belleville
Secaucus
Guttenberg
Florham Park
Union City
Weehawken
East Orange
Kearny
Harrison
Hoboken
Newark
Irvington
Jersey City
Millburn
Summit
Hillside
Union
Elizabeth
Bayonne
Westfield
Cranford
Upper New York Bay
New Jersey
Clark
Linden
Rahway
Plainfield
Carteret
Woodbridge
Metuchen
Perth Amboy
Sayreville
South River
Keansburg
Hazlet
Yonkers
Eastchester
Scarsdale
Harrison
Rye
Mamaroneck
New Rochelle
Glen Cove
Mineola
Floral Park
Garden City
Hempstead
Valley Stream
Rockville Centre
East Rockaway
Long Beach
Gateway National Recreation Area
New York
CR 502
NY 101
NY 878
US 9

6. Conclusions

The data on NYC from its neighbourhoods and combined with the data on restaurants in those, revealed a very diverse culture expressed through its diverse cusine.

However, most probably driven by local density of population groups and clients looking restaurant in specific neighbourhoods, there are very clusters of different cuisines which are focussed also geographically.

It would be clearly recommended to choose a range of cluster based on the cuisine to remain within target group and expectations, (e.g. seafood specialised restaurant in less urban, closer to the sea areas.)

Whilst there are multiple factors to influence the success of a restaurant, as mentioned in the introduction, there are some main location factors.

People's safety and the perception thereof is a big driver of footfall and revenue or lack thereof. it could be demonstrated that there is a great difference between the identified clusters as well as neighbourhoods as far as crime and thus most probably resulting client numbers and revenue are concerned.

Thus the results of this analysis could serve well to anyone planning to open a restaurant in NYC.

Further work and analysis could be undertaken including income, population density and other measure apart from crime and also related them via regression analysis to the ratings and visiting numbers of current restaurants as expression of their success.

This was out-of-scope for this analysis as the Foursquare API as very strict limitations for private accounts on the daily retrievable data.