TASTES OF NEW YORK

IBM Data Science Professional – Capstone Project

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.
- 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?
  - Where?

INTRODUCTION

- 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.
    - 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
    - Dropping since 1991, and, as of 2017, they are among the lowest of major cities in the United States.
    - 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 and avoid the potential pitfalls!
- 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."

DATA

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

1. New York City Dataset (geo.nyu.edu/catalog/nyu_2451_34572)

- 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:

2. Foursquare API: (developer.foursquare.com/docs)

- 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.

3. New York City Crime Database: (www1.nyc.gov/site/nypd/stats/crime-statistics/crime-statistics-landing.page)

- 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

NEIGHBOURHOODS

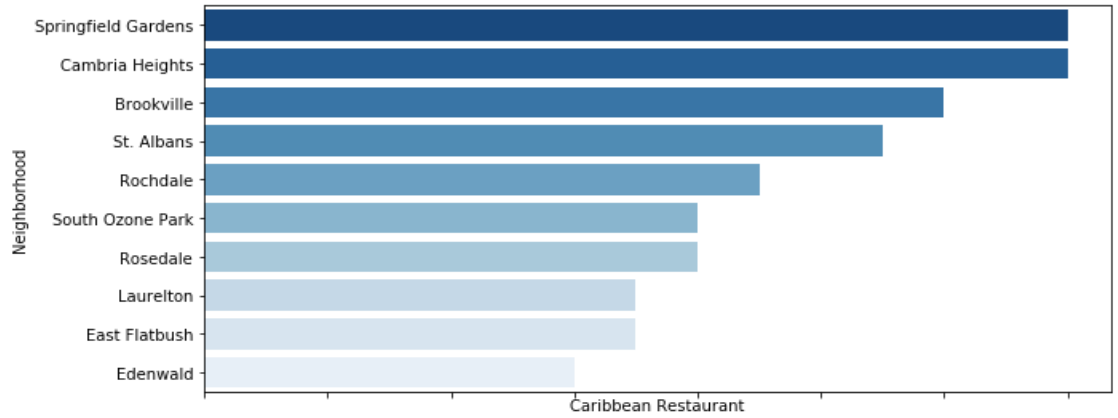
- In order to segment the neighbourhoods of New York City, a dataset is required that contains the 5 boroughs and the neighbourhoods, that exist in each borough, with respective latitude and longitude coordinates.

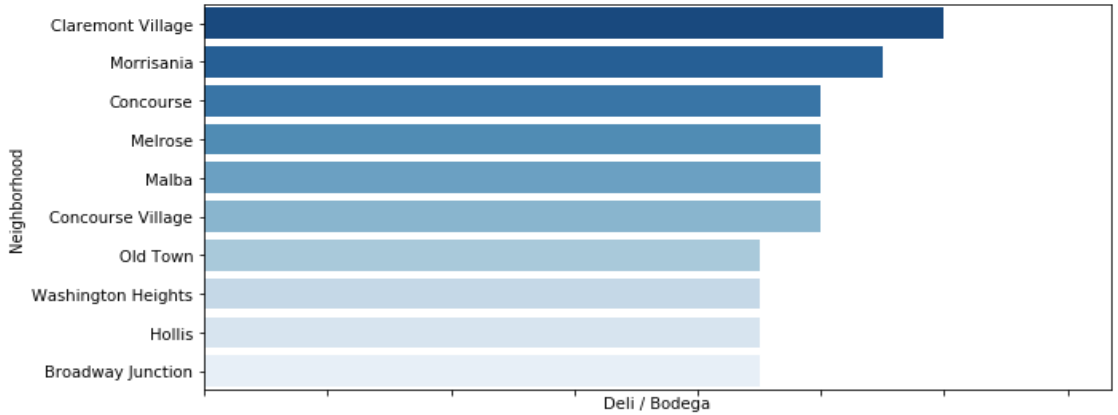
RESTAURANTS

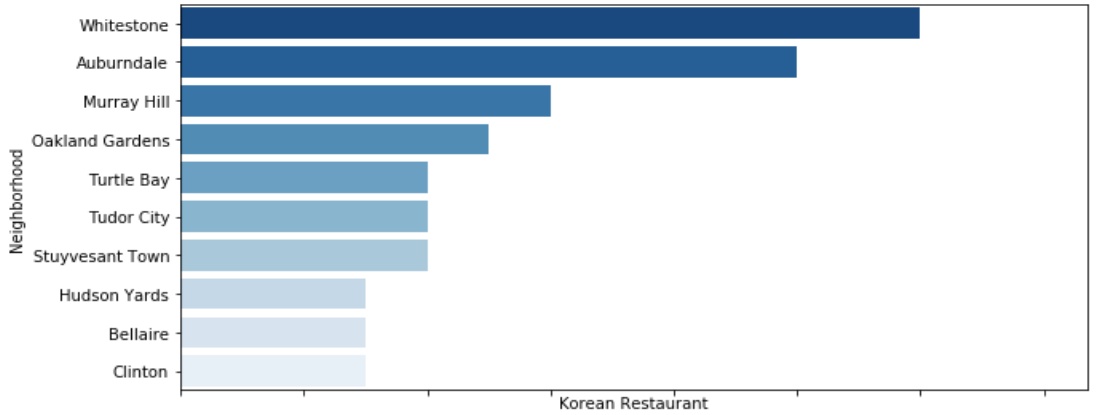
- Let's find out the top 10 food categories in NYC.
- Not surpassingly the most wide-spread categories are Fast Food, Pizza and Deli/Bodega.
- Interestingly the main culturally driven categories are Italian, American followed by Chinese.
- As the mean representation shows a huge deviation from the maximum 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%

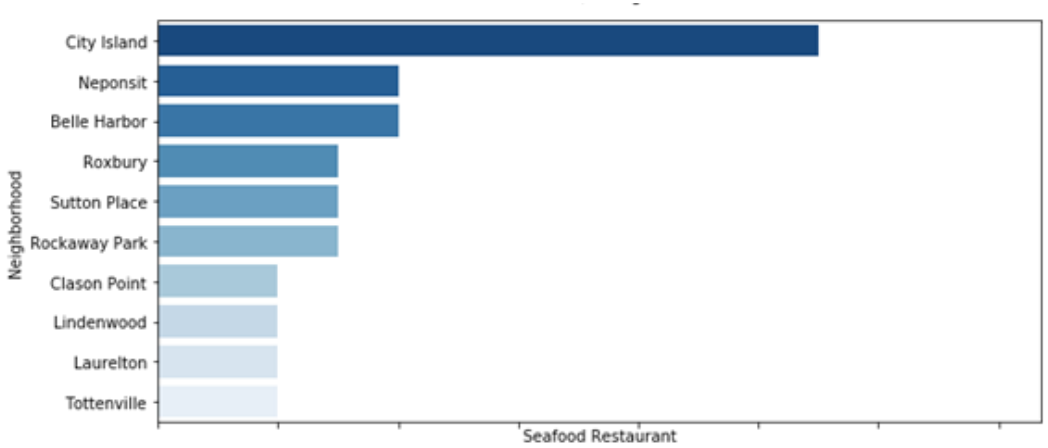
RESTAURANTS CUISINE PER NEIGHBOURHOODS

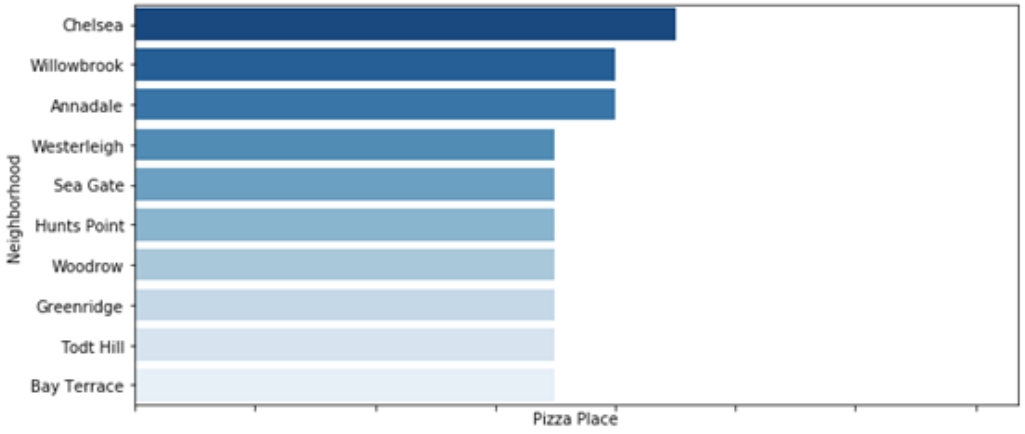


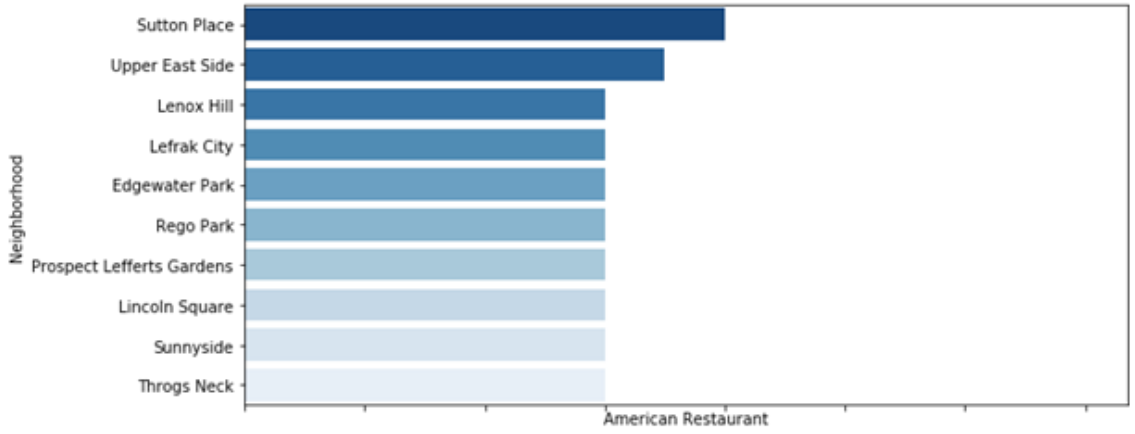




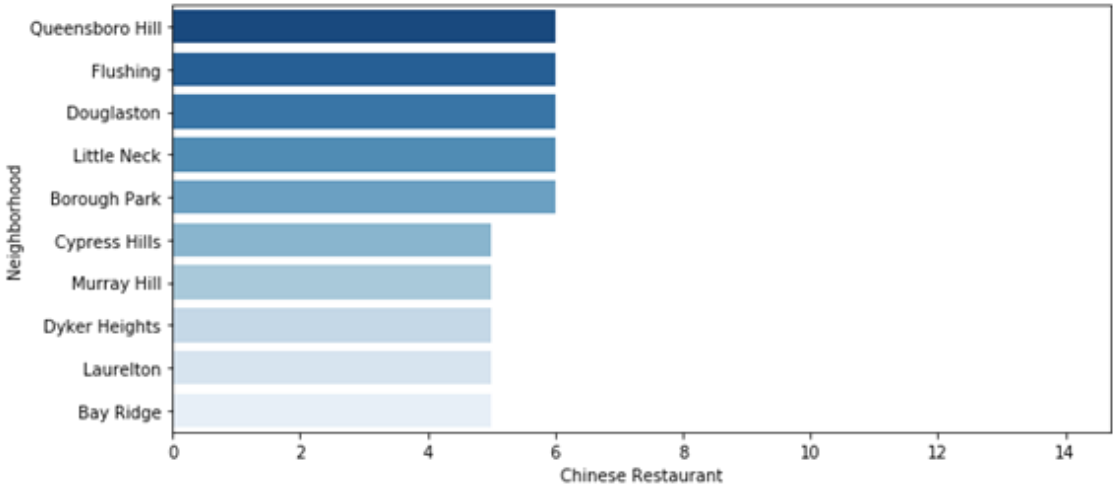
RESTAURANTS CUISINE PER NEIGHBOURHOODS

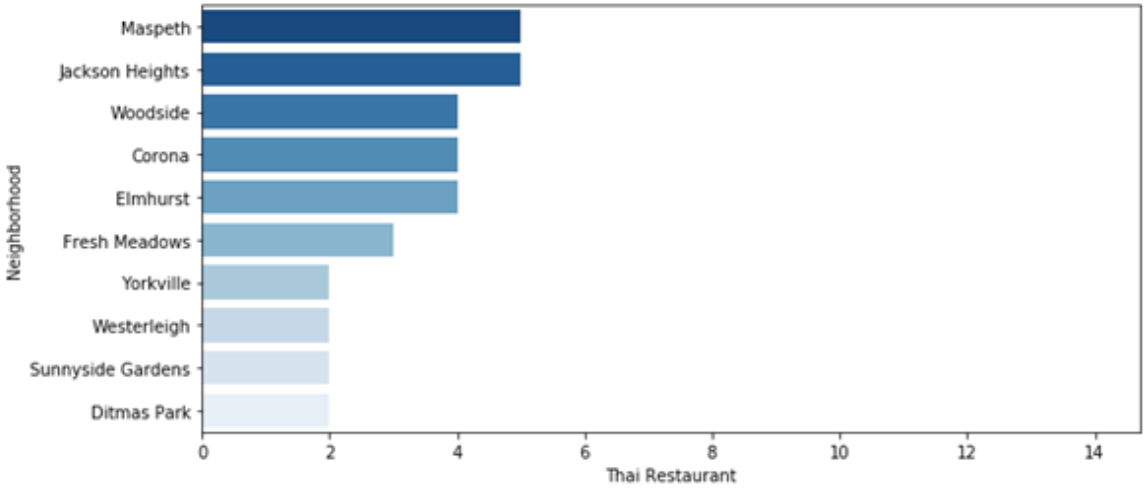




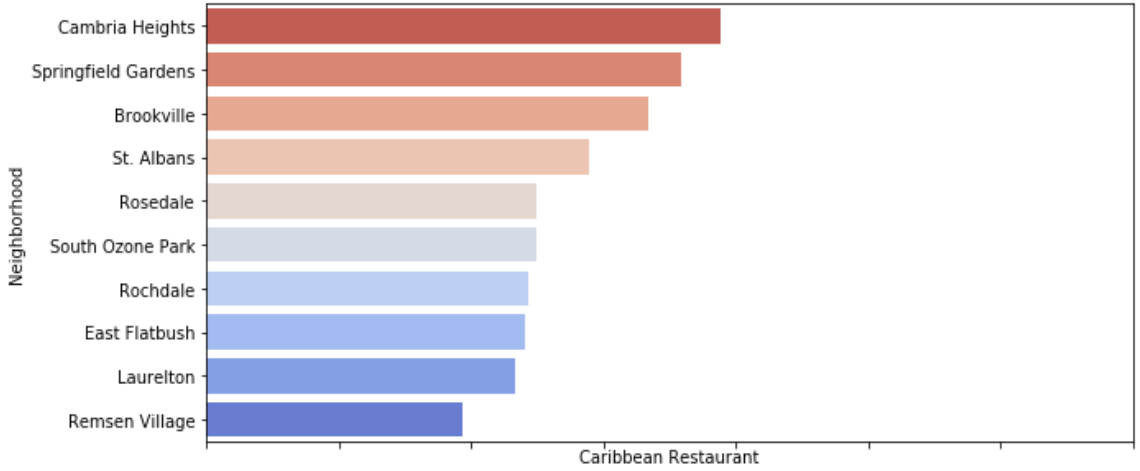


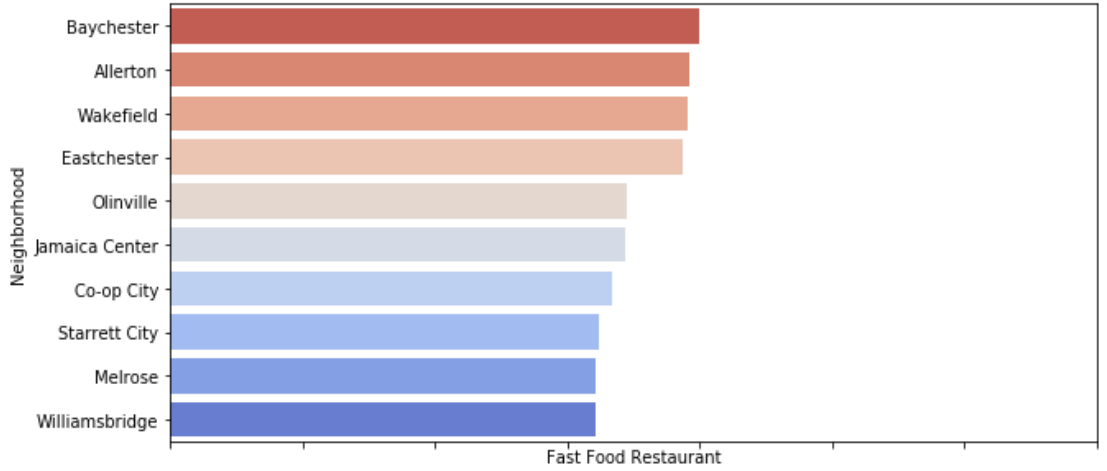
RESTAURANTS CUISINE PER NEIGHBOURHOODS

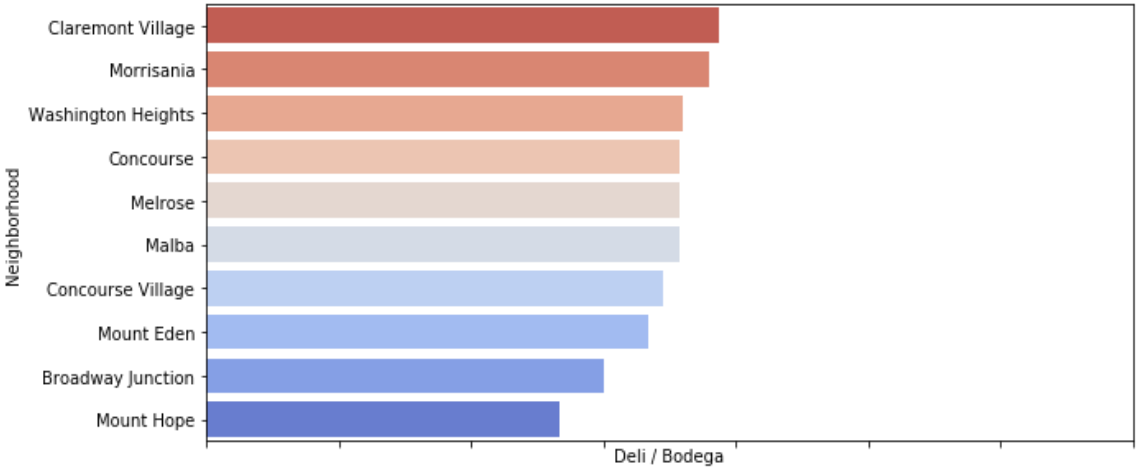


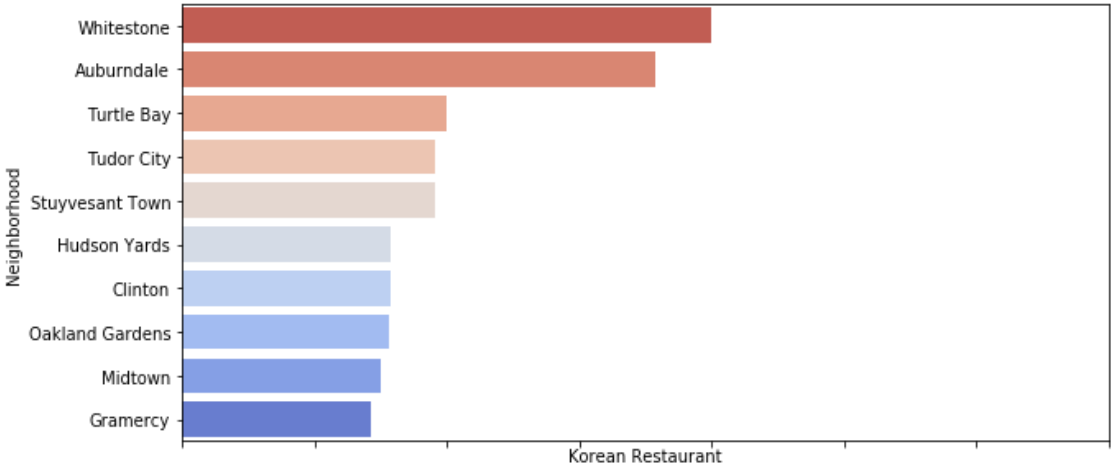


RESTAURANTS CUISINE IN % PER NEIGHBOURHOODS

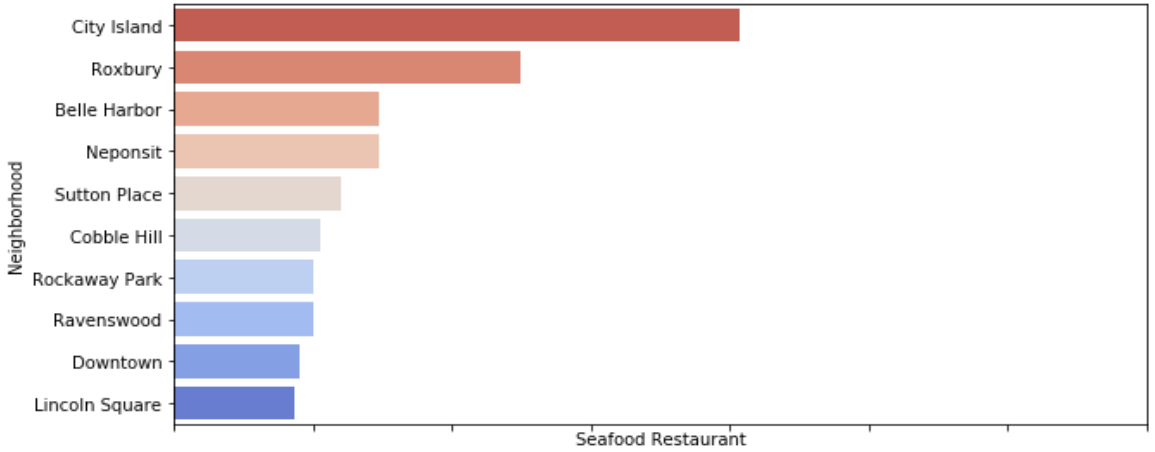


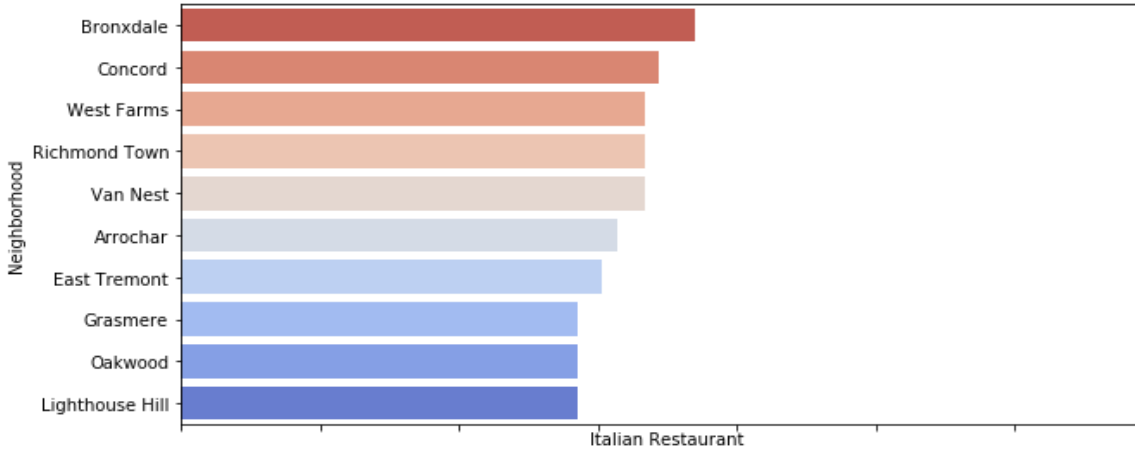


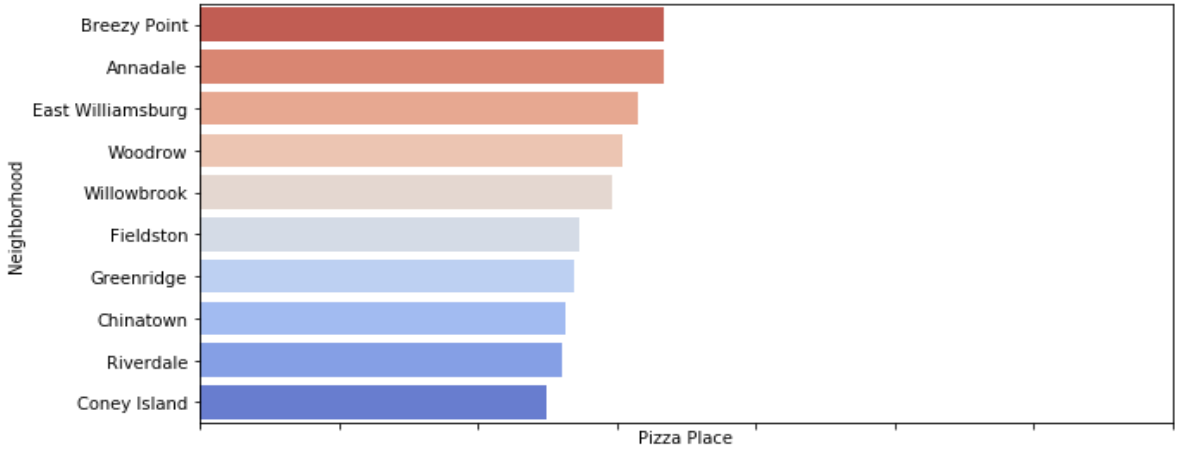


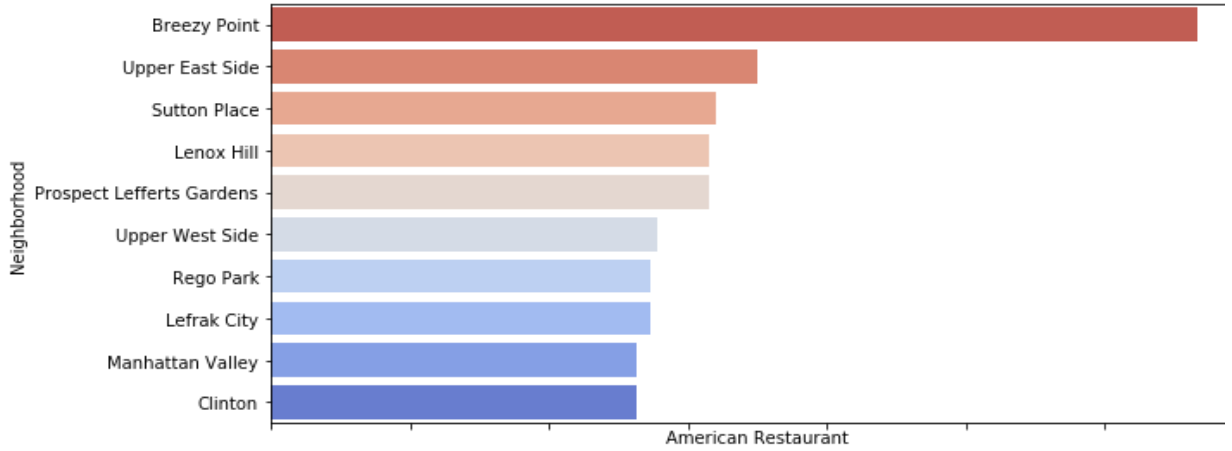


RESTAURANTS CUISINE IN % PER NEIGHBOURHOODS

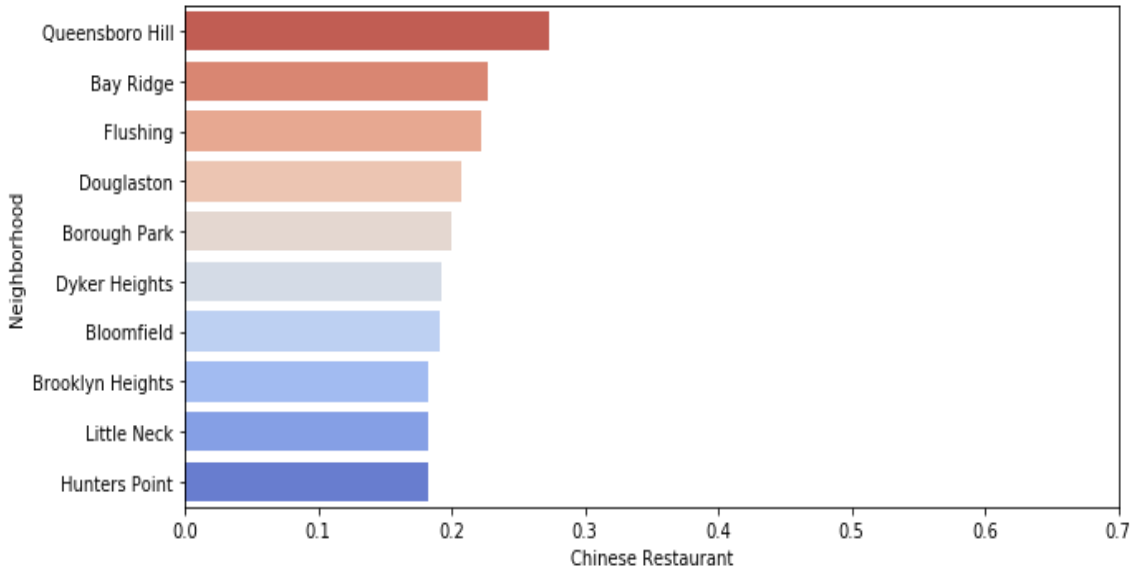


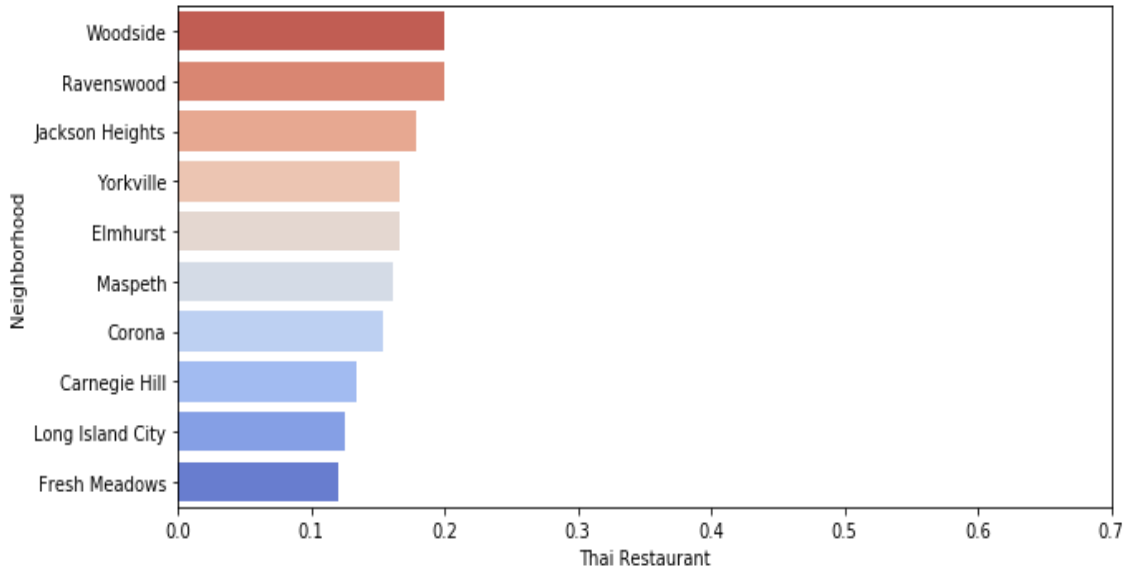






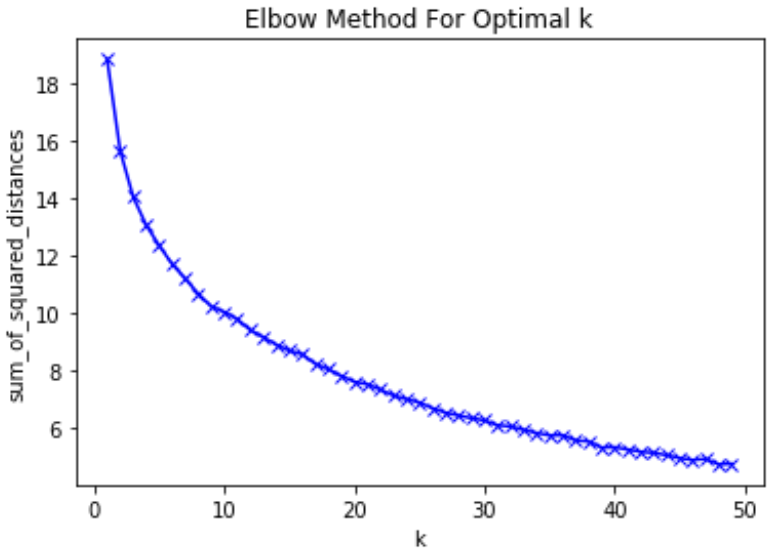
RESTAURANTS CUISINE IN % PER NEIGHBOURHOODS





CLUSTERING RESTAURANTS IN NEIGHBOURHOODS

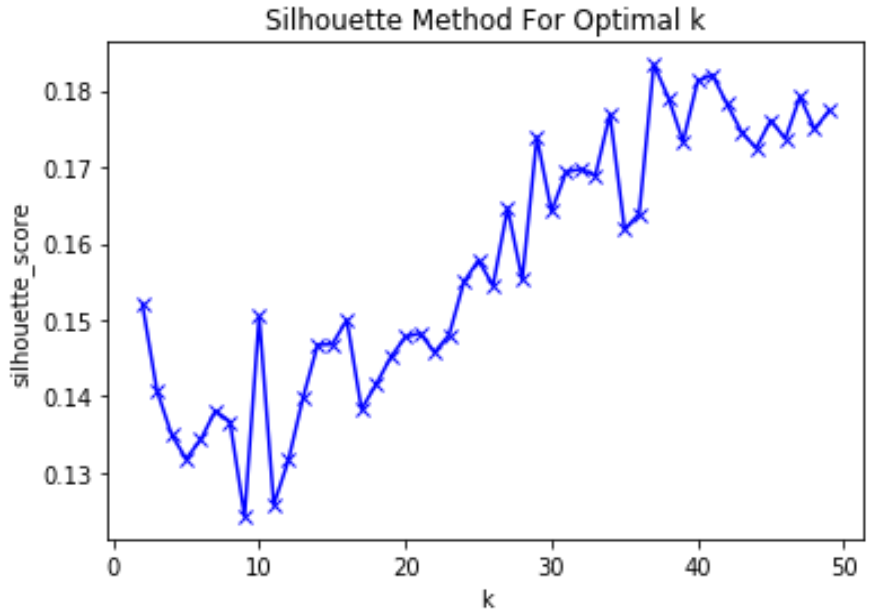
- We run *k*-means clustering to count Neighbourhoods for each cluster label for variable cluster size. This has been run on one-hot-encoded data for the restaurant categories, which have been represented as Euclidian distance in this case.
- We proceed to firstly determine the optimal and most meaningful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



CLUSTERING RESTAURANTS IN NEIGHBOURHOODS

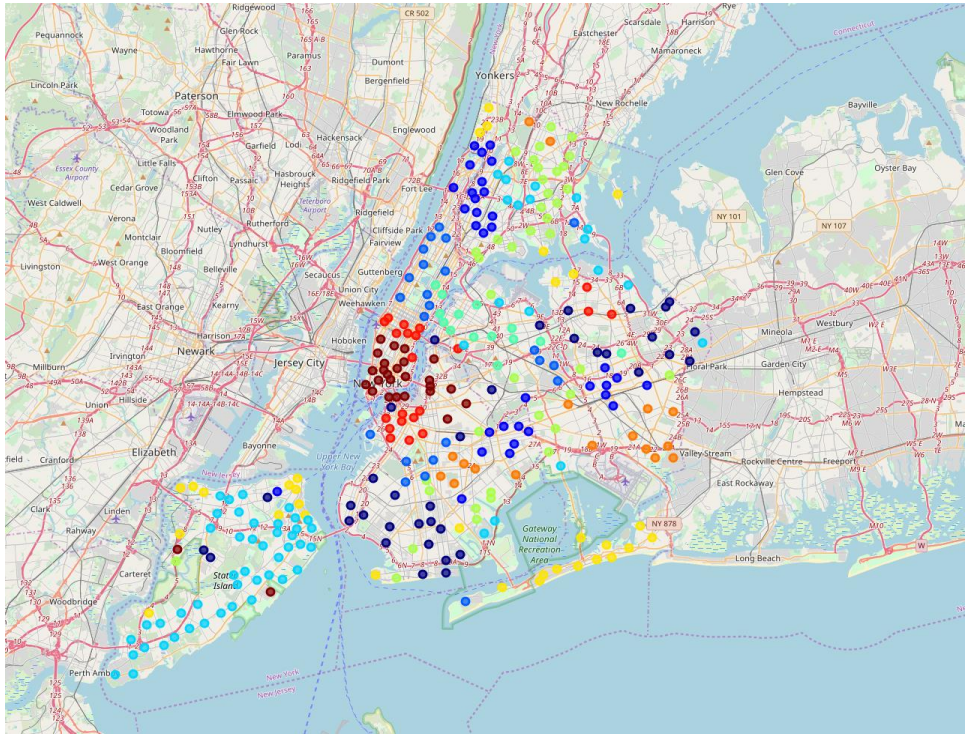
- **The Silhouette Method** - The silhouette value measures how similar a point is to its own cluster (cohesion) compared to other clusters (separation).

- 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.
- We go with a 10 cluster solution



CLUSTER MAPPING

- 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:



CLUSTER DESCRIPTION

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 DESCRIPTION

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 DESCRIPTION

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

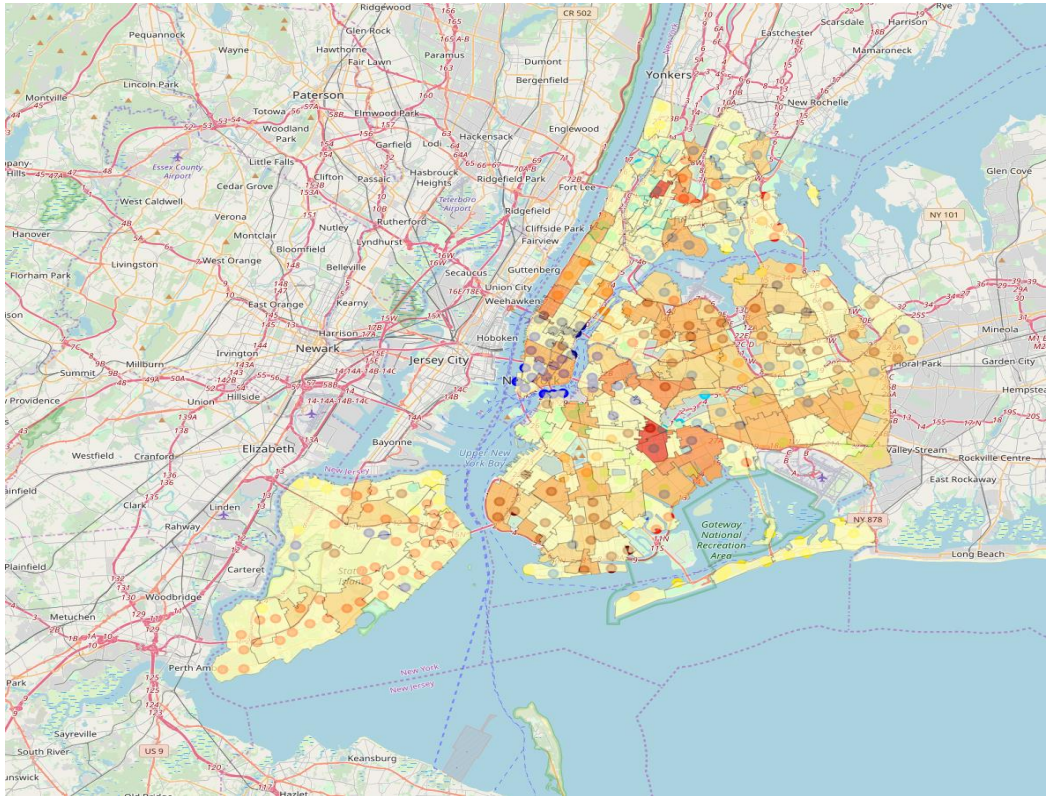
- At this point we want to come back to show how as of this point additional data can be used to further ease the decision which kind of restaurant to open at which location.
- The New Work City Crime database was analysed for this purpose. This was filtered with a condition of committed felonies. Via 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*

CRIME IN CLUSTERS AND NEIGHBORHOODS

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



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 cuisine.
- 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.