

## Introduction |

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an restaurant and school in New York, United States.

Since there are lots of restaurants in New York we will try to detect locations that are not already crowded with restaurants. We choose some candidate location in Queens New York city. We want to get the cluster information about the Center Queens, so that we can analyze the cluster.

Secondly, it is important that analyze the distribution of the restaurant type in each cluster. We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed and get the cluster character, so that best possible final location and restaurant type can be chosen by stakeholders. So, we want to explore the center candidate location that belongs to the restaurant type.

### Data Acquisition And Cleaning

Based on definition of our problem, factors that will influence our decision are:

number of existing restaurants in the neighborhood (any type of restaurant)

number of and distance to Italian restaurants in the neighborhood, if any

distance of neighborhood from city center

number of school in the neighborhood (any type of school)

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods. This is our data flow:

Get the Queens geometry information

Calculate candidate geometry information, under the conditions, like ~6 km from Queens center, and each are has 600 meters each circle apart

Get the detailed information in each circle by using Foursquare API Besides we use other packages, likes: Pandas Numpy Json Geopy Matplotlib Shapely Pyproj Sklearn Folium Requests We get the basic information like:

	Address	Latitude	Longitude	X	Y	Distance from center
0	93-46 210th Pl, Queens Village, NY 11428	40.716520	-73.751049	-5.820665e+06	9.836718e+06	5992.495307
1	211-30 90th Ct, Jamaica, NY 11428	40.720054	-73.750895	-5.820065e+06	9.836718e+06	5840.376700
2	89-28 213th St, Queens Village, NY 11427	40.723589	-73.750740	-5.819465e+06	9.836718e+06	5747.173218
3	214-46 Whitehall Terrace, Jamaica, NY 11427	40.727124	-73.750586	-5.818865e+06	9.836718e+06	5715.767665
4	21B-17 Grand Central Pkwy, Jamaica, NY 11427	40.730659	-73.750431	-5.818265e+06	9.836718e+06	5747.173218
5	220-24 Hartland Ave, Jamaica, NY 11427	40.734194	-73.750276	-5.817665e+06	9.836718e+06	5840.376700
6	220-72 77th Ave, Flushing, NY 11364	40.737730	-73.750122	-5.817065e+06	9.836718e+06	5992.495307

## Methodology

After extraction, we dive into exploring the data. We use the PCA method to reduce dimensions and analyze the two main components KDE information. Finally, we use the KMeans method and Silhouette metric to create the cluster. **Caution**, we preprocessing the feature *Distance from center* by *StandardScaler* method.

Distance from center	Thai Restaurant	Mediterranean Restaurant	Cantonese Restaurant	Sushi Restaurant	Latin American Restaurant	Mexican Restaurant	Southern / Soul Food Restaurant	American Restaurant	Filipino Restaurant	...	New American Restaurant	Chinese Restaurant	Tapas Restaurant	Indian Restaurant	Dumpling Restaurant	Halal Restaurant	Elementary School	Szechuan Restaurant	All Restaurant	AISSc
1.400281	0	0	0	2	1	2	1	0	0	...	0	5	0	0	0	0	0	0	1.099931	
1.292965	0	0	0	2	1	2	1	0	0	...	0	4	0	0	0	1	0	0	0.973840	

## Result And Conclusion

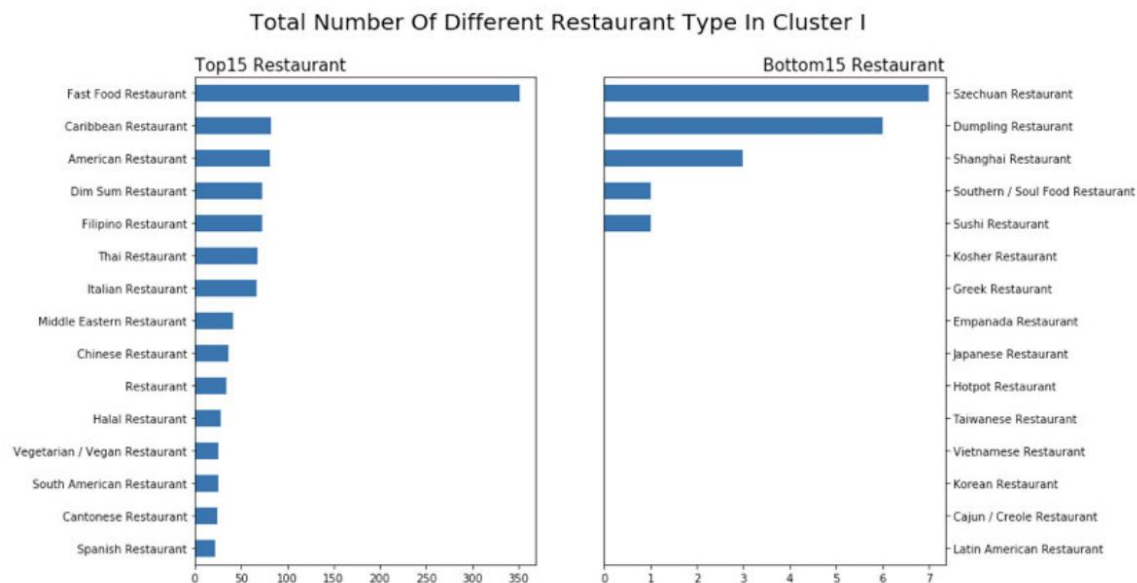
We can explore the cluster information, so that we can get the initial conclusion.

### Cluster I Candidate Location

Fast Restaurant is main type

Western Restaurants have the largest market

Eastern Restaurants have the few market



### Final conclusion

We can make a choice that choose the appropriate restaurant type at appropriate candidate location. Like that the Chinese Breakfast restaurant is good choice at Cluster III candidate location