# 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**, Unite 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:

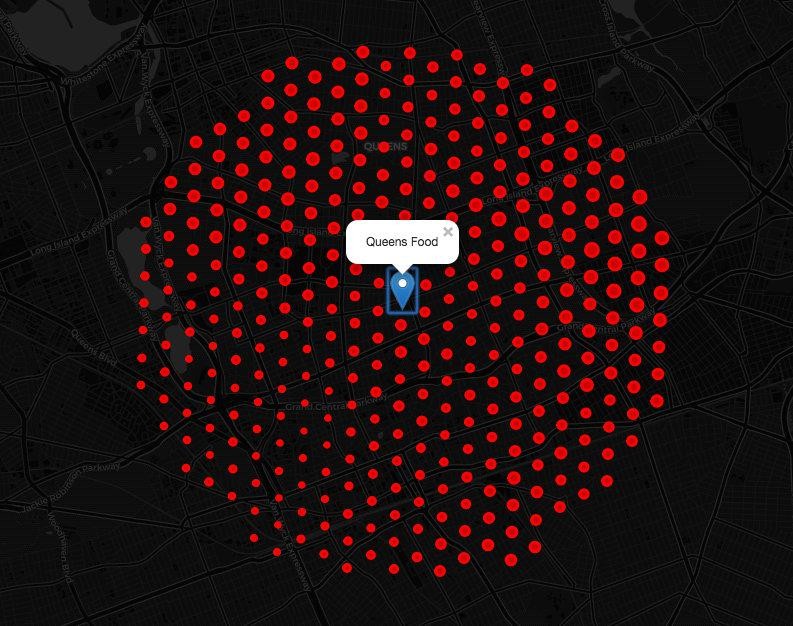
Pandas Numpy Json Geopy Matplotlib Shapely Pyproj Sklearn Folium Requests

We get the basic information like:

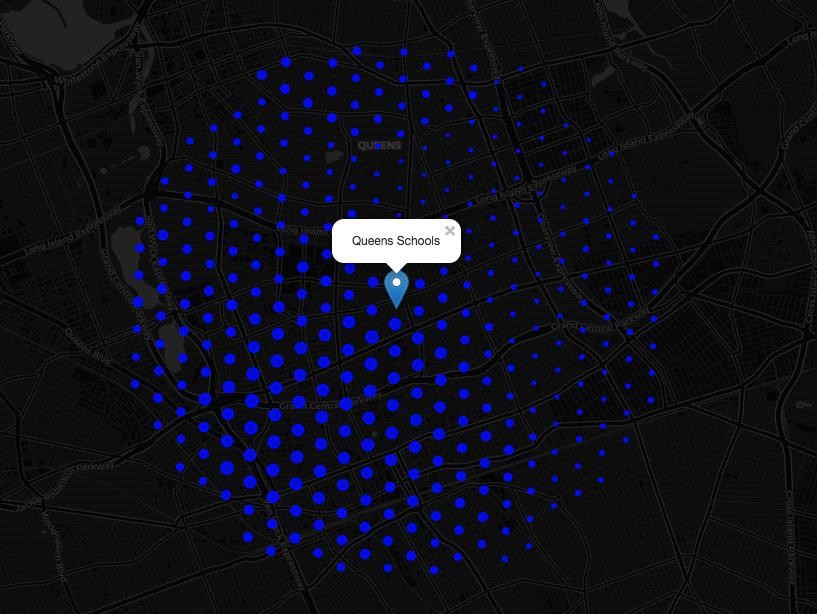


# Methodology

We create the data ﬂow to get the data by using API, then extract the schools and the restaurants information. Right now, we can visualize the distribution of the number of restaurants and schools:



The restaurants is almost uniform in the each candidate location. Wherase, the school is uniform.



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



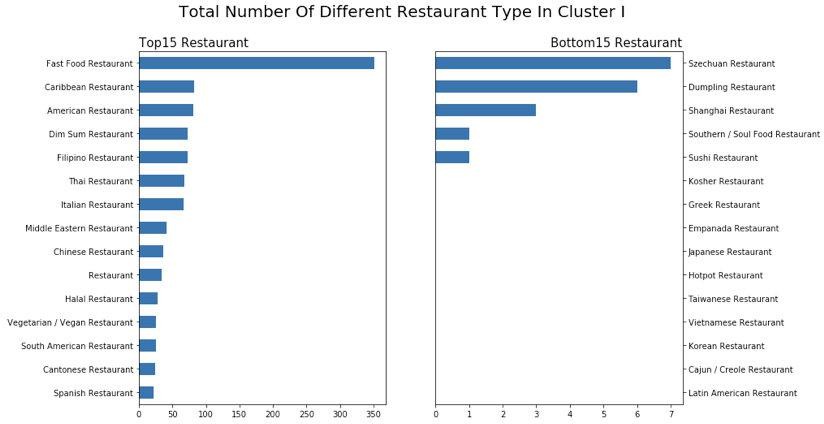
# Result And Conclusion

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

## Cluster I Candidate Location

Fast Restaurant is main type

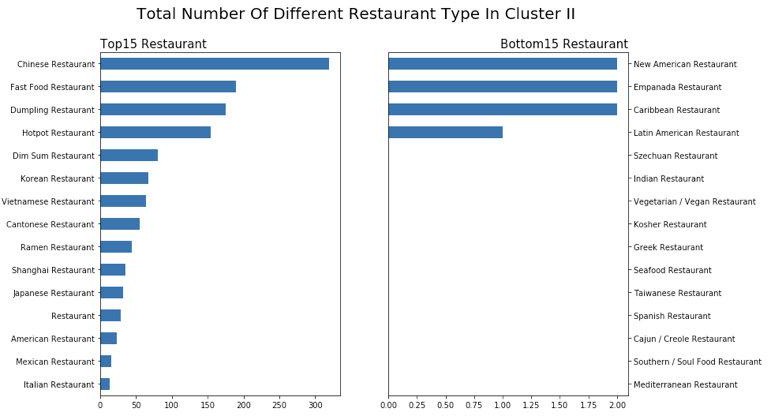
Western Restaurants have the largest market Eastern Restaurants have the few market



## Cluster II Candidate Location

Chinese Restaurant is main type

Asian Restaurants have the largest market, like Chinese Breakfast restaurant Western Restaurants have the few market

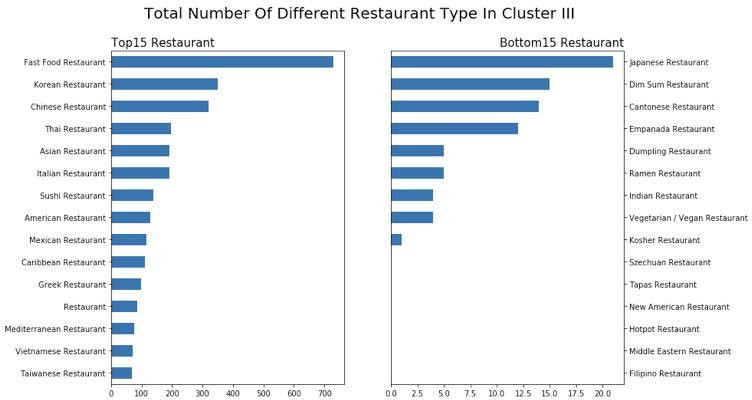


## Cluster III Candidate Location

Fast Food Restaurant is main type

The candidate location has mixture restaurant type.Asian Restaurants and western restaurant are same important type the Asian restaurant is lack of variety

The candidate location is a good choice to open an Asian restaurant. Maybe the Chinese Breakfast restaurant is a good idea

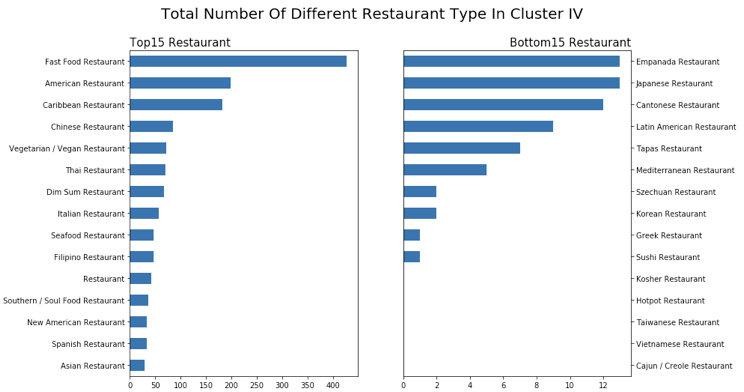


## Cluster IV Candidate Location

Fast Food Restaurant is main type

The candidate location has mixture restaurant type

Asian Restaurants and western restaurant are same important type the Asian restaurant is lack of variety. The candidate location is a good choice to open an Asian restaurant. But the southeastern Asian restaurant is not a good idea



## 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