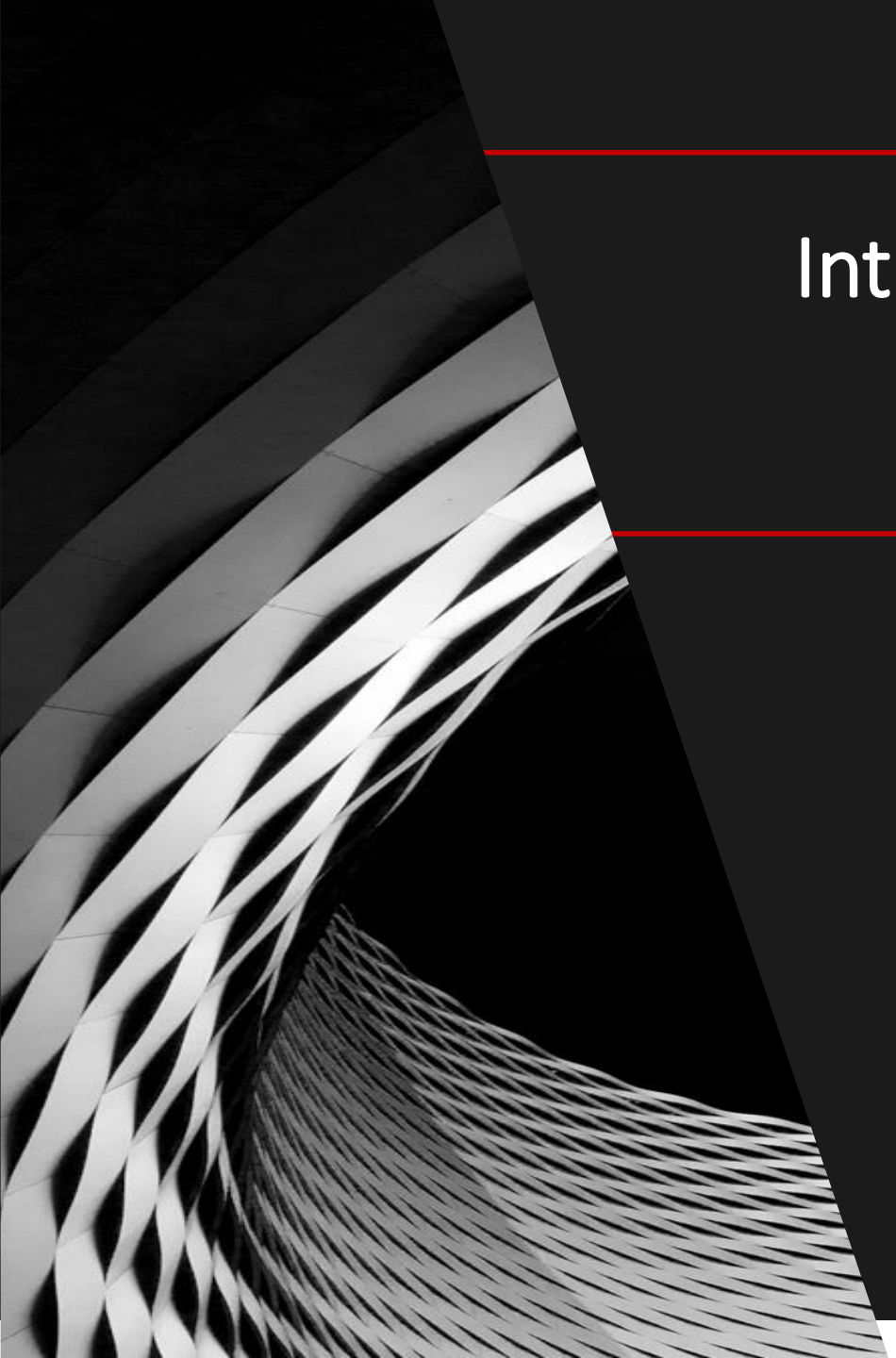
Exploring New York Neighborhoods ——to open a Chinese restaurant

Qingling Shen
qlshen1007@gmail.com



Introduction

New York is the most populous city in the United States. It is the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities. Also, New York is home to the largest ethnic Chinese population outside of Asia, with multiple distinct Chinatowns across the city. Thus, it is one of the best places to start a Chinese restaurant.

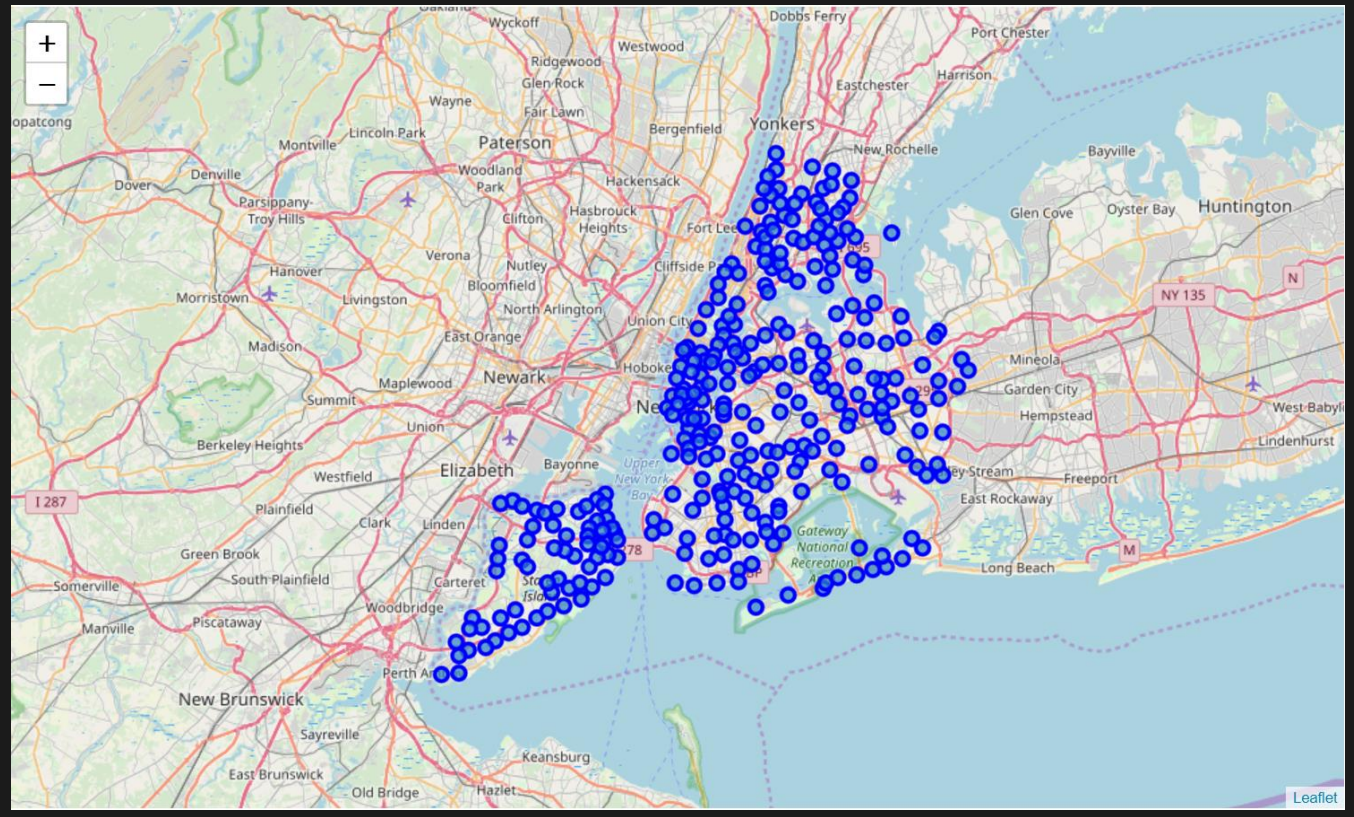
In this project we will go through process to make a decision whether it is a good idea to open a Chinese restaurant. And we also care about which place should we open?

Data acquisition and cleaning

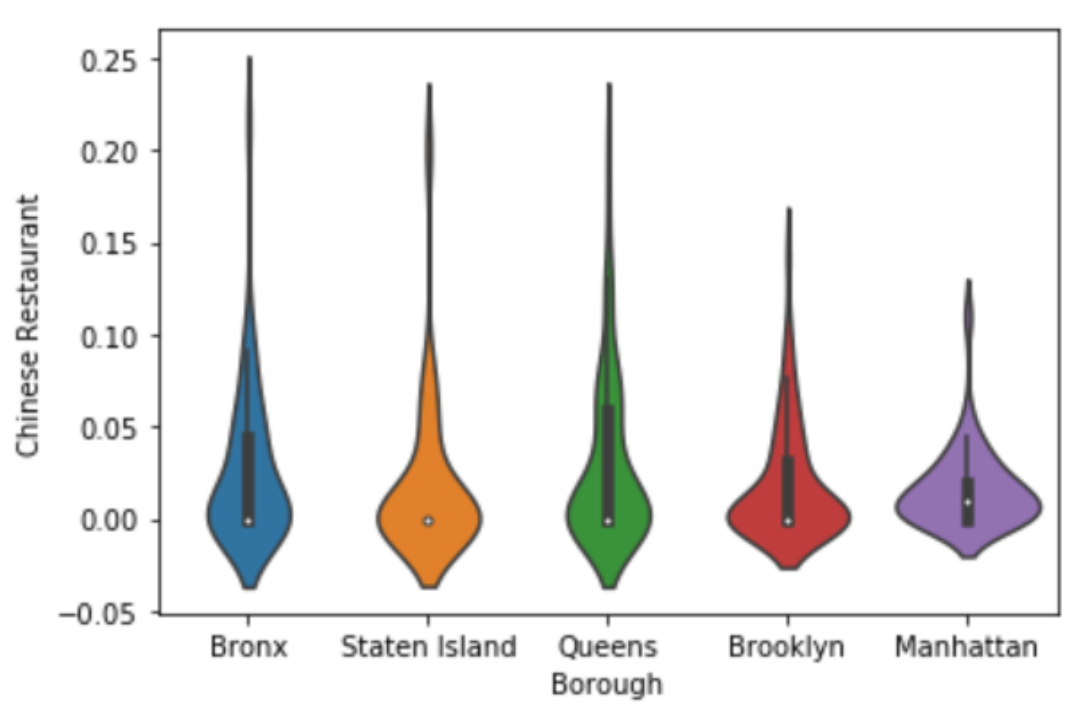
We will use the data mentioned in lab study. Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Here is the link to the dataset: (https://geo.nyu.edu/catalog/nyu_2451_34572).

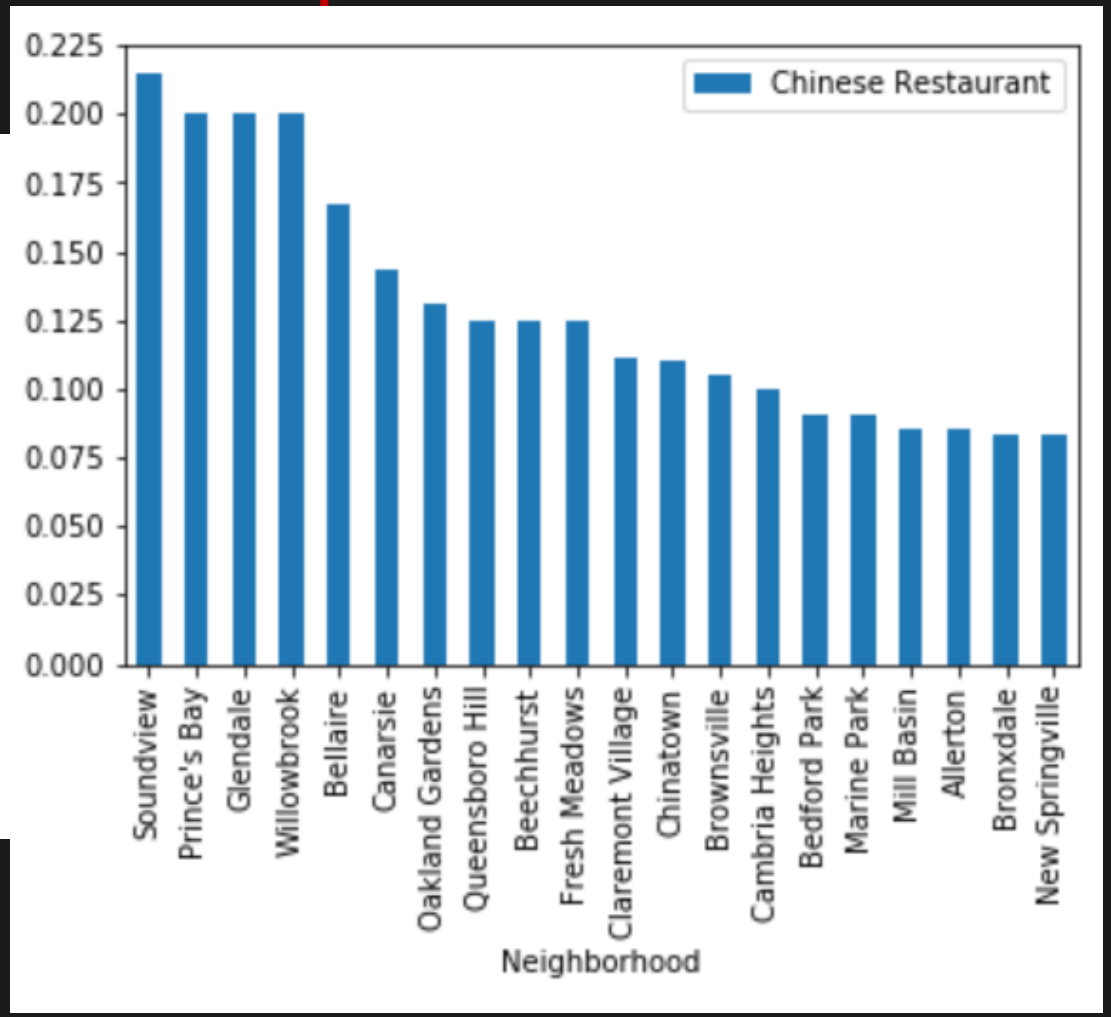
Exploratory Data Analysis

The map of New York with neighborhoods superimposed on top is shown on the right.

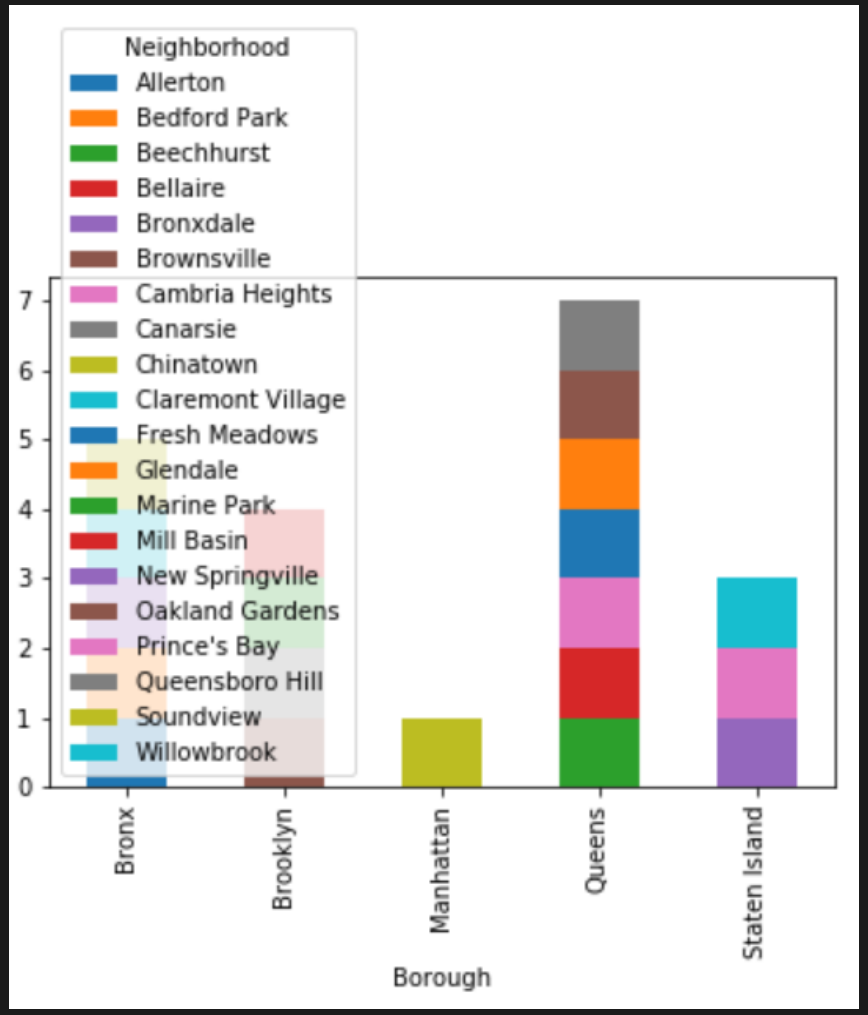


Relationship between neighborhood and Chinese Restaurant



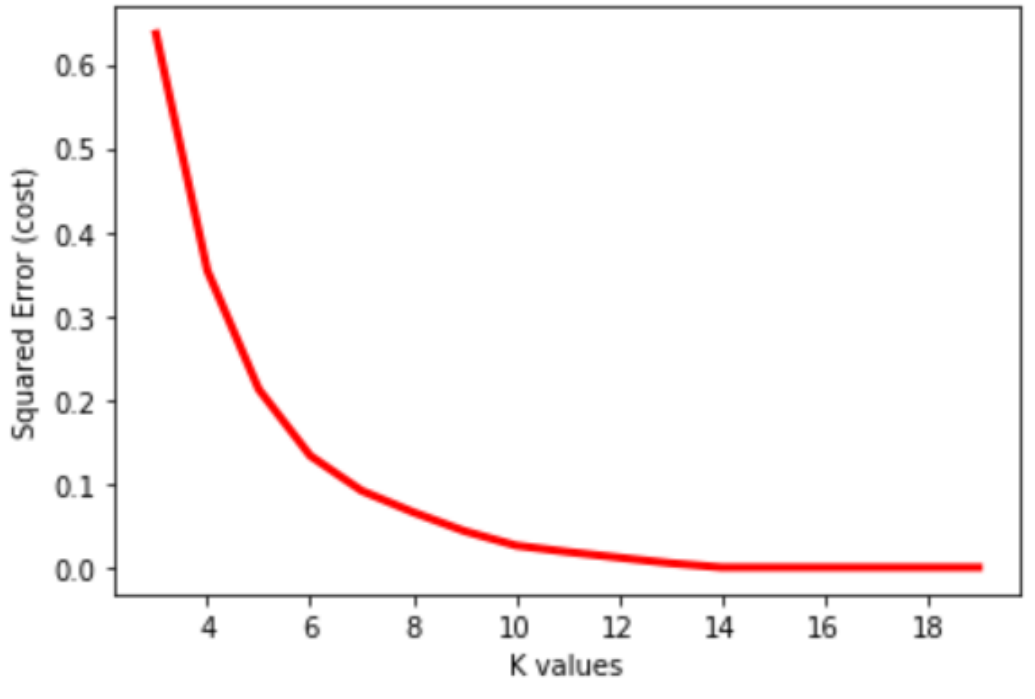


From the plots above we can find that Queens has the largest number of Chinese restaurants. Manhattan and Brooklyn have relatively few Chinese restaurants. Due to the big size of the whole dataset, we only pay attention to Bronx in a follow-up study.



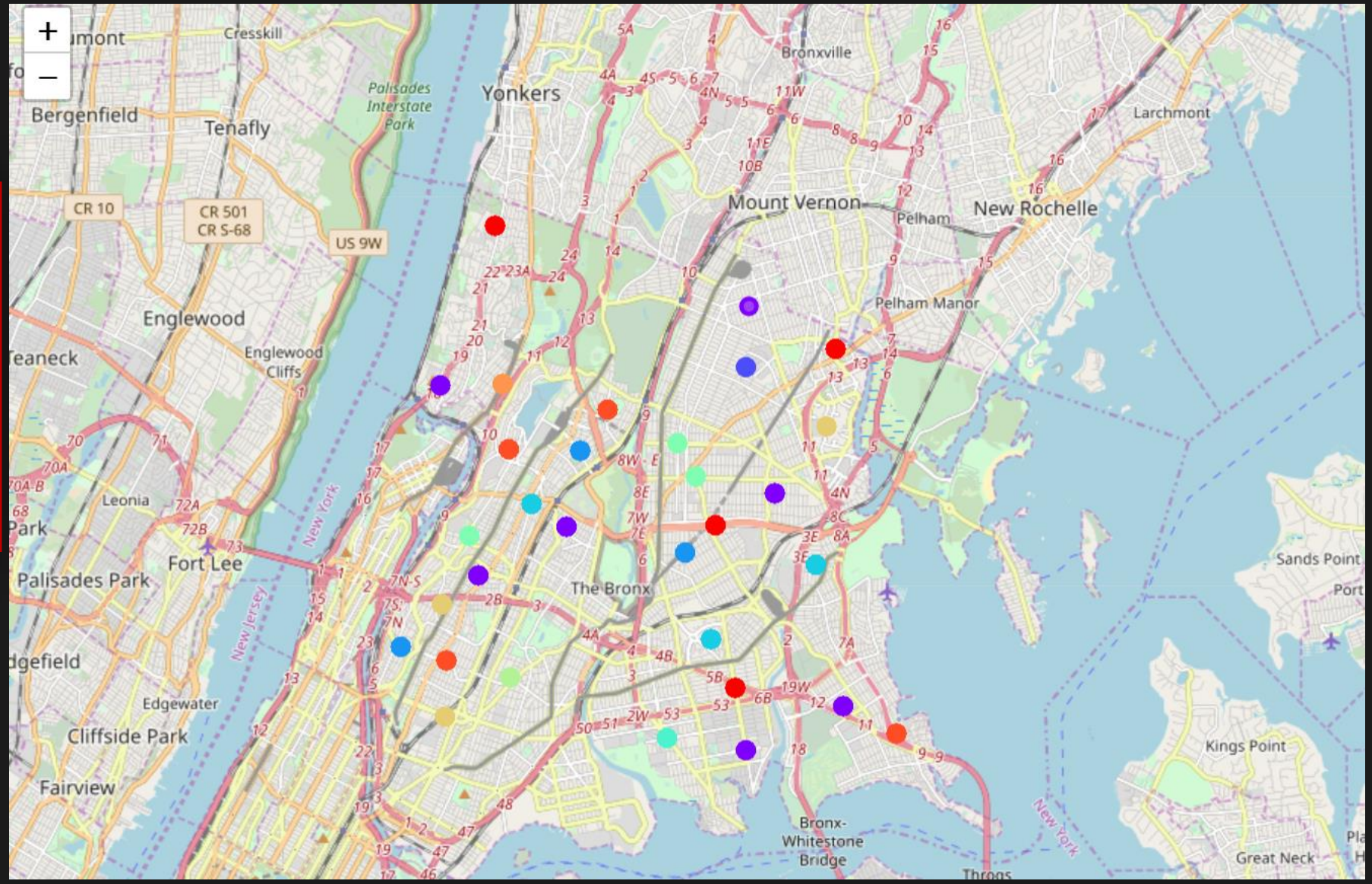
Clustering Neighborhoods of Queens:

First step in K-means clustering is to identify best K value meaning the number of clusters in a given dataset. To do so we are going to use the elbow method on the Bronx borough dataset with Chinese restaurant percentage.



Predictive modelling

The Folium map for the clusters of different neighborhoods is shown on the right: (we just extract the first 800 points with sorted dataset)



```
NY_merged.loc[NY_merged['Cluster Labels'] == 0, NY_merged.columns[[1] + list(range(5, NY_merged.shape[1]))]]
```

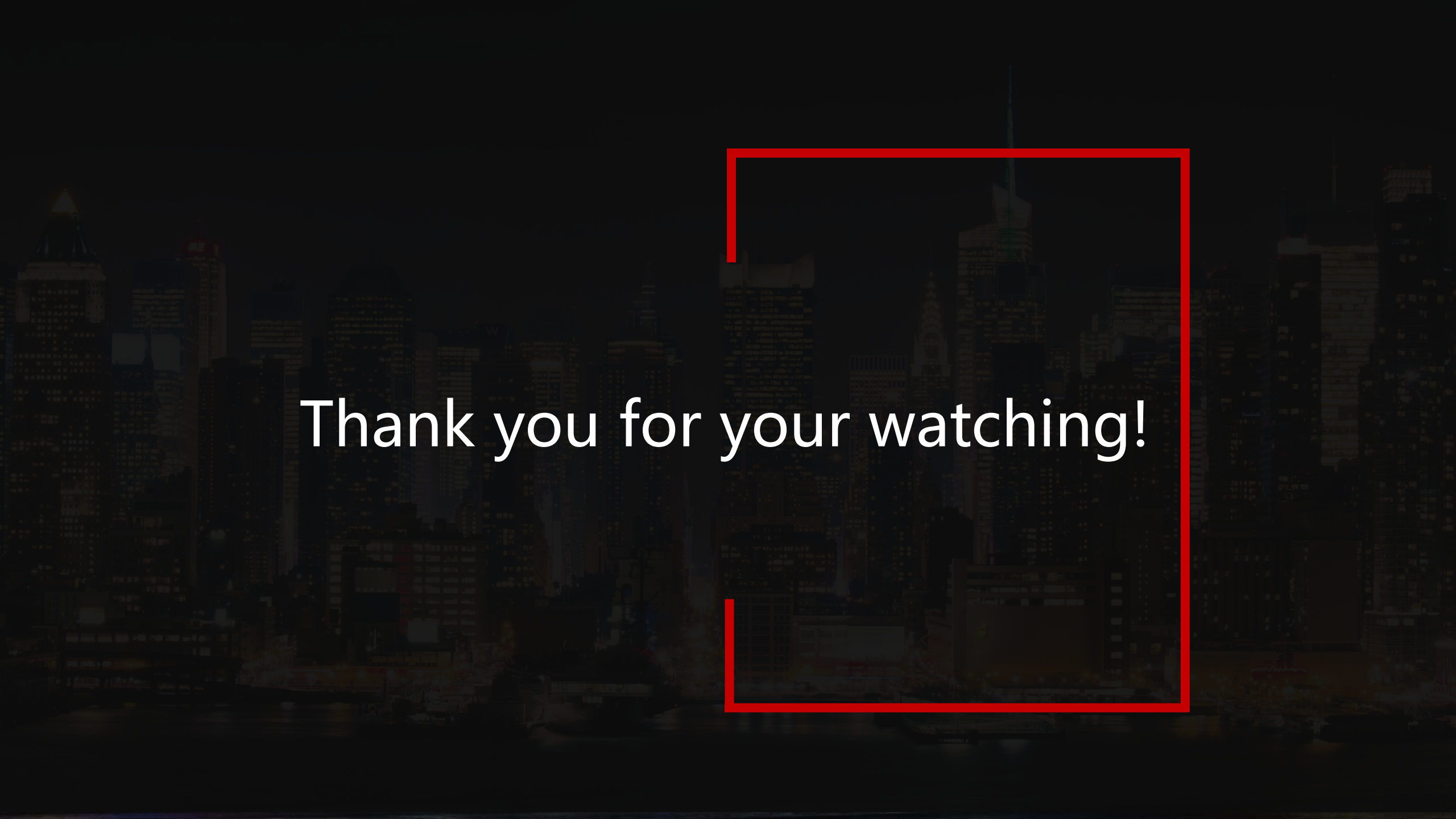
	borough	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	Borough	Chinese Restaurant
26	Bronx	40.885656	-73.829197	Caribbean Restaurant	0.0	Bronx	0.052632
27	Bronx	40.886332	-73.827616	Diner	0.0	Bronx	0.052632
28	Bronx	40.888628	-73.831260	Pizza Place	0.0	Bronx	0.052632
29	Bronx	40.889318	-73.831453	Seafood Restaurant	0.0	Bronx	0.052632
30	Bronx	40.885384	-73.828099	Donut Shop	0.0	Bronx	0.052632
31	Bronx	40.888235	-73.831282	Deli / Bodega	0.0	Bronx	0.052632
32	Bronx	40.888488	-73.831083	Caribbean Restaurant	0.0	Bronx	0.052632

Cluster 0 contains all the neighborhoods which has least number of Chinese restaurants. It is shown in red color in the map.

```
NY_merged.loc[NY_merged['Cluster Labels'] == 10, NY_merged.columns[[1] + list(range(5, NY_merged.shape[1]))]]
```

	borough	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	Borough	Chinese Restaurant
153	Bronx	40.875269	-73.879563	Pizza Place	10.0	Bronx	0.034483
154	Bronx	40.877033	-73.877331	Park	10.0	Bronx	0.034483
155	Bronx	40.874933	-73.879404	Coffee Shop	10.0	Bronx	0.034483
156	Bronx	40.880766	-73.877808	Pizza Place	10.0	Bronx	0.034483
157	Bronx	40.878234	-73.883164	Park	10.0	Bronx	0.034483
158	Bronx	40.880200	-73.883434	Mexican Restaurant	10.0	Bronx	0.034483
159	Bronx	40.881665	-73.879484	Deli / Bodega	10.0	Bronx	0.034483
160	Bronx	40.874727	-73.879660	Bank	10.0	Bronx	0.034483
161	Bronx	40.874499	-73.879515	Pharmacy	10.0	Bronx	0.034483
162	Bronx	40.881566	-73.879299	Restaurant	10.0	Bronx	0.034483

Cluster 14 contains all the neighborhoods which is densely populated with Chinese restaurants. It is shown in orange color in the map.


Thank you for your watching!