

# Identifying best neighborhoods for opening Italian restaurant in New York City

## 1. Introduction

New York is the most populous and most densely populated city in US. It has been described as the cultural, financial, and media capital of the world. An entrepreneur is planning to open an Italian restaurant in this burgeoning city hoping to put a smile on customer's face. He is seeking the help of data science in identifying neighborhoods that has greatest potential.

## 2. Acquiring data

### 2.1 Data Sources

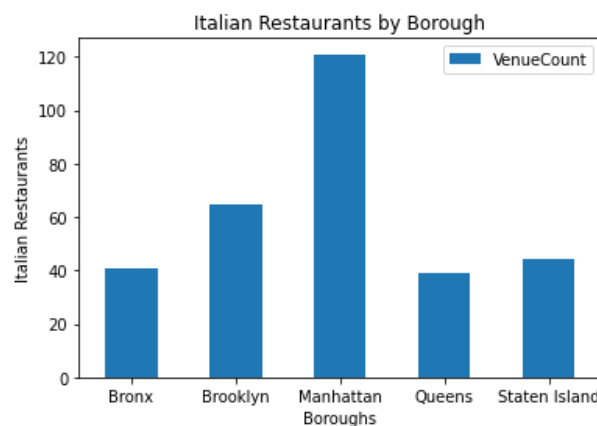
The city's coordinates are downloaded from skillsnetwork as JSON file. This will have the latitude and longitude details of all the neighborhoods of city. The Italian restaurant details in the city are downloaded from FOURSQUARE website along with their coordinates. Population for each Borough are gathered from Wikipedia by web scrapping. This is needed to determine the density of Italian restaurants per Borough.

### 2.2 Processing the data

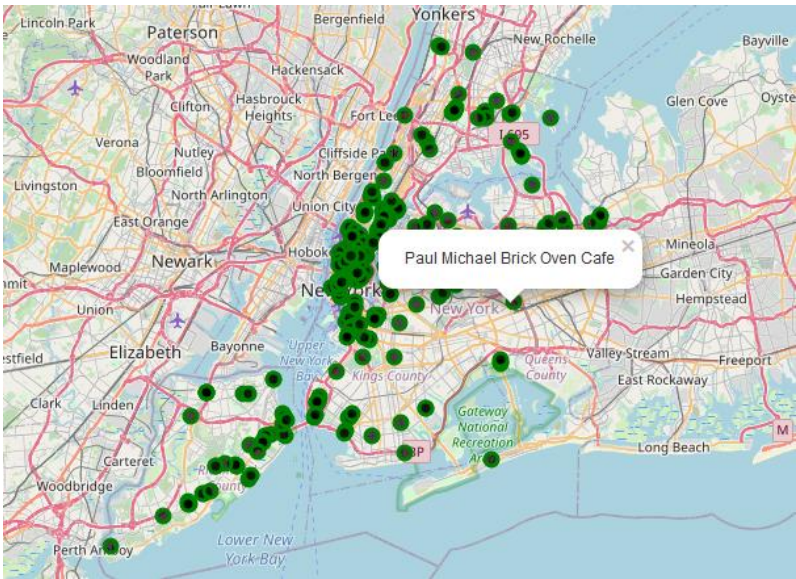
The gathered data are cleaned before processing by either ignoring them if irrelevant or filling with zeros based on the need. The data are loaded into various data frames for processing. We have joined NYC JSON data with FOURSQUARE to list Italian restaurants by neighborhood. Similarly, to calculate the density of Italian restaurants per Borough, we have joined population data with data acquired from other sources.

## 3. Analysis

Firstly, we will see number of Italian restaurants currently open in each Borough.

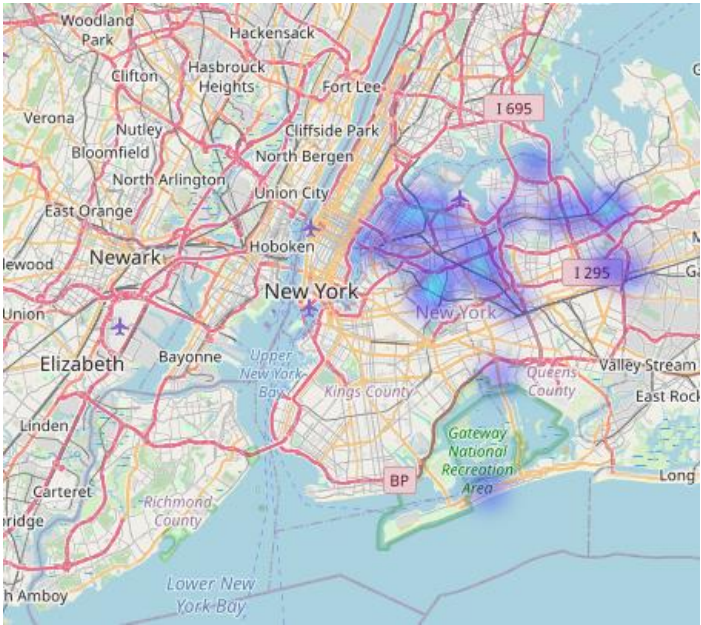


We will also display an interactive map of NYC with locations of existing Italian restaurants and their name in pop-up.

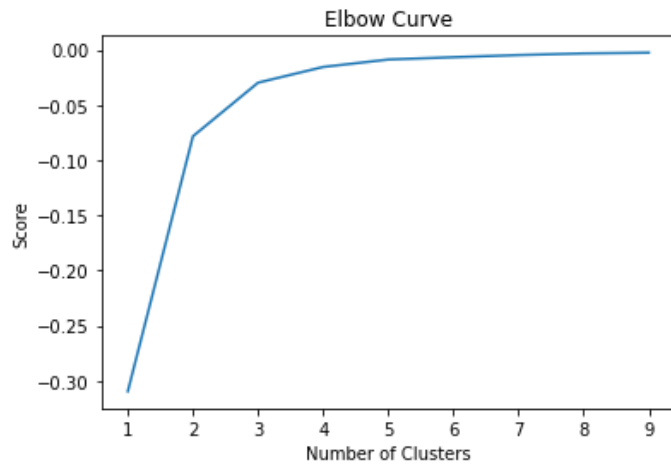


Now, let us narrow down on the Borough that has least density of existing Italian restaurant per 100,000. As we can see below, Queens has lowest density. Let us see the concentration of Italian restaurants in Queens using Heatmap.

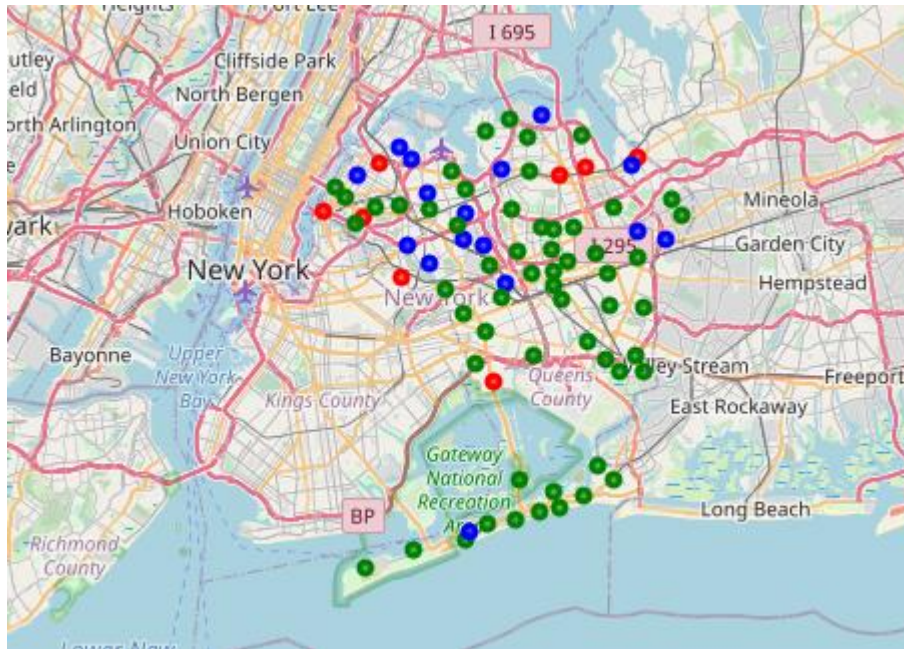
	Borough	VenueCount	Population	Per_100thousand
0	Bronx	41	1418207	2.890974
1	Brooklyn	65	2559903	2.539159
2	Manhattan	121	1628706	7.429211
3	Queens	39	2253858	1.730366
4	Staten Island	44	476143	9.240921



We have used *k-means* for clustering the neighborhoods of Queens and have determined the K based on *Elbow method*.



Next, we will cluster the neighborhoods of Queens into 3 and generate an interactive map based on these clusters.



## 4. Results

The neighborhoods having greatest potentials for opening an Italian restaurant are highlighted in green. We have also listed all these neighborhoods with coordinates below (*full list in Git.*). We suggest opening an Italian restaurant in one of these neighborhoods or in and around the center (40.693714,-73.814812) of all these neighborhoods as shown in Purple in diagram 2 below.

Cluster Labels	Borough	Neighborhood	Latitude	Longitude	VenueCount
1	Queens	Woodside	40.746349	-73.901842	0.0
1	Queens	Elmhurst	40.744049	-73.881656	0.0
1	Queens	Richmond Hill	40.697947	-73.831833	0.0
1	Queens	Long Island City	40.750217	-73.939202	0.0
1	Queens	East Elmhurst	40.764073	-73.867041	0.0
1	Queens	Glendale	40.702762	-73.870742	0.0
1	Queens	Woodhaven	40.689887	-73.858110	0.0
1	Queens	Ozone Park	40.680708	-73.843203	0.0
1	Queens	South Ozone Park	40.668550	-73.809865	0.0
1	Queens	College Point	40.784903	-73.843045	0.0

The center location of all these neighborhoods is highlighted in Purple below (40.693714,-73.814812)

