



# The Battle of Neighborhoods

Xiaojun Ge

## Table of Contents

<b><i>Introduction of business project, target audience</i></b> .....	<b>2</b>
Introduction & Background .....	2
Target Audience.....	2
<b><i>Source of data</i></b> .....	<b>3</b>
<b><i>Exploratory data analysis</i></b> .....	<b>5</b>
Step 1: Download and explore dataset .....	5
Step 2: Explore Neighbourhoods in Queens .....	8
Step 3: Analyse neighbourhood in Queens.....	10
Step 4: Analyse Italian restaurants in Queens .....	12
Step 5: Conclusion.....	16

## Introduction of business project, target audience

### **Introduction & Background**

New York City, which is often called as New York, is the most popular city in the USA, and also it is the most densely populated city in USA. New York City has been described as the cultural, financial and the media capital of the world, which significantly influence the commerce, entertainment, technology, financial of the world.

With the population of almost 20 million people in its metropolitan statistical area, New York City can be described as one of the most competitive city of the world. Due to expensive rental fee and competitiveness, thus the analysis and choice of location is essential for new business need to be carefully.

In this assignment, we will investigate in New York City, how do we choose best location to open a restaurant and get the most benefit by using the Foursquare by providing the location data and venue information. As investigated, Italian families immigrated into the USA in the late 19th and 20th century. So with a huge population of immigration of Italians, they must need to find the restaurant which suitable for them.

In this report, I will focus on exploring on the locations & which neighbourhood and boroughs of New York City have the most Italian restaurants that can suitable for the Italian immigrants' flavour. Moreover, I will investigate if I want to open an Italian restaurant, where should I open?

### **Target Audience**

This project/report aims to provide information and analysis to the New York City restaurant overview and analysis to business stakeholders who may want to invest and open a restaurant in New York.

By looking through this report, business stakeholders will have a virtual overview on which location has the most restaurant and where is the best location for stakeholders to choose to open the restaurant.

## Source of data

1. In order to obtain the basic background and information for New York City, information get from the website: [https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City), where list out the basic history, population, location etc for stakeholders to get a overview of New York City



Figure 1 Basic background and information for NYC

2. Since New York City has 5 boroughs, so in order to segment the neighbourhoods and explore, we need to find the dataset which contains 5 boroughs and neighbourhoods as well within each borough. Latitude and Longitude of each neighbourhood are also needed in the dataset. The available link for this dataset is: [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

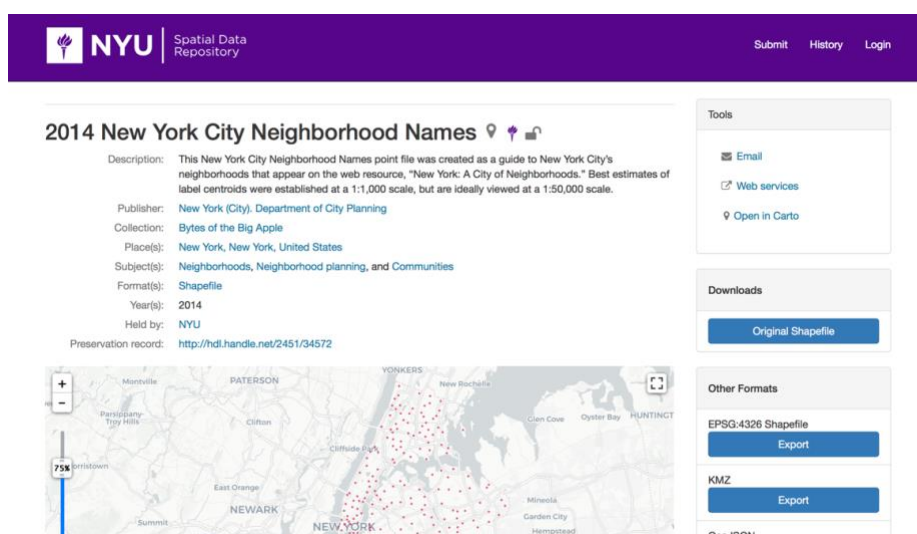


Figure 2 Boroughs and neighbourhoods dataset

3. The link which can directly download the dataset as mentioned in point 2: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

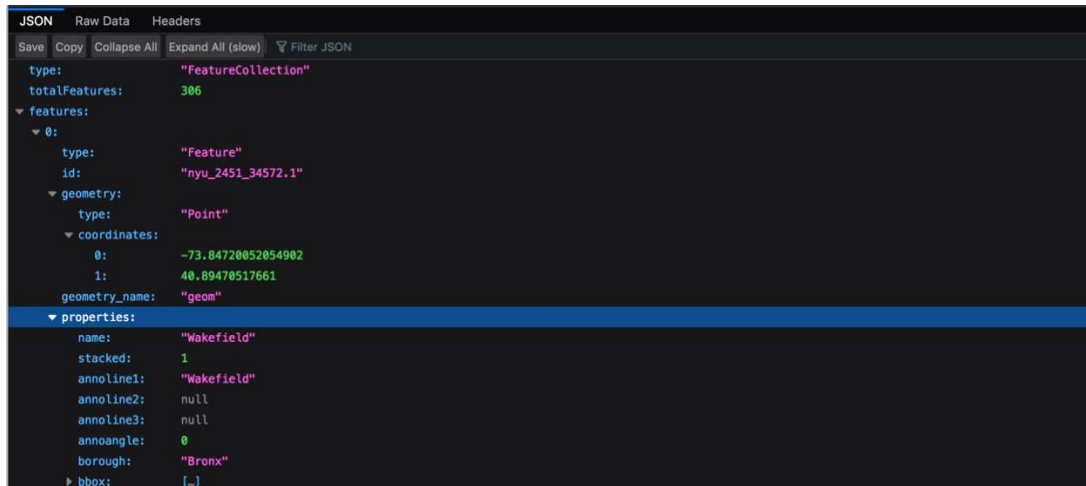


Figure 3 Boroughs and neighbourhood dataset

4. Foursquare API, which has the link of <https://developer.foursquare.com/> will be used as well to investigate the dataset of New York City and neighbourhoods and then will use method of clustering to cluster those neighbourhoods for Italian restaurants in order to get the detailed information.

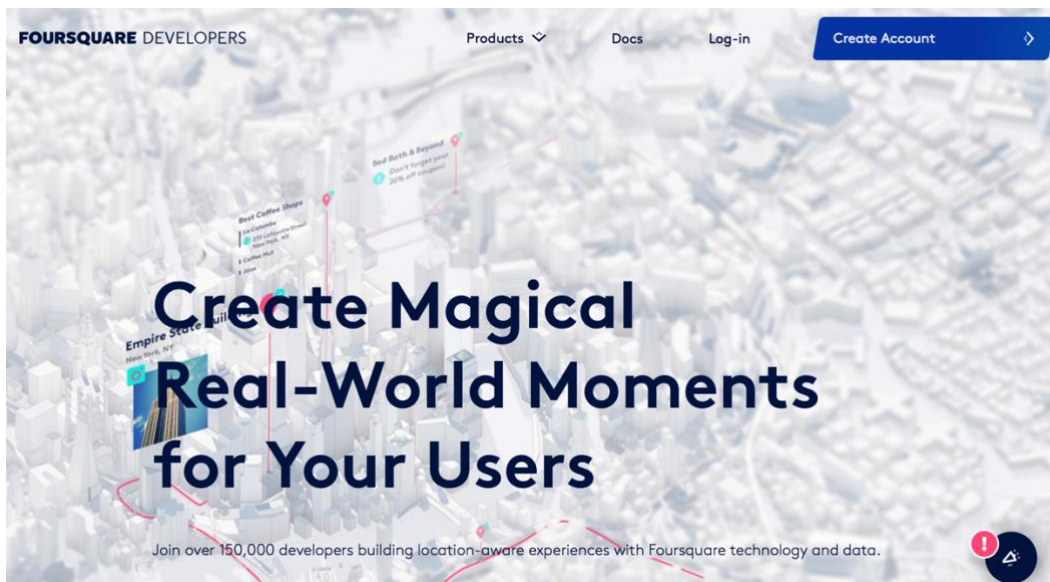


Figure 4 Foursquare API

## Exploratory data analysis

So in order to analyse and find the best location and restaurant in New York City for Italian restaurant opening, we need to analyse detailed information in New York City and its neighbourhoods.

Firstly let us download and import all dependencies we needed.

```
In [1]: import numpy as np # library to handle data in a vectorized manner
import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.pyplot as plt
import matplotlib.colors as colors
import folium
print("All libraries are imported.")

Requirement already satisfied: geopy in /opt/anaconda3/lib/python3.7/site-packages (1.22.0)
Requirement already satisfied: geographiclib<2,>=1.49 in /opt/anaconda3/lib/python3.7/site-packages (from geopy) (1.50)
All libraries are imported.
```

Figure 5 Download and import all libraries and dependencies

## Step 1: Download and explore dataset

Firstly, I perform the action of download the New York City dataset from the website [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) and then I open it in json format.

```
{'type': 'FeatureCollection',
  'totalFeatures': 306,
  'features': [{'type': 'Feature',
    'id': 'nyu_2451_34572.1',
    'geometry': {'type': 'Point',
      'coordinates': [-73.84720052054902, 40.89470517661]},
    'geometry_name': 'geom',
    'properties': {'name': 'Wakefield',
      'stacked': 1,
      'annoline1': 'Wakefield',
      'annoline2': None,
      'annoline3': None,
      'annoangle': 0.0,
      'borough': 'Bronx',
      'bbox': [-73.84720052054902,
        40.89470517661,
        -73.84720052054902,
        40.89470517661]}},
    {'type': 'Feature',
      'id': 'nyu_2451_34572.2',
      'geometry': {'type': 'Point',
        'coordinates': [-73.84720052054902, 40.89470517661]},
      'geometry_name': 'geom',
      'properties': {'name': 'Wakefield',
        'stacked': 1,
        'annoline1': 'Wakefield',
        'annoline2': None,
        'annoline3': None,
        'annoangle': 0.0,
        'borough': 'Bronx',
        'bbox': [-73.84720052054902,
          40.89470517661,
          -73.84720052054902,
          40.89470517661]}}
```

Figure 6 New York dataset in json format

Now we define a function to get the New York city data such as Boroughs, Neighbourhoods along with their latitude and longitude and show 5 example rows of it.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

*Figure 7 Neighbourhood information for New York City*

We use the shape function to see how many neighbourhoods of New York we have and we found that there are 306 neighbourhoods there in New York, which are in 5 borough.

```
neighborhoods.shape
```

```
(306, 4)
```

*Figure 8 Number of neighbourhoods in New York*

Now we print out number of borough and neighbourhoods in a more clear format.

```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
    len(neighborhoods['Borough'].unique()),
    neighborhoods.shape[0]
))
```

```
The dataframe has 5 boroughs and 306 neighborhoods.
```

*Figure 9 Number of boroughs and neighbourhoods in New York*

Then I want to draw a bar chart to visualize these 5 boroughs and 306 neighborhoods.



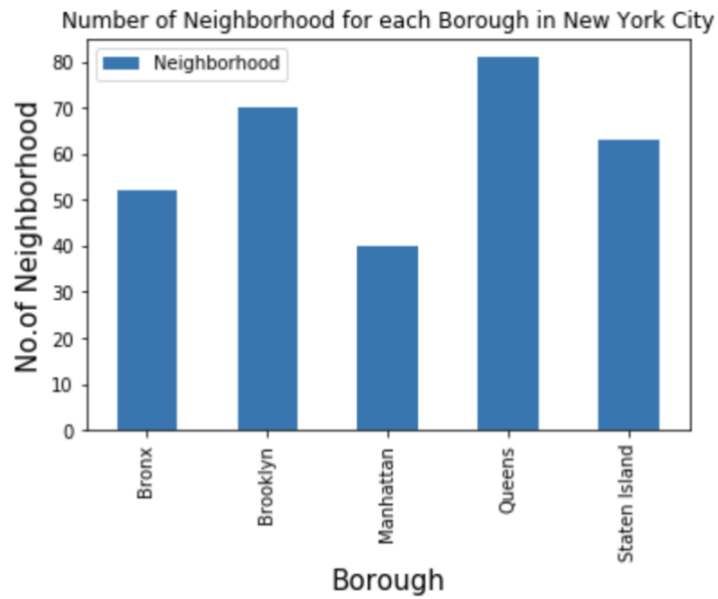


Figure 10 Bar chart of New York borough and neighbourhods

Based on the bar chart, we can clearly see that Queens is the borough which has the most neighborhoods. So, in the next section, I will focus on Queens to analyze into more details. For now, I would like to show the visualization of map of New York and Queens to give a more visual on neighborhoods allocation.

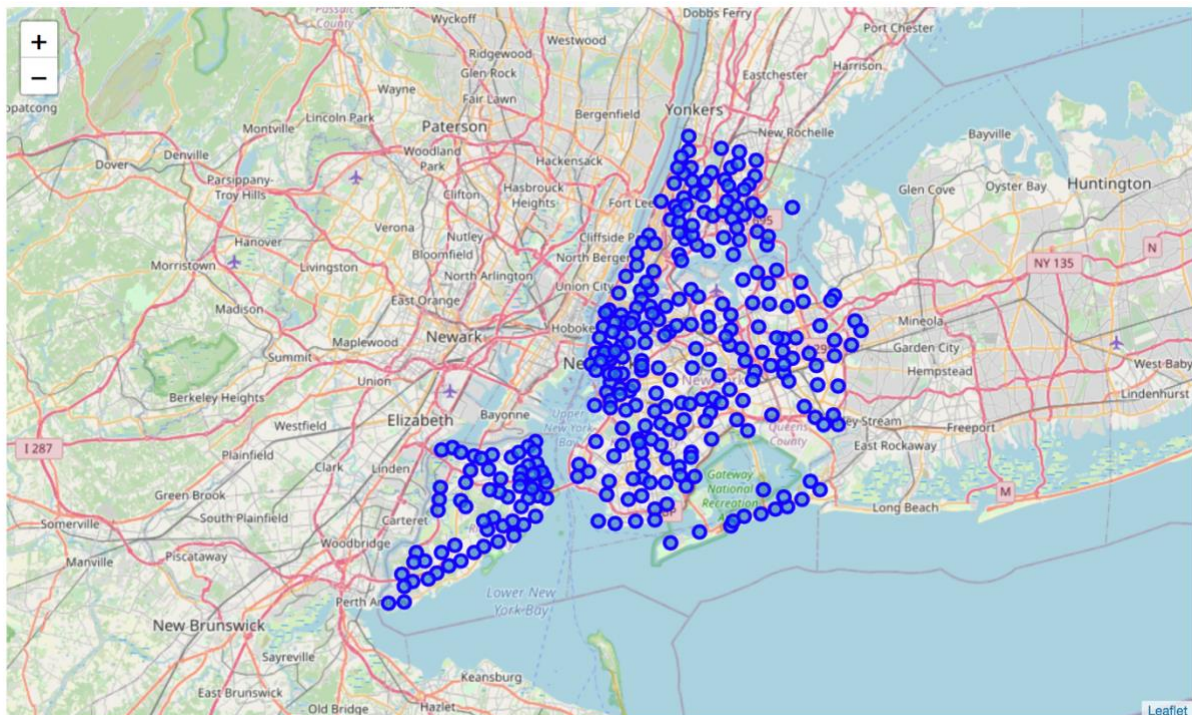


Figure 11 Map of New York City

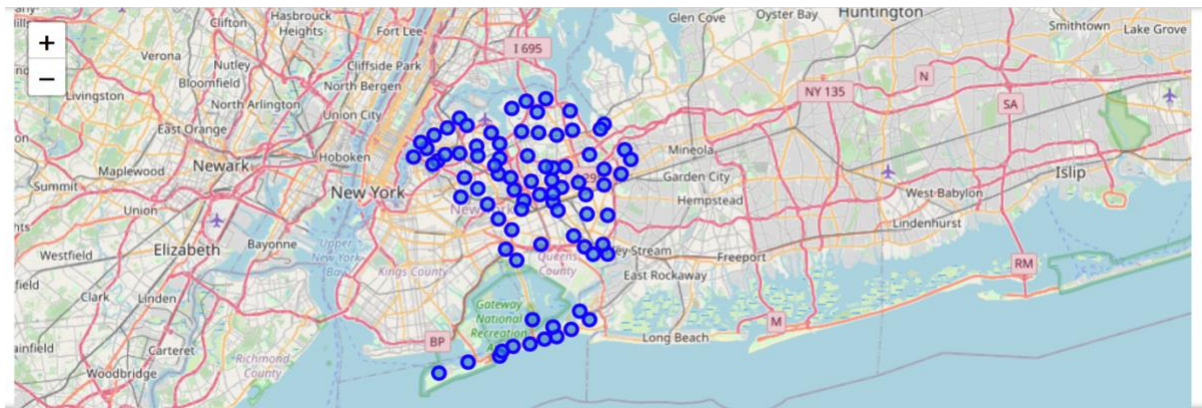
Now we only filter out dataset contains Queens and we create a map to visualize on Queens neighbourhood again.



	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138

*Figure 12 Dataset of Queens*

After we get the geo-location of Queens, I create a map for Queens and its neighbourhoods.



*Figure 13 Map of Queens and its neighbourhoods*

## Step 2: Explore Neighbourhoods in Queens

After we passing through Foursquare API credentials, we want to use it to explore details for Queens, so firstly I print 5 neighbourhoods of Queens as example.

```

0          Astoria
1          Woodside
2    Jackson Heights
3          Elmhurst
4          Howard Beach
5          Corona
Name: Neighborhood, dtype: object

```

*Figure 14 5 examples of neighbourhoods in Queens*

Now we find geo-location of them as well as reference.

```

Latitude and longitude values of Astoria are 40.76850859335492, -73.91565374304234.
Latitude and longitude values of Woodside are 40.74634908860222, -73.90184166838284.
Latitude and longitude values of Jackson Heights are 40.75198138007367, -73.88282109164365.
Latitude and longitude values of Elmhurst are 40.744048505122024, -73.88165622288388.
Latitude and longitude values of Howard Beach are 40.65422527738487, -73.8381376460028.

```

*Figure 15 Geo-location of neighbourhoods in Queens*

Then after we pass through our credentials, fix our limit, radius etc, we get the response from foursquare API in json format.

```
{'meta': {'code': 200, 'requestId': '5f16c56dc92ee105a1926cb3'},
  'response': {'suggestedFilters': {'header': 'Tap to show:',
    'filters': [{'name': 'Open now', 'key': 'openNow'},
      {'name': '$-$$$$', 'key': 'price'}]},
    'headerLocation': 'Old Howard Beach',
    'headerFullLocation': 'Old Howard Beach, Queens',
    'headerLocationGranularity': 'neighborhood',
    'totalResults': 38,
    'suggestedBounds': {'ne': {'lat': 40.65872528188487,
      'lng': -73.83221716558458},
      'sw': {'lat': 40.649725272884865, 'lng': -73.84405812642103}},
    'groups': [{'type': 'Recommended Places',
      'name': 'recommended',
      'items': [{'reasons': {'count': 0,
        'items': [{'summary': 'This spot is popular',
          'type': 'general',
          'reasonName': 'globalInteractionReason'}]}],
      'venue': {'id': '4af23b33f964a520c9e621e3',
        'name': "Lenny's Clam Bar",
```

Figure 16 Foursquare API response

Within radius( I set as 500m) and limit(100) in Queens, I get 5 venues examples to show out with their name, category and their geo-location.

	name	categories	lat	lng
0	Lenny's Clam Bar	Seafood Restaurant	40.655749	-73.839004
1	Brothers Ravioli	Italian Restaurant	40.655282	-73.838991
2	Old Country Bagel	Bagel Shop	40.653484	-73.838576
3	Prima Pasta & Cafe	Italian Restaurant	40.654385	-73.839093
4	Carvel Ice Cream	Ice Cream Shop	40.652120	-73.838437

Figure 17 Venue examples in Queens

After expand the radius to 1000m, there are more and more neighbourhoods and venues appear, using the shape function, there are 5241 venues there and I take 5 examples here.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Astoria	40.768509	-73.915654	Favela Grill	40.767348	-73.917897	Brazilian Restaurant
1	Astoria	40.768509	-73.915654	Titan Foods Inc.	40.769198	-73.919253	Gourmet Shop
2	Astoria	40.768509	-73.915654	CrossFit Queens	40.769404	-73.918977	Gym
3	Astoria	40.768509	-73.915654	Simply Fit Astoria	40.769114	-73.912403	Gym
4	Astoria	40.768509	-73.915654	Al-sham Sweets and Pastries	40.768077	-73.911561	Middle Eastern Restaurant

Figure 18 Examples of venues after radius expand to 1000m

The format just now seems complex to visualize, now I group each venue into neighbourhoods in Queens and show them again.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arverne	35	35	35	35	35	35
Astoria	100	100	100	100	100	100
Astoria Heights	76	76	76	76	76	76
Auburndale	100	100	100	100	100	100
Bay Terrace	63	63	63	63	63	63
Bayside	100	100	100	100	100	100
Bayswater	7	7	7	7	7	7
Beechhurst	47	47	47	47	47	47
Bellaire	56	56	56	56	56	56
Belle Harbor	21	21	21	21	21	21

Figure 19 Group into Queens neighbourhoods

### Step 3: Analyse neighbourhood in Queens

Now as we already identify all neighbourhoods in Queens, we can analyse each neighbourhood in details.

Zoo	ATM	Accessories Store	Afghan Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	At
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 20 Convert categorical into numerical

Then based on the numerical number we are easily to calculate number of occurrence frequency for each category.

Neighborhood	Zoo	ATM	Accessories Store	Afghan Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	At
0	Arverne	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.00	0.0	0.0	0.0	0.0	0.0	0.0
1	Astoria	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.010000	0.01	0.0	0.0	0.0	0.0	0.0	0.0
2	Astoria Heights	0.0	0.0	0.000000	0.0	0.0	0.039474	0.0	0.000000	0.00	0.0	0.0	0.0	0.0	0.0	0.0
3	Auburndale	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.010000	0.00	0.0	0.0	0.0	0.0	0.0	0.0
4	Bay Terrace	0.0	0.0	0.015873	0.0	0.0	0.000000	0.0	0.031746	0.00	0.0	0.0	0.0	0.0	0.0	0.0

Figure 21 Number of occurrence frequency per category

Now we want to target analyse on Italian restaurants in Queens and see how many are there in each neighbourhood in Queens and here is the result.

```

Italian Resturants in Astoria, Queens:3
Italian Resturants in Woodside, Queens:1
Italian Resturants in Jackson Heights, Queens:1
Italian Resturants in Elmhurst, Queens:0
Italian Resturants in Howard Beach, Queens:4
Italian Resturants in Corona, Queens:3
Italian Resturants in Forest Hills, Queens:2
Italian Resturants in Kew Gardens, Queens:1
Italian Resturants in Richmond Hill, Queens:1
Italian Resturants in Flushing, Queens:0
Italian Resturants in Long Island City, Queens:3
Italian Resturants in Sunnyside, Queens:1
Italian Resturants in East Elmhurst, Queens:1
Italian Resturants in Maspeth, Queens:2
Italian Resturants in Ridgewood, Queens:3
Italian Resturants in Glendale, Queens:1
Italian Resturants in Rego Park, Queens:1
Italian Resturants in Woodhaven, Queens:0
Italian Resturants in Ozone Park, Queens:1
Italian Resturants in South Ozone Park, Queens:1
Italian Resturants in College Point, Queens:0
Italian Resturants in Whitestone, Queens:6
Italian Resturants in Bayside, Queens:3
Italian Resturants in Auburndale, Queens:2
Italian Resturants in Little Neck, Queens:6
Italian Resturants in Douglaston, Queens:5
Italian Resturants in Glen Oaks, Queens:1
Italian Resturants in Bellerose, Queens:2

```

*Figure 22 Italian restaurants in Queens*

We show it in a table format with 5 examples.

	Borough	Neighborhood	ID	Name
0	Queens	Astoria	5010b169e4b05edd9983ec7d	Antika Restaurant & Pizzeria
1	Queens	Astoria	4ebf0e380aaf6450ba608920	Via Trenta
2	Queens	Astoria	53a60a88498ef405d8a78217	VITE vinosteria
3	Queens	Woodside	4bc11e07920eb713d6491a2c	Donato's Restaurant
4	Queens	Jackson Heights	4a82d6adf964a5206ff91fe3	Uncle Peter's

*Figure 23 Italian restaurant in Queens*

Write a function to show it in the bar chart format and from the bar chart we can easily identify that Hunters Point has the most Italian restaurant in Queens.

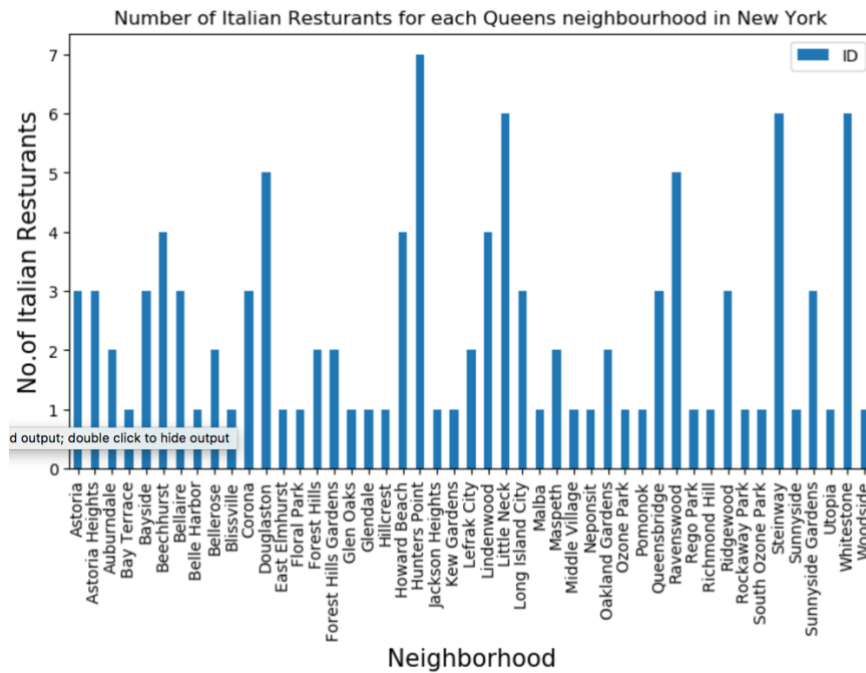


Figure 24 Allocation of Italian restaurants in Queens

Now let us have an overall view of Italian restaurants in Hunters Point.

	Borough	Neighborhood	ID	Name
94	Queens	Hunters Point	463220f3f964a520fb451fe3	Manetta's Ristorante
95	Queens	Hunters Point	4a1a3f6df964a5208b7a1fe3	Bella Via
96	Queens	Hunters Point	49f5bf5ef964a520b46b1fe3	Manducatis Rustica
97	Queens	Hunters Point	556fb3b4498e61f524e70951	Maiella
98	Queens	Hunters Point	4f74ff83e4b09b101271b89b	Manducatis
99	Queens	Hunters Point	53ed8f36498ea1eca81563c1	Il Falco
100	Queens	Hunters Point	540a3547498e9b339b6431ae	Le Fanfare

Figure 25 Italian restaurants in Hunters Point

#### Step 4: Analyse Italian restaurants in Queens

Although Hunters Point has the most Italian restaurants, we still need to analyse overall status in Queens which will have a more accurate evaluation.

I write a function which able to find restaurant's tips, ratings and likes which I put them into a table format and shows here.



	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Queens	Astoria	5010b169e4b05edd9983ec7d	Antika Restaurant & Pizzeria	72	8.2	31
1	Queens	Astoria	4ebf0e380aaf6450ba608920	Via Trenta	57	7.7	28
2	Queens	Astoria	53a60a88498ef405d8a78217	VITE vinosteria	106	8.8	30
3	Queens	Woodside	4bc11e07920eb713d6491a2c	Donato's Restaurant	23	7.2	9
4	Queens	Jackson Heights	4a82d6adf964a5206ff91fe3	Uncle Peter's	42	8.9	11
5	Queens	Howard Beach	4bae3bd0f964a52009963be3	Brothers Ravioli	10	8.1	2
6	Queens	Howard Beach	4bafdb5df964a52086263ce3	Prima Pasta & Cafe	35	7.7	17
7	Queens	Howard Beach	4e6409451495676d566f10b5	Bruno Ristorante Italiano	29	8.7	7
8	Queens	Howard Beach	4ba148f5f964a5201da937e3	Roma View Catering	16	7.0	3
9	Queens	Corona	4b808501f964a5207b7930e3	Park Side	187	8.7	78
10	Queens	Corona	4ea2fc90cc2102b34d3d8dc7	Il Triangolo Ristorante	35	8.6	26
11	Queens	Corona	0	0	0	0.0	0
12	Queens	Forest Hills	4c8d2ce6f87e224bfb6d3b05	Tuscan Hills	46	8.0	15
13	Queens	Forest Hills	4ba0203af964a520105c37e3	Portofino Ristorante	58	7.6	38
14	Queens	Kew Gardens	4c93eb6a72dd224be41d9591	Paul Michael Brick Oven Cafe	16	7.1	14
15	Queens	Richmond Hill	4c93eb6a72dd224be41d9591	Paul Michael Brick Oven Cafe	16	7.1	14
16	Queens	Long Island City	5aa41faf178a2a08410ce198	Beebe's	53	8.6	11
17	Queens	Long Island City	53ed8f36498ea1eca81563c1	Il Falco	38	8.3	15

*Figure 26 Italian restaurants venue details*

Then I display datatypes for each column to show dtype for each. Found that Likes and Tips should convert to float64 for easily process future.

```

Borough      object
Neighborhood  object
ID            object
Name          object
Likes        object
Rating       float64
Tips         object
dtype: object

```

*Figure 27 dtypes for each column*

After converting, it will be easily for us to find the Italian restaurant which has the most tips, highest rating and most likes.



Firstly I would like to find out which Italian restaurant has the most likes in Queens.

```
Borough           Queens
Neighborhood      Corona
ID                4b808501f964a5207b7930e3
Name              Park Side
Likes             187
Rating            8.7
Tips              78
Name: 9, dtype: object
```

*Figure 28 Italian restaurant which has most likes*

Secondly I would like to find out which Italian restaurant has the highest rating in Queens.

```
Borough           Queens
Neighborhood      Sunnyside
ID                555d29f7498eb6112f44ecac
Name              SoleLuna
Likes             119
Rating            9.1
Tips              26
Name: 19, dtype: object
```

*Figure 29 Italian restaurant which has highest rating*

Thirdly I would like to find out which Italian restaurant has the most tips in Queens.

```
Borough           Queens
Neighborhood      Corona
ID                4b808501f964a5207b7930e3
Name              Park Side
Likes             187
Rating            8.7
Tips              78
Name: 9, dtype: object
```

*Figure 30 Italian restaurant which has most tips*

After that, I want to show the average rating of each neighbourhood in Queens which has Italian restaurants.

	Neighborhood	Average Rating
19	Sunnyside	9.100000
9	Jackson Heights	8.900000
15	Rego Park	8.500000
0	Astoria	8.233333
18	South Ozone Park	8.200000
8	Howard Beach	7.875000
12	Long Island City	7.833333
6	Forest Hills	7.800000
17	Ridgewood	7.633333
2	Bayside	7.400000

Figure 31 Average rating of neighbourhoods in Queens

I put them into a bar chart format which will be shown more clearly. Based on bar chart we can see that Sunnyside is the one which shows the highest, followed by Jackson Heights etc.

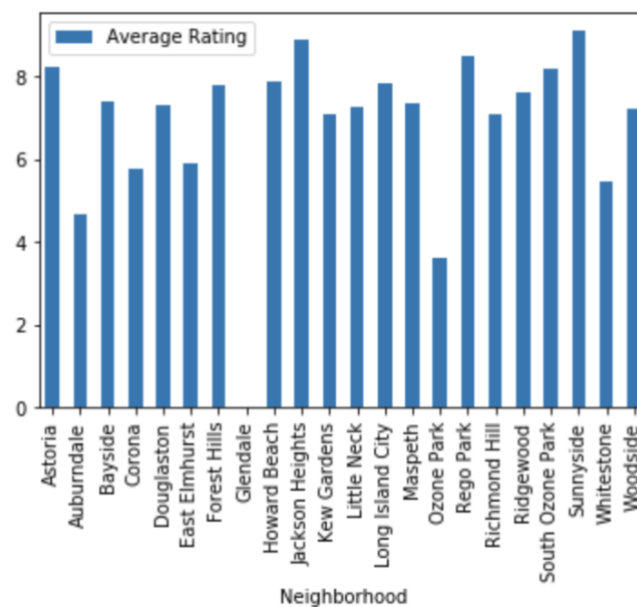


Figure 32 Average rating of neighbourhoods in Queens

Lastly, we filter out the neighbourhoods which has the Italian restaurant which has the rating score  $\geq 8.5$ .

	Neighborhood	Average Rating
9	Jackson Heights	8.9
15	Rego Park	8.5
19	Sunnyside	9.1

Figure 33 Average rating  $\geq 8.5$

## **Step 5: Conclusion**

Some findings and analysis after done the assignment:

1. Queens has the most neighbourhoods among all 5 borough in New York City.
2. Hunters Point has the most Italian restaurant in Queens.
3. Park Side - Corona is the one which has the most LIKES by customers in Queens.
4. SoleLuna - Sunnyside has the highest RATING by customers in Queens.
5. Park Side - Corona has the most TIPS by customers in Queens.
6. Overall in Queens, Sunnyside has the highest average rating regards of Italian restaurants.