

Introduction

New York is one of the world most famous places to live and work. This attracts lot of people to move to the city. The increasing population attracts more options for new entrepreneurs to open business and provide services to the customers. The project elaborates on Data Science attempts to analyze the Queens neighborhood in the city and tries to understand the popular sites in the area.

The deciding factor for most would be how trending the venues are and what kinds of venues are available currently in the selected area. The business problem assumes that the target audiences for the project are the entrepreneur's who will like to review potential for opening a new business in the City of New York or its neighborhood.

Problem Statement

Analyze the data and cluster the data for the neighborhoods to draw a conclusion. The stakeholder will like to know which kinds of restaurant are currently available in the neighborhood. This information will help them on deciding on the new venture.

Data Source and Its Use

We will utilize the Foursquare API to pull the data near Queens area and utilize the data to find the famous venues and popular at that time of the execution of the process. We will clean the data and remove the N/A columns and process the data. We will also create maps to show the venues. We will also present bar graph to help stakeholder visualize the data.

Methodology

- We start by importing the required files for processing the data. We further download the data as json file as show below.

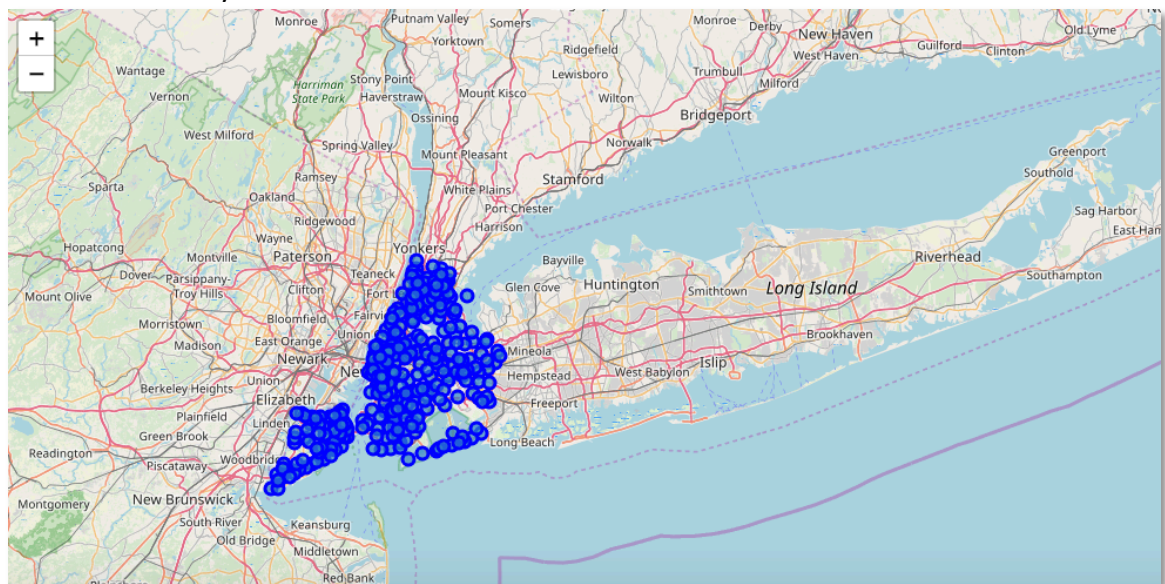
```
Out[19]: {'type': 'FeatureCollection',
          'totalFeatures': 306,
          'features': [{'type': 'Feature',
                        'id': 'nyu_2451_34572.1',
                        'geometry': {'type': 'Point',
                                    'coordinates': [-73.84720052054902, 40.89470517661]},
                        '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]}}],
          'bbox': [-73.84720052054902, 40.89470517661, -73.84720052054902, 40.89470517661]}
```

- Data was organized in Data frame with column names.

Out[24]:

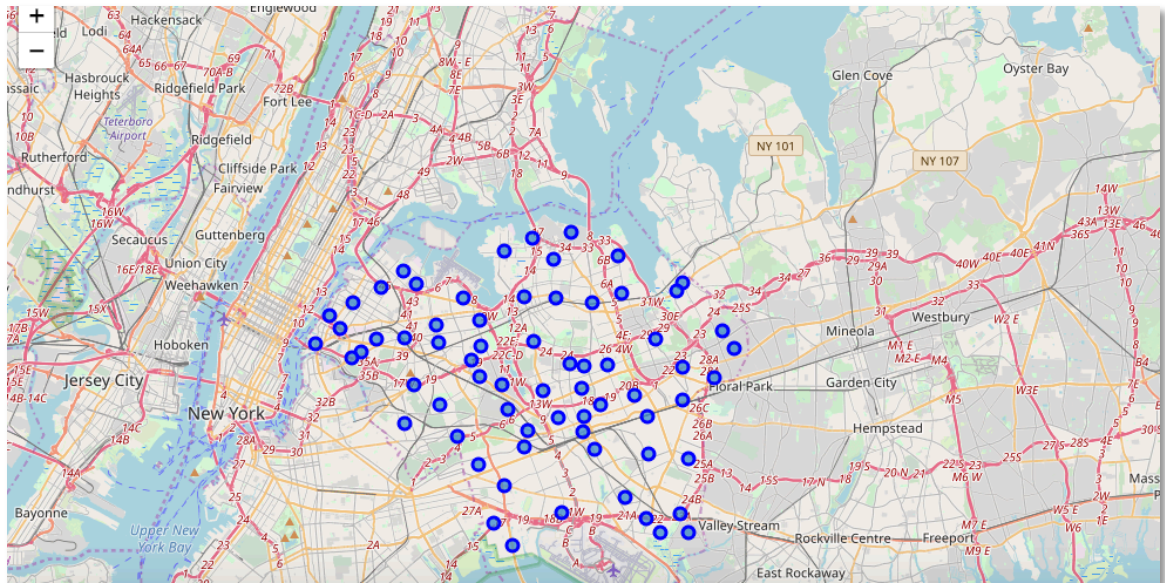
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

- Geocode library was installed which was used to get Geographical code for New York were obtained and then using the data a map of New York was plotted using the Folium library.



- Further, Geographical data for Queens with its neighborhoods was obtained and map was plotted.

	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



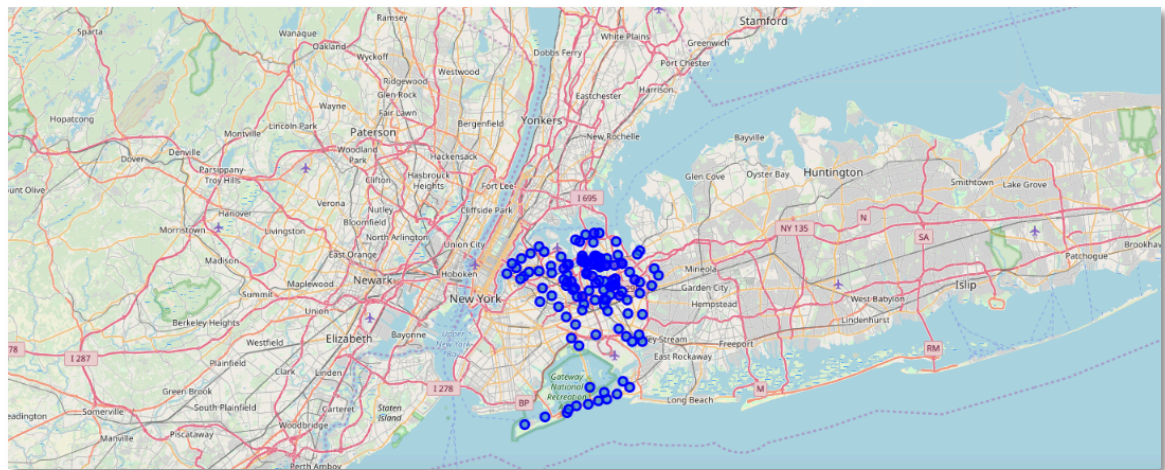
- We utilized Foursquare API using the assigned credential to import the data for Woodside-Queens, which is the most famous area in Queens. We further get the latitude and longitude for the Woodside to pull in the top 50 Venues in radius of 500 for our analysis. We further arrange the data in Data frame as column names, Categories, Lat (Latitude) and Lon (Longitude)

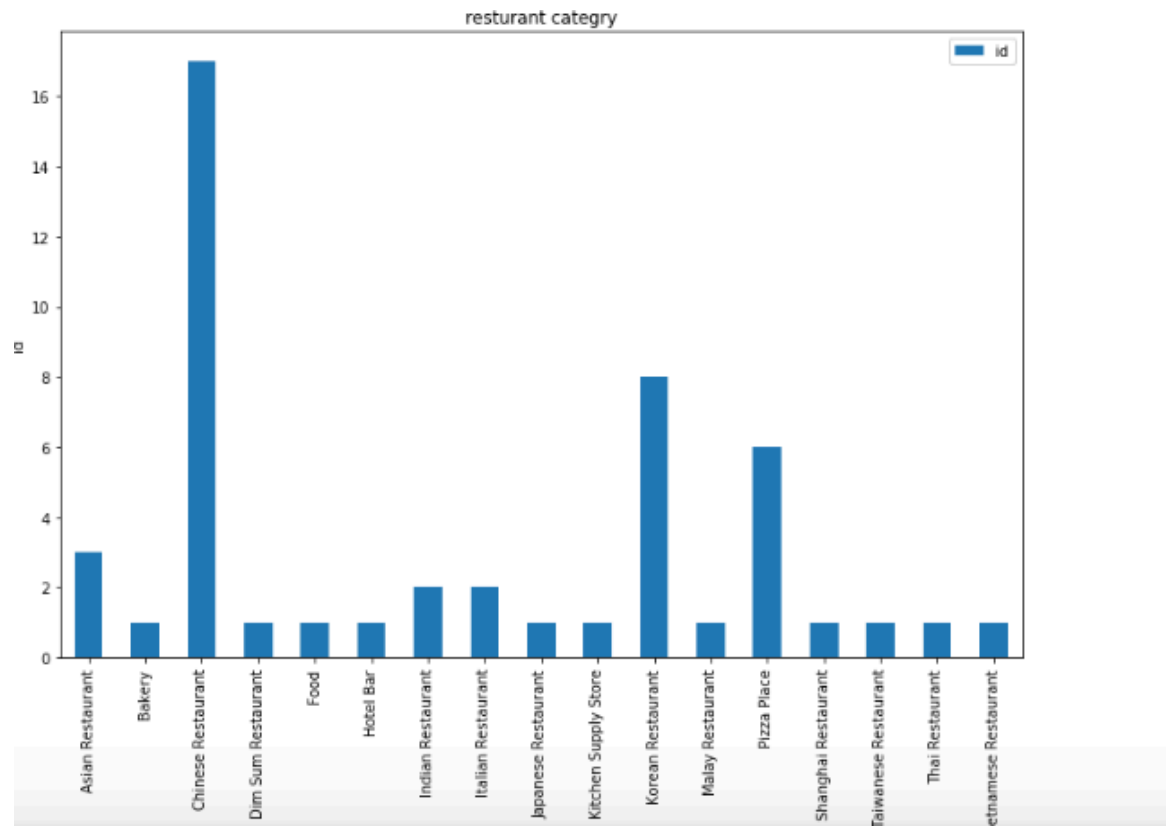
	name	categories	lat	lng
0	SriPraPhai	Thai Restaurant	40.746342	-73.899248
1	Fay Da Bakery	Bakery	40.745836	-73.902115
2	Izalco	Latin American Restaurant	40.746060	-73.899753
3	Jollibee	Fast Food Restaurant	40.745986	-73.901197
4	Rainbow Bakery	Bakery	40.745407	-73.903423

- We further utilize Foursquare API and import more data about the venues as shown in the below table with last column for ID's for each venue.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id
0	Lake Pavilion Restaurant	Chinese Restaurant	60-15 Main St	US	Flushing	United States	at Horace Harding Expressway	2440	[60-15 Main St (at Horace Harding Expressway),...	[[{"label": "display", "lat": 40.742622, "lng": -73.824970}]]	40.742622	-73.824970	NaN	11355	NY	4f844c6fe4b0069705cb042c
1	East Manor Buffet & Restaurant	Chinese Restaurant	4207 Main St	US	Flushing	United States	at Maple Ave.	2525	[4207 Main St (at Maple Ave.), Flushing, NY 11...	[[{"label": "display", "lat": 40.756112, "lng": -73.826404}]]	40.756112	-73.826404	NaN	11355	NY	4a1175a1f964a52018771fe3
2	Punjabi Indian & American Restaurant	Indian Restaurant	160-20 46th Ave	US	Flushing	United States	160st and 46th ave	833	[160-20 46th Ave (160st and 46th ave), Flushing...	[[{"label": "display", "lat": 40.753916, "lng": -73.805918}]]	40.753916	-73.805918	NaN	11358	NY	4c60797a90b2c9b648e03c22
3	Asian Jewels Seafood Restaurant 都城海鮮酒家	Chinese Restaurant	133-30 39th Ave	US	Flushing	United States	btw College Point Blvd & Janet Pl	3180	[133-30 39th Ave (btw College Point Blvd & Janet Pl...	[[{"label": "display", "lat": 40.759209, "lng": -73.833253}]]	40.759209	-73.833253	NaN	11354	NY	4a1845c8f964a52092791fe3
4	New Mulan Seafood Restaurant	Dim Sum Restaurant	136-17 39th Ave	US	New York	United States	at Main St	2986	[136-17 39th Ave (at Main St), New York, NY 11...	[[{"label": "display", "lat": 40.760285, "lng": -73.830250}]]	40.760285	-73.830250	Downtown Flushing	11354	NY	59ecbd4ce97dfb5a1b7e2fd8
5	Tong Sam Gyup Goo Yi Restaurant	Korean Restaurant	16223 Depot Rd	US	Flushing	United States	NaN	1438	[16223 Depot Rd, Flushing, NY 11358, 1 labeled St	[[{"label": "display", "lat": 40.762059, "lng": -73.803123}]]	40.762059	-73.803123	NaN	11358	NY	4bd3578041b9ef3b05f0ffe6

- We had filtered the data for names and located them in map as shown below.
- We further grouped data to utilize for plotting using matplotlib library to show which categories of restaurants are more available in the selected neighborhood.





Result

We can see from the bar chart that Chinese Restaurant followed by Korean and Pizza place are the category, which are more in number in the selected Neighborhood.