

## Datasets used for New York restaurants cuisine analysis project

- 1] Foursquare api from foursquare city guide platform
- 2] New York City neighborhoods data
- 3] Geopy package used to convert address to coordinates.

- 1] Foursquare api from foursquare city guide platform:

We use the foursquare api to acquire the top 20 restaurants in each neighborhood of new York and their cuisine types.

The syntax for fetching for one neighborhood is:

```
url =  
'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, lat, lon, VERSION, search_query, radius, LIMIT)  
  
results = requests.get(url).json()
```

This returns a json object with detailed information of each of the top restaurants. But this project just needs the name and cuisine from this json object.

The name is in `results.response.venues[].name`

The cuisine is in `results.response.venues[].categories[0].shortname`

- 2] New York City neighborhoods data

We download the dataset from [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) or [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) to fetch the names, borough, coordinates of every New York neighborhood.

The dataset looks like this:

	name	borough	lat	lon
0	Wakefield	Bronx	40.8947	-73.8472
1	Co-op City	Bronx	40.8743	-73.8299
2	Eastchester	Bronx	40.8876	-73.8278
3	Fieldston	Bronx	40.8954	-73.9056
4	Riverdale	Bronx	40.8908	-73.9126
...	...	...	...	...
301	Hudson Yards	Manhattan	40.7567	-74.0001
302	Hammels	Queens	40.5873	-73.8055
303	Bayswater	Queens	40.6113	-73.766
304	Queensbridge	Queens	40.7561	-73.9456
305	Fox Hills	Staten Island	40.6173	-74.0817

306 rows × 4 columns

We use the dataset acquired from foursquare api and the above to form the following dataset:

	neighborhood	borough	lat	lon	name	type
0	Wakefield	Bronx	40.894705	-73.847201	Bay 241 Restaurant & Lounge	Caribbean
1	Wakefield	Bronx	40.894705	-73.847201	Kaieteur Restaurant & Bakery	Caribbean
2	Wakefield	Bronx	40.894705	-73.847201	Big Daddy's Caribbean Taste Restaurant	Caribbean
3	Wakefield	Bronx	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	Caribbean
4	Wakefield	Bronx	40.894705	-73.847201	Cool Running Restaurant	Food
5	Wakefield	Bronx	40.894705	-73.847201	Fort Negril Jamaican Restaurant	Food
6	Wakefield	Bronx	40.894705	-73.847201	Allure Restaurant And Bar	Restaurant
7	Wakefield	Bronx	40.894705	-73.847201	Golden Krust Caribbean Restaurant	Caribbean
8	Wakefield	Bronx	40.894705	-73.847201	241 St Cafe & Restaurant	American
9	Wakefield	Bronx	40.894705	-73.847201	Silver Chimes Restaurant	Food

The dataset will further be compressed by grouping rows by neighborhood and one hot encoding it to get no of restaurants in each neighborhood by their cuisines and train the kmeans model to cluster the neighborhoods.

### 3] Geopy package

The geopy package is used to convert an address to its relevant coordinated we can use in folium map plotting or in foursquare api.

The syntax to convert address “New York” to its coordinates is:

```
geolocator = Nominatim(user_agent="NY")
location = geolocator.geocode("New York")
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of New York are 40.7127281, -74.0060152.

These three are what will be used to get all the data needed for the project.