Capstone Project: The Battle of Neighborhoods

Optimal Location for a New African Restaurant in New York City

- New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.
- This final project explores the best locations for African restaurants in New York City whereby the investor or entrepreneur can have great success and consistent profit. So our project will attempt to answer the questions “What type of restaurant would it be?” and “Where would be the best location for it?”

Target Audience

Business personnel who wants to invest or open a restaurant.
Freelancer who loves to have their own restaurant as a side business.

Data Cleaning

For this project, we need the following data:

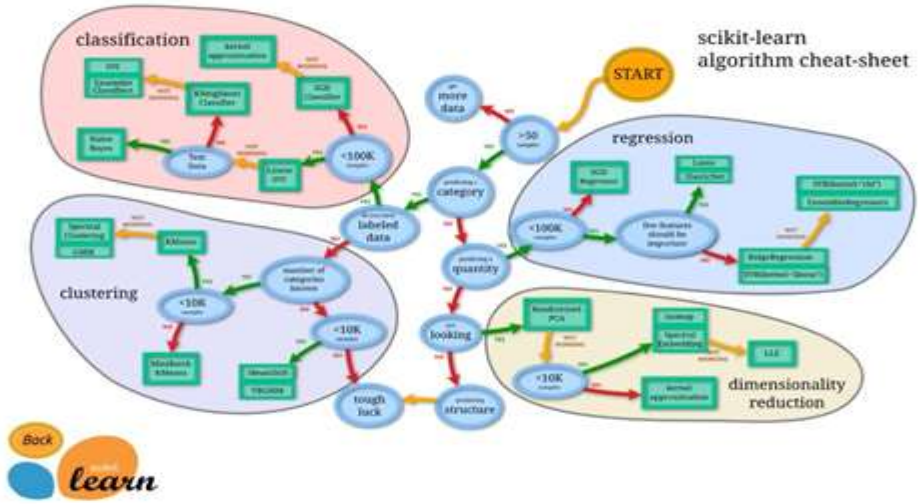
New York City data that contains Borough, Neighborhoods along with their latitudes and longitudes.

Data Source: https://cocl.us/newyorkdataset

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Exploring New York Restaurants

We utilized the Foursquare API to get top 100 restaurants in each neighbourhood.
We applied one-hot encoding and grouped them by taking the mean of the frequency of occurrence of each type.
Then we clustered restaurants using the k-means algorithm based on their types similarity. The k-means is an unsupervised machine learning algorithm for clustering unlabelled data.



Using Foursquare API, collect the top 100 restaurants and their categories for each location within a radius 500 meters.

Now, let's get the top 100 venues that are in New York City within a radius of 500 meters.

First, let's create the GET request URL. Name your URL url.

```
LIMIT = 100
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    latitude,
    longitude,
    radius,
    LIMIT)
url
```

```
'https://api.foursquare.com/v2/venues/explore?&client_id=GNAXGLCD2HMU1QX5CL04JNPAEM2YEJMKTRZCHFDCZICK2OCP&client_secret=C1HCZYM0INCRXSR5K
BPTQGASPD24NDJGDAM3BL2XTGIZCR3O&v=20180604&ll=40.7127281,-74.0060152&radius=500&limit=100'
```

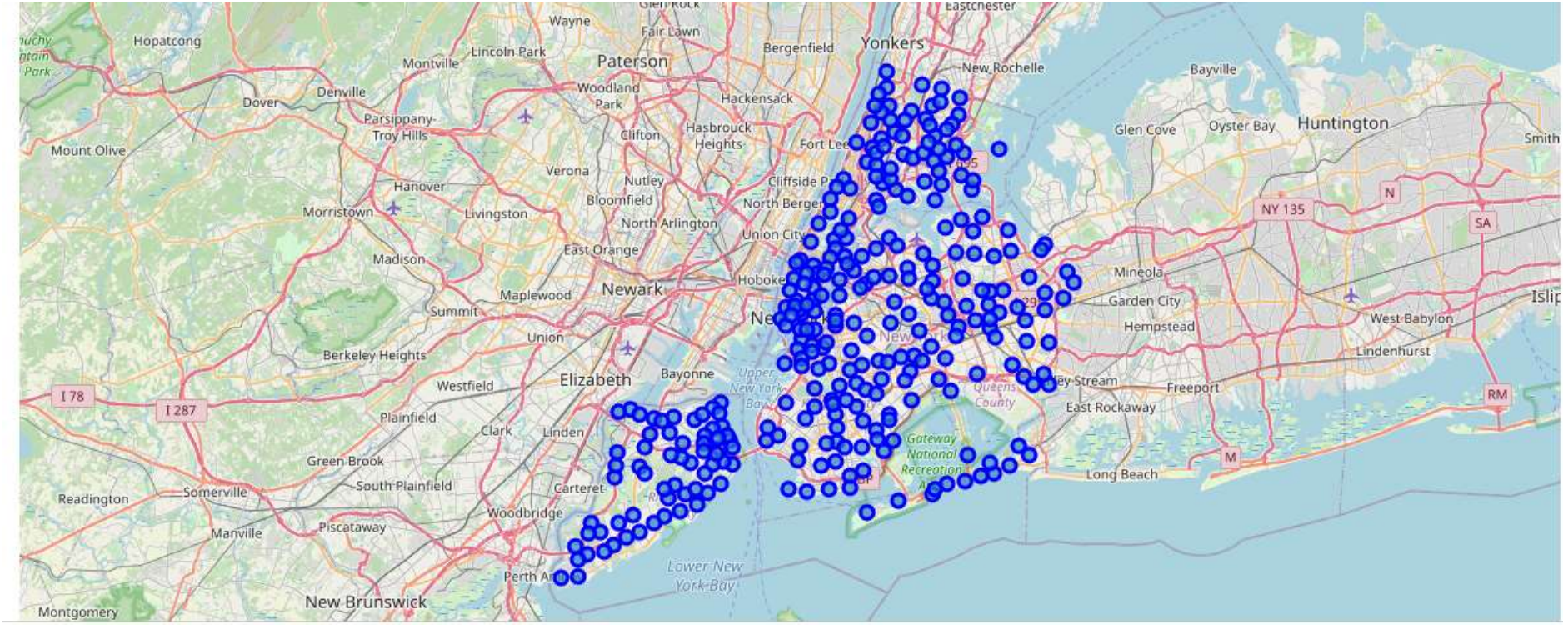
Send the GET request and examine the results

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '5f1097f7d732ec1c90803029'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'},
    {'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Downtown Manhattan',
  'headerFullLocation': 'Downtown Manhattan, New York',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 106,
  'suggestedBounds': {'ne': {'lat': 40.7172281045, 'lng': -74.00008952063419},
   'sw': {'lat': 40.7082280955, 'lng': -74.0119408793658}},
```

Exploring New York City neighborhoods

1. Using Wikipedia, we collected all of New York's neighborhoods.
2. Using geospatial libraries, we added the geographical coordinates.
3. With some cleaning and wrangling, we obtained 288 locations in New York City.



Using the one-hot encoding

4. Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data.

Exploring Restaurants

To begin analysis we need to transform collected information using the one-hot encoding method.

```
# one hot encoding
newyork_onehot = pd.get_dummies(newyork_venues[['Venue Category']], prefix="", prefix_sep="")

# add location column back to dataframe
newyork_onehot['Neighborhood'] = newyork_venues['Neighborhood']

# move location column to the first column
fixed_columns = [newyork_onehot.columns[-1]] + list(newyork_onehot.columns[:-1])
newyork_onehot = newyork_onehot[fixed_columns]

newyork_onehot.head()
```

	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertai
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

Cluster restaurants by k-means algorithm and analyze the top 10 most common restaurants in each cluster.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allerton	Pizza Place	Supermarket	Deli / Bodega	Department Store	Pharmacy	Chinese Restaurant	Bakery	Check Cashing Service	Grocery Store	Martial Arts Dojo
1	Annadale	Pizza Place	Bar	Restaurant	Pub	Train Station	Pharmacy	Food	Cosmetics Shop	American Restaurant	Liquor Store
2	Arden Heights	Bus Stop	Coffee Shop	Pharmacy	Pizza Place	Women's Store	Fast Food Restaurant	Event Service	Event Space	Exhibit	Eye Doctor
3	Arlington	Bus Stop	Intersection	Home Service	Deli / Bodega	Boat or Ferry	American Restaurant	Grocery Store	Falafel Restaurant	Fast Food Restaurant	Farmers Market
4	Arrochar	Bus Stop	Deli / Bodega	Italian Restaurant	Athletics & Sports	Nail Salon	Supermarket	Middle Eastern Restaurant	Mediterranean Restaurant	Outdoors & Recreation	Bagel Shop
5	Arverne	Surf Spot	Sandwich Place	Metro Station	Restaurant	Café	Bed & Breakfast	Coffee Shop	Board Shop	Beach	Donut Shop
6	Astoria	Bar	Middle Eastern Restaurant	Greek Restaurant	Mediterranean Restaurant	Hookah Bar	Seafood Restaurant	Pizza Place	Bakery	Indian Restaurant	Café
7	Astoria Heights	Burger Joint	Playground	Bakery	Bowling Alley	Supermarket	Hostel	Pizza Place	Business Service	Plaza	Italian Restaurant
8	Auburndale	Italian Restaurant	Miscellaneous Shop	Bar	Toy / Game Store	Pet Store	Pharmacy	Korean Restaurant	Supermarket	Fast Food Restaurant	Athletics & Sports
9	Bath Beach	Pizza Place	Pharmacy	Italian Restaurant	Deli / Bodega	Bubble Tea Shop	Fast Food Restaurant	Cantonese Restaurant	Chinese Restaurant	Gas Station	Park

Results

Cluster #1

[38]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	157	157	157	157	157	157	157	157	157	157	157
unique	153	53	69	75	76	74	78	76	89	81	91
top	Chelsea	Italian Restaurant	Coffee Shop	Pizza Place	Pizza Place	Bakery	Pizza Place	Bakery	Exhibit	Exhibit	Eye Doctor
freq	2	26	12	12	12	12	8	9	8	8	8

Chelsea Neighborhood:

The most common venue: Italian Restaurant.
The least common venue: Eye Doctor.
There are three Pizza Places on the 3rd, 4th and 6th venues respectively.
Also, a coffee shop is on the 2nd most common venue.
No African Restaurants in this cluster.

Cluster #2

[40]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	2	2	2	2	2	2	2	2	2	2	2
unique	2	1	1	1	1	1	1	1	1	1	1
top	Todt Hill	Park	Women's Store	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant	Farm
freq	1	2	2	2	2	2	2	2	2	2	2

Todt Hill Neighborhood:

The most common venue: Park.

There's an Ethiopian Restaurant on the 3rd most common venue.

There's also another restaurant on the 9th most common venue.

The least common venue: Farm.

We can see that an African Restaurant exists on Todt Hill.

Cluster #3

[42]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	10	10	10	10	10	10	10	10	10	10	10
unique	10	5	8	8	9	8	6	7	8	8	8
top	Breezy Point	Beach	Beach	Deli / Bodega	Event Service	Bus Stop	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant
freq	1	6	2	3	2	2	3	3	3	3	3

Breezy Point Neighborhood:
The most common venue is a Beach.
The least common venue: Falafel Restaurant.
There is no African Restaurant in this cluster.

Cluster #4

[44]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	1	1	1	1	1	1	1	1	1	1	1
unique	1	1	1	1	1	1	1	1	1	1	1
top	Port Ivory	Bar	Women's Store	Event Service	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant	Farm
freq	1	1	1	1	1	1	1	1	1	1	1

Port Ivory:

The most common venue: Bar.

The least common venue: Farm.

We can clearly observe a restaurant on the 9th most common venue: Falafel Restaurant.

No African Restaurant.

Cluster #5

[46]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	136	136	136	136	136	136	136	136	136	136	136
unique	136	43	52	59	68	66	73	67	75	73	72
top	Crown Heights	Pizza Place	Chinese Restaurant	Pizza Place	Pizza Place	Sandwich Place	Sandwich Place	Pizza Place	Fast Food Restaurant	Bank	Eye Doctor
freq	1	22	11	10	13	11	7	7	6	9	7

Crown Heights Neighborhood:

The most common venue is a Pizza Place.

The least common venue is an Eye Doctor.

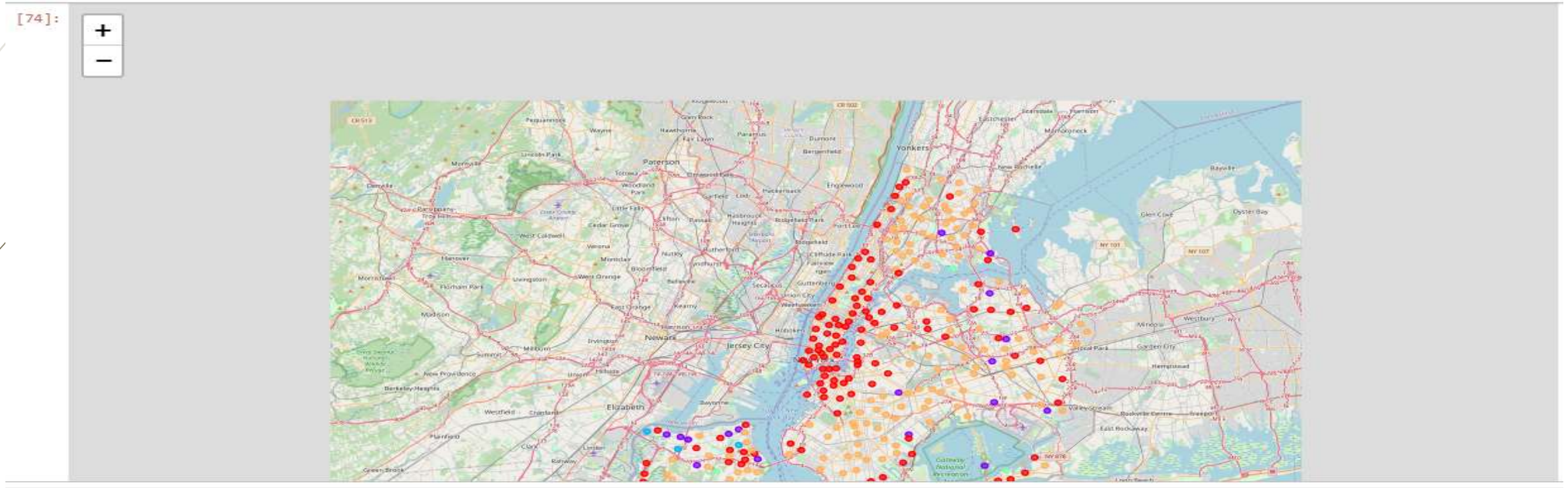
There are many restaurants and pizza spots in this neighborhood.

No African restaurant in this cluster.

Visualizing Clusters

Visualize clusters on the map, thus showing the best locations for opening the chosen restaurant.

[74]:



Cluster 1 – red
Cluster 2 – purple
Cluster 3 – blue
Cluster 4 – green
Cluster 5 - orange

Discussion

Analyzing the most popular restaurants in each cluster, the stakeholder should select the least popular types within the top 10 as a safe choice. In our recommendations, we advise selecting from the 10^{th} or 9^{th} positions. This selection is a reasonable balance between being too popular and having no customers.

Recommendations, based on description of each cluster:

Cluster 1 neighborhood: No most or least common African Restaurant, so it's not advisable to open in Chelsea neighborhood.

Cluster 2 neighborhood: There is an Ethiopian Restaurant in Todt Hill which makes a good location to open an african restaurant as African customers visit this venue.

Cluster 3 to 5 neighbourhood have no African restaurant which is also a bad idea to open a new one at any of the venues.

So cluster 2 is the ideal location to open an African restaurant.

Conclusion

- In this report, we established a methodology to determine what the most promising type of restaurant is, and where it should be opened.
- This type of analysis can be applied to any city of your choice that has available geospatial information.
- This type of analysis can be applied to any type of venue that is available in the Foursquare database.