

# CAPSTONE PROJECT

## The Battle of Neighborhoods

### Select Restaurant by Score

José Gabriel Rosas

Out[1]:

[Click here to toggle on/off the raw code.](#)

Start JGR

# 1. New York Restaurants

When visiting New York, for sure you are going to look for something different to eat and try some international cuisine. But, if you don't know where to look and want to be sure that it is well rated, but not just because of likes in social media, then, there is a study from the Department of Health and Mental Hygiene from New York city, that grades the restaurants.

The final objective looks for the type of cuisine by Neighborhood and its grades, so you can visit with confidence the restaurant you liked.

## DOHMH New York City Restaurant Inspection Results

The following Inspection Results from the DOHMH (Department of Health and Mental Hygiene from New York city), is defined in order to look for the best restaurants by type of cuisine by Borough.

Find below a small description from the DOHMH.

"The dataset contains every sustained or not yet adjudicated violation citation from every full or special program inspection conducted up to three years prior to the most recent inspection for restaurants and college cafeterias in an active status on the RECORD DATE (date of the data pull). When an inspection results in more than one violation, values for associated fields are repeated for each additional violation record. Establishments are uniquely identified by their CAMIS (record ID) number."

## 2. Gathering Data

### 2.1 Data from DOHMH (Department of Health and Mental Hygiene from New York city)

The data has been taken from the link below and saved as an csv file

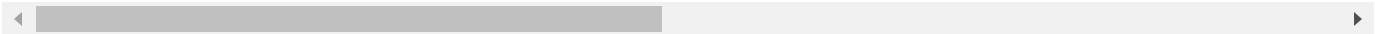
<https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/data>  
(<https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/data>)

#### The file has been downloaded and read as Dataframe

The Dataframe has 384,487 rows and 18 columns, but it is required to do some clean up of the file to get just the information for the purpose of this analysis.

Out[4]:

	CAMIS	DBA	BORO	BUILDING	STREET	ZIPCODE	PHONE	CUISINE DESCRIPTION
0	41638031	MAMAJUANA CAFE	BRONX	3233	EAST TREMONT AVENUE	10461.0	7188248400	Spanish
1	50045811	BROTHERS CREPES	QUEENS	4202	30TH AVE	11103.0	7186060494	Juice, Smoothies, Fruit Salads
2	40739444	DUNKIN' DONUTS	QUEENS	5602	METROPOLITAN AVENUE	11385.0	7183813200	Donuts
3	50058559	PRET A MANGER	MANHATTAN	299	7TH AVE	10001.0	6467280505	American
4	50069855	FERRY HOUSE CAFE	MANHATTAN	PIER 11	SOUTH ST	10005.0	9175363918	CafÃ©/Coffee/Tea



(384487, 18)

Drop the columns that aren't useful for this purpose

Out[6]:

	DBA	BORO	STREET	ZIPCODE	CUISINE DESCRIPTION	SCORE	GRADE
0	MAMAJUANA CAFE	BRONX	EAST TREMONT AVENUE	10461.0	Spanish	9.0	A
1	BROTHERS CREPES	QUEENS	30TH AVE	11103.0	Juice, Smoothies, Fruit Salads	9.0	A
2	DUNKIN' DONUTS	QUEENS	METROPOLITAN AVENUE	11385.0	Donuts	11.0	A
3	PRET A MANGER	MANHATTAN	7TH AVE	10001.0	American	NaN	NaN
4	FERRY HOUSE CAFE	MANHATTAN	SOUTH ST	10005.0	CafÃ©/Coffee/Tea	94.0	C

Out[7]: (384487, 7)

Sort by Borough, Zipcode, Type of Cuisine and Score of Restaurant

Out[8]:

	DBA	BORO	STREET	ZIPCODE	CUISINE DESCRIPTION	SCORE	GRADE
24166	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	55.0	NaN
79128	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	55.0	NaN
125062	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	55.0	NaN
242243	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	55.0	NaN
259089	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	55.0	NaN

Out[9]: (384487, 7)

### Delete restaurant duplicates

Now the Dataframe has been cleaned a little bit, but it can be improved. When some of the columns were deleted, then with the remaining ones, there are many duplicates, so those duplicates are going to be removed. So, from 384,487 rows, now there are 21,292 rows with cleaner data.

Out[10]:

	DBA	BORO	STREET	ZIPCODE	CUISINE DESCRIPTION	SCORE	GRADE
345824	BALIMAYA RESTAURANT	BRONX	3 AVENUE	10451.0	African	10.0	A
124291	RESTAURANT DELICES DES SAVEURS	BRONX	MORRIS AVE	10451.0	African	NaN	NaN
296376	NABAYA AFRICAN AMERICAN RESTUARANT	BRONX	MELROSE AVE	10451.0	African	NaN	NaN
266210	YANKEE TAVERN	BRONX	EAST 161 STREET	10451.0	American	12.0	A
285479	STAN'S SPORTS BAR	BRONX	RIVER AVENUE	10451.0	American	11.0	A

(21292, 7)

### Delete not scored

We don't really want to have in our dataset those restaurants that haven't been rated yet, we do really want to know their grades, so those with NaN in their Scores, are going to be dropped.

So from those 21,292 rows, the dataset now has 12,154. A whole new different quantity from the amount of rows at the beginning.

Out[12]:

	DBA	BORO	STREET	ZIPCODE	CUISINE DESCRIPTION	SCORE	GRADE
345824	BALIMAYA RESTAURANT	BRONX	3 AVENUE	10451.0	African	10.0	A
266210	YANKEE TAVERN	BRONX	EAST 161 STREET	10451.0	American	12.0	A
285479	STAN'S SPORTS BAR	BRONX	RIVER AVENUE	10451.0	American	11.0	A
320150	BULLPEN DELI TWIN DONUT	BRONX	E 161ST ST	10451.0	American	11.0	A
64583	BLUE POINT BLEACHERS BAR 237	BRONX	E 161ST ST	10451.0	American	10.0	A

Out[13]: (11216, 7)

## Rename Columns

This is the Dataframe containing the name of the restaurant, its Borough, Zip code, Type of Cuisine and the score.

It is very useful to be consistent and always use the same way to name the columns and type of data.

### *rest.name* Dataframe

This is the final result of this first step

rest.name Dataframe

Out[14]:

	RESTAURANT	BOROUGH	STREET	ZIPCODE	CUISINE	SCORE	GRADE
345824	BALIMAYA RESTAURANT	BRONX	3 AVENUE	10451.0	African	10.0	A
266210	YANKEE TAVERN	BRONX	EAST 161 STREET	10451.0	American	12.0	A
285479	STAN'S SPORTS BAR	BRONX	RIVER AVENUE	10451.0	American	11.0	A
320150	BULLPEN DELI TWIN DONUT	BRONX	E 161ST ST	10451.0	American	11.0	A
64583	BLUE POINT BLEACHERS BAR 237	BRONX	E 161ST ST	10451.0	American	10.0	A

## 2.2 ZIP Code Definitions of New York City Neighborhoods

The last dataset, rest, has the Borough name and its ZIP codes, but we want to know a little more detail about the neighborhood and the geographical coordinates, so we can use the Foresquare information.

For this purpose, from the Department of Health of New York State, I copied the table, organized it in Excel and the converted it into a CSV file. It has some behind scenes manipulation in Excel to get it converted in this file.

<https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>  
(<https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>)

Below, you'll find the CSV file and how it was manipulated to get the ZIP codes by each Neighborhood, so they can be linked with the Restaurant Names, rest.names, Dataframe

Out[15]:

	Borough	Neighborhood	zip1	zip2	zip3	zip4	zip5	zip6	zip7	zip8	zip9
0	Bronx	Central Bronx	10453	10457.0	10460.0	NaN	NaN	NaN	NaN	NaN	NaN
1	Bronx	Bronx Park and Fordham	10458	10467.0	10468.0	NaN	NaN	NaN	NaN	NaN	NaN
2	Bronx	High Bridge and Morrisania	10451	10452.0	10456.0	NaN	NaN	NaN	NaN	NaN	NaN
3	Bronx	Hunts Point and Mott Haven	10454	10455.0	10459.0	10474.0	NaN	NaN	NaN	NaN	NaN
4	Bronx	Kingsbridge and Riverdale	10463	10471.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Out[16]:

	Borough	Neighborhood	variable	value
373	Queens	West Queens	zip9	NaN
374	Staten Island	Port Richmond	zip9	NaN
375	Staten Island	South Shore	zip9	NaN
376	Staten Island	Stapleton and St. George	zip9	NaN
377	Staten Island	Mid-Island	zip9	NaN

Out[17]:

	Borough	Neighborhood	variable	value
288	Queens	Southwest Queens	zip7	11420.0
289	Queens	West Queens	zip7	11378.0
329	Queens	Southeast Queens	zip8	11428.0
330	Queens	Southwest Queens	zip8	11421.0
371	Queens	Southeast Queens	zip9	11429.0

Out[18]:

	Borough	Neighborhood	variable	ZIPCODE
0	Bronx	Central Bronx	zip1	10453.0
1	Bronx	Bronx Park and Fordham	zip1	10458.0
2	Bronx	High Bridge and Morrisania	zip1	10451.0
3	Bronx	Hunts Point and Mott Haven	zip1	10454.0
4	Bronx	Kingsbridge and Riverdale	zip1	10463.0

## 2.3 Get the coordinates by each ZIP code

Now we have the Borough name, the Neighborhood and the ZIP codes, let's look for the coordinates of each one.

From the US Census Bureau, Gazetteer Files I downloaded the file and save it as csv

ZIP Code Tabulation Areas <https://www.census.gov/geographies/reference-files/2017/geo/gazetter-file.html>  
(<https://www.census.gov/geographies/reference-files/2017/geo/gazetter-file.html>)

zip file [https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2017\\_Gazetteer/2017\\_Gaz\\_zcta\\_national.zip](https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2017_Gazetteer/2017_Gaz_zcta_national.zip)  
([https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2017\\_Gazetteer/2017\\_Gaz\\_zcta\\_national.zip](https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2017_Gazetteer/2017_Gaz_zcta_national.zip))

Out[19]:

	GEOID	ALAND	AWATER	ALAND_SQMI	AWATER_SQMI	INTPTLAT	INTPTLONG
0	601	166659749	799292	64.348	0.309	18.180555	-66.749961
1	602	79307535	4428429	30.621	1.710	18.361945	-67.175597
2	603	81887185	181411	31.617	0.070	18.455183	-67.119887
3	606	109579993	12487	42.309	0.005	18.158327	-66.932928
4	610	93013427	4172059	35.913	1.611	18.295366	-67.125135

Out[20]:

	ZIPCODE	LATITUDE	LONGITUDE
0	601	18.180555	-66.749961
1	602	18.361945	-67.175597
2	603	18.455183	-67.119887
3	606	18.158327	-66.932928
4	610	18.295366	-67.125135

## 2.4 MERGING

Now it's time to merge all the information.

*Merge nyzips with zipcoor*

Out[21]:

	Borough	Neighborhood	variable	ZIPCODE	LATITUDE	LONGITUDE
0	Bronx	Central Bronx	zip1	10453.0	40.852820	-73.912310
1	Bronx	Bronx Park and Fordham	zip1	10458.0	40.862529	-73.888159
2	Bronx	High Bridge and Morrisania	zip1	10451.0	40.820454	-73.925066
3	Bronx	Hunts Point and Mott Haven	zip1	10454.0	40.805492	-73.916604
4	Bronx	Kingsbridge and Riverdale	zip1	10463.0	40.880678	-73.906540

*Merge the last merge with the restaurant dataframe*

Out[22]:

	Borough	Neighborhood	variable	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	BOROUGH	STF
0	Bronx	Central Bronx	zip1	10453.0	40.85282	-73.91231	GALAGALA NY RESTAURANT	BRONX	E 175T
1	Bronx	Central Bronx	zip1	10453.0	40.85282	-73.91231	BEVERLY HILL MANOR	BRONX	JEROME
2	Bronx	Central Bronx	zip1	10453.0	40.85282	-73.91231	MI BAKERY & CAFE	BRONX	W 183R
3	Bronx	Central Bronx	zip1	10453.0	40.85282	-73.91231	MOUNT HOPE BAKERY	BRONX	MC HOP
4	Bronx	Central Bronx	zip1	10453.0	40.85282	-73.91231	MI CASA BAKERY & COFFEE SHOP	BRONX	GR CONCOU

Drop the columns that we don't want. It is twice the Borough column.

The variable column, was just a remaining column from the 'melting' process, but it is useless.

## 2.5 Unified Dataframe and Map

Show the last dataframe, the one that's going to be used in the map.

Now it has 10,884 rows and 10 rearranged columns

Out[24]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCOF
0	Bronx	Central Bronx	10453.0	40.85282	-73.91231	GALAGALA NY RESTAURANT	E 175TH ST	African	3
1	Bronx	Central Bronx	10453.0	40.85282	-73.91231	BEVERLY HILL MANOR	JEROME AVE	American	42
2	Bronx	Central Bronx	10453.0	40.85282	-73.91231	MI BAKERY & CAFE	W 183RD ST	Bakery	6
3	Bronx	Central Bronx	10453.0	40.85282	-73.91231	MOUNT HOPE BAKERY	MOUNT HOPE PL	Bakery	4
4	Bronx	Central Bronx	10453.0	40.85282	-73.91231	MI CASA BAKERY & COFFEE SHOP	GRAND CONCOURSE	Bakery	3

Out[25]: (10884, 10)

The geograpical coordinate of New York City are 40.7308619, -73.9871558.



Out[27]:

### 3. Selecting Cuisine

#### How Many Restaurants, Types of Cuisine and Score Distribution by Borough

A brief analysis of what you can find in each Borough.

First, make a Dataframe to summarize the results, and after that, start to make a decision

Out[28]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCORE
Borough									
	Bronx	1	7	25	25	25	992	327	47
	Brooklyn	1	11	36	36	36	2529	537	73
	Manhattan	1	10	43	43	43	4301	849	74
	Queens	1	10	58	58	58	2587	493	72
	Staten Island	1	4	12	12	12	475	134	48

Out[29]:

	RESTAURANT	CUISINE	Restaurant/Cuisine
Borough			
Bronx	992	47	21.106383
Brooklyn	2529	73	34.643836
Manhattan	4301	74	58.121622
Queens	2587	72	35.930556
Staten Island	475	48	9.895833

So it is showing that Manhattan has the maximum amount of restaurants with 4301 and Staten Island the minimum with 475.

Manhattan, Brooklyn and Queens, mainly have the same amount of types of Cuisines, 74,73,72 respectively.

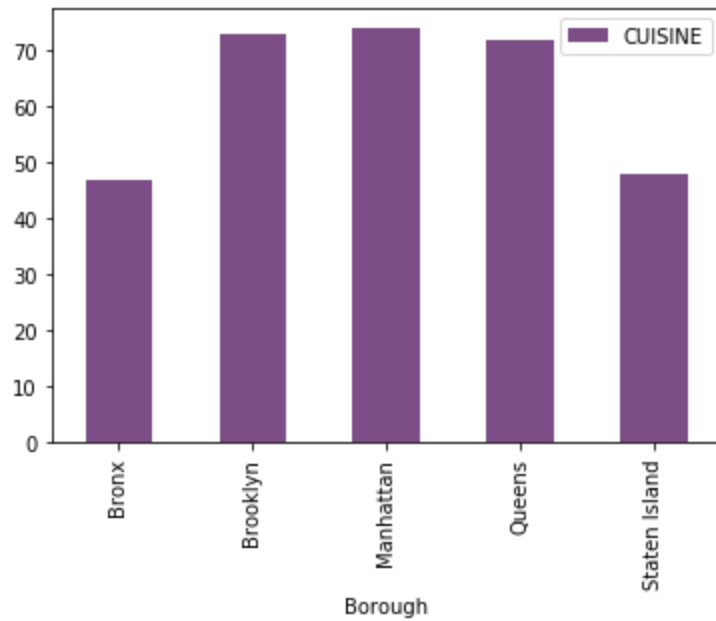
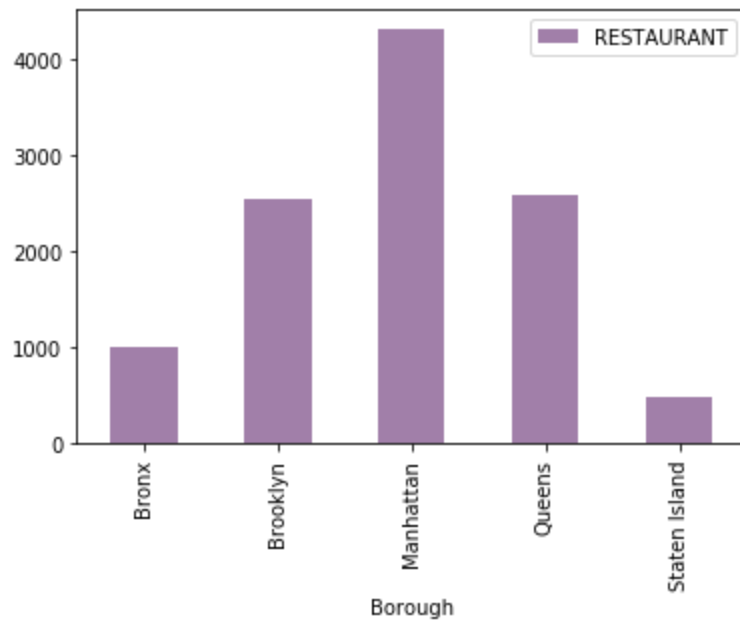
And Bronx and Staten Island have mainly the same amount, 47 and 48

Let's look at those findings in a visual way

How many restaurants per Borough and how many types of cuisines per Borough

Out[31]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eedfc6e4a8>

<Figure size 648x216 with 0 Axes>



### Select the type of Cuisine that we want

List the type of cuisines and select 5 types to analyze the scores of their restaurants.

First identify the different types of cuisines

```
Out[32]: array(['African', 'American', 'Bakery', 'Caribbean', 'Chinese',
'Delicatessen', 'Juice, Smoothies, Fruit Salads',
'Latin (Cuban, Dominican, Puerto Rican, South & Central American)',
'Mexican', 'Pizza', 'Pizza/Italian', 'Seafood', 'Spanish',
'Barbecue',
'Bottled beverages, including water, sodas, juices, etc.',
'Café/Coffee/Tea', 'Chicken', 'Eastern European',
'Ice Cream, Gelato, Yogurt, Ices', 'Italian', 'Japanese', 'Other',
'Sandwiches/Salads/Mixed Buffet', 'Soul Food', 'Bagels/Pretzels',
'Sandwiches', 'Turkish', 'Steak', 'Tex-Mex', 'Asian', 'Greek',
'Indian', 'Jewish/Kosher', 'Thai', 'Hawaiian', 'Chinese/Cuban',
'Hamburgers', 'Irish', 'Czech', 'Donuts', 'Egyptian',
'Mediterranean', 'Middle Eastern', 'Moroccan', 'Polish',
'Chinese/Japanese', 'Russian', 'Peruvian',
'Vietnamese/Cambodian/Malaysia', 'French', 'Cajun', 'Continental',
'Pakistani', 'Salads', 'Tapas', 'Hotdogs', 'Korean', 'Vegetarian',
'Soups & Sandwiches', 'Ethiopian', 'Fruits/Vegetables', 'Armenian',
'Californian', 'English', 'Bangladeshi', 'Brazilian', 'Creole',
'Filipino', 'German', 'Indonesian', 'Not Listed/Not Applicable',
'Nuts/Confectionary', 'Australian', 'Portuguese', 'Afghan',
'Hotdogs/Pretzels', 'Chilean', 'Pancakes/Waffles', 'Creole/Cajun',
'Scandinavian', 'Southwestern', 'Soups'], dtype=object)
```

Selection

If somebody wants to try international cuisines, New York is the place.

But we don't want to choose from ten thousand options, so let's select 5 types that you don't usually eat.

For this case, these are the selected cuisines:

African, Eastern European, Turkish, Creole, Afghan

Create a Dataframe just with the types of cuisines selected.

So, we got just 119 rows

(119, 10)

Out[33]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCOR
0	Bronx	Central Bronx	10453.0	40.852820	-73.912310	GALAGALA NY RESTAURANT	E 175TH ST	African	3
37	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	MB RESTAURANT	WEBSTER AVE	African	5
83	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CAKOR RESTAURANT	EAST 186 STREET	Eastern European	5
84	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CKA KA QELLU	HUGHES AVE	Eastern European	3
120	Bronx	High Bridge and Morrisania	10451.0	40.820454	-73.925066	BALIMAYA RESTAURANT	3 AVENUE	African	10

Grouping chart

How is the distribution amount of restaurants per Borough

Out[34]:

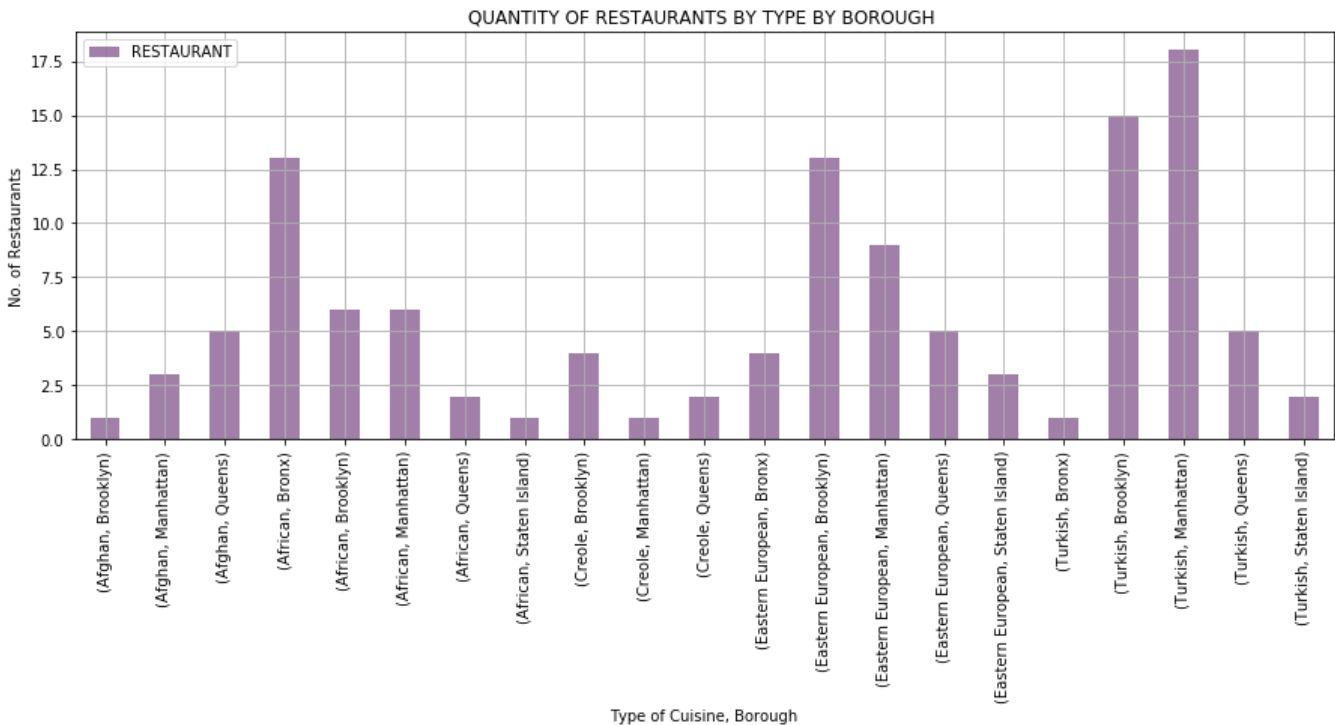
		Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET (
CUISINE		Borough						
Afghan	Brooklyn	1	1	1	1	1	1	1
	Manhattan	1	3	3	3	3	3	3
	Queens	1	4	4	4	4	5	5
African	Bronx	1	4	9	9	9	13	12
	Brooklyn	1	4	6	6	6	6	6
	Manhattan	1	4	6	6	6	6	6
	Queens	1	2	2	2	2	2	2
	Staten Island	1	1	1	1	1	1	1
Creole	Brooklyn	1	1	3	3	3	4	4
	Manhattan	1	1	1	1	1	1	1
	Queens	1	1	2	2	2	2	2
Eastern European	Bronx	1	3	3	3	3	4	4
	Brooklyn	1	3	6	6	6	13	11
	Manhattan	1	5	7	7	7	9	9
	Queens	1	2	3	3	3	5	5
	Staten Island	1	1	1	1	1	3	3
Turkish	Bronx	1	1	1	1	1	1	1
	Brooklyn	1	6	9	9	9	15	11
	Manhattan	1	5	10	10	10	18	17
	Queens	1	4	4	4	4	5	4
	Staten Island	1	2	2	2	2	2	2

Out[35]: count 119.000000  
mean 8.966387  
std 11.184078  
min 0.000000  
25% 5.000000  
50% 7.000000  
75% 10.000000  
max 94.000000  
Name: SCORE, dtype: float64

In which Borough is the majority of each type of the cuisines selected

Out[36]: Text(0.5, 1.0, 'QUANTITY OF RESTAURANTS BY TYPE BY BOROUGH ')

<Figure size 1080x576 with 0 Axes>



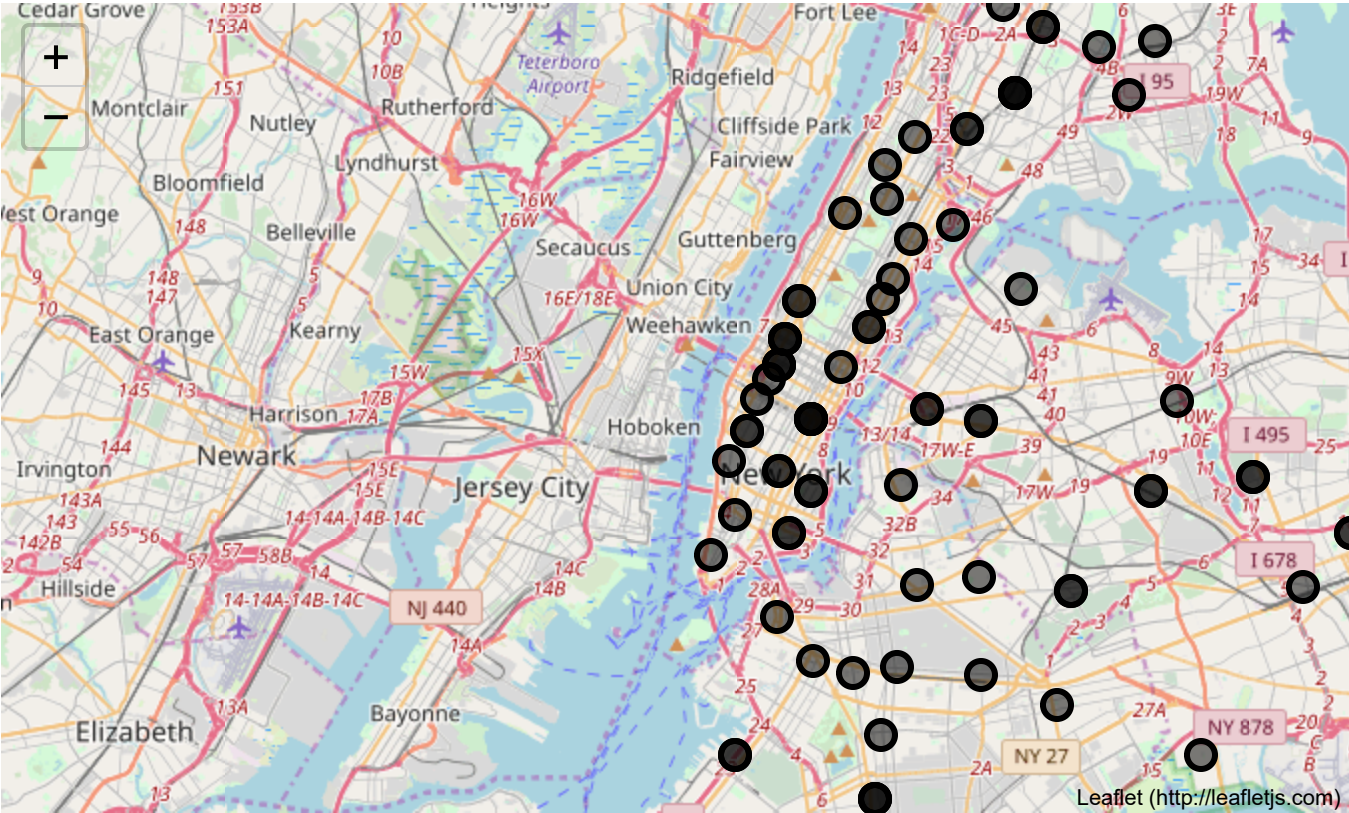
Where are the selected restaurants?

So, based on the last chart, the majority of restaurants of each cuisine are in:

- Afghan in Queens
- African in Bronx
- Creole in Brooklyn
- Eastern European in Brooklyn
- Turkish in Manhattan

Map of New York with the new distribution of restaurants with the selected cuisines

Out[37]:



4. CLUSTERING

Score by Restaurant

Out[38]:

	SCORE
0	3.0
37	5.0
83	5.0
84	3.0
120	10.0

```
Out[39]: array([[-5.35727163e-01],
 [ -3.56145382e-01],
 [ -3.56145382e-01],
 [-5.35727163e-01],
 [ 9.28090720e-02],
 [ 9.28090720e-02],
 [ 3.62181744e-01],
 [ 9.28090720e-02],
 [ 3.01818120e-03],
 [-1.76563600e-01],
 [-1.76563600e-01],
 [-6.25518054e-01],
 [-1.76563600e-01],
 [-3.56145382e-01],
 [ 2.72390854e-01],
 [ 2.72390854e-01],
 [ 9.28090720e-02],
 [-4.45936273e-01],
 [-4.45936273e-01],
 [ 9.28090720e-02],
 [ 3.01818120e-03],
 [-3.56145382e-01],
 [-4.45936273e-01],
 [ 2.72390854e-01],
 [-1.76563600e-01],
 [ 7.63524390e+00],
 [ 9.28090720e-02],
 [ 3.01818120e-03],
 [-1.76563600e-01],
 [ 9.28090720e-02],
 [-5.35727163e-01],
 [-8.67727096e-02],
 [ 9.28090720e-02],
 [ 3.62181744e-01],
 [-1.76563600e-01],
 [-1.76563600e-01],
 [-1.76563600e-01],
 [ 9.28090720e-02],
 [ 3.41507203e+00],
 [-1.76563600e-01],
 [ 9.28090720e-02],
 [-3.56145382e-01],
 [-4.45936273e-01],
 [ 3.62181744e-01],
 [-3.56145382e-01],
 [-2.66354491e-01],
 [ 9.28090720e-02],
 [-4.45936273e-01],
 [-8.67727096e-02],
 [ 1.82599963e-01],
 [-4.45936273e-01],
 [ 3.01818120e-03],
 [-1.76563600e-01],
 [-2.66354491e-01],
 [-4.45936273e-01],
 [-5.35727163e-01],
 [ 3.01818120e-03],
 [-6.25518054e-01],
 [-3.56145382e-01],
 [-8.67727096e-02],
 [ 3.01818120e-03],
```



[ 1.82599963e-01],  
[-8.05099836e-01],  
[ 4.51972635e-01],  
[-4.45936273e-01],  
[-5.35727163e-01],  
[-5.35727163e-01],  
[-3.56145382e-01],  
[-2.66354491e-01],  
[-6.25518054e-01],  
[ 2.72390854e-01],  
[-1.76563600e-01],  
[ 3.01818120e-03],  
[-6.25518054e-01],  
[-1.76563600e-01],  
[-1.76563600e-01],  
[-2.66354491e-01],  
[-8.67727096e-02],  
[-4.45936273e-01],  
[ 9.28090720e-02],  
[-1.76563600e-01],  
[-6.25518054e-01],  
[ 3.62181744e-01],  
[ 2.72390854e-01],  
[ 9.28090720e-02],  
[ 3.01818120e-03],  
[-8.05099836e-01],  
[ 1.82599963e-01],  
[ 1.82599963e-01],  
[-8.67727096e-02],  
[-1.76563600e-01],  
[-6.25518054e-01],  
[ 2.72390854e-01],  
[ 9.28090720e-02],  
[ 9.28090720e-02],  
[-8.67727096e-02],  
[-1.76563600e-01],  
[-3.56145382e-01],  
[-6.25518054e-01],  
[-3.56145382e-01],  
[ 6.01900786e+00],  
[-8.67727096e-02],  
[ 1.82599963e-01],  
[ 9.28090720e-02],  
[-4.45936273e-01],  
[ 2.72390854e-01],  
[ 9.28090720e-02],  
[ 9.28090720e-02],  
[-1.76563600e-01],  
[ 3.62181744e-01],  
[ 1.82599963e-01],  
[-5.35727163e-01],  
[-6.25518054e-01],  
[-6.25518054e-01],  
[ 1.82599963e-01],  
[-1.76563600e-01],  
[-1.76563600e-01],  
[-1.76563600e-01],  
[ 3.01818120e-03]])

```
[2 2 2 2 0 0 0 0 4 4 4 2 4 2 0 0 0 2 2 0 4 2 2 0 4 1 0 4 4 0 2 4 0 0 4 4 4
0 3 4 0 2 2 0 2 4 0 2 4 0 2 4 4 4 2 2 4 2 2 4 4 0 2 0 2 2 2 2 4 2 0 4 4 2
4 4 4 4 2 0 4 2 0 0 0 4 2 0 0 4 4 2 0 0 0 4 4 2 2 2 1 4 0 0 2 0 0 0 4 0 0
2 2 2 0 4 4 4 4]
```

Shape of clustering dataframe (119, 2)

C:\Users\joseg\AppData\Roaming\Python\Python37\site-packages\ipykernel\_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

"""Entry point for launching an IPython kernel.

Out[46]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCOR
0	Bronx	Central Bronx	10453.0	40.852820	-73.912310	GALAGALA NY RESTAURANT	E 175TH ST	African	3
37	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	MB RESTAURANT	WEBSTER AVE	African	5
83	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CAKOR RESTAURANT	EAST 186 STREET	Eastern European	5
84	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CKA KA QELLU	HUGHES AVE	Eastern European	3
120	Bronx	High Bridge and Morrisania	10451.0	40.820454	-73.925066	BALIMAYA RESTAURANT	3 AVENUE	African	10

The new dataframe including a column with the cluster labels.

Now identify the scores by cluster

Out[42]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCOR
0	Bronx	Central Bronx	10453.0	40.852820	-73.912310	GALAGALA NY RESTAURANT	E 175TH ST	African	3
37	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	MB RESTAURANT	WEBSTER AVE	African	5
83	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CAKOR RESTAURANT	EAST 186 STREET	Eastern European	5
84	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CKA KA QELLU	HUGHES AVE	Eastern European	3
120	Bronx	High Bridge and Morrisania	10451.0	40.820454	-73.925066	BALIMAYA RESTAURANT	3 AVENUE	African	10

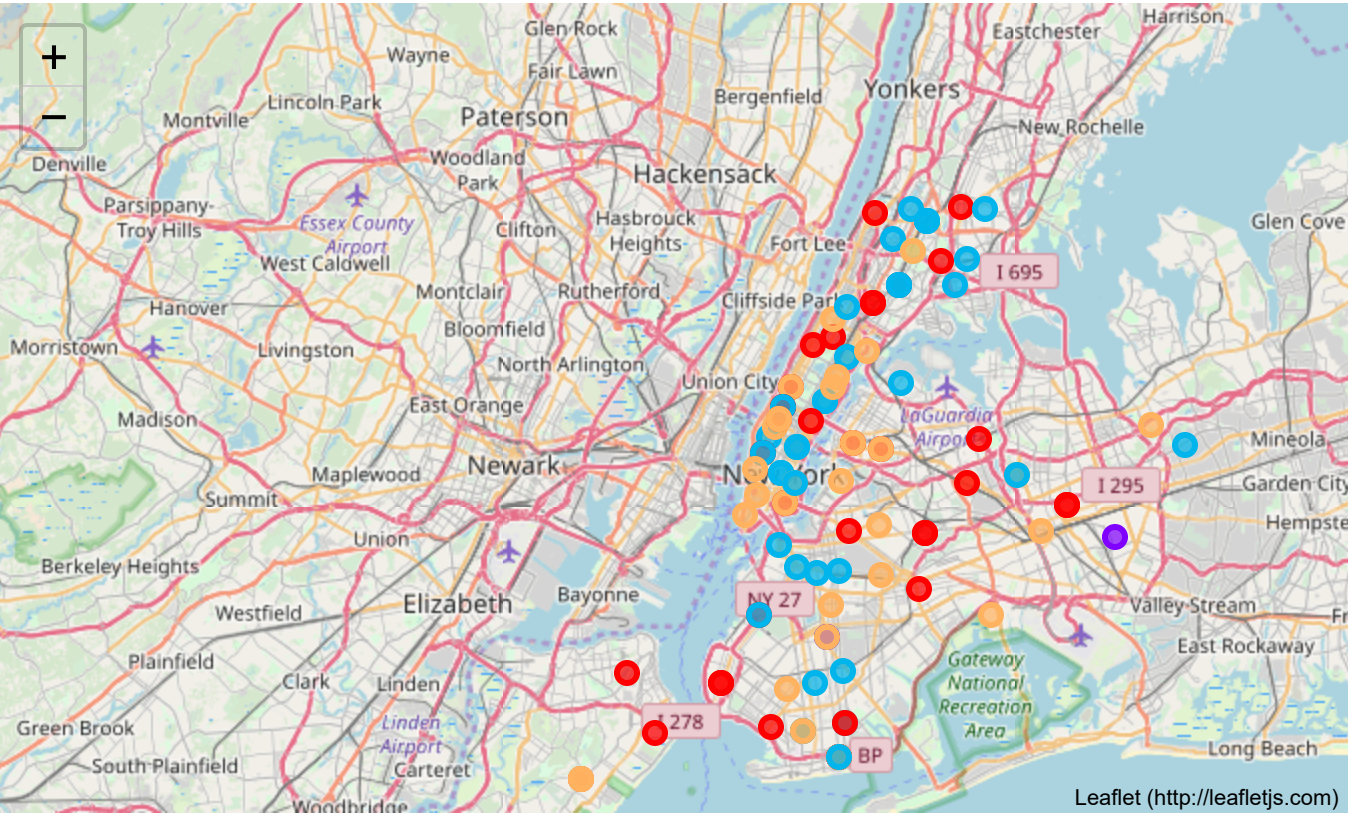
Average score by cluster

Out[47]:

	ZIPCODE	LATITUDE	LONGITUDE	SCORE
Labels				
0	10754.631579	40.708543	-73.949242	11.052632
1	11319.000000	40.672272	-73.857820	85.000000
2	10616.131579	40.745789	-73.934117	3.394737
3	10457.000000	40.847162	-73.898663	47.000000
4	10703.100000	40.692616	-73.956178	7.525000

## Map of Clusters

Out[48]:



## Cluster 0

40 Restaurants

Average Score 7.5

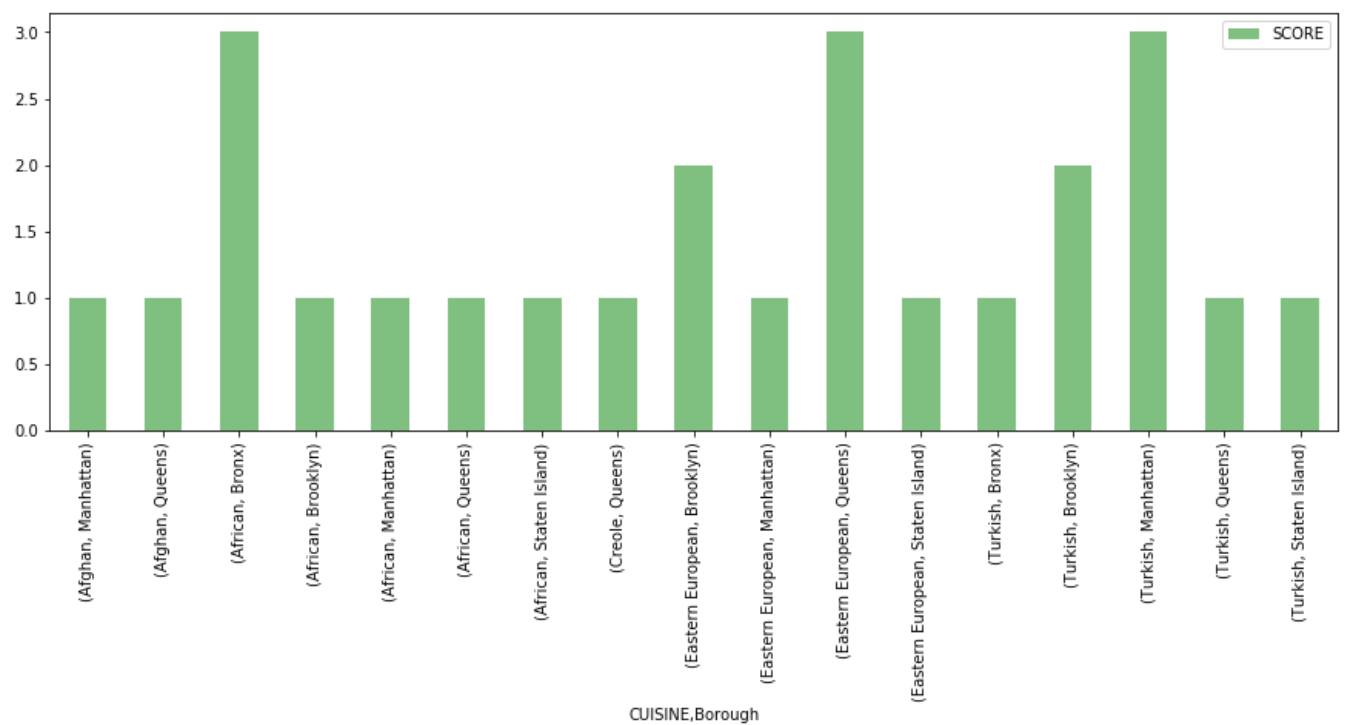
Shape Cluster 0 (38, 11)

Out[49]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCORE
120	Bronx	High Bridge and Morrisania	10451.0	40.820454	-73.925066	BALIMAYA RESTAURANT	3 AVENUE	African	10.0
203	Bronx	High Bridge and Morrisania	10451.0	40.820454	-73.925066	LINCOLN GYRO	E 149TH ST	Turkish	10.0
473	Brooklyn	Southwest Brooklyn	11209.0	40.621993	-74.030134	GEORGIAN DREAM	3RD AVE	Eastern European	13.0
528	Brooklyn	Southwest Brooklyn	11209.0	40.621993	-74.030134	ANTEPLI BAKLAVA	5TH AVE	Turkish	10.0
945	Brooklyn	East New York and New Lots	11207.0	40.670757	-73.894209	FESTAC GRILL AND LOUNGE	HENDRIX ST	African	12.0

Amount of Restaurants by type of cuisine in cluster 0

Out[50]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eedfc43080>



Cluster 1

Average Score 85

2 Restaurants

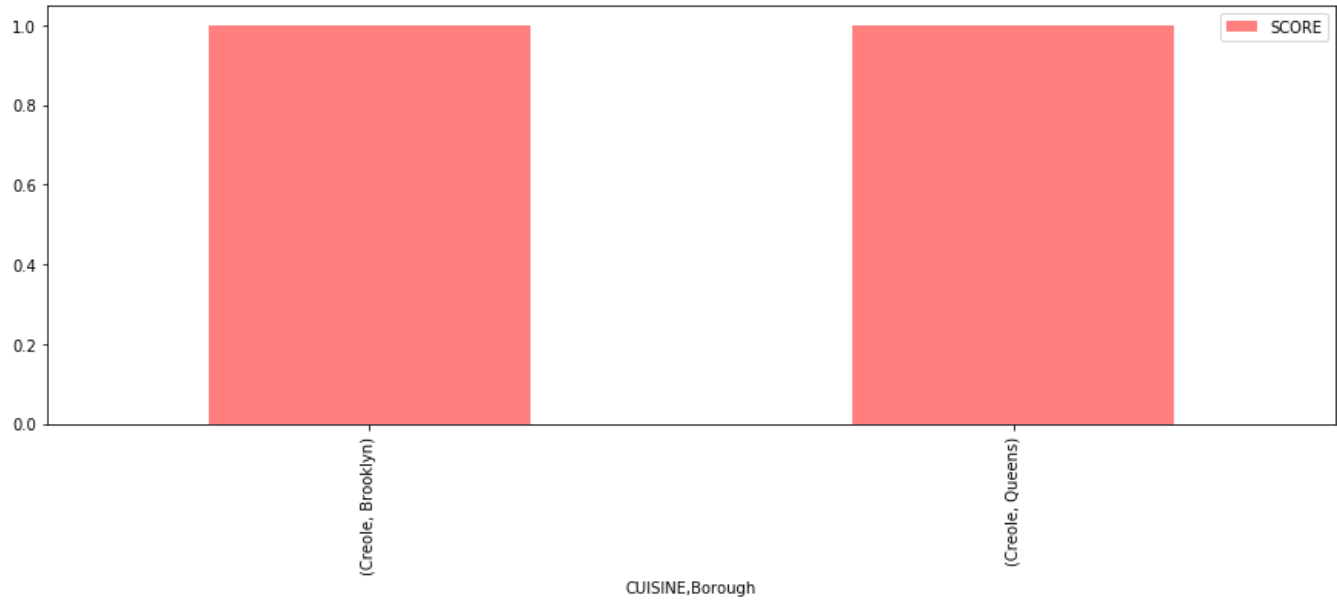
Shape Cluster 1 (2, 11)

Out[51]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCO
2683	Queens	Jamaica	11412.0	40.698096	-73.758990	MAGGY'S PATE KREYOL BAKERY	MURDOCK AVE	Creole	9
7971	Brooklyn	Flatbush	11226.0	40.646448	-73.956649	THE LOFT BKNY RESTAURANT & LOUNGE	NOSTRAND AVE	Creole	7

Amount of Restaurants by type of cuisine in cluster 1

Out[52]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eee045f710>



Cluster 2

Average Score 47

1 Restaurant

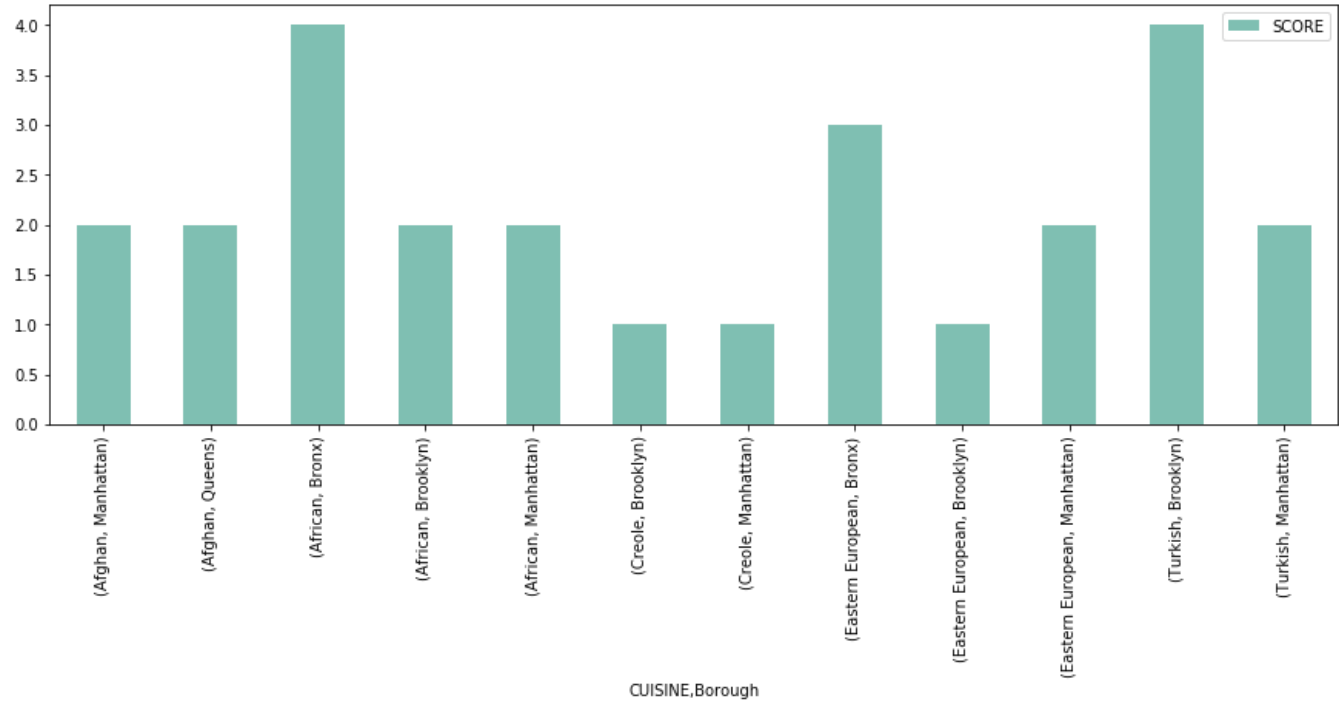
Shape Cluster 2 (38, 11)

Out[53]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCOR
0	Bronx	Central Bronx	10453.0	40.852820	-73.912310	GALAGALA NY RESTAURANT	E 175TH ST	African	3
37	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	MB RESTAURANT	WEBSTER AVE	African	5
83	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CAKOR RESTAURANT	EAST 186 STREET	Eastern European	5
84	Bronx	Bronx Park and Fordham	10458.0	40.862529	-73.888159	CKA KA QELLU	HUGHES AVE	Eastern European	3
692	Brooklyn	Southern Brooklyn	11223.0	40.597143	-73.973426	NY SWEET SPOT CAFE	CONEY ISLAND AVE	Eastern European	2

Amount of Restaurants by type of cuisine in cluster 2

Out[57]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eee0c11f28>



Cluster 3

Average Score 11

38 Restaurants

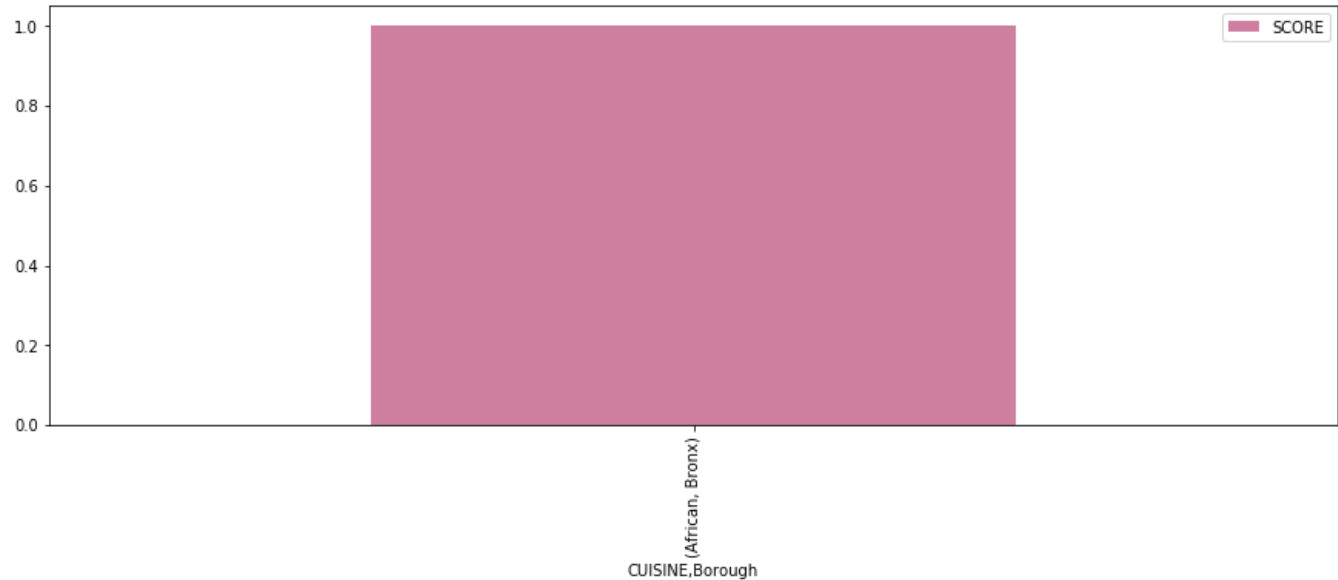
Shape Cluster 3 (1, 11)

Out[58]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCORE
3259	Bronx	Central Bronx	10457.0	40.847162	-73.898663	OSEI-KROM	WEBSTER AVE	African	47

Amount of Restaurants by type of cuisine in cluster 3

Out[59]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eee0839a58>



Cluster 4

Average Score 3.39

38 Restaurants

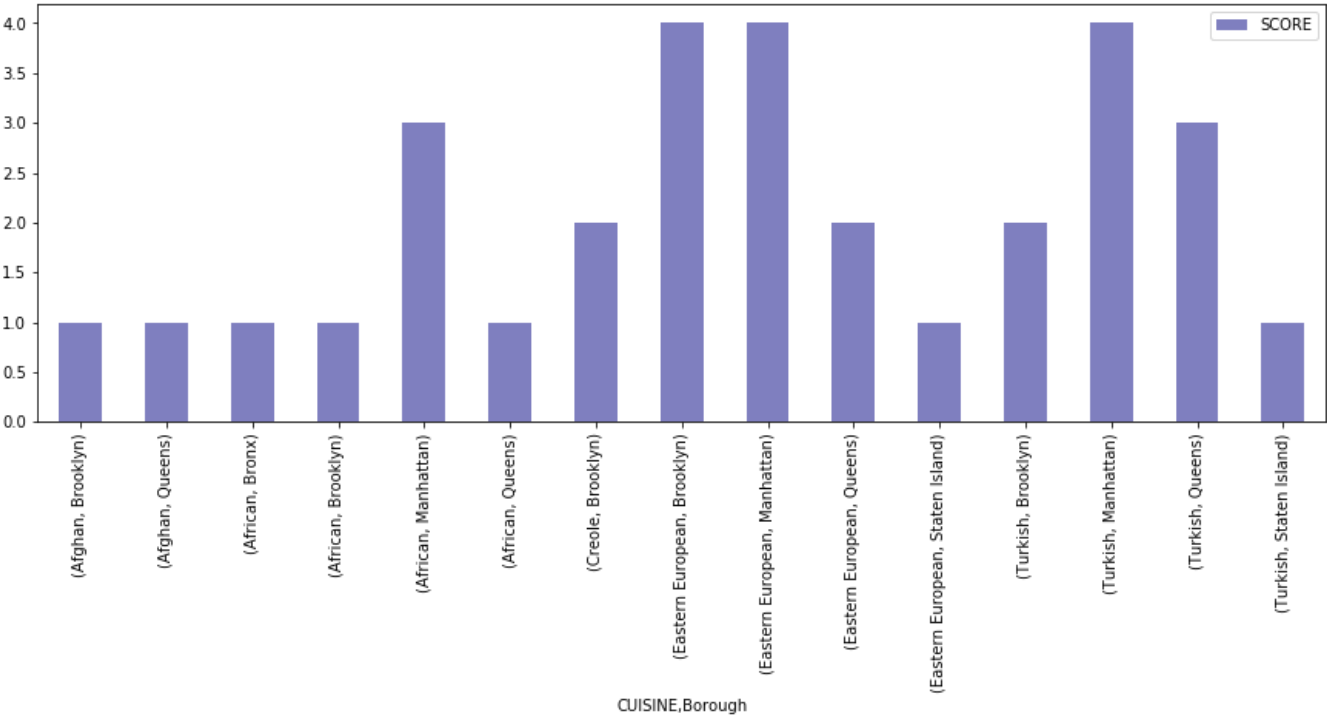
Shape Cluster 4 (40, 11)

Out[61]:

	Borough	Neighborhood	ZIPCODE	LATITUDE	LONGITUDE	RESTAURANT	STREET	CUISINE	SCORE
569	Brooklyn	Borough Park	11204.0	40.618777	-73.984831	Q S PALACE	65TH ST	Eastern European	47
690	Brooklyn	Southern Brooklyn	11223.0	40.597143	-73.973426	CAFE ARBAT	AVENUE U	Eastern European	47
691	Brooklyn	Southern Brooklyn	11223.0	40.597143	-73.973426	MSHOBLIURI/WE ARE GEORGIANS	KINGS HWY	Eastern European	47
741	Brooklyn	Southern Brooklyn	11223.0	40.597143	-73.973426	SAHARA RESTAURANT	CONY ISLAND AVENUE	Turkish	47
2120	Manhattan	Lower East Side	10002.0	40.715776	-73.986211	CAFE KATJA	ORCHARD STREET	Eastern European	47

Amount of Restaurants by type of cuisine in cluster 4

Out[62]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1eee089a908>



Boxplots Charts by Cluster

Out[63]: Text(0.5, 1.0, 'Cluster 4')

