Capstone Project-Battle of Neighbourhoods

By-T.Rukmini Devi

Introduction

- **1.1 Description of the Problem**

New York is known as a state of diverse population. That explains that there are a lot of restaurants with food from different parts of the world. The problem starts with finding a suitable neighborhood place to open a Latin food restaurant.

Description of data

▶ New York Data

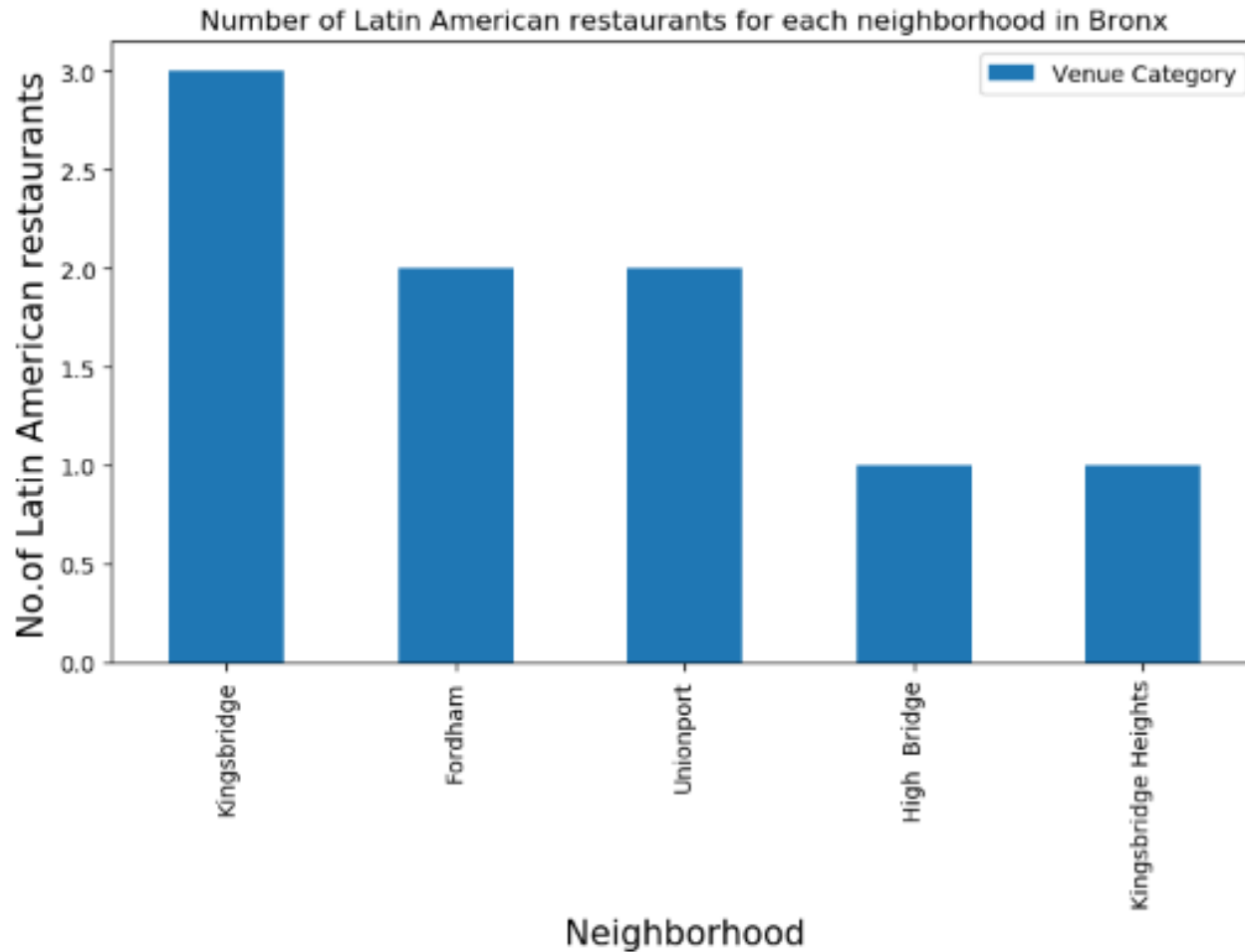The database used is a public one that shows each New York's Borough and neighborhoods with location data.

	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

▶ Foursquare data

With New York data prepared, it is posible to get venues data from each neighborhood with Foursquare API. It is necesary to get a Client ID and password, also it is necesary to define the limit and radius.

Methodology

▶ With both data obtained. First, let's find out which neighborhood has the most latin american restaurants at Bronx district.

Number of Latin American restaurants for each neighborhood in Bronx

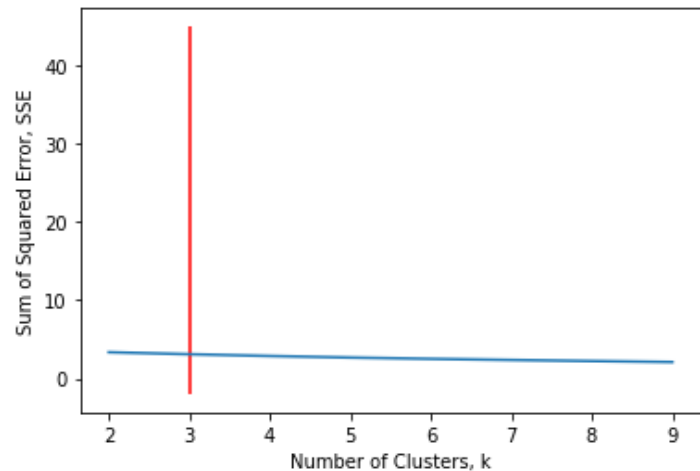▶ Find out the most common venues on each neighborhoods at Bronx district.

	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	Deli / Bodega	Chinese Restaurant	Supermarket	Fast Food Restaurant	Martial Arts Dojo	Breakfast Spot	Spanish Restaurant	Food	Smoke Shop
1	Baychester	Electronics Store	Donut Shop	Pizza Place	Bank	Men's Store	Fast Food Restaurant	Mattress Store	Spanish Restaurant	Sandwich Place	Fried Chicken Joint
2	Bedford Park	Diner	Mexican Restaurant	Pizza Place	Supermarket	Chinese Restaurant	Sandwich Place	Bus Station	Fried Chicken Joint	Deli / Bodega	Train Station
3	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Grocery Store	Dessert Shop	Donut Shop	Spanish Restaurant	Bank	Sandwich Place
4	Bronxdale	Italian Restaurant	Spanish Restaurant	Bank	Pizza Place	Performing Arts Venue	Paper / Office Supplies Store	Chinese Restaurant	Eastern European Restaurant	Mexican Restaurant	Breakfast Spot

Data Clustering

▶ K – Mean method

Finding out the best 'k' value to apply this method.

▶ Elbow Method

The best value of k is 3

Silhouette Coefficient

```
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster2 in range(2, 10):
    kmeans2 = KMeans(n_clusters = n_cluster2, random_state = 0).fit(bronx_grouped_clustering)
    label2 = kmeans2.labels_
    sil_coeff = silhouette_score(bronx_grouped_clustering, label2, metric = 'euclidean')
    print("Where n_clusters = {}, the Silhouette Coefficient is {}".format(n_cluster2, sil_coeff))
```
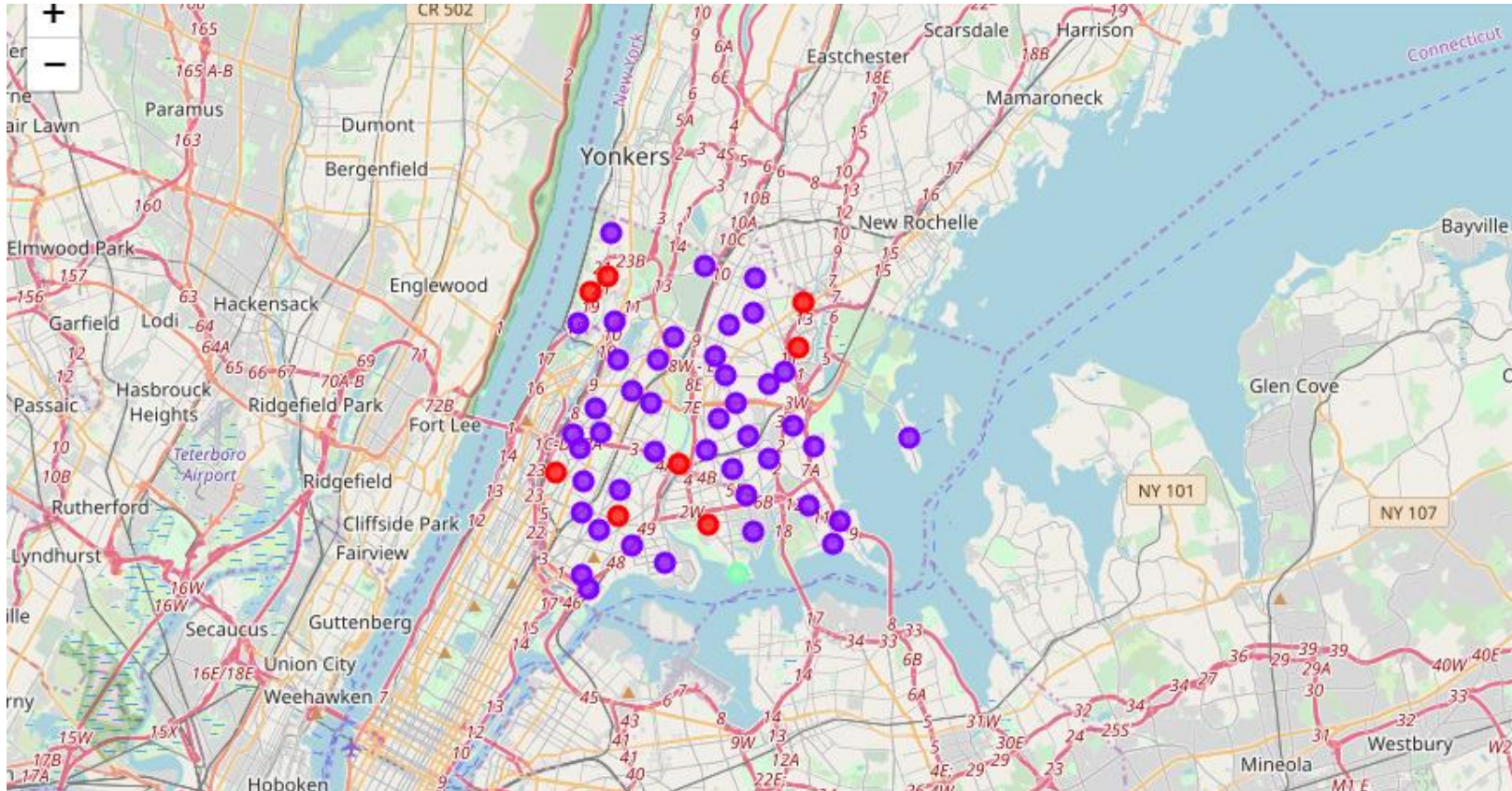
Where n_clusters = 2, the Silhouette Coefficient is 0.6787637908476635
Where n_clusters = 3, the Silhouette Coefficient is 0.6406348689213593
Where n_clusters = 4, the Silhouette Coefficient is 0.6170280718679633
Where n_clusters = 5, the Silhouette Coefficient is 0.6239284792976618
Where n_clusters = 6, the Silhouette Coefficient is 0.650469587277814
Where n_clusters = 7, the Silhouette Coefficient is 0.6537983168861516
Where n_clusters = 8, the Silhouette Coefficient is 0.6647395678798047
Where n_clusters = 9, the Silhouette Coefficient is 0.6550315362017799

The value of k in which the coefficient is higher is the chosen one. In this case k = 3

The method is applied with k = 3

Data Clustering

▶ The result of k-mean method aplication in this case, gives 3 clusters. Those are shown on the next map.

Results

- In resume, according to results shown before:

  - At Bronx district Pizza Place is the most common venue.

  - The neighborhood with the most Latin food restaurants is Kingsbridge. This neighborhood also seems to be a restaurants district and an appropriate place to open a Latin food restaurant.

  - Cluster 2 shows that on those neighborhoods there is predominance of restaurants and food places.

Discussion and Conclusion

- It is very important to note that Clusters 2 is the most viable clusters to create a Latin American Restaurant. Their proximity to other amenities and accessibility to station are paramount.

- In conclusion, this project would have had better results if there were more data in terms of traffic access and the quantity of latin people living on those neighborhoods.