

Battle of the Neighborhoods

Introduction

Chicago is the third largest city in the United States and is home to 2.7 million residents. Nearly 100 neighborhoods make up the city. The city has more than 7,300 restaurants, has over 580 parks with 8300 acres of green spaces, not to mention a beautiful coastline along Lake Michigan, Art attractions, and many more. Chicago is also home to a few fortune 500 companies such as Boeing and United, but the backbone of the economy is small business. When a business considers expansion, the success is dependent on many factors, one of them being location. A business might want to expand into an area similar to where they are currently located.

This analysis will look at the neighborhoods in Chicago and cluster them to help businesses select neighborhoods to expand to. A decision like can be difficult, but hopefully this analysis can help in their decision.

Data

- The venue data from the Chicago neighborhoods. This would include the type of venues a neighborhood has.
 - **Data Observations:**
 - 292 unique venue types
 - 2690 venues near the center of the neighborhood
 - On average nearly 28 venues near the center of the neighborhood
 - **Source:**
Foursquare API
- Chicago neighborhood locations were taken from the following location. This data will be preprocessed and uploaded to IBM Watson for the analysis.
 - **Data Observations:**
 - 98 neighborhoods in Chicago
 - **Source:**
<https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9>
- This venue types of the neighborhood will then be used to cluster the neighborhoods.

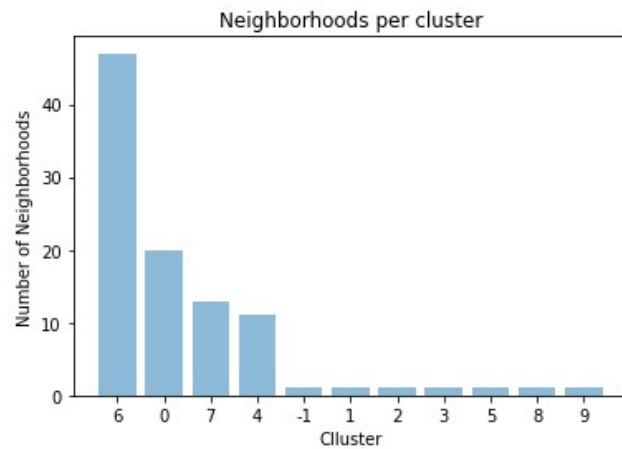
Methodology

- Determine the neighborhoods in Chicago by downloading from the data.cityofchicago.org website.
- The neighborhood data will be cleaned to ensure that there are no duplicate neighborhoods
- Foursquare API will be used to gather the venue data within a 500m radius from the center of the neighborhood
- For each neighborhood, the number of 'Venue Category' instances will be counted.
- The mean of the frequency of occurrence will be calculated for each 'Venue Category' per neighborhood.
- The top 5 venues will be taken per neighborhood.
- The neighborhoods will then be clustered using the K-means clustering Machine learning technique.
- Finally, the data will be visualized using the Folium Package.

Detailed notebook provided in the following link.

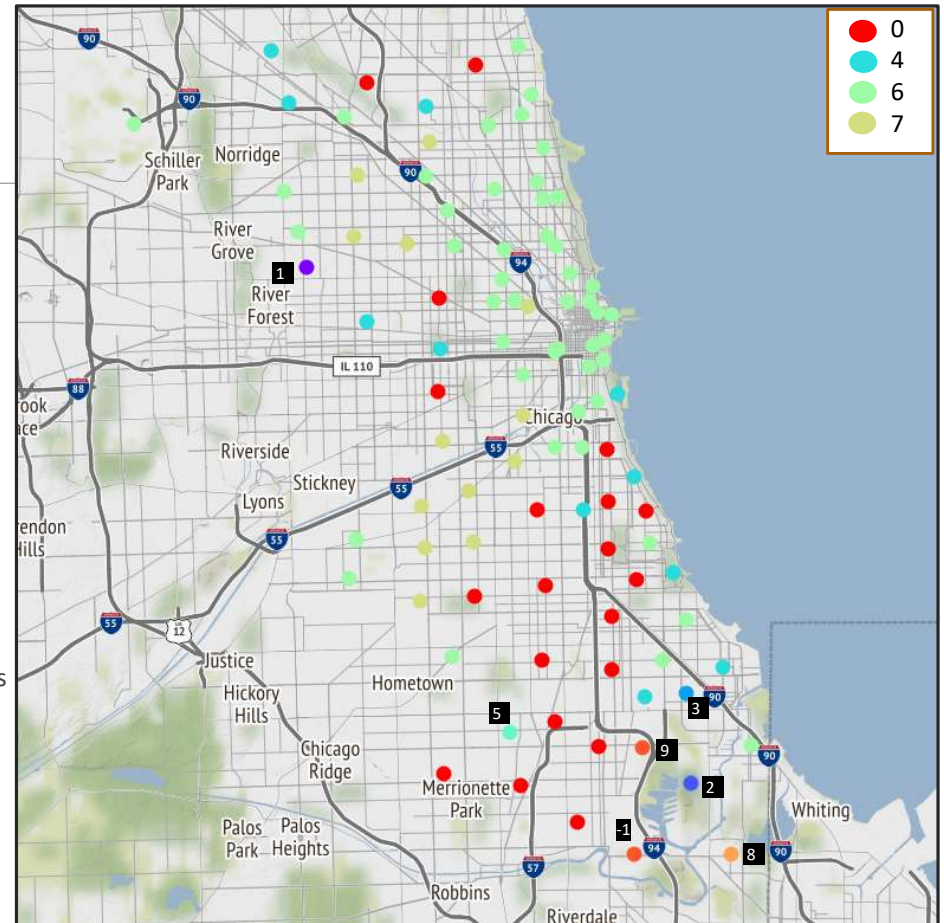
<https://github.com/Ch1nhD/TheCapstoneProject>

Results



- Clustering the Chicago neighborhoods into 10 groups, resulted in 4 neighborhoods that stood out to be similar.
- Cluster 0, 6, and 7 are relatively grouped together based on area of the city, but 4 has a few more pockets around the city.
- Let's dive in further to see how each group is similar.

Note: Cluster -1 did not have Foursquare data



Results Continued 1

Cluster Labels	Venue	Venue Count
6	Pizza Place	23
6	Coffee Shop	17
6	Bar	16
6	Sandwich Place	8
6	Hotel	7
6	Café	7
6	Italian Restaurant	7
6	Mexican Restaurant	6
6	Fast Food Restaurant	6
6	American Restaurant	6

Cluster Labels	Venue	Venue Count
0	Park	6
0	Fast Food Restaurant	5
0	Grocery Store	5
0	Bus Station	4
0	American Restaurant	4
0	Liquor Store	4
0	Discount Store	3
0	Donut Shop	3
0	BBQ Joint	3
0	Currency Exchange	3

Cluster Labels	Venue	Venue Count
7	Mexican Restaurant	13
7	Grocery Store	3
7	Pizza Place	3
7	Chinese Restaurant	3
7	Ice Cream Shop	3
7	Bakery	2
7	Coffee Shop	2
7	Korean Restaurant	2
7	Taco Place	2
7	Supermarket	2

Cluster Labels	Venue	Venue Count
4	Park	11
4	Food	4
4	Train Station	3
4	Burger Joint	2
4	Empanada Restaurant	2
4	Electronics Store	2
4	Eye Doctor	2
4	Grocery Store	2
4	Harbor / Marina	2
4	Women's Store	2

- Taking a closer look at how these areas are similar
 - Cluster 6 has many Pizza places, Coffee shops, Bars, Sandwich shops, etc. Already a heavily invested area and most likely heavy competition. However, if your business is already in this cluster, then these areas should be considered.
 - Cluster 0 and 4 have Parks, Fast Food/Food, and Grocery stores. A Pizza place or Sandwich place could be consider expanding to these areas since there are few in these areas.
 - Cluster 7 is primarily Mexican food, which are areas that serve the Hispanic community. Potential growth area since majority is 1 type of venue.

Results Continued 2

- Why didn't we see any data for Riverdale?
- The center of the neighborhood is within a Water Reclamation Plant and could be why we didn't see any venues here.



Discussion

- The result of the k-Means cluster produced 4 distinct clusters.
- Main cluster locations
 - Cluster 0 was primarily grouped in the South side
 - Cluster 6 was the Loop/North side
 - Cluster 7 was along I55
 - Cluster 4 was sprinkled around the city
- As discussed in the Results section, there are certain businesses that could benefit from expanding to these areas. By looking at the types of businesses a cluster has, a business can see if their business is lacking or overly saturated in an area and decide if expansion would be viable to these areas.
- A limitation of this result is that it only looks at venue types in the area and not other variables such as demographic, available hours of operation, foot traffic, etc.

Conclusion

- This workbook was meant to be a tool to help businesses in their decision process in determining which neighborhoods to expand their business to. The business could expand to a heavily competitive area, similar area, or be in an area where the business could be the first and have little competition.
- In this analysis, we were able to find similarities of each neighborhood in the city and cluster them together using k-means. From the 98 neighborhoods, the analysis was able to produce 4 main clusters.
- This analysis only considered one aspect of many that a business should go through when determining areas to expand their business. Hopefully since the analysis narrowed the city down to 4 clusters, this is one step closer to making a final decision.

Happy mining. Thank you.

A solid orange horizontal bar spanning the width of the slide, located at the bottom.