THE BATTLE OF NEIGHBORHOOD CHICAGO

Capstone Project - The Battle of Neighborhoods

Mohammad AlYousif

February 2021

TABLE OF CONTENTS

The Battle of Neighborhood – Chicago

Introduction

Chicago is the third largest city in population in the United States after New York City and Los Angeles with an estimated population of 2.7 million in 2019. Chicago is also one of the US's most densely populated major cities. It is tourist attraction city and a transportation hub. Moreover, Chicago has a large and ethnically diverse communities.

The city of Chicago is divided into 77 communities as shown in figure 1 for statistical and planning purposes. Each community has a wide variety of ethnically groups. Moreover, each community contains a lot of diversified flavors of shops and restaurants.

Figure 1 Chicago 77 Communities

Problem

If an investor wants to open an American Casual Dining restaurant in one of Chicago communities, which is the best location for a such kind of restaurant?

Restaurant location should be in a place with high traffic of people and low competition from similar kind of restaurants. Also, car parking should be available nearby the restaurant.

Chicago is a big city and choosing the restaurant location will take a lot of time and effort. However, dividing the city into communities and clusters will give investors a better picture on where to open the restaurant. Moreover, after analyzing city clusters, investors can have a good idea what type of investment is suitable in each city cluster.

Data

The list of communities is taken from the following Wikipedia page.

The communities' coordinates are taken from MapQuest https://www.mapquest.com/ through geocoder of GPS visualizer website https://www.gpsvisualizer.com/geocoder/.

The data were stored in a csv file. Pandas data frame were created to store the data.

Then, FourSquare API utilized via the Request library to get the venues details of each community.

Methodology

Chicago Communities List

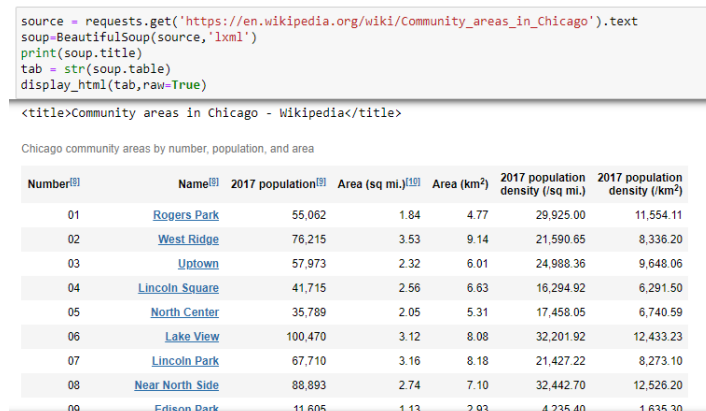Chicago communities list is taken from Wikipedia page. Beautiful soup is used to scrape the web page.

```
source = requests.get('https://en.wikipedia.org/wiki/Community_areas_in_Chicago').text
soup=BeautifulSoup(source,'lxml')
print(soup.title)
tab = str(soup.table)
display_html(tab,raw=True)
```

<title>Community areas in Chicago - Wikipedia</title>

Chicago community areas by number, population, and area

Number[8]	Name[8]	2017 population[9]	Area (sq mi.)[10]	Area (km²)	2017 population density (/sq mi.)	2017 population density (/km²)
01	Rogers Park	55,062	1.84	4.77	29,925.00	11,554.11
02	West Ridge	76,215	3.53	9.14	21,590.65	8,336.20
03	Uptown	57,973	2.32	6.01	24,988.36	9,648.06
04	Lincoln Square	41,715	2.56	6.63	16,294.92	6,291.50
05	North Center	35,789	2.05	5.31	17,458.05	6,740.59
06	Lake View	100,470	3.12	8.08	32,201.92	12,433.23
07	Lincoln Park	67,710	3.16	8.18	21,427.22	8,273.10
08	Near North Side	88,893	2.74	7.10	32,442.70	12,526.20
09	Edison Park	11,605	1.13	2.93	4,235.40	1,635.30

Figure 2 Scraping Wikipedia Page

Then, HTML data is converted to Pandas data frame.

```
In [3]: dfhtml = pd.read_html(tab)
        df=dfhtml[0]
        df.head()
```

Out[3]:

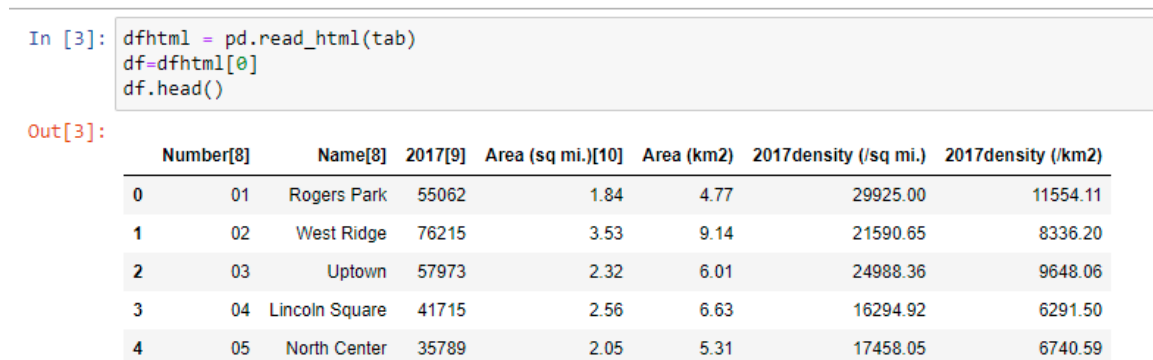|	Number[8]	Name[8]	2017[9]	Area (sq mi.)[10]	Area (km2)	2017density (/sq mi.)	2017density (/km2)
0	01	Rogers Park	55062	1.84	4.77	29925.00	11554.11
1	02	West Ridge	76215	3.53	9.14	21590.65	8336.20
2	03	Uptown	57973	2.32	6.01	24988.36	9648.06
3	04	Lincoln Square	41715	2.56	6.63	16294.92	6291.50
4	05	North Center	35789	2.05	5.31	17458.05	6740.59

Figure 3 Storing Webpage Contents in Data Frame

After that, I did some data cleaning to rename some columns and drop unneeded columns.

Chicago Communities Coordinates

The communities' coordinates are taken from MapQuest https://www.mapquest.com/ through geocoder of GPS visualizer website https://www.gpsvisualizer.com/geocoder/.

I created a MapQuest developer account. Then, I created a MapQuest AppKey. I used the key to get the communities coordinates through GPS visualizer website as shown in the below image.

Figure 4 GPS Visualizer Website

The data extracted from GPS visualizer website were stored in a csv file. Pandas data frame was created to store the data.

Figure 5 Importing CSV File of Coordinates

After that, I merged the communities table and the coordinates table.

Merging Tables

```
In [ ]: df4 = pd.merge(df3,cor, on='Communities')
        df4
```

```
In [9]: df4.drop(['desc', 'color', 'source', 'precision'], axis=1, inplace=True)
        df4
```

Out[9]:

	Communities	Population	latitude	longitude
0	Rogers Park	55062	42.010531	-87.670748
1	West Ridge	76215	42.001567	-87.695137
2	Uptown	57973	41.966063	-87.656105
3	Lincoln Square	41715	41.980215	-87.688119
4	North Center	35789	41.953915	-87.684600
5	Lake View	100470	41.944365	-87.655995
6	Lincoln Park	67710	41.922846	-87.646653
7	Near North Side	88893	41.899620	-87.633350
8	Edison Park	11605	42.007160	-87.814155
9	Norwood Park	37089	41.984944	-87.800635
10	Jefferson Park	26808	41.978694	-87.771047

Figure 6 Merging Communities and Coordinates Tables

Chicago Map

Chicago map was created to show the location of the 77 communities plus the population of each community as shown in the below image.

Create a Map of Chicago

```
In [13]: chicago_map = folium.Map(location=[latitude, longitude],zoom_start=10)

         for lat,lng,population,community in zip(chicago_data['latitude'],chicago_data['longitude'],chicago_data['Population'], chicago_da
             #label = '{}'.format(label)
             label = '{}, population: {}'.format(community, population)
             label = folium.Popup(label, parse_html=True)
             folium.CircleMarker(
                 [lat,lng],
                 radius=5,
                 popup=label,
                 color='blue',
                 fill=True,
                 fill_color='#3186cc',
                 fill_opacity=0.7,
                 parse_html=False).add_to(chicago_map)
         chicago_map
```
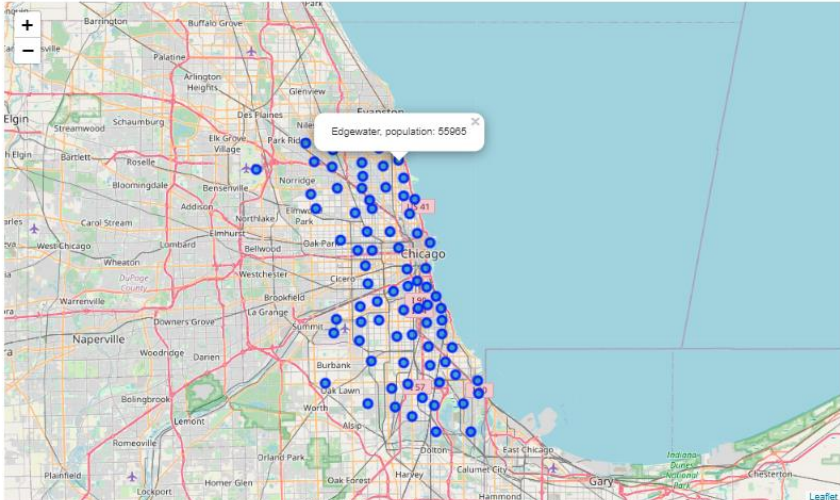
Out[13]:

Figure 7 Chicago Map

5

FourSquare API utilized via the Request library to get the top 200 venues details of each community within a radius of 700 meters.

The top 200 venues of the first community within a radius of 700 meters.

```
LIMIT = 200 # limit of number of venues returned by Foursquare API

radius = 700 # define radius

url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    community_latitude,
    community_longitude,
    radius,
    LIMIT)
url # display URL
```

```
'https://api.foursquare.com/v2/venues/explore?&client_id=AO4YA1HEKGUVSREICL11PBRPPJN2M5DSTDE5NTQBVXNWT1VC&client_secret=WRAUE2G
RMETQ5NC0IS3QZNKTQU5GJPMWDSWAIPP1JEIWMKGM&v=20180605&ll=42.010531,-87.670748&radius=700&limit=200'
```

Figure 8 Creating FourSquare Request URL

Venues by community were grouped by taking the mean of the frequency of occurrence of each category of venues retrieved from FourSquare.

Grouping rows by community and by taking the mean of the frequency of occurrence of each category

In [28]: chicago_grouped = chicago_onehot.groupby('Communities').mean().reset_index()
chicago_grouped

Out[28]:

	Communities	ATM	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Albany Park	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000
1	Archer Heights	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000
2	Armour Square	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.038462
3	Ashburn	0.000000	0.00000	0.000000	0.000000	0.000000	0.076923	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000
4	Auburn Gresham	0.000000	0.00000	0.000000	0.000000	0.000000	0.076923	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000
5	Austin	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000
6	Avalon Park	0.076923	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000

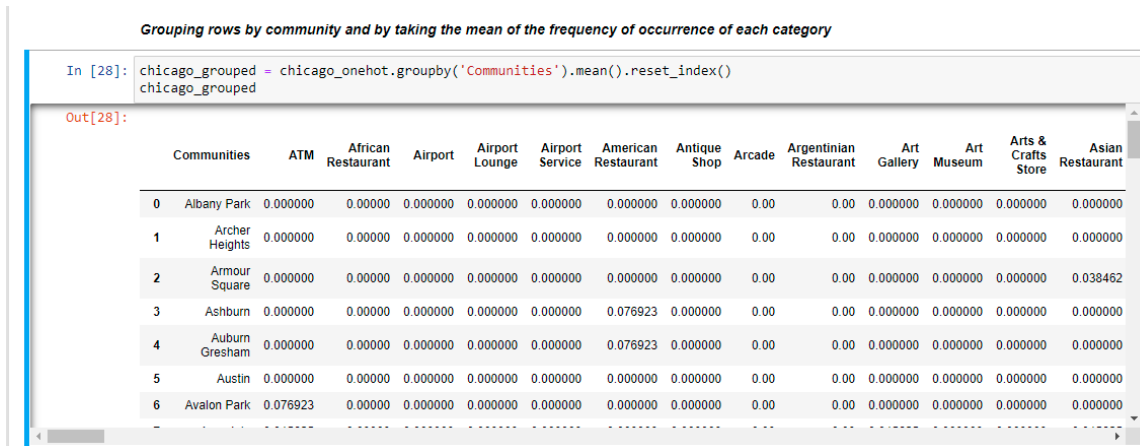*Figure 9 Grouping Venues by Community*

Pandas Data frame created to display the most common venues of each community as shown in the below image.

Creating a dataframe and display the top 10 venues for each community.

```
In [32]: num_top_venues = 10

         indicators = ['st', 'nd', 'rd']

         # create columns according to number of top venues
         columns = ['Communities']
         for ind in np.arange(num_top_venues):
             try:
                 columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
             except:
                 columns.append('{}th Most Common Venue'.format(ind+1))

         # create a new dataframe
         Communities_venues_sorted = pd.DataFrame(columns=columns)
         Communities_venues_sorted['Communities'] = chicago_grouped['Communities']

         for ind in np.arange(chicago_grouped.shape[0]):
             Communities_venues_sorted.iloc[ind, 1:] = return_most_common_venues(chicago_grouped.iloc[ind, :], num_top_venues)

         Communities_venues_sorted
```

Out[32]:

	Communities	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	Albany Park	Mexican Restaurant	Bakery	Grocery Store	Fried Chicken Joint	Park	Hookah Bar	Donut Shop	Pizza Place	Fast Food Restaurant	Mobile Phone Shop
1	Archer Heights	Mobile Phone Shop	Grocery Store	Mexican Restaurant	Bakery	Gas Station	Sandwich Place	Nightclub	Food	Candy Store	Big Box Store
2	Armour Square	Chinese Restaurant	Cosmetics Shop	Bakery	Pizza Place	Mexican Restaurant	Grocery Store	Tailor Shop	Park	Mobile Phone Shop	Seafood Restaurant
3	Ashburn	Liquor Store	Fast Food Restaurant	Train Station	Wings Joint	American Restaurant	Cosmetics Shop	Food	Bus Station	Light Rail Station	Automotive Shop
4	Auburn Gresham	Pharmacy	Southern / Soul Food Restaurant	Discount Store	Bank	Fast Food Restaurant	Seafood Restaurant	Basketball Court	Video Store	Park	Cosmetics Shop
5	Austin	Park	Sandwich Place	Train Station	Discount Store	Seafood Restaurant	Liquor Store	Fast Food Restaurant	Donut Shop	Gym	Business Service

Figure 10 Top Venues for Each Community

Cluster Communities

K-means clustering is unsupervised machine learning algorithms which I used to cluster communities into 5 clusters. K-means will create clusters that have similarities in the most common venues.

Cluster Communities into 5 clusters

```
: # set number of clusters
  kclusters = 5

  chicago_grouped_clustering = chicago_grouped.drop('Communities', 1)

  # run k-means clustering
  kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(chicago_grouped_clustering)

  # check cluster labels generated for each row in the dataframe
  kmeans.labels_[0:10]
```
```
: array([0, 0, 0, 3, 0, 3, 0, 0, 0, 0])
```

Figure 11 K-Mean Clustering

Then, Pandas data frame created to include the cluster label and the top 10 venues for each community.

Creating a Dataframe that includes the cluster as well as the top 10 venues for each community.

```
# add clustering labels

Communities_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

chicago_merged = chicago_data

chicago_merged = chicago_merged.join(Communities_venues_sorted.set_index('Communities'), on='Communities')

chicago_merged
```

	Communities	Population	latitude	longitude	Cluster Labels	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	
0	Rogers Park	55062	42.010531	-87.670748	0	Mexican Restaurant	Sandwich Place	Bar	Park	Bakery	American Restaurant	Pizza Place	E
1	West Ridge	76215	42.001567	-87.695137	0	Indian Restaurant	Grocery Store	Pakistani Restaurant	Boutique	Juice Bar	Electronics Store	Fast Food Restaurant	
2	Uptown	57973	41.966063	-87.656105	0	Coffee Shop	Pizza Place	Diner	Chinese Restaurant	Mexican Restaurant	Bar	Sandwich Place	
3	Lincoln Square	41715	41.980215	-87.688119	0	Karaoke Bar	Korean Restaurant	Brewery	Convenience Store	Pharmacy	Bus Station	South American Restaurant	
4	North Center	35789	41.953915	-87.684600	0	Bar	Bus Station	Coffee Shop	Pharmacy	Pub	Sandwich Place	Bank	Mc
5	Lake View	100470	41.944365	-87.655995	0	Sports Bar	Sandwich Place	Coffee Shop	Taco Place	Gym	Baseball Stadium	Mexican Restaurant	

Figure 12 Pandas Data Frame for Clusters and Top Venues

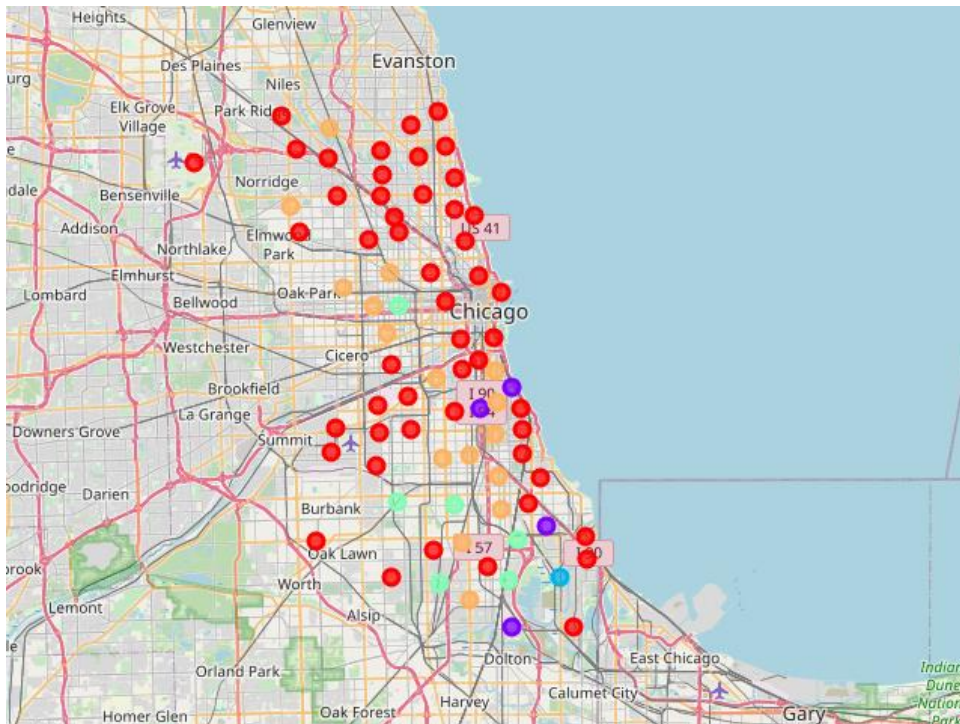Then, Chicago map is created with clusters labels.

Figure 13 Chicago Map with Clusters

8

Results

After clustering the communities into 5 clusters, we got the following details for each cluster:

- Most of the communities fall in one cluster (Cluster 0), which have restaurants, bars and cafes as the most common venues.
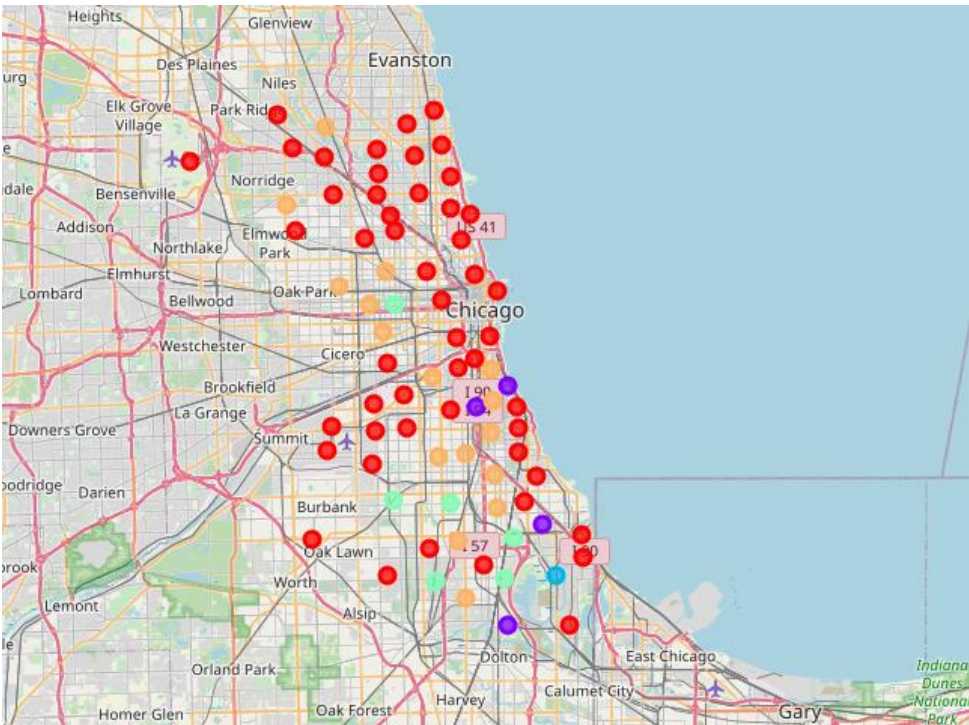

Figure 14 Chicago Communities Clusters

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 0, chicago_merged.columns[[0,1] + list(range(5, chicago_merged.shape[1]))]]
```

	Communities	Population	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 M Comr Ve
0	Rogers Park	55062	Mexican Restaurant	Chinese Restaurant	Sandwich Place	Pizza Place	Bakery	Park	Bus Station	Theater	Bar	Amer Restau
1	West Ridge	76215	Indian Restaurant	Pakistani Restaurant	Grocery Store	Pharmacy	Fast Food Restaurant	Park	Fruit & Vegetable Store	Other Great Outdoors	Sandwich Place	Automc S
2	Uptown	57973	Coffee Shop	Pizza Place	Bar	Diner	Mexican Restaurant	Chinese Restaurant	Sandwich Place	Sushi Restaurant	Supermarket	Playgro
3	Lincoln Square	41715	Brewery	Convenience Store	Karaoke Bar	Bus Station	Market	Pharmacy	Bakery	Bank	Bar	Hot J
4	North Center	35789	Bar	Coffee Shop	Pizza Place	Pharmacy	Video Store	Bank	Dance Studio	Boutique	Thai Restaurant	
5	Lake View	100470	Sandwich Place	Coffee Shop	Sports Bar	Gym	Taco Place	Mexican Restaurant	Baseball Stadium	Gay Bar	Bar	Bal

Figure 15 Cluster 0 Most Common Venues

- Cluster 1 most common venues are parks and beaches.

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 1, chicago_merged.columns[[0,1] + list(range(5, chicago_merged.shape[1]))]
```

	Communities	Population	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
35	Oakland	6645	Park	Beach	Juice Bar	Hotel	Public Art	Poke Place	Noodle House	Other Repair Shop	Other Great Outdoors	Post Office
36	Fuller Park	2439	Park	Baseball Field	Train Station	Bus Station	Food	Fast Food Restaurant	Optical Shop	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors
47	Calumet Heights	13188	Bus Station	Park	Gym / Fitness Center	Food	Office	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors	Organic Grocery	Optical Shop
53	Riverdale	7394	Park	Fast Food Restaurant	Clothing Store	Food	Grocery Store	Organic Grocery	Outdoors & Recreation	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors

Figure 16 Cluster 1 Most Common Venues

- Cluster 2 has only one community with has a variety of common venues.

	Communities	Population	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
50	South Deering	14614	Print Shop	Bar	Mexican Restaurant	Nail Salon	National Park	Nature Preserve	New American Restaurant	Newsstand	Nightclub	Pakistani Restaurant

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 3, chicago_merged.columns[[0,1] + list(range(5, chicago_merged.shape[1]))]
```

Figure 17 Cluster 2 Most Common Venues

- Cluster 3 most common venues are parks and restaurants

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 3, chicago_merged.columns[[0,1] + list(range(5, chicago_merged.shape[1]
```

	Communities	Population	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
26	East Garfield Park	19996	Park	Liquor Store	Train Station	Warehouse Store	Food	Football Stadium	Pharmacy	Supermarket	Bus Line	Bus Station
46	Burnside	2254	Liquor Store	Park	Southern / Soul Food Restaurant	Gym	Train Station	Office	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors	Organic Grocery
49	Pullman	6613	Park	ATM	Train Station	Pharmacy	Sandwich Place	Café	Optical Shop	Food	Food Court	National Park
69	Ashburn	43792	Fast Food Restaurant	American Restaurant	Liquor Store	Train Station	Park	Cosmetics Shop	Light Rail Station	Bus Station	Pizza Place	Other Great Outdoors
70	Auburn Gresham	46278	Pharmacy	Park	Seafood Restaurant	Boutique	American Restaurant	Cosmetics Shop	Discount Store	Southern / Soul Food Restaurant	Video Store	Fast Food Restaurant
74	Morgan Park	22394	Train Station	Park	Currency Exchange	BBQ Joint	Construction &	Discount Store	Track	Home Service	Grocery Store	Pakistani Restaurant

Figure 18 Cluster 3 Most Common Venues

10

- Cluster 4 most common venues are fast food restaurants

```
chicago_merged.loc[chicago_merged['Cluster Labels'] == 4, chicago_merged.columns[[0,1] + list(range(5, chicago_merged.shape[1]))]
```

	Communities	Population	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
11	Forest Glen	19019	Park	Sandwich Place	Currency Exchange	Hobby Shop	Diner	Restaurant	Fast Food Restaurant	Salon / Barbershop	Café	Smoke Shop
16	Dunning	43689	Fast Food Restaurant	Hot Dog Joint	Deli / Bodega	Pizza Place	Donut Shop	Thai Restaurant	Ice Cream Shop	Tattoo Parlor	Bakery	Bank
22	Humboldt Park	56427	Grocery Store	Music Venue	Seafood Restaurant	American Restaurant	Liquor Store	Bus Stop	Art Gallery	BBQ Joint	Video Store	Latin American Restaurant
24	Austin	95260	Grocery Store	Salon / Barbershop	Seafood Restaurant	Sandwich Place	Park	Liquor Store	Gym	Train Station	Donut Shop	Fast Food Restaurant
25	West Garfield Park	17163	Fast Food Restaurant	Shoe Store	Fried Chicken Joint	Clothing Store	Supermarket	Mobile Phone Shop	Food	Taco Place	Sandwich Place	Grocery Store
28	North Lawndale	35947	Seafood Restaurant	Convenience Store	Fast Food Restaurant	Train Station	Optical Shop	Pakistani Restaurant	Outdoors & Recreation	Outdoor Sculpture	Other Repair Shop	Other Great Outdoors
34	Douglas	20781	Park	Fast Food Restaurant	Sandwich Place	Bus Station	Wings Joint	Fried Chicken Joint	Historic Site	Gym / Fitness Center	Pizza Place	Mobile Phone Shop
37	Grand Boulevard	22313	Coffee Shop	Art Gallery	BBQ Joint	Train Station	Restaurant	Boutique	Shoe Store	Bed & Breakfast	Food	Bus Station
39	Washington Park	11502	Fast Food Restaurant	Piercing Parlor	Theater	Park	Breakfast Spot	Lounge	Liquor Store	Train Station	Dog Run	Gas Station
43	Chatham	31120	Fast Food Restaurant	Park	Boutique	Lounge	Donut Shop	Bar	Fried Chicken Joint	Kids Store	Discount Store	Ice Cream Shop

Figure 19 Cluster 4 Most Common Venues

Discussion

After analyzing each cluster and its most common venues, I found that Oakland community in cluster 1 is good place to open an American Casual Dining restaurant. Oakland population is 6645. It has a lot of tourist attraction places such as parks and beaches, so the traffic of people is very high. Also, very few Casual Dining restaurants are located nearby attraction places. Car parking will not be a problem especially in the south of Oakland.

	Communities	Population	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
35	Oakland	6645	Park	Beach	Public Art	Hotel	Trail	Bakery	Juice Bar
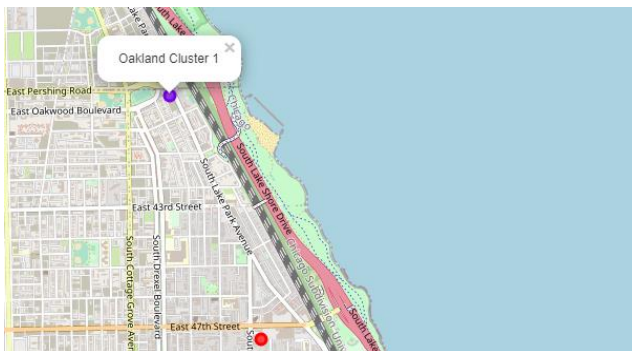
Figure 20 Oakland Community

Moreover, Oakland is north of Kenwood community. Kenwood has a population of 17189. Also, there is few Casual Dining restaurants in Kenwood.

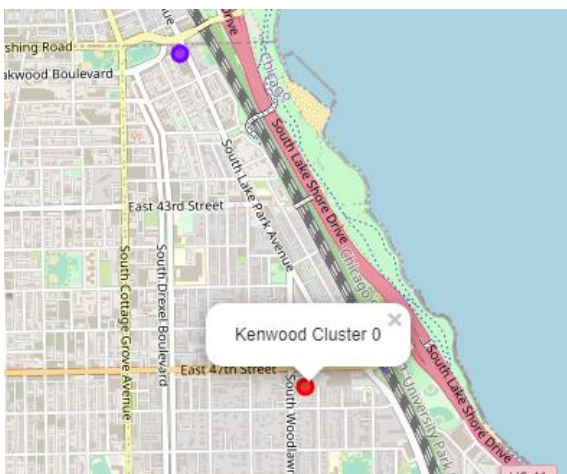| 38	Kenwood	17189	Bus Station	Gym / Fitness Center	Park	Pizza Place	Chinese Restaurant	Shoe Repair	Clothing Store

Figure 21 Kenwood Community

As result of the analysis, I find that the best place for the American Casual Dining restaurant is between south of Oakland community and north of Kenwood community, because of high traffic and availability of car parking. Moreover, opening a restaurant in this location will serve both communities of Oakland and Kenwood and will be in a good spot for tourists.

Conclusion

Chicago is a large city in population and size. It is also a tourist attraction and a transportation hub. There are a lot of business opportunities in the city. Clustering the city communities and analyzing these clusters will give investors a good idea on what type of investment is suitable for each cluster. After analyzing Chicago communities and clusters, I find that the best place for the American Casual Dining restaurant is between south of Oakland community and north of Kenwood community because of high traffic and tourist attraction.

References

https://en.wikipedia.org/wiki/Community_areas_in_Chicago

https://foursquare.com/

https://www.mapquest.com/