IBM Applied Data Science Through Coursera - Capstone

The Battle of Neighborhoods - Week 5 Final Report

Find the Best Neighborhood in Chicago to Open a Boutique Donut Shop

Loren Krokowski - June 2020

Introduction

Being able to foresee this growing trend, a client capitalized on the industry growth by opening his first specialty donut shop in Atlanta, Georgia approximately 10 years ago. This client continued to add additional locations in major metropolitan cities across the United States at a rate of 3 locations per year. Chicago is the next remaining large city that this client is looking to expand operations. It costs over $1,000,000 in starting capital to secure the ideal location, purchase and set up the equipment, hire and train employees, and begin a substantial marketing campaign. Picking the ideal neighborhood is the most critical step in this process because it is the costliest to change at a later date.

Even though the client already has existing locations nationwide, he partners with a local investor who will manage day to day operations and together they focus on creating a local presence where the donut shop has more of a family-owned mom-and-pop feel and not that of a national chain. This focus is strengthened by the locations being in urban areas with more densely populated neighborhoods, higher foot traffic during all times of the day and a mix of business and residential as opposed to suburban outskirts where the national chains reside and foot traffic is nonexistent.

The client has already identified the following parameters that have proven to be successful in locating the most ideal neighborhoods to open their existing locations:

- A large volume of nearby coffee shops - The client's previous research and experience has shown that his target audience is generally the same demographic as those that frequent coffee shops. A larger than average coffee shop presence indicates there will be a larger than average target audience for donuts.
- A competing donut shop - Contrary to most people's belief to open a new retail location far away from existing competition, the client believes it is more beneficial for all donut shops when a handful exist in the same neighborhood. This

helps to reinforce to the community that donuts themselves are popular and desired and revenue for all donut shops can increase.

- Higher than average population density – Communities with higher population densities that also contain foot traffic at varying times of the day generate a consistent ongoing target audience.

The goal of this project is to analyze communities in Chicago, develop an understanding of prominent retail and demographics for each community, and ultimately identify a short list of possibilities for the client to choose from. At that point it will be up to the client to visit each neighborhood, research leasing options, and make the final decision. Our work to narrow his options to just a handful of communities will save him weeks of time by not having to research and drive all over town himself, not to mention the fact that our analysis will be based on hard data to back our selections.

Gather Data

In order to perform this analysis, the project will take place in a Jupyter Notebook running Python programming language and will rely heavily on pandas dataframes. Below is a highlight of the data involved:

1. The Wikipedia page https://en.wikipedia.org/wiki/Community_areas_in_Chicago provides a table of 77 communities in the city of Chicago, including population density. This data will be gathered into a pandas dataframe using the read_html function. The dataframe will be cleaned up to remove unnecessary columns and make the column headers more relevant.
2. The dataframe will be updated to include latitude and longitude coordinates for each community which will then be used in the next step. This will involve using Geopy geocoders through Nominatim to update our dataframe for each community. The updated dataframe will not need to be cleaned more to only add columns for latitude & longitude, removing any additional data brought in by the Geopy process.

3. The Foursquare API will be used to gather venue data for each of the communities based on their coordinates as derived above. Venue data will be limited to a yet to be defined radius to secure the best volume of data.

This combination of gather communities in Chicago, pulling latitude and longitude coordinates for those communities, and then using those coordinates to retrieve venue data from Foursquare will provide the data needed to perform our overall analyses.

Methodology

The data analysis process can be broken down into the following core processes which will be explained in further detail below:

1. Retrieve list of communities in Chicago
2. Retrieve latitude and longitude coordinates for those communities
3. Retrieve venue information in close proximity to each community
4. Aggregate the most common venue categories for each community
5. Perform clustering analysis to differentiate communities with common venue categories
6. Identify preliminary targeted list of communities based on similar retail venues being most common
7. Cross reference preliminary targeted list with population density to reduce list to desired result

<u>Retrieve list of communities in Chicago</u> – A pandas dataframe was used to read the Wikipedia page for communities in Chicago and pull the data into Python. This Wikipedia page already included population density which is an additional piece of data needed for this project. Additional work was necessary to remove unnecessary columns, remove invalid rows and update column headers

List of community areas [edit]

Chicago community areas by number, population, and area

Number[8]	Name[8]	2017 population[9]	Area (sq mi.)[10]	Area (km^2)	2017 population density (/sq mi.)	2017 population density (/km^2)
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
10	Norwood Park	37,089	4.37	11.32	8,487.19	3,276.92
11	Jefferson Park	26,808	2.33	6.03	11,505.58	4,442.33
12	Forest Glen	19,019	3.20	8.29	5,943.44	2,294.78

```
In [74]: df=pd.read_html("https://en.wikipedia.org/wiki/Community_areas_in_Chicago")[0]

In [75]: df
```

Out[75]:

	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
5	06	Lake View	100470	3.12	8.08	32201.92	12433.23
6	07	Lincoln Park	67710	3.16	8.18	21427.22	8273.10
7	08	Near North Side	88893	2.74	7.10	32442.70	12526.20
8	09	Edison Park	11605	1.13	2.93	4235.40	1635.30
9	10	Norwood Park	37089	4.37	11.32	8487.19	3276.92
10	11	Jefferson Park	26808	2.33	6.03	11505.58	4442.33
11	12	Forest Glen	19019	3.20	8.29	5943.44	2294.78

Out[82]:

	Community	2017 Population	Area (sq mi)	2017 Population Density (/sq mi)
0	Rogers Park, Chicago	55062	1.84	29925.00
1	West Ridge, Chicago	76215	3.53	21590.65
2	Uptown, Chicago	57973	2.32	24988.36
3	Lincoln Square, Chicago	41715	2.56	16294.92
4	North Center, Chicago	35789	2.05	17458.05
5	Lake View, Chicago	100470	3.12	32201.92
6	Lincoln Park, Chicago	67710	3.16	21427.22
7	Near North Side, Chicago	88893	2.74	32442.70
8	Edison Park, Chicago	11605	1.13	4235.40
9	Norwood Park, Chicago	37089	4.37	8487.19
10	Jefferson Park, Chicago	26808	2.33	11505.58
11	Forest Glen, Chicago	19019	3.20	5943.44

Retrieve latitude and longitude coordinates for those communities

Geopy.geocoders was used to retrieve location information for each community from Nominatim. This produced a location column with the name of the community, county, zip code, and latitude/longitude coordinates. An additional point column was produced with just the 3-vector coordinates of latitude, longitude and altitude. Separate latitude and longitude columns were produced from the point column after which the location and point columns were removed.

```
In [86]: from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="foursquare_agent")
from geopy.extra.rate_limiter import RateLimiter
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)
chi_coord['location'] = chi_coord['Community'].apply(geocode)
chi_coord['point'] = chi_coord['location'].apply(lambda loc: tuple(loc.point) if loc else None)
chi_coord
```

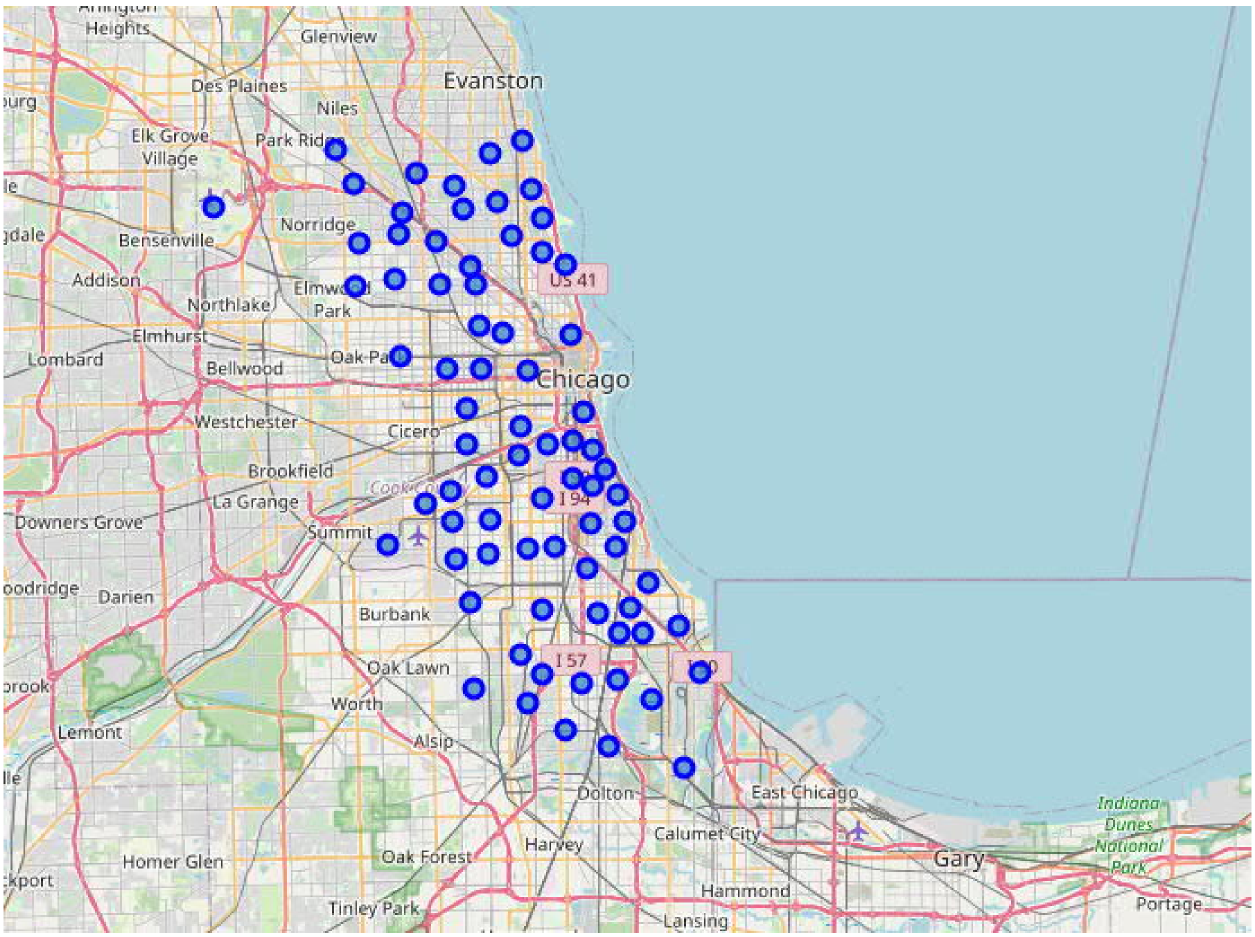
Out[86]:

	Community	2017 Population	Area (sq mi)	2017 Population Density (/sq mi)	location	point
0	Rogers Park, Chicago	55062	1.84	29925.00	(Rogers Park, Chicago, Cook County, Illinois, ...	(42.01053135, -87.67074819664808, 0.0)
1	West Ridge, Chicago	76215	3.53	21590.65	(West Ridge, Chicago, Cook County, Illinois, 6...	(42.0035482, -87.6962426, 0.0)
2	Uptown, Chicago	57973	2.32	24988.36	(Uptown, Chicago, Cook County, Illinois, 60640...	(41.9666299, -87.6555458, 0.0)
3	Lincoln Square, Chicago	41715	2.56	16294.92	(Lincoln Square, Chicago, Cook County, Illinoi...	(41.975989850000005, -87.6896163305115, 0.0)
4	North Center, Chicago	35789	2.05	17458.05	(North Center, Chicago, Cook County, Illinois,...	(41.9561073, -87.6791596, 0.0)
5	Lake View, Chicago	100470	3.12	32201.92	(Lake View, Chicago, Cook County, Illinois, Un...	(41.947050000000004, -87.65542878290054, 0.0)
6	Lincoln Park, Chicago	67710	3.16	21427.22	(Lincoln Park, Boystown, Lake View, Chicago, C...	(41.940297650000005, -87.63811710541756, 0.0)
7	Near North Side, Chicago	88893	2.74	32442.70	(Near North Side, Chicago, Cook County, Illino...	(41.9000327, -87.6344975, 0.0)
8	Edison Park, Chicago	11605	1.13	4235.40	(Edison Park, Chicago, Cook County, Illinois, ...	(42.0057335, -87.81401633833357, 0.0)
9	Norwood Park, Chicago	37089	4.37	8487.19	(Norwood Park, Chicago, Cook County, Illinois,...	(41.9855895, -87.80058173001102, 0.0)

Out[94]:

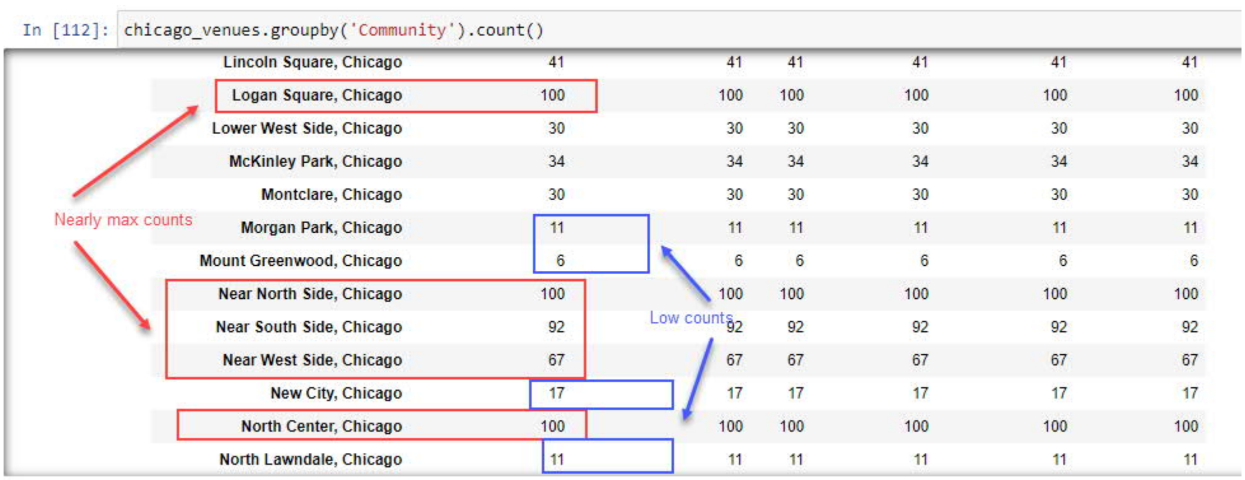
	Community	2017 Population	Area (sq mi)	2017 Population Density (/sq mi)	Latitude	Longitude
0	Rogers Park, Chicago	55062	1.84	29925.00	42.010531	-87.670748
1	West Ridge, Chicago	76215	3.53	21590.65	42.003548	-87.696243
2	Uptown, Chicago	57973	2.32	24988.36	41.966630	-87.655546
3	Lincoln Square, Chicago	41715	2.56	16294.92	41.975990	-87.689616
4	North Center, Chicago	35789	2.05	17458.05	41.956107	-87.679160
5	Lake View, Chicago	100470	3.12	32201.92	41.947050	-87.655429
6	Lincoln Park, Chicago	67710	3.16	21427.22	41.940298	-87.638117
7	Near North Side, Chicago	88893	2.74	32442.70	41.900033	-87.634497
8	Edison Park, Chicago	11605	1.13	4235.40	42.005733	-87.814016
9	Norwood Park, Chicago	37089	4.37	8487.19	41.985590	-87.800582
10	Jefferson Park, Chicago	26808	2.33	11505.58	41.969738	-87.763118
11	Forest Glen, Chicago	19019	3.20	5943.44	41.991752	-87.751674

For visual appeal I then plotted these coordinates of Chicago communities on a map of Chicago using the folium.Map library:



<u>Retrieve venue information in close proximity to each community</u>

Using a Foursquare API, a query was run to retrieve up to 100 venues within 500 meters of a sample community. This sample was run first to obtain an idea of results produces on the one community to use to apply to the set as a whole to maximize results without burdening the system. This first test produced only 33 venue hits, so I increased the radius to 1,000 meters and then received 100 hits. The next step was to run the query on the entire set of communities. The Foursquare API failed repeatedly during the process (approximately 20 attempts, never surpassing the first 3-5 communities). I adjusted the radius down to 500 meters and then Foursquare API process performed successfully on the first try, but only produced an average of 17 venues per community. I then increased the radius to 750 meters. The Foursquare API crashed a few times but made it more than halfway through the list each time and eventually completed the task successfully. This produced an average of 34 venues per community which is less than the 100 limit desired, but I quickly determined by running a .groupby('Community').count() function that the list has some communities with few venue counts, which would not be desired for our project anyway, and many other communities had close to or up to 100 counts which still provided a solid sample set.



```
In [112]: chicago_venues.groupby('Community').count()
```

Lincoln Square, Chicago	41	41	41	41	41	41
Logan Square, Chicago	100	100	100	100	100	100
Lower West Side, Chicago	30	30	30	30	30	30
McKinley Park, Chicago	34	34	34	34	34	34
Montclare, Chicago	30	30	30	30	30	30
Morgan Park, Chicago	11	11	11	11	11	11
Mount Greenwood, Chicago	6	6	6	6	6	6
Near North Side, Chicago	100	100	100	100	100	100
Near South Side, Chicago	92	92	92	92	92	92
Near West Side, Chicago	67	67	67	67	67	67
New City, Chicago	17	17	17	17	17	17
North Center, Chicago	100	100	100	100	100	100
North Lawndale, Chicago	11	11	11	11	11	11

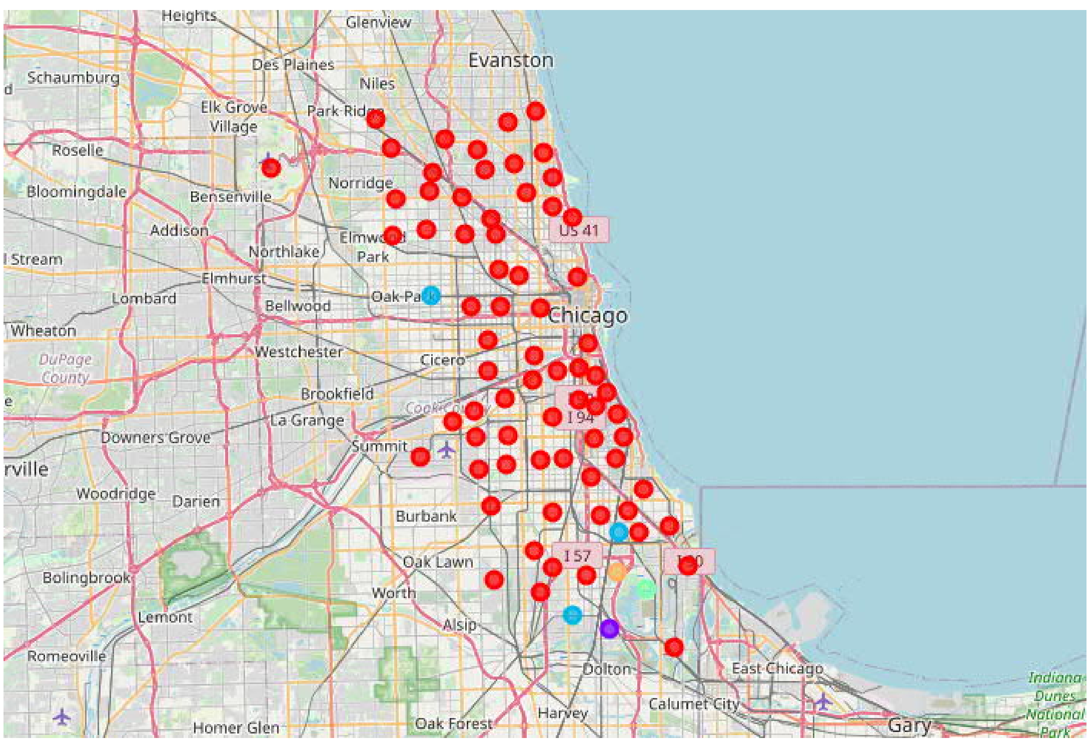
<u>Aggregate the most common venue categories for each community</u>

In order to aggregate most common venue categories for each community, I first calculated the mean frequency of each venue category by community, and then transferred that data into a pandas dataframe while listing 1st, 2nd, 3rd etc. most common venue categories up to 10th most common. At this point I recognized that there is some overlap in descriptions for categories such as Coffee Shop and Café being separate venues, as well as Donut Shop and Bakery, Bar and Pub, etc. I will need to keep this in mind as the process moves closer to analyzing a few final select communities.

	Community	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, Chicago	Mexican Restaurant	Asian Restaurant	Donut Shop	Park	Grocery Store	Bakery	Korean Restaurant	Coffee Shop	Sandwich Place	Chinese Restaurant
1	Archer Heights, Chicago	Mexican Restaurant	Clothing Store	Cosmetics Shop	Mobile Phone Shop	Grocery Store	Arts & Crafts Store	Big Box Store	Nightclub	Seafood Restaurant	Bar
2	Armour Square, Chicago	Chinese Restaurant	Pizza Place	Bar	Bus Station	Asian Restaurant	Sandwich Place	Seafood Restaurant	Storage Facility	Bakery	Bagel Shop
3	Ashburn, Chicago	Train Station	Park	Liquor Store	Cosmetics Shop	Mexican Restaurant	Locksmith	Automotive Shop	Clothing Store	Light Rail Station	Pizza Place
4	Auburn Gresham, Chicago	Grocery Store	Discount Store	Southern / Soul Food Restaurant	Fast Food Restaurant	Basketball Court	Lounge	Park	Boutique	American Restaurant	Department Store

Perform clustering analysis to differentiate communities with common venue categories

A clustering process was then performed relying on the k-means functionality to group communities together into clusters of common popular venues. This process was performed in other projects for other major metropolitan cities and produced more varied responses but for our sample in Chicago, approximately 90% of the communities were aggregated into one cluster, represented by the red circles below:



At first this seemed like data that might now be valuable but as I looked at the other 4 clusters it is easy to see they do not belong. The primary distinguishable difference between the primary cluster and the others has to do with restaurants. Each community in the primary cluster has some combination of food or beverage venue generally as the majority of the top 10 most common venues while the communities in the other clusters generally do not.

	2017 Population	Longitude	Cluster Labels New2	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	55062	-87.670748	0	Mexican Restaurant	Park	Sandwich Place	Pizza Place	Bar	Bakery	American Restaurant	Theater	Chinese Restaurant	Train Station
1	76215	-87.696243	0	Indian Restaurant	Pakistani Restaurant	Grocery Store	Park	Automotive Shop	Convenience Store	Clothing Store	Sandwich Place	Business Service	Fast Food Restaurant
2	57973	-87.655546	0	Coffee Shop	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Diner	Bar	Thai Restaurant	Grocery Store	Sushi Restaurant	Mexican Restaurant
3	41715	-87.689616	0	Bar	Karaoke Bar	Korean Restaurant	Food Truck	Bakery	Thai Restaurant	Bus Station	Juice Bar	Rental Car Location	Sushi Restaurant
4	35789	-87.679160	0	Bar	Coffee Shop	Mexican Restaurant	Pub	Chinese Restaurant	Brewery	Pizza Place	American Restaurant	Dessert Shop	Bank
5	100470	-87.655429	0	Pizza Place	Sports Bar	Bar	Gay Bar	Park	New American Restaurant	General Entertainment	American Restaurant	Coffee Shop	Café

VS

	2017 Population	Longitude	Cluster Labels New2	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
24	95260	-87.764851	2	Park	Intersection	Athletics & Sports	Grocery Store	Flea Market	Fish Market	Flower Shop	Fish & Chips Shop	Filipino Restaurant	Field
45	2254	-87.596714	2	Park	Intersection	Motel	Train Station	Bus Station	Farmers Market	Event Service	Exhibit	Eye Doctor	Fabric Shop
51	27742	-87.637823	2	Convenience Store	Clothing Store	Grocery Store	Train Station	Park	Discount Store	Bank	Burger Joint	Construction & Landscaping	Field

Identify targeted list of communities based on similar retail venues being most common

To begin the process to narrow down the list of communities into a more targeted focus for the client, I ran a query to select only those communities where the Coffee Shop venue was presented within the top 4 most common venues. This produced approximately a dozen communities. The number varied slightly based on re-running this entire analysis on separate days. Look over this preliminary list it was evident that a few communities could be removed right away since they were either near the airport, or Coffee Shop was the 3rd or 4th most common venue and the other top venues were not in a retail category that we desire for our donut shop client.

	Community	2017 Population	Area (sq mi)	2017 Population Density (/sq mi)	Latitude	Longitude	Cluster Labels New2	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
2	Uptown, Chicago	57973	2.32	24988.36	41.966630	-87.655546	0	Coffee Shop	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Diner	Bar	Thai Restaurant	Grocery Store	Sushi Restaurant	Mexican Restaurant
4	North Center, Chicago	35789	2.05	17458.05	41.956107	-87.679160	0	Bar	Coffee Shop	Mexican Restaurant	Pub	Chinese Restaurant	Brewery	Pizza Place	American Restaurant	Dessert Shop	Bank
7	Near North Side, Chicago	88893	2.74	32442.70	41.900033	-87.634497	0	Coffee Shop	Steakhouse	Hotel	Gym	Bar	Pizza Place	Italian Restaurant	Clothing Store	Café	Gym / Fitness Center
12	North Park, Chicago	18842	2.52	7476.98	41.984955	-87.722933	0	Park	Coffee Shop	Sandwich Place	Bus Station	Kitchen Supply Store	Gymnastics Gym	Department Store	Soccer Field	Nature Preserve	Convenience Store
15	Irving Park, Chicago	54606	3.21	17011.21	41.953365	-87.736447	0	Bar	Coffee Shop	Gym	Sandwich Place	Pizza Place	Convenience Store	Donut Shop	Border Crossing	Mexican Restaurant	Pharmacy
21	Logan Square, Chicago	73046	3.59	20347.08	41.928568	-87.706793	0	Bar	Coffee Shop	Cocktail Bar	Café	Latin American Restaurant	Pizza Place	Art Gallery	Mexican Restaurant	Asian Restaurant	Brewery
36	Grand Boulevard, Chicago	22313	1.74	12823.56	41.813923	-87.617272	0	Coffee Shop	Liquor Store	BBQ Joint	Train Station	Art Gallery	Jazz Club	Lounge	Restaurant	Sports Bar	Bakery
39	Hyde Park, Chicago	26827	1.61	16662.73	41.794225	-87.592562	0	Coffee Shop	Sandwich Place	Café	Bookstore	Pizza Place	Thai Restaurant	Spa	Mediterranean Restaurant	Italian Restaurant	Sushi Restaurant
57	McKinley Park, Chicago	15767	1.41	11182.27	41.831700	-87.673664	0	Coffee Shop	Donut Shop	Video Store	Gas Station	Diner	Mexican Restaurant	Bus Station	Chinese Restaurant	Thrift / Vintage Store	Discount Store
75	Edgewater, Chicago	55965	1.74	32163.79	41.983369	-87.663952	0	Sandwich Place	Coffee Shop	Bakery	Mexican Restaurant	Italian Restaurant	Sushi Restaurant	Burger Joint	Theater	Asian Restaurant	Restaurant

<u>Cross reference preliminary targeted list with population density to reduce list to desired result</u>

The final step was to take this list or targeted communities that have already shown to have frequent coffee shops and other similarly desired retail and analyze the population density of each community. Simply looking at the .describe() function in Python gives us the mean, min, max, quartile, and standard deviations:

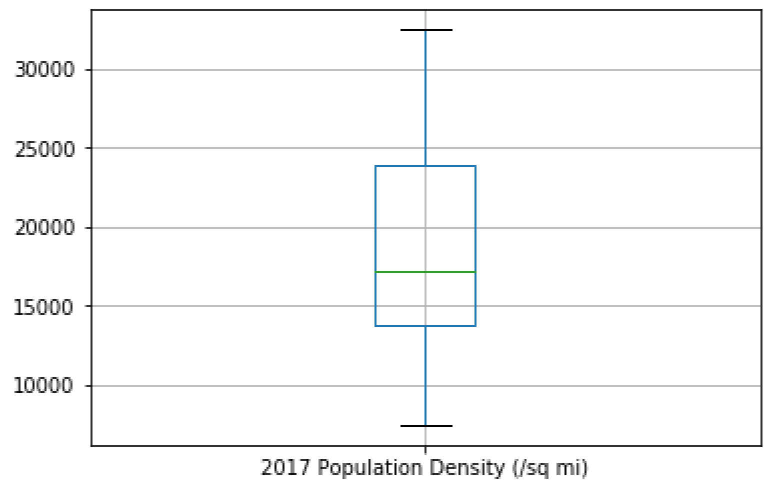
```
In [66]: chi_nearfinal.describe()
```

Out[66]:

	2017 Population	2017 Population Density (/sq mi)	Latitude	Longitude	Cluster Labels New2
count	10.000000	10.000000	10.000000	10.000000	10.0
mean	45002.100000	19255.673000	41.911287	-87.668283	0.0
std	24860.920656	8398.769576	0.072578	0.045852	0.0
min	15767.000000	7476.980000	41.794225	-87.736447	0.0
25%	23441.500000	13783.352500	41.848783	-87.699885	0.0
50%	45197.500000	17234.630000	41.940967	-87.668808	0.0
75%	57471.000000	23828.040000	41.963999	-87.639760	0.0
max	88893.000000	32442.700000	41.984955	-87.592562	0.0

Using this table identifies the mean as 19,256, with a min and max of 7,477 and 32,443 respectively, and a standard deviation of 8,399. There are only two communities, Near North Side and Edgewater, that exceed 1 standard deviation to the upside, and both are nearly the same. Uptown comes next with just inside of one standard deviation but still outside of 75%. A box plot using the matplotlib library in Python illustrates how these three higher population density neighborhoods really stretch out the mean:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff5fc4c65c0>
```



Results

Aggregating venues within a defined radius produced 2,580 venues across our community group. By tallying the mean of each venue group for each community and creating a dataframe of most common venues, I was readily able to filter communities where coffee shops were prevalent. This reduced our initial list of 77 communities down to 12 before performing any additional analysis.

The k-means clustering at first didn't appear to provide much meaning but at second glance it confirms that Chicago as a whole readily embraces a significant volume of restaurant and comparable food & beverage retail that aligns with the metropolitan focus our client is looking for. The clustering map visually identifies the north and central areas of Chicago as geographic areas to focus while and leads us to stray away from the southern side of Chicago.

The population density analysis provided additional key information that readily provided 3 communities as standing out from the rest of the group. This reduction gave us the final information we needed to provide to our client.

Discussion

The ultimate goal of this analysis was to reduce the list of 77 communities down to a handful for the client to then physically inspect and get a feel for each to make his final decision. There are ultimately going to be many other factors that come into this decision that a more detailed analysis could identify but was not included in our initial scope of work. Additional factors that could be considered include income levels, crime reports, oversaturation of existing donut shops, and future plans for each neighborhood. The communities themselves vary in size and using a fixed size for each might not produce the best results. The latitude and longitude coordinates generally pin the center of community while the most nearby retail drag may stretch outside of 750 meters, limiting the data.

The venue data created 302 unique categories. It would be interesting to aggregate these into a smaller number of categories where certain venues such as coffee shops, cafes, bakeries, and donut shops all become one category. A community with one of each of these unique venues might not have a high prevalence of any, and another unrelated venue category may steal the show with multiple venues, but if commonly similar venue categories were aggregates, perhaps it would be easier to differentiate communities.

Another analytical approach would be to calculate a scaled value for desired or undesired venue category and sum the scores of each venue value to compare communities. This approach may produce a more fair result because it would produce values that can be compared against one another versus the approach used in this project where we only know which venues are most common. One community in this project with coffee shop as the most common with 3 looks more appealing than another community where there are 5 coffee shops but the coffee shop venue is only the 3rd or 4th most common because other venue categories have more than 5 each.

Conclusion

The primary objective of this project was to identify ideal locations for our client to invest over $1,000,000 to open a new boutique donut shop. Selecting the wrong location can be costly either by way of reduced success in that location or the physical expenditure to relocate. The analysis process began with an assumption that a plethora of nearby coffee shops in higher population density communities was the target. Our results showed that once we used out final analysis of 10 communities and focused on the 3 with highest population density, those 3 also happen to have coffee shops as their number 1 or 2 most common venue. This is at least a positive indication that our initially focused direction was correct.

Ultimately the population density analysis on the reduced community list based on common coffee shop venues further reduced our search to just three communities. These three communities of Near North Side, Edgewater, and Uptown all have significantly high population densities while having coffee shops as their number one or two most common venue, as well as other retail food and beverage in the other top most common venue categories.