Donut Pick the Wrong Location

Choosing the Ideal Donut Shop Location for a Client
IBM Applied Data Science Capstone
Loren Krokowski – June 2020

Introduction

- Selecting the ideal donut shop location is extremely important. With an expected financial outlay of over $1,000,000, determining the location is a critical first step for the business.
- Based on previous experience, the process will focus on looking for high population density communities in Chicago that specifically have a high incidence of nearby coffee shops.
- Having other nearby donut shops and bakeries is actually a plus

Methodology

- The first step is to retrieve a list of the 77 communities of Chicago from Wikipedia using Python pandas and bring the data into a dataframe and clean the dataframe

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 (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
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

Methodology

- The next step is to utilize geopy.geocoders in conjunction with Nominatim to retrieve latitude and longitude coordinates for the communities and clean the list.

```
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]:

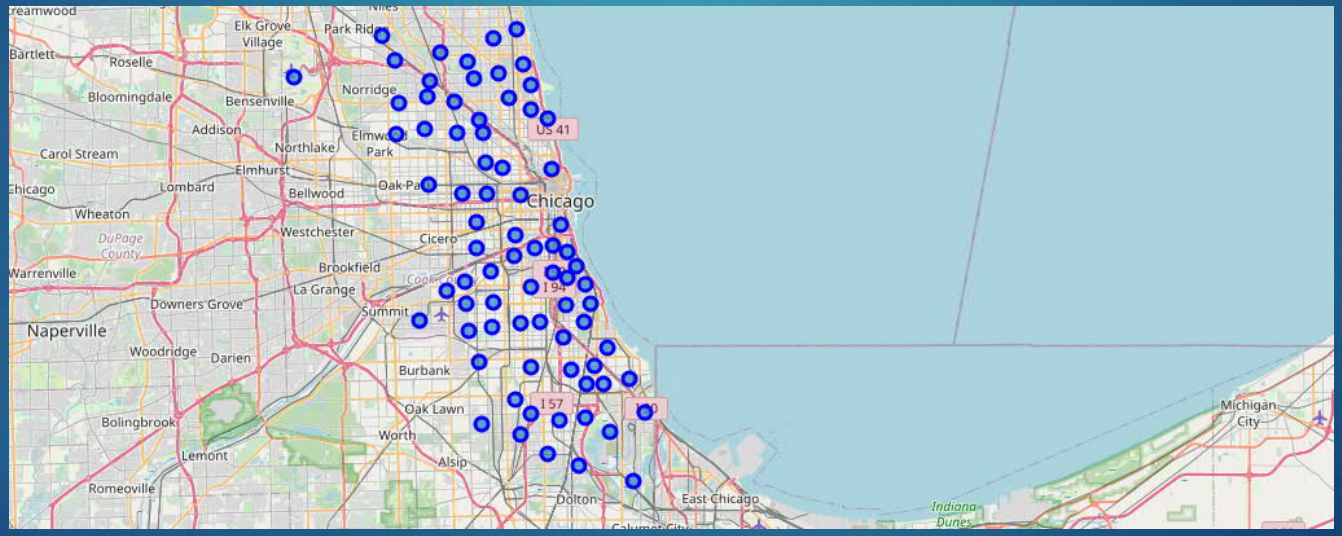
	Community	2017 Population	Area (sq mi)	2017 Population Density (/sq mi)	location	point
0	Rogers Park, Chicago	55062	1.84	29925.00	(Rogers Park, Chicago, Cook County, Illinois, ...	(42.01053135, -87.67074819664808, 0.0)
1	West Ridge, Chicago	76215	3.53	21590.65	(West Ridge, Chicago, Cook County, Illinois, 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, -87.68861633051 15, 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, -87.655428782900054, 0.0)
6	Lincoln Park, Chicago	67710	3.16	21427.22	(Lincoln Park, Boystown, Lake View, Chicago, C...	(41.940297650000005, -87.638117105417 56, 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

Methodology

- Next, we plot the communities onto a map of Chicago based on their latitude & longitude coordinates using folium.Maps in Python



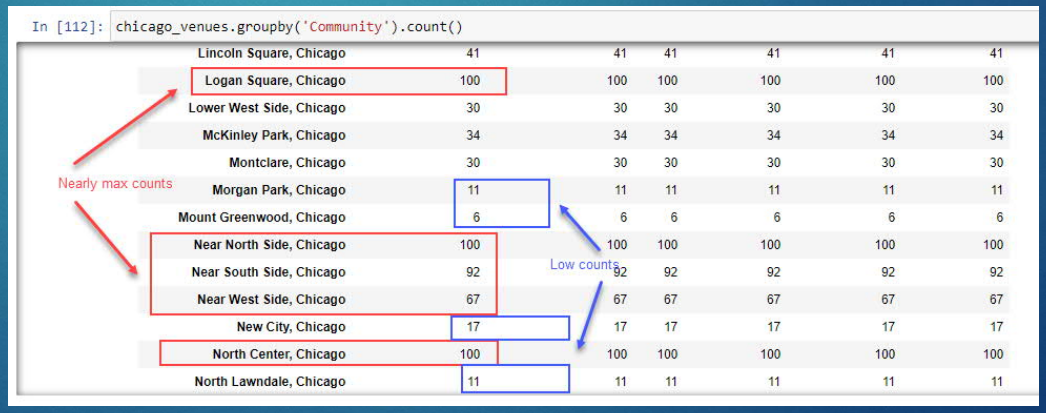
Methodology

- Using the Foursquare API, we then retrieve a list of nearby venues for each based on:
  - Latitude and longitude coordinate of each community
  - Max results per community = 100
  - Radius of each coordinate to retrieve venues = 750 meters

--This produced a list of 2,580 venues for an average of 34 venues per community--

Methodology

- At first glance this might not seem like enough venues. Looking at the groupby('Community").count() we can see there are some communities that indeed do not have many venues in that small of a radius, and those will likely be weeded out in this process. There are however ample communities with higher venue counts, up to the max of 100.



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

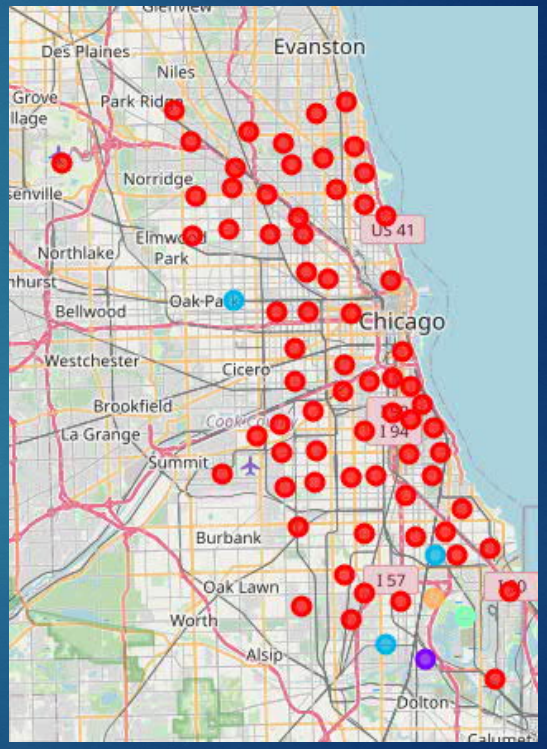
Methodology

- We next aggregate the most common venue categories for each community

	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

Methodology

- A cluster analysis is performed using k-means in Python to cluster communities based on common top venue categories.
- The results show 90% of the communities fall within the same cluster
- Initially this doesn't seem helpful
- We then see the benefit to focus on central and northern areas of Chicago, while not focusing on the southern side of Chicago.
- We also realize Chicago is full of communities that focus heavily on food and beverage retail venues



Methodology

- Confirmation that the primary cluster has an abundance of retail food and beverage venues

	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

Methodology

- Using Python & pandas functionality, filter list of communities to produce a reduced targets list where the Coffee Shop venue is in the top 4 of most common venue categories
- Remove obvious anomalies

	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

Methodology

- Analyze population density as provided in initial data from Wikipedia

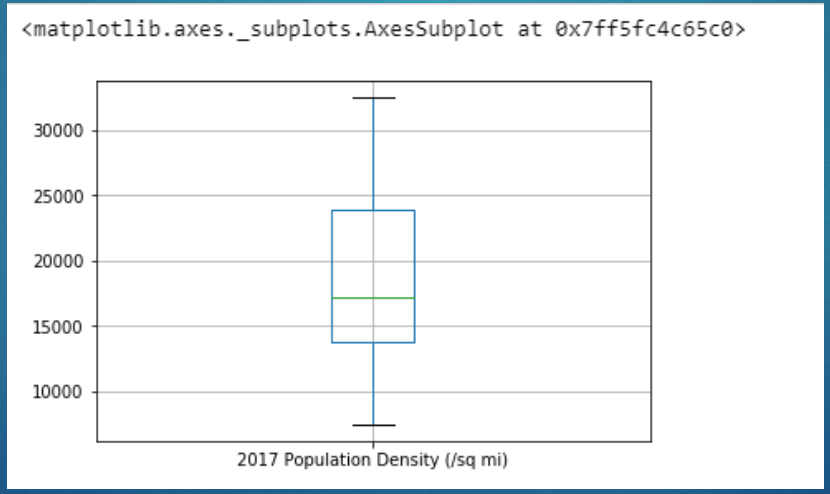
```
In [66]: chi_nearfinal.describe()
```

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

Methodology

- Using a boxplot function in Python through the matplotlib library we can visualize the characteristics of the spread of population densities for our targeted list of 10 communities with Coffee Shop in the top 4 most common venues



Results

- Aggregating venues produced 2,580 venues across our 77 communities
- Use k-means clustering to understand metropolitan geographic and fine tune filtering parameters
- Reduce list of 77 down to 12 communities based on Coffee Shop prevalence
- Further reduce 12 down to 10 based on anomalies that do not fit the target community profile
- Analyze population densities of the 10 remaining communities to find 2 with exceptionally high and one additional with nearly exceptionally high population densities

Conclusion

- Project is to find ideal location for boutique donut shop in Chicago based on prevalence of nearby coffee shops and higher than typical population density

Edgewater

Near North Side

Uptown