Ideal Chicago Neighborhood Brewery Locations

Introduction to the Business Problem Background

Chicago is a city of 2.7M residents with a metro-wide population of over 9.5M. The city is made up of 77 community areas or neighborhoods many with distinct populations, cultures, and attractions.

Additionally, Chicago is a world-wide travel destination that attracts over 50M visitors a year. The city itself has over 130 hotels with over 40,000 hotel rooms. Over 1.6M riders take public transportation on a daily basis and while not one of the top transportation systems in the world, most popular city locations are easily accessible through this system, taxis, or ride-hailing apps.

Possibly at a peak, and with a definite change in environment given world-wide pandemic conditions, the number of breweries in the Chicago-area was at 167 as of 2019, possibly the most in the US.

Introduction to the Business Problem
Problem Description

A group of stakeholders are determined to open a new brewery in the city of Chicago. They believe the current economic conditions will cause many smaller breweries to close and that this will provide an opportunity. They have developed the following location requirements that they want to have met through a detailed analysis:

- Located in a neighborhood that is currently relatively underserved by breweries or brewpubs
- Located near tourist attractions, but especially near existing hotels
- If possible, also serve a local neighborhood population with a higher level of income

Data Requirements

To find a solution to the given problem, a detailed data analysis will be required. One that will need all available data in the following areas:

- Location of all Chicago neighborhoods
- Population of neighborhoods
- Average income levels of neighborhoods
- Tourist levels of neighborhoods(represented by hotel venue locations)
- Existing brewery locations by neighborhood

This data will be used to determine clusters of neighborhoods that meet the requirements of the stakeholders. After scraping website to obtain the data, an analysis using k-means clustering will be performed on the data.

Data Collection

1) Geographical coordinates of Chicago neighborhoods from cityofchicago.org

Example:

Location	PERIMETER	AREA	COMAREA_	COMAREA_ID	AREA_NUMBE	COMMUNITY	AREA_NUM_1	SHAPE_AREA	SHAPE_LEN
-87.60914087617894, 41.84469250265398	0	0	0	0	35	DOUGLAS	35	46,004,621.16	31,027.05

Data Collection (continued)

2) Population of Chicago neighborhoods from chicago.gov

Example:

Num	Community Area	Population 2010	Population 2000	Difference	Percentage
1	Rogers Park	54,991	63,484	-8,493	-13.4%

Data Collection (continued)

3) Income by neighborhood
from http://www.chicagocomputerclasses.com/average-city-chicago-income/

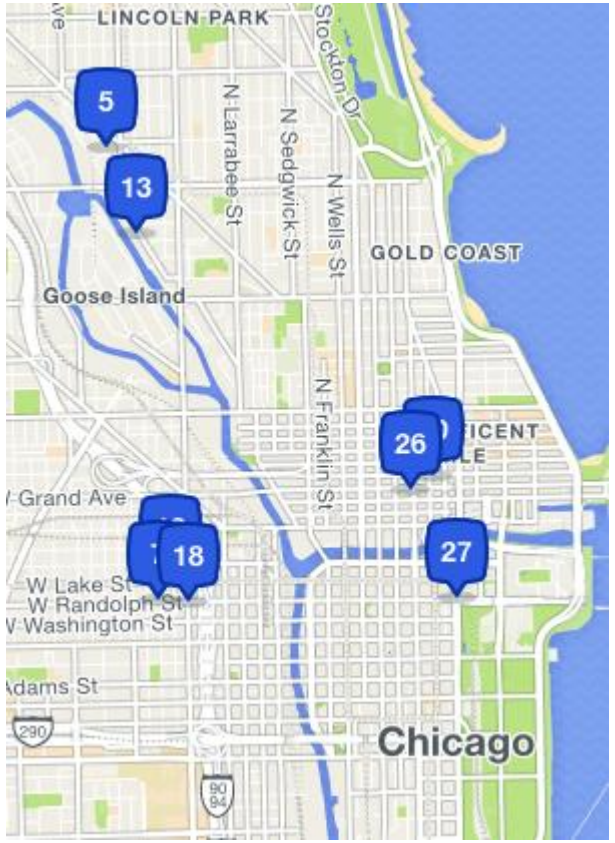
Example:

COMMUNITY AREA NAME	AVERAGE
Near North Side	$88,669.00

Data Collection (continued)

4) Foursquare API

- Brewery venue location data
- Hotel/Tourist venue location data

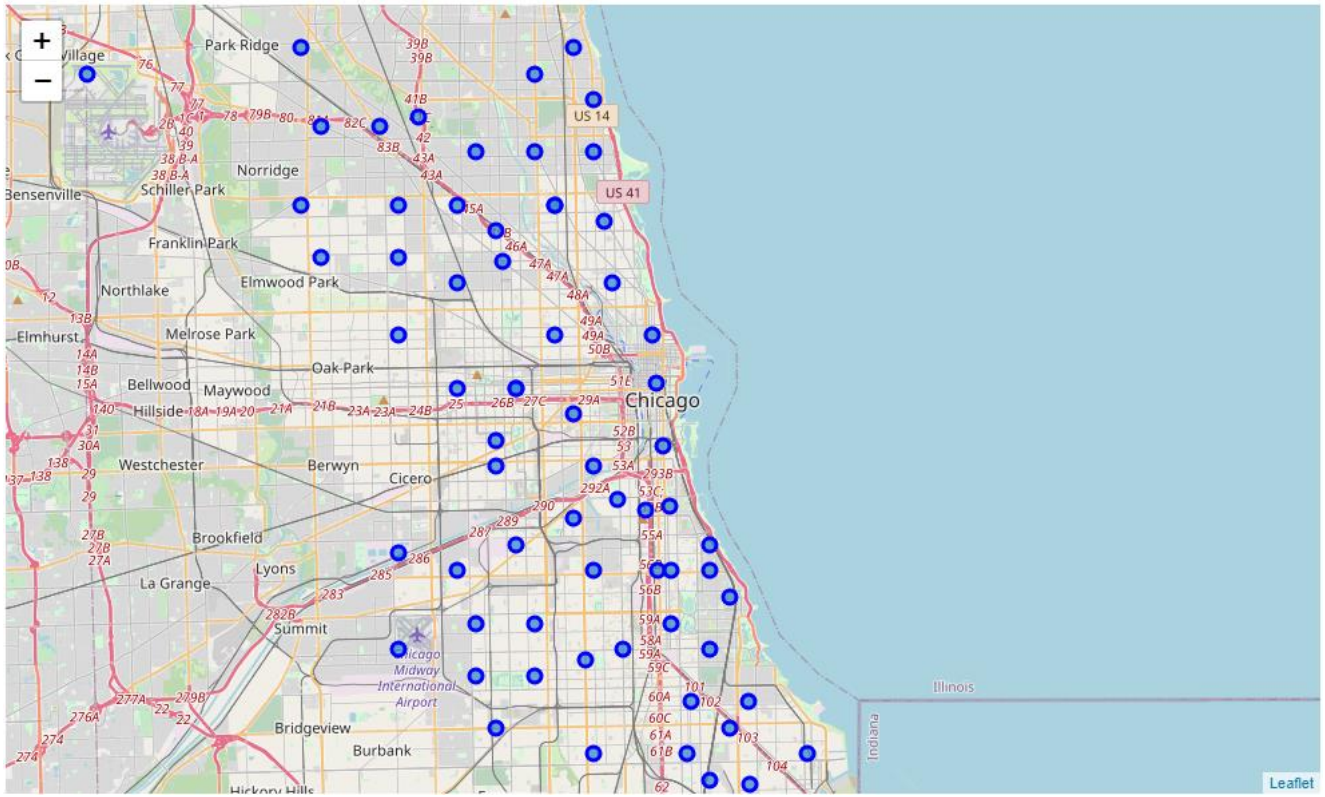


Methodology

- The Analytic Approach used for this scenario is a descriptive approach to show relationships. The goal is to show Chicago neighborhoods that exhibit similar properties based on the data available. There is an expectation that the clusters could represent areas that are of high-value to the success of a brewery, while also showing a lack of currently being served by existing breweries. In this case, k-means clustering will be used to perform the analytics.

Results

- Initially, all of the neighborhood data was pulled as defined in the Data Collection section of this report. In this case, it was discovered that the Chicago neighborhoods are actually Community Areas, but the application is still the same. See a map of all 77 Chicago Community Areas at the right.



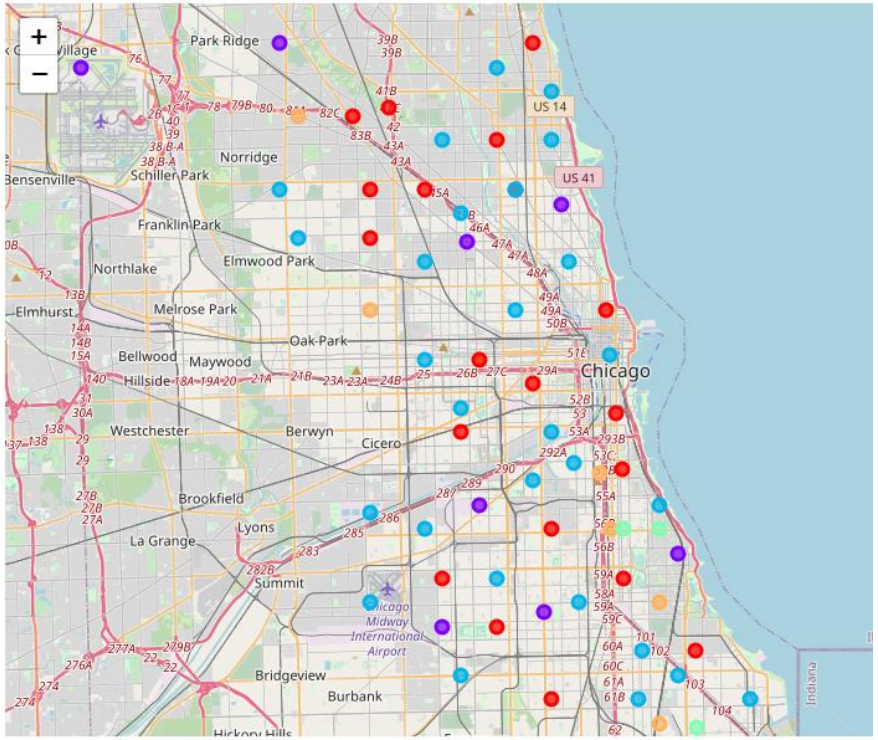
Results (continued)

- Pulled from Foursquare, the data shows that some Community Areas are already served by a number of local breweries and hotels. A sample of the resulting data is shown below. The data was then fitted for clusters by k-means analysis.

	AREA_NUMBER	COMMUNITY	LATITUDE	LONGITUDE	POPULATION	INCOME	BREWERIES	HOTELS
0	14	ALBANY PARK	41.970000	-87.720000	51992	21323	3	1
1	57	ARCHER HEIGHTS	41.810000	-87.730000	13142	16134	0	4
2	34	ARMOUR SQUARE	41.833333	-87.633333	13455	16148	1	4
3	70	ASHBURN	41.750000	-87.710000	43792	23482	0	0
4	71	AUBURN GRESHAM	41.740000	-87.660000	46278	15528	0	0
...	...	...	...	...	...	...	...	...
72	65	WEST LAWN	41.770000	-87.720000	33108	16907	0	6
73	53	WEST PULLMAN	41.680000	-87.630000	27742	16563	0	0
74	2	WEST RIDGE	42.000000	-87.690000	76215	23040	0	3
75	24	WEST TOWN	41.900000	-87.680000	84502	43198	18	9
76	42	WOODLAWN	41.780000	-87.600000	23268	18672	1	0

Results (continued)

- The neighborhoods were grouped into 5 clusters.
- Cluster 2, when filtered to get rid of outliers, shows promise as potential neighborhoods meeting our criteria:
- In particular, two communities show promise; Clearing and Edgewater. Clearing is near Midway Airport, which is the reason for the number of hotels. Would visitors staying near the airport appreciate a local brewery to attend during their stay? Possibly. Edgewater already has a small number of breweries, however, compared to other communities in this list, it has a larger population with a higher income, as well as more local hotels.



AREA_NUMBI	COMMUNITY	LATITUD	LONGITUI	POPULATIC	INCOM	BREWE	HOTELS	CLUSTE
2	WEST RIDGE	42	-87.69	76215	23040	0	3	2
31	LOWER WEST SIDE	41.85	-87.66	32888	16444	2	2	2
36	OAKLAND	41.82	-87.6	6645	19252	0	3	2
57	ARCHER HEIGHTS	41.81	-87.73	13142	16134	0	4	2
60	BRIDGEPORT	41.8375	-87.6475	33637	22694	2	2	2
64	CLEARING	41.78	-87.76	25891	25113	0	15	2
77	EDGEWATER	41.99	-87.66	55965	33385	4	6	2

Discussion

- Further iterations would have to be performed for better results, possibly using better data from Foursquare. Either adjusting the radius of the venue search, or better fitting the results to the boundaries of the neighborhoods, as there was some overlap of venues. Also, suburbs surrounding the neighborhoods were not taken into account.

Conclusion

- Overall, this basic clustering of neighborhoods on a limited data set did not show a perfect example of a cluster that would best be suited for the introduction of a new brewery. Two possibilities in one of the clusters seemed to show the variables were similar for that cluster and better suited for this project. However, increasing the amount of data available, weighing certain variables over others, and better accounting for location information (venue locations within a neighborhood as well as surrounding locations) could yield valuable insight.