What's the best place for the next brewery?

Introduction

The brewery market in Minnesota had exploded over the last years. The total production of beer doubled in the last decade[1]. The twin cities brewery and taproom market is booming but it still has room for more development when compared to other similar metro areas across the US such as Portland, OR, and Denver, CO which have significant more breweries per capita[1]. Since I live in the Twin cities and I enjoy visiting the breweries I thought to use this project and identify which cities in the Twin Cities metro area could be the home to the next brewery. Any number of investors in the metro area or the brewery market can use this report to inform their decision and limit the number of cities they will investigate further as potential brewery location candidates.

Map of the Twin Cities, showing different breweries[2].

---

[1] https://www.mprnews.org/story/2019/05/15/taproom-beer-minnesota-growth-surly
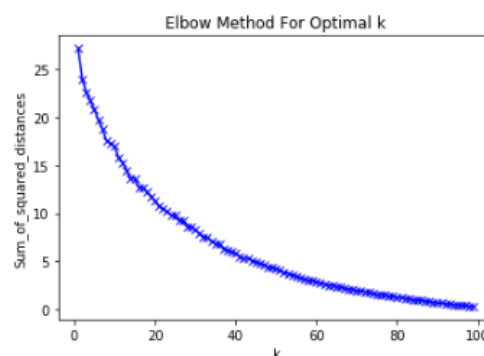[2] https://reneeslimousines.com/tours/brewery-tours/

Data

The Foursquare API was used to retrieve the venue data from all the cities in the Twin Cities metro area. A detailed list of all the cities in the metro was retrieved from the https://www.zipcodestogo.com/Minnesota/ webpage. My assumption is that if a city that does not have as brewery is similar in character (i.e. has similar top venues categories) with a city that has a brewery is a good candidate. Further future analysis that considers financial performances of different breweries could help improve the recommendation.

Methodology

The first step was to read the data that were manually retrieved from the *zipcodestogo* website in a *pandas* data frame. The resulted dataframe includes 141 rows, each with each city in the Twin cities metro area. Then using the *geopy* package I retrieved the geographical coordinates (latitude and longitude) for each city and saved them alongside with their names in a new dataframe named df. I used the *folium* package to create a map of the Twin Cities metro area, marking each city with a blue point, to quickly visualize the city data and look for any obvious mistakes.

Next, I used the *Foursquare* API to retrieve all the venues for each city. I limited the search to only 0.5km from the center because I was limited to the number of requests I could do successfully. Additionally, there is a LIMIT of 100 venues in the API return. I converted the results in a dataframe to continue for further analysis. I performed some descriptive analytics to identify how many venues I retrieved for each city and how many unique categories of venues exist in my data and if brewery were one of them.

In order to cluster the cities, I needed to do some data preprocessing. First, I used the *One Hot Encoding* approach, where each city has a value of 0 for venue categories that are not present and a positive value if venues belonging to the specific category exist in the city. A quick check revealed how many of the cities have breweries to ensure that the key data which in this case is the presence of breweries exist in

our dataset. Then I used k-means to cluster the cities based on the venue categories they have. Initially I tried to find the optimum k using the elbow method, however there was not a clear elbow point indicating that the cities have differences that cannot be grouped completely. Since this is just an initial explorative analysis, we can still use the k-means method for now and in the future add more data or a more complex method for higher definition of the clusters. I proceeded using k values ranging from 4 to 15 to identify the k value that would result in the best

Elbow method showing no clear elbow point.

definition of the cities that would be a good place for a new brewery, i.e. the number of cities left to consider for breweries.

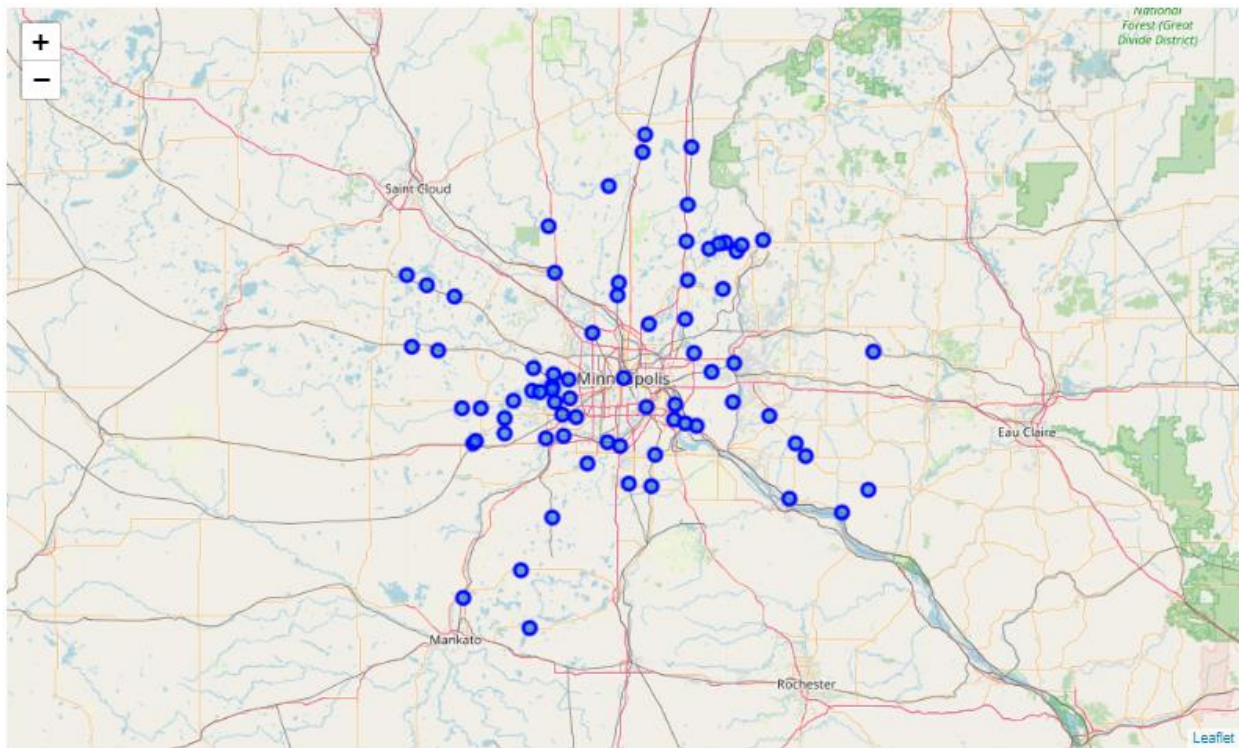After the clustering I identified the cluster that had citied that included breweries. The cities that re in this cluster but do not have breweries themselves are considered good candidates for a new brewery. Finally, I visualized the cluster in the map, showing in blue all the cities in the clyster and in red the ones that already have a brewery.

Results

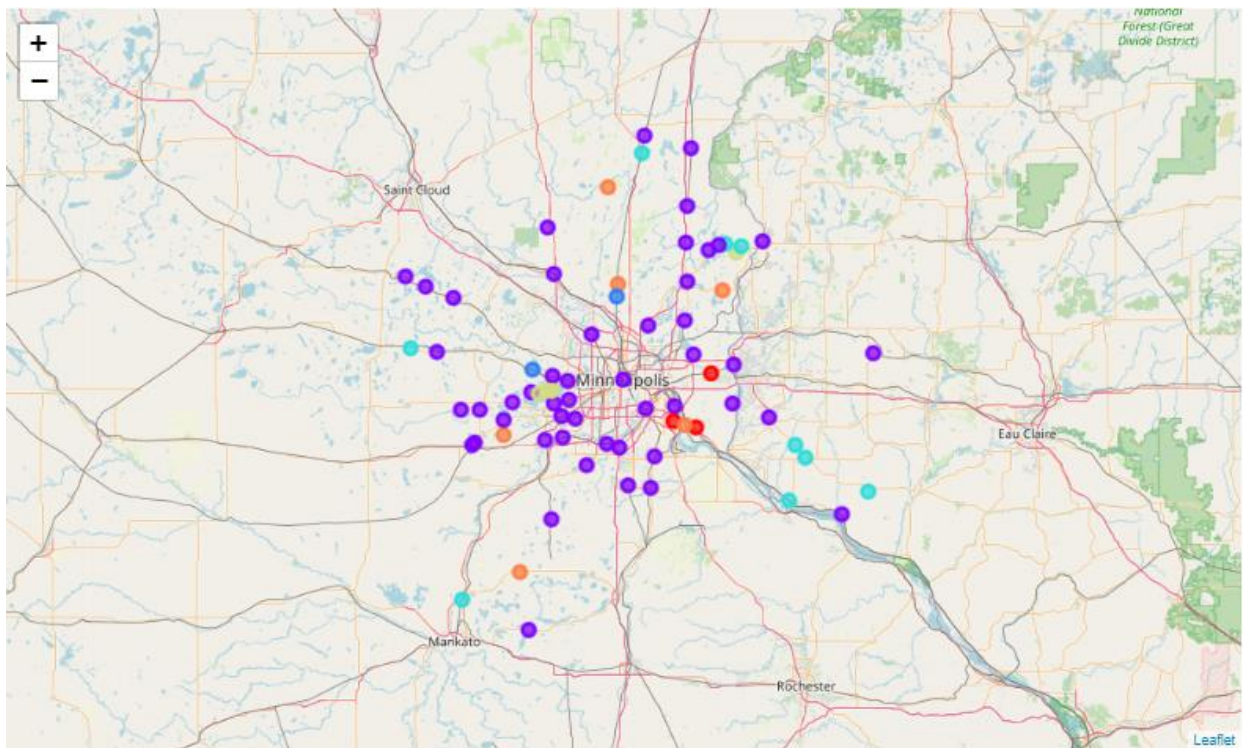There are 142 cities in the Twin Cities Metro area shown in blue below:

Using the Foursquare API I retrieved venues for 113 of them:

City	Number of Venues	City	Number of Venues	City	Number of Venues
Hanover	100	Lakeland	11	Maple Plain	5
Cologne	100	Long Lake	11	Silver Creek	5
Cambridge	100	South Saint Paul	11	Wayzata	4
Newport	64	Prior Lake	11	Plum City	4
Minneapolis	57	Burnsville	11	Mayer	4
Cleveland	50	Braham	10	Lake Elmo	4
Prescott	48	Zimmerman	10	New Richmond	4
Hastings	44	Lindstrom	10	Center City	4
Buffalo	39	Afton	10	Andover	4
Chanhassen	36	Saint Bonifacius	10	Beldenville	3
Rockford	35	North Branch	10	Cedar	3
Dayton	33	Delano	9	Eden Prairie	3
Waterville	32	Rush City	9	Montgomery	3
River Falls	31	Howard Lake	9	Cokato	3
Clearwater	30	Watertown	8	Hammond	3
Hamburg	28	Norwood Young America	8	Hager City	3
Bay City	27	Annandale	8	Hudson	3
Excelsior	26	Rosemount	8	Young America	3
Castle Rock	25	Maple Lake	8	Minnetonka Beach	2
Chaska	24	New Prague	7	Bethel	2
Forest Lake	21	Bayport	7	Saint Michael	2
Osseo	20	Harris	7	Houlton	2
Waverly	19	Belle Plaine	7	Cottage Grove	2
Shakopee	17	Maiden Rock	6	Somerset	2

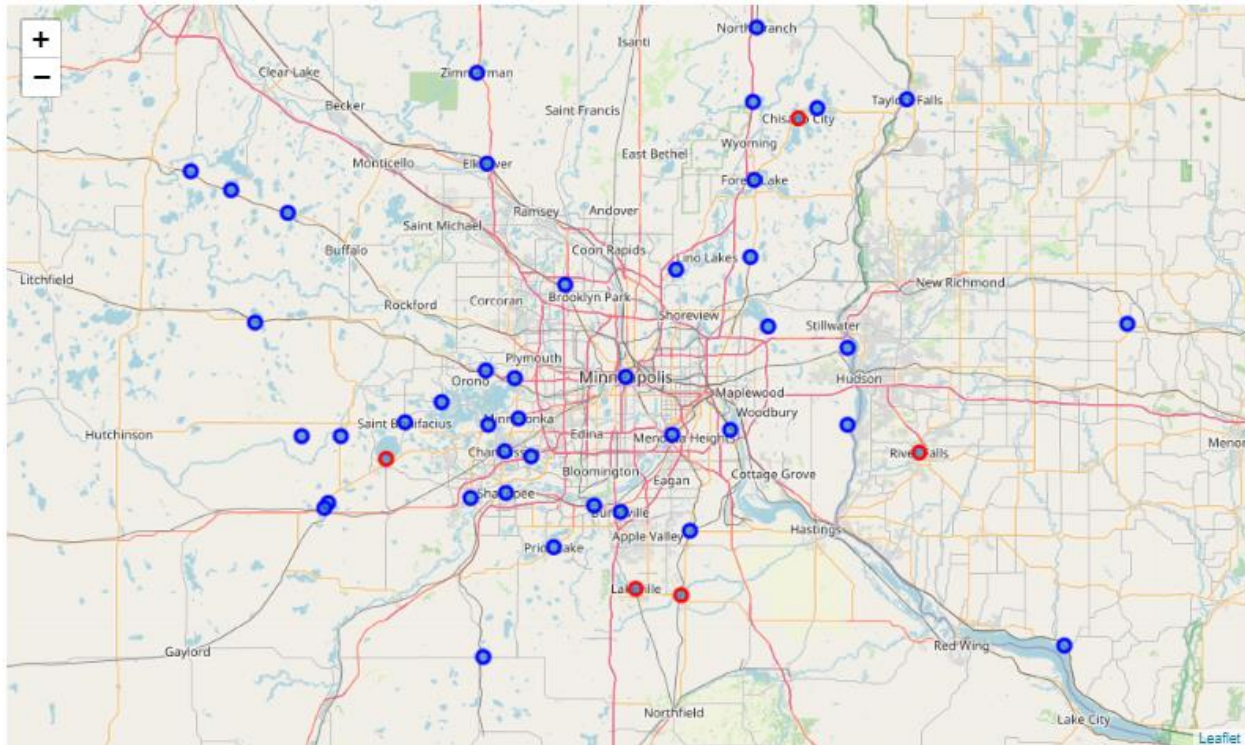|					
Clear Lake	17	Inver Grove Heights	6	Monticello	2
Willernie	17	Hugo	6	Norwood	2
Stillwater	17	Elysian	6	Kasota	1
Waconia	17	Albertville	6	Woodville	1
Mound	15	East Ellsworth	5	Marine On Saint Croix	1
Taylors Falls	14	Spring Park	5	Saint Paul Park	1
Farmington	13	Glenwood City	5	Santiago	1
Elk River	13	Crystal Bay	5	Stanchfield	1
Savage	13	South Haven	5	Spring Valley	1
Chisago City	13	Mendota	5	Dalbo	1
Stacy	12	Big Lake	5	Saint Paul	1
Minnetonka	12	Shafer	5	Circle Pines	1
Hampton	12	Baldwin	5	Saint Francis	1
Lakeville	12	New Germany	5		

There are 296 unique Venue categories, including "Brewery". Specifically, the 20 cities that have a brewery are: Baldwin, Buffalo, Castle Rock, Chisago City, Clear Lake, Cologne, Dayton, Elk River, Excelsior, Farmington, Hampton, Hanover, Lakeville, Prescott, River Falls, Rockford, Shakopee, Stillwater, Waconia, Waterville,

The cities were clustered into 6 clusters shown below. I included an additional 7th cluster for the cites that returned no venues at all.

 Cluster 1 included the cities where breweries were popular. The total number of cities in this cluster is 93, while the cities that have breweries are 14. Therefore the 79 cities that do not have breweries are good candidates for the next brewery.

The cities that re good candidates are visualized bellow in blue, while the cities that already have a brewery are colored red.

Afton, Albertville, Andover, Annandale, Bay City, Bayport, Belle Plaine, Bethel, Big Lake, Braham, Buffalo, Burnsville, Cambridge, Cedar, Chanhassen, Chaska, Circle Pines, Clearwater, Cleveland, Cottage Grove, Crystal Bay, Dalbo, Dayton, Delano, Eden Prairie, Elysian, Forest Lake, Glenwood City, Hamburg, Hammond, Hanover, Hastings, Houlton, Howard Lake, Hudson, Hugo, Inver Grove Heights, Lake Elmo, Lakeland, Lindstrom, Long Lake, Maple Lake, Maple Plain, Mendota, Minneapolis, Minnetonka, Monticello, Mound, New Germany, New Prague, New Richmond, Newport, North Branch, Norwood, Norwood Young America, Osseo, Prior Lake, Rockford, Rosemount, Rush City, Saint Bonifacius, Saint Francis, Saint Michael, Saint Paul, Santiago, Savage, Silver Creek, Somerset, South Saint Paul, Spring Park, Stacy, Stillwater, Taylors Falls, Watertown, Waterville, Waverly, Wayzata, Willernie, Zimmerman.

Through this analysis I was able to achieve a **44%** reduction on the number of candidate cities for a new brewery.

Discussion

As can be seen on the Foursquare data the diversity of the cities in the metro quite big. There are cities that have 100 or more venue, while other have 1 or less. More information on the cities' population and socioeconomical aspects would help normalize the venue distribution and provide more representative data. This is reflected in the absence of an elbow point in the elbow methods indicating that more detail in the data is needed to create more defined k-means clusters. However, the method is sufficient to separate the cities into clusters, with one having 'brewery" as a popular venue. In that cluster there are 79 cities that do not have breweries but are like cities that they have. These 79 cities are good candidates to host the next brewery.

An alternative approach would be to use a supervised method to find cities that would be a good candidate using the Foursquare data and a classification method that would predict if a brewery would be a good venue category for the city, based on cities that already have one.

Conclusion

Breweries in the Twin Cities are a popular venue continuously attracting new customers. An investor will need to consider all possible cities to decide where to build a new brewery. Reducing the number of cites considered by **44%** is a great saving of time, effort and resources that allows the market investigation part of the new business development to move faster and smoother. The next steps would be to continue narrowing down the results till there is only a few cities left where a proper search for local resources (such as a lot or a building) can start.