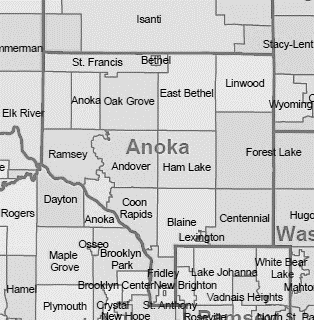
Leveraging Location Data of Anoka County for Optimal Viability of New Storefronts

By: Jonathan Resch



Contact Information: jonathanresch@gmail.com

Published: 12/13/2020

Table of Contents

[Executive Summary/Description of Problem/Introduction 3](#_Toc58794307)

[Background/Introductory Section 3](#_Toc58794308)

[Description of Data 3](#_Toc58794309)

[Methodology 4](#_Toc58794310)

[Exploratory Data Analysis 4](#_Toc58794311)

[Machine Learning algorithms and pre-processing for machine learning 5](#_Toc58794312)

[One Hot Encoding 5](#_Toc58794313)

[K-Means Clustering 5](#_Toc58794314)

[Results 6](#_Toc58794315)

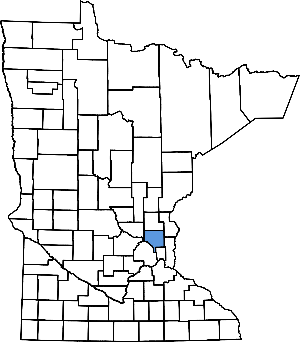
[Discussion 8](#_Toc58794316)

[Conclusion 9](#_Toc58794317)

[Acknowledgements 9](#_Toc58794318)

[References 9](#_Toc58794319)

# Executive Summary/Description of Problem/Introduction

 The Covid-19 pandemic has presented hardship on many people this year. Truly this is an unprecedent time of sadness for many in the forms of health, loss of loved ones, and financial burdens. There are a few people who have been able to adapt to the new pressures of Covid and found opportunity for new business ventures from everything from video chat software, contact-less scheduling, or traveling distances to nurse for hospitals in need. These select people have seen a boom in their personal finances and should they want to invest back into their community want to ensure their new windfall is put in the best place possible. Using location data from Foursquare, this report hopes to elucidate the best location in Anoka county for a new store front. In addition, this report will generate suggestions on what may maximize consumer adoption in that area.

# Background/Introductory Section

  Minnesota borders Canada in the middle of the United States. Anoka county is north of the capital city of Minneapolis. The area can be best described as rural to increasingly more suburban as you approach Minneapolis. The county has a total population of approximately 356,921. (https://www.census.gov/quickfacts/anokacountyminnesota)

 Location is the predominant factor in deciding where the store front is placed as financial data is outside of the scope of this report. Location data is additionally used in this report to imply popularity and success of the corresponding business.

## Description of Data

  Anoka county, MN consists of 20 cities. Using the Foursquare API, venues were searched within a 2500m radius of the center of the latitude and longitude of the listed cities. 580 venues were found. There were 8 cities with less than 10 venues listed with these parameters and 7 cities that generated 68% of the data used. Data includes the city's name, venue's name, latitude and longitude of both city and venue, and a classification of the venue.

  While there are risks in taking into account these cities with few venues listed on Foursquare, these could also be untapped potential. Thankfully there is data supporting popularity of businesses in other cities that can give ideas on what could be successful in the cities of these counties.

# Methodology

The cities of Anoka County were inputted and used with the Foursquare API to explore surrounding venues. The radius of the search was determined to be best suited at a 2500m radius. Some cities were sparsely populated with venue data but this was determined to be for the best in that these cities, such as Nowthen and Bethel, are on the border of Anoka county and extending the radius further would generate more data points but diminish the accuracy of our scope (keeping to Anoka county).

## Exploratory Data Analysis

The data was examined city by city to get an idea of the Foursquare data. Notably, more rural cities tended to have less data whereas moving in closer to the capital city, Minneapolis, there tended to be more data. This trend indirectly effects the analysis in the sense that more rural cities will have more inherent uncertainty to match the scarcity of the data. This can be illustrated well by figure 1 shown below.

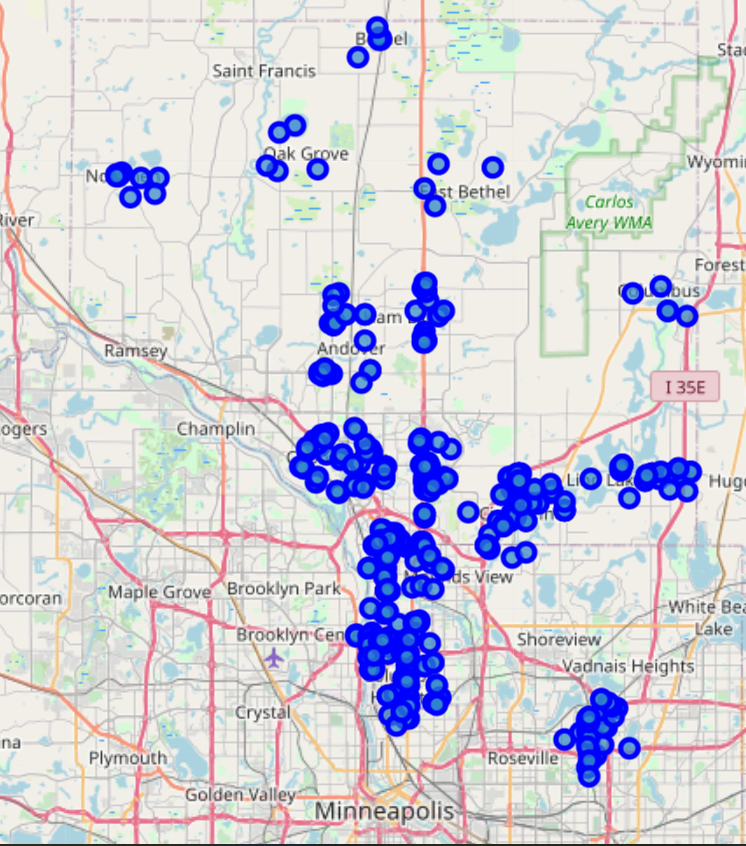


Figure 1. This image generated on the Folium Python package. The blue circles indicate venues. These venues intrinsically appear to make their own clusters but this is a product of the search we put forth to the API. (Radial search around a latitude and longitude associated with the relevant city.)

## Machine Learning algorithms and pre-processing for machine learning

### One Hot Encoding

One hot encoding was performed on the categorical data returned from the Foursquare API. This data was then further processed to generate a top ten category of venues for each city. So that the data was shown as below.



### K-Means Clustering

Using the One-Hot Encoding data, K-means Clustering was performed. Using 4 clusters, the algorithm separated out the cities into different groups. One group was used as the algorithm as a catch-all group that encapsulated 70% of the cities. The other three groups that the algorithm returned showed a good grouping that will be elucidated further.

K-means Clustering was performed a second time (with 4 clusters) exclusively on the catch-all group that was returned in the previous clustering attempt. This yielded further separation of these groupings and allowed for easier classification and descriptions of these groups.

# Results

The clusters generated by K-means are shown below. Further, approximate demographic descriptions for each cluster will be given as well.

|  |  |  |
| --- | --- | --- |
| Cluster | Cities(number of venues) |  |
| 0 | Anoka(6), Columbus(5) |  |
| 1 | Saint Francis(1) |  |
| 2 | East Bethel(4), Bethel(6), Nowthen(9) |  |
| 3 | Circle Pines(56) |  |
| 4 | Columbia Heights(57), Fridley(55), Hilltop(61) |  |

|  |  |  |
| --- | --- | --- |
| Cluster | Cities |  |
| 5 | Blaine(60), Spring Lake Park(55) |  |
| 6 | Ramsey(49) |  |
| 7 | Oak Grove(5), Lino Lakes(8), Lexington(40), Andover(22), Coon Rapids(44), Centerville(18), Ham Lake(19) | These cities were not separated into clusters explicitly. |

# Discussion

The clusters generated can be surmised as approximate demographics of the main customers of that city. For this report, your own opinions on each cluster can be made by looking in the results section. Below are my suggestions as to what these groups are.

Cluster 0 consists of Anoka and Columbus. Parks and repair stores are prominent in the top 3 most common venues listed. I would describe this grouping as outdoors men and women who have a tendency to DIY projects. The racetrack in Anoka additionally gives the subtext that this demographic is what I would call “**Gear-Heads**” or individuals that love to work on or play with automobiles in some capacity.

Cluster 1 consists solely of Saint Francis. The predominant venues are night life spots and yoga studies. This one warranted further examination as an outlier. It turns out there was only one venue listed in this category and it happened to be an Equestrian farm labeled as a night life spot. This cluster can be safely ignored as this makes no sense and a quick google search does not turn up the listed business. This is an **Anomaly**.

Cluster 2 consists of East Bethel, Bethel, and Nowthen. Prominent venues all center around home DIY, construction and landscaping. This demographic I would describe as the DIY type that would flip homes or the type that would enjoy doing projects on their homes in their spare time. The venues also suggest that alcohol is favored in these town with the dive bar and a brewery making an appearance. To give them a label, I would describe them as **Fixer-Uppers**. (In the sense that they like to fix up their homes.)

Cluster 3 consists solely of Circle Pines. With the top three being a pizza joint, video store, and a bar, this sounds like a younger demographic that enjoys taking in a movie at home with pizza. This may be oversimplifying the city of circle pines but any items associated with a movie night at home would probably do well here. I’d call this a **Movie and Pizza** group.

Cluster 4 consists of Columbia Heights, Fridley, and Hilltop. The most frequent venues include a park, a gym, and a pharmacy. These seem to be relatively health conscious individuals who enjoy picnics at the park. The majority of the other popular venues happen to be food service places of some sort. These **Picnic People** have an affinity to parks and good food.

Cluster 5 consists of Blaine and Spring Lake Park. 5 of the top 6 venues happen to be restaurants in this cluster. These are the **Foodies**. People that like to spend their free time going out and eating good food. In addition to all the food, a gym makes a mention in this group. Since one food place frequented is a sports bar, there could be an association in watching sports as well in this group. This could be a great spot for a future restaurant.

Cluster 6 consists of Ramsey. Ramsey is dominated with fast foods and bars but also furniture stores. The people of Ramsey appear to be the **Furnishers** of this county.

Cluster 7 consists of Oak Grove, Lino Lakes, Lexington, Andover, Coon Rapids, Centerville, and Ham Lake. This is sort of the **Catch-All** group that was not separated out. These 7 cities either related too well to the other cities or had too few data points to be separated from the other cities. Oak Grove and Lino Lakes had few associated venues and I would hesitate to pick these as a location without more data. The other 5 cities have a decent amount of data. These 5 cities related very well to all other cities in Anoka county in that they did not have defining characteristics to be separated out using 2 different iterations of K-means clustering. In that vein, these could be great cities to create a generic store front such as a boutique or trying out a different type of store in Anoka county. If your idea feels niche for this area, these 5 cities may be the best melting pots to test your idea.

# Conclusion

In the Discussion section, this report attempts to elucidate the predominant demographics of each city. For a prospective store front, take note of the cities that best suit your interests and try and find real estate in those areas. For an auto shop, try Anoka or Columbus; for new restaurants, try Blaine or Spring Lake Park; and if your idea doesn’t fit into any of the demographics/categories listed above, there are a selection of 5 cities that were unable to be separated out and consisted of a decent amount of data where your ideas could fit in there.

# Acknowledgements

I’d like to thank IBM and Coursera for these wonderful classes they put together and enabled me to learn so much of this information.

I’d also like to thank my wife Mercedes for without her love and support, this would just be a daydream and not something I could explore in the way that I am able to.

# References

1. <https://www.coursera.org/learn/applied-data-science-capstone/>
2. <https://developer.foursquare.com/developer/>