Exploring Boba Tea Shops in Denver Metro Area

IBM Applied Data Science Capstone Project

Kami Rothchild

December 29, 2020

1.Introduction

Boba tea shops are becoming more popular in the USA these days. They originated in Taiwan in the early 1980s. "Boba", "pearl", or "bubble" refers to chewy tapioca balls that are usually served with tea or milk tea. Both come with a variety of flavors. Boba tea shops are unique and hard to find, unlike coffee shops that you can see at almost every corner of the city. Boba tea tends to be more common and popular in Asian countries, so a city like Denver with a large Asian population will be a good place to invest in a boba tea shop.

Denver is one of the fastest growing cities in the USA. More people are moving to Denver every year. From Denver Post on March 2020, "The population for the city and county of Denver was 727,211 on July 1, the date used as the peg for annual estimates the U.S. Census Bureau is releasing Thursday. Denver, the state's most populous county, it grew by nearly 11,000 people, or 1.5%, compared to mid-2018. Since 2010, the city's population has grown by a cumulative 21%, adding more than 127,000 people." So as a resident of this city and a person who is interested in boba tea business, I wonder if someone is looking to open a Boba tea shop in the Denver Metropolitan Area, where would it be the best area to open it? What kind and how popular boba tea shops are in the area?

This study will be helpful for not only an entrepreneur who wants to invest in a boba tea shop, but also an individual who is interested in the boba tea shop business.

2. Data

2.1 Data Sources

Firstly, I have to find a data set that include city, latitude, longitude, and city populations in Denver Metropolitan Area. When I say Denver Metropolitan Area, I mean Denver and the cities around Denver with a relatively high population density. In this case, I will refer to "Denver–Aurora–Lakewood, CO Metropolitan Statistical Area" which consists of ten Colorado counties as following: The City and County of Denver, Arapahoe County, Jefferson County, Adams County, Douglas County, the City and County of Broomfield, Elbert County, Park County, Clear Creek County, and Gilpin County.

- The latitude and longitude data is found on https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=CO.
- For the list of cities in in Denver Metropolitan Area, I scape it from https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Colorado.
- I download city populations data from https://worldpopulationreview.com/states/cities/colorado

Next, I pull the venue data from Foursquare API trying to get a list of boba tea shops that exist in Denver Metropolitan Area. We will use latitude and longitude of Denver as a center of the map and searching for boba tea shops in radius 40 miles or about 65 kilometers. From this step, the features we will get are venue ID, venue name, latitude, longitude, city, zip code, category ID, and category name.

Not only venue data, I can also get venue detail for each boba tea shop to see how they perform via Foursquare API. I pull venue details to see rating, price, and also what are their busy time to determine the best shop hours. I will also use Clustering for this data to see the differences. From this data source, the features of boba tea shop detail we expect to get are rating, price, hours, popular hours, description, and popularity score.

2.2 Data Cleaning

Since data is downloaded and scraped from different sources, there are a couple of problems occurring during the cleaning step.

Firstly, I had a hard time pulling or scraping the data from the website. So, I decide to use a simple method by downloading the data and uploading it back on my Githup instead. That solved the problem very well and the data will stay as long as I have it on my Github repository.

Secondly, there is no latitude and longitude data for every city on the list I have. After I scraped the list of cities in Denver Metropolitan Area, I have 45 cities on the list. However, after I merge the tables from latitude and longitude data and the list of the cities, I only have 30 cities with latitude and longitude. Since the objective of getting latitude and longitude here is just to explore these cities located in the area, that is enough in this case and I did not continue exploring more with city locations.

3. Methodology

First of all, after I have collected the data including cities in Denver Metropolitan Area with latitude and longitude, city populations, and boba tea shops from Foursquare API, I use python *folium* library to visualize location of cities in Denver Metropolitan area.

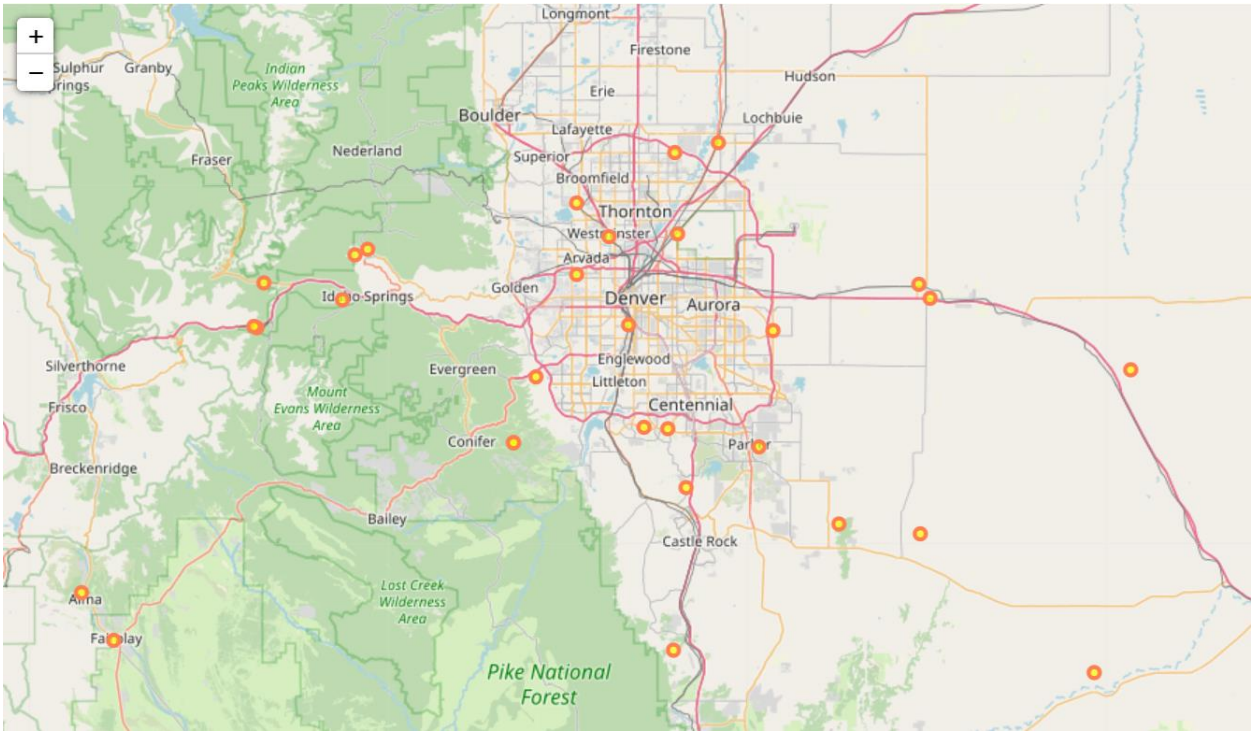
Second step, I start with applying exploratory data analysis to explore and summarize data characteristics. *Bar chart* is used to compare the different numbers of boba tea shops in Denver Metropolitan area. *Scatter plot* and *correlation coefficient* are used to see if there is any linear association between variables.

In the last step, I focus on all 41 boba tea shops in detail. Foursquare API will be used again to pull detail data of each boba tea shop from the list we extract from the first step. After that, we will apply *k-means clustering* to partition boba tea shops into groups that have similar characteristics.

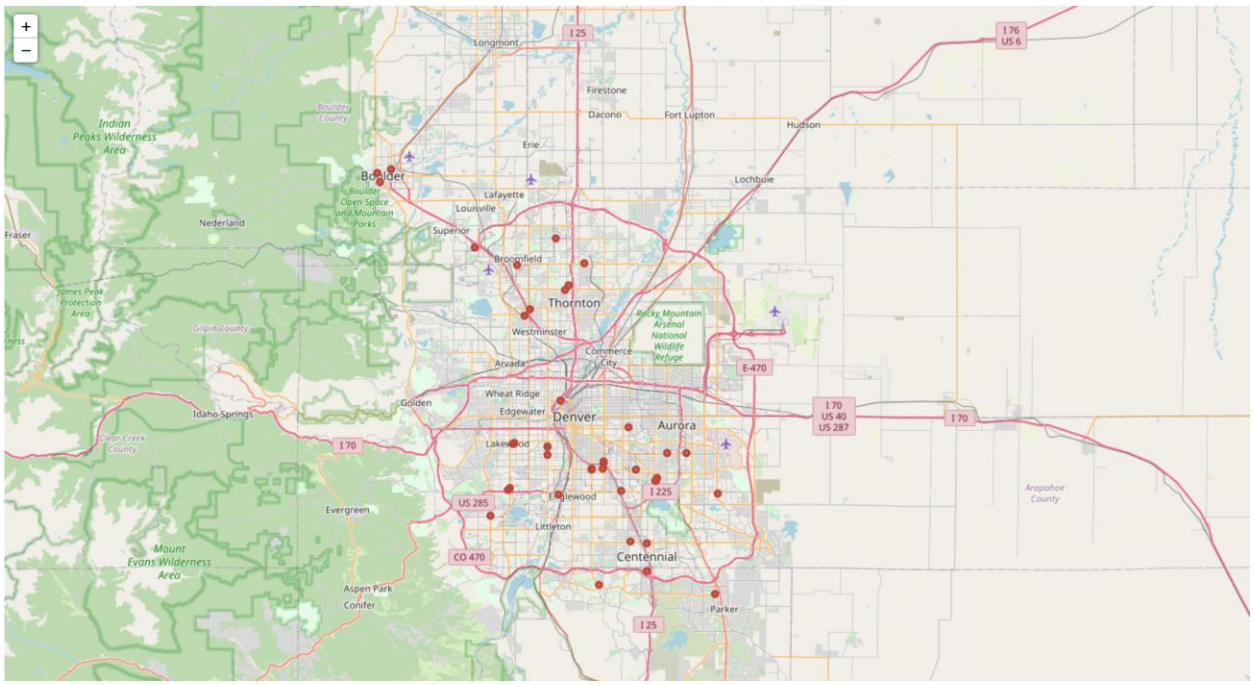
4. Analysis

4.1 Folium Map Visualization

Let's visualize how cities are located throughout the area. From the map, we can see Denver which is a capital, is located at the central area with other cities like Aurora, Arvada, Westminster, Thornton, Broomfield, Englewood, Littleton, and Centennial. So, Denver Metropolitan Area is not just an area close to Denver but also include farther areas down south, in some part of the mountains, and east of Denver.



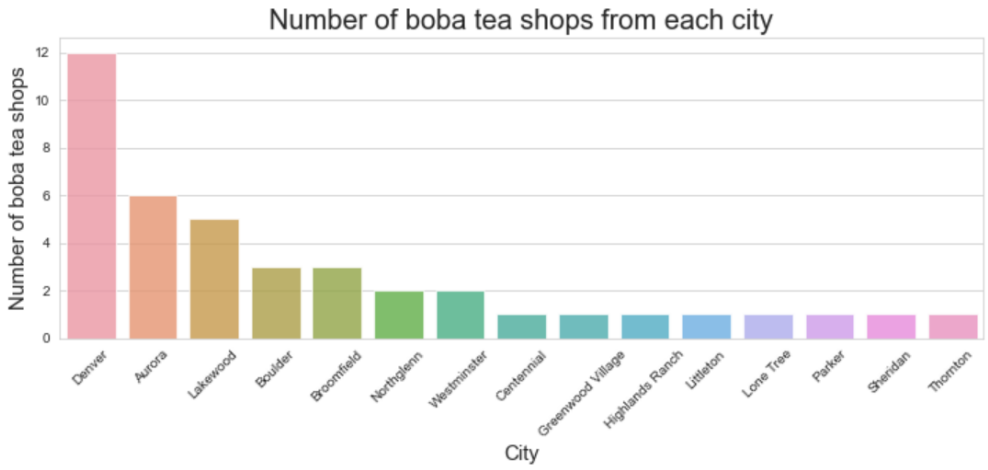
Now, let's visualize locations of boba tea shops in Denver Metropolitan Area. It is obvious that most boba tea shops are located in the south of the area. There are some in the north and interestingly, a few boba tea shops are in just 1 city like Boulder.



4.2 Exploratory Analysis

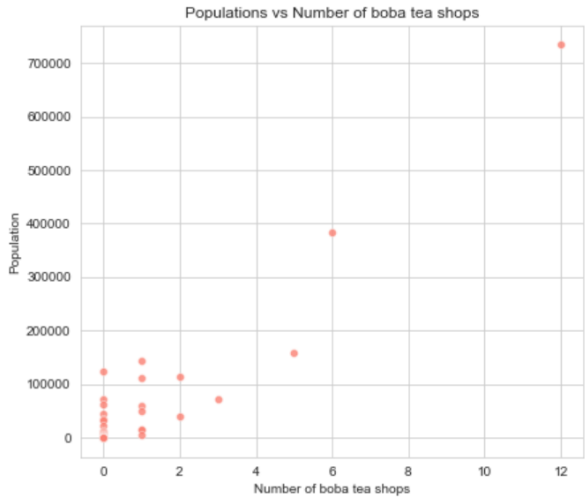
4.2.1 Number of boba tea shops from each city

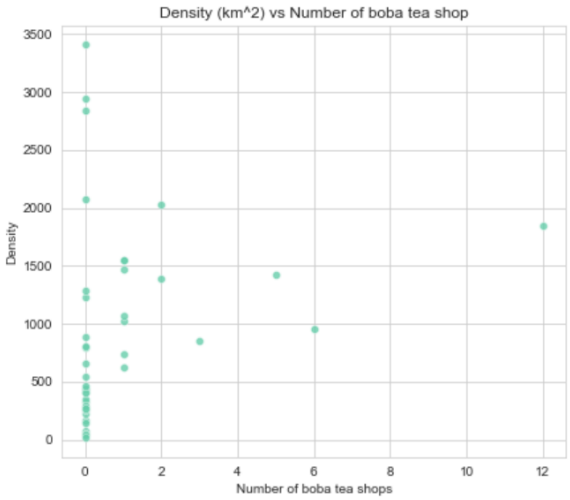
It is predictable that Denver has the biggest numbers of boba tea shops comparing to the other cities and it is confirmed by my data from Foursquare.



4.2.2 Relationship between population, density, and number of boba tea shops

My hypothesis is that the higher number of population that city has, the higher number of boba tea shops it has. As in general, shop owners would prefer to have the shop in a populated area than far away from people. Same as another hypothesis here that the higher density number that city has, the higher number of boba tea shops there is.



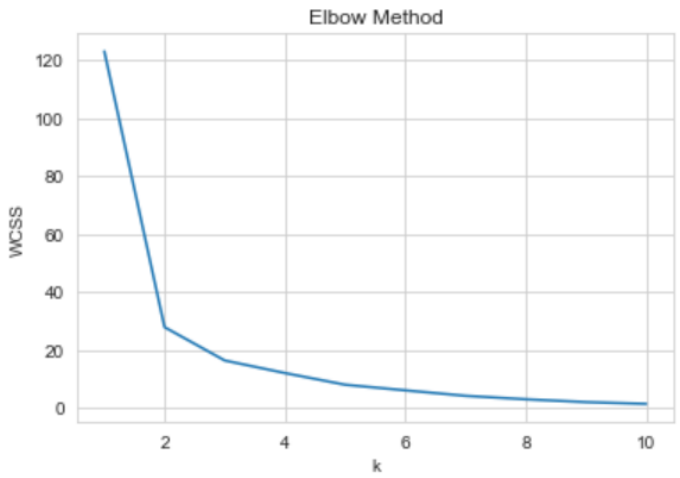


After visualizing the scatter plots, it is still not enough to conclude the relationship between 3 variables. So, I use correlation coefficients to measure how strong the relationships are in the following table. The table shows correlation coefficient 0.94 between number of boba tea shops and populations. Positive number close to 1 means there is positive relationship between population number and number of boba tea shop. On the other hand, it shows Correlation coefficient 0.25 between number of boba tea shops and density of population per square meter. So, there is no significant relationship between these two.

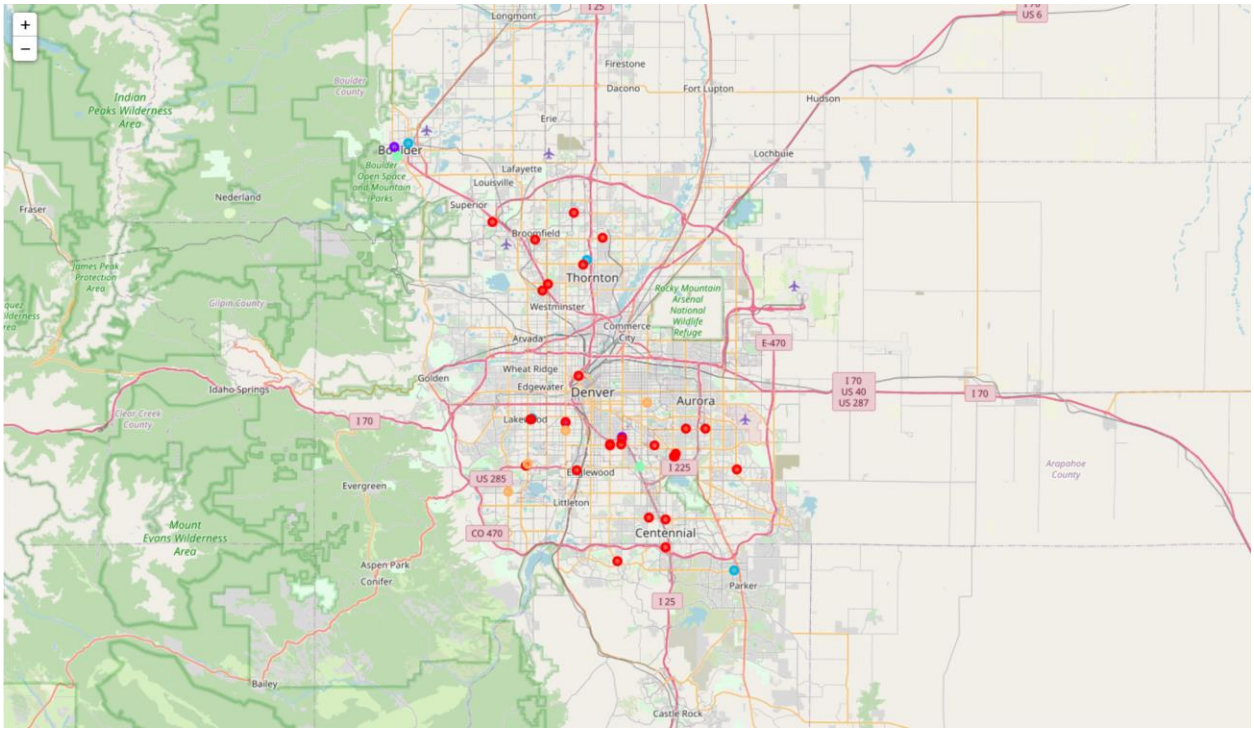
	pop2020	density	Number of boba tea shops
pop2020	1.000000	0.269480	0.939496
density	0.269480	1.000000	0.253816
Number of boba tea shops	0.939496	0.253816	1.000000

4.3 Clustering

I use an unsupervised machine learning technique like k-means clustering to partition boba tea shops we have in the market. Before I start with clustering, I calculate WCSS and use elbow method to choose the best k for this data set. From the following graph, I can see that after k = 5, WCSS is minimal. That means the optimal number of clusters is 5.

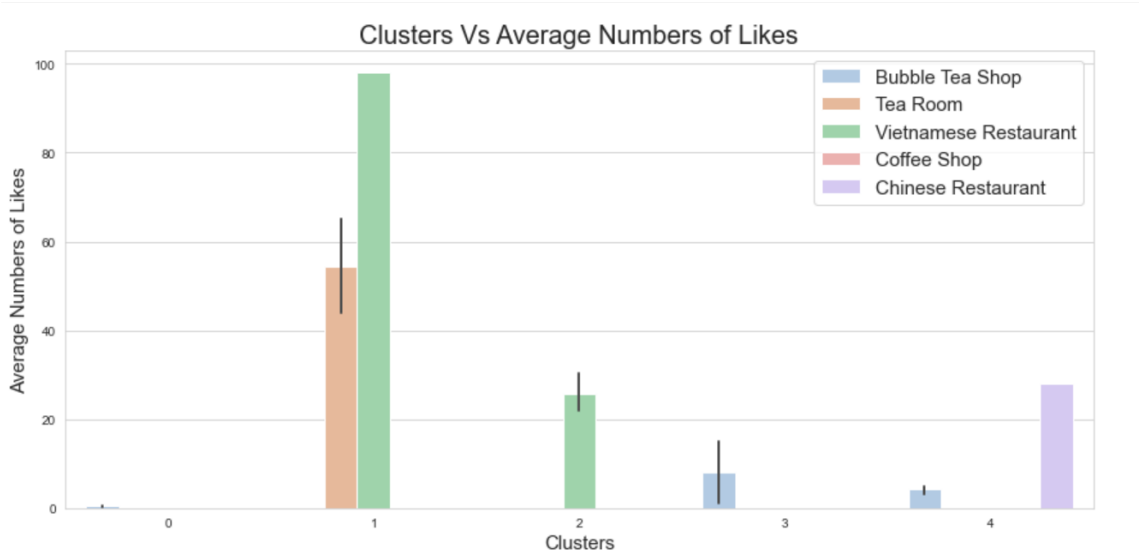


After I apply cluster to the data set, I use folium map again to plot the data. The majority of all clusters is cluster 0 in red color located in the all area. Cluster 1 is in purple which there are 2 shops located in Denver and 1 shop in Boulder. Cluster 2 is in teal located throughout the area. Cluster 3 is in light blue in Denver and Boulder. Cluster 4 is in orange which mostly located in the south area.



The following bar chart shows that cluster 1: tea rooms and a Vietnamese restaurant have the highest numbers of likes among other shops that sell boba teas, followed with a Chinese restaurant in cluster 4, Vietnamese restaurants in cluster 2, and few boba tea shops in cluster 0 only have a few likes.

It is obvious that boba tea shops who specialize and sell only boba teas have less numbers of likes. It is possible that restaurants are easier to gain customers. People need foods but at the same time, they might think that an expensive drink is not necessary.



The bar chart below shows that cluster 1: tea rooms and a Vietnamese restaurant also have the highest rating, followed by boba tea shops in cluster 3, Vietnamese restaurants in cluster 2, Chinese restaurants in cluster 4, and cluster 0 remains no rating.

5. Result and Discussion

The analysis shows that there are only 41 boba tea shops that exist throughout Denver Metropolitan Area in the radius of 40 miles or 65 kilometers from the central of Denver. Since boba teas are not so common in USA like coffee shops, it leads to niche market situation or we can take opportunity to introduce new kind of drink to customers who have not tried it before.

From plotting the map, most of the boba tea shops are located in the north and south part of Denver Metropolitan and Boulder area. The highest density of boba tea shops are in the city of Denver (12 shops), Aurora (6 shops), and Lakewood (5 shops), as we find out that these cities also have the highest populations in Denver Metropolitan area in 2020. Moreover, it has been confirmed with correlation 0.94 that there is significant association between the numbers of boba tea shops and the numbers of populations of each city.

If the investors are looking for unsaturated market, there are many areas that do not have a boba tea shop. For example, Arvada is located northeast of Denver and south of Westminster. The population number is 123,036 in 2020 which is about 18% comparing to Denver. This city will be a good opportunity without a competitor nearby. The following table will show the city population numbers and the numbers of boba tea shops.

	City	pop2020	density	Number of boba tea shops
0	Denver	734134	1849.0876	12
1	Aurora	382742	957.9039	6
2	Lakewood	158660	1423.0873	5
3	Broomfield	72783	851.4743	3
4	Westminster	113919	1392.5473	2
5	Northglenn	38724	2033.9624	2
6	Thornton	143890	1547.9067	1
7	Centennial	112151	1463.9517	1
8	Parker	58578	1018.9020	1
9	Littleton	50507	1544.5755	1
10	Greenwood Village	15749	736.3303	1
11	Lone Tree	15705	618.2243	1
12	Sheridan	6176	1074.6046	1
13	Arvada	123036	1224.8238	0
14	Castle Rock	70687	795.9262	0
15	Commerce City	61313	662.5854	0
16	Brighton	43998	802.1177	0
17	Englewood	35168	2071.8014	0
18	Wheat Ridge	31210	1289.9946	0
19	Golden	22270	891.8024	0
20	Federal Heights	13047	2836.7052	0
21	Castle Pines	10361	416.9618	0
22	Cherry Hills Village	6614	411.8640	0
23	Edgewater	5294	2941.4543	0
24	Glendale	5084	3405.9544	0

In the last part of analysis, all boba tea shops are grouped into 5 clusters by city, number of likes, rating, and category that we pull from Foursquare API. After we examine the features and characteristics of the clusters, it can be said as following:

- Cluster 0 Most Boba tea shops, small number of likes, located all over the area
- Cluster 1 Well known and high rating tea room and Vietnamese restaurant
- Cluster 2 Well known Vietnamese Restaurant, quite high rating
- Cluster 3 Popular boba tea shops "Kung Fu Tea"
- Cluster 4 Boba tea shops with some likes and a well-known Chinese Restaurant

There is an interesting thing we learn from the clusters. Vietnamese and Chinese restaurants are also big competitors to boba tea shops. When people see that these restaurants have a lot of likes and high ratings even just for their foods, the customers will come and the restaurants can have that advantage of selling foods and boba tea altogether. The same as tea shops, they sell all kinds of tea stuff and drink but those people who come for just teas might want to try their boba teas, also.

Another thing we notice is that cluster 3 contains only 1 brand of boba tea franchise which receive a lot numbers of likes and higher rating than any other boba tea shops in Denver Metropolitan area. This franchise of boba tea shops is located in Denver, the capital and economic hub of Colorado, and Boulder which has a large student population at University of Colorado.

6. Conclusions

In this study, I explore how many boba tea shops are located in each city and how they spread out on the Denver Metropolitan area map. We also analyze the relationship between population numbers and numbers of boba tea shops in the cities. We gather data for each boba tea shop as much as it is available on Foursquare and use clustering to partition boba tea shops into groups that all members have similar characteristics such as city, numbers of likes, rating, and categories. This helps the investors to identify what kind of competitors are out there in the market and give them an idea of where their boba tea shop will be. After all, the decision for the best boba tea shop location should be made with not only numbers of populations in the area, but also rent/property cost, taxes, main roads, traffic of the area and etc.