

Location Recommendation for Jadeite Jade Jewelry Store in St. Louis County, MO, United States

Aaron L. Warsaw

July 20, 2019

1. Introduction

1.1 Background

Jadeite is one of two main gemstones, the other being nephrite, that are correctly called jade by the international gem and jewelry industry. High quality jadeite is one of the most expensive gemstones in the world, rivaling that of even diamond. Jade is also one of the most culturally significant gemstones in the world with thousands of years of historical significance in China, areas of North America, Central America, and New Zealand(<https://www.gia.edu/jade-history-lore>). When most people think of jade they think of green jadeite, popular in China and other Asian countries. However, jade comes in other colors as well including: lavender, yellow, blue, black, and white. My favorite color happens to be green, I am fascinated with minerals and gems, and I am thinking about a changing jobs to selling high quality jadeite online and possibly at a physical store location someday. So as the shareholder and the data scientist for this project, the questions I am interested in researching is: where is a good location to open a jewelry store specializing in jadeite, and is there is even a large enough market for a physical store location in the urban area of St. Louis, Missouri, United States of America?

1.2 Problem

Knowing where to locate my store is important because it needs to be in an area close to a customer base that would purchase the product, close to an area where income is high enough to purchase the product, an area with low crime rate, etc. Although westerners are beginning to appreciate jadeite, with such a long history in Asia, and specifically China, the predominate market for high quality jadeite will consist of people of Asian race and ethnicity. Thus, I plan to leverage Foursquare location data for the municipalities (each zip code) in St. Louis County and cluster the different zip codes using K-means to do some preliminary research to see if there is an area with a high number of Asian or Chinese venues (i.e. restaurants, tea houses, cultural centers, Asian markets, etc.). If such an area or cluster can be found then that will give me a good idea of for future investigation. If not, then further research will be needed and it may also indicate that there is not a large enough concentration of people of Asian race and ethnicity to support a physical store. I also plan to use demographic information from (<https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#data-map/tract>) and other maps and data they provide to look at the demographics and areas of Asian race and ethnicity in the areas of St. Louis County. Hopefully, this data will support the results from my Foursquare and clustering analysis. Other

data to consider in clusters are locations with high end shopping and existing jewelry stores.

2. Data Aquisition

I plan to use the same type of analysis that we did for our peer-reviewed assignment in week three of this class. All cities and zip codes in St. Louis County, MO were obtained at(<https://www.zip-codes.com/county/mo-saint-louis.asp>) by copying and pasting into a Microsoft Excel file. All zip codes and their corresponding latitude and longitude coordinates for the United States were downloaded from (www.pier2pier.com/links/files/Countrystate/USA-Zip.xls). Maps and figures from: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#data-map/tract> were captured as screenshots and pasted into this report.

3. Methodology

3.1 Data Pre-processing

Libraries needed for the analysis were installed in a Jupyter notebook. The Excel file containing zip codes, city names, and county was read into the Jupyter notebook and converted into a Pandas data frame. The column name of zip codes was changed to "Postcode" to match the column name of "Postcode" in the data set of all United States postcodes. The Excel file containing all United States postcodes was read into the Jupyter notebook as a csv file and converted into a data frame. Next, a new data frame was made containing only the postcodes for the state of Missouri. The data frame with all postcodes for St. Louis County, Missouri and the data frame with all the postcodes and latitude and longitude coordinates of the state of Missouri were merged on "Postcode" resulting in one data frame with the latitude and longitude coordinates for each postcode in St. Louis County, Missouri. This is the area I want to locate my store and includes the city of St. Louis and surrounding cities or municipalities. The "Postcode" column was converted to type "string" for future analysis.

Geocoder was then used to get the latitude and longitude coordinates of the State of Missouri with output reading: "The geographical coordinates of the state of Missouri, US are 38.7604815, -92.5617875".

Features needed for mapping, segmenting, and clustering were installed so data analysis using Foursquare location data and K-means clustering could begin.

3.2 Mapping and Segmenting.

Folium was used to create a map showing all the postcodes in St. Louis County(Figure 1, below).

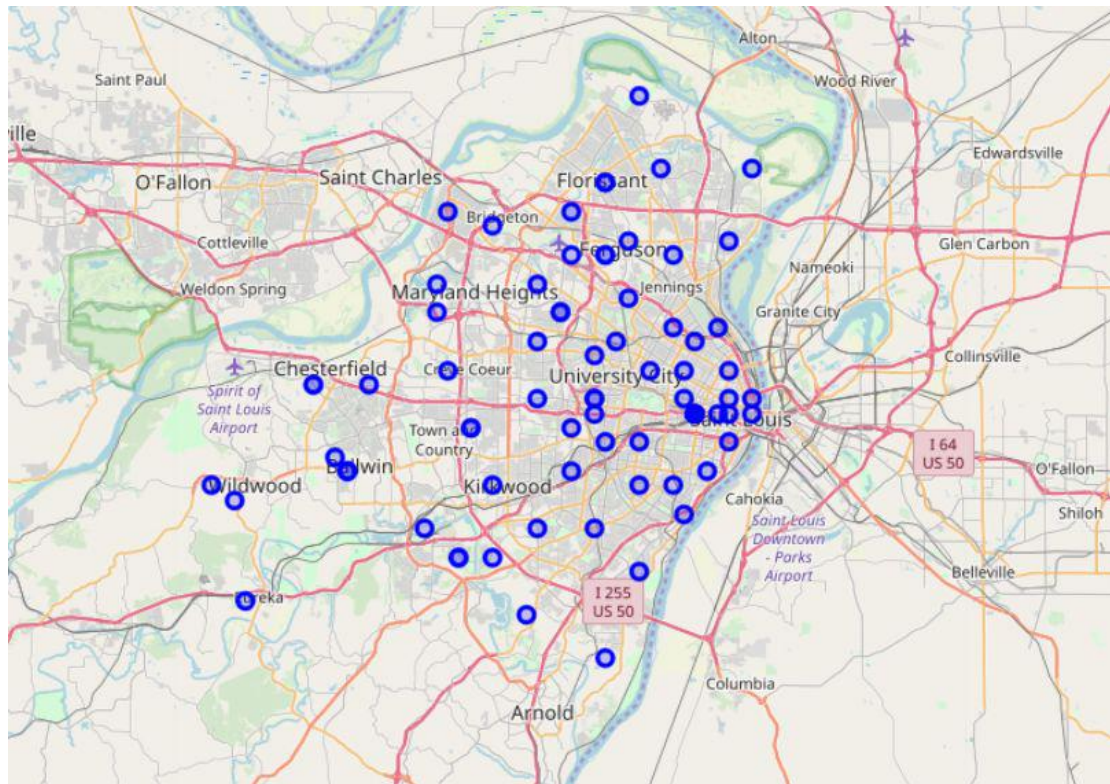


Figure 1. Postcodes in St. Louis County, MO, United States.

Foursquare version and credentials were then defined for searching for venue information and clustering. Following that, the GET request URL for getting all venues in St. Louis County was developed, ran, and put in a new data frame: MO_venues. Means of each frequency of venue per postcode was calculated and then top five types of venues in each postcode were displayed for insights into venue types per postcode before clustering. A new data frame was constructed to show the top 10 most common types of venues in each postcode. The top 10 data frame will be combined with the clustering data frame following clustering.

3.3 Clustering

I used the Elbow method to construct a graph to determine the optimal number of clusters (K) to use. Based on Elbow method results, all postcodes of St. Louis County were clustered into three clusters using unsupervised learning K-means algorithm for clustering. When data frames showing clusters and top 10 most common type of venues were merged on the column name "Postcode", the resulting new data frame showed that four postcodes returned NAN values for cluster number and venues(figure 2, below). These four rows were deleted from the data frame. Results will be discussed later in the report.

Postcode	Municipality	County	Neighbourhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	
50	63127	Sappington	St. Louis County	Saint Louis	MO	38.53	-90.42	NaN	NaN
51	63128	Sappington	St. Louis County	Saint Louis	MO	38.49	-90.39	NaN	NaN
52	63129	St. Louis	St. Louis County	Saint Louis	MO	38.46	-90.32	0.0	Italian Restaurant
53	63130	University City	St. Louis County	Saint Louis	MO	38.67	-90.33	0.0	Park
54	63131	Des Peres	St. Louis County	Saint Louis	MO	38.62	-90.44	1.0	Lounge
55	63131	Frontenac	St. Louis County	Saint Louis	MO	38.62	-90.44	1.0	Lounge
56	63132	Olivette	St. Louis County	Saint Louis	MO	38.68	-90.38	NaN	NaN
70	63146	St. Louis	St. Louis County	Saint Louis	MO	38.70	-90.47	NaN	NaN

Figure 2. Selected cells from data frame showing Postcodes 63127, 63128, 63132, and 63146 with NAN values for cluster number and venue data. These postcodes were deleted from further clustering analysis.

Each cluster was then displayed showing each postcode in the cluster and the cluster's corresponding top 10 most common venues.

3.4 Demographic Statistics.

Demographic maps and statistics from (<https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#data-map/tract>) were observationally evaluated to learn about Asian race and ethnicity distribution in postcodes in St. Louis County(eg. Figure 4, below). Because the majority of potential clients for jadeite jade will most likely be comprised of people of Asian race and ethnicity this data will hopefully support the results of the clustering analysis.

Map of Race and Ethnicity by County Subdivision in St. Louis County

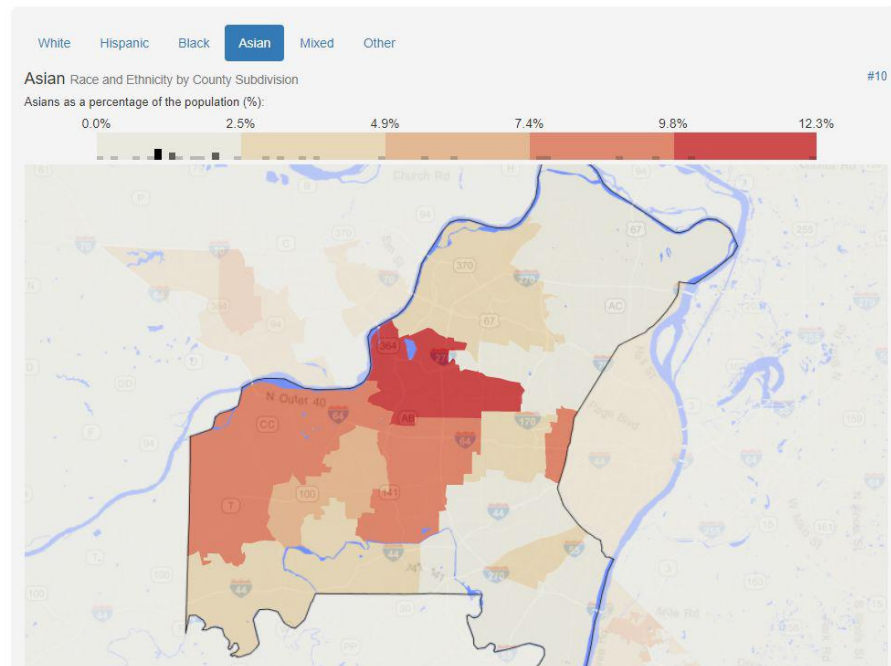


Figure 4. Screenshot of map showing Asian Race and Ethnicity as a percentage of population by County Subdivision in St. Louis County(map from: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#overview>, Accessed on 7/20/2019).

4. Results

4.1 St. Louis County, MO Postcodes.

There were 85 different post codes for the area of St. Louis County, which includes the city of St. Louis, MO, United States. Four postcodes 63127, 63128, 63132, and 63146 corresponding to the municipalities of Sappington, Sappington, Olivette, St. Louis returned NaN values following a Foursquare search for venues and K-means clustering did not return a cluster number for these postcodes. These four postcodes were dropped from the analysis. The reason for no venues being returned is not known, but could be due to the 500 meter radius for venues in the FourSquare API search needing to be expanded. An 800 meter radius search was ran but these four postcodes still failed to return any venues from the FourSquare search. Data from the 500 meter search was used in the analysis.

4.2 Foursquare API St. Louis County, MO Venues.

For my analysis the limit for number of venues returned per postcode was set to = 100. Venues were returned within a radius of 500 meters from the latitude and longitude coordinates for each postcode. For all venues returned from the Foursquare search there were 167 unique categories of venues. Returned categories of venues of interest to this project for finding a jadeite jade jewelry store location include: Accessories Store, **Asian Restaurant**, Bus Station, **Chinese Restaurant**, College Administrative Building, College Cafeteria, College Theatre,

Cosmetics shop, Food, Gourmet Shop, **Health & Beauty Service**, **Japanese Restaurant**, **Jewelry Store**, Light Rail Station, **Sushi Restaurant**, **Tea Room**, **Thai Restaurant**, and Yoga Studio. Venues in bold font are what I really want to see appear in my K-means Clustering analysis as their presence could indicate a high demographic of Asian race and ethnicity and thus identifying a potentially large market for jadeite jade and a possible location for my jewelry store. When looking at the top five types of venues for each postcode, postcodes 63114, 63116, 63144, and 63145 return venues of interest. Output is shown in figure 5 below.

----63114----			----63116----		
	venue	freq		venue	freq
0	Plaza	0.1	0	Zoo	0.25
1	Discount Store	0.1	1	Arcade	0.25
2	Park	0.1	2	Jewelry Store	0.25
3	Mexican Restaurant	0.1	3	Bakery	0.25
4	Chinese Restaurant	0.1	4	Pet Café	0.00

----63144----			----63145----		
	venue	freq		venue	freq
0	Spa	0.1	0	Brewery	0.14
1	Chinese Restaurant	0.1	1	Gym	0.14
2	Sports Bar	0.1	2	Japanese Restaurant	0.14
3	Rental Car Location	0.1	3	Thrift / Vintage Store	0.14
4	Italian Restaurant	0.1	4	Lounge	0.14

Figure 5. Selected postcodes of St. Louis County, MO, United States showing venues of interest(Chinese Restaurant, Jewelry Store, Spa, and Japanese Restaurant) for potential jadeite jade jewelry store location.

4.3 K-means clustering by postcode of St. Louis County.

All postcodes of St. Louis County were clustered into three clusters using unsupervised learning K-means algorithm for clustering. I used the Elbow method to make a graph to determine the optimum number of K's to use(Figure 6).

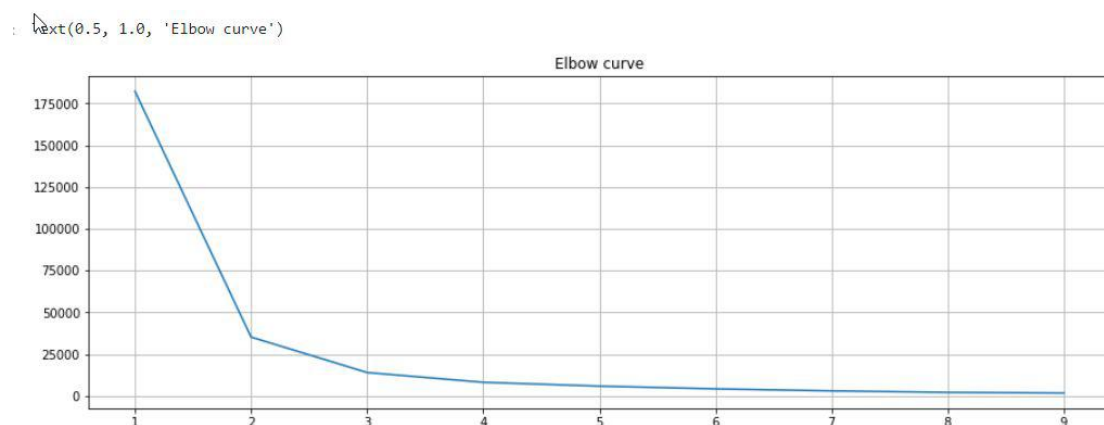


Figure 6. Elbow curve of showing the optimal K number. X axis is number of K's and Y axis is distortion.

The optimal K from the elbow curve in figure 6 is two or three. I would like to have see more than two clusters so I am going to run K-means with three clusters. A Folium map of the clusters was generated and is presented below(Figure 7.)

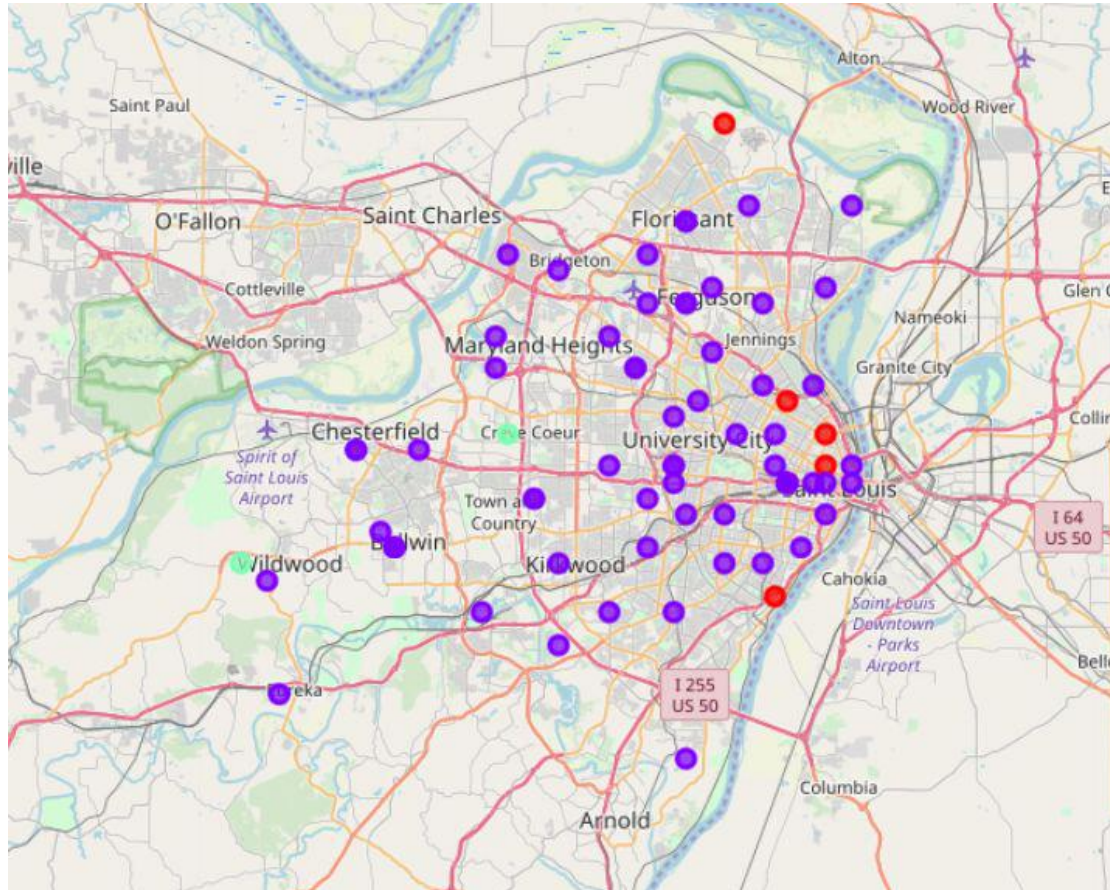


Figure 7. Results from running K-means on all post codes in St. Louis County, MO, United States. Three clusters were ran. Please note light green cluster colors by Creve Coeur and Wildwood, MO.

The first cluster(red color) was comprised five post codes having of a variety of venue types including: gas stations, disc golf, bars, and farm space. Locations were on the East side of St. Louis County. The second cluster(purple color) was comprised of 72 postcodes including a large variety of venue types with postcode locations that were not in one localized area within St. Louis County. The third cluster(light green color) was comprised of two postcodes toward the western part of St. Louis county and share the same top 10 venue types including: scenic overlook, disc golf, fast food restaurant, and farm space. The variety of venues within the clusters was so varied that I am unable to give them a name that would adequately describe the contents or location of the cluster.

5. Discussion

Unfortunately, none of the three clusters resulted in an area where the majority of post codes had many venue types I was looking for with regards to large number or localized area of Asian race and ethnic venues.

I ran a FourSquare API search for venues with a radius of 800 m to see if the postcodes (63127, 63128, 63132, and 63146 corresponding to the municipalities of Sappington, Sappington, Olivette, St. Louis) that returned NaN values under the search radius of 500 m would return venues with the larger search radius. Unfortunately, the larger radius did not result in any venues being returned. I am unable to explain this part of the output.

With the results from the K-means clustering not providing much insight into where to locate the jewelry store, I looked at the four postcodes listed in Figure 5. These postcodes had venues of high interest to me among the top five venues of each post code that were returned from the FourSquare API search with radius of 500 meters. These postcodes are 63114, 63116, 63144, and 63145 and correspond to the municipalities of Breckenridge Hill, St. Louis, Brentwood, and Lambert Airport, respectively. Figure 8 shows selected zip codes(postcodes) in St. Louis County excluding those of St. Louis city. Zip codes 63144(Brentwood) and 63114(Breckenridge Hill) are of special interest to me because they contain venues in the top five venues for each zip code that could indicate a high percentage of people of Asian race and ethnicity(refer to Figure 5 for top five venue types).

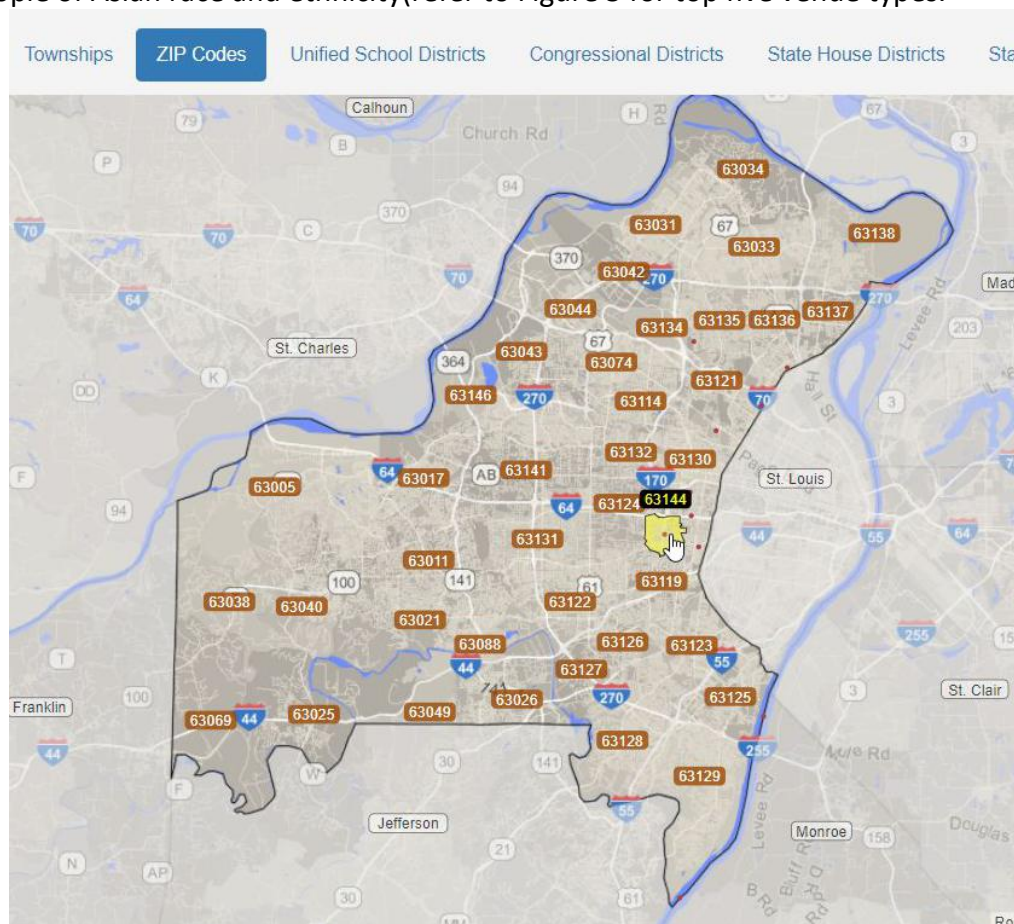


Figure 8. Selected zip codes(postcodes) of St. Louis County excluding zip codes of the city of St. Louis(From: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Ancestry>, Accessed July 22, 2019). Zip codes with venues of interest are 63144(boundary marked in yellow) and 63114(located above zip code 63144).

For the first postcode of interest above, 63144(Brentwood) has Spa and Chinese Restaurant as number one and two, respectively in top five venues as shown in Figure 5. However, in Figure 9 below, Brentwood is number nine on the list in for Asian population as a percentage of total population for places in St. Louis County. The second postcode of interest above 63114(Breckinridge Hill) does not appear on the list in Figure 9, despite having the venue of Chinese Restaurant as the number five spot among the top five venue types as shown in Figure 5. The places with the top five highest percentage of Asian populations are Maryland Hts, Creve Coeur, Clayton, Olivette, and Chesterfield.

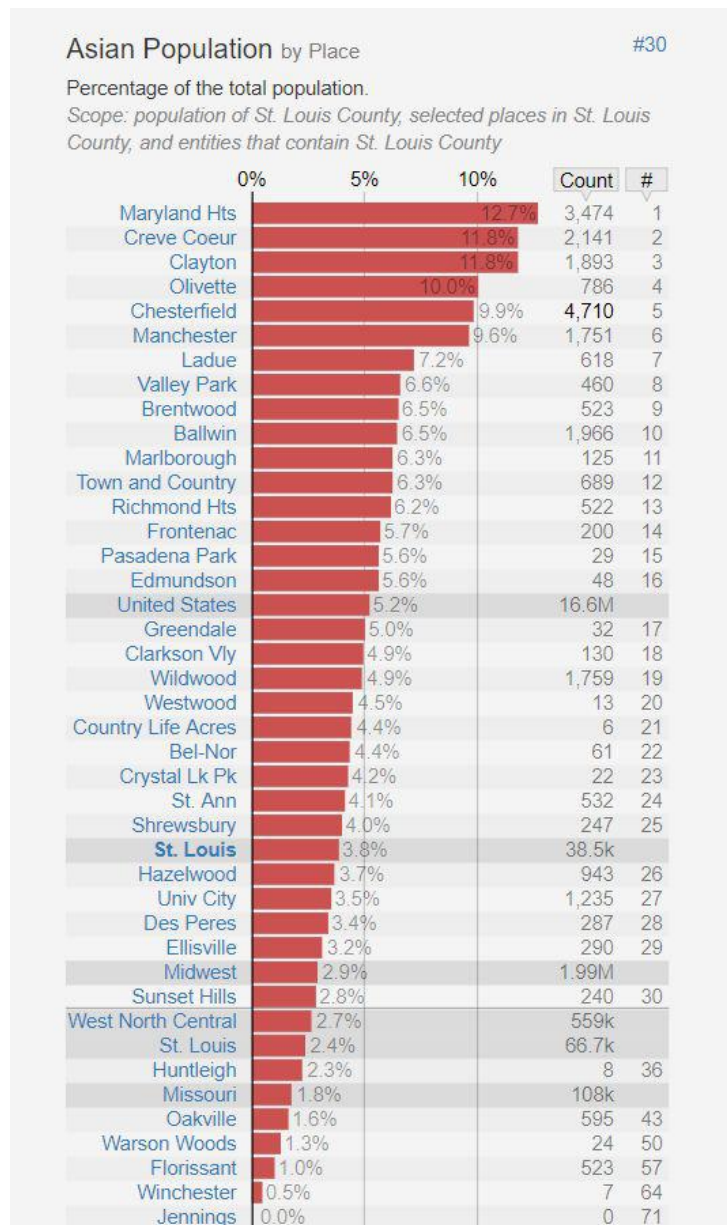


Figure 9. Asian population as a percentage of total population by place in St. Louis County(From: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity>, Accessed July 23, 2019).

The map presented below shows the percentage of Asian race and ethnicity by St. Louis County Subdivisions(also known as Townships; Figure 10.)

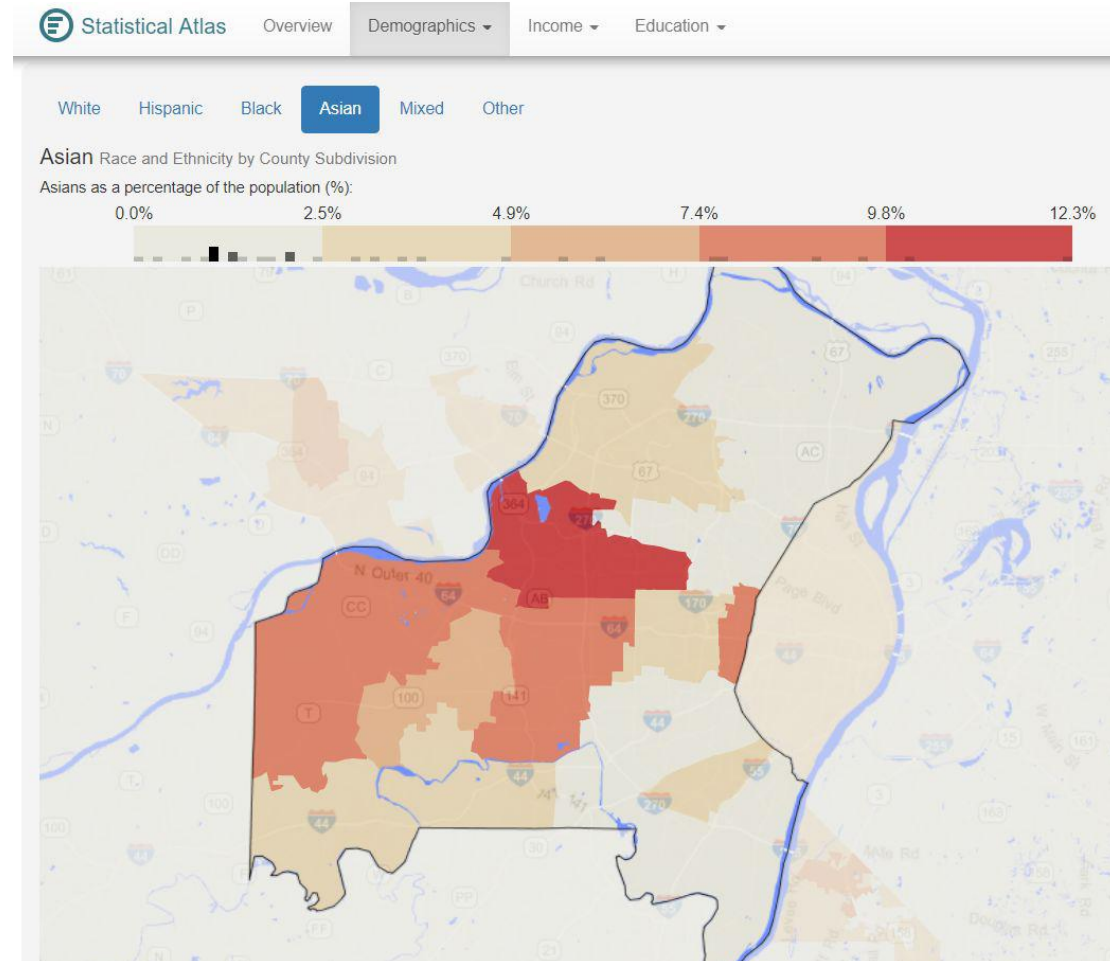


Figure 10. Map showing Asian race and ethnicity of St. Louis County Subdivisions(Townships) as a percentage of the population(From: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity>, Accessed July 23, 2019).

A map like the one shown in Figure 10 divided by postcodes in St. Louis County was not available. Figure 10 shows that the Subdivisions in the 7.4% to 12.3% range border each other with one exception to the east. Subdivisions bordering each other are Maryland Heights, Creve Coeur, Chesterfield, Missouri River, and Queeny Townships. Figure 10 shows the Maryland Heights Township in yellow and the hand and finger pointer at the southern area of the Township points out an area to investigate further for a possible location for my Jadeite jade jewelry shop because it is in the central geographic area in St. Louis County with the highest Asian population as a percentage of the population.

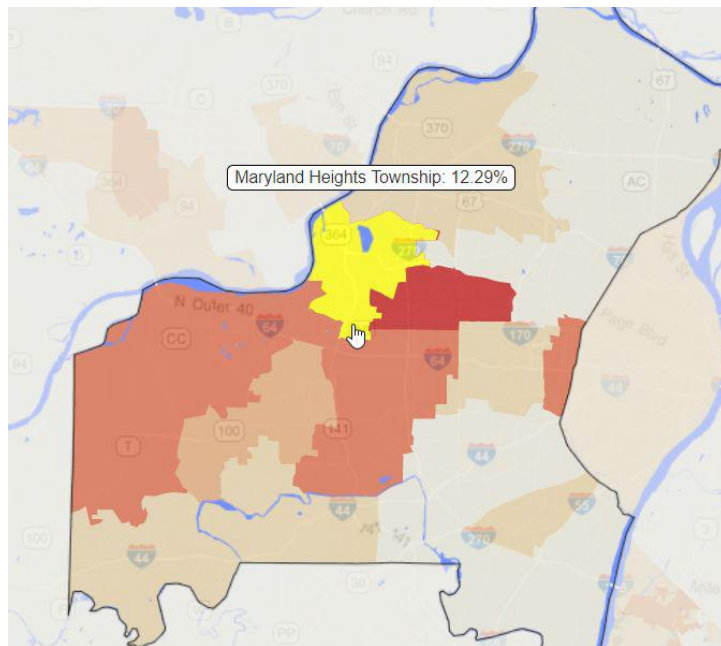


Figure 11. Map of Maryland Heights Township (yellow area) with Asian race and ethnicity of 12.29% of the total population(From: <https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity>, Accessed July 23, 2019).

6. Conclusion

From this analysis using FourSquare API venue data, clustering using K-means, and Asian race and ethnicity data from online sources I was unable to confidently recommend a location for a jadeite jade jewelry store in St. Louis County, MO, United States. However, this analysis was a good starting point because valuable insights into the areas of high Asian demographics within St. Louis County were obtained from observational data from an online source. It should be noted however, the race and ethnicity part of the analysis for St. Louis County did not include the city of St. Louis itself. As a result further data analysis of race and ethnicity demographics of Saint Louis is needed. As a resident of the area, I know that further research is needed to identify venues for the area around Olivette, which did not return any venues in the Foursquare search. Olivette is next to University City, which has a strip along Olive Road that is the closest area to a China Town that the St. Louis Area has. With many Asian stores and restaurants in this area, I am curious why University City did not return more Asian related venues from the FourSquare API search. Future research will also focus on splitting up postcodes in the city of Saint Louis from postcodes in the rest of the county. Postcodes in Saint Louis City returned many of the same venues, which lead to a large cluster when using K = 2, 3, 5, and 7, which I did for exploratory purposes. This is likely due to the density of people and businesses in Saint Louis City. I hypothesize that clustering with a smaller radius within the city of Saint Louis may result in a more diverse set of venues for each postcode within Saint Louis City. Another method may be to cluster based on city instead of post code. For postcodes in Saint Louis County but outside of St. Louis City, clustering on a radius larger than 800 meters may lead to more venues being returned in the Foursquare API search including venues that originally returned no venues. Finally, other

characteristics within Saint Louis County including: age, household income, crime rate, high end shopping, and real estate prices need to be analyzed before a complete story can be told and a location for my jadeite jade jewelry store can be confidently recommended.