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

Aaron L. Warsaw

July 20, 2019

1. **Introduction**
   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 career change 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. **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, etc. Although westerners are beginning to appreciate jadeite, with such a long history is 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 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. This should support the results from my Foursquare and clustering analysis. Other data to consider are locations with high end shopping, existing jewelry stores that have been established and successful for a long duration, crime rate, types of industry in the area, income, etc. These additional data types will be looked at in future research.

1. **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).](http://www.pier2pier.com/links/files/Countrystate/USA-Zip.xls)

1. **Methodology**

**3.1 Data Pre-processing**

Libraries needed for the analysis were installed in a Jupyter notebook.

Excel file containing zip codes, city name, and county were uploaded 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 dataset of all United States postcodes. The xls file downloaded from the link listed above was read into the Jupyter notebook as a csv file. Next, a new dataframe was made containing only the postcodes for the state of Missouri. Now, I have a data frame with all the postcodes for St. Louis County, Missouri and a data frame with all the postcodes and latitude and longitude coordinates of the state of Missouri. These two dataframes were merged on “Postcode” so I now have the latitude and longitude coordinates for each postcode in St. Louis County, Missouri. This is the area I want to locate my store in and includes the city of St. Louis and surrounding cities or municipalities. The “Postcode” column was converted to type “string” for future analysis.

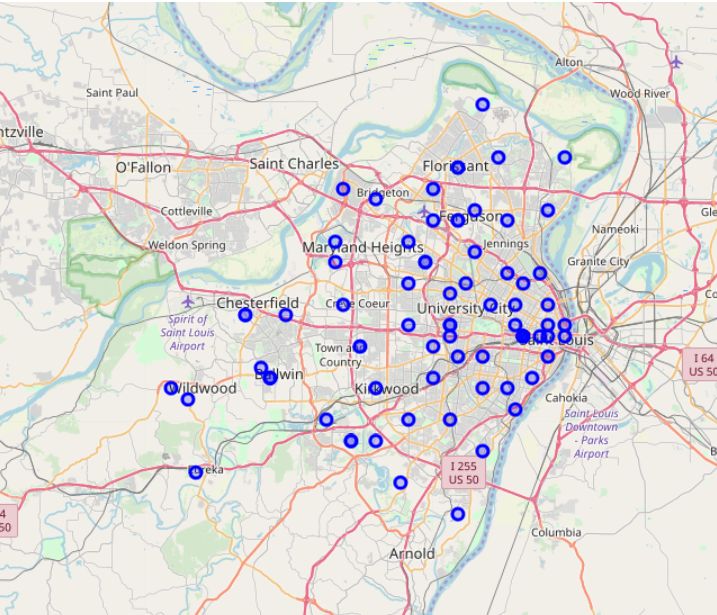
Geocoder was then used to get the latitude and longitude coordinates of the State of Missouri with output reading: “The geographical coordinates of the state of Missouri, US are 38.7604815, -92.5617875”.

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

**2.2 Mapping and Segmenting.**

Folium was used tocreate 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 getting 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. This will be combined with the cluster number following clustering.

**2.3 Clustering**

K-means clustering was ran to divide the postcodes of St. Louis County into 5 clusters. 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.







**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.

**2.4 Demographic Statistics.**

Demographic maps and statistics from (https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#data-map/tract) were looked at to learn about Asian race and ethnicity distribution of 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.

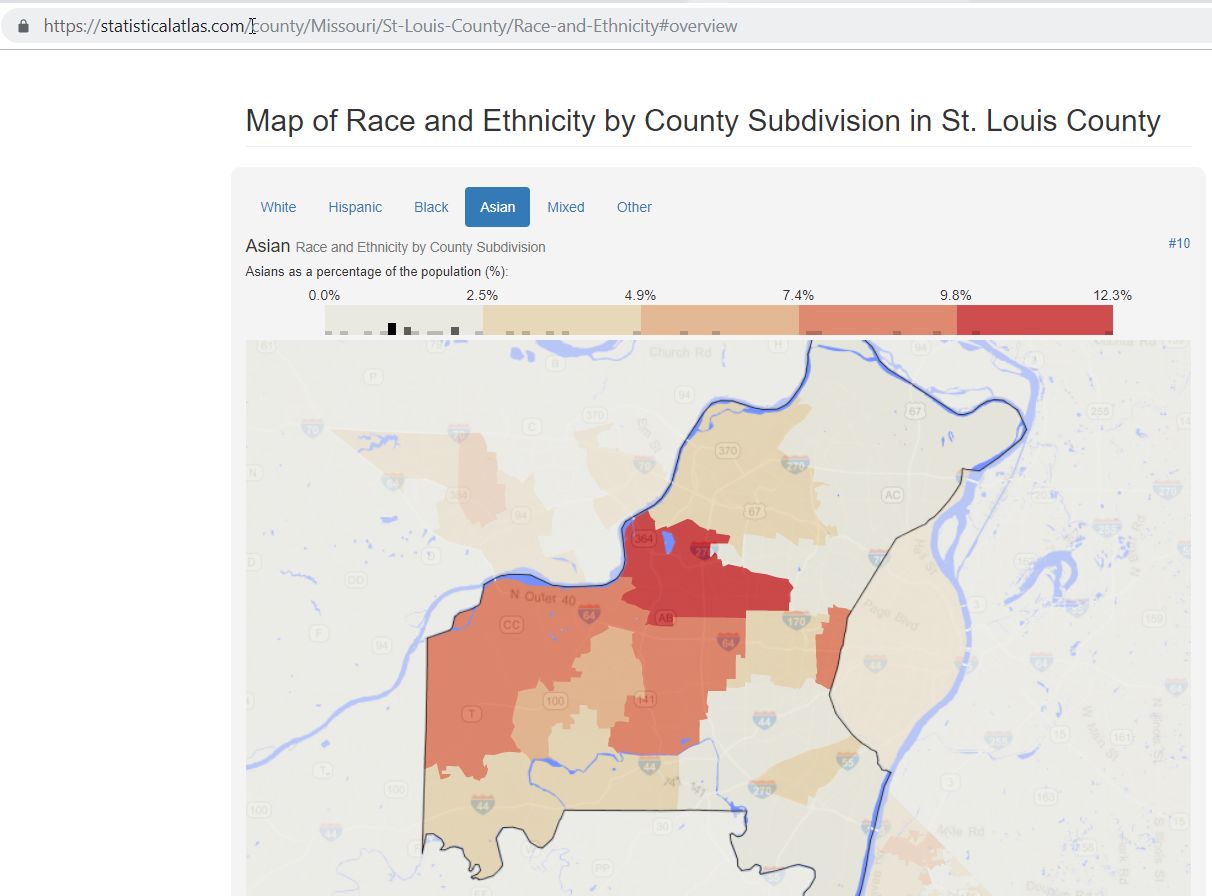


Figure 4. Screenshot of map showing Asian Race and Ethnicity as a percentage of population by County Subdivision in St. Louis County(map and data from: [https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity#overview](https://statisticalatlas.com/county/Missouri/St-Louis-County/Race-and-Ethnicity"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