# **Recommendation of Location for establishing New Jewelry Shop in Bangalore**

# Kishore Budumuru

# **12th July 2020**

1. Introduction

1.1 Background

India has an insatiable appetite to buy gold/new jewelry. Every year tonnes of gold is imported by India to meet the demand of consumers. The bulk of these transactions happen in capital cities like Bangalore. For buying gold, the customers still prefer the brick and mortar shops and do not prefer the ‘online mode’ The reason being the ‘feel and touch’ offered by a physical presence. More often, the customers tend to check in couple of adjacent shops before deciding on such a major expense. So, if a new jewelry shop is to be opened in a vast city like Bangalore where should that be?

1.2 Problem

Inspite of the demand available for jewelry, a bad choice of location for establishing a new shop will dent the profits of the franchises selling gold due to lack of customer footfall, high rents etc. Any franchise desires to do the due diligence before investing in a new location. If we put a new jewelry shop in an already ‘happening’ place, then the franchise is risking ‘extreme competion’ for the same pie of customers. And similarly if the location choses is too far, then customers who desire to check on a couple of shops before deciding to buy may actually never turn-up at such location! The solution lies in striking an optimum balance.

1.3 Interest

Multiple ‘jewelry selling franchises’ like Tanishq, Bhima, Geetanjali would be scouting for new locations to capture the market. To gain the first mover advantage, sophistical tools need to be used to gather accurate information to take informed decision. The current capstone project offers an insight on location choice. The marketing or consulting agencies tasked with such goal would surely be interested in a tool that can analyze neighborhoods quickly and in a visual way. Significant time and resources can be saved by following the techniques learnt in this capstone project (viz. FOURSQUARE API, geopy module application) to map an entire area as large as Bangalore in a matter of hours. So, essentially whoever is trying to putting up a new brick and mortar shop in Bangalore at posh areas will be interested. And they are jewelry selling merchants in India. They will be interested. They can already be having a presence within Bangalore (as with Bhima, Kalyan Jewelers etc) or they are present in rest of India and are attempting to establish a foothold in Bangalore. Example would be Mehrasons Jeweler’s, Khanna Jeweler’s etc.

2. Data acquisition and cleaning

2.1 Data sources

The information required is not directly available and one needs to do web scraping and connect the dots to arrive at the consolidated information.

As our area of interest is the city of Bangalore in India, the first data source needed is the area pin codes within Bangalore. This can be easily obtained from below link:

<https://indiamapia.com/Bangalore.html>

The data will contain the area names and pincodes of Bangalore in a tabular format. We will need geopy package to gather the latitude and longitude of the various pin codes of Bangalore, India. We then further use foursquare API to explore a given pin code (by passing on latitude and longitude information) and gather information on existing venues like jewelry shop, showrooms, banks, malls nearby. We can even search for the presence of a specific brand of jewelers in a particular neighborhood.

2.2 Data Cleaning

While scaping through the above mention weblink, several duplicate entries were noted. This is due to some adjacent areas sharing the same pincode. Using the data cleaning techniques like dataframe.drop\_duplicates command options, the same was fixed.

Missing data entries were also noted and evident through presence of NaN in the data set. By using dataframe.dropna command this can be easily taken care of. The data was also checked for outliers by doing sanity checks on max, min values in a column and whether the areas identified indeed belong to the region in focus. By using foliomaps, the reported latitude and longitude of pincodes is visually verified to be indeed present within the boundary map of Bangalore.

2.3 Feature Selection

Obviously, foursquare API search query for ‘Jewelry Shop’ in a neighborhood is an obvious choice. Considering the fact that most jewelry selling shops tend to be in upscale localities, other thoughtful features could be presence of number of car/bike showrooms in that locality, presence of banks, ice cream shops etc all indicative features of upscale localities.

Using foursquare API calls, all the above information collated against each PIN code and we now have the data set required for further analysis. For a total of ~100 pincodes, with 11 features extraction searches, a total of 1100 API calls were made through a for loop. The information thus obtained is organized in a dataframe for further analysis.

3. Methodology Section

* 1. Exploratory data analysis

After Importing the necessary libraries, followed the below steps

Step1: By webscraping, Get Bangalore pincodes from <https://indiamapia.com/Bangalore.html>

#go to the weblink in crome or explorer and right click on any particular table's cell and click on 'inspect'

# Parse the html content and use BeautifulSoup package to get Area names and corresponding postal codes of Bangalore from above weblink.

Step2: For FOURSQUARE search query to work, we need to provide the latitude and longitude of the area of interest. So, we need to get longitude and latitude from PIN Codes. This can be obtained from the geopy packages (pgeocode) installation.

Clean the data by watching out for duplicates, get rid of rows with NaN.

Step3: Search in Foursquare for neighborhood for a given pincode (by passing on latitude and longitude information) for 'Jewelry', other items of interest as discussed in earlier sections.

Read the response json files as an outcome of the API calls. Organize data into a dataframe.

# Step4: Join earlier data frame (postal code and latitude and longitude) and the foursquare output of count of ‘number of venues’ into a single data frame.

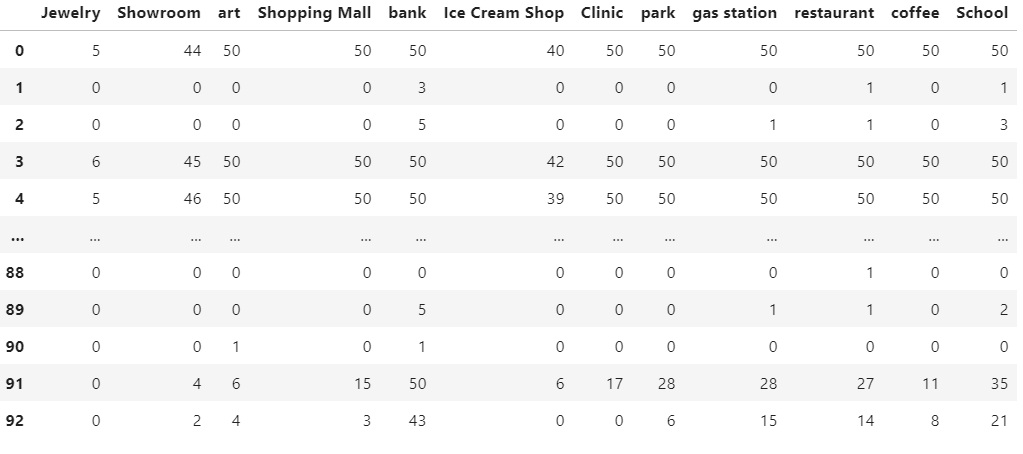
Step5: Cluster Neighborhoods

Run k-means to cluster the neighborhood into 5 clusters

Step:6 visualize in folium map

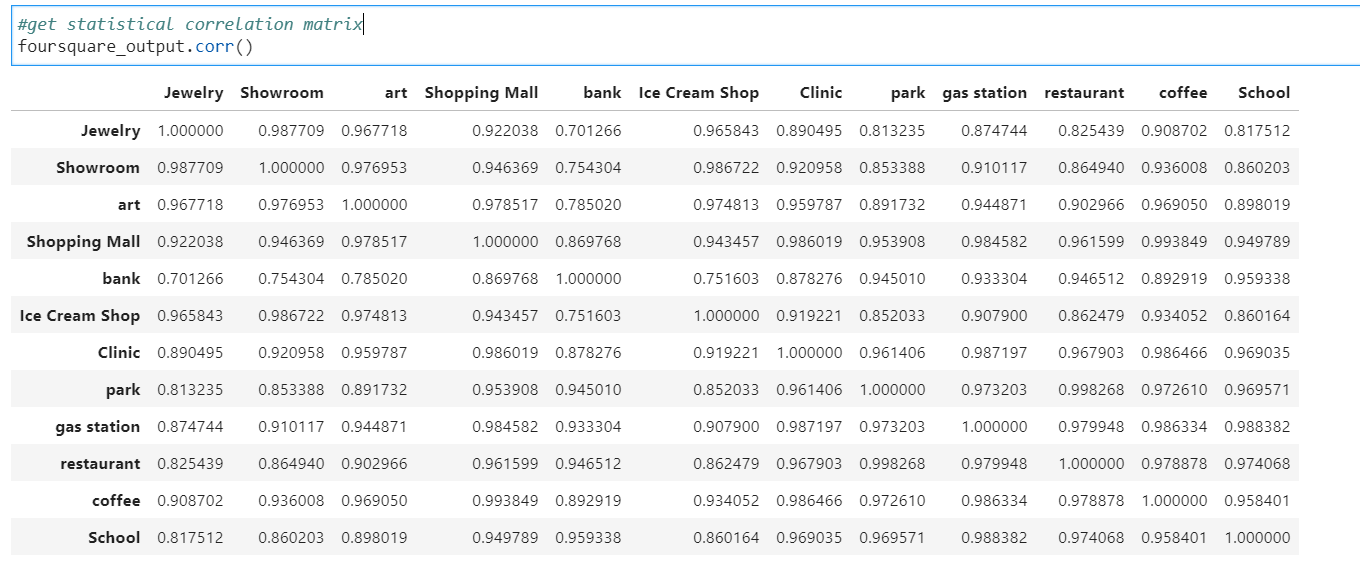
* 1. Inferential statistical testing

To check on the choice of ‘other venues’ to search for in foursquare let us do a correlation check on the presence of car/bike show rooms if a neighborhood has a ‘jewelry show room’. Similarly the correlation check can be done for other search queries I decided upon. This can be done using the statistical technique correlation matrix. The data is present in the dataframe named foursquare\_output shown as below.



**Table1:** Dataframe ‘foursquare\_output’ has above data

Using the statistical command foursquare\_output.corr() we can get the correlation matrix as below.



**Table2:** Correlation matrix.

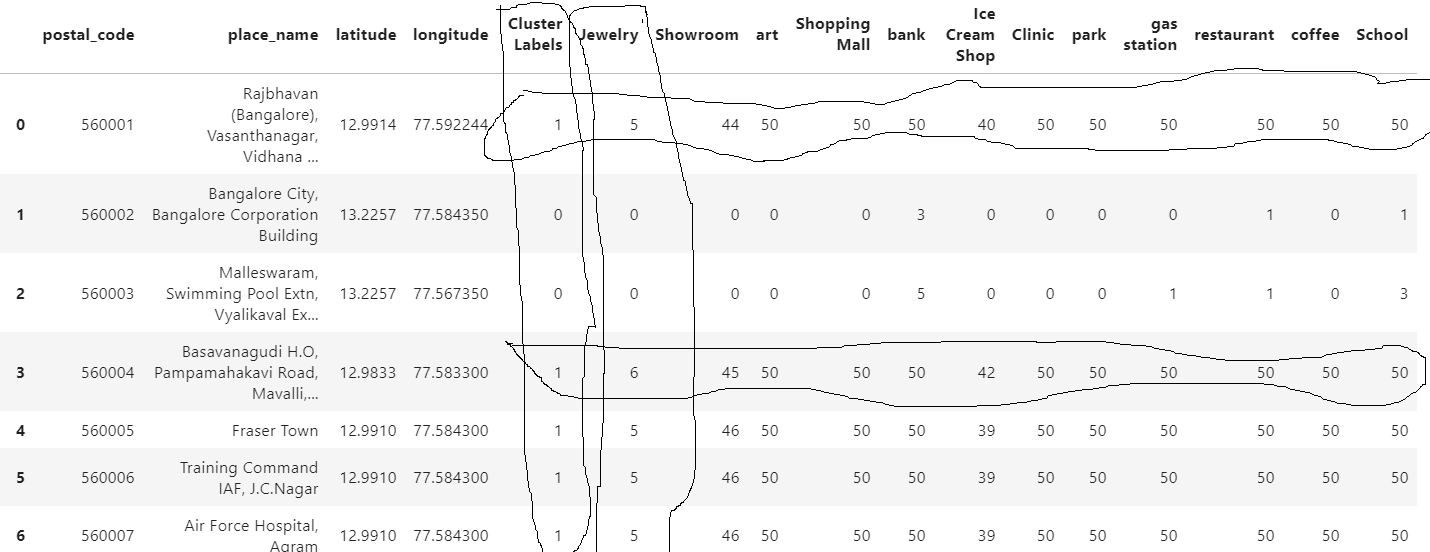
The high values of correlation between number of jewelry shops and number of other venues we chose, point to the right choice of search queries.

3.3 Summary of machine learning techniques used

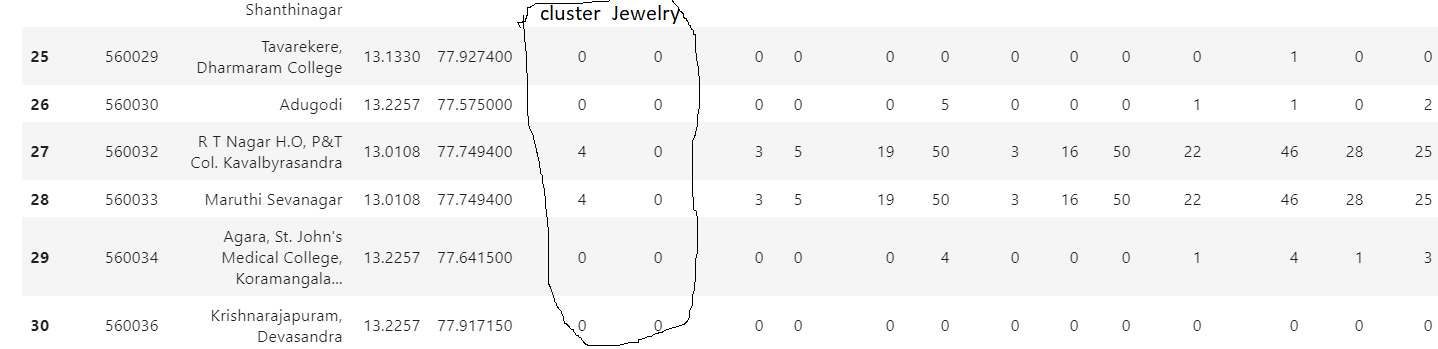
k-means clustering is the machine learning technique used in this project work. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. A cluster refers to a collection of data points aggregated together because of certain similarities. You'll define a target number k, which refers to the number of centroids you need in the dataset.

4. Results Section

The results from this study is summarized in the data frame final\_summary. This contains the column named ‘Cluster Labels’. The K-means clustering technique was applied on the gathered information per pin code (which are # of jewelry shops, # of show rooms, malls, banks, restaurants etc in the vicinity of 1.5KM. By clustering into five distinct zones, one can clearly see that cluster 1 had the greatest number of jewelry shops. It is interesting to note that cluster 1 also happens to be the one with many showrooms in the vicinity. This confirms our initial guess on what constitutes a posh locality and where a jewelry shop needs to be present. Similarly, cluster 3 is made up of pincodes which do have a good number of showrooms but has only one or nil jewelry shops. Thus, our search for a optimal place lies within these pin code areas as well. Below is a screen shot of the summary table capturing the count of items of interest in a neighborhood.



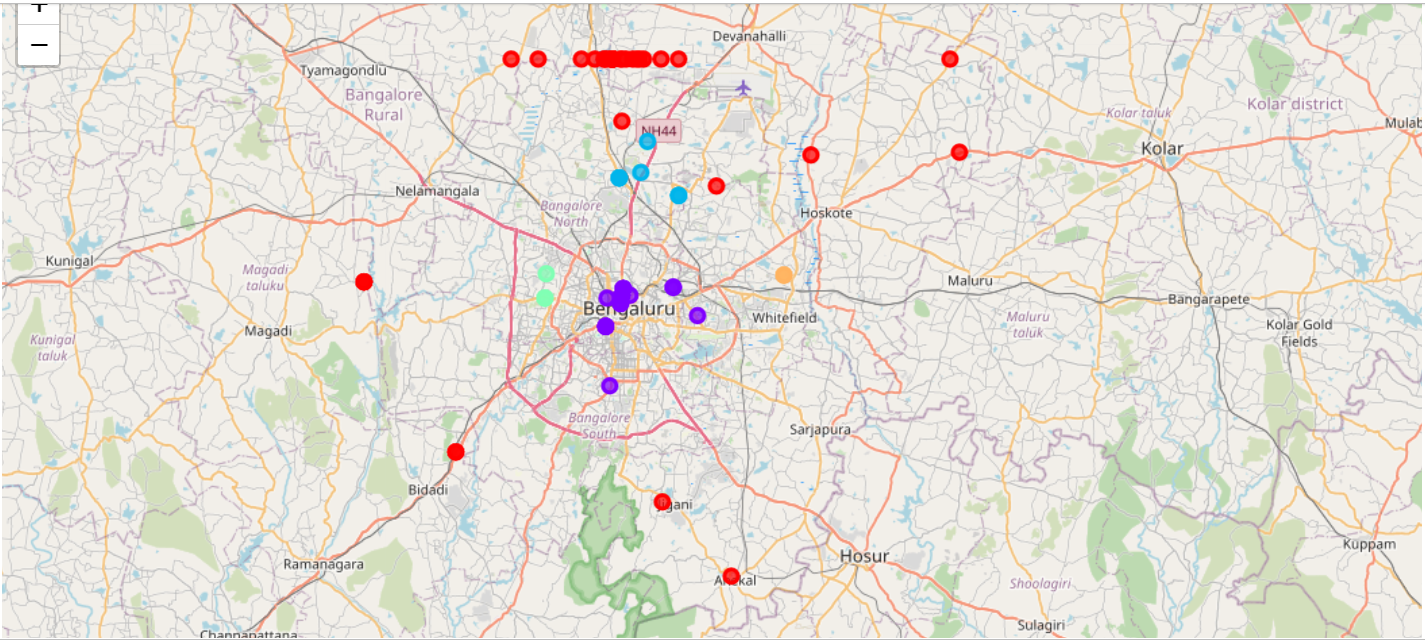
**Table3:** sample extract of final summary table (cluster label-1 clearly identifies all postal codes that have multiple jewelry shops and also has large number of showrooms, art, shopping malls, banks, ice cream shops, parks, gas stations, restaurants, coffee, schools in the vicinity.



**Table4:** sample extract of final summary table (cluster label-0 & 4 clearly identifies all postal codes that have zero jewelry shops and also has significantly lesser (compared to cluster-1) number showrooms, art, shopping malls, banks, ice cream shops, parks, gas stations, restaurants, coffee, schools in the vicinity.



**Table5:** sample extract of final summary table (cluster label-2 & 3 clearly identifies all postal codes that have zero or one jewelry shop). Cluster 2 has very few other ‘venues’ and cluster-3 has quite a good number of other venues like showrooms, art, shopping malls, banks, ice cream shops, parks, gas stations, restaurants, coffee, schools in the vicinity.



**Figure1:** Folio map of clustered pin codes of Bangalore based on #of jewelry shops, car/bike show rooms, restaurants, banks etc in a neighborhood (extracted from thousands of FOURSQUARE search queries)

5. Discussions Section

5.1 Observations from results

# observation: we can clearly see that cluster one has multiple jewelry shops and one among them had just one jewelry shop whereas others in cluster-1 are > 4. This presents an opportunity to have one more jewelry shop in that area pin code.

# index 66, JP Nagar with pincode of 560078 will be a good choice for setting up another gold shop. Notice the strong correlation between the presence of jewelry shop and showroom in the same pincode area. This is successfully captured in the grouping by the clustering algorithm with cluster-label =1.

It is also observed that all the jewelry shops are concentrated in Bengaluru center and almost none away. Cluster 3 aptly captures our needs. It has suitable number of adjacencies in the form of showrooms, malls, ice cream shops indicating an upclass locality and currently this cluster has only one jewelry shop.

5.2 Recommendations based on results

# index 66, JP Nagar with pincode of 560078 will be a good choice for setting up another gold shop. Notice the strong correlation between the presence of jewelry shop and showroom in the same pincode area. This is successfully captutred in the grouping by the clustering algorithm with cluster-label =1.

6. Conclusion

We thus were able to use webscraping, foursquare APIs and clustering techniques to identify potential places where a new jewelry shop can be set up. And the recommendation is JP Nagar with pincode of 560078.