

July 10, 2020

Business Ideas for different areas of Gurugram, India

Milind Kumar Sinha



Introduction

Positioning has always been an important element of setting up a business. Your success as a business depends on how well you are positioned to be found. Positioning includes various factors, from location to the price of your product or service to the message you use to promote the business, online and offline.

A lot of times, a business simply misses out getting the traction it deserves, due to inadequate positioning. For ex. If you set up a high-end clothing apparel's showroom in the middle of a slum, it would never be pinpointed by its target audience. On the other hand, if

you place it next to an exquisite restaurant, it would be easily noticed by the fancy customers who come there to dine.

Hence, it is necessary for every aspiring entrepreneur to pick out the location to attract the appropriate target audience for their business, keeping in mind the nearby points of attraction of the neighborhood.

Data

To address the problem, the following data were used:

1. Areas in Gurgaon represented by different pin codes. (scraped from <https://gurugram.gov.in/std-pin-codes/>)
Ex. Akhera B.O-122107
2. Coordinates (latitudes and longitudes) of the pin codes representing Gurgaon. (extracted from https://www.google.com/search?q=coordinates+of+pincodes+of+gurgaon&rlz=1C1CHBF_enIN883IN883&oq=coordinates+of+pincodes+of+gurgaon&aqs=chrome..69i57.17612j0j7&sourceid=chrome&ie=UTF-8)
Ex. 122001- 28.4555N, 77.0219E
3. Average plot price in different areas of Gurgaon (extracted from 99acres.com, propertywala.com)
Ex. Arjun Colony-Rs. 1,20,000
4. Exploratory data representing top 10 most visited venues around an area. For ex. Cafes, ATM, etc around Palam Vihar. (Data received by making calls to Foursquare API)

Methodology

Web Scraping

The process starts with extracting the details of various pin codes represented under the umbrella of Gurgaon. The data is facilitated by the official government documentation at <https://gurugram.gov.in/std-pin-codes/>

To extract the data, we scraping the data employing a brilliantly useful Python package called **Beautiful Soup**.

Beautiful Soup is a Python library for getting data out of HTML, XML, and other markup languages. Say you have found some webpages that display data relevant to your research, such as date or address information, but that do not provide any way of downloading the data directly. Beautiful Soup helps you pull content from a webpage, remove the HTML markup, and save the information. It is a tool for web scraping that helps you clean up and parse the documents you have pulled down from the web.

One of the primitive things Beautiful Soup can help us with is locating content that is buried within the HTML structure. Beautiful Soup allows you to select content based upon tags. To get a good view of how the tags are nested in the document, we use the method “prettify” on our soup object.

In our code, we extract the data from a table and hence the code segment:

```
locality = soup.select('table')[-1]
locality_rows = locality.find_all('tr')
```

Area:	Akhera B.O
Pin code:	122107
Area:	Arjun Nagar S.O
Pin code:	122001
Area:	Baded B.O
Pin code:	122508
Area:	Badli B.O
Pin code:	122508
Area:	Badshahpur S.O
Pin code:	122101
Area:	Baghanki B.O
Pin code:	122105
Area:	Basai B.O
Pin code:	122006
Area:	Basai Road S.O
Pin code:	122001
Area:	Baslambi B.O
Pin code:	122503

Location Data

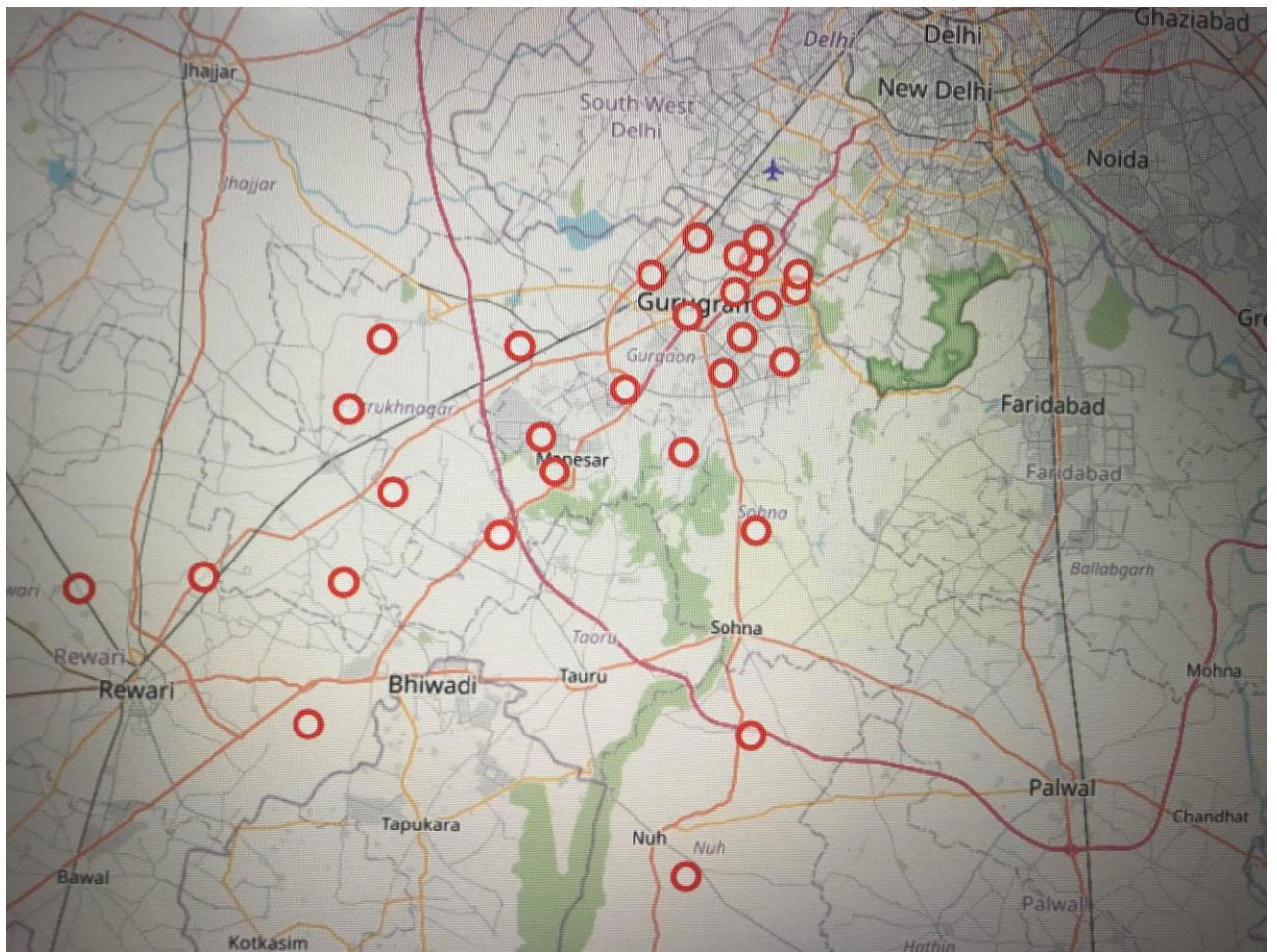
After successfully extracting the data representing all pin codes under Gurgaon, we combine the data with its general coordinates to further plot the various areas on maps and clusters.

Map of Gurugram

To represent the pin codes considered by us, on a map, we use the Python library, **Folium**.

[Folium](#) is a Python library used for visualizing geospatial data. It is easy to use and yet a powerful library. Folium is a Python wrapper for [Leaflet.js](#) which is a leading open-source JavaScript library for plotting interactive maps.

It has the power of Leaflet.js and the simplicity of Python, which makes it an excellent tool for plotting maps. Folium is designed with simplicity, performance, and usability in mind. It works efficiently, can be extended with a lot of plugins, has a beautiful and easy-to-use API.



Foursquare API



To extract the exploratory information regarding 10 most visited venues of our selected areas (represented by respective pin codes), we employ the lovely goldmine of geospatial data, **Foursquare API**.

The Foursquare Places API provides location-based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement. JSON is the preferred response format.

Clustering

Areas with similar characteristics are clustered together to give similar business plans as solution.

This is done by making use of machine learning technique of **clustering**.

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as Euclidean-based distance or correlation-based distance. The decision of which similarity measure to use is application specific.

The features being employed to carry out clustering are the category of top 10 most visited venues in the different areas being considered as well as the normalised plot price in each of these areas.

Silhouette Score

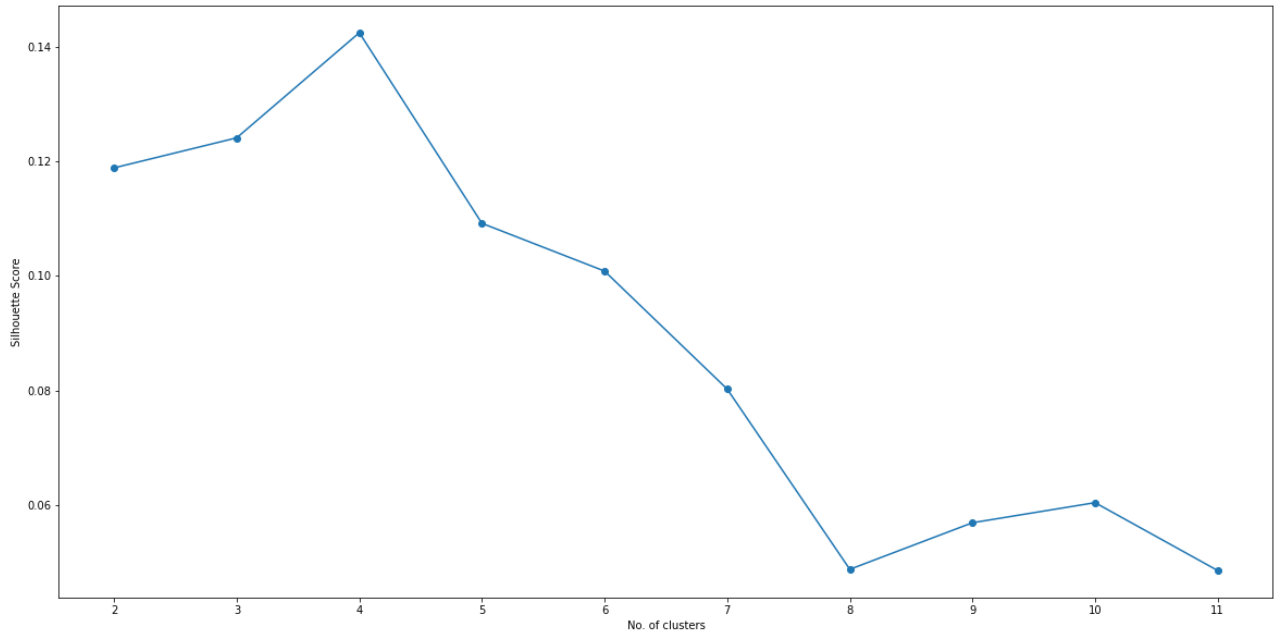
To find out the optimum number of clusters to analyse the data, we find the **silhouette score** of the feature's dataset.

It is calculated for each instance and the formula goes like this:

$$\text{Silhouette Coefficient} = (x-y) / \max(x,y)$$

where, y is the mean intra cluster distance: mean distance to the other instances in the same cluster. x depicts mean nearest cluster distance i.e. mean distance to the instances of the next closest cluster.

The coefficient varies between -1 and 1. A value close to 1 implies that the instance is close to its cluster is a part of the right cluster. Whereas, a value close to -1 means that the value is assigned to the wrong cluster.



As you can see, we get a global maximum at $k=4$. Hence, we form 4 clusters to analyse the areas.

K-means

K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

Results

After applying k-means clustering, we receive four clusters of areas as follows:

Cluster 1

<u>Pin code</u>	<u>Area</u>	<u>1st Most Common Venue</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>	<u>5th</u>	<u>6th</u>	<u>7th</u>	<u>8th</u>	<u>9th</u>	<u>10th</u>	<u>Relative Plot Price</u>
122003	Gurgaon Sector 45 S.O, Gwal Pahari B.O, Hailym..	Food Truck	Women's Store	Café	Garden Center	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Department Store	Coffee Shop	0.848485

Cluster 2

<u>Pin code</u>	<u>Area</u>	<u>1st Most Common Venue</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>	<u>5th</u>	<u>6th</u>	<u>7th</u>	<u>8th</u>	<u>9th</u>	<u>10th</u>	<u>Relative Plot Price</u>
122004	Khandsa B.O, Kherki Kaula B.O, Lakhnaula B.O, ...	Business Service	Shopping Mall	Women's Store	Food Truck	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Department Store	Coffee Shop	0.424242
122007	Industrial Estate S.O (Gurgaon)	Hotel	Garden Center	Indian Restaurant	Bed & Breakfast	Shoe Store	Japanese Restaurant	Pizza Place	Brazilian Restaurant	Chinese Restaurant	Fast Food Restaurant	0.000000
122009	Galleria DLF-IV S.O	Café	Indian Restaurant	Hotel	Bakery	Coffee Shop	Italian Restaurant	Sandwich Place	Restaurant	Bistro	Market	0.248485
122015	Palam Road S.O, Sarhau B.O	ATM	Indian Restaurant	Café	Food Truck	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	Department Store	Coffee Shop	0.000000

122017	Carterpur i B.O, Palam Vihar S.O (Gurgaon)	Tennis Court	Shopping Mall	Market	Café	Wome n's Store	Busin ess Servi ce	Fast Food Restaura nt	Electro nics Store	Donut Shop	Diner	0.280606
---------------	---	-----------------	------------------	--------	------	----------------------	-----------------------------	--------------------------------	--------------------------	---------------	-------	----------

Cluster 3

<u>Pin code</u>	<u>Area</u>	<u>1st Most Commo n Venue</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>	<u>5th</u>	<u>6th</u>	<u>7th</u>	<u>8th</u>	<u>9th</u>	<u>10th</u>	<u>Relative Plot Price</u>
123401	Kankarw ali Rewari S.O	Women's Store	Café	Garden Center	Food Truck	Fast Food Restaur ant	Electr onics Store	Donut Shop	Diner	Departm ent Store	Coffee Shop	0.515152

Cluster 4

<u>Pin code</u>	<u>Area</u>	<u>1st Most Commo n Venue</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>	<u>5th</u>	<u>6th</u>	<u>7th</u>	<u>8th</u>	<u>9th</u>	<u>10th</u>	<u>Relative Plot Price</u>
122001	Arjun Nagar S.O, Basai Road S.O, Gurgaon H.O, ...	Electroni cs Store	Shopping Plaza	Lightin g Store	Gift Shop	Wome n's Store	Café	Food Truck	Fast Food Restaur ant	Donut Shop	Diner	0.484848
122002	Chakkar pur B.O, DLF QE S.O, Nathupur B.O	Indian Restaura nt	Café	Sandwi ch Place	Chines e Restaur an	Office	Phar macy	Pizza Place	Liquor Store	Bakery	Coffee Shop	0.763636
122010	DLF Ph- III S.O	Café	Asian Restaura nt	Depart ment Store	Mediterranean Restaur ant	Wome n's Store	Food Truck	Fast Food Restaura nt	Electro nics Store	Donut Shop	Diner	0.569697

122011	Gurgaon Sector	Gym	Market	Snack Place	Bakery	Pizza Place	Miscellaneous Shop	Fast Food Restaurant	Electronics Store	Donut Shop	Diner	0.909091
122016	Dundahe ra B.O, Indir Mewat B .O, Industrial Com...	Pizza Place	Fast Food Restaurant	Thai Restaurant	Tea Room	Donut Shop	Indian Restaurant	Diner	Sandwich Place	Café	Business Service	0.515152
122018	Gurgaon South City II S.O	Shopping Mall	BBQ Joint	Department Store	Club House	Women's Store	Café	Food Truck	Fast Food Restaurant	Electronics Store	Donut Shop	1.000000

Discussions

Based on the data presented in the above section, following observations can be made for the clusters:

Cluster 1

- The relative plot price is high; hence the area experiences a great footfall.
- The most frequently visited venues are women's stores and food stores, hence making women- the target audience.

Cluster 2

- The relative plot prices are way lower.
- The cluster reflects an “idealistic local shopping complex with ATM's and stores of all kinds” scenario, which isn't the highlight of the city (but frequent enough) and hence offers great opportunities of establishment.

Cluster 3

- The cluster is very similar to Cluster 1, but with lower plot price.

Cluster 4

- The cluster represents an ideal “on-the-run” characteristics with venues representing the necessity visits as well as the lifestyle visits of the younger generation (electronics stores, unconventional food stores, etc), hence making the age group 16-30 as the target audience.

Conclusion

After carefully examining the presented characteristics of each cluster, the following business plans can be recommended for areas lying in the cluster:

Cluster 1: Salon, Spa, beauty parlors, etc.

Cluster 2: Jewelry store, high-end bakery, etc.

Cluster 3: Salon, Spa, beauty parlors, bakery etc.

Cluster 4: Gaming arena, Clubs, Party halls, etc.