

Capstone Project - The Battle of Neighborhoods - Bangalore, India

– Week 1

1. Introduction

1.1. Background

Bangalore, officially known as Bengaluru, is the capital of the Indian state of Karnataka. It has a population of over ten million, making it a megacity and the third-most populous city and fifth-most populous urban agglomeration in India.

Bangalore is known as the "Garden City of India" because of its greenery, broad streets and the presence of many public parks. It is sometimes called as the "Pub Capital of India" and the "Rock/Metal Capital of India" because of its underground music scene and it is one of the premier places to hold international rock concerts.

Bangalore also has a very active and vibrant theatre culture

The diversity of cuisine is reflective of the social and economic diversity of Bangalore. Bangalore has a wide and varied mix of restaurant types and cuisines and Bangalorean's deem eating out as an intrinsic part of their culture.

1.2. Problem Description:

Now let me explain the context of this Capstone project through a scenario. Say you live on the South side of the city of Bangalore in India. You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, such as fast food joints, pharmacies, parks, schools and so on. Now say you receive a job offer from a great company on the other side of the city with great career prospects. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer.

Wouldn't it be great if you are able to determine neighborhoods on the other side of the city that are the same as your current neighborhood, and if not perhaps similar neighborhoods that are at least closer to your new job?

1.3. Objective

The aim of this report is to study and analyze the neighborhoods of Bangalore city and group them into similar clusters and, to analyze those clusters to gather meaningful information. That information can be used to find out neighborhoods that are same as your current neighborhood or at least similar.

1.4. Target Audience

This information provided by this report would be useful for people who are interested in relocating to a different part of the city and are interested in finding new neighborhoods that are highly similar to their existing neighborhood.

2. Data Description:

To consider the objective stated, we can list the below data sources used for the analysis.

2.1. Bangalore Neighborhood Data:

The following page was scraped to pull out the necessary information: <http://www.geonames.org/export/zip/>

The information obtained had to be cleansed and necessary data such as 'Pin Code', 'Neighborhood', 'State', 'District', 'Borough', 'Latitude', 'Longitude' of India was transformed into a pandas' data frame for further analysis. For this particular, Business Problem, the necessary data of 'Bangalore' – Borough was extracted.

2.2. Using Foursquare Location Data:

Foursquare data is very comprehensive and it powers location data for Apple, Uber etc. The Foursquare API is to retrieve information about the popular spots in Bangalore. The popular spots returned depends on the highest foot traffic and thus it depends on the time when the call is made. So we may get different popular venues depending upon different time of the day. The call returns a JSON file and we need to turn that into a data-frame.

We will then leverage this data in order to determine which locality is the most appropriate in order to locate the similar neighborhoods.

3. C. Methodology:

3.1. Scraping the Geonames page and gathering data into a Pandas dataframe

To start with our analysis, we used the **Beautiful Soup** package to transform the data in the text file of India(IN) on the Geonames page into the pandas dataframe and it was cleansed to the information required as seen below.

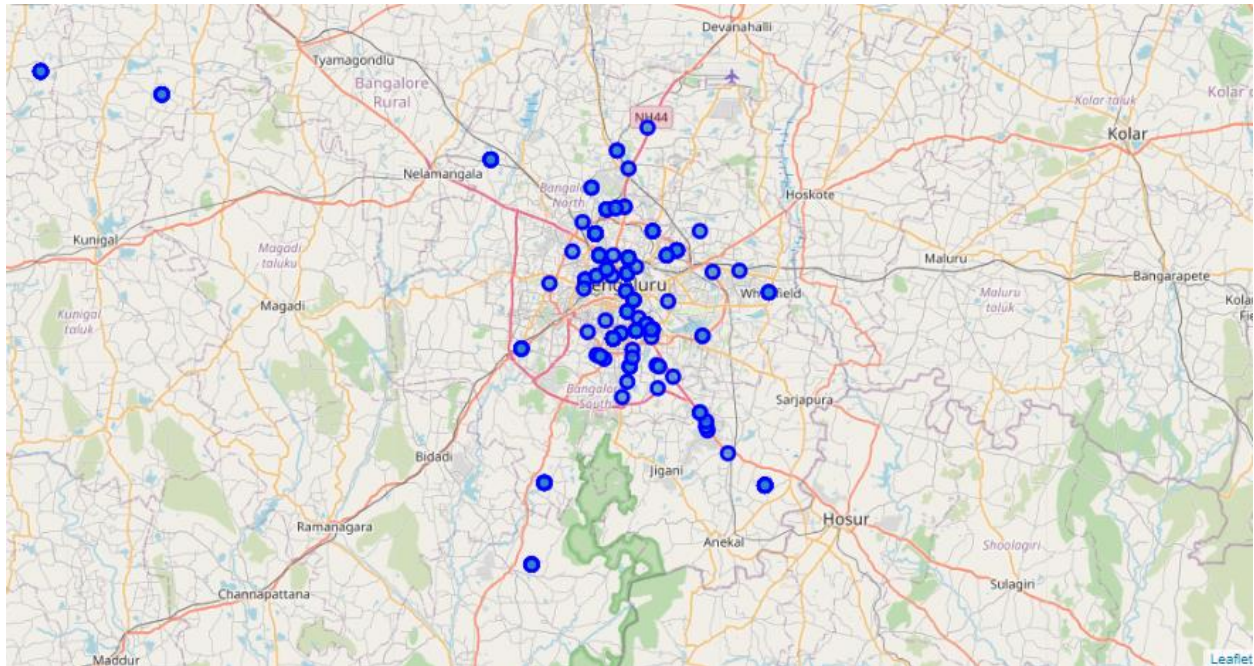
	Pin Code	Neighborhood	State	District	Borough	Latitude	Longitude
0	744101	Marine Jetty	Andaman & Nicobar Islands	South Andaman	Portblair	11.6667	92.7500
1	744101	Port Blair	Andaman & Nicobar Islands	South Andaman	Port Blair	11.6667	92.7500
2	744101	N.S.Building	Andaman & Nicobar Islands	South Andaman	Portblair	11.6667	92.7500
3	744102	Haddo	Andaman & Nicobar Islands	South Andaman	Port Blair	11.6833	92.7167
4	744102	Chatham	Andaman & Nicobar Islands	South Andaman	Portblair	11.7000	92.6667

3.2. Generating a map of Bangalore and plotting the Neighborhood data on it

We first filter the data to find boroughs containing the word “Bangalore”,

	Pin Code	Neighborhood	State	District	Borough	Latitude	Longitude
0	560001	Bangalore G.P.O.	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
1	560001	Legislators Home	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
2	560001	Mahatma Gandhi Road	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
3	560001	Vidhana Soudha	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
4	560001	Rajbhavan (Bangalore)	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
...
253	560110	Ullalu Upanagar	Karnataka	Bangalore	Bangalore	13.1791	77.0942
254	562130	Tavarekere (Bangalore)	Karnataka	Bangalore	Bangaloresouth	13.9288	77.0614
255	562157	Bettahalsur	Karnataka	Bangalore	Bangalore North	13.1438	77.6169
256	562162	Madanayakanahalli	Karnataka	Bangalore	Bangalore North	17.8647	77.8229
257	562162	Dasanapura	Karnataka	Bangalore	Bangalore North	17.8647	77.8229

We then use the python **folium** library to visualize geographic details of Bangalore and its boroughs. I created a map of Bangalore with boroughs superimposed on top using the latitude and longitude values to get the visual as below:



3.3. Utilizing Foursquare API to explore the neighborhoods

Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. We set the LIMIT parameter to **100**, which would limit the number of venues returned by the Foursquare API and the radius of 500 meter. Here is a head of the list of Nearby Venues for the first neighborhood i.e. The Bangalore GPO.

	name	categories	lat	lng
0	Ujwal Bar & Restaurant	Indian Restaurant	12.992280	77.594473
1	Millers 46	Steakhouse	12.991666	77.594207
2	Alliance Française	Concert Hall	12.991232	77.596723
3	Infinitea	Tea Room	12.987157	77.594835
4	Imperial Restaurant	Indian Restaurant	12.991150	77.593837

We create a new function that will repeat the process above for all the neighborhoods in Bangalore. This function will give us a list of all venues present in Bangalore city. Here is a head() value of this dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bangalore G.P.O.	12.9914	77.5944	Ujwal Bar & Restaurant	12.992280	77.594473	Indian Restaurant
1	Bangalore G.P.O.	12.9914	77.5944	Millers 46	12.991666	77.594207	Steakhouse
2	Bangalore G.P.O.	12.9914	77.5944	Alliance Française	12.991232	77.596723	Concert Hall
3	Bangalore G.P.O.	12.9914	77.5944	Infinitea	12.987157	77.594835	Tea Room
4	Bangalore G.P.O.	12.9914	77.5944	Imperial Restaurant	12.991150	77.593837	Indian Restaurant

3.4. Analyze each neighborhood

We use One Hot Encoding, use the neighborhood to group data, and find out the top ten venues present in each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A F Station Yelahanka	IT Services	Creperie	Food & Drink Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Event Service	Electronics Store	Eastern European Restaurant	Dor Sho
1	Achitnagar	Gym / Fitness Center	Health & Beauty Service	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Event Service	Electronics Store	Eastern European Restaurant	Dor Sho
2	Adugodi	Indian Restaurant	Fast Food Restaurant	Design Studio	Bus Station	Zoo Exhibit	Eastern European Restaurant	Financial or Legal Service	Farmers Market	Event Service	Elec Stor
3	Agara	Café	Indian Restaurant	Coffee Shop	Pub	Ice Cream Shop	Lounge	Pizza Place	Bar	Paper / Office Supplies Store	Mex Res
4	Agram	Home Service	Zoo Exhibit	Donut Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Event Service	Electronics Store	Eastern European Restaurant	Dive

We have some common venue categories in the neighborhoods. We use the unsupervised learning **K-means algorithm** to cluster the neighborhoods. K-Means algorithm is one of the most common method for clustering in unsupervised learning.

We use a **k_cluster** value of 11 to split the neighborhoods into 11 different clusters based on the similarity they have concerning the venues they contain.

4. Results:

4.1. Adding the Cluster Labels to the Venue Data

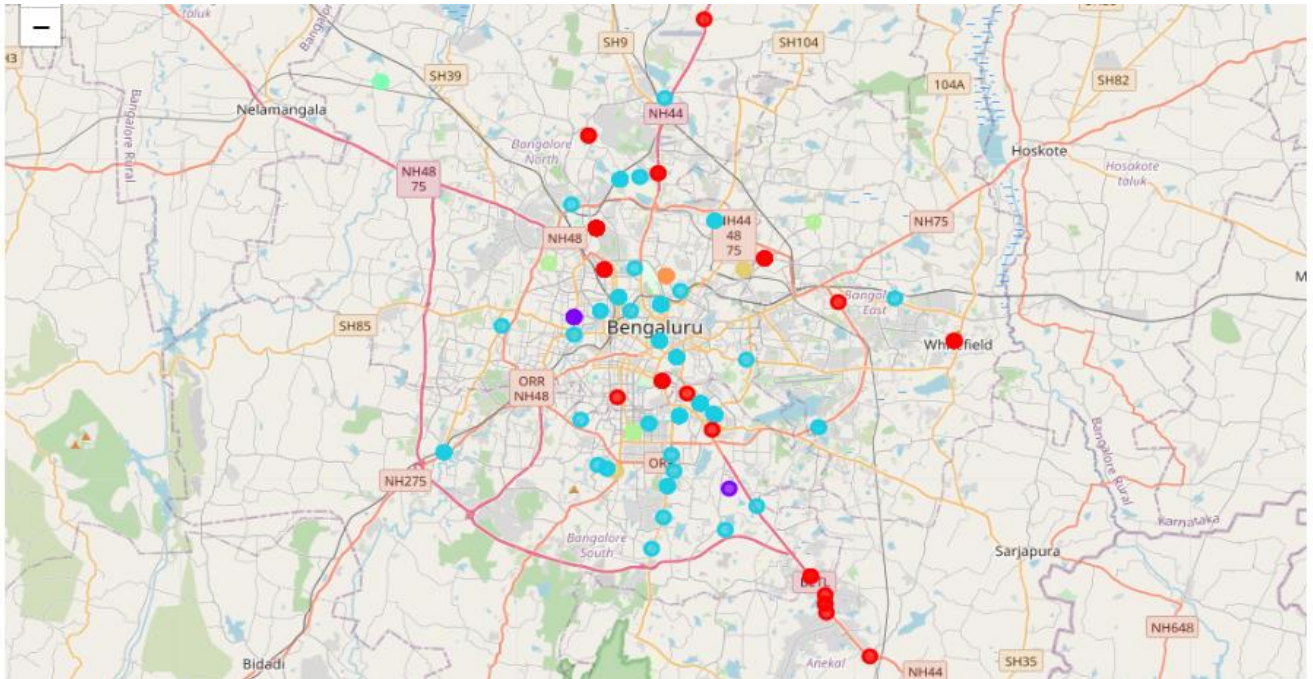
The below table depicts the clustered data along with the top 10 most common venues in that cluster.

:

Neighborhood	State	District	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Bangalore G.P.O.	Karnataka	Bangalore	Bangalore North	12.9914	77.5944	4	Indian Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant
Legislators Home	Karnataka	Bangalore	Bangalore North	12.9914	77.5944	4	Indian Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant
Mahatma Gandhi Road	Karnataka	Bangalore	Bangalore North	12.9914	77.5944	4	Indian Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant
Vidhana Soudha	Karnataka	Bangalore	Bangalore North	12.9914	77.5944	4	Indian Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant
Rajbhavan (Bangalore)	Karnataka	Bangalore	Bangalore North	12.9914	77.5944	4	Indian Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant

4.2. Visualizing the resulting Clusters

We use the matplotlib and folium packages to visualize the clusters on a map of Bangalore.



5. Discussion:

The intent with which analysis was carried out was to find out similar neighborhoods for a person relocating within the city.

As we analyze the results section, we can analyze the clusters and see similar neighborhoods in different parts of the city. For example, if we compare the different neighborhoods clustered in cluster 7.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
157	Tyagrajnagar	Bangalore South	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
158	B Sk II Stage	Bangalore South	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
159	Padmanabhnagar	Bangalore South	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
160	Jayanagar West	Bangalore South	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
163	Nagasandra (Bangalore)	Bangalore North	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
164	Nelakadirahalli	Bangalore North	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop
165	Bagalgunte	Bangalore North	12.9229	77.5798	7	Indian Restaurant	Coffee Shop	Hookah Bar	Salad Place	Fast Food Restaurant	Ice Cream Shop

As seen in the table above, if a person wished to move from North Bangalore to South Bangalore. If a person's current location were in the Neighborhood of Nagasandra (Bangalore) in North Bangalore, which has venues like Indian Restaurant, Coffee shop and Hookah bar nearby, the person, would like to relocate to a neighborhood like Jayanagar West in Bangalore South which also has venues like Indian Restaurant, Coffee shop and Hookah bar. This is just one example of how our data analysis can help people relocate from one part of the city to another which similar to their current localities.

6. Conclusion:

In a fast moving world, there are many real life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in Bangalore based on the most common venues in those neighborhoods. Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.

7. References:

- Geonames content: <http://www.geonames.org/export/zip/>
- Foursquare API