Capstone Project - The Battle of Neighborhoods

DATE: 08/10/2020
AUTHOR: Mercy Paul

Introduction

This project is the final part of the IBM Data Science Professional course in Coursera. The project requirement is to leverage the Foursquare location data to explore or compare neighborhoods or cities of choice or to come up with a problem that can use the Foursquare location data to solve.

This project will cover all phases in the data science life cycle to resolve a problem. The following components will be used:

- Folium library, including choropleth map, heatmap in map view
- Foursquare APIs
- K-Means Clustering Algorithm
- Pandas, Numpy

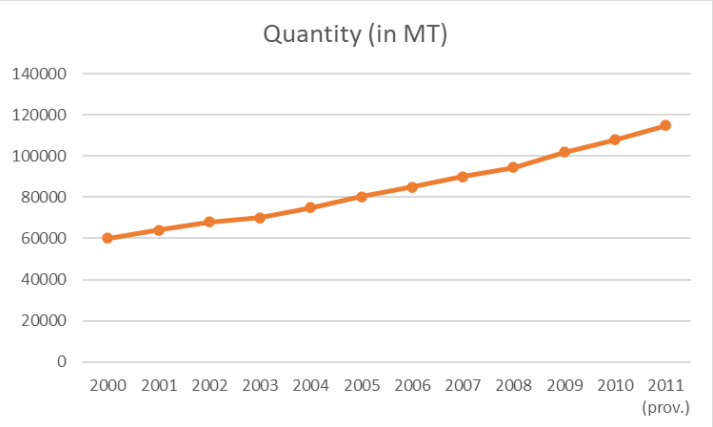
We will implement this project using the following steps/phases:

1. Define Business Objective
2. Data understanding
3. Modeling
4. Operationalize and acceptance

Business Objective/Problem Statement

This project involves a hypothetical scenario to explore the neighborhoods in Bangalore city to find the best location to establish a café. Bangalore is one of the largest capital cities in Southern India with a population of about 15 million. Bangalore is also one of the most ethnically diverse cities in the country. Bangalore is a hub for entrepreneurs and there is an opportunity in the quick service cafés catering to a young and aspiring population. With this idea in mind the objective is to explore the neighborhood and assess where the competition is and find the most suitable location to establish a café.

Bangalore is a fast-growing cosmopolitan from a diverse background. Since Bangalore is considered the IT hub of the country people from all over the country travel to Bangalore to do business. The Coffee Board of India is an organization managed by the Ministry of Commerce and Industry of the government of India to promote coffee production in India.



Source: https://www.indiacoffee.org/coffee-statistics.html?page=CoffeeData#cof

According to the board the domestic consumption of coffee in India has been steadily growing and expected to continue. All though there is no dearth to the number of cafes present in Bangalore the ever growing traffic congestion introduces the need for customers to find cafes or coffee places without having to commute for long hours and find a quiet spot to unwind.

The target audience for this use case will be any entrepreneur looking to open a café in the city of Bangalore. The objective hence is to find the most preferred area using scientific methods to open a café in Bangalore city.

Understanding Data

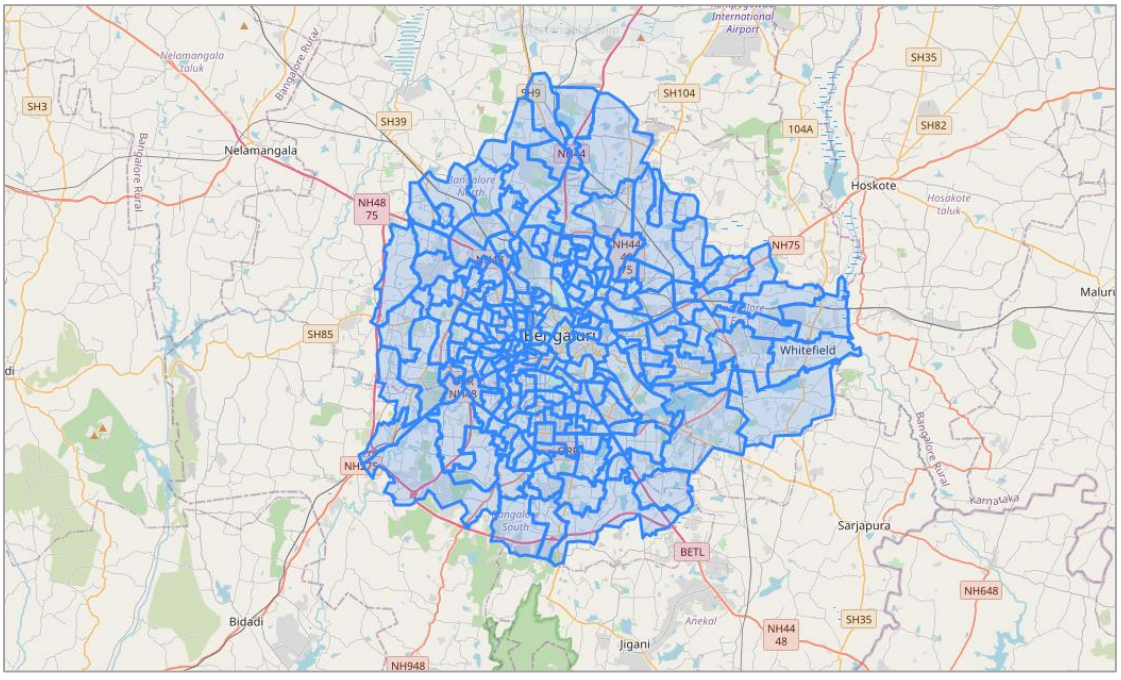
Data Sources

To address the problem statement, we will utilize the following data:

1. Bangalore city ward data:

   The district of Bangalore is the capital city of the state Karnataka. The city of Bangalore is divided into 198 wards for administrative purposes. The Bangalore city dataset provides the shape information of every ward in Bangalore and population density for every ward. This data is made available by datameet.org in GeoJSON format.

	OBJECTID	ASS_CONST_	ASS_CONST1	WARD_NO	WARD_NAME	POP_M	POP_F	POP_SC	POP_ST	POP_TOTAL	AREA_SQ_KM	LAT	LON	RESERVATIO	geometry
0	1	150	Yelahanka	2.0	Chowdeswari Ward	10402.0	9224.0	2630.0	286.0	19626.0	7.06	13.121709	77.580422	General	MULTIPOLYGON (((77.59229 13.09720, 77.59094 13...
1	2	150	Yelahanka	3.0	Atturu	13129.0	10891.0	2921.0	665.0	24020.0	10.15	13.102805	77.560038	General (Women)	MULTIPOLYGON (((77.56862 13.12705, 77.57064 13...
2	3	150	Yelahanka	4.0	Yelahanka Satellite Town	13457.0	12325.0	3687.0	601.0	25782.0	4.90	13.090987	77.583925	Backward Category - A	MULTIPOLYGON (((77.59094 13.09842, 77.59229 13...
3	4	151	K.R. Puram	51.0	Vijnanapura	18118.0	16969.0	6454.0	228.0	35087.0	2.05	13.006063	77.669565	Scheduled Caste	MULTIPOLYGON (((77.67683 13.01147, 77.67695 13...
4	5	151	K.R. Puram	53.0	Basavanapura	11494.0	10518.0	4115.0	325.0	22012.0	6.28	13.016847	77.715456	General	MULTIPOLYGON (((77.72899 13.02061, 77.72994 13...



2. Foursquare API:

Foursquare location data will be used for to retrieve the venue details for every ward. The API details will tell us the location of existing cafes and their popularity, this information will also help us identify similar locations with an opportunity to establish a new café.

	index	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	7	Hudi	13.022376	77.705493	Cafe Coffee Day	13.020088	77.709071	Café
1	29	Kengeri	12.915069	77.477528	Cafe Coffee Day	12.915288	77.481766	Café
2	38	Rajarajeshwari Nagar	12.918757	77.522142	Kaapi Katte	12.916458	77.520567	Café
3	39	Rajarajeshwari Nagar	12.918757	77.522142	Cafe Coffee Day	12.920846	77.520570	Café
4	111	Mattikere	13.032590	77.561034	Cafe Coffee Day	13.031717	77.559406	Café
...	...	...	...	...	...	...	...	...
116	1073	Pattabhiram Nagar	12.924545	77.587545	Starbucks	12.924624	77.583805	Coffee Shop
117	1125	Bellanduru	12.922874	77.680209	Starbucks	12.922508	77.680960	Coffee Shop
118	1150	Dodda Nekkundi	12.968183	77.707824	Cafe Coffee day	12.966507	77.709676	Coffee Shop
119	1185	Hongasandra	12.896769	77.627517	Jaiswal Coffee Roaster	12.893406	77.624910	Coffee Shop
120	1254	Marathahalli	12.950743	77.691495	Hatti Kaapi	12.948150	77.689790	Coffee Shop

We will combine the ward level population and Foursquare location data to identify areas where the current cafes are located. This gives an idea as to the potential locations to avoid for establishing the new café.

Methodology

Exploring Data

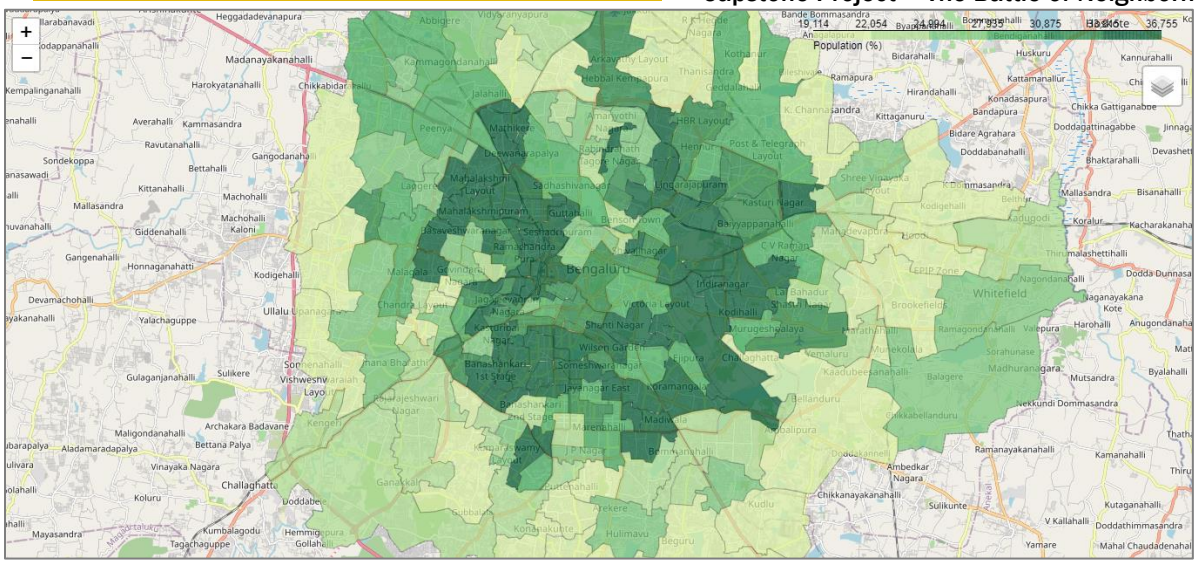
The Bangalore city source data has the following columns:

```
OBJECTID        int64
ASS_CONST_      object
ASS_CONST1      object
WARD_NO         float64
WARD_NAME       object
POP_M           float64
POP_F           float64
POP_SC          float64
POP_ST          float64
POP_TOTAL       float64
AREA_SQ_KM      float64
LAT             float64
LON             float64
RESERVATIO      object
geometry        geometry
```

The columns that we will use are:

```
WARD_NAME
POP_TOTAL
LAT
LON
geometry
```

Using the above data, we can identify the densely populated areas of Bangalore denoted by heat map as shown below. We will want to establish our café in potential densely populated or areas with potential growth.

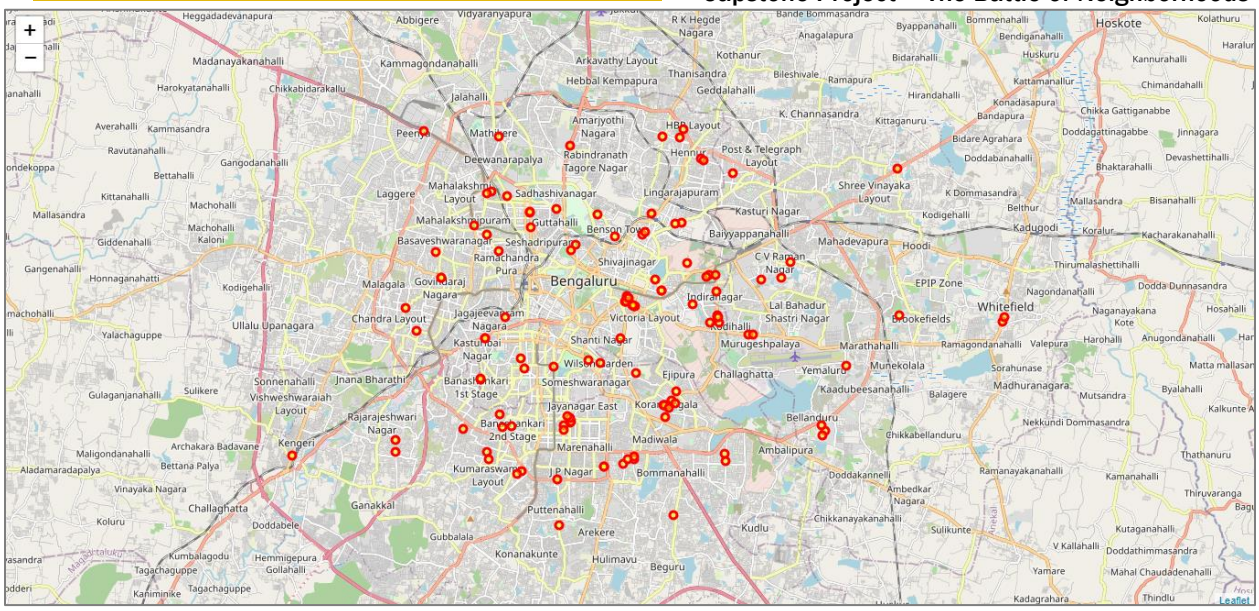


We will use the foursquare API data to explore Bangalore city to identify venues and filter out cafes. On exploring all the venues in Bangalore, we found about 183 unique venues. Out of the 183 unique values we found that cafes were categorized as Cafes or Coffee Shop. On filtering out coffee shops we identified about 122 cafes in Bangalore as shown below.

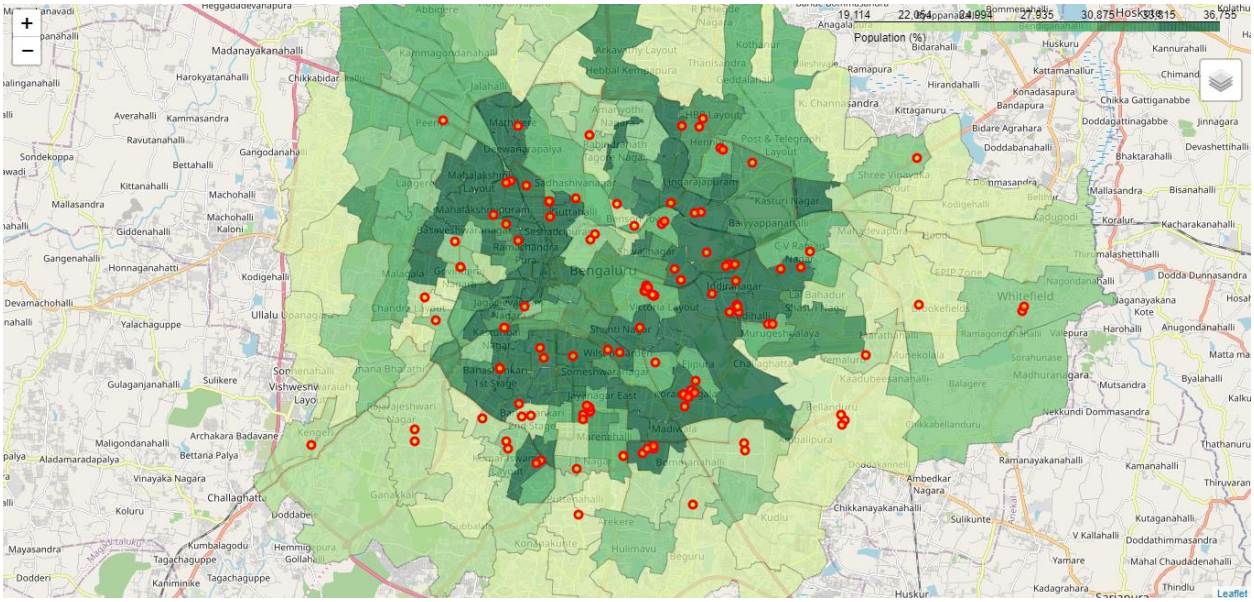
	index	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	4	Yelahanka Satellite Town	13.090987	77.583925	Vishwa Bakery	13.089937	77.583161	Café
1	11	Hudi	13.022376	77.705493	Cafe Coffee Day	13.020088	77.709071	Café
2	33	Kengeri	12.915069	77.477528	Cafe Coffee Day	12.915288	77.481766	Café
3	39	Rajarajeshwari Nagar	12.918757	77.522142	Kaapi Katte	12.916458	77.520567	Café
4	40	Rajarajeshwari Nagar	12.918757	77.522142	Cafe Coffee Day	12.920846	77.520570	Café
...	...	...	...	...	...	...	...	...
117	1082	BTM Layout	12.911838	77.609345	Lassi Shop	12.913702	77.607090	Coffee Shop
118	1098	Pattabhiram Nagar	12.924545	77.587545	Starbucks	12.924624	77.583805	Coffee Shop
119	1147	Bellanduru	12.922874	77.680209	Starbucks	12.922508	77.680960	Coffee Shop
120	1171	Dodda Nekkundi	12.968183	77.707824	Cafe Coffee day	12.966507	77.709676	Coffee Shop
121	1271	Marathahalli	12.950743	77.691495	Hatti Kaapi	12.948150	77.689790	Coffee Shop

122 rows × 8 columns

Visualizing the cafes on map.

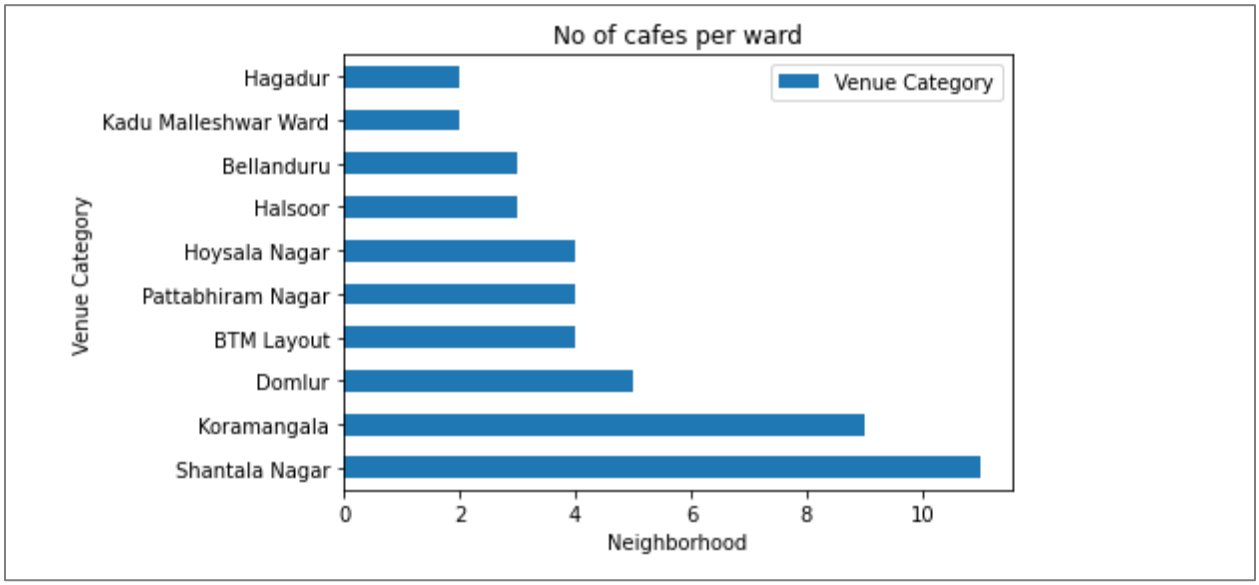


Superimposing the café locations on the ward level population density identifies the areas with cafés already established

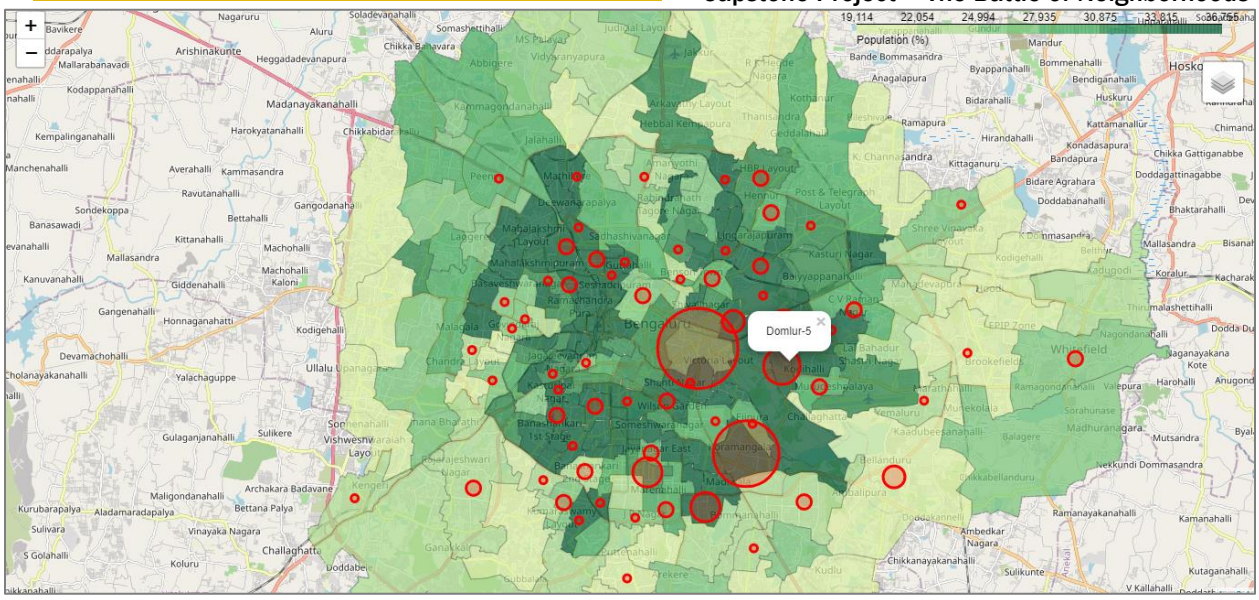


We clearly see that darker the population density more the number of cafes and also that most the cafes are concentrated towards the heart of the city. Now let's see if we can identify wards with enough cafes and possible areas to skip when scouting for locations.

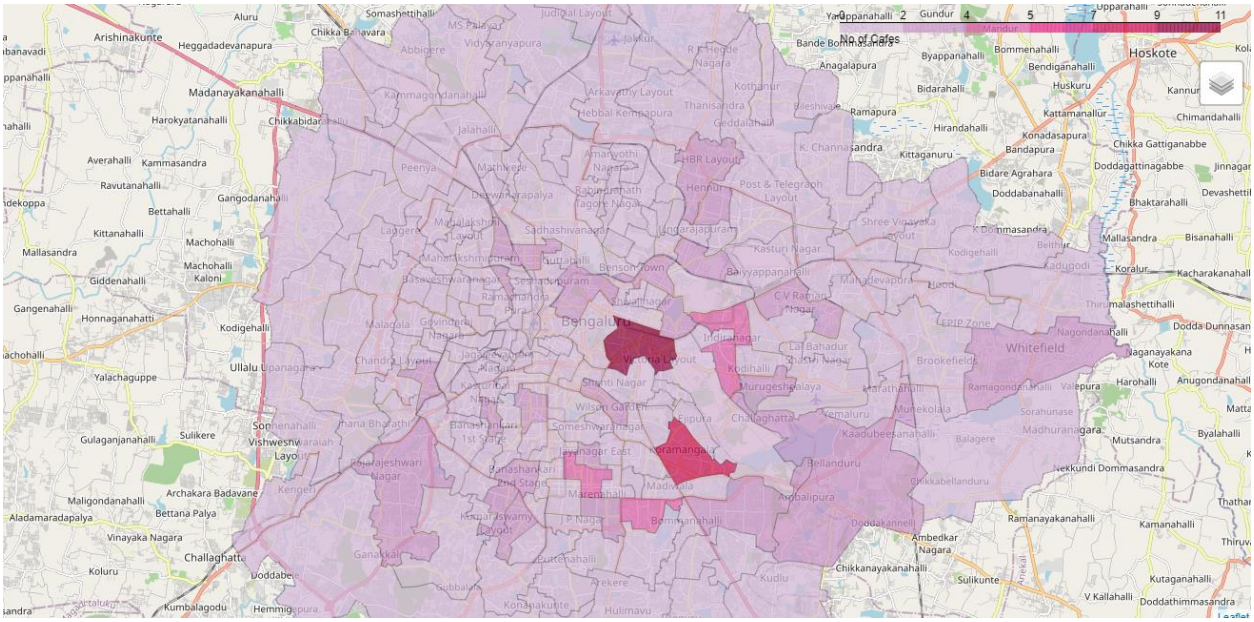
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Cafe_Count
Shantala Nagar	12.969867	77.606464	11
Koramangala	12.931340	77.624649	9
Domlur	12.963361	77.638040	5
BTM Layout	12.911838	77.609345	4
Hoysala Nagar	12.978456	77.638783	4
...	...	...	...
Agrahara Dasarahalli	12.980497	77.541535	1
Kengeri	12.915069	77.477528	1
Konankunte	12.885775	77.579901	1
Kottegepalya	12.982456	77.514090	1
Yelahanka Satellite Town	13.090987	77.583925	1



We can see that Shantala Nagar (also popularly known as M.G. Road) and Koramangala are the most popular location for cafes and there are around 10 cafes in each of the areas. Let's see what insights are gained by superimposing the café locations on the population density view.



We can see the wards with large number of cafes and possibly avoid reducing competition. Let's see another view of the cafe's distribution in Bangalore City.



We will further explore these neighborhoods to explore by using data modeling technique to find the most appropriate location.

Also let's analyze the data to see the top 10 common venues in each neighborhood. This will help us at a later stage to pick the most appropriate location.

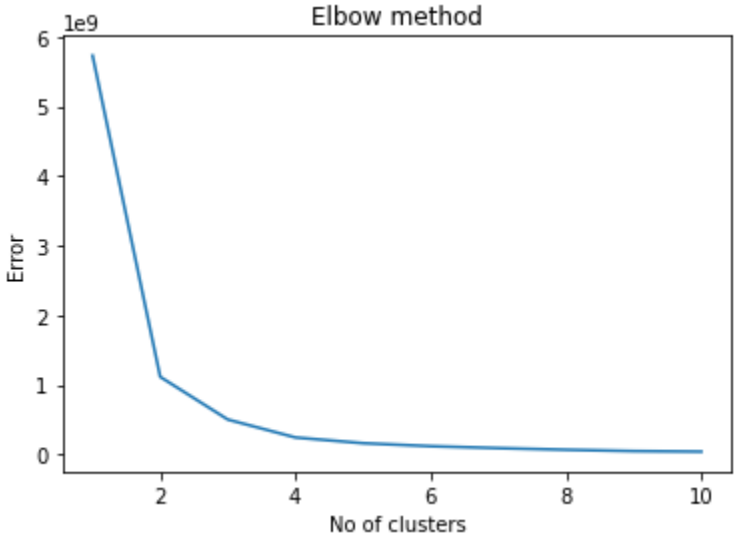
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A Narayanapura	Movie Theater	Bus Stop	Fish Market	Electronics Store	Bus Station	Fast Food Restaurant	Farmers Market	Event Space	Eastern European Restaurant	Dumpling Restaurant
1	Adugodi	Indian Restaurant	Café	Yoga Studio	Donut Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
2	Agaram	Burger Joint	Yoga Studio	Dry Cleaner	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant
3	Agrahara Dasarahalli	Bagel Shop	Indian Restaurant	Garden	Café	Yoga Studio	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space
4	Anjanapura	ATM	Campground	Donut Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant

Modeling using K-Means

We will use K-means to segment the data using the ward population and number of cafes features.

	Neighborhood	POP_TOTAL	Cafe_Count
0	Chowdeswari Ward	19626.0	0.0
1	Atturu	24020.0	0.0
2	Yelahanka Satellite Town	25782.0	1.0
3	Vijnanapura	35087.0	0.0
4	Basavanapura	22012.0	0.0
...	...	...	...
193	Madivala	35155.0	0.0
194	Ramamurthy Nagar	21999.0	0.0
195	Horamavu	28167.0	0.0
196	Marathahalli	22489.0	1.0
197	Hemmigepura	24311.0	0.0

To determine K we will use the elbow method on the above data set.

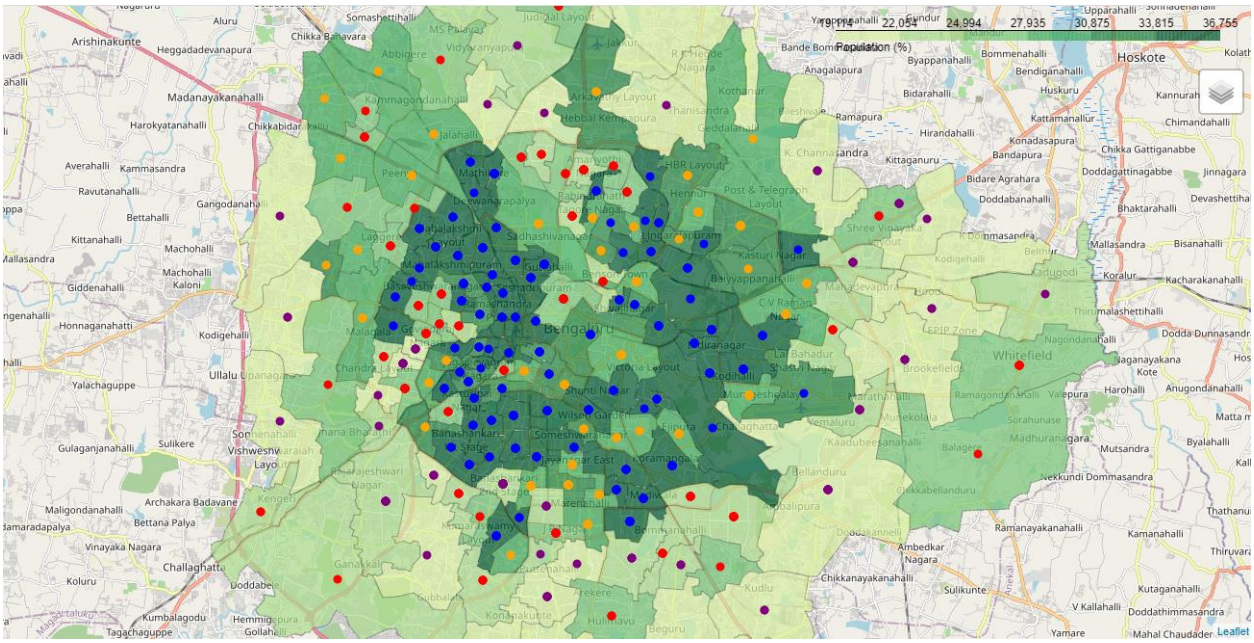


The point after which the distortion/inertia start decreasing in a linear fashion is k=4. The results of algorithm is the assignment of a label to a neighborhood as shown below.

	Neighborhood	POP_TOTAL	Cafe_Count	Cluster Labels	LAT	LON
0	Chowdeswari Ward	19626.0	0.0	0	13.121709	77.580422
1	Atturu	24020.0	0.0	3	13.102805	77.560038
2	Yelahanka Satellite Town	25782.0	1.0	3	13.090987	77.583925
3	Vijnanapura	35087.0	0.0	1	13.006063	77.669565
4	Basavanapura	22012.0	0.0	0	13.016847	77.715456
...	...	...	...	...	...	...
193	Madivala	35155.0	0.0	1	12.920018	77.614418
194	Ramamurthy Nagar	21999.0	0.0	0	13.033613	77.676539
195	Horamavu	28167.0	0.0	2	13.044561	77.653271
196	Marathahalli	22489.0	1.0	0	12.950743	77.691495
197	Hemmigepura	24311.0	0.0	3	12.891903	77.505013

198 rows × 6 columns

Let's run the K-Means algorithm on the dataset using k=4 and visualize the cluster results on the maps in different colors.



We observe that the clusters are concentric in nature and moves away from the heart of the city and very closely tied to the population density of the neighborhoods.

Results

Let's examine each of the clusters.

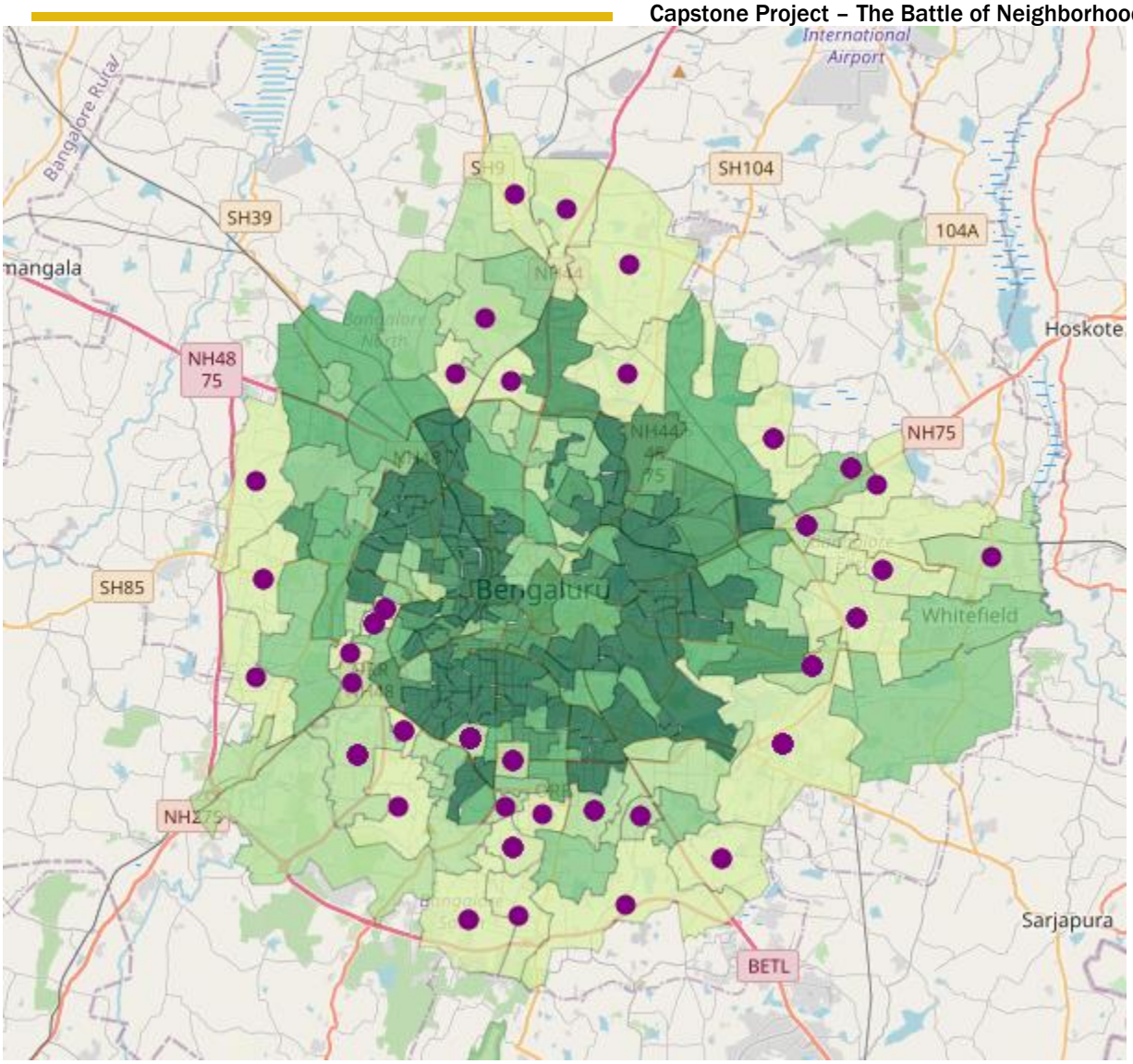
Cluster 0

We will first list all the neighborhoods in the first cluster and analyze them.

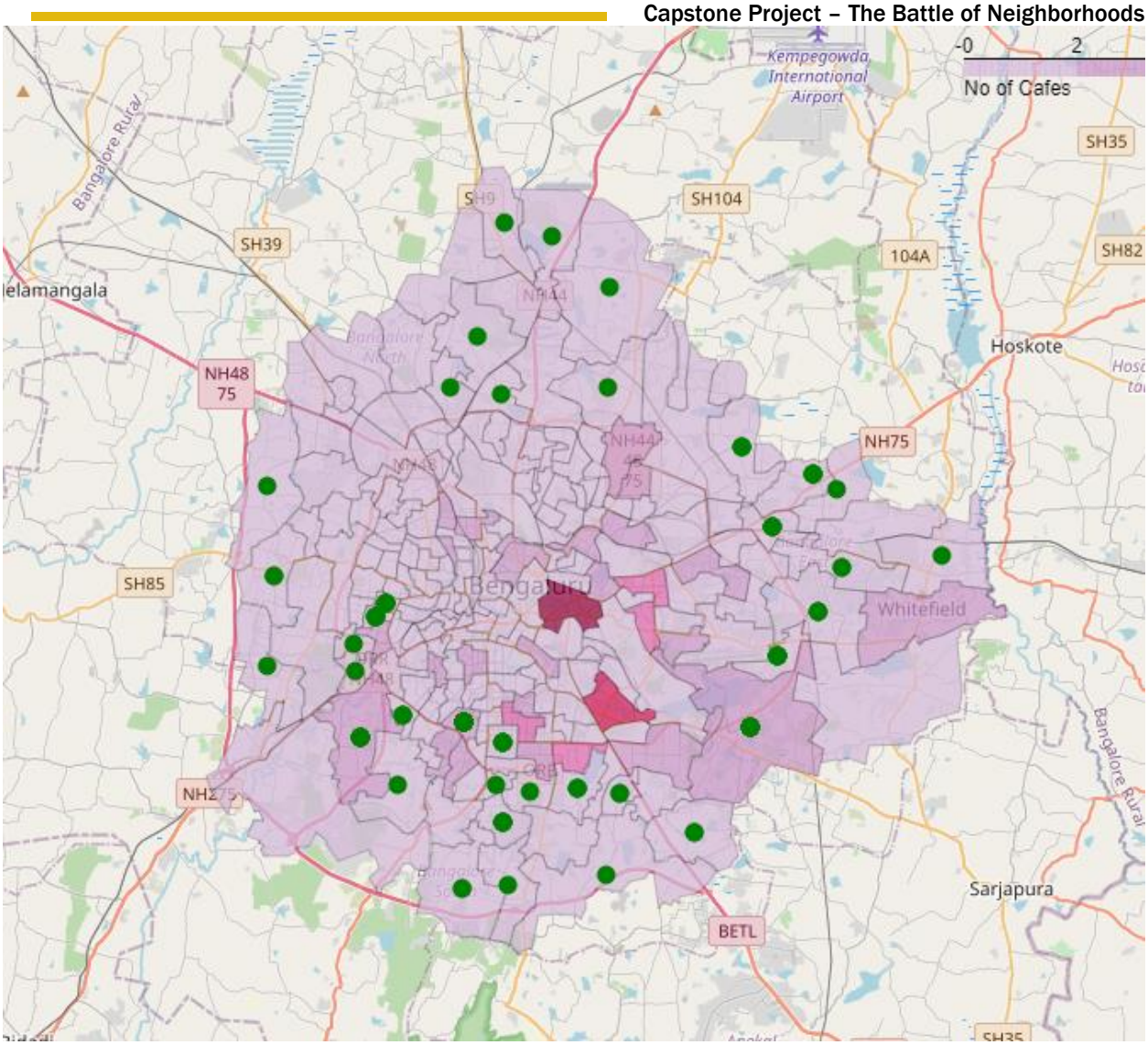
	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Chowdeswari Ward	0	Department Store	Yoga Studio	Donut Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant
1	Marenahalli	0	Indian Restaurant	Food Truck	Arts & Crafts Store	Bus Station	Yoga Studio	Dumpling Restaurant	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market
2	Maruthi Mandir ward	0	ATM	American Restaurant	Snack Place	Gym	Pharmacy	Dog Run	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
3	Nayandahalli	0	Southern / Soul Food Restaurant	Yoga Studio	Food	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant

There is a total of 23 neighborhoods in this cluster, by looking at the top 10 common venues in the cluster it is obvious that cafes are not very popular in this cluster.

Let's analyze this cluster from a location perspective.



It is obvious from this illustration that these are the less densely populated neighborhoods on the outskirts of the city.

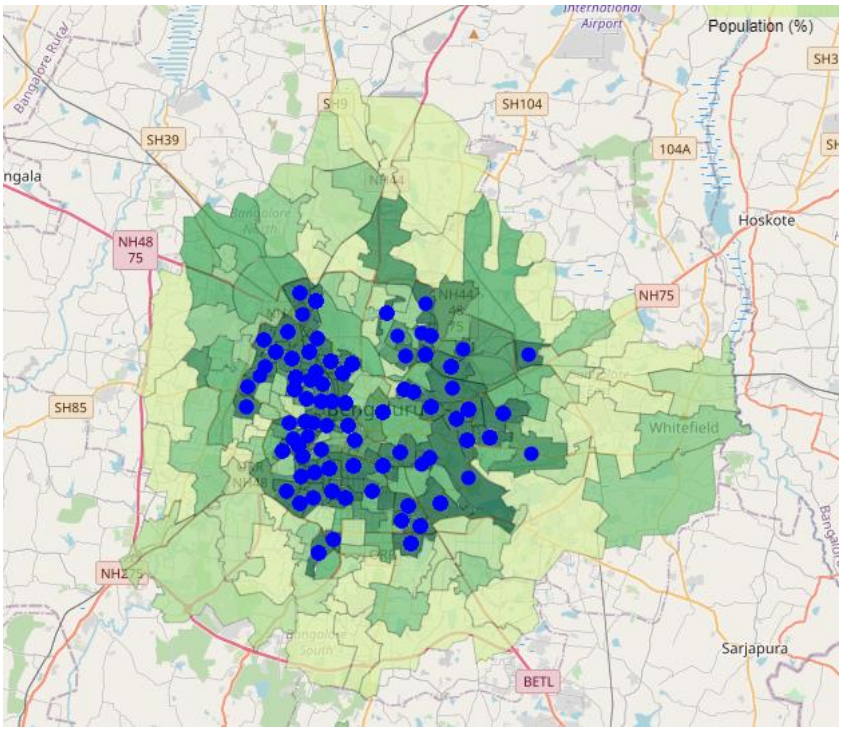


Also, it is very clear that cafes are not the most popular venues in these neighborhoods and not the most suitable location to establish the new café.

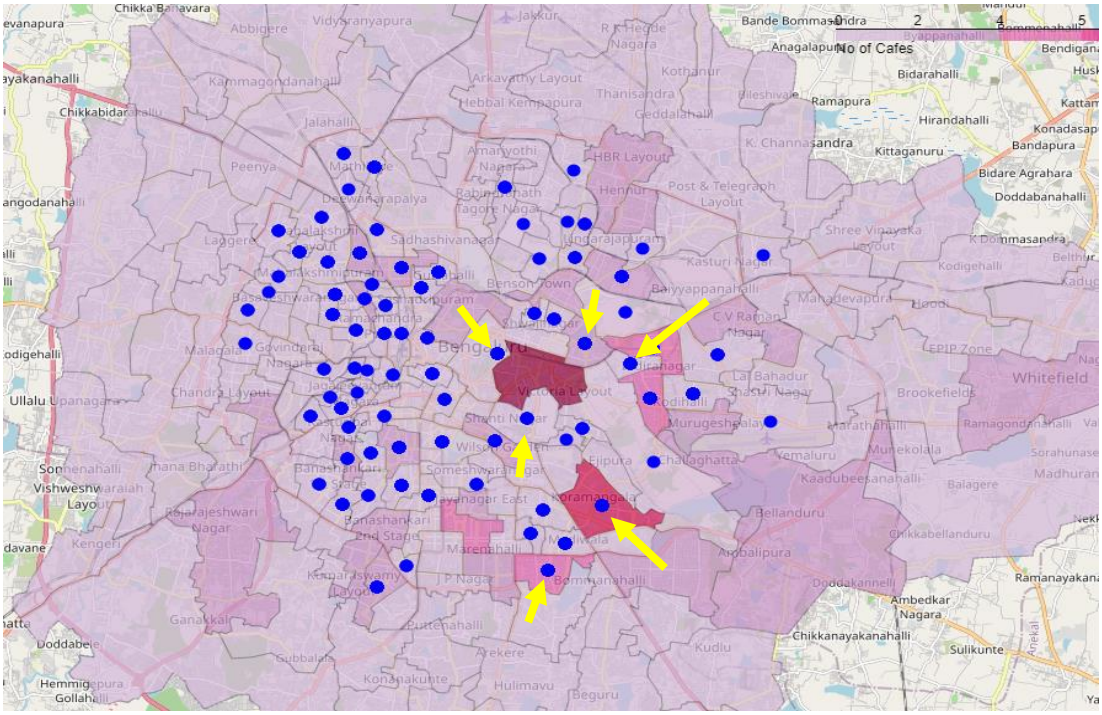
Cluster 1

	Neighborhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Azad Nagar	1	Shopping Mall	Fast Food Restaurant	Café	Multiplex	Donut Shop	Fish Market	Farmers Market	Event Space	Electronics Store
1	Rajaji Nagar	1	Indian Restaurant	Park	Chinese Restaurant	Clothing Store	Café	Burger Joint	Fast Food Restaurant	Bakery	Seafood Restaurant
2	Dharmaraya Swamy Temple	1	Theater	Market	Food Truck	South Indian Restaurant	Yoga Studio	Donut Shop	Fast Food Restaurant	Farmers Market	Event Space
3	Sunkenahalli	1	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Ice Cream Shop	Juice Bar	Department Store	Pizza Place	Sandwich Place	Bakery
4	Prakash Nagar	1	Café	Motorcycle Shop	Bakery	Yoga Studio	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space
...	...	...	...	...	...	...	...	...	...	...	...
72	Vijnanapura	1	ATM	Donut Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant
73	J P Park	1	Park	ATM	Chinese Restaurant	Donut Shop	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store

There is a total of 76 neighborhoods in this cluster, by looking at the top 10 common venues in the cluster we see cafes are popular in some wards.



We see that this cluster coincides with most populated wards in the city.



We also see that there are quite a few wards large number is cafes in this cluster. It would be wise to pick a few potential locations from this cluster close to or in wards with a few cafes already established as highlighted above.

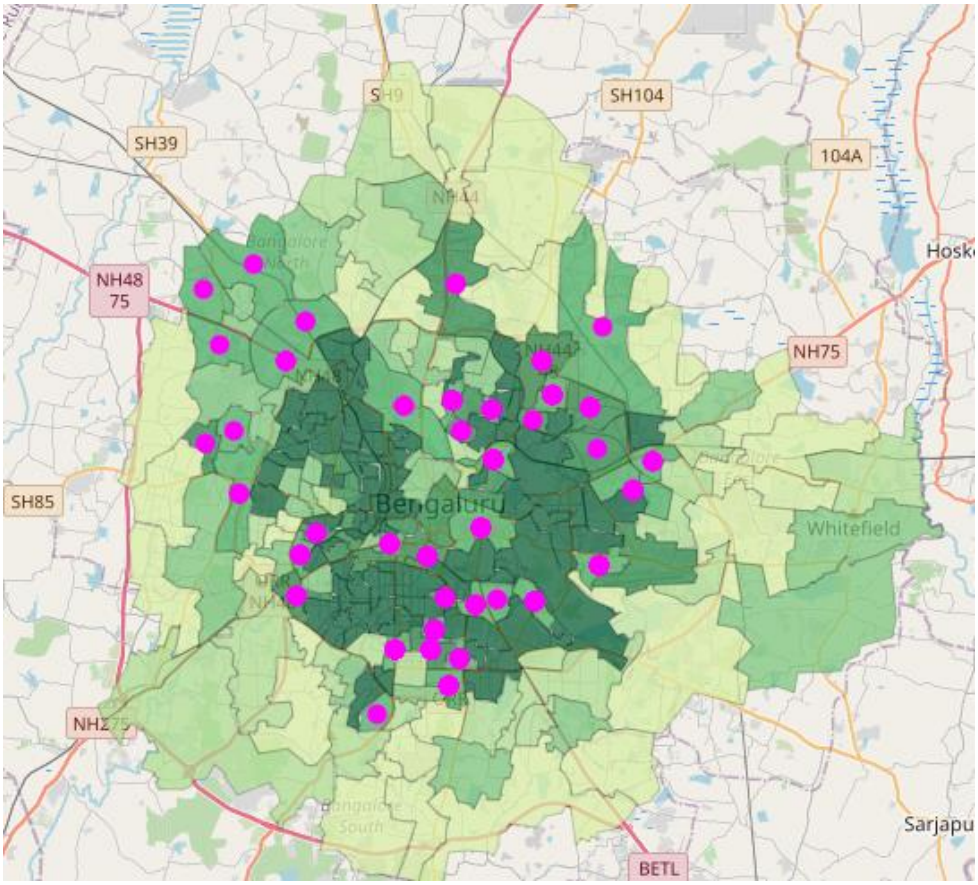
Potential Locations to consider are Koramangala, Shanthi Nagar, BTM Layout, Sampangiram Nagar, Halsoor, Domlur

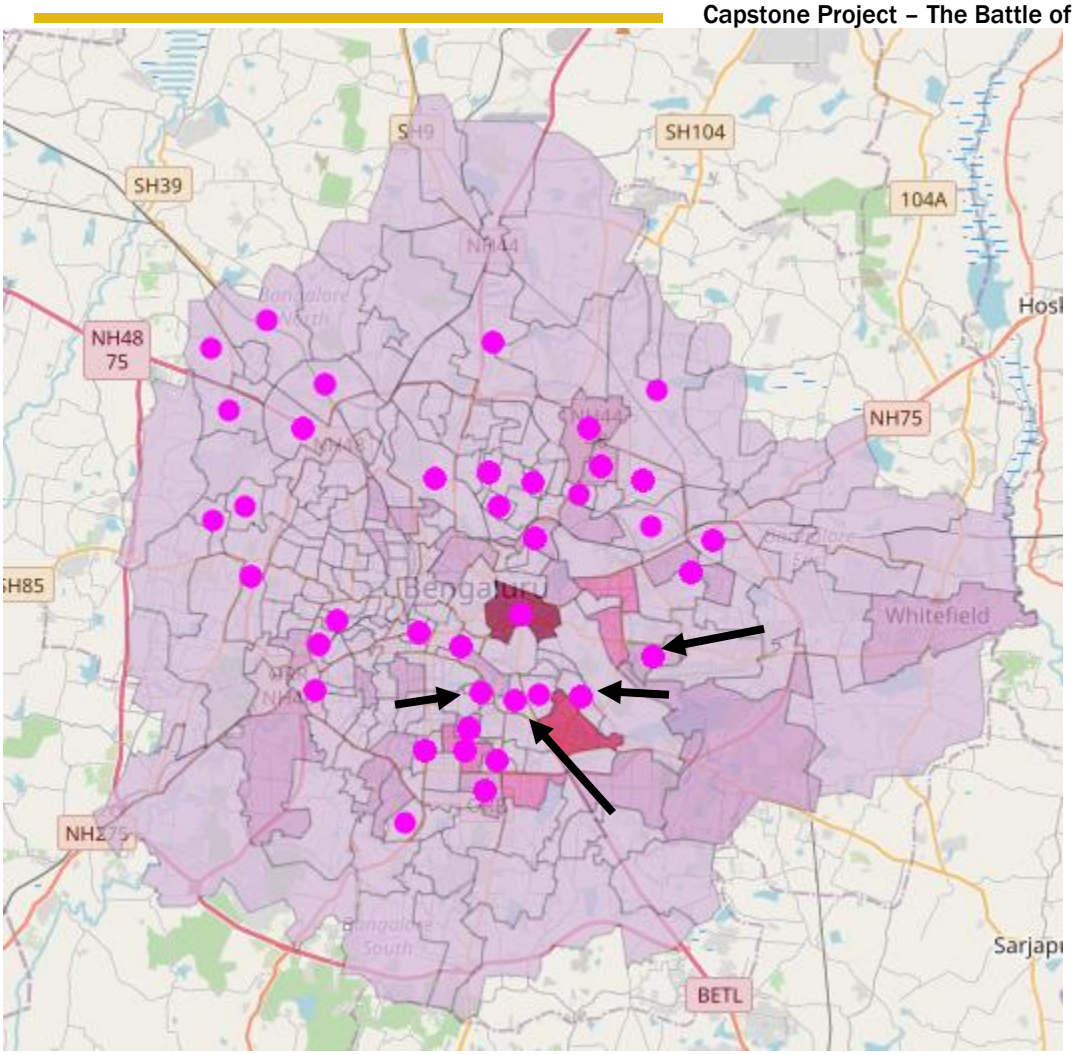
Cluster 2

The neighborhoods in cluster 2 are shown below

	Neighborhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Byatarayanapura	2	Print Shop	Restaurant	Indian Restaurant	Dog Run	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
1	Karisandra	2	Ice Cream Shop	Indian Restaurant	Italian Restaurant	Diner	Restaurant	Gym	Snack Place	South Indian Restaurant	Salad Place	Dry Cleaner
2	Jalahalli	2	Soccer Field	Basketball Court	Yoga Studio	Food	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
3	Jayanagar East	2	Indian Restaurant	Asian Restaurant	Chinese Restaurant	Radio Station	Yoga Studio	Donut Shop	Fish Market	Fast Food Restaurant	Farmers Market	Event Space
4	HMT Ward	2	Coffee Shop	Fast Food Restaurant	Train Station	Indian Restaurant	Department Store	Donut Shop	Fish Market	Farmers Market	Event Space	Electronics Store
5	Kottegepalya	2	Breakfast Spot	Indian Restaurant	Café	Yoga Studio	Dry Cleaner	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space
6	J P Nagar	2	Indian Restaurant	Café	Fast Food Restaurant	Snack Place	South Indian Restaurant	Diner	Yoga Studio	Dry Cleaner	Fish Market	Farmers Market

There are 30 neighborhoods in this cluster and again cafes are not the most common venues here either.





We see that these cluster are in the medium populated neighborhoods and close to the densely populated neighborhoods. Except one ward there are not too many cafes in the areas. We also notice that the cafes are open more towards the center of the city and in the south and south east direction.

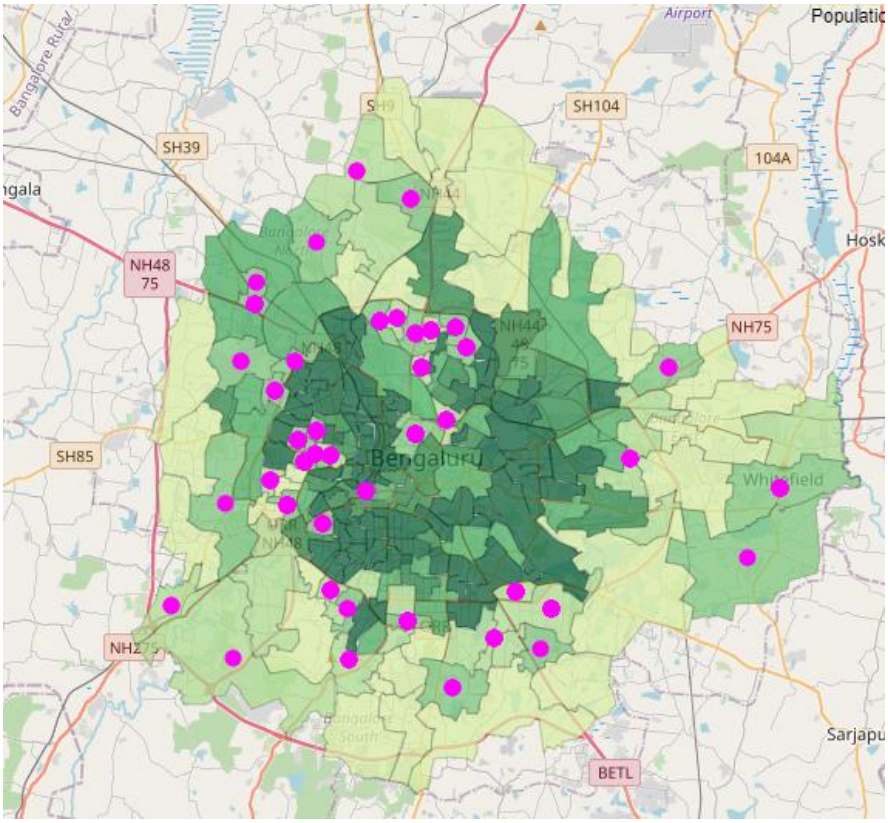
Potential Locations to consider are Ejipura, Jayanagar, Siddapura.

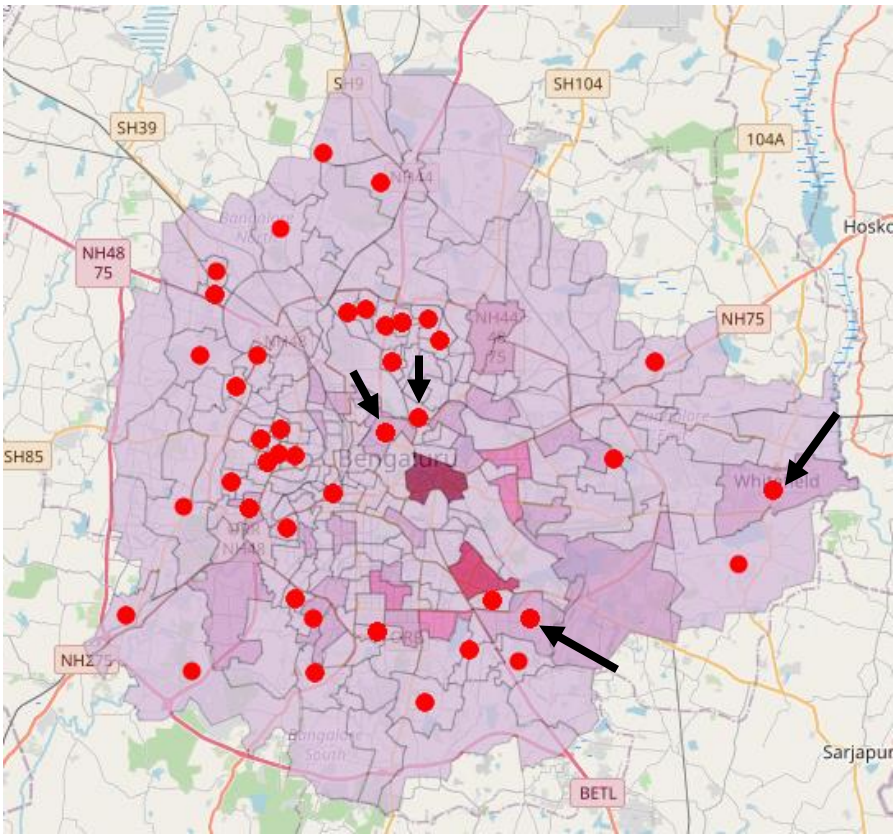
Cluster 3

The neighborhoods in cluster 3 are shown below

	Neighborhood	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	HSR Layout	3	Indian Restaurant	Café	Department Store	Gym / Fitness Center	Snack Place	Food Court	Restaurant	Badminton Court	Pizza Place	Kebab Restaurant
1	Hagadur	3	Café	Brewery	Bakery	Restaurant	Kerala Restaurant	Eastern European Restaurant	Hotel	Hotel Bar	Ice Cream Shop	Department Store
2	Gangenahalli	3	Indian Restaurant	Hotel	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Pizza Place	Dog Run	Fish Market	Farmers Market	Event Space	Electronics Store
3	T Dasarahalli	3	Indian Restaurant	Metro Station	Resort	Karnataka Restaurant	Indian Sweet Shop	Yoga Studio	Dry Cleaner	Fast Food Restaurant	Farmers Market	Event Space
4	Mallasandra	3	ATM	Donut Shop	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
5	Arakere	3	ATM	Dog Run	Badminton Court	Dry Cleaner	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
6	Bommanahalli	3	Indian Restaurant	Hotel	South Indian Restaurant	Department Store	Donut Shop	Fish Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store

There are 38 neighborhoods in this cluster and again cafes are common is some of the areas. Let's analyze them geographically.





We see that this cluster looks very similar to the first segment where the locations are towards the outskirts of the cities, however the difference here is the population density and café population. These areas have a higher population and some places have high café count. Some of the locations in this cluster like Whitefield, HSR layout, Vasanth Nagar and Jayamahal are potential spots for a new café.

Discussion

We have observed that cluster 1, 2 and 3 have potential locations. Cluster 1 and 2 locations are densely populated and located closer to the heart of the city and also more competition from existing cafes around. But these locations are a safe bet as generally we see that cafes are not the most common venues in this location. However, a couple of challenges that we can foresee in choosing a location close to the center of the city is the commercial property rental and traffic. Rent tend to get higher in populous location and also as these locations are dense there is high traffic. Cluster 3 locations are not too populated but are still ever growing and a few cafes are already established, and rents could be lower and traffic better. I would recommend a location in cluster 2 and 3 as the most appropriate for a café establishment. Location recommendations – Whitefield, HSR Layout, Koramangala, BTM Layout.

Conclusion

Cafes are a retreat one can escape from the daily rut or a peaceful place to connect and get creative. A lot of factors impact the selection of a location for a café. The café could be part of a franchise or an independent business. The entrepreneur could be looking at a busy high footfall location or a more laidback environment. We have used scientific methods using data science using a couple of factors to help entrepreneur pick at optimum location.