Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

In this project we will try to find a location for a restaurant. This report will provide recommendation to the stakeholders interested in opening a **restaurant** in **Bangalore**, India.

As part of this analysis, I will try to find **locations with low density of restaurants** and which are **close to city**.

we will utilize data science to find few neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of problem, factors that will influence our decission are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- distance of neighborhood from city

We have decided to use regularly spaced grid of locations, centered around city center, to define neighborhoods.

Following data sources were used to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **geopy geocoder API**
- number of restaurants and location in every neighborhood will be obtained using **Foursquare API**

- coordinate of Bangalore city will be obtained using **geopy geocoder API** of well known location (Koramangala)

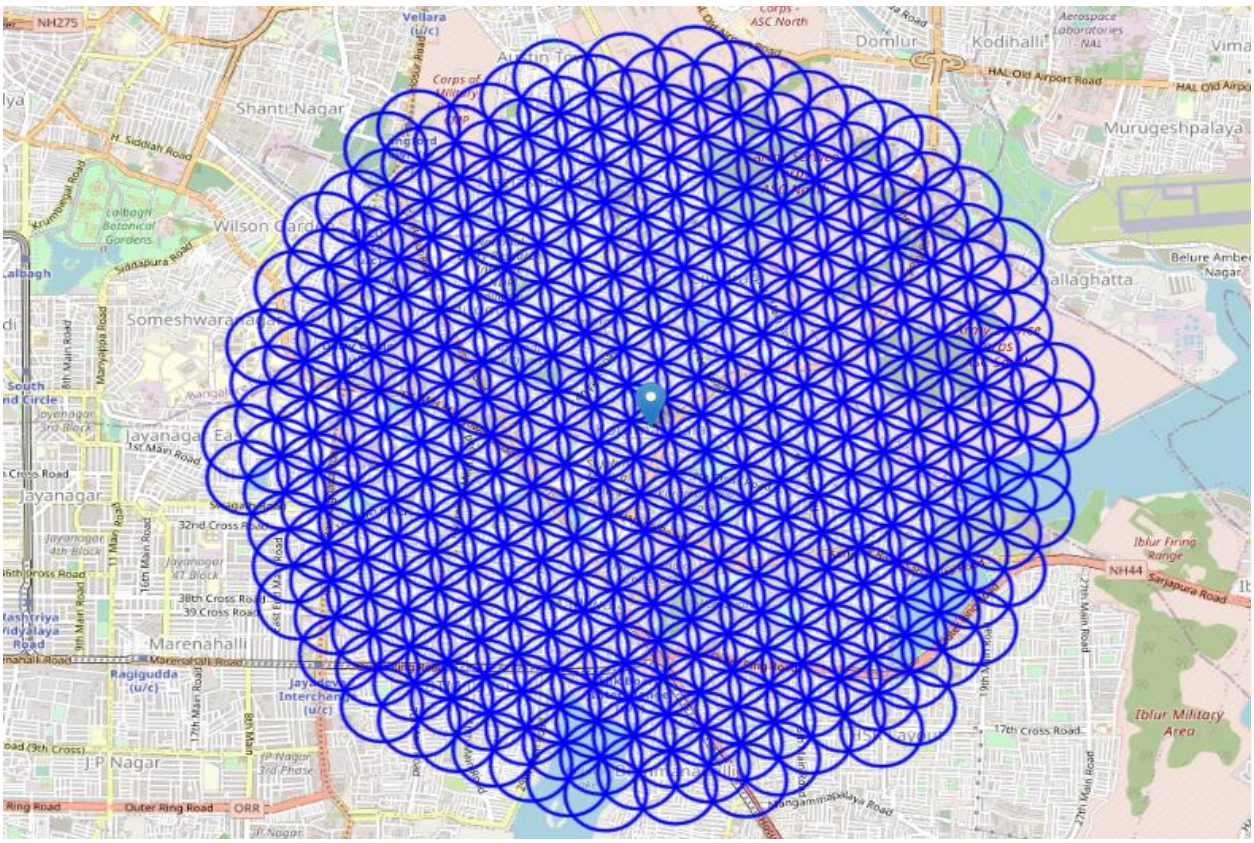
Data Acquisition and Cleaning

We created latitude & longitude coordinates for centroids of our candidate neighborhoods. We created a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Bangalore city center.

We first found the latitude & longitude of Bangalore city, using specific, well known address and geopy geocoder API.

Then we created a grid of area candidates, equally spaced, centered around city center and within ~6km from Koramangala. Neighborhoods were defined as circular areas with a radius of 300 meters, so our neighborhood centers were 600 meters apart.

To accurately calculate distances we needed to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we projected those coordinates back to latitude/longitude degrees to be shown on Folium map.
After that we created a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.



After that we used Geopy geaocoder API to get approximate addresses of those locations.

	Address	Latitude	Longitude	X	Y	Distance from center
0	6th Cross Road, MICO Layout, BTM Layout Ward, ...	12.913728	77.604103	8.895760e+06	2.955396e+06	5992.495307
1	Sri Devi Baker's, 10th Cross Road, BTM Layout ...	12.912640	77.606624	8.896360e+06	2.955396e+06	5840.376700
2	IAS Officers Colony, BTM Layout Ward, South Zo...	12.911553	77.609144	8.896960e+06	2.955396e+06	5747.173218
3	IAS Officers Colony, BTM Layout Ward, South Zo...	12.910466	77.611664	8.897560e+06	2.955396e+06	5715.767665
4	29th Main Road, Mahadeshwara Nagara, BTM Layou...	12.909379	77.614184	8.898160e+06	2.955396e+06	5747.173218
5	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.908292	77.616704	8.898760e+06	2.955396e+06	5840.376700
6	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.907206	77.619223	8.899360e+06	2.955396e+06	5992.495307
7	4th Main Road, Gurappana Palya, Jayanagar East...	12.917500	77.601282	8.894860e+06	2.955915e+06	5855.766389
8	BTM Layout 2nd Stage, BTM Layout Ward, South Z...	12.916413	77.603803	8.895460e+06	2.955915e+06	5604.462508
9	8th Main Road, BTM Layout 2nd Stage, BTM Layou...	12.915325	77.606323	8.896060e+06	2.955915e+06	5408.326913

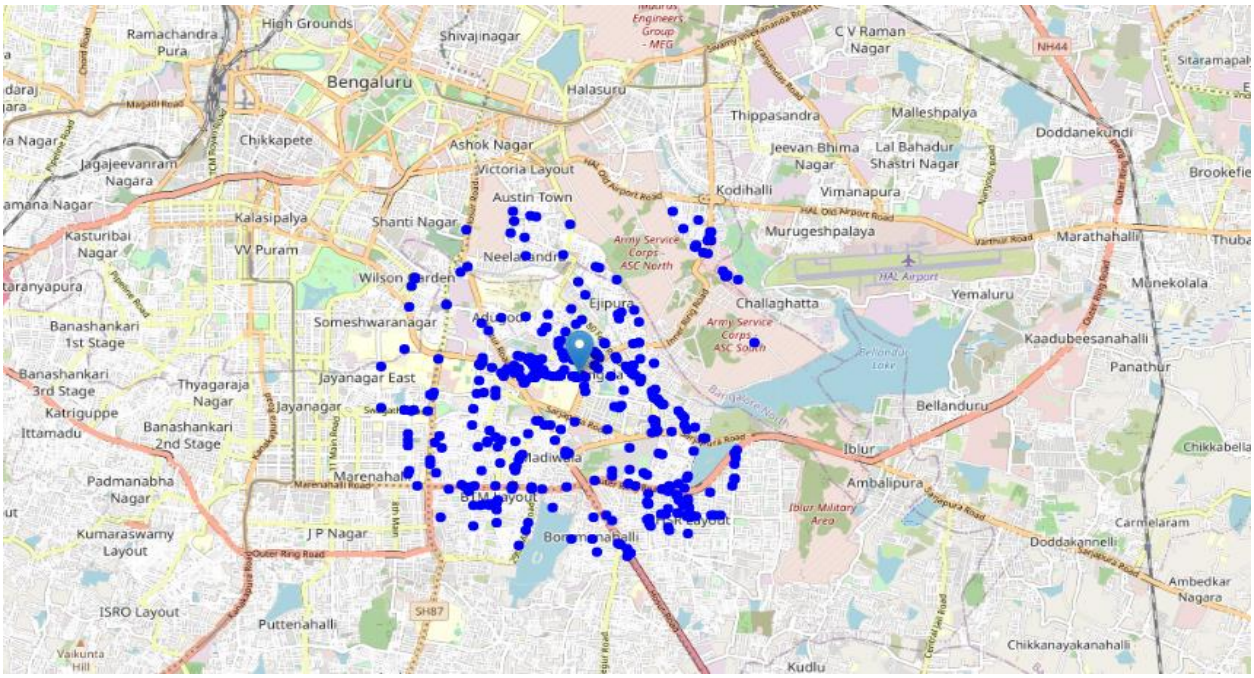
Now that we have our location candidates, we used Foursquare API to get info on restaurants in each neighborhood.

Since we were interested in venues in 'food' category, We included in our list only venues that have 'restaurant' in category name.

Following is a sample of Restaurants in each of the candidate location

Restaurants around location

Restaurants around location 2: Flavours of China, Balaji's Veg, Momoz, The Royal Tandoori, Namaste, Golmaal Paratha
Restaurants around location 3: Flavours of China, Balaji's Veg, Momoz, Beijing Bites, Shaan-e-Punjab, Mast Kalandar, Golmaal Paratha, Swadishta Aahar
Restaurants around location 4: Mast Kalandar, Kacha Papad
Restaurants around location 5: Tongue Mridang, Kacha Papad, Biryani Bowl
Restaurants around location 6:
Restaurants around location 7: gugabana
Restaurants around location 8: Udupi Gokul Cafe
Restaurants around location 9: Punjabi Sanjhaa Chullah, Udupi Gokul Cafe, Narmadha's Hyderabad Biriyani, Emerald Restaurant
Restaurants around location 10: Flavours of China, Balaji's Veg, Punjabi Sanjhaa Chullah, Narmadha's Hyderabad Biriyani, Daana Paani, The Royal Tandoori, Emerald Restaurant, The Barbeques



Now we had all the restaurants in area within few kilometers from Koramangala. We also know which restaurants exactly are in vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new restaurant!

Methodology

we directed our efforts on detecting areas of Bangalore that have low restaurant density. We limited our analysis to area ~6km around center.

In data gathering step we had collected the required **data: location of every restaurant within 6km from Bangalore center** (Koramangala).

Second step in our analysis we calculated and explored '**restaurant density**' across different areas of Bangalore - we used **heatmaps** to identify a few promising areas close to center with low number of restaurants and focussed our attention on those areas.

In third and final step we focussed on most promising areas and within those created **clusters of locations that meet some basic requirements** established in discussion with stakeholders:

we took into consideration locations with **no more than two restaurants in radius of 250 meters**. We presented map of all such locations but also created clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and searched for optimal venue location by stakeholders.

Analysis

We performed some basic explanatory data analysis and derived some additional info from our raw data. such as count the **number of restaurants in every area candidate**.
Following is a sample

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	6th Cross Road, MICO Layout, BTM Layout Ward, ...	12.913728	77.604103	8.895760e+06	2.955396e+06	5992.495307	1
1	Sri Devi Baker's, 10th Cross Road, BTM Layout ...	12.912640	77.606624	8.896360e+06	2.955396e+06	5840.376700	6
2	IAS Officers Colony, BTM Layout Ward, South Zo...	12.911553	77.609144	8.896960e+06	2.955396e+06	5747.173218	8

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
3	IAS Officers Colony, BTM Layout Ward, South Zo...	12.910466	77.611664	8.897560e+06	2.955396e+06	5715.767665	2
4	29th Main Road, Mahadeshwara Nagara, BTM Layou...	12.909379	77.614184	8.898160e+06	2.955396e+06	5747.173218	3
5	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.908292	77.616704	8.898760e+06	2.955396e+06	5840.376700	0
6	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.907206	77.619223	8.899360e+06	2.955396e+06	5992.495307	1
7	4th Main Road, Gurappana Palya, Jayanagar East...	12.917500	77.601282	8.894860e+06	2.955915e+06	5855.766389	1
8	BTM Layout 2nd Stage, BTM Layout Ward, South Z...	12.916413	77.603803	8.895460e+06	2.955915e+06	5604.462508	4
9	8th Main Road, BTM Layout 2nd Stage, BTM Layou...	12.915325	77.606323	8.896060e+06	2.955915e+06	5408.326913	9

After that calculated the **distance to nearest restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it was).

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to restaurants
0	6th Cross Road, MICO Layout, BTM Layout Ward, ...	12.913728	77.604103	8.895760e+06	2.955396e+06	5992.495307	1	526.124573

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to restaurants
1	Sri Devi Baker's, 10th Cross Road, BTM Layout ...	12.912640	77.606624	8.896360e+06	2.955396e+06	5840.376700	6	237.878044
2	IAS Officers Colony, BTM Layout Ward, South Zo...	12.911553	77.609144	8.896960e+06	2.955396e+06	5747.173218	8	223.577744
3	IAS Officers Colony, BTM Layout Ward, South Zo...	12.910466	77.611664	8.897560e+06	2.955396e+06	5715.767665	2	382.827769
4	29th Main Road, Mahadeshwara Nagara, BTM Layou...	12.909379	77.614184	8.898160e+06	2.955396e+06	5747.173218	3	96.581417
5	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.908292	77.616704	8.898760e+06	2.955396e+06	5840.376700	0	635.228301
6	Madivala Lake, Kodichikkanahalli Road, Vakil M...	12.907206	77.619223	8.899360e+06	2.955396e+06	5992.495307	1	493.203597
7	4th Main Road, Gurappana Palya, Jayanagar East...	12.917500	77.601282	8.894860e+06	2.955915e+06	5855.766389	1	272.299559
8	BTM Layout 2nd Stage, BTM Layout Ward, South Z...	12.916413	77.603803	8.895460e+06	2.955915e+06	5604.462508	4	168.612974

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to restaurants
9	8th Main Road, BTM Layout 2nd Stage, BTM Layou...	12.915325	77.606323	8.896060e+06	2.955915e+06	5408.326913	9	145.353117

Then we created a map showing **heatmap / density of restaurants** and tried to extract some meaningful info from that. Also, we showed a few circles indicating distance of 1km, 2km and 3km from Koramangala.

Based on this we focussed our analysis on areas *south-west, south, south-east and east from Bangalore center* - we moved the center of our area of interest and reduced it's size to have a radius of **2.5km**.

Above map showed all the pockets of low restaurant density in closest to Bangalore City center.

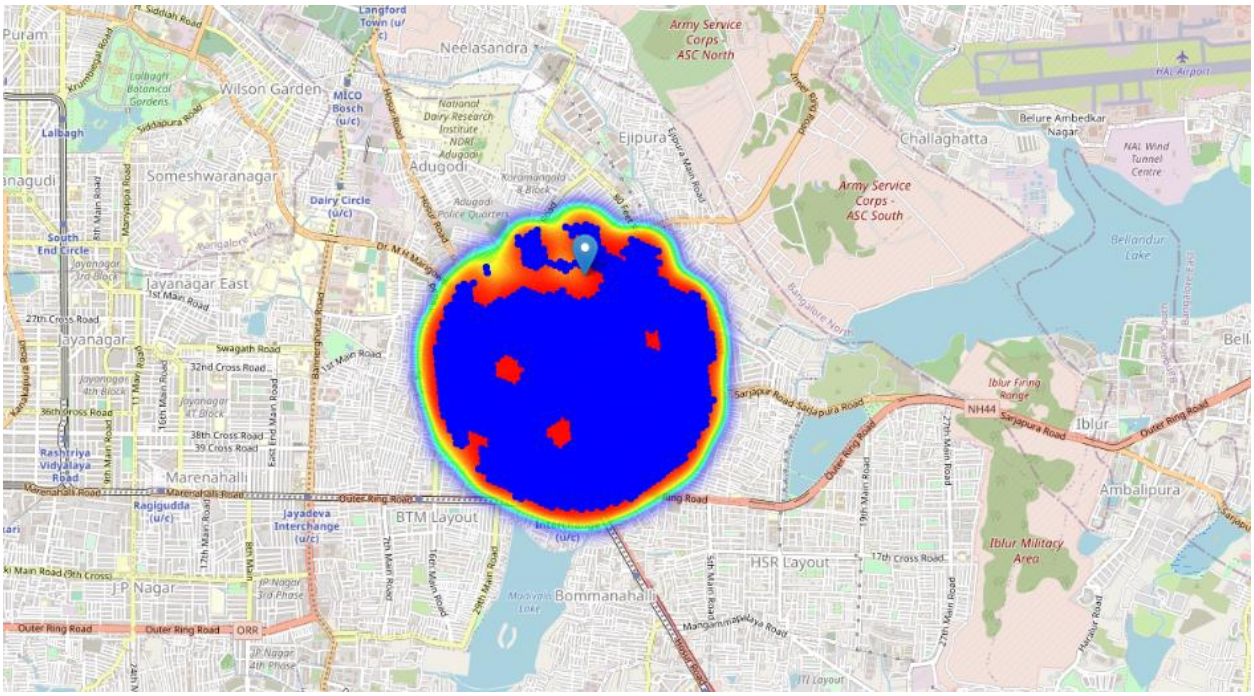
We also created new, more dense grid of location candidates restricted to our new region of interest

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to restaurant
0	12.916719	77.616723	8.898010e+06	2.957111e+06	1	114.966556
1	12.916538	77.617143	8.898110e+06	2.957111e+06	1	117.038991
2	12.918073	77.614574	8.897460e+06	2.957198e+06	0	359.662992
3	12.917892	77.614994	8.897560e+06	2.957198e+06	0	302.901380
4	12.917710	77.615414	8.897660e+06	2.957198e+06	0	272.288494
5	12.917529	77.615834	8.897760e+06	2.957198e+06	0	276.645626
6	12.917348	77.616254	8.897860e+06	2.957198e+06	0	274.995012

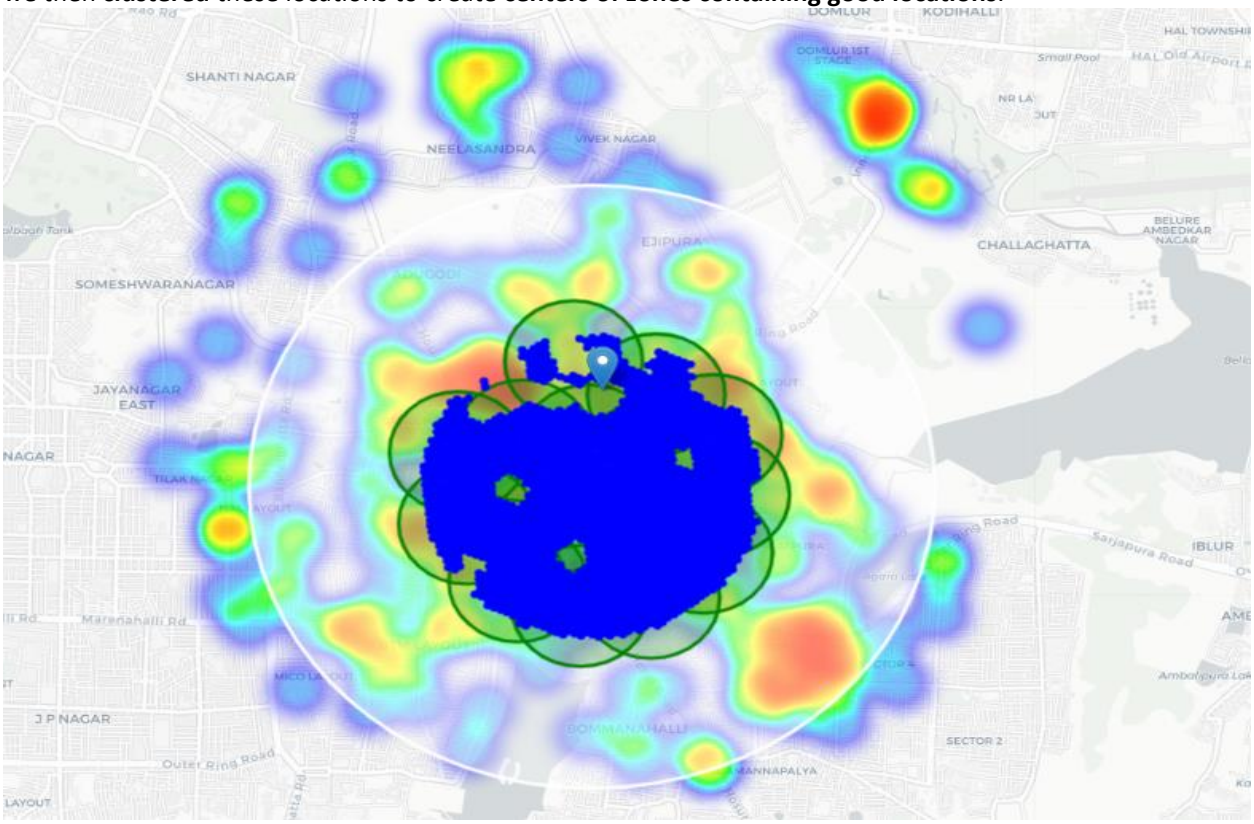
	Latitude	Longitude	X	Y	Restaurants nearby	Distance to restaurant
7	12.917167	77.616674	8.897960e+06	2.957198e+06	1	214.716262
8	12.916985	77.617093	8.898060e+06	2.957198e+06	1	191.269155
9	12.916804	77.617513	8.898160e+06	2.957198e+06	1	216.944016

Now we **filtered** these locations to show **locations with no more than two restaurants in radius of 250 meters** and showed these locations on a map.

We now had a bunch of locations close to Koramangala, and we knew that each of those locations have no more than two restaurants in radius of 250m. These locations were potential candidates for a new restaurant

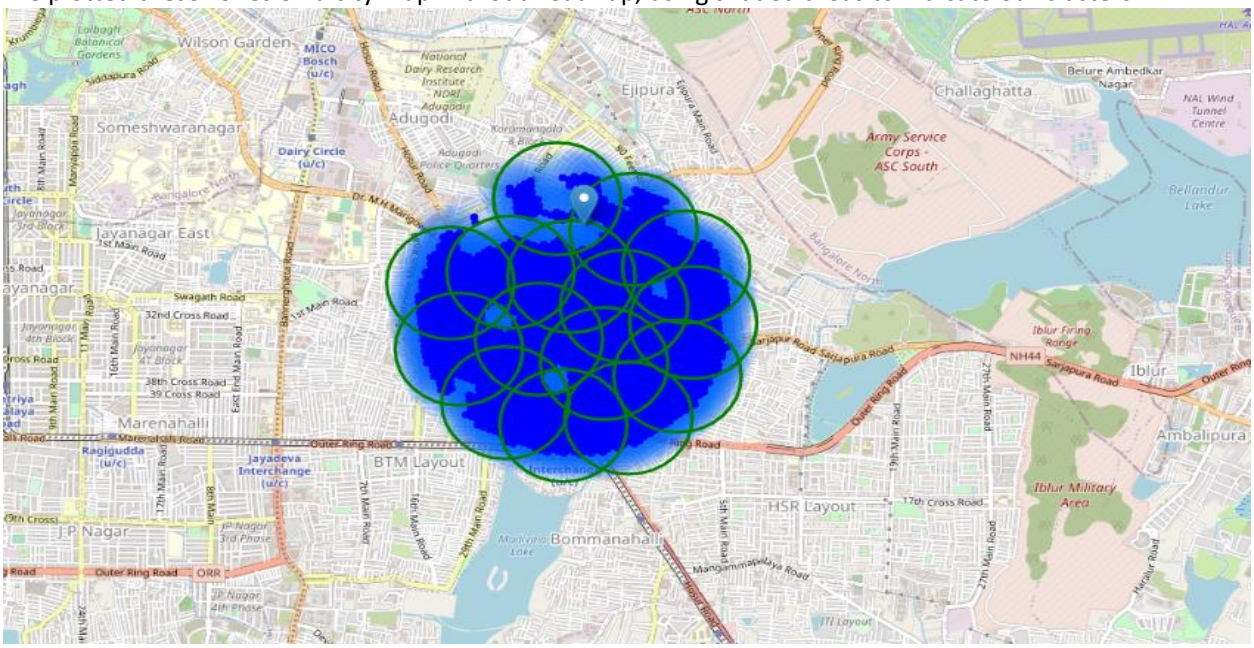


we then **clustered** these locations to create **centers of zones containing good locations**.



Addresses of these cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

We plotted these zones on a city map without heatmap, using shaded areas to indicate our clusters:



Finally, we had **the addresses of these candidate centers using reverse geocode** which could be presented to stakeholders.

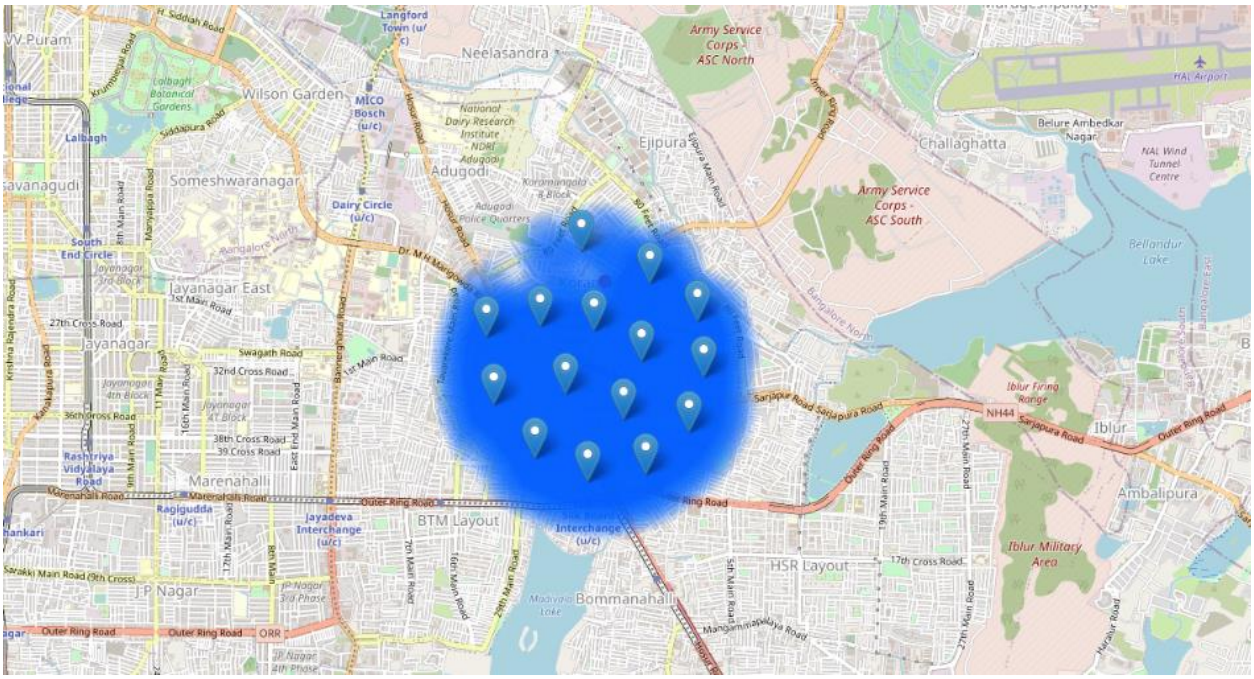
==========

Addresses of centers of areas recommended for further analysis

==========

St Johns Medical College, Sarjapur Road, Jakkasandra, South Zone, Bengaluru 560034 => 1.4km from Koramangala
Madiwala Sarjapura Road, Jakkasandra, South Zone, Bengaluru - 560034 => 2.5km from Koramangala
2nd Cross, 4th C block, Maistripalaya, Koramangala, South Zone, Bengaluru 5560034 => 1.8km from Koramangala
Maruthi Nagara, Madivala, South Zone, Bengaluru - 560034 => 3.4km from Koramangala
Maistripalaya, Koramangala, South Zone, Bengaluru 560095 => 0.8km from Koramangala
Krupanidhi College, Sarjapur Road, Jakkasandra, South Zone, Bengaluru 5560034 => 3.1km from Koramangala
Jakkasandra, South Zone, Bengaluru - 560034 => 3.5km from Koramangala
St Johns Medical College, 4th Cross Road, Suddagunte Palya Ward, South Zone, Bengaluru - 560034 => 2.3km from Koramangala
Indian Institute of Astrophysics, Mahayogi Vemana Road, Koramangala 2nd Block, Koramangala, South Zone, Bengaluru - 560034 => 2.1km from Koramangala
St Johns Medical College, Sarjapur Road, Jakkasandra, South Zone, Bengaluru 560034 => 0.9km from Koramangala
Swadist, 4th A Cross Road, Maistripalaya, Koramangala, South Zone, Bengaluru 560095 => 0.7km from Koramangala
Venkatpura, Koramangala, South Zone, Bengaluru 5560034 => 2.4km from Koramangala
3rd Cross Road, Koramangala 2nd Block, Koramangala, South Zone, Bengaluru 5560034 => 1.6km from Koramangala
Suddagunte Palya Ward, South Zone, Bengaluru - 560034 => 2.9km from Koramangala
Maruthi Nagara, Madivala, South Zone, Bengaluru - 560034 => 3.6km from Koramangala

This concludeed our analysis. We had created 15 addresses representing centers of zones containing locations with low number of restaurants, all zones being close to Bangalore center (all less than 4km from Koramangala, and about half of those less than 2km from Koramangala). Although zones were shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.



Results and Discussion

Our analysis shows that although there is a great number of restaurants in Bangalore, there are pockets of low restaurant density close to city center. Highest concentration of restaurants was detected north and west from Koramangala, so we focused our attention to areas south, south-east and east.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Koramangala) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually

optimal locations for a new restaurant. Purpose of this analysis was to only provide info on areas close to Bangalore center but not crowded with existing restaurants - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Bangalore areas close to center with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general areas that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.