Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Introduction: Business Problem **1**

Data **1**

Neighborhood Candidates 2

Foursquare 3

Methodology **3**

Analysis **3**

Results and Discussion **10**

Conclusion **10**

Introduction: Business Problem

In this project we worked towards finding an optimal location for a breakfast point. Specifically, this report will be targeted to stakeholders interested in opening a **breakfast spot** near richmond circle in Bangalore. Since there are lots of restaurants near richmond circle, we detected **locations that are not already crowded with restaurants and eating joints**. We were also particularly interested in **areas with no breakfast spots in the vicinity**. We preferred locations **as close to richmond circle as possible**, assuming that first two conditions are met. We used our data science powers to generate a few most promising 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 our problem, factors that influenced our decisions are:

- Number of existing breakfast spots in the neighborhood

- Number of and distance to breakfast spots in the neighborhood, if any
- Distance of neighborhood from richmond circle

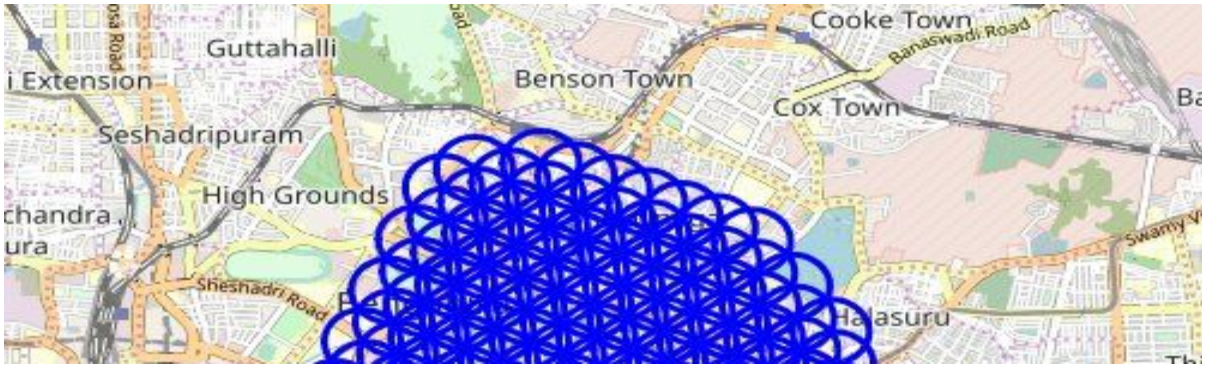
We decided to use regularly spaced circular grids of locations, centered around Richmond Circle center, to define our neighborhoods. Following data sources were needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of breakfast spots and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of center will be obtained using **Google Maps API geocoding** of well known Richmond Town in Bangalore.

Neighborhood Candidates

We created latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approx. 12x12 kilometers centered around Richmond Circle, Bangalore. We found the latitude & longitude of Richmond Circle, using specific, well known addresses and Google Maps geocoding API. We created a grid of area candidates, equally spaced, centered around city center and within ~3km from Richmond Circle. Our neighborhoods were defined as circular areas with a radius of 300 meters, so our neighborhood centers were 600 meters apart. Reason of choosing 3 km vicinity was because of the limit on calls that can be made to FourSquare API for number of generated neighborhoods. Only 950 calls / day can be done to Foursquare Api using a free developers account. 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. We created functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters). 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**. we now had the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors is exactly the same) and within ~3km from Richmond Circle.

Then we used Google Maps API to get approximate addresses of those locations.



Foursquare

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

We were interested in venues in 'nightlife' category, 'shop and services' category and 'breakfast spot' category. So we included in out list only venues that have these in category name. Finding out the number of spots for all these mentioned categories helped in instilling more confidence in stakeholders regarding the choice of business they want to start with. So now we have all the breakfast spots in area within few kilometers from Richmond Circle, and we know their number! We also know which breakfast points 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 breakfast spot/point near Richmond Circle.

Methodology

In this project we directed our efforts on detecting areas of Bengaluru that have low breakfast spot density. We will limit our analysis to area ~3km around richmond circle.

In first step we collected the required **data: location and type (category) of every nightlife and service spots within 3km from Richmond Circle** (Under the Fly Over). We have also **identified breakfast spots** (according to Foursquare categorization).

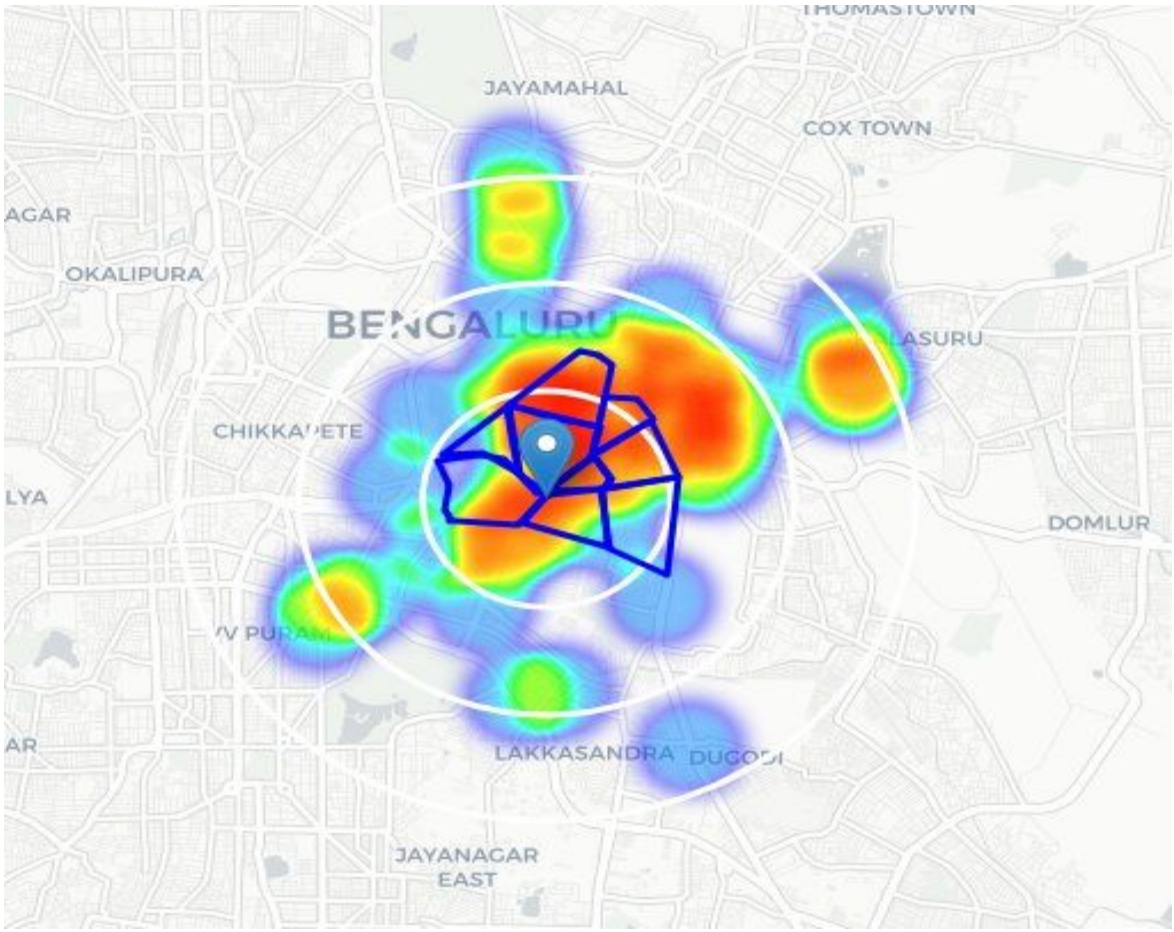
Second step in our analysis we calculated and exploration of '**breakfast spot density**' across different areas of Bengaluru - we used **heatmaps** to identify a few promising areas close to center with low number of breakfast spot in general (*and* less breakfast spot in vicinity) and focus our attention on those areas.

In third and final step we focused 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 breakfast spot in radius of 250 meters**, and we wanted locations **without breakfast spot in radius of 400 meters**. We present map of all such locations but also create 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 search for optimal venue location by stakeholders.

Analysis

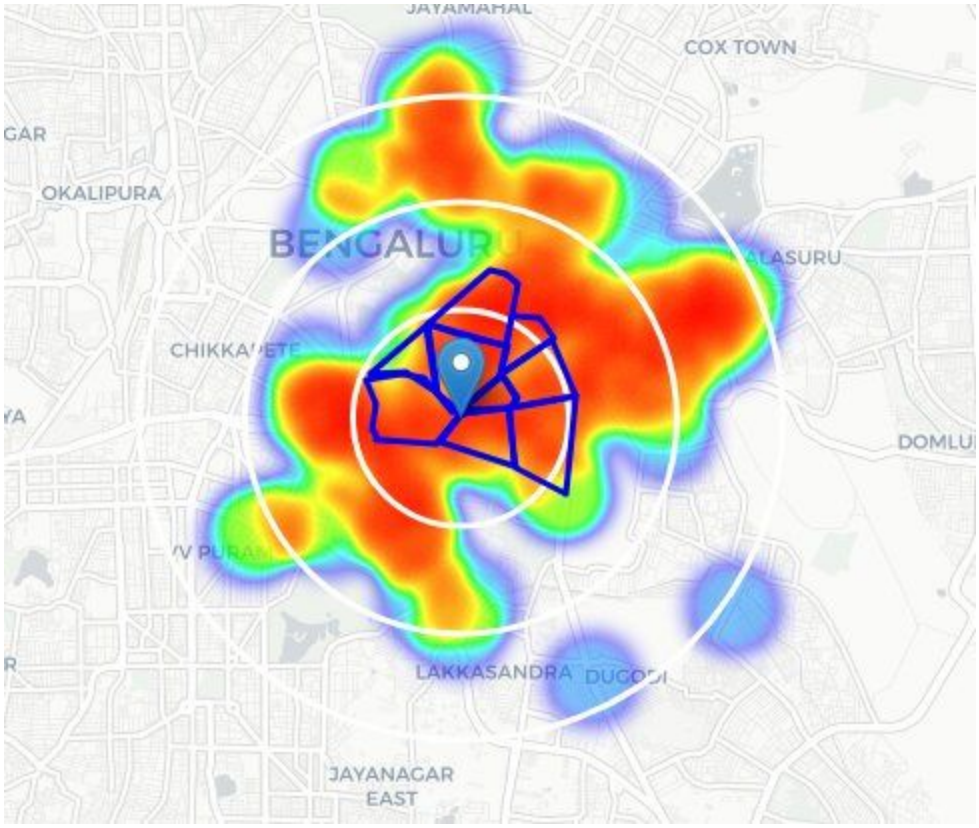
First we performed some basic exploratory data analysis and derived some additional info from our raw data. We counted the **number of breakfast spot, nightlife spots and shop and services spots in every area candidate**. Then we calculated the **distance to nearest breakfast spot from every area candidate center** (not only those within 300m - we wanted distance to closest one, regardless of how distant it is). so **on average Breakfast spot could be found within ~600m** from

every area center candidate. That's fairly close, so we needed to filter our areas carefully! We created a map showing **heatmap / density of breakfast spots** and extracted some meaningful info from that. Also, we show **borders of Richmond boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Richmond Circle.



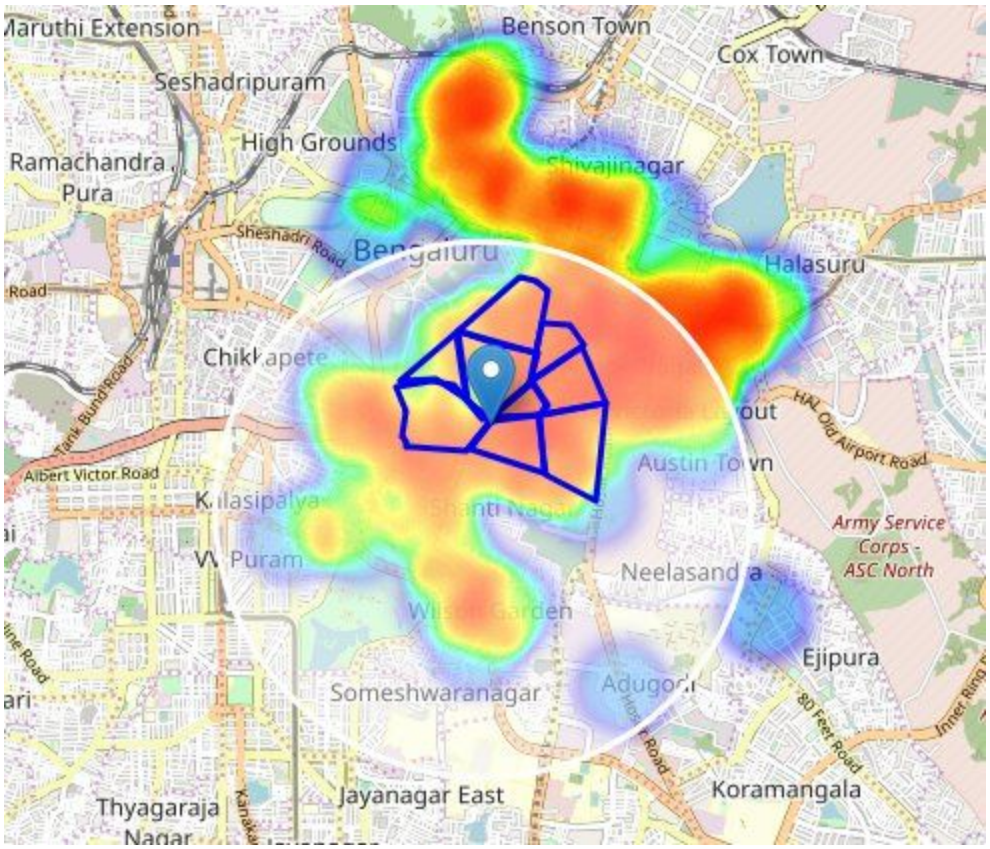
As it is visible in the map above a few pockets of low nightlife spots density closest to city center can be found **south, south-east and west from Richmond Circle**.

on average Breakfast spot was found within ~600m from every area center candidate. That's fairly close, so we needed to filter our areas carefully! We created a map showing **heatmap / density of breakfast spots** and tried to extract some meaningful info from that. Also, we show **borders of Richmond boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Richmond circle. Let's create another heatmap map showing **heatmap/density of breakfast spot** only.

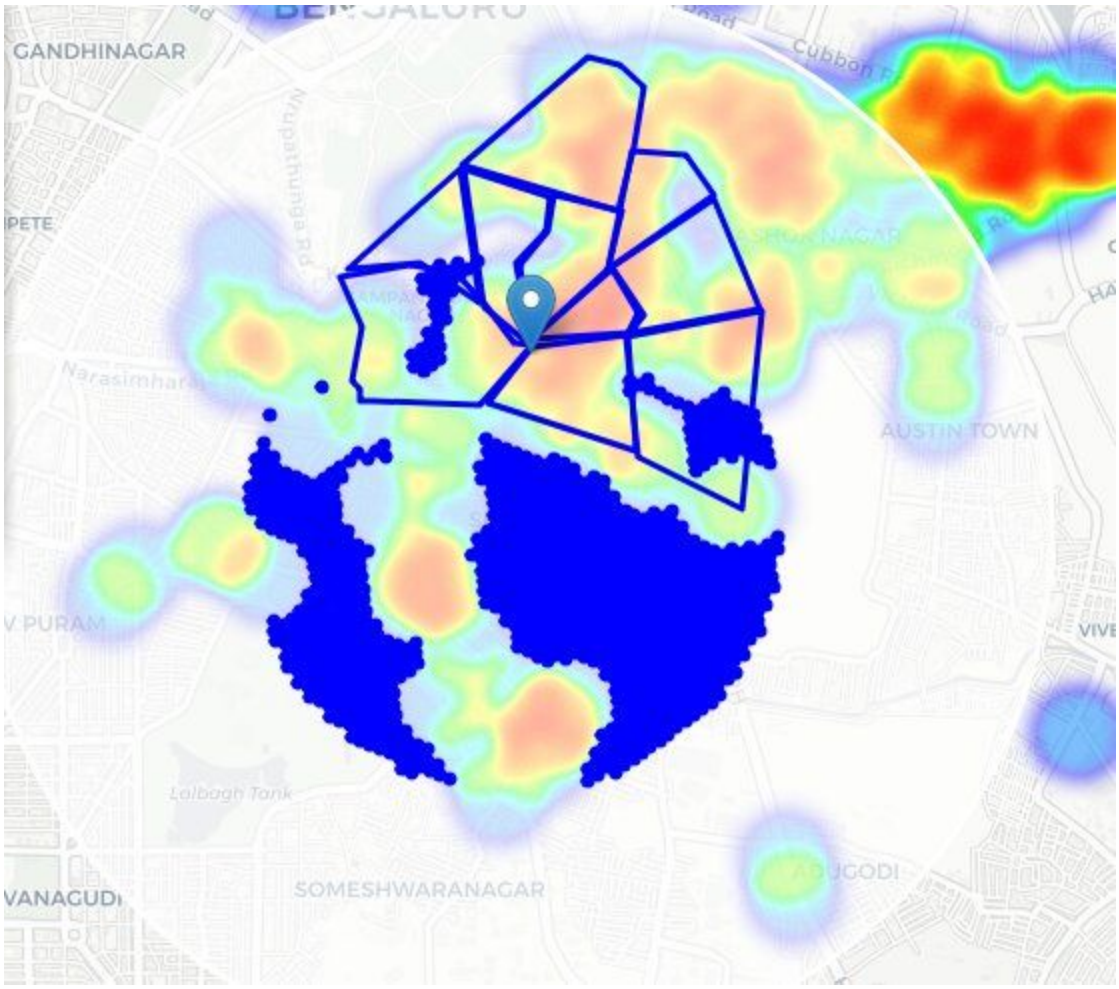


This map is so 'hot' (breakfast spot are quite densely distributed near RIchmond Circle) it also indicates higher density of existing breakfast spot directly north and west from Richmond Circle, with closest pockets of **low breakfast spot density positioned east, south-east and south from richmond circle**.

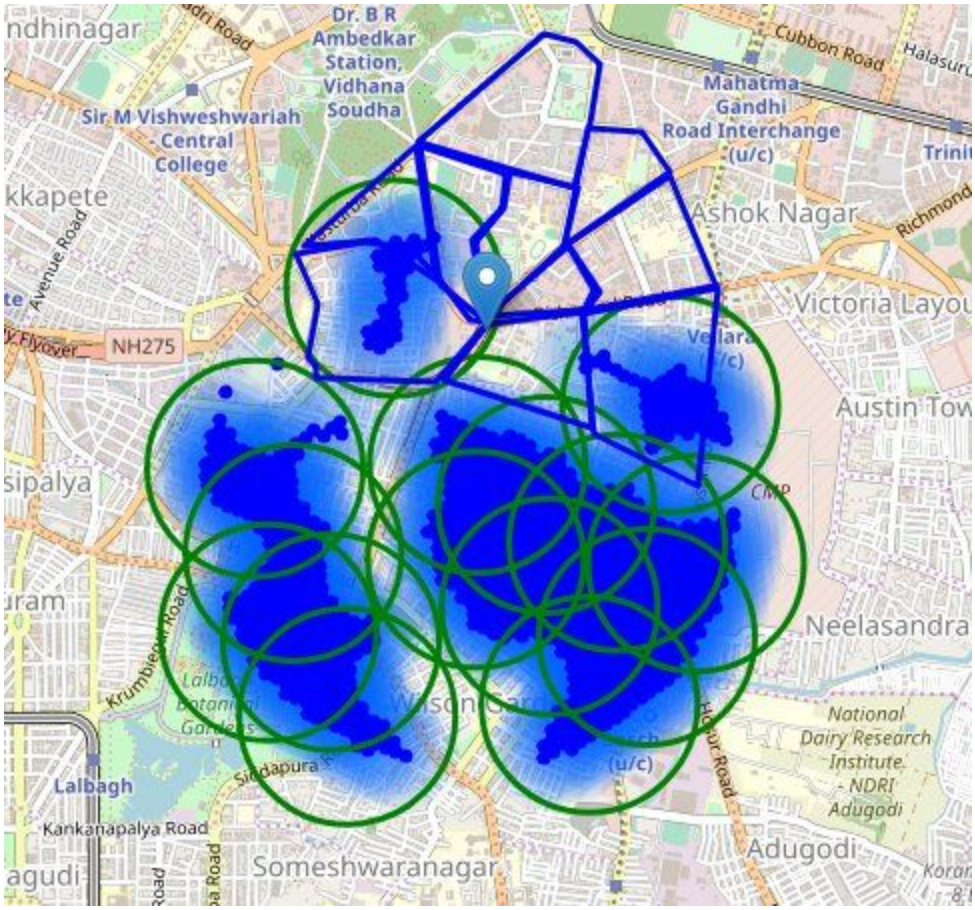
Based on this we will now focus our analysis on areas *south-west, south, south-east and east from Richmond Circle center* - we moved the center of our area of interest and reduced its size to have a radius of **2.5km**. This places our location candidates mostly in boroughs with large low restaurant density south and south west from richmond circle, however this borough was less interesting to stakeholders as it's mostly residential and less popular with tourists). Not bad - this nicely covers all the pockets of low breakfast spot density in nearby boroughs. Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m apart).



We calculated two most important things for each location candidate: **number of breakfast spot in vicinity** (we'll use radius of **250 meters**) and **distance to closest breakfast spot**. The we **filter** those locations: we were interested only in **locations with no more than two breakfast spot in radius of 250 meters**, and **no breakfast spot in radius of 400 meters**. Let's see how this looks on a map. Looking good. We now have a bunch of locations fairly close to Richmond Circle (mostly in south and south east of Richmond borough), and we know that each of those locations has no more than two restaurants in radius of 250m, and no breakfast spot closer than 400m. Any of those locations is a potential candidate for a new breakfast spot, at least based on nearby competition.



What we now have is a clear indication of zones with low number of breakfast spots in the vicinity. Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis. Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates. Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics. Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:



Finally, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

==========
Addresses of centers of areas recommended for further analysis
==========

10/2, Alfred St, Richmond Town, Bengaluru, Karnataka 560025 => 1.8km from Richmond Circle

Hosur Main Road, Sudhama Nagar, Bengaluru, Karnataka 560004 => 3.2km from Richmond Circle
147, 2nd Main Rd, Jalakanteshwara Nagar, Shanti Nagar, Bengaluru, Karnataka 560027 => 2.6km from Richmond Circle
Laxmi Road, Shantinagar => 1.2km from Richmond Circle
77, SUPREM KUTIR, 5th Cross Rd, Bheemanna Garden, Ayappa Garden, Shanti Nagar, Bengaluru, Karnataka 560027 => 1.7km from Richmond Circle
9, 1st Cross Rd, K.S. Garden, Sudhama Nagar, Bengaluru, Karnataka 560027 => 2.5km from Richmond Circle
27/1, Hosur Rd, Langford Town, Shanti Nagar, Bengaluru, Karnataka 560025 => 2.9km from Richmond Circle
13, 5th Cross Rd, Sudhama Nagar, Bengaluru, Karnataka 560027 => 2.7km from Richmond Circle

16, Muniyappa Rd, Sampangi Rama Nagara, Sampangi Rama Nagar, Bengaluru, Karnataka 560027 => 0.9km from Richmond Circle
KSRTC Bus Depot 2, 4th Cross Lakshmi Rd, Shanti Nagar, Bengaluru, Karnataka 560027 => 2.1km from Richmond Circle
81A, BTS Main Rd, Chinnayanpalya, Wilson Garden, Bengaluru, Karnataka 560027 => 3.6km from Richmond Circle
Mallige Hospital, Siddapura Rd, Mavalli, Bengaluru, Karnataka 560004 => 3.8km from Richmond Circle
Al-ameen towers, Hosur Main Road, Sudhama Nagar, Bengaluru, Karnataka 560027 => 3.5km from Richmond Circle
2A, Cemetry Rd, Bhetal Layout, Langford Town, Shanti Nagar, Bengaluru, Karnataka 560025 => 2.3km from Richmond Circle
132, J P Nagar, Jalakanteshwara Nagar, Shanti Nagar, Bengaluru, Karnataka 560027 => 3.2km from Richmond Circle

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of breakfast spot and no breakfast spot nearby, all zones being fairly close to city center (all less than 4km from Richmond Circle, and about half of those less than 2km from Richmond circle). Although zones are 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. Most

of the zones are located around nearby boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a great number of breakfast spot in Bengaluru , there are pockets of low breakfast spot density fairly close to richmond circle center. Highest concentration of breakfast spot was detected north and east from Richmond Circle, so we focused our attention to areas south, south-west and west, corresponding to boroughs near by.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Richmond Circle) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two breakfast spots in radius of 250m and those with breakfast spot closer than 400m 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 breakfast spot 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 breakfast spot! Purpose of this analysis was to only provide info on areas close to Richmond circle center but not crowded with existing breakfast spot - it is entirely possible that there is a very good reason for small number of breakfast spot in any of those areas, reasons which would make them unsuitable for a new breakfast spots 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 Bengaluru areas close to center with low number of breakfast spot in order to aid stakeholders in narrowing down the search for optimal location for a new breakfast spot. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs 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 (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal breakfast spot 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.