# Capstone Project - The Battle of Neighborhoods

DATE: 08/10/2020
AUTHOR: Mercy Paul

# 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.

## Exploring Data

The Bangalore city source data has the following columns:

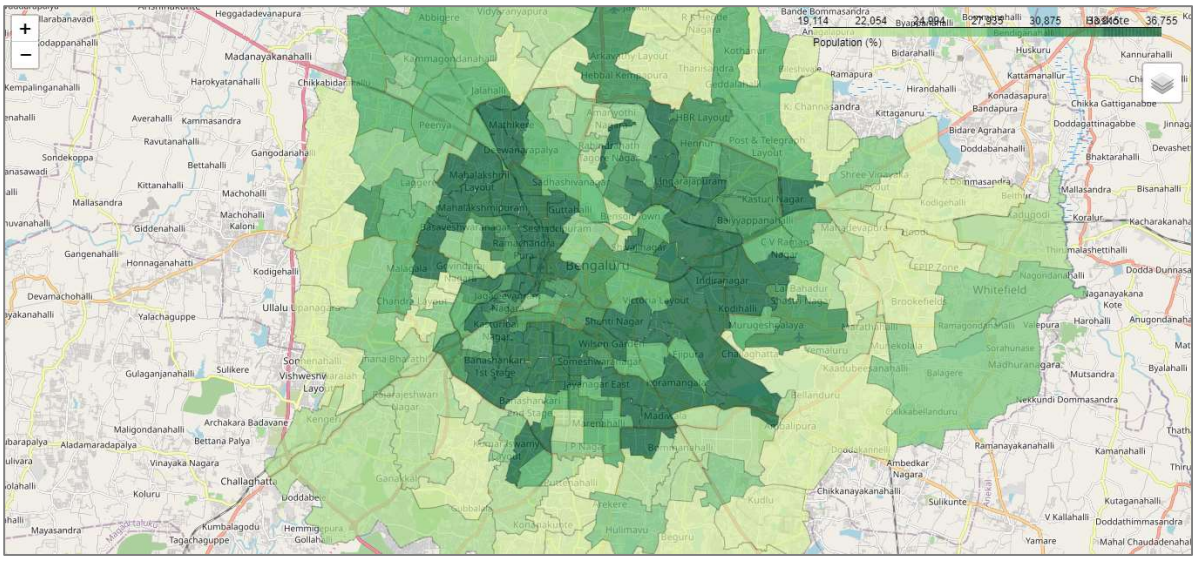
OBJECTID int64
ASS_CONST_ object
ASS_CONST1 object
WARD_NO float64
WARD_NAME object
POP_M float64
POP_F float64
POP_SC float64
POP_ST float64
POP_TOTAL float64
AREA_SQ_KM float64
LAT float64
LON float64
RESERVATIO object
geometry geometry

The columns that we will use are:

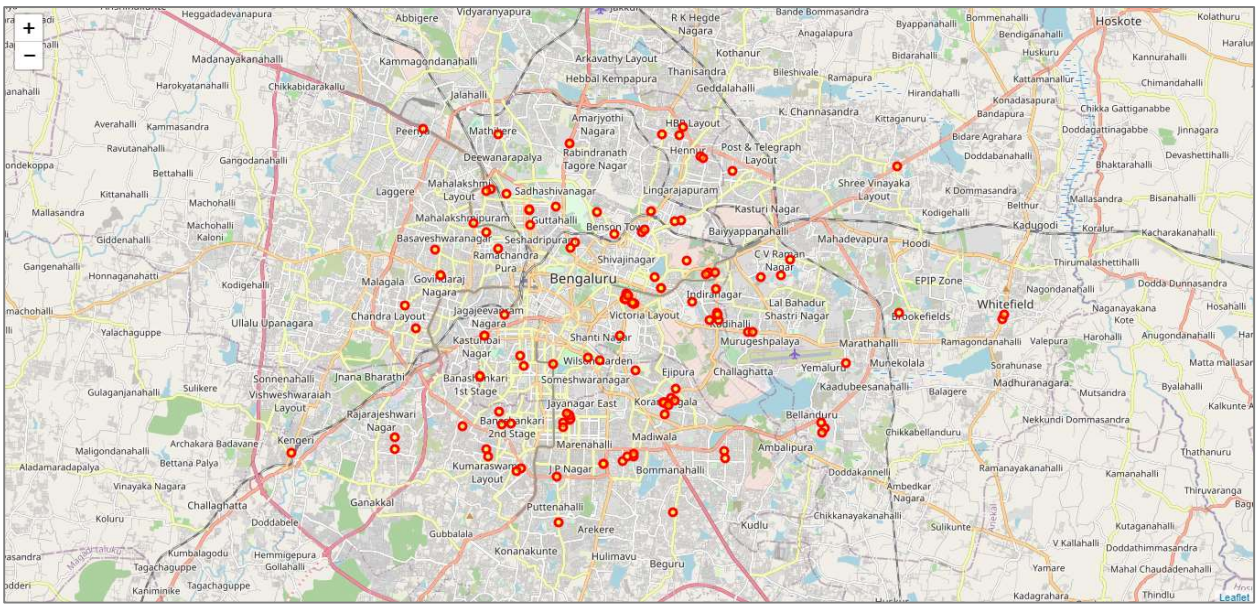
WARD_NAME

POP_TOTAL
LAT
LON
geometry

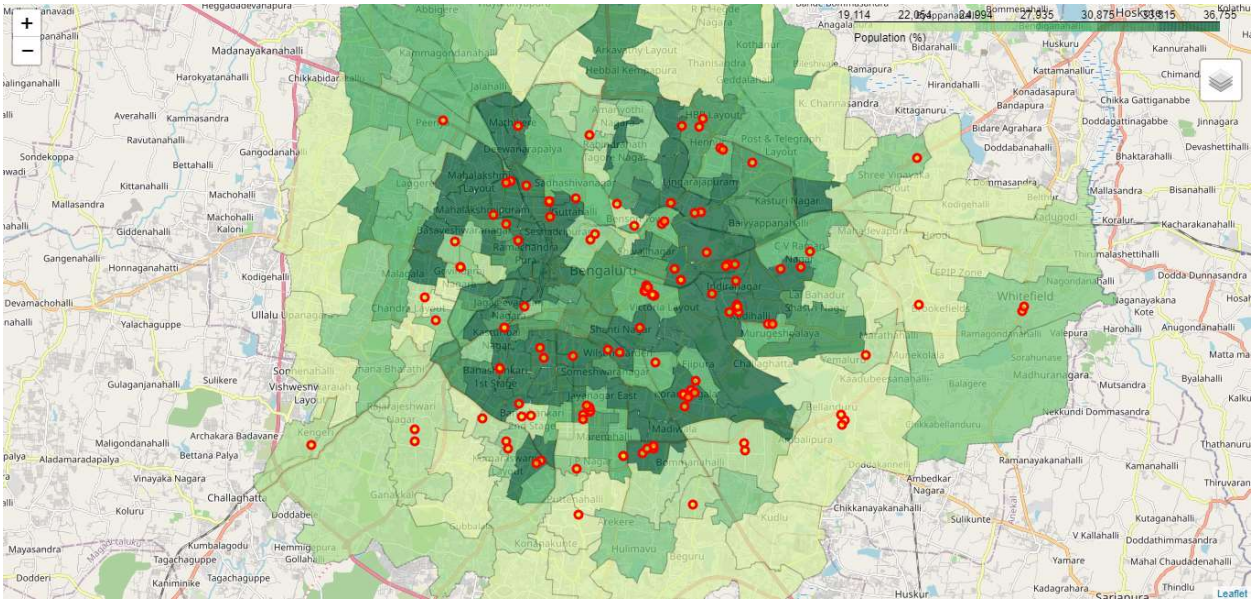
Using the above data we can identify the densely populated areas of Bangalore denoted by heat map as shown below. We will want to establish our café in potential densely populated or areas with potential growth.



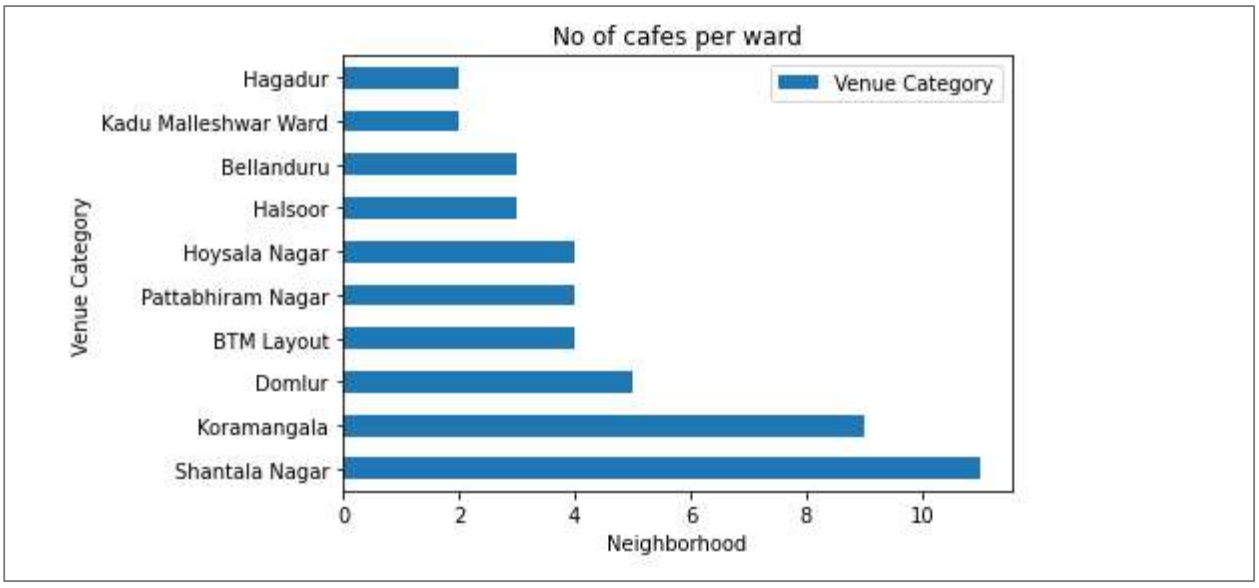

We will use the foursquare API data to explore Bangalore city to identify venues and filter out cafes. On exploring all the venues in Bangalore, we found about 183 unique venues. Out of the 183 unique values we found that cafes were categorized as Cafes or Coffee Shop. On filtering out coffee shops we identified about 121 cafes in Bangalore as shown below.



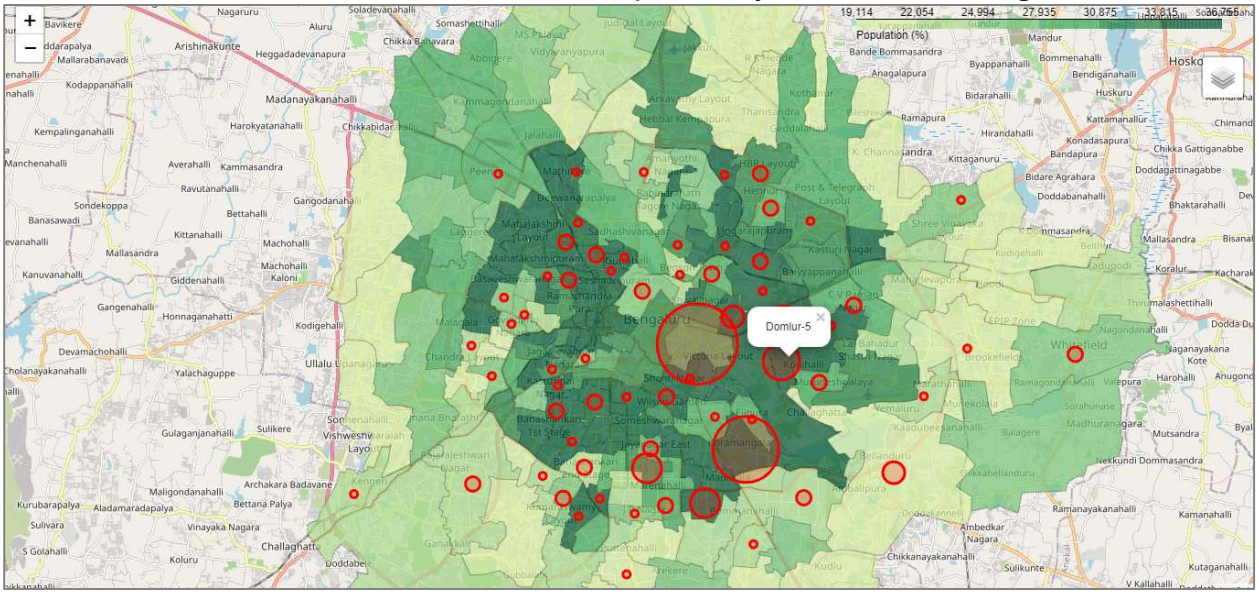
Superimposing the café locations on the ward level population density identifies the areas with cafés already established




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.




We can see that Shantala Nagar 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. We will further explore these neighborhoods to explore by using data modeling technique to find the most appropriate location.