

# APPLIED DATA SCIENCE CAPSTONE – REPORT

## OPENING A NEW MICROBREWERY IN BANGALORE, INDIA

DIPTARKA GUPTA

MAY 2020

### 1. Introduction

Bangalore is one of the largest cities in India. The city has population of around 12 million and the city is expanding rapidly over the last couple of decades due to the Information Technology industry. Due to this reason the city is also known to be the 'Silicon Valley of India'. But, over the ages Bangalore is popularly known across India to be the Pub Capital of country. And it continues to be so. With the rapid increase in population, especially many young people moving here job prospects there has been a growth in the Pub and Restaurant scene in the city. Now, Craft Beer is trending all over the world, and Bangalore is too it is a same case. As of 2019, there has been more than 60+ microbreweries spread across the city and the number is going to keep on increasing. According to the current trends it is likely that more Microbreweries are going to pop up across the city.

### 2. Business Problem

In this capstone project, we are going to analyse and select the best neighbourhoods in Bangalore, India to open a Microbrewery. We will use Data Science Methodologies and Machine Learning techniques such as Clustering to find the best location to set up a Microbrewery. The results will help someone planning to set up a Microbrewery find the perfect location. There are many socio-economic, geographical and political reasons of setting up microbreweries in certain parts of the city.

### 3. Data

We used Data Scraping techniques to obtain the list of all the wards(neighbourhoods) in Bangalore, India from the Wikipedia page: [https://en.wikipedia.org/wiki/List\\_of\\_wards\\_in\\_Bangalore](https://en.wikipedia.org/wiki/List_of_wards_in_Bangalore)

Wards are actually local authority areas(like small neighbourhoods) used for electoral purposes. In Bangalore the civic authority BBMP( Bruhat Bengaluru Mahanagara Palike ) administers these 198 wards.

In the table we also get the Assembly Constituency(for state electoral) and Lok Sabha Constituency(for national electoral) information as well. This is helpful in finding the correct geolocation since some place might have same names and this can help in differentiating. We will incorporate data cleaning and data wrangling on this data obtained and finally, get the geolocation data (latitude and longitude) using the python geocoder.

After this, we used Foursquare API to get all the venue information of each neighbourhood. We will filter this venue data to with the relevant venue information which relates to a Microbrewery.

We are going to form Clusters based on the venue information and visualize it on a map of Bangalore using Folium API.

### 4. Methodology

- Using Data Scraping Techniques we obtain the list of all the wards(neighbourhoods) in Bangalore, India from the Wikipedia page: [https://en.wikipedia.org/wiki/List\\_of\\_wards\\_in\\_Bangalore](https://en.wikipedia.org/wiki/List_of_wards_in_Bangalore).
- We use the Python Geocoder to obtain the latitude and longitude of each neighbourhood.
- Using the Foursquare API we obtain the first 200 venues for each neighbourhood with a radius of 1km.
- We perform One Hot Encoding on the Dataframe having the venue information.
- The One Hot Encoded Dataframe is further filtered for only venues 'Brewery'.

- We perform a k-means Cluster on the Filtered Dataframe, with a Cluster Size of 3. We find the neighbourhoods where Microbreweries are already present and opening a new microbrewery might be a less risk option in those neighbourhoods
  - Cluster 1: Neighbourhood without any Breweries/Microbreweries within 1 km radius.
  - Cluster 2: Neighbourhood already having some or at least a Microbrewery within 1km radius
  - Cluster 3: Neighbourhoods having a high concentration of Microbreweries within the 1km radius.
- We analyse each Clusters.
- Same method of k-means Clustering is applied with number of Clusters Again being 3, with venues values of 'Pub','Beer Garden','Beer Bar','Food & Drink Shop' from the One Hot Encoded Dataframe. This analysis will help us find neighbourhoods with good potential. We will also find some overlapping from the last analysis as well as some new neighbourhoods. These new neighbourhoods have good potential of opening a Microbrewery. It can be more rewarding with less competition but the risk involved is more.
- We map the clusters of the respective analysis using Folium API.

## 5. Results

From our initial analysis, we find that most of the Microbreweries are present in the Southern and Eastern and some Central regions of the city. The neighbourhoods in and around JP Nagar, Jayanagar, Koramangala, Central Bangalore District, Indiranagar, CV Raman Nagar and Hoodi seem to be the places having most concentrations of Microbreweries. These neighbourhood already have a good thriving business of Microbreweries in the city and opening a new one would mean less risk but a higher competition.

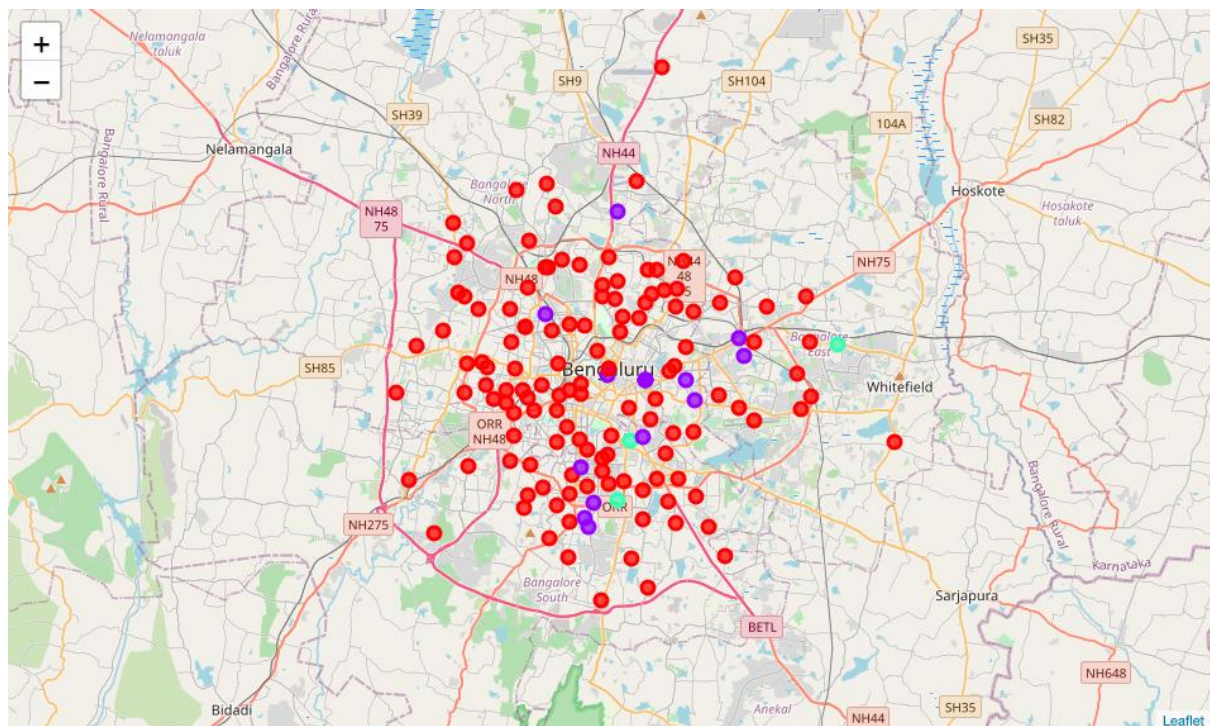


Fig 1. Cluster 1: Red, Cluster 2: Purple, Cluster 3: Green



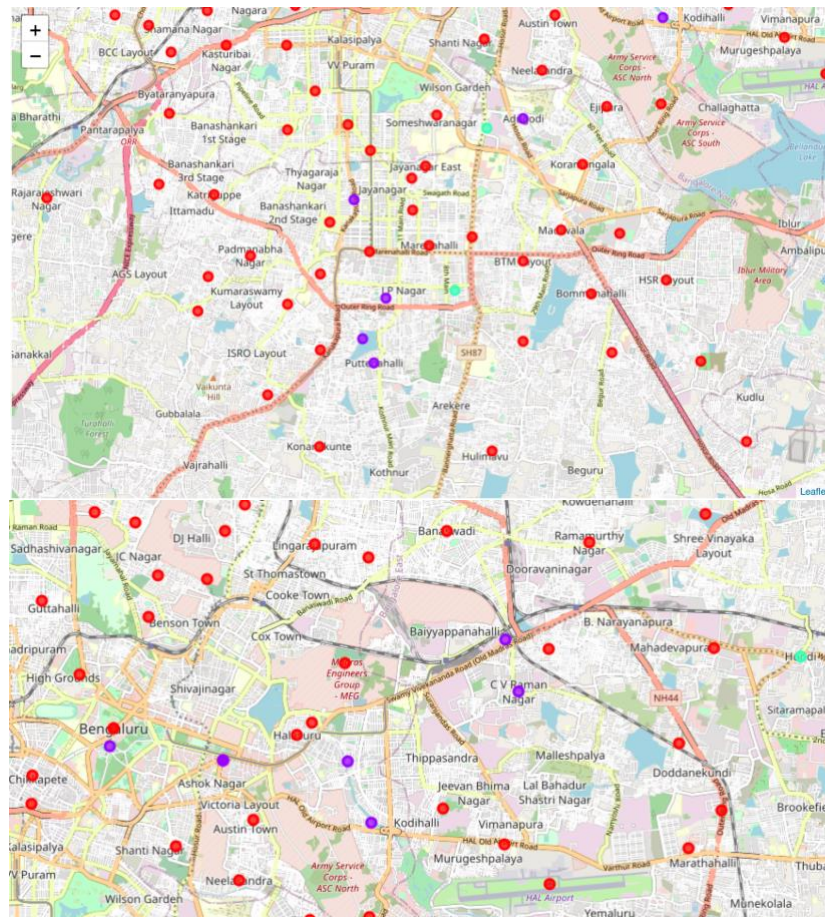


Fig 1.1. Neighbourhoods with more concentration of Microbreweries

From our second analysis, we find that the potential neighbourhood have a huge amount of overlapping with our previous analysis. We find further more neighbourhood in the central part of the city having a good potential.

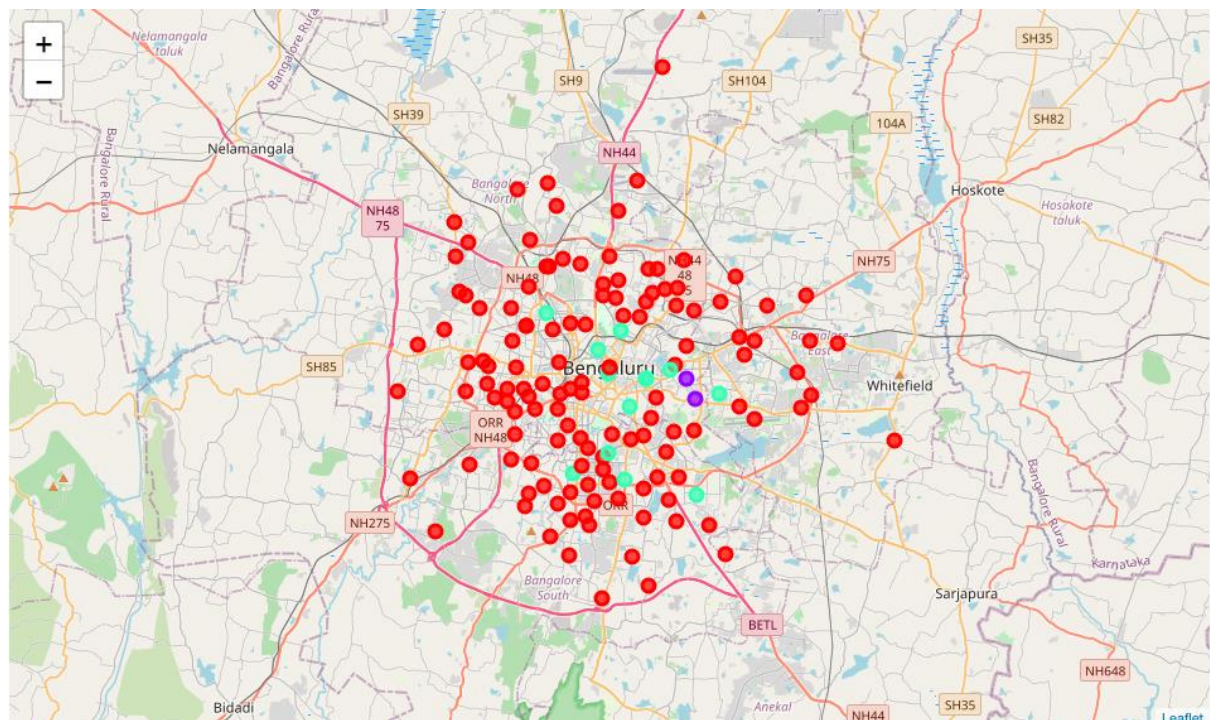


Fig 2. Cluster 1: Red, Cluster 2: Green, Cluster 3: Purple

## **6. Conclusion**

In this study, we find out which are the neighbourhoods suitable for opening a new Microbrewery in Bangalore, India. We also find out the Neighbourhoods with good potential. Our entire analysis is based on the Foursquare API data.

## **7. Future Direction**

We have only considered the Foursquare Venue information to find the out the suitable neighbourhood. We have not considered any other information such as socio-economic information, population density of those neighbourhood for the analysis. In future we can add these and make our study more comprehensive.