

# Coursera Capstone Project: Battle of Neighborhoods

Sunidhi Hegde

hsunidhi8@gmail.com

## 1. Introduction

Bangalore is the largest and the capital city of the state of Karnataka, India. Touted as the 'Garden City' for its beautiful parks and green landscapes, it is now regarded as one of the largest IT hubs in the country. This boom in the information technology industry has garnered immense revenues and in addition to that, has also led to the growth of various other industries which includes the organized food service market, one of the largest services in the country after retail and insurance.

As revealed by the latest NRAI India Food Services Report, Bangalore's food service market stands third, at INR 20,014 Crores (2.8 Billion USD), among other major metropolitan cities in India. The average expenditure per month per household on eating out is INR 3,586 in Bengaluru, which is higher than the national average of INR 2,500. The city has approximately 42,000 restaurants and employs over one lakh people. The growth of Bengaluru's food service industry has been accredited to the cosmopolitan culture and IT sector jointly sparking a revolution in the eating and cooking habits of the residents.

Various neighborhoods in Bangalore are slowly becoming hotspots for culinary indulgences and for a person looking to open a food joint or a café, making an informed decision is necessary to reap maximum benefits in terms of customer loyalty and revenue. As too many neighborhoods are vastly commercializing their spaces, it is quite difficult for a potential café owner to decide on an optimal location and hence, a quantitative approach is needed to choose the right location.

## 2. Business Problem

Bangalore has been lauded for embracing the start-up culture and various notable start-ups have thus originated here ever since. This also means potential start-ups that are yet to make it big struggle to rent office space and most of their collaborations are done ad-hoc. This is where the quaint cafes the city has to offer save the day. This also explains the sudden surge in the number of cafes in certain neighborhoods that have IT offices or office spaces up for rent. Neighborhoods that border around major IT hubs of the city are also not far behind – one can see numerous food joints spread across these neighborhoods and being profitable. The main aim of this project is to find such ideal spots across the city to open a café or a restaurant and gain maximum profit by catering to the needs of the demographic discussed above.

### 3. Data

Data for this project has been obtained through three major sources as discussed below.

#### 3.1 List of neighborhoods

The list of neighborhoods in Bangalore was obtained by Web Scraping a Wikipedia page. This was accomplished through a Python library called BeautifulSoup and later stored as a csv file.

```
df.head()
```

Neighbourhood	
0	Adugodi
1	Agara, Bangalore
2	Ananthnagar
3	Anjanapura
4	Arekere

#### 3.2 Latitudes and longitudes

The comprehensive mapping of latitudes and longitudes of all the neighborhoods was accomplished through Nominatim Geocoding API. This data was later stored in a pandas Dataframe.

	Neighbourhood	Address	Latitude	Longitude
0	Adugodi	(Adugodi, South Zone, Bengaluru, Bangalore Sou...	12.942847	77.610416
1	Agara, Bangalore	(Agara, Kanakapura taluk, Ramanagara district,...	12.620112	77.479307
3	Anjanapura	(Anjanapura, Bommanahalli Zone, Bengaluru, Ban...	12.858081	77.558071
4	Arekere	(Arekere, Bommanahalli Zone, Bengaluru, Bangal...	12.887209	77.596049
5	Austin Town	(Austin Town, East Zone, Bengaluru, Bangalore ...	12.961274	77.615294

#### 3.3 Venue Details

The Dataframe containing the neighborhoods and their locations is used to obtain the details of venues in that neighborhood. This was done using Foursquare API after creating a developer account which allows for a limited number of free API calls per

day. Below is the dataset after obtaining location and venue data for the list of neighborhoods.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue Name	Venue Category	Venue Latitude	Venue Longitude
0	Adugodi	12.942847	77.610416	PVR IMAX	Movie Theater	12.934595	77.611321
1	Adugodi	12.942847	77.610416	Lot Like Crêpes	Creperie	12.936421	77.613284
2	Adugodi	12.942847	77.610416	Zingron - Naga Kitchen	Indian Restaurant	12.936271	77.615051
3	Adugodi	12.942847	77.610416	Koramangala Social	Lounge	12.935518	77.614097
4	Adugodi	12.942847	77.610416	Tommy Hilfiger	Clothing Store	12.934552	77.611347

## 4. Methodology

The steps and methodologies implemented during the course this project can be summarized as given below. Each step was carried out based on careful consideration of the requirements for the analysis.

### 4.1 Nominatim Geocoding API

The Nominatim Geocoding API is a free geocoding API which can be used instead of Google Geocoding API as they have made it mandatory to have a premium account. However, Nominatim has its own drawbacks – the predominant one being frequent timeout errors. The key is to be persistent and make as many API calls as required until the entire neighborhood dataset has been traversed.

```
latitudes = []
longitudes = []
neighborhood = []
for nbd in df["Neighbourhood"] :
    address = nbd
    geolocator = Nominatim(user_agent="blr_explorer")

    location=geolocator.geocode(address)
    print(location)
    if location!= None:
        lat = location.latitude
        lng = location.longitude
        neighbourhood.append(location)
        latitudes.append(lat)
        longitudes.append(lng)
```

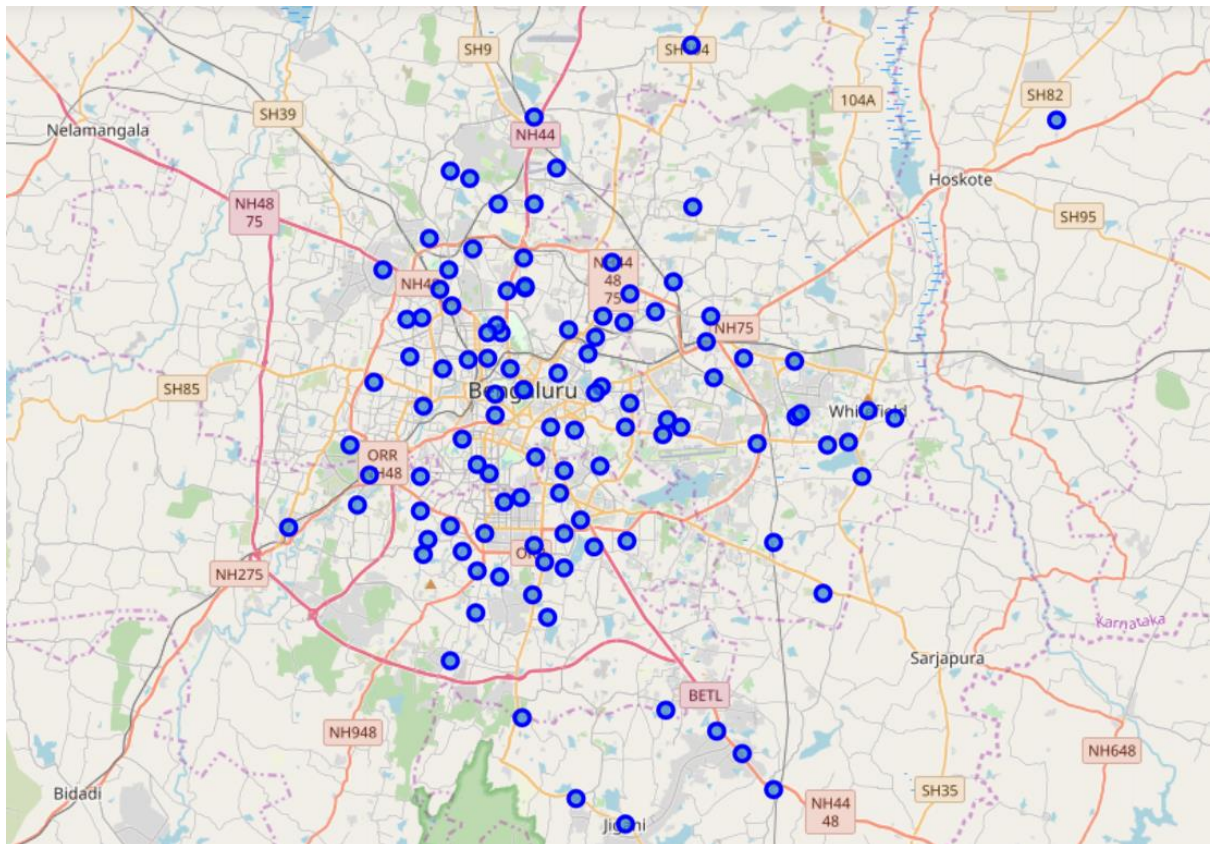
## 4.2 Plotting a map using Folium

Folium is a handy, comprehensive library in Python to visualize co-ordinates of a location on a map. Using the latitudes and longitudes obtained through Nominatim, a map with markers pointing to specific neighborhoods can be plotted.

```
# create map of Bangalore using Latitude and Longitude values
map_bangalore = folium.Map(location=[lat, long], zoom_start=11) #Passing location info of Bangalore

# adding markers to map
for lat, lng, label in zip(df['Latitude'], df['Longitude'], df['Neighbourhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_bangalore)

map_bangalore
```



## 4.3 Obtaining Venue details

Once the neighborhoods can be visualized accurately on a map, they can be used to glean necessary details of the venues in and around each location on the map. This can be done using Foursquare API and for this project, only venue categories were used.

```
explore_df_list = []

for i, nbd_name in enumerate(df3['Neighbourhood']):
    try :

        nbd_name = df3.loc[i, 'Neighbourhood']
        nbd_lat = df3.loc[i, 'Latitude']
        nbd_lng = df3.loc[i, 'Longitude']
        radius = 1000
        LIMIT = 30
        url = 'https://api.foursquare.com/v2/venues/explore?client_id={} \
&client_secret={} &ll={},{}&v={} &radius={} &limit={}\' \
.format(CLIENT_ID, CLIENT_SECRET, nbd_lat, nbd_lng, VERSION, radius, LIMIT)

        results = json.loads(requests.get(url).text)
        results = results['response']['groups'][0]['items']
        nearby = json_normalize(results)

        filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
        nearby = nearby.loc[:, filtered_columns]
        columns = ['Name', 'Category', 'Latitude', 'Longitude']
        nearby.columns = columns
        nearby['Category'] = nearby.apply(get_category_type, axis=1)

        for i, name in enumerate(nearby['Name']):
            s_list = nearby.loc[i, :].values.tolist()
            f_list = [nbd_name, nbd_lat, nbd_lng] + s_list
            explore_df_list.append(f_list)

    except Exception as e:
        pass
```

## 4.4 One-hot encoding

Venue categories are successfully obtained using Foursquare API. After this step, one-hot encoding is performed to normalize data and pivot the categories to enable easy analysis for clustering.

	Neighbourhood	ATM	Accessories Store	American Restaurant	Andhra Restaurant	Animal Shelter	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	...	Tennis Court	Theater	Toy / Game Store	Trail	Train Station	Ud Restaur
0	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

## 4.5 Top 10 venues

After one-hot encoding is performed, insights like top 10 venues around each neighborhood can be found. The code snippet that demonstrates this process is as shown below.

```
num_top_venues = 10
indicators = ['st', 'nd', 'rd']
columns = ['Neighbourhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{} {} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighbourhoods_venues_sorted = pd.DataFrame(columns=columns)
neighbourhoods_venues_sorted['Neighbourhood'] = blr_grouped['Neighbourhood']

for ind in np.arange(blr_grouped.shape[0]):
    neighbourhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(blr_grouped.iloc[ind, :], num_top_venues)

neighbourhoods_venues_sorted.head()
```

## 4.6 K-means clustering

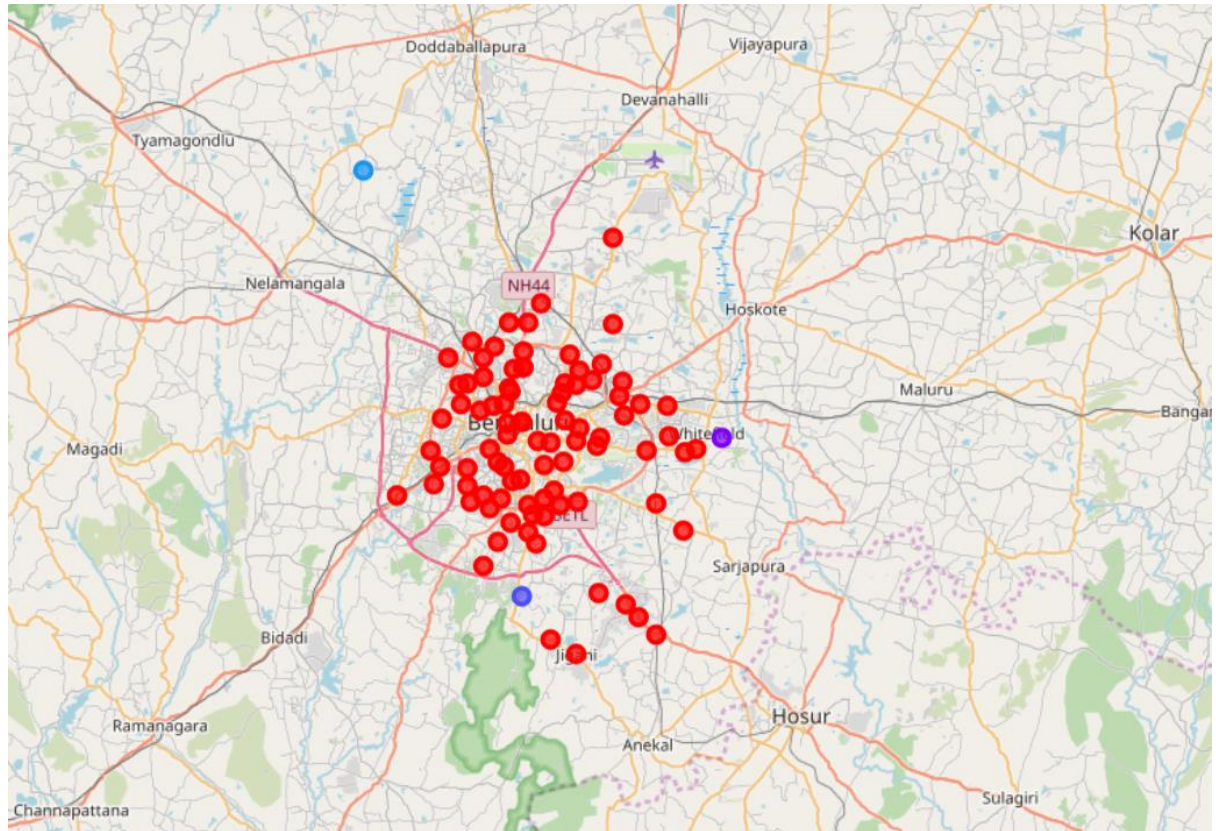
One of the best unsupervised learning algorithms is K-means clustering owing to its simplicity and fast output. K-means clustering is employed to garner insights about similar groups of neighborhoods - finding groups of neighborhoods that share certain characteristics. K-means was specifically chosen as the features are large in number and require fast computation to yield results. An optimal value of K was found by iterating from 2 to 20 and using silhouette score as a parameter for accuracy.

```
K = 4
bgc = blr_grouped_clustering
kmeans = KMeans(n_clusters = K, init = 'k-means++', random_state = 0).fit(bgc)
```

## 5. Results

The neighborhoods are segmented into K clusters depending on the value of K assigned while training the model. This can be visualized on a map where each cluster is indicated by a different color. As can be seen from the map, cluster 0 in red is the most predominant.





## 6. Discussion

After a thorough analysis of the clusters, cluster 0 seems to be the best fit containing the target venue category – Café.

	Neighbourhood	Latitude	Longitude	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
26	Cooke Town	13.002785	77.624747	0	Café	Bakery	Department Store	Fast Food Restaurant	Shopping Mall	Chinese Restaurant	Indian Restaurant	Clothing Store	Coffee Shop
30	Devarachikkanahalli	12.902105	77.601220	0	Café	Multiplex	Pizza Place	Indian Restaurant	Italian Restaurant	Coffee Shop	Clothing Store	Mediterranean Restaurant	Garage
33	Ejipura	12.945245	77.626914	0	Café	Pizza Place	Indian Restaurant	Gym / Fitness Center	Clothing Store	Ice Cream Shop	Department Store	Fast Food Restaurant	Fast Food Restaurant
47	HSR Layout	12.911623	77.638862	0	Café	Indian Restaurant	Ice Cream Shop	Pizza Place	Snack Place	Coffee Shop	Kebab Restaurant	Liquor Store	Burrito Restaurant
48	Hulimavu	12.877349	77.602803	0	Café	Movie Theater	Multiplex	South Indian Restaurant	Department Store	Badminton Court	Lake	Fast Food Restaurant	Bowl Restaurant
101	Sadashivanagar	13.007708	77.579589	0	Café	Coffee Shop	Indian Restaurant	Department Store	Dessert Shop	Gym	Ice Cream Shop	Women's Store	Fitness Center
105	Shikaripalya	12.835985	77.657181	0	Café	Indian Restaurant	Fast Food Restaurant	Hotel	Coffee Shop	Juice Bar	Department Store	Pizza Place	Chinese Restaurant

Neighborhoods namely Cooke Town, Devarachikkanahalli, Ejipura, HSR Layout, Hulimavu, Sadashivanagar and Shikaripalya look like the prime locations to start a food service like a café. These neighborhoods also border major IT hubs in the city and some of them even have office spaces coming up which implies that the business perspective is fulfilled.

Recent studies have shown that the IT boom in Bangalore is going to increase exponentially and choosing to start a café in the above neighborhood would prove beneficial in the long run.

## **7. Conclusion**

The organized food service industry will continue to flourish given the penchant for exploring various cuisines amongst the residents of metropolitan cities. Keeping an eye out for promising locations to open a food business can be beneficial but weighing the pros and cons of any venture quantitatively is a need of the hour.

This project has successfully devised a quantitative approach for a potential café owner to make an informed decision about the location to start their business. This could save a lot of costs and help the business thrive when a consistent customer base has developed.