***Recommending neighborhood options for a leading business workspace developer & manager (ABC) in Bengaluru***

Kabilan Sundararaj

June 28, 2020

**Table of contents**

|  |  |
| --- | --- |
| Topic | Page number |
| Introduction section | 2 |
| Data section | 2 |
| Methodology section | 3 |
| Result section | 5 |
| Discussion section | 5 |
| Conclusion section | 8 |

1. **Introduction & Interest:**

ABC is contemplating the decision of selecting neighborhood options for their upcoming workspace development project in Bengaluru, India. They have been well known in workspace development & managing industry. **2 out of top 10 workspaces** in Bengaluru belongs them. Given the success those two business parks enjoy, they are considering choosing a neighborhood very similar to the ones they already developed. They believe that if they find a similar neighborhood, they don't have to scrutinize much about the other parameters that might lead to the success of the planned workspace. They firmly believe in that success of their existing workspaces is because of the neighborhood & composition of venues & facilities around them. The client is reaching out to us to come up top 5 similar neighborhoods for them. Based on the top recommendations, they are planning further to scrutinize the demand and other business factors to choose the final neighborhood for the planned workspace development

1. **Data:**

**How Foursquare location data helps?**

Since ABC believes that the success of the workspace depends on the neighborhood & compositions of nearby venues & facilities, foursquare can provide us with the venue data in & around most of the neighborhoods in Bengaluru

**Data that will be used in the project:**

There are three data types would be required to execute our solution, they are as follows:

1. All available boroughs & neighborhood names in Bengaluru
2. Geographical coordinates of the above neighborhoods
3. Borough, neighborhood data of existing ABC workspaces in Bengaluru
4. Venue details, categories & other venue related info

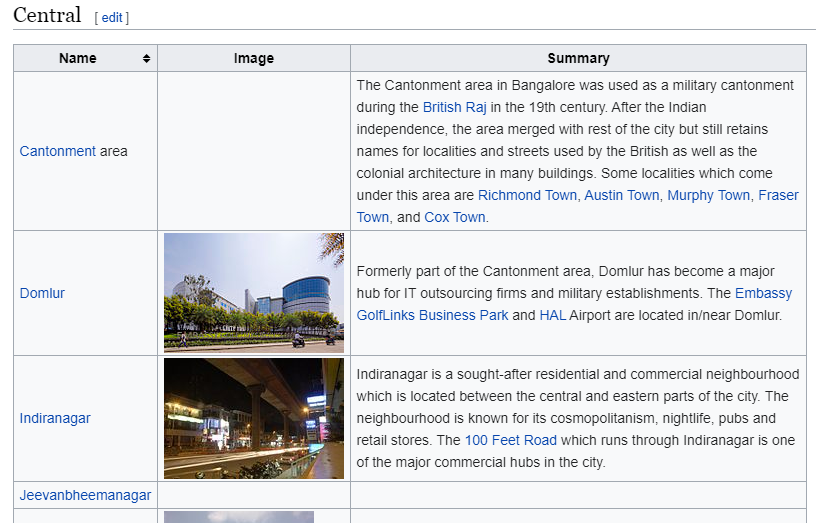
**Data sources for the above:**

1. [https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Bengaluru](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)
2. The geographical coordinates were not readily available. Since the number of neighborhoods was less, manually collected the data from web.
3. Manually googled & found the same
4. Foursquare venue details to be fetched through the neighborhood names

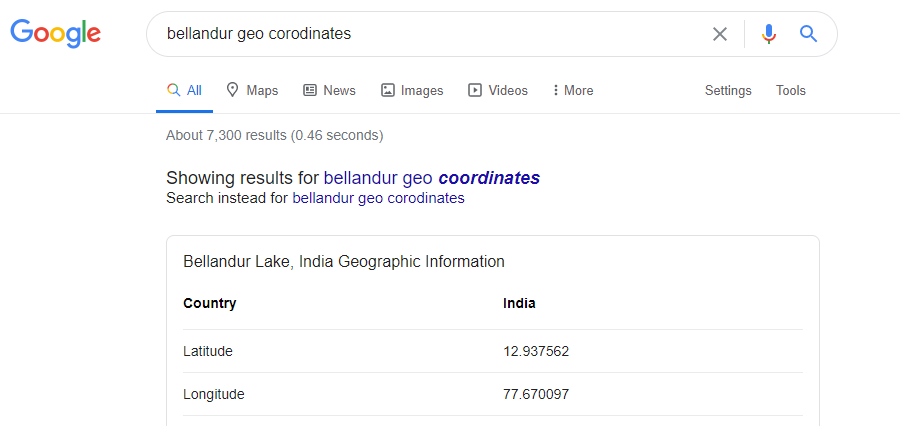
**Data prep:**

* Performed web scraping on Wikipedia page for getting neighborhoods
* Manually added the coordinates after researching them for each row

Before scrapping the neighborhood, data looks like below



Scrapping the geo coordinates manually.



After prep the data looks something like this

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Borough** | **Neighborhoods** | **Latitude** | **Longitude** |
| 0 | Central | Cantonment area | 12.972442 | 77.580643 |
| 1 | Central | Domlur | 12.9609917 | 77.6387256 |
| 2 | Central | Indiranagar | 12.971891 | 77.641151 |

.. and so, on

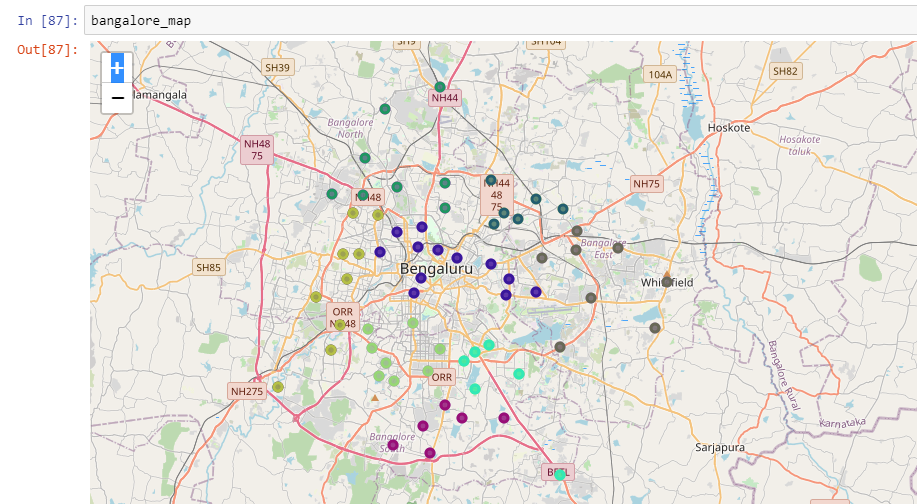
1. **Methodology:**

**Summarized approach:**

* The neighborhood where the existing ABC workspaces are found first.
* Most frequent venues are listed based on frequency for both neighborhoods
* If there exists good similarity between the two, we can try clustering the neighborhoods based on the top venues data to find similar neighborhoods
* The venue details for each neighborhood is gathered using foursquare
* For each neighborhood the top venue categories are calculated & listed
* The above data is fed into a clustering algorithm (preferably KMeans algorithm, as are to find a neighborhood with similar composition of venues)
* If both neighborhoods fall under a same cluster then, the average top venue frequency is averaged for both – this later form the weights for ordering the top 5 options under the same cluster
* The weights got above is multiplied with the other neighborhoods in the same cluster
* They are ranked based on the sum product of weights with their top venue frequencies
* The top 5-10 neighborhoods is present as options for them to further choose

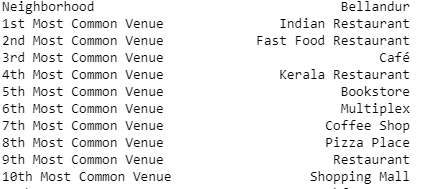
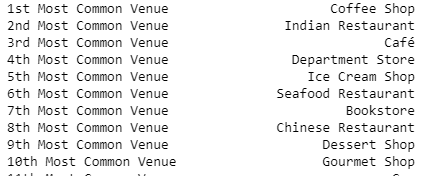
**Exploring the neighborhoods:**

Since the client requires a very similar neighborhood to their existing workspace neighborhoods. First the exploration of all available neighborhoods is done through geo-based visualization based on folium-based maps. The basic location understanding of the neighborhoods is understood as below.



**Exploring neighborhoods of existing workspaces:**

Then further research was done to explore the composition of venues around neighborhoods Sadashivanagar & Bellandur. This is done by getting venue details under Foursquare & grouped under each neighborhood. The similarity/dissimilarity between the two neighborhoods is explored by listing their top locations. PFB the comparison

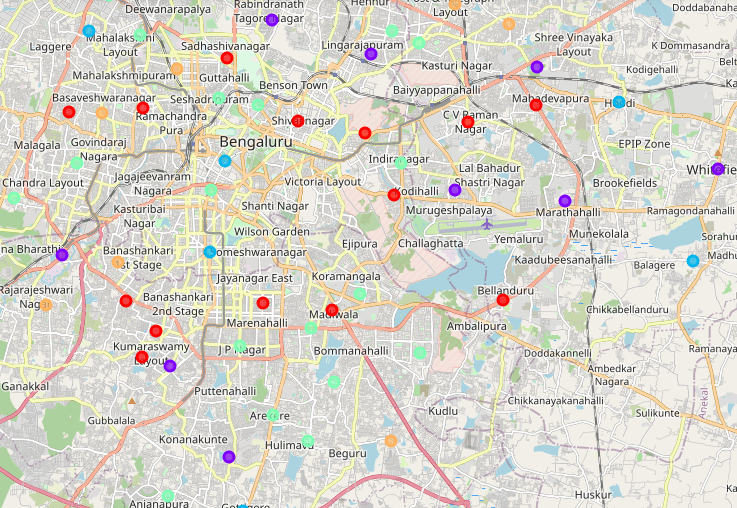
 

From the above comparison , we easily find that both of the neighborhoods has 4-5 common venue types which are the most frequently found venues. This insight pointed us forward in the clustering approach. Where the same common venue details is done for all the neighborhoods in Bangalore.

**Clustering Neighborhoods:**

Since we are checking for similarity of neighborhoods to the known two, KMeans algorithm is the most obvious algorithm to calculated the multidimensional distance between each neighborhoods

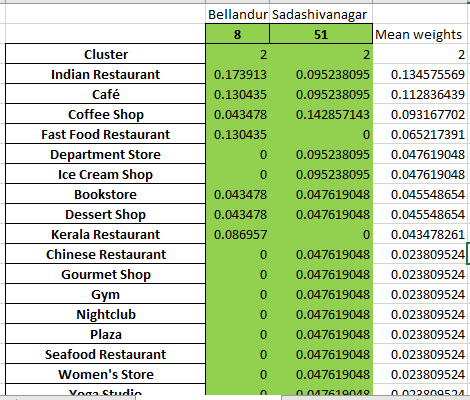
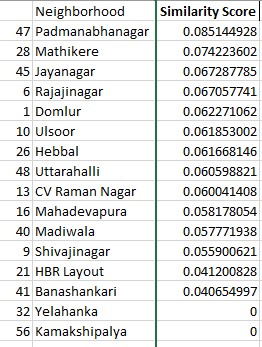
Post feeding the data, it was found that both neighbordhoods fall under the same cluster. Both Bellandur aand Sadashivanagar are highlighted in red below



1. **Results**

**Selecting top 5-10 similar options from 16 neighborhoods in the cluster:**

Bellandur & Sadashivanagar top venue frequencies are taken as weights & multiplied with the frequencies of other neighborhood venue frequencies. The sum product of the same is done. Ideally this is to be divided by the weigths, this was not done as the sum of weights is 1. Then the options are sorted from largest to smallest, thereby listing the top 5-10 options for ABC to choose from

We can cleary see the top 10 neighborhood options based on the similarity score above from the cluster level options

1. **Discussion:**

The other clusters are also explored to understand the classification trend in neighborhoods. We can clearly see clusters in pratical sense as well. Before that we will look at the other clusters & understand the insights

**Cluster 1 insights:** Looking at Whitefield, Electronic city, Marathahalli & KR Puram we can cleary group them together as **IT hubs** which houses and **flaunts the most mature, modern facilities & venues**. This can be clearly below in the table figure.



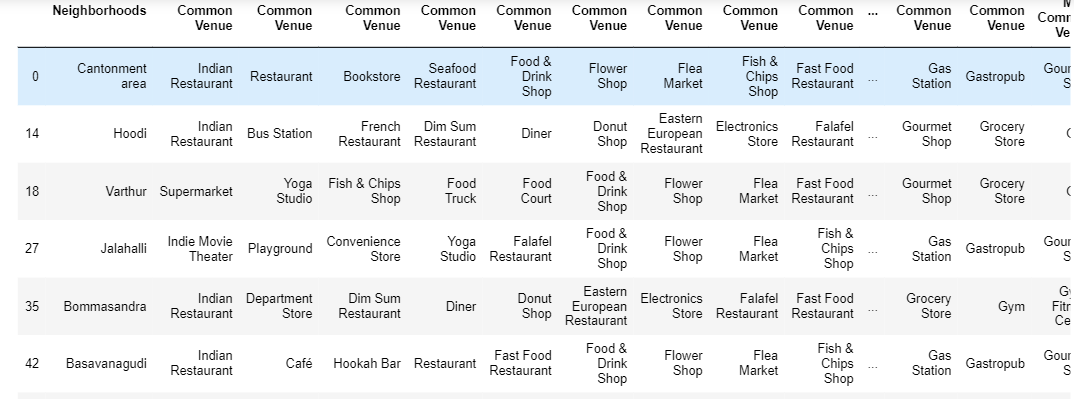
**Cluster 3 insights:**

This cluster houses neighborhoods such as Koramangala, Indiranagar, HSR, BTM etc which are known for the **nightlife & youth residential areas.** These neighborhoods is very big on food & related venues as well, but less on office & It spaces.



**Cluster 2 insights:**

The cluster here is not big on migrant population, so it is mainly dominated by native Karnataka crowd, developling an workspace for MNCs is not a good option here. They also do not house any big workspaces



**Cluster 0 insights:**

If one had lived here Bengaluru, we can clearly understand that these neighborhoods enjoy a unique combination of big workspaces, IT hubs & as well as residential places. There is a equal mix of migrant & native population around these neighborhoods, which is one . Along with all recreational & food venues, they also enjoy a host of facilities & shops. We can see Bellandur & Sadasivanagar here rightly so. This shows that the model has performed really well in clustering the results

1. **Conclusion:**

As already shown in the results section, where the above cluster 0 is ranked based on similarity scores of other workspace neighborhoods. The top 10 options are very good in terms of the intial shortlist, but selecting the final one for the planned construction should be done with more scrutiny & focussed on business implications. However this project easily does 70 % of the ask.