

New Brewery Opening Location

1. Introduction:

As part of the IBM capstone project, I have chosen an idea related to *opening a new Brewery* in a location of my preference i.e. Bangalore.

Bangalore is famous for its pubs, bars and brewery culture. So, this got me interested in finding out what locations/neighborhoods in Bangalore would be

suitable for opening a new brewery or a branch of an existing brewery.

One of the motivations for choosing this problem statement is related to my preference for brewed beer vs bottled beer – I occasionally do enjoy bottled beer,

but given a choice I would choose freshly brewed. Also, in case of brewed beer

there are many options in terms flavors and concentrations to choose from.

And I am seeing an increasing trend in this behavior and will result in many new breweries opening with more variations in their menu.

From the business point of view, the location of an outlet can make or break decision. Many factors contribute to the success of running bar, pub or brewery.

Primary being the location/neighborhood, and others include surrounding venues, type of crowd in the neighborhood, main landmarks etc.

So, in this project, I will using clustering concept to find the best possible locations in Bangalore to open a Brewery.

2. Data acquisition:

For any data science project, the quality of data is more important than the quantity of data. The key data that I will be using in this project is the location data of Bangalore

neighborhood i.e. longitude and latitude of key places. Along with this to fetch

venue details of places around these neighborhoods, I will be using the Foursquare APIs.

Below is snapshot of the location data of Bangalore neighborhood. It contains neighborhood name, pin code, zone (*east, west etc.*), longitude and latitude.

I have acquired this data by web scraping mechanism.

Neighbourhood	Pincode
Mahatma Gandhi Road S.O	
Malleswaram S.O	
Basavanagudi H.O	
Fraser Town S.O	
H.A.L II Stage H.O	
Rajajinagar H.O	
Jayangar III Block S.O	
Jalahalli H.O	
Jalahalli East S.O	
Vimanapura S.O	
Gayathrinagar S.O	
Yeswanthpura S.O	
Dharmaram College S.O	
St. John's Medical College S.O	
Carmelram S.O	
Marathahalli Colony S.O	
Jayanagar H.O	
Arabic College S.O	
Ashoknagar S.O (Bangalore)	
Mathikere S.O	
Malleswaram West S.O	
Yelahanka S.O	
G.K.V.K. S.O	

Along with this, I would be using the venues information in and around these neighborhoods. That will look like something like shown below.

Using the Foursquare API, I will be getting all the venues belonging to different categories like parks, temples, schools, restaurants, café, bars etc.

	Neighborhood	Neighborhood Latitude	N
0	Mahatma Gandhi Road S.O	12.9739	
1	Mahatma Gandhi Road S.O	12.9739	
2	Mahatma Gandhi Road S.O	12.9739	
3	Mahatma Gandhi Road S.O	12.9739	
4	Mahatma Gandhi Road S.O	12.9739	
5	Mahatma Gandhi Road S.O	12.9739	
6	Mahatma Gandhi Road S.O	12.9739	
7	Mahatma Gandhi Road S.O	12.9739	
8	Mahatma Gandhi Road S.O	12.9739	

One important consideration to keep in mind is that I am looking for neighborhoods suitable to open new beer brewery. As a result, below aspects have been

kept in mind while data cleaning and filtering.

- Beer Brewery business can thrive in places where there is already a hub of eateries, bars, pubs and other breweries.
- This indicates that people these neighborhoods are already having

preference for eating and drinking out.

- Or these neighborhoods have other aspects, like a shopping mall, IT parks or movie halls, that is attracting people.
- Adding to this, people would want to have options in terms of their dining/drinking experience. So, looking out for such neighborhoods can add value since a new place might get a preference.
- Also starting a brewery in a place that doesn't have any good concentration of food and drink places might be risky.
- So, I would be filtering data to remove other category venues.
- As a result, in each neighborhood, I would be interested in venues that belong these venue categories: restaurants, bars, cafes, pubs, breweries etc...

3. Methodology

The problem I am trying to solve here is identifying the suitable location in a city to

open a new beer brewery. Whether it is a restaurant or café or any other eatery

location matters a lot. Using the location data of neighborhoods and venue information, I would be able to build clusters of neighborhoods having similar characteristic.

As discussed in the data acquisition section, I would be filtering to acquire right set of venue categories. The right way to go about analyzing the neighborhoods of a city

is to use the similarity in the venues hosted by them. Using the presence of venues of different categories and their mean occurrence in the neighborhood I can build cluster

of neighborhood with similar venue categories.

K-means algorithm is best suited for this problem with such requirement. Using this algorithm, I have created clusters with similar characteristics in terms of the venues hosted

by them. I have used the Elbow method to find optimal K in K-means, and it comes to be 4.

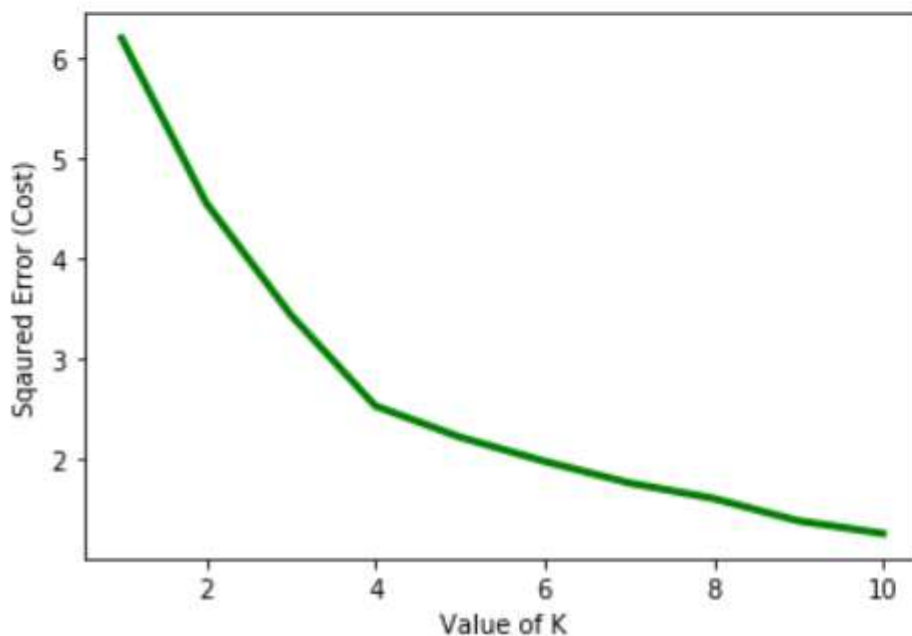
```

: cost=[]
  # set number of clusters
  kclusters = 5
  for i in range(1,11):

      # run k-means clustering
      kmeans = KMeans(n_clusters=i, random_state=0).fit(bangalore_venues_grouped)
      # check cluster labels generated for each row in the dataframe
      kmeans.labels_[0:10]
      cost.append(kmeans.inertia_)

: plt.plot(range(1, 11), cost, color='g', linewidth='3')
  plt.xlabel("Value of K")
  plt.ylabel("Sqaured Error (Cost)")
  plt.show()

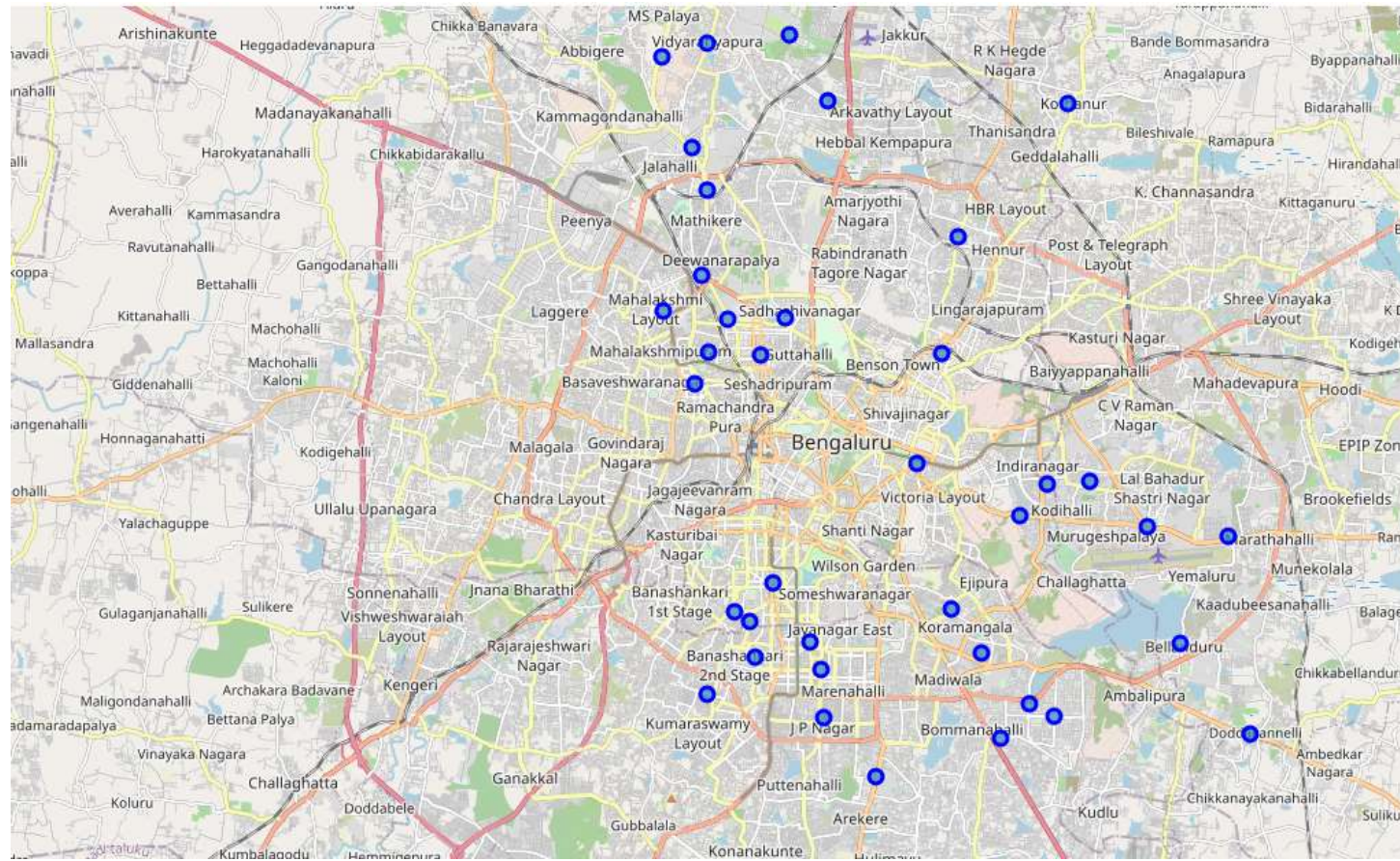
```



The below maps show the neighborhood before and after the cluster.

Before

After



have a brewery. And that neighborhood would one of the ideal candidate for opening

new brewery.

```
: cluster0 = clusters[clusters['Cluster_Labels']==0]
```

Clearly Cluster 0 is not a suitable choice

```
: cluster0.groupby(['Neighborhood', 'Venue_Category']).count()
```

Cluster Labels		
	Neighborhood	Venue_Category
	Bommasandra Industrial Estate S.O	Fast Food Restaurant
	Sarjapur	Fast Food Restaurant
		Restaurant

```
: cluster1 = clusters[clusters['Cluster_Labels']==1]
```

In Cluster 1 none of the neighborhood have

```
: cluster1.groupby(['Neighborhood', 'Venue_Category']).count()
```

Cluster Labels		
	Neighborhood	Venue_Category
	Gayathrinagar S.O	Italian Restaurant
		Lounge
		Vegetarian / Vegan Restaurant
	Gottigere S.O	American Restaurant
		Dive Bar
		Fast Food Restaurant
		Italian Restaurant
		Rajasthani Restaurant
	Koramangala VI Bk S.O	Shopping Mall
		Coffee Shop
		Italian Restaurant
		Brewery
		Coffee Shop
	Kothanur S.O	Italian Restaurant
		Restaurant
		Vegetarian / Vegan Restaurant


```
: cluster2 =clusters[clusters['Cluster Labels']==2]
```

Similar to Cluser 0

```
: cluster2.groupby(['Neighborhood','Venue_Category']).count()
```

```
: Cluster Labels
```

Neighborhood	Venue_Category	
Domlur S.O	Indian Restaurant	1
Jeevanbhimannagar S.O	Indian Restaurant	1
Sadashivanagar S.O	Indian Restaurant	1
Sahakaranagar P.O S.O	Fast Food Restaurant	1
	Indian Restaurant	3

```
1]: cluster3 =clusters[clusters['Cluster_Labels']==3]
```

In Cluster 3 , we see that there is rich mix of brewery, ba

```
.]: cluster3.groupby(['Neighborhood','Venue_Category']).count()
```

		Cluster_Labels
Neighborhood	Venue_Category	
Arabic College S.O	Brewery	1
	Chinese Restaurant	1
	Coffee Shop	2
	Falafel Restaurant	1
	Fast Food Restaurant	6
	Indian Chinese Restaurant	1
	Indian Restaurant	9
	Italian Restaurant	1
	Korean Restaurant	3
	Lounge	2
	Pub	1
	Restaurant	1
	Seafood Restaurant	1
	Shopping Mall	1
	South Indian Restaurant	1
	Vegetarian / Vegan Restaurant	1
	Andhra Restaurant	1
	Bar	1
	Bengali Restaurant	1
	Brewery	1
Bannerghatta Road S.O	Chinese Restaurant	1
	Coffee Shop	2
	Eastern European Restaurant	1
	Fast Food Restaurant	1
	Indian Restaurant	10
	Italian Restaurant	1
	Lounge	1
	Middle Eastern Restaurant	1
	Restaurant	1
	South Indian Restaurant	2
	Asian Restaurant	3

On further analysis I found that there are two neighborhoods in this cluster that doesn't have a brewery but have a pubs , bars and restaurants.

Now lets see if any neighborhood has bars/pu

```
] cluster3_bars_pubs = cluster3_group_clustered[(cluster3_group_clustered
```

```
] cluster3_bars_pubs[['Neighborhood', 'Bar', 'Pub', 'Brewery']]
```

```
] Neighborhood Bar Pub Brewery
```

5	Carmelram S.O	2	1	1
7	Fraser Town S.O	1	2	0
8	HAL II Stage H.O	4	5	3
14	Mahatma Gandhi Road S.O	1	5	4
15	Malleswaram S.O	1	2	0
19	St. John's Medical College S.O	2	1	2

5. Conclusion

As discussed above we can see that using the location data of Bangalore neighborhoods and also the venue information from Foursquare API,

I was able to create cluster of neighborhood that have mix of restaurants, bars, cafe, pubs etc..

And on further analyzing , I was able to get to the neighborhoods that are suitable to host new Brewery.

