The Battle of the Neighborhoods - Chennai

by,

Lingesh R

1. Introduction:

Chennai, Tamil Nadu is home to a large number of business, from micro to big companies..It is also a great tourist destination. In this project we will determine the optimal location to open a restaurant in the Chennai district, and will be targeted at stakeholders interested in opening a restaurant in Chennai. Since we are not aware of what type of restaurant is popular in Chennai, we will also determine the most popular type(s) of restaurants in Chennai. We will utilize data science and machine learning techniques in order to do this. This project will be of interest to anyone wanting to open a restaurant in Chennai, and as a secondary application can also be useful to anyone wanting to visit or live in Chennai.

2. Data

In this project we will utilize three datasets in order to solve our problem. The first dataset is a list of municipalities. This data set was scraped from Wikipedia using BeautifulSoup. The next data set is one contain the latitude and longitude for each municipality. This dataset was obtaining using Geocoder package. The file containing the list of venues was obtained using the Foursquare API. We selected all venues not just restaurants in order to see where restaurants were more popular. These are the top 5 rows of these data sets.

Neighborhood data:

	Neighbourhood
0	Red Hills
1	Royapuram
2	Korukkupet
3	Vyasarpadi
4	Tondiarpet

Neighborhood with latitude and longitude coordinates:

	Neighbourhood	Latitude	Longitude
0	Red Hills	13.19543	80.18431
1	Royapuram	13.11394	80.29420
2	Korukkupet	13.11680	80.27726
3	Vyasarpadi	13.11778	80.25168
4	Tondiarpet	13.12923	80.28955

Combined with Foursquare API data:

	Municipalities	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Red Hills	13.19543	80.18431	Hotel Balaji Bavan	13.193716	80.185292	Indian Restaurant
1	Red Hills	13.19543	80.18431	Pvr Cinemas Skls Galaxy Mall	13.191586	80.186100	Movie Theater
2	Red Hills	13.19543	80.18431	Skls Galaxy Mall	13.191500	80.186038	Shopping Mall
3	Red Hills	13.19543	80.18431	Lakshmi Theatre	13.189894	80.188758	Multiplex
4	Red Hills	13.19543	80.18431	Universal Gym	13.190710	80.177946	Gym

Some of the features extracted from the data were, the latitude and longitude of each municipality, and from the foursquare API we extracted the venues, the venue category, and its geo-coordinates. All of this data can be used to help us in our analysis and problem solving, which will be discussed more thoroughly in the methodology and analysis sections.

3. Methodology

After cleaning the data, we began the analysis of the data by utilizing one hot encoding on the venues data, this makes it easier to work with categorical variables, in this case the venue categories, which are needed in order to solve this problem and do the analysis. After this we will do a frequency analysis using the mean of the frequency of occurrence of each category. This will aid us in when we utilize clustering, as well as in identifying the ideal locations for a restaurant. After doing this we sort the data into a data frame displaying the top 10 venues, in order to make it easier to work with.

The next step is to begin the clustering process, for this project we have chosen to use k-means clustering. We have chosen this over other methods for several reasons. First of all, it is easier to implement, and is relatively efficient for our data set size. Second of all we are not aware of any arbitrary clusters in our dataset, so we rule out using DBSCAN.

The first step in k-means clustering is to choose the optimal or best value for ‘k’, i.e. the number of clusters. For this project we have chosen to do so using the elbow method. Using this method, we plot the values for ‘k’ on the X axis and the distortion on the Y axis (the values calculated with the cost function). When K increases, the centroids are closer to the cluster’s centroids. The improvements will decline, at some point rapidly, creating the elbow shape. The point located on the ‘elbow’ is our optimal ‘k .

After we choose the optimal ‘k’ we will create the model and fit it to the data and create a new data frame that includes the cluster and the top 10 venues. Using this we create a map to display the clusters and display the individual clusters.

4. Results and Discussion

One-hot encoded Dataframe:

	Municipalities	Accessories Store	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Amphitheater	Andhra Restaurant	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Astrologer	Athletics & Sports	BBQ Joint	Bakery	Bank
0	Adyar	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.014706	0.0	0.000000	0.044118	0.0	0.000000	0.014706	0.014706	0.000000
1	Alandur	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.125000	0.000000
2	Alwarthirunagar	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
3	Ambattur	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
4	Aminjikarai	0.055556	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
5	Anna Nagar	0.000000	0.000000	0.000000	0.00	0.014706	0.000000	0.000000	0.000000	0.0	0.000000	0.014706	0.0	0.000000	0.014706	0.029412	0.000000
6	Annanur	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
7	Arumbakkam	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
8	Ashok Nagar	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.040000	0.000000

Frequency analysis:

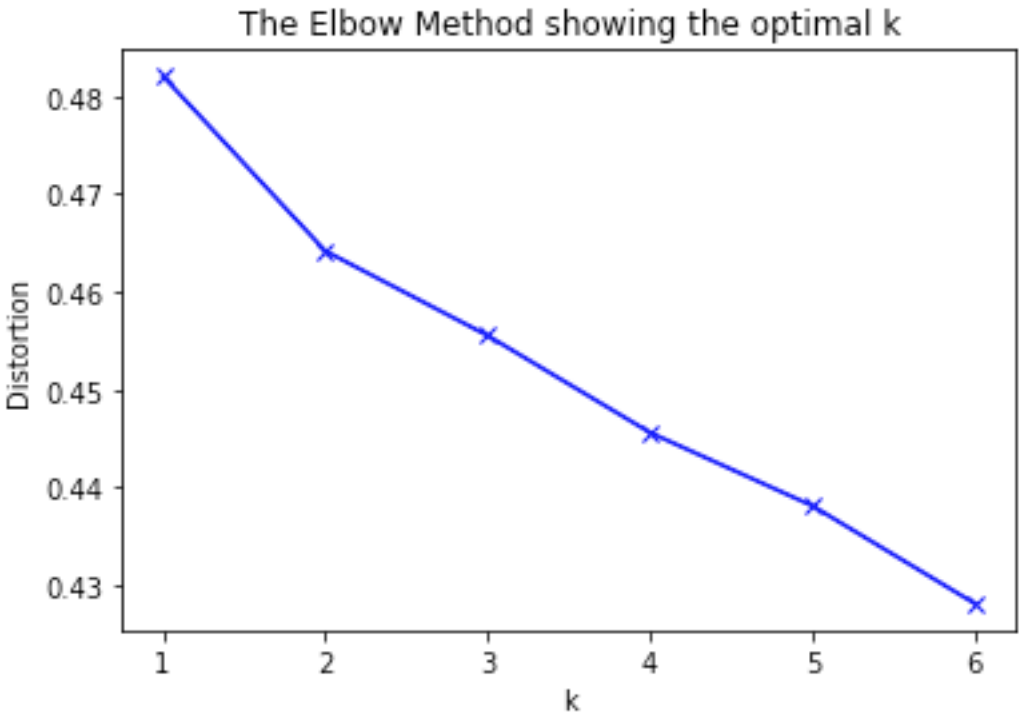
```
----Adyar----
                venue  freq
0  Indian Restaurant  0.25
1               Café  0.06
2   Asian Restaurant  0.04
3  Fast Food Restaurant  0.04
4        Pizza Place  0.04
```

```
----Alandur----
                venue  freq
0  Indian Restaurant  0.19
1            Bakery  0.12
2      Metro Station  0.12
3          Hotel Bar  0.06
4               Café  0.06
```

```
----Alwarthirunagar----
                venue  freq
0       Clothing Store  0.23
1  Fast Food Restaurant  0.15
2        Pizza Place  0.08
3                Gym  0.08
4               Café  0.08
```

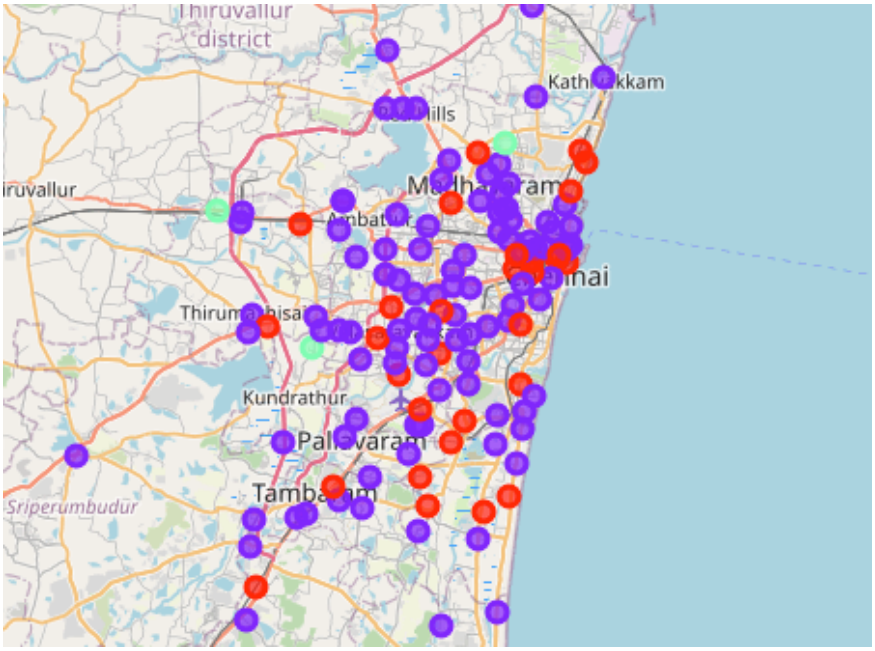
```
----Ambattur----
                venue  freq
0          Food Court  0.50
1               Café  0.25
2  Fast Food Restaurant  0.25
3   Accessories Store  0.00
4  New American Restaurant  0.00
```

Using Elbow method to find optimal value of k:



Based on the above graph, the optimal value for ‘k’ was chosen as 3.

Map of clusters formed:



Cluster 1:

	Neighbourhood	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	10th Most Common Venue
5	Tiruvottiyur	Indie Movie Theater	Indian Restaurant	Beach	Historic Site	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food
12	Sowcarpet	Indian Restaurant	Restaurant	Market	Food Truck	Harbor / Marina	Video Store	Dessert Shop	Fish Market	Cricket Ground	Food Court
15	Central	Indian Restaurant	South Indian Restaurant	Movie Theater	Fish Market	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
20	Parry's Corner	Indian Restaurant	Market	Video Store	Museum	Fish Market	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food
21	Purasawalkam	Indian Restaurant	Hotel	Italian Restaurant	Pizza Place	South Indian Restaurant	Hostel	Juice Bar	Historic Site	Fast Food Restaurant	Farmers Market
25	New Washermenpet	Pharmacy	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
30	Periamet	Indian Restaurant	Bookstore	Pizza Place	Hotel	Vegetarian / Vegan Restaurant	Soccer Stadium	Nightclub	Farmers Market	Movie Theater	Fish Market
31	Choolai	Indian Restaurant	Ice Cream Shop	Soccer Stadium	Hotel	Clothing Store	Women's Store	Fast Food Restaurant	Food Truck	Food Court	Food & Drink Shop
37	Manjambakkam	Indian Restaurant	Lake	Women's Store	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
40	T.V.K. Nagar	Indie Movie Theater	Indian Restaurant	Historic Site	Coworking Space	Cricket Ground	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
42	Lakshmipuram	Indian Restaurant	Intersection	Department Store	Hotel	Photography Studio	American Restaurant	Flea Market	Furniture / Home Store	Fruit & Vegetable Store	Fried Chicken Joint
47	Avadi	Indian Restaurant	Department Store	Shopping Mall	Movie Theater	Fish Market	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
51	Broadway	Indian Restaurant	Market	Food Truck	Restaurant	Video Store	Department Store	Dessert Shop	Fried Chicken Joint	Cricket Ground	Food Court
55	Kosapet	Indian Restaurant	Department Store	Hotel	Women's Store	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food
58	manapakkam	Indian Restaurant	Trail	Juice Bar	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop
62	Defence Colony	Indian Restaurant	Market	Movie Theater	Middle Eastern Restaurant	Hotel Pool	Fruit & Vegetable Store	Food Truck	Food Court	Ice Cream Shop	Food & Drink Shop
70	Maduravoyal	Indian Restaurant	Movie Theater	Fast Food Restaurant	Pizza Place	Hotel Pool	Food Truck	Food & Drink Shop	IT Services	Food	Flower Shop
74	Karambakkam	Indian Restaurant	Lake	Burger Joint	Bakery	Restaurant	Donut Shop	Dessert Shop	Fried Chicken Joint	Food Truck	Department Store
76	Saligramam	Indian Restaurant	Pizza Place	Hotel	Clothing Store	Movie Theater	Bookstore	Ice Cream Shop	Hotel Bar	Farmers Market	Food & Drink Shop
103	Gopalapuram	Indian Restaurant	Restaurant	Hotel	Café	Ice Cream Shop	Juice Bar	Concert Hall	Tea Room	Grocery Store	Dessert Shop
110	Adyar	Indian Restaurant	Café	Juice Bar	Fast Food Restaurant	Asian Restaurant	Pizza Place	Bookstore	Ice Cream Shop	Electronics Store	North Indian Restaurant
117	Madipakkam	Indian Restaurant	Convenience Store	Department Store	Women's Store	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food
119	Velachery	Indian Restaurant	Fast Food Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Pizza Place	Sandwich Place	Ice Cream Shop	Bakery	Accessories Store
123	Kovilambakkam	Indian Restaurant	Ice Cream Shop	Food Truck	Shipping Store	Hotel	Women's Store	Fast Food Restaurant	Food Court	Food & Drink Shop	Food
126	Injambakkam	Indian Restaurant	Restaurant	Indie Movie Theater	Bistro	Burger Joint	Food & Drink Shop	Café	Chinese Restaurant	Women's Store	Food Truck
134	Kadaperi	Indian Restaurant	Light Rail Station	Jewelry Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
136	Pazhavanthangal	Indian Restaurant	Hotel	Department Store	Supermarket	Café	Multiplex	Grocery Store	Restaurant	Breakfast Spot	Comfort Food Restaurant

Cluster 2:

	Neighbourhood	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	10th Most Common Venue
0	Red Hills	Indian Restaurant	Shopping Mall	Movie Theater	Gym	Multiplex	Fish Market	Food Court	Food & Drink Shop	Food	Flower Shop
1	Royapuram	Indian Restaurant	Ice Cream Shop	Convenience Store	Dessert Shop	Spa	Movie Theater	Fish Market	Food Truck	Food Court	Food & Drink Shop
2	Korukkupet	Vegetarian / Vegan Restaurant	Indie Movie Theater	Boutique	Department Store	Dessert Shop	Fruit & Vegetable Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop
3	Vyasarpadi	Pharmacy	Department Store	Bakery	Chinese Restaurant	Fish Market	Fruit & Vegetable Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop
4	Tondiarpet	Harbor / Marina	Historic Site	Pharmacy	Beach	Fried Chicken Joint	Diner	Flea Market	Cricket Ground	Food Truck	Food Court
7	Minjur	Park	Women's Store	Fast Food Restaurant	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
8	Old Washermenpet	Women's Store	Department Store	Fish Market	Fruit & Vegetable Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop

Cluster 3:

	Neighbourhood	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	10th Most Common Venue
18	Mathur MMDA	Pharmacy	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market
80	Thirunindravur	Pharmacy	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market
90	Mangadu	Pharmacy	Fruit & Vegetable Store	Coworking Space	Cricket Ground	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop

Based on the above results, the most popular restaurant in Chennai is the Indian restaurant, being the most common venue restaurant for 21 neighborhoods and showing as the 2nd most common venue for other neighborhoods in cluster 1. It should also be noted that restaurants in general are in the top 10 for all districts in cluster 2 as well. Based on these results the recommended type of restaurant to open is an Indian restaurant, based one where they are most common, we feel they would have most success in any district in cluster 1, where it is not shown as 1st most common venue. If stakeholders wish to avoid competition, they can choose a district in cluster 2 where Indian restaurants do not show up in the top 10, such as Minjur or Old Washermenpet, as it can

be assumed, they will have success there due the similarities they share with other clusters. We do not recommend opening an Indian restaurant in cluster 3, as there are no restaurants appearing in the top 10. The decision however is ultimately left to the stakeholder's discretion.

5. Conclusion

The purpose of this project was to find the best type of restaurant to open in Chennai, Tamil Nadu and a suitable location for it. Based on the analysis we conducted we recommend opening an Indian restaurant in any one of the neighborhoods located in cluster 1, based on how common it is in this area. We leave the decision of which district specifically up to stakeholders. It should be noted however, that this report should not be the sole criterion for making this decision. Analyses such as these are only a tool to aid the decision-making process and are only as good as the data and methods (which are never perfect) used in them. We recommend using other methods, tools and forms of decision-making criteria before reaching a final decision.