

Coursera Capstone Project

In accordance with *IBM Data Science Professional Certificate*, Course:9, *Applied Data Science Capstone*

Opening up of a new Mexican Restaurant in the city of *Bangalore, India.*

by: T. Neethi Thevan

-July 2019-



Aerial view of UB City at night. Bangalore, India

About the city,

Bangalore, officially known as Bengaluru, is the capital of the Indian state of Karnataka. It has a population of about 10 million and a metropolitan population of about 8.52 million, making it the third most populous city and fifth most populous urban agglomeration in India. Located in southern India on the Deccan Plateau, at a height of over 900 m (3,000 ft) above sea level, Bangalore is known for its pleasant climate throughout the year. Its elevation is the highest among the major cities of India. Bangalore is widely regarded as the "**Silicon Valley of India**" (or "IT capital of India") because of its role as the nation's leading information technology (IT) exporter. Indian technological organisations such as ISRO, Infosys, Wipro and HAL are headquartered in the city. A demographically diverse city, Bangalore is the second fastest-growing major metropolis in India. Recent estimates of the metro economy of its urban area have ranked Bangalore either the fourth or fifth-most productive metro area of India. It is home to many educational and research institutions in India, such as Indian Institute of Science (IISc), Indian Institute of Management (Bangalore) (IIMB), International Institute of Information Technology, Bangalore (IIITB), National Institute of Fashion Technology, Bangalore, National Institute of Design, Bangalore (NID R&D Campus), National Law School of India University (NLSIU) and National Institute of Mental Health and Neurosciences (NIMHANS). Numerous state-owned aerospace and defence organisations, such as Bharat Electronics, Hindustan Aeronautics and National Aerospace Laboratories are located in the city. The city also houses the Kannada film industry.

Business Problem Statement,

My client, a multinational Mexican Restaurant chain, *Tacolicious* is planning to open up a restaurant in the city of Bangalore. They prefer that location of the restaurant should be within the city limits as well as should be as far as it can be from any other Mexican restaurant. This is to attract more customers and to reduce the competency.

Target Audience of this project,

The analysis can be very help to particular person or an organisation that are keen on opening a restaurant in the city of Bangalore. They can be befitted by the knowledge of different hotels in a particular locality and which can be used to make important decision on the setting up of the restaurant.

Data

- **Data Requirement**

1. The name of all neighbourhoods in the city of Bangalore.
2. The latitude and longitude of all the neighbourhoods.
3. Information about the type of restaurant/hotels in a particular locality.

- **Data collection**

1. For administrative purposes, the city of Bangalore is divided into nine zones, which are further subdivided into a total of 198 wards administered by the Bruhat Bengaluru Mahanagara Palike (BBMP). The names of localities of Bangalore can be found in terms of city corporation wards. This data is available at the link given below,
url: '<http://bbmp.gov.in/ward-maps>'
2. For finding the latitude and longitude of these neighbourhood was be found by scrapping the google web search results.
3. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests. The API accepts coordinates of a place as request and returns the details of top 100 popular venues surrounding the place.
The results include a large variety of venues including parks, ATM, shopping malls etc. We can filter out the names and type of restaurants from them easily.

- **Data Analysis**

After the procurement of the data EDA and data wrangling operations are done. Basically, all the data is converted in to a data frame and ***K-means clustering algorithm*** is used to see how the different neighbourhood clusters. ***Folium map*** is used to visualize the locations of all neighbourhood in Bangalore city. ***Wordcloud*** visualization is used to examine the properties of different clusters.

Methodology

- **Neighbourhood names**

The names of the neighbourhood are available in the above mentioned url . Python library Pandas was used to read the html content of the web page and convert the data into a dataframe.

- **Latitude and Longitude of neighbourhoods**

The latitude and longitude of each neighbourhood was found by web scrapping *Google* search. For the missing values the latitude and longitude data we added manually. Python library *Beautifulsoup* was used to scrape the google search results. A dataframe containing neighbourhood names and their corresponding latitude and longitude is formed.

- **Bangalore Map**

folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the *leaflet.js* library. Manipulate your data in Python, then visualize it in on a Leaflet map via *folium*.

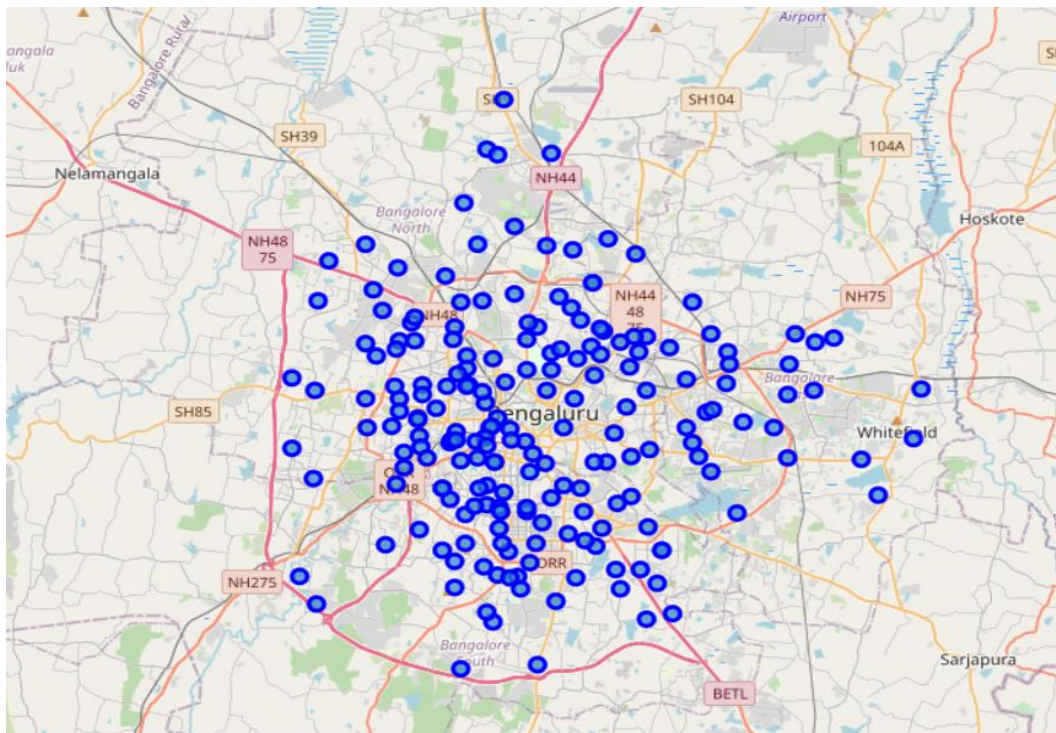


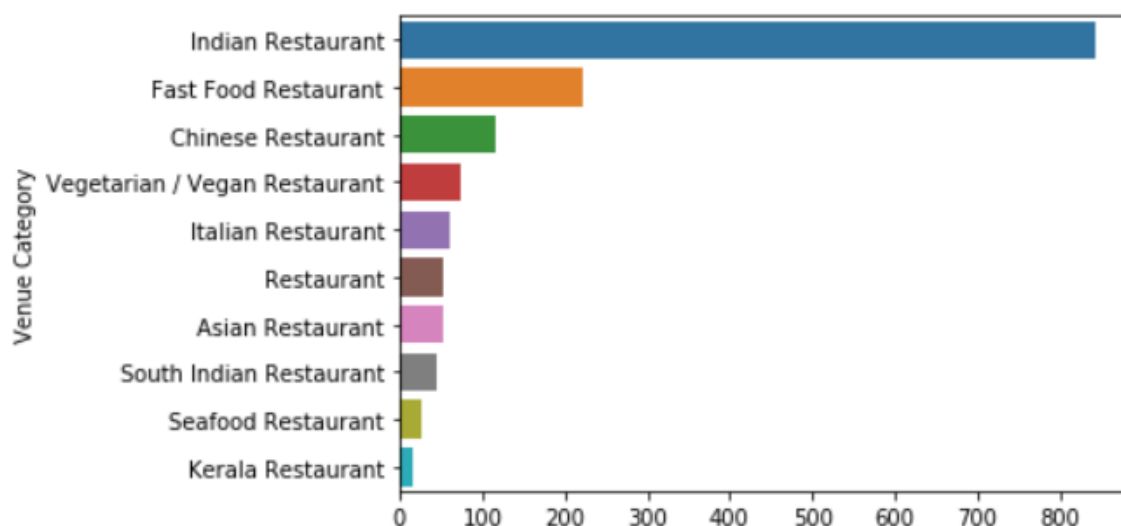
Fig: The generated Bangalore city map with neighbourhoods marked

• Venue Details

FourSquare API was used to retrieve details of top 100 venue in terms of popularity within a radius of 1 km surround each neighbourhood. This data involves the details of places such as ATM, park, hospital, educational institutes etc which are not important for our analysis. Then we filter out only the details of restaurant separately.

• Data Analysis

The number of restaurants in whole Bangalore was found and comparison was based on the top 10 restaurants types based on its number.



We can infer that substantially there are many of number of Indian Restaurants in Bangalore followed by fast food restaurants.

• One hot encoding

As categorical data is of no use for cluster analysis. We have to use one hot encoding method to convert the restaurant names to either 0 or 1 based on their presence in the particular neighbourhood.

›J:

	Neighbourhood	Afghan Restaurant	American Restaurant	Andhra Restaurant	Asian Restaurant	Bengali Restaurant	Chettinad Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Eastern European Restaurant	Rest
0	Kempegowda	0	0	0	0	0	0	0	0	0	0	0	
1	Kempegowda	0	0	0	0	0	0	0	0	0	0	0	
2	Kempegowda	0	0	0	0	0	0	0	0	0	0	0	
5	Kempegowda	0	0	0	0	0	0	0	0	0	0	0	
7	Kempegowda	0	0	0	0	0	0	0	0	0	0	0	

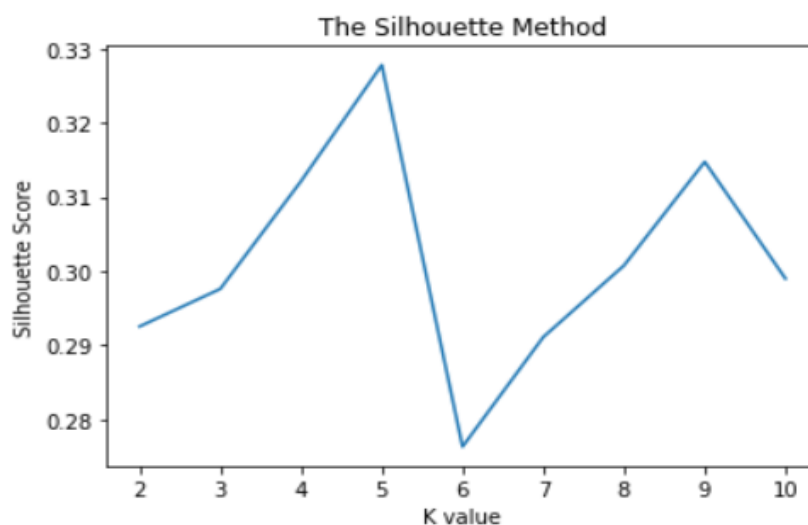
• Top 10 restaurants

All there are many restaurants available in a particular neighbourhood. Top 10 restaurants in terms of frequency was found concerning with a particular neighbourhood.

	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	A. Narayanapura	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	French Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant
1	Adugodi	Indian Restaurant	Kerala Restaurant	Punjabi Restaurant	Tibetan Restaurant	Eastern European Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant
2	Agaram	Indian Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant	Hyderabadi Restaurant
3	Agrahara Dasarahalli	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant
4	Arakere	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	South Indian Restaurant	Vietnamese Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant

• Silhouette score analysis

The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from -1 to $+1$, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. The silhouette can be calculated with a distance metric, such as the Euclidean distance . The silhouette score for different number of centroids were calculated and plotted in a graph.



The Optimum value of K is 5

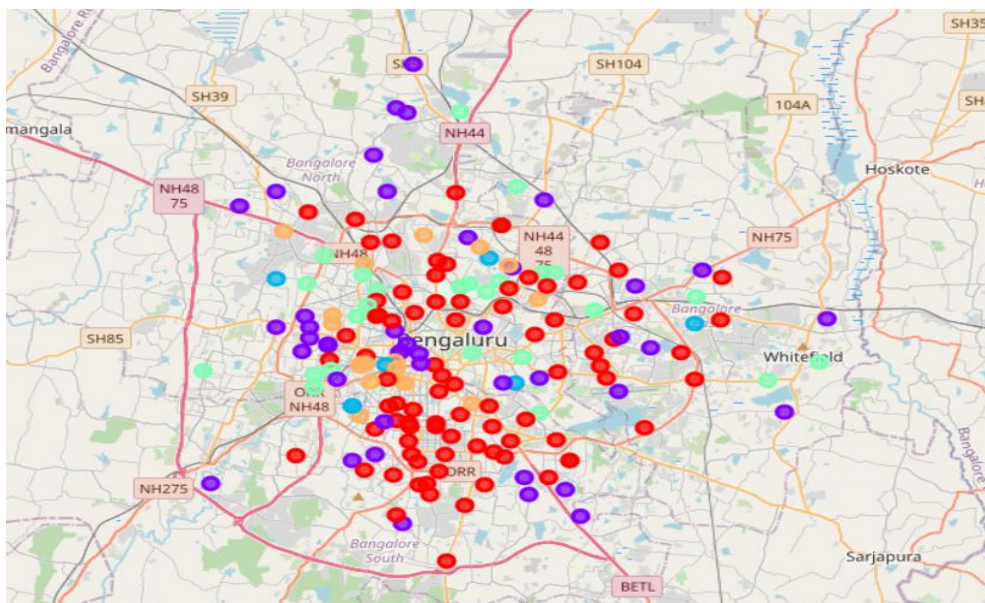
- **K – means clustering**

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes. The objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset. We use this algorithm to identify to which cluster a particular neighbourhood belongs.

Cluster Labels		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	0	A. Narayanapura	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	French Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant
1	0	Adugodi	Indian Restaurant	Kerala Restaurant	Punjabi Restaurant	Tibetan Restaurant	Eastern European Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant
2	1	Agaram	Indian Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant	Hyderabadi Restaurant
3	0	Agrahara Dasarahalli	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Kerala Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant
4	0	Arakere	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	South Indian Restaurant	Vietnamese Restaurant	Kebab Restaurant	Karnataka Restaurant	Japanese Restaurant	Italian Restaurant	Indian Chinese Restaurant

- **Mapping the clusters**

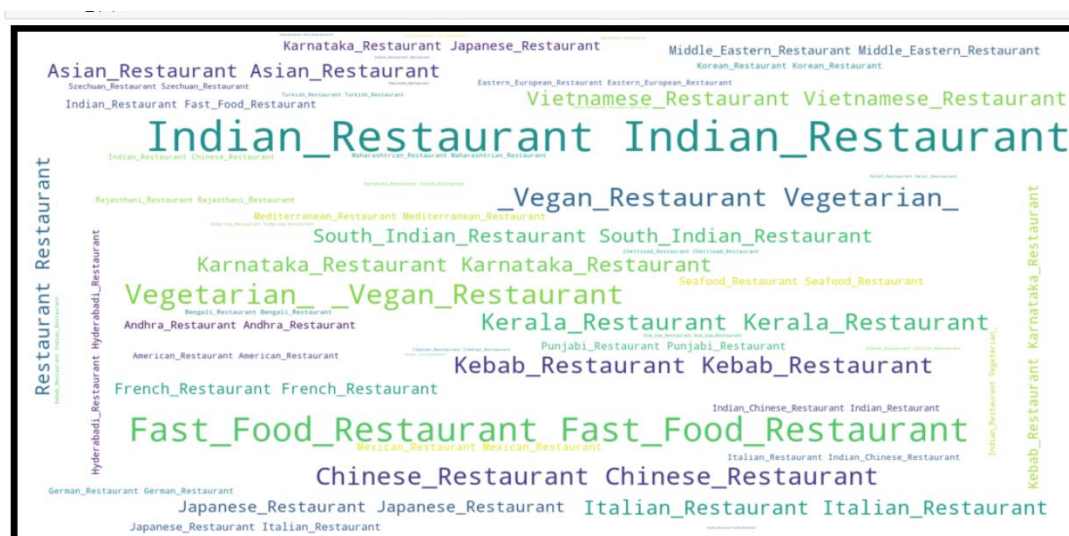
The neighbourhoods were mapped in folium leaflet with different colour in according to their cluster.



Results

All the clusters were examined to find their characteristics. *Wordcloud* was used to visualize the clusters. *Wordcloud* is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. The restaurant names available under each neighbourhood was taken and the frequency of the names were modified based on their position.

- **Cluster 0**



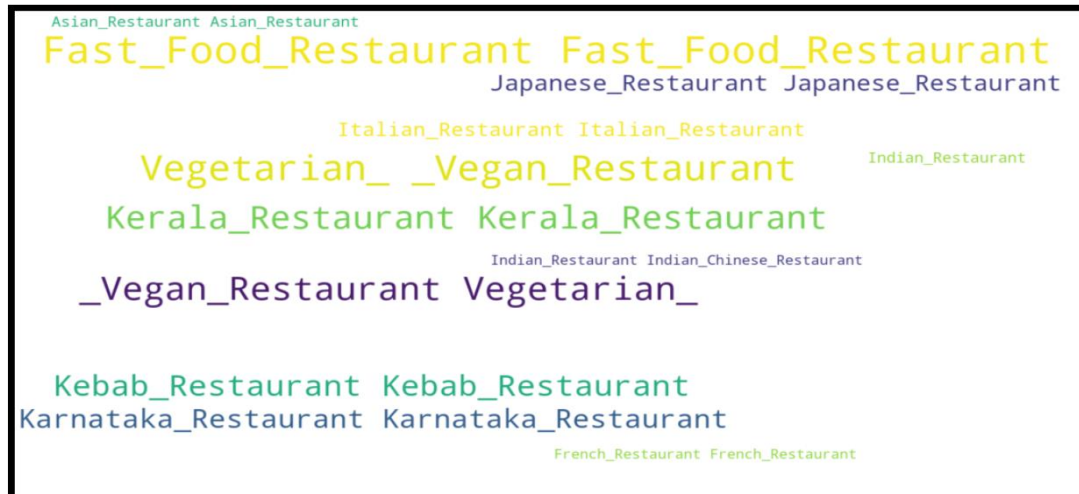
From the wordcloud we can infer that the neighbour hood under **cluster 0** has more number of *Indian and Fast Food Restaurants* followed by *Kerala Restaurant*

- **Cluster 1**



From the wordcloud we can infer that the neighbour hood under **cluster 1** has more number of *Indian Restaurant* and *Vietnamese Restaurant*

- Cluster_2



From the wordcloud we can infer that the neighbour hood under **cluster 2** has more number of **Fast Food Restaurants** followed by **Vegan/Vegetarain Restaurant** and **Kerala_Restaurant**

- Cluster_3



From the wordcloud we can infer that the neighbourhood under **cluster 3** has more types of restaurant ,then any other cluster. The neighbourhood falling under this cluster mainly a part of city centre.

- **Cluster_4**



From the wordcloud we can infer that the neighbourhood under cluster 3 has more number of *Indian Restaurants* followed by *Fast Food Restaurant* and *Kebab Restaurant*

Conclusion

Analysing all the clusters tells us that the best place to open a Mexican Restaurant could be one of the places that falls under cluster_3 as there are no hotels that serve Mexican foods in those neighbourhood.

The suggested neighbourhoods are,

1. Kaval Byrasandar
2. Rajagopalanagar
3. Garudacharpalya
4. Vannarpet
5. Binnypet
6. Srinagara
7. Girinagara