

# **Using Foursquare API to analyze Chennai's neighbourhoods**

## **Background**

Chennai, formerly known as Madras, is the capital of the Indian state of Tamil Nadu. Located on the Coromandel Coast of the Bay of Bengal, it is one of the largest cultural, economic and educational centres of south India. According to the 2011 Indian census, it is the sixth-most populous city and fourth-most populous urban agglomeration in India. The city together with the adjoining regions constitutes the Chennai Metropolitan Area, which is the 36th-largest urban area by population in the world. The traditional and de facto gateway of South India, Chennai is among the most-visited Indian cities by foreign tourists because of its culture and famous gastronomy. Each neighbourhood have different number of restaurants/food stalls/cafés.

## **Problem**

Data that might contribute to determining the number of restaurants/food stalls/cafés repartition among the neighbourhoods of Chennai. This project aims to describe the neighbourhoods which the highest number of restaurants/food stalls/cafés.

## **Interest**

I would be interesting to understand the repartition of restaurants/food stalls/cafés among the neighbourhoods to move there to eat easily when we are too busy to cook.

## **Data**

### **Data source**

The data for Chennai city was sparse online:

*url = 'https://en.wikipedia.org/wiki/Areas\_of\_Chennai'*

### ***Data cleaning***

I've manually scraped the list of neighbourhoods from this Wikipedia page using bs4, and then grabbed all the hyperlinks. Using urllib, these links are visited individually and the coordinates and pincodes are scraped and put into a pandas dataframe.

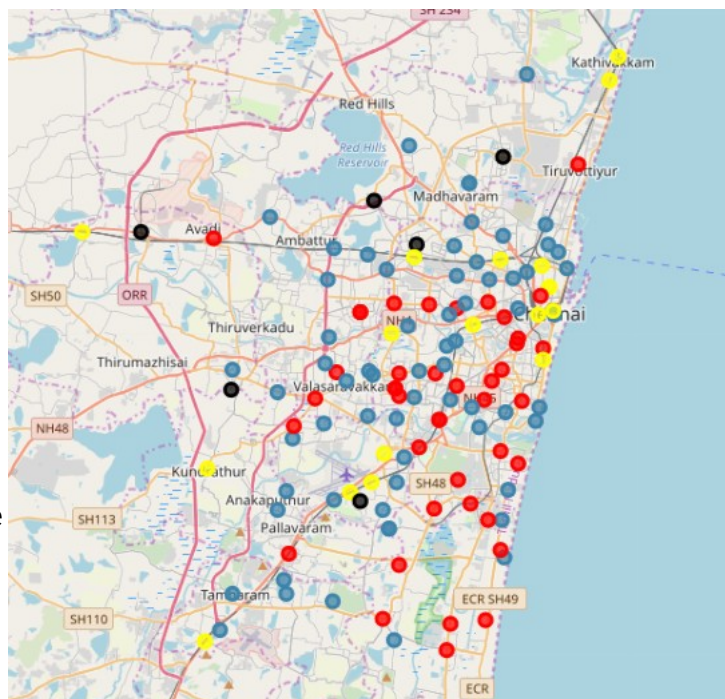
The next step is to get all the venues in each neighbourhood within a specified radius, in this case, 500 metres. For that i used Foursquare

```
url = 'https://api.foursquare.com/v2/venues/explore'
```

After doing this for each neighbourhood, the resulting venues are again put into a pandas dataframe.

### **Methodology**

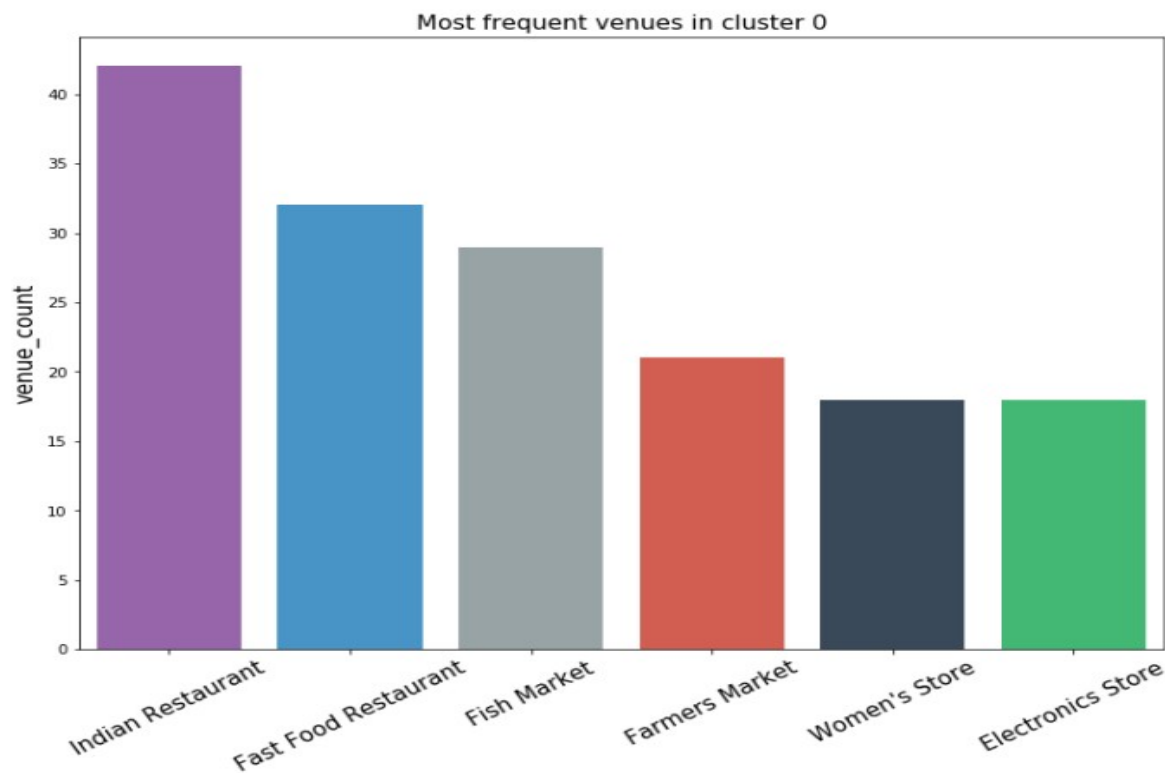
After this, the venue categories are one-hot encoded and then the 10 most frequent venues in each neighbourhood are found using code that I totally did not copy-paste from the tutorial notebooks. Finally, k-means clustering is used to group these neighbourhoods. I forgot to choose the best k for this algorithm. I didn't notice until after I published this. Oh well. After they are clustered, we can use Folium (which is a map visualization library for Python), to see all the neighbourhoods and the cluster they belong to on a nice map:



Cluster 0 (Red) have Indian Restaurants as their most frequent venue.

Cluster 1 (Blue)'s frequent venues are mostly related to food. Evident by counting the occurrences.

Ice Cream Shop	9
Fast Food Restaurant	5
Vegetarian / Vegan Restaurant	5
Pizza Place	5
Coffee Shop	4



Not much can be inferred from this map. Due to the fact that the clusters are grouped together not due to their Euclidean distances on the map, but due to properties of the venues themselves. Inspecting the dataframe which has the cluster labels along with the most frequent venue in each neighbourhood, we can see that:

Cluster 0 (Red) have Indian Restaurants as their most frequent venue.

Cluster 1 (Blue)'s frequent venues are mostly related to food. Evident by counting the occurrences of each.

Use these results, we can solve our trivial problem. Neighbourhoods in either Cluster 0 or Cluster 1, are the ones with the highest concentration of restaurants and other food-related venues.

## Results

While it looks like we have solved our problem, there is one flaw. Of course, Foursquare data for the city of Chennai is considerably sparse compared to other well developed cities. It also probably does not take into factor the ton of small food shops scattered throughout the city. Using Foursquare, the individual ratings for each venue could also be retrieved, but it did not seem to have rating data for Chennai. This would've helped pick out individual restaurant

suggestions and propose a neighbourhood with highly rated restaurants.

## **Conclusion**

Anybody would not have trouble finding food in the city of Chennai. For lesser travel times, we can choose any of the neighbourhoods in Cluster 0, however we'll find that most of them are Indian restaurants. Independent of travel distance, the cluster choice does not matter much as there are more restaurants than any other venue. There is simply not enough data to do an in-depth analysis. However, individually marking the venues which are food-related is also a possibility to do for the future.

The outcome of the project was impacted by the limited effectiveness of the Foursquare API for a city like Chennai. Using some other more developed city would've probably yielded better results. The way the neighbourhoods were suggested were majorly due to Indian restaurants alone.

Further analysis could've been done by using the rating of each venue however yet again ratings were not available for venues in Chennai. If ratings were available, individual restaurants in each locality could've been suggested.