

June 25,2020

Setting up an Indian Restaurant in Chennai

Jawahar RB

Note: This project is a capstone project and is part of the Coursera IBM Data Science Certification.

Introduction

Background

Recent Demographic surveys have shown that Chennai Metropolitan Area is one of the most densely packed regions in India. Interestingly due to immense job opportunities in Chennai, it is also one of the most culturally diverse regions with people from all over India. Throughout the years we have seen a strong increase in ethnic food in the county itself which correlates with the increase in a diverse population. Studies show that there is a shortage of number of restaurants to population of the city.

The reason for choosing an Indian restaurant is mainly because people in Chennai mostly prefer Indian cuisine and there is a huge shortage of restaurants in the densely populated city with a density of 26,553 people per square kilometres.

When deciding the most prominent location for a Indian Restaurant, it is important to consider some important factors:

Demand

For this problem, we can choose a neighborhood that contains people frequently visiting Indian restaurants compared to other places.

Location

To have an effective restaurant business, it is always a beneficial step to set it up in a densely populated location. By observing metropolitan demographics in general there are strong consensus and statistical evidence that shows population density is higher closer to the downtown area of a city.

Rent Price

To make sure the new business can financially be at an advantage we will choose the area that has the decent rent.

June 25,2020

Data and Resources:

To achieve our target, we use the following sources and services:

BeautifulSoup

This Wikipedia page will be the source of identifying all the neighborhoods that exist in the Chennai area: https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Chennai

GeoPy

We will be using GeoPy to map our neighborhoods. This API allows us to locate the precise coordinates of a location using numerous data sources.

Folium

To map our coordinates on a map that can be visualized, we will be using the folium API.

Foursquare API

The [Foursquare API](#) gives us access to the most popular locations in a given area. Using the above-mentioned services and the API we can pin-point the restaurant demographics in the Neighborhood

Methodology

Setup

The first and foremost step taken for this project is to collect a list of all the Neighborhoods of Chennai. To achieve this the most helpful source was Wikipedia itself. Therefore, we used BeautifulSoup to scrape that data. Some neighborhoods which weren't prominent parts of Chennai area were removed from the scraping. Getting the names is enough but the next important step is to be able to visualize those neighborhoods on a map to determine their exact locations. For this, we used GeoPy Nominatim, a geolocator that reads the list of Neighborhood names and found their respective geo coordinates. Next, we used folium to map out the exact coordinates.

The next step taken was to analyze the neighborhoods and see the activity present in the respective areas. We took advantage of the Foursquare API. The API allows us to see the most popular venues in a given location. The venues are then analyzed in detail to find existing patterns. For example, coffee shop density and recreational activity venues in a given area.

June 25,2020

These patterns are more visible when we group neighborhoods in a cluster. Therefore, we will be using K-Means clustering to find suitable clusters to our goal. The “n” number of clusters will be analyzed, and we will be choosing the cluster that fits our desired goal.

Note: One method of picking out the best cluster is to see the spread of restaurants.

After the cluster is selected, we will further analyze it and find the number of Indian Restaurants in the area. This helps us further shortlist our possible locations. As mentioned above we will be keeping in context the distance from the Center of Chennai. As a bonus, to ensure the restaurant can be financially stable we will pick the area with decent rent price.

Note: It is very important to understand that when it comes to setting up a restaurant strategically there are a lot of other factors involved. This project serves to target the fundamentals.

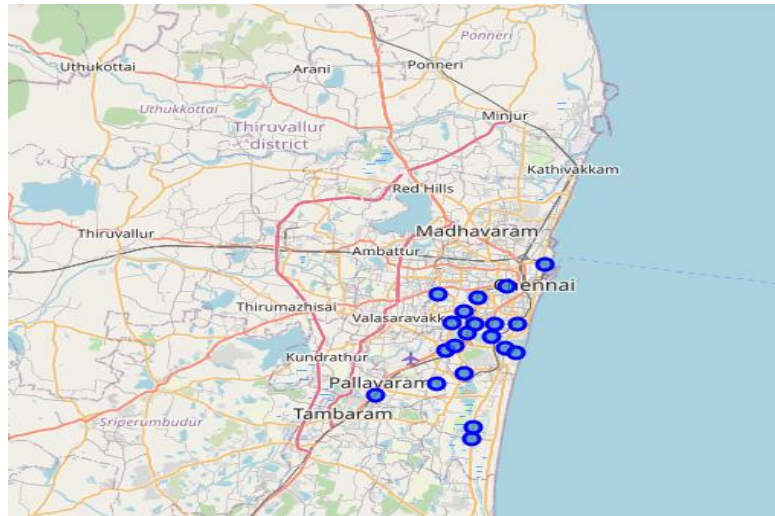
Data Analysis

To get a good list of the Neighborhoods in Los Angeles we will be using Wikipedia, “List of neighborhoods of Chennai”. Using BeautifulSoup and GeoPy we get a list of Neighborhoods and their coordinates. This has become our finalized neighborhood Data Frame before we further process it using Foursquare:

	City	Neighborhood	Latitude	Longitude
0	Chennai	Adyar	13.006300	80.257400
1	Chennai	Alandur	13.003000	80.204000
2	Chennai	Alwarpet	13.033900	80.248600
3	Chennai	Ashok Nagar	13.035100	80.209500
4	Chennai	Basant Nagar	13.000200	80.266800
5	Chennai	Chrompet	12.951610	80.140970
6	Chennai	Egmore	13.078000	80.259000
7	Chennai	Guindy	13.008413	80.212688
8	Chennai	Karapakkam	12.914136	80.229285
9	Chennai	Kotturpuram	13.019600	80.245800

Using the coordinates and Folium we project these coordinates on a geographic map of Chennai with the markers in blue point towards each location.

June 25,2020

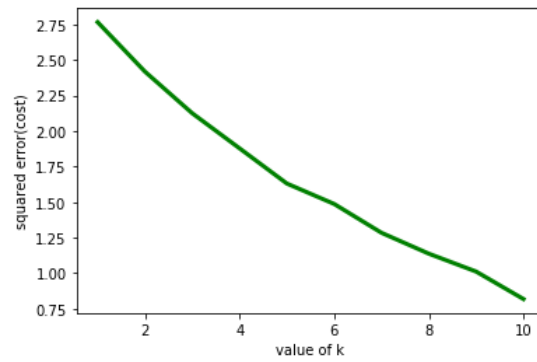


The next step is to use the previously acquired credentials and get access to the Foursquare API. As previously mentioned, the API provides us with location data of all the venues in their respective neighborhoods. We can get access, to restaurant names, the category, and the category counts to understand the commercial demographics better. Over here we can see the most popular venues from the venue 'Adyar'.

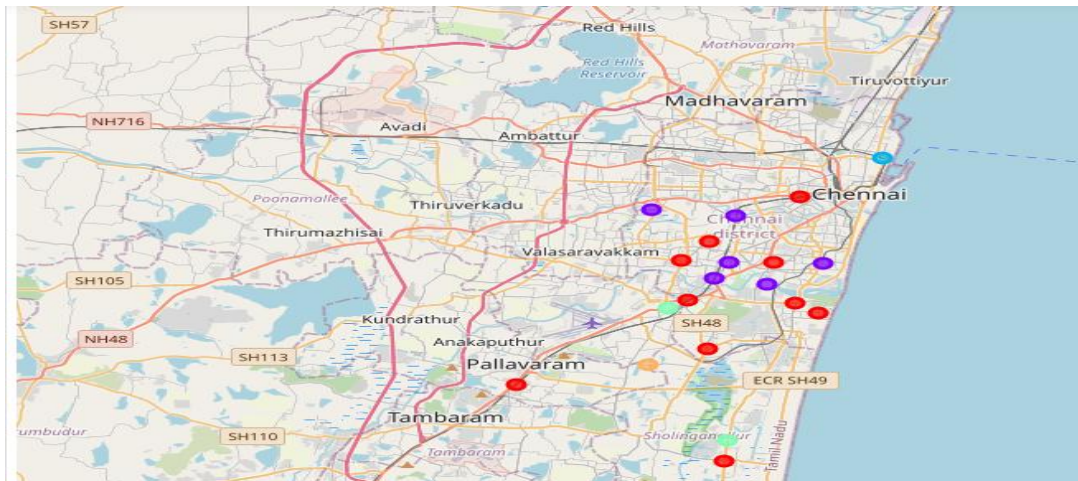
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adyar	13.0063	80.2574	Bombay Brassiere	13.006961	80.256419	North Indian Restaurant
1	Adyar	13.0063	80.2574	Cool Biz	13.006632	80.257275	Juice Bar
2	Adyar	13.0063	80.2574	McDonald's	13.006632	80.254255	Fast Food Restaurant
3	Adyar	13.0063	80.2574	Prem's Graama Bhojanam	13.006345	80.253995	Vegetarian / Vegan Restaurant
4	Adyar	13.0063	80.2574	Adyar Ananda Bhavan	13.005824	80.257368	Indian Restaurant

The next step is to create clusters. Using the K-means clustering algorithm we check for the value of K clusters. This was done by using elbow method.

June 25,2020



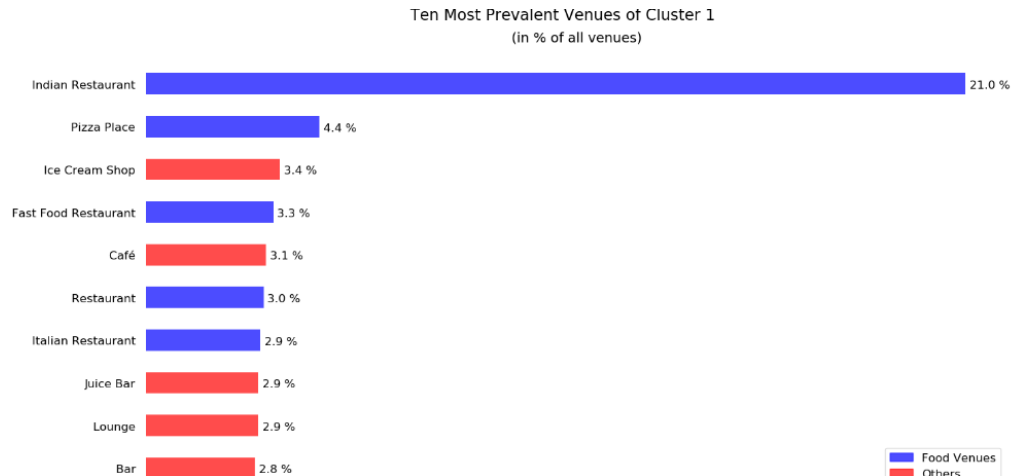
As we can see above K=5 achieved the best coefficient therefore we use this value as our K-cluster. Now we can easily separate the neighborhoods by clusters using Folium and our K value. We will also be able to see the top 10 venues for each neighborhood.



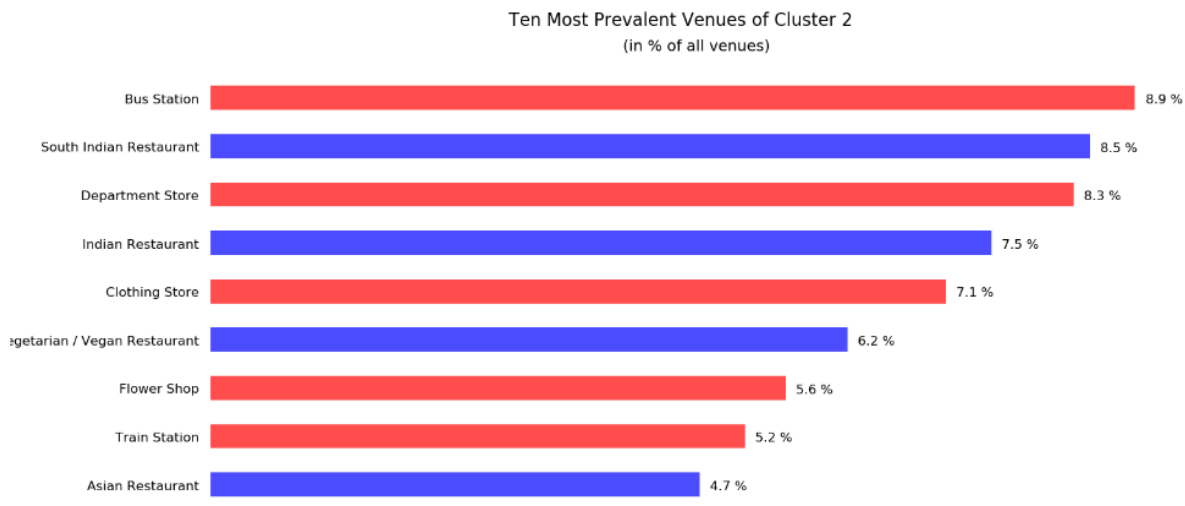
	City	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Chennai	Adyar	13.0063	80.2574	0	Indian Restaurant	Coffee Shop	North Indian Restaurant	Electronics Store	Movie Theater	Grocery Store	Gym / Fitness Center	Fast Food Restaurant	Ice Cream Shop
1	Chennai	Alandur	13.0030	80.2040	3	Hotel Bar	Hotel	Fish Market	Indian Restaurant	Gym	Grocery Store	Gourmet Shop	Gift Shop	Food
2	Chennai	Alwarpet	13.0339	80.2486	0	Indian Restaurant	Lounge	Japanese Restaurant	Restaurant	Thai Restaurant	Hotel	Italian Restaurant	Pizza Place	Burger Joint
3	Chennai	Ashok Nagar	13.0351	80.2095	0	Indian Restaurant	Fast Food Restaurant	Movie Theater	Diner	Vegetarian / Vegan Restaurant	Metro Station	Café	Sculpture Garden	Pizza Place
4	Chennai	Besant Nagar	13.0002	80.2668	0	Ice Cream Shop	Indian Restaurant	Juice Bar	Bistro	New American Restaurant	Beach	Coffee Shop	Chinese Restaurant	Café

June 25,2020

To further examine the clusters, we will observe the 10 venues in each cluster using a bar plot. This will help us see what the density of the venues are in a given cluster.

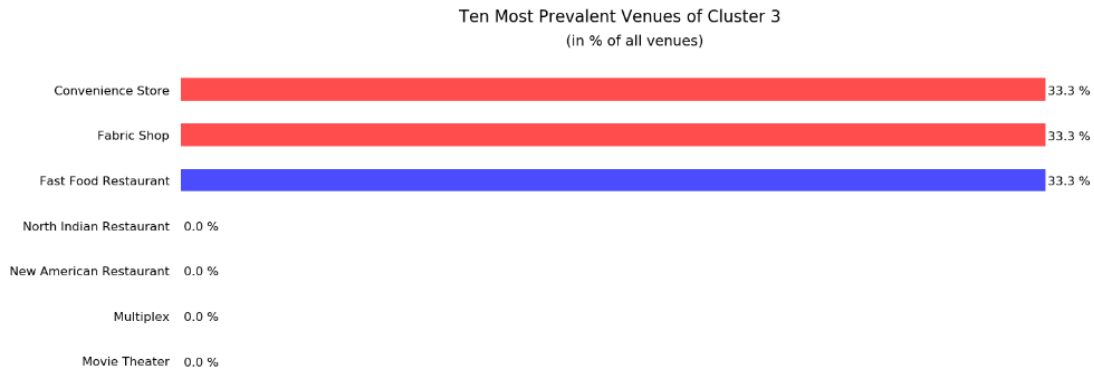


Neighborhoods in cluster 1 have Indian restaurants as their most preferred venue. It tops the top 10 venues. So, considering cluster 1 might be a good idea.

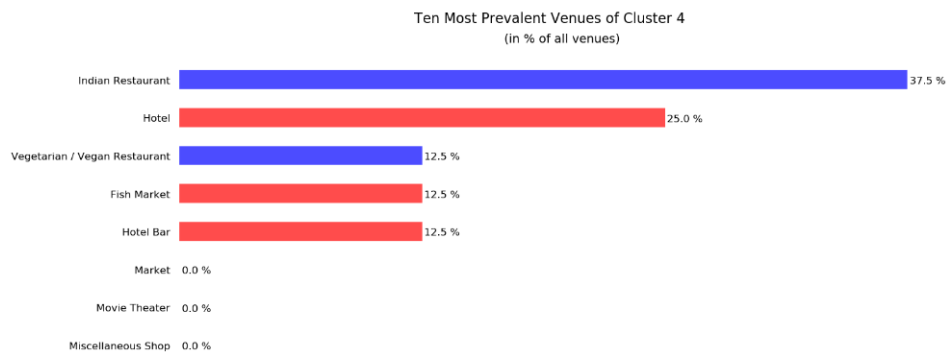


Cluster 2 has bus station as most visited venue. When compared to cluster 1, no high preference for Indian restaurant found in cluster 2.

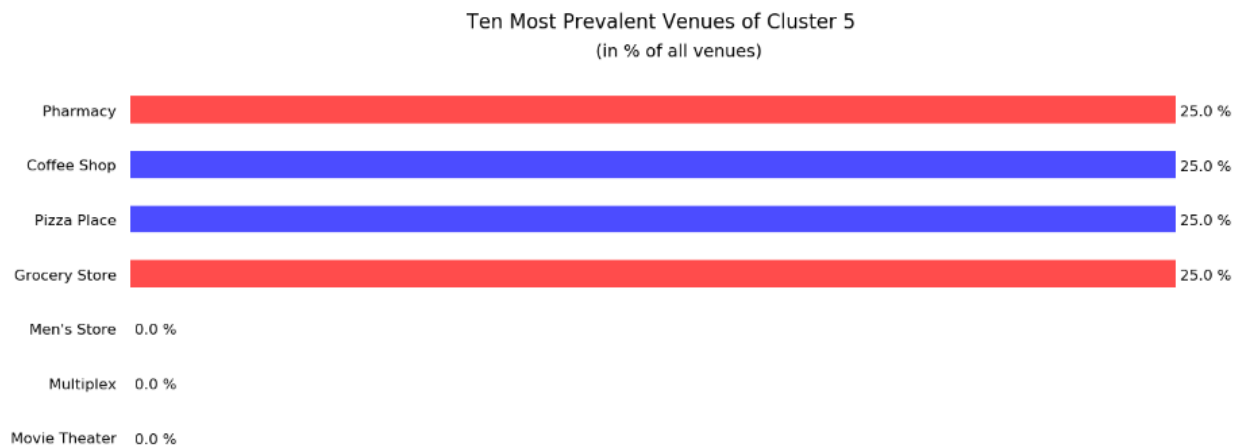
June 25,2020



From above bar plot, it is clear that Cluster 3 is not suitable for opening a restaurant.



Cluster 4 also has Indian restaurant a most preferred venue. But it doesn't have theatres, metro stations in their surroundings like cluster 1 has.



From above bar plot, it is clear that cluster 5 won't be suitable for opening a restaurant as people in that cluster wouldn't hang out to restaurants.

Results

From the bar plot analysis, it is clear that opening an Indian restaurant in cluster 1 areas will be a good idea. In that cluster, the neighbourhood Ashok Nagar will be a better place since it is a densely populated area and theatres, cafe and metro station are present in their surroundings.

Discussion

To reach our desired outcome and to have a concise analysis of the datasets used we have to go through numerous steps. The first step was being able to identify the correct neighborhoods and map them. To use the Foursquare API more effectively we limited our Neighborhoods to the ones that have more 10 or more venues. The reason for doing this is because, it would practically be more effective to open a restaurant in an area that is popular and has more venues that people might visit.

One of the main deciding factors for this analysis was the number of clusters we used. In our case we chose $K=5$ using elbow method. It is also important to note a high k value is also not recommended as it would diversify the neighborhood clusters and risks us not being able to identify a suitable cluster correctly.

The next step taken was to find the optimum cluster which has Indian restaurant as most visited venue. Using bar plot, we analyzed the top 10 venues for each neighbourhood present in each cluster.

One last step is selecting the most preferred neighbourhood in the optimal cluster. The neighbourhood Ashok Nagar is selected based on its dense population and their surrounding venues.

Conclusion

This project focused on highlighting the basic research needed to find potential neighborhoods to open an Indian restaurant in Chennai. It follows a few simple procedures which were, finding neighborhood data of Chennai. This can be done by either web scrapping or finding a usable *.json* file. The next step conducted was to obtain the coordinates of the respective neighborhoods using a geolocator and then mapping them over a map of Chennai using Folium. Our main instrument in this project was the use of Foursquare API which gives us access to the commercial traffic in the region. Commercial traffic, in this case, is all the venues listed in their respective locations and their frequency. The API helps us identify hotspots are more occurring more frequently and plan accordingly. Using the Foursquare API we selected the neighborhoods with at least 10 venues and used K-means clustering ($K = 5$) to further examine our data frame. According to our needs, Cluster 1 was chosen to be the best option. Finally, our neighborhood choice was trimmed down to Ashok Nagar since it is a densely populated area and surrounded with venues like theatres, cafe and metro station.