Predicting suitable Location for Restaurants

Shashi Kumar Sharma

April 25, 2020

1. Introduction

1.1 Background

Despite the substantial risks and the tough hours, owning a restaurant business is one of the most rewarding experiences of one's life for anyone that have nurtured a lifelong ambition of starting up a restaurant. Though the future looks bright for the food-service industry overall, there are no guarantees in this business. A hard reality is that many restaurants fail during their first year, frequently due to a lack of planning. But that doesn't mean your food-service business has to be an extremely complex operation. In fact, the more streamlined you can make it, the better your chances for success.

Some of the factors that affects the success of any restaurants are:

- concept of restaurant
- location of restaurant
- licenses required to start the restaurant business
- availability of workforce for restaurant business
- Arranging vendors and suppliers for restaurant and more.

1.2 Problem

The location is an essential factor to consider while discussing how to start a restaurant business, as it can determine the success of any restaurant. While choosing restaurant's location, it is a good idea to identify ones competitor in that area and gauge their progress and understand their business model. Evaluating the competition would help restaurant understand the customer base that hovers in that area, their pocket size, and their preferences. That's why this Project aims to predict a proper location for success of any restaurant (For this project specifically in Chennai, India).

1.3 Interest

Anyone who interests in starting up a restaurant, and don't know how to go about it, or are currently in the middle of setting up your first restaurant. This may help them with one of the most crucial factor for their Business Success.

2. Data Acquisition and Cleaning

2.1 Data Source

Most of the Location data is acquired from online scraping of sites. I got required data from finkado here. The site contains Pin Codes (Specific code to a Location) of many regions of Chennai.

2.2 Data Cleaning

The webpage data was downloaded using inbuilt request package of Python. After which the required table of pin-codes of regions was excluded from the webpage data. This was the first step on data cleaning.

Secondly, table had just pin-codes and location name in it. I needed location's Latitude and Longitude as well so that this can be used in Folium package for plotting the locations on map. So I used geopy package to extract the latitude and longitude data of the given location in the dataset.

While extracting the geo-coordinate data form the dataset, many of the location didn't return any coordinate (None). The reason may be due to closeness of the location to each other or lack of data in Geopy database. So these data rows had to be removed for further processing.

Lastly, I used the dataset in foursquare API for getting popular venue list on each and every location.

2.3 Feature Selection

While getting the location data from the site, the total number of observation was 67 i.e. 67 Location was left after generating the location coordinate with 5 Features (Area, Pin-code, Location, Latitude, and Longitude). Since 'Location and Pin-code' feature was redundant for further processing, these feature was dropped at this stage.

When this dataset w further forwarded to Foursquare API, the final dataset had 1883 observation i.e. Venue objects with 6 Features (Latitude, Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category). Thus the final Feature had these below mentioned categories.

	Area	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aminjikarai S.O, Chennai	13.07214	80.220545	McDonald's	13.073759	80.221388	Fast Food Restaurant
1	Aminjikarai S.O, Chennai	13.07214	80.220545	PVR	13.073799	80.221392	Multiplex
2	Aminjikarai S.O, Chennai	13.07214	80.220545	Westside	13.073687	80.221380	Clothing Store
3	Aminjikarai S.O, Chennai	13.07214	80.220545	@home	13.071499	80.222309	Furniture / Home Store
4	Aminjikarai S.O, Chennai	13.07214	80.220545	Landmark	13.073781	80.221393	Bookstore

Table 1. Selected Features during Data cleaning

3. Exploratory Data Analysis

3.1 Visualizing number of venues and splitting as per categories for each location

Once the required Features was selected, I tried getting the total numbers of Venue for each location. So that the represented data can be understood more easily.

Area	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Aminjikarai S.O, Chennai	20	20	20	20	20	20
Anna Nagar S.O (Chennai), Chennai	2	2	2	2	2	2
Anna Road H.O, Chennai	6	6	6	6	6	6
Chepauk S.O, Chennai	15	15	15	15	15	15
Chintadripet S.O, Chennai	23	23	23	23	23	23
Choolai S.O, Chennai	5	5	5	5	5	5
Choolaimedu S.O, Chennai	15	15	15	15	15	15

Table 2. Number of venues for each Location

After this, it was time to view the mean of each venue categories for each location. The mean was taken to understand the data more correctly and to normalize it. This resulted in 165 columns of unique Venue.

	Area	ATM	African Restaurant	Airport	American Restaurant	Amphitheater	Antique Shop	Arcade	Arts & Crafts Store	Asian Restaurant	...	Thai Restaurant	Theater	Toy / Game Store	Train Station	Vegetarian / Vegan Restaurant
0	Aminjikarai S.O, Chennai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.000000
1	Anna Nagar S.O (Chennai), Chennai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.500000	0.000000
2	Anna Road H.O, Chennai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.166667	0.000000
3	Chepauk S.O, Chennai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.066667	0.066667
4	Chintadripet S.O, Chennai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.086957	0.086957

5 rows × 166 columns

Table 3. Mean for each Venue per location

Now, I viewed each location along with the top 5 most common venues. For understanding how much cluster I should take for proper clustering of data. For the data it was pre-assumed, that there will be two clusters with Restaurants in it. One would be having majority of Indian Themed Restaurants and other will be with other venues as well.

When we made a new dataset for arranging top 10 venues for each location, It became more clear that the above assumption is going to effect the clustering pattern.

	Area	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	Aminjikarai S.O, Chennai	Fast Food Restaurant	Pizza Place	Event Space	Furniture / Home Store	Clothing Store	Men's Store	Bakery	Playground	Burger Joint	Electronics Store
1	Anna Nagar S.O (Chennai), Chennai	Burmese Restaurant	Train Station	Yoga Studio	Farmers Market	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Field
2	Anna Road H.O, Chennai	Indian Restaurant	Department Store	Train Station	Movie Theater	Farmers Market	Food & Drink Shop	Food	Flower Shop	Flea Market	Field
3	Chepauk S.O, Chennai	Indian Restaurant	Memorial Site	Beach	Music Store	Platform	Middle Eastern Restaurant	Light Rail Station	Seafood Restaurant	Cricket Ground	Vegetarian / Vegan Restaurant
4	Chintadripet S.O, Chennai	Hotel	Indian Restaurant	Movie Theater	Bookstore	Vegetarian / Vegan Restaurant	Train Station	Sandwich Place	Flea Market	Fast Food Restaurant	Farmers Market

Table 4. Top 10 Venues for each location

3.2 Predictive model used

As the project was to use Foursquare API along with Clustering algorithm for clustering data. Hence I used K-means clustering algorithm with number of clusters to be 5.

4. Result

After applying the K-means algorithm on the dataset, Data was split into 5 clusters. As pre-assumed, two clusters resulted in containing maximum restaurant in them. One cluster had majority of Indian themed restaurants while other cluster had mix of each restaurants along with public facilities as well. In the below figure, the Purple circles are for location having majority of Indian themed restaurants while red circles are for the other one.
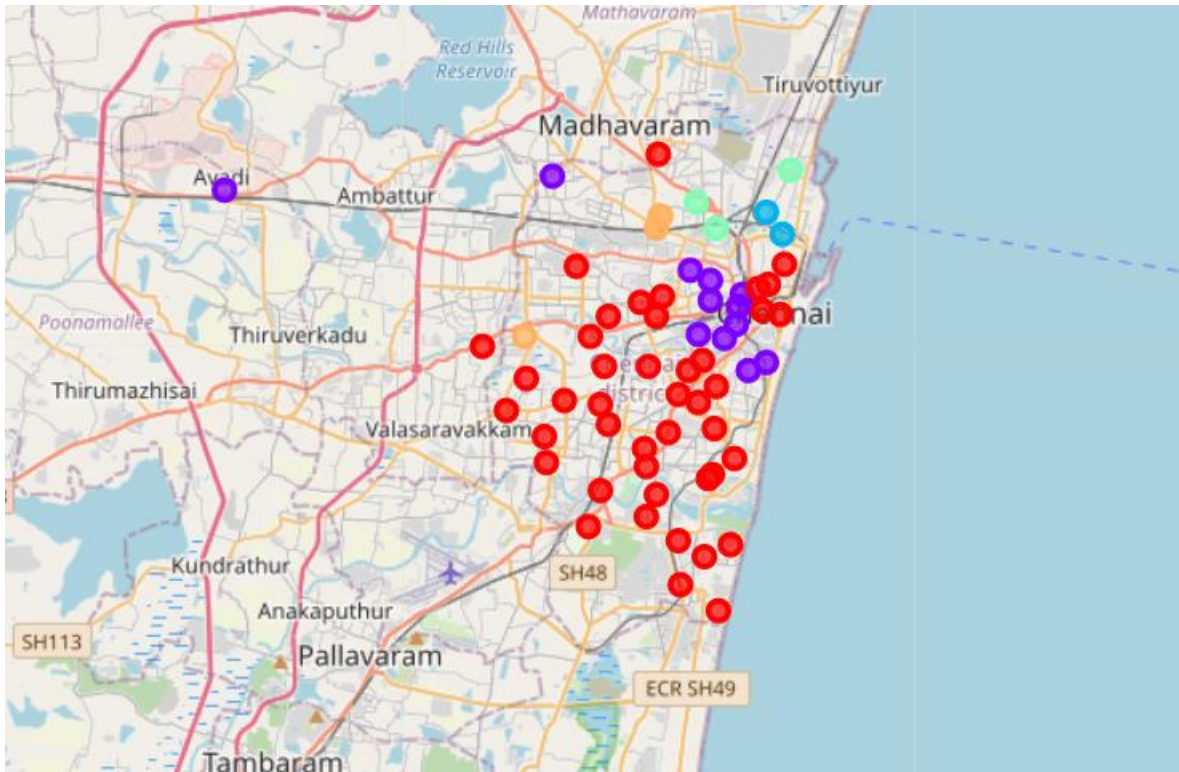
Figure 1. Map showing Clusters

Cluster 1

```
In [29]: Chennai_merged.loc[Chennai_merged['Labels'] == 0, Chennai_merged.columns[[1] + list(range(5, Chennai_merged.shape[1]))]]
```

46	Kilpauk Medical College S.O, Chennai	0	Indian Restaurant	Bakery	Café	Fast Food Restaurant	Italian Restaurant	Juice Bar	Frozen Yogurt Shop	Food Truck	South Indian Restaurant	Multiplex
47	Kilpauk S.O, Chennai	0	Café	Indian Restaurant	Juice Bar	Ice Cream Shop	Frozen Yogurt Shop	Gym / Fitness Center	Daycare	Food	Coffee Shop	Fast Food Restaurant
48	Kodambakkam S.O, Chennai	0	Indian Restaurant	Juice Bar	Jewelry Store	Electronics Store	Clothing Store	Chinese Restaurant	Food Court	Lounge	Bakery	Furniture / Home Store
49	Kodungaiyur S.O, Chennai	0	Ice Cream Shop	Fast Food Restaurant	Coffee Shop	Currency Exchange	Farmers Market	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
52	Kotturpuram S.O, Chennai	0	Bakery	Department Store	Park	Sandwich Place	Café	Convenience Store	Chinese Restaurant	College Library	College Cafeteria	Boat or Ferry
60	Madras Medical College S.O, Chennai	0	Indian Restaurant	Train Station	Bookstore	Market	Historic Site	Fast Food Restaurant	Electronics Store	Museum	Nightclub	Platform
61	Madras University S.O, Chennai	0	Chinese Restaurant	Park	College Cafeteria	Department Store	Sandwich Place	Café	Clothing Store	Market	Field	Light Rail Station

Figure 2. Cluster having other themed restaurants with public facilities near them

Cluster 2

```
In [30]: Chennai_merged.loc[Chennai_merged['Labels'] == 1, Chennai_merged.columns[[1] + list(range(5, Chennai_merged.shape[1]))]]
```

Out[30]:

	Area	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	10th Most Common Venue
4	Anna Road H.O, Chennai	1	Indian Restaurant	Department Store	Train Station	Movie Theater	Farmers Market	Food & Drink Shop	Food	Flower Shop	Flea Market	Field
13	Chepauk S.O, Chennai	1	Indian Restaurant	Memorial Site	Beach	Music Store	Platform	Middle Eastern Restaurant	Light Rail Station	Seafood Restaurant	Cricket Ground	Vegetarian / Vegan Restaurant
15	Chintadripet S.O, Chennai	1	Hotel	Indian Restaurant	Movie Theater	Bookstore	Vegetarian / Vegan Restaurant	Train Station	Sandwich Place	Flea Market	Fast Food Restaurant	Farmers Market
16	Choolai S.O, Chennai	1	Indian Restaurant	Hotel	Pizza Place	Department Store	Juice Bar	Event Space	Food & Drink Shop	Food	Flower Shop	Flea Market
20	Dr. Ambedkar Nagar S.O (Chennai), Chennai	1	Indian Restaurant	Intersection	Department Store	Electronics Store	Hotel	ATM	Train Station	Watch Shop	Food	Flower Shop
23	Egmore S.O, Chennai	1	Indian Restaurant	Hotel	Café	Seafood Restaurant	Pizza Place	Farmers Market	Donut Shop	Middle Eastern Restaurant	Mexican Restaurant	Coffee Shop
51	Kosapet S.O (Chennai), Chennai	1	Indian Restaurant	Hotel	Vegetarian / Vegan Restaurant	Department Store	Farmers Market	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market

Figure 3. Cluster having majority of Indian themed restaurants

5. Discussion

In this study, I analyzed the location data and pre-assumed surety of two cluster in the result, and the predictive model also suggested the same. I analyzed with the above cluster that major Indian themed restaurants are near the railway station. As this seems fair since many people from various cities in India migrate to Chennai for job opportunities.

Apart from these, different location has various themed restaurant around them. This is all because of diversity in citizen of these places.

6. Conclusion

Coming back to the result. The owner of the restaurant may choose any of the location in two cluster depending on his/her requirements. If he choose to open another themed restaurant with public facilities near them. Then he/she can go with cluster with different themed restaurant.

If he choose to open specifically Indian themed restaurant. Then he/she can go with cluster with Majority of Indian themed restaurant.