



Coursera Capstone project

Finding a location to start a Chinese Restaurant



Problem description

- The client wants to start a Chinese restaurant in Chennai in Tamil Nadu, India.
- Doesn't want a place where there are many already an established Chinese restaurant
- It should be near to places where it is often crowded such as malls, markets, multiplexes, train and bus stations etc.
- The main goal of the project is to find an area in Chennai which has most of the crowded places
- And also the least number of Chinese type restaurants.



Selecting, cleaning and analyzing data

- The latitudes and longitudes of all the places in Chennai is needed.
- pgeocode library returns the location details if given the pincodes of the areas.
- Pincodes are taken from the following website url by web scraping.
- <https://www.mapsofindia.com/pincode/india/tamil-nadu/chennai/>
- Using location details, the venues and places in those locations are obtained using the Foursquare API

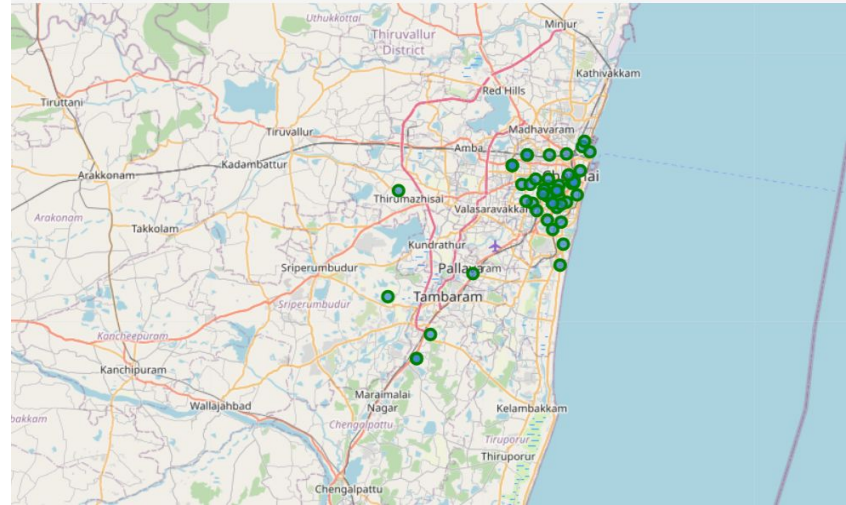
Data analyzing

Missing data

- The rows with the missing data were removed.

Data with location

- The new dataframe is formed with the places, pincodes and their respective latitudes and longitudes.



	Pincode	Places	Latitude	Longitude
0	600018	Pr. Accountant General, Teynampet	13.0433	80.2528
1	600020	Shastri Nagar (Chennai), Kasturibai Nagar, Ady...	12.9967	80.2603
2	600082	Periyar Nagar, Kumaran Nagar, G K M Colony	13.0572	80.2554



Foursquare API

- A full dataframe is created which has all the retrieved venues with its location for all of the areas.
- All the venues and its locations are retrieved by requesting the foursquare api.

	Places	Latitude of the place	Longitude of the place	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Pr. Accountant General, Teynampet	13.0433	80.2528	Stix	13.042898	80.248640	Chinese Restaurant
1	Pr. Accountant General, Teynampet	13.0433	80.2528	Hyatt Regency Chennai	13.042874	80.248593	Hotel
2	Pr. Accountant General, Teynampet	13.0433	80.2528	Ente Keralam	13.042188	80.255880	Kerala Restaurant
3	Pr. Accountant General, Teynampet	13.0433	80.2528	Z The Tapas Bar & Restaurant	13.045808	80.258013	Bar
4	Pr. Accountant General, Teynampet	13.0433	80.2528	Nando's	13.046237	80.256267	African Restaurant

Duplicate values

- Duplicate values are created because of not using radius.
- Duplicate values are deleted.

```
chennai_venues.loc[chennai_venues['Venue'] == 'Subway'].count()
```

```
1]: Places          44
   Latitude of the place  44
   Longitude of the place  44
   Venue                44
   Venue Latitude        44
   Venue Longitude       44
   Venue Category        44
   dtype: int64
```

As we can see the subway duplicates are removed and the n removed.

As the radius was not given in the url request section, there may be duplicate values in the dataset. So it must be cleaned.

```
In [27]: chennai_venues.loc[chennai_venues['Venue'] == 'Subway'] #checking for duplicate values for a the venue Subway
```

Out[27]:

	Places	Latitude of the place	Longitude of the place	Venue	Venue Latitude	Venue Longitude	Venue Category
38	Pr. Accountant General, Teynampet	13.0433	80.2528	Subway	13.054787	80.249618	Sandwich Place
96	Pr. Accountant General, Teynampet	13.0433	80.2528	Subway	13.049846	80.256958	Sandwich Place
162	Shastri Nagar (Chennai), Kasturibai Nagar, Ady...	12.9967	80.2603	Subway	12.989579	80.248830	Sandwich Place
176	Shastri Nagar (Chennai), Kasturibai Nagar, Ady...	12.9967	80.2603	Subway	12.999040	80.254743	Sandwich Place
193	Shastri Nagar (Chennai), Kasturibai Nagar, Ady...	12.9967	80.2603	Subway	13.017396	80.271303	Sandwich Place
...
4200	Valmiki Nagar, Tiruvannmiyur, Palavakkam (Kanch...	12.9695	80.2561	Subway	12.999040	80.254743	Sandwich Place
4276	Vadapalani	13.0511	80.2125	Subway	13.038853	80.212274	Sandwich Place
4279	Vadapalani	13.0511	80.2125	Subway	13.047275	80.194957	Sandwich Place
4348	Velacheri	13.0647	80.2523	Subway	13.061403	80.248521	Sandwich Place
4372	Velacheri	13.0647	80.2523	Subway	13.054787	80.249618	Sandwich Place

85 rows × 7 columns

It's been confirmed that there are duplicate values for Subway. There must be many like this. The dataframe currently contains 4570 samples.

One-Hot encoding

- Dataset encoded to view the most common places in all areas of Chennai.
- Dataframe was grouped and sorted to view all areas top 10 venues.

```
chennai_venues_sorted
```

Out[37]:

	Places	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	Anna Nagar (Chennai)	Indian Restaurant	Market	Coffee Shop	Pizza Place	Indie Movie Theater	Zoo Exhibit	Sports Club	Sandwich Place	Movie Theater	Multicuisine Indian Restaurant
1	Anna Nagar East	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Multiplex	Café	Clothing Store	Bus Station	Bakery	Pizza Place	Chinese Restaurant
2	Anna Nagar Western Extn	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Multiplex	Café	Clothing Store	Bus Station	Bakery	Pizza Place	Chinese Restaurant
3	Arumbakkam, D G Vaishnav College	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Clothing Store	Multiplex	Electronics Store	Shopping Mall	Hotel	Department Store

K-means Clustering

- Similar areas are grouped which will be easy for us to identify
- Clusters which are favourable to our criteria can be grouped together.
- Others can be ignored.

Out[58]:

	Places	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	Anna Nagar East	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Multiplex	Café	Clothing Store	Bus Station	Bakery	Pizza Place	Chinese Restaurant
6	Anna Nagar Western Extn	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Multiplex	Café	Clothing Store	Bus Station	Bakery	Pizza Place	Chinese Restaurant
8	Arumbakkam, D G Vaishnav College	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Clothing Store	Multiplex	Electronics Store	Shopping Mall	Hotel	Department Store
21	Venkatesapuram, Puliyanthope, Perambur Barracks	Indian Restaurant	Bus Station	Train Station	Café	South Indian Restaurant	Fast Food Restaurant	Racetrack	Bakery	Sporting Goods Shop	Theme Park
32	Saidapet (Chennai), Guindy North	Indian Restaurant	Fast Food Restaurant	Bakery	Vegetarian / Vegan Restaurant	Café	Restaurant	Chinese Restaurant	Coffee Shop	Park	Middle Eastern Restaurant
33	High Court Building (Chennai)	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Multiplex	Café	Clothing Store	Bus Station	Bakery	Pizza Place	Chinese Restaurant
34	Thygarayanagar North ND, Thygarayanagar South ...	Indian Restaurant	Fast Food Restaurant	Bakery	Vegetarian / Vegan Restaurant	Café	Restaurant	Chinese Restaurant	Coffee Shop	Park	Middle Eastern Restaurant
42	Nerkundram, Koyambedu Wholesale Market Com, Ko...	Indian Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Clothing Store	Café	Hotel	Burger Joint	Restaurant
49	Shenoy Nagar, Aminjikarai	Indian Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Chinese Restaurant	Café	Fast Food Restaurant	Bakery	Restaurant	Clothing Store	Bookstore

The third one has more Fast food restaurants and also less number of crowded places. So this one can be ignored.

Results

- Dataframe was clustered into 5 different groups.
- Clusters most favourable to criteria are grouped and others ignored.

In [57]: c2

Out[57]:

	Places	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 Nagar (Chennai)	Indian Restaurant	Market	Coffee Shop	Pizza Place	Indie Movie Theater	Zoo Exhibit	Sports Club	Sandwich Place	Movie Theater	Multicuisine Indian Restaurant
11	Besantnagar, Rajaji Bhavan	Indian Restaurant	Market	Indie Movie Theater	Train Station	Fast Food Restaurant	Coffee Shop	Café	Zoo	Juice Bar	Multicuisine Indian Restaurant
14	Tidel Park, TTTI Taramani	Indian Restaurant	Indie Movie Theater	Sports Club	Zoo	Food	Multicuisine Indian Restaurant	Department Store	Pizza Place	Racetrack	Coffee Shop
20	Choolaimedu	Indian Restaurant	Market	Indie Movie Theater	Train Station	Fast Food Restaurant	Coffee Shop	Café	Zoo	Juice Bar	Multicuisine Indian Restaurant
37	Kalaighnar Karunanidhi Nagar	Indian Restaurant	Pizza Place	Fast Food Restaurant	Ice Cream Shop	Movie Theater	Café	Bookstore	Department Store	Pub	Road

As for the 2nd cluster, it has more crowded places like theatres, zoo and markets. So we'll take this into consideration

Conclusion

- Final dataframe analyzed for most crowded places and least number of Chinese and Fast Food restaurants.
- Most satisfying result chosen.

```
In [68]: final_df[final_df['Key counts']>3]
```

```
Out[68]:
```

	Places	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	Key counts
13	Washermanpet, Washermanpet East	Indian Restaurant	Train Station	Hotel	Multiplex	Farmers Market	Coffee Shop	Sandwich Place	Bus Station	Nightclub	Restaurant	4.0
38	Rayapuram	Indian Restaurant	Train Station	Italian Restaurant	Hotel	Farmers Market	Multiplex	Electronics Store	Market	Snack Place	Motel	4.0
54	Tondiarpet, Tondiarpet West, Tondiarpet Bazaar	Indian Restaurant	Hotel	Train Station	Multiplex	Farmers Market	Nightclub	Pizza Place	Museum	Market	Coffee Shop	4.0
58	Vyasarpadi, Vyasar Nagar Colony	Indian Restaurant	Coffee Shop	Italian Restaurant	Train Station	Hotel	Multiplex	Farmers Market	Fast Food Restaurant	Food	Market	4.0

Finally the dataframe is narrowed down to four areas in Chennai which are the best suitable for the client to start a Chinese Restaurant.

Washermanpet area would be the best according to the data as it has Train Station, Multiplex and Market in the top 5. And also the bus station when considering top 10.



Future directions

- Project was done only based on the number of certain types of venues in Chennai.
- The trending places in a location was difficult to get from the foursquare api.
- If this footfall numbers had been used in this project, it would have given more accurate results than using the most common venues.