Coursera Capstone

IBM Applied Data Science Capstone

Opening a Restaurant near by IT Parks in Chennai, India

By: Koteswara Rao Nataru

May 2020

Introduction:

For many employees, visiting restaurants is a great way to relax and enjoy themselves during weekdays(work hours), weekends and holidays. People will choose the nearest the restaurants because they must continue their deliverables and to reduce the travelling time and cost. Many of IT employees often visit restaurants if you provide the good quality of food and different taste than others.
For better business choosing the best location and type of style like North Indian, Chinese and South Indian is major part. Our business will get succeeded if you we choose these parameters correctly.

So, we will try to find the optimal location with our data and technology.

Business Problem:

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a restaurant near by IT Parks in Chennai, India.
Since there are lots of restaurants in Chennai, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no North Indian and/or Chinese restaurants in vicinity. We would also prefer locations as close to city centre as possible, if first two conditions are met. We will use our data science powers to generate a few most promising neighbourhoods based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: If a Investor is looking to open a new restaurant, where would you recommend that they open it?

Data:

Based on definition of our problem, factors that will influence our decision are:

- List of neighbourhoods i.e. IT Parks in Chennai
- List of existing restaurants in the neighbourhood (any type of restaurant)
- List of and distance to North Indian and/or Chinese restaurants in the neighbourhood, if any
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and to get the venue data.
- Venue data, particularly data related to restaurants.
- We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them:

This **Wikipedia** page (**https://en.wikipedia.org/wiki/List_of_tech_parks_in_Chennai**) contains a list of neighbourhoods in Chennai. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python Pandas packages. Then we will get the geographical coordinates of the neighbourhoods using **Python Geocoder** package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use **Foursquare API** to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Food category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

Methodology

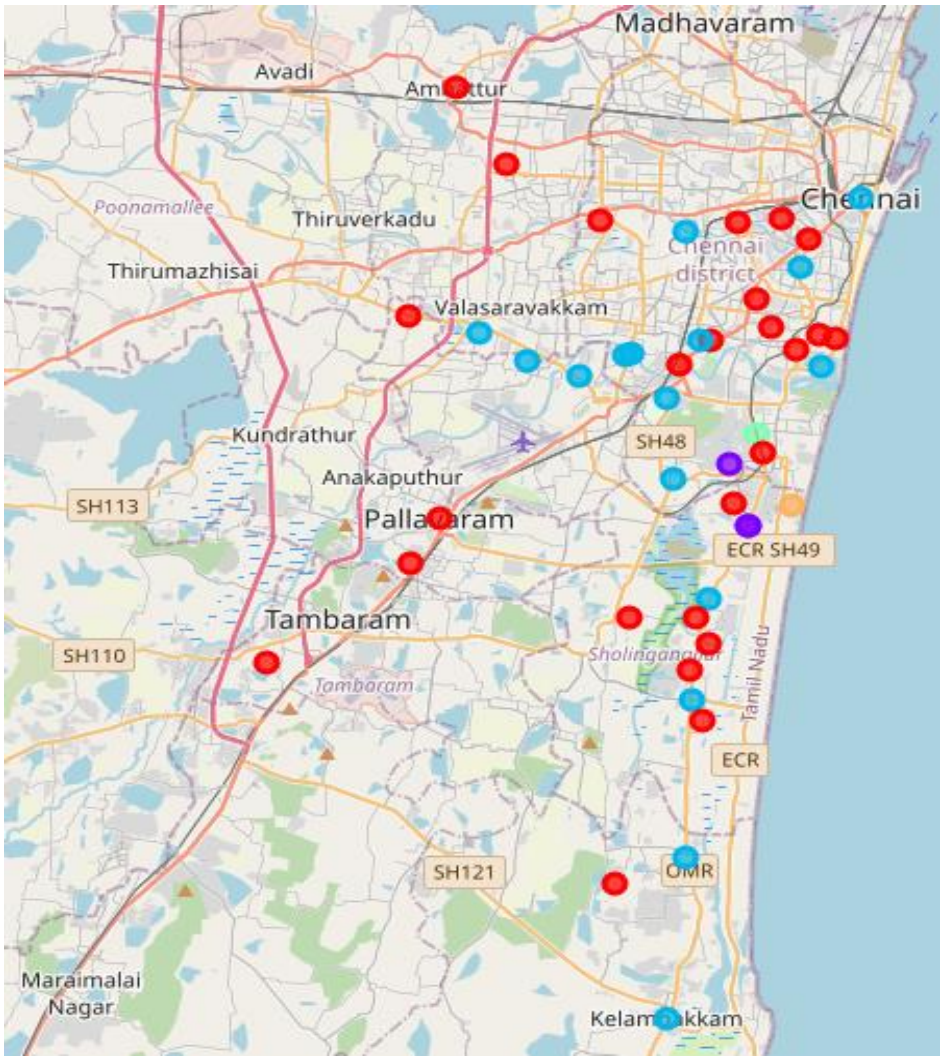
In this project we will direct our efforts on detecting areas of Chennai IT parks that have low restaurant density, particularly those with low number of Special restaurants.

In first step we have collected the required data: location and type (category) of every restaurant. Second step in our analysis will be calculation and exploration of restaurant density across different areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders.

Results

- Categorized the neighbourhoods into 5 clusters:
    - Cluster 0: Neighbourhoods with high number of authentic restaurants
    - Cluster 1: Neighbourhoods with moderate number of authentic restaurants
    - Cluster 2: Neighbourhoods with high number of authentic restaurants
    - Cluster 3: Neighbourhoods with low number of authentic restaurants
    - Cluster 4: Neighbourhoods with low number of authentic restaurants



Discussion:

- Highest number in cluster 0 & 2 and moderate number in cluster 1
- Cluster 3 & 4 has very low number to no Restaurants in the neighbourhoods

Conclusion:

- ❑ Answer to business question: The neighbourhoods in cluster 3 are the most preferred locations to open a new authentic restaurant

- ❑ Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurant

References

List of neighbourhoods in Chennai Wikipedia. Retrieved from **https://en.wikipedia.org/wiki/List_of_tech_parks_in_Chennai**

Foursquare Developers Documentation. Foursquare. Retrieved from https://developer.foursquare.com/docs