

Food Investment

CourseRA Capstone Project

Jul’20

# Introduction

### Background

Delhi is a big city and the capital of India. It is one of most populous cities and hence has a huge number of attractions amongst various investment firms from within or outside of India. Some of these have been IT companies/MNCs or top of the shelf restaurants, high-class pubs, living spaces, car showrooms, etc. to name a few. This project includes finding areas or neighborhoods in Delhi, where an investor can look to open a new food joint.

### Problem Statement

As we are dealing with food joint, it could be anything – be it restaurant, food truck or a café. This will be solemnly based on the results of this project. In this project, we will be focusing on areas where food joints are not so common and have good opportunity for the firm to invest in any food ventures due to lesser competition as compared to areas where there are number of places to visit and eat.

We will be using the data analysis and clustering to get the required areas/neighborhoods.

# Data Handling

### Data required and its sources

For this project, we would need various areas of New Delhi along with their latitudinal and longitudinal information. This would help us spread around various areas of the city and explore the parts which are most suited for the need.

Source for this information is from a repository:

<https://raw.githubusercontent.com/sanand0/pincode/master/data/IN.csv>

The second set of data, where we will be using the Foursquare API – is the dataset that will provide us the different types of venues present in an area(captured through the above dataset).

Foursquare API – which we all are familiar with and as necessary requirement for the Capstone project.

[https://api.foursquare.com/v2/venues/search?client\_id={}&client\_secret={}&ll={},{}&v={}&radius={}&limit={}](https://api.foursquare.com/v2/venues/search?client_id=%7b%7d&client_secret=%7b%7d&ll=%7b%7d,%7b%7d&v=%7b%7d&radius=%7b%7d&limit=%7b%7d)

The above two datasets should be enough for us to determine the neighborhoods where there will be less competition due to either lack of or having less frequent restaurants/food joints in the vicinity.

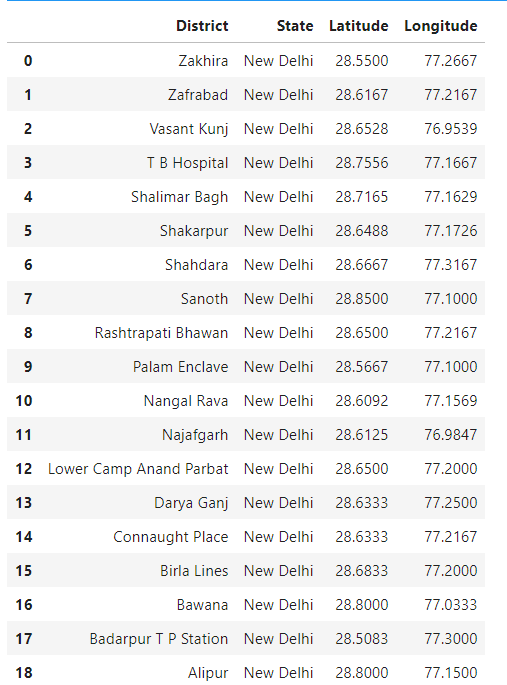
### Data Preparation

The data received from the repository was not exactly in the required format and need some tweaks, including the renaming of columns – as below

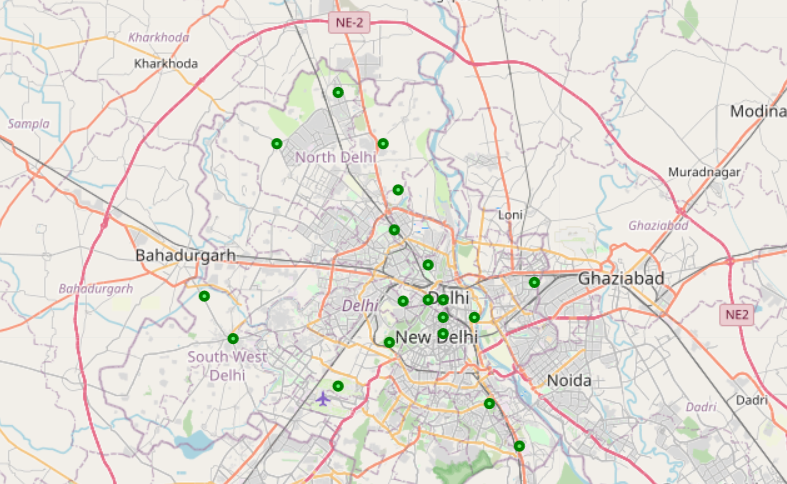
data = data.rename(columns={'key' :'Zip','place\_name':'District','admin\_name1':'State','latitude':'Latitude','longitude':'Longitude'})

We also had few districts that had exact same Latitude/Longitude due to closer vicinity and for the same purpose, we had to take only one district per Latitude and Longitude for our analysis – as there will be exact duplication of venues and hence clustering for the areas/districts falling under same lat and lang.

Final data looked like this was put in a dataframe.



Plotting these areas on the map using geolocator library.



The Foursquare API data returned didn’t have any issue in it – only preparation needed was to create a separate dataframe for it and extract the venue categories along with the necessary information.

# Methodology

The algorithm that we have gone ahead with is the k-clustering model.

This method was selected as it would help us find cluster of areas with similar venues/categories and identify or separately cluster few areas which may not have food joints as the common venues.

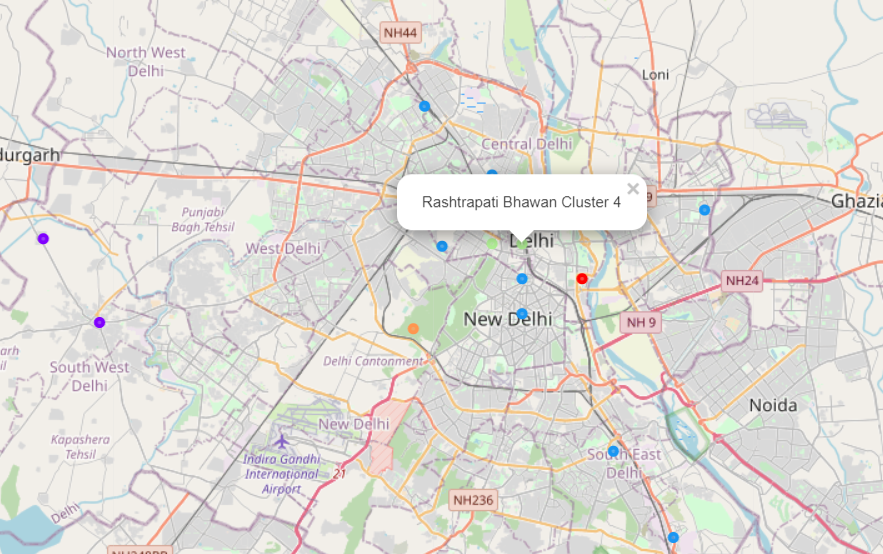
# Results

For the clustering, I wanted to go and separate as many as possible to exactly distinguish the areas for us but given the records were not too high, I had picked the value to get 6 clusters.

Sample results as in below grid where the most common venues for each area will help us identify.



From the location perspective too, we had put these areas on the map and we can see the different clusters – in our case, the clusters are separated by different colors(please see the purple, red, blue, orange and green dots).



# Conclusion

We were able to identify that the areas like in Cluster Label as ‘zero -0’ or cluster ‘5’ would be best suited for an investment as the places to eat are the 4th most common venues – as compared to either the 1st, 2nd or 3rd in the rest of the clusters.

