IBM Coursera Data Science Capstone

Opening a Coffee Shop in New Delhi

By Ananya Kumar

Introduction and Idea

Idea : To open up a coffee shop in a neighborhood in New Delhi, based on the number of coffee already in the neighborhood.

Inspiration : There is great disparity in the number of coffee shops in a neighborhood as some neighborhoods have too many coffee shops while others have none. Hence, there will be a great demand for coffee shops in neighborhoods that don’t have one. People will support this as this will reduce their travelling time to get coffee.

Business Problem

Location of a coffee shop is one of the primary factors which determine its success as opening a coffee shop in a neighborhood which does not have one will obviously translate to greater profits.

Business Question :

If any potential property developers are looking to build a coffee shop, where should they build it based on location of nearby coffee shops ?

Data Requirements

- A list of all the neighborhoods in New Delhi
- The latitude and longitude of each neighborhood
- Venue data that has information about the coffee shops in a locality

Sources of Data

- Wikipedia page with a list of all the recognized neighborhoods in New Delhi
- Foursquare of venue data about coffee shops
- Geocoder package for the latitudes and longitudes of each neighborhood

Methodology

First, we'll scrape the Wikipedia page for a comprehensive list of neighborhoods and convert it into a pandas DataFrame.

Then, we'll use the geocoder package to get the latitudes and longitudes for each of the neighborhoods and add them to our DataFrame.

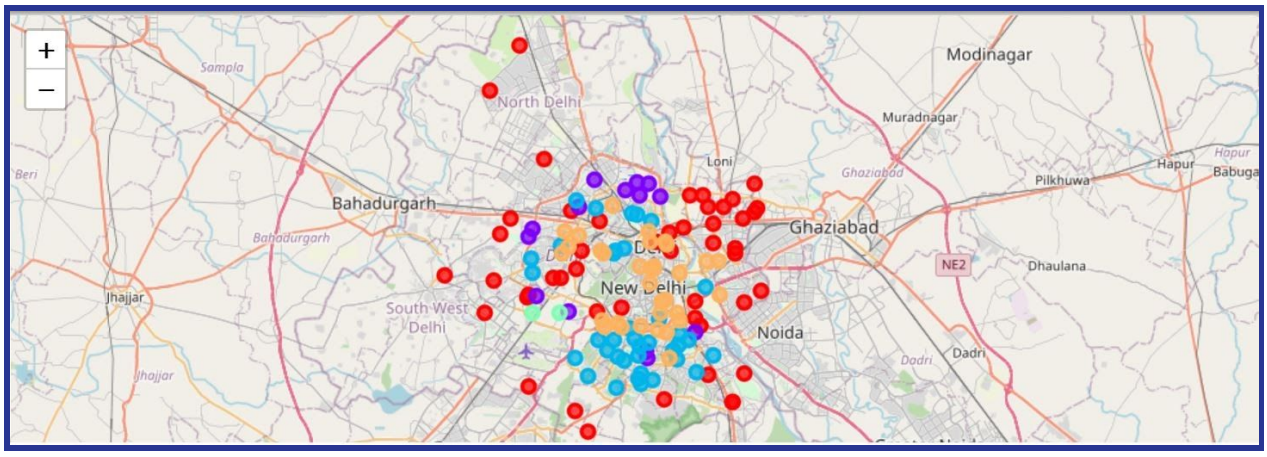
We'll use the foursquare API to get the venue data for each location and group the data by neighborhood. We'll also take the mean of the frequency of occurrence of each value in the venue category (eg : Grocery, Mall, etc) to prepare the data for clustering.

Finally, we will filter the "Coffee Shop" data from the DataFrame and add it in a new DataFrame which we'll use for K-Means clustering. Along with being the simplest and easiest method to cluster data, it is also very effective and suitable for this kind of problem.

The neighborhoods will be clustered into 5 categories based on the number of coffee shops in the neighborhood where the first cluster had no coffee shops and the fifth neighborhood has the most.

Based on the occurrence of the coffee shops in a neighborhood, we will answer the business question : where should we open a coffee shop ?

Results

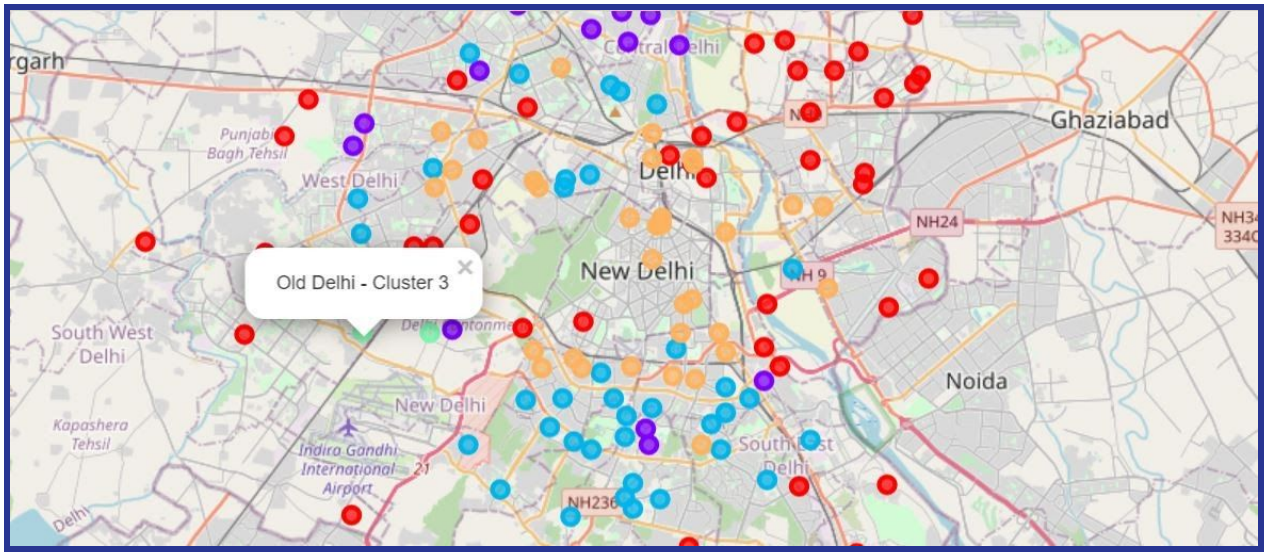


The results from the KMeans clustering partitioned New Delhi neighborhoods into five parts :

1. Cluster 0 : Red
2. Cluster 1 : Purple
3. Cluster 2 : Teal
4. Cluster 3 : Green
5. Cluster 4 : Orange

Discussion

We can see that neighborhoods that have the most number of coffee shops are at the heart of the city and the number of coffee shops per neighborhood decreases as we move away from the centre.



There are :

1. Maximum number of neighborhoods in the first cluster which don't have any coffee shops.

2. Almost no coffee shops in clusters 1 too while there is a moderate number of coffee shops in the second cluster

There is also a high number of neighborhoods which have too many coffee shops. Hence, there is great disparity in the distribution of coffee shops across neighborhoods. This also shows that any potential investors would be better suited investing in clusters 0, 1 or 3.

Limitations and Suggestions for Future Research

In this project, we've only considered one factor: the frequency of occurrence of coffee shops. In any future research, other features like population and income of the neighborhood should be considered as that will provide a comprehensive result. Also, as this project made use of the free SandBox Tier of the Foursquare API, it was limited to a specific number of API calls and results that could be returned. With a paid account we can bypass these limitations and obtain much more holistic results.

Conclusion

Why was this relevant ?

Most people have to travel to another neighborhood each day just to get a cup of coffee. And with the added strain of traffic congestion, opening a coffee shop in a neighborhood that does not have almost none will obviously lead to better profits

Answer to the business question :

Potential investors should open the coffee shop in Clusters 0, 1 or 3. Most preferably, the third cluster should be used as it is near enough to the heart of the city and most neighborhoods in that area show successful coffee shops, so people travelling to those neighborhoods to get coffee will come to a coffee shop in their neighborhood instead.

References

Category: Neighbourhoods in Delhi Wikipedia retrieved from https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi

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