Predicting the best location to open a restaurant

Ankit Pranay

December 30, 2020

1. Introduction

1.1 Background

Location analysis is a technique for finding the best location for your new restaurant. So, choosing a good location for your business might be the single most effective thing you can do to succeed. There are many factors to consider when performing a location analysis and looking for a good location for your new restaurant, such as accessibility, zoning, crime rates, and local demographics. Performing a location analysis can help you find the best spot for your business, ensuring that your new business starts off on the right foot.

1.2 Business Problem

We need data that might help us in deciding the location for your new restaurant so that it starts off on the right foot. This project aims to predict the location best fit for a particular kind of restaurant.

2. Data Description

2.1 Data Sources

- I am using the Delhi Neighbourhood dataset from Kaggle to get the different city locations along with its latitude and longitude.
- I am using Delhi Metro dataset from Kaggle to know about the accessibility for a location.
- I am using the Restaurant dataset from Kaggle to know more the different types of restaurant.
- Using Foursquare API to know about the different venues available for a location.

3. Methodology

Folium

Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. All cluster visualization are done with the help of Folium which in turn generates a Leaflet map made using OpenStreetMap technology.

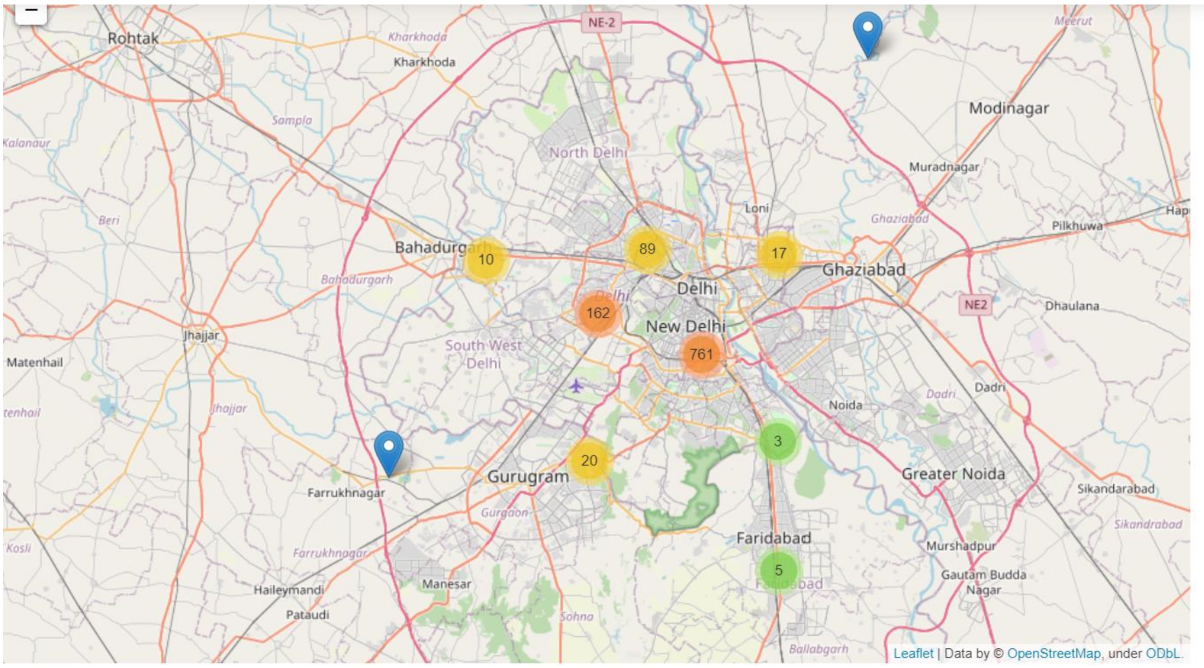

Fig.: Delhi Restaurants and Metro Stations.

One hot encoding

One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. For the K-means Clustering Algorithm, all unique items under Category are one-hot encoded.

	Neighborhood	Afghan Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Awadhi Restaurant	BBQ Joint	Bagel Shop	Bakery	Bengali Restaurant	Bistro	Breakfast Spot	Buffet	Burger Joint
0	Adarsh Nagar	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Adarsh Nagar	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Adarsh Nagar	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Adarsh Nagar	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Ashok Vihar	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig.: One-hot encoded data.

Top 10 most common venues

Due to high variety in the venues, only the top 10 common venues are selected and a new DataFrame is made, which is used to train the K-means Clustering Algorithm.

	Neighborhood	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	Adarsh Nagar	Fast Food Restaurant	Pizza Place	Indian Restaurant	Vegetarian / Vegan Restaurant	Dumpling Restaurant	Dhaba	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop
1	Alaknanda	BBQ Joint	Indian Restaurant	New American Restaurant	Restaurant	Middle Eastern Restaurant	Pizza Place	Steakhouse	Deli / Bodega	Dhaba	Dim Sum Restaurant
2	Anand Vihar	Indian Restaurant	Pizza Place	Indian Sweet Shop	Soup Place	Punjabi Restaurant	Vegetarian / Vegan Restaurant	Donut Shop	Deli / Bodega	Dhaba	Dim Sum Restaurant
3	Ashok Vihar	Indian Restaurant	Bakery	Diner	Falafel Restaurant	Dhaba	Dim Sum Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
4	Azadpur	Café	Argentinian Restaurant	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop

Fig.: Top 10 most common venues.

K-means clustering

The venue data is then trained using K-means Clustering Algorithm to get the desired clusters to base the analysis on. K-means was chosen as the variables (Venue Categories) are huge, and in such situations K-means will be computationally faster than other clustering algorithms.

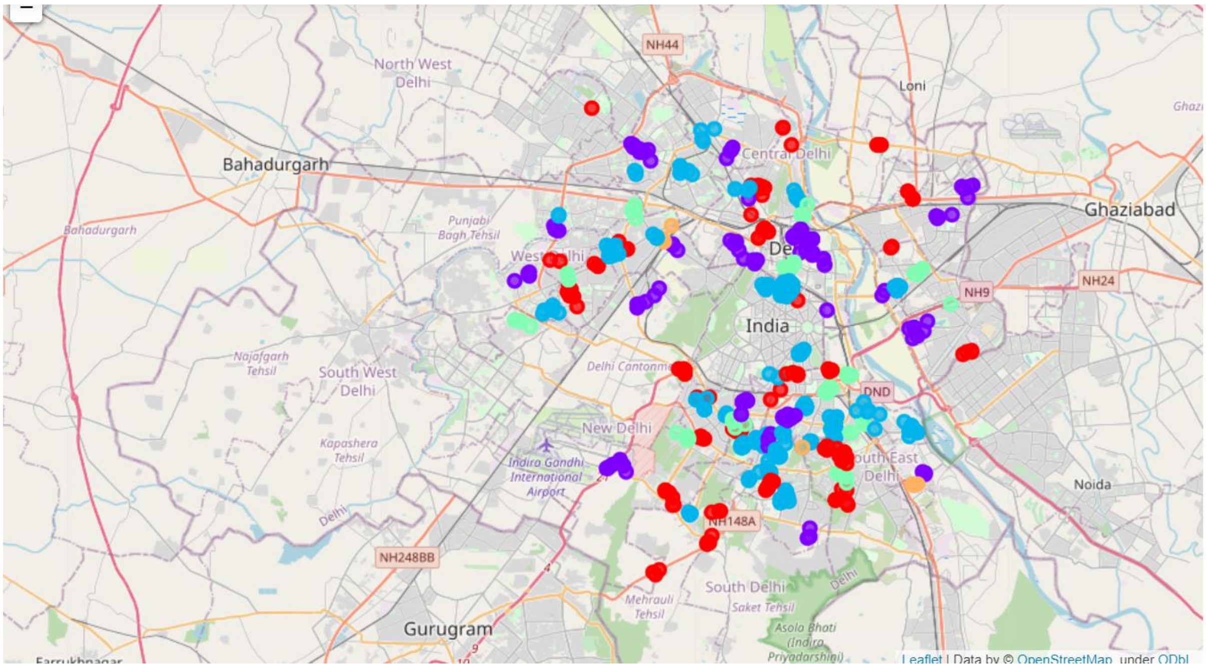

Fig.: Restaurants of Delhi (Clustered)

Analysis of Clusters

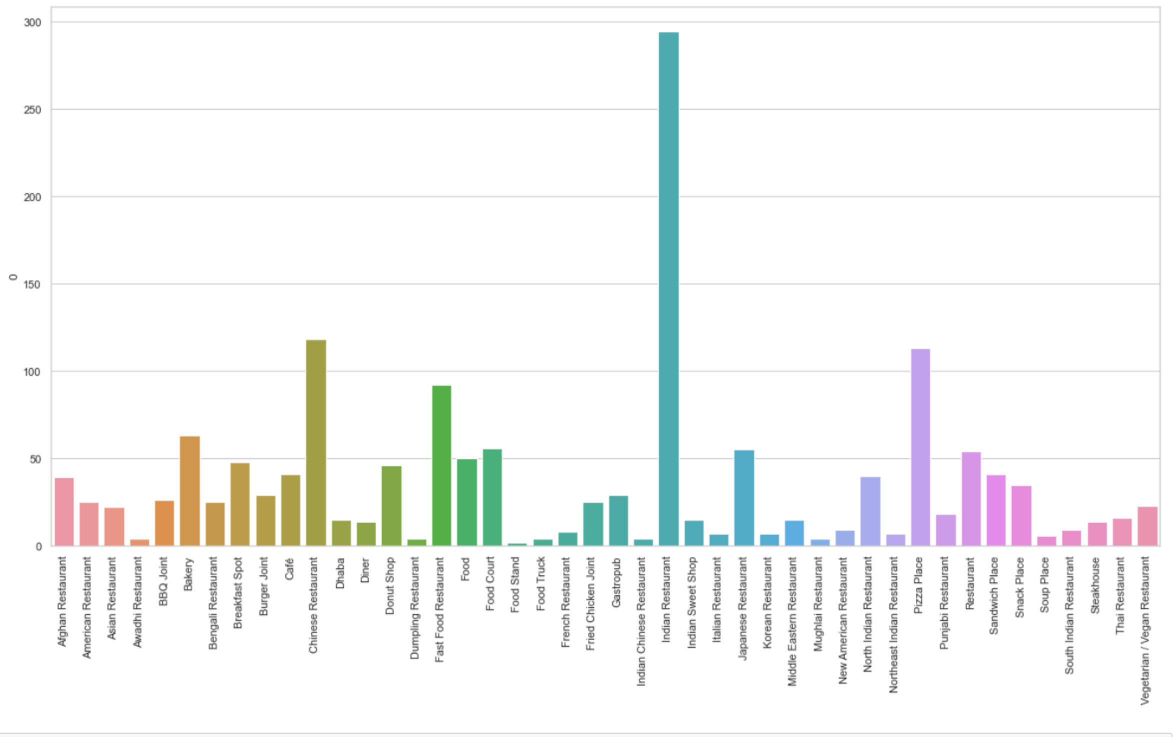

Fig.: Cluster Label 0

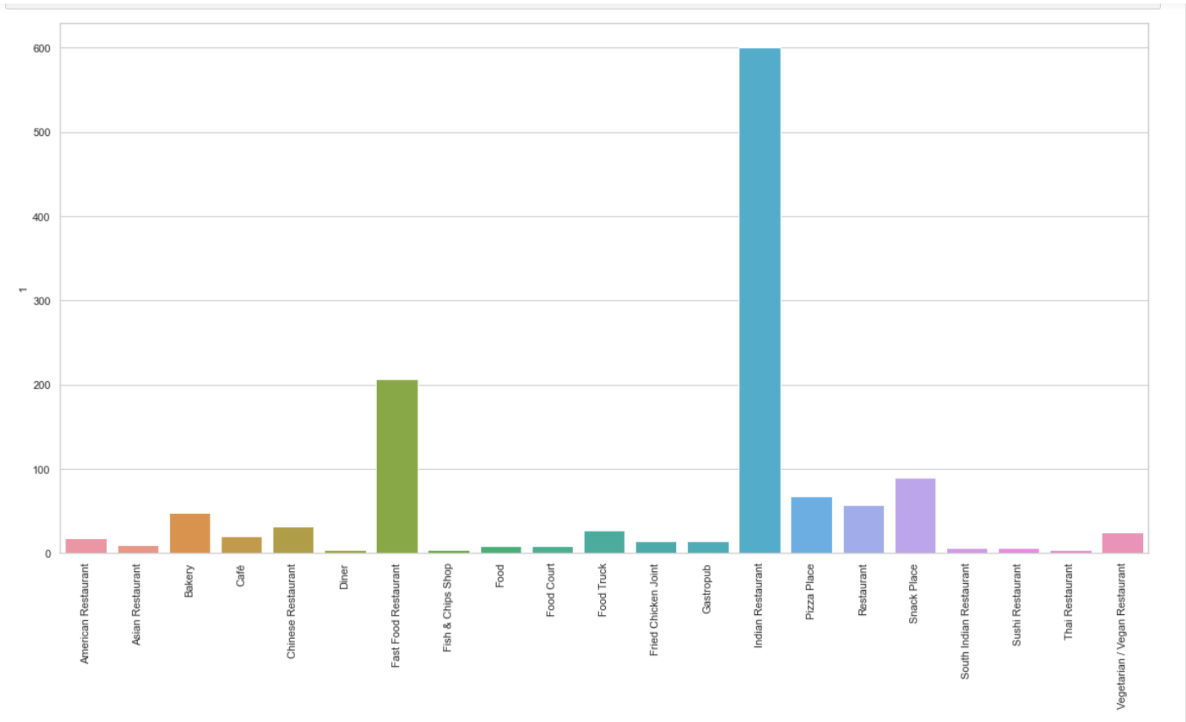

Fig.: Cluster Label 1

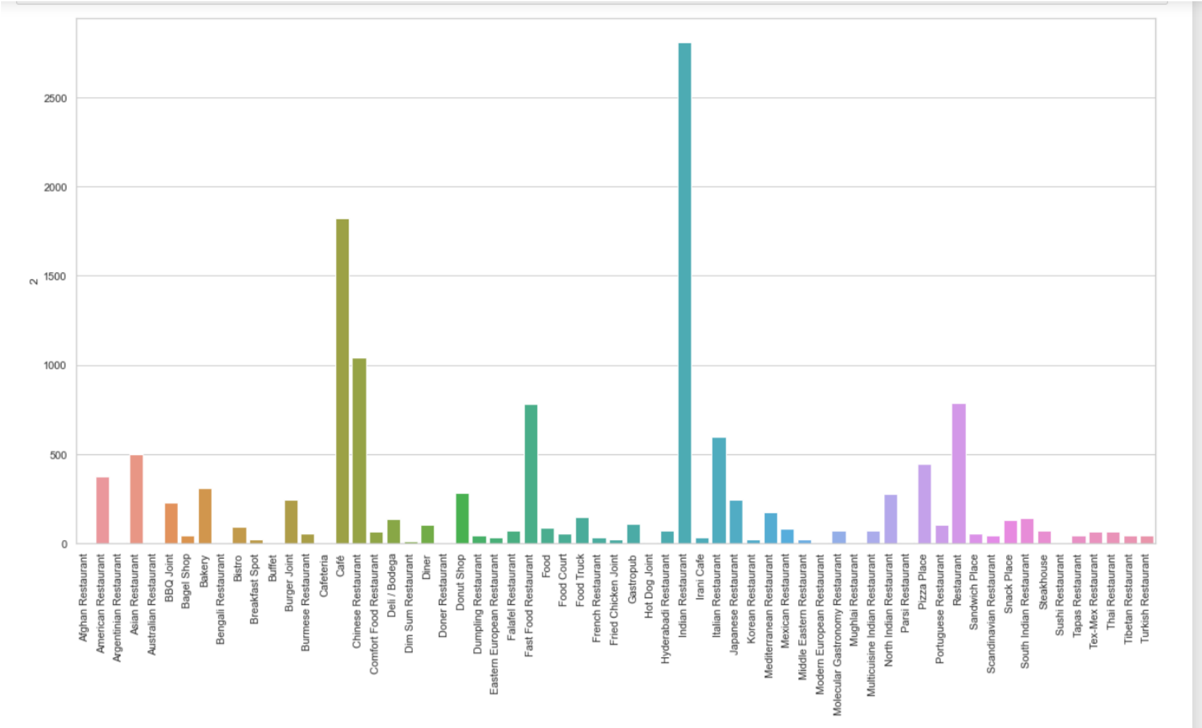

Fig.: Cluster Label 2

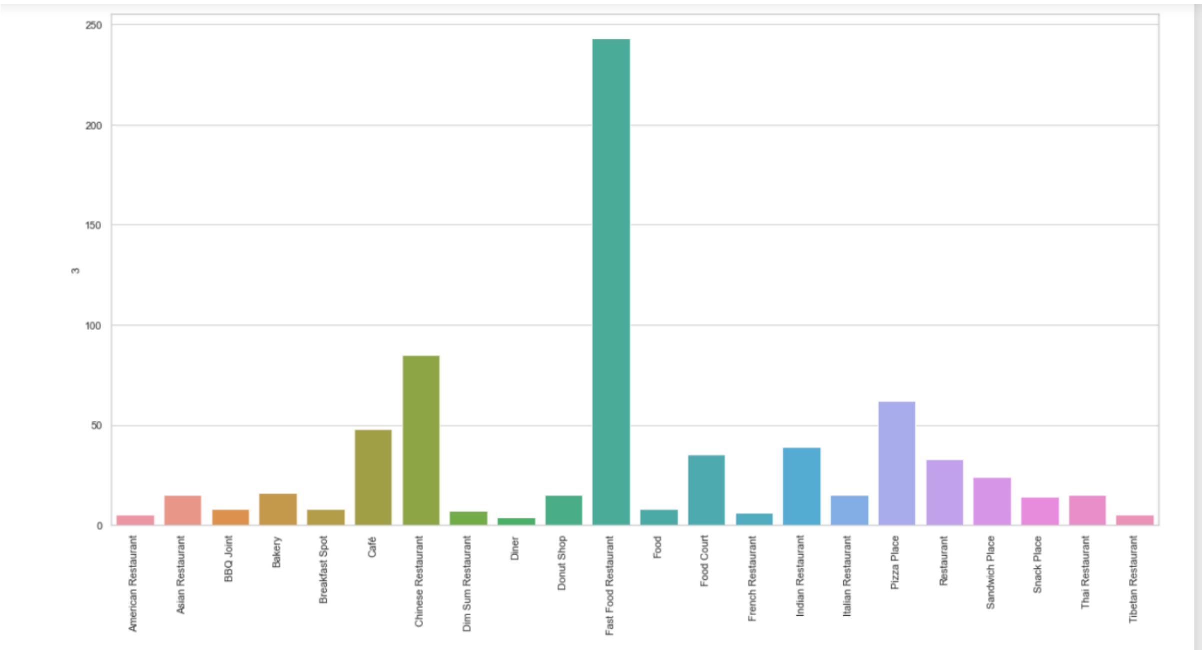

Fig.: Cluster Label 3

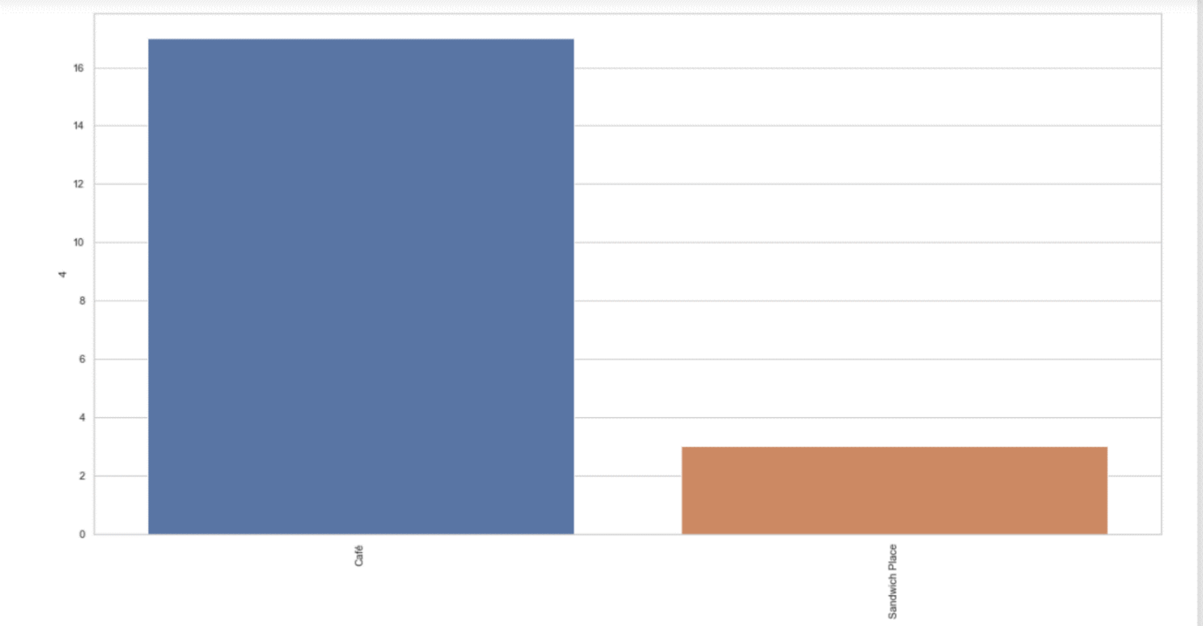

Fig.: Cluster Label 4

4. Result & Discussion

Our Analysis was done on over 186 neighborhoods, containing over 848 restaurants within 2km radius of every neighborhood. We segragated these neighborhoods on the basis of types and amounts of restaurants. Five clusters were obtained, each having a unique collection of restaurants. Since, we were focused on finding optimal neighborhoods for opening Indian restaurants, we selected cluster 2 and 3 which had the highest number of Indian restaurants. The above actions left us with the only those neighborhoods that had a shared characteristics of and that had a high demand for Indian restaurants.

The neighborhoods recommendation obtained here are not completely accurate. This is due to the limitations in the dataset used in the project. Due to lack of cross referencing sources, we may have missed a few neighborhoods from our consideration. The foursquare API does not contain, or does not rely, a comprehensive dataset about the restaurants present in delhi. Surely, in a city like Delhi with a population of over 19 million, there are much more restaurants than 848.

5. Conclusion

In this study, I analyzed the Delhi neighborhood having different types of restaurants across the state. I analyzed different clusters where Indian restaurants were present in huge numbers.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.