

Starting a Wholesale Coffee Supplier Business in Delhi, India

Kanishk Kumar

February 28, 2021

1. Introduction

1.1 Background

Delhi is India's second largest city with a population of more than 11 million and a density of 11,312 people per square kilometer. It is one of the oldest cities in the world, and has been continuously inhabited since the 6th century BCE serving as a capital of various kingdoms and empires. For many centuries Delhi has been a dominant trading and commercial center in northern India, and after 1990s it has emerged as an important node in the international corporate and financial network. This leads the city to become a world leader among other metropolitan and cosmopolitan cities from many sectors, including business.

1.2 Problem

Stardoves, a hypothetical wholesale coffee beans supplier is looking for the right neighborhood in Delhi for setting up their services. To ensure this project's success, the team requires insights into the demographics and neighboring businesses. For each neighborhood, we can ask: Which has the highest coffee store and cafe density? Are its surrounding neighborhoods compatible and fall in the same segment? Thus, the project goal is to figure out the best locations for setting up the agency in Delhi that meets the criteria.

1.3 Interest

Obviously, entrepreneurs who are passionate about wholesale businesses in a metropolitan city would be very interested in this project. The project is also for business owners and stakeholders who want to expand their businesses and wonder how data science could be applied to the questions at hand.

2. Data

2.1 Data Requirements and Collection

Most of the neighborhood data can be found on Wikipedia. [This](#) Wikipedia page will be scraped for Borough and Neighborhood details of Delhi. Coordinates of the neighborhoods from the

scraped data will be obtained using [Geopy](#). The number of venues, their type and location in every neighborhood will be obtained using [Foursquare API](#).

2.2 Data cleaning and Feature Extraction

The first data is a Wikipedia page about the neighborhoods of Delhi. We will scrape the page and create a data frame consisting of two columns; Borough, and Neighborhood. We remove any rows that do not have borough assigned. Then, we will be using the Geocoder python package to retrieve the neighborhood's coordinates. It will return 108 rows and 4 columns.

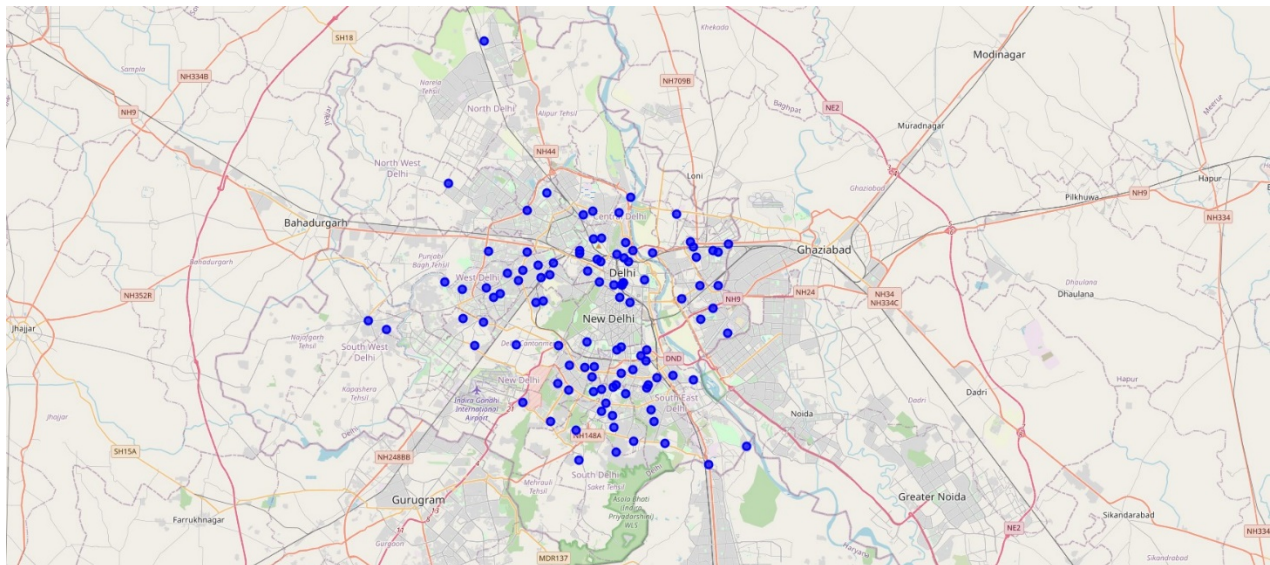
The second data is stored inside Foursquare Location Data, and we will use Foursquare API to access it. We utilize the neighborhood names to retrieve popular venues around a specific radius. As a result, the same venue categories will be returned to different neighborhoods. We can use this idea to cluster the neighborhoods based on their venues representing services and amenities.

We will run the k-Means algorithm to perform this clustering with different number of clusters (k). The features will be the mean of the frequency of occurrence of each venue category. Finally, we can visualize the cluster model using the Folium module.

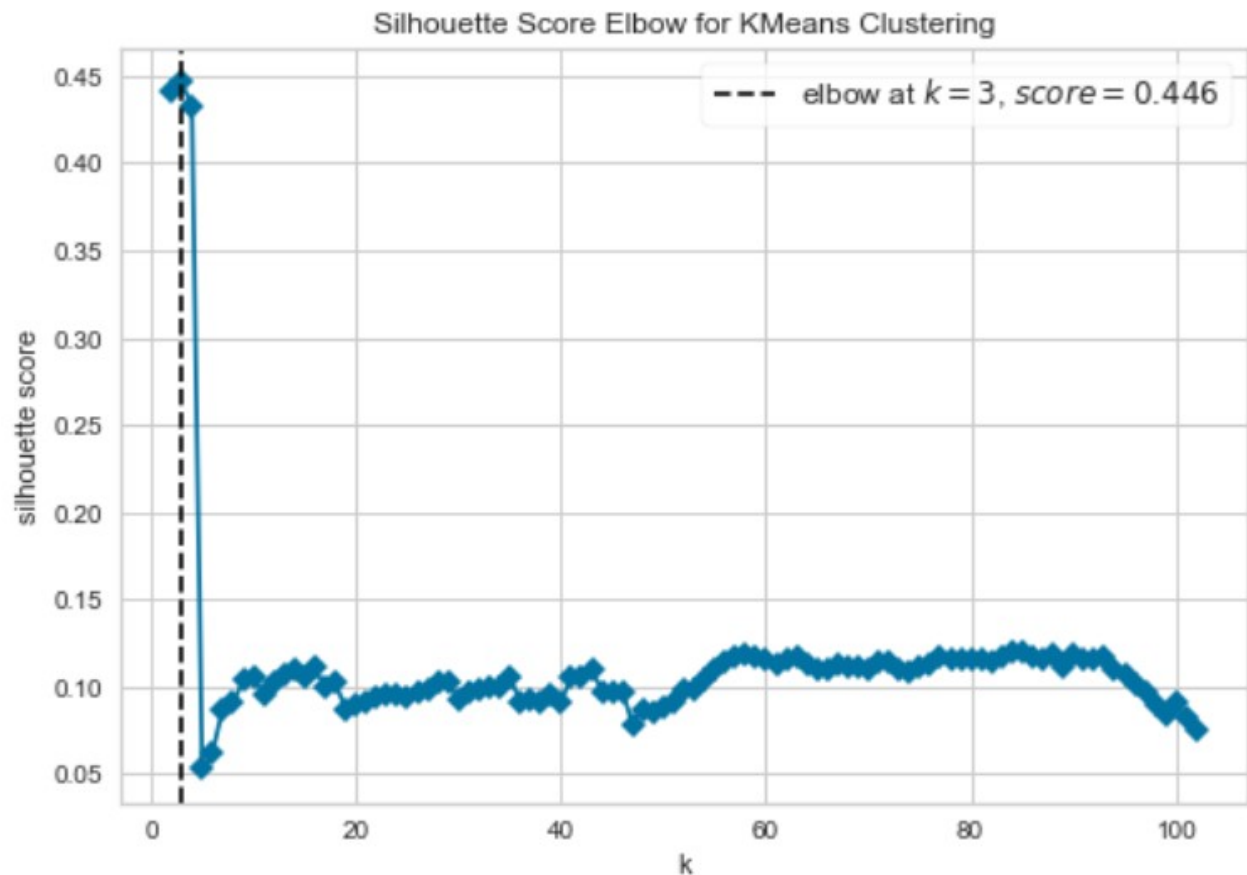
To sum up, we will use the 1st data to obtain the exact coordinates for each neighborhood based on their names, allowing us to explore and map the city. We will then use the coordinates and Foursquare credentials to access the 2nd data source through its API and retrieve the popular venues along with their details, especially for Cafes and Coffee Shops.

3. Exploratory Data Analysis

I plotted the neighborhood data on the map of Delhi using the Folium library.



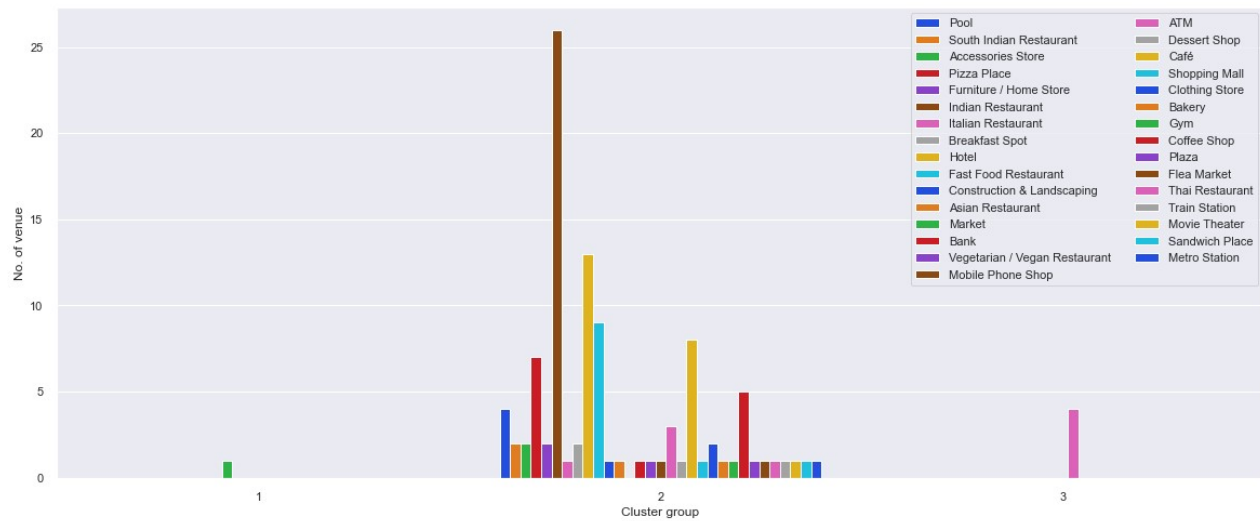
I used the Silhouette Coefficient Method to find the optimal numbers of clusters and visualized it using KElbowVisualizer.



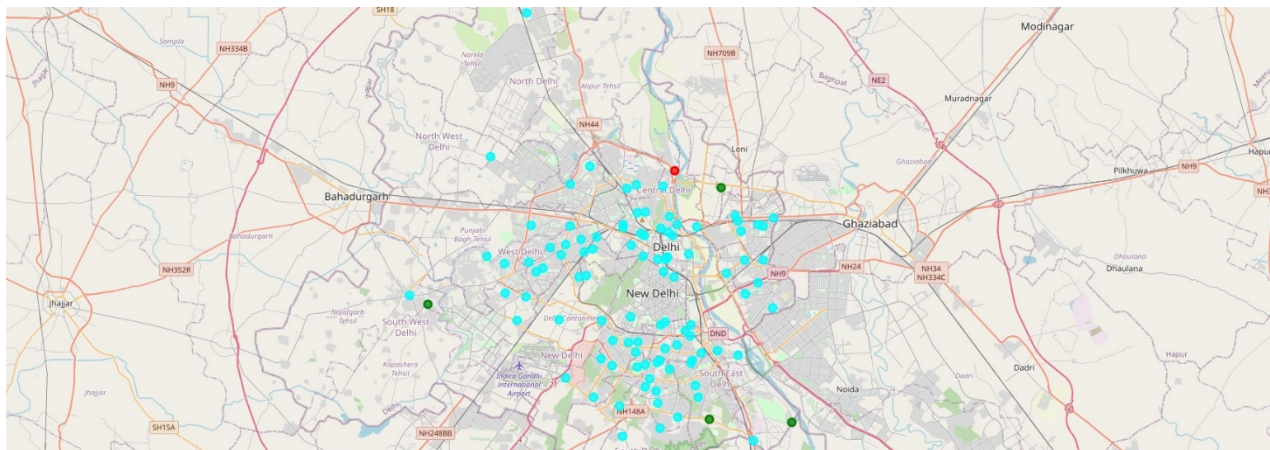
From the above plot, I observed that 3 is the optimal number of clusters. So, I applied K-Means Clustering method to segment the neighborhoods into three clusters.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	North West Delhi	Adarsh Nagar	28.614193	77.071541	1	Pool	Indian Restaurant	Gym
1	North West Delhi	Ashok Vihar	28.699453	77.184826	1	South Indian Restaurant	Asian Restaurant	Donut Shop
2	North West Delhi	Begum Pur	28.725503	77.058371	1	Accessories Store	Women's Store	Flower Shop
3	North West Delhi	Model Town	28.702714	77.193991	1	Pizza Place	Department Store	Track Stadium
4	North West Delhi	Narela	28.842610	77.091835	1	Furniture / Home Store	Mobile Phone Shop	Women's Store
...
102	West Delhi	Rajouri Garden	28.645112	77.123933	1	Fast Food Restaurant	Indian Restaurant	Café
103	West Delhi	Tihar Village	28.634636	77.107112	1	Fast Food Restaurant	Restaurant	Donut Shop
104	West Delhi	Tilak Nagar	28.639650	77.094039	1	Clothing Store	Fast Food Restaurant	Donut Shop
105	West Delhi	Vikas Nagar	28.644009	77.054470	1	Pool	Business Service	Music Store
106	West Delhi	Vikaspuri	28.638419	77.070836	1	Pizza Place	Café	Fast Food Restaurant

I visualized the neighborhood clusters with Countplot using the Seaborn library to check which cluster has the most number of Cafes and Coffee Shops.



In Cluster 2, we have the highest number of Cafes and Coffee Shops, in Cluster 1, we mostly have Markets and Women's Stores and in Cluster 3, we have ATMs and Flower Shops. I plotted and examined the above data on map.



4. Results

So I filtered out and created two separate data frames using the segmented data from Cluster 2.

I found that the best neighborhoods are Sarvodaya Enclave and Munirka as they both have either Cafés or Coffee Shops as their most common venue in the segmented cluster.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Alaknanda	Coffee Shop	Restaurant	Asian Restaurant
Hauz Khas	Coffee Shop	Indian Restaurant	Café
Neeti Bagh	Coffee Shop	Pizza Place	Café
Sarvodaya Enclave	Coffee Shop	Café	Indian Restaurant
Rama Krishna Puram	Coffee Shop	Thai Restaurant	Convenience Store
East Vinod Nagar	Café	Dessert Shop	Sandwich Place
Hauz Khas Village	Café	Indian Restaurant	Coffee Shop
New Friends Colony	Café	Hotel	Indian Restaurant
Safdarjung Enclave	Café	Park	Lounge
Siri Fort	Café	Indian Restaurant	Market
Dhaura Kuan	Café	Indian Restaurant	Bakery
Munirka	Café	Coffee Shop	Indian Restaurant
Vasant Vihar	Café	Chinese Restaurant	Pizza Place

5. Discussion

I found the best neighborhoods for opening wholesale coffee supplier business by comparing the most common venues between the neighborhoods which had Coffee Shops and Cafes in Cluster 2. It was important that the neighborhoods belong to the same cluster.

6. Conclusion

In this study, I analyzed the neighborhoods of Delhi. I identified and segmented its neighborhoods based on the number and types of venues in them. I built clustering models and classification models to find the best neighborhoods for a Wholesale Coffee Supplier Business. These models can be very useful for entrepreneurs who are passionate about wholesale business in a metropolitan city.