DATA SCIENCE CAPSTONE PROJECT

BATTLE OF NEIGHBORHOODS

INTRODUCTION

Mr. X, a food technologist, manufactures and supplies tea and coffee dispenser machine to food eateries in the most common venues in the Central district of Delhi and makes good profit. He supplies to the most common venues as they attract the maximum customers resulting in higher profits and good branding for his products for future growth.

To play safe he is thinking of checking the similarity between different districts of Delhi & expand his business in the district that is most similar to Central Delhi.

So to check the similarity between different districts of Delhi, considering both the population as well as the most common venues is important.

Aim: To perform K Means to cluster different districts based on their population & most common food venues & select the most appropriate district for business expansion.

The target audience is Mr. X.

DATA

The data on population in each of the 11 districts has been extracted from https://www.nriol.com/india-statistics/delhi/districts.asp using BeautifulSoup library & the unnecessary columns are dropped from the table. The population is normalized to get the **population fraction**.

Foursquare API is used to get the nearby venues in each of the districts with the help of the location coordinates. The top 20 most common venues for each district are then displayed using one-hot encoding.

	Name of District	Population	Latitude	Longitude
0	Central Delhi	0.032668	28.6643	77.2167
1	North Delhi	0.049871	28.7886	77.1412
2	South Delhi	0.154328	28.4817	77.1873
3	East Delhi	0.096406	28.6280	77.2956
4	North East Delhi	0.126496	28.7184	77.2580
5	South West Delhi	0.129410	28.5929	77.0346
6	New Delhi	0.007548	28.6139	77.2090
7	North West Delhi	0.206124	28.7186	77.0685
8	West Delhi	0.142915	28.6663	77.0680
9	Shahdara	0.018230	28.6894	77.2919
10	South East Delhi	0.036004	28.5630	77.2611

Table 1: Districts of Delhi dataframe

	Venue ID	Venue Latitude	Venue Longitude	Venue Name	Venue Category	Name of District	Population
1	4d9d759348b6224b70c2249f	28.658050	77.223377	Kake Di Hatti | काके दी हट्टी	Indian Restaurant	Central Delhi	0.032668
2	5662936e498e19a9801a663f	28.657325	77.224138	Amritsari Lassi Wala	Snack Place	Central Delhi	0.032668
4	5654478f498e405c3e4e0d6a	28.641451	77.215506	The Drunkyard Cafe	Tibetan Restaurant	Central Delhi	0.032668
6	54783eab498e910f8bd1781d	28.634455	77.222139	Naturals Ice Cream	Ice Cream Shop	Central Delhi	0.032668
7	513a1684e4b0581da224735e	28.693245	77.204948	Big Yellow Door	Café	Central Delhi	0.032668

Table 2: List of nearby venues

METHODOLOGY

I have used K Means for making clusters based on population & most common venues. The elbow point is determined to get the optimal value of K.

I have plotted these clusters on the map of Delhi using Folium.
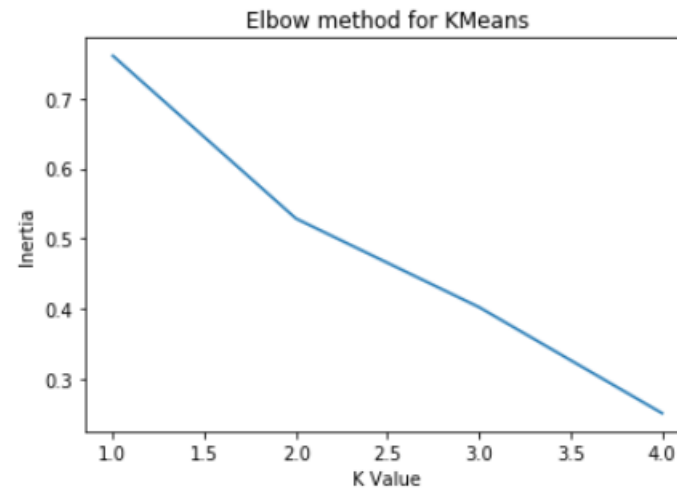
K Means Elbow Point

	Name of District	Latitude	Longitude	Cluster Label
0	Central Delhi	28.6643	77.2167	0
2	South Delhi	28.4817	77.1873	0
3	East Delhi	28.6280	77.2956	1
4	North East Delhi	28.7184	77.2580	0
5	South West Delhi	28.5929	77.0346	1
6	New Delhi	28.6139	77.2090	0
7	North West Delhi	28.7186	77.0685	1
8	West Delhi	28.6663	77.0680	1

Table 3: Cluster Labels using K Means
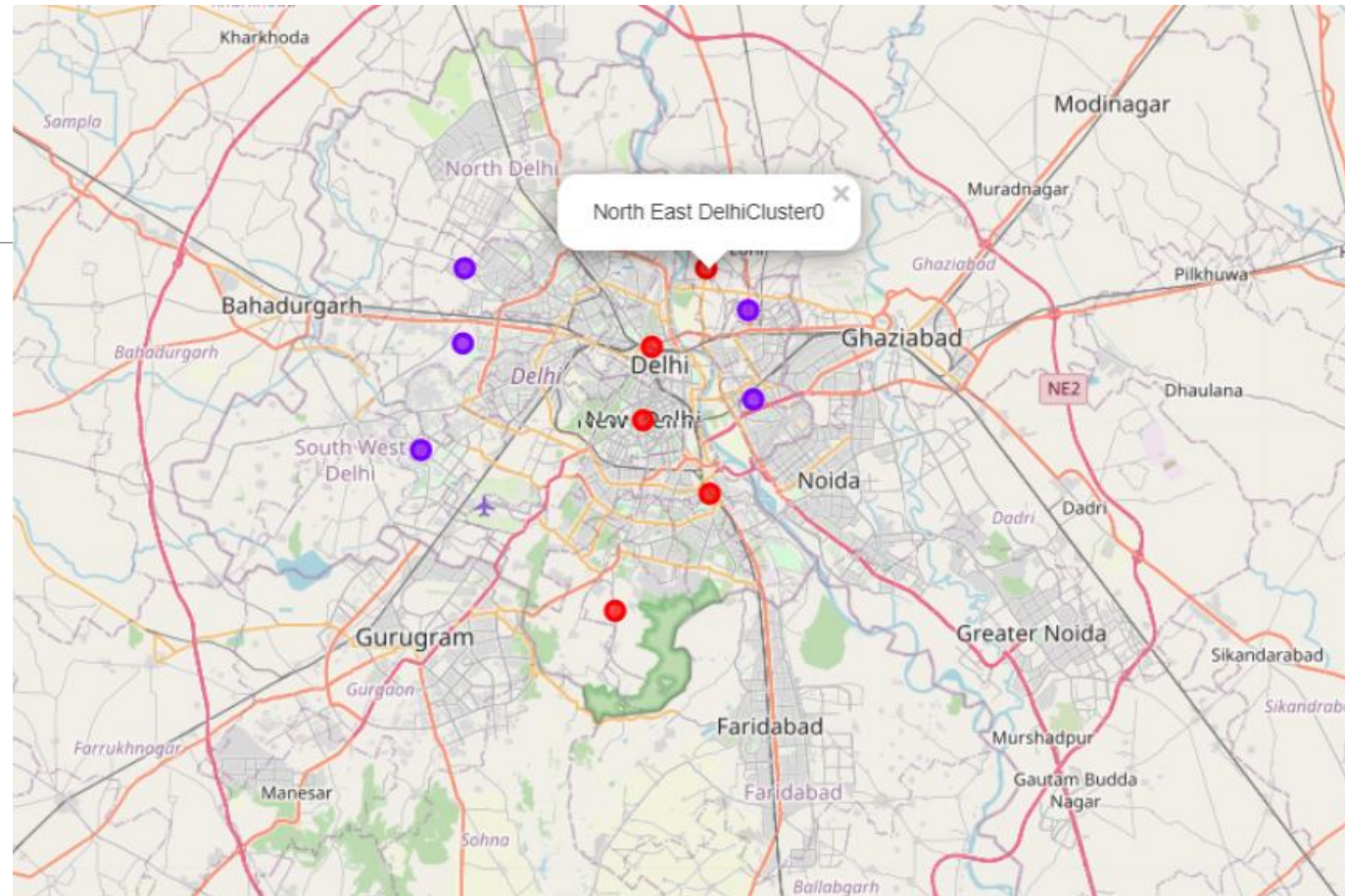
Map of Delhi specifying the clusters

RESULT

The 1st cluster consists of the following districts:- Central Delhi, South Delhi, North East Delhi, New Delhi & South East Delhi.

The 3rd cluster consists of the following districts:- East Delhi, South West Delhi, North West Delhi, West Delhi & Shahdara.

	Name of District	Population	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	Central Delhi	0.032668	Indian Restaurant	Café	Fast Food Restaurant	Snack Place	Lounge	Coffee Shop	Bakery	Donut Shop	Pizza Place	Tibetan Restaurant
2	South Delhi	0.154328	Italian Restaurant	Café	Coffee Shop	Thai Restaurant	Pub	Indian Restaurant	Donut Shop	Food Truck	Asian Restaurant	Breakfast Spot
4	North East Delhi	0.126496	Tibetan Restaurant	Restaurant	Café	Coffee Shop	Asian Restaurant	Sandwich Place	Food Court	Food	Fast Food Restaurant	English Restaurant
6	New Delhi	0.007548	Indian Restaurant	Café	Lounge	Bar	Mediterranean Restaurant	Coffee Shop	Hotel Bar	North Indian Restaurant	Chinese Restaurant	Restaurant
10	South East Delhi	0.036004	Indian Restaurant	Restaurant	Café	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Bakery	Dessert Shop	Japanese Restaurant

Table 4: Cluster 1

	Name of District	Population	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
3	East Delhi	0.096406	Pizza Place	Indian Restaurant	Café	Fast Food Restaurant	Coffee Shop	Asian Restaurant	Food Court	Food Truck	Dessert Shop	Department Store
5	South West Delhi	0.129410	Fast Food Restaurant	Indian Restaurant	Café	Pizza Place	Fried Chicken Joint	Chinese Restaurant	Cocktail Bar	Coffee Shop	Asian Restaurant	Restaurant
7	North West Delhi	0.206124	Indian Restaurant	Fast Food Restaurant	Department Store	Pizza Place	Tibetan Restaurant	Food Court	Food & Drink Shop	Food	English Restaurant	Dumpling Restaurant
8	West Delhi	0.142915	Fast Food Restaurant	Coffee Shop	Pizza Place	Indian Restaurant	Café	Restaurant	Donut Shop	Department Store	BBQ Joint	Sandwich Place
9	Shahdara	0.018230	Pizza Place	Food Truck	Dessert Shop	BBQ Joint	Fast Food Restaurant	Café	Italian Restaurant	Food Court	Food & Drink Shop	Food

Table 6: Cluster 2

RESULT

We are interested in the cluster having "Central Delhi".

In **Central Delhi**, most common are Indian Restaurant, Café & Fast Food Restaurant.

In **South Delhi**, most common are Italian Restaurant, Café & Coffee Shop.

In **North East Delhi**, most common are Tibetan Restaurant, Restaurant & Café.

In **New Delhi**, most common are Indian Restaurant, Café & Lounge.

In **West Delhi**, most common are Fast Food Restaurant, Coffee Shop & Pizza Plaza.

In **South East Delhi**, most common are Indian Restaurant, Restaurant & Café.

Also, the district with the highest population among these 5 is **South Delhi**.

Now, **New Delhi District** has same top 2 most common eateries but the population is way lesser than **Central Delhi**. This means that even the most common venues would not be having enough footfall.

DISCUSSION

If we look at all the common venues, it is clearly evident that **South East Delhi** has 5 similar venues as **Central Delhi**. Also it has the **similar population** as Central Delhi.

So, Mr. X can expand his business in **South East Delhi**.

CONCLUSION

The purpose of this project was to explore the different districts of Delhi keeping in mind population & nearby food eateries.

One of the limitations of this project is that I have assigned the population in the district to each food outlet in the district. However, each district of Delhi is quite big in size and needs to be broken down to neighborhoods to get a more accurate representation but neighborhood level population data was not available for Delhi.

Therefore, with the available information, it is safe to conclude that **South East Delhi** would be the best option for Mr. X.