The Battle of the Neighborhoods

1.Introduction & Business Problem

Problem Background:

The Indian Retail Industry has emerged as one of the most dynamic and fast – growing industries due to the entry of several new players in the recent times along with rising income levels, growing aspirations, favorable demographics and easy credit availability. It constitutes over 10% of the country's Gross Domestic Product (GDP) and around 8% of the employment and is valued at USD 672 billion, Globally, India is fifth – largest global destination in the rectal space which growing at a rate of 12% per annum. The major Popular Indian retailers are Shopper Stop, Reliance fresh, Big Bazaar. Pantaloons Retail India Ltd, Ebony, Piramyd, Globus, Vishal Mega Mart, More and Subhiksha.

The city of Delhi continues to be enamored with fashion and glamour. The people of the city are known for their flamboyant taste in food, clothing, colors and weddings. The city is a mixed bag of ethnicities and cultures, which is evident in the varied preferences in terms of product and price, among the consumers. Because of this varied consumer mix, the shopping choices and destinations available in the city are massive, which makes Delhi the perfect shopping destination. The city provides shopping opportunities to all consumer segments irrespective of age, gender, budget, preference of brand, and retail experience.

Problem Description:

Retail is the process of selling consumer goods or services to customers through multiple channels of distribution to earn a profit. Retailers satisfy demand identified through a supply chain. The term "retailer" is typically applied where a service provider fills the small orders of many individuals, who are end-users, rather than large orders of a small number of wholesale, corporate or government clientele.
Retail business is highly profitable. There is an ever increasing demand for retail items from consumers. But there is a tough competition from new as well as established retail store in the retail market. So, if any company is planning to launch a retail store, know that it will run successfully depending on various factors. Following points should be considered before starting retail store.
1.Create your mission statement
2.Know how to present your business
3.Figure out your finances
4.Plan for customer service

5.Investigate your space
6.Find a location with good traffic
7.Design the right store layout
8.Take care of legal paperwork
9.Finalize your products
10.Network with potential customers

In this project we will be focusing on finding a right location with good traffic for any company who is planning to open a retail store in Delhi, India. Store location should be such that it attracts a lot of traffic all the time. A place where people like to hang out, eat and chill is ideal for retail stores. Thereby choice of location is very important.

2. Data Description

In this project, we will analyze National Capital of India: Delhi. We will be using the below datasets for analyzing Delhi, India.

Data 1: Population Data

City has a total of 11 Boroughs/District and 27 tehsils/neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 11 boroughs and the neighborhoods that exist in each borough.

Details about Population, Literacy Rate and Sex Ratio of people living in each Tehsil of each District of Delhi is collected from the website of [Census India]
https://www.censusindia.co.in/states/delhi

Data 2: Postal Code Data

To get the coordinates (Latitude, Longitude) of each tehsil/neighborhood of District/Borough, we will use geocoder class library which takes query as address and postal codes and gives output as coordinates of neighborhood. To get postal codes of each tehsil, we will use data set available from website:
http://www.egreenwatch.nic.in/Masters/Public/TehsilView.aspx?SID=10

Data 3: Foursquare API

Delhi City geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in Delhi City. Foursquare API is used to get location coordinates of each Tehsil, venues in the neighborhood of each Tehsil, user ratings of those venues etc. This data will be used to define clusters of Tehsils in which a retail store can be set up.

3. Methodology

In the first step data is collected for getting postal code of all tehsils through which we can visualize all neighborhoods in Delhi and further nearby venues can be explored using Foursquare. In the next step population data is collected and is further explored and analyzed based on Literacy Rate, Sex Ratio In the next step we will use coordinates of each tehsil to explore nearby venues of different category

- Fetching Data
- Visualizing Delhi on Map
- Exploring Population, Literacy Rate and Sex Ratio by Districts of Delhi
- Exploring Venues in the Tehsils

In the last step we will focus on analyzing top ten venues of each neighborhood and based on unique category we will group all neighborhood based on **k-means clustering** algorithm. We will present map of all clusters of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final ' exploration and search for optimal venue location by stakeholders.

3.1 Fetching Data

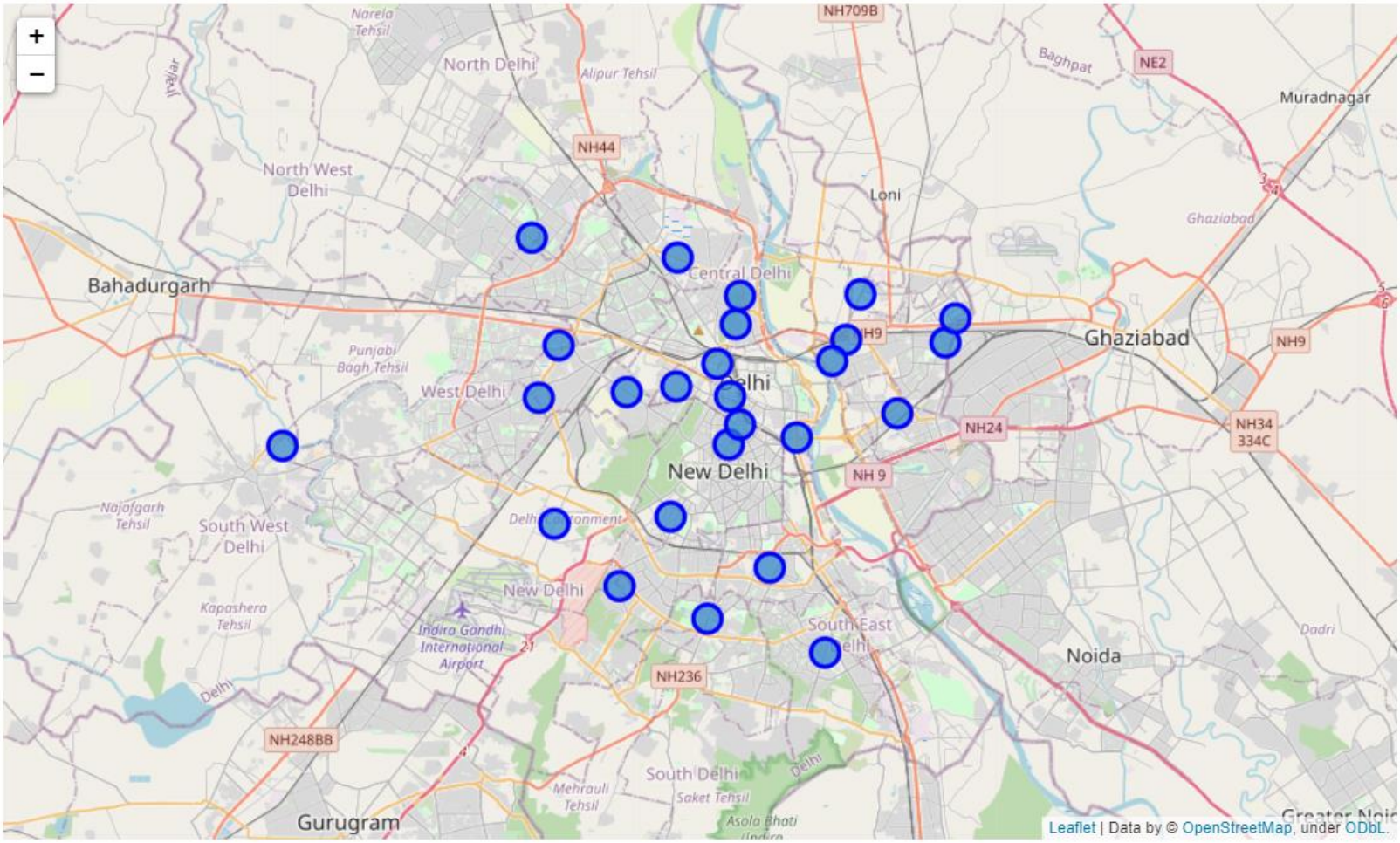
Firstly, data is fetched for getting postal code of all tehsils through which we can visualize all neighborhoods in Delhi and further nearby venues can be explored using Foursquare. then Population data is fetched from census 2011website
https://www.censusindia.co.in/states/delhi

After this both data is merged into common data frame having information of postal code and population of each tehsil of Delhi.

	Borough	Neighborhood	Postal Code	Population
0	North West	Narela	2071	809913
1	North	Civil Lines	2074	688616
2	North East	Seelam Pur	2077	1378779
3	East	Gandhi Nagar	2080	395342
4	New Delhi	Parliament Street	2083	52394
5	Central	Karol Bagh	2086	136599
6	West	Punjabi Bagh	2089	799453
7	South West	Najafgarh	2092	1365152
8	South	Defence Colony	2095	637775
9	South	Hauz Khas	2096	1231293
10	South West	Delhi Cantonment	2093	286140
11	West	Patel Nagar	2090	1262158
12	Central	Pahar Ganj	2087	174613
13	New Delhi	Connaught Place	2084	28228
14	East	Vivek Vihar	2081	247906
15	North East	Shahdara	2078	322931
16	North	Sadar Bazar	2075	130188
17	North West	Saraswati Vihar	2072	2250816
18	North West	Model Town	2073	595810
19	North	Kotwali	2076	69174
20	North East	Seema Puri	2079	539914
21	East	Preet Vihar	2082	1066098
22	New Delhi	Chanakya Puri	2085	61382
23	Central	Darya Ganj	2088	271108
24	West	Rajouri Garden	2091	481632
25	South West	Vasant Vihar	2094	641666
26	South	Kalkaji	2097	862861

3.2 Visualizing Delhi on Map

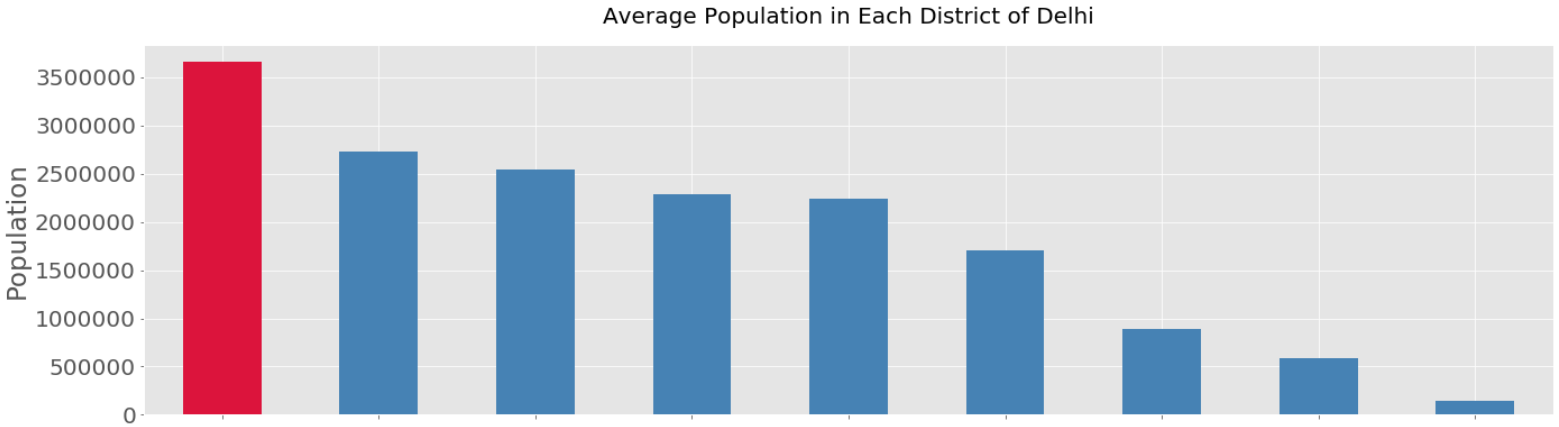
After getting coordinate of each tehsil using geocoder library, we use folium maps to generate each tehsil of Delhi.

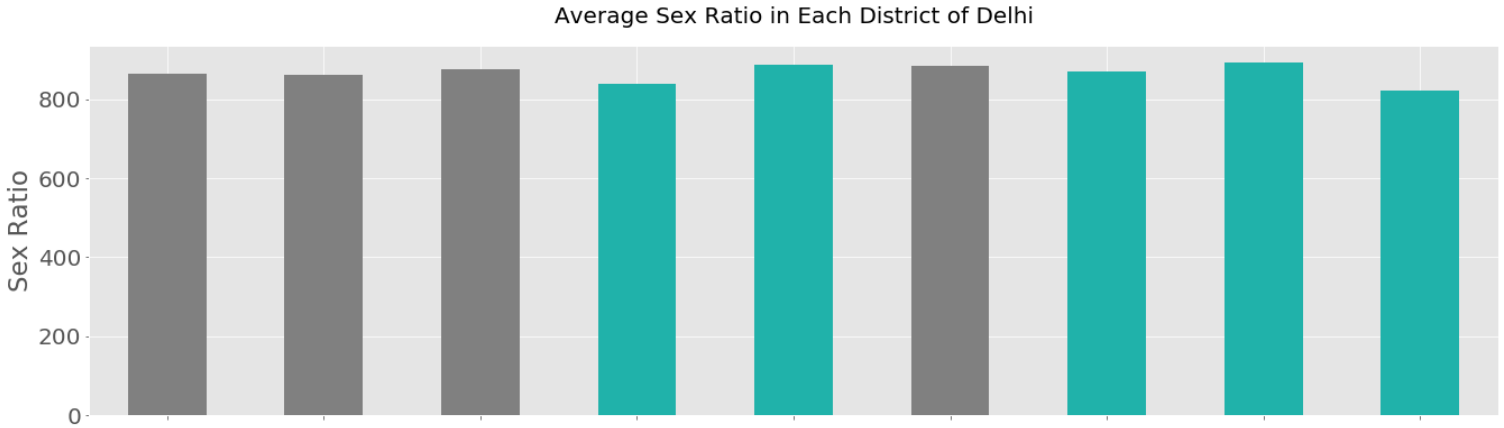


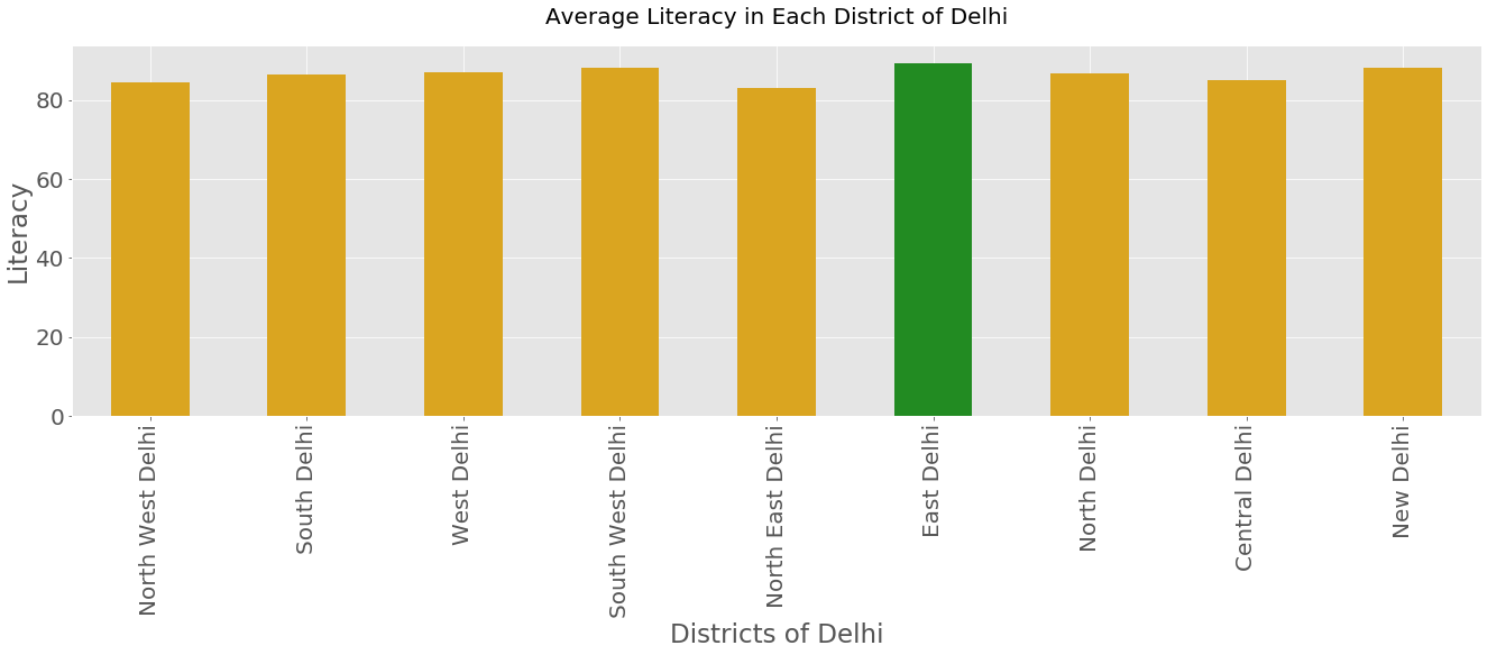
3.3 Exploring Population, Literacy Rate and Sex Ratio by Districts of Delhi

Again fetching Population, Literacy Rate and Sex Ratio by Districts of Delhi from census 2011 and analyzing district by district will give following bar plots.

District Population Sex Ratio Literacy







Observations:

- North West Delhi has highest population followed by South Delhi and then, West Delhi.
- Central Delhi, East Delhi and North East Delhi have highest Sex Ratio
- New Delhi has highest Literacy Rate followed by East Delhi

Inference:

From The Above Plots, we can infer Tehsils in West Delhi can be considered to be good locations for retail store as they have good population size, good sex ratio and descent literacy rate. But for those stores that are aiming for niche market, Tehsils in East Delhi can be better locations to target as it has low population size, high literacy rate and moderate sex ratio.

3.4 Exploring Venues in the Tehsils

Explore top 100 venues in each tehsil using Foursquare API to fetch venues in any Tehsil. For this, we will create GET request URL: After this, we will fetch URL from Foursquare API in JSON format and convert it into data frame. Now, we will get venues in all Tehsils of Delhi.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Narela	28.83979	77.07696	Axis Bank ATM	28.839769	77.076927	ATM
1	Narela	28.83979	77.07696	Axis Bank ATM	28.839769	77.076927	ATM
2	Narela	28.83979	77.07696	Satyam Multiplex	28.846124	77.083340	Multiplex
3	Narela	28.83979	77.07696	vicky traders	28.846180	77.083427	Furniture / Home Store
4	Civil Lines	28.67671	77.21767	moets	28.678015	77.224739	Chinese Restaurant
5	Civil Lines	28.67671	77.21767	Dunkin'	28.680002	77.208363	Donut Shop
6	Civil Lines	28.67671	77.21767	24SEVEN	28.678732	77.208037	Grocery Store
7	Civil Lines	28.67671	77.21767	Maidens Hotel	28.674377	77.226452	Hotel
8	Civil Lines	28.67671	77.21767	Domino's Pizza	28.679000	77.208000	Pizza Place
9	Civil Lines	28.67671	77.21767	The Embassy	28.676556	77.224969	Indian Restaurant
10	Civil Lines	28.67671	77.21767	SUBWAY	28.679507	77.208324	Sandwich Place
11	Civil Lines	28.67671	77.21767	Kamla Nehru Ridge Forest	28.681929	77.216098	Trail
12	Civil Lines	28.67671	77.21767	Coffee Day Express	28.678072	77.224295	Coffee Shop
13	Civil Lines	28.67671	77.21767	Civil Lines Metro Station	28.677042	77.224920	Light Rail Station
14	Civil Lines	28.67671	77.21767	Jain Pan House	28.677842	77.224902	Smoke Shop
15	Civil Lines	28.67671	77.21767	The Exchange Store	28.676799	77.225188	Convenience Store
16	Civil Lines	28.67671	77.21767	Civil Lines | सिविल लाइंस Metro Station	28.675264	77.225098	Light Rail Station
17	Civil Lines	28.67671	77.21767	Nukkad Wala	28.679703	77.208280	Snack Place
18	Seelam Pur	28.67011	77.27122	Metro Wholesale Mall	28.670565	77.268190	Shopping Mall
19	Seelam Pur	28.67011	77.27122	Dilshad Garden Metro Station	28.670221	77.267216	Platform
20	Seelam Pur	28.67011	77.27122	Seelampur Metro Station	28.669805	77.266846	Train Station
21	Seelam Pur	28.67011	77.27122	Welcome Metro Station	28.671902	77.277772	Light Rail Station
22	Seelam Pur	28.67011	77.27122	Axis Bank ATM	28.672850	77.278490	ATM
23	Seelam Pur	28.67011	77.27122	McDonald's	28.663433	77.278282	Fast Food Restaurant
24	Gandhi Nagar	28.66091	77.26432	Rupa Ice Cream Parlour	28.659367	77.264051	Dessert Shop
25	Gandhi Nagar	28.66091	77.26432	Mother Sweets	28.659192	77.264066	Dessert Shop
26	Gandhi Nagar	28.66091	77.26432	Tandoori Night	28.659179	77.264052	Indian Restaurant
27	Gandhi Nagar	28.66091	77.26432	Jheel chowk	28.656600	77.272665	Burger Joint
28	Gandhi Nagar	28.66091	77.26432	Seelampur Metro Station	28.669805	77.266846	Train Station

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Chanakya Puri	14	14	14	14	14	14
Civil Lines	14	14	14	14	14	14
Connaught Place	93	93	93	93	93	93
Darya Ganj	12	12	12	12	12	12
Defence Colony	50	50	50	50	50	50
Delhi Cantonment	4	4	4	4	4	4
Gandhi Nagar	5	5	5	5	5	5
Hauz Khas	92	92	92	92	92	92
Kalkaji	12	12	12	12	12	12
Karol Bagh	37	37	37	37	37	37
Kotwali	6	6	6	6	6	6
Model Town	13	13	13	13	13	13
Najafgarh	1	1	1	1	1	1
Narela	3	3	3	3	3	3
Pahar Ganj	45	45	45	45	45	45
Parliament Street	74	74	74	74	74	74
Patel Nagar	11	11	11	11	11	11
Preet Vihar	18	18	18	18	18	18
Punjabi Bagh	18	18	18	18	18	18
Rajouri Garden	39	39	39	39	39	39
Sadar Bazar	4	4	4	4	4	4
Saraswati Vihar	13	13	13	13	13	13
Seelam Pur	6	6	6	6	6	6
Seema Puri	5	5	5	5	5	5
Shahdara	5	5	5	5	5	5
Vasant Vihar	38	38	38	38	38	38
Vivek Vihar	8	8	8	8	8	8

4.Analysis

4.1 Analyzing Each Neighborhood

Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

	Neighborhood	ATM	Airport Terminal	American Restaurant	Arcade	Art Gallery	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	...	Tea Room	Theater	Tibetan Restaurant	Track	
0	Chanakya Puri	0.000000	0.000000	0.000000	0.000000	0.000000	0.071429	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
1	Civil Lines	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.07
2	Connaught Place	0.000000	0.000000	0.000000	0.000000	0.000000	0.021505	0.000	0.021505	0.00000	...	0.010753	0.00000	0.010753	0.000	0.00
3	Darya Ganj	0.000000	0.000000	0.000000	0.000000	0.083333	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
4	Defence Colony	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
5	Delhi Cantonment	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
6	Gandhi Nagar	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
7	Hauz Khas	0.000000	0.000000	0.021739	0.010870	0.021739	0.032609	0.000	0.000000	0.01087	...	0.010870	0.01087	0.010870	0.000	0.00
8	Kalkaji	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00
9	Karol Bagh	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.027027	0.00000	...	0.000000	0.00000	0.000000	0.000	0.00

Creating data frame consisting of top 10 most common venues in each tehsil

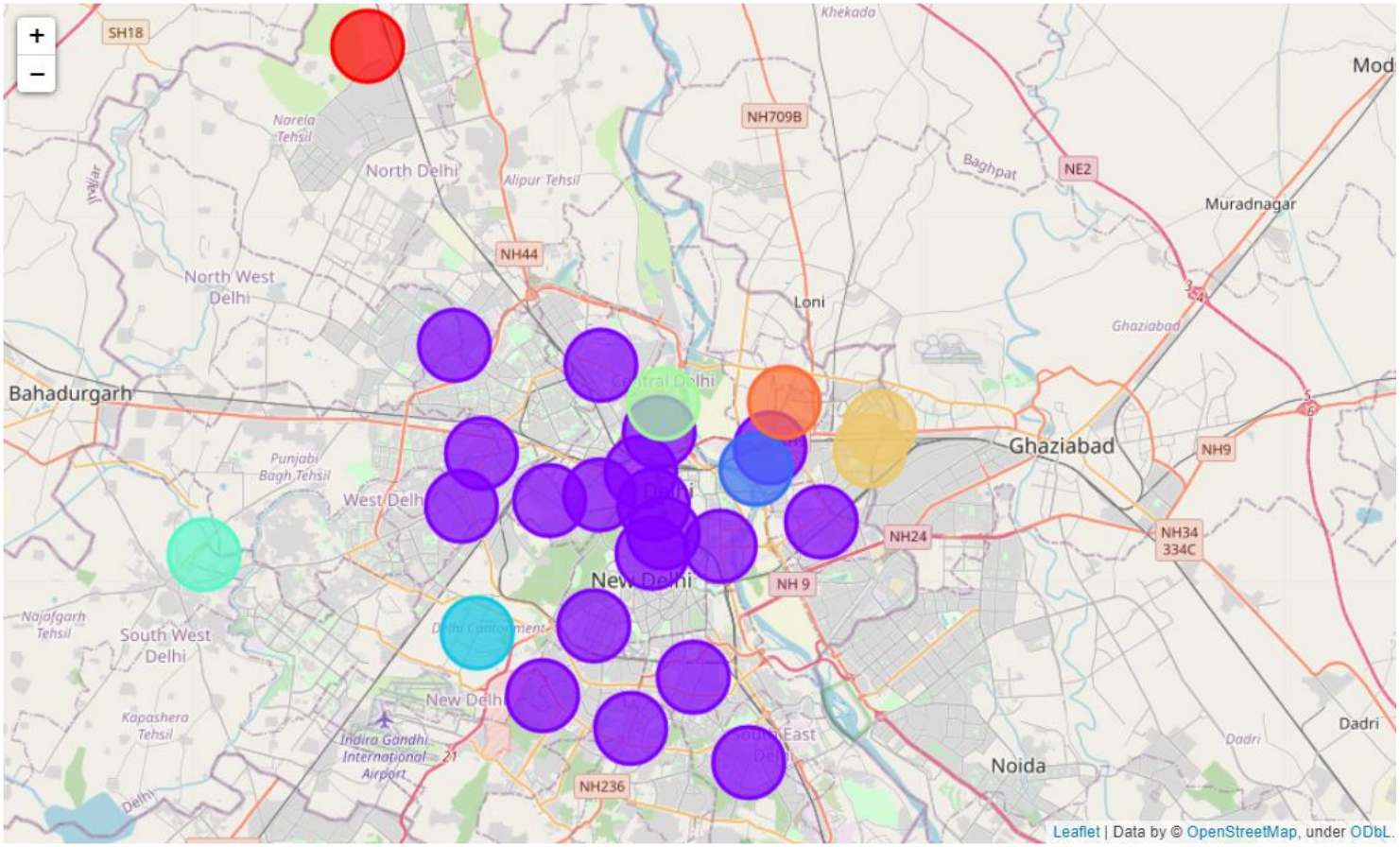
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Chanakya Puri	Indian Restaurant	Hotel	Café	Park	Moroccan Restaurant	Restaurant	Karnataka Restaurant	Pub	Asian Restaurant	Nightclub
1	Civil Lines	Light Rail Station	Pizza Place	Snack Place	Indian Restaurant	Grocery Store	Donut Shop	Convenience Store	Coffee Shop	Chinese Restaurant	Sandwich Place
2	Connaught Place	Indian Restaurant	Café	Hotel	Bar	Chinese Restaurant	Coffee Shop	Lounge	Pub	Deli / Bodega	BBQ Joint
3	Darya Ganj	Café	Art Gallery	Clothing Store	Light Rail Station	Stadium	Historic Site	Fast Food Restaurant	Plaza	Train Station	Hotel
4	Defence Colony	Indian Restaurant	Italian Restaurant	Café	Market	Fast Food Restaurant	Pizza Place	Bakery	Hotel	Sandwich Place	Dessert Shop
5	Delhi Cantonment	Convenience Store	Coffee Shop	Historic Site	Shopping Mall	Yoga Studio	Garden	Event Space	Falafel Restaurant	Fast Food Restaurant	Flea Market
6	Gandhi Nagar	Dessert Shop	Burger Joint	Indian Restaurant	Train Station	Event Space	Falafel Restaurant	Fast Food Restaurant	Flea Market	Food	Food & Drink Shop
7	Hauz Khas	Indian Restaurant	Coffee Shop	Café	Bar	Dessert Shop	Chinese Restaurant	Market	Bakery	Asian Restaurant	Restaurant
8	Kalkaji	Hotel	Market	Fast Food Restaurant	Pizza Place	Train Station	Convenience Store	Restaurant	Indian Restaurant	Sandwich Place	Dumpling Restaurant
9	Karol Bagh	Indian Restaurant	Hotel	Fast Food Restaurant	Coffee Shop	Snack Place	Bakery	Food & Drink Shop	Pizza Place	BBQ Joint	Café

4.2 Clustering Tehsils

Using k means clustering algorithm all tehsils are clustered based on unique category features which is generated using one hot encoding. Total 8 clusters were created which further reduced to 6 based on some common values.

	Borough	Neighborhood	Postal Code	Population	Latitude	Longitude	Cluster Labels	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
0	North West	Narela	2071	809913	28.8398	77.077	0	ATM	Multiplex	Furniture / Home Store	Donut Shop	Event Space	Falafel Restaurant	Fast Food Restaurant
1	North	Civil Lines	2074	688616	28.6767	77.2177	1	Light Rail Station	Pizza Place	Snack Place	Indian Restaurant	Grocery Store	Donut Shop	Convenience Store
2	North East	Seelam Pur	2077	1378779	28.6701	77.2712	1	ATM	Shopping Mall	Light Rail Station	Fast Food Restaurant	Platform	Train Station	Arcade
3	East	Gandhi Nagar	2080	395342	28.6609	77.2643	2	Dessert Shop	Burger Joint	Indian Restaurant	Train Station	Event Space	Falafel Restaurant	Fast Food Restaurant
4	New Delhi	Parliament Street	2083	52394	28.6255	77.2142	1	Hotel	Indian Restaurant	Chinese Restaurant	Bar	Café	Coffee Shop	Spa

This clusters is then plotted on map, where each colored circle represents different cluster.



4.3 Examining Clusters

Cluster 1: North West Delhi
In cluster 1, tehsils that have ATM, Multiplex, Shop and Service are clubbed together.

Cluster 2: New Delhi, Central Delhi, North Delhi, East Delhi, South-West Delhi, West Delhi, South Delhi and North-West Delhi
In cluster 2, Tehsils with Indian/Fast Food/Italian Restaurant, cafe/coffee shop, hotel and pubs are clubbed together. This is the biggest cluster found through k-means algorithm.

Cluster 3: East Delhi
In Cluster 3, tehsils with Dessert Shop, Event Space, Falafel Restaurant and Flea Market are clubbed together.

Cluster 4: South-West Delhi
This Cluster has almost same venue categories as in cluster 1,3. So, Cluster 1,3 and Cluster 4 can be clubbed together.

Cluster 5: North Delhi, West Delhi and South Delhi
In Cluster 5, Tehsils with Resort and Yoga Studio are clubbed together.

Cluster 6:
In Cluster 6, Tehsils with Restaurant and Pizza Place are clubbed together.

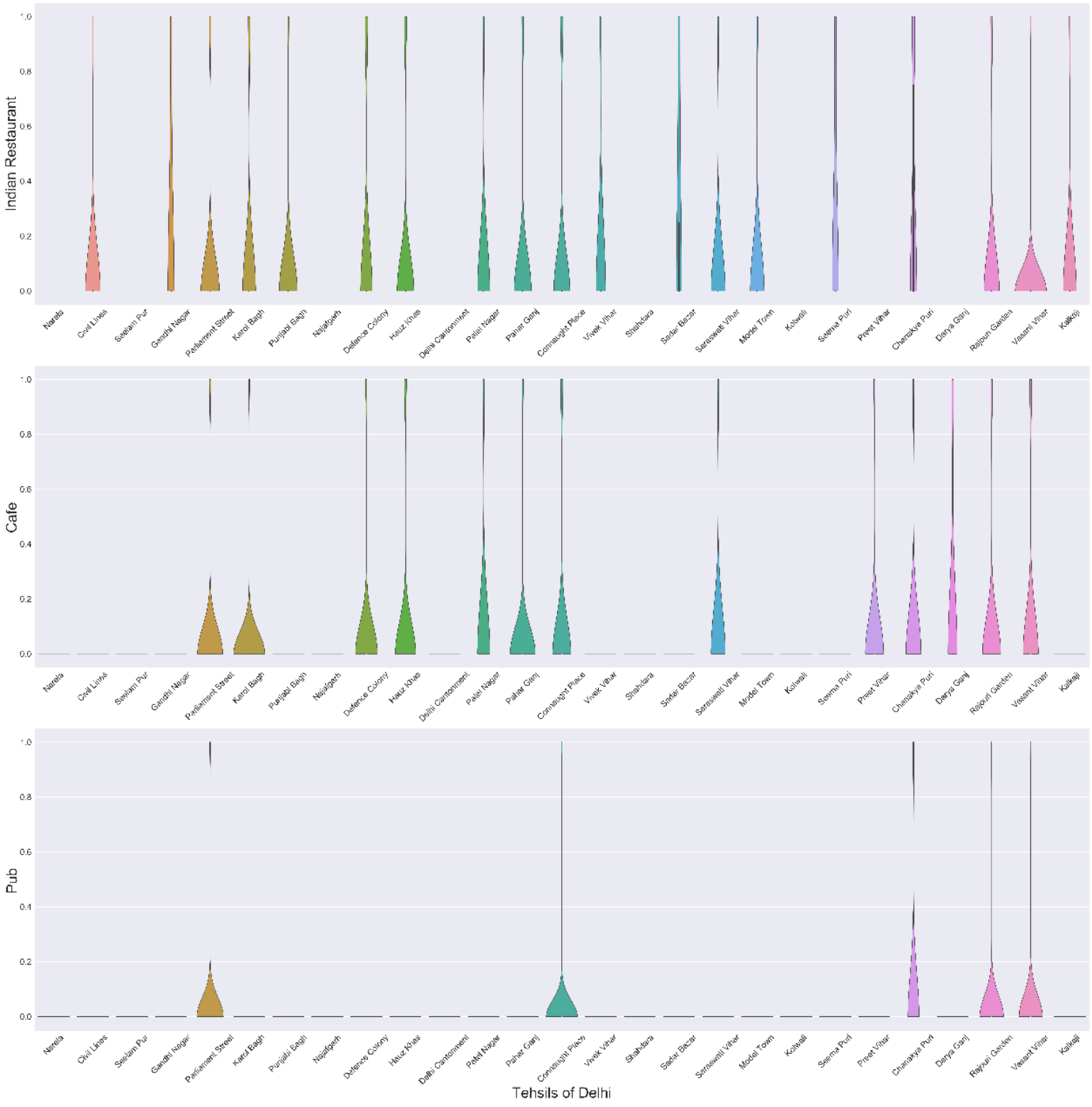
Cluster 7:
In Cluster 7, Shop, Service and Light Rail Station are clubbed together.

Cluster 8:
In Cluster 8 all kinds of Shops, Park and Garden are clubbed together.

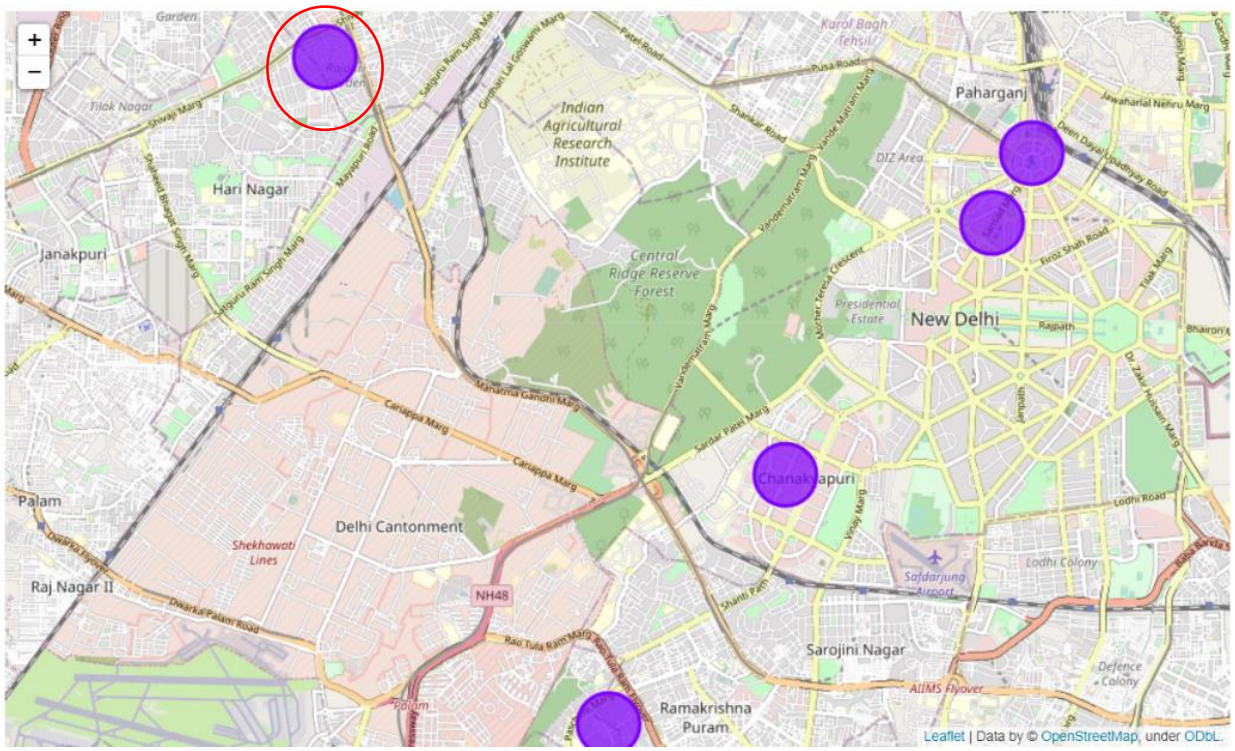
The Clusters 1,3 and 4 are almost similar. So, they can be clubbed together. Hence, there are 6 clusters in total found after clubbing similar clusters.

4..4 Finding and Visualizing Prospective Tehsils Based on a Criteria

As already stated, preferred locations to set up retail stores in Tehsils which has Indian Restaurant, Cafe and Pubs as the most common value as lots of people come at these places for having fun and would also like to do shopping if any retail stores will be present nearby. Violin plot from the seaborn library can be used to analyze venue data - it is a great way to visualize frequency distribution datasets, they display a density estimation of the underlying distribution.



Based on above violin plots, **Chanakya Puri**, **Connaught Place**, **Parliament Street**, **Vasant Vihar** and **Rajouri Garden** can be considered for opening retail store.



5.Results and Discussion

Based on above analysis we can say that West Delhi shows good potential for retail market as West Delhi has good population size, good literacy rate and descent gender ratio, we can narrow down our choice for prospective locations for a retail store to Rajouri Garden as it is in West Delhi District. (Represented by red circle on map)

We can say that based on certain parameters or criteria, we can select one or two clusters of Tehsils of Delhi through the above methodology for the establishment of a retail store.There may be other ways this analysis could have been performed based on different methodology, different algorithm and perhaps different data sources. But I chose this method as the dataset related to tehsils in Delhi is readily available through Census India, and also, K-Means algorithm is easily able to be applied on this dataset.

6.Conclusion

In the analysis performed based on population data, we can improve the analysis by performing various measures. Above analysis have some limitation which we can overcome by performing following measures

1. Area of each Tehsil can also be used for analysis.
2. Data related to centennials and millennials can be employed for decision-making and, thus can be used for further analysis.
3. Store Brands such as Shopper Stop, Reliance fresh, Big Bazaar etc. famous or trending in a particular tehsil can also be helpful to decide better locations for establishment of retail stores.