Most liked Restaurant Cluster in Delhi

Description:

Delhi is the capital of India. The National Capital Region is one of the most famous tourist spot in India. The city on an average has tourist footfall of over 100,000. Delhi has people from different cultures, religions, countries. This makes a hotspot to host various restaurants offering different cuisine food.

In this project, I have taken two popular regions in Delhi namely, New Delhi and South Delhi. These two regions are one of the posh locations in Delhi and food hub with various cuisines. This project clusters the venues and neighborhoods based on number of likes for venues and then highly liked clusters are found in these two boroughs.

Stakeholder's Interest:

This problem can be used by travel agencies when planning packages for tourists, making sure to include most liked food hotspots in a city. Food industry is growing and so its demand. This kind of clustering can be used by companies offering delivery services. They can identify regions in cities with most liked food joints and make sure to offer delivery services in those regions from restaurants.

Data Set Used:

The dataset used to solve the above problem is Neighborhoods dataset from https://www.kaggle.com/shaswatd673/delhi-neighborhood-data. This dataset contains the neighborhoods information for Delhi city.

The dataset contains Neighborhood, Borough, Latitude, Longitude values for Delhi.

	Borough	Neighborh	latitude	longitude
0	North We	Adarsh Na	28.61419	77.07154
1	North We	Ashok Vih	28.69945	77.18483
2	North We	Azadpur	28.70766	77.17555
3	North We	Bawana	28.79966	77.03288
4	North We	Begum Pur		
5	North We	Dhaka	39.03171	-90.2612
6	North We	Jahangirpi	28.72597	77.16266
7	North We	Karala	28.73514	77.03251
8	North We	Keshav Pu	28.68893	77.16168
9	North We	Kingsway	28.61426	77.20155
10	North We	Kohat Enc	28.69804	77.14054
11	North We	Model Tov	39.03171	-90.2612
12	North We	Narela	28.84261	77.09184
13	North We	Pitam Pur	28.70327	77.13225
14	North We	Rani Bagh	28.68598	77.13252
15	North We	Rithala	28.72081	77.10718

Fig 1: Data set snapshot

Methodology:

Once the dataset is loaded, rows with NULL values are removed, a map is created using Folium library in Python. This library is useful for creating interactive maps. In this project, I have folium to create maps for New Delhi, India. The input to folium function is latitude and longitude of the location.
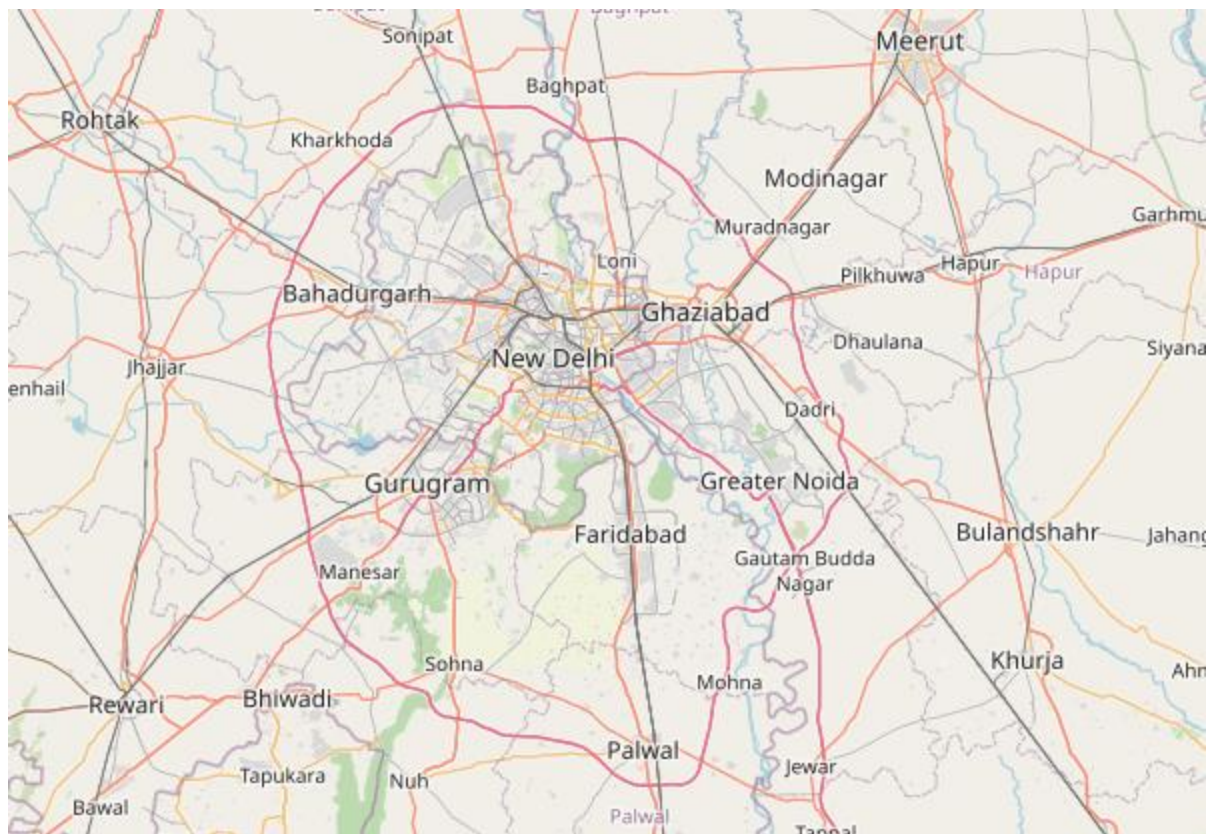
Fig 2: New Delhi Map

Once, New Delhi map is created, Neighborhoods and Borough were plotted on the same map using their respective Latitude and Longitude information. These are plotted as blue dots on the map (Fig 3).
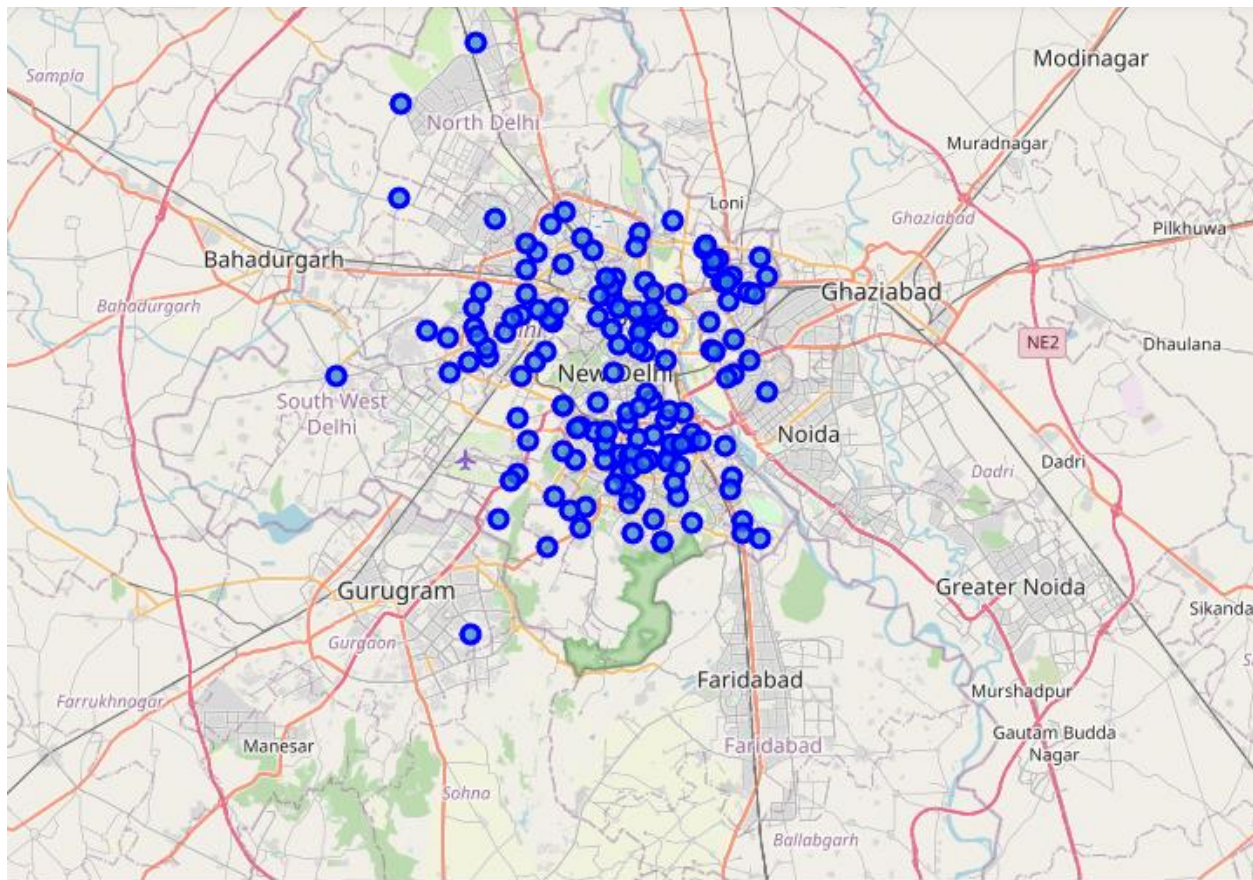
Fig 3: Neighborhoods plotted

Out of these neighborhoods, only New Delhi and South Delhi Borough is considered. These two boroughs in total have around 44 neighborhoods.

Using FourSquare API:

Now that a dataframe with New Delhi and South Delhi is obtained, Foursquare API is used to extract the venues details using explore endpoint.
'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'
Using the above link, information for venues at these locations is obtained. The output of this request is processed in a json format and relevant information of interest is extracted. Venue name, Venue ID, Venue Latitude, Venue Longitude and Venue category is extracted in a new data frame.
The next API being requested is to get likes for these venues. The api is as follows:
'https://api.foursquare.com/v2/venues/{}/likes?&client_id={}&client_secret={}&v={}'. The field of interest is count and we get it for each venue. The venues being considered for the purpose of this project are food venues. All other category venues are removed.
Venues are grouped based on Neighborhood and then data frame is joined with likes dataframe.

K- means clustering:

K- means clustering method is used to perform clustering of places based on the likes. Neighborhoods with similar likes are clustered together. To pursue this, scikit learn's KMeans library is used.

To identify the ideal number of clusters, elbow method is used. Kmeans clusters are obtained for a range of values for k. Here, we take k from 1 to 10. Using the chart, suitable number of cluster is obtained and this is the value of k for our dataset.

Again model is trained using this value of k and clusters are obtained.

Results:

To identify a suitable number of clusters, elbow technique is used. From this method, it can be seen that K = 3 is a suitable value / number of clusters for our dataset.

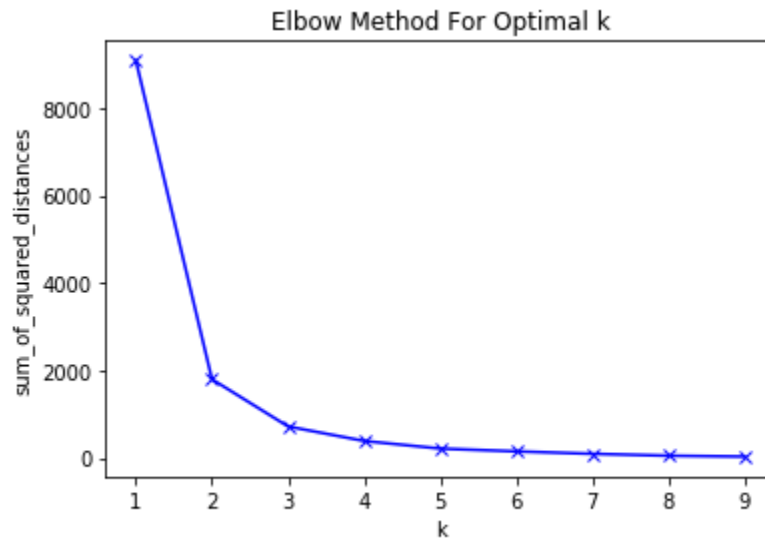
Fig 4: Elbow method to get optimal value for K

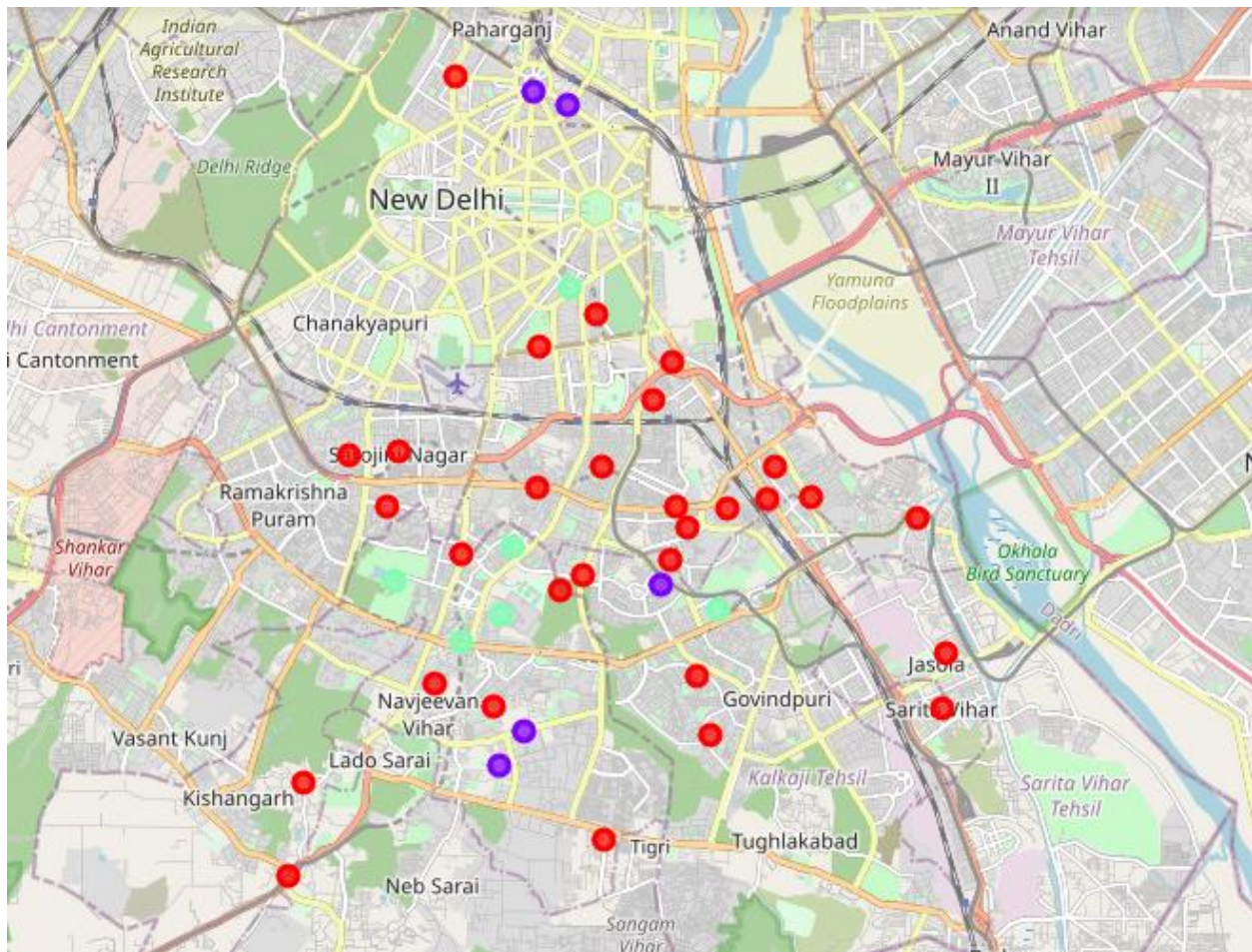Once, KMeans is run with K = 3, we obtain three clusters as follows:

Fig 5: Clusters in New Delhi and South Delhi Borough

Discussion:

From the results above, it can be seen that most of the highly liked clusters are in New Delhi borough. The most liked cluster is cluster 2.


```
# Cluster 2
delhi_grouped.loc[delhi_grouped['Cluster']
```

	Neighborhood	Venue Longitude	Likes Count
1	Barakhamba Road	77.222245	40.000000
4	Connaught Place	77.219801	49.880952
17	Kailash Colony	77.242078	40.400000
20	Khirki Village	77.218093	41.621622
33	Saket	77.216635	63.333333

Fig 6: Cluster 2

Further, this can be extended to get trending venues and a correlation can be drawn between most liked venues and trending venues.

Conclusion:

This project has clustered highly likes neighborhoods in Delhi. The project uses FourSquare API to obtain details about the venues in these neighborhoods. Two API endpoints have been used i.e. explore and likes. Further, with likes available for all food veneus, K means clustering is performed. K means clustering cluster all neighborhoods in three clusters and from the results Cluster 2 has the most liked places in New Delhi.

Therefore, this shows that most liked places or neighborhoods are like Saket, Barakhamba Road, Connaught Place etc.

This can be performed for various cities and towns across various countries.