Capstone Project: Battle of Neighborhoods

Neighborhood Analysis and Segmentation of city of Lucknow

Sharad Agarwal

July 27, 2020

1. Introduction

1.1 Background

Lucknow is one of the largest cities in India and the capital city of Uttar Pradesh, situated in the north eastern part of India. It is an old settlement established at least in the 14th century and developing as a cultural and commercial center of the region. It is considered to be the original home for the Nawab people.

Lucknow is a historically multi-national and multi-cultural city, with plenty of tourist attractions and interesting places to visit. Lucknow is one of the important centers of business and IT technology, with a large number of the related companies and international organizations.

There are over 3 million people living in the city, and Lucknow has a special culture, cuisine, traditions, and lifestyle, which is worth seeing and getting to know.

1.2 Business Problem

Main objective of this project to is to explore the city of Lucknow for venues and neighborhoods. Find out the categories of the venues, find most common venues in each neighborhood. Create the clusters of neighborhoods for segmentation.

Use latitudes and longitudes information wherever needed.

1.3 Interest

Analysis done in this project will help anyone who is planning to move to Lucknow, or looking to explore the city of Lucknow as tourist, places to eat & visit, shopping or even if to start a new business and is looking for options.

2. Data

Basic information about Lucknow is taken from https://www.latlong.net/place/lucknow-uttar-pradesh-india-27621.html

List of neighborhoods in Lucknow is obtained from Wikipedia and Google.

Latitudes and longitudes of the neighborhoods are taken from https://www.gps-latitude-longitude.com/address-to-longitude-latitude-gps-coordinates and written in a csv file.

Foursquare location data is used to explore the city of Lucknow neighborhood, find details about each neighborhood, venues present and the category of each venue. Foursquare provides details about neighborhood using latitude and longitude.

3. Methodology

3.1 Data Acquisition

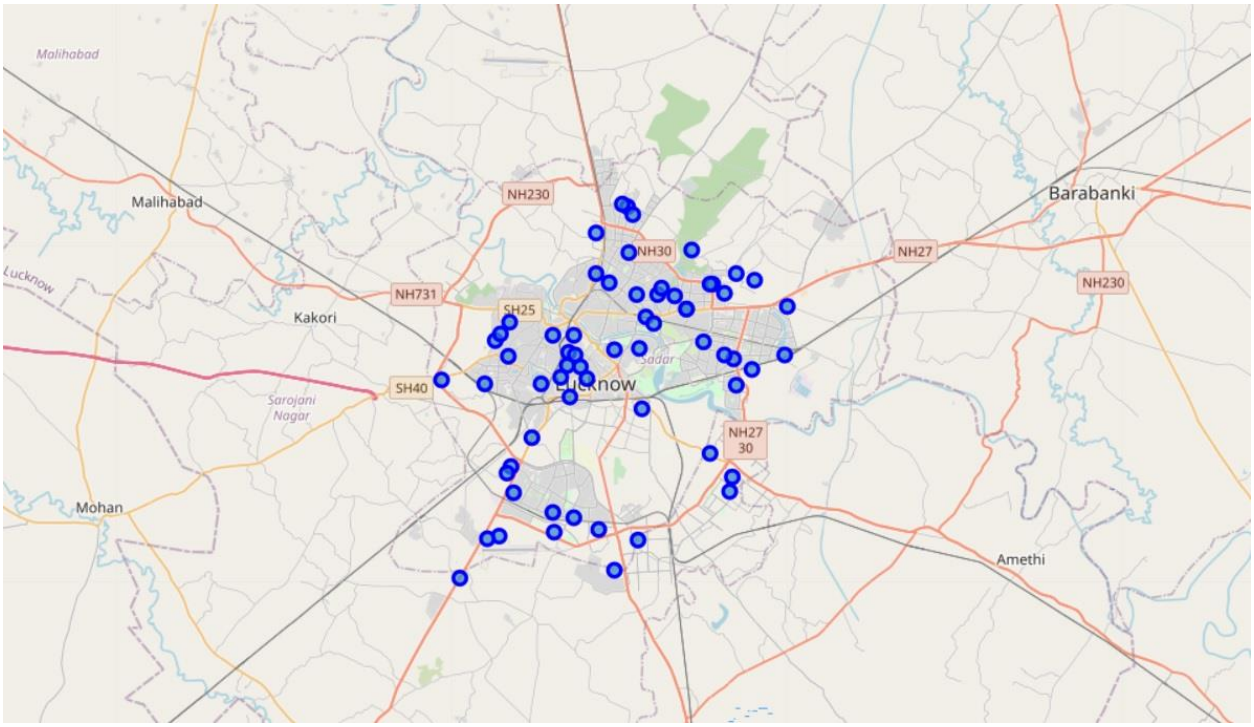
First the geographical coordinates of Lucknow were obtained using geolocator. And after that latitudes and longitudes of each neighborhood were read from a csv file and put into a dataframe.

	Neighborhood	Latitude	Longitude
0	Aishbagh	26.838404	80.906328
1	Alambagh	26.812744	80.901280
2	Aliganj	27.492207	79.168962
3	Aliganj Extension	26.889694	80.934943
4	Amausi	26.767357	80.884090
...	...	...	...
59	Vijay Khand	26.857908	80.991251
60	Vikram Khand	26.851657	81.001956
61	Vimal Nagar	26.886420	81.017961
62	Vrindavan Yojna	26.765091	80.957140
63	Yaseen Ganj	26.861359	80.884561

64 rows × 3 columns

3.2 Visualization of neighborhoods

Neighborhoods were rendered on a map using Folium package



3.3 Explore neighborhoods using Foursquare API

Each neighborhood is explored using Foursquare API. Radius of 5000 and limit of 100 is considered for the exploration.

3.3.1 Data Cleaning

Duplicate venues were dropped that had same latitude and longitude.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aishbagh	26.838404	80.906328	Bara Imam Bara	26.870001	80.913005	History Museum
1	Aishbagh	26.838404	80.906328	Sharma Tea Corner	26.847726	80.940747	Tea Room
2	Aishbagh	26.838404	80.906328	The Mughal's Dastarkhwan	26.844244	80.940310	Indian Restaurant
3	Aishbagh	26.838404	80.906328	Rajajipuram E-Block	26.843061	80.880187	Flea Market
4	Aishbagh	26.838404	80.906328	Dastarkhwan	26.852692	80.936927	Indian Restaurant
...	...	...	...	...	...	...	...
124	Tedhi Pulia	26.917039	80.954327	Ice N Spice	26.877524	80.970634	Vegetarian / Vegan Restaurant
125	United City Colony	26.922149	80.948533	Pizza Hut	26.896358	80.941474	Pizza Place
126	United City Colony	26.922149	80.948533	Swad	26.901243	80.946290	Breakfast Spot
127	United City Colony	26.922149	80.948533	Aryan Family's Delight	26.881949	80.946905	Restaurant
128	Vimal Nagar	26.886420	81.017961	Lohia Park	26.858393	80.983198	Park

129 rows × 7 columns

3.4 Analysis of neighborhoods

Neighborhood venue and its categories data was transformed into a dataframe such that venue categories were used as columns of a dataframe. And then data was grouped for each neighborhood with each venue category column representing the mean frequency of that category.

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

```
#lucknow_grouped = lucknow_onehot.groupby('1_Neighborhood').sum().reset_index()
lucknow_grouped = lucknow_onehot.groupby('1_Neighborhood').mean().reset_index()
lucknow_grouped.head()
```

	1_Neighborhood	ATM	Airport	Airport Service	American Restaurant	Art Gallery	Asian Restaurant	BBQ Joint	Bakery	Breakfast Spot	...	Pizza Place	Platform	Plaza	Restaurant	Sandwich Place
0	Aishbagh	0.0	0.0	0.0	0.022727	0.0	0.0	0.000000	0.022727	0.000	...	0.045455	0.0	0.0	0.000	0.022727
1	Aliganj	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.000	...	0.000000	0.0	0.0	0.000	0.000000
2	Aliganj Extension	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.250000	0.125	...	0.250000	0.0	0.0	0.125	0.000000
3	Amausi	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.000	...	0.000000	0.0	0.0	1.000	0.000000
4	Aminabad	0.0	0.0	0.0	0.000000	0.0	0.0	0.111111	0.000000	0.000	...	0.000000	0.0	0.0	0.000	0.000000

5 rows × 48 columns

3.5 Segmentation

Neighborhoods were segmented using K-Means clustering technique. Data was divided in five clusters.

4. Results

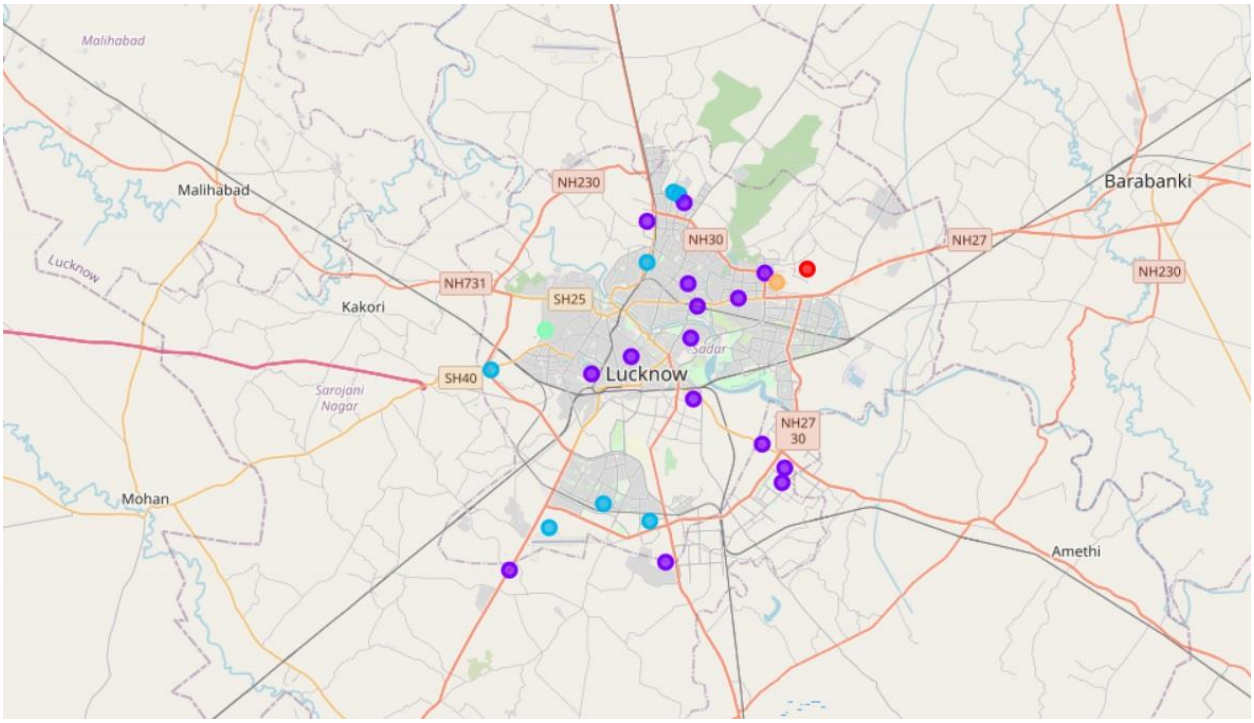
There were 47 unique venue categories across 26 unique neighborhoods

```
There are 47 uniques venue categories.
array(['History Museum', 'Tea Room', 'Indian Restaurant', 'Flea Market',
       'Ice Cream Shop', 'Café', 'Neighborhood', 'Shopping Mall',
       'Fast Food Restaurant', 'Hotel', 'Hookah Bar', 'Bakery', 'Market',
       'Pizza Place', 'Coffee Shop', 'Sandwich Place', 'Train Station',
       'Multiplex', 'American Restaurant', 'Food Court', 'Shop & Service',
       'Clothing Store', 'Restaurant', 'Breakfast Spot', 'Nightclub',
       'BBQ Joint', 'Chinese Restaurant', 'Park', 'Electronics Store',
       'Department Store', 'Plaza', 'Lounge', 'Fried Chicken Joint',
       'Indian Sweet Shop', 'Grocery Store', 'ATM', 'Platform',
       'Building', 'Bus Station', 'Convenience Store', 'Garden',
       'Food Truck', 'Airport', 'Airport Service', 'Asian Restaurant',
       'Art Gallery', 'Vegetarian / Vegan Restaurant'], dtype=object)
```

Top 10 common venues for first 5 neighborhoods

	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	Aishbagh	Indian Restaurant	Fast Food Restaurant	Hotel	Shopping Mall	Café	Train Station	Hookah Bar	Ice Cream Shop	Pizza Place	Multiplex
1	Aliganj	Clothing Store	Vegetarian / Vegan Restaurant	Café	Food Court	Flea Market	Fast Food Restaurant	Electronics Store	Department Store	Convenience Store	Coffee Shop
2	Aliganj Extension	Pizza Place	Bakery	Restaurant	Coffee Shop	Breakfast Spot	Market	Vegetarian / Vegan Restaurant	Café	Flea Market	Fast Food Restaurant
3	Amausi	Restaurant	Vegetarian / Vegan Restaurant	Café	Food Court	Flea Market	Fast Food Restaurant	Electronics Store	Department Store	Convenience Store	Coffee Shop
4	Aminabad	Multiplex	Hotel	BBQ Joint	Park	Nightclub	Indian Restaurant	Chinese Restaurant	Café	Fast Food Restaurant	Electronics Store

Neighborhood segmentation on a map is shown below. Each color represents a cluster.



5. Discussion

1. Though there were 64 neighborhoods in the beginning, number of neighborhoods were reduced to 26 after dropping duplicate venues based on latitude and longitude.
2. Food related venues and Flea Market were among top venue categories among most of the neighborhoods
3. Out of total five clusters, three clusters had only one neighborhood and other two clusters had 16 and 7 neighborhoods

6. Conclusion

Neighborhood analysis of the city of Lucknow done in this project gives us an insight about the city. Results obtained give us the details about its various neighborhoods, venues in each, and top 10 most common venues.

Decision about visiting the city, or a particular neighborhood, for business, tourist, or living purpose can be taken using these results.