The Battle of Neighborhoods

Vishal Jotwani

14th December 2020

Table of Contents

1. Introduction

Kanpur and Lucknow are the two most important cities of Uttar Pradesh. Lucknow is the capital city of the Indian state of Uttar Pradesh, whereas is also the second largest city proper and the largest urban agglomeration in Uttar Pradesh. The two cities being in Uttar Pradesh have various similarities and dissimilarities. This study focuses on finding the similarities within neighborhoods of each city. It also tries to understand the tradition and culture of each neighborhood.

Based on the venues and culture for which neighborhoods are famous for, this project helps business persons can set up their ventures like restaurants, hotels etc.

2. Data Acquisition

The data of neighborhoods is collected through https://data.gov.in.

All India Pin code Directory contains all the pin-code list across India with geocodes and other relevant information like Office Name, Office Type, Delivery Status, Division, Region, Circle, Taluk, Districts, States, Telephones, Related Sub office and Related Head office etc. Postal Index Number (PIN) or PIN Code is a 6 digit code of Post Office numbering used by India Post.

From the above dataset, appropriate rows with District name Kanpur Nagar and Lucknow will be fetched for the purpose of analysis. The attributes like office name, pin code, District name, will be fetched.

3. Methodology

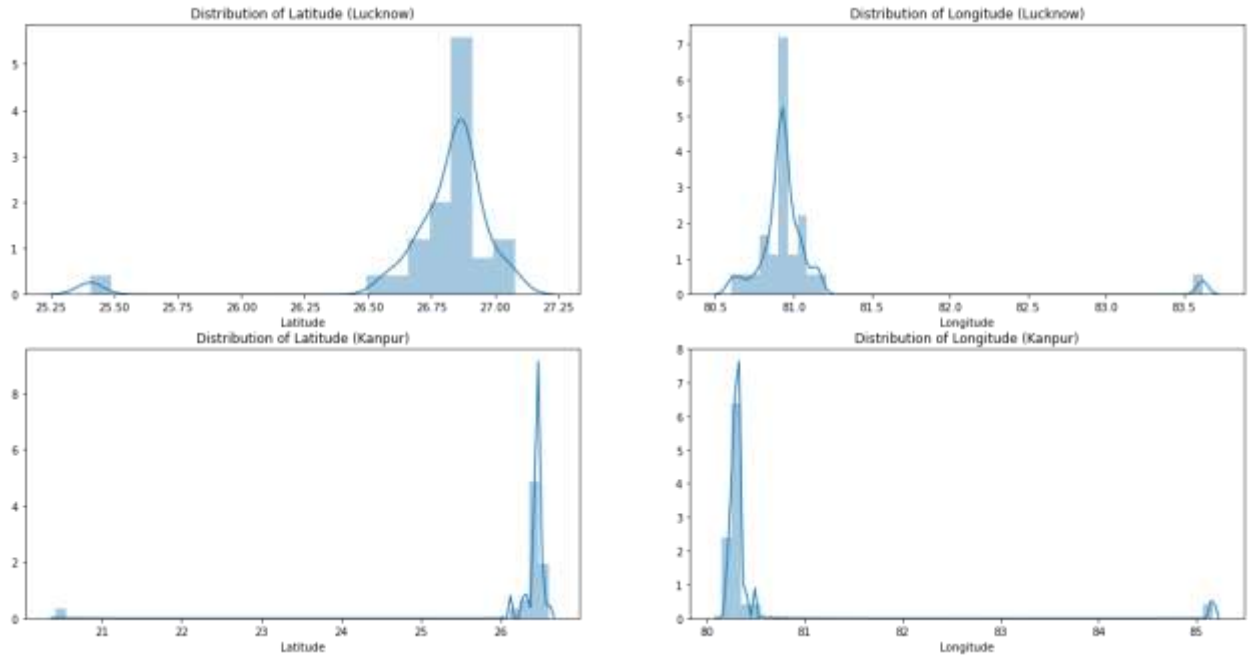
3.1 Data Preprocessing

Data Preprocessing is the process of converting raw data into the form which is suitable for further analysis and machine learning modeling. Appropriate columns like officename, pincode, districtname are fetched from the initial dataset. The rows with districtname Kanpur Nagar and Lucknow are drawn from the dataset. The new dataset has 469 rows and 3 columns.

The new names are given to the columns officename is converted into Neighborhood, pincode is given Pincode and districtname is given City. The column Neighborhood contains additional information attached with it in the form of S.O., B.O., H.O. and G.P.O which is removed. There are no null values in the dataset. The dataset is grouped on the column Pincode.

New columns Latitude and Longitude are added to the dataset. The dataset column Latitude and Longitude contains outliers as shown in figure 1 which are removed. Foursquare API is used to collect location data for each neighborhood of both cities. The dataset is prepared for the analysis.

Figure 1



3.2 Modeling

K-Means Clustering machine learning technique is used for finding similarities in neighborhoods of each city. K-means clustering is a method of **vector quantization**, originally from signal processing that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.

4. Result

The neighborhoods of Lucknow obtained through K-Means Clustering technique are as follows-

Table 1

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
25	A N L Colony , Aishbagh , Arya Nagar , Blunt S...	0	Train Station	Shopping Mall	Market	Food Court

Table 2

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
26	A P Sabha , Canal Colony , CPMG Campus , Darul...	1	Indian Restaurant	Fast Food Restaurant	Bakery	Hookah Bar
28	Adarsh Nagar , Ain , Alambagh , Anandnagar , B...	1	Food Court	Fast Food Restaurant	Pizza Place	Bus Station
31	Alamnagar , Avas Vikas Colony , R G C Mill , R...	1	Flea Market	Asian Restaurant	Platform	Pharmacy
38	Aminabad Park , Balrampur Hospital , Ganesh Ga...	1	Indian Restaurant	Hotel	Café	Fast Food Restaurant
39	Amraigaon , Anaura Kalan , Chinhat , Golla , J...	1	Indian Restaurant	Fast Food Restaurant	Clothing Store	Train Station
50	Gomtinagar , Gomtinagar Vistar , Laulai , Male...	1	Fast Food Restaurant	Café	Multiplex	Shopping Mall
53	Nadwa , New Hyderabad , R P Line	1	Indian Restaurant	Fast Food Restaurant	Hookah Bar	Ice Cream Shop

Table 3

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
32	Aliganj Extension , Aliganj , Chandganj	2	Dessert Shop	Bank	Breakfast Spot	Pizza Place
46	D M Road , Daliganj , Diguria , Ghaila , Lonik...	2	History Museum	Fast Food Restaurant	Bakery	Coffee Shop
51	H E School , M N Colony , Mahanagar , Sant Mar...	2	Bakery	Gym	Restaurant	Plaza

Table 4

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
35	Amberganj , Chaupatiyan , Hasnain Market , Hus...	3	ATM	Pizza Place	Fast Food Restaurant	Indian Restaurant
44	Batha Sabauli , Jankipuram , Madiyaon	3	ATM	Bank	Business Service	Chinese Restaurant
52	Kalyanpur , Vikas Nagar	3	ATM	IT Services	Indian Restaurant	Building

Table 5

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
40	Antagarhi , Atari , Badiyan , Bazargaon , Bee...	4	Hotel	Train Station	Fast Food Restaurant	Chinese Restaurant

The neighborhoods of Kanpur obtained through K-Means Clustering technique are as follows-

Table 6

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
7	Barra , Barra Vishwa Bank Sector-A	0	ATM	Fast Food Restaurant	Business Service	Paper / Office Supplies Store
11	Binour , Bisayakpur , Fatehpur Roshnai , Jugra...	0	ATM	Pizza Place	Lighting Store	Market

Table 7

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
13	COD , Dmsrde , Shyam Nagar	1	Fast Food Restaurant	Shopping Mall	Park	Coffee Shop
17	Juhi Colony , Nirala Nagar	1	Fast Food Restaurant	Shopping Mall	Park	Coffee Shop
18	Kasigaon , Kathongar , N H Road , Naubasta , P...	1	Fast Food Restaurant	IT Services	Dessert Shop	Train Station
19	Munshipurwa , T P Nagar	1	Fast Food Restaurant	Shopping Mall	Park	Coffee Shop

Table 8

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Agriculture College , Arya Nagar , Employeemen...	2	Shopping Mall	Bakery	Hotel	Multiplex
2	Anand Nagar , Kidwai Nagar , RBI Colony , Yash...	2	Multiplex	Shopping Mall	Pizza Place	Café
5	B N Shukla & Sons , Birhana Raod , Colonelganj...	2	Fast Food Restaurant	Hotel	Multiplex	Furniture / Home Store
12	Cambridge Road , Gandhi Grame , Hal Colony , H...	2	ATM	Airport	Shopping Mall	Pharmacy
14	D O Oil Mill , Jawahar Nagar , Kaushalpuri , R...	2	Fast Food Restaurant	Bakery	Gym	Coffee Shop
15	Govind Nagar , Govind Nagar West	2	Multiplex	Shopping Mall	Indian Restaurant	Coffee Shop
16	Iit	2	Performing Arts Venue	ATM	Hostel	Hotel

Table 9

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
4	Avas Vikas Yojna No. 3 , Bairl , Kalyanpur , K...	3	Food Court	Clothing Store	Snack Place	Hostel

Table 10

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
3	Anwarganj , Chauki Jarib	4	Train Station	Shopping Mall	Bus Station	Coffee Shop

5. Discussion

In each of the neighborhood the most common venues include hotels, restaurants and cafes indicating that these are the most visited places in those neighborhoods.

6. Conclusion

In this study, similar neighborhoods of two largest cities of Uttar Pradesh, Kanpur and Lucknow are found. The dataset is collected through portal of Government of India. Location data is collected with the help of Foursquare API. For each neighborhood four most common venues are found which can help the business person to set up their business in that neighborhood by providing an idea for which that neighborhood is famous.