Applied Data Science Capstone

IBM

Capstone Project - The Battle of neighbourhoods (Week 1)

Finding the best areas in Kanpur, India to open a Shopping Mall

Introduction

Mall industry has emerged as one of the growing sector in India and also, visiting shopping malls has become a great way to relax and enjoy for the visitors during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. Currently there are not many shopping malls in Kanpur but opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

The objective of this capstone project is to analyse and select the best locations in the city of Kanpur, India to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kanpur, India, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Data

→To solve the problem, we will need the following data:

• List of neighbourhoods in Kanpur. This defines the scope of this project which is confined to the city of Kanpur, the eighth-most populous metropolitan area in India

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

→Sources of data and methods to extract them:

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Kanpur) contains a list of neighbourhoods in Kanpur, with a total of 21 neighbourhoods.

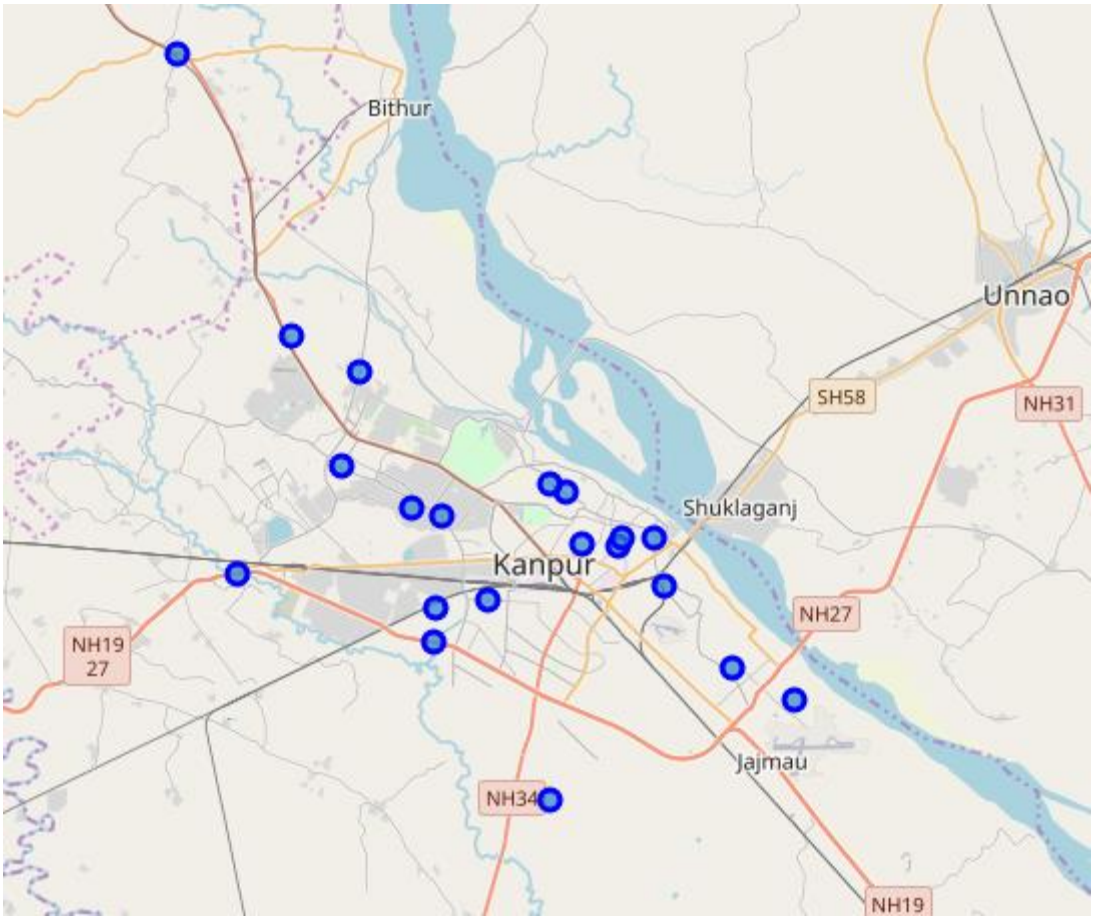
- We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages.
- Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.
- After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000

developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward.

This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

Methodology

Possible locations for the shopping mall in the neighbourhoods in Kanpur:



Using FourSquare API, it was found that 41 unique venue categories in Kanpur, with 'Shopping Mall' being one of them:

```
array(['Breakfast Spot', 'Business Service', 'ATM', 'Diner', 'Bar', 'Spa',
       'Bowling Alley', 'Shopping Mall', 'Bookstore', 'Hotel',
       'Indian Restaurant', 'Golf Course', 'Convenience Store',
       'Food Court', 'Cricket Ground', 'Dessert Shop', 'Café',
       'Fast Food Restaurant', 'Clothing Store', 'Pizza Place',
       'Multiplex', 'Coffee Shop', 'Market', 'Bus Station',
       'Jewelry Store', 'Train Station', 'Airport', 'Snack Place',
       'Bakery', 'Pool', 'Garden Center', 'Cosmetics Shop', "Men's Store",
       'Chinese Restaurant', 'Camera Store', 'Furniture / Home Store',
       'Accessories Store', 'Park', 'Tea Room', 'Gym', 'Ice Cream Shop'],
      dtype=object)
```

Data was prepared using OneHotEncoding to explore each neighbourhood across all the venue categories

K-Means was applied to get different clusters based on frequency of shopping mall in the neighbourhood:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Barra, Kanpur	0.0000	0	26.436512	80.284110
1	Birhana Road, Kanpur	0.1250	2	26.428658	80.386283
2	Bithoor	0.0000	0	26.519671	80.258794
3	Chaman Ganj, Kanpur	0.0625	2	26.467179	80.334996
4	Govind Nagar	0.0000	0	26.449520	80.302740

Results

Cluster -1:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Barra, Kanpur	0.0	0	26.436512	80.284110
14	Rawatpur	0.0	0	26.477770	80.276340
13	Ratan Lal Nagar	0.0	0	26.447290	80.284820
12	Padri Lalpur	0.0	0	26.388380	80.323780
7	Kalyanpur, Uttar Pradesh	0.0	0	26.490680	80.252150
8	Kanpur Cantonment	0.0	0	26.453790	80.362830
4	Govind Nagar	0.0	0	26.449520	80.302740
2	Bithoor	0.0	0	26.519671	80.258794

Cluster - 2:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
6	Kakadeo	0.25	1	26.47543	80.28700
5	Jajmau	0.25	1	26.41890	80.40744

Cluster – 3:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
15	The Mall, Kanpur	0.050000	2	26.468932	80.359776
9	Latouche Road, Kanpur	0.047619	2	26.466097	80.347061
10	McRobertganj	0.052632	2	26.482900	80.329010
11	Meston Road, Kanpur	0.050000	2	26.468861	80.348330
3	Chaman Ganj, Kanpur	0.062500	2	26.467179	80.334996
1	Birhana Road, Kanpur	0.125000	2	26.428658	80.386283
16	VIP Road, Kanpur	0.083333	2	26.485630	80.323988

Discussion

As per the clusters observed after K-Means, cluster 1 is the best candidate to open a shopping mall.

Since, the data is only for 21 neighbours, it will be a good test to see how the model performs when trained with data of more neighbours. This can be seen has future enhancement in the current project.

Conclusion

Most of the shopping malls in Kanpur City are concentrated in cluster 2 with moderate in cluster 2 and none in cluster 1, hence for the new builders, it will be a good opportunity to open shopping mall in cluster 1 since there will be no competition which lead to benefit for both the vendors and for the people living in the nearby areas.