Segmenting and Clustering Neighborhoods in Varanasi for modernization

Reetu Singh

Introduction

- In this project, we are going to explore Varanasi city, the cultural hub of India. In particular, we are going to explore various venues categories from foursquare, specifically ATMs in the area. We would look at the ATM distribution and come up with the possible places as a suggestion to install/set up new ATM machines. Varanasi is a place represented by our Hon PM of India and it has also been selected to develop Varanasi as a Smart Heritage City. Inspired by Hon Prime Minister Narendra Modi's vision, the VAKYO initiative aims to connect the heritage cultural cities of Varanasi and Kyoto through the India Japan Global Partnership. (http://www.vakyo.org/)

Data

- To work on the same, we have downloaded pin code data from the govt website for all over India (without Latitude/Longitude) https://data.gov.in . We cleaned and extracted data for Varanasi alone. Using geocoder, Latitude and longitude of the pin code is extracted. Every pin code has more than one post office under it. And address of each Post office is included as well, however, it is not possible to retrieve geocoordinates for these offices using geocode. Hence, we are sticking to geocoordinate by pin code data.

Methodology

- To work on the same we have taken pin-code data from the govt website for all over India. We have cleaned and extracted data for Varanasi alone. Using gecoder, we extracted Latitude and longitude of the pin-code. Geocoder doesn't return geocoordinates for all the pin-code, while it returned wrong geocoordinate eg. China location. Hence, we had to clean the data further, dropping the empty location and selecting only those locations that fall within India.
- Please note that we are using pin-code as a neighborhood as we couldn't get geo coordinates for the places/post offices, which was the ideal case. The city of Varanasi encompasses a total area of **1550.3** sq. km and area under Pin-code are of a size up to 48 sq. km. There are 30 pin-codes retrieved for Varanasi with Lat/Long data, and 303 post offices falling under the pin-codes (without Lat/Long info).
- For the 30 pin-code, we would use foursquare credentials to fetch venues and explore the neighborhoods. Since area under pin-codes are large, we are going to use radius of up to 4km and Limit up to 200.
- We would further analyze the venue category, and learn about various current ATMs location. We would also find the pin-code and the area where there are no ATMs. The recommendation would be made in terms of list of the pin-code along with areas which are potential candidate for ATMs machine setup.

Result & discussion: 1/5

There are around 30 pin-code found with geocoordinates for Varanasi. When we overlaid the pin-code on the map using Folium, this is what we get:

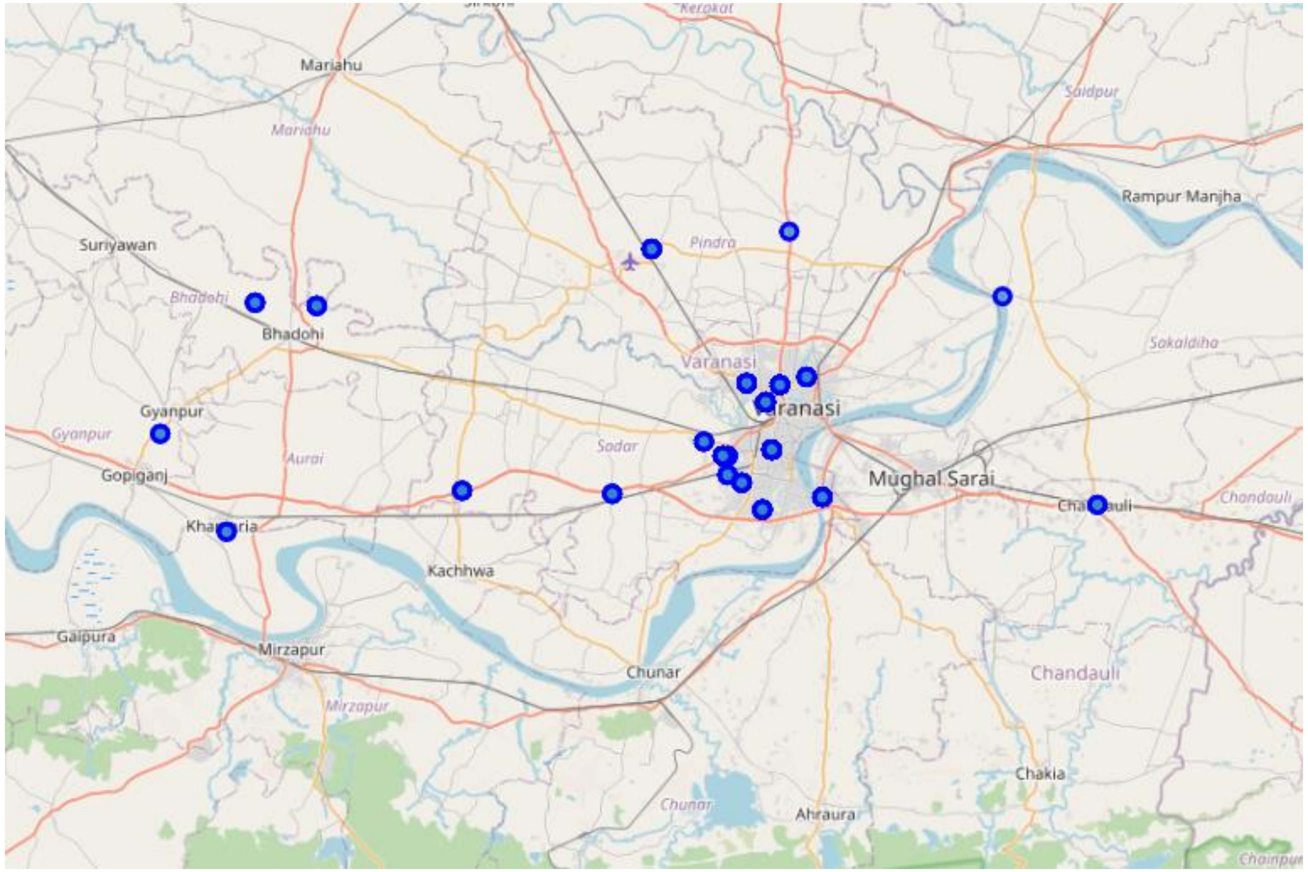

Figure 0 Varanasi Map with pincode (30)

Result & discussion: 2/5

There are around 171 venues fetches which we tried to capture top 200 venues by pin-code in 4000 meters, since areas under pin-code are bigger in size and our venue are coming out much smaller in numbers.

When we overlaid the venues on the map using Folium, this is what we get:

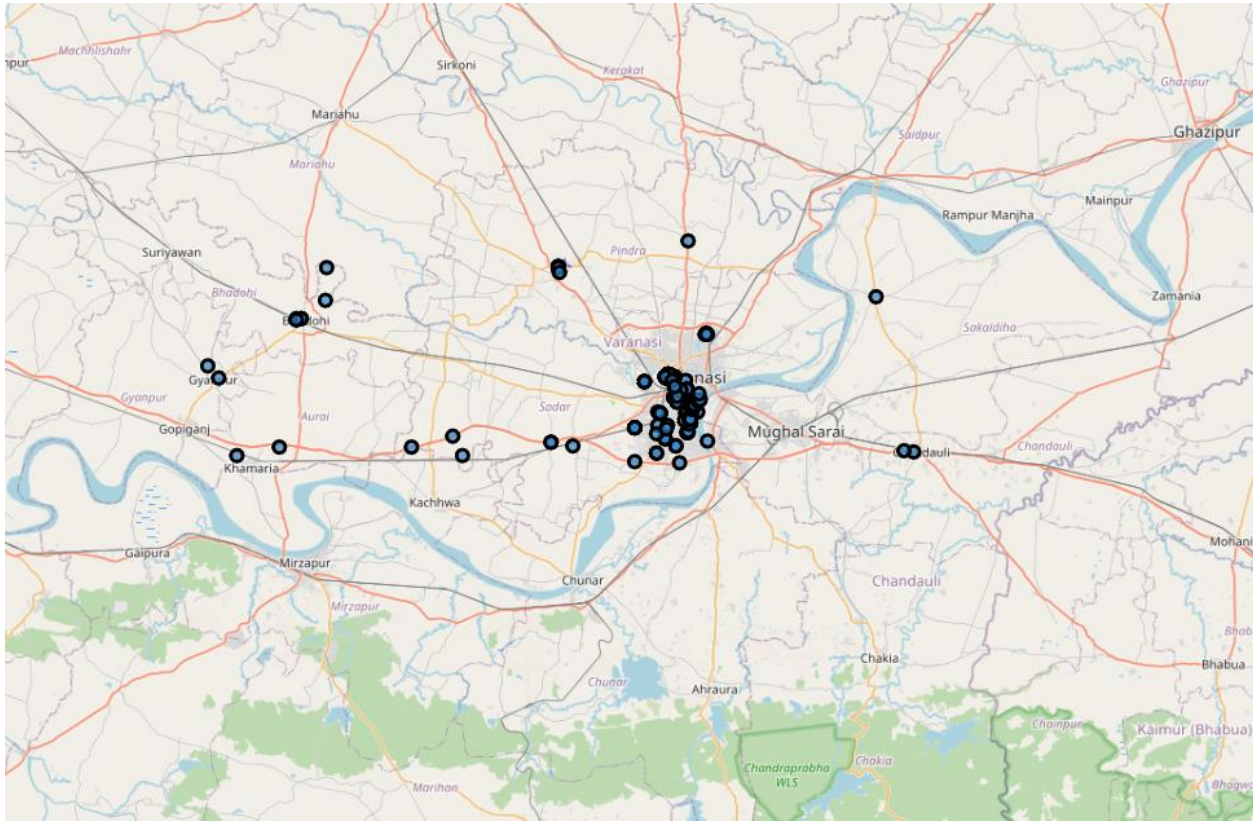

Figure 1 Varanasi Map with venues (171)

Result & discussion: 3/5

We analyze the venues by venue category that gives 46 unique categories. List here shows categories sorted in descending order:

In top 5, the uniques categories are :

23	Hotel	48
0	ATM	15
9	Café	9
26	Indian Restaurant	9
32	Pizza Place	9

Figure 2 Top 5 categories of venue item

Now, analyzing the top neighborhood/pin-code, with highest venues, we get this.

	Venue Category	Venue
Neighborhood		
221010	20	20
221008	8	8
221011	8	8
221002	8	8
221001	7	7
221302	7	7

Figure 3 Top 6 Pin code/Neighborhood with highest number of venues

Result & discussion: 4/5

Next, we explore how many neighbor/pin-code has "ATM" in the vicinity. We see that 11 pin-code has ATMs within its periphery. There are 4 pin-code that has 2 ATMs each.

```
city_venues[city_venues['Venue Category']=='ATM'].groupby(['Neighborhood','Venue Category']).count().reset_index()
```

	Neighborhood	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	221101	ATM	1	1	1	1	1
1	221104	ATM	1	1	1	1	1
2	221106	ATM	2	2	2	2	2
3	221107	ATM	1	1	1	1	1
4	221108	ATM	2	2	2	2	2
5	221302	ATM	1	1	1	1	1
6	221304	ATM	2	2	2	2	2
7	221307	ATM	1	1	1	1	1
8	221311	ATM	1	1	1	1	1
9	221402	ATM	1	1	1	1	1
10	232104	ATM	2	2	2	2	2

Figure 4 Pin-code/neighborhood with ATM in vicinity

Now let's again visualize the Varanasi map and display the neighborhood overlaid with 'ATM' venue.

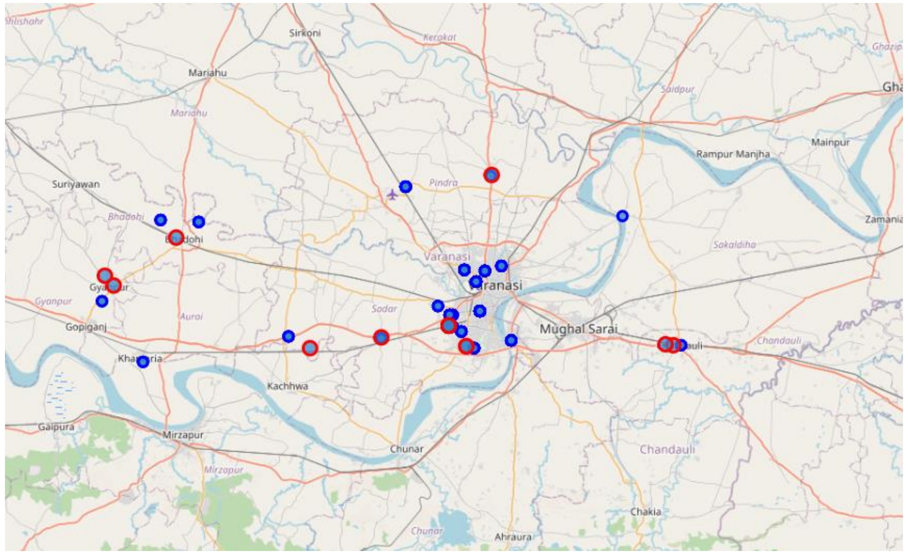

Figure 5 Varanasi map with pin-code(circled in blue) and ATMs location(circled in red)

Result & discussion: 5/5

In the end, we see all the three Varanasi pin-code (blue circle), Venues within pin- codes (black circled) and ATM location (circled with red)¶.

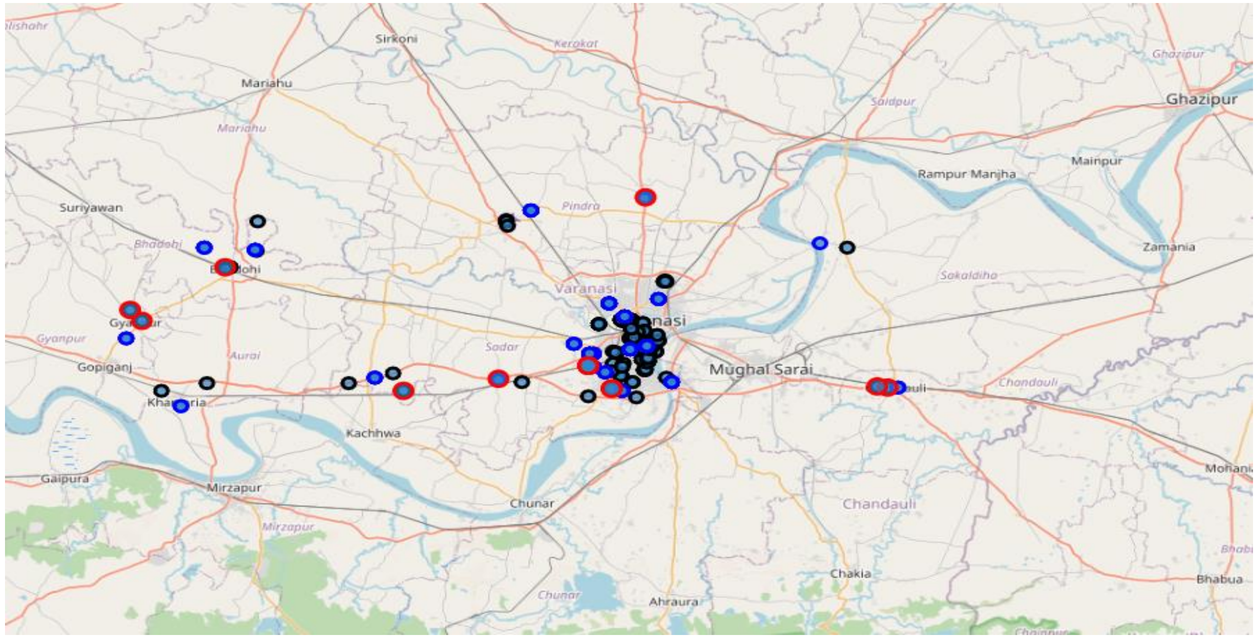

Figure 6 Varanasi map with three markings: pin-code (blue circle), Venues within pin- codes (black circled) and ATM location (circled with red)¶.

Finally, we discovered the areas within pin-codes that have no ATMs. We can apply further clustering method, given that number of venues per pin-codes are very small, visual inspection is enough to conclude.

Find the areas within pincodes that have no ATMs.

```
df_noatm =pd.merge(city_venuesATM, city_venues, how='left', on='Neighborhood')['Neighborhood'].unique()
print(df_noatm)
```

```
[221101 221104 232104 221311 221402 221106 221307 221108 221302 221107
 221304]
```

Conclusion

Conclusion: As we can see, even after increasing the radius and limit of search, we could find only 197 venues from 45 distinct Venue categories. Our top 5 highest categories are from 'Hotel':46, 'ATM': 15, 'Indian Restaurant': 14, 'Café':12, 'Pizza Place':11. The venues are sparse and distributed more around Ganga river. There are 11 pin code under which 75 areas which should be considered to set up ATMs.¶