Classification of Kathmandu and Lalitpur Wards

Saurav Bhattarai

18/01/2021

Contents

Introduction

Kathmandu and Lalitpur are two biggest cities of Nepal. Although being two different cities they are almost a combined city for residents of these two metropolitan cities as they are only divided by river Bagmati and daily thousands of people cross these two cities for as their normal routine.

For this project we want to classify different wards of these two cities on the basis of venues to study both city as one and look out the places which mirrors each other or acts as one unit. We are also taking look at the population density of these wards for detail information. This project would be really insightful for inhabitants who want to explore different wards of both cities on basis of their desire venues and population.

Data Collection

We need to collect data regarding different wards, their locations, population density, area for Kathmandu and Lalitpur city

As we are collecting data for two cities Kathmandu and Lalitpur, we will at first individually collect data for each city and later merge them as a single dataset for two cities. Data for individual city are scrapped and collected from data provided from respective metropolitan city websites. As most data about wards areas, population and location were in Nepali language we have to translate it to English using Google translate saved the site and then scrap the data importing the saved site.

Later on we have to find latitude and longitude of wards by using location and geocoder library.

Venues details can be obtained from four square API for all wards.

My data sources are:

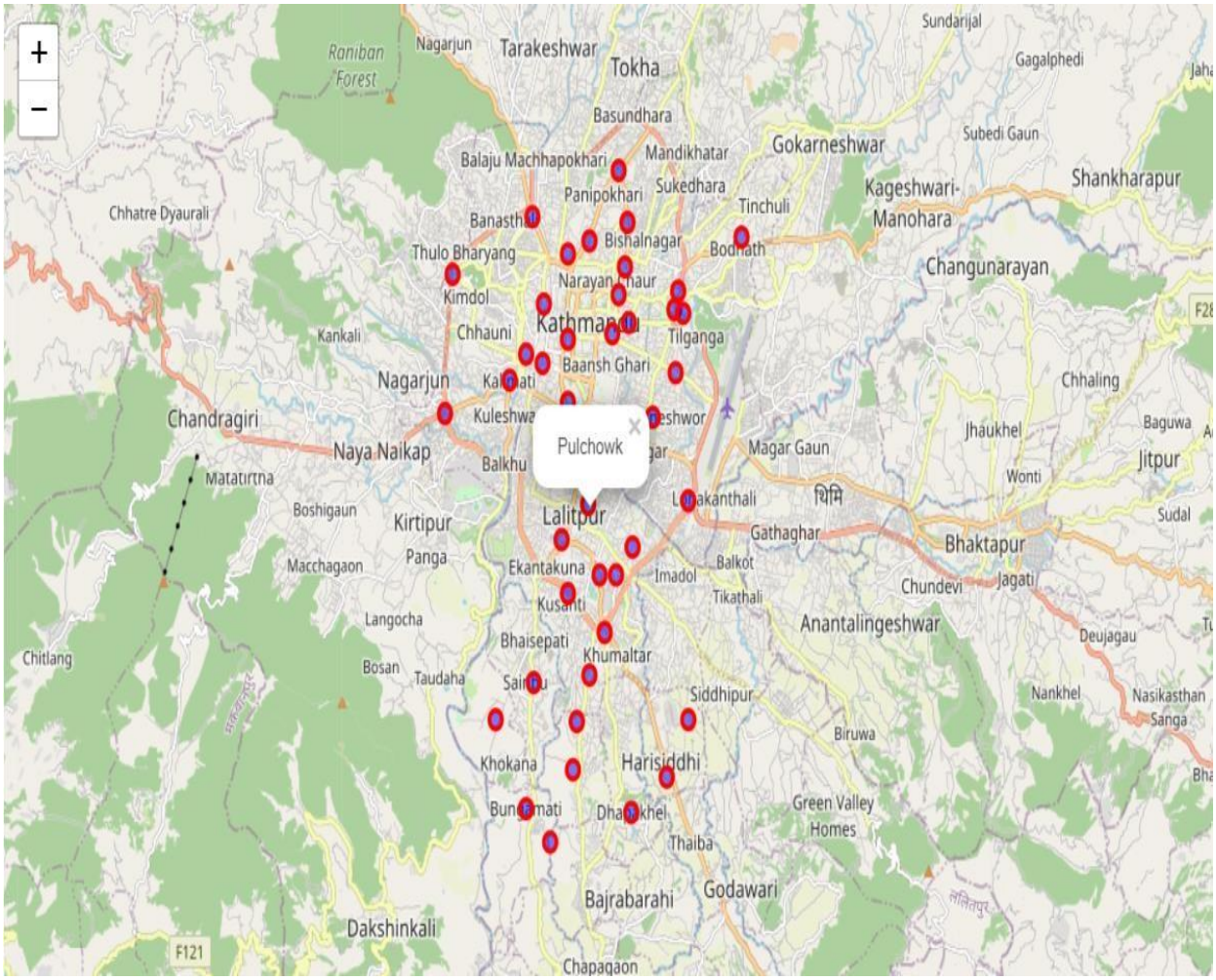
1. Kathmandu wards and their offices location from [1] and respective wards population was collected from [2]. The wards area imported from a paper [3].
2. Similarly Lalitpur wards and their offices location was collected from [4] and respective wards population and area was collected from [5].
3. The longitude and latitudes of wards on basis of location was searched using geocoder library and missing values were manually placed using Google map.[6].
4. The data regarding venues list, categories, details and numbers was collected from four square API [7].

Methodology

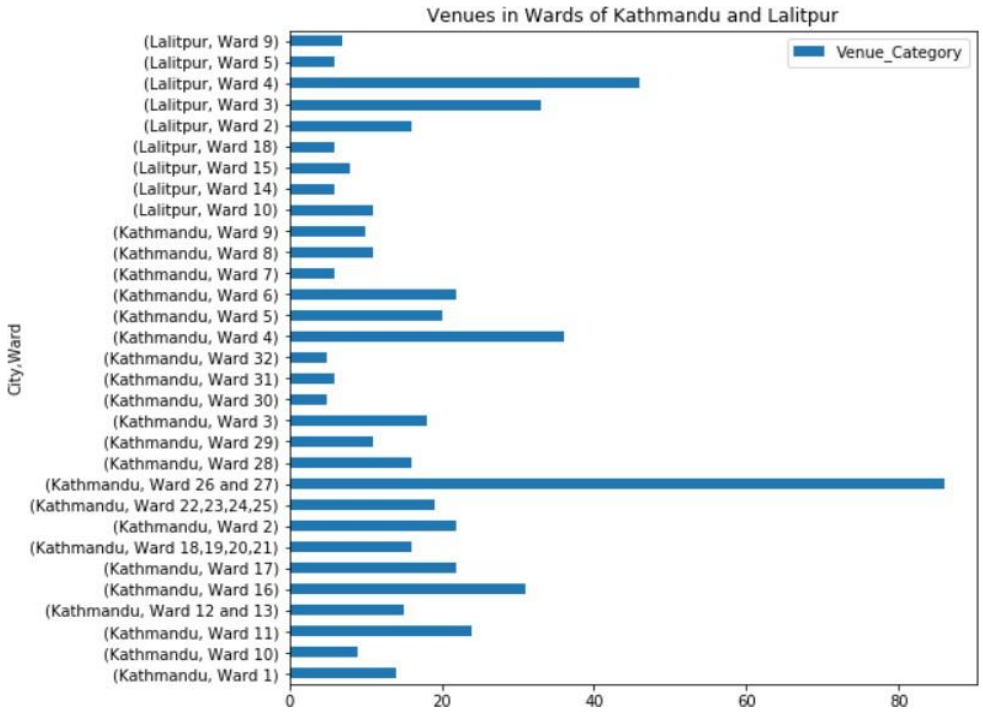
Data Preparation

We have only selected wards with area greater than 0.5 sq. km and merge the wards with area less than 0.5 sq. km sq to reduce the number of number of smaller wards as many venues will overlap if the wards are small which could affect our study later on. The data of both Kathmandu and Lalitpur was merge to get a final data set for both cities.

A folium map was created using this dataset to study the location of wards in both cities in our dataset.



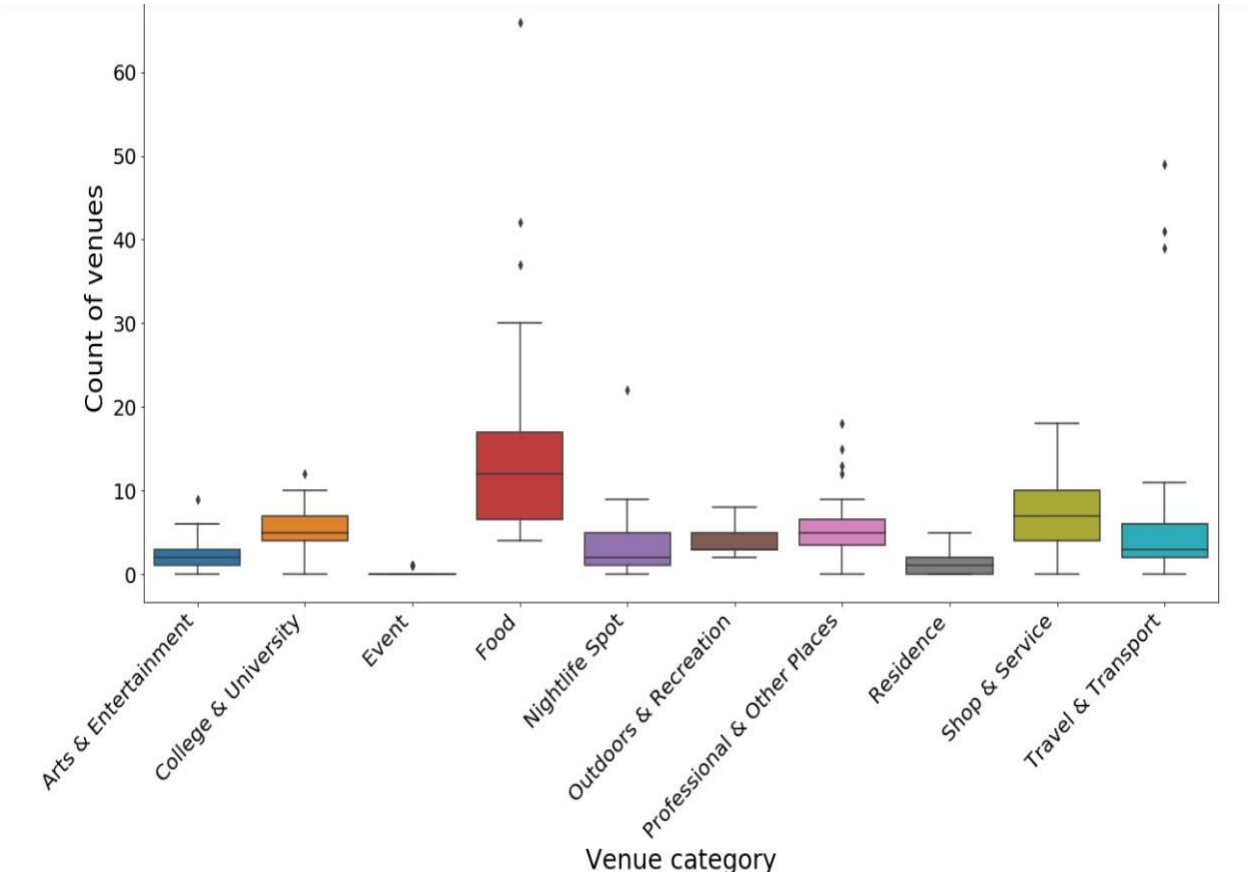
Venues for each wards were collected using four square API and the radius was selected as the radius obtained from the area of respective ward. For better analysis and data study only those wards were selected whose number of venues were at least 5.



As there were about 94 different venues categories we decided to use 10 main categorical venues provided by four square API rather than sub categories venues . The 10 categorical venues and their respective IDs were obtained from four square API .

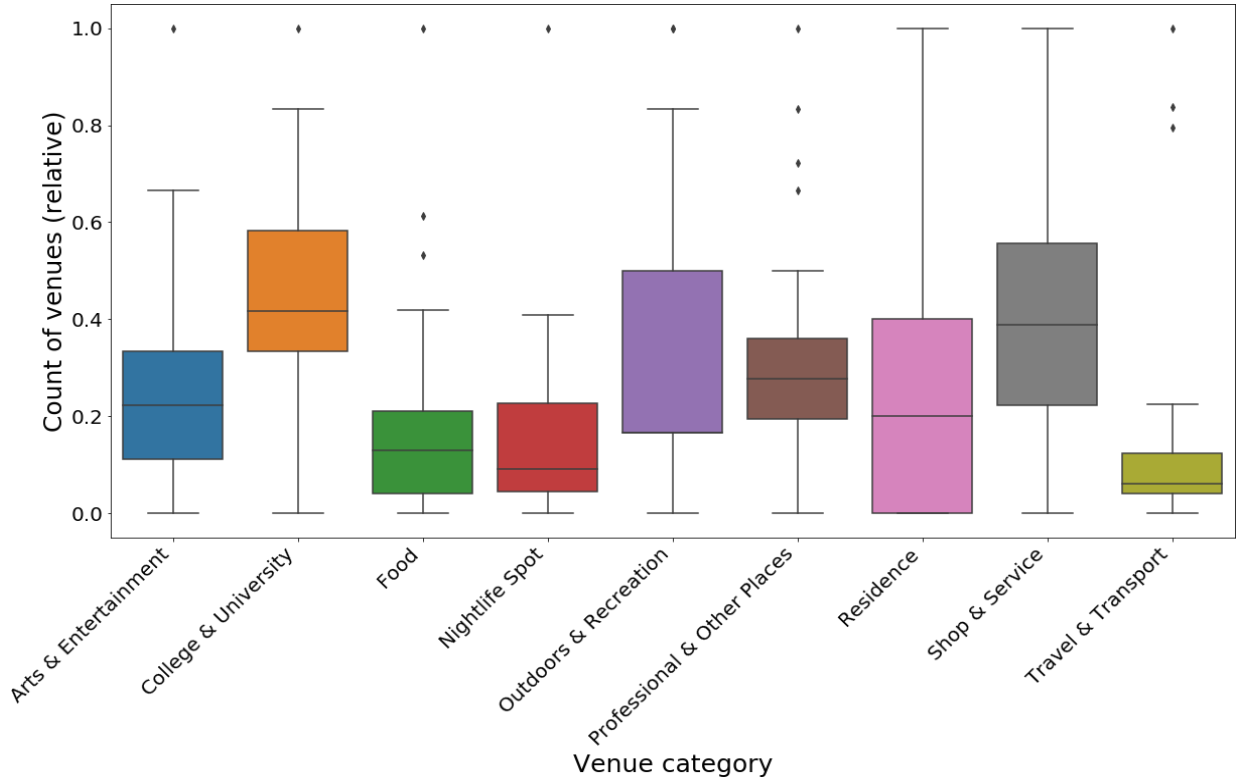
	name	Id
0	Arts & Entertainment	4d4b7104d754a06370d81259
1	College & University	4d4b7105d754a06372d81259
2	Event	4d4b7105d754a06373d81259
3	Food	4d4b7105d754a06374d81259
4	Nightlife Spot	4d4b7105d754a06376d81259
5	Outdoors & Recreation	4d4b7105d754a06377d81259
6	Professional & Other Places	4d4b7105d754a06375d81259
7	Residence	4e67e38e036454776db1fb3a
8	Shop & Service	4d4b7105d754a06378d81259
9	Travel & Transport	4d4b7105d754a06379d81259

The number of each categorical venues from this ten categories were obtained from four square API for each ward and following box plot describes how these categories are distributed in wards of Lalitpur and Kathmandu.



We can observe that Food dominates the category whereas Shop and services, outdoors and Creations are leading behind Food. As event venue has no significant data we dropped event category.

We normalized the data using min-max scaling (1: highest value, 0: lowest value).This normalizes the data and makes it easy to interpret different venues .The scaled data looks like



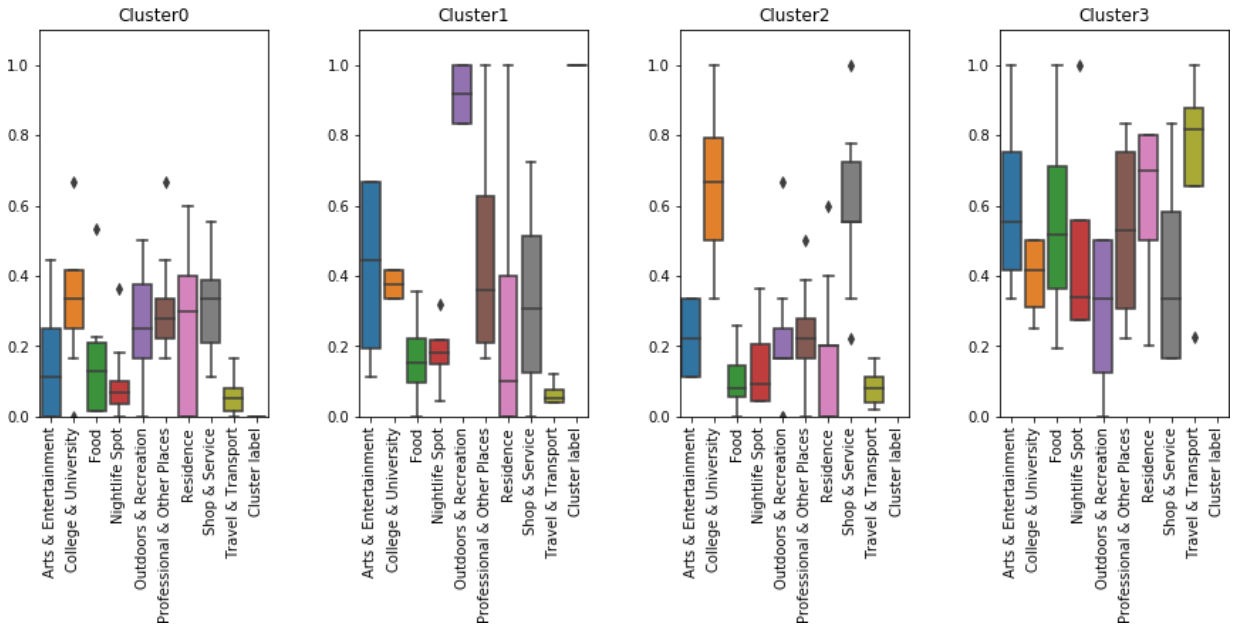
Clustering

We used KMeans Clustering for clustering the wards of Kathmandu and Lalitpur based on the scaled categories data of each ward.

Preliminary results for different number of clusters (K).

1. K=2 : shows only high number and low number of venues
2. K=3 : clusters with high number of colleges, low number of venues and third one is hard to interpret
3. K=4 : clusters with high number of colleges, another one with high number of outdoors and Remaining two with high and low number venues
4. K=5 : hard to interpret

We chose number of cluster four for our analysis of wards clusters as it gives the most insight.



We can observe from box plot that when clustering data in four clusters we get following clusters.

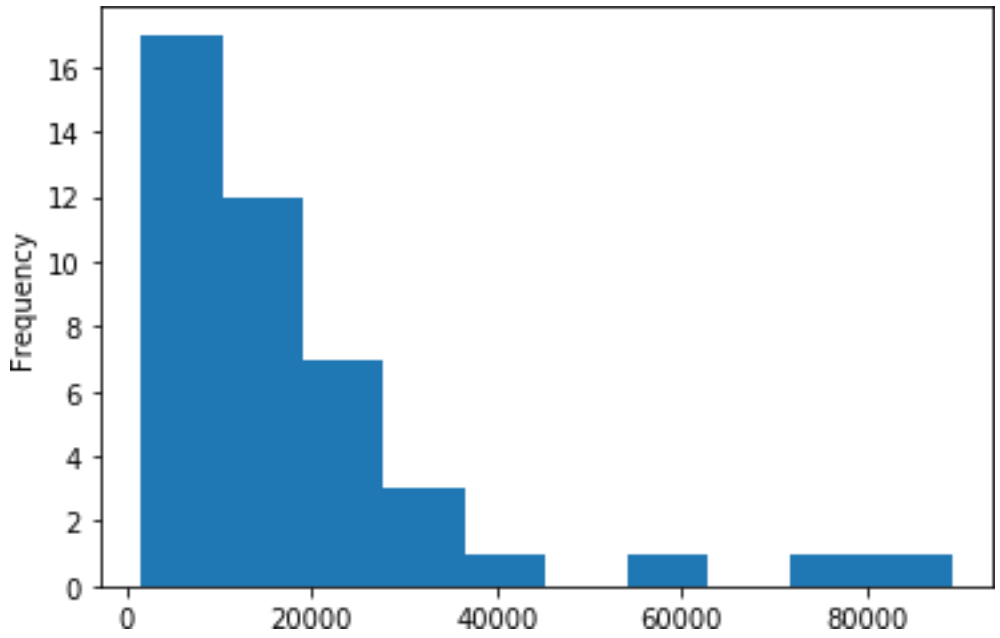
a. Cluster 0 : Diverse but low venues cluster
b. Cluster 1: Outdoors, Creation , Arts and Entertainment Cluster
c. Cluster 2: Colleges, universities, shops and services Cluster
d. Cluster 3: Diverse and high venues cluster

Top five venues

Next we ranked top five individual sub venue category for each ward to get proper insight in what venues are popular in the particular ward.

	Ward	City	1st Most common venue	2nd Most common venue	3rd Most common venue	4 th Most common venue	5 th Most common venue
0	Ward 1	Kathmandu	Asian Restaurant	Hotel	Shopping Mall	Bar	Burger Joint
1	Ward 10	Kathmandu	Fast Food Restaurant	Restaurant	Café	Bus Stop	Hotel
2	Ward 11	Kathmandu	Café	Shopping Mall	Multiplex	Indian Restaurant	Coffee Shop
3	Ward 12 and 13	Kathmandu	Hotel	Asian Restaurant	Café	Bus Station	Indian Restaurant
4	Ward 16	Kathmandu	Hotel	Asian Restaurant	Restaurant	Bus Station	Soccer Stadium

Labelling



This histogram represents the distribution of Kathmandu and Lalitpur wards population density.

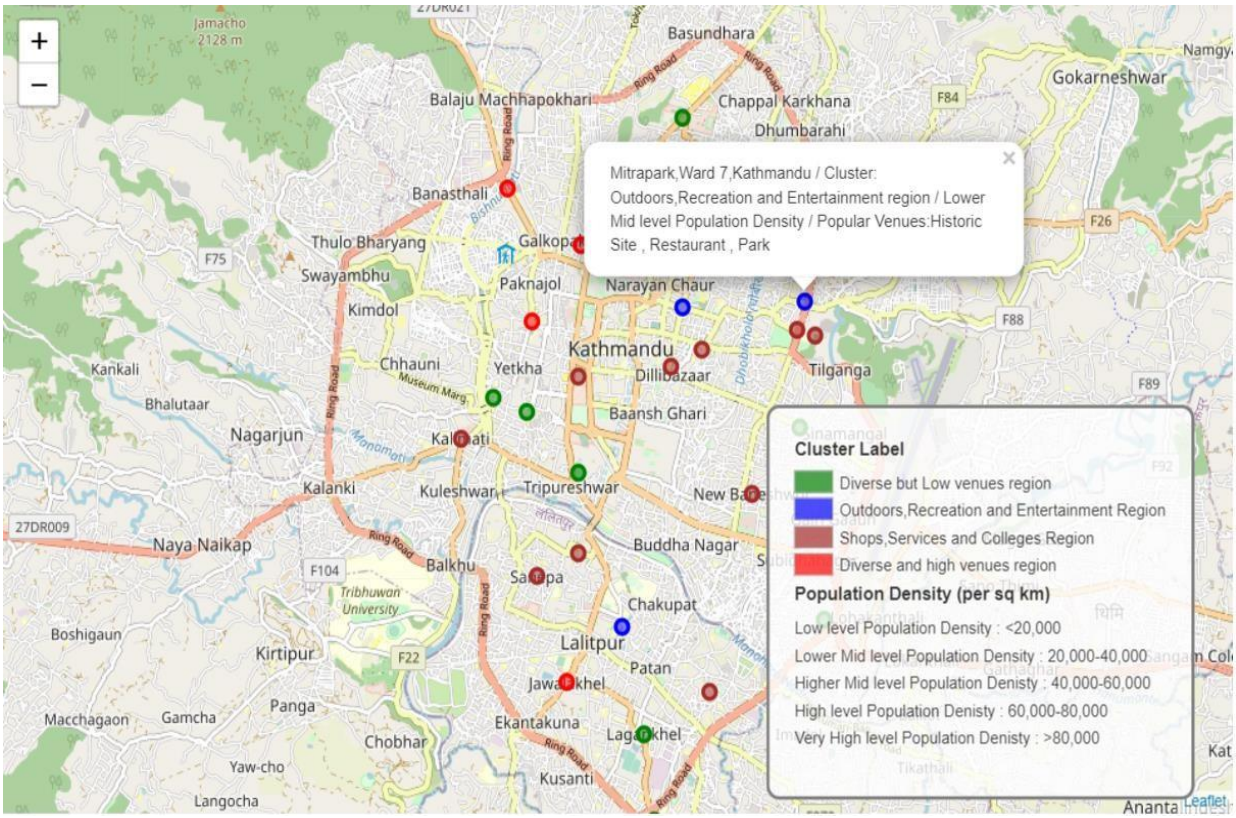
Based on this histogram we can label each wards with following label respect to their population density.

a. >20,000=Low level Population Density
b. 20000-40000= Lower Mid-level 1 Population Density
c. 40,000-60,000= Higher Mid-level 2 Population Density
d. 60,000-80,000=High level Population Density
e. >80,000=Very high level Population Density

Merging the all the details we have obtained for each ward with population density label and cluster label we get following final dataset.

	Ward	City	Location	latitude	longitude	Area(sq km)	Population Density (per sq km)	1st Most common venue	2nd Most common venue	3rd Most common venue	4 th Most common venue	5 th Most common venue	Cluster label	Cluster	Population Density
0	Ward 1	Kathmandu	Naxal	27.712678	85.328703	1.3818	5795.0	Asian Restaurant	Hotel	Shopping Mall	Bar	Burger Joint	2	Outdoors,Recreation and Entertainment region	Low level Population Density
1	Ward 10	Kathmandu	New Baneshwor	27.692621	85.338072	1.5643	25455.0	Fast Food Restaurant	Restaurant	Café	Bus Stop	Hotel	1	Shops, Services and Colleges region	Lower Mid level Population Density
2	Ward 11	Kathmandu	Tripureshwor	27.695000	85.314900	1.8126	9801.0	Café	Shopping Mall	Multiplex	Indian Restaurant	Coffee Shop	0	Diverse but Low venues region	Low level Population Density
3	Ward 12 and 13	Kathmandu	Kalimati	27.698583	85.299157	2.7843	19293.0	Hotel	Asian Restaurant	Café	Bus Station	Indian Restaurant	1	Shops, Services and Colleges region	Low level Population Density
4	Ward 16	Kathmandu	Sohrakhutte	27.725406	85.305457	4.6335	18224.0	Hotel	Asian Restaurant	Restaurant	Bus Station	Soccer Stadium	3	Diverse and High venues region	Low level Population Density

Finally we achieved a plot of Kathmandu and Lalitpur Wards on basis of their cluster with information on top 3 venues, cluster label and population density.



Results

Clustering different clustered wards in map shows us that most of the diverse but low venues region lies in the outskirt of combined Lalitpur and Kathmandu city whereas center of both cities and combined city do consist high venues and specific venues region. We can observe both cities wards mirroring each other at many mirror geographical locations .The population density gives us further insight on the wards residential and market situation.

Discussion

The four square data are insightful but not all-encompassing especially for country like Nepal where digital platform use and accessibility has just begin to boom . We can observe that many wards have low numbers of venues in four square API which certainly doesn't coincide with the actual number of venues. Also the four square venue data focus on category it doesn't include the magnitude of venue. As we can see that a park certainly attracts more magnitude of in comparison to a restaurant.

Conclusion

We can deploy tools like this in far greater scale with more features like housing rate, air quality rate to study our cities and other regions in greater details.

Not only for people seeking out their favorite neighborhood, these sort of analysis can even significantly benefit parties like administration, businesses groups and real estate investors.

References

[1] Kathmandu Ward Offices Location

[2] Kathmandu Ward Profiles

[3] Chhetri, Sachin & Kayastha, P... (2015). Manifestation of an Analytic Hierarchy Process (AHP) Model on Fire Potential Zonation Mapping in Kathmandu Metropolitan City, Nepal. ISPRS International Journal of Geo-Information. 4. 400-417. 10.3390/ijgi4010400.

[4] Lalitpur Wards Offices

[5] Lalipur Wards Details

[6] Google Maps

[7] Four Square