

Classifying the tourist places in Kathmandu

Table of content

Table of content	2
Introduction and problem statement	3
Data Description	3
Methodology	5
Results	7
Discussion	8
Conclusion	8

Introduction and problem statement

Kathmandu is one of the most popular tourist destinations in the world. Millions of tourists visit the small capital of Nepal every year to witness the culture and nature. For an outsider, information regarding the different places is crucial to make important decisions for food, stay, and travel.

In this project, I have tried to cluster several neighborhoods of Kathmandu so that each cluster is suitable for some common purpose. Tourists are the main target of this information and I believe it will help them in decision making.

Foursquare API is used to explore the neighborhoods and obtain the most popular places in each neighborhood. Then, a cluster analysis (K-means analysis) on the popular places of Kathmandu is performed to group the neighborhood.

Data Description

Since there is no government information on the classification of neighborhoods, I have first created a list of popular neighborhoods. In addition to this, I have also created lists of their latitude and longitude and combined all the lists into a single pandas data frame.

The kinds of data used in the data frame and their source are mentioned below:

- a. Neighborhood name (Obtained from Wikipedia and local government websites)
- b. Latitude and longitude of such neighborhood (obtained from google map)

Example of the used dataset is shown below:

```
[197] df.head()
```



	Neighborhood	Latitude	Longitude
0	Baluwatar	27.7255	85.3298
1	Baneshwor	27.6915	85.3420
2	Chabahil	27.7166	85.3485
3	Dillibazaar	27.7054	85.3267
4	Gairidhara	27.7191	85.3262

Figure: Sample of data

FourSquare API is used to explore the places and obtain categorical features regarding the venues popular within a 300m radius of the neighborhood.

Methodology

Following steps are followed to obtain the clustering of the neighborhood in Kathmandu:

- a. Data collection: Collecting the data on tourist neighborhoods of Kathmandu using Wikipedia, government websites, and Google Maps.
- b. Data conversion: Data collected in list is converted to the pandas' data frame for analysis.
- c. Exploring neighborhood: Using a python function and FourSquare API, each neighborhood stored in the data frame is explored, their popular venues with their categorical values are obtained and stored.
- d. One hot encoding: For clustering, the neighborhood, the categorical variables obtained from the foursquare API are converted to a binary class. New columns are added and appended to the data frame. This is the final data frame suitable for clustering analysis.
- e. Finding the suitable k: Before applying the K-means clustering to the dataset, I have used the Sillhouette method to find the best value of k suitable for the dataset.

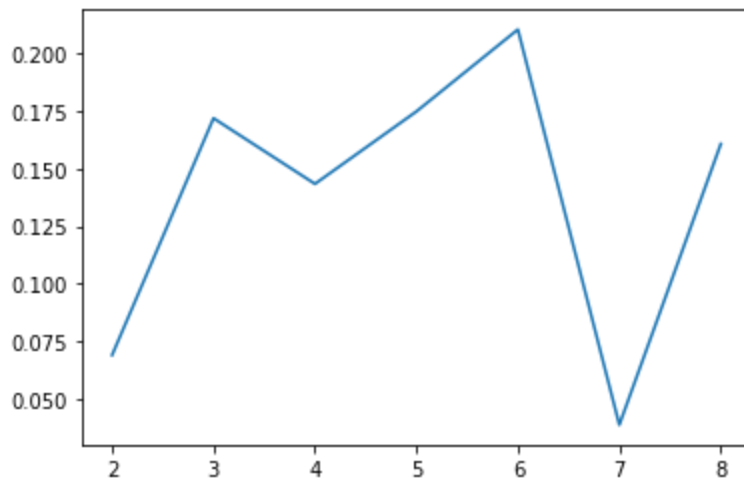
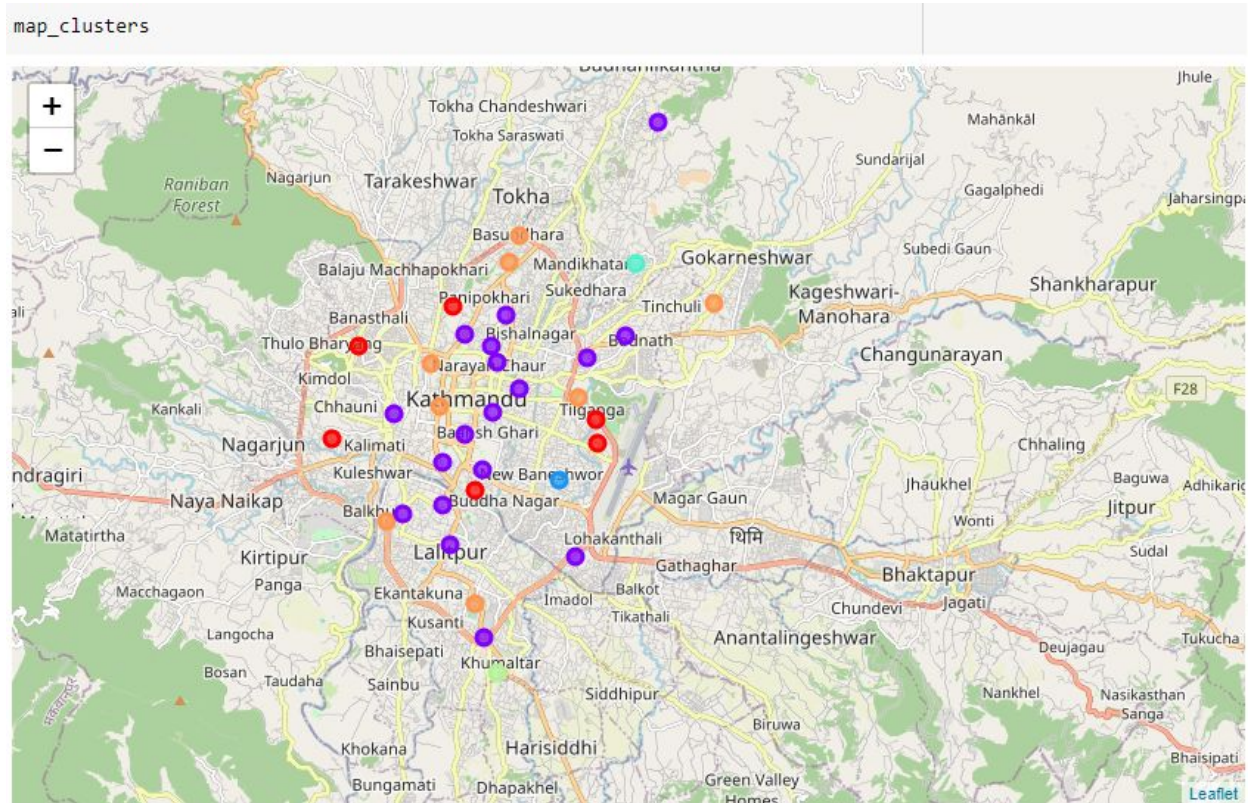


Figure: Silhouette score vs. k

- f. Clustering: After finding the value of k , k -means clustering is performed on the dataset.

Results



After examining each cluster, I have reached to following conclusions:

1. Each cluster of the neighborhood has a peculiar venue popular. Which are stated below:
 - a. Cluster 1 is suitable for hotels and is closer to the airport.
 - b. Cluster 2 is suitable for restaurant and dining purposes.
 - c. Cluster 3 is most popular for recreational activities.
 - d. Cluster 4 is popular for sports.
 - e. Cluster 5 is popular for coffee shops.
 - f. Cluster 6 is popular for shopping and recreational activities. The places in this cluster also has a high density of population.

Discussion

The data obtained from the FourSquare API was very small. Maybe this is because people are not very active contributors to the platform. Hence, the dataset couldn't yet clearly indicate the clusters. There are issues of different places (location-wise) falling in the same cluster. The motive of clustering is also not clear. Maybe, with a larger dataset, the clustering would have been effective.

Conclusion

With more data on neighborhoods, this clustering data can be effective in providing important information to tourists regarding the features and popularity of different places in Kathmandu. This will save time and help them in quick and best decision making.