

Classifying the tourist places in Kathmandu

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 will be used to explore the neighborhoods and obtain the most popular places in each neighborhood. Then, a cluster analysis (K-means analysis) will be performed to group the neighborhood.

Data Description

Since there is no government information on the classification of neighborhoods, I will be first creating a list of popular neighborhoods. I will also create lists of their latitude and longitude and combine all the lists into a single pandas data frame.

As mentioned earlier, the foursquare API will provide me with the places exploration information.

Adequate preprocessing of datasets will be performed to obtain clean data for analysis. Unnecessary columns will be dropped, na values will be ignored and rows will be deleted.

While performing K-Means clustering, I will use the Sillhouette Method to find the best k value for the clustering.

The kinds of data used in the data frame:

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

```
[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