



The Battle of Neighborhood

Comparing similarity of neighbourhoods in the cities of India

Introduction

India is the seventh largest country in the world by geographical area and world's second-most populous country with more than 1.38 crore people. According to Census of India 2011, there are 7935 cities.

The number of metropolitan cities having million plus population has also increased from 35 to 53 as per 2011 census.

In the current modern world citizens migrate from one place to another for a better opportunity.

The world is facing an unprecedented threat from the COVID-19 pandemic. It is difficult for a citizen who wants to explore the whole new city and choose the locality in order to find a place of reside.

The aim of the project is to compare the cities of India to find the similarities based on the venues in the neighbourhood. This helps the working class people who may have to move to new city and are looking for the comfortable locality to stay. The locality they would want to choose is dependent on the lifestyle they carry that is in turn dependent majorly on the venues in the nearby locality. For example some people prefer to stay in a neighbourhood which has restaurants, shopping malls in the near vicinity whereas some people prefer staying near to parks, recreational zones etc. In this project, I will be making an attempt to compare the similarities of cities Pune and Bangalore City, and this methodology can be extended to compare any cities of India. Both these cities are culturally diversified, vibrant and full of opportunities for a working professional.

Data Info



Post office is a Indian public facility that provides mail services, including accepting of letters and parcels, providing post office boxes to every nook and corner of the country, a dataset All India Pin code Dataset from the Open Government Data(OGD) Platform India is used to derive the neighbourhoods of cities.



Geopy client to fetch the coordinates of the neighbourhood.



Foursquare API to get 100 venues in the 1500 meters radius of the neighbourhood.

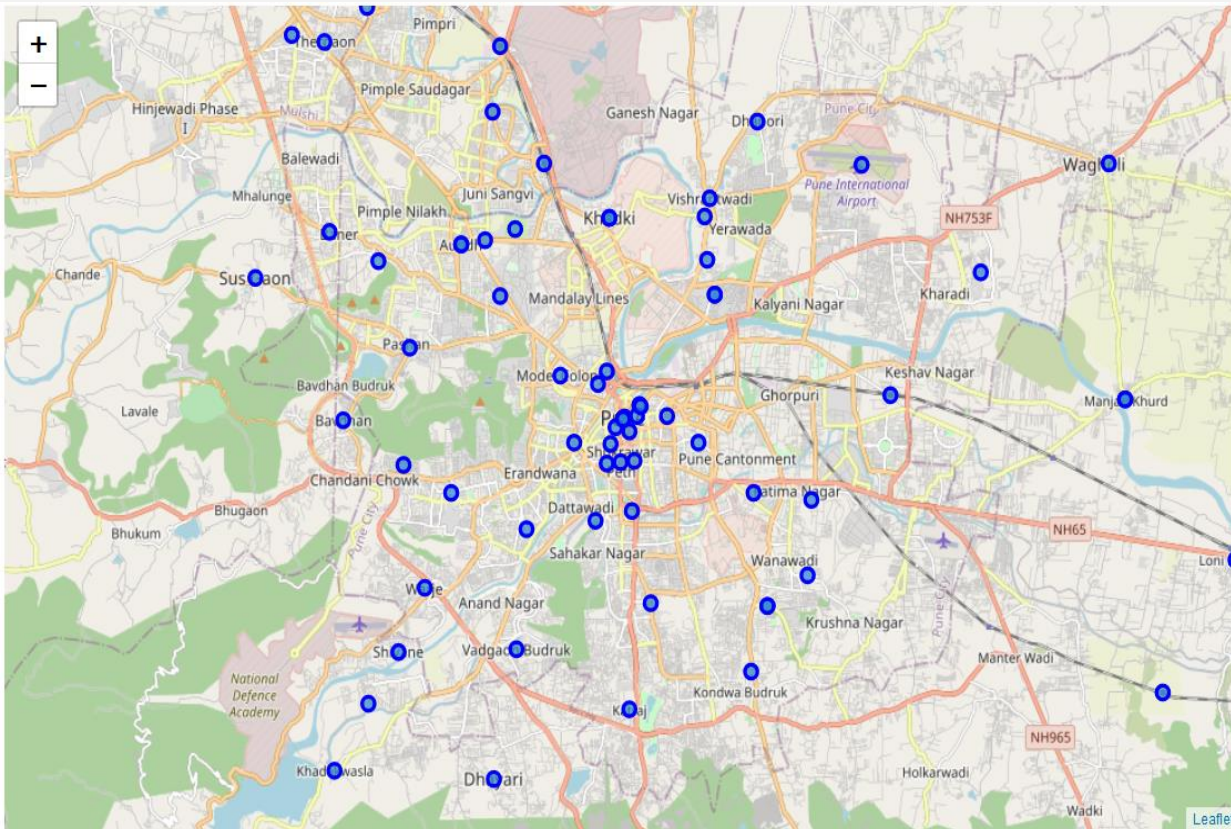


Data cleaning to discard unnecessary features, duplicate records on the filtered dataset of Pune and Bangalore.

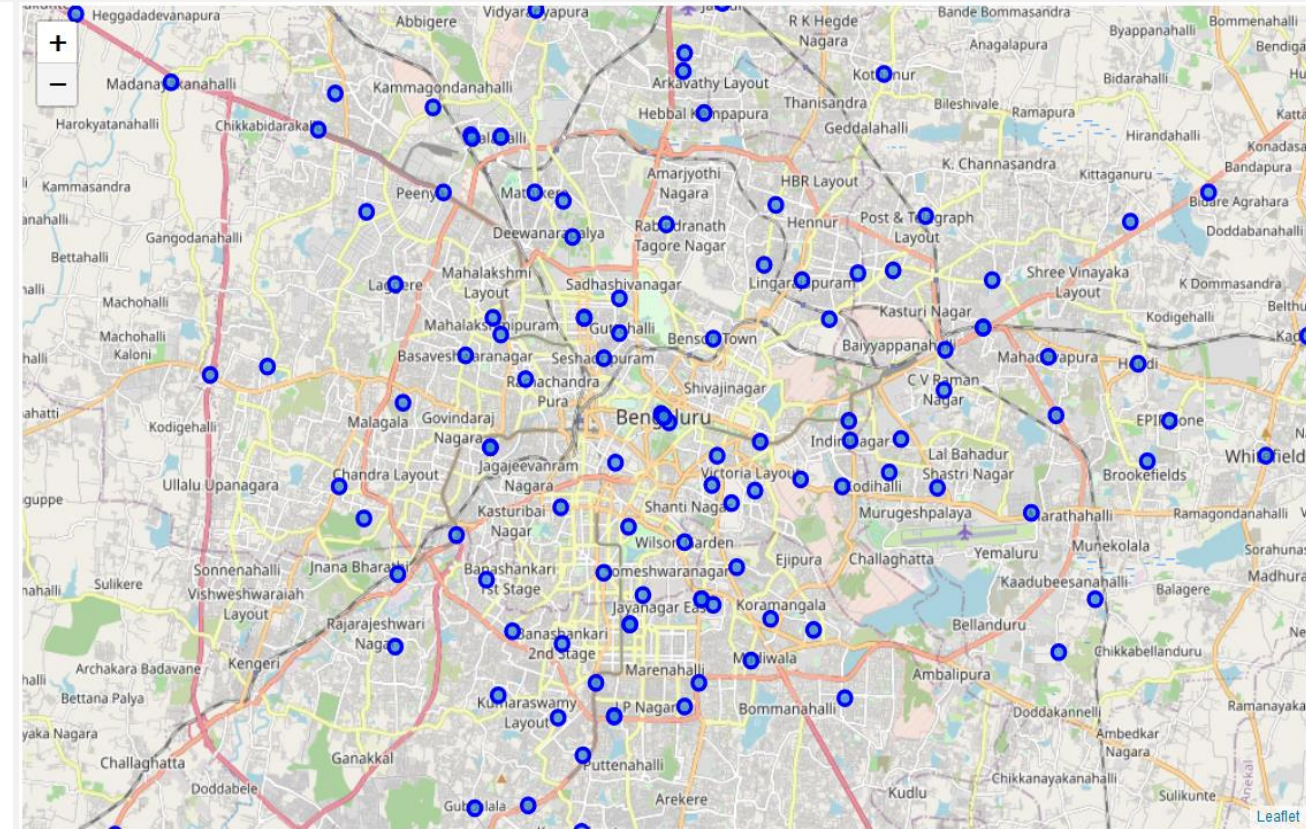


We have 105 and 147 neighbourhoods with their geographical coordinates.

Visual Analysis



Neighborhoods of Pune

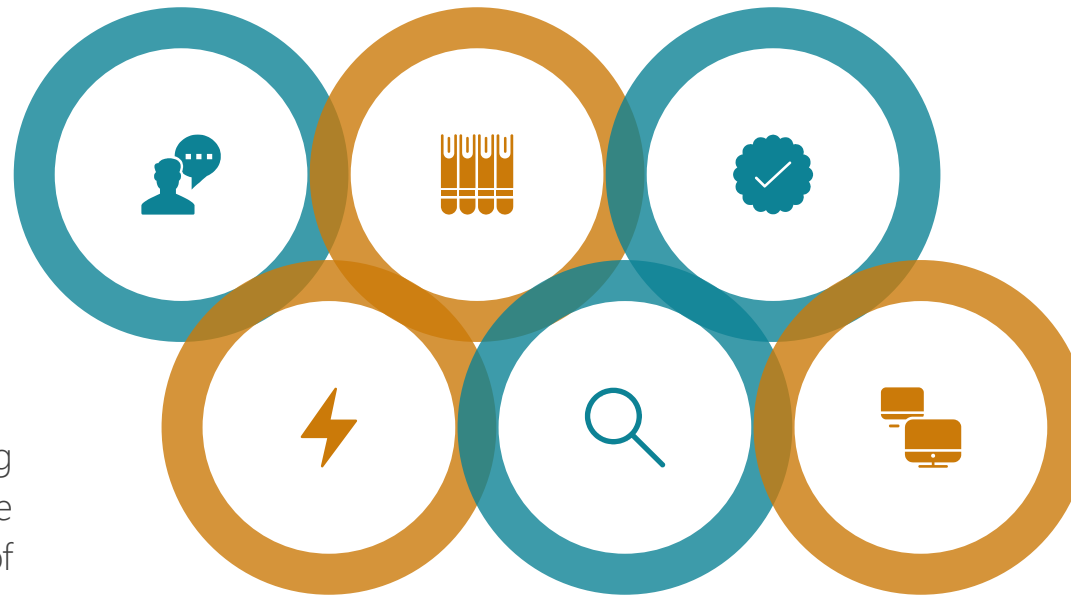


Neighborhoods of Bangalore

Project Analysis

The neighbourhood points on the map are closely and uniformly placed. Hence the radius of 1500 meters is chosen to get the venues in the neighbourhoods.

Unsupervised learning methodology to be applied to identify the existing relationship(similarity) between two cities.



To apply the machine learning technique, the data is required to be meaningful and in terms of the type of venues that exist in the neighbourhood.

K-means clustering algorithm identifies the k number of centroids, and then allocates every data point to the nearest cluster centroid. K-means clustering is one of the simplest, popular unsupervised machine learning algorithms and is suited to solve the problem of this project

One-hot encoding is used on the categorical data venue type and the mean of the values is taken to represent the neighbourhood in terms of venue type.

Results and Analysis

Cluster 0

- › Many restaurants, departmental stores, café, coffee shop, brewery.

Cluster 3

- › ATMs, Auto workshops, Restaurants, Pizza shops.

Cluster 1

- › Scenic outlook, zoo.

Cluster 4

- › Resorts, Lake, Trails, Farms

Cluster 2

- › Indian restaurants, café, coffee shop, fast food, dessert shops

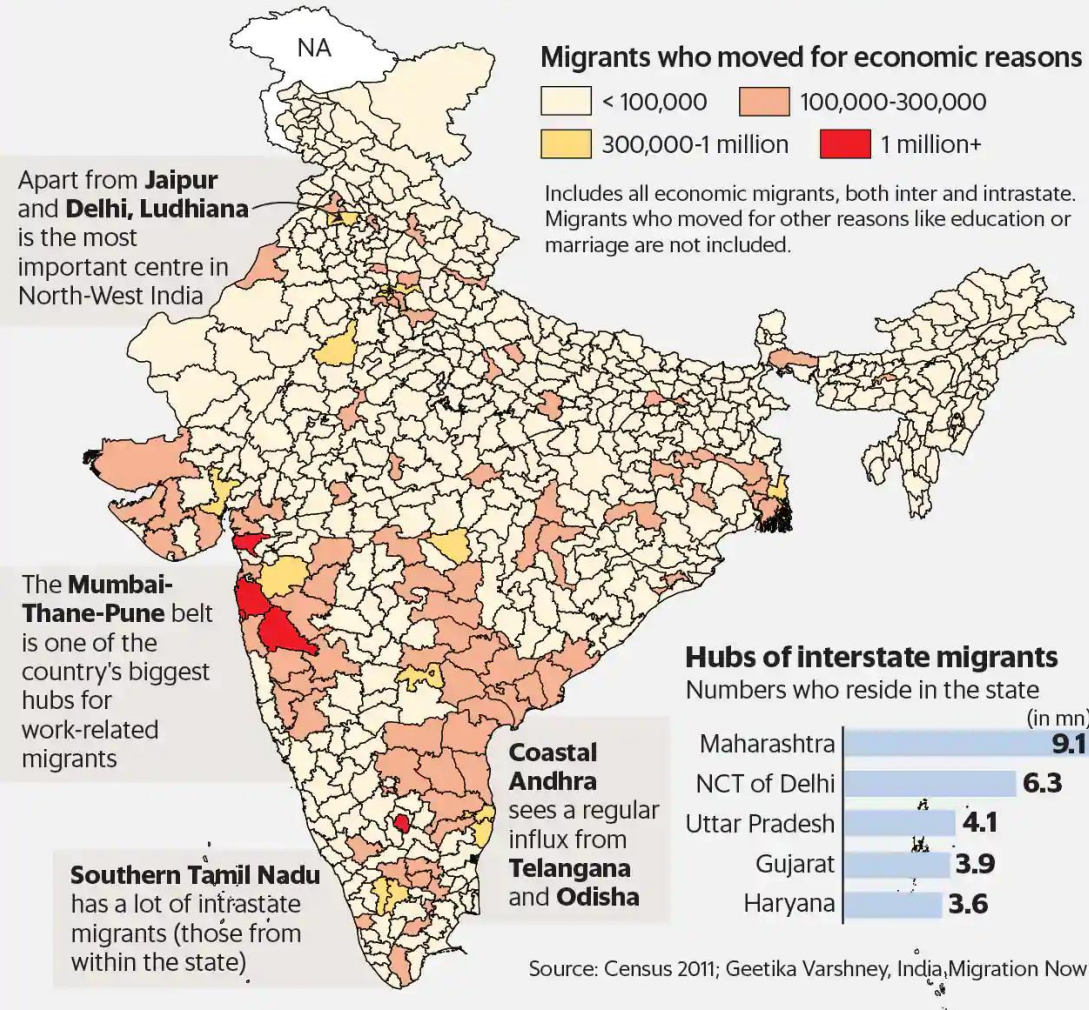
Cluster 5

- › Indian restaurants, flea market, Gym, bakery, garden , super markets.

Conclusion

THE MIGRANT HOTSPOTS

Some districts, like Thane or Mumbai suburban, house over a million economic migrants. Just 22 districts in the country account for a third of all migrant workers.



The purpose of this project was to demonstrate the methodology to explore and find similarities of the neighbourhoods based on the venues in the neighbourhoods in any two cities.



This idea can be applied to any other cities of India, the all India pin code dataset provided by OGD India platform has neighbourhoods of all the cities in India.



Further we could include the cost of living by including the rental data for creating clusters, there are lot of websites which advertise and display the rental price of the properties.



Thank You