

Governing Bengaluru (Bangalore)

1.Introduction/Business Problem

1.1. Background

Bengaluru (formerly and commonly known as Bangalore) is one of the fastest growing cities in the world. Many people move to Bengaluru to pursue career opportunities, a grand city life, and the fantastic weather. People often refer to it as the 'Silicon Valley of India' as it is a hotbed for Information Technology, Artificial Intelligence, and Data Science. As the infrastructure and population continue to grow, efficient governance has become a problem. Many argue over the split of the Bruhat Bengaluru Mahanagara Palike (BBMP) and whether splitting up Bengaluru would promote smoother administration.

1.2. Problem

Bengaluru is a very diverse place, containing different cultures and spanning over both urban and rural territory. Splitting up Bengaluru, so that similar policies/projects can be efficiently co-implemented is a difficult problem to pursue intuitively. Within a small radius, one can find small fish markets and some of the most modern malls. Thus, a data science based approach may prove useful to solve this issue.

1.3. Interest

The policy-makers in Bengaluru would be the interested party in such an analysis. Splitting up Bengaluru into similar neighborhoods would aid in smoother administration, and policies can be geared to solve problems that are likely similar within similar neighborhoods.

2. Data

2.1. Source

Kaggle is a community based environment for data scientists and machine learning enthusiasts. Google created this environment, and it is a fantastic place for people to share data and projects. The data for Bengaluru neighborhoods was obtained from the site, and the original data source would be a central Indian website 'data.gov.in'.

2.2. Data Cleaning

The data obtained look to have some obviously incorrect outliers. Places that are in Bengaluru cannot have such a wide range of latitudes and longitudes. So outliers are

removed and replaced by values found online. In addition, duplicate neighborhoods and excessive columns were removed to obtain a clean dataset.

2.3. Data Usage

The data contains latitudes and longitudes that can be used in foursquare to procure nearby venues. With this data, one can cluster similar neighborhoods together. This can be used to answer the question of how to split up Bengaluru into similar neighborhoods for effective governance. Table 1 is an example of how the foursquare data is received and input into a dataframe. Knowing the most common venues can predict which neighborhoods are similar to each other. A specific example would be places with many shopping malls and pubs might be a determining factor in grouping similar neighborhoods.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Achitnagar	Recreation Center	Bakery	Asian Restaurant	Restaurant	Dry Cleaner	Eastern European Restaurant	Electronics Store	Event Service	Event Space	Yoga Studio
1	Adugodi	Indian Restaurant	Café	Pizza Place	Coffee Shop	Gym	Lounge	Chinese Restaurant	Clothing Store	Tea Room	Dessert Shop
2	Agram	Indian Restaurant	Hotel	Restaurant	Ice Cream Shop	Asian Restaurant	Pub	Clothing Store	Bar	Pizza Place	Café
3	Akkur	Fast Food Restaurant	Bus Station	Yoga Studio	Duty-free Shop	Flea Market	Fishing Spot	Financial or Legal Service	Field	Farmers Market	Farm
4	Alahalli	Food & Drink Shop	Indie Movie Theater	Duty-free Shop	Food	Flea Market	Fishing Spot	Financial or Legal Service	Field	Fast Food Restaurant	Farmers Market

Table 1: Most Common Venues for Bengaluru Neighborhoods