

Governing Bengaluru

(Bangalore)



Introduction

- 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.

Problem and Interest

- 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.
- 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.

Data Source and Cleaning

- The data was obtained from Kaggle, which is a community based environment for data scientists and machine learning enthusiasts.
- 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.

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. The table below is an example of how the foursquare data is received and input into a dataframe.

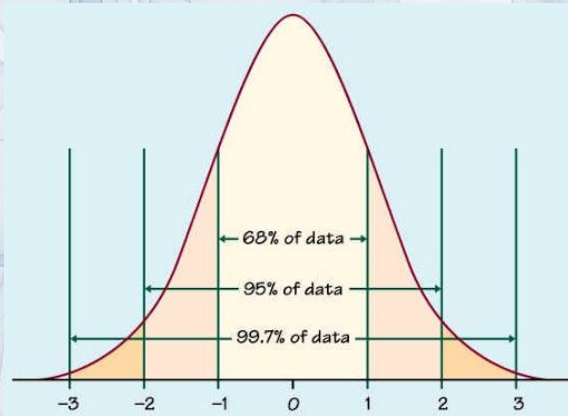
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Achitnagar	Recreation Center	Bakery	Asian Restaurant	Restaurant	Dry 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

Methodology

- Looking into the data and the history of the city it is expected that there will be a large number of neighborhoods in the center and the number of neighborhoods that surround will be of lower concentrations.
- Furthermore, with the airport located far from the city center, it is expected that there will be a cluster of neighborhoods that have accumulated near that location.
- Analyzing the data backs these claims with many neighborhood locations (latitude and longitude) clustered around the city center. We also see a higher than usual number of neighborhoods near the airport compared to locations equally distant from the city center.

Removing Outlier using Z-Score

- The z-score represents how close or far an observation is from the overall mean. For a normal distribution as shown in the Figure below, 99.7% of the data is within three standard deviations of the mean.
- Using python's scipy library, the z-scores of the locations were retrieved. After the outliers were highlighted, they could be replaced manually with correct latitudes and longitudes found online.



Data Science and Machine Learning Techniques

- First, foursquare data was used to collect venues nearby to each neighborhood.
- These categorical variables can be used to identify similar neighborhoods.
- For a numerical algorithm to be used, the categorical variables need to be converted to numerical values, which can be done by using dummy variables. Each venue type is put into a separate column, and if the location has the venue type it will be recorded as one, and if it isn't present it is recorded as zero.

One-Hot Encoding Table

	Neighborhood	ATM	Accessories Store	Afghan Restaurant	Airport	Airport Service	Airport Terminal	American Restaurant	Andhra Restaurant	Arcade	...	Turkish Coffeehouse	Udupi Restaurant	Vegetarian / Vegan Restaurant
0	Achitnagar	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000
1	Adugodi	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000
2	Agram	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.01	0.000000
3	Akkur	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000
4	Alahalli	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000
...
208	Whitefield	0.0	0.0	0.0	0.000000	0.0	0.0	0.027027	0.000000	0.0	...	0.0	0.00	0.013514
209	Yadavanahalli	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000
210	Yelachenahalli	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.021505
211	Yelahanka	0.0	0.0	0.0	0.032258	0.0	0.0	0.032258	0.032258	0.0	...	0.0	0.00	0.032258
212	Yeliyur	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.00	0.000000

K-Means Clustering and Folium

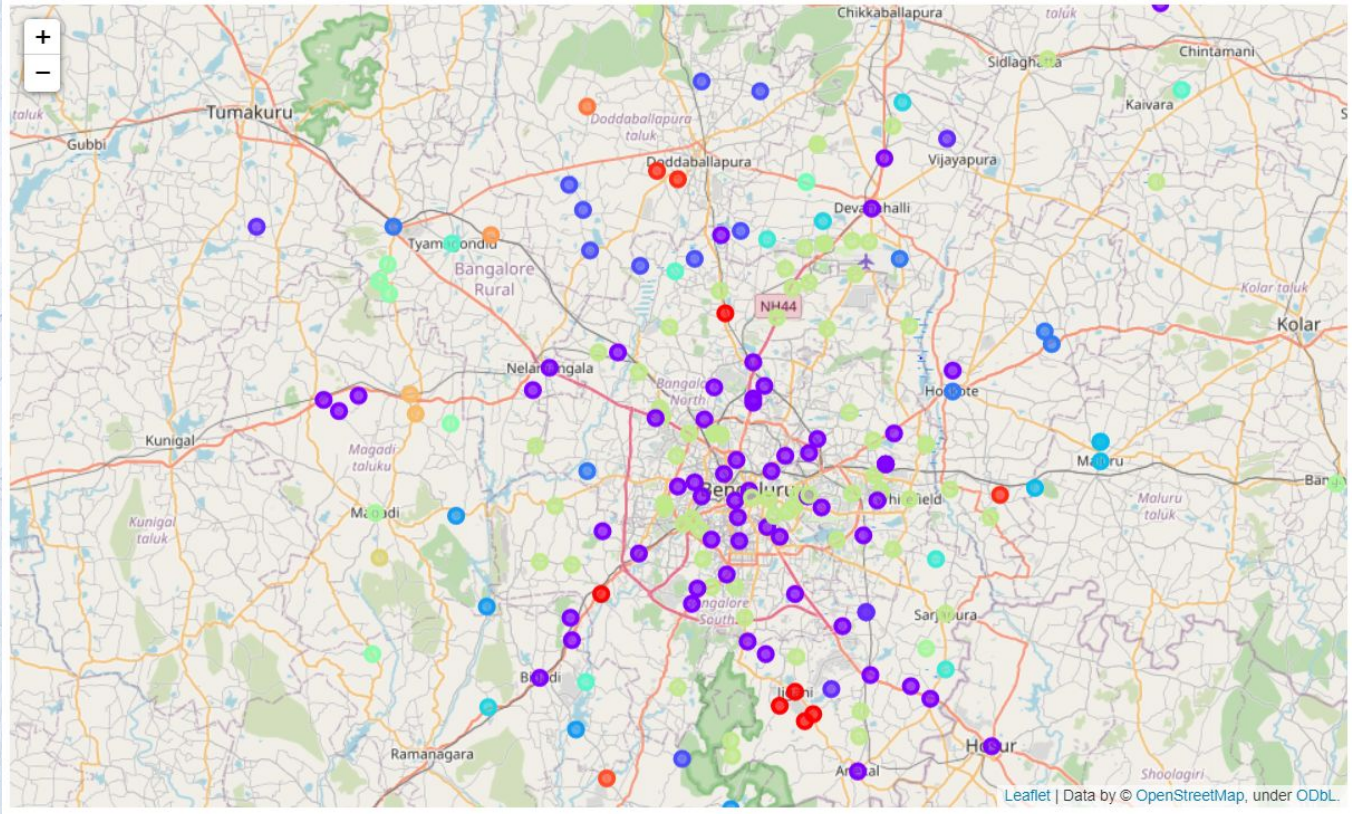
- K-means clustering was the primary technique used to analyze the data.
- This technique is an unsupervised machine learning methodology that uses attempts to cluster similar data points together in a way that minimizes the intra-distance within a cluster and maximize the inter-distance between clusters.
- Finally using Folium, the map with color coded clusters is obtained.
- This visually shows the clustering for Bengaluru neighborhoods.

Results of the Three Main Clusters

- Cluster 1 seems to be citizen hubs.
- Cluster 13 is the travel heavy neighborhoods.
- Cluster 6 has the train station as the primary venue. The daily wage workers who ply their trade in flea markets, fishing spots and fields encompass these neighborhoods.

	Neighborhood	Latitude	Longitude	Cluster Labels	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
14	Fraser Town	12.997000	77.614400	1.0	Indian Restaurant	Café	Tea Room	Middle Eastern Restaurant	Bakery	Pub	Ice Cream Shop	Shopping Mall	Coffee Shop	Fast Food Restaurant
54	Chandapura	12.801700	77.711600	1.0	Indian Restaurant	Asian Restaurant	Coffee Shop	Train Station	Yoga Studio	Farm	Electronics Store	Event Service	Event Space	Falafel Restaurant
165	Doddajala	13.176735	77.652050	13.0	Fast Food Restaurant	Hotel	Airport Terminal	Bike Shop	Farm	Toll Booth	Food Truck	Airport	Event Service	Food & Drink Shop
241	Korati	12.971600	77.594600	13.0	Indian Restaurant	Hotel	Lounge	Brewery	Pub	Café	Ice Cream Shop	Italian Restaurant	Sushi Restaurant	Japanese Restaurant
264	Malur	13.006034	77.938284	6.0	Train Station	ATM	Men's Store	Duty-free Shop	Flea Market	Fishing Spot	Financial or Legal Service	Field	Fast Food Restaurant	Farmers Market
274	Marasandra	12.980402	77.873983	6.0	Train Station	Yoga Studio	Duty-free Shop	Food	Flea Market	Fishing Spot	Financial or Legal Service	Field	Fast Food Restaurant	Farmers Market

Map of Bengaluru Neighborhoods



There is significant clustering near the city center, the airport, and near the highways.

Concentration of neighborhoods decrease as you move away from the city centre.

Discussion

- Cluster 1 would benefit most from housing projects. These are places with shopping malls, bakeries and yoga studios for the people who have settled down in Bengaluru.
- As the population grows, housing prices will spike. Cluster 13 would benefit from infrastructure rebuilding. This cluster is for people who live intermittently in Bengaluru. Thus connecting the metro from these neighborhoods to the airport could be enormously beneficial.
- Cluster 6 would benefit from cheap transportation options as they are daily wage earners working in the flea market and in fishing. A greater number of buses and subsidized ticket prices would improve the standard of living immensely.

Conclusion

- Data on Bengaluru neighborhoods were cleaned and outliers were corrected.
- Using calls to foursquare, the most common venues surrounding a neighborhood were obtained.
- With a k-means clustering approach, the neighborhoods were divided into clusters.
- This could help the government allocate funding in the most impactful manner. This kind of data- driven approach can aid in making accurate investment to neighborhoods rather than a one size fits all approach.
- Governing a city like Bengaluru is a difficult task and data science may prove useful in a successful outcome.