Bangalore Neighborhood Shootout

Table of Contents

[Introduction 2](#_Toc43495556)

[Problem Description 3](#_Toc43495557)

[Target Audience 3](#_Toc43495558)

[Data 4](#_Toc43495559)

[Methodology 7](#_Toc43495560)

[RESULT and discussion 9](#_Toc43495561)

[CONCLUSION 9](#_Toc43495562)

# **Introduction**



**Bangalore** is the capital city of the state of Karnataka.

**Bangalore** consists of the closely built old town, together with a number of modern outlying areas (former suburbs) laid out in a gridiron pattern to the north and south, with many parks and wide streets. A large military area is situated just southeast of central Bengaluru. Outlying areas incorporated into the city in 2007 include large green spaces as well as farmland and are sites of population and ICT-related growth.

In today’s technology driven world people move from one place to another on account of new jobs, better amenities, schooling facilities, better work life balance etc. In a city like Bangalore the lives of people revolve around traffic which the city is known for. Whenever people move on account of any of the reasons mentioned above, they try to explore the place they like about and try to dig as much information as possible about it. It can be about the neighborhood, locality, market, price of the place and many more factors including neighborhood analysis.

In such a scenario it would be good to have as much data analysis which could make a person’s life easy by considering a comparative analysis between the neighborhoods with a host of predetermined factors.

# **Problem Description**

The objective of this project would be to help the end user or anyone who wishes to

1. Choosing a rental apartment/house
2. Buy a house/apartment
3. Start a Super market
4. Start a Playschool
5. Start a gymnasium

These could be based on the distribution of various facilities available around the neighborhood.

As an example, this project would compare 2 randomly picked neighborhoods and analyze the top 10 most common venues in each of those two neighborhoods

The project proposes to use K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc.

This would give a better understanding of the pros and cons between the two chosen neighborhoods and present the much wanted insights into the neighborhoods and conclude which one is better and which has the edge over the other.

# Target Audience

Target audiences for this are not limited to an individual or a group of people who are looking for the information. It could be anyone who wishes to rent a place, buy a house or anyone who wishes to open a new business.

# Data

**Data requirements:**

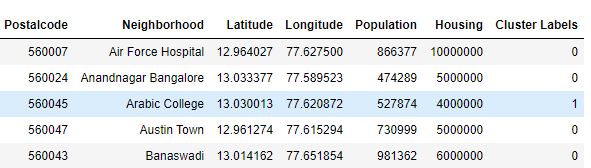
To achieve the said objective data is crucial.

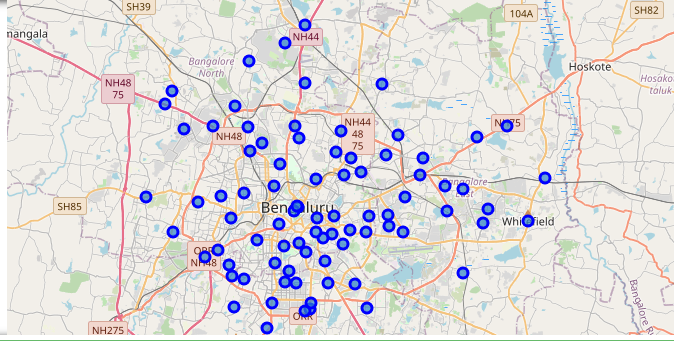
The data that is needed to build a robust recommendation system should address some of the things that are not visible to the human eye, meaning some of the data needs to be extracted from data sources which may not be immediately available.

For this purpose the following sources of data were explored

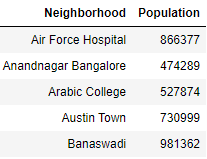
* Kaggle Dataset – By using this data set, neighbourhoods, co-ordinates of each neighbourhood
* Geocoder python - Neighbourhood coordinates of each locality of Bangalore city
* Foursquare API - By using this API, we will get all the venues in each neighbourhood
* Population data for each neighbourhood in Bangalore – This was a difficult one to get but have taken approximate values given the size of Bangalore today which is around 12 million
* Housing prices in Bangalore (this too is approximate given the way Bangalore prices fluctuate)

1. From the dataset available I extracted the neighborhood data along with the latitude and longitude along with population and housing. There were close to 100 neighborhoods whose data has been analyzed

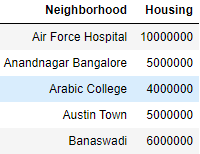




* 2. Population - This was a difficult one to get but have taken approximate values given the size of Bangalore today which is around 12 million



1. Housing – It was tough to gauge this one given the way the real estate industry is so fragmented between small time builders and large ones. I had to sift through real estate websites and then fill out some of the housing prices for each of the areas which may be inaccurate sometimes. All amounts are in rupees.



4. Foursquare API:

Foursquare API helped me to arrive at the different nearby venues and then use them to form a cluster.

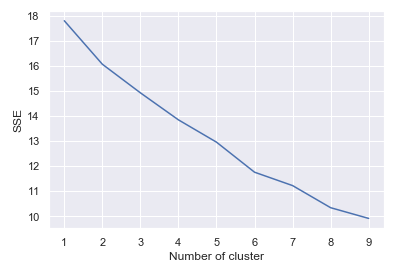
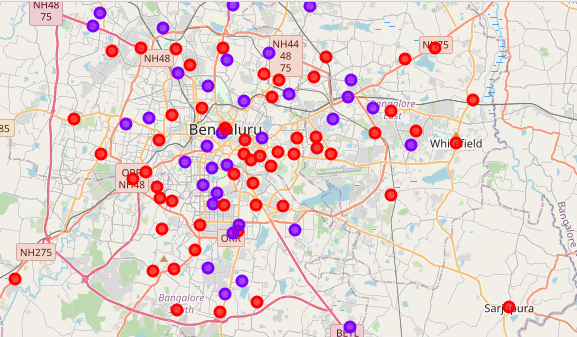


# Methodology

**Exploratory analysis:**

I scrapped the data from different data sources and then combined them into a single data set which had the required features. So I zeroed in on 2 main features which always comes to one’s mind when choosing a place to live, rent an apt etc on what are the different amenities in a neighborhood, what is the population, price of housing etc.

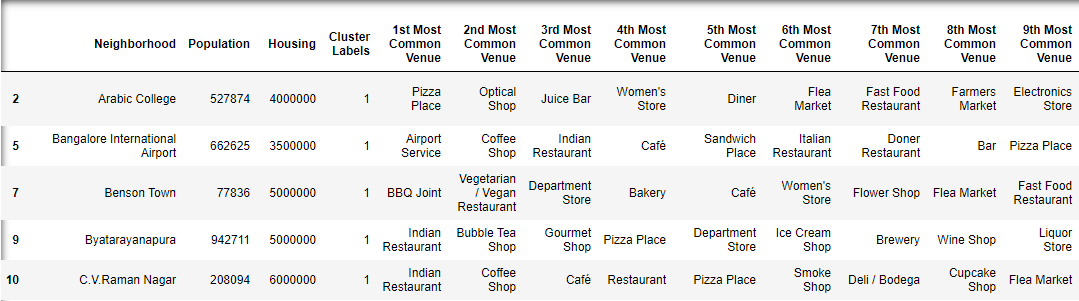
For producing the no.of clusters, a graph was plotted to find the no.of clusters using the elbow K-means algorithm.

Cluster 1



Cluster 2

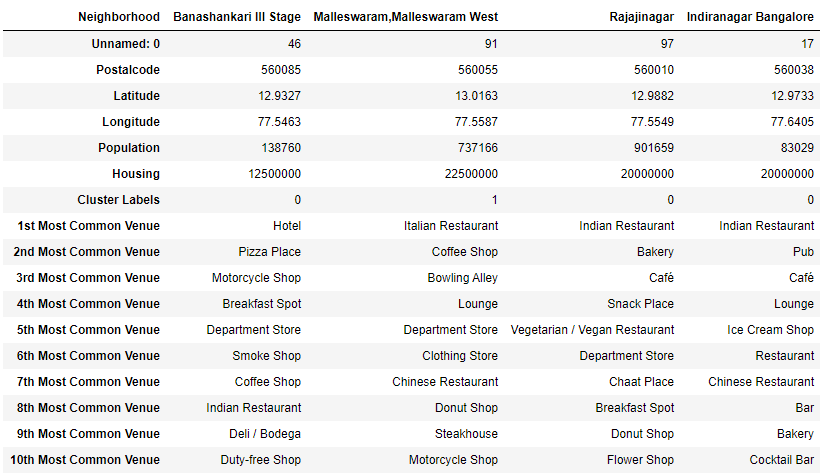


**Inferential analysis:**

To choose a particular neighborhood I have used only 2 features in population and housing which I felt are very important criteria which speaks volumes about the area in terms of amenities, value, how livable it is etc.

I chose a neighborhood from all 4 corners of Bangalore south, east, north and west and

compared them for all of the amenities in addition to the population and housing prices.



# RESULT and discussion

The aim of this project was to provide credible information on the best neighborhoods based on data analysis. The relationship between the best neighborhoods and the features in terms of venues it offers along with lesser population with affordable housing prices were some of the correlations that I wanted to analyze which I think came up quite reasonably well. The clustering exercise done here will be useful for anyone who vies to know about the neighborhoods and probably shift or setup shop for business.

# CONCLUSION

The recommendation about the best neighborhood made here is based on a powerful data driven model. As Bangalore expands and more neighborhoods come up the efficiency may decrease with more data but accuracy will increase which will provide more data for people to move around neighborhoods.