

SETTING UP A MALL IN BANGALORE, INDIA

Submitted by
Netra Chandrasekhar

Business Problem

Visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers.

The objective of this capstone project is to analyse and select the best locations in the city of Bangalore, India to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Bangalore, India, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Data

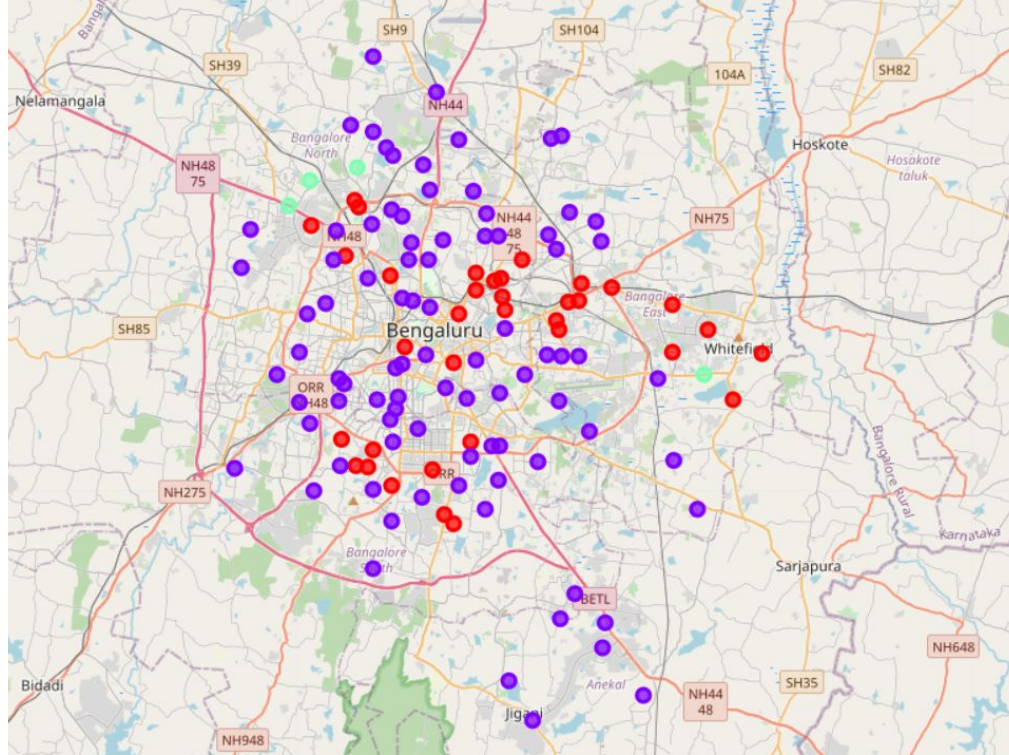
To solve the problem, we will need the following data:

- List of neighbourhoods in Bangalore. This defines the scope of this project which is confined to the city of Bangalore, the capital of TamilNadu in India
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Source of Data

- This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Bangalore) contains a list of neighbourhoods in Bangalore. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using the Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.
- We will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward.

K-means to cluster the neighborhoods in Bangalore into 3 clusters.



Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0: Neighbourhoods with moderate number to no existence of shopping malls
- Cluster 1: Neighbourhoods with high number of shopping malls
- Cluster 2: Neighbourhoods with low concentration of shopping malls

The results of the clustering are visualized in the map below with cluster 1 in red colour, cluster 0 in purple colour, and cluster 2 in mint green colour.

Conclusion

Most of the shopping malls are concentrated in the central area of the Bangalore city, with the highest number in cluster 1 and moderate number in cluster 0. On the other hand, cluster 2 has very low number to totally no shopping mall in the neighborhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, this also shows that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 2 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.