

OPENING A NEW SHOPPING MALL IN BANGALORE, INDIA

31/10/2020

IBM Capstone Project

Opening a new shopping mall in Bangalore, India is a machine learning model, developed to demonstrate as a capstone project to IBM through Coursera. It recommends best location for construction of a new mall based on analysis of already available details of locations of shopping malls.

Table of contents

Topic	Page number
Introduction section	3
Data section	4
Methodology section	5
Result section	7
Discussion section	8
Conclusion section	8

Introduction:

Bangalore is the capital and largest city of the Indian state of Karnataka. Visiting shopping malls is a great way for people to relax and enjoy themselves during weekends and holidays. Shopping malls tend to be a major tourist attraction. The mall are often more convenient, for a tourist, to possess one central location to try to do all their shopping; instead of driving many miles just to buy different types of products for their personal needs. They can do grocery shopping, dine at restaurants, try various fashion outlets, watch movies and perform more activities.

Shopping malls are sort of a one-stop destination for all kinds of shoppers. For retailers, the central location and therefore the large crowd at the shopping malls provides an excellent channel to advertise their products and services. Property developers also are taking advantage of this trend to create more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Bangalore and many more are being built. Opening shopping malls allows property developers to earn consistent income. Of course, like any business decision, opening a replacement mall requires serious consideration and may be a lot more complicated than it seems. Particularly, the location of the mall is one among the foremost important factor which will determine whether the mall is going to be a hit or a failure.

Business Problem:

The objective of this capstone project is to analyze 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?**

Target Audience:

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Karnataka state, Bangalore. Since shopping centres are the most sought-after shopping destinations, and with so many benefits to shoppers as well as businessmen, shopping centres will only rise in popularity with time. Shoppers prefer shopping centres/malls to stand-alone shops for various reasons, main reason being a fun filled experience.

And retailers can have the shop in a prime location with a minimal investment since shopping malls are usually located in prime locations which are easily accessible.

Data:

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

- List of neighborhoods in Bangalore.
- Latitude and longitude coordinates of those neighborhoods. 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 be using this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them:

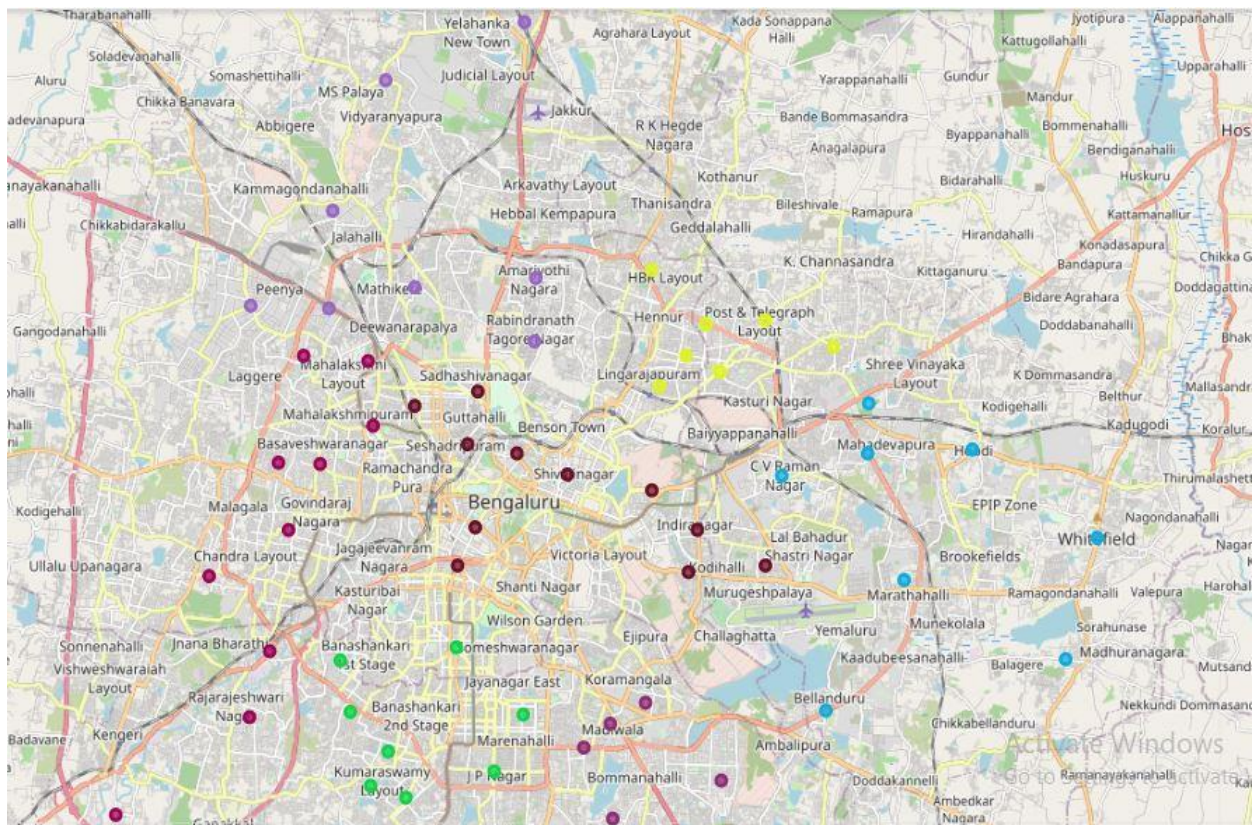
This Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore) contains a list of neighborhoods in Bangalore, with a total of 65 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods.

Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a specified radius and also corresponding coordinates, venue location and names. Our interest lies in the Shopping Mall category in order to help us to solve the business problem put forward. This project makes use of data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

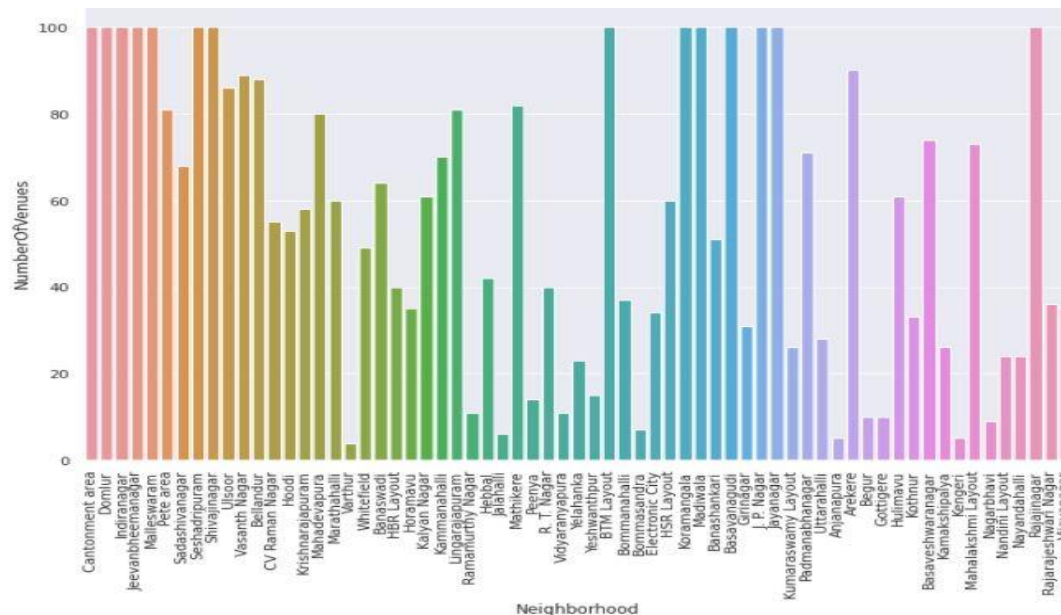
Methodology:

Firstly, we need to get the list of neighborhoods in the city of Bangalore. The list is available in the Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore). We will do web scraping using Python requests to extract the list of neighbourhoods. This is just a list of names. The table headings becoming the boroughs and data becoming the neighborhoods. Bangalore has 8 boroughs and 65 neighborhoods. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. The Geocoder package allows us to convert address into geographical coordinates in the form of latitude and longitude. Since I did not get proper coordinates for the neighborhood, (Map created using Folium package did not depict the original plot of neighborhood in Bangalore city), I manually googled each neighborhood to find its corresponding latitude and longitude and created a CSV file with the column names Borough, Neighborhood, Latitude and Longitude.

After gathering the data, we populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package.



Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data. While exploring the dataset, I found out that few neighborhoods has most number of venues while Varthur has the least.



Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the “Shopping Mall” data, we will filter the “Shopping Mall” as venue category for the neighbourhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular

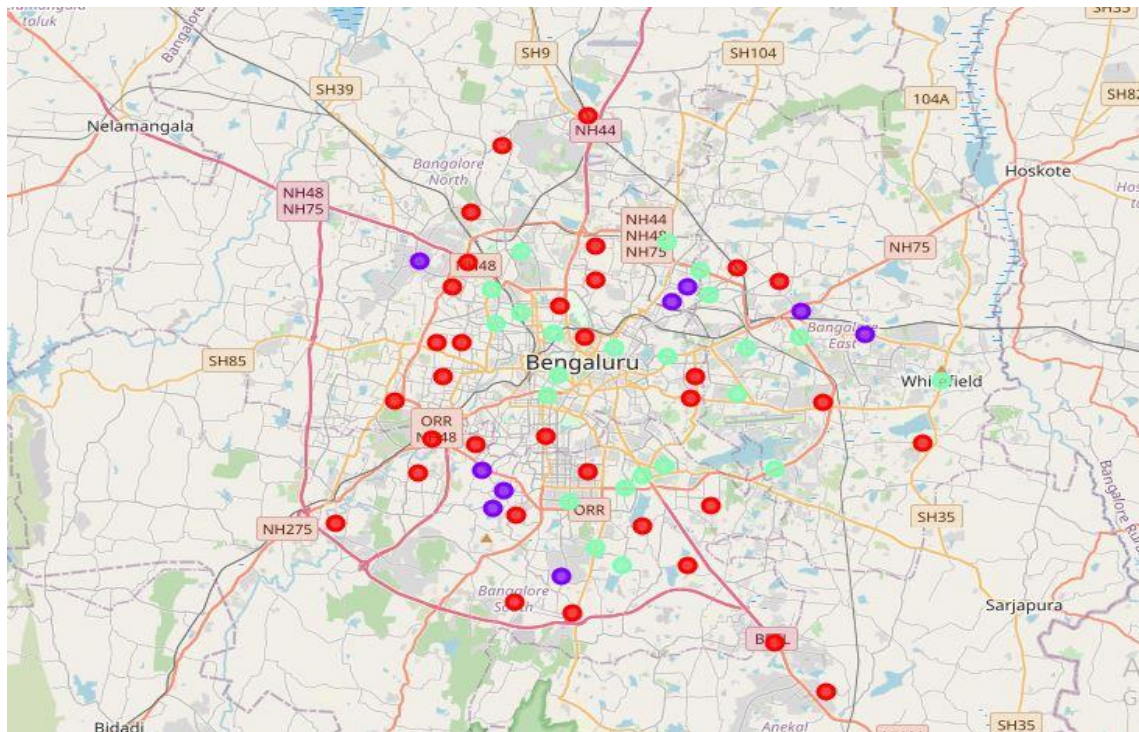
unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “Shopping Mall”. The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

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

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



Discussion:

As observations noted from the map in the Results section, most of the shopping malls are concentrated in the Central and Eastern area of Bangalore city, with the highest number in cluster 1 and moderate number in cluster 2. On the other hand, cluster 0 has very low number to no shopping mall in the neighbourhoods. 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. The results also show that the oversupply of shopping malls mostly happened in the central area of the city, with the Northern and Western area having very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 0 with little to no competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

We only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall which can be used for further research.

Conclusion:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new shopping mall. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.