

IBM Data Science Capstone

Analyzing the Neighborhoods for Opening a Shopping Mall in Bengaluru, India

Sathvik Prabhu

July 2021



Introduction:

The urban population loves visiting shopping malls as it is a great way to relax and enjoy themselves during weekends and holidays. They can go grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies, and engage in arcade games and many more. Shopping malls are like a one-stop destination for all shopping needs. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. Opening shopping malls allows property developers to earn consistent rental income.

As a result, there are many shopping malls in the city of Bengaluru and many more are being built. Bengaluru also known as "*The Silicon Valley of India*" is the IT hub of India . It is the second fastest-growing major metropolis in India. Bangalore is a vibrant city which is always up and alive with its streets packed with people from all backgrounds. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem:

The objective of this capstone project is to analyse and select the best locations in the city of Bengaluru, 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 question : Which is the best location to open a shopping mall in Bengaluru, India?

Target Audience of this project:

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in Bengaluru , India.

Data Collection:

The data required for this project is as follows has been collected from multiple sources. The following data is required for the project:

1. Neighborhood data of Bengaluru.
2. Geographical coordinates of Bengaluru and all neighborhoods in Bengaluru.
3. Venue data for neighborhoods in Bengaluru.

Neighborhood Data :

The data of the neighborhoods in Bengaluru was scraped from https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore . The data is read into a pandas data frame using the `read_html()` method. This is done so that the Wikipedia page provides a comprehensive and detailed table of the data which can easily be scraped using the `read_html()` method of pandas.

Geographical Coordinates :

The geographical coordinates for Bengaluru, has been obtained from the GeoPy library in python. This data is relevant for plotting the map of Bengaluru using the Folium library in python. The geocoder library in python has been used to obtain latitude and longitude data for various neighborhoods in Bengaluru. The coordinates of all neighborhoods in Bengaluru are used to check the accuracy of coordinates given on Wikipedia and replace them in our data frame if the absolute difference is more than 0.001. These coordinates are then further used for plotting using the Folium library in python.

Venue Data:

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Bengaluru and is used to study the popular venues of different neighborhoods. We are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many 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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.