New Shopping Mall Bengaluru, India

- Nithin Balaji

IBM - Applied Data Science Capstone

Capstone Project

May 2020

TABLE OF CONTENTS

1.0 INTRODUCTION

The city of Bengaluru, in Karnataka is called as the Silicon Valley of India. This city is flooded with IT and Tech companies which boosts the GDP per capita of the city marking its place at fifth position in India [1]. Therefore, the buying potential of individuals is growing rapidly in this city.

For individuals, visiting shopping malls are like a one-stop destination for satisfying all their needs such as grocery shopping, dine at restaurants, fashion outlets, play games and watch movies. Retailers and Property developers are taking advantage of this trend to build more shopping malls to balance and meet the demand and supply.

As a result, there are many shopping malls in Bengaluru and many under construction. Yet, the demand of the consumers is not meet. Therefore, property developers are willing to construct more shopping malls in this city as this gives them a constant income. So now the property developers are left with a simple question - "Where should we construct a mall?". The is an important business decision to be taken as the location of the property decides its success or a failure.

2.0 TARGET AUDIENCE

The target audience of this project is anyone who interested in constructing a shop mall but are facing difficulties in which neighbourhood the mall to is to be constructed to attract maximum customers.

3.0 DATA COLLECTION

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

• The geographical coordinates are collected using Python Geocoder package

• The details for the mall location is scrapped from the FourSquare developers website.

The data required for this project is scrapped the the following sources (links),

1.https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore

2.https://foursquare.com/explore?ll=10.466705%2C77.533295&mode=url&near=Bangalore%2C%20Karn%C4%81taka&q=Shopping%20Mall

the following attributes are collected from this website,

    2.1 Neighbourhood

    2.2 Latitude

    2.3 Longitude

    2.4 Venue Name

    2.5 Venue Latitude

    2.6 Venue Longitude

    2.6 Venue Category

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 Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods.

4.0 METHODOLOGY

For the Capstone project I have analysed and selected the best location for construction of a mall in the Bengaluru city, Karnataka. This being a Data science project the methodology followed is,

1. Understanding the business problem

2. Data collection

3. Data cleaning

4. Modelling

5. Data Visualization

6. Business insights

7. Data Driven decision

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 neighbourhood and examine how many unique categories can be curated from all the returned venues.

Then, we will analyse each neighbourhood by grouping the rows by neighbourhood 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 analysing 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.

5.0 RESULTS

The results from the k-means clustering show that we categorize the neighbourhoods into 3

Clusters are based on the frequency of occurrence for "Shopping Mall":

- ➢   Cluster 0 [Red]: Areas with moderate number of shopping malls
- ➢   Cluster 1 [Purple]: Areas with low number to no existence of shopping malls
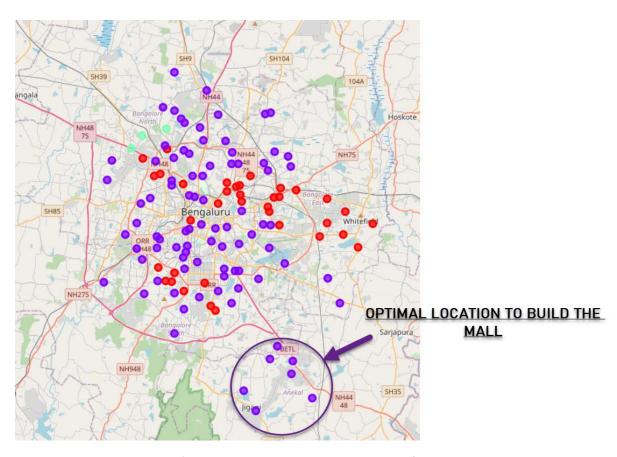- ➢   Cluster 2 [Green]: Areas with high concentration of shopping malls

Fig 1 – Map showing clusters (k - means Clustering)

6.0 CONCLUSION

1. From the map it can be seen that the Eastern and the North-Western parts of the city is flooded with malls

2. The count of malls in the southern part of the city is relatively less.

3. This part of Bengaluru being newly developed is the optimal location for building a new Shopping Mall

4. The optimal areas include,
   4.1 Electronic city phase - 1
   4.2 Electronic city phase - 2
   4.3 Anekal
   4.4 Jigani
   4.5 Surya Nagar phase - 1
   4.6 Bommasandra

This represents a great opportunity and high potential areas to open new shopping malls as there is no competition.

7.0 LIMITATIONS AND FUTURE SCOPE

- ➢ In this project the shopping malls data is taken from Foursquare website which is not up to date. Therefore, the model predicted is not 100% accurate.
- ➢ In future along with the geographical distributions we can also add the population attribute to the model.

REFERENCE

[1] https://en.wikipedia.org/wiki/List_of_Indian_cities_by_GDP_per_capita