# New Shopping Mall Bengaluru, India

IBM - Applied Data Science Capstone

May 2020

---

- NITHIN BALAJI

# 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.

# 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.

# 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.

# DATA ATTRIBUTES

- Neighbourhood
- Latitude
- Longitude
- Venue Name
- Venue Latitude
- Venue Longitude
- Venue Category

# METHODOLOGY

1. Understanding the business problem
2. Data collection
3. Data cleaning
4. Modelling
5. Data Visualization
6. Business insights
7. Data Driven decision

# K – Means Clustering

- The k-means clustering algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**.

- The way kmeans algorithm works is as follows:
  - Specify number of clusters K.
  - Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
  - Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

# RESULTS

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

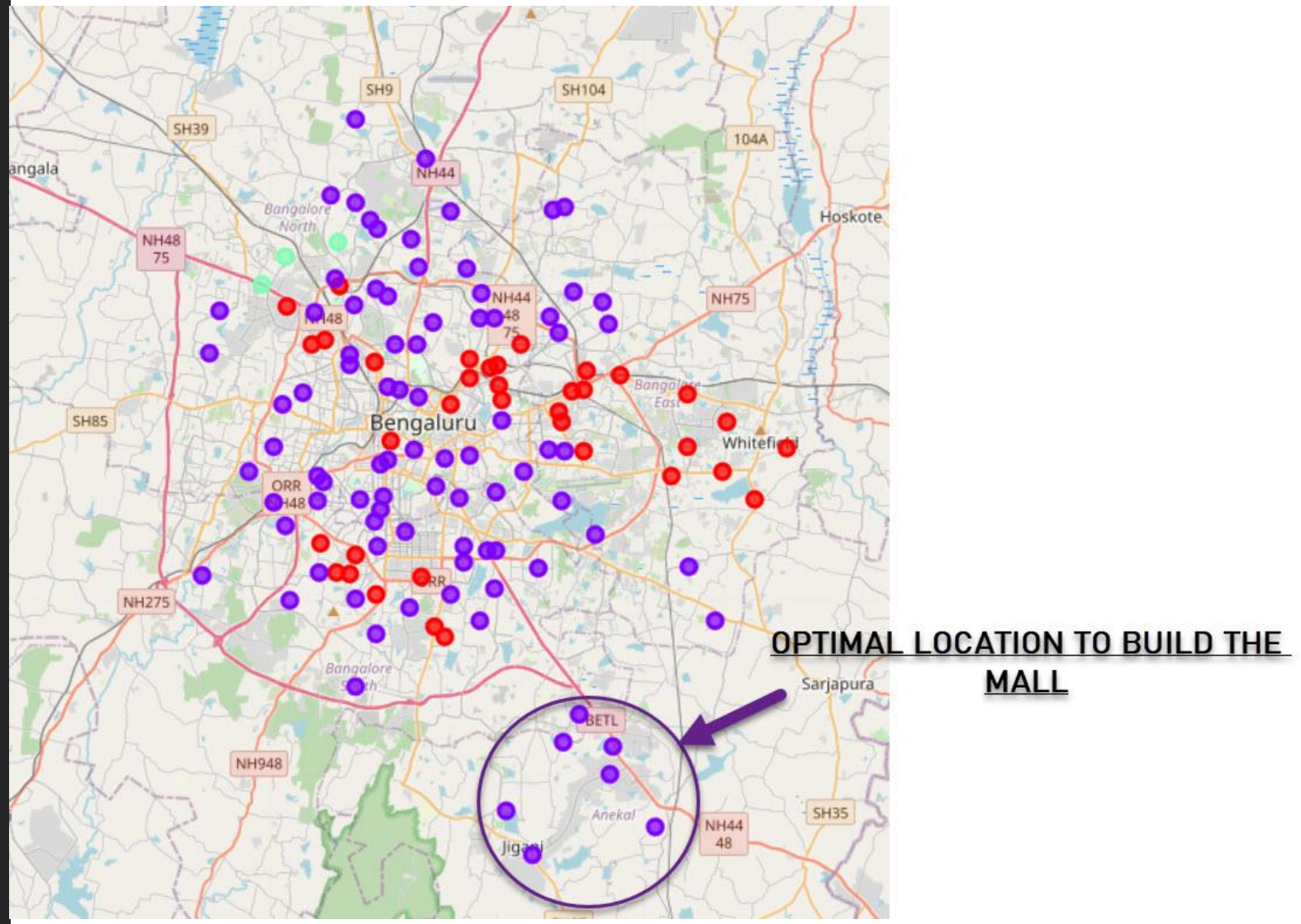


*Fig 1 – Map showing clusters (k - means Clustering)*

# 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.