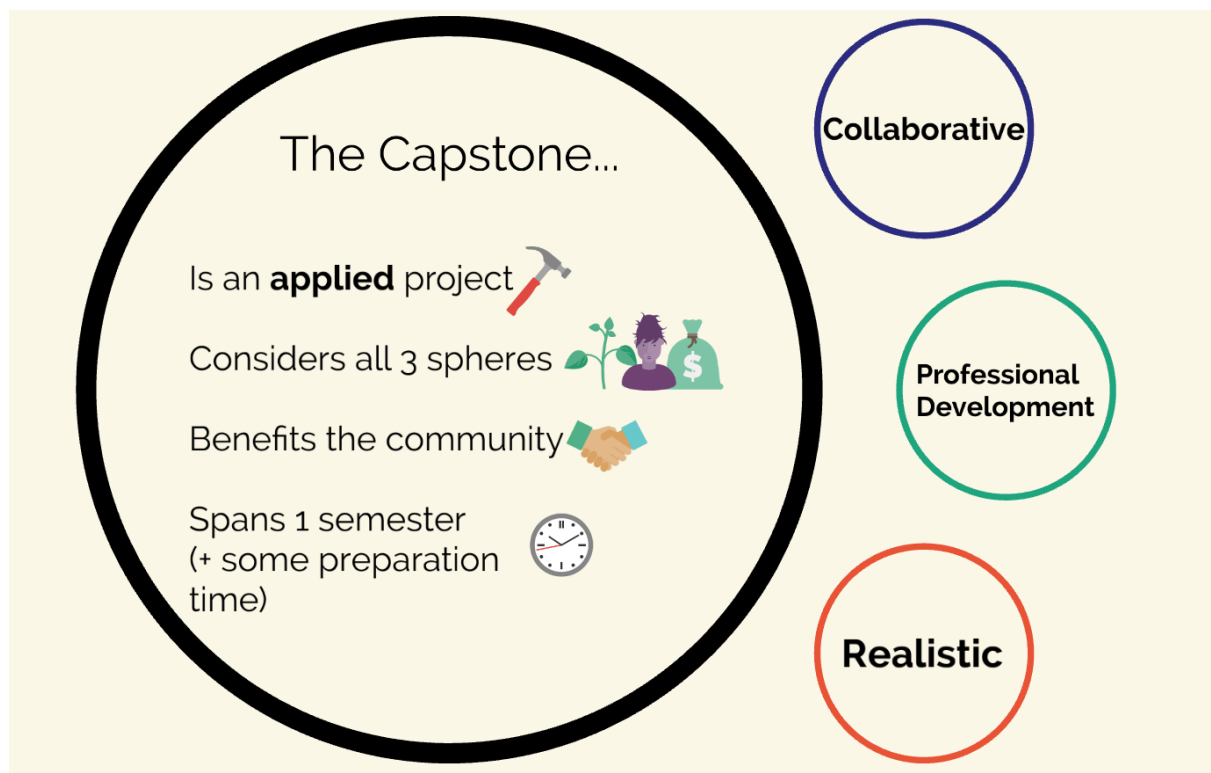# Opening new Shopping Mall

Coursera Capstone

Archit Jhingan

Jan 2020

Introduction

In a growing city such as Bangalore, many people go to mall and shopping complexes to find do shopping, watch movies and various other activities. New demand of malls is increasing day by day. This project will suggest the various areas in the city which are good for making shopping malls. New players in market want to know the location which will suit the best for opening a shopping mall in the area. Which area will suit them the best will be analyzed in the project?

Business Problem

Best area for opening a new mall will be identified in the city of Bangalore. Using techniques of clustering we will identify the various places that are best for property builder to open a mall. It will be beneficial for new players in market who wants to open a shopping complex in the area.
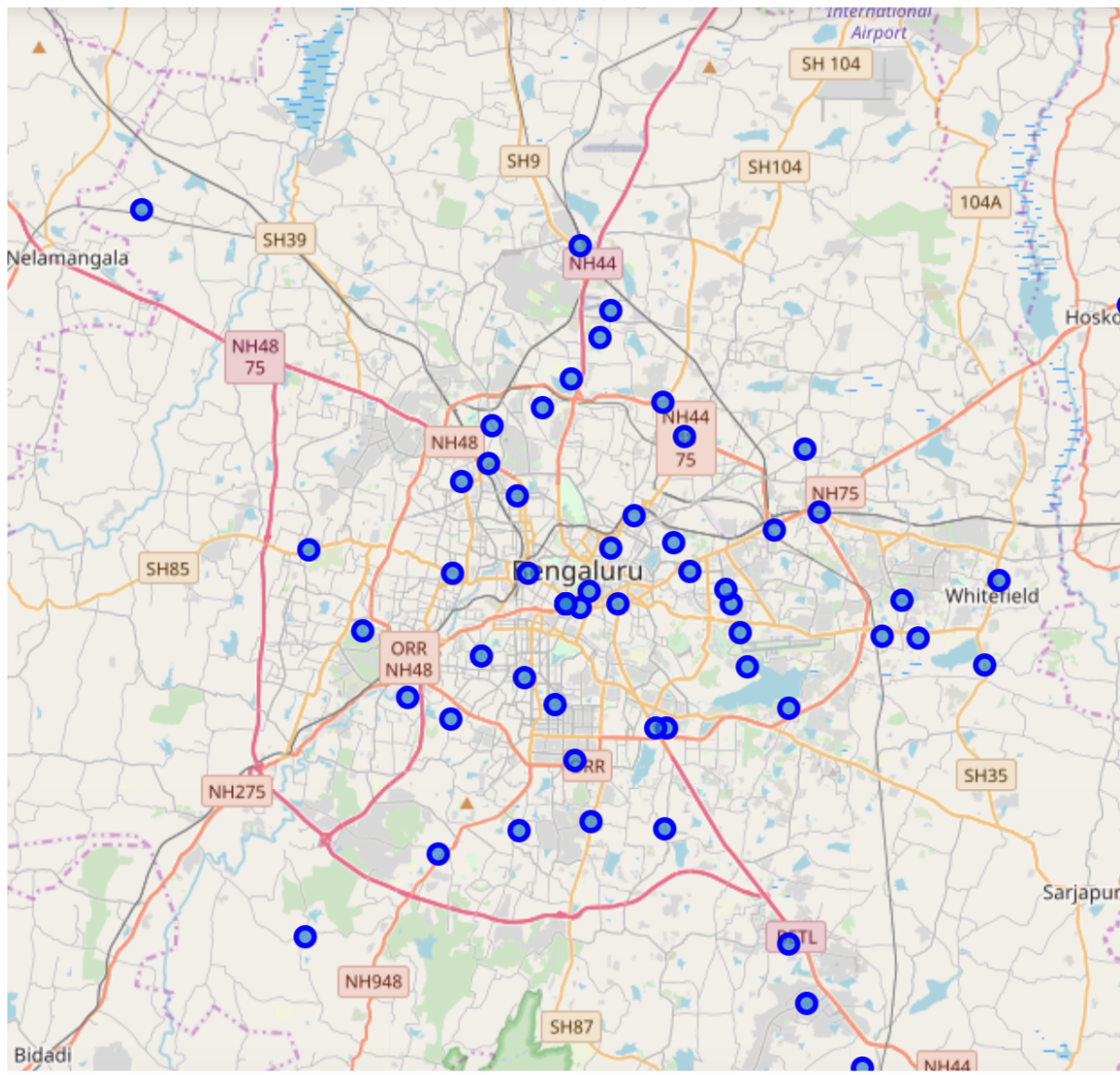
Data Collection

Data will be picked up from https://commons.wikimedia.org/wiki/Category:Suburbs_of_Bangalore which will have all then neighborhoods of Bangalore. In order to get the latitude and longitude of the data those neighborhood will be using geocoder. This will give us all the latitude and longitude of that area. Venue details of that particular area using foursquare API will give us all the shopping malls present in that area. We can use that data to work on our analyses.

Methodology

So for this project we need to get the different location in the city of Bangalore that are required for setting up a new mall. For this we need data regarding neighborhoods of Bangalore. We collected the data regarding all the neighborhood of Bangalore from the website : "https://commons.wikimedia.org/wiki/Category:Suburbs_of_Bangalore". After collecting the list the data will then be cleaned using regular expression all the unwanted data is removed using regular expression.

Now we got the neighborhood of Bangalore, but it won't be enough to get the info related to venues at the location. For that we will be using geocoder that will convert neighborhood location in latitude and longitude.

After getting details of latitude and longitude we will be using the data to get the details of different venues or mall located at different neighborhood of Bangalore. All the venues details will be collected and mapped with the current data set version

After getting details of all the venues data we will be only concentrating on shopping malls , thereby filtering out only shopping mall from the data collected.

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.

Result

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 of shopping malls

- Cluster 1: Neighbourhoods with low number to no existence of shopping malls

- Cluster 2: Neighbourhoods with high concentration of shopping malls
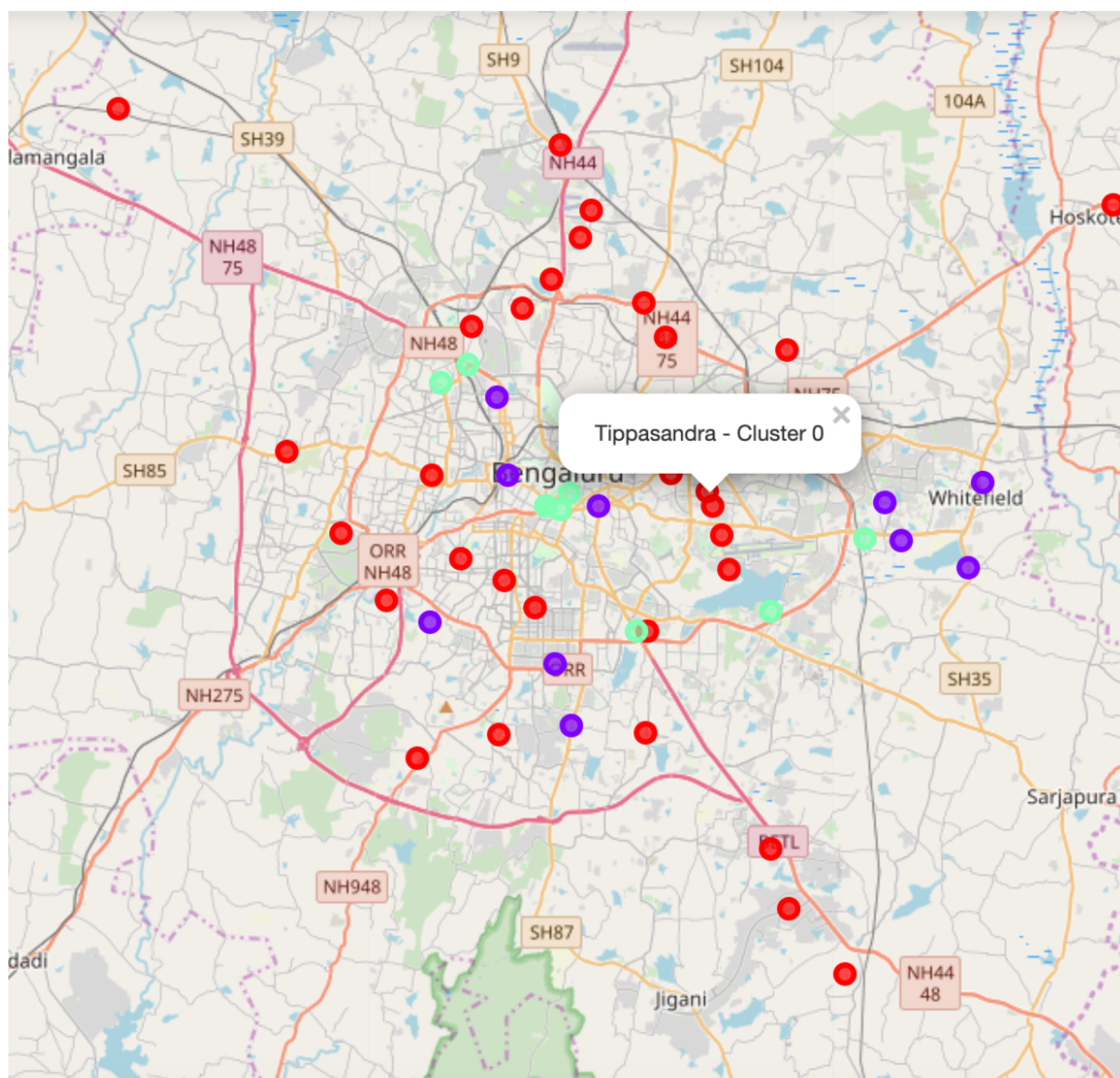
Cluster 0: Red
Cluster 1: Purple
Cluster 2: Green

Observation

It is suggested to open mall in area with less or no malls. As there will be no business competition it will be better opportunity for the business to run in less or no shopping mall areas.ß