The Battle of Neighborhoods – Opening new shopping mall in Bangalore, India.

Rupali Panda

(IBM Data Science Capstone Project)

Introduction

- Shopping malls provide a wider variety of good and services.
- An important part of modern city life
- Bangalore city, with a population of 10 million, has an increasing demand for shopping malls.

Business Problem

- Developers must review and strategically target a suitable location for a new shopping mall to maximize profit.
- Analysis to select the best locations in the neighbourhoods of Bangalore
- Data science methodology & machine learning techniques (clustering) used.

Target Audience

- Property developers and investors
- Businesses /store owners
- Local people (consumers)

Data and Methodology

- List of all Bangalore neighbourhoods: Webscrapping from Wikipedia (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore)
- Latitude and Longitude data: Geocoder package.
- Venue data: Foursquare API (top 100 venues in a 1000 m radius).
- Data extraction: beautifulsoup package.
- k-means clustering: Unsupervised clustering with 5 centroids, based on the concentration of shopping malls

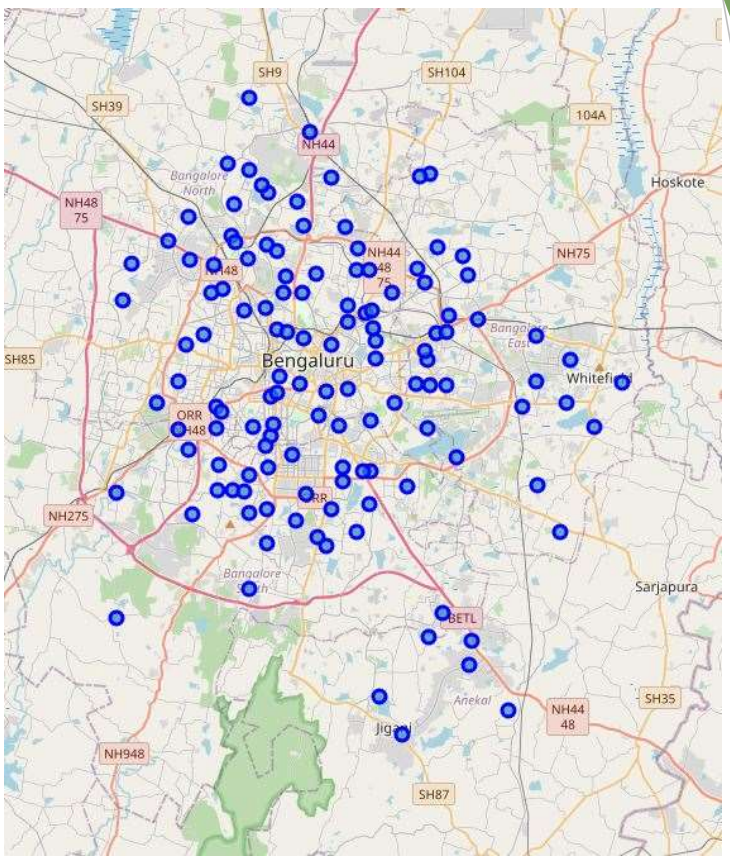

Fig. 1: Neighbourhoods of metropolitan Bangalore city.

Results

- 5 clusters of neighbourhoods by k-means clustering
- Cluster 1 (Red): lowest concentration of shopping malls
- Cluster 2 (Purple): low to moderate concentration of shopping malls
- Cluster 3 (Blue): high concentration of shopping malls
- Cluster 4 (Green): 2nd highest concentration of shopping malls
- Cluster 5 (Orange): highest number of shopping malls

Discussion

- Most of the shopping malls are in the central area of Bangalore (cluster 4 and 5) - developers should avoid these neighbourhoods
- cluster 1: very low number of shopping malls, suitable for opening new shopping malls.
- This project considers only one aspect (i.e., concentration of shopping malls) of the neighbourhoods
- Other factors such as demographic data and real estate value affect the decision as well.
- Advanced user account of the Foursquare API (instead of the free Sandbox Tier account used in this report) may yield better results.
- Different clustering algorithms could also be used to compare the results.

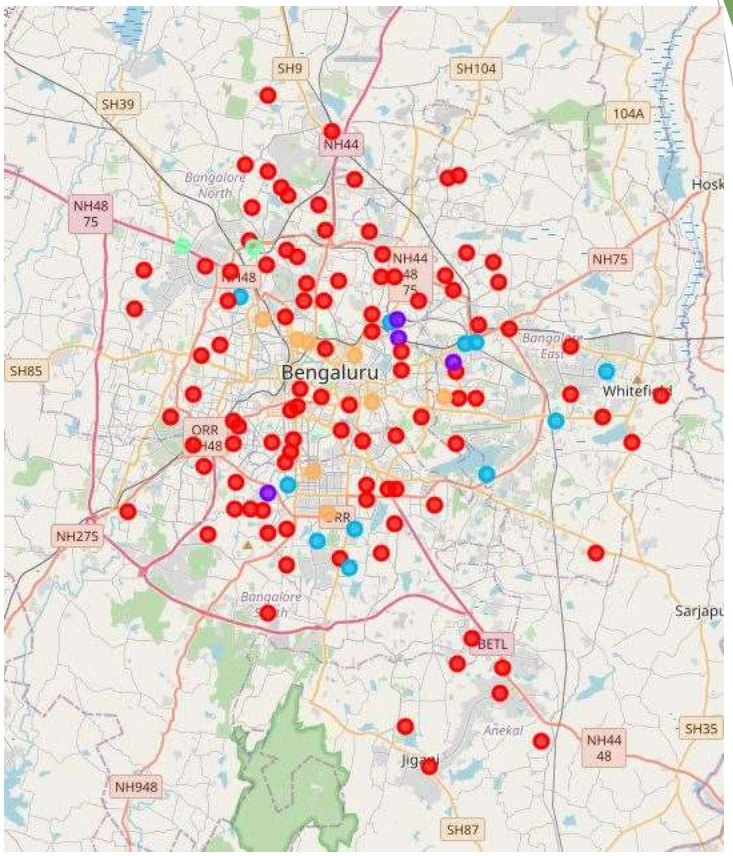

Fig. 2: Clustering of Shopping Mall categories in Bangalore

THANK YOU !