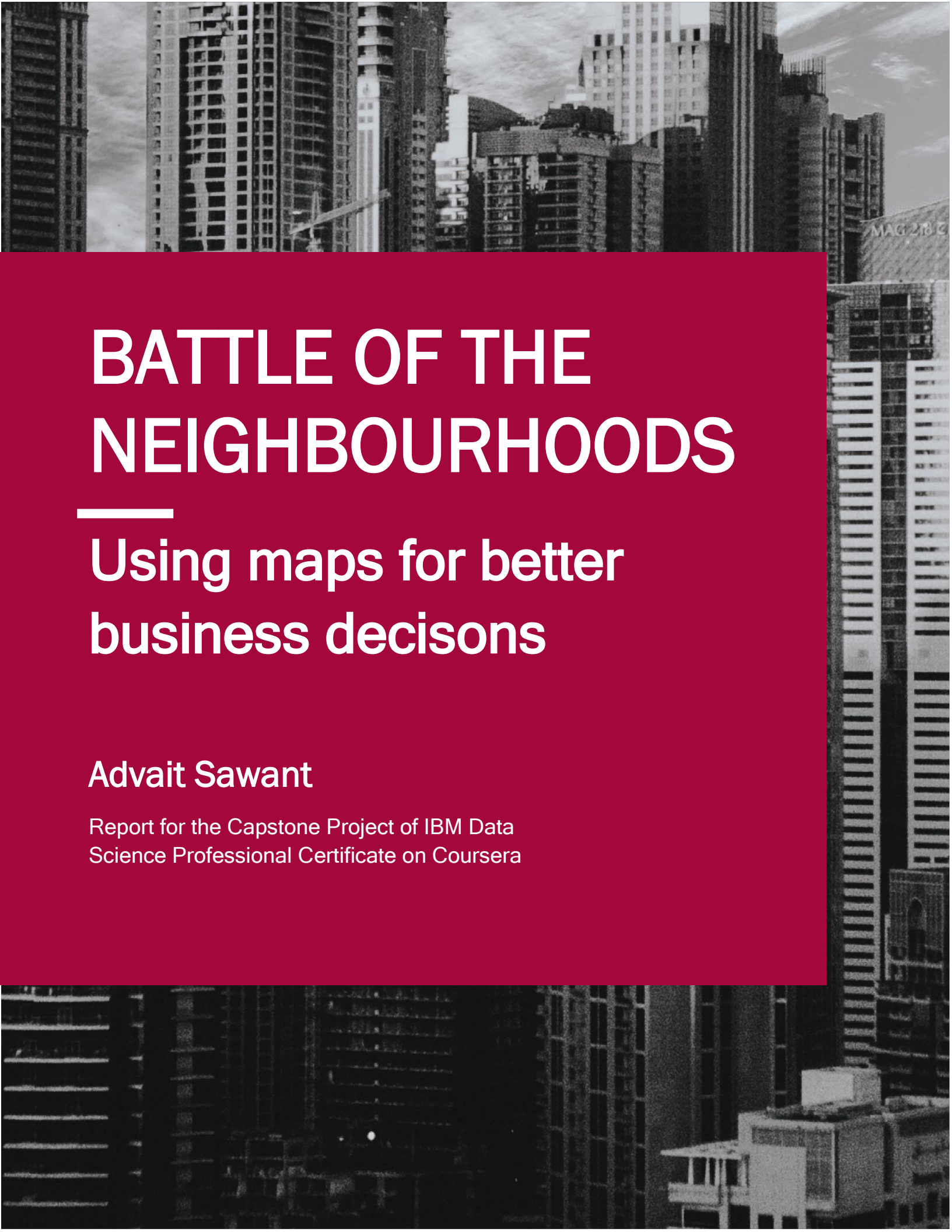


BATTLE OF THE NEIGHBOURHOODS

Using maps for better business decisons

Advait Sawant

Report for the Capstone Project of IBM Data Science Professional Certificate on Coursera

Introduction to the opportunity

Bangalore, officially called Bengaluru, is the capital of Karnataka. It is the second fastest growing metropolis of India and also the fourth highest GDP contributor after cities like Mumbai, Delhi, and Chennai. Bangalore is popularly known as the ‘Silicon Valley’ of India for being a major IT hub of the nation.

Due to the opportunities available here, it attracts people from all over the nation herein search of jobs. This growth has also created opportunities for many entrepreneurs to set up shop here. Our project looks from the perspective of an entrepreneur who wants to open a coffee shop in Bangalore.

Business Problem

Location is one of the most important factors of a business. Whether it is a Real Estate company or a simple coffee shop, location can make or break a business.

The attractiveness of a location for doing business itself depends upon a number of factors like Quality of Living (QoL), Accessibility, Business Competitors, Crime Rates and even Taxes.

In this project we will be focusing on one factor: Business Competitors.

In markets of low-cost goods such as coffee, competition has a huge impact on profits. Areas with low competition can lead to higher footfall at the business and thus more profits.

Target Audience

This project is made for entrepreneurs who want to open a coffee shop in Bangalore and also coffee connoisseurs who want variety.

However, with minimal modifications, this project can be utilized by any entrepreneur who wants to find a suitable location for his business. It can also be used by consumers to find places the most suited for their purposes.

Data Description

For this project, we will require 3 types of data:

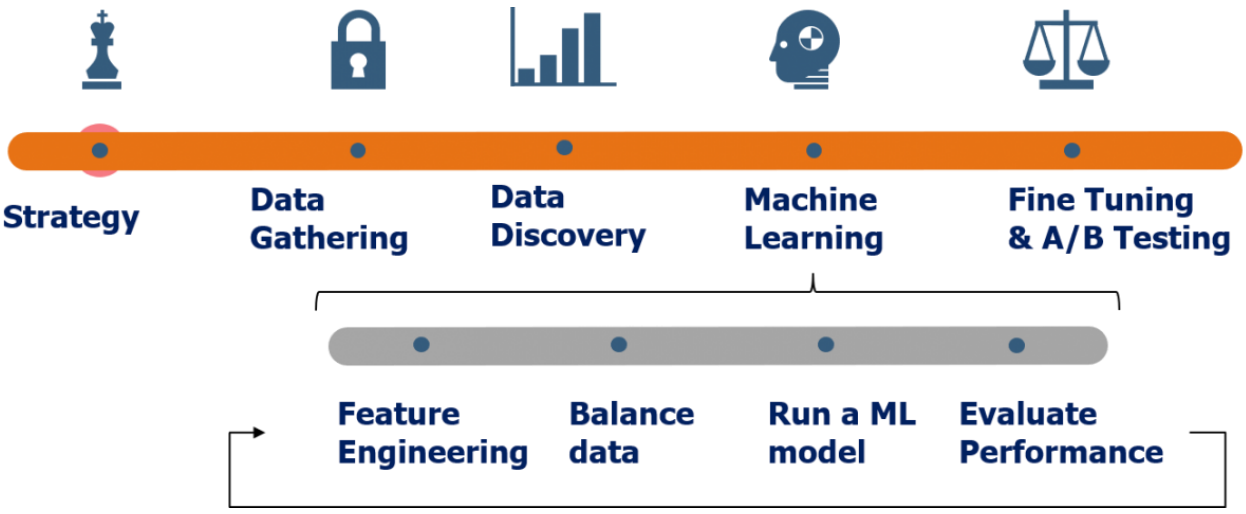
1. Pin code Data of Bangalore.
2. Location coordinates (Latitude & Longitude) of the pin codes.
3. Venue data for venues around the location coordinates.

Pin code Data: We will obtain the required pin code data from https://www.onlinebangalore.com/guide/pincodes/pincode.html and scrape it into a Pandas Data-frame.

Location Coordinates: We will use the pin code data that we have scraped and feed it into the GeoSpace module. We will get a data-frame that contains latitude and longitude coordinates. (https://github.com/geospace-code/pymap3d)

Venue Data: We will get the venue data from the Foursquare API that we are using. Since we are using a free-tier Sandbox account, we will have to restrict the number of queries.

Methodology

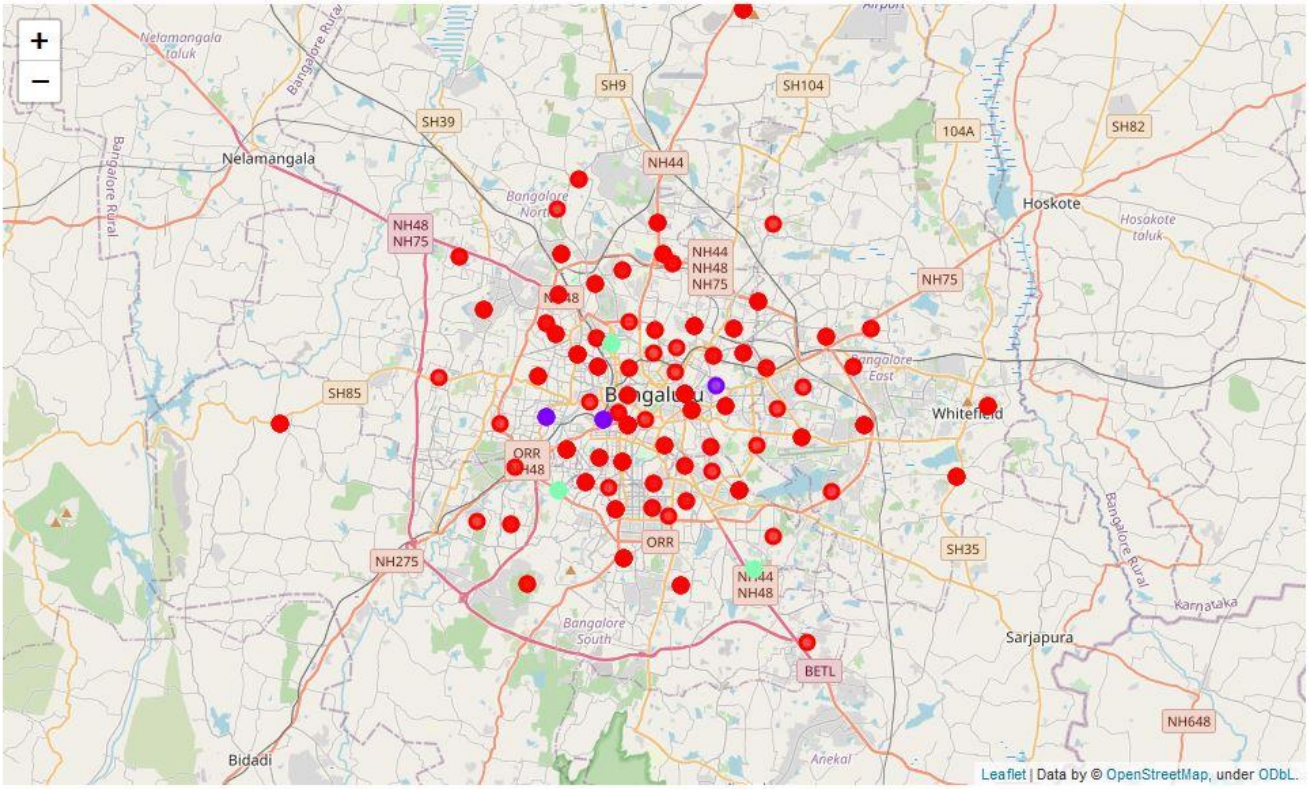


The methodology followed in this project is as follows:

1) Installing & Importing Python libraries and dependencies.
2) Scraping pin code data from a webpage into a Pandas Data-frame.
   a) Data pre-processing.
   b) Output as csv file (.csv).
3) Making a map of Bangalore.
   a) Obtaining geographical coordinates for the pin codes.
   b) Making a map of the different pin codes.
4) Using Foursquare to get venue data.
   a) Setting up Foursquare account credentials.
   b) Listing the top 100 venues within a radius of 500 metres from each of the coordinates.
   c) Pre-processing Data.
   d) Output as csv file (.csv).
5) Feature engineering for the selected business Problem
   a) Simplification.
   b) Feature Selection.
   c) Handling categorical data through one-hot encoding.
6) Clustering and Cluster Visualization.
7) Examining Clusters.
8) Making observations.

Results

We divided the locations into 3 clusters using the KMeans clustering module from Scikit-Learn depending upon the number of coffee shops. The results are as follows:



Cluster 0 (Red colour): Cluster 0 has very few to no coffee shops.

Cluster 1 (Violet colour): Cluster 1 has a relatively high density of coffee shops.

Cluster 2 (Mint colour): Cluster 2 has a moderate density of coffee shops.

Recommendations

Most of the Coffee Shops are concentrated in the a few parts of Bangalore city, with the highest number in cluster 1 and a moderate number in cluster 2.

On the other hand, cluster 0 has very few to no coffee shops in the neighbourhoods.

This represents a great opportunity and high potential areas to open new Coffee Shops as there is very little to no competition from existing coffee shops.

Meanwhile, coffee shops in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of coffee Shops.

From another perspective, this also shows that the oversupply of coffee shops mostly happened in only a few parts of the city, and most of the city is open for business.

Therefore, this project recommends entrepreneurs to capitalize on these findings to open new coffee shops in neighbourhoods in cluster 0 with little to no competition.

Entrepreneurs with unique selling propositions can stand out from the competition and open new coffee shops in neighbourhoods in cluster 2 with moderate competition.

Lastly, entrepreneurs are advised to avoid neighbourhoods in cluster 1 which already have a high concentration of coffee shops and are likely suffering from intense competition.

Limitations

In this project, we have considered only one variable: competition. In reality there are many other factors which can make or break a business. Multiple other factors need to be considered to make a business decision like crime rate, tax rates, ease of availing credit, socio-economic condition of the target population.

There are also other limitations that we have faced in the project due to using a free sandbox account, which has limited functionality. In a large metropolis like Bangalore, the number of venues can easily exceed 100 and lead to missing out of some competitors. This can skew the results that we have found.

Future Work

If this project is taken up in the future, it should attempt to mitigate the limitations mentioned above. Using a premium Foursquare account should allow us to mitigate the problems associated with lower data resolution. Other factors affecting business should also be considered.

Other clustering methods like DBSCAN should also be used for more accurate results.

Conclusions

In this project we have extracted data from websites through webscraping, cleaned the data using Pandas, called Foursquare APIs, clustered data using SciKit-learn and created maps using Folium.

We have found out through clustering that Cluster 0 has very low competition, Cluster 1 has a large number of competitors and Cluster 2 has a moderate number of competitors. We have found that cluster 0 is the most lucrative cluster due to low competition and will recommend any entrepreneurs to open their coffee shops in cluster 0. With differentiation, a coffee shop in cluster 2 can also be viable.

References

1. Pin codes of Bangalore: https://www.mapsofindia.com/pincode/india/karnataka/bangalore/
2. GeoSpace: https://github.com/geospace-code/pymap3d
3. Foursquare: https://foursquare.com/