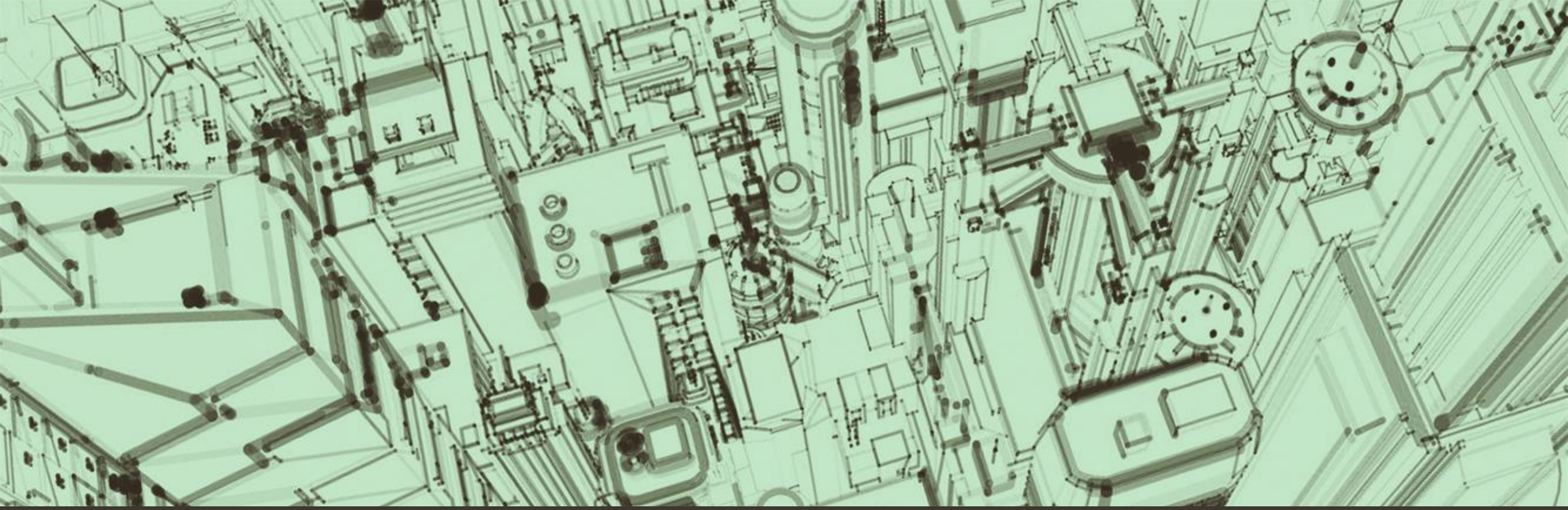


BATTLE OF THE NEIGHBOURHOODS

Using maps for better business decisions

By Advait Sawant (10/02/2021)

Business Problem

- Location is a very important factor which can make or break a business.
- Viability of a business depends upon a number of factors like competitors, tax rates, accessibility, crime rates and Quality of Life (QoL).
- For this project, we will be focusing on one factor: Competitors.
- More competitors → Slimmer margins → Less Profits.
- Fewer competitors → Larger margins → More Profits.

- Business Question: Where do I as an entrepreneur set up a coffee shop in Bangalore to maximize my profits?

Data Used

- Pin Code Data: Pin codes were scraped from https://www.onlinebangalore.com/guide/pincodes/pincode.html and loaded into a Pandas Data-Frame.
- Location Coordinates: Pin codes were fed into the GeoSpace module to get the location coordinates (Latitude & Longitude). https://github.com/geospace-code/pymap3d
- Venues Data: Venue data was obtained from Foursquare by using their Places API. Location coordinates were fed into the API to obtain the details of nearby venues.

Methodology

- Installing & Importing Python libraries and dependencies.
- Scraping pin code data from a webpage into a Pandas Data-frame.
  - a) Data pre-processing.
  - b) Output as csv file (.csv).
- Making a map of Bangalore.
  - a) Obtaining geographical coordinates for the pin codes.
  - b) Making a map of the different pin codes.
- Using Foursquare to get venue data.
  - a) Setting up Foursquare account credentials.
  - b) Listing the top 100 venues within a radius of 500 meters from each of the coordinates.
  - c) Pre-processing Data.
  - d) Output as csv file (.csv).

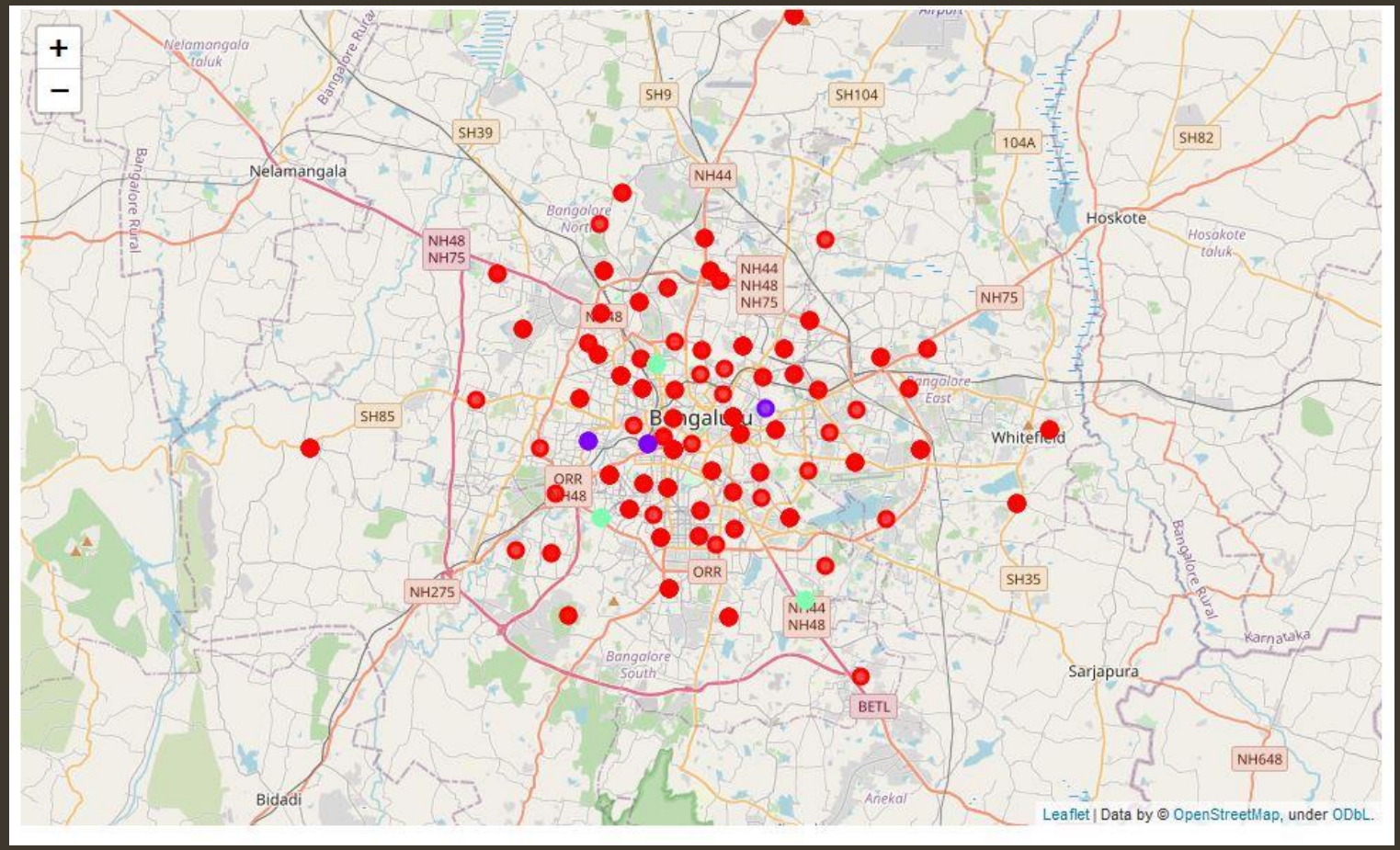
Methodology

- Feature engineering for the selected business Problem
  - a) Simplification.
  - b) Feature Selection.
  - c) Handling categorical data through one-hot encoding.
- Clustering and Cluster Visualization.
- Making a map of Bangalore.
  - a) Obtaining geographical coordinates for the pin codes.
  - b) Making a map of the different pin codes.
- Examining Clusters.
- Making observations.

Results

Data divided into 3 clusters using KMeans clustering.

- Cluster 0 (Red colour): Low density of coffee shops
- Cluster 1 (Violet colour): High density of coffee shops.
- Cluster 2 (Mint colour): 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.

- Cluster 0 has very few to no coffee shops in the neighborhoods. Great opportunity and high number of potential areas to open new Coffee Shops as there is very little to no competition from existing coffee shops.

- Coffee shops in cluster 1 likely suffering from intense competition due to high competition.
- Oversupply of coffee shops in only a few parts of the city, most of the city is open for business.

Recommendations

- Therefore, this project recommends entrepreneurs to capitalize on these findings to open new coffee shops in neighborhoods 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 neighborhoods in cluster 2 with moderate competition.
- Lastly, entrepreneurs are advised to avoid neighborhoods in cluster 1 which already have a high concentration of coffee shops and are likely suffering from intense competition.

Limitations

- We have considered only one variable: Competition. Other variables need to be considered before making business decisions.
- Data used in this is of low resolutions due to limits on Foursquare API calls placed on the Sandbox account.
- K-Means clustering isn't the most accurate clustering approach.

Future Work

- Use of more variables to make business decisions and recommendations.
- Use of a paid Foursquare account to improve data resolution.
- Use of other clustering approaches like DBSCAN.

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.

- Cluster 0 has very little competition, Cluster 1 has very high competition and Cluster 2 has moderate competition
- Cluster 0 seems most profitable to open a coffee shop, cluster 2 can be viable with differentiation.

References

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

Thank You!