

Coursera Capstone

Set up Café in Bangalore

IBM DATA SCIENCE CERTIFICATION

Amitabha Kanjilal
September 2019

Introduction

In the modern days that we live in, cafés have become very popular, and is a great place to hang out with friends or family. Any major city would usually have a bunch of cafés around, and many investors still look out for business opportunities for opening brand-new cafés in town. However, the location is a vital factor for getting profits out of it.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Bangalore, India to open a new café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Bangalore, if an entrepreneur is looking to set up a new café, where would you recommend that they open it?

Data Collection

We need the following data to solve the problem –

- List of neighbourhoods in Bangalore
- Latitude and Longitude for those neighbourhoods
- Venues, especially cafés in 10 km radius of these co-ordinates

We are extracting the neighbourhood data by web-scraping the Wikipedia link for localities in Bangalore. After we get the neighbourhood names, we are searching for the co-ordinates of those neighbourhood in google and again scraping for the latitude and longitude values.

Once we have the co-ordinates, we are using the foursquare API to fetch top 100 venues within a 10 km radius of each neighbourhood.

Methodology

Once we have collected the data as explained above, we will transform the data into another data to check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the Café data, we will filter the Café as venue category for the neighbourhoods.

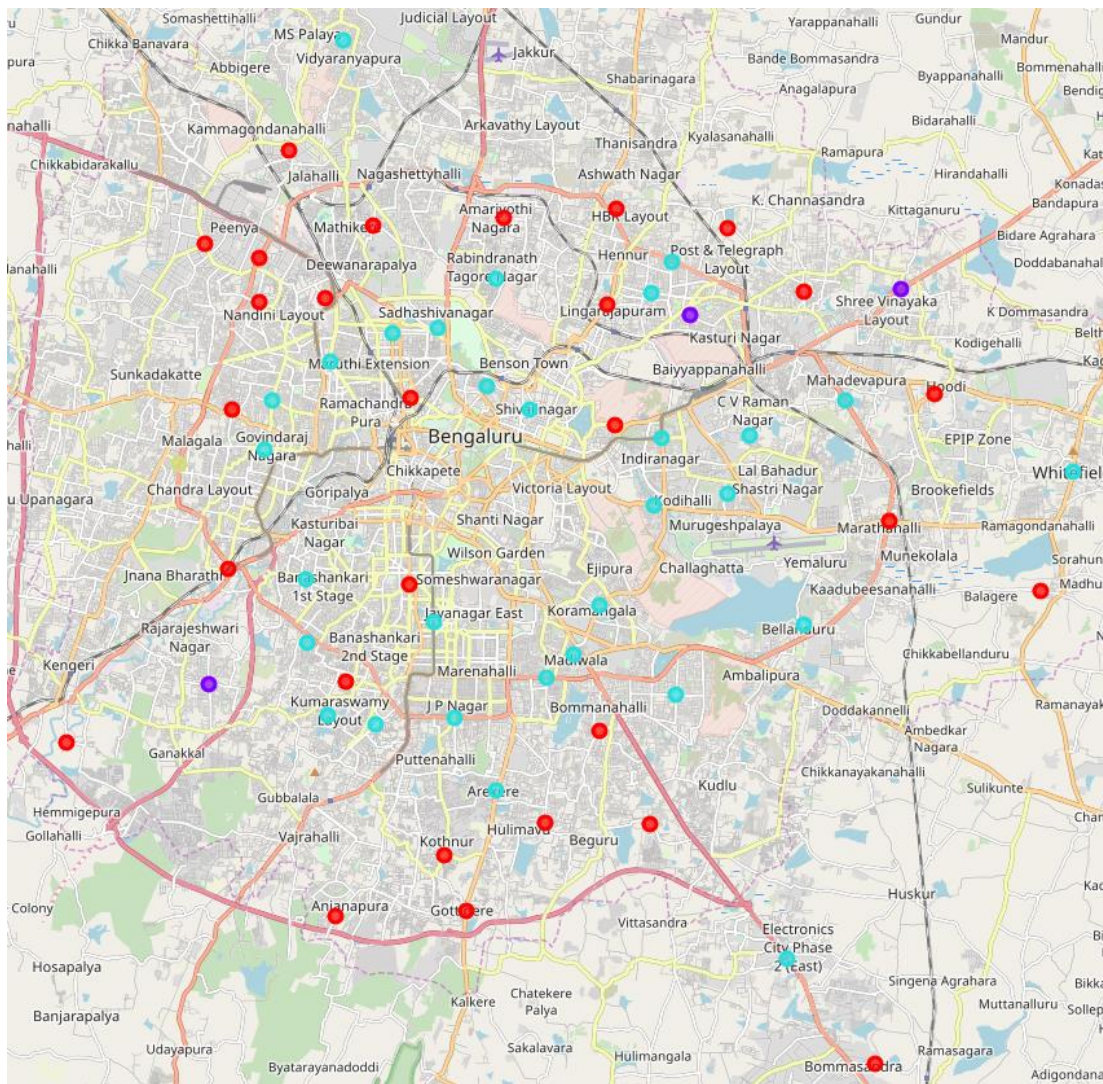
Finally, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping

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 4 clusters based on their frequency of occurrence for Café. The results will allow us to identify which neighbourhoods have higher concentration of cafés while which neighbourhoods have fewer number of cafés. Based on the occurrence of cafés in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new cafés.

Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for Cafés:

- Cluster 0 (Red): Neighbourhoods with low number to no existence of cafés
- Cluster 1 (Purple): Neighbourhoods with moderately high number of cafés
- Cluster 2 (Blue): Neighbourhoods with moderately low concentration of cafés
- Cluster 3 (Yellow): Neighbourhoods with very high concentration of cafés



Discussion

Most of the cafés are concentrated in the central area of Bangalore, most of which belong to cluster 2. Cluster 1 has very less concentration of cafés and are mostly away from the centre of the city. However, looking at the plotted map, we can observe 2 neighbourhoods, which are in the central part of the city but belong to cluster 1. We can deduce that these are 3 neighbourhoods that haven't been commercialized, so this project recommends that Ulsoor, Basavagudi and Sheshadripuram would be an ideal place to set up a new café.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. entrepreneurs regarding the best locations to open a new café. To answer the business question that was raised in the introduction section, the answer proposed by this project is that the neighbourhoods in central Bangalore in cluster 0 are the most preferred locations to open a new café. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new café.

References

Localities in Bangalore:

- https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Foursquare Developers Documentation:

- <https://developer.foursquare.com/docs>

Appendix

- Cluster 0

Anjanapura, Hulimavu, Jalahalli, Kamakshipalya, Kothnur, Lingarajapuram, Mahalakshmi Layout, Marathahalli, Horamavu, Mathikere, Nayandahalli, Padmanabhanagar, Peenya, Ramamurthy Nagar, **Seshadripuram**, **Ulsoor**, Varthur, Nandini Layout, Hoodi, Kengeri, Bommasandra, **Basavanagudi**, Begur, Bommanahalli, Hebbal, Yeshwanthpur, Gottigere, HBR Layout

- Cluster 1

Rajarajeshwari Nagar, Banaswadi, Krishnarajapuram

- Cluster 2

Jeevanbheemanagar, R. T. Nagar, Rajajinagar, Sadashivanagar, Shivajinagar, BTM Layout, Uttarahalli, Arekere, Vasanth Nagar, Vidyaranyapura, Vijayanagar, Banashankari, Basaveshwaranagar, Bellandur, Jayanagar, Kalyan Nagar, Girinagar, Kammanahalli, Whitefield, Koramangala, Electronic City, Indiranagar, Kumaraswamy Layout, Madiwala, Mahadevapura, CV Raman Nagar, Malleswaram, J. P. Nagar, Domlur, HSR Layout

- Cluster 3

Nagarbhavi