Optimal location for opening a new cafe alongside the Bangalore Metro

Muhammed Salih

29 April 2020

1. Introduction

1.1. Background

Bangalore is one of India's biggest cities with a population of over 10 million. It is often called the 'Silicon Valley' of India as it is the hub for India's IT companies.

The Bangalore Metro Rail Corporation (BMRC) operates the Namma Metro (translated as Our Metro in English) which serves as a key player in the city's public transit. The Namma Metro which started operations in 2011 currently has 40 stations and sees a daily ridership in the range of four hundred and fifty thousand. Due to its high ridership, the metro stations and its vicinity have become centres of commercial activity, especially for food and beverage outlets.

1.2. Problem

This project tries to suggest the suitable locations near metro stations for opening new cafes for a client who is planning to expand its operations in the city over the next few years.

1.3. Interest

Our client who owns a cafe business with more than 50 cafes across the country currently, is executing this study as part of its plans to expand its operations in the city of Bangalore.

2. Data Description

2.1. Data Sources

In order to do the analysis, we need the location data of each metro station which can be readily acquired from Wikipedia. Once the location data is ready, the Foursquare API can be used to get a detailed overview of the venues that are currently operational in the vicinity of each metro station. So to summarise, the required data for this project will be acquired from the following sources.

Table 1 - Data Sources

Item	Source
Geographical locations of the metro stations	Wikipedia Pages
Venue data for each stations	Foursquare API

2.2. Data Cleaning and Formatting

Since our dataset will primarily contain the venue data from Foursquare API, missing values or incorrect data will not be a big concern. The data will be stored as a pandas dataframe in Python and appropriate Python packages will be used for EDA, visualisation, and modelling.

	Venue Name	Category	Latitude	Longitude
0	Hotel Fishland	Seafood Restaurant	12.975569	77.578592
1	Udupi Sri Krishna Bhavan	Indian Restaurant	12.971563	77.574158
2	SGS Donne Biriyani	Indian Restaurant	12.970325	77.572648
3	Sangam Sweets	Dessert Shop	12.976924	77.577891
4	Sapna Book House	Bookstore	12.976355	77.578461

Fig 1 - Sample dataframe with the venues near Majestic metro Station

2.3. Approach towards using the data

The Foursquare API can give a detailed insight into the composition of the type of businesses that are currently operating near the metro stations. By using proper transformations like one-hot encoding and other operations, the relative frequency for each business type (say coffee shops, cafes, restaurants, cinema etc.) can be found out for the metro stations.

As far as the aim of the current project is concerned, a neighborhood currently overcrowded with cafes might not offer a good prospect when it comes to opening a new cafe. And at the same time, a neighborhood that has very few food and beverage outlets currently might represent a lower potential for the success of such a venture as well. So our best bet for opening a new cafe would be neighborhoods that already have an active set of food and beverages businesses running currently, but the number of cafes are not very high.

For our analysis, the metro stations would be clustered using the k-means clustering algorithm based on the neighborhood venue makeup as obtained from the Foursquare API data. Once the clustering is done, the characteristic features of the clusters (like the top most occurring type of venue) will be used to finalise the clusters that are good candidates for opening a new cafe for our client.

3. Methodology

3.1. Exploratory Data Analysis

Initially, the data pertaining to the geographical locations of the metro stations was imported as a pandas dataframe and their locations were visualised using the Folium package. The metro stations are divided into two lines (Green Line and Purple Line) which are represented by the marker colours in the generated map.

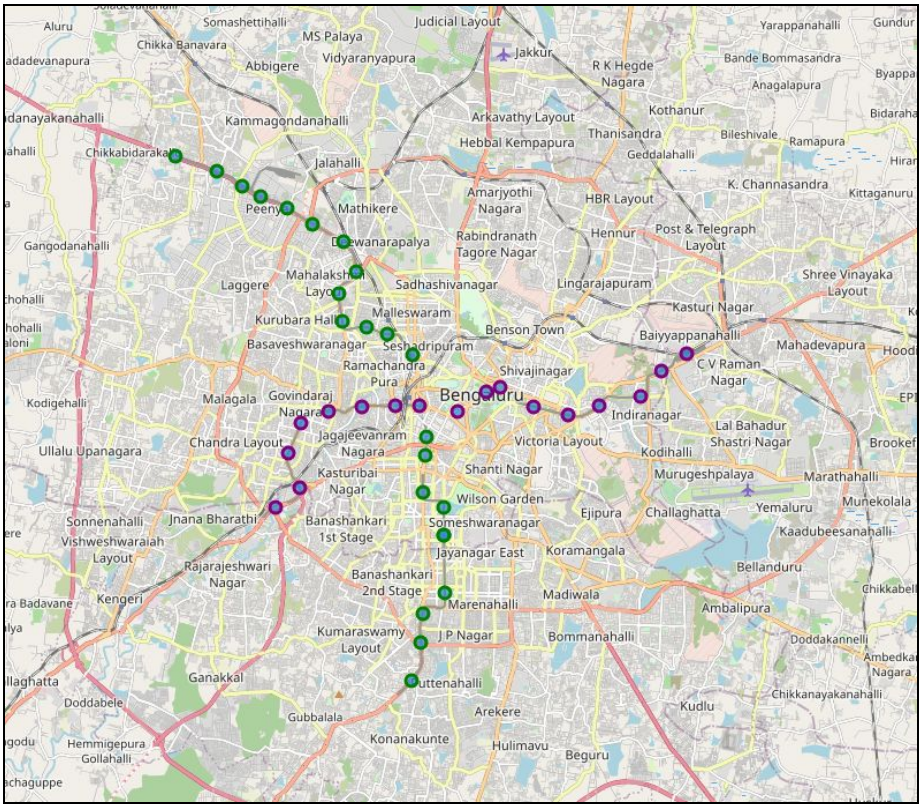

Fig 2 - Locations of the stations in Bangalore Metro

Distribution of the top venue categories across stations in each line

To get a deeper insight into how the composition of venues change across the two lines, a seaborn boxplot was used to plot the variation of the number of venues in each category (for the top 10 venue categories) for each line.

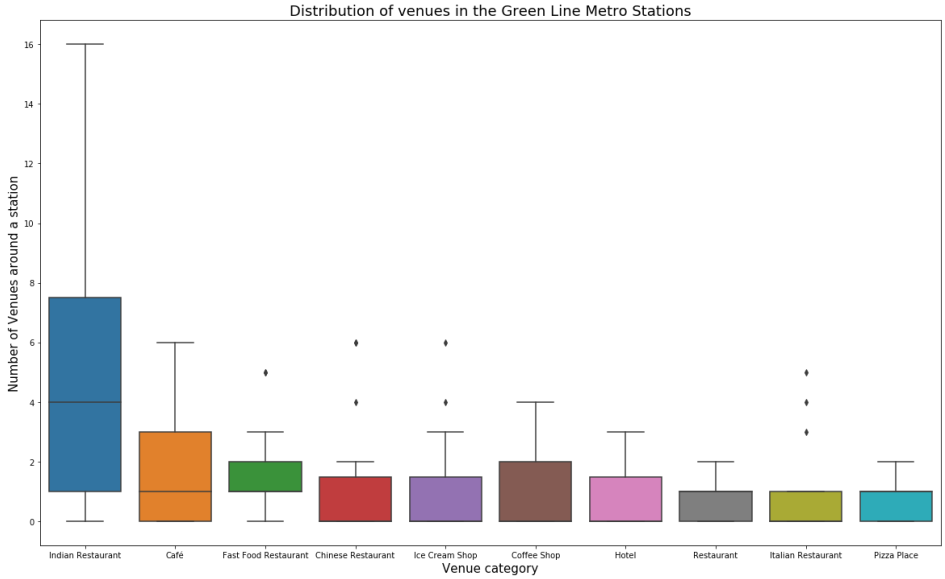

Fig 3 - Distribution of venues in the Green Line Metro Stations

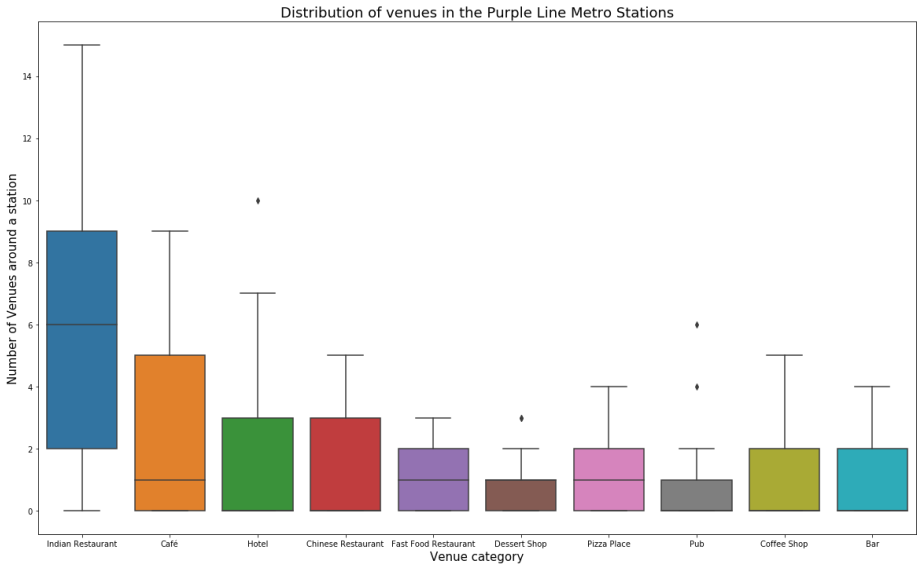

Fig 4 - Distribution of venues in the Purple Line Metro Stations

Following that, each metro station was analysed individually and the top 5 venue categories for each station was listed.

```
----Jayaprakash Nagar----
                  venue  freq
0           Food Truck  0.14
1     Department Store  0.10
2          Pizza Place  0.10
3    Indian Restaurant  0.05
4  Fast Food Restaurant  0.05


----Krishna Rajendra Market----
                          venue  freq
0            Indian Restaurant  0.41
1                       Market  0.06
2                         Park  0.06
3  Paper / Office Supplies Store  0.06
4          Fast Food Restaurant  0.06
```

Fig 5 - Top 5 venue categories for two of the metro stations

3.2. Clustering Algorithm

To execute this project, we need to find patterns among the distribution of venues that are already present in the vicinity of each metro station and then group the metro stations based on their similarities. Such a process can help us find out which set of metro stations would represent good prospects for the success a new cafe opened in its vicinity. For this project, we

used the k-means clustering algorithm and the scikit-learn package in Python was used for implementing it.

3.3. Finding the optimal value for the number of clusters

In k-means clustering algorithm, the number of clusters is an important hyperparameter that needs to be chosen properly so that we get the best results out of the algorithm. The most common method employed is to plot the inertia value (i.e. the sum of the squared distance of each point from the centre of the cluster it belongs to) for various values of k and then choosing the elbow point in the graph as the optimal value for k. This method was implemented in our analysis for choosing the optimal value for the number of clusters.

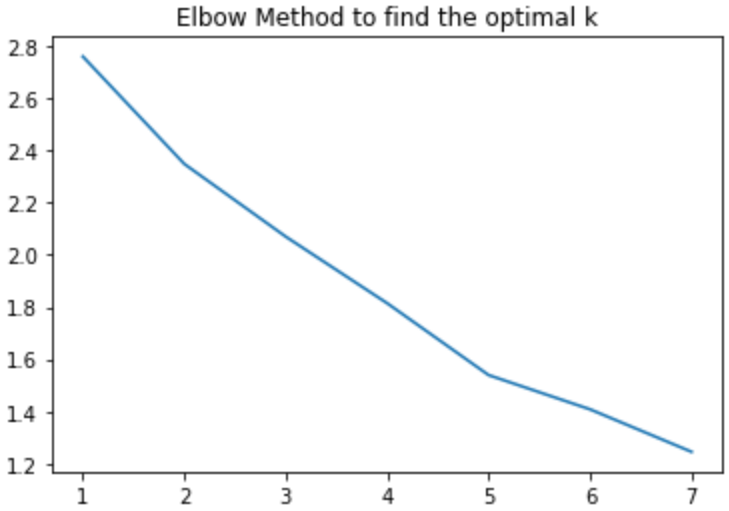

Fig 6 - Elbow method to find the optimal value of k

Although there is no definitive elbow, the slope of the curve decreases from n = 5. So let's choose the number of clusters for our analysis as 5.

3.4. Implementing the k-means clustering machine learning algorithm

The venue data was collected for each metro station using the Foursquare API and then the data was processed in such a way that the relative frequency of occurrence for each venue category was calculated for the metro stations. This data was then used for clustering the metro stations into 5 different groups with similar characteristics.

4. Results

On running the k-means algorithm, the metro stations were clustered into 5 groups with each cluster containing the following number of metro stations.

Table 2 - Number of metro stations in each cluster

Cluster	Number of Stations
Cluster 1	4
Cluster 2	5
Cluster 3	27
Cluster 4	1
Cluster 5	3

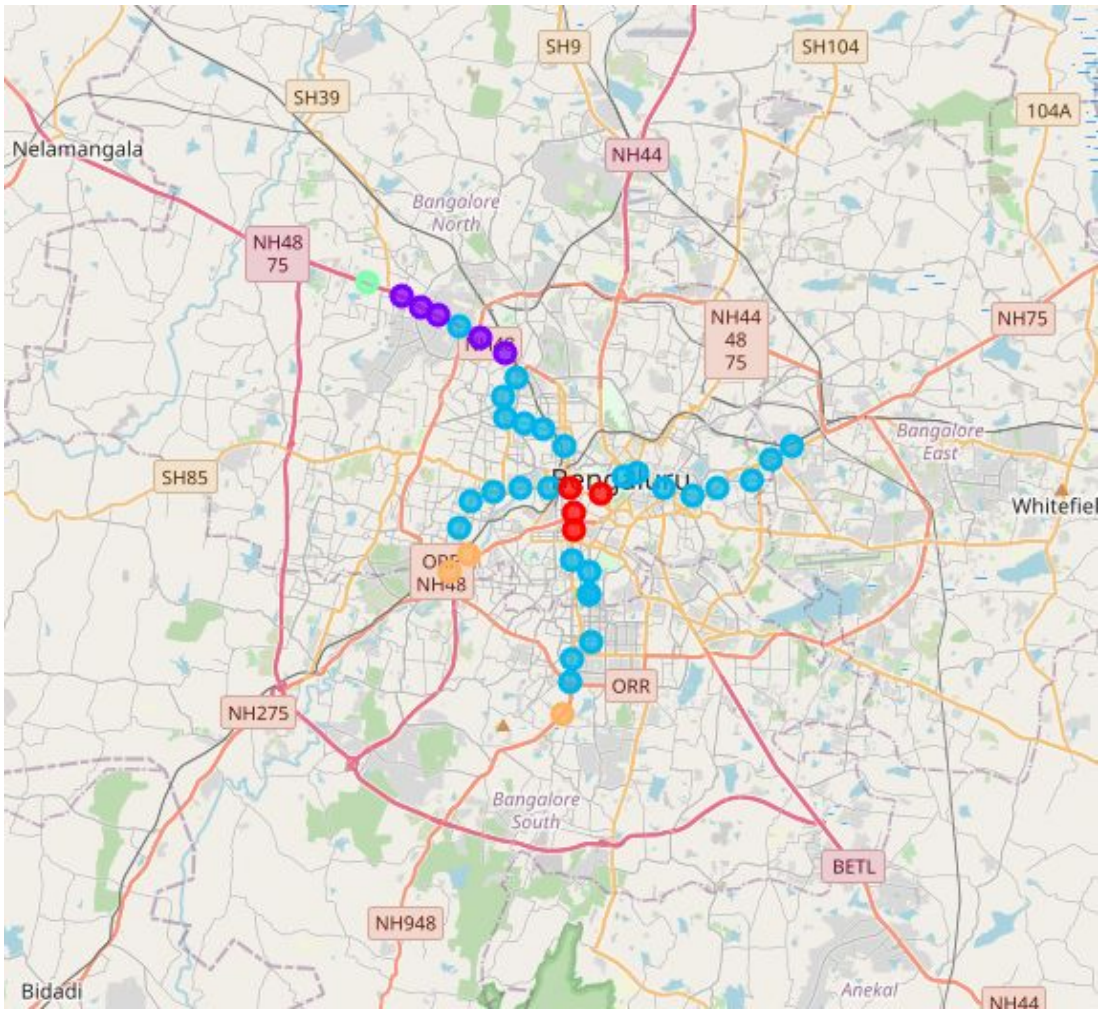

Fig 6 - Results of the clustering represented in a folium map

The metro stations in each cluster are given in the following tables.

Cluster 1

	Metro Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Chickpete	Indian Restaurant	Market	Miscellaneous Shop	Rest Area	Historic Site
14	Krishna Rajendra Market	Indian Restaurant	Flower Shop	General Entertainment	Market	South Indian Restaurant
18	Majestic	Indian Restaurant	Hotel	Bed & Breakfast	Platform	Bus Station
30	Sir M. Visveshwarya	Indian Restaurant	Hotel	Karnataka Restaurant	Bus Station	Dessert Shop

Cluster 2

	Metro Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	Dasarahalli	Shopping Mall	Fast Food Restaurant	Multiplex	Resort	Karnataka Restaurant
11	Jalahalli	Light Rail Station	Resort	Karnataka Restaurant	Metro Station	Multiplex
24	Peenya Industry	Fast Food Restaurant	Indian Restaurant	Light Rail Station	Bus Station	Train Station
38	Yeshwantpur	Hotel	Bus Station	Fast Food Restaurant	Market	Multiplex
39	Goraguntepalya	Fast Food Restaurant	Hotel	Bar	Punjabi Restaurant	Multiplex

Cluster 3 (Not all stations are included)

	Metro Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Attiguppe	Coffee Shop	Pizza Place	Seafood Restaurant	Ice Cream Shop	South Indian Restaurant
1	Baiyappanahalli	Coffee Shop	Café	Platform	Light Rail Station	Cafeteria
2	Banashankari	Indian Restaurant	Ice Cream Shop	Café	Women's Store	Hookah Bar
3	Bangalore City Railway Station	Platform	Indian Restaurant	Indie Movie Theater	Metro Station	Electronics Store
5	Cubbon Park	Indian Restaurant	Hotel	Chinese Restaurant	Café	Italian Restaurant
8	Halasuru	Café	Hotel	Department Store	Chinese Restaurant	Bar
9	Hosahalli	Indian Restaurant	Fast Food Restaurant	Snack Place	Department Store	Ice Cream Shop
10	Indiranagar	Indian Restaurant	Café	Pub	Chinese Restaurant	Bar
12	Jayanagar	Indian Restaurant	Chinese Restaurant	Café	Hotel	Lounge
13	Jayaprakash Nagar	Food Truck	Pizza Place	Department Store	Café	Liquor Store
15	Lalbagh	Indian Restaurant	Café	Ice Cream Shop	Electronics Store	Snack Place
16	Magadi Road	Indian Restaurant	Department Store	Hotel	Stadium	Shopping Mall
17	Mahalakshmi	Italian Restaurant	Lounge	Clothing Store	Bowling Alley	Department Store
19	Mahatma Gandhi Road	Indian Restaurant	Café	Clothing Store	Pub	Lounge
22	National College	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Bakery	Park

Cluster 4

	Metro Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
21	Nagasandra	Gas Station	Light Rail Station	Athletics & Sports	Bakery	Women's Store

Cluster 5

	Metro Station Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Deepanjali Nagar	Metro Station	Indian Restaurant	Smoke Shop	Travel & Transport	Department Store
20	Mysore Road	Metro Station	Travel & Transport	Train Station	Road	Deli / Bodega
37	Yelachenahalli	Department Store	Indian Restaurant	Pizza Place	Metro Station	Café

5. Discussion

It can be seen that out of the 5 clusters, cluster 4 is a single member cluster which is more or less an outlier that is very different from the other metro stations. On analysing its characteristics, cluster 4 doesn't seem to be a good candidate for opening a new cafe. Cluster 5 also represents a less outgoing populace and is not a great fit for opening a new cafe.

Now coming to the other clusters, Cluster 3 which is the largest cluster with 27 members represent a group of metro stations that already has a large number of cafes in its vicinity (with cafes featuring as 2nd - 5th most occuring venue in almost all its members). So this cluster is also not a great candidate for our new cafe owing to the high competition that is already present.

The remaining 2 clusters (Cluster 1 and Cluster 2) have a thriving food culture but are not congested with a lot of cafes. And hence would be good candidates for opening a new cafe. Amongst the two clusters, cluster 2 represents a more urban and outgoing community (as evident from the type of venues that are prevalent) and hence would be the best choice for our new cafe.

Cluster	Description	Remarks
Cluster 1	Not congested with cafes, but represent a less outgoing community	Second most preferred
Cluster 2	Urban centres with a thriving food culture, not congested with cafes	Most preferred for opening a new cafe
Cluster 3	Saturated with cafes	Not a viable choice for opening a new cafe
Cluster 4	Outlier case with a single member	-
Cluster 5	Lacks a strong food culture with very few F&B venues currently	-

Fig 7 - Summary of the project findings

6. Conclusion

This project was aimed at finding the best location for opening a new cafe alongside the stations of Bangalore metro. The approach taken was to cluster the stations on the basis of venues that are already functioning in the vicinity of each station. The clustering was done using the k-means clustering machine learning algorithm. The characteristics of the resulting clusters were analysed and the cluster that represented the best prospects for the success of a new cafe was suggested.

The suggested clusters should be considered as a primary recommendation. A more detailed analysis including a market analysis and financial analysis should be conducted for each metro station in the recommended cluster so that the locations for opening the new cafe can be finalised.