Applied Data Science Capstone

Finding the best location for a coffee shop in Bangalore

- Himanish Prakash

1. Introduction

1.1 Background

Bangalore or Bengaluru is the capital of the Indian State Karnataka is known for its pleasant climate throughout the year. Bangalore is widely regarded as the "Silicon Valley of India" (or "IT capital of India") because of its role as the nation's leading information technology (IT) exporter. A demographically diverse city, Bangalore is the second fastest growing major metropolis in India. Alot of people shifts to Bangalore because of its advancement in IT sector and is also the hub of lot of startups. Bengaluru doesn't just have India's most knowledgeable coffee drinkers but also has more people who experiment with home brewing and roasting. India's tryst with coffee began in the 16th Century, when Baba Budan, a Sufi cleric smuggled seven coffee beans through the port of Mocha in Yemen. He sowed these near Chikmagalur, in an area that came to be known as Baba Budan Hills, and got India hooked to this wonder-bean. Today, Karnataka produces more than 70% of India's coffee and the state's key coffee cultivation centres– Coorg, Chikmagalur and Sakleshpur—are all popular weekend getaways from the city. In such a large and rich coffee culture city of Bangalore, it will be competitive to start up coffee business. We need to to find the hotspot to lay our coffee shop and get the maximum foot traffic and potential revenue.

1.2 Problem

The objective of this capstone project is to analyze and select the best neighborhood in the city of Bangalore to open a new coffee shop. Using data science methodology and machine learning techniques like clustering, this project will aim to provide a solution to the question: Which neighborhood in the city of Bangalore would a stakeholder of a coffee shop look to open a business.

1.3 Interest

Target audience of the project will be all those locals, outsiders, tourist who are in Bangalore and don't start there day without a coffee. From yougsters to senior citizens who love coffee comes under the target audience.2. Data Section

2.1 Data Source

For this project we need the following data:

Bangalore Coffee shops data that contains list Locality, name along with their latitude and longitude.

▪ Data source : Bangalore Zomato kaggle dataset

▪ Description : This data set contains the required information. And we will use this data set to explore various locality of city.

Nearby places in each locality of Bangalore city.

▪ Data source : Foursquare API

▪ Description : By using this api we will get all the venues in each neighborhood.

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined in one table. There were lot of columns which were of no use to us so we removed them. There were lot of missing values as well for that I removed the rows which had missing values. The dataset wasn't that complicated hence it was easy to do data wrangling.

2.3 Feature Selection

Initially there were 17 columns in the Zomato Dataset of Bangalore after cleaning 8 columns were there. They were address, name, rate, votes, location, rest_type, cuisines, and listed_in.

From that columns and rows we discovered the count of maximum restaurant that are presently in Bangalore. And we discovered that Café Coffee Day was having 89 branches in total in Bangalore. Onesta, Empire restaurant, kanti sweets and five star chicken were after CCD.

Then I visualized the above in the form of bar graph to get better understanding of it. 3. Methodology Section

3.1 Data Analysis

From the required dataset I took out the rows which only have coffee shops detail so that I can get the 5 most crowded areas where potential traffic is highest. From the analysis we found that MG road, Jayanagar, Banashankari, Indiranagar, Kormangala 4th block, and church street are the top 5 locations which have maximum number of coffee shops located. To depict it I used pie chart to get actual percentage of each area.

3.2 Foursquare API

Using foursquare API I got the the longitude and latitude of the coffee location. Also along with longitude and latitude will explore all the nieghborhoods in Bangalore. I used the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. Used K-mean clustering algorithm to cluster the neighborhood and with the help of folium library visualized the neighborhoods in Bangalore and their emerging clusters.

I grouped the coffee shops on the basis of there location, using one hot encoding it was more easy to analysis and observe.

Then with each location I got the 10 most common coffee shops present their on the basis of that I clustered them into 5 clusters using K- means clustering algorithm. Dataframe along with location and their 10 most common venues(coffee shops)

Putting into pandas Dataframe

4. Result

From the following graph you can see MG Road, Church Street, Banshakari, VV puram, Kormangala, shivajinagar and Ashok Nagar have maximum number of coffee shops in Bangalore. So the potential traffic for coffee is maximum here. Someone who is entrepreneur and who is looking for a location to open coffee shop in Bangalore he/she can check the availability of space in these areas or around this year. The maximum competition will be from café coffee shop as it is has the maximum rating and maximum potential traffic in Bangalore. Using folium library we visualized the clustered. Further we can examine each

cluster individually.

5. Conclusion

In this report I tired to give a glimpse of what I did in my capstone project. I managed to find the best location to lay down a coffee shop in Bangalore by analyzing zomato dataset and foursquare API. We could see that each of the neighbourhoods in the city have a wide variety of experiences to offer which is unique in it's own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion. Bangalore seem to offer a vacation stay or a romantic getaway with a lot of places to explore, beautiful landscapes, amazing food and a wide variety of culture. Overall, it's upto the stakeholders to decide which experience they would prefer more and which would more to their liking