# Project — The Battle of Neighbourhoods

# Narasimha Rao

# July 2020

# Table of contents

# 1. Introduction: Business Problem

## 1.1 Background

Bangalore , officially known as Bengaluru, is the capital of the Indian state of Karnataka. It has a population of about 10 million and a metropolitan population of about 8.52 million, making it the third most populous city and fifth most populous urban agglomeration in India.Located in southern India on the Deccan Plateau, at a height of over 900 m (3,000 ft) above sea level, Bangalore is known for its pleasant climate throughout the year. Its elevation is the highest among the major cities of India.

Bengaluru 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.Indian technological organisations such as ISRO, Infosys, Wipro and HAL are headquartered in the city. A demographically diverse city, Bangalore is the second fastest-growing major metropolis in India.

 Bangalore is also home to many vegan-friendly restaurants and vegan activism groups, and has been named as India's most vegan-friendly city by PETA India.

 With it's diverse culture , comes diverse food items. There are many restaurants in Bangalore City, each belonging to different categories like South Indian, North Indian, Andhra, Chinese , Italian , French etc. So as part of this project , we will list and visualise all major parts of Bangalore City.

## 1.2  Problem

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry has'nt been saturated yet and the demand is increasing day by day. Inspite of increasing demand it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality?. Which are locations has more restaurants? Do the entire locality loves Chinese or Andhra  food. If yes then is that locality populated by a particular sect of people for eg. Andhra, Foreigners, Jain, Marwaris, Gujaratis,etc. These kind of analysis can be done using the data, by studying different factors.

## 1.3  Interest

The project aims to select the best location to stay in Bangalore if someone likes Chinese or Andhra cuisines and Best locations to start a new Chinese and Andhra Cuisines or the best location to stay if someone prefer Chinese Cuisine based on the total restaurents, locality, riviews and neighborhoods.

- ➢ What places are having best restaurant in Bangalore?
- ➢ Which place are suitable for edible person in Bangalore city
- ➢ Which place are not suitable for edible person in Bangalore city
- ➢ Which is the best place to stay if I prefer Chinese Cuisine  or Which areas have large number of Chinese Resturant Market ?
- ➢ Which is the best place to stay if I prefer Andhra Cuisine ?
- ➢ What is best location in Bangalore City for Chinese Cuisine ?
- ➢ What is best location in Bangalore City for Andhra Cuisine ?
- ➢ Which all areas have less number of resturant ?

# 2.Data Acquisition and Cleaning

# 2.1 Data Acquisition

The data acquired for this project from the following data sources:

Bangalore Resturants data that contains list Locality, Resturant name,Rating along with their latitude and longitude.

- ➢ Data source : Zomato kaggel dataset
- ➢ Description : This data set contains the required information. And we will use this data set to explore various locality of Bangalore city.

Nearby places in each locality of Bangalore city.

- ➢ Data source : Fousquare API
- ➢ Description : By using this api we will get all the venues in each neighborhood.

Latitude and Longitude values from the geolocater.geocodes.

The Zomoto data set contains the following columns:

*url:* URL of the restaurant in the zomato website.

*address:* address of the restaurant in Bengaluru.

*name:* Name of the restaurant.

*online_order:* whether online ordering is available in the restaurant or not.

*book_table:* table book option available or not.

*rate:* contains the overall rating of the restaurant out of 5.

*votes:* total number of votes.

*phone:* phone number of the restaurant.

*location:* the neighborhood in which the restaurant is located.

*rest_type:* restaurant type.

**dish_liked:** dishes people liked in the restaurant.

**cuisines:** food styles, separated by comma.

**approx_cost:** approximate cost for meal for two people.

**reviews_list:**list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer.

**menu_item:** list of menus available in the restaurant.

**listed_in(type):** type of meal.

**listed_in(city):** contains the neighborhood in which the restaurant is listed.

# 2.2 Data Analysis and Cleansing

Data preparation for all the phases done separately.

Dataset is not clean and contains redundant as well as unnecessary data. So, done the cleaning process and deleted the unnecessary or redundant features.

For data analysis, we do not need the contact details of the restaurant so, deleting the following features :

> ➢ url
> ➢ address
> ➢ phone

The menu_item feature contains the names of the dishes available in the restaurant. This will hardly impact our analysis, as it is highly mathematics driven. Also, we have other features like rest_type, cuisines, listed_in(type) and dish_liked which give us a fair idea about what the restaurants have to offer, as we do not need to be very specific about what all dishes are available in the restaurant for the analysis. So, droping this feature as well.

The reviews_list feature also will hardly contribute for the analysis as it is pure text and we also have features like rate and votes that cover the necessary information. But, one can still extract this feature and play around like running an NLP algorithm. As far as the data analysis is considered deleting this feature as well.

Deleted some of the features and lets clean the remaining.

The rate feature is a string and it contains the character '/'.This is not required, and so I will remove the '/5' from each entry in the column, changing the datatype of the feature to float.

Renamed the some features, for simplicity sake.

The following features were renamed :

➢ approx_cost(for two people) : **avg_cost**
➢ listed_in(type) : **meal_type**
➢ listed_in(city) : **city_loc**

The feature description given on kaggle, makes one think about the two features : location and city. Both have the same description, so this is redundant information. Now, which one to keep?

In this approach I kept the city feature(0 null values) and deleting the location feature.

Now Data is ready, lets have a look at it.

| | name | online_order | book_table | rate | votes | location | rest_type | dish_liked | cuisines | avg_cost | meal_type | city_loc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Jalsa | Yes | Yes | 4.1 | 775 | Banashankari | Casual Dining | Pasta, Lunch Buffet, Masala Papad, Paneer Laja... | North Indian, Mughlai, Chinese | 800 | Buffet | Banashankari |
| 1 | Spice Elephant | Yes | No | 4.1 | 787 | Banashankari | Casual Dining | Momos, Lunch Buffet, Chocolate Nirvana, Thai G... | Chinese, North Indian, Thai | 800 | Buffet | Banashankari |
| 2 | San Churro Cafe | Yes | No | 3.8 | 918 | Banashankari | Cafe, Casual Dining | Churros, Cannelloni, Minestrone Soup, Hot Choc... | Cafe, Mexican, Italian | 800 | Buffet | Banashankari |
| 3 | Addhuri Udupi Bhojana | No | No | 3.7 | 88 | Banashankari | Quick Bites | Masala Dosa | South Indian, North Indian | 300 | Buffet | Banashankari |
| 4 | Grand Village | No | No | 3.8 | 166 | Basavanagudi | Casual Dining | Panipuri, Gol Gappe | North Indian, Rajasthani | 600 | Buffet | Banashankari |

| | name | online_order | book_table | rate | votes | rest_type | dish_liked | cuisines | avg_cost | meal_type | city_loc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 51672 | 51672 | 51672 | 51672.000000 | 51672.000000 | 51466 | 23639 | 51672 | 51352 | 51672 | 51672 |
| unique | 8785 | 2 | 2 | NaN | NaN | 93 | 5271 | 2723 | 70 | 7 | 30 |
| top | Cafe Coffee Day | Yes | No | NaN | NaN | Quick Bites | Biryani | North Indian | 300 | Delivery | BTM |
| freq | 96 | 30428 | 45223 | NaN | NaN | 19129 | 182 | 2913 | 7576 | 25917 | 3278 |
| mean | NaN | NaN | NaN | 2.983062 | 283.920073 | NaN | NaN | NaN | NaN | NaN | NaN |
| std | NaN | NaN | NaN | 1.515475 | 804.138258 | NaN | NaN | NaN | NaN | NaN | NaN |
| min | NaN | NaN | NaN | 0.000000 | 0.000000 | NaN | NaN | NaN | NaN | NaN | NaN |
| 25% | NaN | NaN | NaN | 3.000000 | 7.000000 | NaN | NaN | NaN | NaN | NaN | NaN |
| 50% | NaN | NaN | NaN | 3.600000 | 41.000000 | NaN | NaN | NaN | NaN | NaN | NaN |
| 75% | NaN | NaN | NaN | 3.900000 | 198.000000 | NaN | NaN | NaN | NaN | NaN | NaN |
| max | NaN | NaN | NaN | 4.900000 | 16832.000000 | NaN | NaN | NaN | NaN | NaN | NaN |

Extracted the Latitude and longitude values from Geolocator and added both the columns to the Data frame.

| | name | online_order | book_table | rate | votes | rest_type | dish_liked | cuisines | avg_cost | meal_type | city_loc | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Jalsa | Yes | Yes | 4.1 | 775 | Casual Dining | Pasta, Lunch Buffet, Masala Papad, Paneer Laja... | North Indian, Mughlai, Chinese | 800 | Buffet | Banashankari | 12.915221 | 77.573598 |
| 1 | Spice Elephant | Yes | No | 4.1 | 787 | Casual Dining | Momos, Lunch Buffet, Chocolate Nirvana, Thai G... | Chinese, North Indian, Thai | 800 | Buffet | Banashankari | 12.915221 | 77.573598 |
| 2 | San Churro Cafe | Yes | No | 3.8 | 918 | Cafe, Casual Dining | Churros, Cannelloni, Minestrone Soup, Hot Choc... | Cafe, Mexican, Italian | 800 | Buffet | Banashankari | 12.915221 | 77.573598 |
| 3 | Addhuri Udupi Bhojana | No | No | 3.7 | 88 | Quick Bites | Masala Dosa | South Indian, North Indian | 300 | Buffet | Banashankari | 12.915221 | 77.573598 |
| 4 | Grand Village | No | No | 3.8 | 166 | Casual Dining | Panipuri, Gol Gappe | North Indian, Rajasthani | 600 | Buffet | Banashankari | 12.915221 | 77.573598 |

With the help of city_loc, rate, name, Latitude and Longitude features we can predict the Best restaurants and locations in Bangalore city and we can make the clustering.