

Data Description

Bangalore Restaurants Analysis

Himanshu Poddar

1 Introduction

The main data used for this project will be from two sources:

- The restaurants in each neighborhood scraped from the Zomato website.
- Explore trending venues in a neighborhood particularly restaurant (FourSquare API).

Other supporting data:

- Coordinates (Geocoder Python)
- GeoJson (<https://github.com/openbangalore>)

2 Data Collection Process

The data was scraped from Zomato in two phase. After going through the structure of the website I found that for each neighborhood there are 6-7 category of restaurants viz. Buffet, Cafes, Delivery, Desserts, Dine-out, Drinks & nightlife, Pubs and bars.

2.1 Phase I

In Phase I of extraction only the URL, name and address of the restaurant were extracted which were visible on the front page. The URL's for each of the restaurants on the zomato were recorded in the csv file so that later the data can be extracted individually for each restaurant. This made the extraction process easier and reduced the extra load on my machine. The data for each neighborhood and each category can be found here

2.2 Phase II

In Phase II the recorded data for each restaurant and each category was read and data for each restaurant was scraped individually. 15 variables were scraped in this phase. For each of the neighborhood and for each category their online_order, book_table, rate, votes, phone, location, rest_type, dish_liked, cuisines, approx_cost(for two people), reviews_list, menu_item was extracted. See section 5 for more details about the variables.

2.3 Foursquare API

For each neighborhood, geopy module to convert an address into latitude and longitude values. For each neighborhood's coordinate, call FourSquare API to get the trending venues in that location.

3 Using data to solve the problem

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the establishment of different types of restaurant at different places in Bengaluru, aggregate rating of each restaurant, 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't 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. Do the entire locality loves vegetarian food. If yes then is that locality populated by a particular sect of people for eg. Jain, Marwaris, Gujaratis who are mostly vegetarian. These kind of analysis can be done using the data, by studying the factors such as

- Approx Price of food
- Location of the restaurant
- Theme based restaurant or not
- Which locality of that city serves that cuisines with maximum number of restaurants
- The needs of people who are striving to get the best cuisine of the neighborhood
- Is a particular neighborhood famous for its own kind of food.

“Just so that you have a good meal the next time you step out”