

Capstone Project - Exploring and Clustering Restaurants in Pune, India

Shardul Samdurkar
May 02, 2020

1. Introduction

1.1 Background

In this project I will explore and group the restaurants in the city of Pune in few categories based on the Ratings and Price range of the Restaurants.

Pune is a large city with an area of 331 sq.km. Pune is also home to an eclectic crowd of food-lovers and the city serves up everything from Western favourites to Asian delicacies. Whether you're looking for a heaven of fusion foods or a traditional taste of lip-smacking snacks—there's no dearth of restaurants to satisfy the discerning palate. Pune, being a metropolis and centre of prominent education institutes, is a melting pot of cultures which is reflected in its food too. The food-lovers are always in search of Restaurants with Good Ratings and also Restaurants that suits there Pocket.

With the help of machine learning and Data science I will try to achieve the objective of this project.

1.2 Interested audience

The target audience for such a project is of two types. This project will help the food-loving people of Pune to shortlist and plan their visit to the restaurants accordingly.

This project will also help businesses who are looking to invest in eateries in the Pune city.

.

2. Data

2.1 Data Source

Based on definition of our problem, factors that will influence our decision are:

- The location of the Restaurants in Pune
- The categories of these restaurants
- The Ratings of these restaurants
- The Price range of these restaurants

The data source that I will use to fetch the location information, the categories, ratings and price range of restaurants will be **the FourSquare API**

2.2 Data Cleaning

The Foursquare data contains information of each type of venue for example: historic places, museums, temples, hospitals. But in this project we are only interested in places like the restaurants, cafes, fast-food corners, etc.

So the Foursquare data was filtered by giving a category Id which brings only those venues which are related to food.

The data of restaurants obtained from the Foursquare API had several entries or records , where the column had 'NaN' values. The rows containing 'NaN' values were removed.

The data of restaurants obtained from Foursquare API corresponding to the Rating and Price range, had many records where the Price range was not specified or the Price range column was not specified.

The restaurants where the Price range was not specified were excluded from the final Dataset.

As a result the final dataset had 64 restaurants in total.

The dataset looked like this:

	id	Rating	Price_Tier	Price_Range	name	categories	lat	lng	crossStreet
0	4b9b583cf964a520aa0136e3	8.5	1	Cheap	Café GoodLuck	Café	18.517388	73.841485	Corner of FC Road
1	4de24c587d8b2547eaffe58b	8.4	1	Cheap	Tilak Hotel	Snack Place	18.509603	73.846883	Tilak road
2	4f016c4629c201c6fec99a78	8.3	2	Moderate	Barbecue Nation	BBQ Joint	18.516237	73.842223	Jangali Road
...
62	4d0a55daacd4a1cdfc6f8193	7.6	1	Cheap	Richie Rich	Café	18.457362	73.852508	Katraj
63	51205bb2e4b0a1ec80282126	7.7	1	Cheap	Khasbaug Misal	Breakfast Spot	18.560823	73.787599	Baner Road

64 rows × 9 columns

3. Methodology and Exploratory Data Analysis

In this project our objective is to explore and group the restaurants in the city of Pune in few categories based on the Ratings and Price range of the Restaurants.

3.1 Step 1

In first step I have collected the required **data: location, type (category), rating, price range of every restaurant within 20km from Pune center** by using only the Foursquare API

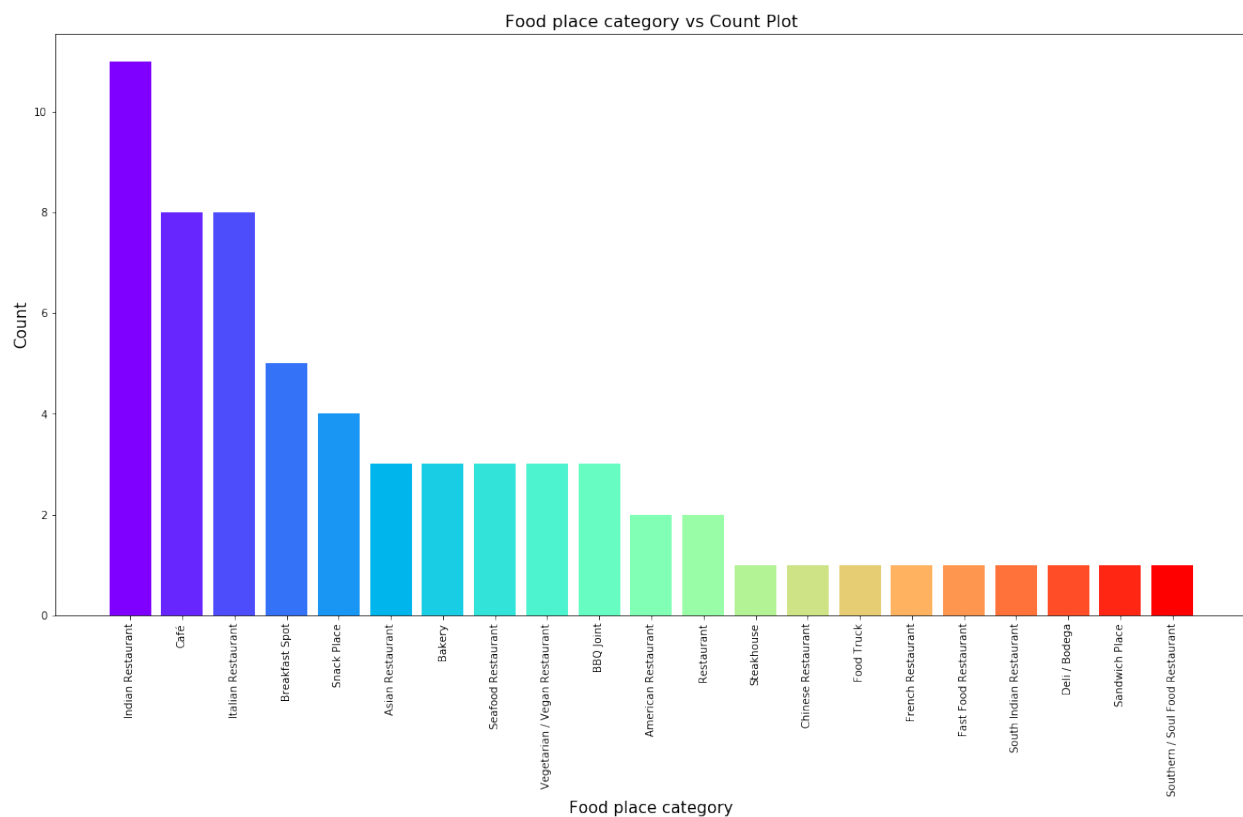
3.2 Step 2

In the Second step, I analyzed the data that we created based on the ratings and price of each restaurants.

I analyzed the restaurants based on their **Category distribution** and also analyzed restaurants based on their **Price range**

3.2.1 Plot of Category Distribution

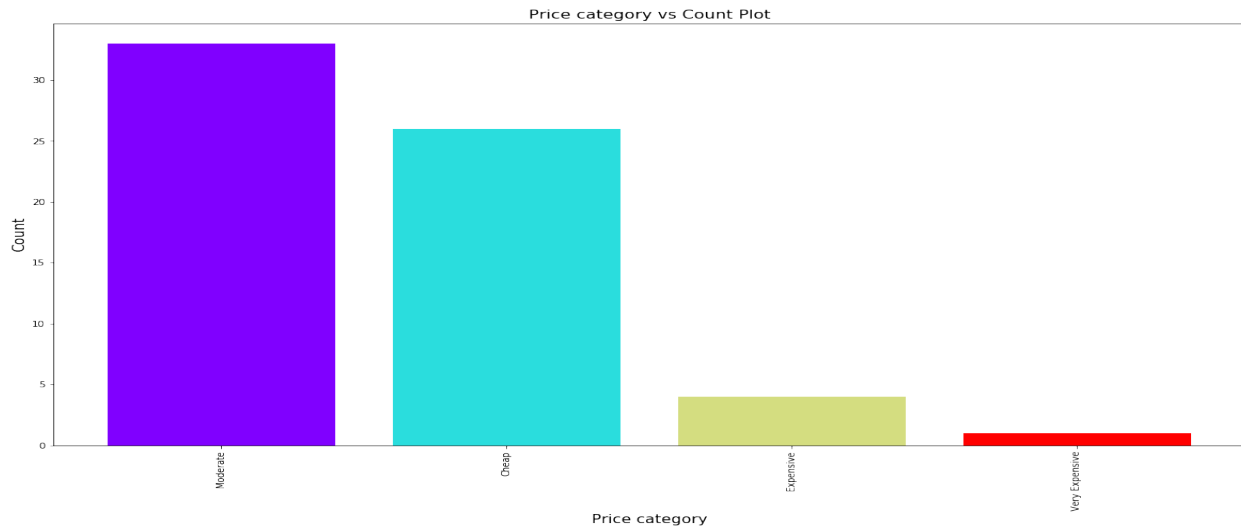
There are restaurants of various categories in Pune. We plotted it and checked which is the **most popular category**



From this plot we understood that the 3 most popular categories of Eateries in Pune are **Indian Restaurants, Cafe** and **Italian Restaurants**

3.2.2 Price Category Plot

I plotted the Price Category vs Count plot to get idea of Price category of majority of restaurants.



From this plot we understood that the **majority** restaurants in Pune are **Moderately priced** or **Cheaply priced**. There are very few Expensive/Very Expensive restaurants in Pune

3.3 Step 3

In the next step I preprocessed the data so that it becomes ready to be put into the Clustering Model and also I Normalized the dataset. Normalization is a statistical method that helps mathematical-based algorithms interpret features with different magnitudes and distributions equally. I used **StandardScaler()** function to normalize our dataset.

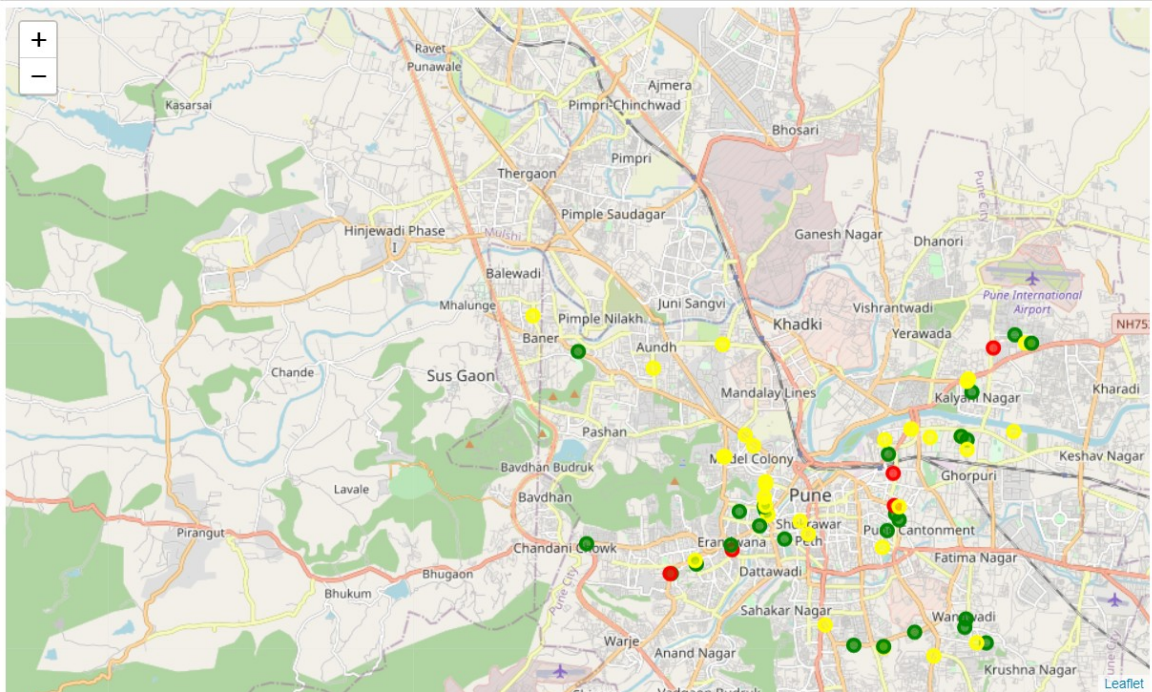
3.4 Step 4

Lastly in the fourth and final step, I **clustered the restaurants based on the available information of each**. This allowed us to clearly identify which restaurants can be recommended and with what features.

The Clustering was done on the basis of two features of the dataset namely, the **Rating** and **Price Range**

The Map obtained after Clustering the data into three categories looked like this :

Out[21]:



4. Results and Discussion

Based on our analysis above, we draw a number of conclusions that will be useful to any food lover in Pune

After collecting data from the **Foursquare API** we got a total of **100** restaurants. However after Data cleaning it resulted in a total restaurant count of **64**.

We identified that from the total number of eateries in Pune the most popular eateries are **Indian Restaurants, Cafes and Italian restaurants** respectively.

We also understood that most of the eateries in Pune are **Moderately** and **Cheaply Priced** only few eateries fall in the **Expensive/ Very Expensive** Category

Finally through Clustering we came to know that most of the restaurants in Pune have a **good Rating** and also based on the **Price range** they can be categorized into three types i.e.:

- Cheaply priced Restaurants
- Moderately priced Restaurants
- Expensive/Very Expensive Restaurants

With this information and by looking at the map we can conclude that :

- **Cheaply priced** Restaurants are located mainly in the Wanawadi area and the Pune Cantonment area
- **Moderately priced** Restaurants mainly are located near Deccan Gymkhana and Bund Garden road
- **Expensive/Very Expensive** Restaurants are located mainly near Kothrud and Pune Junction area

5. Conclusion

The purpose of this project was to explore and group the restaurants in Pune based on their Ratings and Price range in order to help people locate the restaurants according to their needs. In this project we collected the required data, processed it, analyzed it and performed clustering of these restaurants and identified the places where we can find the restaurants of desired categories

We found that Cheaply priced Restaurants are located mainly in the Wanawadi area and the Pune Cantonment area, while the Moderately priced Restaurants mainly are located near Deccan Gymkhana and Bund Garden road and the Expensive/Very Expensive Restaurants are located mainly near Kothrud and Pune Junction area.