Exploring venues in Pune, India using Google, Foursquare and Zomato API

Arshit Patel
November 21, 2019

1. Introduction

1.1 Background

Pune is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense.

1.2 Problem

When we think of it by the investor, we expect from them to prefer the districts where there is a lower real estate cost and the type of business they want to install is less intense. Thus, our aim here is to identify places that will guide investors in this direction, nowadays.

1.3 Interest

Here, we'll identify places that are fit for various individuals based on the data collected from the Google, Foursquare and Zomato APIs and information retrieved from the Data Science application.

2. Data acquisition and cleaning

2.1 Data sources

The venue data has been collected from two APIs, Foursquare API, and Zomato API. The first step was to search for venues within a radius of 6 Kilometers from Pune's center point. After extracting over 208 locations using the Foursquare API, the latitude and longitude values were used to fetch the venue details using Zomato API. The Google API is used to get the latitude and longitude of the target place center for e.g. Pune here.

2.2 Data cleaning

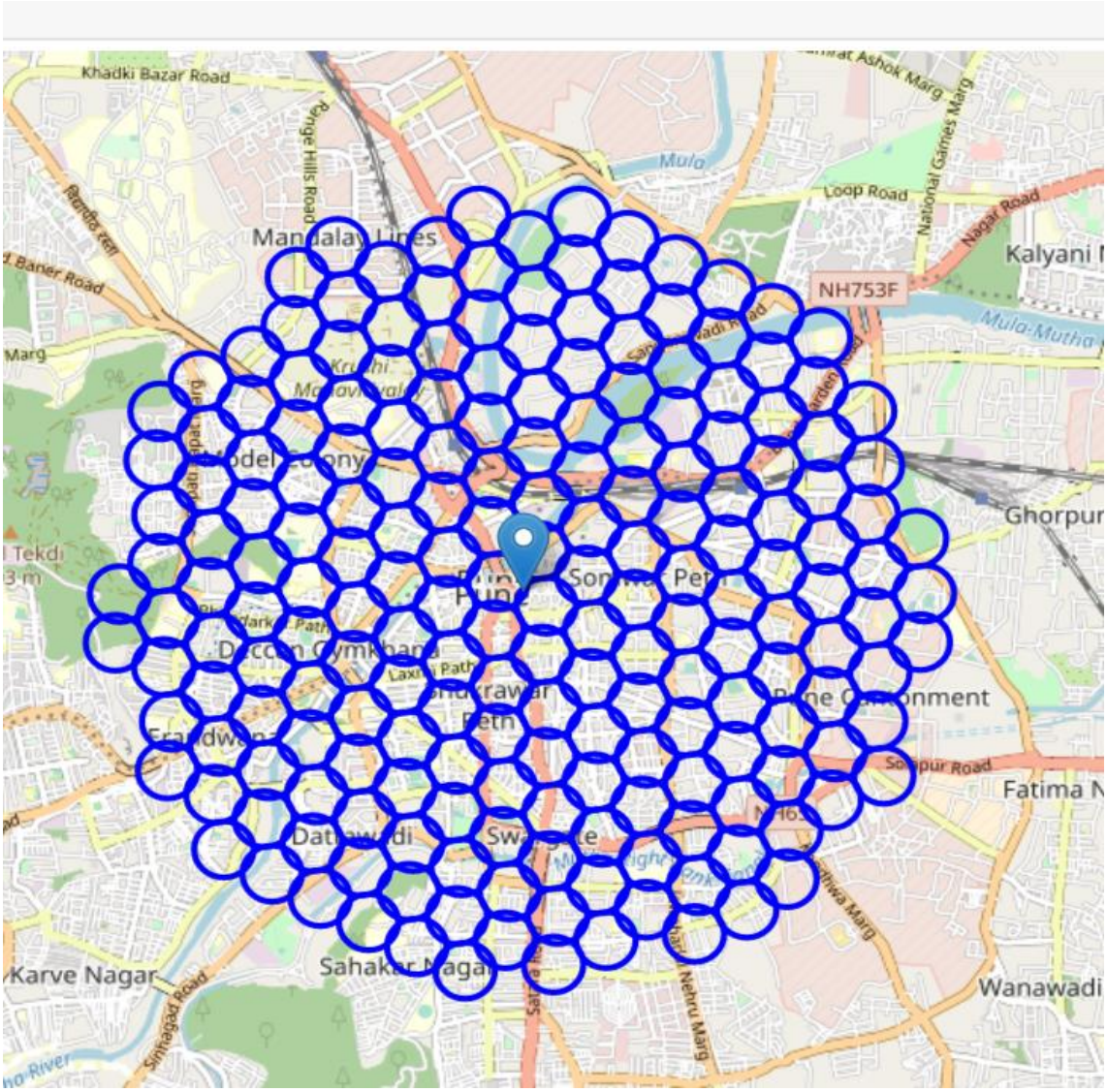
The Data collected using Zomato API returns all venues which are a cafe where The venues retrieve from the Foursquare API is mixed like restaurant, cafe, bar, etc. So, I classify that data in the column **isCafé** which contains Boolean values and the classification is done base on the service the venue provided. Over 400 data in the radius of 6 km we took 208 data and after redundancy removing like very close places that their coordinates are same we got final 168 data to process further.

2.3 Feature selection

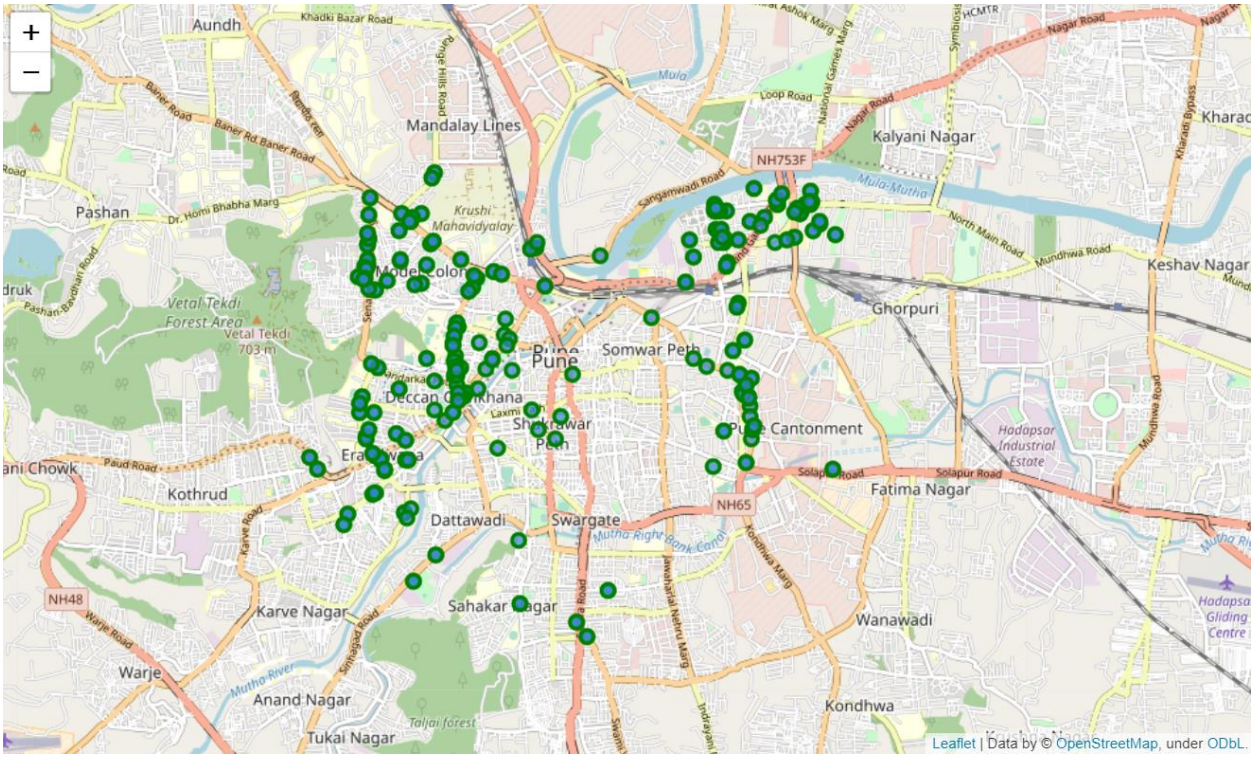
After getting data from both Foursquare and Zomato APIs with the feature like venue name, coordinates, rating, price, cuisine, services provided, we select name, latitude, longitude, type of venues for analysis.

3. Exploratory Data Analysis

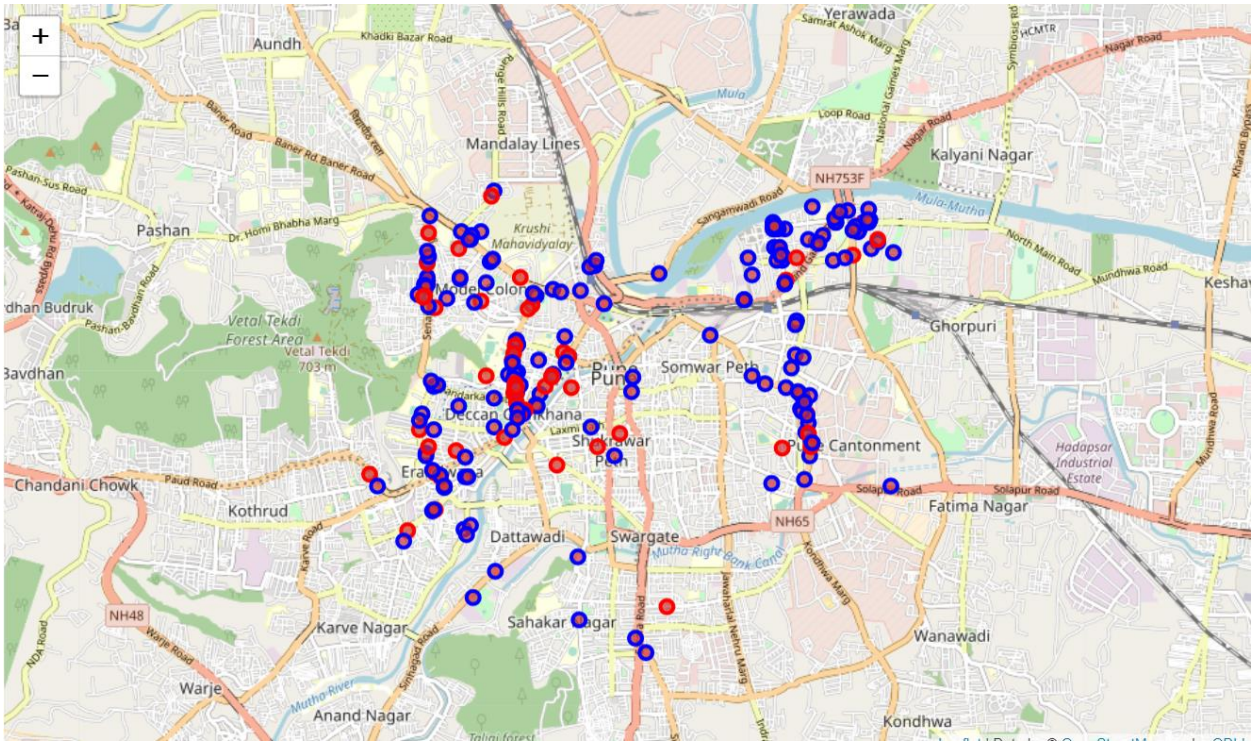
After collecting data from the Foursquare and Zomato APIs, we got a list of 208 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as their names to combine them and remove all the outliers. This resulted in a total venue count of 168.



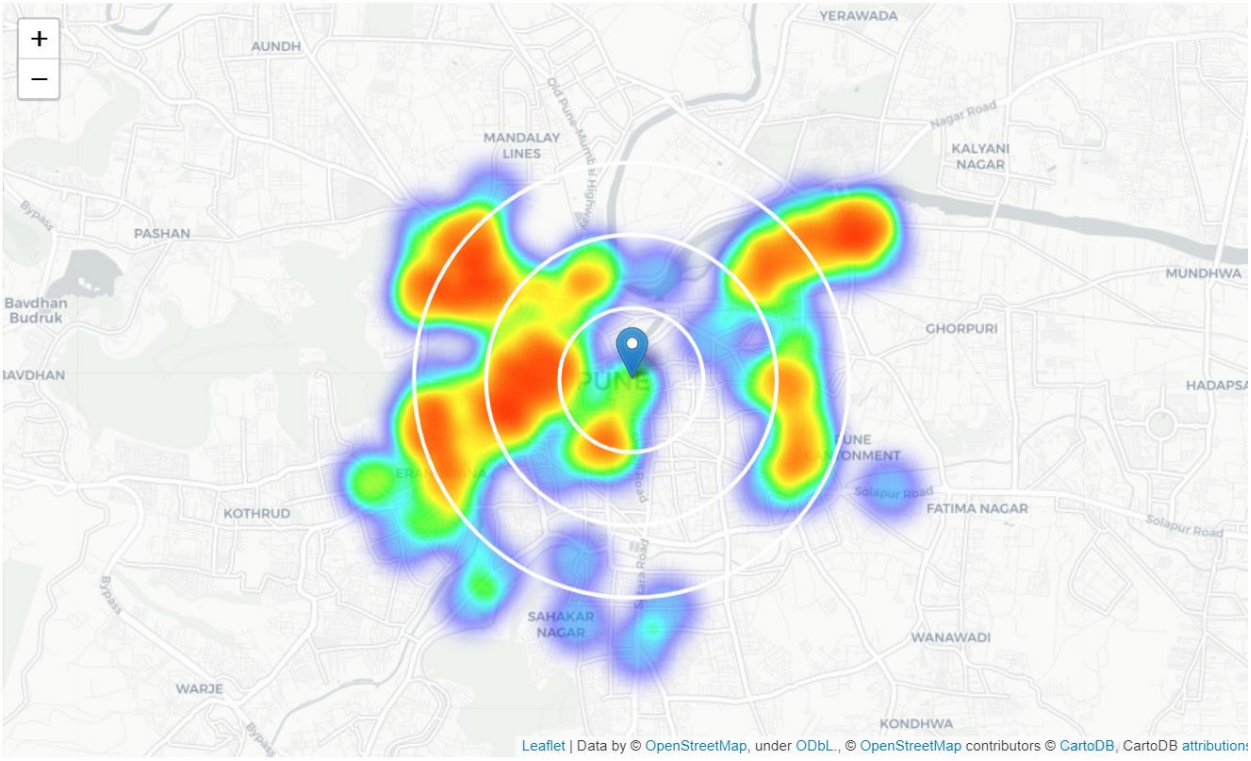
Our Area of Interest



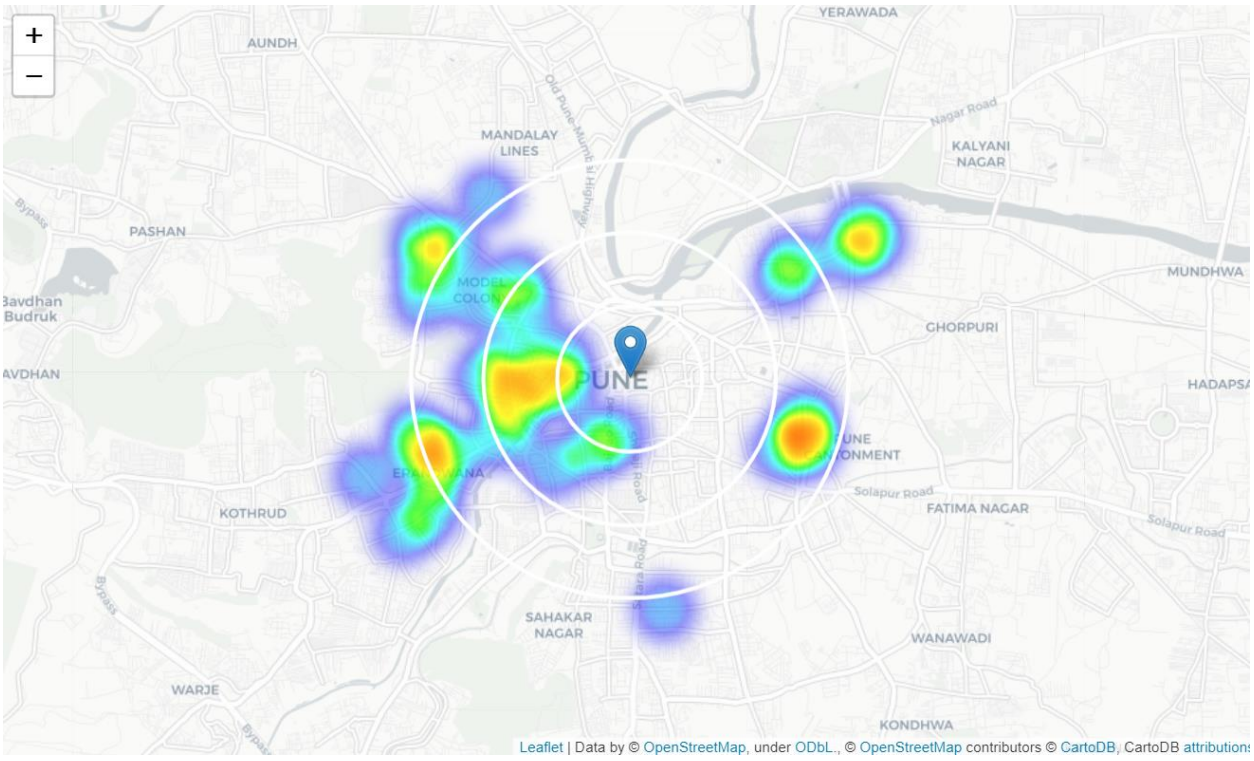
Some of the Food Provider of our selected Area



Red ones are the Cafes and blue ones are other.



Density of all Food Provider



Density of Café

We identified that from the total set of venues, the majority of them were Cafes are near the Deccan area. An investor who loves Cafes would surely benefit from coming to Pune.

4. Predictive Modeling

The field Food is very vast ,with thousand choice of eating preference ,the one with clear mindset can set their business well. The Factor like local famous food, Locality, Profession, etc., all are matters when it comes to the term business.

With the help of different algorithms like K-mean, KNN, decision tree we can cluster the group of people who likes to often visit the coffee shop and also can predict other customer behavior based on the previous data we have collected.

For E.g. with the help of previous behavior of several customer our prediction model classify that who often buy cappuccino also buy garlic bread ,based on this prediction we can provide some offer to the new customer to make the bond strong.