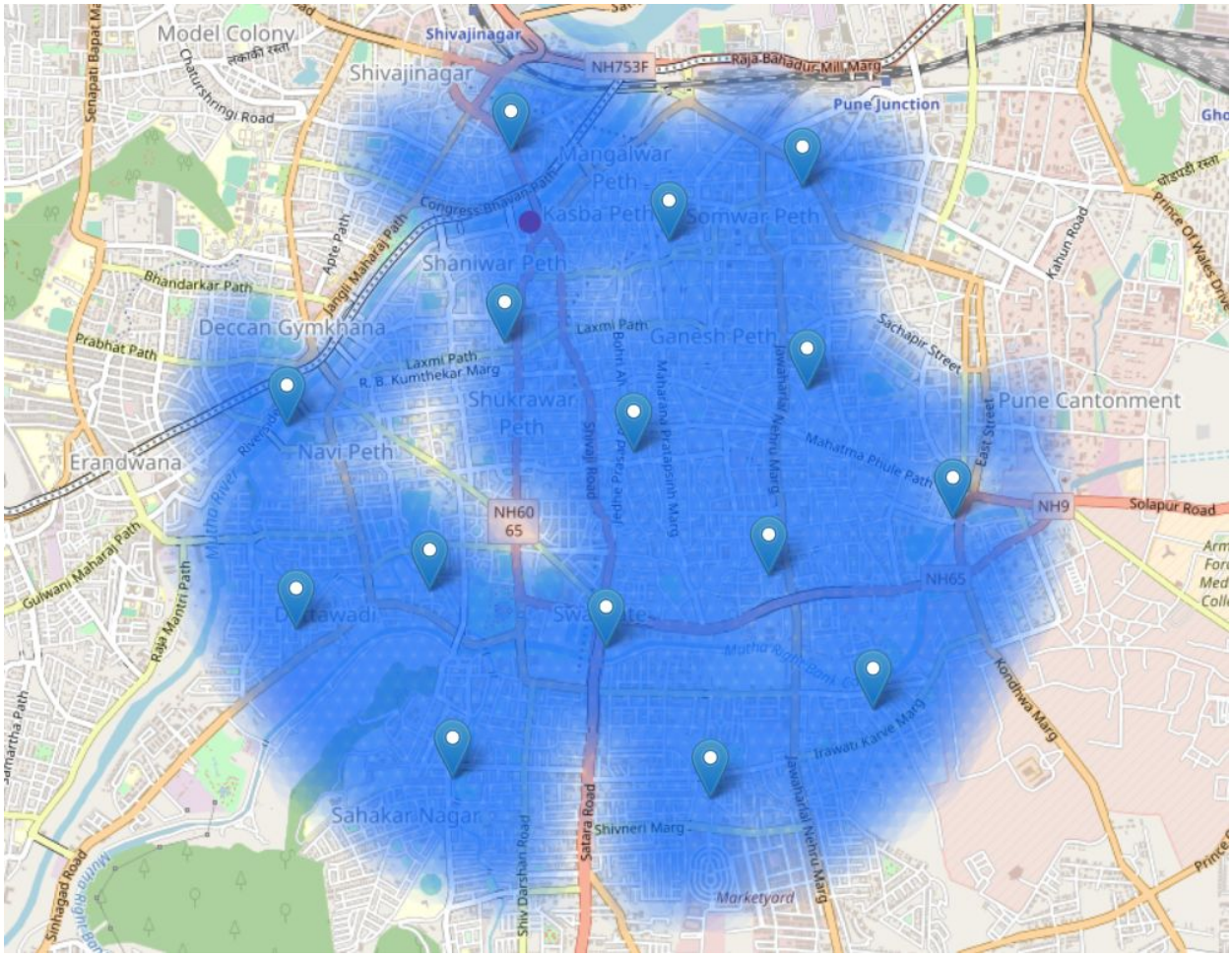

The Battle of Neighborhoods

Applied Data Science Capstone Project

18.03.2020

—

Kunal Gauraw

Table of Contents

Introduction

Background

Pune is the second largest city in the Indian state of Maharashtra, after Mumbai. The city has emerged as a major educational hub in recent decades, with nearly half of the total international students in the country studying in Pune. The city has the eighth largest metropolitan economy and the sixth highest per capita income in the country. The key sectors of the local economy are education, manufacturing and information technology. The educational and economic ecosystem of the city attracts a young and vibrant metropolitan crowd.

An Italian restaurant chain wants to open a restaurant in Pune city. There are lots of restaurants in Pune. They want to find an optimal location for opening an Italian restaurant near to the city center.

Business Problem

The business problem here is to generate a few most promising neighborhoods in Pune to open an Italian restaurant based on some defined criterias. There are three important criterias regarding neighborhood locations:

- The locations should not already be crowded with restaurants.
- There should not be any Italian restaurants in the vicinity.
- The locations should be as close to the city center as possible.

We will use our data science powers to find optimal neighborhood locations based on defined criterias. We will collect, explore and analyze neighborhood locations and restaurants data to answer following four questions:

1. How many restaurants are present in a radius of 6 km around the city center?
2. How many of the identified restaurants are Italian restaurants?
3. What are distances to the identified restaurants from the city center?
4. What are the locations with no more than two restaurants in a radius of 250 meters and without Italian restaurants in a radius of 400 meters?

Stakeholders

The stakeholder for this project is the Italian restaurant chain who wants to open an Italian restaurant in Pune.

Data

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

1. number of existing restaurants in the neighborhood (any type of restaurant)
2. number of and distance to Italian restaurants in the neighborhood, if any
3. distance of neighborhood from city center

We decided to use a regularly spaced grid of locations, centered around the city center, to define our neighborhoods.

Data Source

Following data sources will be needed to extract/generate the required information:

- **Centers of candidate areas** will be generated algorithmically and approximate addresses of centers of those areas will be obtained using the GeoPy python library.
- **Number of restaurants and their type and location in every neighborhood** will be obtained using **Foursquare API**
- **Coordinate of Pune center** will be obtained using GeoPy python library.

Data Preparation

First, we find geographical coordinates of the Pune city center by converting the address into latitude and longitude values using the GeoPy python library.

```
The geographical coordinate of Pune, India: [18.521428, 73.8544541]
```

We create a grid of neighborhood area candidates which is equally spaced and centered around the city center. These candidate areas are within 6km from Pune city center. We

defined neighborhoods as circular areas with a radius of 300 meters. Distance between two neighborhood centers is 600 meters.

To accurately calculate distances we convert our grid of locations in the Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees).

```
Pune center longitude=73.8544541, latitude=18.521428

Pune center UTM X=379084.4523924125, Y=2048259.8861715223
```

Then we project those coordinates back to latitude/longitude degrees to be shown on the Folium map. Total 364 candidate neighborhood centers are generated.

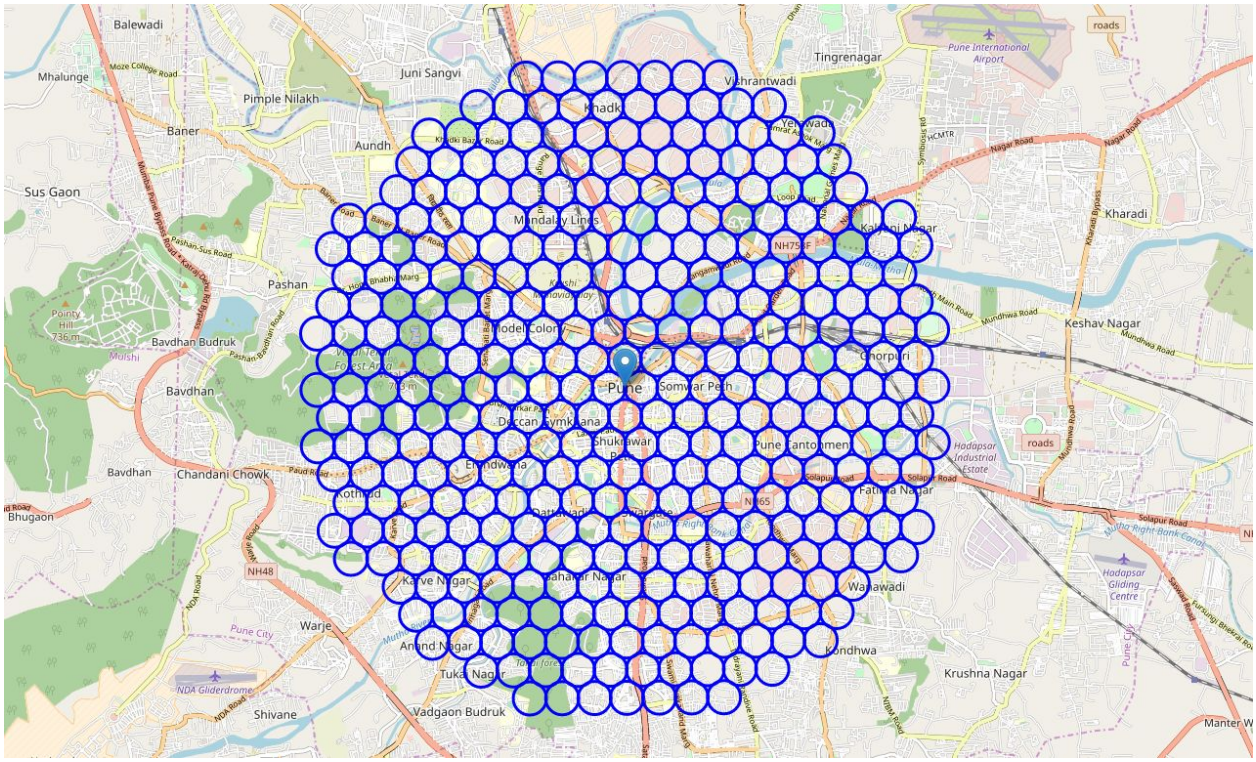

Figure 1: Map of 364 generated candidate neighborhoods

Next, we use the GeoPy python library to reverse geocode and find addresses of each center of the neighborhood grid cell. We also find the distance to each neighborhood center from Pune city center. The list of addresses and its corresponding latitude, longitude, UTM x, UTM y, and distance from center are placed into a tabular form of Pandas data frame.

	Address	Latitude	Longitude	X	Y	Distance from center
0	Sinhagad College Wadgaon Campus, Sinhgad Cante...	18.469675	73.837752	377284.452392	2.042544e+06	5992.495307
1	Tukai Nagar, Pune, 411001	18.469709	73.843434	377884.452392	2.042544e+06	5840.376700
2	Sahakar Nagar, Pune, 411001	18.469744	73.849116	378484.452392	2.042544e+06	5747.173218
3	Sahakar Nagar, Pune, 411001	18.469778	73.854797	379084.452392	2.042544e+06	5715.767665
4	KK Market, KK Market road, Sahakar Nagar, Pune...	18.469813	73.860479	379684.452392	2.042544e+06	5747.173218
5	Sahakar Nagar, Pune, 411001	18.469847	73.866161	380284.452392	2.042544e+06	5840.376700
6	Kondhwa, Pune, 411018	18.469881	73.871843	380884.452392	2.042544e+06	5992.495307
7	Prosperity Society, Tukai Nagar, Pune, 411001	18.474317	73.829198	376384.452392	2.043064e+06	5855.766389
8	Prosperity Society, Tukai Nagar, Pune, 411001	18.474352	73.834879	376984.452392	2.043064e+06	5604.462508
9	Prosperity Society, Tukai Nagar, Pune, 411001	18.474387	73.840561	377584.452392	2.043064e+06	5408.326913

Figure 2: Table of candidate locations containing addresses, location and distance from center

Now we use **Foursquare API** to get venues near the center of candidate location. From the list of venues we filter out restaurants and then narrow down filtering to Italian restaurants. After statistical analysis we get following information about existing restaurants:

```
Total number of restaurants: 552

Total number of Italian restaurants: 24

Percentage of Italian restaurants: 4.35%

Average number of restaurants in neighborhood: 1.3214285714285714
```

We also get a list of restaurants around each candidate location.

```
Restaurants around location
---------------------------
Restaurants around location 101: Abhishek Veg, Purepur Kolhapur, Manali, Pavitra pure veg, Abhishek Non Veg
Restaurants around location 102: Harvest Club, Multi Spice, Vishnu Ji Ki Rasoi, Aswan
Restaurants around location 103: Grace Biryani House
Restaurants around location 104:
Restaurants around location 105: Atharva hotel, sarasbaug chowpati
Restaurants around location 106: Sahil Hotel, Hotel Natraj
Restaurants around location 107:
Restaurants around location 108: Chandan restaurant
Restaurants around location 109: In Green
Restaurants around location 110:
```

Figure 3: Sample list of restaurants around candidate locations

We visualize all the collected restaurants in our area of interest on a map in blue color and Italian restaurants in red color.

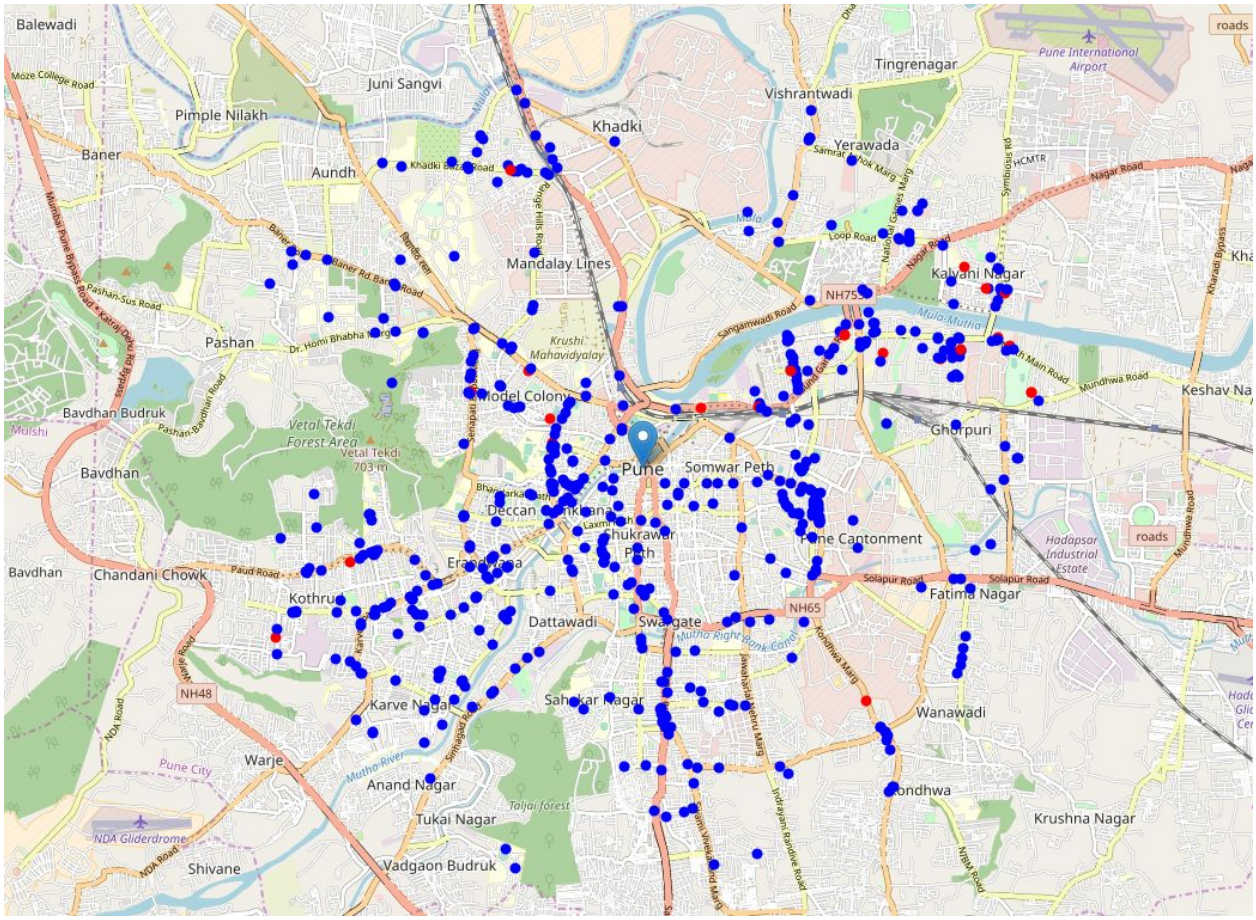

Figure 4: Map of all restaurants including Italian restaurants around Pune city center

So now we have all the restaurants in an area within 6 kilometers from Pune city center and which ones are Italian restaurants! We also know which restaurants exactly are in the vicinity of every neighborhood candidate center.

This concludes the data preparation phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Italian restaurant in Pune!

Methodology

To be updated

Results

To be updated

Discussion

To be updated

Conclusion

To be updated