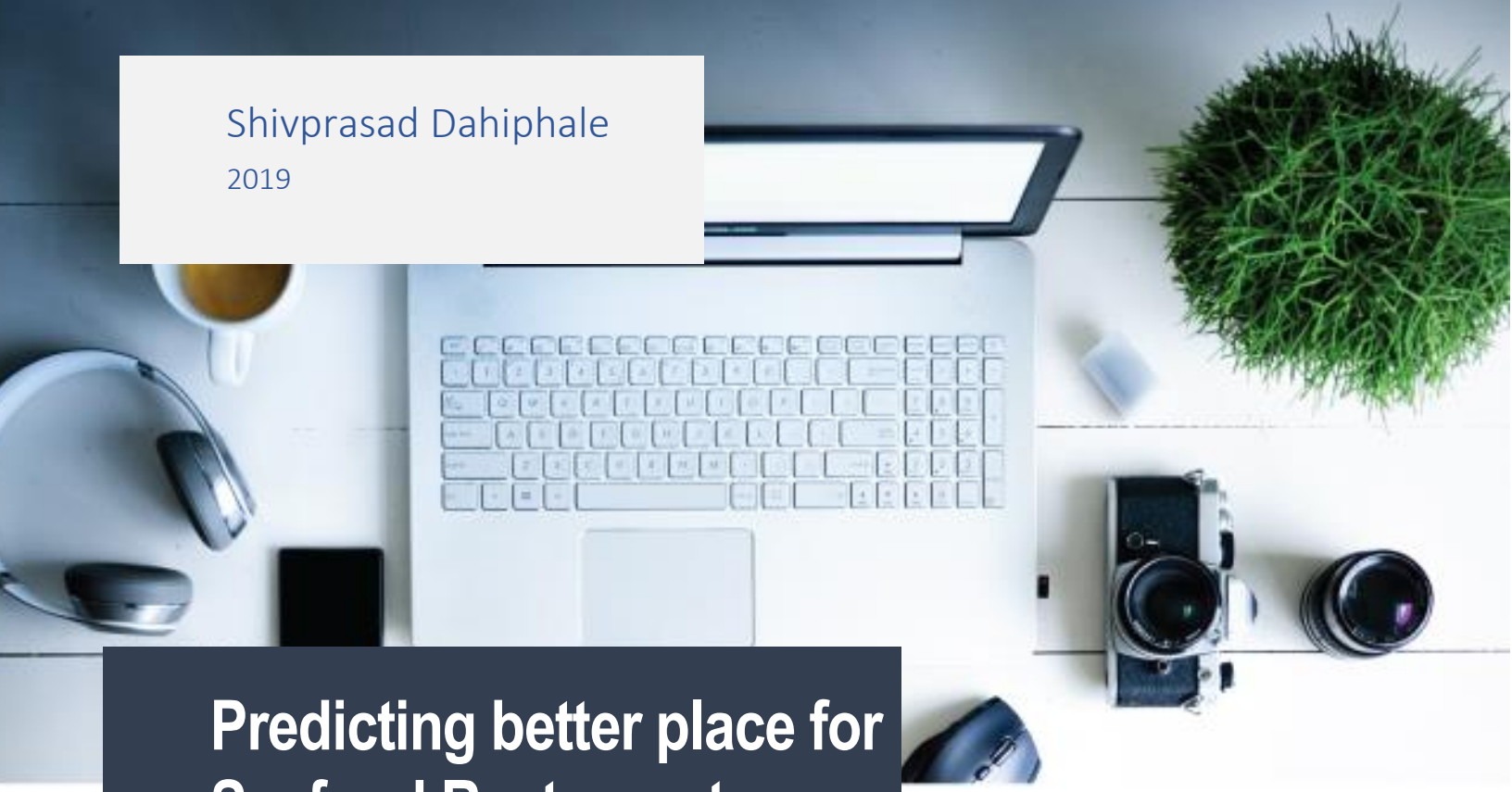


Shivprasad Dahiphale
2019

Predicting better place for Seafood Restaurant

1. Introduction

1.1. Background

Pune, is India's biggest and traditionally most-important city. With biggest IT Industry and a large population of diverse ethnicities, Pune is a fertile ground for opening restaurants and trying out new concepts. This Traditional Indian city houses a large number of vegetarians with Non- vegetarians. Successful restaurant concepts include Seafood, high-end cafes, vegan and vegetarian restaurants etc.

1.2. Problem

Location selection is bigger problem for people who want to start new restaurant in this city. Our few customers want to start new seafood restaurant in Pune city but they don't know the better place to open restaurant. We need solution for such kind of problems, where we can compare similar locations and their nearby area Venus. Similar locations should be grouped and application should have functionality to display those location on map.

1.3. Interest

Obviously new restaurant owners are will be interested in knowing such areas for better places where they can start category restaurant and get more profit and provide better service to nearby people with their delicious variety of food.

2. Data acquisition and cleaning

2.1. Data Source

We have location data available on Google maps. We can get that data to find Pune City areas and coordinates for each area.

That data will be useful to find all category restaurant from Foursquare. Foursquare is important data source for us. From Foursquare we will get all restaurant with their categories and their tips and rating information.

Foursquare have all the updated data and we can access that data easily using foursquare APIs.

2.2. Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. We need to join data from Google API data with Foursquare data for future analysis. We will fetch Location names data from Google APIs and suing that Names Foursquare APIs will give nearby Venues including restaurant data.

There is issue with Foursquare data set response which will be in JSON format which we have to convert into tabular format by parsing that data. We want to go deep in restaurant data and other Venue data is not required so that is not useful for us so we need to filter out unwanted data from response dataset.

We have restaurant categories in string list format. For our analysis we need that column with Boolean value so we need to transform categories list to Boolean(TRUE/FALSE) values for is_seafood column.

2.3. Feature Selection

After data cleaning, there were more than 500 records with 12 feature columns. We have sample data of our feature selection.

```
List of all restaurants
-----------------------
('4c8cfd090bcbbfb742a6e4fe', 'Raj Restaurant', 18.520096523157108, 73.85024257438126, 'Shaniwar peth (Near Omkareshwar
('4fa8d86fe4b09ce16568787a', 'Shivsagar', 18.520151692146218, 73.84589420653944, 'JM road, Pune, Mahārāshtra', 257, Tr
('4cdd04494006a143b792e2b2', 'Pride Of Hyderabad', 18.51715297926503, 73.87466669082642, 'Synagogue Street (Off Moledi
('4cd62284a42b236a9978f408', 'Sheetal Hotel', 18.502301724622903, 73.81611918532329, 'Near Karve Statue (Karve Rd) (Ka
('512cd69ee4b07767b1b66b48', 'Deep Restaurant (Bengali Cuisine)', 18.562018808071436, 73.83598777440325, 'Aundh Road,
('4b1f6478f964a520132624e3', 'Mayur Pure Veg', 18.515904104184575, 73.87961615242538, 'M.G.Road, Pune, Mahārāshtra', 2
('50deb44ce4b0f6dd06fa9181', "McDonald's", 18.510699219829938, 73.83865174277742, 'Pune Central (Karve Road), Pune 411
('5053248fe4b0e5434bc49d47', 'KFC (central)', 18.50591758750925, 73.8423112434891, 'India', 252, True, 7721831.7639481
('4d53b70ec8d0f04d67cbb9d4', 'Agriculture College Chowpaty', 18.53315513677621, 73.84637745443582, 'Next to rahul thea
('4f3fd8c1e4b0a5f25242a725', 'Raj Restaurant', 18.50568552899937, 73.82436263116742, 'India', 280, True, 7718980.03149
...
```

3. Exploratory Data Analysis

3.1. Calculation of target variable

Restaurant distance from Pune City center is not a feature in dataset that need to be calculated using coordinates.

We're interested in venues in 'food' category, but only those that are Seafood restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Seafood restaurant' category, as we need info on Seafood restaurants in the neighborhood.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from Pune center. We have also identified Seafood restaurants (according to Foursquare categorization).

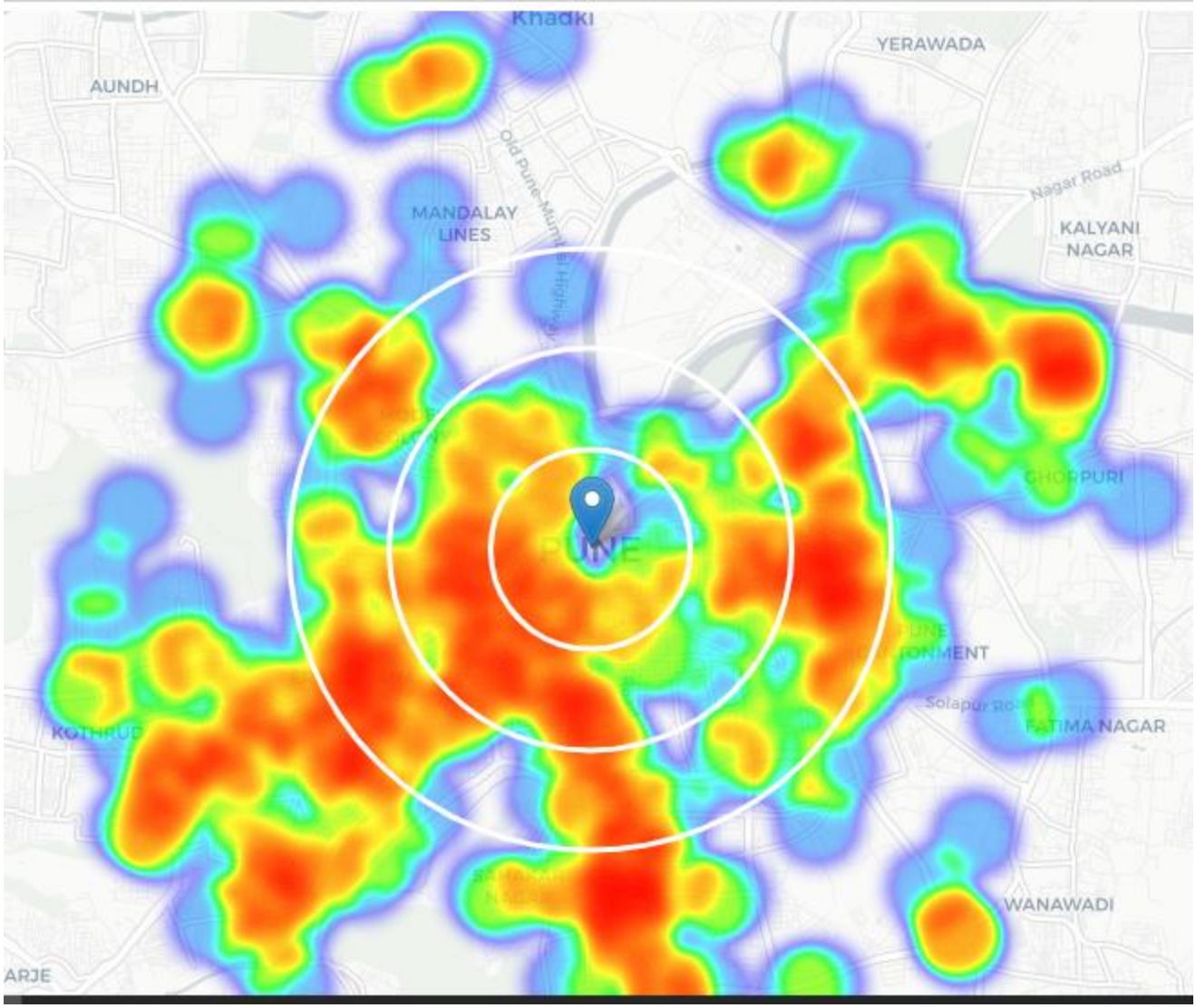
Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Pune - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Seafood restaurants in North Pune) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Seafood restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Seafood restaurant
0	19, Bhagyoday Colony, Suvarn Colony, Hingne Bu...	18.487654	73.819275	7719770.39	3654053.926	8876.556780	1	946.581553
1	33, Sridhar Colony, Karve Nagar, Pune, Maharas...	18.485435	73.823663	7720670.39	3654053.926	8645.418455	3	888.899138
2	411051, 24/4/2, Late Narayanrao Kondiba Jagtap...	18.483215	73.828050	7721570.39	3654053.926	8503.720378	0	1429.838697
3	23/1/A, Anand Vihar Colony Rd, Mahadev Nagar, ...	18.480995	73.832437	7722470.39	3654053.926	8455.960044	0	1872.907960
4	2, Anand Vihar, Mahadev Nagar, Hingne Khurd, P...	18.478776	73.836823	7723370.39	3654053.926	8503.720378	0	2412.579822
5	Unnamed Road, Taljai Forest Area, Pune, Mahara...	18.476557	73.841208	7724270.39	3654053.926	8645.418455	0	3123.027148
6	Sr. No. 8/2+3, Plot No.10, Vivekanand Nagar, T...	18.474338	73.845593	7725170.39	3654053.926	8876.556780	0	3258.601411
7	20, Happy Colony, Kothrud, Pune, Maharashtra 4...	18.494981	73.813976	7718270.39	3654833.349	8750.384168	3	1356.523822
8	Rokdoba Super Market,mavale aali Chowk, Karve ...	18.492761	73.818364	7719170.39	3654833.349	8355.789794	0	550.104500
9	Vitthal Mandir Karvenagar, Karve Nagar, Pune, ...	18.490540	73.822753	7720070.39	3654833.349	8042.961089	3	360.944408

3.2. Relationship between all restaurants and Seafood restaurants

After finding all category restaurants and seafood restaurants average of Seafood restaurants having in Pune city and area wise average seafood restaurants.



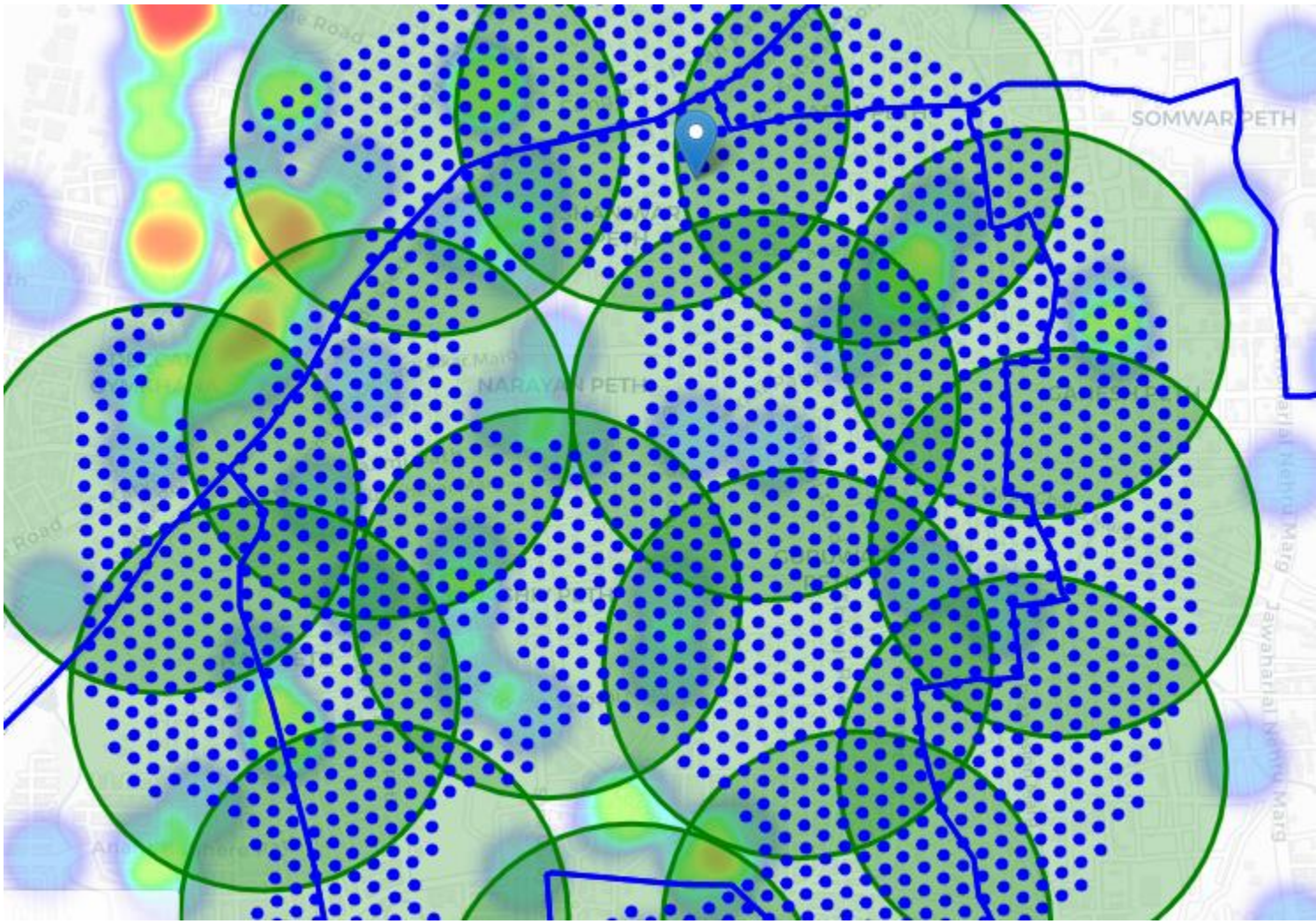
Looks like a few pockets of low restaurant density closest to city center can be found **North from Pune Center**.

4. Predictive Modeling

4.1. Applying Clustering algorithms and their problems

Clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.



4.2. Solution to the problems

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Seafood restaurants nearby, all zones being fairly close to city center (all less than 4km from Pune Center, and about half of those less than 2km from Pune Center). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Wakhad and Model Colony boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

==
Addresses of centers of areas recommended for further analysis
==

216-225, Maharana Pratap Singh Rd, Ganj Peth, Pune, Maharashtra 411002 => 3.0km from pune
Balgandharv Rangamandir Premises, Shaniwar Peth, Pune, Maharashtra 411004 => 1.2km from pune
180, Lal Bahadur Shastri Rd, Navi Peth, Sadashiv Peth, Pune, Maharashtra 411030 => 3.0km from pune
247, Shukrawar Peth, Pune, Maharashtra 411002 => 2.2km from pune
P216, Tulshibaug, Budhwar Peth, Pune, Maharashtra 411002 => 1.0km from pune
705, Kasba Peth, Pune, Maharashtra 411011 => 0.8km from pune
357, Ganesh Peth, Pune, Maharashtra 411011 => 1.6km from pune
216, Thorale Madhavarao Peshve Path, Dandekar Pool, Municipal Colony, Pune, Maharashtra 411030 => 3.5km from pune
297, VG Kothari Path, Narayan Peth, Pune, Maharashtra 411030 => 1.8km from pune
1117, Sathe Colony, Shukrawar Peth, Pune, Maharashtra 411002 => 3.3km from pune
Tilak Bridge, जयंतराव टिळक पूल, Shaniwar Peth, Pune, Maharashtra 411030 => 0.4km from pune
1405, Krantiveer Vasudev Phadake Rd, Perugate, Sadashiv Peth, Pune, Maharashtra 411030 => 2.0km from pune
Pune, Khilarewadi, Erandwane, Pune, Maharashtra 411004 => 2.7km from pune
1750, Raviwar Peth, Pune, Maharashtra 411002 => 2.3km from pune
Unnamed Road, Saras Baug, Pune, Maharashtra 411030 => 3.7km from pune

5. Conclusions

Purpose of this project was to identify Pune areas close to center with low number of restaurants (particularly Seafood restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (North and West Pune), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.