Predicting the optimal location for opening new restaurant in the city of Pune India

Pooja Agarwal
March 6th 2020

Table of contents

1. Introduction

1.1 Background:

Pune is one of the fastest growing cities of India. The city has seen a major growth in population as it is moving ahead in its path of being a hub for IT and Education services. It is also in the process of becoming a SMART city attracting even more people. The property costs in Pune are not as high as other major cities of India. But with good average income, this city is a good candidate for business. Specially Restaurant business as the city has people from all over India as well as people from outside India residing in the city working for the IT companies. It seems to be a good time to open a multi-cuisine or franchise (McD, KFC etc) restaurant in the growing city of Pune.

1.2 Problem:

Best location to open a new restaurant in Pune. The success of a restaurant is affected by its location, area population and other attractions in the area. Hence, it is very important to do a thorough analysis of the optimal location for opening the new restaurant.

1.3 Audience:
Someone looking to open a new restaurant in the city.

1.4 Solution Approach:
In this project we will try to find an optimal location for a restaurant in the city of Pune India.

Using data analysis techniques we will find out the density of restaurants in various neighborhoods of the city. We will also find the top venues of the area.

Then we will find areas with low density of restaurants and some new attractions coming in like a big IT firm opening their office or a new multiplex being constructed in the area etc.

2. Data Acquisition and Cleaning

2.1 For the problem discussed above, I have taken below **data** for the city of Pune, India:

- Neighborhoods of the city
- Map coordinates of the neighborhoods
- Various types of venues in 500m of the neighborhoods
- Current attraction - a new/undergoing project

2.2 Sources:

- Neighborhoods data is taken from Wikipedia
- Coordinates are taken using Google Maps API geocoders
- Venues data is taken from Foursquare API

2.3 Further I will be using Google to check for news of any new/ongoing projects, IT offices, multiplexes being constructed in the areas of choice we get. So that we are able to make an even better suggestion.

2.4 For Example - Lets say after our analysis, we get two neighborhoods n1 and n2 that seems like optimal options for opening a restaurant. Then using the additional data, we can if there is a new/ongoing IT SEZ office with a capacity of 10,000 people is being/newly constructed in (say) n1. On the other hand there are no such activities going on in n2. According to this, we will be able to recommend n1 over n2.

The dataset created contains information about the venues (from Foursquare.com) under 64 unique venue categories for all neighborhoods in the city.

2.5 Additional data points

I will analyze the data further on below points:

- number of venue in restaurants/food category in the neighborhoods
- neighborhoods with nearby venues like IT Center, Malls, Multiplexes or Cineplex, Education centers will be given preference in selection
- neighborhoods with high rate of housing sales will be further given preference

3. Methodology

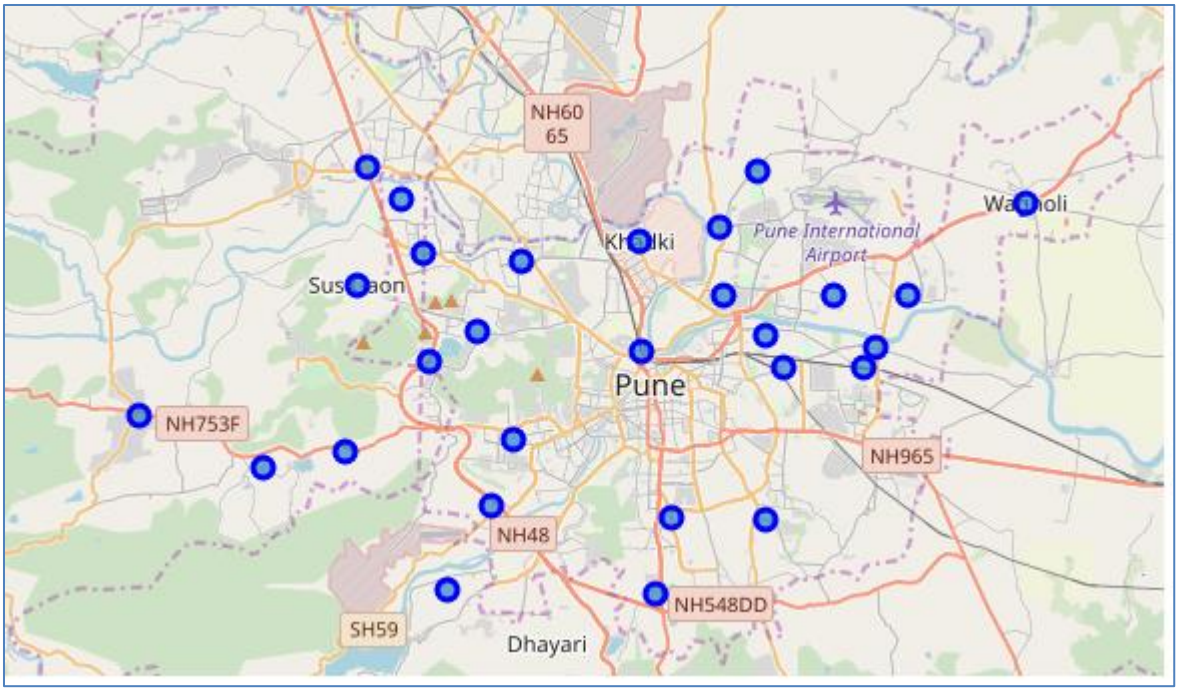
- Using the above data, we will try to find the area in the city which will be optimal for opening a restaurant. This area is expected to have **fewer numbers of restaurants** as compared to other areas.
- Another factor to keep check of is the location of the area in the city. A restaurant closer to cities center has higher chances of success than an area in the outskirts.
- To prepare the data, we will take all neighborhoods and their coordinates using **Foursquare API**.
- Further we will take **trending venues** in the neighborhoods within a radius of 0.5 km. We will categories the neighborhoods with the venue density to get an idea of the neighborhoods with most and least number of Restaurants.
- Then we will take the top 10 most common venues for each neighborhood. Next we will merge it with location coordinates. Then using this data we will create clusters using **k-means clustering** and will plot the clusters on a map using **folium**.
- Finally we will recognize the neighborhoods with low density of restaurants and further add the **Property Rates** of the neighborhoods. To find for a good solution, we will identify the area with optimal venue density and Property Rate. As both these are factors in selecting an area for a new restaurant.

4. Exploratory Data Analysis

- We collected the Pune city neighborhood data from Wikipedia. Then we used Google geocoder the fetch latitude and longitude coordinates for each neighborhood.

	Neighborhood	Latitude	Longitude
0	Aundh	18.561883	73.810196
1	Baner	18.564243	73.776857
2	Bavdhan Khurd	18.313881	74.023109
3	Bavdhan Budruk	18.529135	73.778700
4	Balewadi	18.582027	73.768983

- Next, we plotted the neighborhood data using a folium map



- Then we fetched venues nearby all the neighborhoods in the city using Foursquare explore API
- From the Foursquare lab in the previous module, we know that all the information is in the items key. We used the get_category_type function from the Foursquare lab.
- We created a function to clean the json returned from Foursquare api and structure it into a pandas dataframe for all the neighborhoods in Pune India
- We provided the Foursquare credentials for making a connection for API calls
- Then we declared LIMIT=100 and radius=500m for the Foursquare API call
- Next we have written the code to run the above function on each neighborhood and create a new dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aundh	18.561883	73.810196	Westend mall	18.561814	73.807220	Shopping Mall
1	Aundh	18.561883	73.810196	Cinepolis IMAX	18.561756	73.807192	Multiplex
2	Aundh	18.561883	73.810196	Picantos Mexican Grill	18.560654	73.812447	Mexican Restaurant
3	Aundh	18.561883	73.810196	Venkys Xprs	18.560550	73.808964	Fast Food Restaurant
4	Aundh	18.561883	73.810196	Café Maroo	18.564801	73.809141	Korean Restaurant

- Then we mapped the data to find 5 areas with most and least venues count

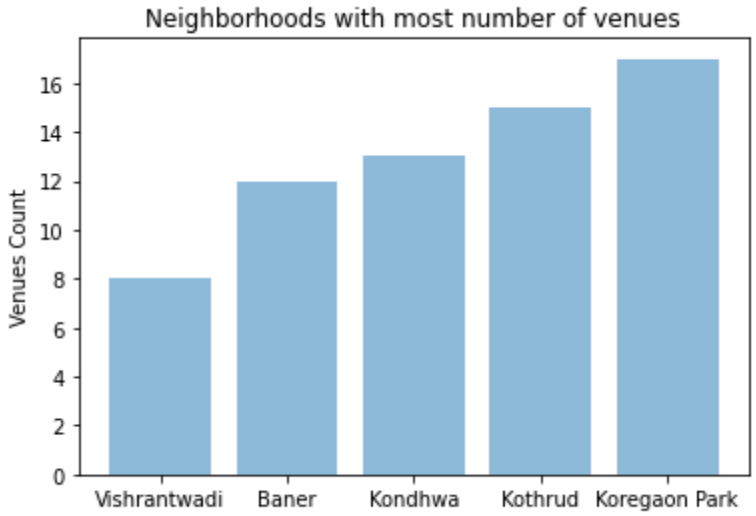
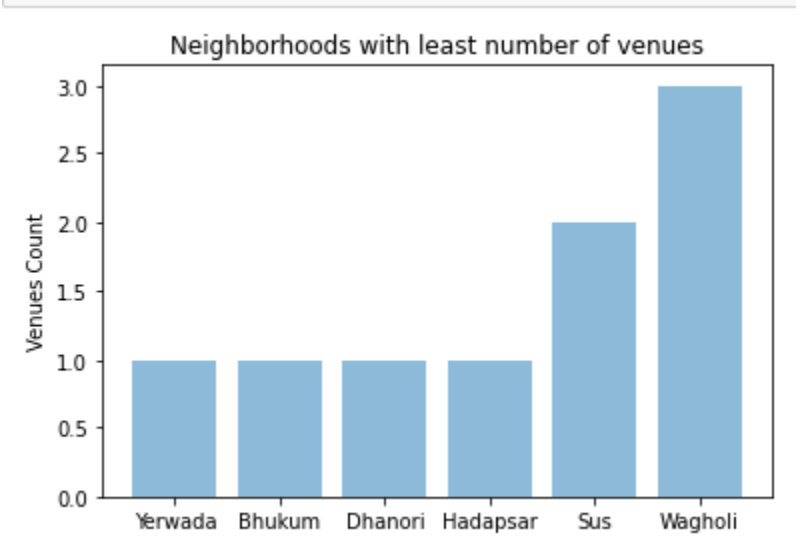


Chart1: Neighborhoods with most number of venues

Chart2: Neighborhoods with least number of venues



- After that we checked the unique categories returned for the city. And there were 68 unique categories.
- Then we found every venue in the neighborhood using **one hot encoding**

	Neighborhood	ATM	Art Gallery	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Breakfast Spot	Bus Line	...	Shoe Store	Shopping Mall	Snack Place	South Indian Restaurant	Sporting Goods Shop	Tea Room	T Restaur
0	Aundh	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	
1	Aundh	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
2	Aundh	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
3	Aundh	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
4	Aundh	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	

5 rows × 69 columns

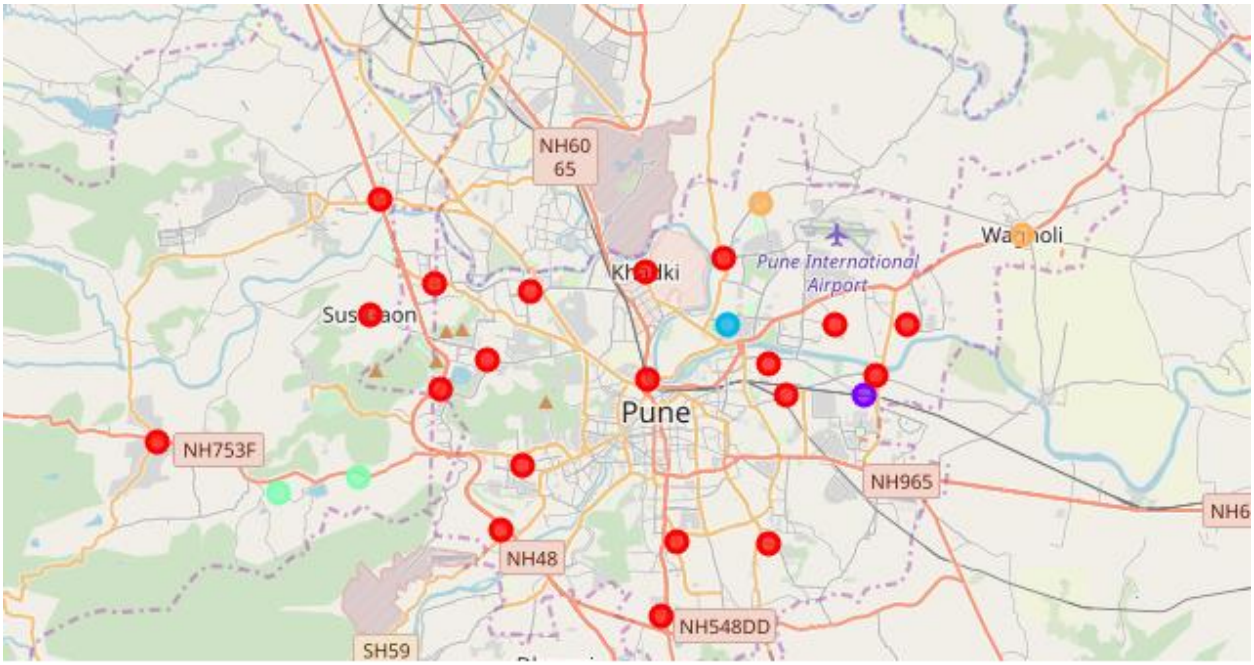
- Next we created the new dataframe and display the top 10 venues for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aundh	Fast Food Restaurant	Indian Restaurant	Ice Cream Shop	Clothing Store	Sporting Goods Shop	Shopping Mall	Vegetarian / Vegan Restaurant	Jewelry Store	Juice Bar	Kids Store
1	Baner	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Motorcycle Shop	Grocery Store	Gourmet Shop	Café	Seafood Restaurant	Flea Market	Exhibit
2	Bavdhan Budruk	Lake	Indian Restaurant	Fast Food Restaurant	Gourmet Shop	Exhibit	Farmers Market	Flea Market	Food	Fruit & Vegetable Store	Zoo
3	Bhugaon	ATM	Seafood Restaurant	Lake	Eastern European Restaurant	Exhibit	Farmers Market	Fast Food Restaurant	Flea Market	Food	Fruit & Vegetable Store
4	Bhukum	ATM	Fruit & Vegetable Store	Eastern European Restaurant	Exhibit	Farmers Market	Fast Food Restaurant	Flea Market	Food	Gourmet Shop	Vegetarian / Vegan Restaurant
5	Bibvewadi	Ice Cream Shop	Seafood Restaurant	Gym	Bakery	Café	Exhibit	Farmers Market	Fast Food Restaurant	Flea Market	Food

- After this we created cluster on Neighborhoods. We used k-means clustering approach to cluster the neighborhood into 5 clusters.
- Then we created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 Mo Commo Venu
0	Aundh	18.561883	73.810196	0	Fast Food Restaurant	Indian Restaurant	Ice Cream Shop	Clothing Store	Sporting Goods Shop	Shopping Mall	Vegetarian / Vegan Restaurant	Jewelry Store	Juice Bar	Kid Sto
1	Baner	18.564243	73.776857	0	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Motorcycle Shop	Grocery Store	Gourmet Shop	Café	Seafood Restaurant	Flea Market	Exhib
2	Bavdhan Budruk	18.529135	73.778700	0	Lake	Indian Restaurant	Fast Food Restaurant	Gourmet Shop	Exhibit	Farmers Market	Flea Market	Food	Fruit & Vegetable Store	Zo
3	Shivajinagar	18.532591	73.851311	0	Department Store	Restaurant	Indian Restaurant	Hospital	Hotel	BBQ Joint	Bakery	Art Gallery	Asian Restaurant	Hooka B
4	Bibvewadi	18.478174	73.862105	0	Ice Cream Shop	Seafood Restaurant	Gym	Bakery	Café	Exhibit	Farmers Market	Fast Food Restaurant	Flea Market	Foo

- Finally, we visualize the resulting clusters



- From above it was seen that we are able to select some neighborhoods that look potential optimal options. Like Hadapsar, Yerwada, Dhanori and Wagholi.
- Then we included the Property Rates in price per sq ft. for the shortlisted areas.
- Finally we added the distance of these neighborhoods to Pune cities center. To the dataframe.

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Property_Rates	Distance from center of City
0	Dhanori	18.590671	73.891319	4	Restaurant	Zoo	Food	4632	0.078445
1	Hadapsar	18.526967	73.927825	1	Train Station	Zoo	Fruit & Vegetable Store	5100	0.073580
2	Wagholi	18.580630	73.983310	4	Indian Restaurant	Pizza Place	Restaurant	3782	0.141805
3	Yerwada	18.550361	73.879746	2	Donut Shop	Fruit & Vegetable Store	Eastern European Restaurant	6758	0.038429

5. Results and Discussion

- Our analysis shows that above are the areas with low density of venues as compared to other areas of the city. Interesting to note that three of these areas are not very far from the city center.
- On further analyzing these areas on factors like population density, and other major attractions in the area, we see that the area 'Yerwada' contains a lot of Business Parks, big Hotels, golf course ground and more. Even though property rates are a nit high but this area also has least distance from city center. This looks like a good trade off.
- Another location that looks promising is 'Hadapsar' where on further analysis we found that a lot of new construction is going on in this area. And it is nearby to international airport.

6. Conclusion

- The purpose of this project was to identify an optimal location to open a new restaurant in the city of Pune India.
- The above data has been created using the neighborhoods of the city, the coordinates and the various venues density in the neighborhoods. Using the method of clustering, the above locations were divided in to clusters depending on most famous venues in the areas. As expected the old and highly developed/expensive parts of the city were found part of one cluster. That proved the accuracy of the measures. Above areas listed as optimal options were not parts of such clusters, hence good candidates for the problems solution.
- The results can help an interested audience to take an informed decision. As there is no limitation to data, more factors can be taken into account along with above information to finalize a location. Details like new/ongoing projects, Business Parks or Cineplex in the area, connectivity of metro lines to the area, population of the area etc. are few factors that can add value to take a decision.

7. References

- Google geocoder API
- Foursquare API
- https://www.99acres.com/property-rates-and-price-trends-in-pune
- https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Pune