Recommending a Business in Pune, India

Vijay Kumar, Jan 30,2021

1. Introduction

1.1 Business Problem

A person or businessman searches for a venue in a new city or want to open a new business, they're highly interested in the best places that the city has to offer. The person might want to know what kind of business they can start in a particular area or what are the most common and best business present in a location. If they know this valuable information then it certainly would help decide which business to open in a neighbourhood around the city.

Pune is known for its educational facilities, having more than hundred educational institutes and nine universities. Pune has well-established glass, sugar and metal forging industries since the 1950s. Pune is also home to numerous small and medium-sized enterprises. Pune Infotech Park has emerged as the leading IT centre in not just Pune, but the entire country. Pune is also emerging as a prominent city for Business Process Outsourcing (BPO) companies. Today, Pune has many complexes, shopping malls, restaurant and multiplexes, if you are the person thinking of opening a business then Pune will be the right place.

1.2 Interested audience

Anyone who is thinking to open a business or wants to explore the business can use this project, it would certainly help him to find the best business option or even the best venues.

2. Data Gathering

Using the BeautifulSoup library, I scarped all the neighbourhood form Wikipedia (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Pune).

First 5 scraped data from Wikipedia.

	Negihborhod
0	Aundh
1	Balewadi
2	Baner
3	Bavdhan
4	Bhavani Peth

Next from Geocoder python library, I extracted the latitude and longitude of all neighbourhoods. Below is the snapshot.

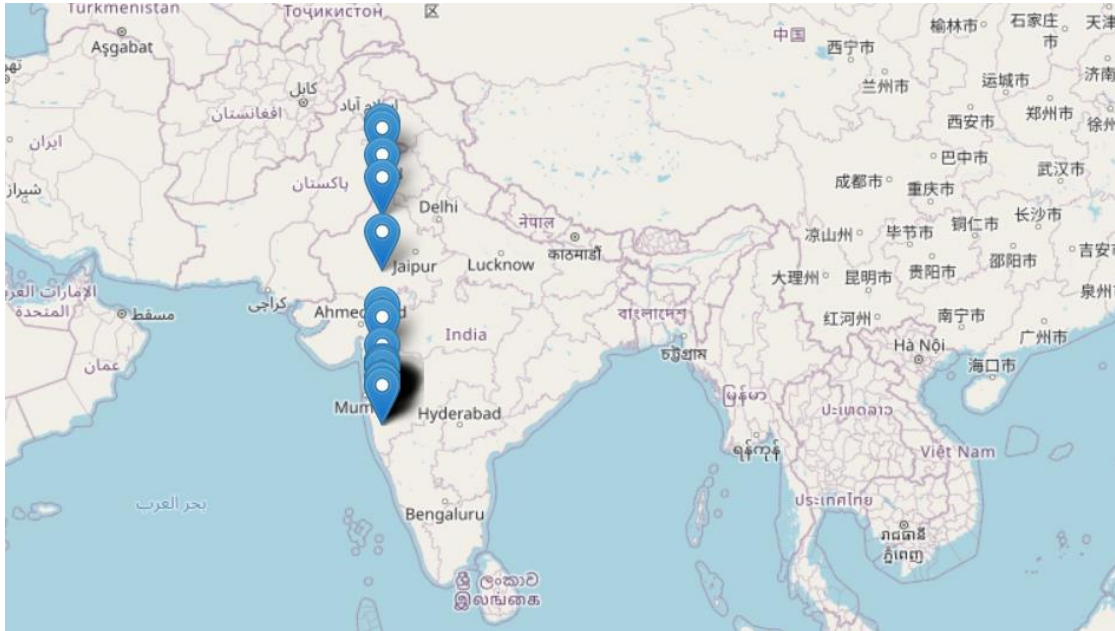|	Neighborhood	Latitude	Longitude
0	Aundh	18.561883	73.810196
1	Balewadi	18.582027	73.768983
2	Baner	19.532282	82.808697
3	Bavdhan	18.520954	73.778087
4	Bhavani Peth	17.689937	75.905524

And finally in this section I utilized Foursquare API to get all venues around all neighbourhoods. Below is the snapshot of final Dataframe.

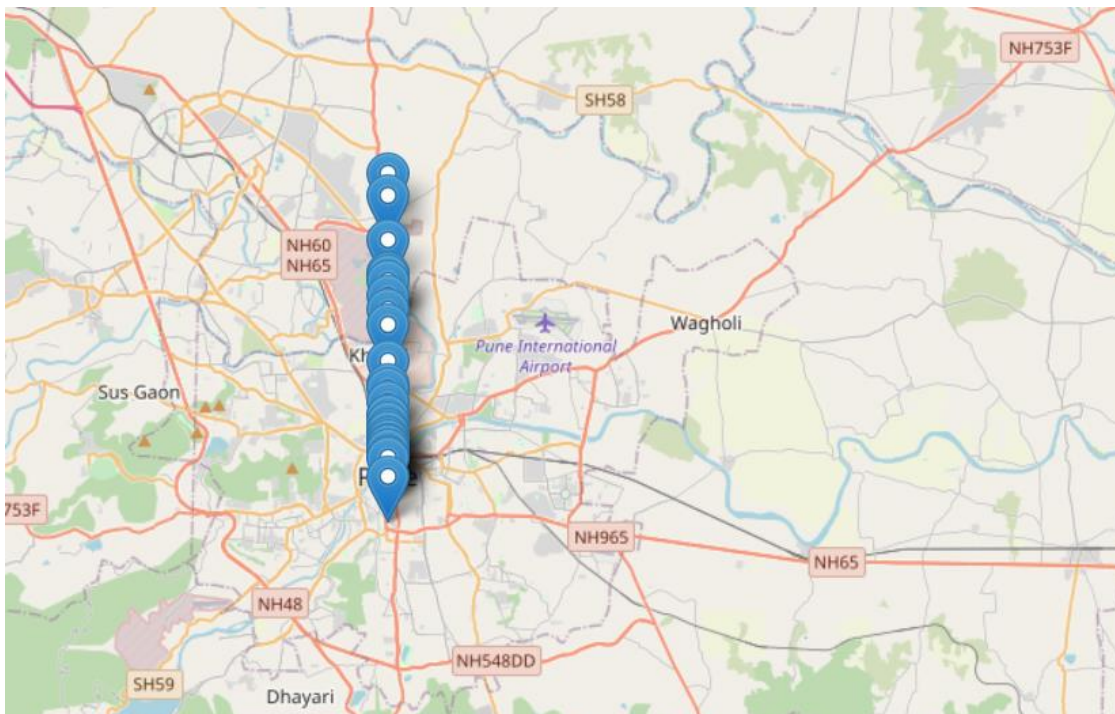|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aundh	18.561883	73.810196	Crosswords	18.556177	73.809131	Bookstore
1	Aundh	18.561883	73.810196	Westend mall	18.561814	73.807220	Shopping Mall
2	Aundh	18.561883	73.810196	Yolkshire	18.553576	73.806888	English Restaurant
3	Aundh	18.561883	73.810196	Starbucks	18.556595	73.809153	Coffee Shop
4	Aundh	18.561883	73.810196	Naturals Icecream	18.556192	73.809169	Ice Cream Shop

3. Data Cleaning

From figure 1 we can see there are some outliers in the dataset. This dataset need be filtered out, result which have latitude greater than 18 and longitude greater than 73.

Figure 2 is the filtered dataset.

Figure 1: Coordinates Returned by Foursquare

Figure 2: Filtered result map

4. Methodology and Exploratory Data Analysis

In this section I analysed the top venue types that exits in Pune using the filtered dataset from previous step. Next I explored the venues to better understand the location of various venues and places.

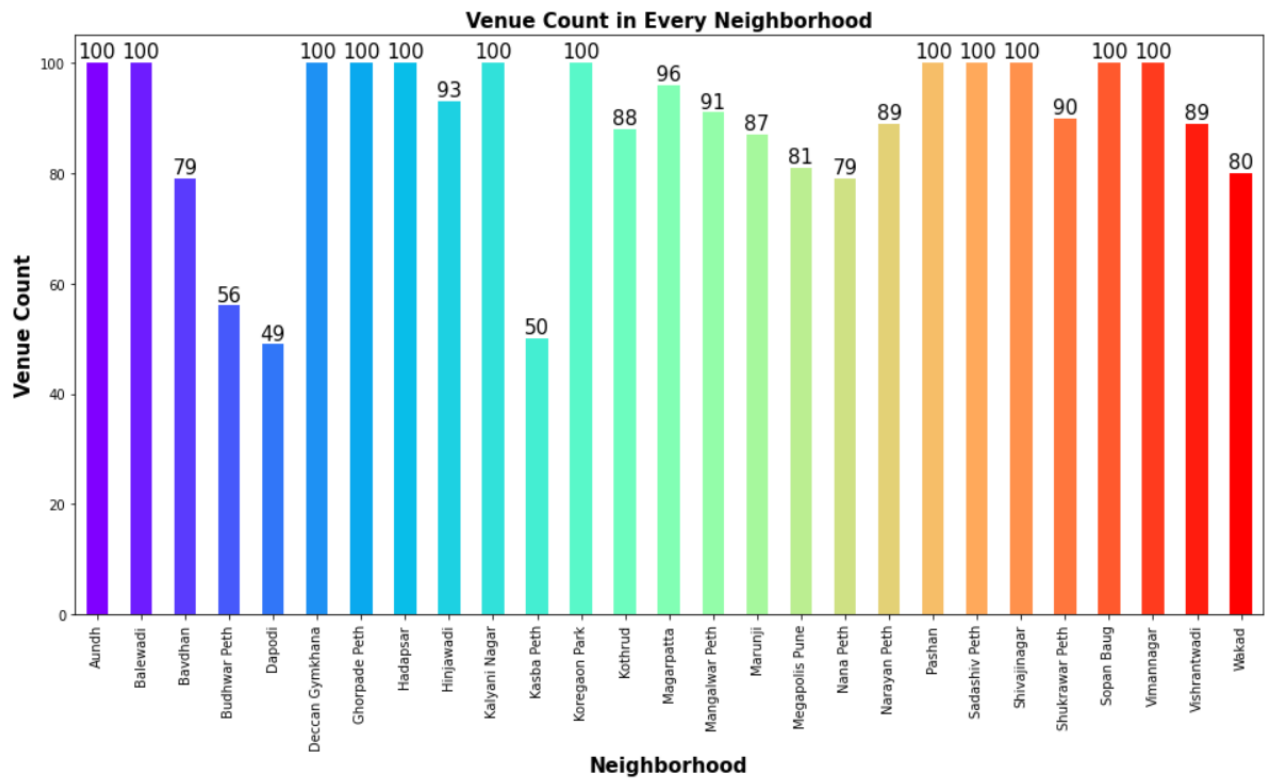First, I plotted all the neighbourhoods and venue counts for every neighbourhoods

Figure 3: Venue Count of neighbourhoods

From Fig. 3 we can see many of the neighbourhood have 100 venues, so I analysed the single neighbourhood which have 100 venues. The reason for choosing the neighbourhood having 100 venues is there will be more number of categories to find.

I started analysing the single neighbourhood 'Pashan' which have total 100 venues. Using foursquare API to get all venues of 'Pashan'.

I used Pie chart and a bar chart to visualize the data and the venues of 'Pashan'.
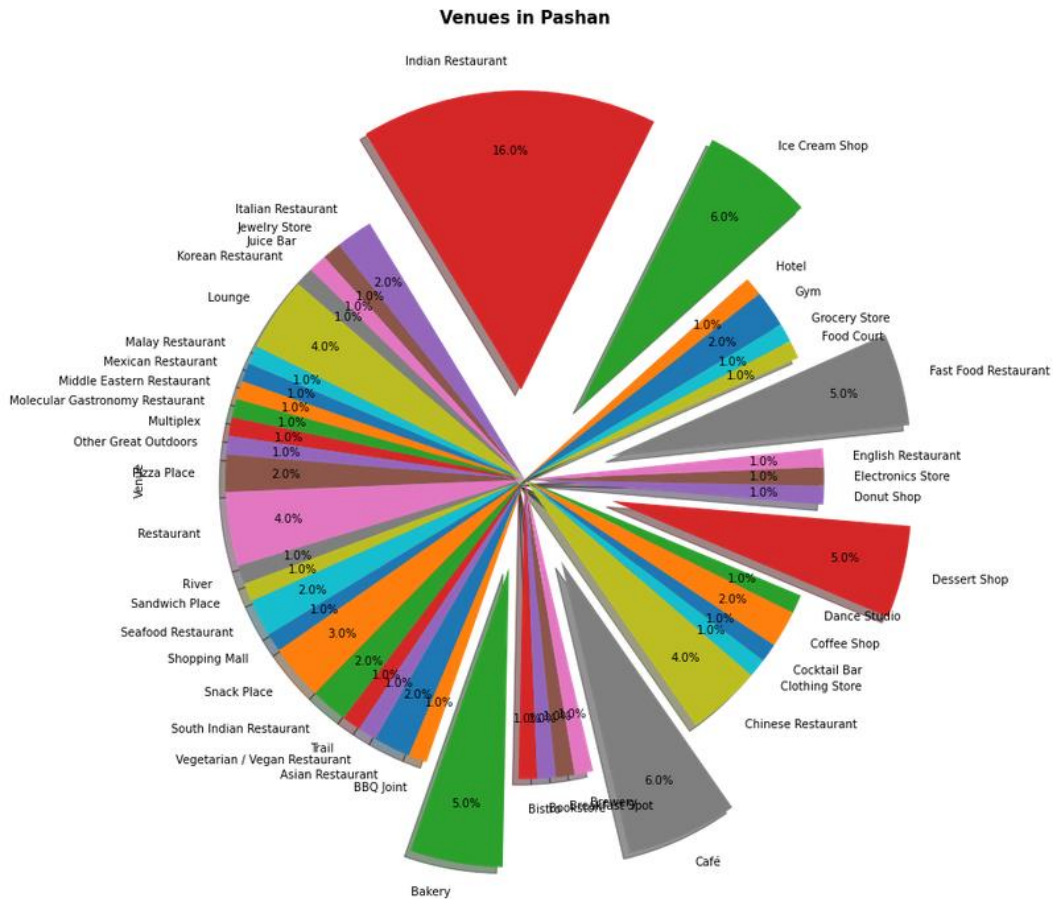
Figure 4: Venue Count in Pashan neighbourhoods

As we can see 16% of them are the Indian Restaurant followed by Bakery, Ice Cream Shop, Café and fast-food Restaurant.
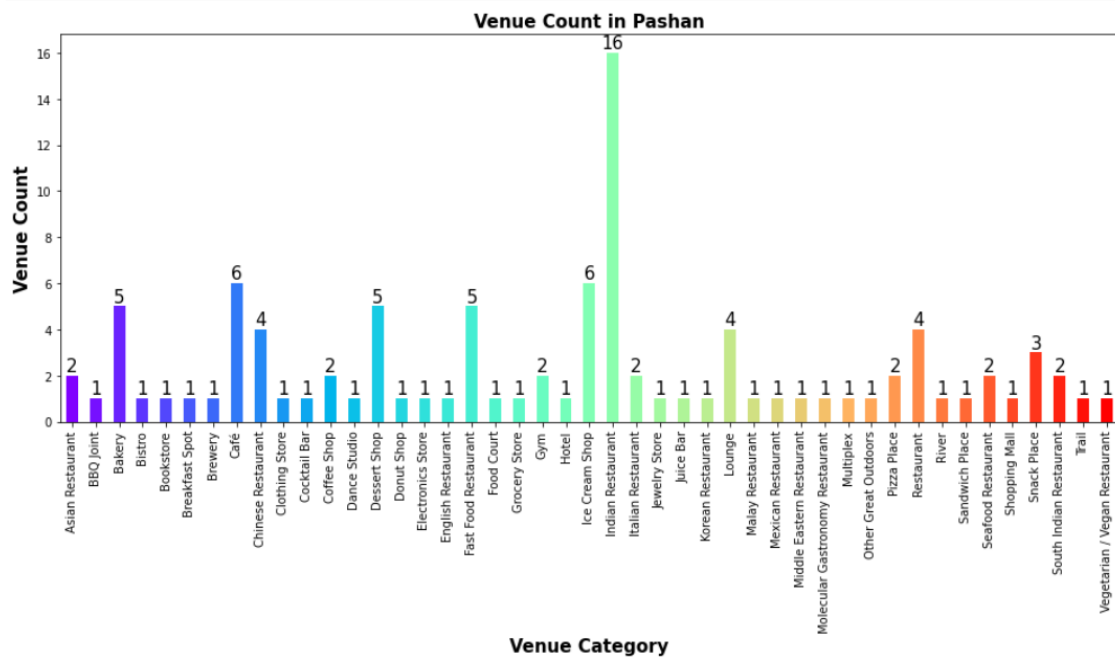

Figure 5: Venue Category Count in Pashan

Fig 5 shows Indian Restaurant is maximum in Pashan, also we can see there are some Multicuisine Restaurant present. If you are the kind of person who likes to try different cuisines , Pashan would be the right place for you.

Foursquare.com API can also be able to leverage on the data to find out the top common nearby venues and their categories in each of these neighbourhoods. This is critical as we want to recommend a neighbourhood whereby the category of a particular venue is low (lower competition).

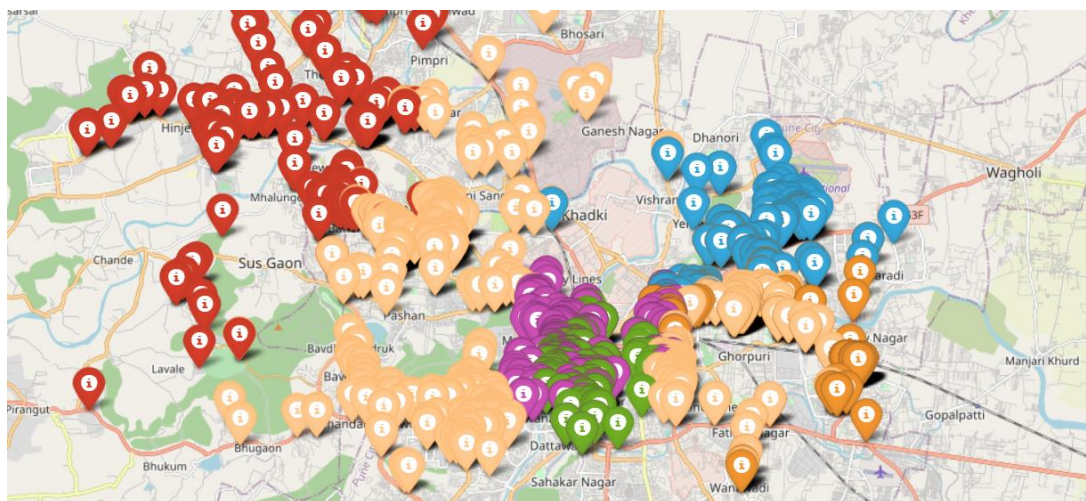|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aundh	Indian Restaurant	Café	Ice Cream Shop	Restaurant	Fast Food Restaurant	Lounge	Dessert Shop	Bakery	Snack Place	Bus Station
1	Balewadi	Indian Restaurant	Fast Food Restaurant	Café	Hotel	Sandwich Place	Lounge	South Indian Restaurant	Pizza Place	Chinese Restaurant	Ice Cream Shop
2	Bavdhan	Indian Restaurant	Café	Coffee Shop	Fast Food Restaurant	Bakery	Dessert Shop	Sandwich Place	Chinese Restaurant	Lounge	Snack Place
3	Budhwar Peth	Indian Restaurant	Snack Place	Vegetarian / Vegan Restaurant	Seafood Restaurant	Dessert Shop	Ice Cream Shop	Fast Food Restaurant	Tea Room	Coffee Shop	Convenience Store
4	Dapodi	Indian Restaurant	Ice Cream Shop	Bakery	Chinese Restaurant	Gym	Fast Food Restaurant	Snack Place	Café	Coffee Shop	Dessert Shop

Figure 6: Most Common venues in all Neighbourhood

5. Clustering

To identify the similarities in all venues, K-Mean clustering algorithm is used to cluster them in 6 different groups.

I used colour code to distinguish the cluster from other clusters.

1. Colour Orange represent Cluster 0
2. Colour Green represent Cluster 1
3. Colour Beige represent Cluster 2
4. Colour Red represent Cluster 3
5. Colour Blue represent Cluster 4
6. Colour Purple represent Cluster 5

Figure 7: Clusters of Neighbourhoods

I further analysed each cluster to find out which venue category is common in each cluster and from that it has been found in all clusters Indian Restaurant is the most common venue.

```
Most Comon venue in Clusters 2 is Indian Restaurant

Most Comon venue in Clusters 5 is Indian Restaurant

Most Comon venue in Clusters 3 is Indian Restaurant

Most Comon venue in Clusters 0 is Indian Restaurant

Most Comon venue in Clusters 1 is Indian Restaurant

Most Comon venue in Clusters 4 is Indian Restaurant
```
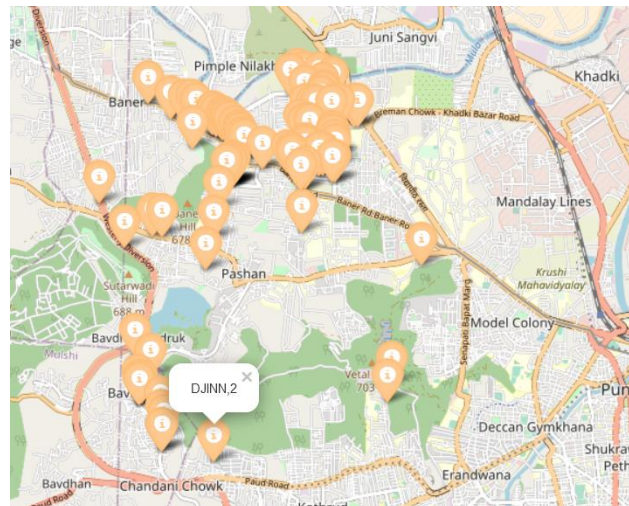
Figure 8: Common Venues in all Clusters

I analysed a single neighbourhood to see most common and least common venue.

```
1st Most Common venues in Pashan is Indian Restaurant.
```

Figure 9: Common Venues in Pashan

```
6th Most Common venues in Pashan is Fast Food Restaurant.
```

Figure 10: 6th Common Venues in Pashan

Figure 11: Venues in Pashan

6. Results and Discussion

After collecting data from the Foursquare API and Wikipedia, got a list of 56 different neighbourhoods. However, not all venue has the correct coordinates so 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 neighbourhood count of 27.

Based on the result above, in all clusters *Indian Restaurant* is the most common followed by *Fast-food*. So, if you like to open any business you certainly don't want to open an *Indian restaurant* specifically in neighbourhood which have 100 venues unless you have a backup budget.

You can think of a business which is neither least common nor most common in any of the neighbourhood, for example in Hinjewadi Most Common venue is Indian Restaurant and the lest common venue is a food court, reason may be because there are multiple Restaurant present in the area. But A Café, Sandwich place and a Pizza Place are neither most common nor least common, so opening a Sandwich place or a Café place might be a good idea.

7. Conclusion

The purpose of this project is to explore different kind of venues and business in Pune city. I have concluded that if you focus on opening a particular business and if you have a backup budget then you can certainly open a business which is most popular in a particular area and if not, you can think of a business which is neither most common nor least common.

Project can be used to explore venues of any city and can take an idea what kind of business he/she can run in the city.