

Capstone Project

The Battle of Neighborhoods

Finding best places for Italian restaurants in Pune, India

Date: 19th August, 2020
Author: Shubhranshi Kapoor

1. Introduction

1.1 Background :

Pune, also known as Poona, is a bustling metropolis in the Indian State of Maharashtra and the eighth most populous city in India, with an estimated population of 7.4 million as of 2020. It has been ranked as "the most livable city in India" several times. Pune is widely regarded as the second major "IT hub of India" and the top "automobile and manufacturing hub of India". It is also known as the "Oxford of the East" due to the presence of several well-known educational institutions. 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. Research institutes of information technology, education, management and training attract students and professionals from India and overseas.

With its diverse culture, come diverse food items. There are many restaurants in Pune City, each belonging to different categories like Chinese, Italian, and French etc. apart from wide variety of Indian cuisines. So as part of this project, we will list and visualize all major parts of Pune City.

1.2 Problem :

The problems and questions we will be addressing in this project are:

- Which places have best restaurants in Pune?
- Which places have worst restaurants in Pune?
- Which place has highest number of restaurants in Pune City?
- Which place has lowest number of restaurants in Pune City?
- What area has most number of Italian restaurants in Pune?
- What area has highest rating of Italian restaurant in Pune?

2. Data Acquisition and Preprocessing

2.1 Data Collection :

For this project we need the following data :

1. Pune Restaurants data that contains list Locality, Restaurant name, Rating along with their latitude and longitude.
 - Data source : Zomato kaggel dataset : <https://www.kaggle.com/shrutimehta/zomato-restaurants-data> (<https://www.kaggle.com/shrutimehta/zomato-restaurants-data>)
 - Description : This data set contains the required information. And we will use this data set to explore various locality of Pune city.
2. Nearby places in each locality of Pune city.
 - Data source : Fousquare API : <https://developer.foursquare.com/> (<https://developer.foursquare.com/>)
 - Description : By using this api we will get all the venues in each neighborhood.

2.2 Approach :

1. Collect the Pune city data from Zomato kaggel dataset.
2. Using FourSquare API we will find all venues for each neighborhood.
3. Filter out all venues that are nearby by locality.
4. Using aggregative rating for each resturant to find the best places.
5. Visualize the Ranking of neighborhoods using folium library(python).

2.3 Data Cleaning :

The dataset consists of restaurants from many cities in India.

1. So first step is to filter out data for only ‘Pune’ city.

In [40]:

```
df_PUNE = df_data[df_data['city'] == 'Pune']
df_PUNE.reset_index(drop=True, inplace=True)
df_PUNE.head()
```

Out[40]:

	zomato_url	name	city	area	rating	rating_count	telephone	cusine
0	https://www.zomato.com/pune/indo-western-food-station	Indo Western Food Station	Pune	Akurdi	3.5	20.0	+91 7020558356	Cafe, Chinese
1	https://www.zomato.com/pune/gangotree-akurdi	Gangotree	Pune	Akurdi	3.9	141.0	020 27640469 +91 9096554201	South Indian, North Indian, Street Food
2	https://www.zomato.com/pune/shri-krishna-pure-veg-restaurant	Shri Krishna Pure Veg Restaurant	Pune	Akurdi	3.4	55.0	+91 7517529996 +91 7517529997	North Indian, Chinese

2. Remove the unwanted columns and rows from dataset

In [41]:

```
#df_Res= df_PUNE[df_PUNE.Longitude !=0.000000][['Restaurant Name', 'Locality', 'Longitude', 'Latitude', 'Cuisines', 'Aggregate rating', 'Rating text', 'Votes']]
df_PUNE = df_PUNE.dropna()
df_PUNE = df_PUNE.reset_index(drop=True)

df_Res= df_PUNE[df_PUNE.longitude !=0.000000]
df_Res= df_PUNE[df_PUNE.latitude !=0.000000]
df_Res = df_Res[df_Res.rating !=0.00]
df_Res= df_Res[['name', 'area', 'longitude', 'latitude', 'cuisine', 'rating', 'rating_count']]
df_Res.rename(columns={'name': 'Restaurant Name', 'area': 'Locality', 'longitude': 'Longitude', 'latitude': 'Latitude', 'cuisine': 'Cuisines', 'rating': 'Aggregate rating', 'rating_count': 'Votes'}, inplace=True)

df_Res.head()
```

Out[41]:

	Restaurant Name	Locality	Longitude	Latitude	Cuisines	Aggregate rating	Votes
0	Gangotree	Akurdi	73.765289	18.648924	South Indian, North Indian, Street Food	3.9	141.0
1	Expresso Royle	Akurdi	73.766440	18.645427	Fast Food, Cafe	3.7	90.0
2	Cafe Hashtag	Akurdi	73.767460	18.650769	Cafe	2.5	51.0
3	Sangeeta Restaurant & Bar	Akurdi	73.783322	18.650569	Chinese, North Indian	3.7	90.0
4	Naivedyam	Akurdi	73.786870	18.649210	Maharashtrian	3.7	105.0

3. Use Foursquare API to explore all the locality and venues in Pune.

In [57]:

```
pune_venues.groupby('Locality').count()
```

Out[57]:

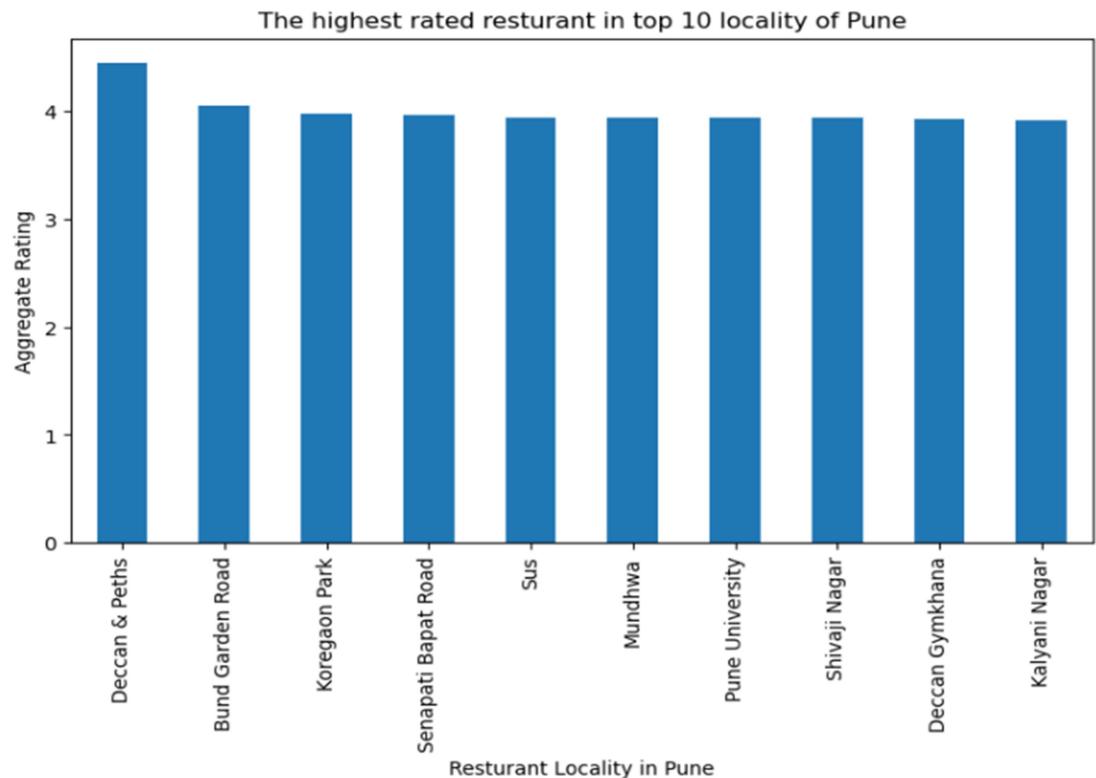
Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Locality						
Akurdi	9	9	9	9	9	9
Aundh	40	40	40	40	40	40
B.T. Kawade Road	8	8	8	8	8	8
Balewadi	4	4	4	4	4	4
Baner	19	19	19	19	19	19

3. Methodology

3.1 Exploratory Analysis:

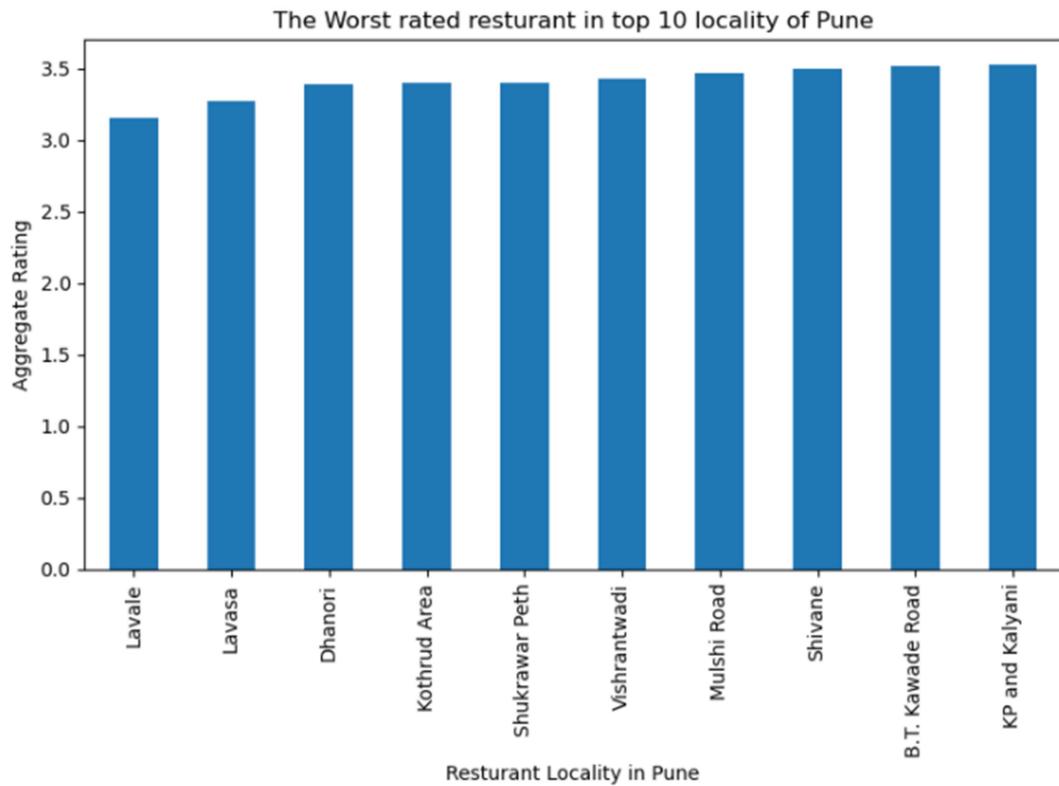
After cleansing the datasets to more useful forms, I created some visualizations to interpret the data we have better. I used bar charts and maps to explore the datasets to address questions in this project.

1. Which places have best restaurant in Pune



The best restaurants are available in **Deccan & Peths**

2. Which places have worst restaurants in Pune



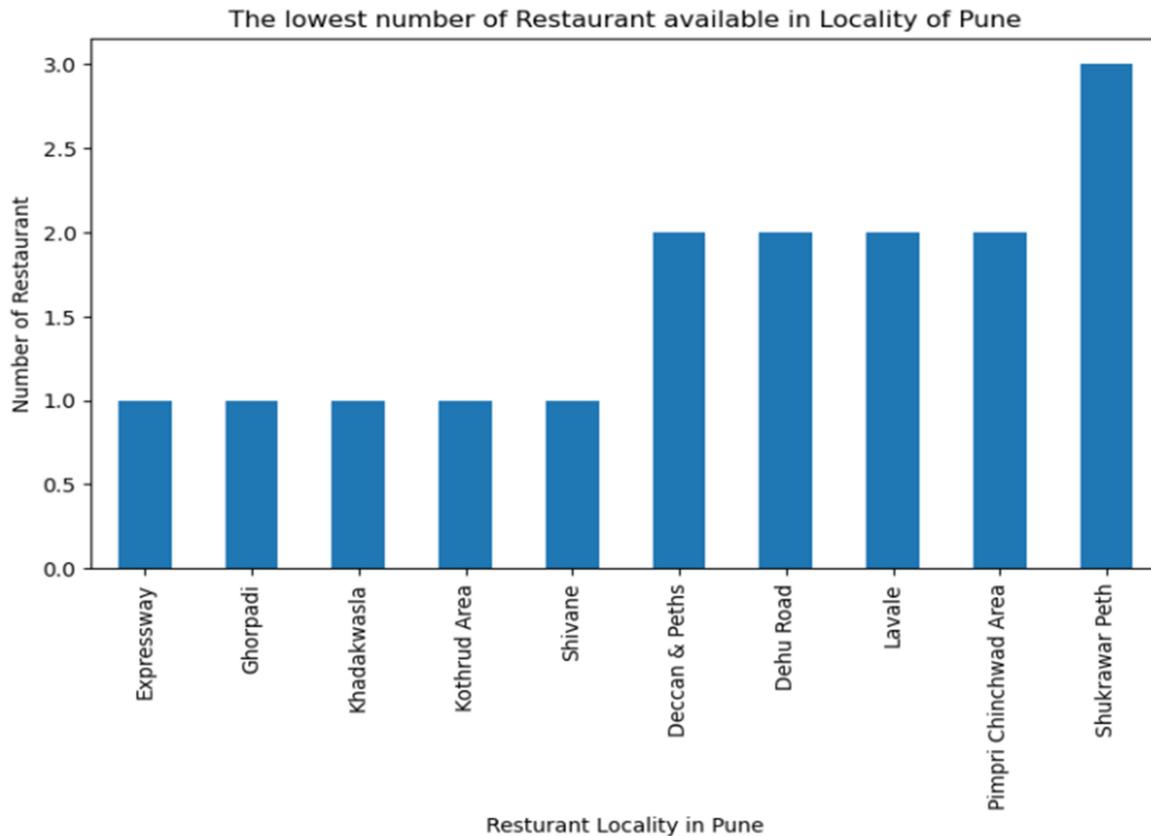
The worst restaurants are in **Lavale**

3. Which place has highest number of restaurants in Pune city



Baner and Viman Nagar has highest number of restaurants in Pune City

4. Which place has lowest number of restaurants in Pune city



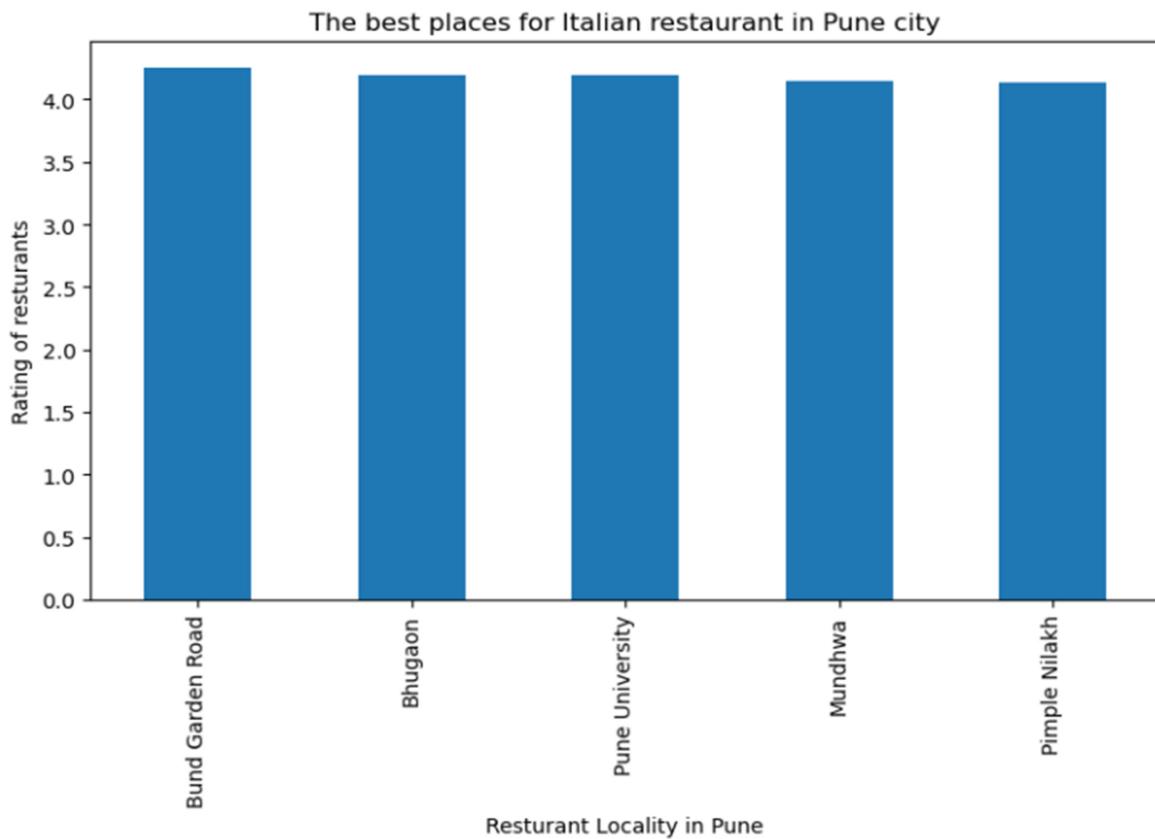
Expressway, Khadakwasla , Kothrud Area have lowest number of restaurants

5. Which area has most number of Italian restaurants in Pune city



Baner is the best place for Italian restaurant.

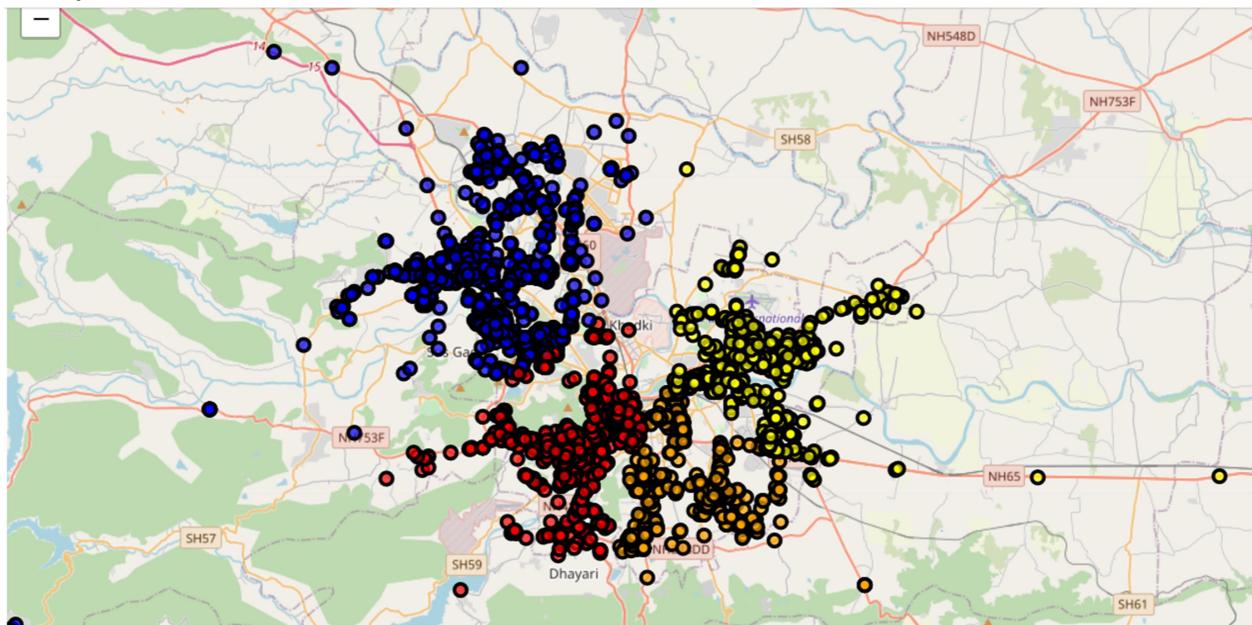
6. Which area has highest rating of Italian restaurants in Pune



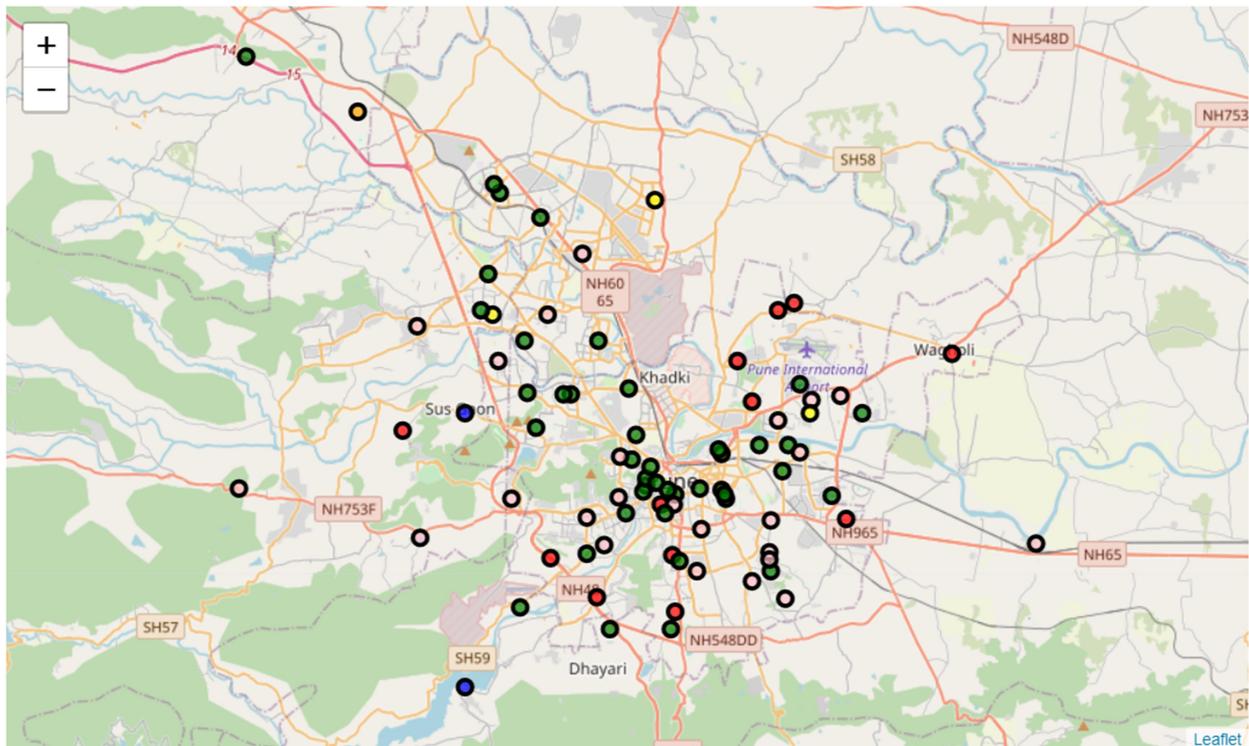
Bund Garden Road has best Italian restaurants.

3.2 Cluster Analysis :

1. Map to show the restaurant clusters in Pune



2. Map to show clusters of Pune localities based on various venues



3. K-means clustering was conducted in order to group the localities according to what venues they have using Foursquare data in order to feel the atmosphere of each locality.

As a first step of cluster analysis, one hot encoding was conducted to give binary values to each venue categories. Let's observe each cluster and the characteristics they possess.

Cluster 1

```
## Examine Clusters
## Cluster 1
pune_merged.loc[pune_merged['Cluster Labels'] == 0, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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
18	18.462694	3.692308	2013.0	0	Indian Restaurant	Snack Place	Chinese Restaurant	Yoga Studio	Farmers Market	Food Truck	Food Court	Food
19	18.598300	3.393333	2261.0	0	Afghan Restaurant	Indian Restaurant	Concert Hall	Yoga Studio	Farmers Market	French Restaurant	Food Truck	Food Court
27	18.504538	3.649565	34862.0	0	Coffee Shop	Farmers Market	Indian Restaurant	Bar	French Restaurant	Food Truck	Food Court	Food
41	18.544079	3.150000	114.0	0	Indian Restaurant	Bar	Food Court	Burger Joint	Coffee Shop	Yoga Studio	Frozen Yogurt Shop	French Restaurant
44	18.601209	3.670000	2030.0	0	Indian Restaurant	Yoga Studio	Falafel Restaurant	French Restaurant	Food Truck	Food Court	Food	Flea Market
54	18.488272	3.777778	2126.0	0	Indian Restaurant	Coffee Shop	Ice Cream Shop	Bistro	Multiplex	Southern / Soul Food Restaurant	Breakfast Spot	Dessert Shop

Cluster 2

```
## Examine Clusters

## Cluster 2
pune_merged.loc[pune_merged['Cluster Labels'] == 1, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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 M Comr Ver
0	18.650478	3.704348	4310.0	1	Coffee Shop	Gym	Mexican Restaurant	Food Truck	Fast Food Restaurant	Thai Restaurant	Ind Restaur
1	18.560537	3.721053	48882.0	1	Indian Restaurant	Dessert Shop	Bakery	Shopping Mall	Coffee Shop	Grocery Store	Chin
2	18.525643	3.520000	374.0	1	Juice Bar	Shopping Mall	Indian Restaurant	Hostel	Cupcake Shop	Bakery	Snack Pl
4	18.560805	3.821074	121456.0	1	Indian Restaurant	South Indian Restaurant	Ice Cream Shop	Fast Food Restaurant	Gym	Restaurant	Chin Restaur
9	18.515551	3.807692	2345.0	1	Indian Restaurant	Ice Cream Shop	Bakery	Road	Flea Market	Miscellaneous Shop	Snack Pl
10	18.533483	4.050000	6979.0	1	Indian Restaurant	Chinese Restaurant	Bakery	Fast Food Restaurant	Bistro	Kebab Restaurant	Restaur

Cluster 3

```
## Examine Clusters

## Cluster 3
pune_merged.loc[pune_merged['Cluster Labels'] == 2, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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 Co
34	18.429091	3.70	169.0	2	Resort	Restaurant	Yoga Studio	English Restaurant	Food Truck	Food Court	Food	Flea Market	Fast Rest
74	18.552177	3.94	3991.0	2	Resort	Yoga Studio	English Restaurant	Food Truck	Food Court	Food	Flea Market	Fast Food Restaurant	Fast

Cluster 4

```
## Examine Clusters

## Cluster 4
pune_merged.loc[pune_merged['Cluster Labels'] == 3, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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	9 C
6	18.647716	3.733333	2465.0	3	Pizza Place	Falafel Restaurant	French Restaurant	Food Truck	Food Court	Food	Flea Market	Fast Food Restaurant	
60	18.596027	3.650000	104.0	3	Ice Cream Shop	Coffee Shop	Pizza Place	American Restaurant	Fruit & Vegetable Store	French Restaurant	Food Truck	Food Court	
79	18.551980	3.644828	3856.0	3	Ice Cream Shop	Department Store	Afghan Restaurant	Garden	Frozen Yogurt Shop	French Restaurant	Food Truck	Food Court	

Cluster 5

```
## Examine Clusters
## Cluster 5
pune_merged.loc[pune_merged['Cluster Labels'] == 4, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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
17	18.687418	3.75	120.0	4	Trail	Yoga Studio	Falafel Restaurant	French Restaurant	Food Truck	Food Court	Food	Flea Market	Fast Food Restau

Cluster 6

```
## Examine Clusters
## Cluster 6
pune_merged.loc[pune_merged['Cluster Labels'] == 5, pune_merged.columns[[1] + list(range(5, pune_merged.shape[1]))]]
```

	Lat	Agg_Rating	No_of_Votes	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
3	18.575478	3.644444	7997.0	5	Restaurant	Chinese Restaurant	Asian Restaurant	Tea Room	Yoga Studio	Food Truck	Food Court		
5	18.513202	3.651429	18895.0	5	Dessert Shop	Chinese Restaurant	Snack Place	Café	Shop & Service	Seafood Restaurant	Food Court		
7	18.495588	3.857143	4287.0	5	Hookah Bar	Café	Resort	Yoga Studio	Falafel Restaurant	Food Truck	Food Court		
8	18.481024	3.769048	8384.0	5	Snack Place	Bakery	Smoke Shop	Café	Seafood Restaurant	Ice Cream Shop	Indian Restaurant		
12	18.560001	3.581818	1948.0	5	Restaurant	Electronics Store	Men's Store	Market	Gym / Fitness Center	Snack Place	Dumpling Restaurant		
25	18.503680	3.733333	751.0	5	Multiplex	Shopping Mall	Wedding Hall	Gym	Café	Breakfast Spot	Motorcycle Shop	Electric	

4. Results

Upon different analysis, I was able to discover the best neighborhood on the criteria of highest number of restaurants and quality of restaurants. Also, I have tried to identify most livable neighborhood that have most of the basic venues. Let's review the analysis made in this project and then we will summarize the project findings in the conclusion area.

1. Finding aggregate rating of all restaurants in each neighborhood/locality.

```
df_final = df_final[df_final['Aggregate rating'] != 0.00000]
df_final.columns = ['Locality', 'Lat', 'Lng', 'No_of_Restaurant', 'Cusines', 'Agg_Rating', 'No_of_Votes']
df_final.head()
```

	Locality	Lat	Lng	No_of_Restaurant	Cusines	Agg_Rating	No_of_Votes
0	Akurdi	18.650478	73.772272	23	South Indian, North Indian, Street Food, Fast ...	3.704348	4310.0
1	Aundh	18.560537	73.806087	133	Korean, Japanese, Asian, North Indian, Mughlai...	3.721053	48882.0
2	B.T. Kawade Road	18.525643	73.906212	5	North Indian, Maharashtrian, Desserts, Chinese...	3.520000	374.0
3	Balewadi	18.575478	73.772069	36	Chinese, North Indian, North Indian, Chinese, ...	3.644444	7997.0
4	Baner	18.560805	73.785572	242	North Indian, Continental, Asian, Italian, Mah...	3.821074	121456.0

2.Exploring 10 common venues in each neighborhood/locality.

	Locality	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	Akurdi	Coffee Shop	Gym	Mexican Restaurant	Food Truck	Fast Food Restaurant	Thai Restaurant	Indian Restaurant	Donut Shop	Dumpling Restaurant	Electronics Store
1	Aundh	Indian Restaurant	Dessert Shop	Bakery	Shopping Mall	Coffee Shop	Grocery Store	Café	Restaurant	Department Store	Diner
2	B.T. Kawade Road	Juice Bar	Shopping Mall	Indian Restaurant	Hostel	Cupcake Shop	Bakery	Snack Place	Gym	Farmers Market	Food Truck
3	Balewadi	Restaurant	Chinese Restaurant	Asian Restaurant	Tea Room	Yoga Studio	Food Truck	Food Court	Food	Flea Market	Fast Food Restaurant
4	Baner	Indian Restaurant	South Indian Restaurant	Ice Cream Shop	Fast Food Restaurant	Gym	Restaurant	Chinese Restaurant	Shopping Mall	Breakfast Spot	Lounge

5. Conclusion

To summarize, these are the findings:

1. The best restaurants are available in Deccan & Peths.
2. The worst restaurants are in Lavale.
3. Baner and Viman Nagar have highest number of restaurants in Pune City.
4. Expressway, Khadakwasla , Kothrud Area have lowest number of restaurants.
5. Baner has the most number of Italian restaurants.
6. Bund Garden Road has best Italian restaurants according to ratings.
7. Places in Cluster 1 & 2 has many Indian restaurants.
8. Places in Cluster 2 has many cafes, coffee shops and dessert shops.
9. Places in Cluster 3 has resorts and yoga studios.
10. Places in Cluster 6 is most livable as it has - restaurants, dessert shops, hotels, cafes , shopping malls etc.