Restaurants in Famous Indian Cities

Riona Almeida | Applied Data Science Capstone Project (IBM/Coursera)
29th April, 2020

Content

- A description of the problem and a discussion of the background.
- A description of the data and how it will be used to solve the problem.

1.1 Background

Being a citizen of India, I have always been intrigued with what the various cities of India has to offer. With our wide range of cuisines and our very large appetite to try different delicacies, restaurants and cafés around this country are not only well versed in serving Indian cuisine but are highly recommended for international dishes too.

It is not an uncommon fact that there are some densely populated cities dispersed around Indian soil. Financial superpowers like Mumbai, technical giants like Bangalore and emerging cities like Pune – each have their own ___. Then again there are some religious and historic sites like ____ that have managed to uphold their legacies in terms of maintaining livelihoods. What each city has to offer to arouse our taste buds and how they are connected to each other? That is the question the natural foodie in me seeks to answer.

1.2 Problem

Clustering restaurants from different cities in India will help determine some common characteristics between them. Analyzing these different venues, we can visualize the following data:

- Groups of cities similar to each other w.r.t restaurants
- Why these places are so populated?
- What characteristics do they share?

1.3 Interest

Say if one were to visit a famous Indian city and would want to acquire knowledge of the best food spots. Or better still maybe some entrepreneur wants to set up base with a restaurant, he/she will need to know the kind of cuisines and eateries the general public picks. With the attained information from this study, these individuals will be able to make their stay and capital more profitable.

This data can also be useful for culinary experts and food critics to categorize and rate famous places based on their similarities.

2. Data Acquisition

For famous Indian cities, I found a dataset of 200 odd famous cities with respect to population. It has the following attributes:

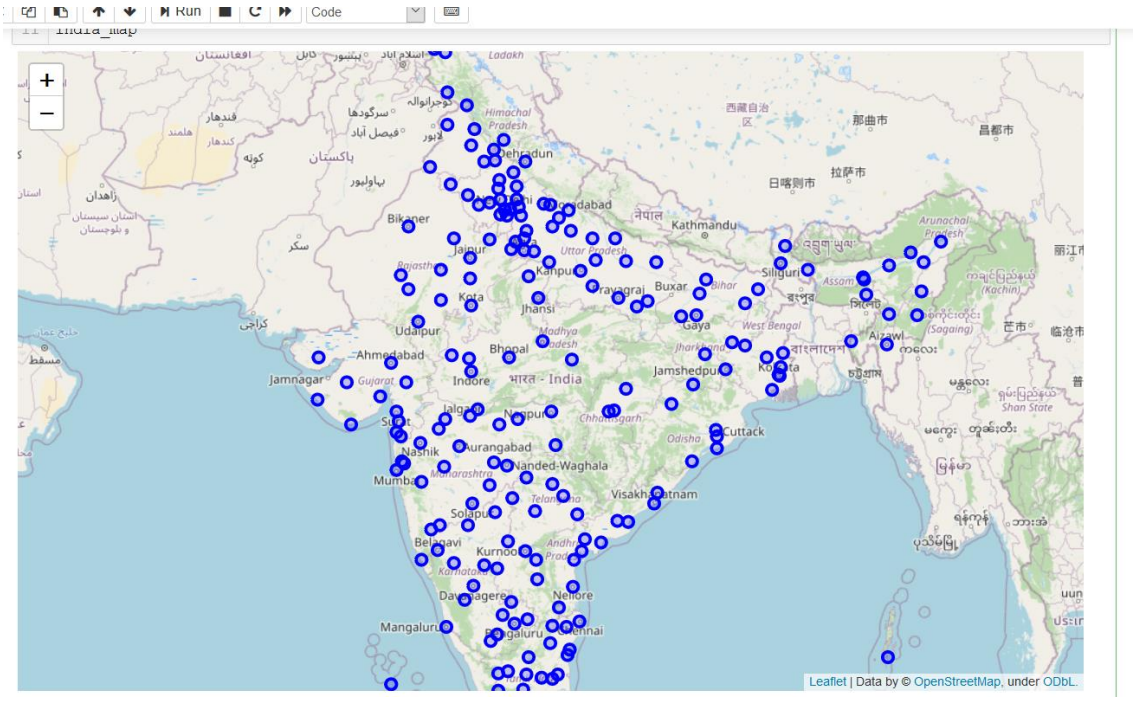
- Latitude
- Longitude
- State
- Population
- City name

Source : https://simplemaps.com/data/in-cities

urban population of 18,978,000.

Download Data in Spreadsheet (213 cities) | Get all cities in India (478,635 cities) | Looking for an India map? | Other Countries

city	lat	lng	country	iso2	admin	capital	population	population_proper
Mumbai	18.987807	72.836447	India	IN	Mahārāshtra	admin	18978000	12691836
Delhi	28.651952	77.231495	India	IN	Delhi	admin	15926000	7633213
Kolkata	22.562627	88.363044	India	IN	West Bengal	admin	14787000	4631392
Chennai	13.084622	80.248357	India	IN	Tamil Nādu	admin	7163000	4328063
Bengalūru	12.977063	77.587106	India	IN	Karnātaka	admin	6787000	5104047
Hyderabad	17.384052	78.456355	India	IN	Andhra Pradesh	admin	6376000	3597816
Ahmadābād	23.025793	72.587265	India	IN	Gujarāt	minor	5375000	3719710
Hāora	22.576882	88.318566	India	IN	West Bengal		4841638	1027672
Pune	18.513271	73.849852	India	IN	Mahārāshtra		4672000	2935744
Sūrat	21.195944	72.830232	India	IN	Gujarāt		3842000	2894504
Mardānpur	26.430066	80.267176	India	IN	Uttar Pradesh		3162000	2823249
Rāmpura	26.884682	75.789336	India	IN	Rājasthān		2917000	2711758
Lucknow	26.839281	80.923133	India	IN	Uttar Pradesh	admin	2695000	2472011
Nāra	21.203096	79.089284	India	IN	Mahārāshtra		2454000	2228018
Patna	25.615379	85.101027	India	IN	Bihār	admin	2158000	1599920
Indore	22.717736	75.85859	India	IN	Madhya Pradesh		2026000	1837041
Vadodara	22.299405	73.208119	India	IN	Gujarāt		1756000	1409476
Bhopal	23.254688	77.402892	India	IN	Madhya Pradesh	admin	1727000	1599914
Coimbatore	11.005547	76.966122	India	IN	Tamil Nādu		1696000	959823
Ludhiāna	30.912042	75.853789	India	IN	Punjab		1649000	1545368
Āgra	27.187935	78.003944	India	IN	Uttar Pradesh		1592000	1430055
Kalyān	19.243703	73.135537	India	IN	Mahārāshtra		1576614	1576614
Vishākhapatnam	17.704052	83.297663	India	IN	Andhra Pradesh		1529000	1063178
Kochi	9.947743	76.253802	India	IN	Kerala		1519000	604696
Nāsik	19.999963	73.776887	India	IN	Mahārāshtra		1473000	1289497
Meerut	28.980018	77.706356	India	IN	Uttar Pradesh		1398000	1223184
Farīdābād	28.411236	77.313162	India	IN	Haryāna		1394000	1394000



Now that we have the details of each city, the essential information i.e. the venue data can be extracted using the FourSquare API using request methods.

To get venues and their categories from these cities, I will use the FourSquare API. The latitude and longitude values can be retrieved from the above table. For each city the range will be around 5 km and limit of up to 100 most popular venues.

From the FourSquare API, the data retrieved will have the following attributes:

- Venue
- Venue Category
- Latitude
- Longitude

From this data, for analysis I will only require restaurants/ café/ hotels to determine a food roadmap of the country.

	City	Neighborhood Group Latitude	Neighborhood Group Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mumbai	18.987807	72.836447	The St. Regis Mumbai	18.993652	72.825220	Hotel
1	Mumbai	18.987807	72.836447	Smoke House Deli	18.994478	72.824400	Restaurant
2	Mumbai	18.987807	72.836447	Jai Hind Lunch Home	19.002183	72.829512	Seafood Restaurant
3	Mumbai	18.987807	72.836447	Ladu Samrat	18.994526	72.836574	Maharashtrian Restaurant
4	Mumbai	18.987807	72.836447	The Irish House	18.994416	72.824721	Pub
5	Mumbai	18.987807	72.836447	Indigo Delicatessen	18.994498	72.823760	Deli / Bodega
6	Mumbai	18.987807	72.836447	Kebabs & Kurries	18.997938	72.837639	Indian Restaurant
7	Mumbai	18.987807	72.836447	The Sahib Room & Kipling Bar	18.994333	72.823850	Indian Restaurant
8	Mumbai	18.987807	72.836447	ITC Grand Central	18.998469	72.838433	Hotel
9	Mumbai	18.987807	72.836447	Francesco's	18.995435	72.824964	Pizza Place
10	Mumbai	18.987807	72.836447	Four Seasons	18.994356	72.820319	Hotel
11	Mumbai	18.987807	72.836447	Li Bai - St. Regis	18.994186	72.823795	Hotel Bar
12	Mumbai	18.987807	72.836447	Le 15 Patisserie	18.995091	72.823735	Cupcake Shop
13	Mumbai	18.987807	72.836447	Bombay Canteen	19.003410	72.827816	Indian Restaurant
14	Mumbai	18.987807	72.836447	The Tasting Room	18.992942	72.823103	Restaurant
15	Mumbai	18.987807	72.836447	Tote On The Turf	18.980266	72.820294	Nightclub

Fig. Final data obtained from all venues

3. Methodology and Exploratory Data Analysis

As a first step, I retrieved the venues in India from Foursquare. I extract the location data from the Foursquare API for all venues up to a distance of 5 kilometers from the center of each city location.

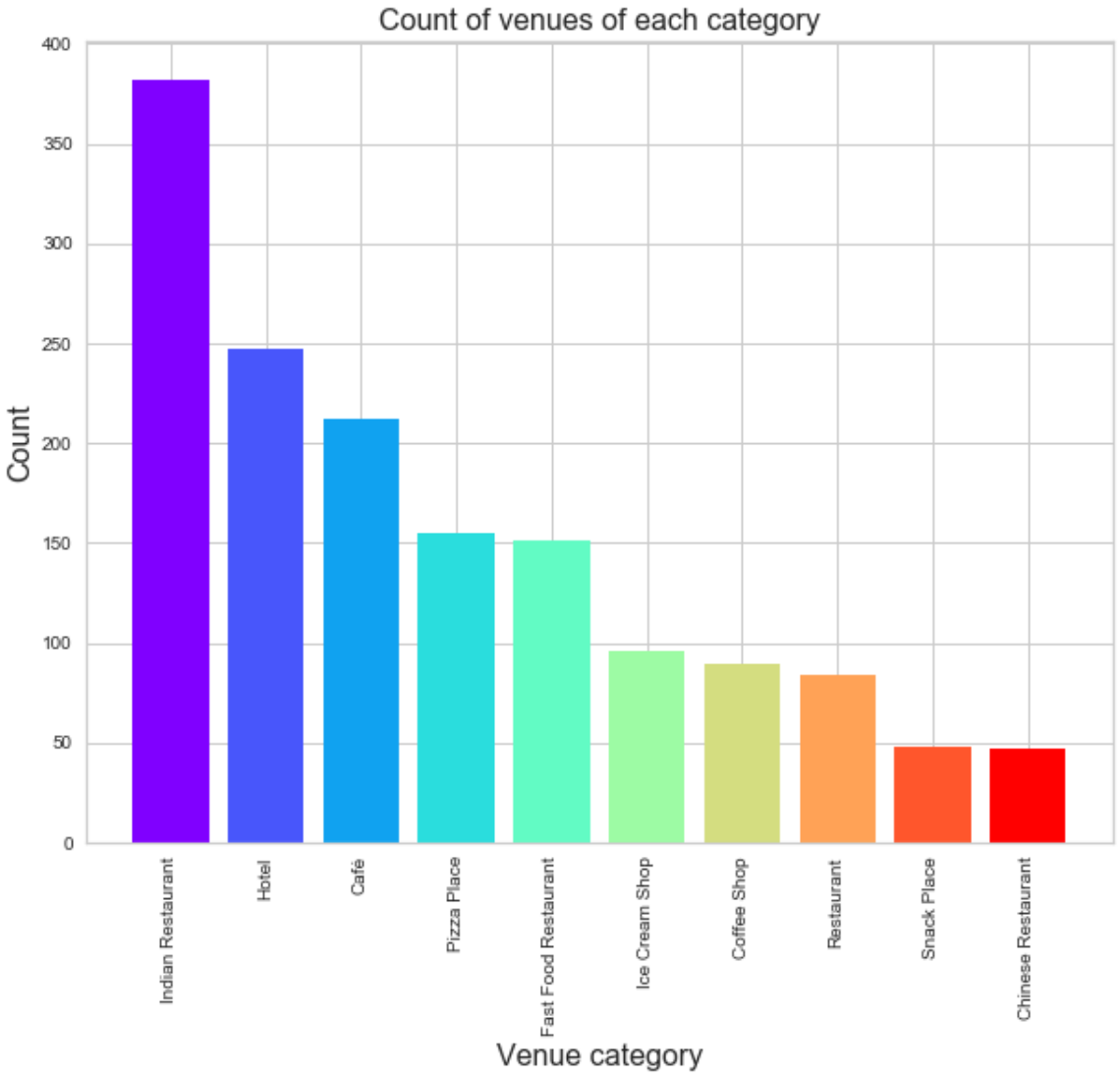
Using data cleaning, the dataset from the API is analysed based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the dataset. The final data will include the venue name, category, address, latitude, longitude.

Using this dataset, I begin by analyzing the top venue types that exist in India. I will then explore the venues on maps. This will allow us to better understand the location of various venues and the places where many venues co-exist and create place worth visiting. I'll also explore the venues based on the ratings and price range of various venues. The venues will be plot using proper color coding such that a simple glance at the map would reveal the location of the venues as well as give information about them. I aim to identify places which can be recommended to visitors based on their price and rating preferences. I'll also cluster the venues and see if we can draw meaningful information out of what kind of venues exist in India.

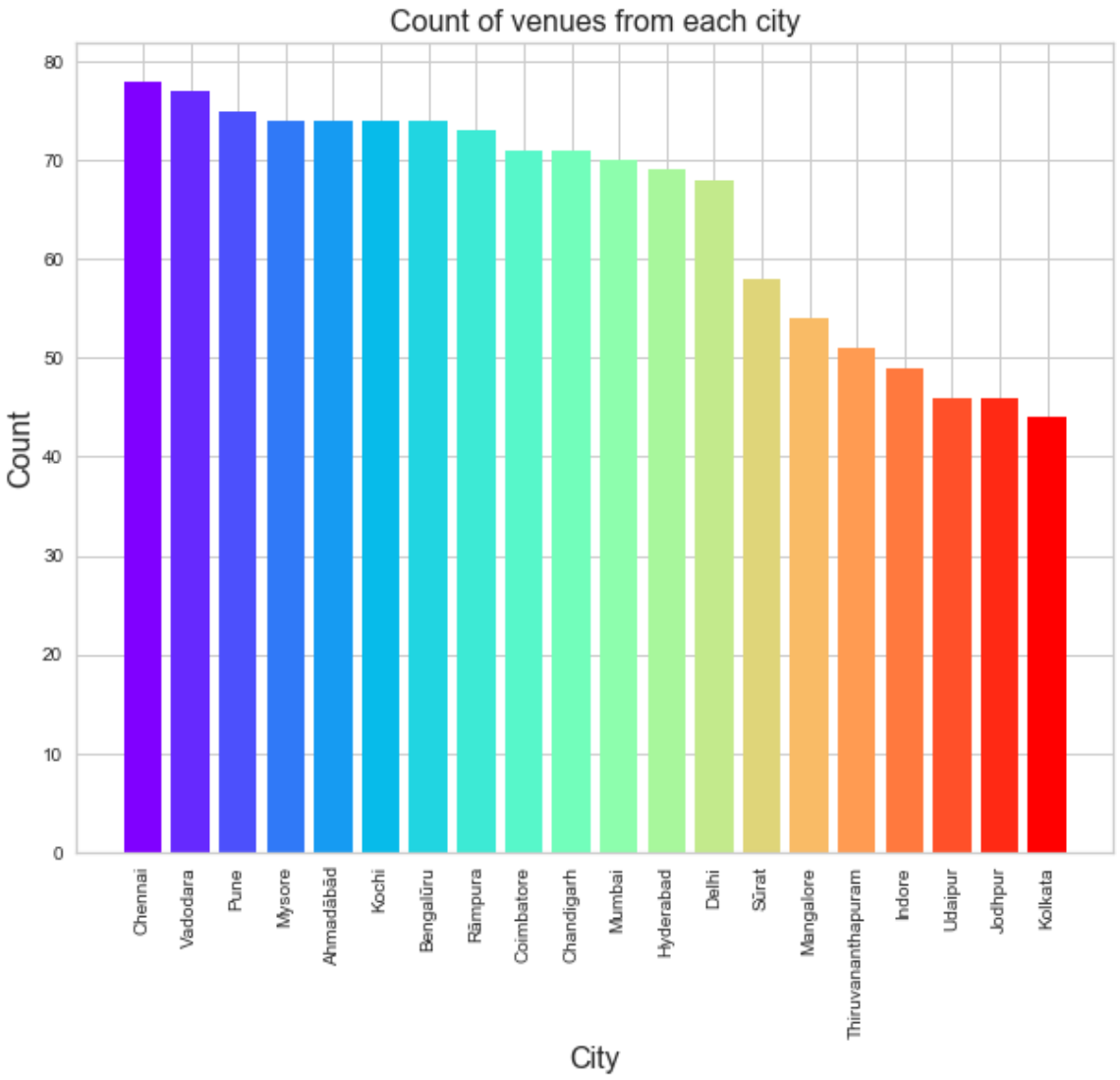
As a final step, I will analyze these plots and try to draw conclusions on what places can be recommended to visitors. I'll discuss my findings and any inferences I can draw.

3.1 Categories

I begin my analysis by taking a look at the various restaurant categories that exist in India.



From this chart we can see that Indian Restaurants, Cafes, Pizza Places, Fast Food Restaurants and Ice Cream Shops are some of the most popular categories across the country among others



From this chart, we understand that most venues were acquired from more populated cities like Chennai, Vadodara, Pune, Mysore, Mumbai, Hyderabad etc.

4. Clustering

Finally, I cluster all the venues based on their price range, location and more to identify similar venues and the relationship amongst them. I used KMeans clustering and decided to cluster the venues into two separate groups.

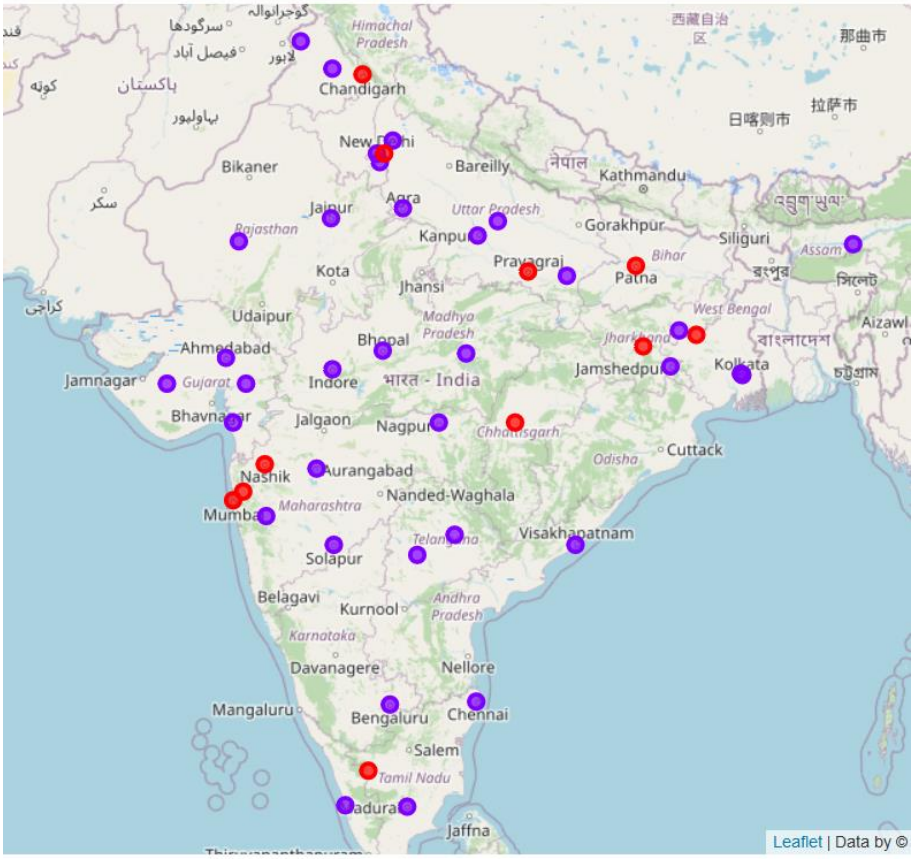


Fig. Clustered cities based on popular eating spots

From the map, we see the two clusters:

1. The first cluster is very sparsely spread and has very limited venues.
2. The second cluster is spread across country and includes the major cities.

I'll also check the venues of the two clusters.

Cluster 1 : This cluster seems to contain cities that have Indian Restaurants, Fast Food Restaurants, Pizza Places and Hotels as their most popular venues. In terms of population, apart from Mumbai, they are not that populated. Seems like the inhabitants don't explore much variety.

	city	lat	population	population_proper	Cluster Labels	Most Common Venue no. 1	Most Common Venue no. 2	Most Common Venue no. 3	Most Common Venue no. 4	Most Common Venue no. 5
0	Mumbai	18.987807	18978000.0	12691836.0	0	Indian Restaurant	Restaurant	Hotel	Ice Cream Shop	Café
14	Patna	25.615379	2158000.0	1599920.0	0	Indian Restaurant	Fast Food Restaurant	Pizza Place	Hotel	Ice Cream Shop
18	Coimbatore	11.005547	1696000.0	959823.0	0	Indian Restaurant	Hotel	Pizza Place	Fast Food Restaurant	Café
21	Kalyān	19.243703	1576614.0	1576614.0	0	Indian Restaurant	Hotel	Fast Food Restaurant	Café	Restaurant
24	Nāsik	19.999963	1473000.0	1289497.0	0	Indian Restaurant	Hotel	Restaurant	Pizza Place	Fast Food Restaurant

Cluster 2 : In this second cluster, other restaurant categories like Chinese restaurants, Italian restaurants, Sandwich Places are seen to be popular options among the citizens. This cluster mostly contains most populated cities of India and we can clearly see that people here are very diverse in choosing their food eating hotspots. Even though Indian Restaurants take the major chunk as most popular venue in both clusters, this particular cluster has a bit more variety.

	city	lat	population	population_proper	Cluster Labels	Most Common Venue no. 1	Most Common Venue no. 2	Most Common Venue no. 3	Most Common Venue no. 4	Most Common Venue no. 5
1	Delhi	28.651952	15926000.0	7633213.0	1	Indian Restaurant	Hotel	Chinese Restaurant	Italian Restaurant	Café
2	Kolkata	22.562627	14787000.0	4631392.0	1	Indian Restaurant	Sandwich Place	Chinese Restaurant	Fast Food Restaurant	Café
3	Chennai	13.084622	7163000.0	4328063.0	1	Indian Restaurant	Ice Cream Shop	Hotel	Pub	Café
4	Bengalūru	12.977063	6787000.0	5104047.0	1	Indian Restaurant	Café	Hotel	Fast Food Restaurant	Coffee Shop
5	Hyderabad	17.384052	6376000.0	3597816.0	1	Indian Restaurant	Hotel	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Pizza Place
6	Ahmadābād	23.025793	5375000.0	3719710.0	1	Indian Restaurant	Fast Food Restaurant	Pizza Place	Café	Tea Room
7	Hāora	22.576882	4841638.0	1027672.0	1	Hotel	Indian Restaurant	Café	Restaurant	BBQ Joint
8	Pune	18.513271	4672000.0	2935744.0	1	Indian Restaurant	Coffee Shop	Pizza Place	Café	Fast Food Restaurant
9	Sūrat	21.195944	3842000.0	2894504.0	1	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Coffee Shop
10	Mardānpur	26.430066	3162000.0	2823249.0	1	Pizza Place	Sandwich Place	Hotel	Ice Cream Shop	Indian Restaurant

Results

Indian Restaurants continue to be the heart of most Indian cities around the country, serving as the most popular choice. International cuisine options though present are very rare. Cafes and Fast Food restaurants are also some options frequented upon. So even though Indian Restaurants are top priority, we can easily say judging from the clusters that the cities are seeing a moving trend in making cafes and fast food restaurants among the more popular options.

A large number of the most populated cities appear in the second cluster where a variety of eating options ensue compared to the first cluster that is mostly stuck on Indian restaurants being the most popular choices. The first cluster contains more traditional spots like Coimbatore, Patna and Nasik thus Indian restaurants are easily the most loved option there. The second cluster majorly contains emerging cities like Bangalore, Delhi, Chennai that embark newer trends and hence have a wave of popular options ranging from Multi cuisine spots, sandwich spots and delis.