

# Capstone Project: Clustering top ten Indian cities based on venues and analyzing business scopes in each city.

**Business Problem:** The business opportunities in the top ten populated Indian cities.

**Introduction:** The current GDP growth rate of India is 7.4 percent, so it is great opportunity of investors to invest in India and leverage the potential of Indian people. In my project I would like to demonstrate the most categories of places in top ten Indian cities. This project will give analytical insight to the investors to invest in different businesses and will help understanding the choice of people in different cities.

**Data:** I have collected the dataset containing the data of top populated cities in India from Internet sources like Wikipedia. I have collected the data of postal codes of each city from "www.mapsofindia.com" website. I have collected the data related to different places from foursquare.

These datasets have helped to analyze the venues spread across top 10 Indian cities. We have looked on statistical views of different venues in each city. We have looked on cumulative statistical report after merging all the datasets. Which helped to have better idea of culture in the cities. We have clustered the cities across India and got to know which cities are similar or different to each other.

The Cities has been considered are:

- i) Ahmedabad
- ii) Bangalore
- iii) Chennai
- iv) Delhi
- v) Hyderabad
- vi) Jaipur
- vii) Kolkata
- viii) Mumbai
- ix) Pune
- x) Surat

**Methodology:** We have collected top 10 most populated cities in India as per 2017 report published in Wikipedia. We have created bar charts for each 10 cities which helps to have very detailed idea of investing in different cities. We have depicted the top 20 categories of venues as per foursquare datasets which helps to have business decision. We have created cumulative plot of these cities and plotted the numbers of the venues in a single bar plot which helps investors to choose business domain. We have clustered these top 10 cities using k-means algorithm to know the difference among the cities and business distribution can done based on that report.

## Results:

### i) Statistical view of top 20 venues in Ahmedabad.

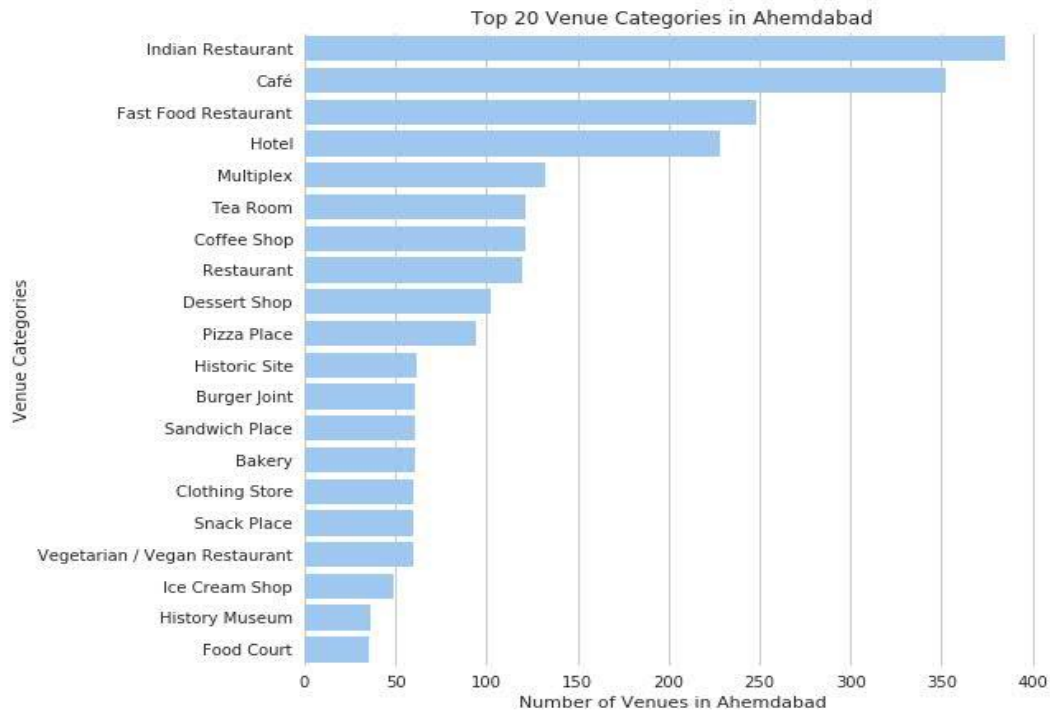


Fig :1

### ii) Statistical view of top 20 venues in Bangalore.

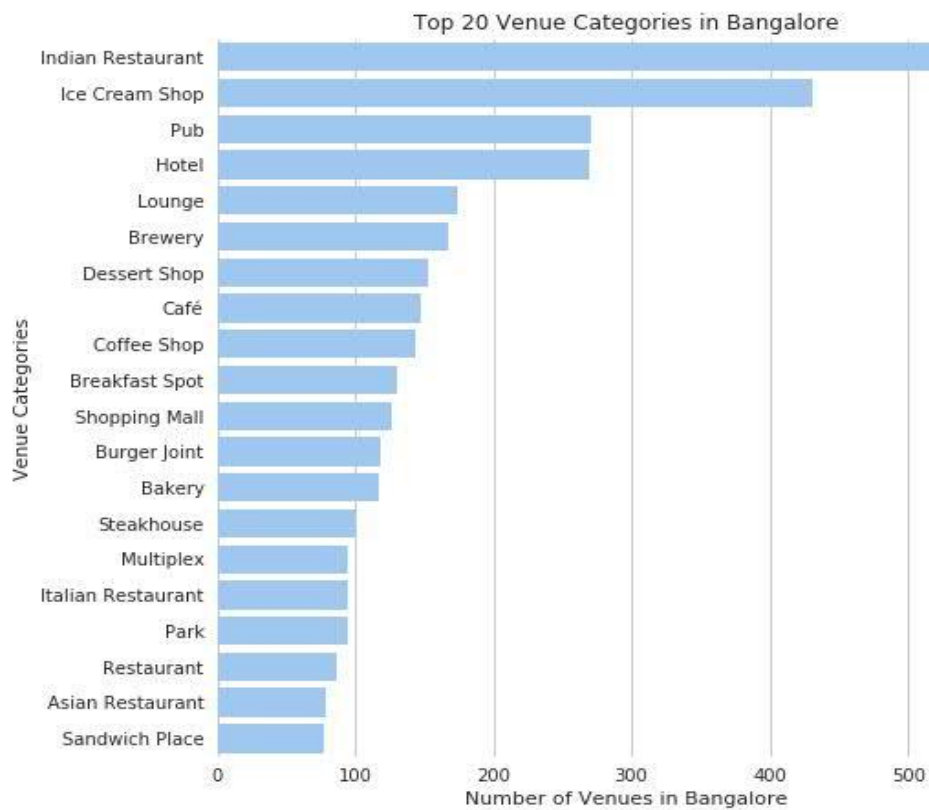


Fig :2

iii) Statistical view of top 20 venues in Chennai.

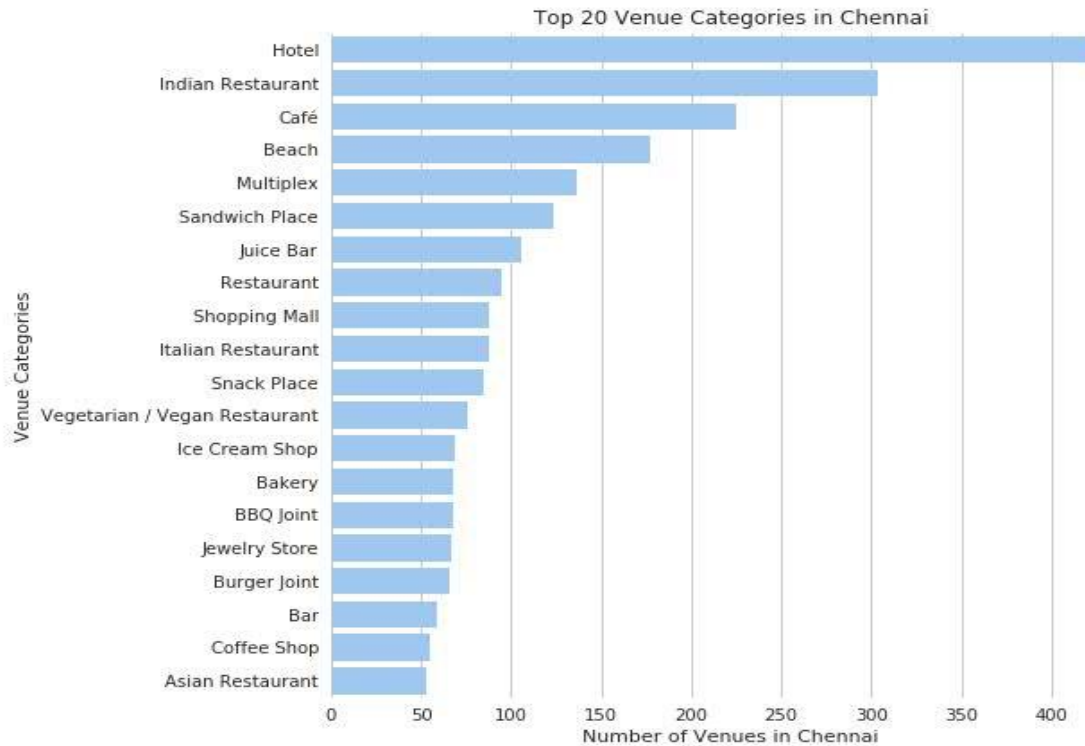


Fig :3

iv) Statistical view of top 20 venues in Delhi.

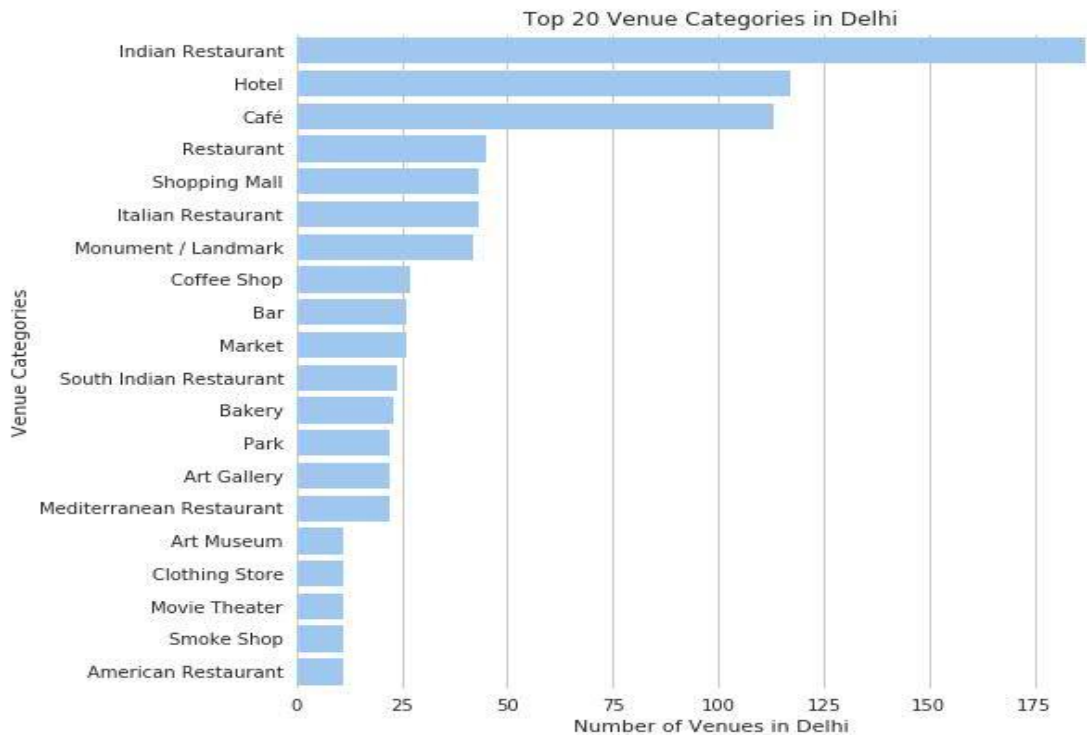


Fig :4

v) Statistical view of top 20 venues in Hyderabad.

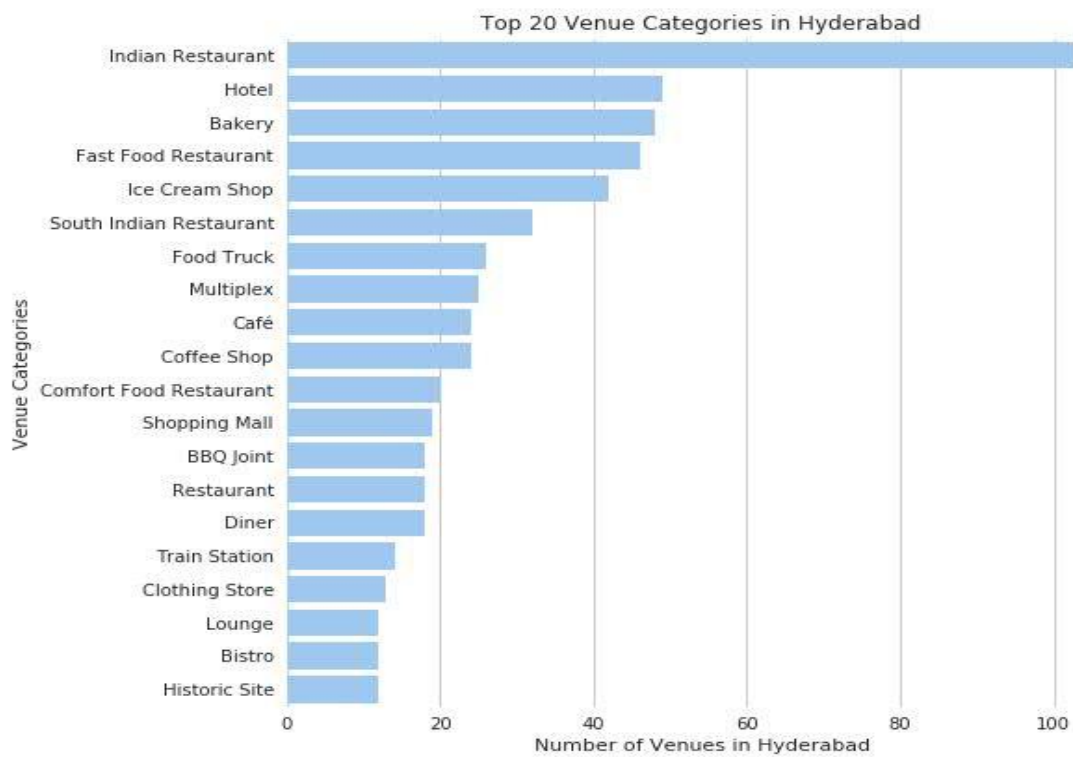


Fig :5

vi) Statistical view of top 20 venues in Jaipur.

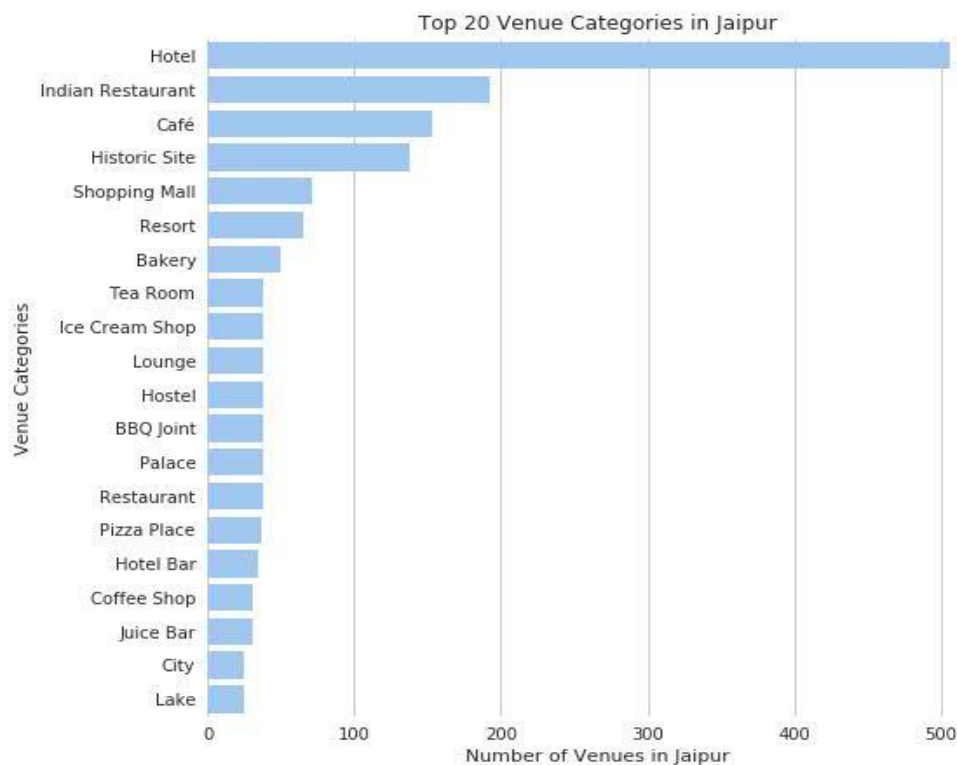


Fig :6

vii) Statistical view of top 20 venues in Kolkata.

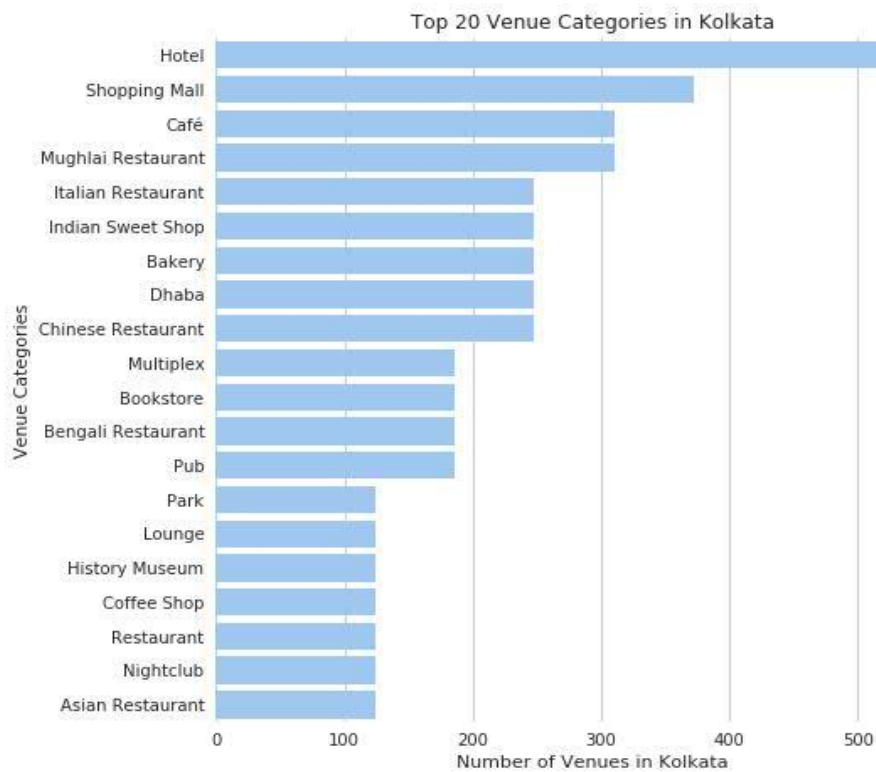


Fig :7

viii) Statistical view of top 20 venues in Mumbai.

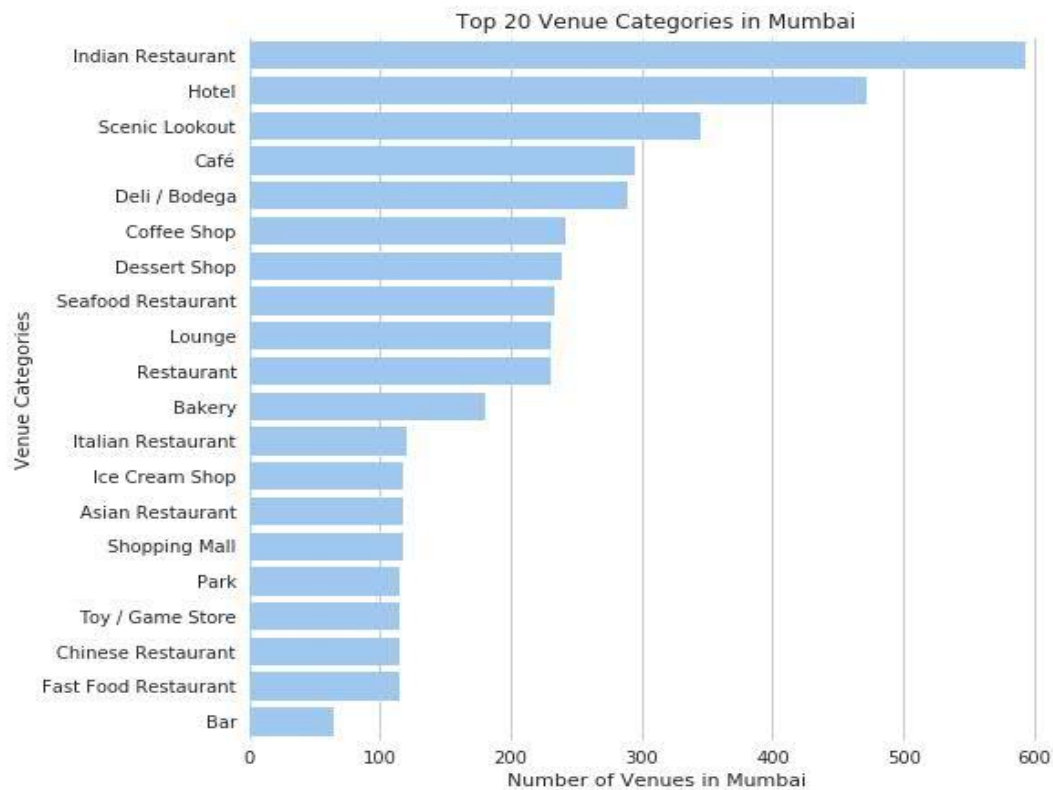
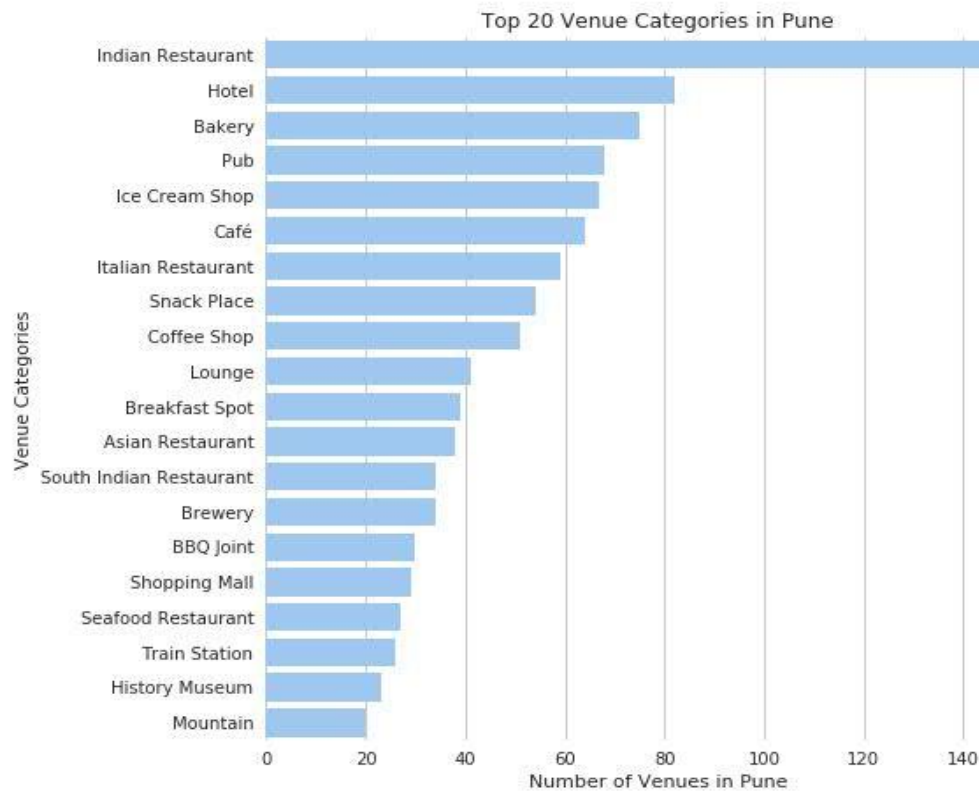


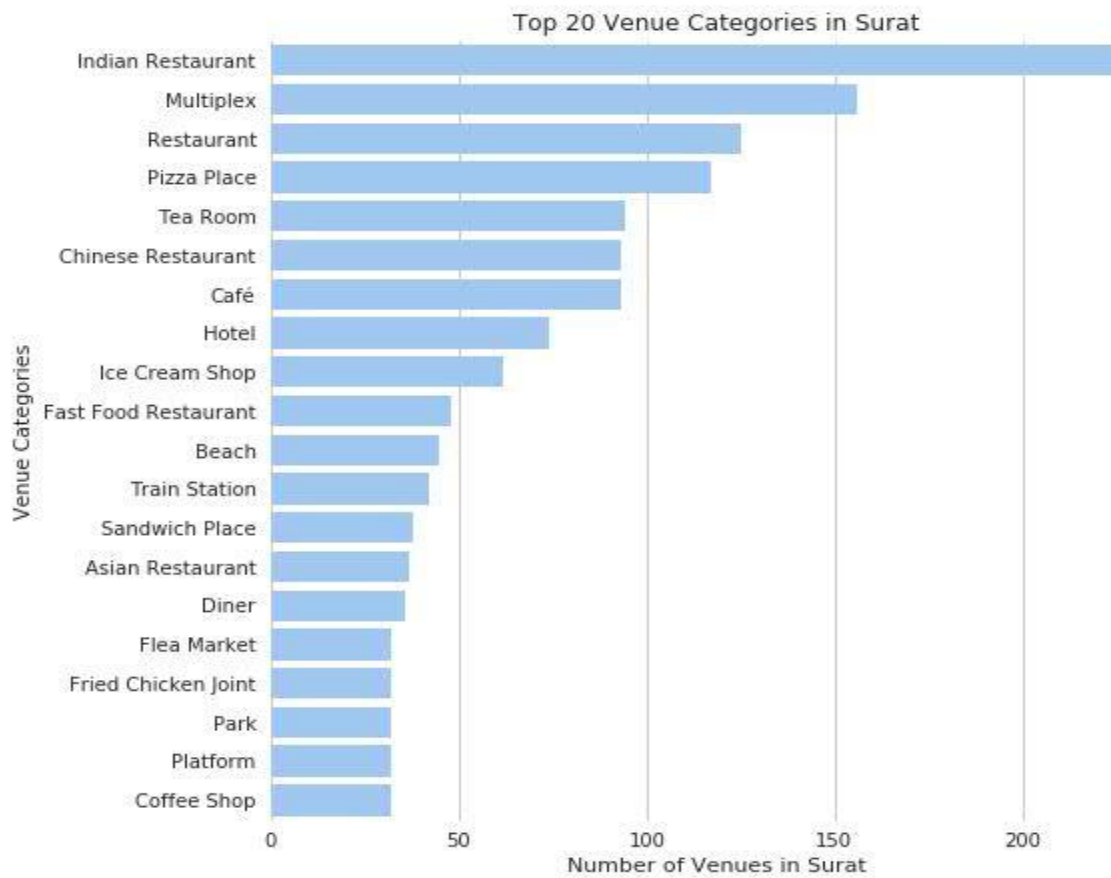
Fig :8

ix) Statistical view of top 20 venues in Pune.



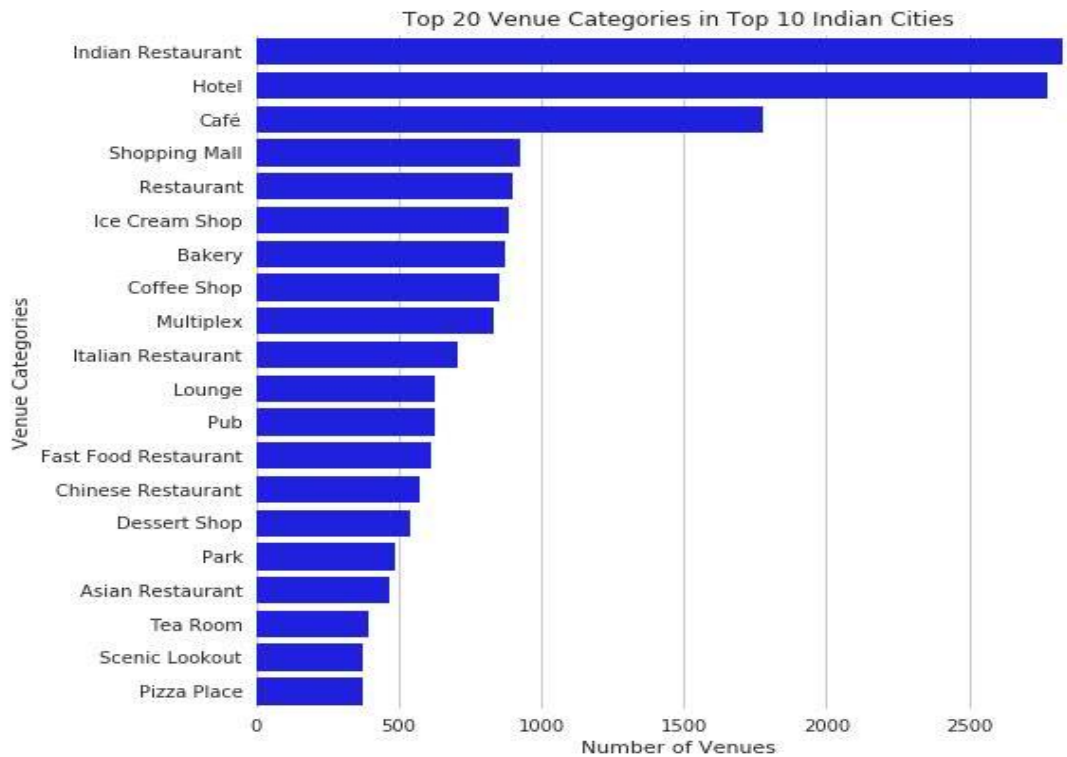
**Fig :9**

**x) Statistical view of top 20 venues in Surat.**



**Fig :10**

**Xi) Cumulative statistical view of top 20 venues.**



**Fig :11**

**Xii) Table containing top ten cities and top ten venues of each city.**

City	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 Ahmedabad	Indian Restaurant	Café	Fast Food Restaurant	Hotel	Multiplex	Tea Room	Coffee Shop	Restaurant	Dessert Shop	Pizza Place
1 Bangalore	Indian Restaurant	Ice Cream Shop	Pub	Hotel	Lounge	Brewery	Dessert Shop	Café	Coffee Shop	Breakfast Spot
2 Chennai	Hotel	Indian Restaurant	Café	Beach	Multiplex	Sandwich Place	Juice Bar	Restaurant	Italian Restaurant	Shopping Mall
3 Delhi	Indian Restaurant	Hotel	Café	Restaurant	Shopping Mall	Italian Restaurant	Monument / Landmark	Coffee Shop	Bar	Market
4 Hyderabad	Indian Restaurant	Hotel	Bakery	Fast Food Restaurant	Ice Cream Shop	South Indian Restaurant	Food Truck	Multiplex	Coffee Shop	Café
5 Jaipur	Hotel	Indian Restaurant	Café	Historic Site	Shopping Mall	Resort	Bakery	Hostel	Palace	Ice Cream Shop
6 Kolkata	Hotel	Shopping Mall	Café	Mughlai Restaurant	Italian Restaurant	Dhaba	Indian Sweet Shop	Chinese Restaurant	Bakery	Bengali Restaurant
7 Mumbai	Indian Restaurant	Hotel	Scenic Lookout	Café	Deli / Bodega	Coffee Shop	Dessert Shop	Seafood Restaurant	Lounge	Restaurant
8 Pune	Indian Restaurant	Hotel	Bakery	Pub	Ice Cream Shop	Café	Italian Restaurant	Snack Place	Coffee Shop	Lounge
9 Surat	Indian Restaurant	Multiplex	Restaurant	Pizza Place	Tea Room	Chinese Restaurant	Café	Hotel	Ice Cream Shop	Fast Food Restaurant

**Fig :12**



xiii)Top ten cities and how they are different as per foursquare dataset.

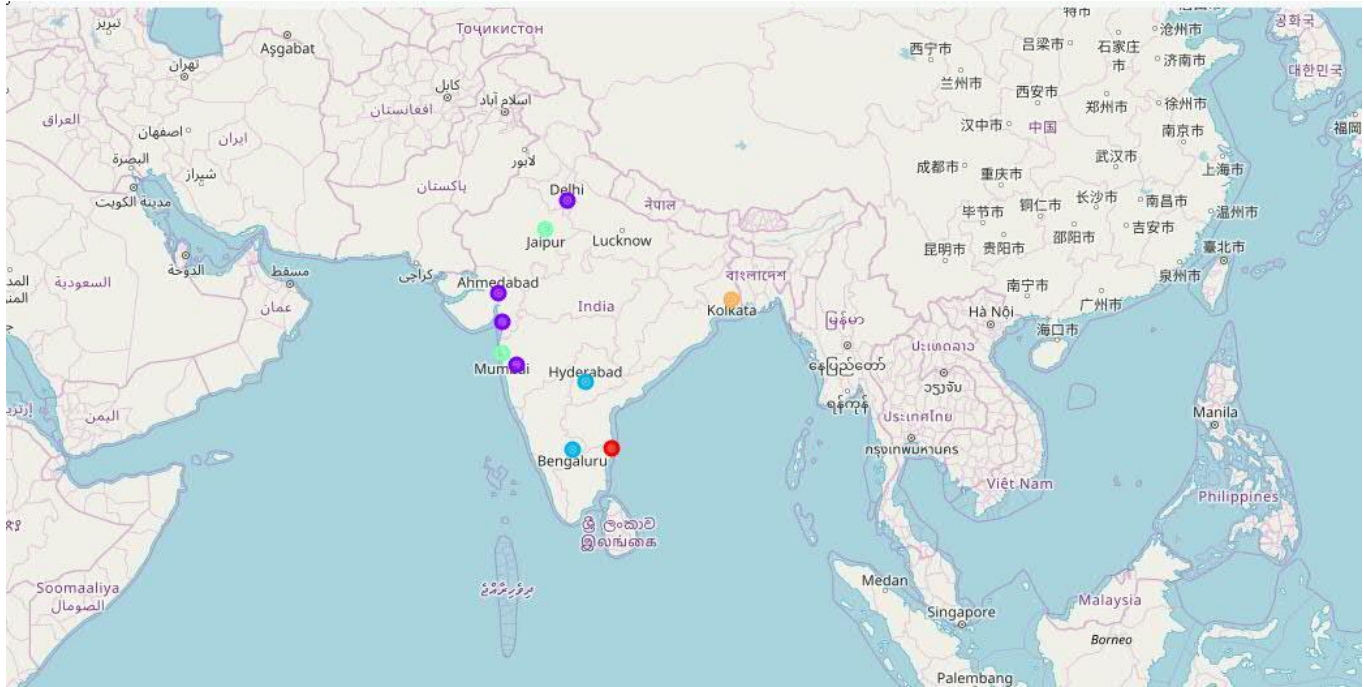


Fig :13

xiv) Detailed View of clusters:

Cluster1:

States	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
5 Tamil Nadu	Hotel	Indian Restaurant	Café	Beach	Multiplex	Sandwich Place	Juice Bar	Restaurant	Italian Restaurant	Shopping Mall

Cluster2:

States	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
1 Delhi	Indian Restaurant	Hotel	Café	Restaurant	Shopping Mall	Italian Restaurant	Monument / Landmark	Coffee Shop	Bar	Market
4 Gujarat	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7 Gujarat	Indian Restaurant	Multiplex	Restaurant	Pizza Place	Tea Room	Chinese Restaurant	Café	Hotel	Ice Cream Shop	Fast Food Restaurant
8 Maharashtra	Indian Restaurant	Hotel	Bakery	Pub	Ice Cream Shop	Café	Italian Restaurant	Snack Place	Coffee Shop	Lounge

Cluster3:

States	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
2 Karnataka	Indian Restaurant	Ice Cream Shop	Pub	Hotel	Lounge	Brewery	Dessert Shop	Café	Coffee Shop	Breakfast Spot
3 Telangana	Indian Restaurant	Hotel	Bakery	Fast Food Restaurant	Ice Cream Shop	South Indian Restaurant	Food Truck	Multiplex	Coffee Shop	Café



#### Cluster4:

	States	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	Maharashtra	Indian Restaurant	Hotel	Scenic Lookout	Café	Deli / Bodega	Coffee Shop	Dessert Shop	Seafood Restaurant	Lounge	Restaurant
9	Rajasthan	Hotel	Indian Restaurant	Café	Historic Site	Shopping Mall	Resort	Bakery	Hostel	Palace	Ice Cream Shop

#### Cluster5:

	States	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
6	West Bengal	Hotel	Shopping Mall	Café	Mughlai Restaurant	Italian Restaurant	Dhaba	Indian Sweet Shop	Chinese Restaurant	Bakery	Bengali Restaurant

**Discussion:** From the above-mentioned results we can understand the most of these cities has top dominated venues are Indian Restrained, Hotels, and shopping malls. After clustering we can see that Delhi, Ahmedabad, Surat and Pune have similar venues, Jaipur and Mumbai have similar venues in to categories, Bangalore and Hyderabad have similar top venues, but Chennai and Kolkata are all different from all other cities. This project would be very helpful to have business decisions and would be good choice for investors as well.

**Conclusion:** In project we have discussed all the detailed insights of important different venues spread across different cities in India. We have found different clusters of Indian cities using machine learning algorithm. We have got to know about top venues in different cities. This analysis gives better insights to all in very simple manner.