

# **Indian City Business Opportunity Analysis**

Akshay Cholapurath

July 20, 2020

## **Introduction**

### **Background**

India is on its road to become a manufacturing hub for the international companies as a part of Make in India campaign. Make in India was launched by the Government of India to encourage companies to manufacture their products in India and incentivize with dedicated investments into manufacturing. The policy approach was to create a conducive environment for investments, develop modern and efficient infrastructure, and open up new sectors for foreign capital. The initiative targeted 25 economic sectors for job creation and skill enhancement and aimed to transform India into a global design and manufacturing hub. So huge amount of investments from different companies are expected in the near future. Along with this, there will be huge potential for many business opportunities to make use of.

### **Problem**

It is very difficult to gain a bird's eye view of the insights into business opportunities and the location where it will make a huge impact. The opportunity arises mainly due to lack of competition. This project is intended to address this problem and gain many business opportunities insights based on its type, location, competition and the population which we can address with a perfect business and increase the chances of its success.

### **Interest**

There will be a lot of interested parties who would like to gain such insights on the business opportunities. Many companies, business firms, individuals who are looking to start a new business venture or who are already into business and looking to expand their business to different locations across India will be very much interested in this project.

## **Data Source and Cleaning**

The data for this Project is mainly acquired from the following link

<https://simplemaps.com/data/in-cities>

It provides a list of 213 prominent cities in India. Each row includes a city's latitude, longitude, state and other variables of interest such as population. This is a subset of all 478,635 places in India that you'll find in our World Cities Database. They have released this data subset for free under the MIT license. So because of this, we are free to use it for personal or commercial applications. And it can easily be downloaded in .csv, .json and .xlsx format directly from the website. A brief sample of the data is attached below.

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	Karnāṭaka	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

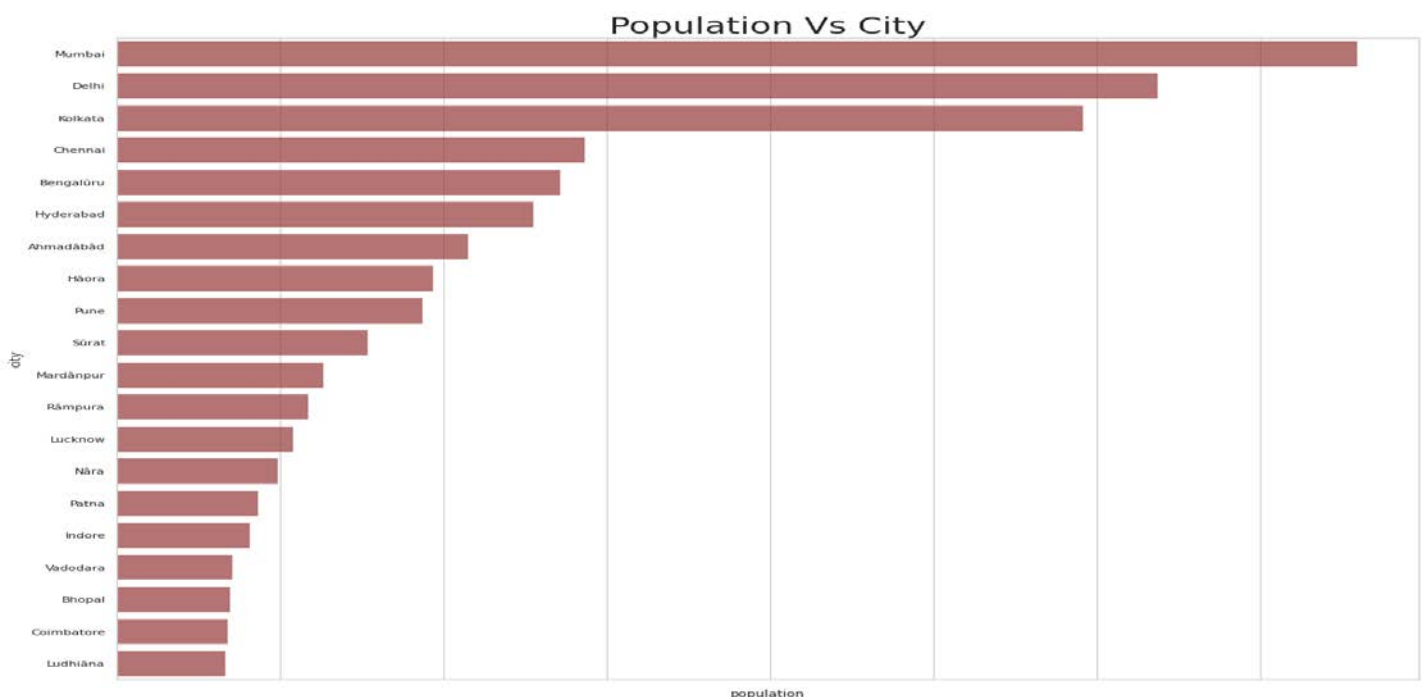
Since the data is mostly in a proper data frame format, there is not much cleaning to do on this data. I have cleared some of the unwanted columns from the table such as Country as it is obvious that we are analysis the country India, iso2 which is another short form of the country, capital and population proper.

We will then use the lat, log data of different cities and use it to the Foursquare API and get details about the venues which are surrounding the cities with in 10 km. Then those data which we acquire from the Foursquare will be merged with this data and we will make a detailed Exploratory Data Analysis to address our problem.

## Methodology

### General Comparison between Cities based on its Population

I have loaded the data into a pandas Data frame and removed the unwanted columns from the df and the made a basic Bar plot of top 20 most populated cities in India.

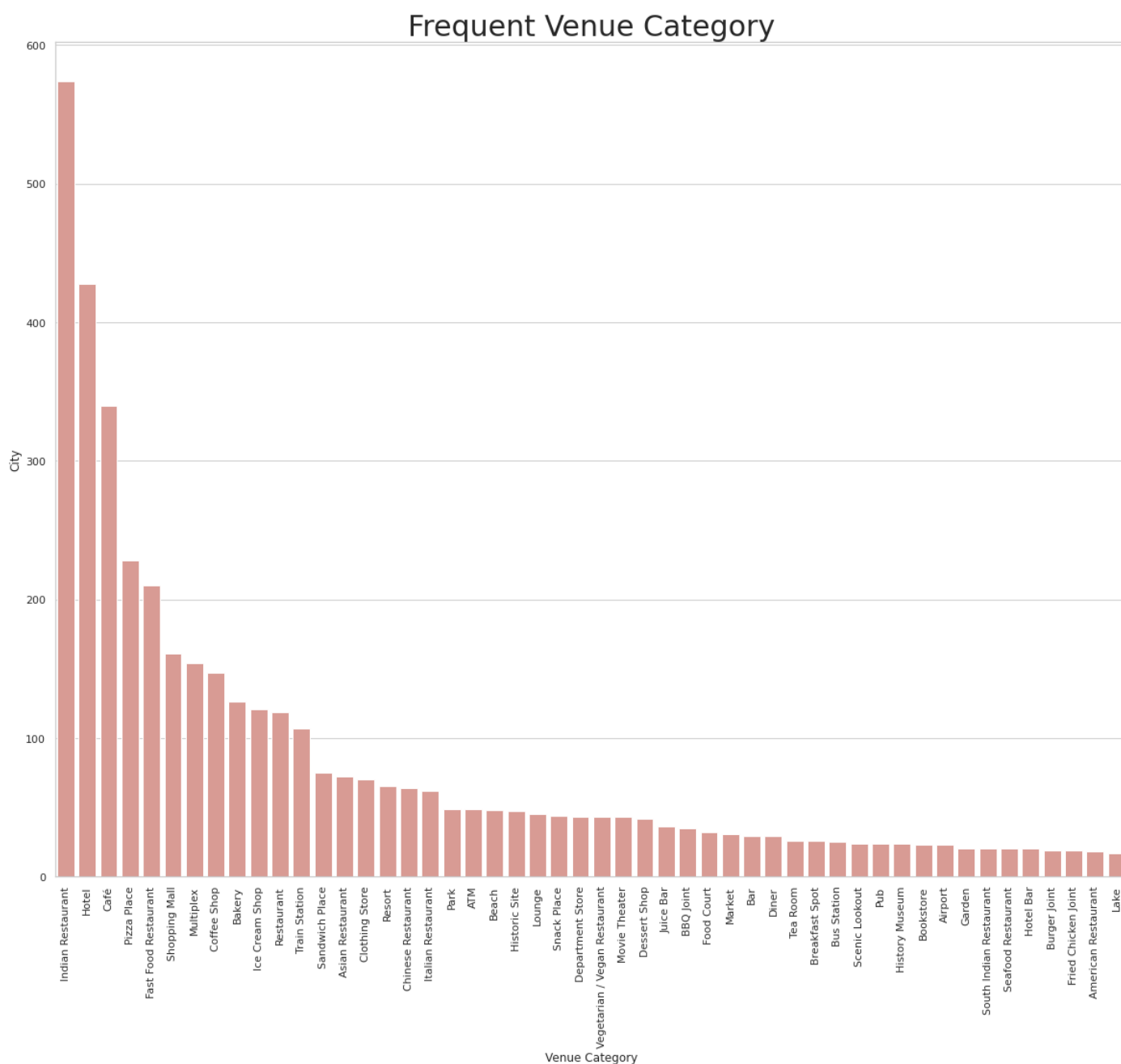


As we can see that the population of Mumbai, Delhi and Kolkata is significantly larger than the rest of the cities as there is a huge jump in the bar chart. The detailed values corresponding to these cities can be seen in the Notebook file.

## Venue Category Analysis

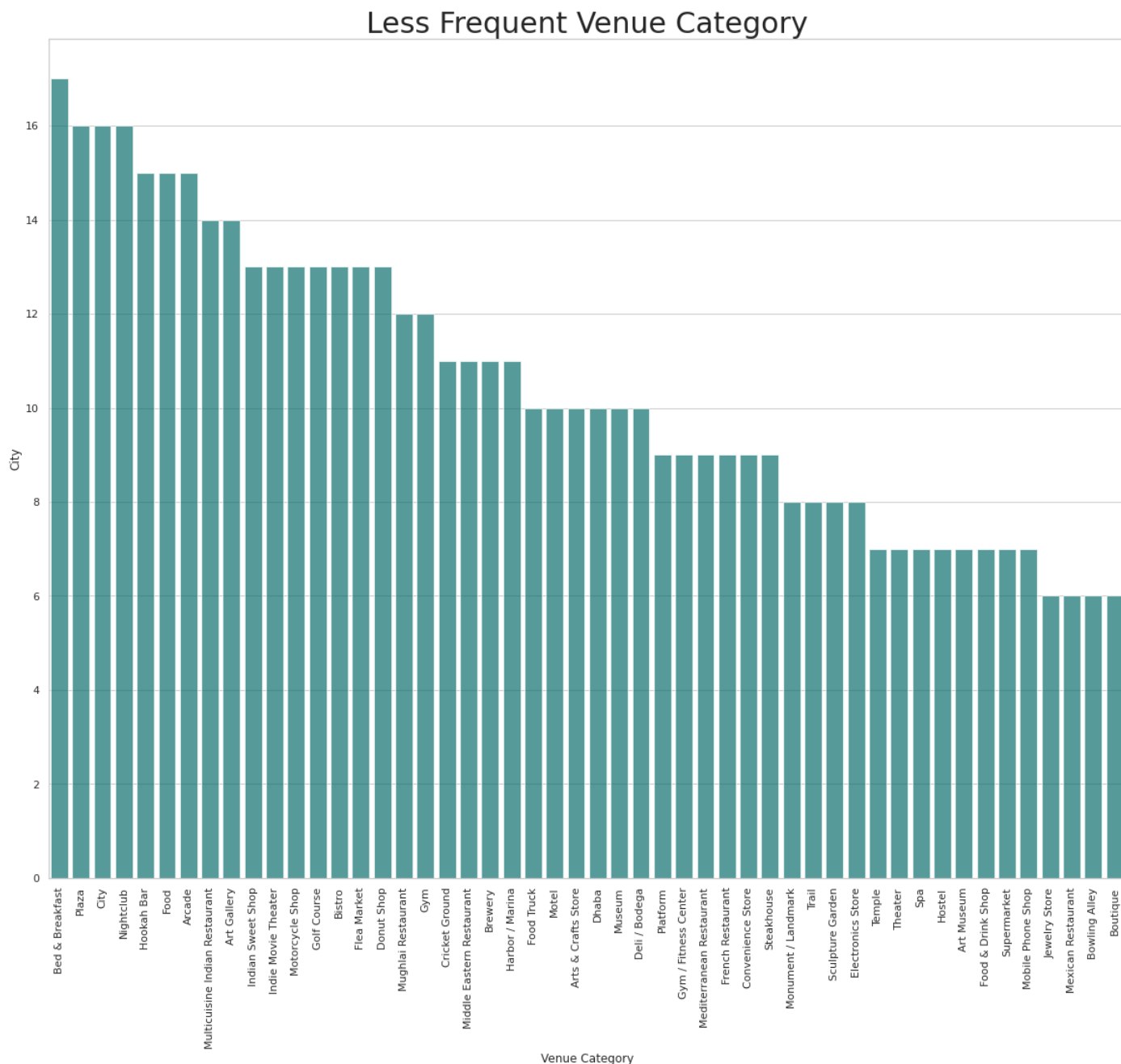
Next I have used the Foursquare API and was able to acquire Foursquare Data of the venue and venue category based on these cities with in radius of 10km. I have set the limited venue to be 100 in order to make the analysis. The resulted data frame was grouped based on the count of venue category and I have split the data into 2 categories for easy analysis.

### Most Frequent Business



Here when we analyse the Most Frequent Venue category, we can easily understand that the competition in the food business is very high. Different segments of food business such as Restaurants, Hotels, cafes etc are very much frequent in most of the cities. So we can infer that the Food business is mostly saturated.

## Less Frequent Business

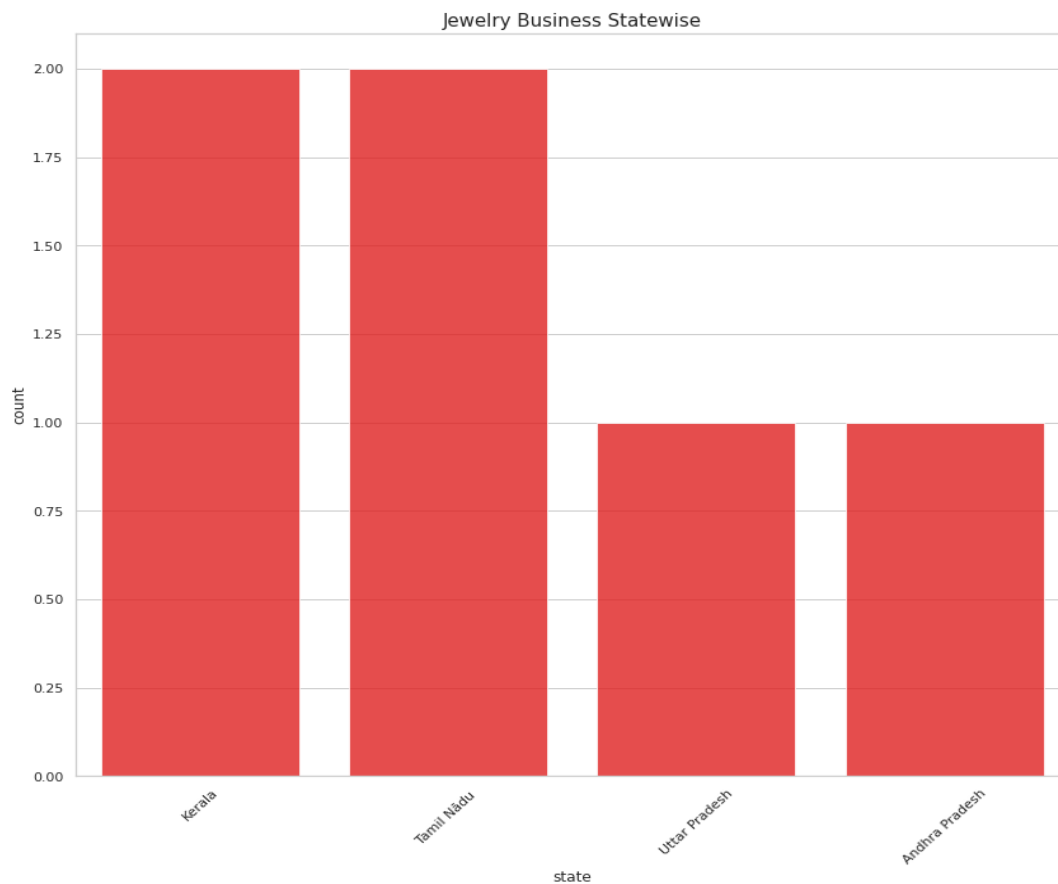


This is the category which we are interested in. We can see that there are valuable business which are having comparatively lesser competition and some of the locations does not even have these business with in the core city based on our Foursquare Data.

Now we will explore some of the Business Opportunities from the above plot.

- Jewellery Business State wise Comparison

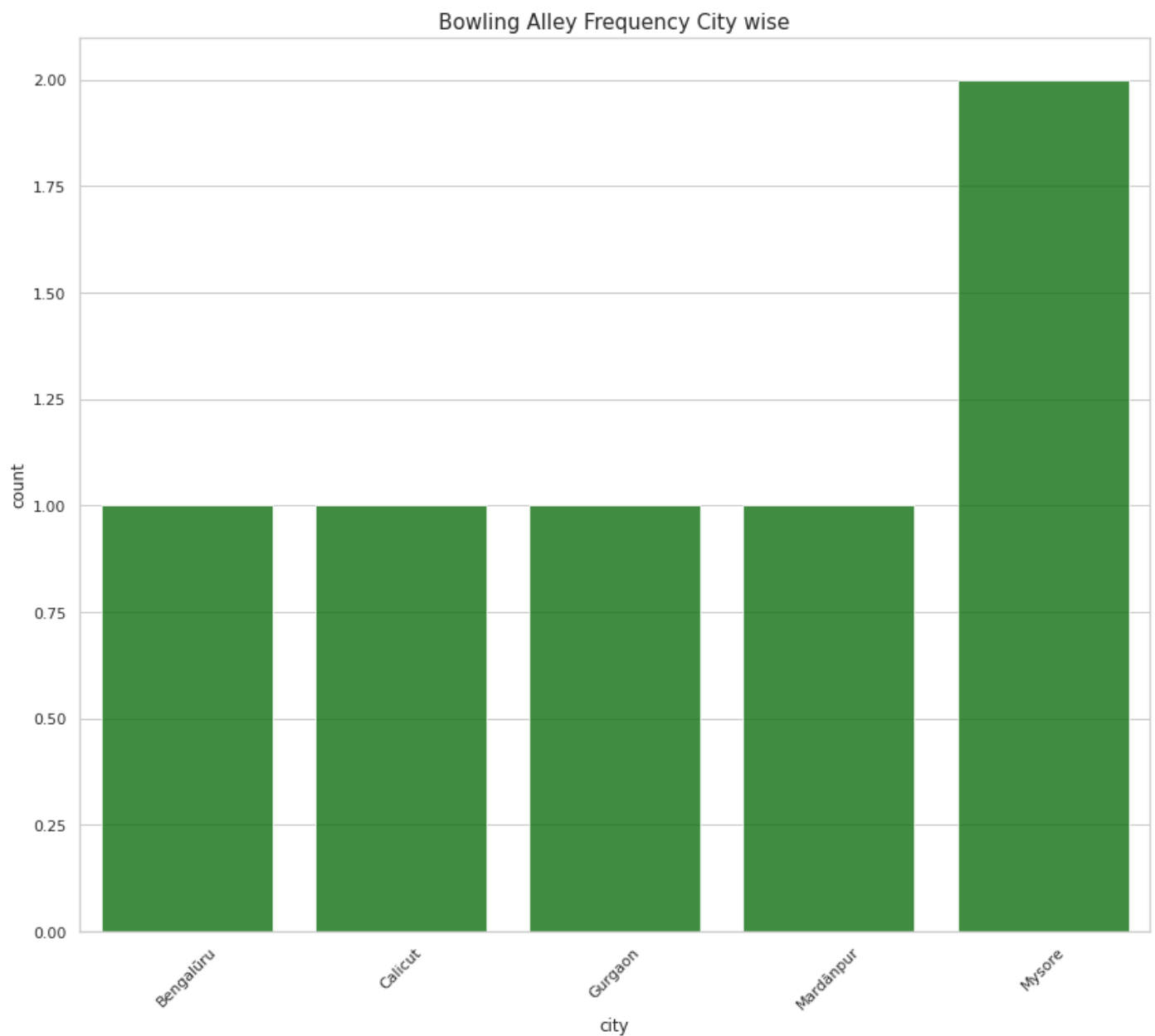
Based on the Foursquare Data, we can organize and arrange the data and plot it in its Frequency based on the state.



Here we can see that these are the states which has reputed Jewellery business within the main city. We can even go further and list the cities of these states which are having the business. This is given in the notebook. So there is a huge opportunity for all other states to have a Jewellery business within the city as all these states have relatively lesser population while comparing it with the first plot.

- Bowling Alley Business

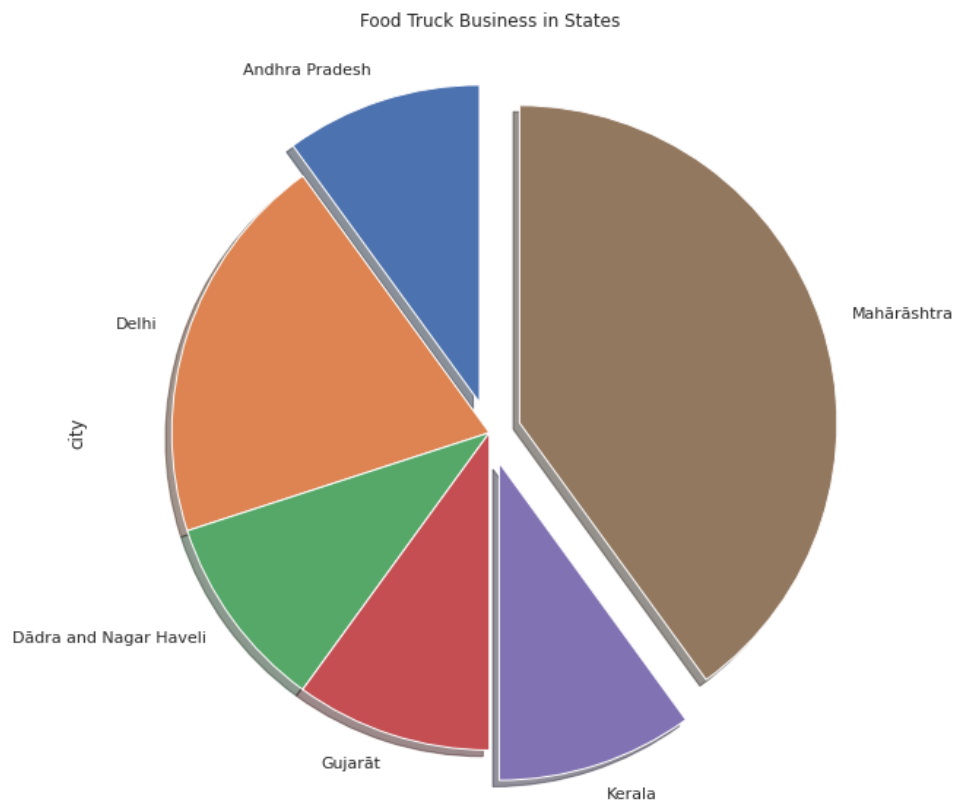
Here we can see some of the states which has reputed Bowling Alley business within the main city. We can infer that the city Mysore is saturated on this particular business. But it gives a huge insight that there are many other major cities with high population who doesnot have access to this entertainment and will be a very good business opportunity for the cities except the ones mentioned below. We can even go further and list the cities of these states which are having the business. This is given in the notebook.



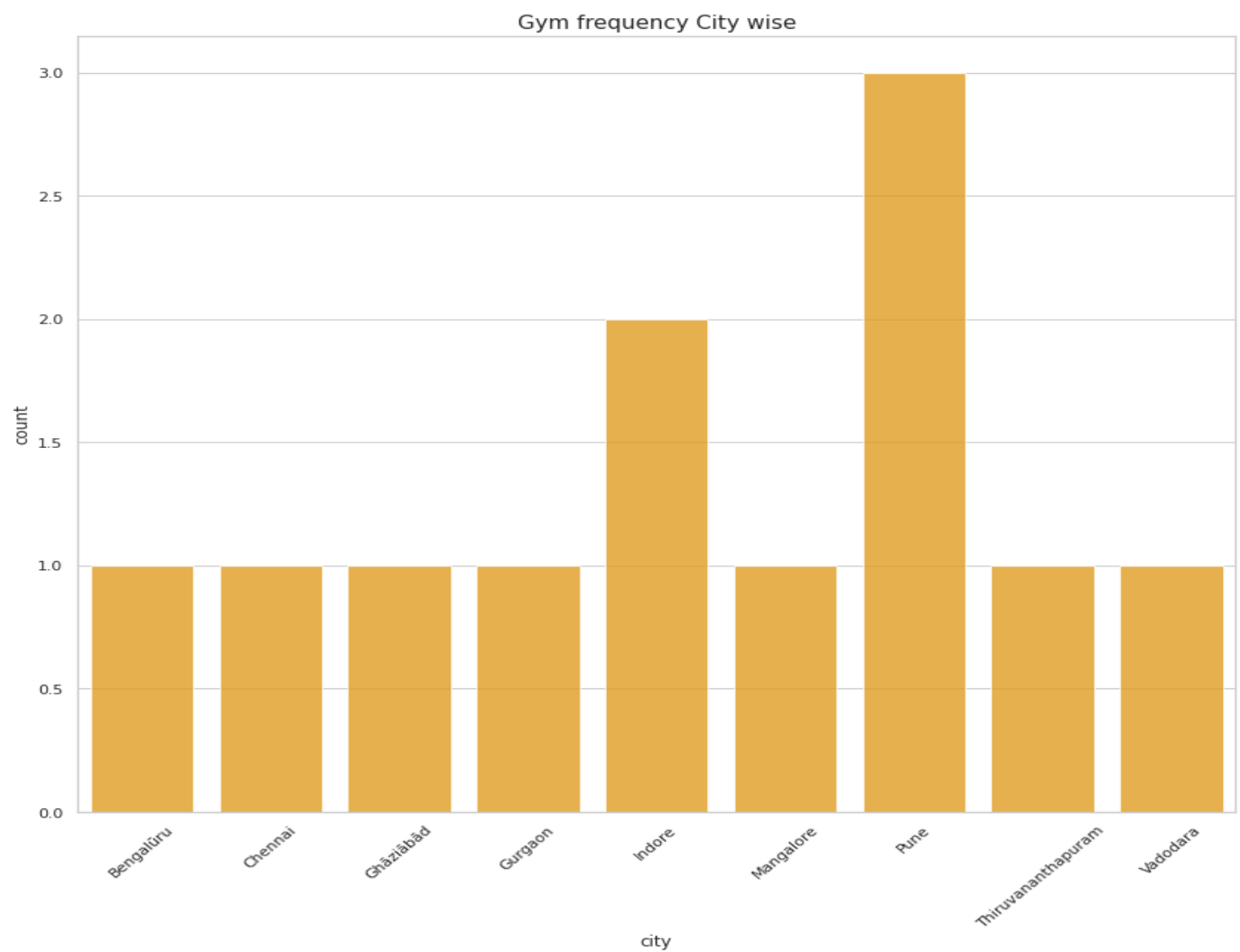
- Food Truck Business

Even though food business is saturated in India, the culture of Food Trucks have not yet caught up in the city. People will be interested to check these out as they provide food at a very fast pace and also will not be that expensive. It is mobile which gives a very good advantage over the rest of the food industry and can be parked near the parks in major cities.

Some of the major states which are having Food trucks are mentioned below. It will be a very good business opportunity for the rest of the cities. In depth city comparison is provided in the notebook.



- Physical Training Business

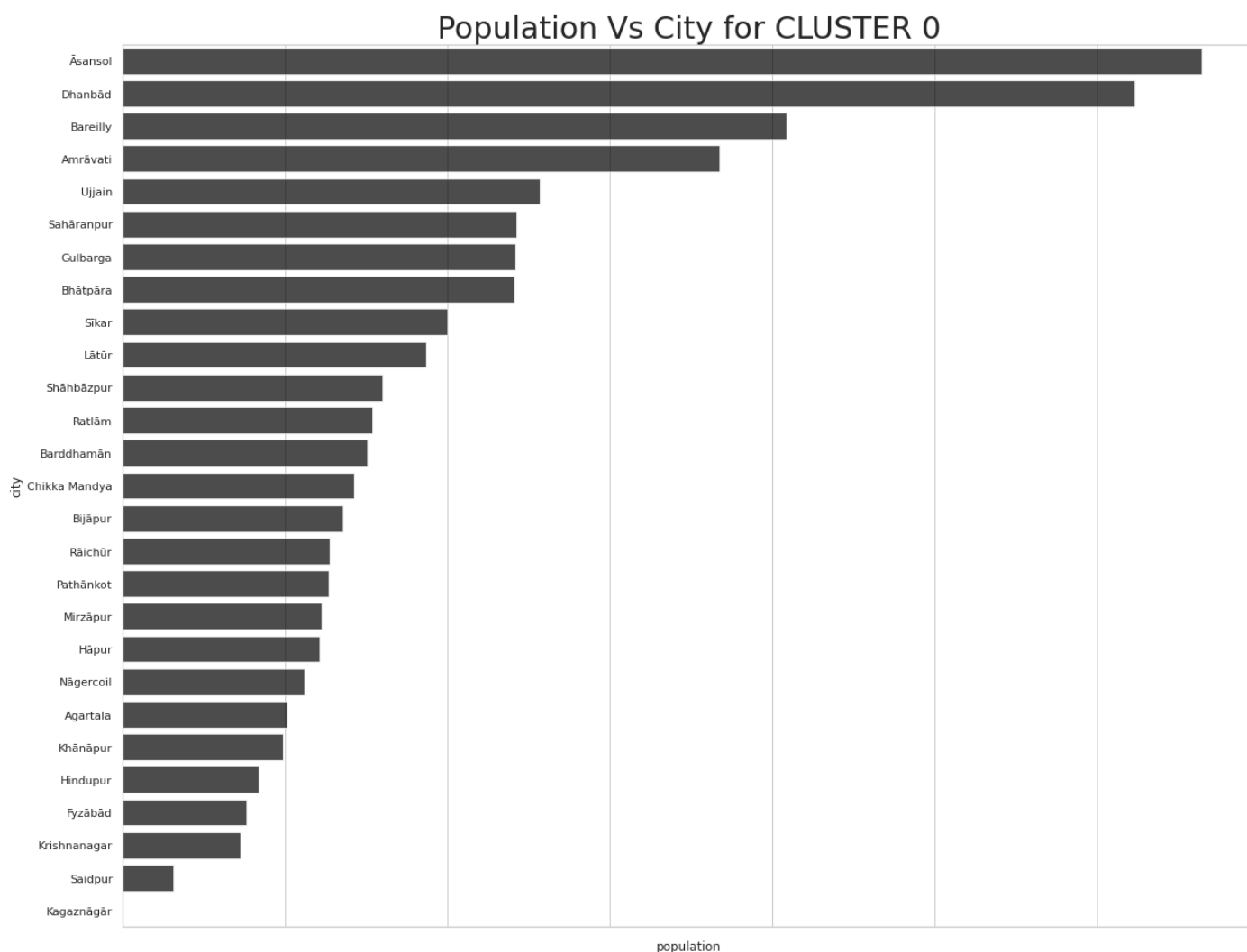


Here we can see that some of the major cities already have Gym facilities. But there are many other highly populated cities which does not have access to the Gym in the core city. It will be a great opportunity for a Gym chain business to open its outlets lesser competition cities and expand their business.

## City Cluster Analysis (K means clustering)

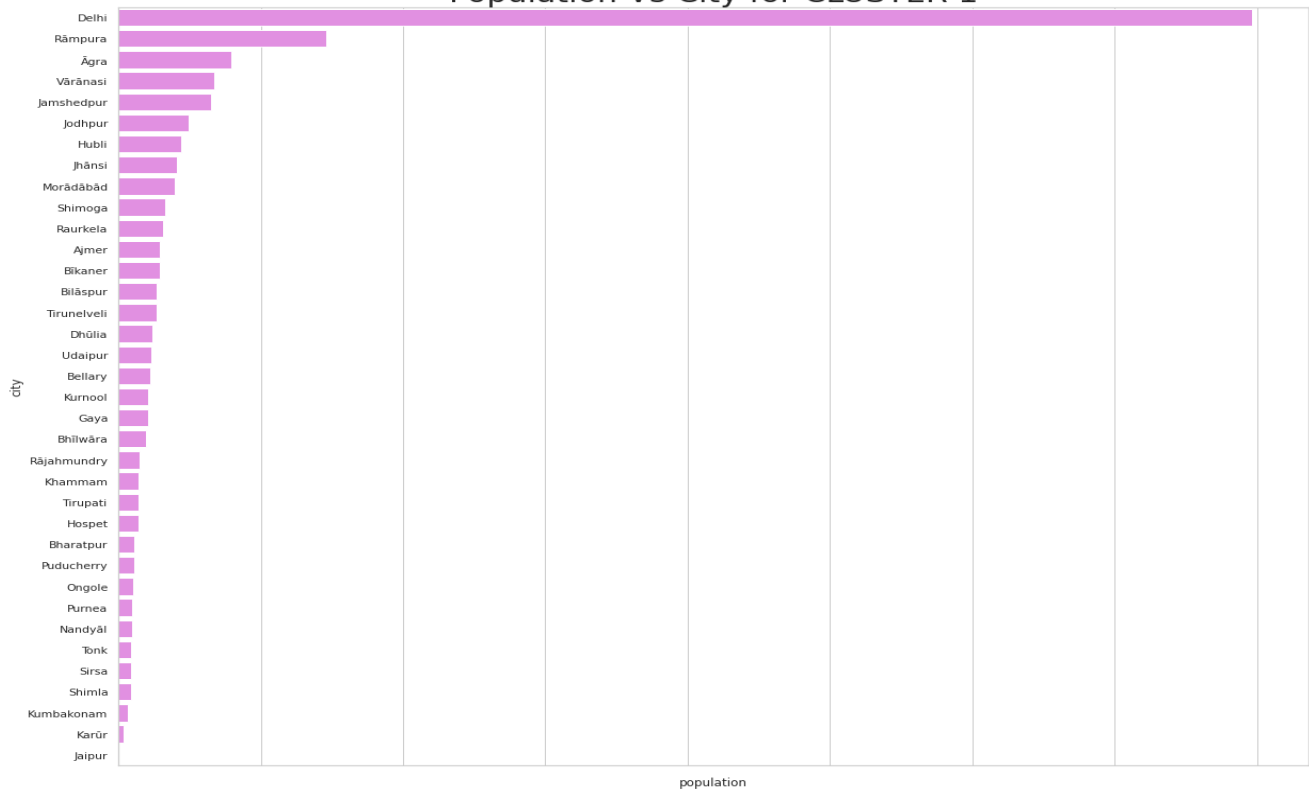
Once I have checked into some of the opportunities, next I have clustered the city into 5 major groups using the K means machine learning algorithm based on the similarities of Businesses which are in the core city. This will sort similar cities in which people are interested in similar venues and entertainments. Then I have merged the groups with cities and its population and have grouped the cities based on the cluster.

What this does is that businesses which are successfully running in major cities can be implemented into cities which have lesser competition and it can be implemented into the same cluster cities. We have made a plot of different Clusters against the city population. So it will be easier to spot the location to address a wider population in which the cities are similar. The detailed clusters are provided in the Notebook so that cities with lesser population can also be targeted based on certain type of people such as North Indian restaurants in South Indian cities and vice versa for a different food experience.

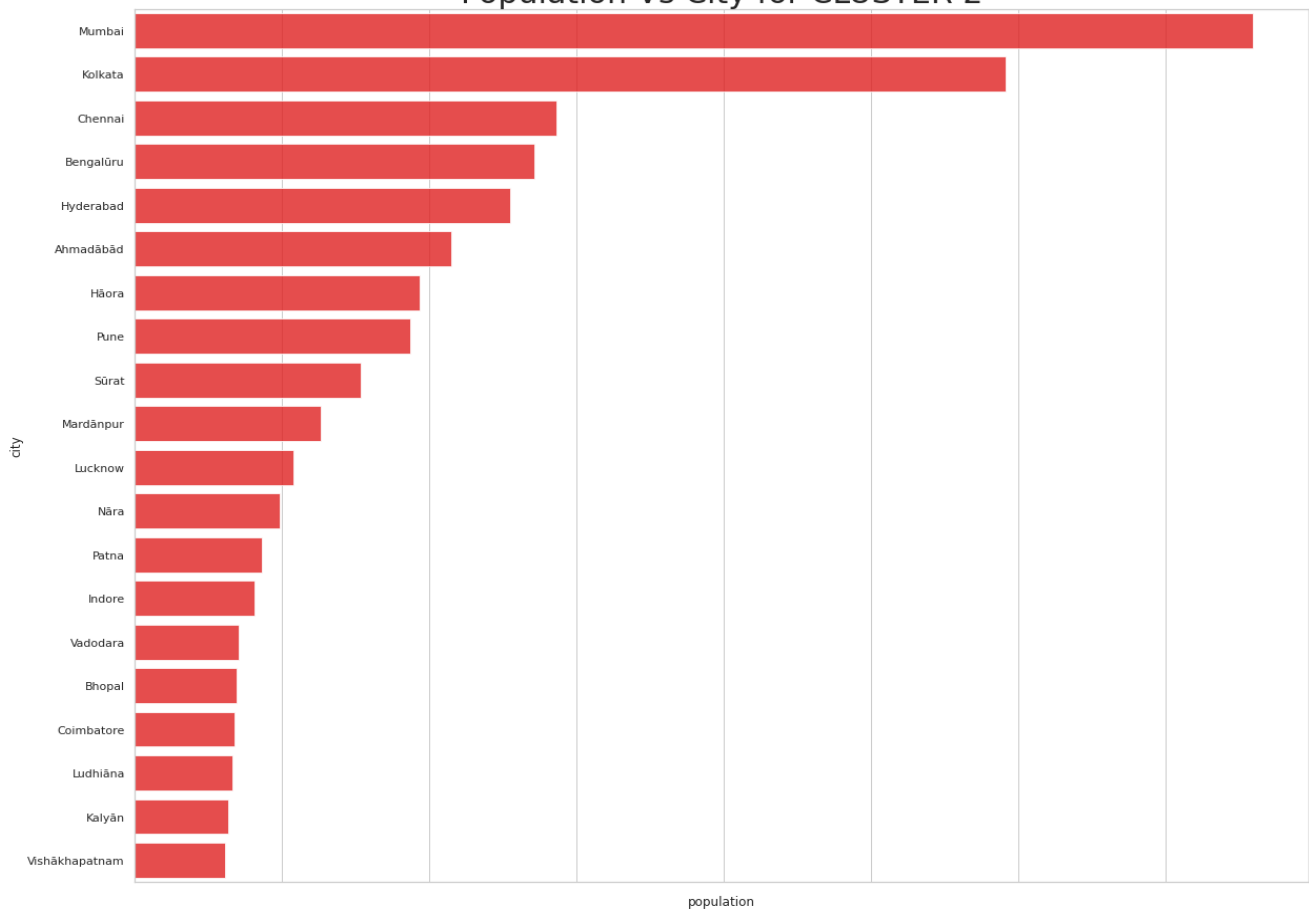




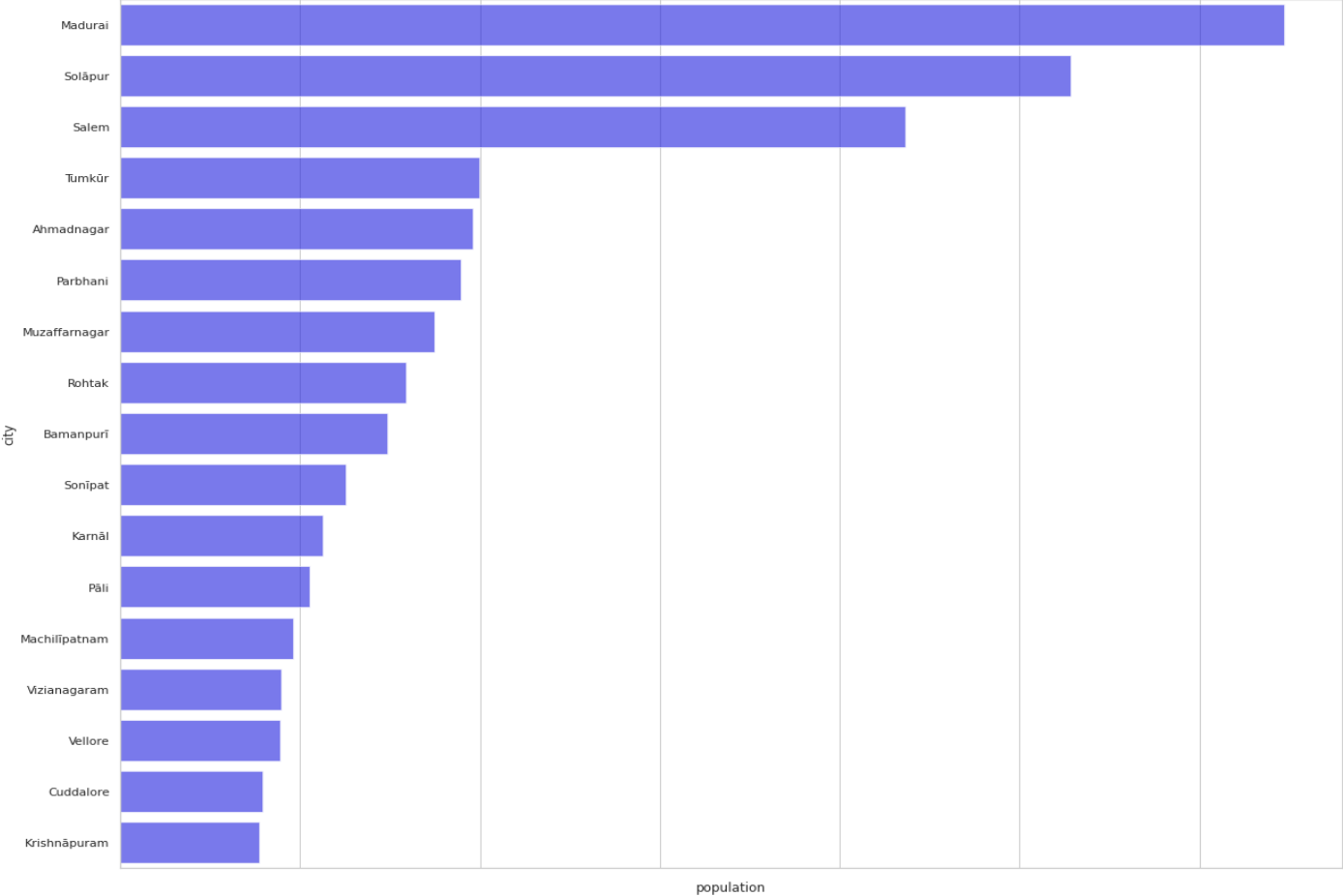
Population Vs City for CLUSTER 1



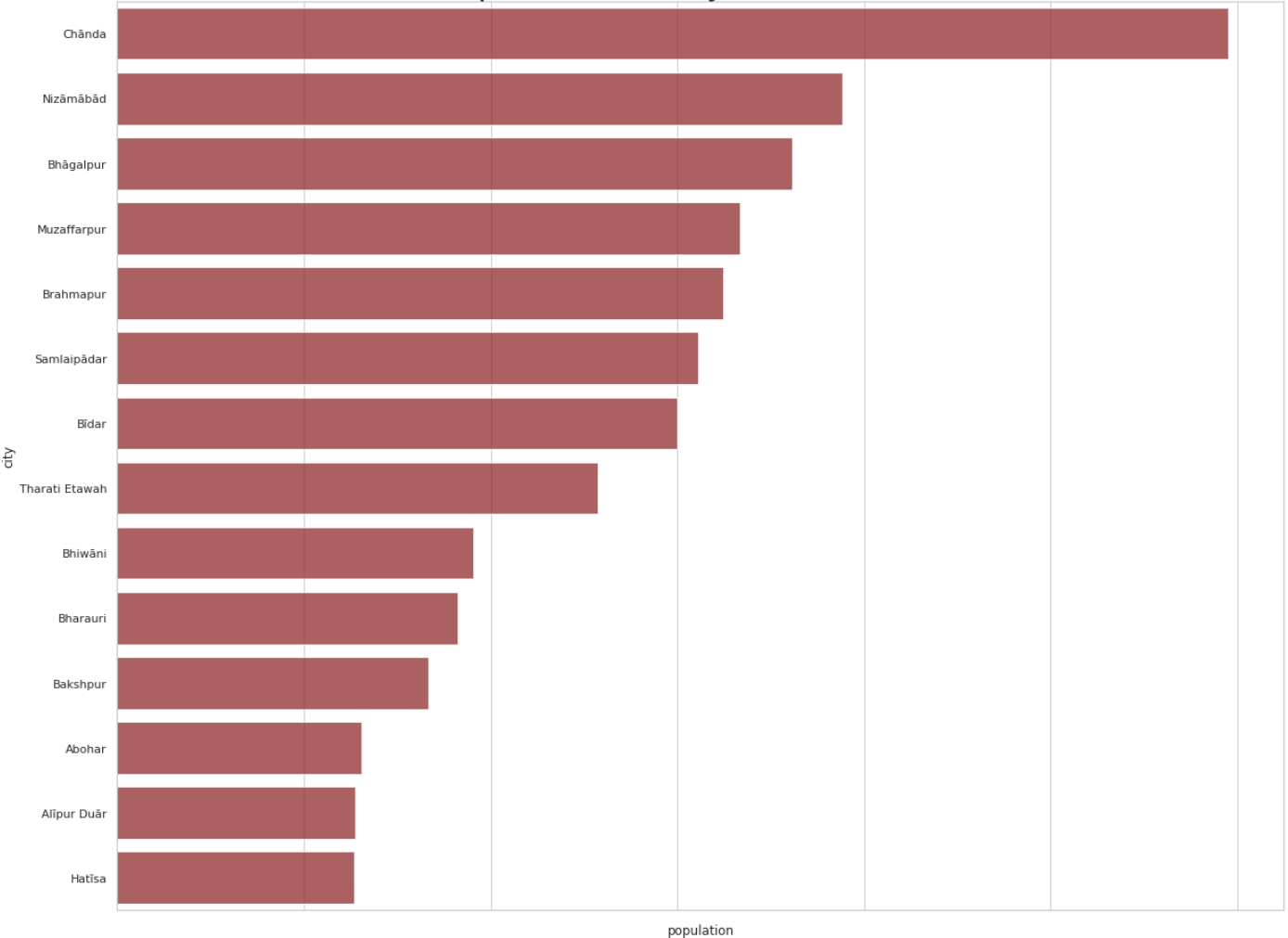
Population Vs City for CLUSTER 2



Population Vs City for CLUSTER 3



Population Vs City for CLUSTER 4



## **Result and Discussion**

With the above Exploratory Data Analysis, we can get many useful insights in starting a business, its location, the population we can address and the competition. By analysing the foursquare data, we found some of the valuable insights into many different opportunities and we have discussed some of it above in detail.

We have analysed the Jewellery business, Bowling Alley, Gym and Food truck business. What we have plotted are the cities and states which have these businesses. What we can suggest is that these businesses can be adopted in the similar cities which belong to the same cluster of the given city. There will be many cities similar to those cities while referring to those clusters.

While starting a business there will be many constraints and along with it, we can analyse the locations of the similar cities which will have a better chance of success. The details regarding the clusters are provided in depth in the notebook and it will be of a very great help.

## **Conclusion**

In this study, I analysed the base data along with the foursquare data in order to find some very interesting insights which will be very useful for the business startups. We know that these are some of the most important criteria which determine the success of a business venture.

Many companies, business firms, individuals who are looking to start a new business venture and who are already into business and looking to expand their business to different locations across India will be very much interested in this project.

Seize the opportunity when it comes and make the best use of it. This is a very important time in India and the opportunity to invest and explore is huge. We must make the best use out of this situation and be able to make a successful business which leaves a legacy.