# Using Data Analysis to find the best state to live and business in India.

## A.1 Introduction

India is having huge population which is 2nd in the world. This gives numbers of opportunities to those who want to start a new business in India.

Someone from outside of India who want to stay and start a new business in India will look after certain parameters to choose the best state to live. Three main parameters to consider are **Literacy rate**, **Crime rate** and **Per Capita NSDP**(Net state Domestic Product). Now not any single state will have highest literacy rate and Per capita NSDP along with lowest crime rate. So here we will use the Data Analysis methods to find the best state which has balances between all these three parameters.

After having the best state to live in we will find the best business options and in which city one of that businesses can be started to have the successful business.
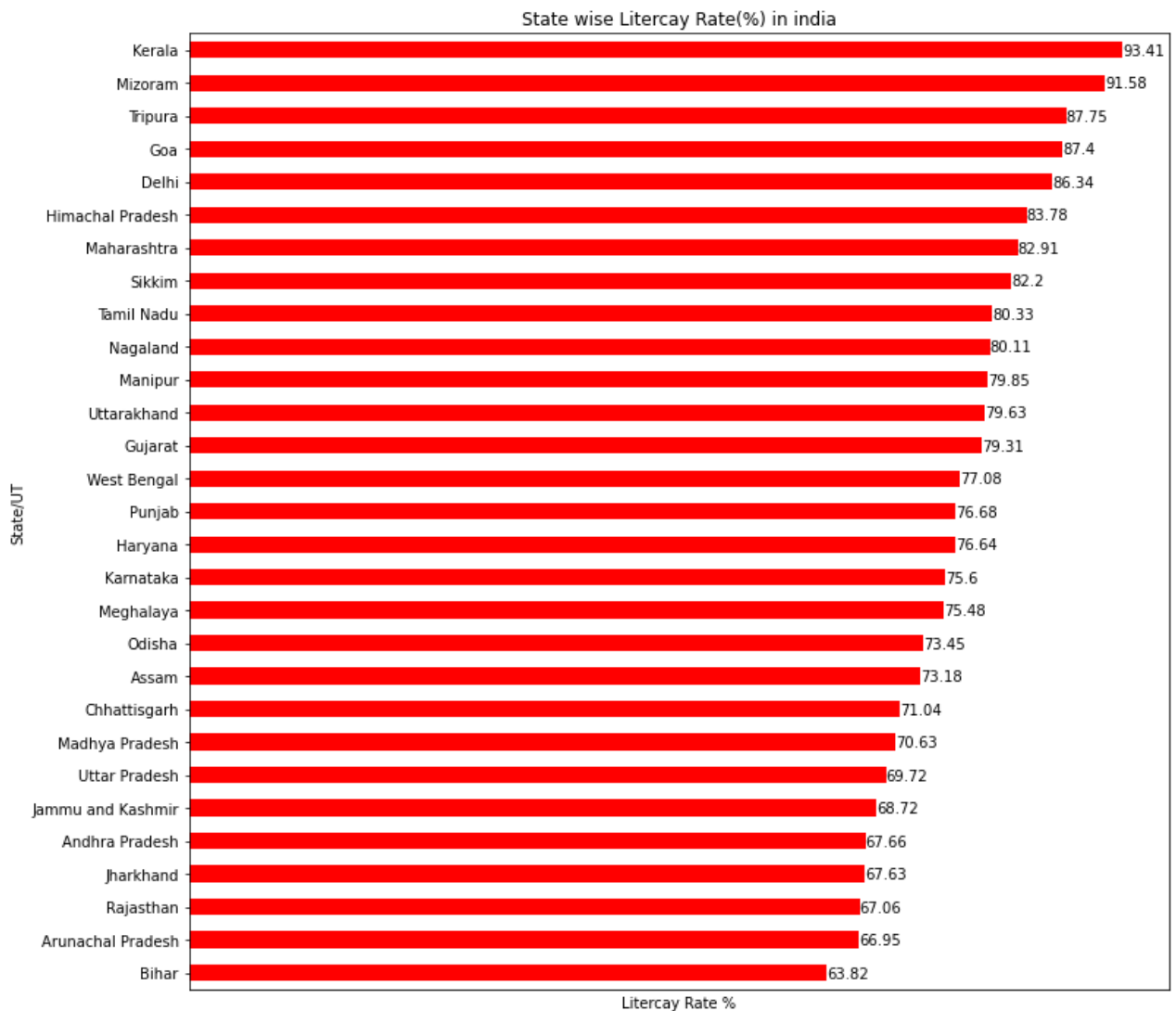
## A.2 Data Description

- I have used following state wise (latest available) data from the "Wikipedia" to find the best state to live in.

1. **literacy Rate(2011) :** percentage of the population of a given age group that can read and write.

1. **Crime Rate(2018) :** Number of crimes happening per population of one lac.

1. **Per Capita NSDP(2018) :** Gross domestic product (GDP) minus depreciation on a country's capital goods.

- I have used **FourSquare API** to retrieve business venues data to find the best business option and along with best business location.

## B. Methodology

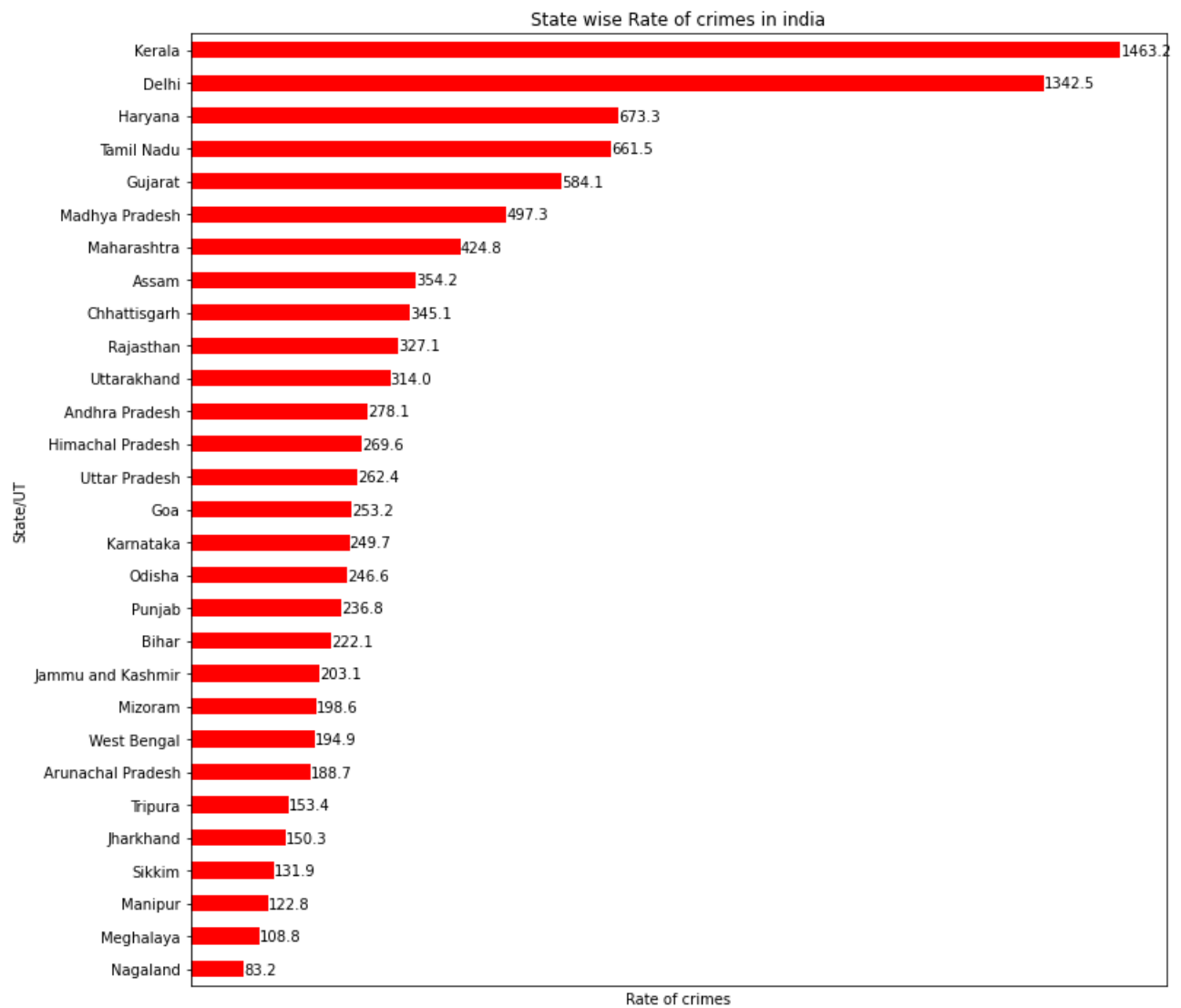- I first used *Pandas* library to retrieve all above data about the states. It is like below.

| State/UT | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) |
|---|---|---|---|
| Andhra Pradesh | 278.1 | 67.66 | 151173 |
| Arunachal Pradesh | 188.7 | 66.95 | 139588 |
| Assam | 354.2 | 73.18 | 82078 |
| Bihar | 222.1 | 63.82 | 43822 |
| Chhattisgarh | 345.1 | 71.04 | 96887 |
| Delhi | 1342.5 | 86.34 | 365529 |
| Goa | 253.2 | 87.40 | 458304 |
| Gujarat | 584.1 | 79.31 | 197447 |
| Haryana | 673.3 | 76.64 | 236147 |
| Himachal Pradesh | 269.6 | 83.78 | 183104 |
| Jammu and Kashmir | 203.1 | 68.72 | 91882 |
| Jharkhand | 150.3 | 67.63 | 76019 |
| Karnataka | 249.7 | 75.60 | 210887 |
| Kerala | 1463.2 | 93.41 | 204105 |
| Madhya Pradesh | 497.3 | 70.63 | 90998 |
| Maharashtra | 424.8 | 82.91 | 191736 |
| Manipur | 122.8 | 79.85 | 69978 |
| Meghalaya | 108.8 | 75.48 | 89024 |
| Mizoram | 198.6 | 91.58 | 138401 |
| Nagaland | 83.2 | 80.11 | 116882 |
| Odisha | 246.6 | 73.45 | 95164 |
| Punjab | 236.8 | 76.68 | 154996 |
| Rajasthan | 327.1 | 67.06 | 110606 |
| Sikkim | 131.9 | 82.20 | 357643 |
| Tamil Nadu | 661.5 | 80.33 | 193750 |
| Tripura | 153.4 | 87.75 | 113102 |
| Uttar Pradesh | 262.4 | 69.72 | 66512 |
| Uttarakhand | 314.0 | 79.63 | 198738 |
| West Bengal | 194.9 | 77.08 | 109491 |

- let's visualize *State Vs. Literacy* Rate %.

**State wise Litercay Rate(%) in india**

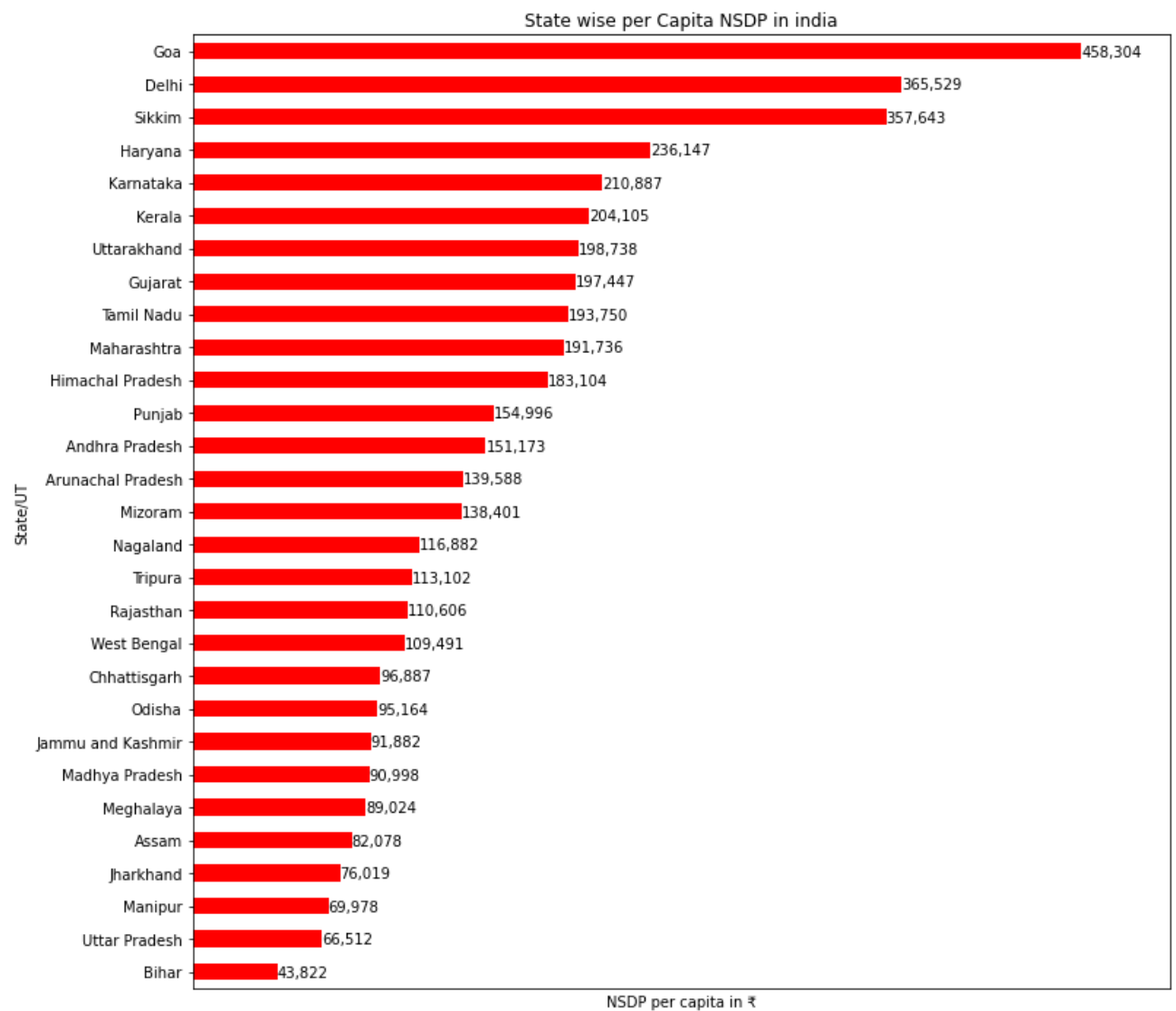| State/UT | Litercay Rate % |
| --- | --- |
| Kerala | 93.41 |
| Mizoram | 91.58 |
| Tripura | 87.75 |
| Goa | 87.4 |
| Delhi | 86.34 |
| Himachal Pradesh | 83.78 |
| Maharashtra | 82.91 |
| Sikkim | 82.2 |
| Tamil Nadu | 80.33 |
| Nagaland | 80.11 |
| Manipur | 79.85 |
| Uttarakhand | 79.63 |
| Gujarat | 79.31 |
| West Bengal | 77.08 |
| Punjab | 76.68 |
| Haryana | 76.64 |
| Karnataka | 75.6 |
| Meghalaya | 75.48 |
| Odisha | 73.45 |
| Assam | 73.18 |
| Chhattisgarh | 71.04 |
| Madhya Pradesh | 70.63 |
| Uttar Pradesh | 69.72 |
| Jammu and Kashmir | 68.72 |
| Andhra Pradesh | 67.66 |
| Jharkhand | 67.63 |
| Rajasthan | 67.06 |
| Arunachal Pradesh | 66.95 |
| Bihar | 63.82 |

*Here we can see that **Kerala** is highly literate state while **Bihar** is less literate state.*

- let's visualize *State Vs. Rate of crimes %*.



State wise Rate of crimes in india

| State/UT | Rate of crimes |
|---|---|
| Kerala | 1463.2 |
| Delhi | 1342.5 |
| Haryana | 673.3 |
| Tamil Nadu | 661.5 |
| Gujarat | 584.1 |
| Madhya Pradesh | 497.3 |
| Maharashtra | 424.8 |
| Assam | 354.2 |
| Chhattisgarh | 345.1 |
| Rajasthan | 327.1 |
| Uttarakhand | 314.0 |
| Andhra Pradesh | 278.1 |
| Himachal Pradesh | 269.6 |
| Uttar Pradesh | 262.4 |
| Goa | 253.2 |
| Karnataka | 249.7 |
| Odisha | 246.6 |
| Punjab | 236.8 |
| Bihar | 222.1 |
| Jammu and Kashmir | 203.1 |
| Mizoram | 198.6 |
| West Bengal | 194.9 |
| Arunachal Pradesh | 188.7 |
| Tripura | 153.4 |
| Jharkhand | 150.3 |
| Sikkim | 131.9 |
| Manipur | 122.8 |
| Meghalaya | 108.8 |
| Nagaland | 83.2 |

*Here we can see that **Kerala** is having highest crime rate while **Nagaland** is having lowest crime rate.*

- let's visualize *State Vs. NSDP %*.



State wise per Capita NSDP in india

| State/UT | NSDP per capita in ₹ |
|---|---|
| Goa | 458,304 |
| Delhi | 365,529 |
| Sikkim | 357,643 |
| Haryana | 236,147 |
| Karnataka | 210,887 |
| Kerala | 204,105 |
| Uttarakhand | 198,738 |
| Gujarat | 197,447 |
| Tamil Nadu | 193,750 |
| Maharashtra | 191,736 |
| Himachal Pradesh | 183,104 |
| Punjab | 154,996 |
| Andhra Pradesh | 151,173 |
| Arunachal Pradesh | 139,588 |
| Mizoram | 138,401 |
| Nagaland | 116,882 |
| Tripura | 113,102 |
| Rajasthan | 110,606 |
| West Bengal | 109,491 |
| Chhattisgarh | 96,887 |
| Odisha | 95,164 |
| Jammu and Kashmir | 91,882 |
| Madhya Pradesh | 90,998 |
| Meghalaya | 89,024 |
| Assam | 82,078 |
| Jharkhand | 76,019 |
| Manipur | 69,978 |
| Uttar Pradesh | 66,512 |
| Bihar | 43,822 |

Here we can see that **Goa** is having highest per Capita NSDP while **Bihar** is having lowest per Capita NSDP.

**Now let's make clusters of the states to identify which states is having balance between all these three parameters. I used "*K-means*" algorithm to make clusters of states.**

To make clusters of the states according to these parameters. We find the four optimal number of cluster using *Elbow method*.
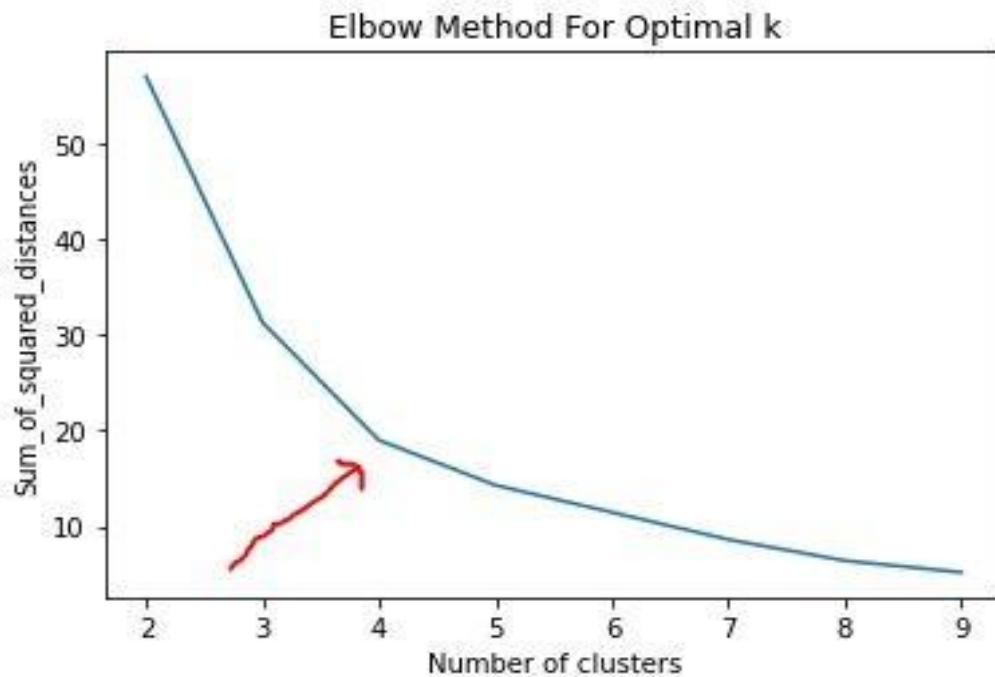
Here is the visualization of the clustered states where each colour shows the one cluster.



| State | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) |
|---|---|---|---|
| Andhra Pradesh | 278.1 | 67.66 | 151173 |
| Uttar Pradesh | 262.4 | 69.72 | 66512 |
| Rajasthan | 327.1 | 67.06 | 110606 |
| Odisha | 246.6 | 73.45 | 99196 |
| Meghalaya | 108.8 | 75.48 | 84725 |
| Manipur | 122.8 | 79.85 | 75226 |
| Jharkhand | 150.3 | 67.63 | 73155 |
| Jammu and Kashmir | 203.1 | 68.72 | 92347 |
| Madhya Pradesh | 497.3 | 70.63 | 90165 |
| West Bengal | 194.9 | 77.08 | 101138 |
| Arunachal Pradesh | 188.7 | 66.95 | 139588 |
| Assam | 354.2 | 73.18 | 82837 |
| Chhattisgarh | 345.1 | 71.04 | 92413 |
| Bihar | 222.1 | 63.82 | 40982 |
| Kerala | 1463.2 | 99.41 | 204105 |
| Delhi | 1342.5 | 86.34 | 358430 |
| Tripura | 153.4 | 87.75 | 112849 |
| Tamil Nadu | 661.5 | 80.33 | 193964 |
| Punjab | 236.8 | 76.68 | 154313 |
| Himachal Pradesh | 269.6 | 83.78 | 183104 |
| Haryana | 673.3 | 76.64 | 236147 |
| Maharashtra | 424.8 | 82.91 | 191736 |
| Uttarakhand | 314 | 79.63 | 198738 |
| Karnataka | 249.7 | 75.6 | 212477 |
| Gujarat | 584.1 | 79.31 | 195845 |
| Nagaland | 83.2 | 80.11 | 116882 |
| Mizoram | 198.6 | 91.58 | 147602 |
| Sikkim | 131.9 | 82.2 | 380926 |
| Goa | 253.2 | 87.4 | 430081 |

We can find the centroid of the clusters by taking mean value of each parameter grouped by cluster label. On the basis of the centroid values we can differentiate between the clusters.

| clusterLabel | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) |
|---|---|---|---|
| 0 | 1402.85 | 92.88 | 281267.50 |
| 1 | 349.91 | 81.30 | 176696.09 |
| 2 | 253.72 | 71.04 | 92901.23 |
| 3 | 192.55 | 84.80 | 405503.50 |

Let us standardize the value of the cluster means with their maximum values to have better comparison among clusters.

| clusterLabel | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) |
|---|---|---|---|
| 0 | 1.00 | 1.00 | 0.69 |
| 1 | 0.25 | 0.88 | 0.44 |
| 2 | 0.18 | 0.76 | 0.23 |
| 3 | 0.14 | 0.91 | 1.00 |

As can be seen in above centroid values maximum Per Capita NSDP(₹) lies in cluster 3, maximum Litercay Rate lies in cluster 0 while lowest Rate of crimes lies in cluster 3.

So to find the relatively best cluster which balances between all three parameters we define a score which multiplies the positive parameters (Litercay Rate % & Per Capita NSDP(₹)) and divides the result with the negative parameters (Rate of crimes). It gives following table.

| clusterLabel | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) | Score |
|---|---|---|---|---|
| 3 | 0.14 | 0.91 | 1.00 | 6.50 |
| 1 | 0.25 | 0.88 | 0.44 | 1.55 |
| 2 | 0.18 | 0.76 | 0.23 | 0.97 |
| 0 | 1.00 | 1.00 | 0.69 | 0.69 |

Based upon that score, here we find two worst states **Delhi** & **Kerala** to live in. Both states are having high rate of literacy and High NSDP but at the same time both states are having high rate of crimes which makes them states not to choose to live in.

| State/UT | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) | clusterLabel |
|---|---|---|---|---|
| Delhi | 1342.5 | 86.34 | 358430 | 0 |
| Kerala | 1463.2 | 99.41 | 204105 | 0 |

And we find two best states **Goa** & **Sikkim** to live in which balances between all those three parameters.

| State/UT | Rate of crimes | Litercay Rate % | Per Capita NSDP(₹) | clusterLabel |
|---|---|---|---|---|
| Goa | 253.2 | 87.4 | 430081 | 3 |
| Sikkim | 131.9 | 82.2 | 380926 | 3 |

*One can choose between these two states to live there and start a new business.  Let's assume that our client chooses **Goa**.*

**Now Using FOURSQUARE API, we will explore Neighborhoods and venues in GOA state to find the best business category and the best location for that business so that business can be more profitable.**

FourSquare API returns data with following cities in Goa along with its business venues

| City |
| --- |
| Agonda |
| Anjuna |
| Bambolim |
| Bardez |
| Benaulim |
| Betalbatim |
| Candolim |
| Cavelossim |
| Colva |
| Madgaon |
| Old Goa |
| Palolem |
| Panjim |
| Ponda |
| Salcete |
| Vagator |
| Varca |
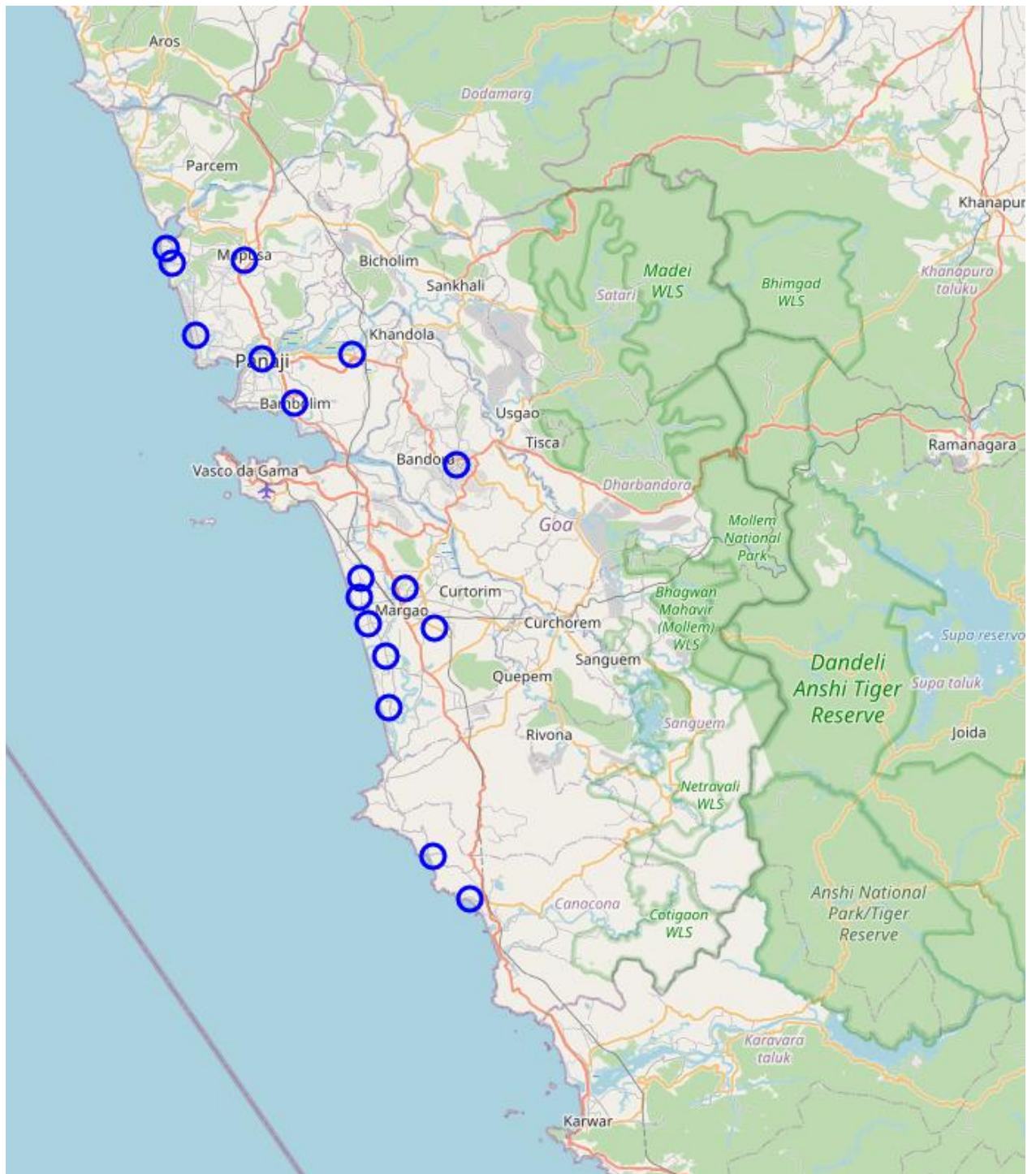
With the venues data provided by the FourSqaure API and after doing the exploratory data analysis I found the following table which shows 1st less frequent to 10th less frequent business in the respective city.

| | 1st less freq. | 2nd less freq. | 3rd less freq. | 4th less freq. | 5th less freq. | 6th less freq. | 7th less freq. | 8th less freq. | 9th less freq. | 10th less freq. |
|---|---|---|---|---|---|---|---|---|---|---|
| Agonda | American Restaurant | Portuguese Restaurant | Nightclub | Mughlai Restaurant | Lounge | Juice Bar | Hostel | Historic Site | French Restaurant | Food Court |
| Anjuna | Vegetarian / Vegan Restaurant | Swiss Restaurant | Jazz Club | Gastropub | Food Court | Food | Farm | Eastern European Restaurant | Deli / Bodega | Grocery Store |
| Bambolim | Grocery Store | Food Court | Deli / Bodega | Convenience Store | Swiss Restaurant | Chinese Restaurant | Motel | Caribbean Restaurant | Mughlai Restaurant | Pool |
| Bardez | Vegetarian / Vegan Restaurant | Swiss Restaurant | Hotel Bar | Gastropub | Food Court | Food | Farm | Eastern European Restaurant | Deli / Bodega | Grocery Store |
| Benaulim | Greek Restaurant | Swiss Restaurant | Hostel | Hotel Bar | Convenience Store | Juice Bar | Motel | Lounge | Breakfast Spot | Brazilian Restaurant |
| Betalbatim | Juice Bar | Rest Area | Swiss Restaurant | Brazilian Restaurant | Breakfast Spot | Motel | Deli / Bodega | Convenience Store | Hostel | Hotel Bar |
| Candolim | Vegetarian / Vegan Restaurant | Gastropub | Food Court | Food | Farm | Eastern European Restaurant | Hotel Bar | Deli / Bodega | Grocery Store | Chinese Restaurant |
| Cavelossim | Greek Restaurant | Fish & Chips Shop | Historic Site | Hostel | Convenience Store | Juice Bar | Swiss Restaurant | Motel | Lounge | Burger Joint |
| Colva | Juice Bar | Rest Area | Nightclub | Swiss Restaurant | Brazilian Restaurant | Breakfast Spot | Motel | Convenience Store | Hotel Bar | Hostel |
| Madgaon | Lounge | Rest Area | Swiss Restaurant | Nightclub | Brazilian Restaurant | Breakfast Spot | Motel | Hostel | Juice Bar | Convenience Store |
| Old Goa | Vegetarian / Vegan Restaurant | Gastropub | Hotel Bar | Food Court | Food | Eastern European Restaurant | Deli / Bodega | Convenience Store | Swiss Restaurant | Chinese Restaurant |
| Palolem | American Restaurant | Portuguese Restaurant | Nightclub | Mughlai Restaurant | Motel | Lounge | Juice Bar | Hostel | Historic Site | French Restaurant |
| Panjim | Vegetarian / Vegan Restaurant | Food Court | Food | Grocery Store | Eastern European Restaurant | Deli / Bodega | Hotel Bar | Swiss Restaurant | Caribbean Restaurant | Chinese Restaurant |
| Ponda | Hotel Bar | Deli / Bodega | Motel | Lounge | Breakfast Spot | Juice Bar | Swiss Restaurant | Hostel | French Restaurant | Gastropub |
| Salcete | Greek Restaurant | Fish & Chips Shop | Hostel | Hotel Bar | Convenience Store | Juice Bar | Swiss Restaurant | Motel | Lounge | Burger Joint |
| Varca | Greek Restaurant | Fish & Chips Shop | Hostel | Hotel Bar | Convenience Store | Juice Bar | Swiss Restaurant | Motel | Lounge | Burger Joint |

Now I find the most frequent category at every position which would be likely most profitable business as it would be also absent in nearby cities also.

| 1st less freq. | 2nd less freq. | 3rd less freq. | 4th less freq. | 5th less freq. | 6th less freq. | 7th less freq. | 8th less freq. | 9th less freq. | 10th less freq. |
|---|---|---|---|---|---|---|---|---|---|
| Vegetarian / Vegan Restaurant | Swiss Restaurant | Nightclub | Hotel Bar | Convenience Store | Juice Bar | Swiss Restaurant | Convenience Store | Lounge | Chinese Restaurant |
| 5 | 3 | 3 | 3 | 4 | 6 | 4 | 3 | 3 | 3 |

Upper data shows that in total 17 nos. of cities Vegetarian / Vegan Restaurant is lacking(Having less frequency) in 5 nos. of cities, Swiss Restaurant is lacking in 3 nos. of cities and so on...

## C. Results

Finally following is the list of cities in GOA state for each business category where the starting one of that business will be more profitable. Which is our main aim of this business problem.

**Vegetarian / Vegan Restaurant**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Anjuna |
| 1 | Bardez |
| 2 | Candolim |
| 3 | Old Goa |
| 4 | Panjim |

**Swiss Restaurant**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Anjuna |
| 1 | Bardez |
| 2 | Benaulim |

**Nightclub**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Agonda |
| 1 | Colva |
| 2 | Palolem |

**Hotel Bar**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Benaulim |
| 1 | Salcete |
| 2 | Varca |

**Convenience Store**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Benaulim |
| 1 | Cavelossim |
| 2 | Salcete |
| 3 | Varca |

**Juice Bar**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Agonda |
| 1 | Benaulim |
| 2 | Cavelossim |
| 3 | Ponda |
| 4 | Salcete |
| 5 | Varca |

**Swiss Restaurant**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Cavelossim |
| 1 | Ponda |
| 2 | Salcete |
| 3 | Varca |

**Convenience Store**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Betalbatim |
| 1 | Colva |
| 2 | Old Goa |

**Lounge**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Cavelossim |
| 1 | Salcete |
| 2 | Varca |

**Chinese Restaurant**
Business can be started in following cities

|   | City |
|---|---|
| 0 | Candolim |
| 1 | Old Goa |
| 2 | Panjim |

## D. Discussion and Conclusion:

So after using various libraries and API in python We can do some data analysis and run some machine learning algorithm to extract out useful information from the data available in the web.

Here we found the GOA as a best state to live and list of business which can be stared in respective listed cities to have maximum benefits from that business.