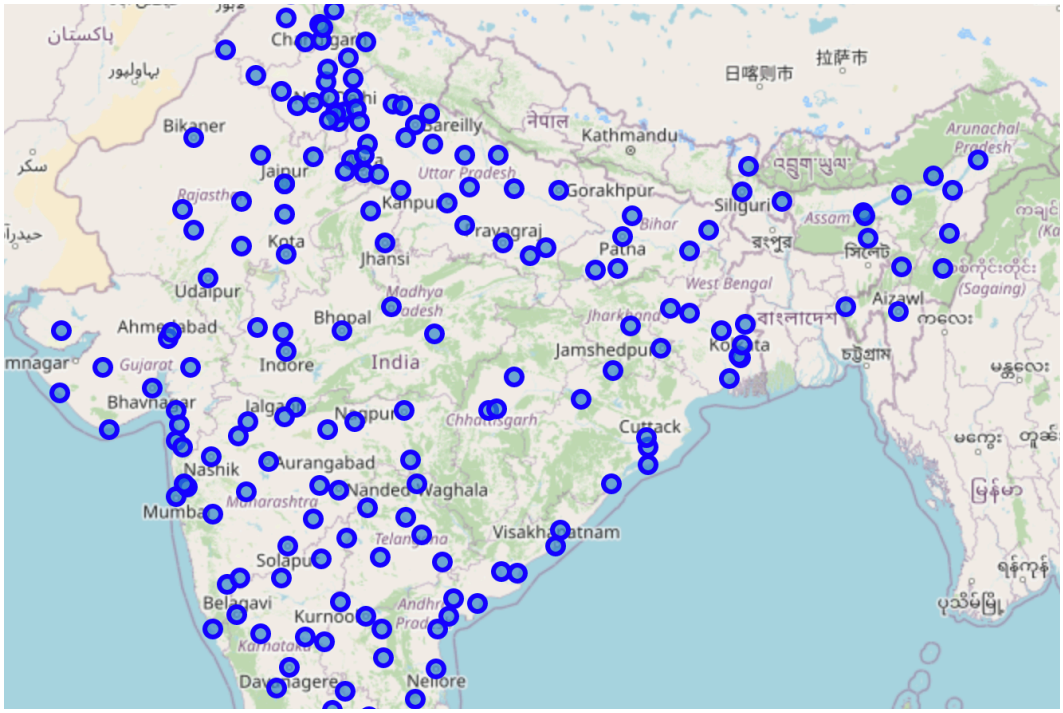
OPENING A RESTAURANT IN INDIA

ABSTRACT



Raj is going to India to open a new restaurant, where should he go, which City is best to invest ?

IBM DATA SCIENCE - FINAL CAPSTONE

Introduction

1.1 Background

India is one of the country's with the one of the best work-life balance, most of Middle East Asian people move out to places like Pakistan or Nepal because their balance. Exhausted of working 60 hours in Nepal, getting poor salaries at the end of the month they go abroad to find new opportunities and start their new life.

1.2 Problem

Raj is thinking to open a new restaurant in India and cook typical dishes in some city in India, but the obviously question is, where in India?

1.3 Interest / Target Audience

This will be interesting for any person planning to move to India for opening a restaurant, but also will be interesting for those persons thinking to open any business related thing in the India because I will display the most common venues in each of the different cities in India, so maybe someone is planning to open a Supermarket, or an IT business and they can find interesting where it is the best place for their purpose.

Data acquisition and cleaning

2.1 Data sources

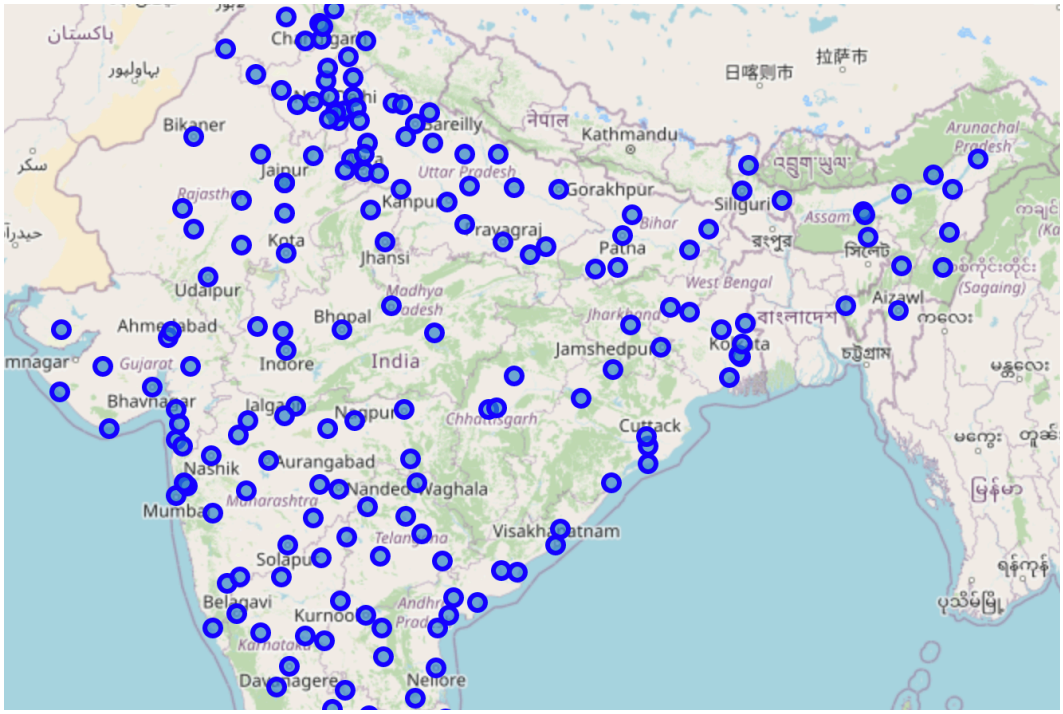
I will use the .csv file provided by the following page: https://simplemaps.com/data/in-cities consisting on all the different cities within India. In addition, this dataset will provide me the different latitude and longitude of all of those cities, which will be useful for printing them in a map using Foursquare API. Also, the dataset will provide the population of the 14 first cities (or the ones with more population like The Mumbai or Delhi).

2.2 Data cleaning and how will be processed

	City	Latitude	Longitude
0	Mumbai	18.987807	72.836447
1	Delhi	28.651952	77.231495
2	Kolkata	22.562627	88.363044
3	Chennai	13.084622	80.248357
4	Bengalūru	12.977063	77.587106

First of all, I sorted the different cities by their population to have the most populated at first because those are usually the most common chosen by the people who is going to India. After, I will delete the rows that will not be useful for this purpose, or at least for this project, having at the end only a table with the name of the city, their latitude and longitude that will be necessarily for showing into the map

Once I have the different cities with their latitude and longitude will be the time to print them in the map, for this purpose I will use the folium library of Python to display it in the map.



India Map

Note: At this point I have not worked with the data yet, this is only for showing the different cities into the map.

Once I have the different data displayed into the map, now is time to start applying the Data Analysis and fetch the data in different Clusters with "Similar data"

Exploratory Data Analysis

Now that we have all the cities represented in the map with their latitude and longitude, I will use my Foursquare credentials to be able to fetch all the venues surrounding that city with an radius of 500, so I will be only focused in the centre on the city because there is where usually people go for restaurants.

Once I have every venue in the city, I will use One Hot Encoding to extract all the dummy variables that will be necessarily to implement in the future the K means clustering.

	Yoga Studio	ATM	Accessories Store	Adult Boutique	American Restaurant	Andhra Restaurant	Arcade	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Thrift / Vintage Store	Tourist Information Center	S
0	0	0	0	0	0	0	0	0	0	0	...	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	

Now that I have the different venues like Hotels, Ice Cream shops, Restaurants and so on, I will obtain the mean of all of them to extract the mean of each venue in that city.

With this data we can observe that Delhi the most common thing is Restaurant, but also I can see that Chennai is also Restaurant, that means that probably in the future when I will apply clustering, those cities will be in the same cluster.

```
----Abohar----
                venue  freq
0   Convenience Store  0.17
1               Café  0.17
2                ATM  0.17
3  Photography Studio  0.17
4       Train Station  0.17


----Agartala----
           venue  freq
0  Historic Site  0.14
1    Coffee Shop  0.14
2      Multiplex  0.14
3           Lake  0.14
4          Hotel  0.14


----Ahmadnagar----
                         venue  freq
0            Indian Restaurant   1.0
1                 Yoga Studio   0.0
2  Northeast Indian Restaurant   0.0
3               Movie Theater   0.0
4           Mughlai Restaurant   0.0
```

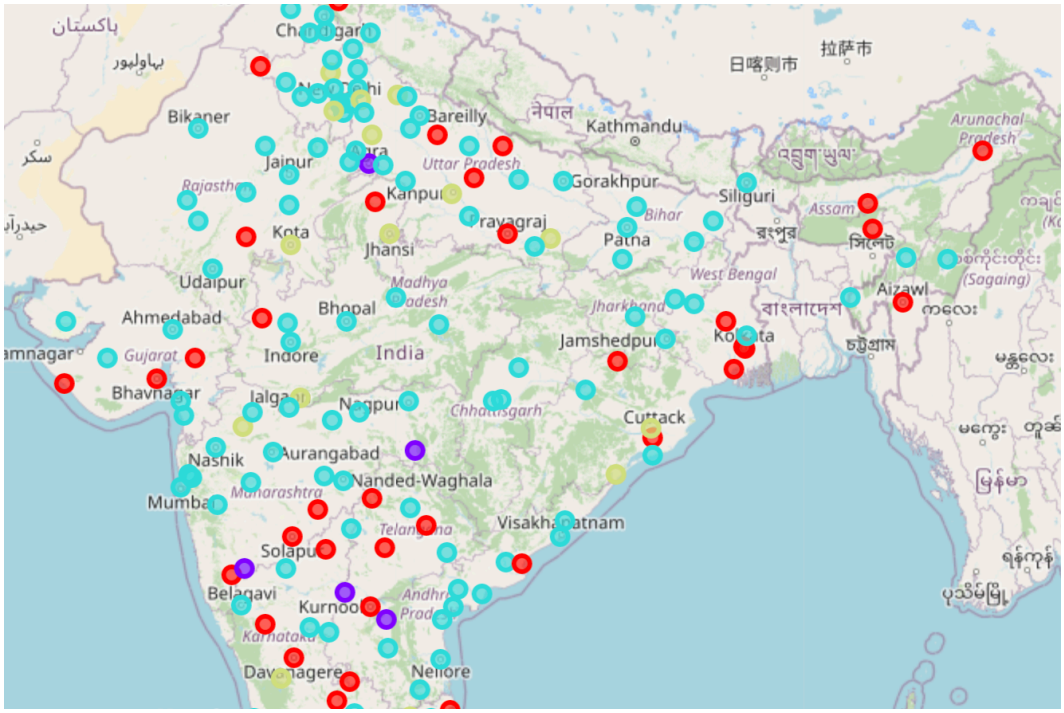
Once I get the frequency of each venue in each city, I will make a table to make it more user friendly and easily understandable.

	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
0	Abohar	Café	Train Station	Fast Food Restaurant	Convenience Store	Photography Studio	ATM	Dairy Store
1	Agartala	Lake	Coffee Shop	Hotel	Historic Site	Science Museum	Multiplex	Salad Place
2	Ahmadnagar	Indian Restaurant	Dumpling Restaurant	Food	Flea Market	Field	Fast Food Restaurant	Farmers Market
3	Ahmadābād	Snack Place	Historic Site	Fast Food Restaurant	Indian Restaurant	Hotel Bar	Hotel	Flea Market
4	Aizawl	Hotel	Indian Chinese Restaurant	Shopping Mall	Electronics Store	Food	Flea Market	Field

With that table, I will Cluster the data in 4 different tables (the optimum number of clusters) and finally, to be able to represent it again I will need to join the latitude and longitude in that table, having as a result the following one:

	City	Latitude	Longitude	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
0	Mumbai	18.987807	72.836447	2.0	Multiplex	Snack Place	Vegetarian / Vegan Restaurant	Plaza	Fast Food Restaurant	Train Station	Women's Store
1	Delhi	28.651952	77.231495	0.0	Indian Restaurant	Snack Place	Market	Flea Market	Hotel	Historic Site	Paper / Office Supplies Store
2	Kolkata	22.562627	88.363044	2.0	IT Services	Video Store	Hotel	Park	Market	Multiplex	Dumpling Restaurant
3	Chennai	13.084622	80.248357	0.0	Indian Restaurant	Italian Restaurant	Fast Food Restaurant	Yoga Studio	Dairy Store	Deli / Bodega	Department Store
4	Bengalūru	12.977063	77.587106	2.0	Vineyard	Plaza	Coffee Shop	Hotel	Metro Station	Women's Store	Electronics Store

Note: As you can see, the cluster labels are floats, I will need to make them int for be able to represent them with different colors in the map with each city separated by their cluster. Now that I have the most common venue, the City, the longitude and which Cluster is inside in, it is time to plot it in the map to see the results and the different clusters created in India.



Different Regions in India Marked

And if I have a look to the data:

Cluster 0 (Brown cities)

These cities are mostly the touristic ones and as you can see, they have plenty of restaurants, Bars or food related business. In addition, Chennai is also in this cluster as we predicted.

	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
1	Delhi	Indian Restaurant	Snack Place	Market	Flea Market	Hotel	Historic Site	Paper / Office Supplies Store
3	Chennai	Indian Restaurant	Italian Restaurant	Fast Food Restaurant	Yoga Studio	Dairy Store	Deli / Bodega	Department Store
6	Ahmadābād	Snack Place	Historic Site	Fast Food Restaurant	Indian Restaurant	Hotel Bar	Hotel	Flea Market
8	Pune	Indian Restaurant	Ice Cream Shop	Women's Store	Jewelry Store	Seafood Restaurant	Road	Coffee Shop
12	Lucknow	Indian Restaurant	ATM	Market	Electronics Store	Food Court	Food	Flea Market
18	Coimbatore	Indian Restaurant	Stadium	Ice Cream Shop	Hotel	Chinese Restaurant	Travel & Transport	Train Station
19	Ludhiāna	Indian Restaurant	Fast Food Restaurant	Plaza	Shopping Mall	Electronics Store	Food	Flea Market
24	Nāsik	Bus Station	Indian Restaurant	Hotel	Motel	Electronics Store	Food	Flea Market
26	Farīdābād	Indian Restaurant	Convenience Store	Shopping Mall	Electronics Store	Food	Flea Market	Field
31	Madurai	Indian Restaurant	Department Store	Clothing Store	IT Services	Hotel Bar	Flea Market	Field
35	Amritsar	Indian Restaurant	Pizza Place	Punjabi Restaurant	Coffee Shop	Fast Food Restaurant	Museum	Diner
45	Chandigarh	Indian Restaurant	Dessert Shop	Café	Theater	Deli / Bodega	Department Store	Dairy Store
47	Thiruvananthapuram	Indian Restaurant	Clothing Store	Multiplex	Shopping Mall	Arcade	Snack Place	Beach
49	Hubli	Food Court	Indian Restaurant	Café	Bus Station	Electronics Store	Food	Flea Market

Cluster 1 (Red cities)

This cluster seem to have the cities that are in between tourist places, or probably for working companies or fabrics because most of them have lot of Clothing Store.

	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
9	Sūrat	Clothing Store	Indian Restaurant	Ice Cream Shop	Food Court	Flea Market	Field	Fast Food Restaurant
40	Bhilai	Clothing Store	Electronics Store	Food Court	Food	Flea Market	Field	Fast Food Restaurant
134	Imphal	Clothing Store	Electronics Store	Food Court	Food	Flea Market	Field	Fast Food Restaurant
141	Hāpur	Clothing Store	Electronics Store	Food Court	Food	Flea Market	Field	Fast Food Restaurant
187	Pīlibhīt	Clothing Store	Electronics Store	Food Court	Food	Flea Market	Field	Fast Food Restaurant

Cluster 2 (Blue cities)

About the Blue cities seem to be the same city and in the data that we fetched they had different name because they have the same Venues with multiplexes mostly.

	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
0	Mumbai	Multiplex	Snack Place	Vegetarian / Vegan Restaurant	Plaza	Fast Food Restaurant	Train Station	Women's Store
2	Kolkata	IT Services	Video Store	Hotel	Park	Market	Multiplex	Dumpling Restaurant
4	Bengalūru	Vineyard	Plaza	Coffee Shop	Hotel	Metro Station	Women's Store	Electronics Store
5	Hyderabad	Home Service	Women's Store	Dumpling Restaurant	Food	Flea Market	Field	Fast Food Restaurant
7	Hāora	Indian Sweet Shop	Business Service	Electronics Store	Food	Flea Market	Field	Fast Food Restaurant
11	Rāmpura	Café	ATM	IT Services	Dumpling Restaurant	Food	Flea Market	Field
14	Patna	Board Shop	Pizza Place	Men's Store	Fabric Shop	Food Court	Food	Flea Market
15	Indore	Historic Site	Café	Hotel Bar	Donut Shop	Flea Market	Field	Fast Food Restaurant
16	Vadodara	Ice Cream Shop	Lake	Market	Clothing Store	Hotel	Dumpling Restaurant	Flea Market

Cluster 3 (Violet cities)

Finally, on this cluster it is mainly ATMS & Women Stores.

	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
41	Solāpur	Business Service	ATM	Women's Store	Food Court	Food	Flea Market	Field
46	Gwalior	ATM	Women's Store	Electronics Store	Food Court	Food	Flea Market	Field
67	Gopālpur	ATM	Hotel	Women's Store	Dumpling Restaurant	Food	Flea Market	Field
79	Hisar	ATM	Train Station	Health & Beauty Service	Women's Store	Dumpling Restaurant	Food	Flea Market
81	Tirunelveli	ATM	Women's Store	Electronics Store	Food Court	Food	Flea Market	Field
89	Bhātpāra	Platform	ATM	Electronics Store	Food Court	Food	Flea Market	Field
95	Gaya	ATM	Fried Chicken Joint	Women's Store	Electronics Store	Food	Flea Market	Field
102	Parbhani	ATM	Women's Store	Electronics Store	Food Court	Food	Flea Market	Field
108	Muzaffarpur	Health & Beauty Service	ATM	Electronics Store	Dumpling Restaurant	Food	Flea Market	Field

CONCLUSIONS

The question is:

Where Raj should open a Restaurant? Answering this question I would say that Raj should go to a tourist place like Mumbai, Chennai, Ahmedabad or one of those because it is inside the Cluster 0 (Brown color) meaning that it is plenty of restaurants (Which means that they really work there) And also those cities because is where there is more population. Also, I would suggest to Raj to go to a city plenty of Supermarkets or markets like in the last cluster because probably there he can finds better quality food, and way cheaper.

FUTURE IMPROVEMENTS.

- Fetching Data about rent prices for each city and compare them
- Fetching Data about benefits of different restaurants in each city
- Fetching Data about cost of living in each city
- Analyze each city by their cities surrounding