PROJECT REPORT

By: P N M MRUDU

Analysis of Populous Cities of India to find cities with similar restaurant culture

Introduction/Problem Statement Section

Indian Restaurant Industry has been one of the fastest-growing industries in the country. The market size of the Indian restaurants and food services industry stands at Rs 3.7 trillion as of 2018 registering a year-on-year growth of about 10% and a CAGR growth of 8.4% between 2013 and 2018(Information source: http://www.careratings.com/upload/NewsFiles/Studies/Restaurants%20%20QSRs%20May%202019.pdf).

There are many things one has to consider and plan accordingly when venturing into the Restaurant Business. Some of the important steps involved are deciding the concept of your restaurant, getting investors to fund your restaurant business, evaluating restaurant costs involved, and deciding the location of your restaurant.

Selecting a suitable city for a particular type of restaurant or expansion of an existing restaurant group is a crucial step. The existing restaurant culture of that city influences the acceptance of new restaurants, the inflow of customers, and local competition. Although the restaurant business is growing in India, eating-out culture is not so popular in villages and less populous cities.

So, through this project, we are aiming to find a solution to the **problem** – "which cities of India are suitable for establishing a particular type of chain of restaurants or for expansion of an already existing restaurant group to other cities and restaurant categories."

Solution to this problem is benefited for Businesses thinking to start a restaurant chain or Businesses trying to expand their restaurant business to other potential cities.

Data Section

Two Datasets have been utilized for this project. Densely populated cities of India have a broader scope for Restaurant business, for this purpose, a list of 45 most populous cities of India has been collected from Wikipedia.

(Source link: https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population).

For further analysis of these cities on the basis of restaurants and common restaurant categories operating in these cities, a JASON file containing the details for each city has been extracted using the Foursquare API.

Combining these two datasets, we get information about the number of restaurants under different venue categories and the most common venues in these areas. Using unsupervised Data Science algorithms, we can group similar cities, this will help us determine which cities are similar on the basis of restaurant culture and therefore, helps us choose a city for expansion of existing business or for the establishment of a chain of restaurant groups.

Methodology Section

For populous cities data, Beautifulsoup and geopy packages have been utilized.

Using BeautifulSoup populous cities data has been extracted from the Wikipedia page and external links in the city's column have been removed. For this project, only the top 45 cities list has been extracted from the cleaned data. Using geopy package latitude and longitude values of each of the 45 cities has been collected and appended to the previous dataset. Latitude and longitude values of the city are useful for passing them to the Foursquare API for venue details extraction.

Using Foursquare API all the restaurant venue details have been collected for each city, with restaurant limit of "100" and radius 60000 meters. The resulting JASON file has been converted to pandas DataFrame for further analysis. Using this data all the venues from each city has been grouped together by the City and mean of the data has been calculated. For the final dataset, each city along with 20 most common venue categories has been curated.

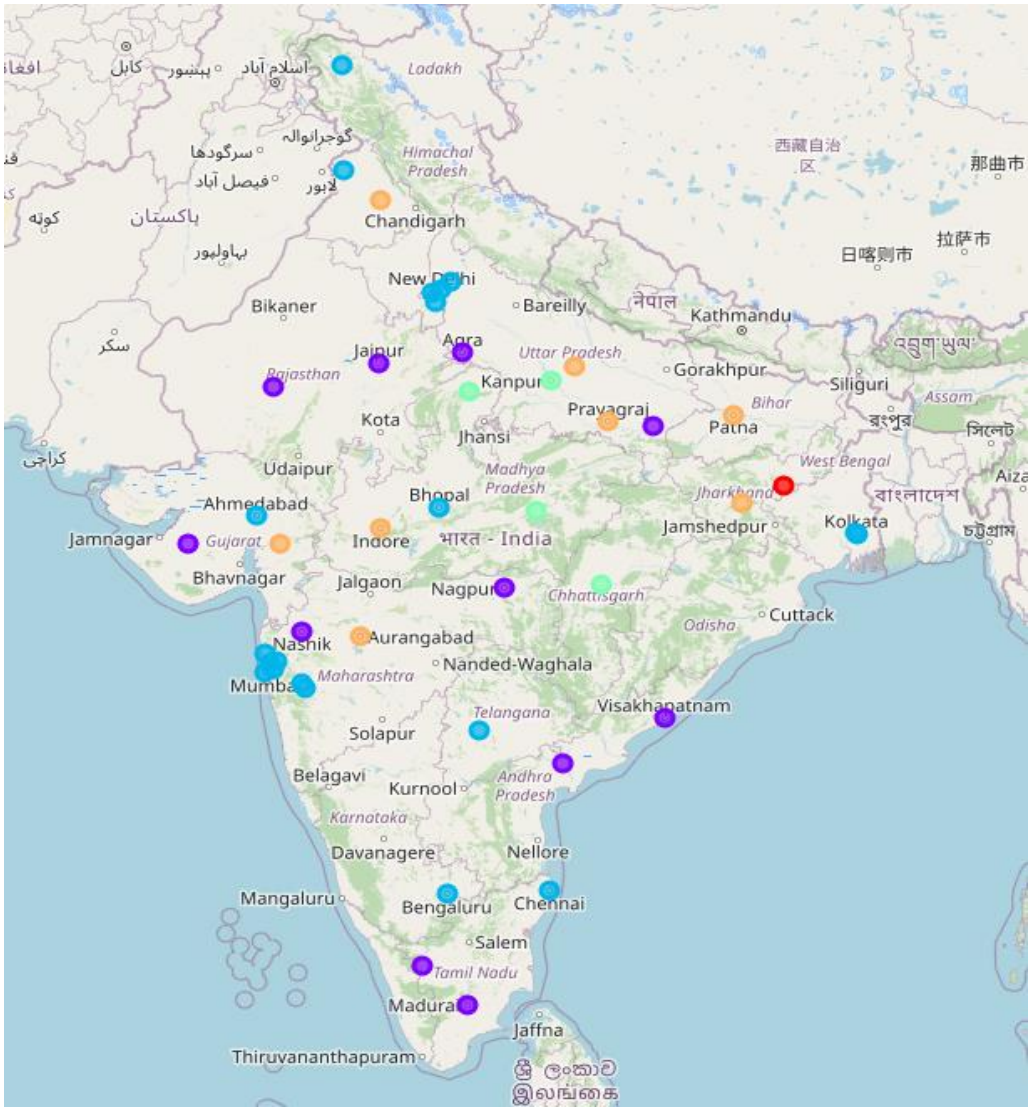
Unsupervised algorithm clustering has been chosen for this project as it will aid in finding similar cities without the help of labelled data.

To cluster similar cities, the k-means algorithm has been run on the final dataset with k value 5. For better understanding, the clusters have been visualized on the map.

Results Section

Running k means algorithm on the final dataset has resulted with cities grouped into 5 clusters.

The cluster data has been mapped on an India map for better visualization of the resulted clustering.



Number of cities in each cluster:

Cluster Number	Number of cities
Cluster 1	1
Cluster 2	11
Cluster 3	21
Cluster 4	4
Cluster 5	8

List of Cluster 1 Cities:

	latitude	City	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
32	23.795281	Dhanbad	23.795281	86.430964	0	Pizza Place	Indian Restaurant	Asian Restaurant

List of Cluster 2 Cities:

	latitude	City	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	26.916194	Jaipur	26.916194	75.820349	1	Café	Indian Restaurant	Bakery
1	21.149813	Nagpur	21.149813	79.082056	1	Indian Restaurant	Fast Food Restaurant	Sandwich Place
2	17.723128	Visakhapatnam	17.723128	83.301284	1	Indian Restaurant	Café	Restaurant
3	27.175255	Agra	27.175255	78.009816	1	Indian Restaurant	Café	Fast Food Restaurant
4	20.011247	Nashik	20.011247	73.790236	1	Indian Restaurant	Café	Pizza Place
5	22.305199	Rajkot	22.305199	70.802834	1	Indian Restaurant	Fast Food Restaurant	Pizza Place
6	25.335649	Varanasi	25.335649	83.007629	1	Indian Restaurant	Café	Pizza Place
7	11.001812	Coimbatore	11.001812	76.962842	1	Indian Restaurant	Vegetarian / Vegan Restaurant	Café
8	16.508759	Vijayawada	16.508759	80.618510	1	Indian Restaurant	Fast Food Restaurant	Café
9	26.296772	Jodhpur	26.296772	73.035143	1	Indian Restaurant	Restaurant	Café
10	9.926115	Madurai	9.926115	78.114098	1	Indian Restaurant	Café	Restaurant

List of Cluster 3 Cities:

	latitude	City	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	18.938771	Mumbai	18.938771	72.835335	2	Indian Restaurant	Restaurant	Café
1	28.651718	Delhi	28.651718	77.221939	2	Indian Restaurant	Café	Italian Restaurant
2	12.979120	Bangalore	12.979120	77.591300	2	Indian Restaurant	Café	Burger Joint
3	17.388786	Hyderabad	17.388786	78.461065	2	Indian Restaurant	Bakery	Café
4	23.021624	Ahmedabad	23.021624	72.579707	2	Indian Restaurant	Café	Fast Food Restaurant
5	13.080172	Chennai	13.080172	80.283833	2	Indian Restaurant	Café	Seafood Restaurant
6	22.545412	Kolkata	22.545412	88.356775	2	Chinese Restaurant	Indian Restaurant	Café
7	45.938300	Surat	45.938300	3.255300	2	French Restaurant	Restaurant	Italian Restaurant
8	18.521428	Pune	18.521428	73.854454	2	Café	Indian Restaurant	Fast Food Restaurant
9	19.194329	Thane	19.194329	72.970178	2	Indian Restaurant	Restaurant	Seafood Restaurant
10	23.253092	Bhopal	23.253092	77.396272	2	Indian Restaurant	Pizza Place	Bakery
11	18.627929	Pimpri-Chinchwad	18.627929	73.800983	2	Café	Fast Food Restaurant	Indian Restaurant
12	28.711241	Ghaziabad	28.711241	77.444537	2	Indian Restaurant	Café	Italian Restaurant
13	28.402837	Faridabad	28.402837	77.308563	2	Indian Restaurant	Café	Italian Restaurant
14	28.916667	Meerut	28.916667	77.683333	2	Indian Restaurant	Café	Fast Food Restaurant
15	19.233487	Kalyan-Dombivli	19.233487	73.118246	2	Indian Restaurant	Restaurant	Café
16	19.425879	Vasai-Virar	19.425879	72.822490	2	Indian Restaurant	Restaurant	Seafood Restaurant
17	34.074744	Srinagar	34.074744	74.820444	2	Café	Indian Restaurant	Fried Chicken Joint
18	31.634308	Amritsar	31.634308	74.873679	2	Café	Pakistani Restaurant	Fast Food Restaurant
19	19.030826	Navi Mumbai	19.030826	73.019854	2	Indian Restaurant	Fast Food Restaurant	Restaurant
20	22.588222	Howrah	22.588222	88.323139	2	Indian Restaurant	Chinese Restaurant	Café

List of Cluster 4 Cities:

	latitude	City	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	26.460914	Kanpur	26.460914	80.321759	3	Fast Food Restaurant	Café	Pizza Place
1	26.203725	Gwalior	26.203725	78.157363	3	Fast Food Restaurant	Café	Pizza Place
2	23.160894	Jabalpur	23.160894	79.949770	3	Café	Chinese Restaurant	Asian Restaurant
3	21.237947	Raipur	21.237947	81.633683	3	Café	Pizza Place	Fast Food Restaurant

List of Cluster 5 Cities:

	latitude	City	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	26.838100	Lucknow	26.838100	80.934600	4	Indian Restaurant	Fast Food Restaurant	Café
1	22.720362	Indore	22.720362	75.868200	4	Indian Restaurant	Fast Food Restaurant	Café
2	25.609324	Patna	25.609324	85.123525	4	Café	Pizza Place	Indian Restaurant
3	22.297314	Vadodara	22.297314	73.194257	4	Indian Restaurant	Café	Fast Food Restaurant
4	30.909016	Ludhiana	30.909016	75.851601	4	Fast Food Restaurant	Indian Restaurant	Café
5	19.877263	Aurangabad	19.877263	75.339024	4	Café	Indian Restaurant	Pizza Place
6	25.438130	Allahabad	25.438130	81.833800	4	Fast Food Restaurant	Pizza Place	Indian Restaurant
7	23.370035	Ranchi	23.370035	85.325013	4	Indian Restaurant	Café	Pizza Place

Discussion Section

Cluster 1 data (01 City):

from this data we can observe that Dhanbad is the only City among all the 45 cities with “Pizza Place” as most common venue.

Cluster 2 data (11 Cities):

Most common venue in this category of cities is Indian restaurant.

Cluster 3 data (21 Cities):

Cities in this category are mostly populated with Indian Restaurants and Cafes.

Cluster 4 data (4 Cities):

Cities in this category are mostly populated with Fast Food Restaurants and Cafes.

Cluster 5 data (08 Cities):

Cities in this category are mostly populated with Indian Restaurants, Fast Food Restaurants and Cafes.

From all the results, we can observe that Indian Restaurants, Cafes and Fast Food Restaurants are the most common Venues in India. Venturing into any one of these or a combination of these venue categories is advisable for starting a new Restaurant chain.

Existing Restaurants who want to expand their business to other cities can look up the cluster data:

(i) Restaurants already in Indian restaurant category can expand to cities in cluster 2 or to cluster 3 cities if they want to start Café venue category along with Indian restaurants category.

(ii) Pizza seems to be very common venue in Dhanbad, if any pizza restaurant which doesn't have a branch in Dhanbad can definitely try to start one.

(iii) From the Data we can observe that Chinese, seafood and French restaurants are not so common in India.

(iv) Restaurants with a combination of Fast food and café venue category can target cluster 4 cities.

Conclusion Section

In this project, two datasets- populous cities of India and their common restaurant venue categories have been combined to identify similarity between cities in terms of their restaurants. For identifying similar groups among the cities K-means algorithm has been used. From the resulting clusters we could observe that Indian restaurants, cafes and Fast food restaurants are the three most common venue categories in India. There is only one city Dhanbad with pizza as the 1st most common venue in the city, there are 11 cities out of 45 with Indian restaurants occupying their majority share of restaurants. All this data and insights could be utilized by both new and existing restaurant chains. For example, Indian restaurant category and cafés appears to be the common restaurant culture in India, therefore, venturing into any one of these venue categories reduces the possibility of poor in-flow of customers.