Coursera Capstone

IBM Applied Data Science Capstone

Opening a cricket stadium in India

By: Mihir M Kestur

Introduction

Cricket is a sport that's watched and cherished by many in the world. The craze for the sport, in the Indian subcontinent, is arguably unparalleled. The country hosts a variety of international tournaments ranging from the shortest format of the game to the longest 5-day test match series'. It also hosts a number of domestic tournaments, amongst which the Indian premier league (IPL) is a world-class, highly anticipated, annual festival. As per https://en.wikipedia.org/wiki/Indian_Premier_League, the Indian cricket industry is easily a billion dollar industry thus having a major positive impact on the economy. Therefore it is easy to realize the importance of having sufficient cricket stadiums to cater to the demand the sport creates in the country.

Business problem

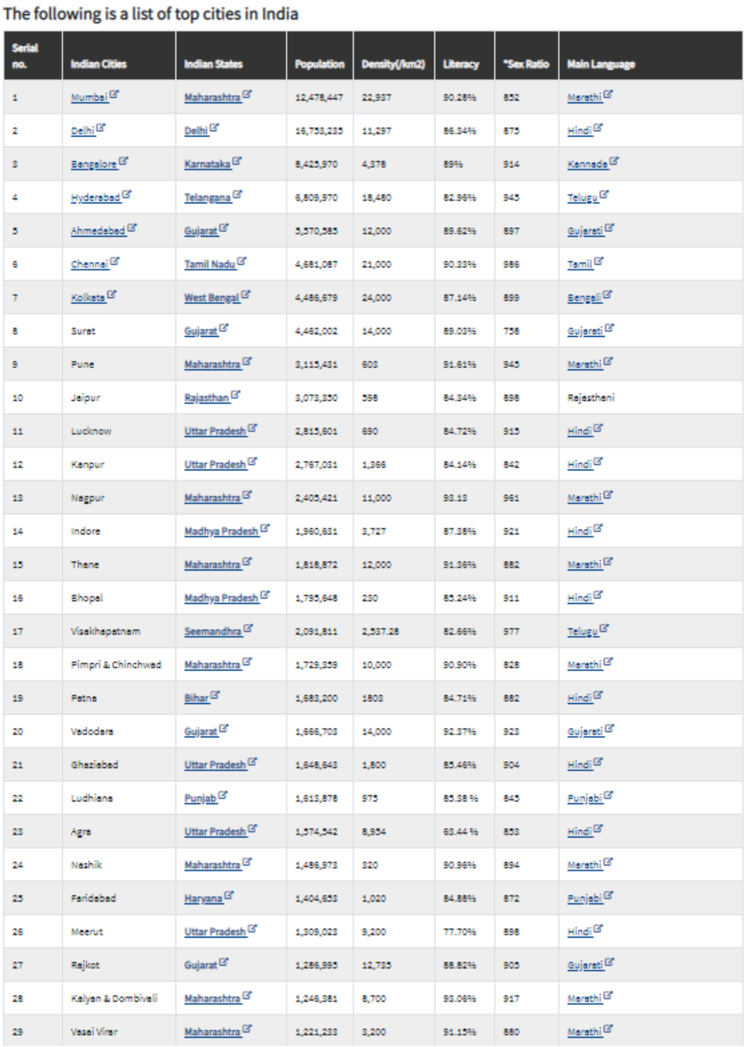
The objective of this capstone project is to analyse and select, using data science methodologies and machine learning techniques like clustering, the best location to open a new cricket stadium in the country of India.

Data Requirements

Some parameters that affect the selection of the city are proximity/existence of an airport, restaurants and other attractive venues for the players and spectators. It is also important to exclude the cities in which the cricket stadiums already exist.

- List of top 100 cities in the country web scrapped from https://www.nriol.com/india-statistics/biggest-cities-india.asp
- Cities in which cricket stadiums already existing scrapped from https://en.wikipedia.org/wiki/List_of_international_cricket_grounds_in_India
- Airport in the city scrapped from
  1. https://www.mapsofindia.com/air-network/international-airport-map.htm
  2. https://www.mapsofindia.com/air-network/domestic-airport-map.htm
- Restaurants and other eateries in the city obtained from foursquare API.

Methodology



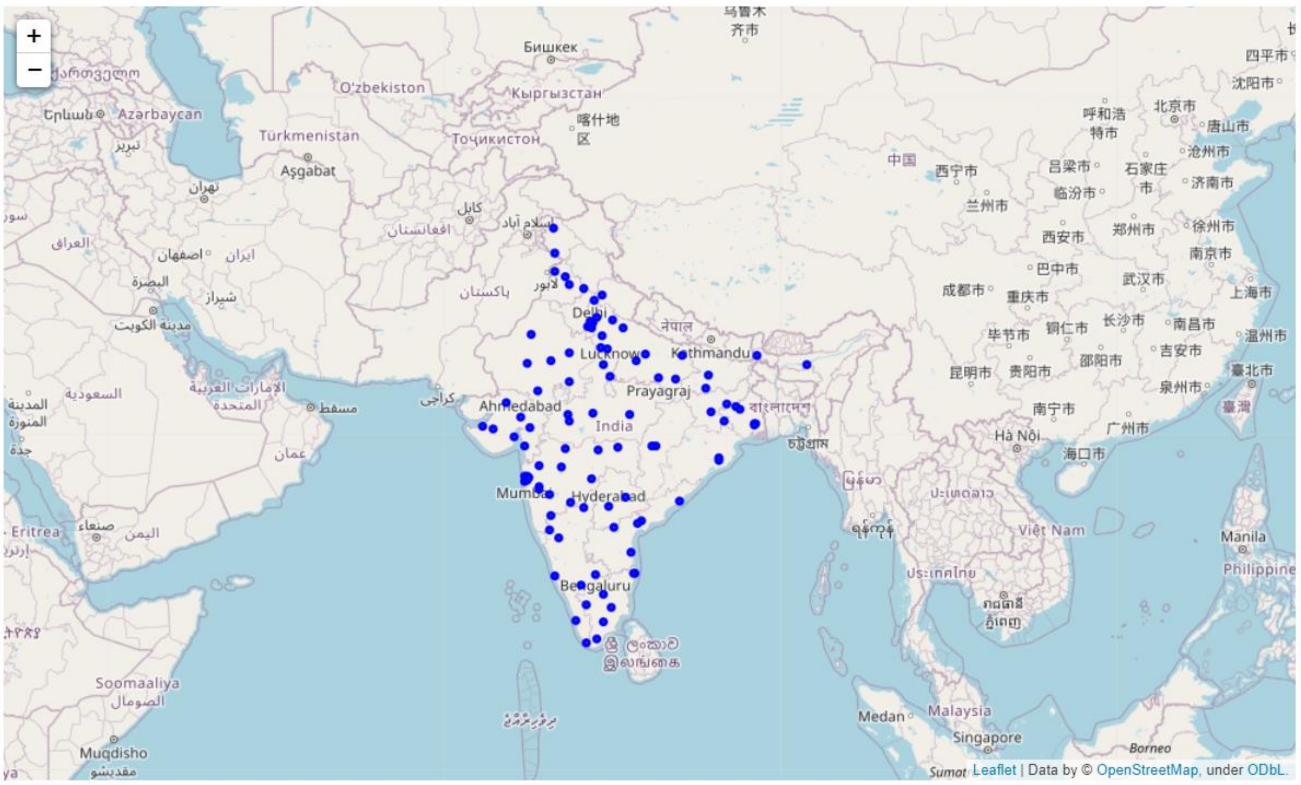
The following is a list of top cities in India

Serial no.	Indian Cities	Indian States	Population	Density(/km2)	Literacy	*Sex Ratio	Main Language
1	Mumbai	Maharashtra	12,478,447	22,937	90.28%	852	Marathi
2	Delhi	Delhi	16,753,235	11,297	86.34%	875	Hindi
3	Bangalore	Karnataka	8,425,970	4,378	89%	914	Kannada
4	Hyderabad	Telangana	6,809,970	18,480	82.96%	945	Telugu
5	Ahmedabad	Gujarat	5,570,585	12,000	89.62%	897	Gujarati
6	Chennai	Tamil Nadu	4,681,087	21,000	90.33%	986	Tamil
7	Kolkata	West Bengal	4,486,679	24,000	87.14%	899	Bengali
8	Surat	Gujarat	4,462,002	14,000	89.03%	758	Gujarati
9	Pune	Maharashtra	3,115,431	603	91.61%	945	Marathi
10	Jaipur	Rajasthan	3,073,350	598	84.34%	898	Rajasthani
11	Lucknow	Uttar Pradesh	2,815,601	690	84.72%	915	Hindi
12	Kanpur	Uttar Pradesh	2,767,031	1,366	84.14%	842	Hindi
13	Nagpur	Maharashtra	2,405,421	11,000	93.13	961	Marathi
14	Indore	Madhya Pradesh	1,960,631	3,727	87.38%	921	Hindi
15	Thane	Maharashtra	1,818,872	12,000	91.36%	882	Marathi
16	Bhopal	Madhya Pradesh	1,795,648	230	85.24%	911	Hindi
17	Visakhapatnam	Seemandhra	2,091,811	2,537.28	82.66%	977	Telugu
18	Pimpri & Chinchwad	Maharashtra	1,729,359	10,000	90.90%	828	Marathi
19	Patna	Bihar	1,683,200	1803	84.71%	882	Hindi
20	Vadodara	Gujarat	1,666,703	14,000	92.37%	923	Gujarati
21	Ghaziabad	Uttar Pradesh	1,648,643	1,800	85.46%	904	Hindi
22	Ludhiana	Punjab	1,613,878	975	85.38 %	845	Punjabi
23	Agra	Uttar Pradesh	1,574,542	8,954	63.44 %	853	Hindi
24	Nashik	Maharashtra	1,486,973	320	90.96%	894	Marathi
25	Faridabad	Haryana	1,404,653	1,020	84.88%	872	Punjabi
26	Meerut	Uttar Pradesh	1,309,023	9,200	77.70%	898	Hindi
27	Rajkot	Gujarat	1,286,995	12,735	88.82%	905	Gujarati
28	Kalyan & Dombivali	Maharashtra	1,246,381	8,700	93.06%	917	Marathi
29	Vasai Virar	Maharashtra	1,221,233	3,200	91.15%	880	Marathi

To obtain cities that are good candidates in which a new stadium can be opened, top 100 cities in the country are scrapped from https://www.nriol.com/india-statistics/biggest-cities-india.asp. It is done by using the BeautifulSoup library.

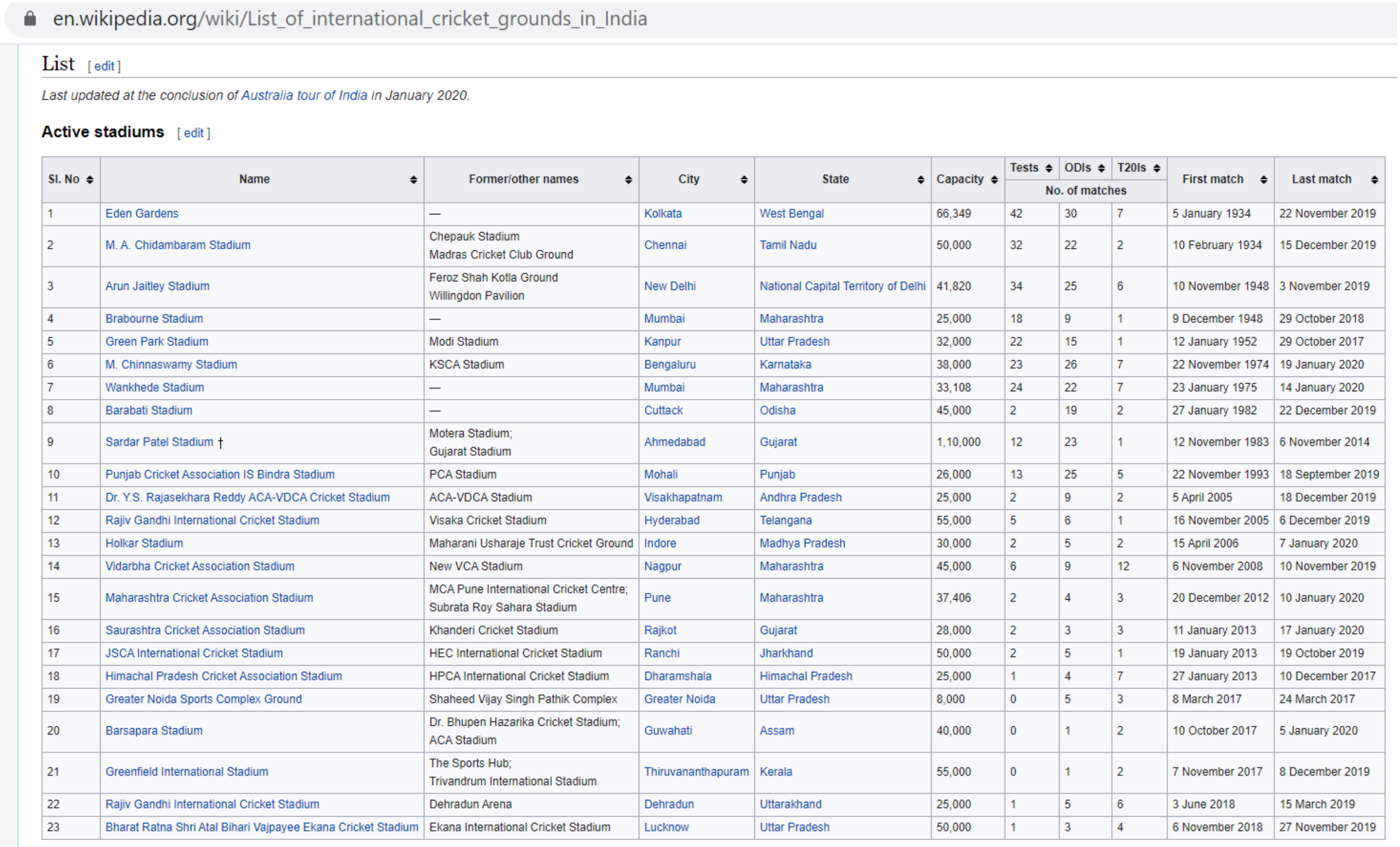
It is then plotted on a map to visualize the data. This is done by using the folium library that provides responsive maps with many other features.

Out[10]:



Cities in which cricket stadiums already exist is obtained by web craping a Wikipedia page https://en.wikipedia.org/wiki/List_of_international_cricket_grounds_in_India

This is done to filter the potential cities and avoid a city to have two fully functional cricket stadiums.



en.wikipedia.org/wiki/List_of_international_cricket_grounds_in_India

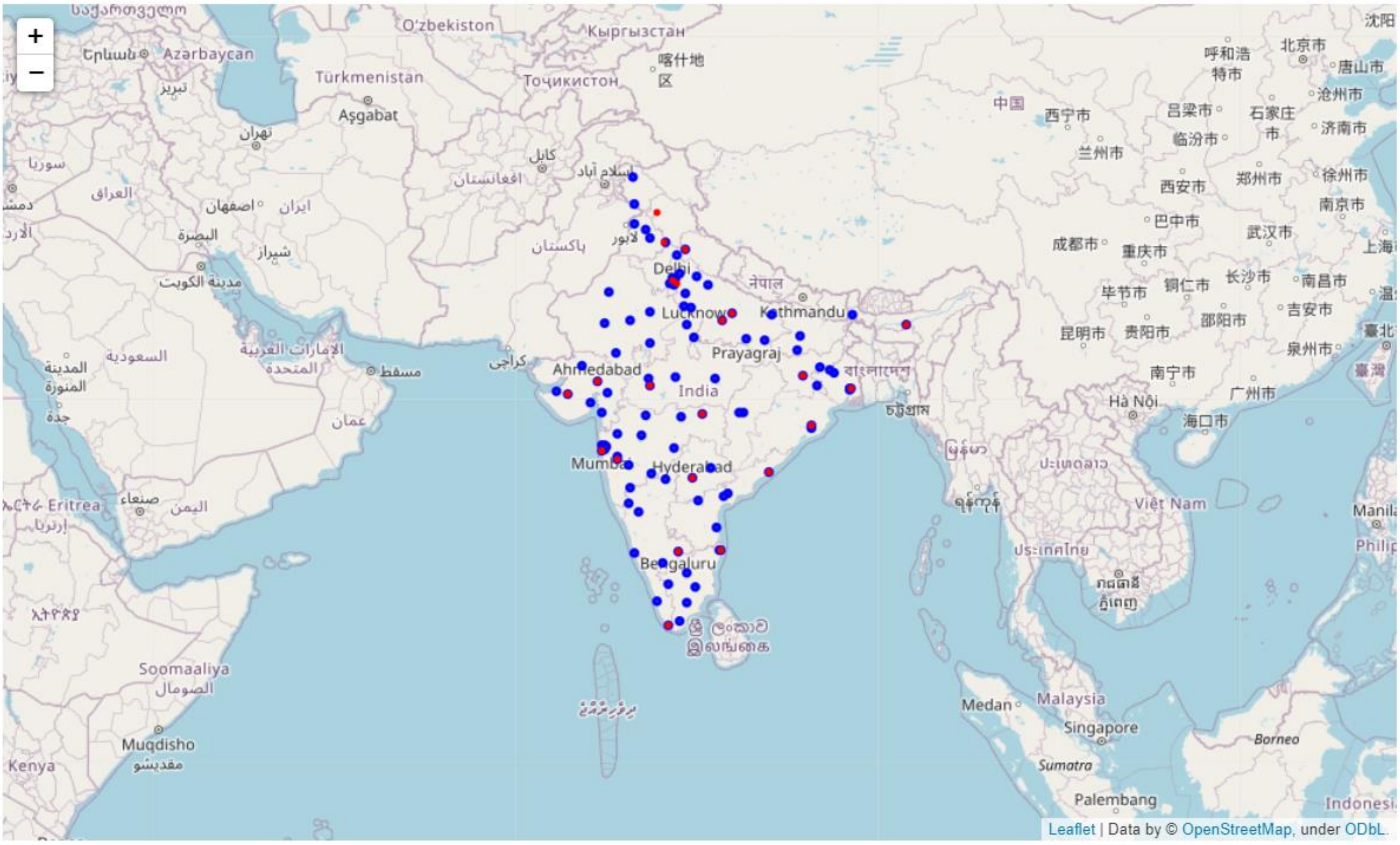
List [edit]

Last updated at the conclusion of Australia tour of India in January 2020.

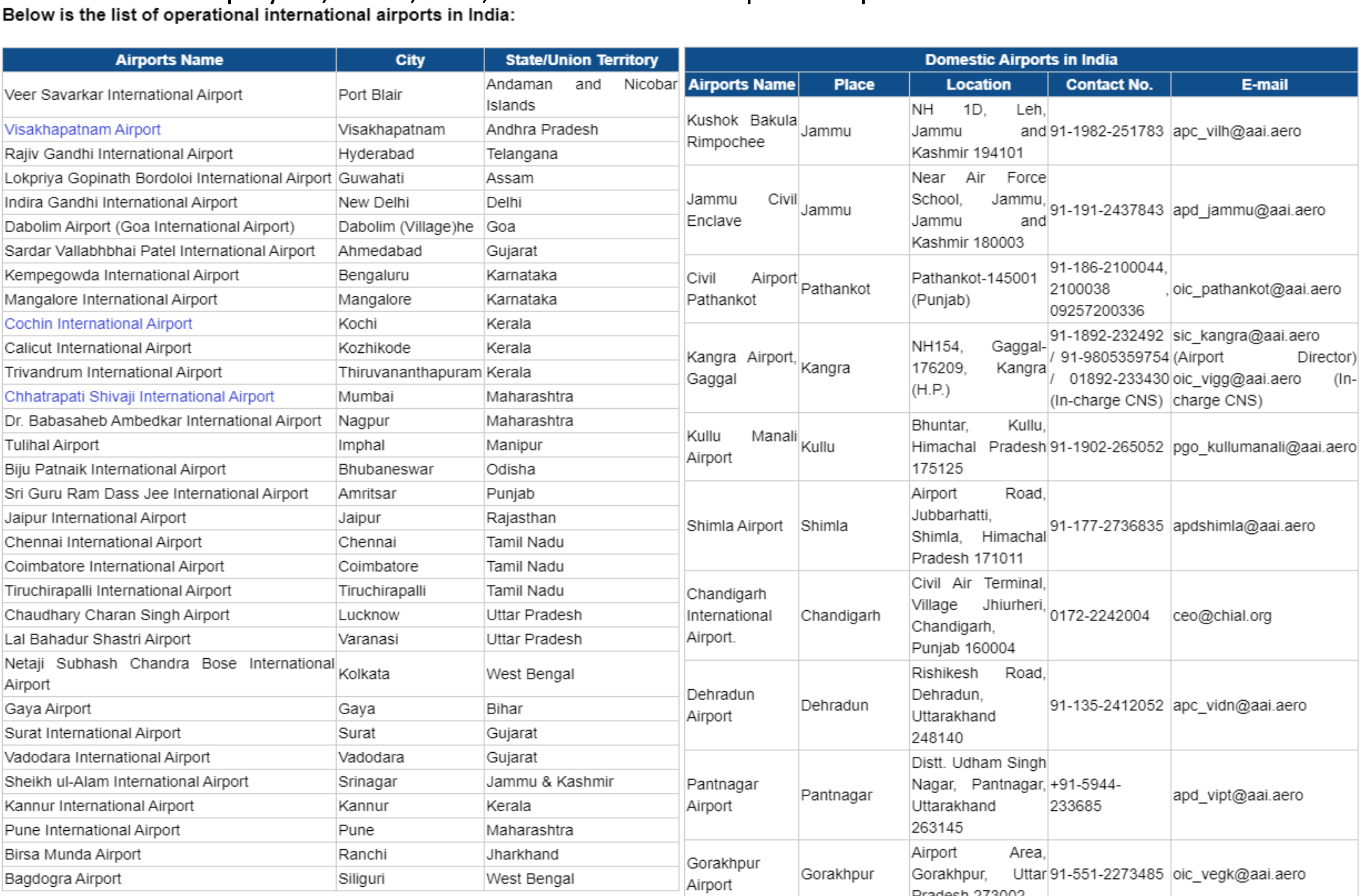
Active stadiums [edit]

Sl. No	Name	Former/other names	City	State	Capacity	Tests	ODIs	T20Is	First match	Last match
						No. of matches				
1	Eden Gardens	—	Kolkata	West Bengal	66,349	42	30	7	5 January 1934	22 November 2019
2	M. A. Chidambaram Stadium	Chepauk Stadium Madras Cricket Club Ground	Chennai	Tamil Nadu	50,000	32	22	2	10 February 1934	15 December 2019
3	Arun Jaitley Stadium	Feroz Shah Kotla Ground Willingdon Pavilion	New Delhi	National Capital Territory of Delhi	41,820	34	25	6	10 November 1948	3 November 2019
4	Brabourne Stadium	—	Mumbai	Maharashtra	25,000	18	9	1	9 December 1948	29 October 2018
5	Green Park Stadium	Modi Stadium	Kanpur	Uttar Pradesh	32,000	22	15	1	12 January 1952	29 October 2017
6	M. Chinnaswamy Stadium	KSCA Stadium	Bengaluru	Karnataka	38,000	23	26	7	22 November 1974	19 January 2020
7	Wankhede Stadium	—	Mumbai	Maharashtra	33,108	24	22	7	23 January 1975	14 January 2020
8	Barabati Stadium	—	Cuttack	Odisha	45,000	2	19	2	27 January 1982	22 December 2019
9	Sardar Patel Stadium †	Motera Stadium; Gujarat Stadium	Ahmedabad	Gujarat	1,10,000	12	23	1	12 November 1983	6 November 2014
10	Punjab Cricket Association IS Bindra Stadium	PCA Stadium	Mohali	Punjab	26,000	13	25	5	22 November 1993	18 September 2019
11	Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium	ACA-VDCA Stadium	Visakhapatnam	Andhra Pradesh	25,000	2	9	2	5 April 2005	18 December 2019
12	Rajiv Gandhi International Cricket Stadium	Visaka Cricket Stadium	Hyderabad	Telangana	55,000	5	6	1	16 November 2005	6 December 2019
13	Holkar Stadium	Maharani Usharaje Trust Cricket Ground	Indore	Madhya Pradesh	30,000	2	5	2	15 April 2006	7 January 2020
14	Vidarbha Cricket Association Stadium	New VCA Stadium	Nagpur	Maharashtra	45,000	6	9	12	6 November 2008	10 November 2019
15	Maharashtra Cricket Association Stadium	MCA Pune International Cricket Centre; Subrata Roy Sahara Stadium	Pune	Maharashtra	37,406	2	4	3	20 December 2012	10 January 2020
16	Saurashtra Cricket Association Stadium	Khanderi Cricket Stadium	Rajkot	Gujarat	28,000	2	3	3	11 January 2013	17 January 2020
17	JSCA International Cricket Stadium	HEC International Cricket Stadium	Ranchi	Jharkhand	50,000	2	5	1	19 January 2013	19 October 2019
18	Himachal Pradesh Cricket Association Stadium	HPCA International Cricket Stadium	Dharamshala	Himachal Pradesh	25,000	1	4	7	27 January 2013	10 December 2017
19	Greater Noida Sports Complex Ground	Shaheed Vijay Singh Pathik Complex	Greater Noida	Uttar Pradesh	8,000	0	5	3	8 March 2017	24 March 2017
20	Barsapara Stadium	Dr. Bhupen Hazarika Cricket Stadium; ACA Stadium	Guwahati	Assam	40,000	0	1	2	10 October 2017	5 January 2020
21	Greenfield International Stadium	The Sports Hub; Trivandrum International Stadium	Thiruvananthapuram	Kerala	55,000	0	1	2	7 November 2017	8 December 2019
22	Rajiv Gandhi International Cricket Stadium	Dehradun Arena	Dehradun	Uttarakhand	25,000	1	5	6	3 June 2018	15 March 2019
23	Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium	Ekana International Cricket Stadium	Lucknow	Uttar Pradesh	50,000	1	3	4	6 November 2018	27 November 2019

The cities are plotted on the map to visualize. Red points are cities with existing cricket stadiums.



Next, cities which have operational airports are obtained as this an essential infrastructure that allows players, staffs, fans, etc. to move from place to place.

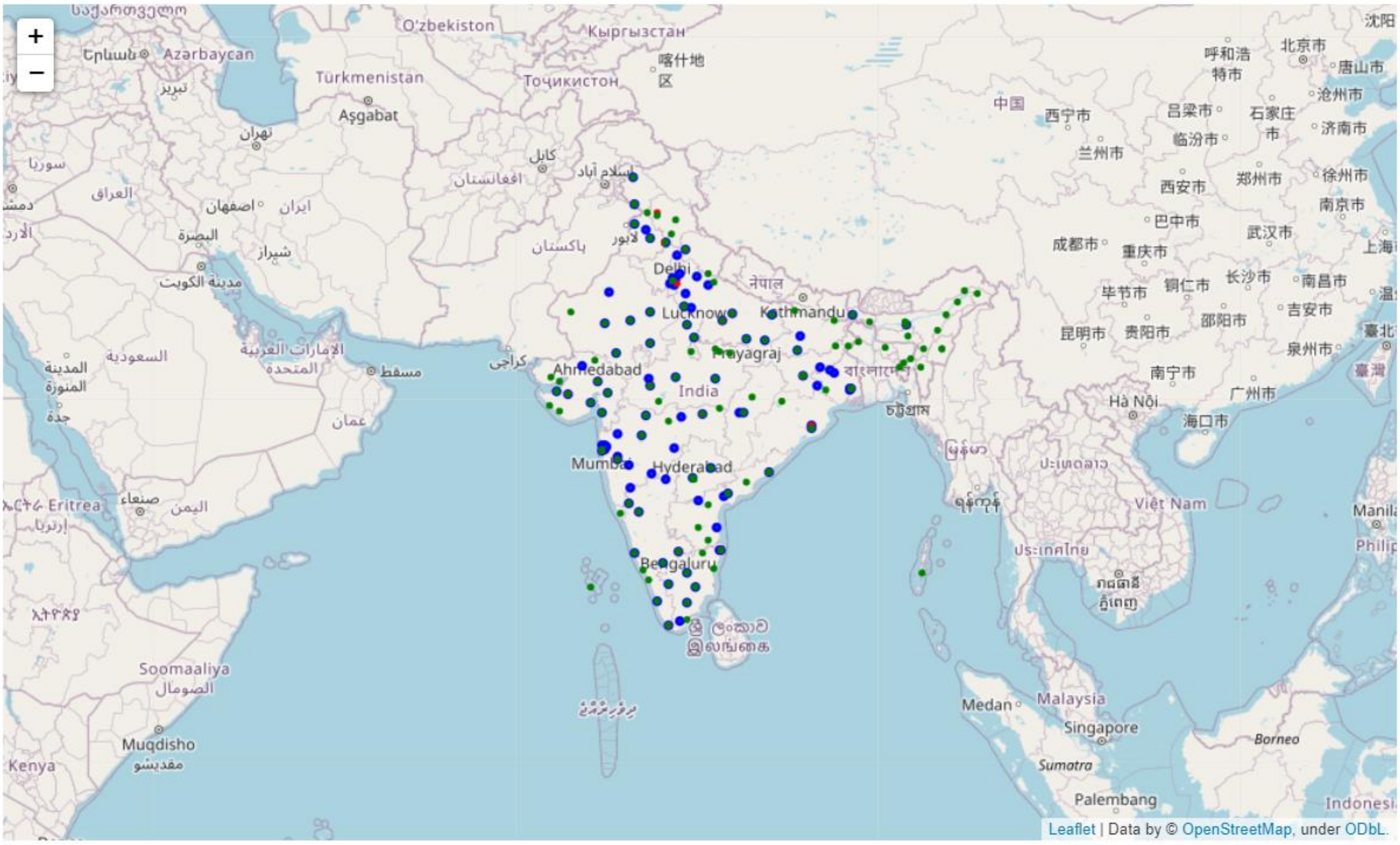


Below is the list of operational international airports in India:

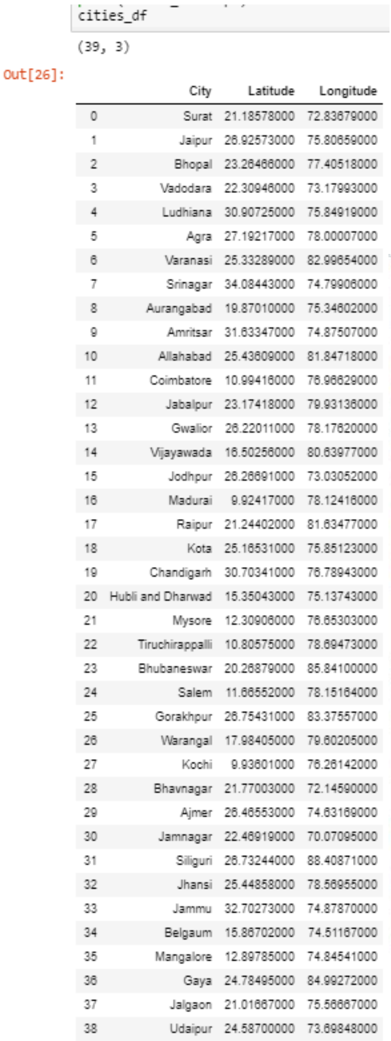
Airports Name	City	State/Union Territory
Veer Savarkar International Airport	Port Blair	Andaman and Nicobar Islands
Visakhapatnam Airport	Visakhapatnam	Andhra Pradesh
Rajiv Gandhi International Airport	Hyderabad	Telangana
Lokpriya Gopinath Bordoloi International Airport	Guwahati	Assam
Indira Gandhi International Airport	New Delhi	Delhi
Dabolim Airport (Goa International Airport)	Dabolim (Village)he	Goa
Sardar Vallabhbhai Patel International Airport	Ahmedabad	Gujarat
Kempegowda International Airport	Bengaluru	Karnataka
Mangalore International Airport	Mangalore	Karnataka
Cochin International Airport	Kochi	Kerala
Calicut International Airport	Kozhikode	Kerala
Trivandrum International Airport	Thiruvananthapuram	Kerala
Chhatrapati Shivaji International Airport	Mumbai	Maharashtra
Dr. Babasaheb Ambedkar International Airport	Nagpur	Maharashtra
Tulihal Airport	Imphal	Manipur
Biju Patnaik International Airport	Bhubaneswar	Odisha
Sri Guru Ram Dass Jee International Airport	Amritsar	Punjab
Jaipur International Airport	Jaipur	Rajasthan
Chennai International Airport	Chennai	Tamil Nadu
Coimbatore International Airport	Coimbatore	Tamil Nadu
Tiruchirapalli International Airport	Tiruchirapalli	Tamil Nadu
Chaudhary Charan Singh Airport	Lucknow	Uttar Pradesh
Lal Bahadur Shastri Airport	Varanasi	Uttar Pradesh
Netaji Subhash Chandra Bose International Airport	Kolkata	West Bengal
Gaya Airport	Gaya	Bihar
Surat International Airport	Surat	Gujarat
Vadodara International Airport	Vadodara	Gujarat
Sheikh ul-Alam International Airport	Srinagar	Jammu & Kashmir
Kannur International Airport	Kannur	Kerala
Pune International Airport	Pune	Maharashtra
Birsa Munda Airport	Ranchi	Jharkhand
Bagdogra Airport	Siliguri	West Bengal

Domestic Airports in India				
Airports Name	Place	Location	Contact No.	E-mail
Kushok Bakula Rimpochee	Jammu	NH 1D, Leh, Jammu and Kashmir 194101	91-1982-251783	apc_vilh@aai.aero
Jammu Civil Enclave	Jammu	Near Air Force School, Jammu, Jammu and Kashmir 180003	91-191-2437843	apd_jammu@aai.aero
Civil Airport Pathankot	Pathankot	Pathankot-145001 (Punjab)	91-186-2100044, 2100038 , 09257200336	oic_pathankot@aai.aero
Kangra Airport, Gaggal	Kangra	NH154, Gaggal-176209, Kangra (H.P.)	91-1892-232492 / 91-9805359754 / 01892-233430 (In-charge CNS)	sic_kangra@aai.aero (Airport Director) oic_vigg@aai.aero (In-charge CNS)
Kullu Manali Airport	Kullu	Bhuntar, Kullu, Himachal Pradesh 175125	91-1902-265052	pgo_kullumanali@aai.aero
Shimla Airport	Shimla	Airport Road, Jubbarhatti, Shimla, Himachal Pradesh 171011	91-177-2736835	apdshimla@aai.aero
Chandigarh International Airport.	Chandigarh	Civil Air Terminal, Village Jhiurheri, Chandigarh, Punjab 160004	0172-2242004	ceo@chial.org
Dehradun Airport	Dehradun	Rishikesh Road, Dehradun, Uttarakhand 248140	91-135-2412052	apc_vidn@aai.aero
Pantnagar Airport	Pantnagar	Distt. Udham Singh Nagar, Pantnagar, Uttarakhand 263145	+91-5944-233685	apd_vipt@aai.aero
Gorakhpur Airport	Gorakhpur	Airport Area, Gorakhpur, Uttar Pradesh 273002	91-551-2273485	oic_vegk@aai.aero

The cities obtained are plotted on the map for visualization and Green markers depict the cities with airports.



This is followed by filtering out cities. Cities which do not have airports and cities in which cricket stadiums exist is removed from the main dataframe that consists of the 100 potential cities, to obtain a filtered dataframe that has 39 potential cities.

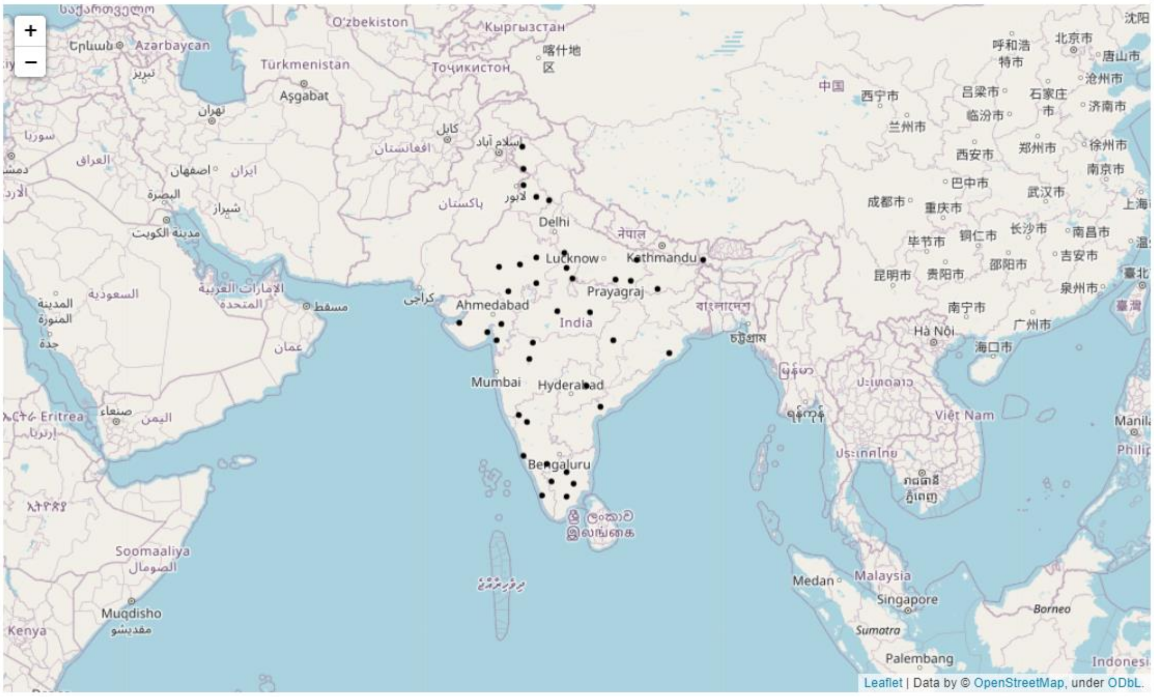


```
cities_df
```

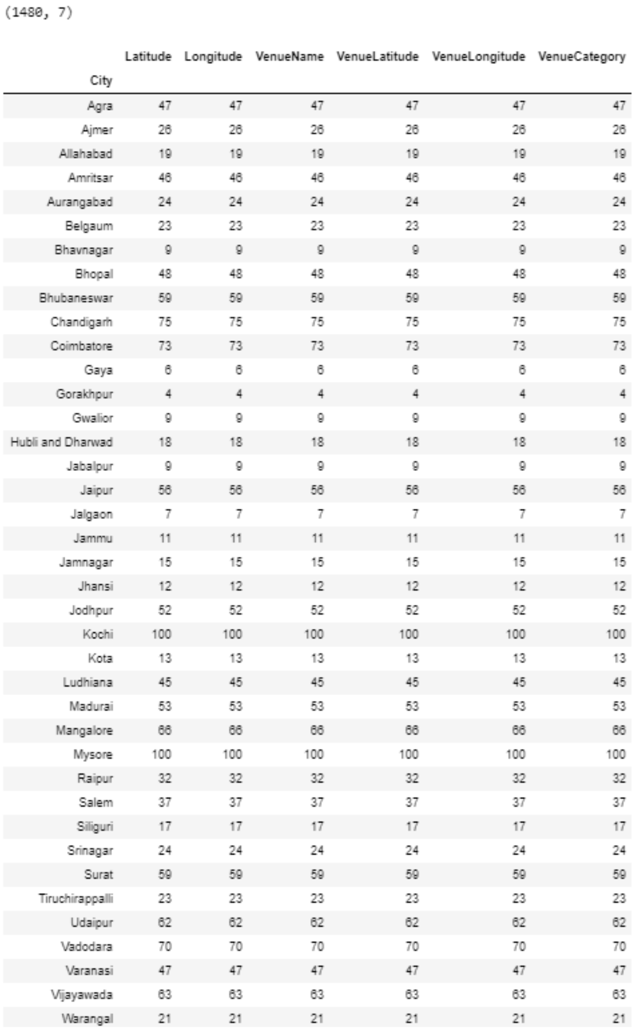
```
(39, 3)
Out[26]:
```

	City	Latitude	Longitude
0	Surat	21.18578000	72.83679000
1	Jaipur	26.92573000	75.80659000
2	Bhopal	23.26466000	77.40518000
3	Vadodara	22.30946000	73.17993000
4	Ludhiana	30.90725000	75.84919000
5	Agra	27.19217000	78.00007000
6	Varanasi	25.33289000	82.99654000
7	Srinagar	34.08443000	74.79906000
8	Aurangabad	19.87010000	75.34602000
9	Amritsar	31.63347000	74.87507000
10	Allahabad	25.43609000	81.84718000
11	Coimbatore	10.99416000	76.96629000
12	Jabalpur	23.17418000	79.93136000
13	Gwalior	26.22011000	78.17620000
14	Vijayawada	16.50256000	80.63977000
15	Jodhpur	26.26691000	73.03052000
16	Madurai	9.92417000	78.12416000
17	Raipur	21.24402000	81.63477000
18	Kota	25.16531000	75.85123000
19	Chandigarh	30.70341000	76.78943000
20	Hubli and Dharwad	15.35043000	75.13743000
21	Mysore	12.30906000	76.65303000
22	Tiruchirappalli	10.80575000	78.69473000
23	Bhubaneswar	20.26879000	85.84100000
24	Salem	11.66552000	78.15164000
25	Gorakhpur	26.75431000	83.37557000
26	Warangal	17.98405000	79.60205000
27	Kochi	9.93601000	76.26142000
28	Bhavnagar	21.77003000	72.14590000
29	Ajmer	26.46553000	74.63169000
30	Jamnagar	22.46919000	70.07095000
31	Siliguri	26.73244000	88.40871000
32	Jhansi	25.44858000	78.56955000
33	Jammu	32.70273000	74.87870000
34	Belgaum	15.86702000	74.51167000
35	Mangalore	12.89785000	74.84541000
36	Gaya	24.78495000	84.99272000
37	Jalgaon	21.01667000	75.56667000
38	Udaipur	24.58700000	73.69848000

It is then plotted on the map for visualization. Black points represent the filtered potential cities.



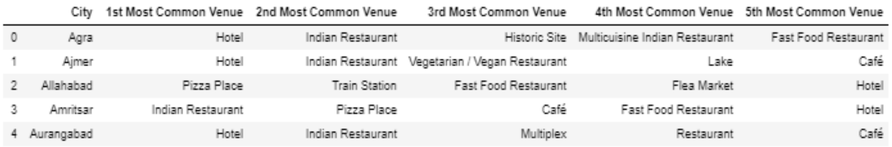
To further filter the cities, we obtain restuarants and other popular venues using the foursquare API for each city.



```
(1480, 7)
```

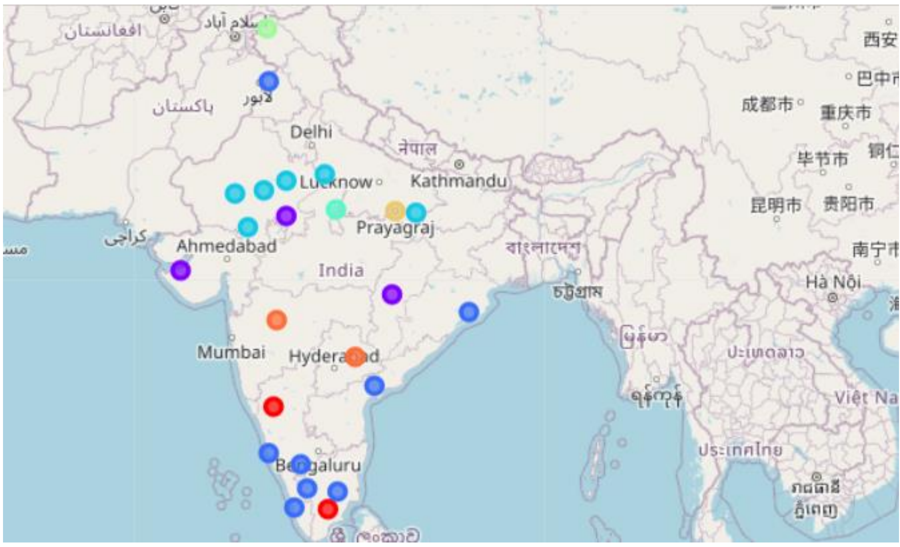
City	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Agra	47	47	47	47	47	47
Ajmer	26	26	26	26	26	26
Allahabad	19	19	19	19	19	19
Amritsar	46	46	46	46	46	46
Aurangabad	24	24	24	24	24	24
Belgaum	23	23	23	23	23	23
Bhavnagar	9	9	9	9	9	9
Bhopal	48	48	48	48	48	48
Bhubaneswar	59	59	59	59	59	59
Chandigarh	75	75	75	75	75	75
Coimbatore	73	73	73	73	73	73
Gaya	6	6	6	6	6	6
Gorakhpur	4	4	4	4	4	4
Gwalior	9	9	9	9	9	9
Hubli and Dharwad	18	18	18	18	18	18
Jabalpur	9	9	9	9	9	9
Jaipur	56	56	56	56	56	56
Jalgaon	7	7	7	7	7	7
Jammu	11	11	11	11	11	11
Jamnagar	15	15	15	15	15	15
Jhansi	12	12	12	12	12	12
Jodhpur	52	52	52	52	52	52
Kochi	100	100	100	100	100	100
Kota	13	13	13	13	13	13
Ludhiana	45	45	45	45	45	45
Madurai	53	53	53	53	53	53
Mangalore	66	66	66	66	66	66
Mysore	100	100	100	100	100	100
Raipur	32	32	32	32	32	32
Salem	37	37	37	37	37	37
Siliguri	17	17	17	17	17	17
Srinagar	24	24	24	24	24	24
Surat	59	59	59	59	59	59
Tiruchirappalli	23	23	23	23	23	23
Udaipur	62	62	62	62	62	62
Vadodara	70	70	70	70	70	70
Varanasi	47	47	47	47	47	47
Vijayawada	63	63	63	63	63	63
Warangal	21	21	21	21	21	21

This data is preprocessed via one hot encoding and and top 5 venues in each city is obtained.



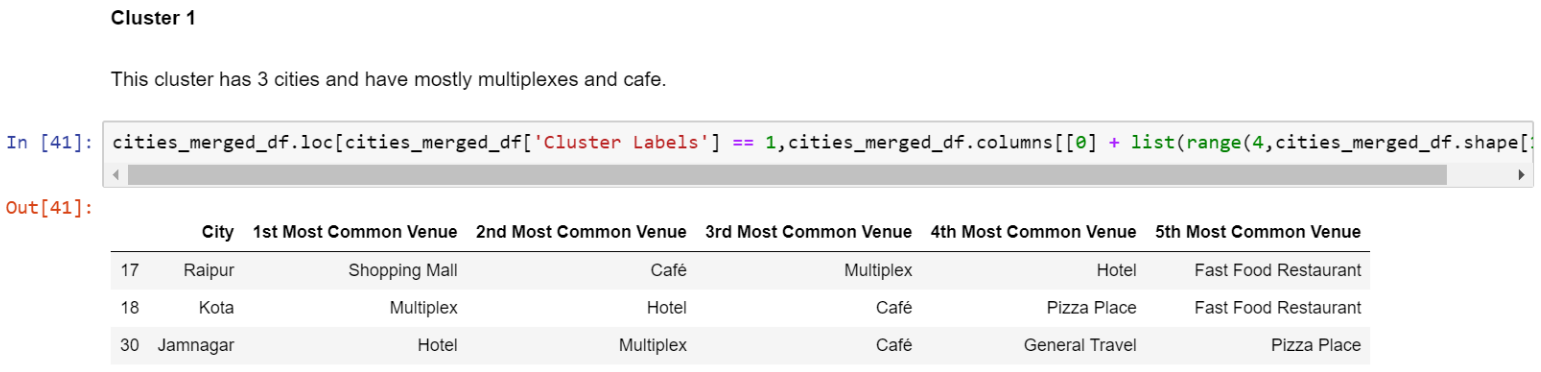
	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agra	Hotel	Indian Restaurant	Historic Site	Multicuisine Indian Restaurant	Fast Food Restaurant
1	Ajmer	Hotel	Indian Restaurant	Vegetarian / Vegan Restaurant	Lake	Café
2	Allahabad	Pizza Place	Train Station	Fast Food Restaurant	Flea Market	Hotel
3	Amritsar	Indian Restaurant	Pizza Place	Café	Fast Food Restaurant	Hotel
4	Aurangabad	Hotel	Indian Restaurant	Multiplex	Restaurant	Café

The data thus obtained is clusteres via K-means algorithm into 8 different clusters and is visualized on the map.



Observations:

The obtained clusters is now analyzed individually.



Cluster 1

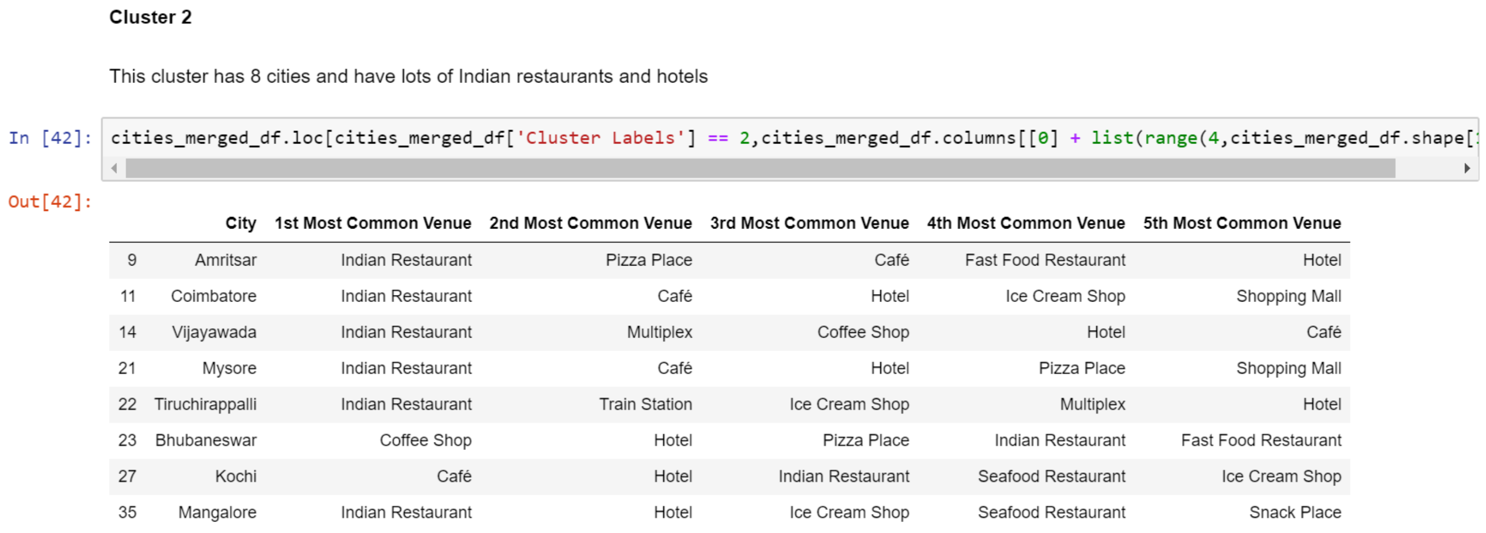
This cluster has 3 cities and have mostly multiplexes and cafe.

```
In [41]: cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 1,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[41]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	Raipur	Shopping Mall	Café	Multiplex	Hotel	Fast Food Restaurant
18	Kota	Multiplex	Hotel	Café	Pizza Place	Fast Food Restaurant
30	Jamnagar	Hotel	Multiplex	Café	General Travel	Pizza Place

The first cluster has lot of multiplexes and shopping malls, hence it is not ideal to set up a stadium which requires mostly hotels and different eateries.



Cluster 2

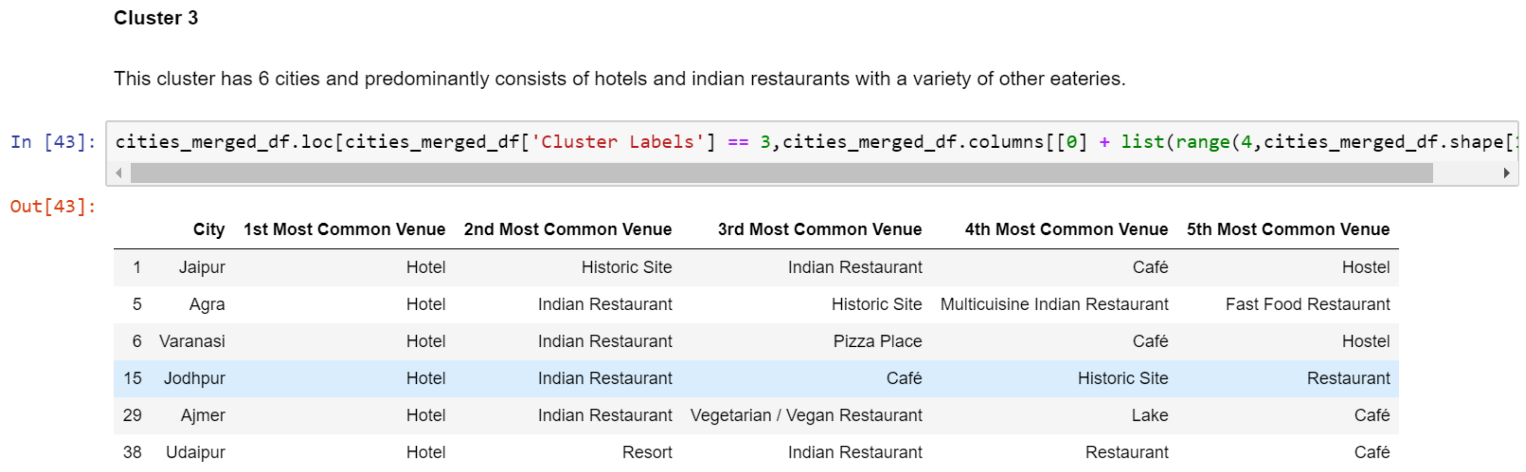
This cluster has 8 cities and have lots of Indian restaurants and hotels

```
In [42]: cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 2,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[42]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Amritsar	Indian Restaurant	Pizza Place	Café	Fast Food Restaurant	Hotel
11	Coimbatore	Indian Restaurant	Café	Hotel	Ice Cream Shop	Shopping Mall
14	Vijayawada	Indian Restaurant	Multiplex	Coffee Shop	Hotel	Café
21	Mysore	Indian Restaurant	Café	Hotel	Pizza Place	Shopping Mall
22	Tiruchirappalli	Indian Restaurant	Train Station	Ice Cream Shop	Multiplex	Hotel
23	Bhubaneswar	Coffee Shop	Hotel	Pizza Place	Indian Restaurant	Fast Food Restaurant
27	Kochi	Café	Hotel	Indian Restaurant	Seafood Restaurant	Ice Cream Shop
35	Mangalore	Indian Restaurant	Hotel	Ice Cream Shop	Seafood Restaurant	Snack Place

In the second cluster, cities mainly have Indian resaurants and hotels which makes it a great cluster to look for a potential city.



Cluster 3

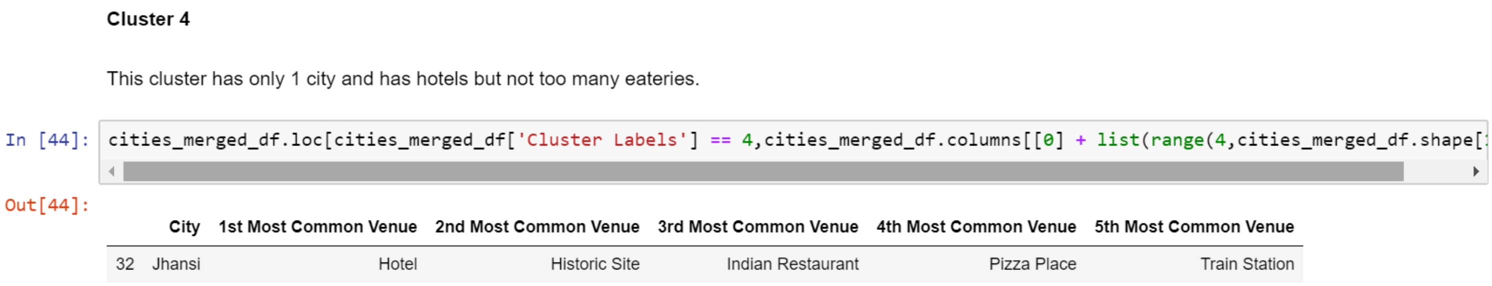
This cluster has 6 cities and predominantly consists of hotels and indian restaurants with a variety of other eateries.

```
In [43]: cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 3,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[43]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Jaipur	Hotel	Historic Site	Indian Restaurant	Café	Hostel
5	Agra	Hotel	Indian Restaurant	Historic Site	Multicuisine Indian Restaurant	Fast Food Restaurant
6	Varanasi	Hotel	Indian Restaurant	Pizza Place	Café	Hostel
15	Jodhpur	Hotel	Indian Restaurant	Café	Historic Site	Restaurant
29	Ajmer	Hotel	Indian Restaurant	Vegetarian / Vegan Restaurant	Lake	Café
38	Udaipur	Hotel	Resort	Indian Restaurant	Restaurant	Café

The third cluster has many hotels and a variety of different restaurants. This makes the cluster a strong candidate for building cricket stadiums.



Cluster 4

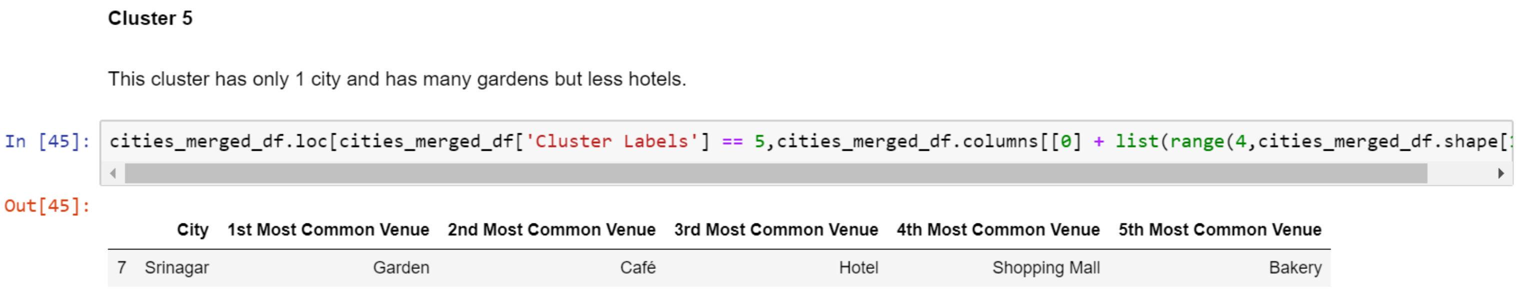
This cluster has only 1 city and has hotels but not too many eateries.

In [44]:
```
cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 4,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[44]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
32	Jhansi	Hotel	Historic Site	Indian Restaurant	Pizza Place	Train Station

The fourth cluster has one city in which there aren’t many eateries hence this cluster isn’t suitable.



Cluster 5

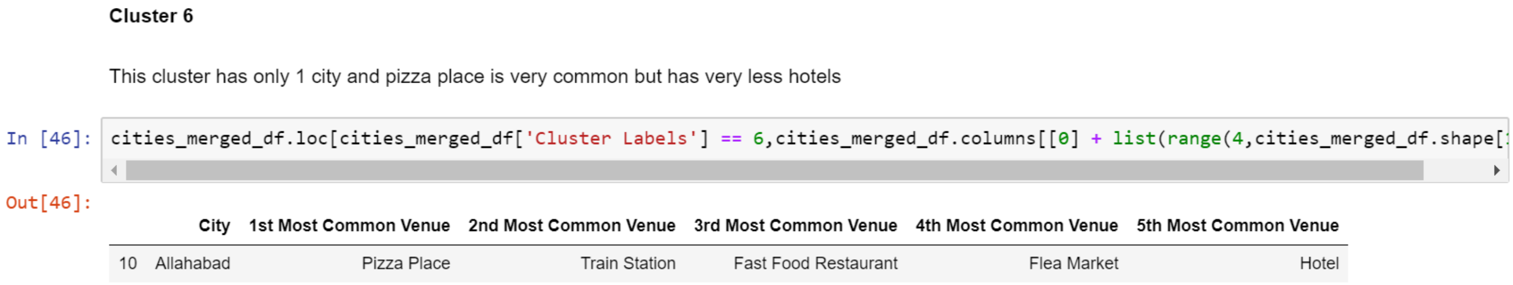
This cluster has only 1 city and has many gardens but less hotels.

In [45]:
```
cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 5,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[45]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Srinagar	Garden	Café	Hotel	Shopping Mall	Bakery

The fifth cluster too has only 1 city in which there are very few hotels and eateries which makes it a non-preferable candidate.



Cluster 6

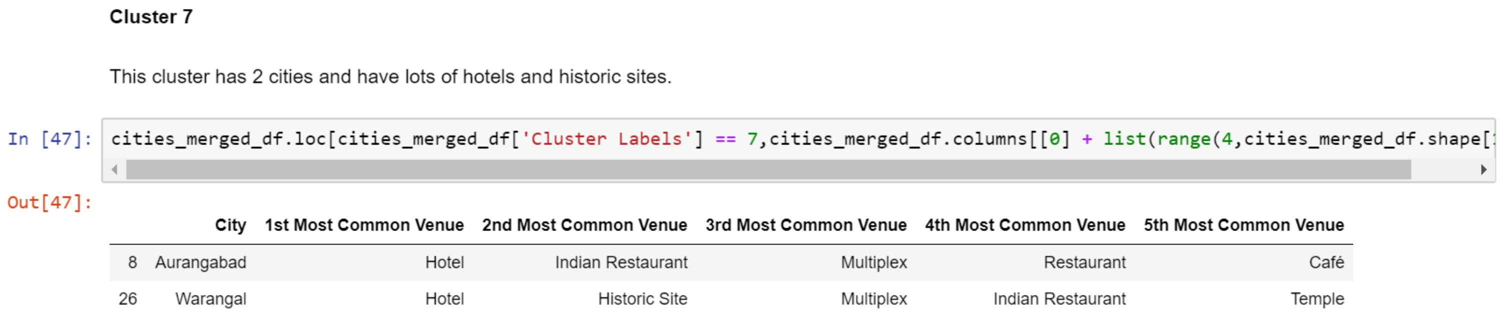
This cluster has only 1 city and pizza place is very common but has very less hotels

In [46]:
```
cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 6,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[46]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Allahabad	Pizza Place	Train Station	Fast Food Restaurant	Flea Market	Hotel

The sixth cluster containing only 1 city has very few hotels (being the 5th most common venue) makes it an undesirable candidate.



Cluster 7

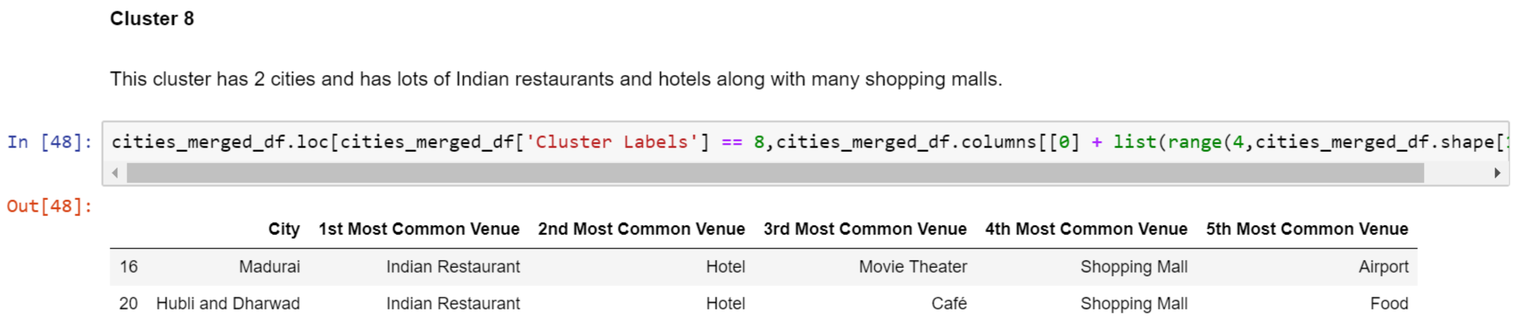
This cluster has 2 cities and have lots of hotels and historic sites.

In [47]:
```
cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 7,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[47]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Aurangabad	Hotel	Indian Restaurant	Multiplex	Restaurant	Café
26	Warangal	Hotel	Historic Site	Multiplex	Indian Restaurant	Temple

The seventh cluster has two cities in which hotels and multiplexes are common. This is a fair candidate and can be considered.



Cluster 8

This cluster has 2 cities and has lots of Indian restaurants and hotels along with many shopping malls.

```
In [48]: cities_merged_df.loc[cities_merged_df['Cluster Labels'] == 8,cities_merged_df.columns[[0] + list(range(4,cities_merged_df.shape[
```

Out[48]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
16	Madurai	Indian Restaurant	Hotel	Movie Theater	Shopping Mall	Airport
20	Hubli and Dharwad	Indian Restaurant	Hotel	Café	Shopping Mall	Food

The eighth cluster too has two cities in which Indian restaurants and hotels are common. This makes it a good candidate for the construction of the cricket stadiums.

Conclusion:

Thus, we have obtained *8 cities from cluster 2*, *6 cities from cluster 3*, *2 cities from cluster 7*, *2 cities from cluster 8,* that makes 18 potential cities in which a new cricket stadium can be built. We have narraowed down potential candidates from a 100 cities to 18 using clustering algorithm and other exploratory analytic techniques.

Please note that this analysis is a very primitive and crude form of analysis. Many other parameters like infrastructure, population density, availabilty of technical staff, etc. have not been considered.