

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# ENTERTAINMENT CENTER CLASSIFICATION IN INDIAN CITIES

# Business Need

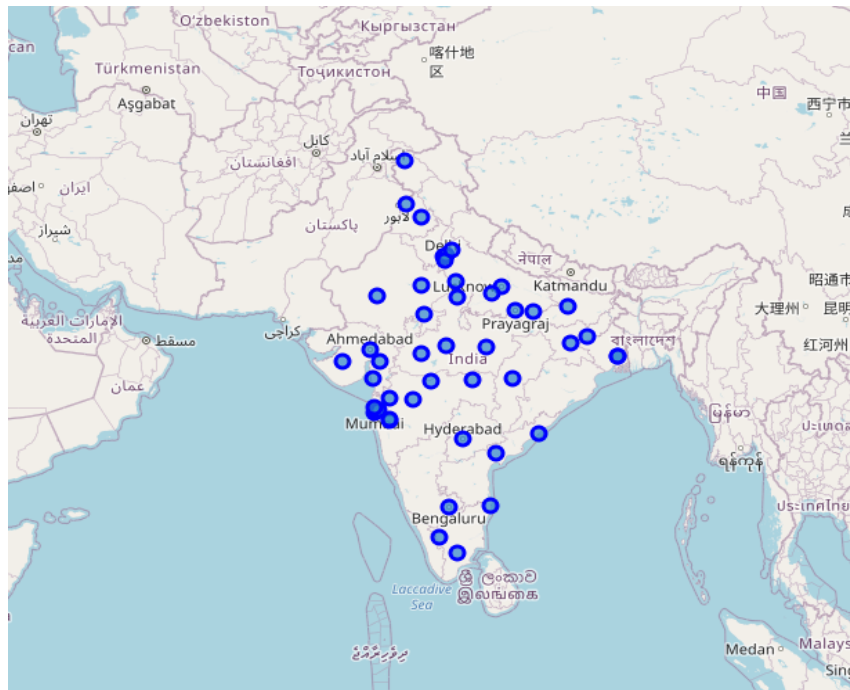
- ❖ Investing in real estates in Indian cities can be mind-bogglingly difficult given the diverse nature of the cities in general
- ❖ A proper data classification model for exploratory data analysis based on given criteria can be a good solution
- ❖ Different cities grouped into criteria based clusters can also allow comparisons and the best possible choice taken

# Data Acquisition

- ❖ The business requirement is to help the requestor (the fictional Mr Sharma) with deciding which are the best cities to open and invest in a chain of nightclubs.
- ❖ The city and population data is webscraped from [https://en.wikipedia.org/wiki/List\\_of\\_cities\\_in\\_India\\_by\\_population](https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population)
- ❖ The prerequisite criteria of cities with population above 1,000,000 is maintained while adjusting dataset
- ❖ Python's Nominatim is used to populate our dataset with corresponding coordinates

Rank ↕	City ↕	Population (2011) <sup>[3]</sup> ↕	Population (2001) ▼	State or union territory ↕
1	<a href="#">Mumbai</a>	12,442,373	11,978,450	<a href="#">Maharashtra</a>
2	<a href="#">Delhi</a>	11,007,835	9,879,172	<a href="#">Delhi</a>
7	<a href="#">Kolkata</a>	4,486,679	4,572,876	<a href="#">West Bengal</a>
6	<a href="#">Chennai</a>	4,681,087	4,343,645	<a href="#">Tamil Nadu</a>
3	<a href="#">Bangalore</a>	8,436,675	4,301,326	<a href="#">Karnataka</a>
4	<a href="#">Hyderabad</a>	6,809,970	3,637,483	<a href="#">Telangana</a>
5	<a href="#">Ahmedabad</a>	5,570,585	3,520,085	<a href="#">Gujarat</a>
12	<a href="#">Kanpur</a>	2,767,031	2,551,337	<a href="#">Uttar Pradesh</a>
9	<a href="#">Pune</a>	3,115,431	2,538,473	<a href="#">Maharashtra</a>

A small clipping of Wikipedia data source to be used



Map of India generated by code with the cities plotted

# Data Cleaning

- ❖ Foursquare is used to provide us with location data. Foursquare docs are very resourceful and maintained in a timely fashion
- ❖ We use <https://developer.foursquare.com/docs/build-with-foursquare/categories/> to provide is with the master(Nightlife Spot) and child categories we would require for creating our venues dataset.
- ❖ After one hot encoding our dataset, getting the unique venue categories and removing the unnecessary rows we get our venues dataframe
- ❖ We sort the dataframe to get the top 10 venues from each city only.



A word cloud generated from the common venue categories

# Data Clustering

- ❖ The data is clustered on the basis of their venue categories and the top ten most common venues are chosen as a base
- ❖ Using a *kcluster* value of 5 to set number of clusters, the clustering algorithm is used to generate a set of cluster labels
- ❖ Using the cluster labels we can plot on map and visualize the 5 different clusters of cities (0-4)



Cluster 0



Cluster 1



Cluster 2



Cluster 3



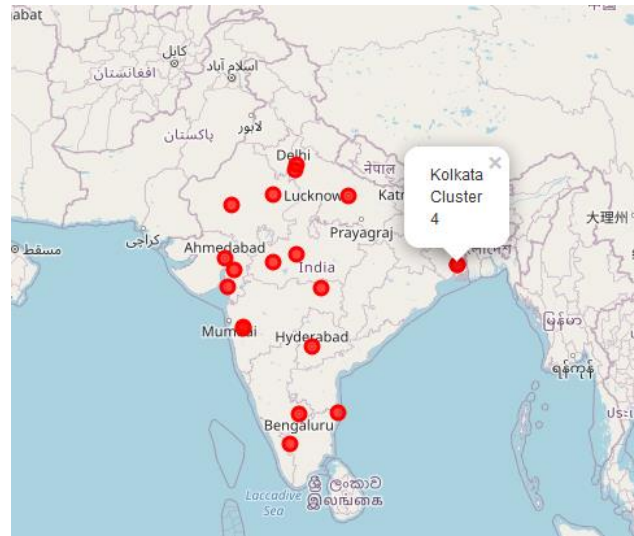
Cluster 4



# Data Interpretation

- ❖ As visualized 5 different clusters of varying common venues have been formed
- ❖ Cluster 0-2 have been eliminated from consideration due to factors of not having nightclubs as important venues or less population
- ❖ Cluster 3 is a cluster of high importance due to it having important, economically strong cities which are close to each other and have a vibrant entertainment center culture(albeit nightclubs are a bit less popular)
- ❖ Cluster 4 is the cluster recommended due to having nightclubs as common entertainment venues and represented by economically important cities across India. But drawback is that the cities are spread out over India's massive landscape.

A	B	C	D	E	F	G	H	I	J	K	L	M
	City	1st Most Com	2nd Most Com	3rd Most Com	4th Most Com	5th Most Com	6th Most Com	7th Most Com	8th Most Com	9th Most Com	10th Most Common	Venue
2	Bangalore	Brewery	Lounge	Pub	Bar	Hookah Bar	Wine Bar	Cocktail Bar	Dive Bar	Gastropub	Hotel Bar	
3	Hyderabad	Lounge	Hotel Bar	Nightclub	Brewery	Gastropub	Hookah Bar	Pub	Beer Garden	Cocktail Bar	Sports Bar	
4	Ahmedabad	Hotel Bar	Lounge	Speakeasy	Nightclub	Hookah Bar	Bar	Whisky Bar	Sports Bar	Sake Bar	Pub	
5	Chennai	Lounge	Bar	Pub	Nightclub	Whisky Bar	Cocktail Bar	Hotel Bar	Sports Bar	Speakeasy	Sake Bar	
6	Kolkata	Nightclub	Pub	Lounge	Hookah Bar	Bar	Gastropub	Karaoke Bar	Brewery	Dive Bar	Cocktail Bar	
7	Surat	Sports Bar	Cocktail Bar	Lounge	Brewery	Pub	Night Market	Bar	Whisky Bar	Speakeasy	Sake Bar	
8	Pune	Lounge	Pub	Bar	Nightclub	Brewery	Gastropub	Hookah Bar	Cocktail Bar	Beach Bar	Whisky Bar	
9	Jaipur	Lounge	Bar	Hotel Bar	Hookah Bar	Nightclub	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	Pub	
10	Lucknow	Lounge	Hookah Bar	Nightclub	Bar	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	Pub	Night Market	
12	Nagpur	Lounge	Bar	Pub	Nightclub	Beach Bar	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	Night Market	
13	Indore	Lounge	Pub	Hookah Bar	Bar	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	Nightclub	Night Market	
15	Bhopal	Cocktail Bar	Pub	Hookah Bar	Hotel Bar	Nightclub	Lounge	Bar	Night Market	Whisky Bar	Sports Bar	
17	Pimpri-Ch	Lounge	Pub	Bar	Brewery	Nightclub	Cocktail Bar	Gastropub	Hookah Bar	Beach Bar	Whisky Bar	
19	Vadodara	Lounge	Sports Bar	Bar	Speakeasy	Hotel Bar	Whisky Bar	Sake Bar	Pub	Nightclub	Night Market	
20	Ghaziabad	Lounge	Sports Bar	Hookah Bar	Bar	Pub	Nightclub	Hotel Bar	Whisky Bar	Speakeasy	Sake Bar	
25	Faridabad	Lounge	Nightclub	Bar	Pub	Gastropub	Karaoke Bar	Beer Garden	Dive Bar	Cocktail Bar	Hookah Bar	
37	Howrah	Nightclub	Pub	Lounge	Hookah Bar	Bar	Brewery	Gastropub	Night Market	Whisky Bar	Sports Bar	
40	Coimbatore	Nightclub	Bar	Lounge	Beach Bar	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	Pub	Night Market	
42	Jodhpur	Hotel Bar	Lounge	Cocktail Bar	Hookah Bar	Bar	Night Market	Whisky Bar	Sports Bar	Speakeasy	Sake Bar	



Cluster 4 dataset and plotted on map respectively





THANK YOU