Introduction

To compare the localities of the four most populated cities in India - Mumbai, Delhi, Bangalore, and Hyderabad, and determine how similar or dissimilar they are.
This study will be relevant to any company which has a presence in any of these cities and like to expand the business to any other city among these cities.

Data

1. Locality based PIN mapping data for acquiring Pincode and address of all the localities within these cities.
    a. Source:
       https://data.gov.in/resources/villagelocality-based-pin-mapping-16th-march-2017
2. Google Geocoding API for converting address to latitude and longitude.
3. Foursquare API to collect venues info of localities.

Methodology

Once the data is gathered, cleansed and stored in the data frame, the following step-by-step procedure is followed throughout this analysis.
1. Draw a map and show all the localities
       a) Create a map of India with all localities superimposed on top.
2. Explore the neighborhoods
       a) Explore all the localities in the cities - Delhi, Bangalore, Mumbai and Hyderabad; using the latitude & longitude data.
       b) Using Foursquare API, obtain a list of all the venues available in these cities.
       c) Check number of venues and unique categories returned for each locality.
       d) Group the data by locality and by taking the mean of the frequency of occurrence of each category.
       g) Display each locality along with the top 5 most common venues.
       h) Identify the top 10 venues for each locality. This data will be further used for clustering.
3. Cluster neighborhoods
       a) With an assumption of 5 clusters, use K-Cluster algorithm to come up with 5 different clusters in India with a similar set of venues.
       b) Map all the defined clusters.
4. Examine clusters and predict
       a) Examine each cluster and determine the discriminating venue categories that distinguish each cluster.
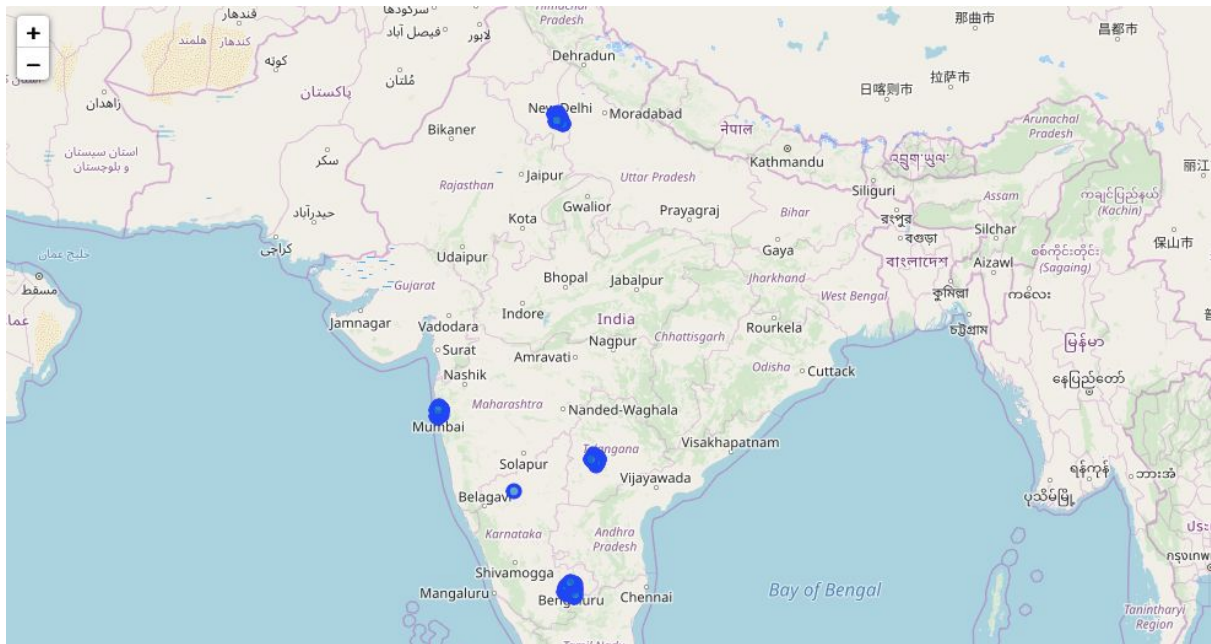
Results

Results below are based on the steps defined in the Data and Methodology sections.

Localities data cleansed, processed and merged with geospatial coordinates:
The following image shows this information.

	borough	pincode	neighbourhood	latitude	longitude
0	BANGALORE, KARNATAKA	560001	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.9816	77.5943
1	BANGALORE, KARNATAKA	560002	Bangalore North, Bangalore City S.O, Bangalore...	12.9608	77.5842
2	BANGALORE, KARNATAKA	560003	Bangalore North, Malleswaram S.O, Bangalore No...	12.9989	77.5761
3	BANGALORE, KARNATAKA	560004	Bangalore South, Basavanagudi H.O, Bangalore S...	12.9526	77.5722
4	BANGALORE, KARNATAKA	560005	Bangalore North, Fraser Town S.O, Bangalore North	12.998	77.6227
5	BANGALORE, KARNATAKA	560006	Bangalore North, J.C.Nagar S.O, Bangalore Nort...	13.0166	77.5836
6	BANGALORE, KARNATAKA	560007	Bangalore North, Air Force Hospital S.O, Banga...	12.9639	77.6308

Drawing a map and showing the neighborhoods

Explore the neighborhoods

- Using Foursquare API, Total 2655 venues have been identified in India (all 4 cities), out of which there are 245 unique categories of the venues.

	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.981603	77.594306	Chevron Hotel	12.983619	77.594379	Hotel
1	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.981603	77.594306	South Indies	12.983459	77.594437	Indian Restaurant
2	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.981603	77.594306	Vidhana Soudha	12.979027	77.591881	Capitol Building
3	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.981603	77.594306	Oriental Spice	12.983893	77.593591	Chinese Restaurant
4	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.981603	77.594306	Nisarga Garden Restaurant	12.983332	77.594431	Indian Restaurant

- Based on one hot encoding, calculating the frequency of venue categories for all localities.

- Displaying the top 5 venue categories for localities with frequency.

```
----Aliganj, Kasturba Nagar, Lodi Road H.O, Defence Colony----
                    venue  freq
0       Indian Restaurant   0.2
1        Pedestrian Plaza   0.1
2  Performing Arts Venue   0.1
3      Chinese Restaurant   0.1
4                   Diner   0.1


----Amberpet, Malakpet Colony S.O, Amberpet, Amberpet, Sripuram Colony B.O, Amberpet, Saidabad,
abad), Saidabad----
                          venue  freq
0                          Café  0.50
1                   Pizza Place  0.25
2             Indian Restaurant  0.25
3                           ATM  0.00
4     Middle Eastern Restaurant  0.00


----Ameerpet, Erragadda S.O, Ameerpet----
                     venue  freq
0                Bookstore  0.25
1            Movie Theater  0.25
2              Pizza Place  0.25
3      American Restaurant  0.25
4                    Motel  0.00
```
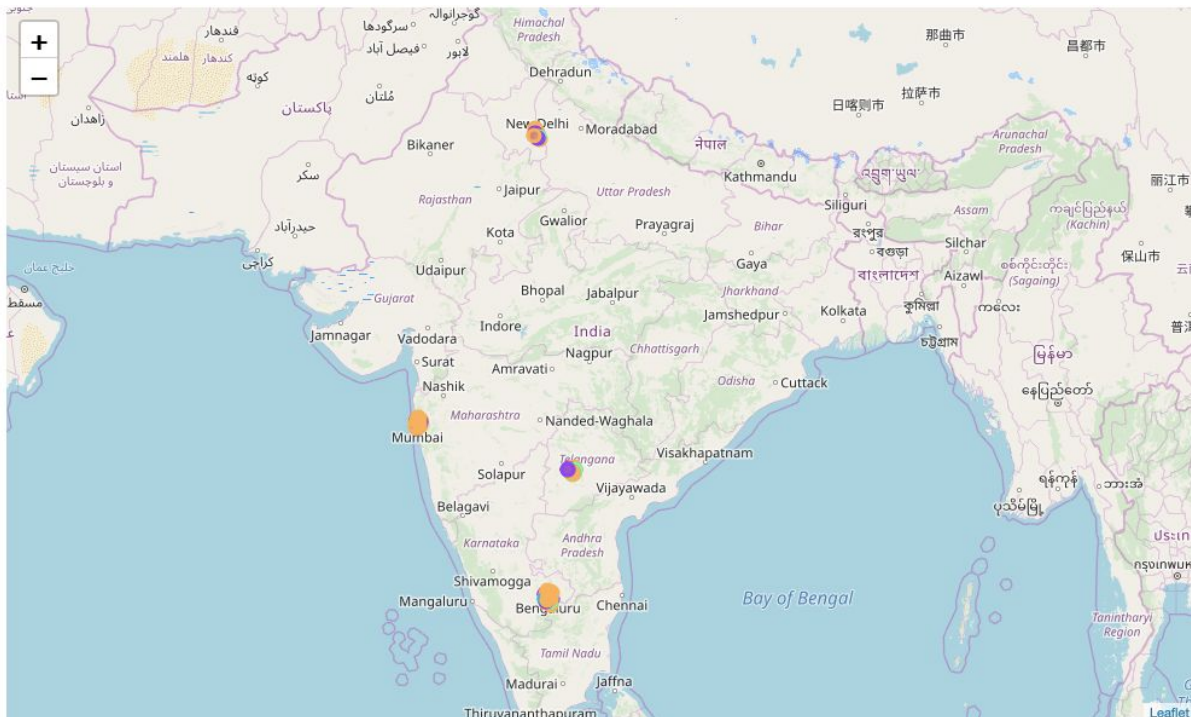
- The top 10 venues for each locality

	Locality	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aliganj, Kasturba Nagar, Lodi Road H.O, Defenc...	Indian Restaurant	Food Court	Temple	Diner	Pedestrian Plaza	Performing Arts Venue	Chinese Restaurant	Eastern European Restaurant	Theater	Dessert Shop
1	Amberpet, Malakpet Colony S.O, Amberpet, Amber...	Café	Indian Restaurant	Pizza Place	Duty-free Shop	Dessert Shop	Diner	Dog Run	Doner Restaurant	Donut Shop	Zoo
2	Ameerpet, Erragadda S.O, Ameerpet	American Restaurant	Movie Theater	Pizza Place	Bookstore	Farm	Falafel Restaurant	Event Space	Electronics Store	Farmers Market	Fast Food Restaurant
3	Ameerpet, Gopanpally S.O, Ameerpet	Athletics & Sports	Convenience Store	Indian Restaurant	Food Truck	Zoo	Eastern European Restaurant	Doner Restaurant	Donut Shop	Duty-free Shop	Event Space
4	Ameerpet, Sanjeev Reddy Nagar S.O, Ameerpet	Indian Restaurant	Pizza Place	Juice Bar	Bakery	Café	Breakfast Spot	Sandwich Place	Falafel Restaurant	Event Space	Electronics Store

Cluster Neighborhoods

- The K-mean clustering algorithm resulted in 5 clusters.

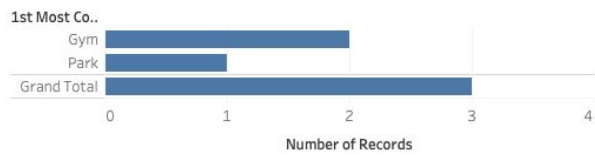|	borough	pincode	Locality	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	BANGALORE, KARNATAKA	560001	Bangalore North, Bangalore Bazaar S.O, Bangalo...	12.9816	77.5943	1.0	Café	Indian Restaurant	Hotel	Metro Station	Intersection
1	BANGALORE, KARNATAKA	560002	Bangalore North, Bangalore City S.O, Bangalore...	12.9608	77.5842	4.0	Theater	Other Nightlife	Indian Restaurant	Electronics Store	Hotel
2	BANGALORE, KARNATAKA	560003	Bangalore North, Malleswaram S.O, Bangalore No...	12.9989	77.5761	3.0	Café	Department Store	Gym	Clothing Store	Plaza
3	BANGALORE, KARNATAKA	560004	Bangalore South, Basavanagudi H.O, Bangalore S...	12.9526	77.5722	4.0	Sandwich Place	Breakfast Spot	Bakery	Pool	Garden Center
4	BANGALORE, KARNATAKA	560005	Bangalore North, Fraser Town S.O, Bangalore North	12.998	77.6227	1.0	Flea Market	Coffee Shop	Shopping Mall	Chinese Restaurant	New American Restaurant

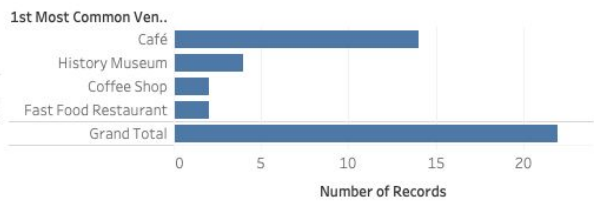- All the defined clusters are mapped on India
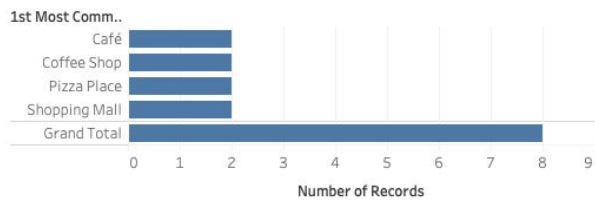
Examine Clusters on Tableau

Examine Clusters

Cluster-0

1st Most Co..

Gym	
Park	
Grand Total	

Number of Records

Cluster-1

1st Most Comm..

Café	
Coffee Shop	
Pizza Place	
Shopping Mall	
Grand Total	

Number of Records

Cluster-2

1st Most Co..

Restaurant	
Grand Total	

Number of Records

Cluster-3

1st Most Common Ven..

Café	
History Museum	
Coffee Shop	
Fast Food Restaurant	
Grand Total	

Number of Records

Cluster-4

1st Most Common Ven..

Fast Food Restaurant	
Coffee Shop	
Café	
Hotel	
Grand Total	

Number of Records
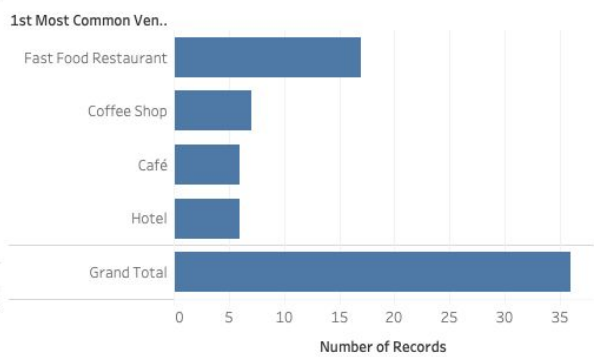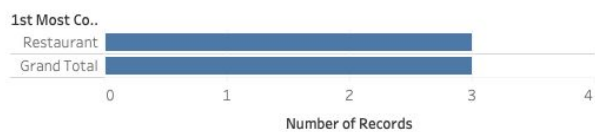
Conclusion

The cluster analysis is used to classify localities into relatively homogeneous groups based on a set of venue categories considered. These categories can be chosen according to the market research objectives; which problem need to be solved and which hypotheses need to be proven or debunked. The clusters are not predefined, but rather naturally suggested by the data, revealing any similarities or dissimilarities.