

Similarity and Difference between major Indian cities

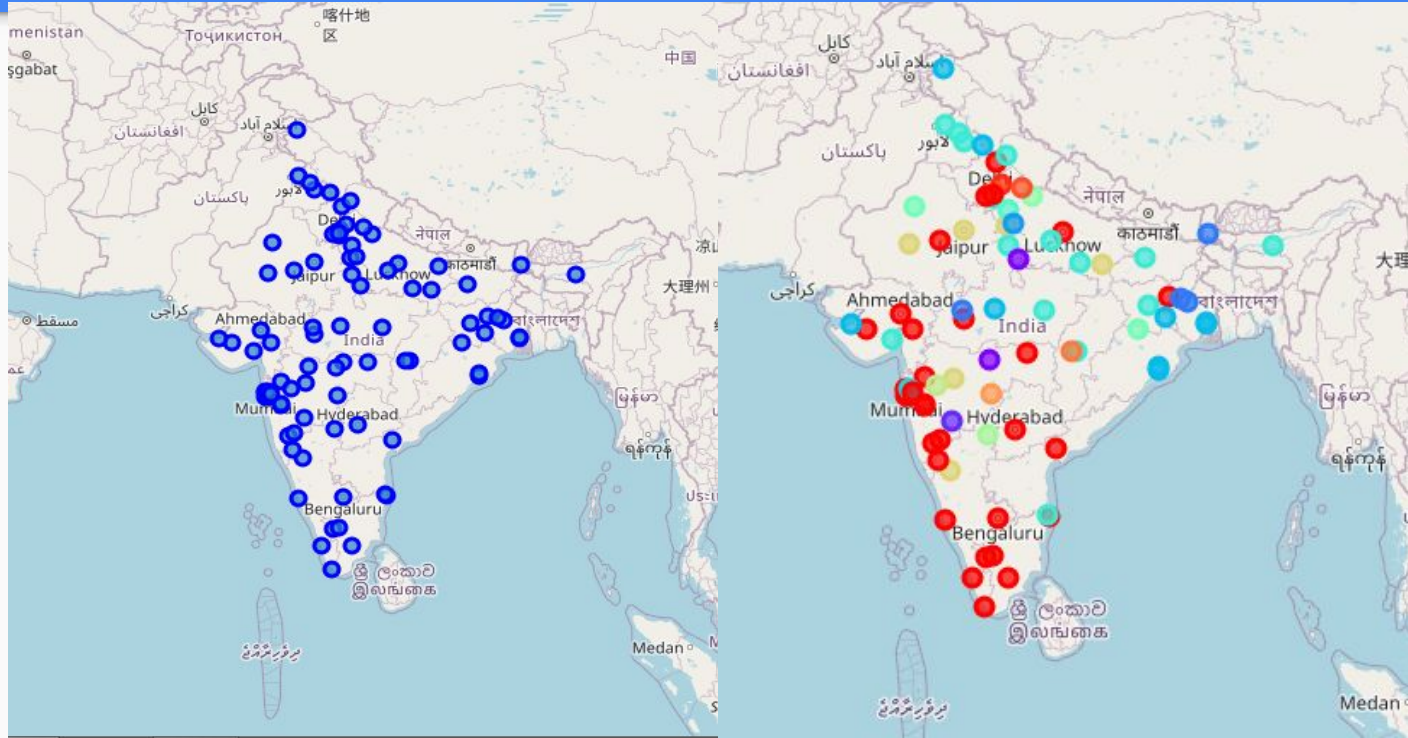
Are all the cities are same?

- Is it possible to determine whether there is any similarity between different cities
- What are the factor to look for when doing exploratory data analysis
- Choosing which city is best for expansion

Data collection and cleaning

- Web scraping data from Wikipedia to collect city information and saving it as pandas dataframe and removing unnecessary data to get city and states name and limiting the data to 100 cities
- Using geocoder to get location data of each city
- Foursquare API to get venues data and adding it to our dataframe to get venue's location and category for each city

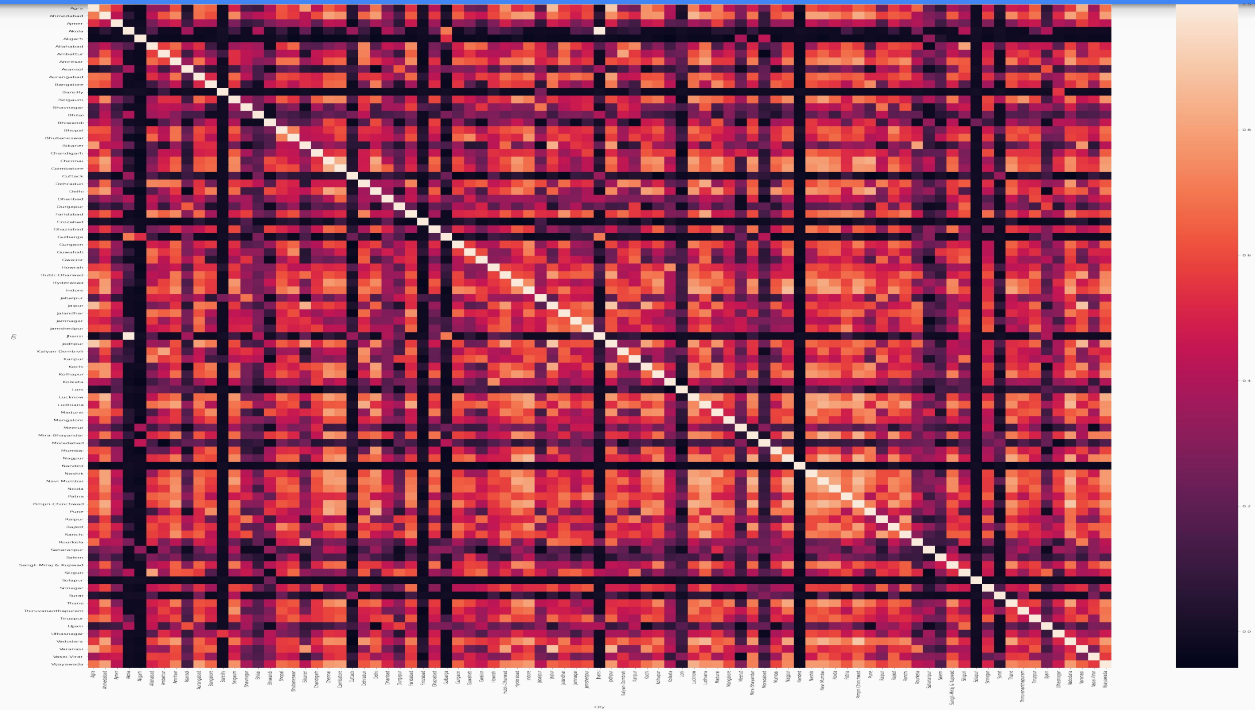
Visualizing raw data and Cluster data



Correlation matrix for every city

City	Agra	Ahmedabad	Ajmer	Akola	Aligarh	Allahabad	Ambattur	Amritsar
City								
Agra	1.000000	0.707224	0.536000	-0.013823	-0.026260	0.337391	0.331171	0.609927
Ahmedabad	0.707224	1.000000	0.436395	0.072793	-0.032926	0.431695	0.620302	0.738768
Ajmer	0.536000	0.436395	1.000000	-0.010889	-0.020686	0.261886	0.420562	0.481071
Akola	-0.013823	0.072793	-0.010889	1.000000	-0.007723	0.180471	-0.015778	-0.016284
Aligarh	-0.026260	-0.032926	-0.020686	-0.007723	1.000000	-0.023748	-0.029975	-0.030935
Allahabad	0.337391	0.431695	0.261886	0.180471	-0.023748	1.000000	0.571356	0.670224
Ambattur	0.331171	0.620302	0.420562	-0.015778	-0.029975	0.571356	1.000000	0.585241
Amritsar	0.609927	0.738768	0.481071	-0.016284	-0.030935	0.670224	0.585241	1.000000

Correlation Heatmap



Conclusion

- Most of the cities are similar to one another
- Cluster shows that there are three main clusters
- There are 13 cities which are different from all other cities
- Heatmap shows correlation with each other, black strips shows outliers

Thank You