# Exploration of Indian Cities

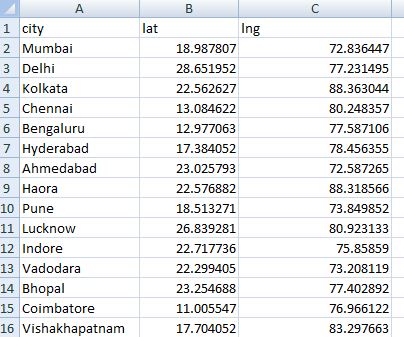
****

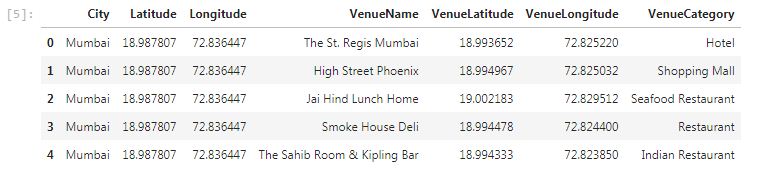
IntroductionIndia is the 2nd most populous country in the world. It is also 5th largest economy and also vibrant hub  
for entrepreneurship activity. It is diverse, multicultural and provides a secure, safe and comfortable  
business environment. But unfortunately the data available about different cities is minimal. So, data  
about various places needs to be created for businesses, citizens, government etc to make better  
decisions.

Business problemParsing the data available about Indian cities and presenting it in a more understandable format along  
with clustering cities based on number of popular places have many applications including  
establishment of businesses, selecting relocation cities etc

Target AudienceBusinesses/ Investors/ Entrepreneurs: Decision upon where to start a new business like restaurant  
across India. What business is missing in which city?  
Citizens: Decide upon relocation based on his preferences for food, museums, entertainment, shopping  
etc  
Travelers: Decisions about tourism plans becomes easier with detailed sorting of information about  
destination, ccultural places, food availability etc  
Government: Policy decisions about incentives to businesses, tourism promotion etc can be made

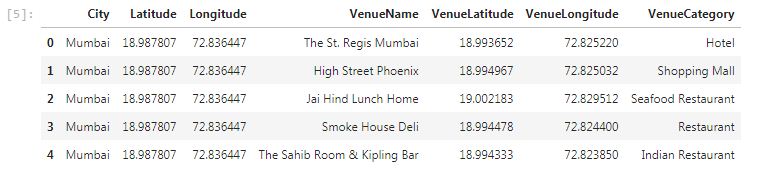
**Data**

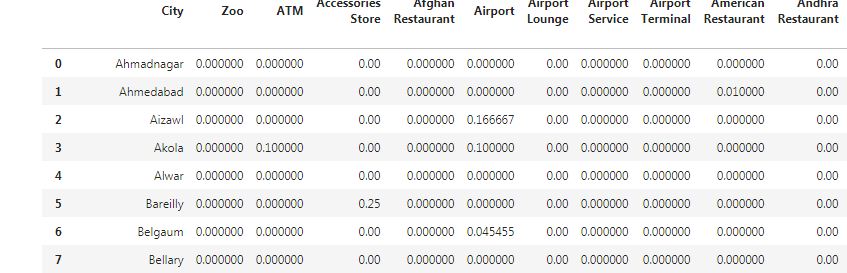
* Location of various cities across India along with latitude and longitudes is needed. Data(CSV) file from <https://simplemaps.com/data/in-cities> is used for this. This CSV file is read in the project and converted to a data frame. Here is a sample of CSV
* Venue data of various Indian cities is extracted using Foursquare Explore venue API. It returns a list of different kind of venues like Hotels, Restaurants, Museums etc. More details can be explored in <https://developer.foursquare.com/docs/api>. Below is a data frame constructed with this data.

****

**Methodology**

The project is performed based on learnings in IBM Data Science professional course.

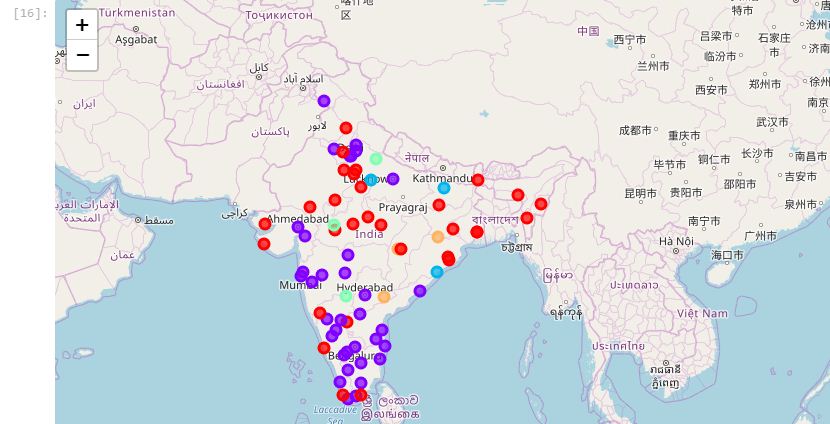
* Information about various Indian cities like latitude and longitude is extracted from the CSV file mentioned in data section above.
* Then we will use the Foursquare API to explore neighborhoods in these cities and create data frame about various venues



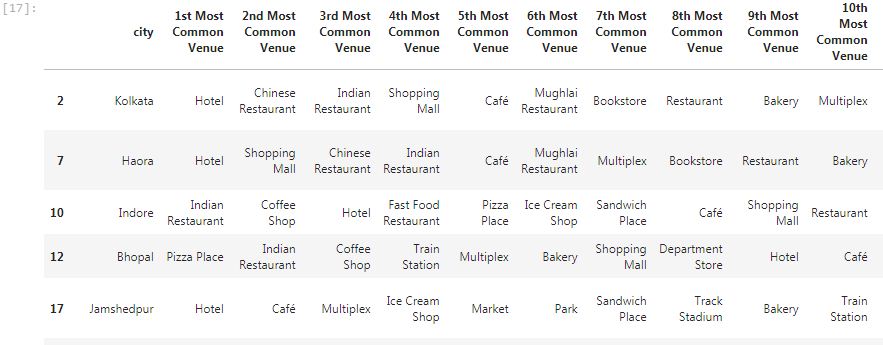
* Group data about venues based on Venue Categories per city
* Use KMeans cluster for clustering of various Indian cities.



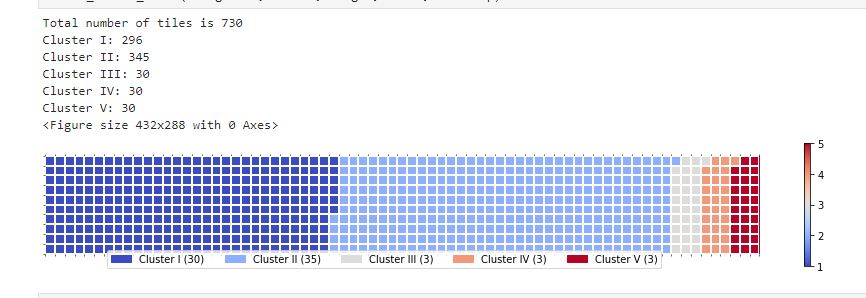
* Map these cities using Folium library



* For city in each cluster, find information about most occurring venues and store it in a dataframe.



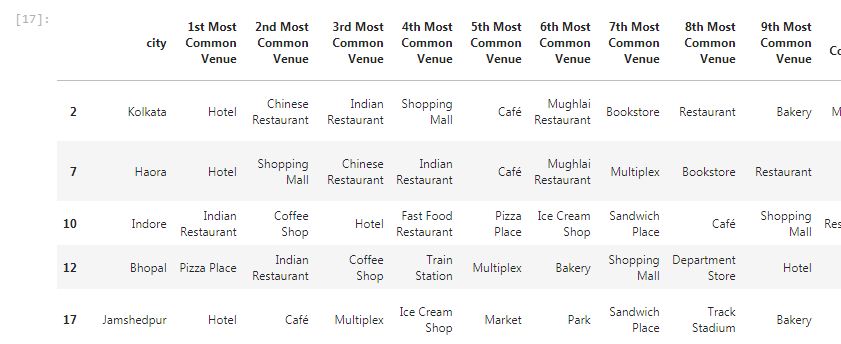
* General Waffle Charts for better representation of number of cities in each cluster to get a comprehensive overview.



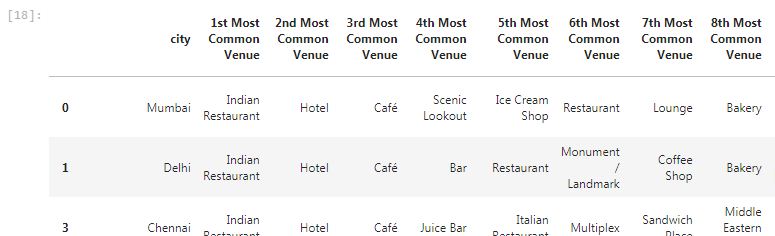
**Results**

Details about the most common places in various cities

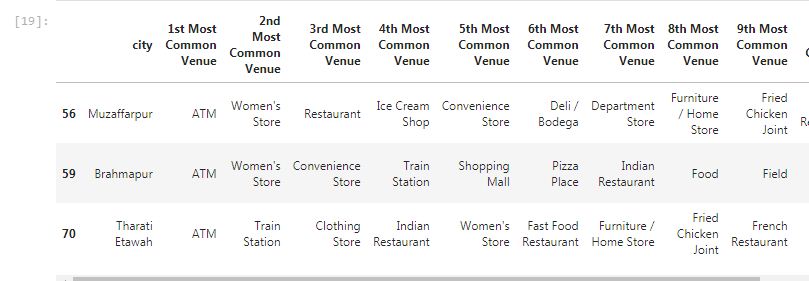
Cluster 1

****

Cluster2

****

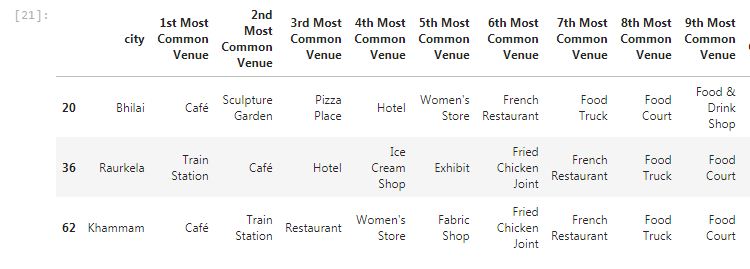
Cluster 3

****

Cluster 4

****

Cluster 5

****

**The information about most common places and clustering of cities will be very helpful for various stakeholders like businesses, citizens, government etc**

**Discussion**

In this section, I would be discussing the observations I have noted and the recommendation that I can make based on the results.

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

* A look at the most common venues shows the richness of Indian cities
* It shows many results
* For instance, Government can evaluate the effectiveness of their financial inclusion programmes as ATMs formed the most common in cluster 3.
* Tourists can diversify their choices like Cluster 5 provides different experiences including Sculpture garden
* Many Indian cities are dominated by hotels, restaurants etc. So, there is no dearth of cuisines
* It can be done further enhanced through addition of data about factors such as transportation, demographics of inhabitants etc to create Standard of Living frameworks too

Finally, FourSquare proved to be a good source of data but it would have been great if there is a way to access more diverse and quantity of data.

**Conclusion**

Although all of the goals of this project were met there is definitely room for further improvement and development as noted above. Addition of further data related to transportation, crimes, education, hospitals etc can create an overall data base. Through the analysis, let us hope that it will be used in future for better policy decisions, business establishment, tourist arrivals to India etc.