

# Presentation

## Objective

In this project, we will study area classification using Foursquare API data and ML segmentation and clustering. The aim of this project is to segment areas of Delhi and Mumbai based on the most common places captured from Foursquare in India.

Using segmentation and clustering, we hope we can determine:

the similarity or dissimilarity of both cities classification of area located inside the city whether it is residential, tourism places, or others

## Data

Data is acquired from following two -

-- For Mumbai (<https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/>)

-- For Delhi (<https://www.whatsuplife.in/delhi/blog/zip-pin-postal-code-pincodes-delhi/>)

and these will be converted to csv by parsing the html text

Data is in form of Area along with their Pincodes for each city. We will further fetch the latitude and longitude for each area and store to a DataFrame for analysis and also to a separate CSV file to avoid scrapping again.

This data (Area, Pincode, City, Latitude, Longitude) will be helpful to identify common places using FS API.

## Methodology

Above, we have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in both cities, Mumbai and Delhi

After that, explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters

K-means clustering algorithm will be used to complete this task. And also, the Folium library to visualize the neighborhoods in Mumbai and Delhi and their emerging clusters.

Based on dataframe analysis above, we found out that 400003 area in Mumbai and 110001 area in Delhi are both have the highest number of area within itself.

## Discussion

Based on cluster for each city above, we believe that classification for each cluster can be done better with calculation of venue categories (most common) in each city. Referring to each cluster,

we can't determine clearly what represent in each cluster by using Foursquare - Most Common Venue data.

However, for the sake of this project we assumed each cluster as follow:

-- Cluster 1: Mumbai: Tourism -- Cluster 2: Mumbai: Residential -- Cluster 3: Mumbai: Mix -- ster 1: Delhi: Residential -- ster 2: Delhi: Tourism -- ster 3: Delhi: Sport

What is lacking at this point is a systematic, quantitative way to identify and distinguish different district and to describe the correlation most common venues as recorded in Foursquare. The reality is however more complex: similar cities might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

We believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the different cities. Further studies are indeed needed in order to relate the data acquired, then observe it to more meaningful and objective results.

## **Conclusion**

Using Foursquare API, we can captured data of common places all around the world. Using it, we refer back to our main objectives, which is to determine;

the similarity or dissimilarirty of both cities classification of area located inside the city whether it is residential, tourism places, or others

In conclusion, both cities Mumbai and Delhi are the center of attraction among Indians. However, to declare both cities are similar or dissimilar base on common venues visited is quite difficult. Both cities is similar in some venues also dissimilar in certain venues. And for classitification based on common venues, again we must have more systematic or quantitative way to identify and declare this. Comparison can be made, but no such method or quantitative data to determine this. We hope in the future, a method to determine it can be establish and explore for references.

Thank you.