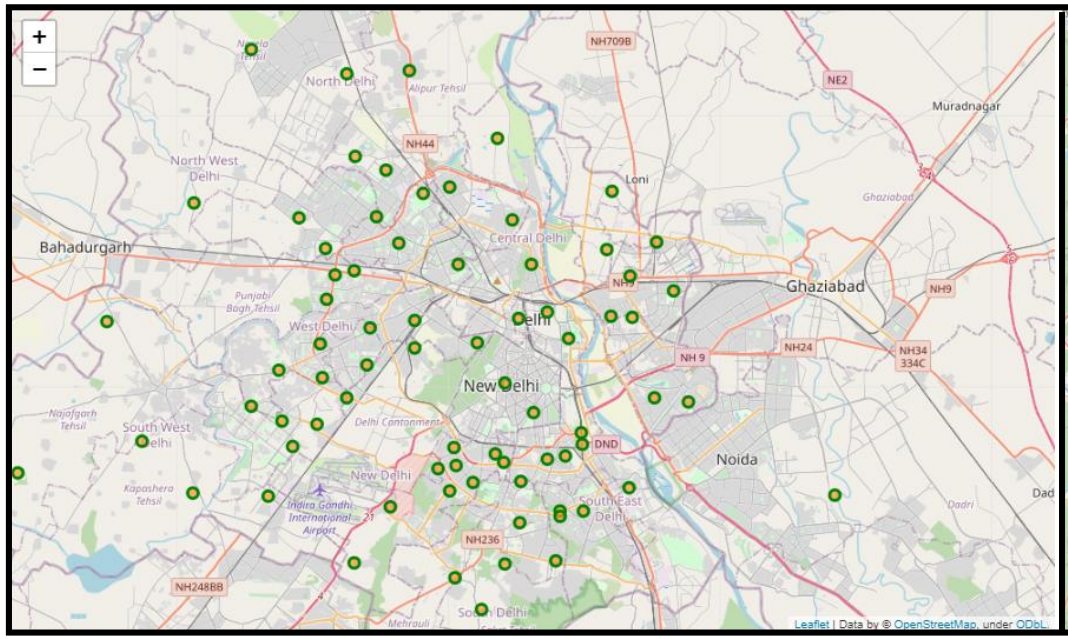
The Battle Of Neighborhoods Delhi vs Pune

Introduction

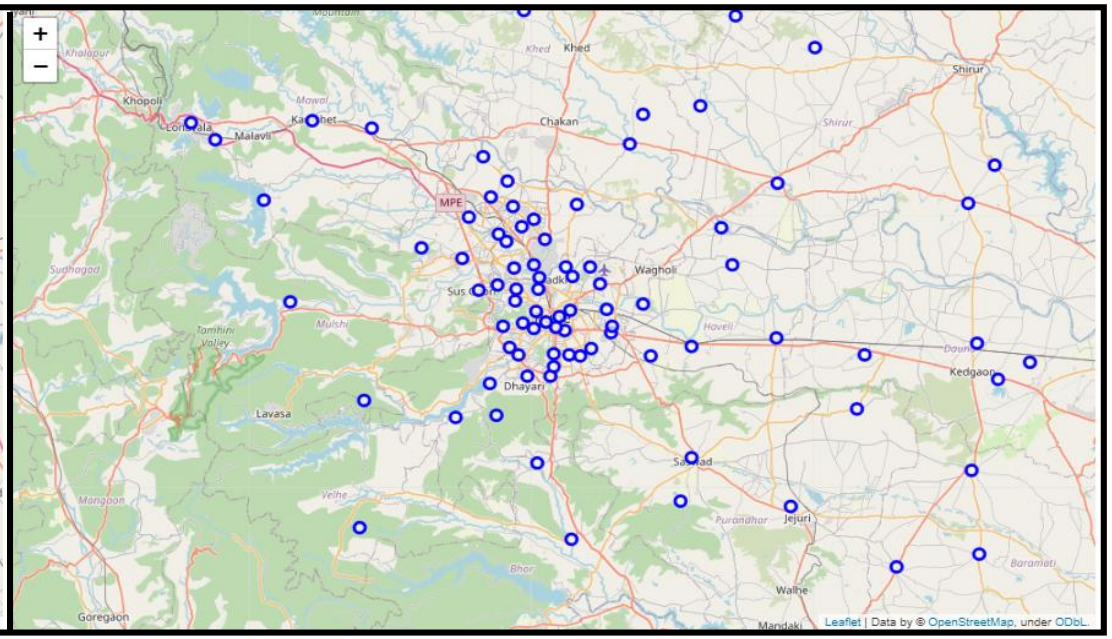
- It is tough for people who leave their native place and shift to a new place in order to pursue their career. This is a guide for such people.
- Here we consider that a newly Graduated Engineer has left his native place in Delhi and has come to Pune.
- Similarity between Dilshad Garden, Delhi and Infotech Park (Hinjawadi), Pune.
- Analyzing, clustering and comparing neighborhoods of Delhi and Pune.

Data Acquisition and Cleaning

- For this Project two type of Data is required:
  - **Neighborhood Data** - Datasets that lists the names of the neighborhoods of Delhi and Pune and their latitude and longitude coordinates. Data is acquired by Web Scraping.
    Delhi - https://www.mapsofindia.com/pincode/india/delhi/
    Pune - https://www.mapsofindia.com/pincode/india/maharashtra/pune/
  - **Venues Data** - Data that describes the top 50 venues (restaurants, cafes, parks, museums, etc.) in each neighborhood of the two cities. This data will be retrieved from Foursquare which is one of the world largest sources of location and venue data. Foursquare API will be utilized to get and download the data.
    An example URL is the following:
    https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

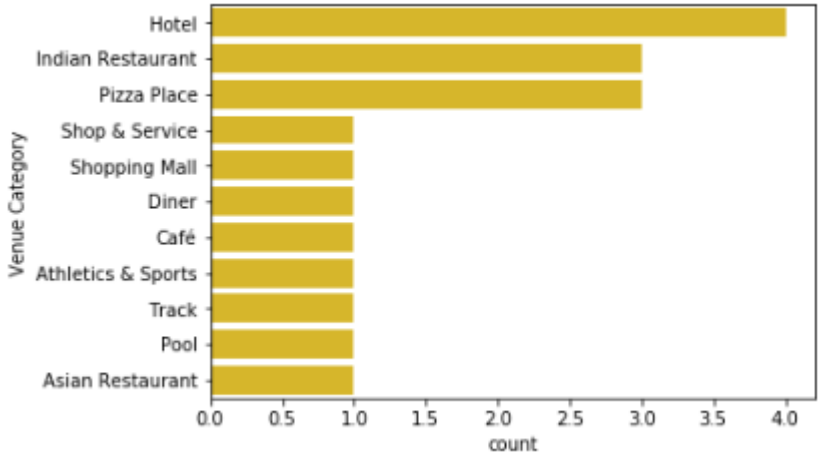


Map of Delhi With Neighborhoods

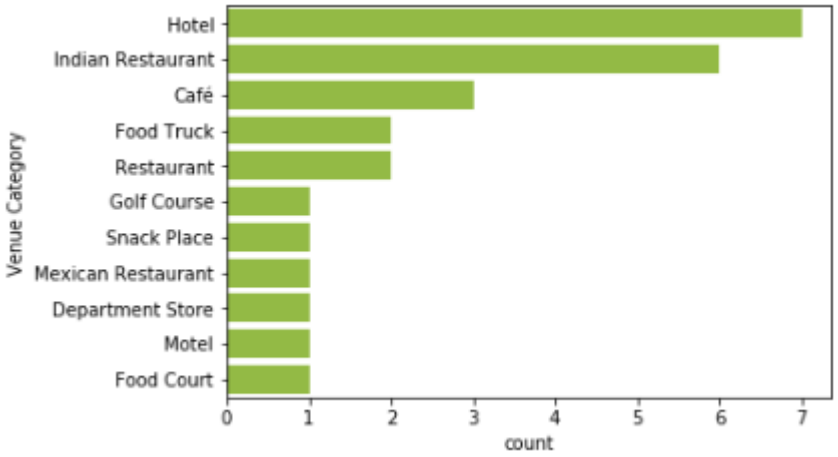


Map of Pune With Neighborhoods

Dilshad Garden vs Rajiv Gandhi Infotech Park



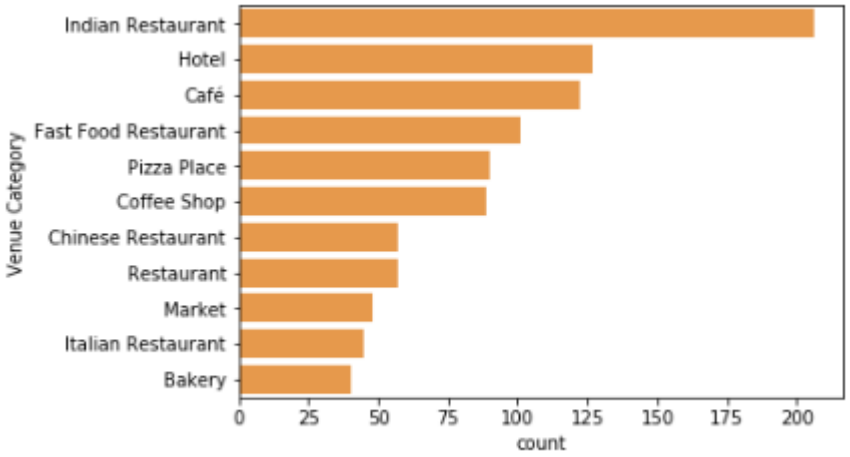
Top 10 Venue Categories of Dilshad Garden



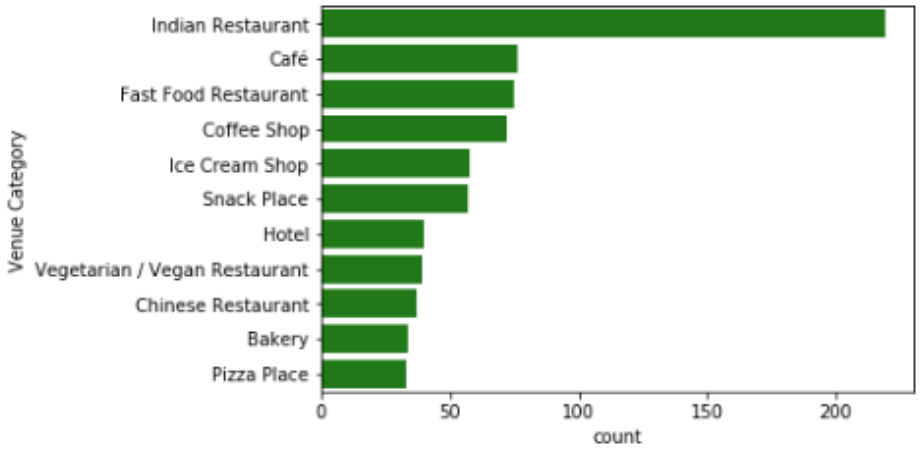
Top 10 Venue Categories of Infotech Park

- "Hotel" is the most popular place in both Postal Codes, but Infotech Park (Hinjawadi) has a greater number of hotels as compared to Dilshad Garden. Infotech Park (Hinjawadi) has a greater number of Food Venues as compared to Dilshad Garden.
- One possible reason could be that Infotech Park (Hinjawadi) is an IT hub. Thisjustifies a greater number of hotels, multicuisine restaurants, cafe etc.

Delhi vs Pune



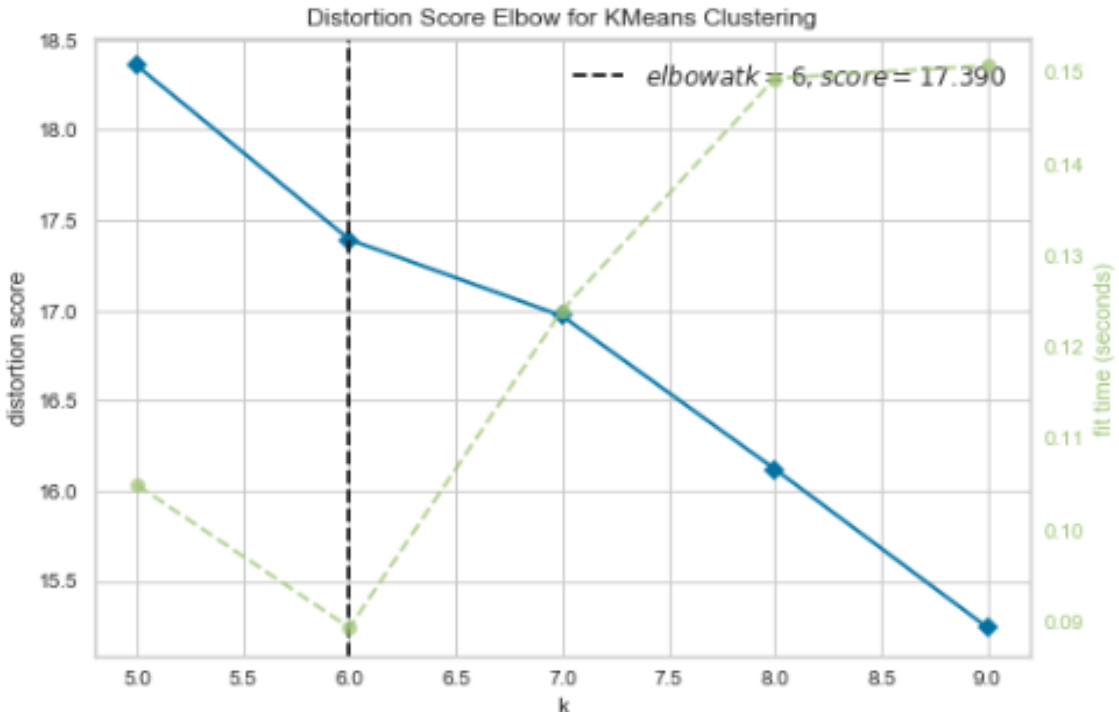
Top 10 Venue Categories of Delhi



Top 10 Venue Categories of Pune

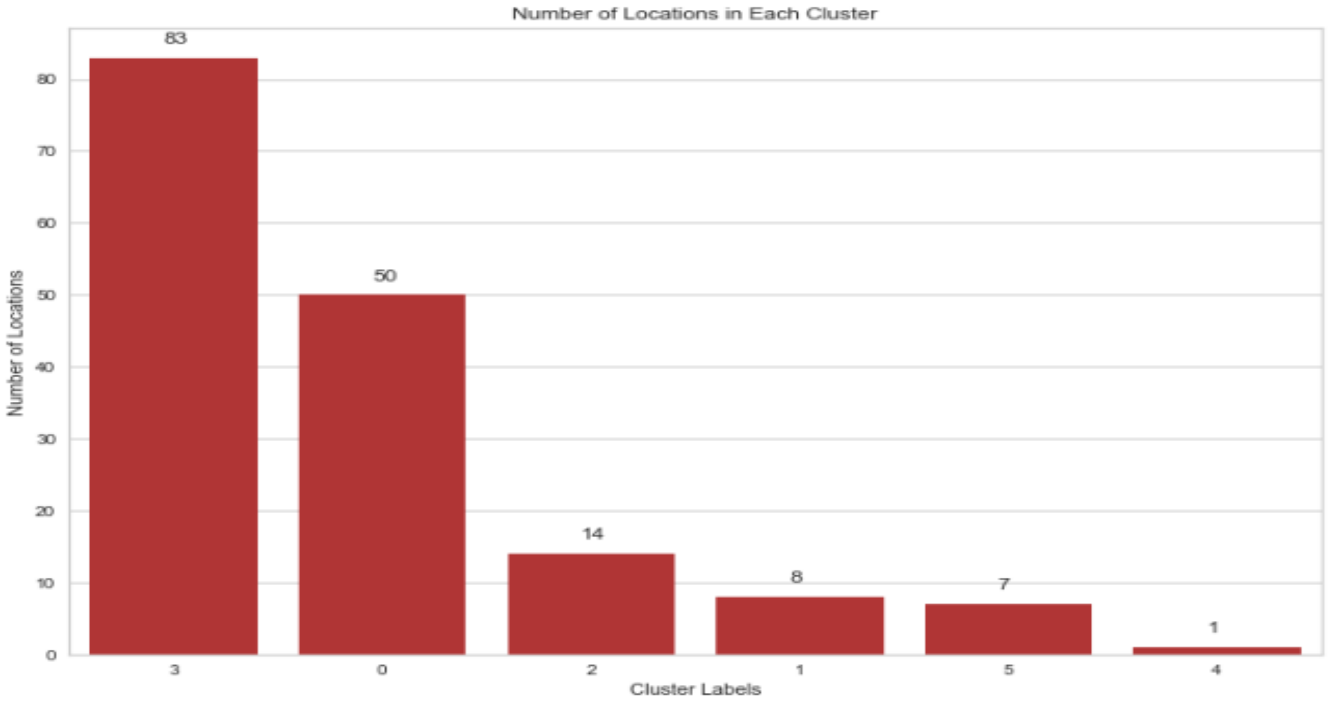
- "Indian Restaurant" is clearly a favorite choice in both Delhi and Pune.
- In Delhi, second most opened place is "Hotel" whereas in Pune its "Cafe". One of the Possible reasons could be that Delhi is capital of India so it's more likely to have tourists. Hence a greater number of Hotels.
- Also, both locations have tremendous number of food options. Which is one of the similarities between two places.

KMEANS



- K-Means clustering algorithm identifies k number of centroids and allocates every data point to nearest clusters while keeping the centroids as small as possible.
- It is one of the simplest and popular unsupervised machine learning algorithm and apt for this Project.
- Value of k from KELBOW_VISUALIZER() is 6.

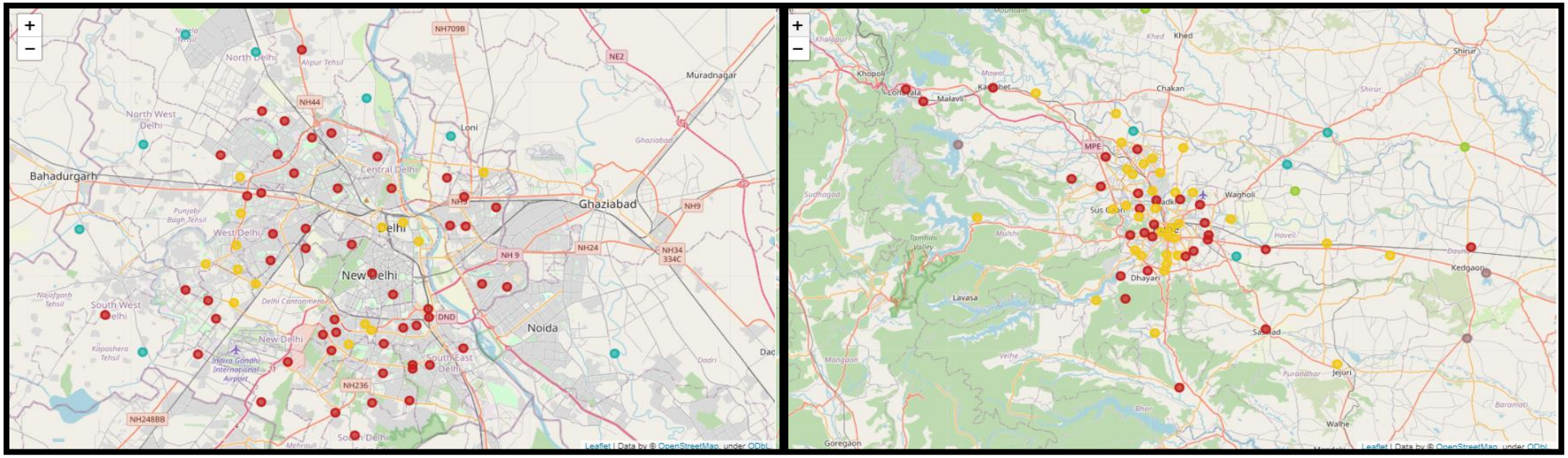
Clustering Results



Number of Locations in Each Locations.

The output of the clustering operation is 6 clusters with cluster labels 0, 1, 2, 3, 4and 5.

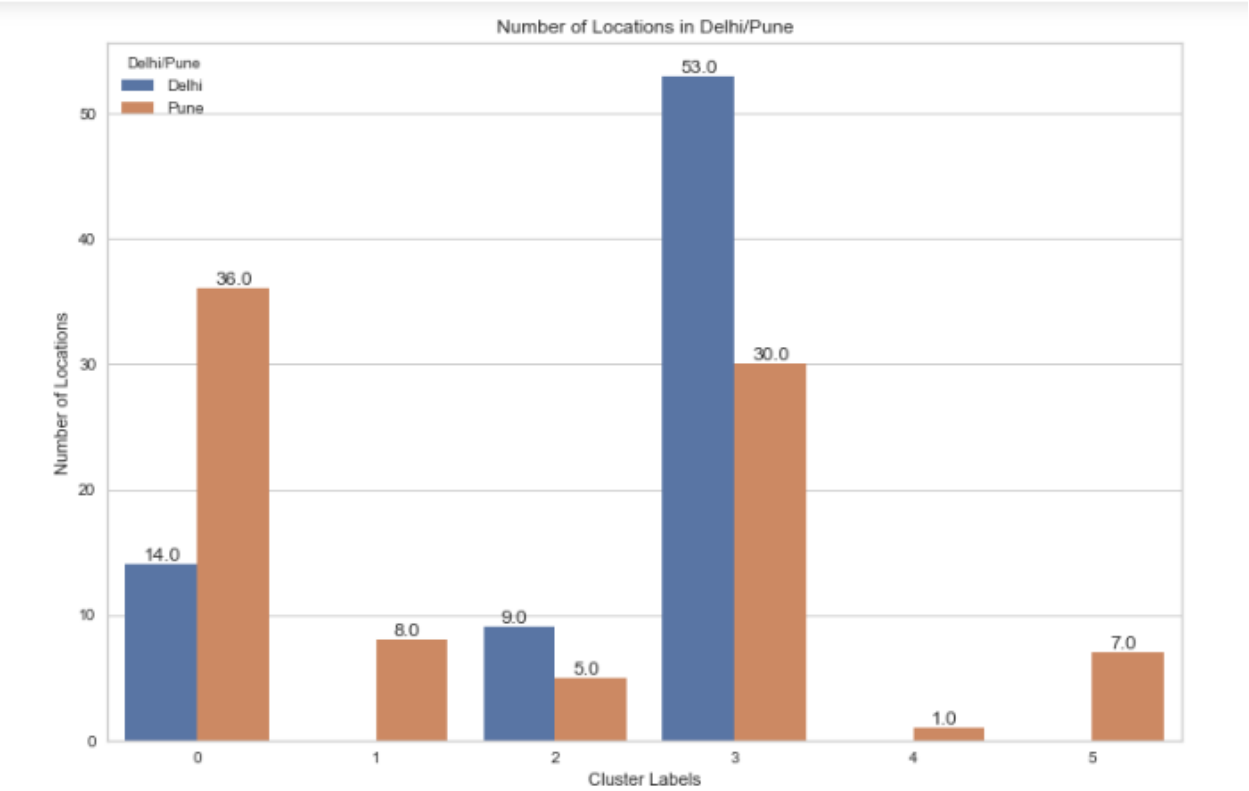
Each cluster is expected to contain a group of similar neighborhoods based on the categories of the venues in each neighborhood.



Delhi Map with Labeled Neighborhoods

Pune Map with Labeled Neighborhoods

- From above maps, clearly both locations have neighborhoods in common. Specially in cluster 3 and cluster 0.
- Coming to Dilshad Garden and Infotech park, both are clustered under cluster 3. This implies someone whose native place is Dilshad Garden, Delhi and comes to Pune, Maharashtra to pursue his/her career will not face many challenges in terms of adjusting to new place. Both places are almost similar.



- Cluster 0 is dominated by Locations Pune with 36 Locations and 14 Locations from Delhi.
- Cluster 3 is dominated by Location Delhi with 53 Locations and 30 locations from Pune.

- Out of 163 neighborhoods collectively, 143 neighborhoods are similar Between Delhi and Pune I.e. 90.19% similarity rate.
- 16 neighborhoods (cluster 1, 4 and 5) in Pune do not have any similar neighborhoods in Delhi.

Limitations and Recommendations

- In this project, we only considered clustering locations on the similarity of Venues. We didn't consider Population Density, Standard of living, Accessibility, Connectivity etc. Which can be crucial for clustering Locations.
- It will be recommended for future research purpose to devise a methodology to estimate such data to be used in clustering algorithm.
- This project made use of Sandbox Tier account of Foursquare API that came with limitations as to the API calls and results returned. Future research can make use of paid account to bypass these limitations.

Conclusion

In this project, the neighborhoods of Delhi and Pune were clustered into multiple groups based on the categories (types) of the venues in these neighborhoods.

The results showed that there are venue categories that are more common in some cluster than the others; the most common venue categories differ from one cluster to the other. If a deeper analysis—taking more aspects into account—is performed, it might result in discovering different style in each cluster based on the most common categories in the cluster.