

**Introduction:**

Mumbai is a densely populated city on the west coast of India. Capital city of the Indian state of Maharashtra, it is the financial, commercial and entertainment capital of the country. The city houses important financial institutions and the corporate [headquarters](https://en.wikipedia.org/wiki/List_of_Indian_companies_headquartered_in_Mumbai) of numerous [Indian companies](https://en.wikipedia.org/wiki/List_of_Indian_companies) and [multinational corporations](https://en.wikipedia.org/wiki/Multinational_corporations). It is also home to some of India's premier scientific and nuclear institutes. The city offers myriad employment opportunities and is the destination of choice for people from all over the country who relocate here seeking better livelihoods, economic prospects and growth opportunities.

**Problem Statement:**

This project aims to analyze various neighbourhoods in the city of Mumbai and offer a comparison between them in terms of amenities available. This can be extremely useful for individuals or families looking to relocate to the city and provides them with an objective analysis across multiple parameters to facilitate their choice of neighbourhood aligned with their needs and priorities.

**Data Source:**

All India Pin-code directory with contact details along with Latitude and longitude from the Open Government Data (OGD) Platform India. This contains all the pin-code list across India with geocodes and other relevant information. Latitude and longitude information for Mumbai neighbourhoods from the dataset is then used along with the Foursquare API to obtain information on venues within them.

**Methodology:**

We carry out clustering of the neighbourhoods using K-Means algorithm. The neighbourhoods are grouped into 3 clusters based on similarity of types of venues within them. Cluster labels are assigned to them and the data is loaded into a new dataframe including top 10 venues for each neighbourhood.

Finally, each of the clusters are examined. Depending on the types of venues within each cluster, final conclusions are drawn with regards to suitability of the neighbourhood depending on the requirements of the persons looking to relocate. Final results are then summarised.

**Results:**

The neighborhoods of Mumbai were grouped into 3 sample clusters on the basis of most common venue types within them.

Cluster 1 was observed to have more individual lifestyle options such as cafes, restaurants, bars as well as gyms and flea markets. These comprised of neighborhoods located more towards South Mumbai.

Cluster 2 was observed to have more commercial establishments. These comprised of neighborhoods located more towards Central Mumbai.

Cluster 3 was observed to have more retail stores, markets and fast food restaurants. These comprised of neighborhoods located more towards North Mumbai.

**Discussion: Observations & Recommendations**

Individuals or families can shortlist neighborhoods based on increased presence / availability of their preferred venues. Individuals may prefer the option of staying in Cluster 1 neighborhoods with greater options for eating out such as cafes, restaurants, bars as well as lifestyle options such as gyms and flea markets.

On the other hand, families may prefer to go with Cluster 3 neighborhoods that have more stores and fast-food restaurants.

**Conclusion:**

This methodology helps provide interested parties with the ability to review respective neighborhoods in a structured and grouped manner, thereby giving them the ability to narrow down on neighborhoods that suit their lifestyle requirements.

**Notebook URL:**

<https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/b3f3241d-426a-4b17-bdd4-4083d59095c3/view?access_token=f14e12844500e09200819e08eb18262c9b63d812d6d7dbabb070b1c21798db36>

**Presentation:**

****

**Technologies used:**

* IBM Cloud Watson Studio
* Jupyter notebook
* Foursquare API
* Language: Python 3.7

**Packages/Libraries used :-**

* numpy - (to handle data in a vectorized manner)
* pandas - (for data analysis)
* requests - (to handle requests)
* KMeans(sklearn.cluster) - (for K-means clustering)