

Battle of Neighbourhoods -PUNE & MUMBAI in India

1. Introduction / Business Problem

When thinking about relocating to a new city or country for work purposes or to start a new life, or to go for a holiday destination people tend to research areas before moving. This research includes population rate, average house price, school ratings, crime rates, weather conditions, recreational facilities, holiday destinations-tourism, Carnivals and Sports events/activity, Etc.

Based on the above, a search engine algorithm would be an efficient tool to use that will allow users to enter cities and get the neighbourhood name that best suits their lifestyle or living conditions.

In this project, we will study in detail the area classification using foursquare data and machine learning segmentation and clustering. And segment areas of two cities based on the most common places captured from Foursquare.

This could be done as the aim of this Project using an algorithm (Using segmentation and clustering) that will perform an extensive analysis on

1. The similarities and dissimilarities between the two cities of the user's search criteria,
2. Determine which city is best suited for lifestyle.

This project is based on a recommendation system using the Pune and Mumbai cities in India as my search criteria

1.Introduction of the cities

1. PUNE:

Pune formerly called Poona (1857–1978), is the second largest city in the Indian state of Maharashtra, after Mumbai. It is the ninth most populous city in the country with an estimated population of 3.13 million. In the 18th century, the city was the seat of the Peshwas, the prime ministers of the Maratha Empire and so was one of the most important political centres on the Indian subcontinent. Pune is ranked the number one city in India in the ease of living ranking index. The city is considered to be the cultural capital of Maharashtra. It is also known as the "Oxford of the East" due to the presence of several well-known educational institutions. The city has emerged as a major educational hub in recent decades, with nearly half of the total international students in the country studying in Pune. Research institutes of information technology, education, management and training attract students and professionals from India and overseas. Several colleges in Pune have student-exchange programs with colleges in Europe. Pune is also an important centre for civil services training.

2. MUMBAI

Mumbai is the capital city of the Indian state of Maharashtra. As of 2011 it is the most populous city in India with an estimated city proper population of 12.4 million. The larger Mumbai Metropolitan Region is the second most populous metropolitan area in India, with a population of 21.3 million as of 2016. Mumbai is the financial, commercial and entertainment capital of India. It is also one of the world's top ten centres of commerce in terms of global financial flow, generating 6.16% of India's GDP and accounting for 25% of industrial output, 70% of maritime trade in India (Mumbai Port Trust and JNPT), and 70% of capital transactions to India's economy. The city houses important financial institutions such as the Reserve Bank of India, the Bombay Stock Exchange, the National Stock Exchange of India, the SEBI and the corporate headquarters of numerous Indian companies and multinational corporations. It is also home to some of India's premier scientific and nuclear institutes like Bhabha Atomic Research Centre, Nuclear Power Corporation of India, Indian Rare Earths, Tata Institute of Fundamental Research, Atomic Energy Regulatory Board, Atomic Energy Commission of India, and the Department of Atomic Energy. The city also houses India's Hindi

(Bollywood) and Marathi cinema industries. Mumbai's business opportunities, as well as its potential to offer a higher standard of living, attract migrants from all over India, making the city a melting pot of many communities and cultures.

2,Data Requirements

Required data can be gathered from: - Pune and Mumbai City information, including districts and neighbourhoods, can be obtained from Wikipedia:

- (Source - <https://en.wikipedia.org/wiki/Mumbai>)
- (Source - <https://en.wikipedia.org/wiki/Pune>)

- Photos and Picture of Pune and Mumbai City used for Presentation from (Source <https://afrotourism.com/travelogue/>)

-The data used for this project will be acquired from <http://www.sapostalcodes.info>. The datasets consists of the postal codes and Location names of each city.

- In order to obtain venues and their categories we will use Foursquare API search feature [FOURSQUARE] (<https://foursquare.com/>) will be used to collect neighbourhood venue information as well as the longitude and latitude details of each suburb. Details about local venues and locality will provide insight into the qualities of a neighbourhood.

- In addition to Foursquare, various python packages will be used to create maps and machine learning models to gather further insights and provide efficient recommendations and results into our neighbourhood battle project.

This package includes:

1. Pandas - Library for Data Analysis
2. NumPy – Library to handle data in a vectorized manner
3. JSON – Library to handle JSON files
4. Geopy – To retrieve Location Data
5. Geocoder - For geolocation of neighborhoods

6. Requests – Library to handle http requests

7. Matplotlib – Python Plotting Module

8. Sklearn – Python machine learning Library

9. Folium – Map rendering Library

Basic Work Flow followed as:

♣ HTTP requests would be made to this Foursquare API server using postal codes of Port Elizabeth Suburbs and Durban Suburbs to pull out the latitude and longitude which will be used for creation of the map as well data analysis.

♣ Using credentials Foursquare API search feature would be enabled to collect the nearby places of the suburbs. Due to http request limitations, the number of places per suburb parameter would be set to 100 and the radius parameter would be set to 700.

♣ Folium- Python visualization library would be used to visualize the suburbs cluster distribution of Port Elizabeth and Durban over an interactive leaflet map.

♣ Extensive comparative analysis of two suburbs would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

♣ Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighbourhoods. These clusters from each of those two chosen suburbs would be analysed individually collectively and comparatively to derive the conclusions.

3. Methodology

A Jupyter Notebook developed in order to process data and segment the neighborhoods. Following steps are implemented:

Import Libraries

The notebook requires the following libraries. And we have installed it

1. Pandas - Library for Data Analysis

2. NumPy – Library to handle data in a vectorized manner

- 3.JSON – Library to handle JSON files
- 4.Geopy – To retrieve Location Data
- 5.Geocoder - For geolocation of neighborhoods
- 6.Requests – Library to handle http requests
- 7.Matplotlib – Python Plotting Module
- 8.Sklearn – Python machine learning Library
- 9.Folium – Map rendering Library 2.

1. Build neighbourhoods

List of Location and Postal code information is obtained from <http://www.sapostalcodes.info/> for Pune, and Mumbai city of India That list contains the names of the neighbourhoods for both the cities. As output a dataset containing a list of "city, suburb" is build.

2. Neighbourhoods geo graphical location

Every element in the neighbourhoods dataset is geographically located using Python Geolocator and two columns are updated Containing the latitude and the longitude coordinates of each city, neighbourhood. Also the Geographical coordinate of Pune, found out.

3. Find Geographical Coordinates and No of Locations

As a next stage, the Geographical coordinate of Pune and Mumbai city found out. Also the no of suburbs are found out for both the cities.

4. Venues compilation

As next step Foursquare services are used for obtaining venues for every neighbourhood. The output is a new dataset with many records for every neighbourhood containing the venues found for every one of them.

5. Neighbourhoods Segmentation

The problem in hand is a case of unsupervised segmentation and, from the possible machine learning algorithms, K-means was choosen. Taking in account that the venues information obtained from Foursquare is categorical, it must be previously processed in order to be handled by K-means algorithm. For this `_pandas.get dummies` is used for dummies variables. Therefore For this unsupervised machine learning algorithm K-mean clustering applied to form the clusters of different categories of places residing in and around the neighbourhoods.

Next step is built the segmentation data frame, composed of the top venues for every neighbourhood plus a segment label determined by K-means.

6. Segments analysis

Every segment is printed individually, were different characteristics can be observed for each group. These clusters from each of those two chosen suburbs would be analysed individually collectively and comparatively to derive the conclusions. Next section describes the results.

4. Results

1. Outcomes – Pune, India

The K-means method was used to cluster the Location of Pune city into 5 clusters. The details of the clusters are as follows:

1. Cluster 1: It contains 395 locations.

In this cluster coffee shops, pizza place, farm, Wine shop ,Departmental shops, Donut Shop, Electronica stores are present.

2. Cluster 2: It contains 143 locations.

In this cluster veg restaurants, Dance studios, coffee shops, Dessert Shop, Lake, Fast food, Tea room are available.

3. Cluster 3: It contains 13 locations.

In this very less options are available such as Departmental store,snack shop

4. Cluster 4: It contains 91 locations.

This Cluster includes Movie Theatres, Ice Cream Shops, Multiplex, Coffee Shops, Dance studio, Clothing Studio

5. Cluster 5: It contains 156 locations.

This cluster includes maximum things like Movie Theatres, Ice Cream Shops, Multiplex, Coffee Shops, Dance studio, Clothing shop, Departmental store,snack shop, pizza place, farm, Wine shop

2. Outcomes – Mumbai, India

The K-means method was used to cluster the Location of Mumbai city into 5 clusters. The details of the clusters are as follows:

1. Cluster 1: It contains 10 locations.

In this cluster coffee shops, pizza place, farm, Wine shop ,Departmental shops, Fish shop ,Café, Asian restaurants, Chinese Restaurants, farmer Market, Smoke Shop Donut Shop, Electronica stores are present.

2. Cluster 2: It contains 20 locations.

In this cluster veg restaurants, Dance studios, coffee shops, Dessert Shop, Lake, Fast food, Tea room, Gym, Cocktail bar, Resort are available.

3. Cluster 3: It contains 10 locations.

In this very less options are available such as Airport, Smoke Shop, Italian Restaurant, Departmental store,snack shop

4. Cluster 4: It contains 9 locations.

This Cluster includes Movie Theatres, Ice Cream Shops, Multiplex, Coffee Shops, Dance studio, Clothing Store, Pizza Shop, Bakery, Music Venue,Indian Restaurants, Juice Bar, Irani café,

5. Discussions

Pune has 47 locations with 15 venues. In addition, the geographical coordinate of Pune, India are 18.5203062, 73.8543185. and The best Location to stay in is ANDGAON with the following venues:

Neighborhood	Andgaon
1st Most Common Venue	Asian Restaurant
2nd Most Common Venue	Fast Food Restaurant
3rd Most Common Venue	Bank
4th Most Common Venue	Shoe Store
5th Most Common Venue	Donut Shop
6th Most Common Venue	Wine Shop
7th Most Common Venue	Electronics Store
8th Most Common Venue	Concert Hall
9th Most Common Venue	Dance Studio
10th Most Common Venue	Deli / Bodega

Mumbai has 47 locations with 15 venues. In addition, the geographical coordinate of Mumbai , India are 18.9387711, 72.8353355. And the best Location to stay in is LAL BANGH wh o's pincode is 400012 with the following venues:

Neighborhood	Lal Bangh
1st Most Common Venue	Indian Restaurant
2nd Most Common Venue	Gym / Fitness Center
3rd Most Common Venue	Chinese Restaurant
4th Most Common Venue	Stadium
5th Most Common Venue	Recording Studio
6th Most Common Venue	Bus Station
7th Most Common Venue	Bar
8th Most Common Venue	Playground
9th Most Common Venue	Sporting Goods Shop
10th Most Common Venue	Restaurant

6. Conclusions

Mumbai has a significant more number of Locations and venues than Pune therefore it would be the better option to relocate to Mumbai, specifically LAL BANGH as the most efficient choice. Mumbai offers a variety in choices for restaurants, gyms, grocery stores, and Water Park, golf , Sporting shops, playground, Bus Station , Stadium, course extracurricular activities for individuals and families.