DETERMINING THE PLACES FOR ACCOMODATION IN MUMBAI

1. Introduction

1.1 Background

Mumbai is one of the most vibrant cities of India. As of 2018, Mumbai was the most populous city of India. Furthermore, Mumbai ranks seventh in the world among the cities with most population. With around twenty million population, everything gets crowded. Also, Mumbai is the financial, commercial and entertainment capital of India.

This leads to the city having numerous corporates settling themselves in the city.

Moreover, Mumbai is situated on a narrow peninsula. This leads to it not being able to expand proportionally to the needs. This results in a much higher population density.

Hence the need for living peacefully and being able to obtain a liveable and healthy environment becomes one of the major priorities when searching for a proper neighbourhood to live in.

1.2 Problem

The problem addressed is the choice to find the correct neighbourhood to live in the city of Mumbai. Since Mumbai is a diverse city, one can easily be misled and find himself/ herself in an area not suitable to live in. The proposed solution is an analysis which identifies the neighbourhoods as one where one can find all the necessary areas in the surroundings of the house. Also, those areas are found where there isn't an environment to live there, rather they are commercial or entertainment areas. These areas would be more polluted, traffic prone and one might struggle to find a peaceful environment.

1.3 Interest

The target audience for this analysis are those who are looking for houses in various neighbourhoods, and want to know the areas which are popular in the neighbourhoods. Also, this analysis can be useful for existing residents, who want to shift or keep their house for sale. People could also use this to set or estimate the prices according to the luxuries and necessities present in the neighbourhood.

2. Data Overview

2.1 Data Source

The data was obtained by scraping the web page of Wikipedia. Link: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

2.2 Data Cleaning

The data was cleaned and preprocessed. All the boroughs were combined. The neighbourhoods were combined, and listed as a single comma separated attribute.

Since there was no latitude or longitude value for the boroughs, and no geojson file available, the mean of the latitude and longitude values of the neighbourhoods lying in the same borough were taken and listed as the latitude and longitude value of the borough.

2.3 Feature Selection

Feature selection was done by converting the popular areas around a neighbourhood in a 500m radius through one-hot encoding. Then the top ten locations were selected for each borough and used as the features for clustering.