

IBM Applied Data Science Capstone

INTRODUCTION:

Mumbai is the capital of India and can also be referred to as the city that never sleeps. Mumbai has the complete package such as its customs, lifestyles, and blend of culture. Mumbai is one of India's cosmopolitan cities because it is a financial powerhouse. Additionally, it is the epicenter of India's fashion industry and the origin of Indian Cinema.

Mumbai is located on Maharashtra's coast, and it is one of the most populated areas in the world. Mumbai has a diversified infrastructure. Mumbai has its advantages and disadvantages. Mumbai undergoes perennial problems such as air pollution, water pollution, overcrowding and much more. With such as diverse population, Mumbai has a diverse quality of living. The major problem that arises when one moves to a new city is finding a good area to prosper and thrive.

BUSINESS PROBLEM:

I aim to answer the below questions in this project:

1. Show and visualize the major parts of Mumbai with the top existing infrastructure.
2. Where are the finest locations in Mumbai as per infrastructure?
3. What areas have the most potential for the development of infrastructures?
4. Which areas in Mumbai lack infrastructure facilities?
5. Within Mumbai what is the finest place to stay for all vital infrastructure facilities?

TARGET AUDIENCE:

The main goal of this project is to assist individuals in exploring better facilities around Mumbai. Additionally, assisting people with making an efficient and smart selection on picking a good neighborhood out of the various postal codes within Mumbai.

There are tons of people migrating from many states of India and they need research for great housing prices, children friendly, new business etc. This project targets the individuals looking for a more efficient neighborhood; thus, helping to make people aware of the neighborhood prior to them moving to Mumbai for work or area of opportunity.

DATA DESCRIPTION:

Mumbai's demographics demonstrate that it is a very large and ethnically diverse city. With such a diverse culture, leads to a diverse infrastructure. There are many infrastructure types in Mumbai such as hospitals, colleges, supermarkets, hotels, etc.

Within this project I used the following data:

- Mumbai Pin codes were scraped from the below website:

Data Source: <https://mumbai7.com/postal-codes-in-mumbai/>

This website contains a list of the different pin codes, postal office names, and division name.

- Various types of infrastructures in each neighborhood within Mumbai.

Data Source: Foursquare API

Using the API to retrieve the different venues in each postal office. Then I will filter all the venues to obtain the different infrastructures and venues.

- GeoSpace Data:

Data Source: <https://github.com/geospace-code/pymap3d>

Using the Geospace data will provide the longitude and latitude coordinates of the postal office within Mumbai.

This project will demonstrate data exploration and provide the answers to the five above questions. Additionally, this project will showcase my data science skills which range from web scraping, working with API, data cleaning, data wrangling, data visualization and machine learning.