

Coursera Capstone

IBM Applied Data Science Capstone

Buying/Building an Independent Home in Hyderabad, India

By: Akshay Agnihotri
January 2021

Introduction

Hyderabad is one of the fastest developing cities of India, with a booming IT industry and various other opportunities. In such a scenario, it attracts lot of people from various other parts of the country, who come for work and want to settle in the city. At first, they start out in rented apartments, and then would want to either buy a flat or build their own house. Building a residential home, also has the added advantage of rental income and as the land value increases over time, independent housing property also can be seen as a good investment option. Of course, as with any business decision, buying a house/land requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the house is one of the most important decisions that will determine by how much the house/land value will increase over time.



Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Hyderabad, India to build a new house/land. Using data science methodology and machine learning techniques like clustering, this project aims to solve this and get an answer to the business question: In the city of Hyderabad, India, if a property developer is looking to build/buy a new house/land, where would you recommend that they do so?

Target Audience of this project

This project is particularly useful to individual working professionals, property developers, investors looking to buy or invest in independent house/land in the capital city of Hyderabad in the state of Telangana. This project is timely as the city's IT sector is booming.

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Hyderabad. This defines the scope of this project which is confined to the city of Hyderabad, the capital city of the state of Telangana in India.
- Latitude and longitude of these neighbourhoods. This is required for plotting the map and also to get the venue data.
- Venue data, particularly data related to Home (private) and Housing Development. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

- This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) contains a list of neighbourhoods in Hyderabad, with a total of 200 neighbourhoods. We will use web scraping techniques for extracting the data from the Wikipedia page, with the help of Python requests and beautiful soup packages.
- Then we will use Python Geocoder package to get the geographical coordinates of the neighbourhoods which will give us the latitude and longitude coordinates of the neighbourhoods.
- After that, we will be using Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Home (private) and Housing Development category in order to help us to solve the business problem put forward.
- This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).