

Capstone - The Battle of Neighborhoods Report

Contents

- 1 Introduction 2
 - 1.1. Background 2
 - 1.2. Problem..... 2
 - 1.3. Interest..... 3
- 2 Data Acquisition & Cleaning..... 3
 - 2.1 Data Acquisition 3
 - 2.2 Data Cleaning 4

1 Introduction

1.1. Background

I am an Indian and lot of my friends live in various parts of the world. I am taking this specific scenario based on my own experience because not just one or two, many friends asked me about places to live in Hyderabad since many of them wanted to move to the city. Of course, they are all in Information technology industry.

Since I have been living in Hyderabad for more than 10 years now, I have seen the surge in IT industries and number of small, mid and big IT companies like Amazon, Google, Microsoft, Deloitte, OnePlus to name a few etc. are expanding their footprint in the city of pearl and this is continuing in 2019 -2020.

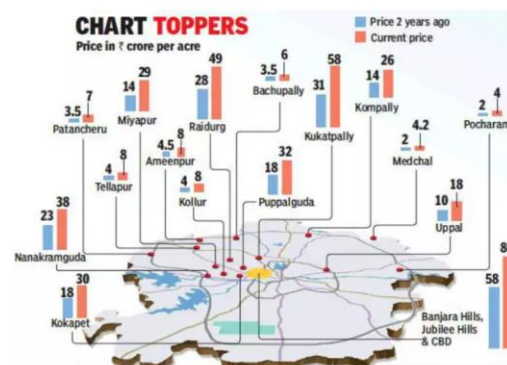
Now, because of this many people have either moved to planning to moving to Hyderabad for work and people who don't know much about the city either rely on their friends, if any, to know about the city and place they could reside which is close to office or having amenities of their interest close by or look for specific information on web site or even come to the city first and then start searching or rely upon brokers.

What if you are abroad and want to move to the city, wouldn't it be nice to have some key information about the neighborhoods in Hyderabad before hand to help you taking the decision on moving to the neighborhood of your interest?

Should you also invest in the city or may be open a business here....? Hyderabad has grown tremendously in last 10 years as evident from this chart (source – times of India) that shows the 100% growth in land prices in last two years and price movement published by 99acres.com.

Price movement in key micro-markets			
LOCALITY	AVERAGE 'ASK' RATE (PER SQ FT)	QOQ CHANGE	YOY CHANGE
Beeramguda	2,600-3,100	3%	15%
Manikonda	3,600-4,400	4%	14%
Kukatpally	4,100-5,400	4%	14%
Nizampet	2,800-3,500	2%	12%
Miyapur	3,700-4,600	4%	13%
Gachibowli	4,800-5,800	4%	10%
Kondapur	4,900-5,800	4%	15%
Chandanagar	3,600-4,800	4%	16%
Pragati Nagar	3,100-3,700	4%	7%
Nallagandla	4,900-5,700	5%	10%

Source – www.99acres.com



Source – timesofindia.indiatimes.com

1.2. Problem

This project aims to find specific places of interest in and around various neighborhoods of Hyderabad to enable a family or individuals moving to Hyderabad either from within India or from

any part of the world. We are going to need neighborhoods listing of Hyderabad and their coordinates information in order to call foursquare API for listing specific areas of interest. I'll be considering specific categories places of which would be listed in an around neighborhoods.

Category	Category ID
School	4bf58dd8d48988d13b941735
Hospital	4bf58dd8d48988d196941735
Pharmacy	4bf58dd8d48988d10f951735
Stationery Store	52f2ab2ebcbc57f1066b8b21
Indian Restaurant	4bf58dd8d48988d10f941735
Convenience Store	4d954b0ea243a5684a65b473
Chinese Restaurant	4bf58dd8d48988d145941735

These categories ID are taken from foursquare API documentations from here : <https://developer.foursquare.com/docs/resources/categories> . More categories can be added if needed in the code for listing addition

1.3. Interest

People who are considering relocating to Hyderabad will be interested to identify & explore its neighborhoods and common venues around each neighborhood in order to decide on a location based on categories listed.

2 Data Acquisition & Cleaning

2.1 Data Acquisition

The data required for this project is a list of Hyderabad neighborhoods with their latitude and longitude information.

The first source of this dataset is a scraped from Wikipedia page that contains the listing of neighborhoods in Hyderabad.

@link1 - https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India

Each neighborhood listed on this page points to neighborhood information page having more information about the neighborhood including its latitude and longitude information. For example, more information of "Abids" neighborhood is linked from the above to its own page –

@link2 - <https://en.wikipedia.org/wiki/Abids>.

Using these two web pages data will be scrapped and dataset with the following columns would be created for Hyderabad city.

- **Neighborhood** – Name of the neighborhood in Hyderabad
- **Latitude** – Latitude of the neighborhood
- **Longitude** – Longitude of the neighborhood

2.2 Data Cleaning

The data cleaning would be done in two steps using **Beautiful Soup** library in python.

The first step would be to scrap the data from @link1 and stored in tabular format. After the web scrapping, string manipulation would be performed to get the neighborhoods names and their corresponding links information.

The second step would involve going through each neighborhood one by one and web scrapping the linked page to extract latitude and longitude information. For each neighborhood information page, we'll extract the page content and perform string manipulation to extract Latitude and Longitude Information for the final dataset.

Using the final dataset containing the neighborhoods in Hyderabad along with latitude and longitude information, we can find specific venues within 3km meter radius of each neighborhood by connecting to the Foursquare API. This would return a json file containing venues of all categories taken up to be listed in each neighborhood and converted to a data frame. The data frame will be used to create clusters using K-mean clustering algorithm to cluster similar neighborhoods together.