Coursera capstone

IBM Applied Data Science Capstone

Comparing and exploring neighbourhood of Bengaluru and Hyderabad

By

Ankit Kashyap

April (2020)

Introduction

This is the report for the capstone project for IBM Data Science Specialization containing 9 courses. This reports is composed of basic background and problem statement as required by week 4 assignment.

Both Bengaluru and Hyderabad are two major IT hub of India. Both of them house some of the biggest names in IT sector globally and are most sought out for start-ups. Also, these two are major metropolitan cities of South India beside Chennai. Obviously there have been war of supremacy about these two cities. This project would explore both of these cities and will analyse what is popular in these two cities. Also comparing them on different aspects of neighbourhood.

Business Problem

This post elaborates on a data science project that attempts to analyse the neighbourhoods in each of these two cities and tries to understand what is popular in them and what they have to offer to someone who is contemplating to make a choice on seeking a life in either of the metro cities.

The other idea is to compare the neighbourhoods about how similar or dissimilar are the.

The deciding factor can be how unique, serene are the neighbourhoods of these cities. This study can be used to decide place to start any particular business by comparing the vibrant environment for that particular business in given cities. The venue to open the business can also be inferred from this study of both cities.

Data

Nothing can be implied without the usage of data. So, to begin with this project w e need data about both of these cities. We need the data about the neighbourhoods of both these cities. Fortunately, Government of India provides us data with postal code for all India. The CSV file is provided under https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude.

After reading into pandas dataframe we will need to get only relevant data of these two cities from all India postal codes. After getting the required we may need to do some clean up, like dropping the unnecessary columns. Neighbourhood's with same pin codes will be combined. Also, there maybe row with *NaN* values which will be dropped as they aren't relevant.

The *Neonatim* library of *geopy* in python would help us to get the latitude and longitude of the neighbourhoods.

Foursquare API will be used to explore the neighbourhood of these two cities. Folium library will be used for visualization.