Bengaluru Real Estate Data Analysis

A Capstone Project for Applied Data Science Capstone – IBM/Coursera

Done by: Shravan Bharadwaj

Version: Week 4 Version 2

Data

Based on the problem stated in the Introduction file, following are the factors which influence in buying any plot of land in Bengaluru:

1. Neighborhood Site Price Data in Rupees per Square Feet for prices estimation.

2. Kaveri River Water Availability based on the supply of water in number of days in a week in each neighborhood.

   Note: River Kaveri is the major river which supplies water to Bengaluru. Other supplies of water would be small lakes and groundwater. Here Kaveri water availability is used in the analysis.

3. The question of how close are important venues like metro, hospital, grocery store, clothing store, bank etc. to the neighborhood, is a very important factor which comes in play while buying any plot of land.

Following are the data sources where data will be extracted:

1. Real Estate price data from websites like Makaan, MagicBrics, 99 Acres and Housing.com.

2. Kaveri River water availability based on the supply of water in number of days in a week from the website of Bangalore Water Supply and Sewage Board.

3. Bengaluru ward data from BBMP for names of the wards, latitudes and longitudes, population residing.

4. Venue data from Foursquare API to get the minimum distance of venues like metro, hospital, clothing store etc. from a given ward/neighborhood.

Data description:

1. House prices are estimated in Rupees per square feet in India. If a plot of land has 5000 Rs/ Sq. Ft. and it estimated area is 2400 Sq. Ft., then the total cost of land would be 2400 times 5000 which yields 12 Million rupees.

2. Water Supply of Kaveri River is based on the supply of number of days in a week. The Water board supplies the water for about 1-2 Hours on those particular days in Bengaluru. Water is stored in tanks and used for domestic purposes.

3. The wards data of Bengaluru contains the names of 198 wards/ neighborhoods. It has the names of the wards, coordinates of the wards and the population residing in the ward.

4. The Foursquare API will be used for getting the minimum distances of venues like metro, hospital etc. which will be used for clustering.

5. Bengaluru Geojson file is used for plotting choropleth maps. It contains the geographic details of all the wards in Bengaluru.

Given below is a brief description to solve the Problem:

1. Collect Data from all sources.
2. Clean the data.
3. Extract the venue details from Foursquare API.
4. Analyze the data.
5. Cluster using K-Means Clustering.
6. Plot choropleth maps.
7. Create Dashboard.
8. Analyze the clusters and get proper inferences.