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.

	Borough	Neighborhood	Latitude	Longitude	Population	Price	Kaveri
0	Yelahanka	Chowdeswari Ward	13.12171	77.58042	19626	3644	2
1	Yelahanka	Atturu	13.102805	77.560036	24020	3295	2
2	Yelahanka	Yelahanka Satellite Town	13.090986	77.58392	25782	3900	2
3	K.R. Puram	Vijnanapura	13.006063	77.66956	35087	5840	0
4	K.R. Puram	Basavanapura	13.016847	77.715454	22012	4000	0
5	K.R. Puram	Hudi	13.022375	77.70549	20700	3484.778	0
6	K.R. Puram	Devasandra	13.001798	77.689125	22057	3986	0
7	K.R. Puram	A Narayanapura	12.994474	77.672585	29420	5100	0
8	K.R. Puram	Vijnana Nagar	12.978493	77.68177	24757	4500	0
9	K.R. Puram	HAL Airport	12.956536	77.6715	33066	8783	0
10	Byatarayanapura	Dodda Bommasandra	13.056595	77.55869	21640	7500	2

Figure 1. Initial Cleaned Data in Excel. It contains Boroughs, Neighborhoods, Coordinates, Population, Price in Rs. /SQ. FT. and Kaveri water facility based on no. of days per week.

	Borough	Neighborhood	Latitude	Longitude	Population	Price	Kaveri
0	Yelahanka	Chowdeswari Ward	13.121710	77.580420	19626	3644	2
1	Yelahanka	Atturu	13.102805	77.560036	24020	3295	2
2	Yelahanka	Yelahanka Satellite Town	13.090986	77.583920	25782	3900	2
3	K.R. Puram	Vijnanapura	13.006063	77.669560	35087	5840	0
4	K.R. Puram	Basavanapura	13.016847	77.715454	22012	4000	0

Figure 2. Initial Data frame with Boroughs, Neighborhood, Coordinates, Population, Price in Rs. /SQ. FT. and Kaveri water facility based on no. of days per week.

	Unnamed: 0	Neighborhood	Latitude	Longitude	Price	Metro	Hospital	Grocery	School	University	Movie	Clothing	Sports	Bank	Kaveri
0	0	Chowdeswari Ward	13.121710	77.580420	3644.0	5000	1842	3133	525	1007	3427	2617	2408	1272	2
1	1	Atturu	13.102805	77.560036	3295.0	5000	1380	1441	1501	1942	2653	1788	2110	1879	2
2	2	Yelahanka Satellite Town	13.090986	77.583920	3900.0	5000	758	1250	975	807	1271	337	794	524	2
3	3	Vijnanapura	13.006063	77.669560	5840.0	1655	623	744	910	930	1568	727	854	929	0
4	4	Basavanapura	13.016847	77.715454	4000.0	5000	949	682	1065	1132	3026	2189	760	789	0

Figure 3. Data of minimum distances (in meters) of venues extracted from Foursquare API. After dropping 'Neighborhood', 'Latitude' ,'Longitude' and 'Unnamed: 0', the data frame will be used for clustering.

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.