DATA DICTIONARY

House Pricing Dataset

The data set contains 506 observations of house prices from different towns. Corresponding to each house price are data of 18 other features/variables on which the price of an observation is suspected to be dependent on.

The data set contains both categorical and Numerical variable data types.

NUMERICAL VARIABLES

There are eleven (16) columns having numerical values including the target variable.

Column Names	*Description*
price	Value of the house
crime_rate	Crime rate in that neighbourhood
resid_area	Proportion of residential area in the town
air_qual	Quality of air in that neighbourhood
room_num	Average number of rooms in houses of that locality
age	How old the house is in construction years
dist1	Distance from employment hub 1
dist2	Distance from employment hub 2
dist3	Distance from employment hub 3
dist4	Distance from employment hub 4
teachers	Number of teachers per thousand population in the town
poor_prop	Proportin of poor population in the town
n_hos_beds	Number of hospital beds per 1000 population in the town
n_hot_rooms	Number of hotel rooms per 1000 population in the town
rainfall	The yearly average rainfall in centimeters
parks	Proportion of land assigned as parks and green areas in the town

CATEGORICAL VARIABLES

There are three (3) columns having categorical values.

Column Names	*Description*
airport	Is there an airport in the city? (Yes/No)
waterbody	What type of natural fresh water source is there in the city (lake/ river/ both/ none)
bus_ter	Is there a bus terminal in the city? (Yes/ No)