

Dataset

Your client is the city council of Boston, MA. The council shared with you the attached data set and a data dictionary. The Boston city council leadership is interested in understanding the drivers behind the value of houses in Boston and are looking for data-driven recommendations on how they can increase the value of housing. They would like you to come up with hedonic models. They insist that the analysis be done in R.

Please prepare PPT clearly mentioning about the approach you have taken, insights derived and what the final recommendations are?

Variable	Description
CRIM	per capita crime rate by town
ZN	proportion of residential land zoned for lots over 25,000 sq.ft.
INDUS	proportion of non-retail business acres per town.
CHAS	Charles River dummy variable (1 if tract bounds river; 0 otherwise)
NOX	nitric oxides concentration (parts per 10 million)
RM	average number of rooms per dwelling
AGE	proportion of owner-occupied units built prior to 1940
DIS	weighted distances to five Boston employment centres
RAD	index of accessibility to radial highways
TAX	full-value property-tax rate per \$10,000
PTRATIO	pupil-teacher ratio by town
B	$1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town
LSTAT	% lower status of the population
MEDV	Median value of owner-occupied homes in \$1000

Hints

1) Try to clean the data by removing outliers ($\text{mean} \pm 3 \cdot \text{sd}$) or any other measure to decide on the boundary.