Website: <https://www.kaggle.com/c/boston-housing/overview>

Overview

[Description](https://www.kaggle.com/c/boston-housing/overview/description)[Evaluation](https://www.kaggle.com/c/boston-housing/overview/evaluation)

Housing Values in Suburbs of Boston

The medv variable is the target variable.

Data description

The Boston data frame has 506 rows and 14 columns.

This data frame contains the following columns:

*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*  
nitrogen 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 mean of 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.

*black*  
1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.

*lstat*  
lower status of the population (percent).

*medv*  
median value of owner-occupied homes in \$1000s.

Source

Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.

Belsley D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.