

Used Dataset Schema

Boston dataset - <https://www.kaggle.com/datasets/vikrishnan/boston-house-prices>

Feature	Description	Data Type
CRIM	Per capita crime rate by	Float
ZN	proportion of residential land zoned for lots over 25,000 sq.ft.	Integer
INDUS	Proportion of non-retail business acres per town	Float
CHAS	Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)	Boolean
NOX	Nitric oxides concentration (parts per 10 million)	Float
RM	Average number of rooms per dwelling	Float
AGE	Proportion of owner-occupied units built prior to 1940	Float
DIS	Weighted distances to five Boston employment centers	Float
RAD	Index of accessibility to radial highways	Integer
TAX	Full-value property-tax rate per \$10,000	Float
PTRATIO	Pupil-teacher ratio by town	Float
B	$1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town	Float
LSTAT	% lower status of the population	Float
MEDV	Median value of owner-occupied homes in \$1000s (The label)	Float

The French Motor Claims dataset - <https://www.kaggle.com/datasets/floser/french-motor-claims-datasets-fremtpl2freq>

Feature	Description	Data Type
IDpol	IDpol The policy ID (used to link with the claims dataset)	Integer
Exposure	The exposure period	Float
Area	The area code	Categorical
VehPower	The power of the car (ordered categorical)	Categorical
VehAge	The vehicle age, in years.	Integer
DrivAge	The driver age, in years (in France, people can drive a car at 18)	Integer
BonusMalus	Bonus/malus, between 50 and 350: <100 means bonus, >100 means malus in France.	Integer
VehBrand	The car brand (unknown categories)	Categorical
VehGas	The car gas, Diesel or regular	Categorical
Density	The density of inhabitants (number of inhabitants per km2) in the city the driver of the car lives in.	Integer
Region	The policy regions in France (based on a standard French classification)	Categorical
ClaimNb	Number of claims during the exposure period (The label)	Integer

