**Automobile.csv Metadata**

This data set consists of three types of entities: (a) the specification of an auto in terms of various characteristics, (b) its assigned insurance risk rating, (c) its normalized losses in use as compared to other cars. The second rating corresponds to the degree to which the auto is more risky than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is more risky (or less), this symbol is adjusted by moving it up (or down) the scale. Actuarians call this process "symboling". A value of +3 indicates that the auto is risky, -3 that it is probably pretty safe. The third factor is the relative average loss payment per insured vehicle year. This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/speciality, etc...), and represents the average loss per car per year.

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **symboling** | Risk factor rating of the car (from -3 to 3); higher values indicate higher risk. |
| **normalized-losses** | Normalized insurance risk losses; may have missing values. |
| **make** | Name of the car manufacturer (e.g., Toyota, Honda, BMW). |
| **fuel-type** | Type of fuel used (e.g., gas, diesel). |
| **aspiration** | Type of engine aspiration (e.g., standard, turbocharged). |
| **num-of-doors** | Number of doors on the car (e.g., two, four); may have missing values. |
| **body-style** | The general shape of the car (e.g., sedan, hatchback, wagon, convertible). |
| **drive-wheels** | Type of drivetrain (e.g., front-wheel drive, rear-wheel drive, all-wheel drive). |
| **engine-location** | Location of the engine (e.g., front, rear). |
| **wheel-base** | Distance between the front and rear axles (in inches). |
| **length** | Length of the car (in inches). |
| **width** | Width of the car (in inches). |
| **height** | Height of the car (in inches). |
| **curb-weight** | Total weight of the car without passengers or cargo (in pounds). |
| **engine-type** | Type of engine (e.g., dohc, ohcv, ohc, l, rotor). |
| **num-of-cylinders** | Number of cylinders in the engine (e.g., four, six, eight). |
| **engine-size** | Engine displacement (in cubic centimeters, cc). |
| **fuel-system** | Type of fuel injection system (e.g., mpfi, 2bbl, idi). |
| **bore** | Diameter of the engine cylinder bore (in inches). |
| **stroke** | Length of the piston stroke (in inches). |
| **compression-ratio** | Ratio of cylinder volume at bottom and top of the piston stroke. |
| **horsepower** | Power output of the engine (in horsepower, hp); may have missing values. |
| **peak-rpm** | Engine revolutions per minute (rpm) at peak horsepower; may have missing values. |
| **city-mpg** | Fuel efficiency in miles per gallon (MPG) for city driving. |
| **highway-mpg** | Fuel efficiency in miles per gallon (MPG) for highway driving. |
| **price** | Price of the car (in dollars); may have missing values. |

cols=['symboling', 'normalized\_losses', 'make', 'fuel\_type', 'aspiration', 'num\_of\_doors', 'body\_style', 'drive\_wheels', 'engine\_location', 'wheel\_base', 'length', 'width', 'height', 'curb\_weight', 'engine\_type', 'num\_of\_cylinders', 'engine\_size', 'fuel\_system', 'bore', 'stroke', 'compression\_ratio', 'horsepower', 'peak-rpm', 'city\_mpg', 'highway\_mpg', 'price']