

EDA: Assignment 1

Data Set Information

The car dataset contains information about automobiles in different categories. You will perform EDA on this dataset. The main aim of this analysis is to find the important features that help find a car's cost.

[Cars Dataset](#)

This data set consists of three types of entities:

- the specification of an auto in terms of various characteristics,
- its assigned insurance risk rating,
- its normalized losses in use as compared to other cars.

The second rating corresponds to the degree to which the auto is riskier than its price indicates. Cars are initially assigned a risk factor symbol associated with their price. Then, if it is riskier (or less), this symbol is adjusted by moving it up (or down) the scale. Actuaries call this process "symboling". A value of +3 indicates that the auto is risky, and -3 indicates it is probably 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/specialty, etc...) and represents the average loss per car per year.

Note: Several attributes in the database could be used as a "class" attribute.

Attribute Information

1. **symboling**: -3, -2, -1, 0, 1, 2, 3.
2. **normalized-losses**: continuous from 65 to 256.
3. **Make**: alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
4. **fuel-type**: diesel, gas.
5. **aspiration**: std, turbo.
6. **num-of-doors**: four, two.
7. **body-style**: hardtop, wagon, sedan, hatchback, convertible.
8. **drive-wheels**: 4wd, fwd, rwd.
9. **engine-location**: front, rear.
10. **wheel-base**: continuous from 86.6 to 120.9.
11. **length**: continuous from 141.1 to 208.1.
12. **width**: continuous from 60.3 to 72.3.
13. **height**: continuous from 47.8 to 59.8.
14. **curb-weight**: continuous from 1488 to 4066.
15. **engine-type**: dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
16. **num-of-cylinders**: eight, five, four, six, three, twelve, two.
17. **engine-size**: continuous from 61 to 326.
18. **fuel-system**: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
19. **bore**: continuous from 2.54 to 3.94.
20. **stroke**: continuous from 2.07 to 4.17.
21. **compression-ratio**: continuous from 7 to 23.
22. **horsepower**: continuous from 48 to 288.
23. **peak-rpm**: continuous from 4150 to 6600.
24. **city-mpg**: continuous from 13 to 49.
25. **highway-mpg**: continuous from 16 to 54.
26. **price**: continuous from 5118 to 45400.