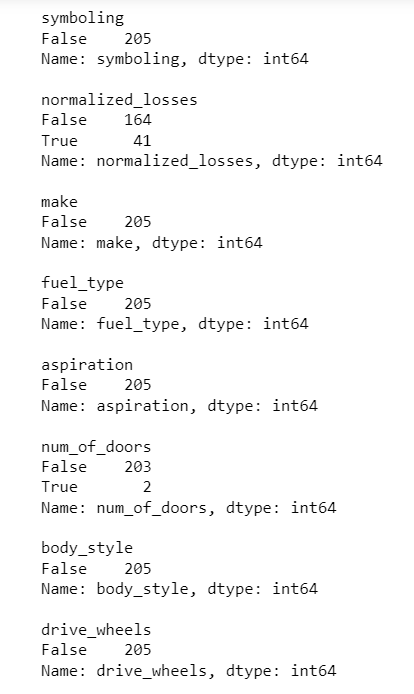
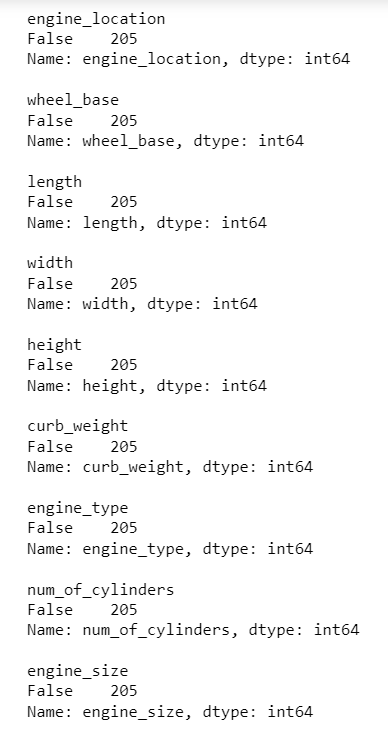
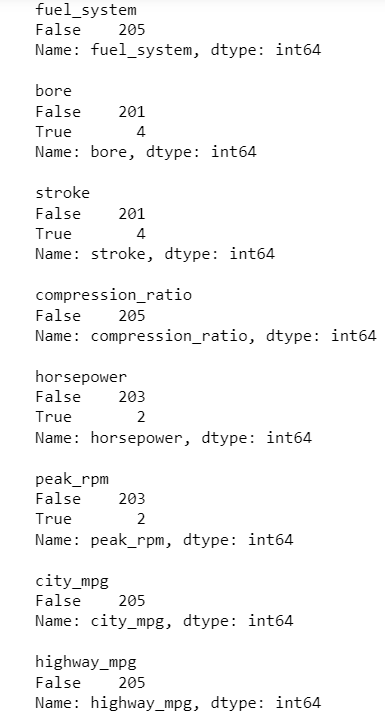
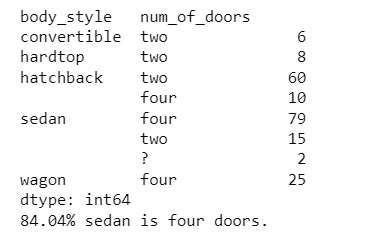
Аналіз даних

Під час дослідження отриманих даних, було вирішено видалити ціни з нульовим значенням, деякі втрачені дані замінити на середнє значення, а рядкові значення обробляти окремо.

Далі проаналізувавши типи кузова машин, ми знайшли втрачені дані для Седана, та вирішили замінити зниклу кількість дверей на 4. Отримали результат, що 84.4% Седанів мають 4 двері.



Кореляція

Після обробки та кореляції даних, із 26 значень залишилось 21 значення. Також можем виділити найбільш впливові на ціну значення.

* *Symboling has correlation 0.036, p-value is 0.6331842154850545*
* *Normalized losses has correlation 0.317, p-value is 1.4352759788069227e-05*
* *Make has correlation 0.005, p-value is 0.9465789197062073*
* *Num of doors has correlation 0.122, p-value is 0.10254970734462912*
* *Body style has correlation 0.085, p-value is 0.25421650817357294*
* *Drive wheels has correlation 0.565, p-value is 1.5008084787375338e-16*
* *Wheel base has correlation 0.612, p-value is 7.210669291835463e-20*
* *Length has correlation 0.716, p-value is 1.3237649365440924e-29*
* *Width has correlation 0.745, p-value is 4.483940735629761e-33*
* *Height has correlation 0.187, p-value is 0.012055937204577285*
* *Curb weight has correlation 0.838, p-value is 1.3532161301065677e-48*
* *Engine type has correlation 0.053, p-value is 0.4819564774757745*
* *Num of cylinders has correlation 0.155, p-value is 0.03756021162069747*
* *Engine size has correlation 0.725, p-value is 1.1540207363299491e-30*
* *Fuel system has correlation 0.671, p-value is 7.219421668533878e-25*
* *Compression ratio has correlation 0.022, p-value is 0.766261136862567*
* *Horsepower has correlation 0.769, p-value is 2.0073201202419802e-36*
* *Peak rpm has correlation 0.046, p-value is 0.5410315199111934*
* *City mpg has correlation 0.714, p-value is 2.591260715545443e-29*
* *Highway mpg has correlation 0.699, p-value is 1.0327686400384447e-27*

Із нашого результату отримаємо найбільш впливові на ціну значення.

*Curb weight has correlation 0.838, p-value is 1.3532161301065677e-4*

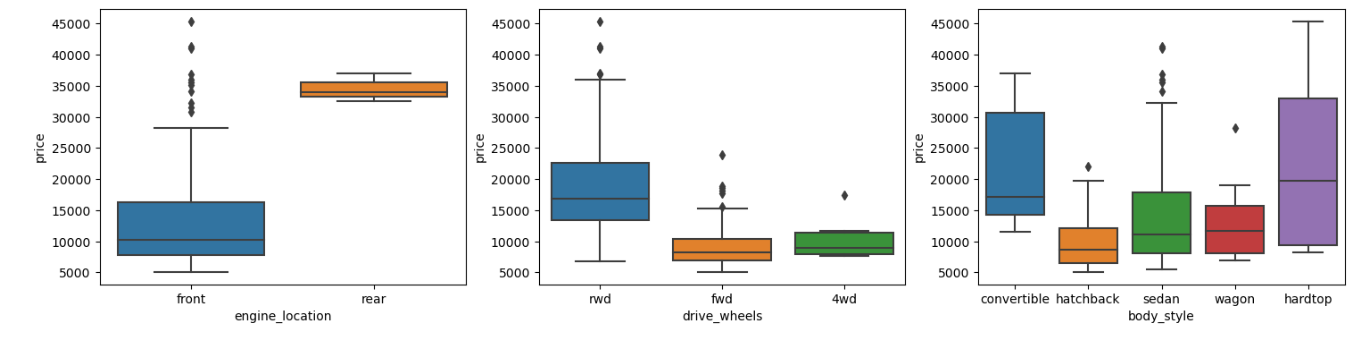
*Horsepower has correlation 0.769, p-value is 2.0073201202419802e-36*

*Engine size has correlation 0.725, p-value is 1.1540207363299491e-30*

*Highway mpg has correlation 0.699, p-value is 1.0327686400384447e-27*

*Fuel system has correlation 0.671, p-value is 7.219421668533878e-25*

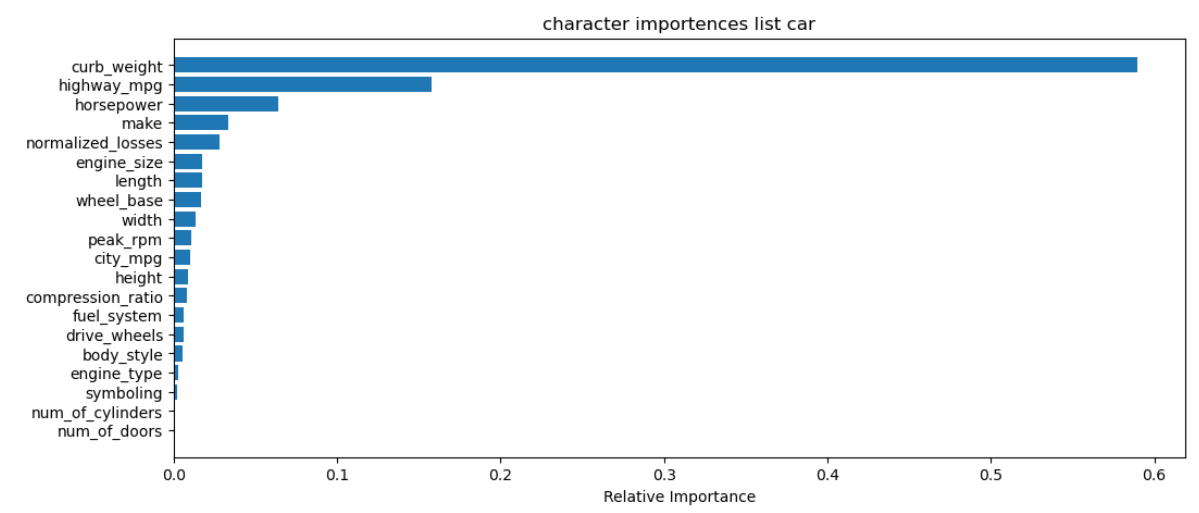
Візуалізація



Проаналізувавши графіки маємо висновки:

1. Машини, у яких двигун знаходиться спереду, знаходяться в ціновій категорії з 6 000 до 27 000. А більш всього машин мають ціновий діапазон із 8 000 до 16 000.
2. Машини, у яких двигун знаходиться позаду, знаходяться в ціновій категорії з 32 000 до 37 000. Та такі машини в більшості мають ціну від 34 000 до 36 000.

Використовуючи потужні деревовидні ML-моделі (RandomForest), розраховано показники важливості.



Проведено кореляцію даних, які ми відзначили як важливі

