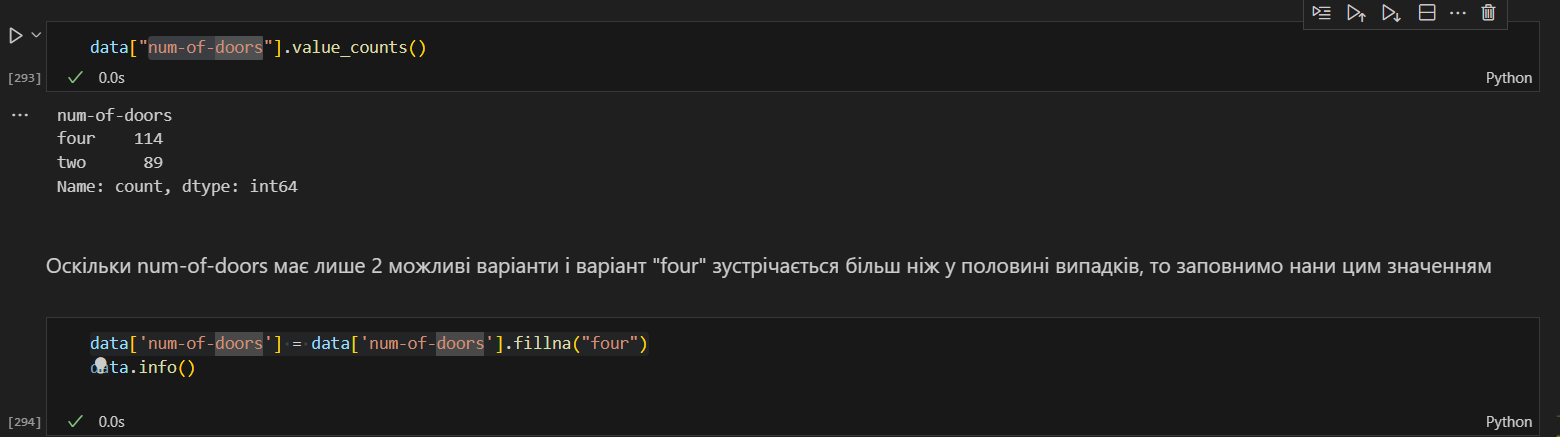
**Лабораторна робота №3**

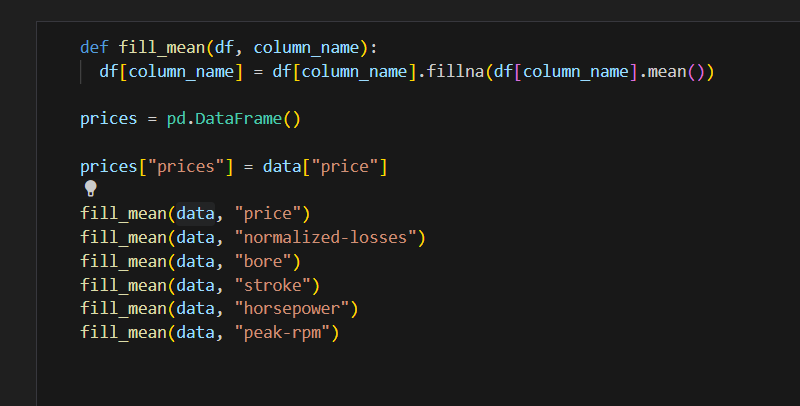
Методи аналізу та вибору значущих ознак

(Features’ Selection Procedures)

Перед початком роботи було очищено дані. Для цього у відсутніх значеннях числових стовбців було внесено середнє значення з усіх з цього стовбця, навіть для значення ціни. Для текстового стовбця було вибрано найбільш поширений варіант.

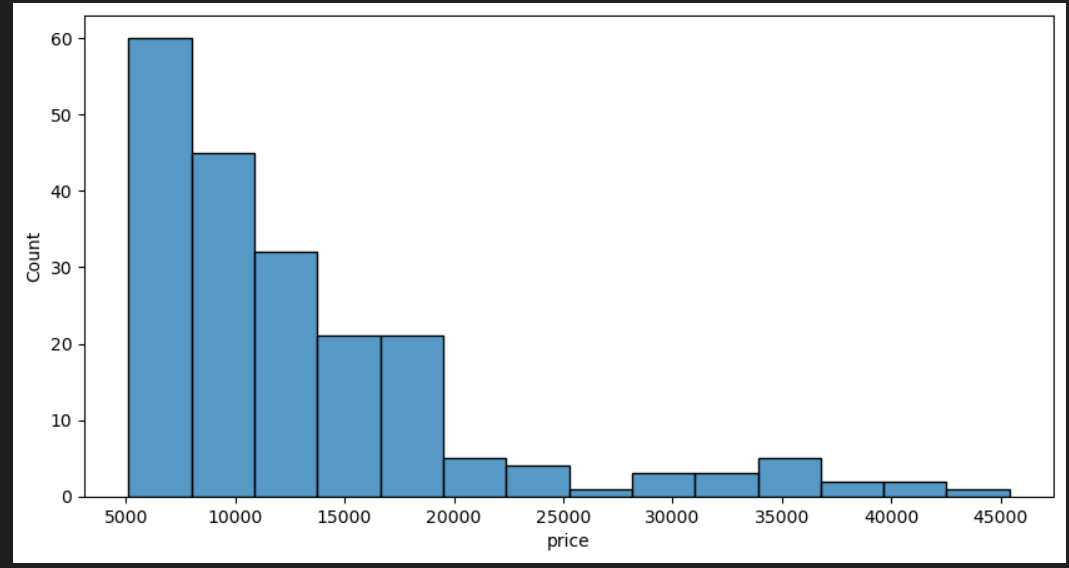


Заповнення стовбця з текстовими даними



Заповнення числових стовбців середніми значеннями

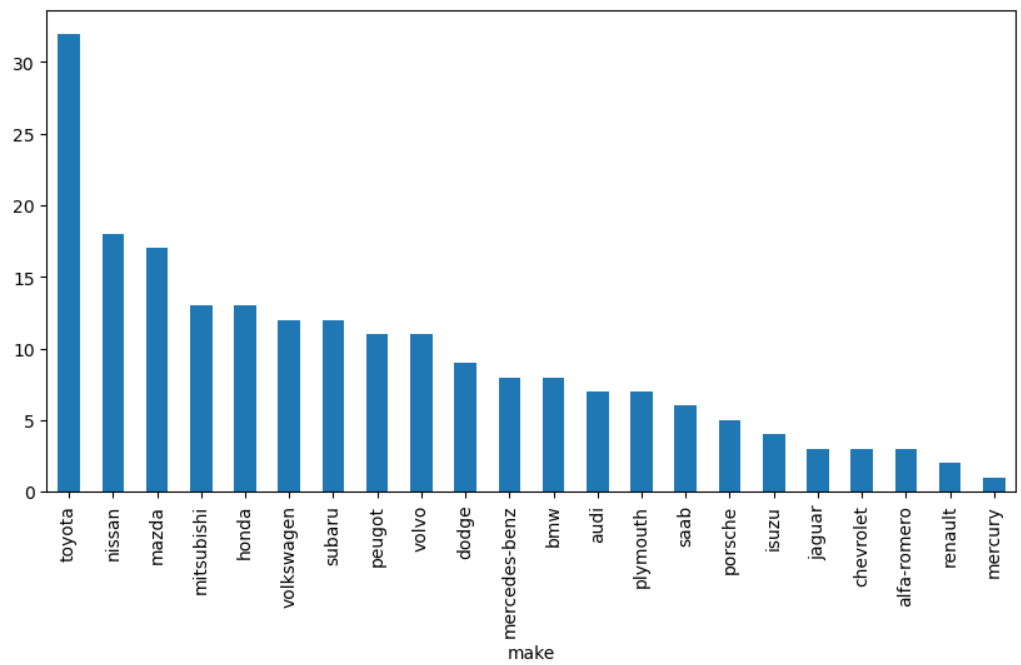
Аналізуємо ціну моделей автомобілів



Кількість моделей автомобілів від ціни

Оскільки в нас дані по середній ціні за деякі моделі автомобілів, а не по кожному окремому, то ми можемо з цього графіка зробити висновок, що найбільше дешевих моделей, а зі зростанням ціни їх кількість зменшується. Найбільший спад з 15к до 20к, до того ж моделей у ціновому сегменті з 25к до 30к найменша кількість (якщо не враховувати що моделей ціною у 45к і більше немає)

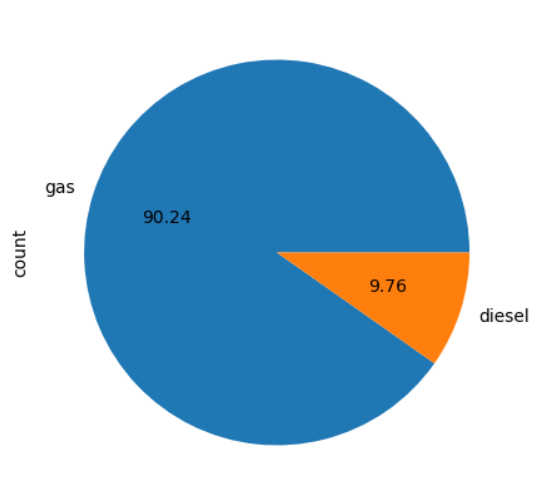
Проаналізуємо кількість моделей автомобілів за виробниками



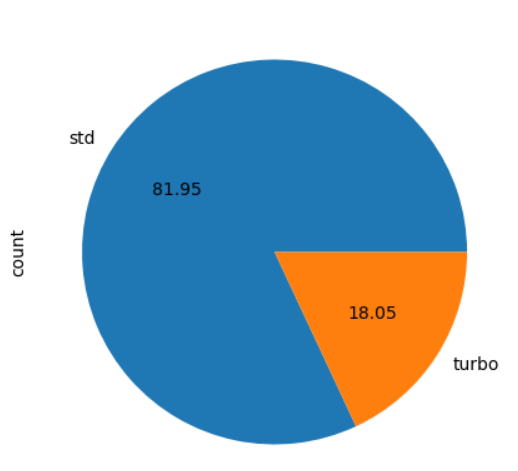
Кількість моделей автомобілів за виробниками

Найбільше моделей автомобілів виробила тойота, трохи менше виробили нісан і мазда. Найменшу кількість автомобілів виробили меркурі, ренаулт та альфа-ромеро.

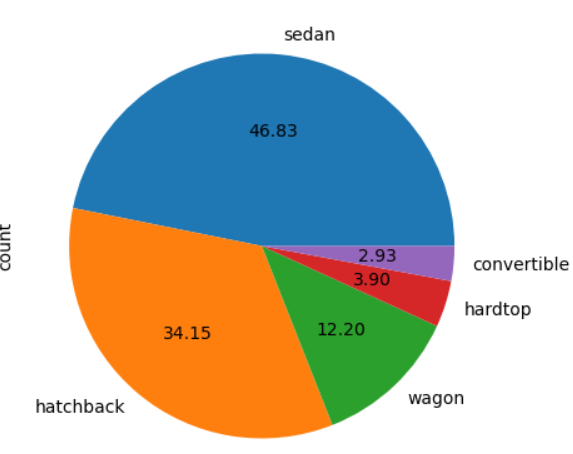
Кількість моделей з різними видами палива (у % від усього):



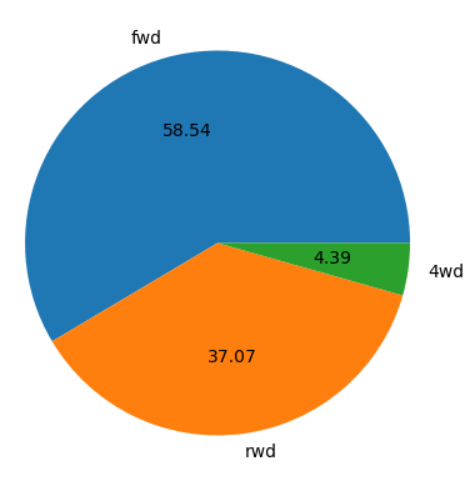
Моделі з видами aspiration (піддув? Наддув?):



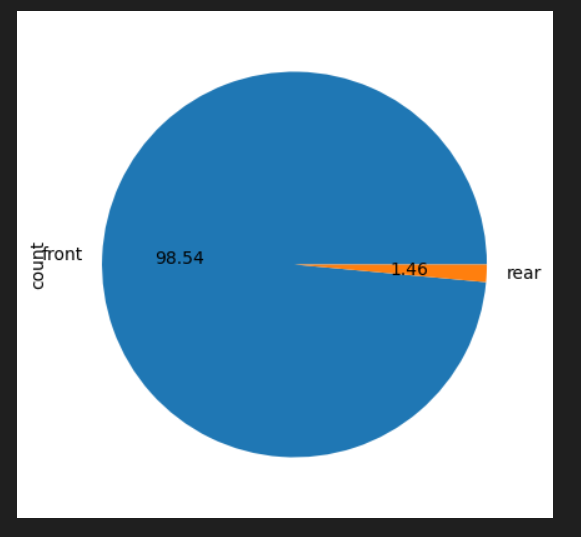
Моделі за типом кузова:



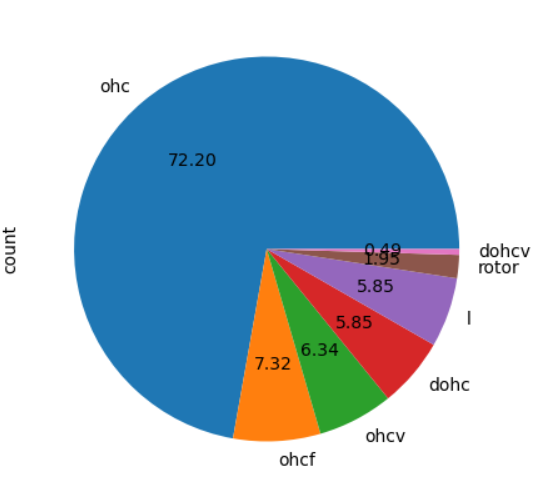
Моделі за керуючими колесами:



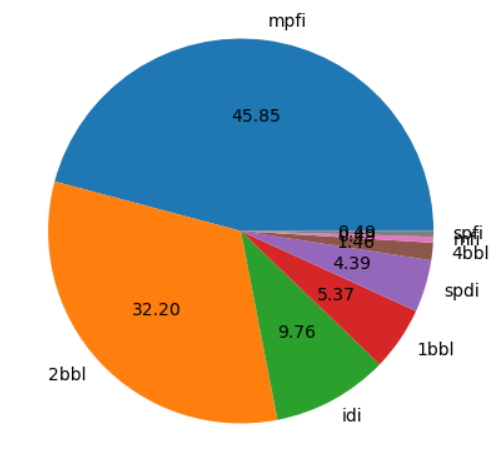
За розташуванням двигуна:



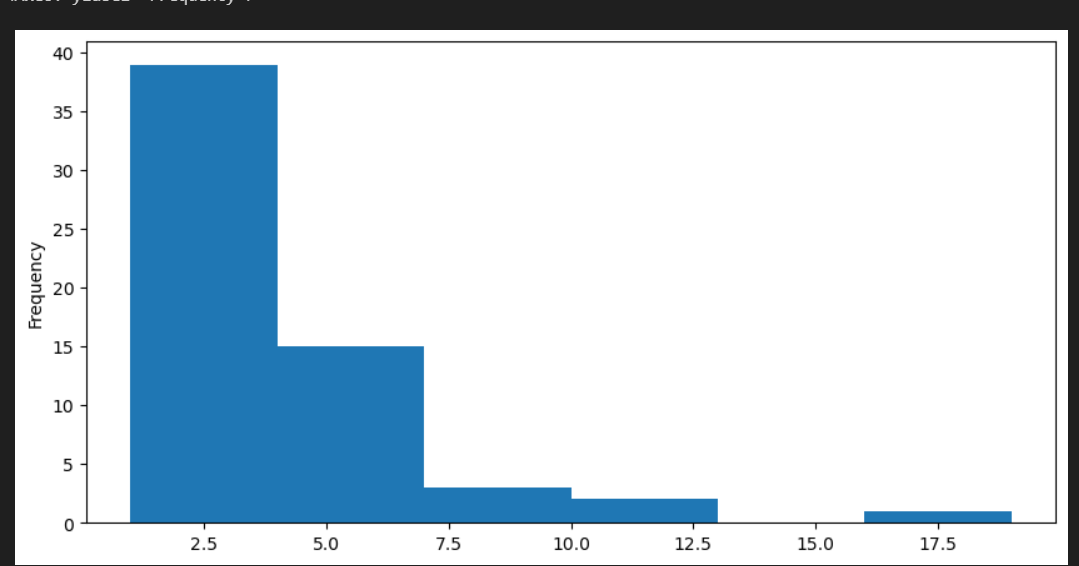
За типом двигуна:



За топливною системою?

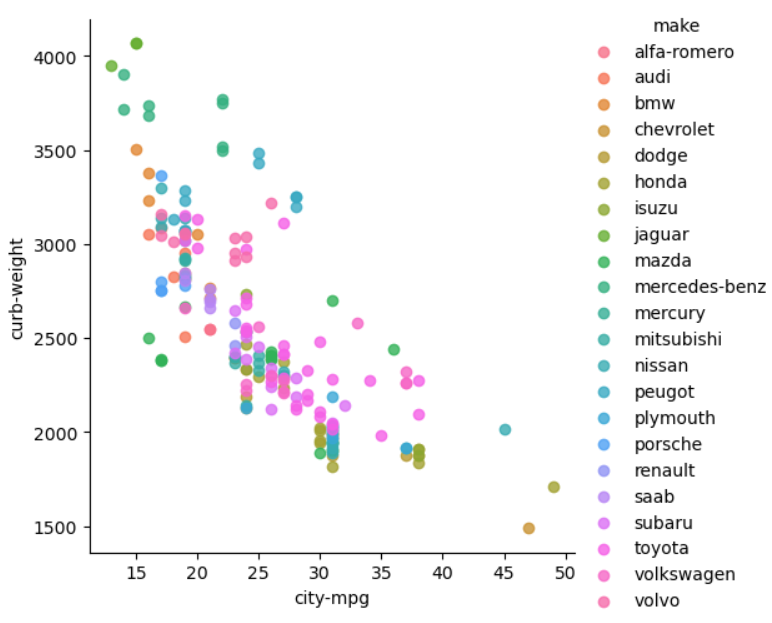


Частота, з якою автомобіль буде належати до одного з 6 класів за кількістю кінських сил:

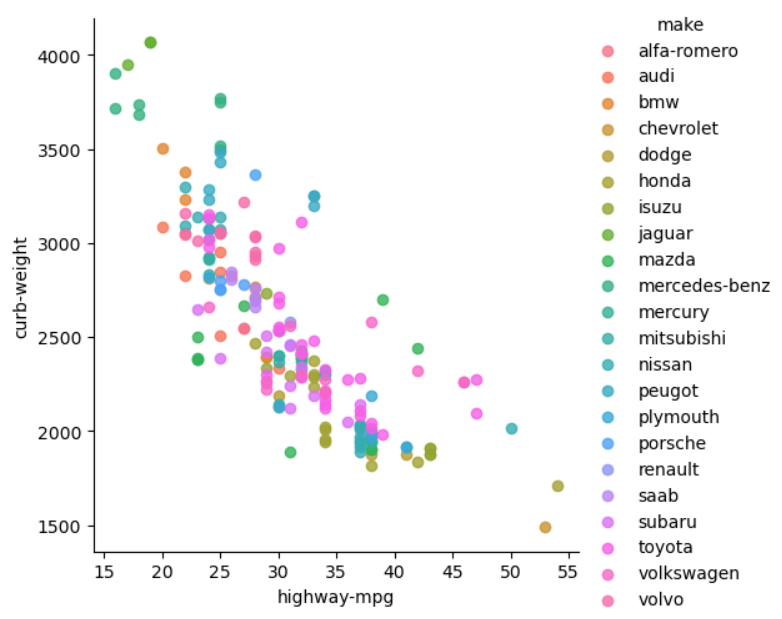


Як бачимо, найбільше моделей з найменшою кількістю кінських сил.

Графіки залежності максимальної швидкості від маси автомобіля:



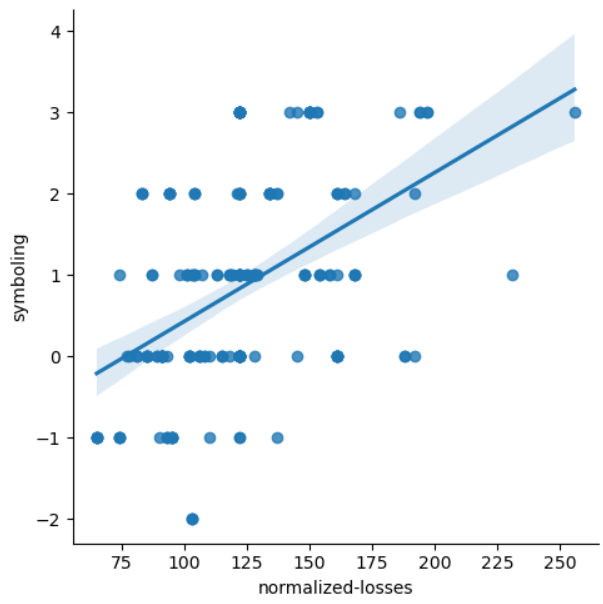
У місті



На швидкісних дорогах

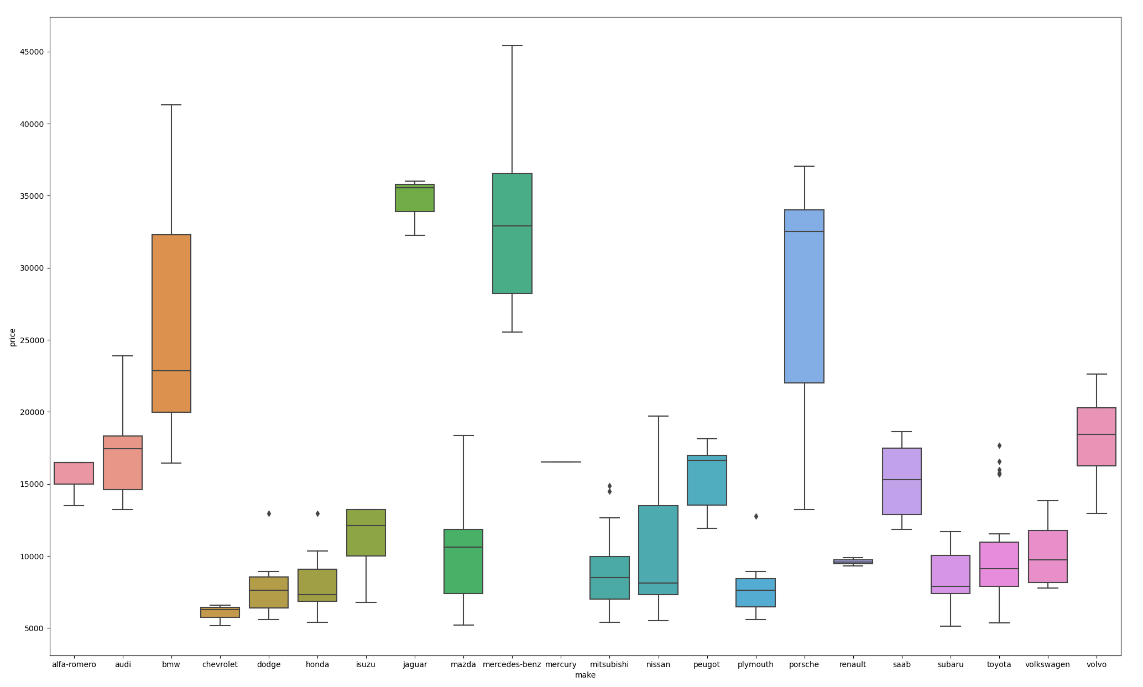
Видно, що чим менша вага, тим більша максимальна швидкість.

Графік витрат на автомобіль від класу страховки:



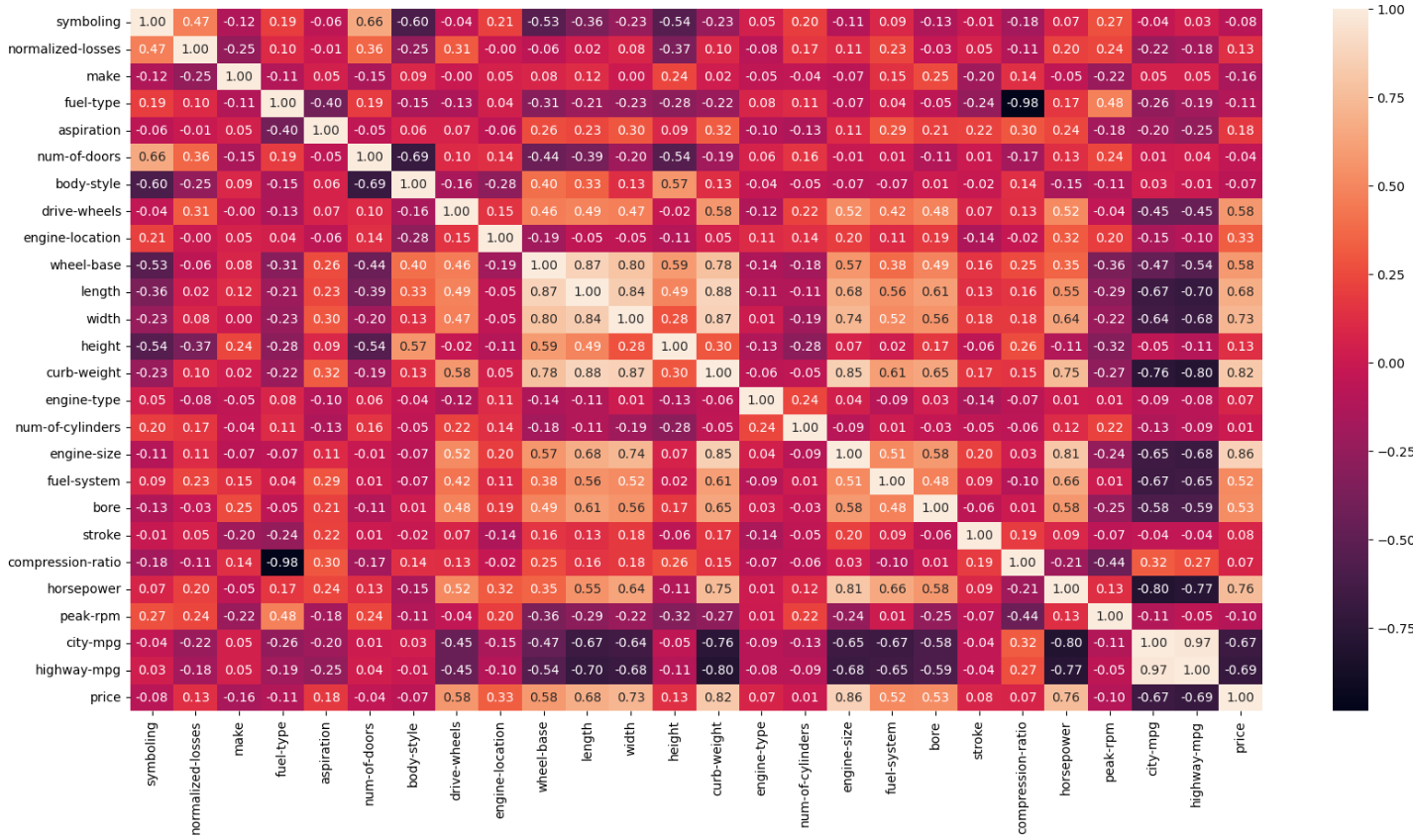
На графіку видно що чим більші витрати, тим більший клас страховки.

Графік залежності ціни автомобіля від його виробника:



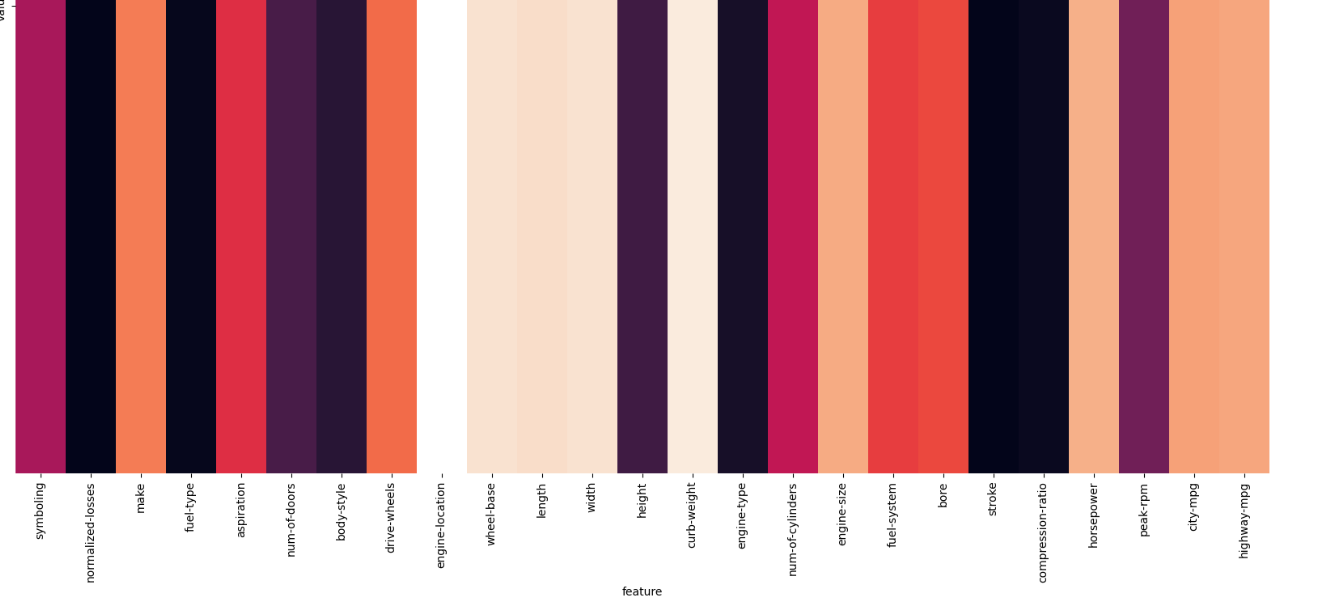
Меркурі виробляє автомобілі в одній категорії, судячи по графіку (або відсутні дані про інші моделі). Тойота виробляє автомобілі в ніші недорогих автомобілів. Бмв, порше, мерседес та ягуар виробляють дорогі моделі автомобілів.

Графік кореляції між параметрами:



Ми можемо побачити що "fuel-type" незалежить від "compression-ratio", "wheel-base", "length", "width", "height" та "curb-weight" корелюють між собою. "horsepower" залежить від "engine-size", "curb-weight", "price", "width" та "height" параметрів. "city-mpg" та "highway-mpg" дуже сильно корелюють одне з одним.

Аналіз Pearson Correlation по стовбцю та його впливу на ціну:

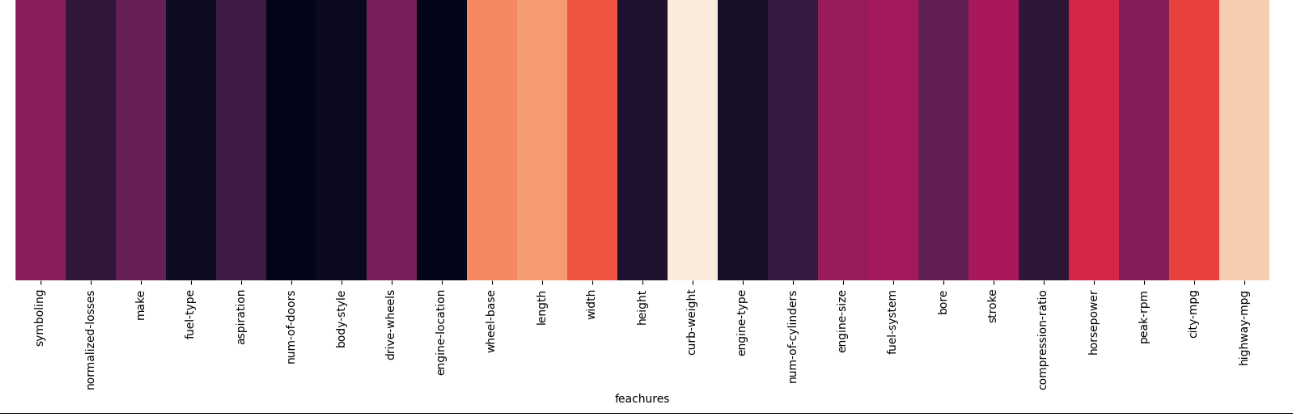


За результатом аналізу:

Найбільший вплив на ціну мають параметри wheel-base, length, width, curb-weight, engine-size, horsepower, city-mpg та highway-mpg

Найменше на ціну впливають peak-rpm, compression-ratio, stroke, engine-type, height, body-style, num-of-doors, fuel-type, make, normalized-losses, symboling

Mutual Information аналіз



За цим аналізом найбільше на ціну впливають параметри curb-weight, width, wheel-base, length, highway-mpg.

Найменше впливають на ціну параметри fuel-type, body-style, num-of-doors, engine-location