

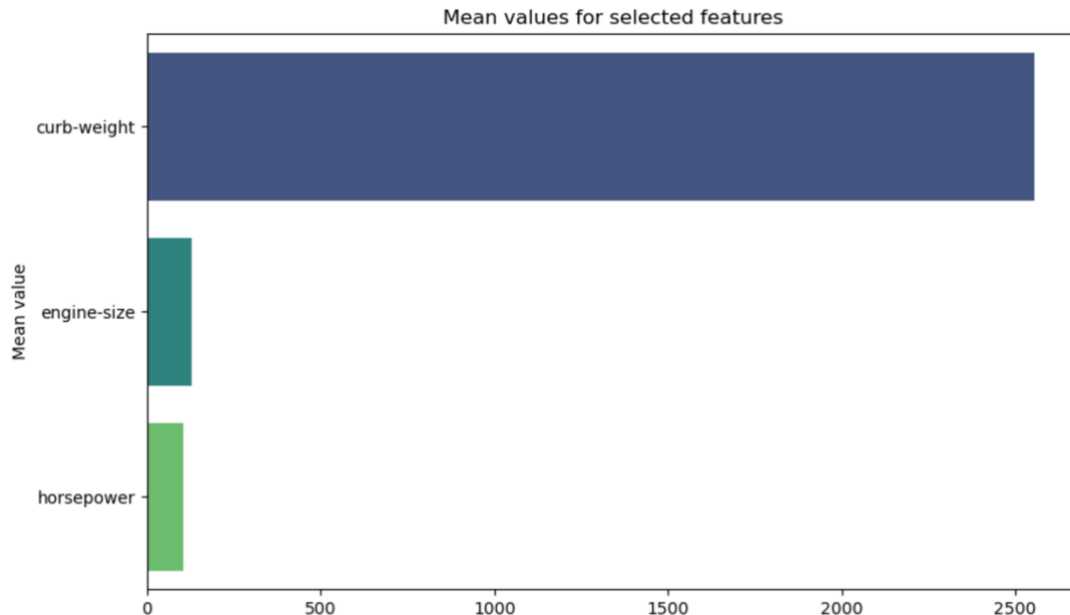
# Аналітичний звіт щодо факторів, які впливають на ціну підтриманих автомобілів в Америці.

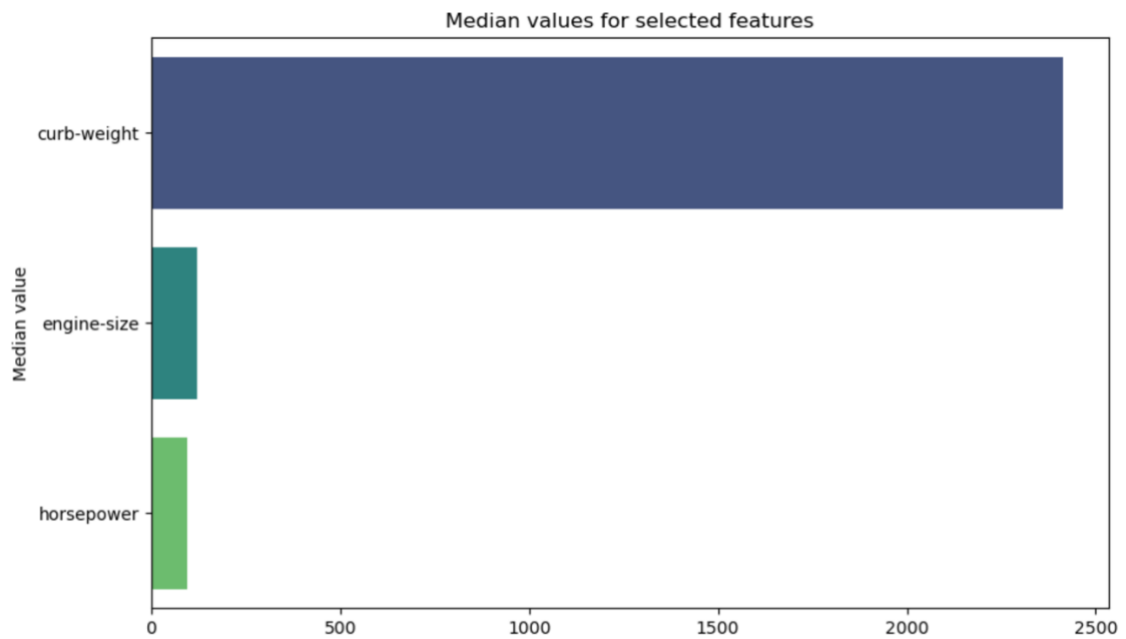
## *Вступ.*

Цей звіт аналізує дані з автомобільного щорічника Ward's Automotive Yearbook за 1985 рік для виявлення ключових ознак, які впливають на ціну підтриманих автомобілів в Америці. Для вибору значущих ознак були використані такі методи: однофакторний вибір ознак – кореляція Пірсона (задача регресії), багатофакторний вибір ознак – ReliefF, та Tree Based Method. Звіт містить також аналіз ознак на взаємозалежність, що відображено відповідними heatmap по кожному з трьох використаних методів дослідження.

## *Однофакторний вибір ознак.*

Метод кореляції Пірсона виділив наступні три ключові ознаки: curb weight, engine size, та horsepower. Всі три ознаки є кількісними, тому проаналізуємо їх mean, median:

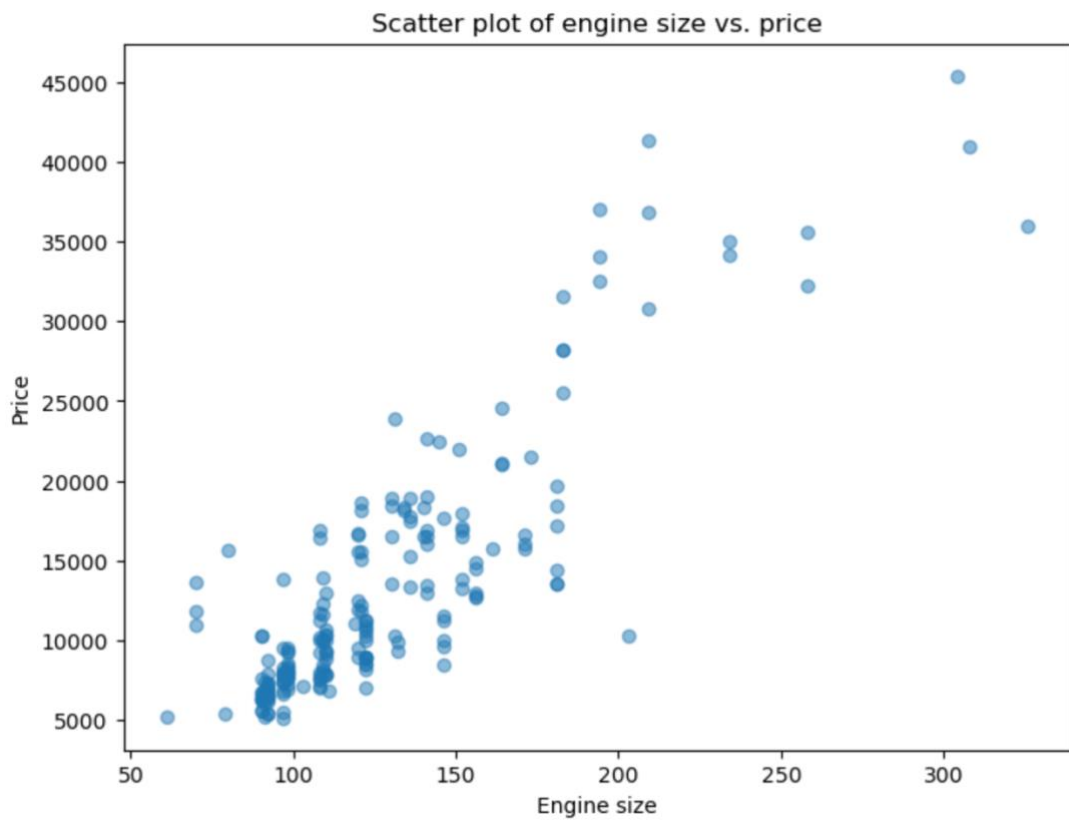




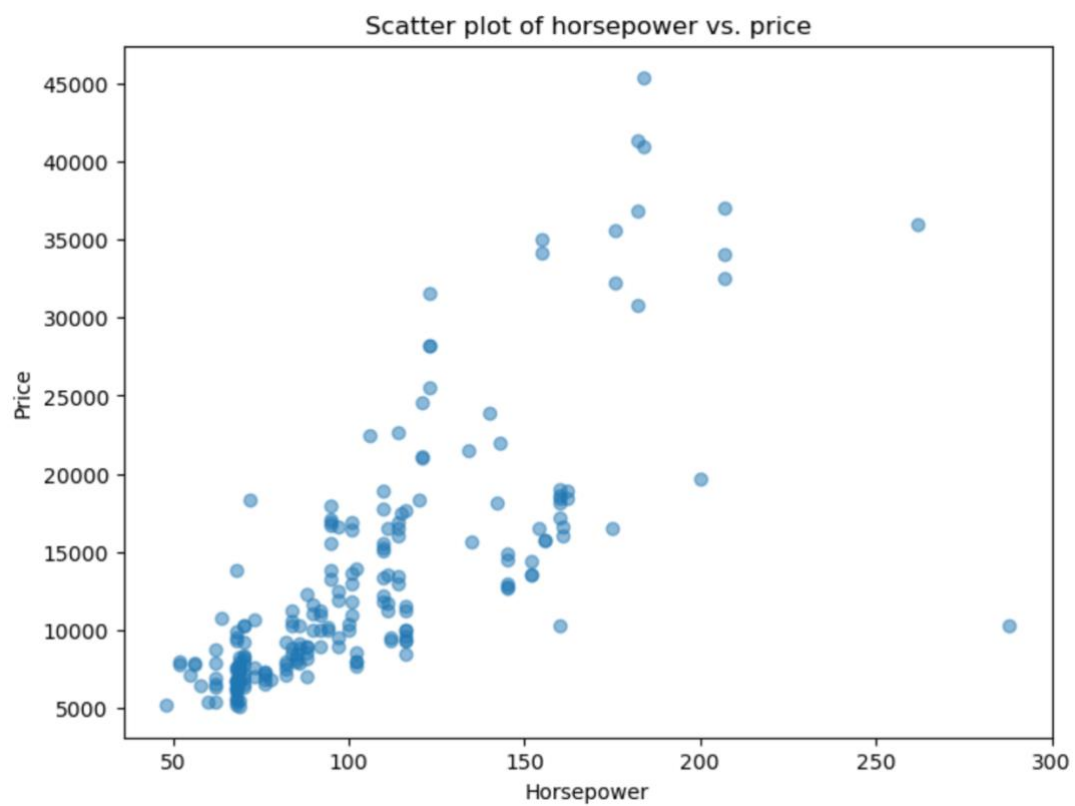
Тепер візуалізуємо кореляцію між кожним з ключових ознак та ціною.



Бачимо, що загалом кореляція між curb weight та ціною позитивна, що означає, що зі збільшенням ваги автомобіля, ймовірно, збільшиться його ціна.



Кореляція між engine size та ціною також загалом позитивна.



Також, кореляція між horsepower та ціною також позитивна.

Побудуємо heatmap, який показує аналіз ознак на взаємозалежність:

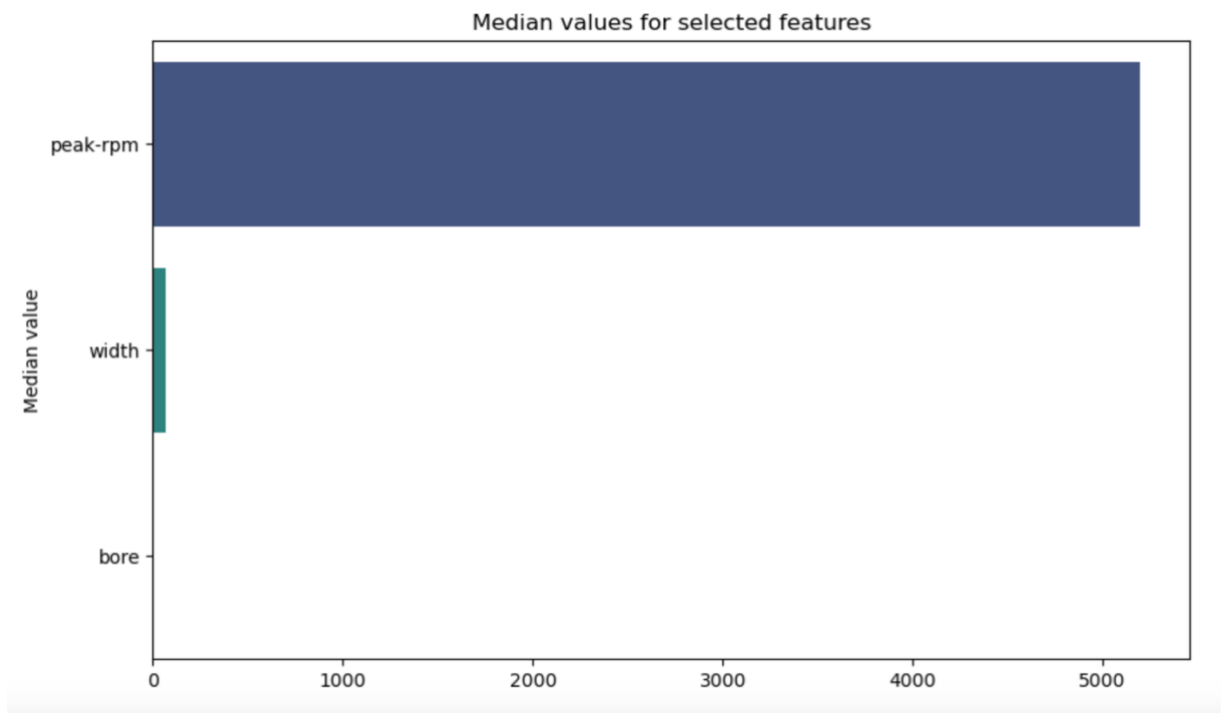
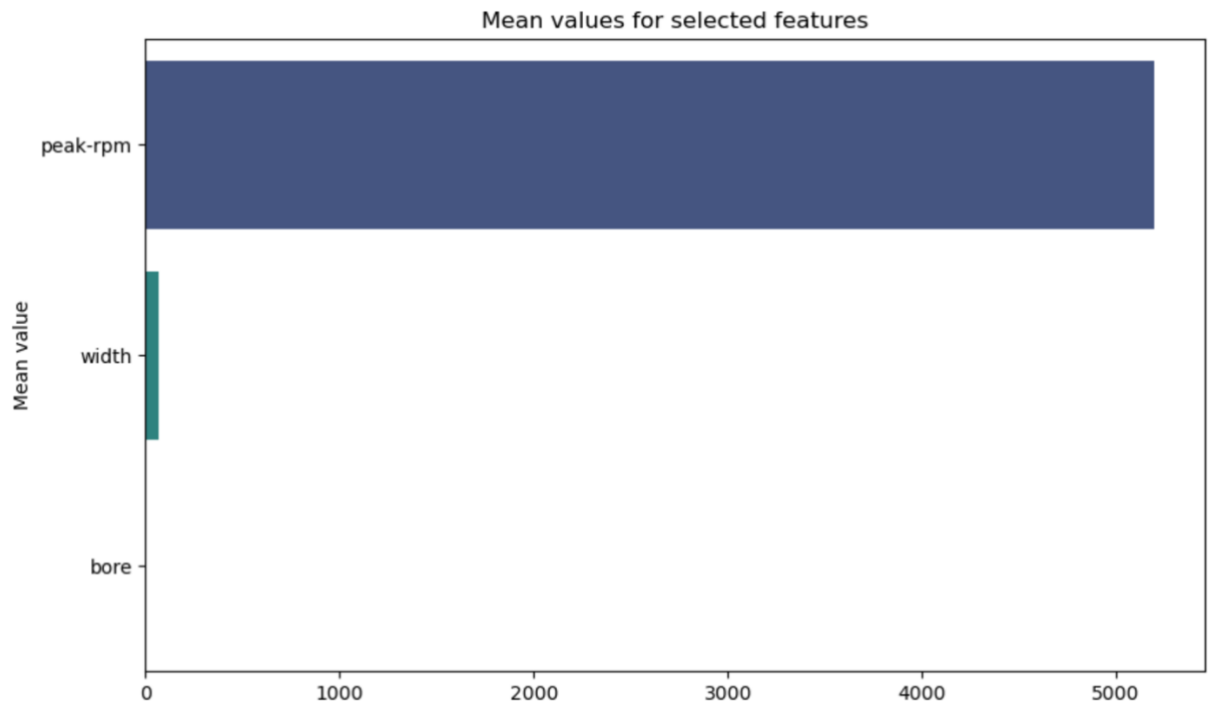


Бачимо, що всі ознаки мають позитивну кореляцію. Найбільшу з них має engine size (0.86), після чого йде curb weight (0.82) та horsepower (0.75).

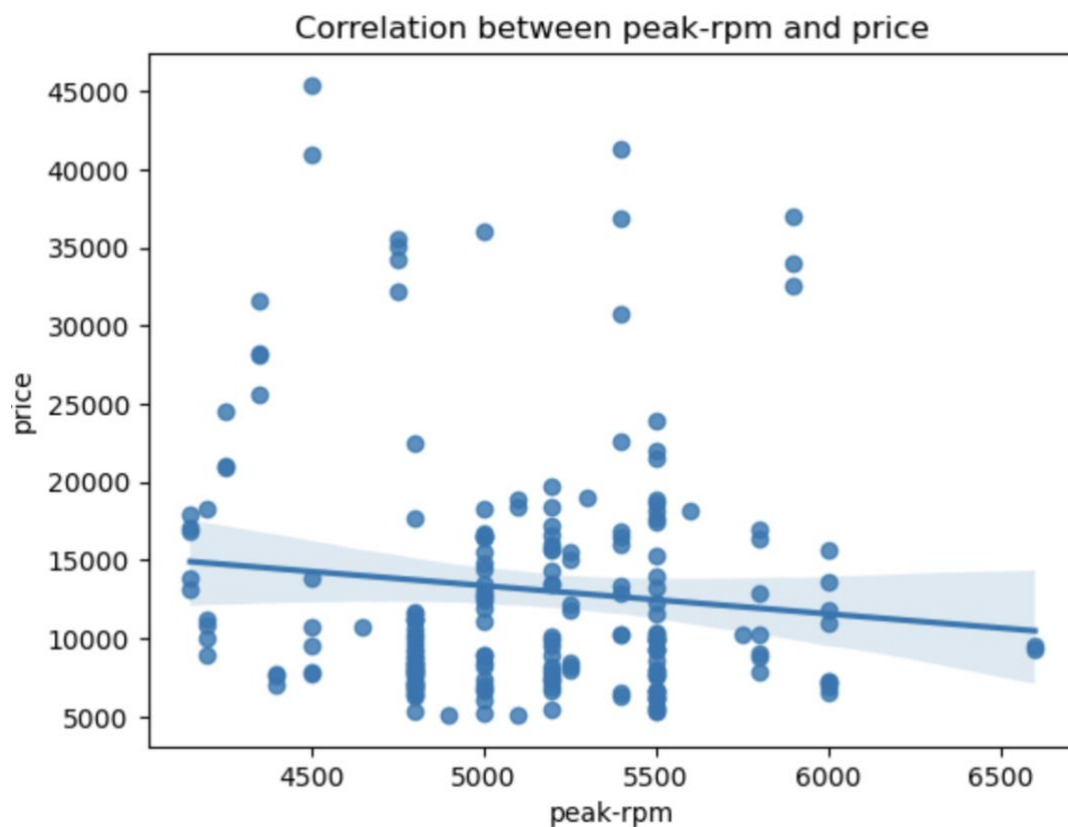
*Висновки:* Однофакторний вибір ознак виявив, що ключовими факторами, що впливають на ціну, є: curb weight, engine size, та horsepower. Після аналізу кореляції кожної ознаки з ціною було виявлено, що кореляція позитивна. Але, слід зазначити, що хоча позитивна кореляція вказує на те, що зі збільшенням однієї змінної імовірність збільшення іншої збільшується, це не означає причинно-наслідковий зв'язок. Наприклад, позитивна кореляція між вагою автомобіля і ціною може бути зумовлена тим, що важкі автомобілі часто мають більше додаткових функцій або є вищого класу, що може призвести до вищих цін. Однак це не завжди є причиною вищої ціни.

### ***Багатофакторний вибір ознак.***

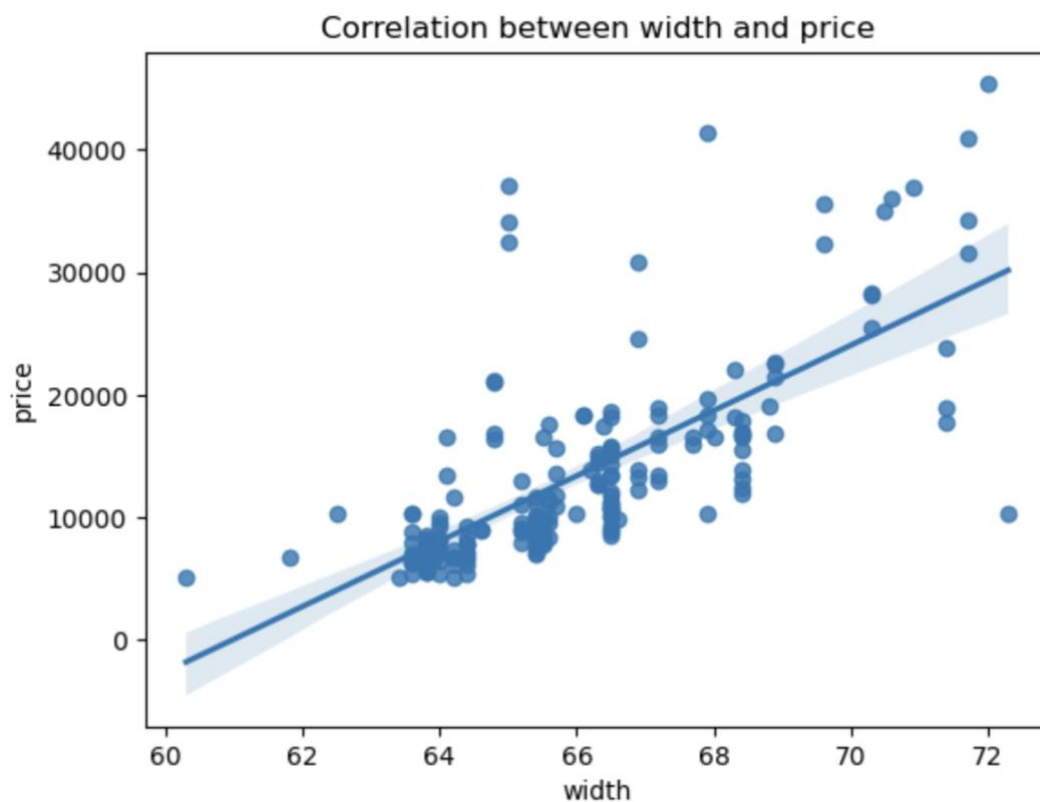
Метод ReliefF виділив наступні три ключові ознаки: peak rpm, width, bore. Всі три ознаки є кількісними, тому проаналізуємо їх mean, median:



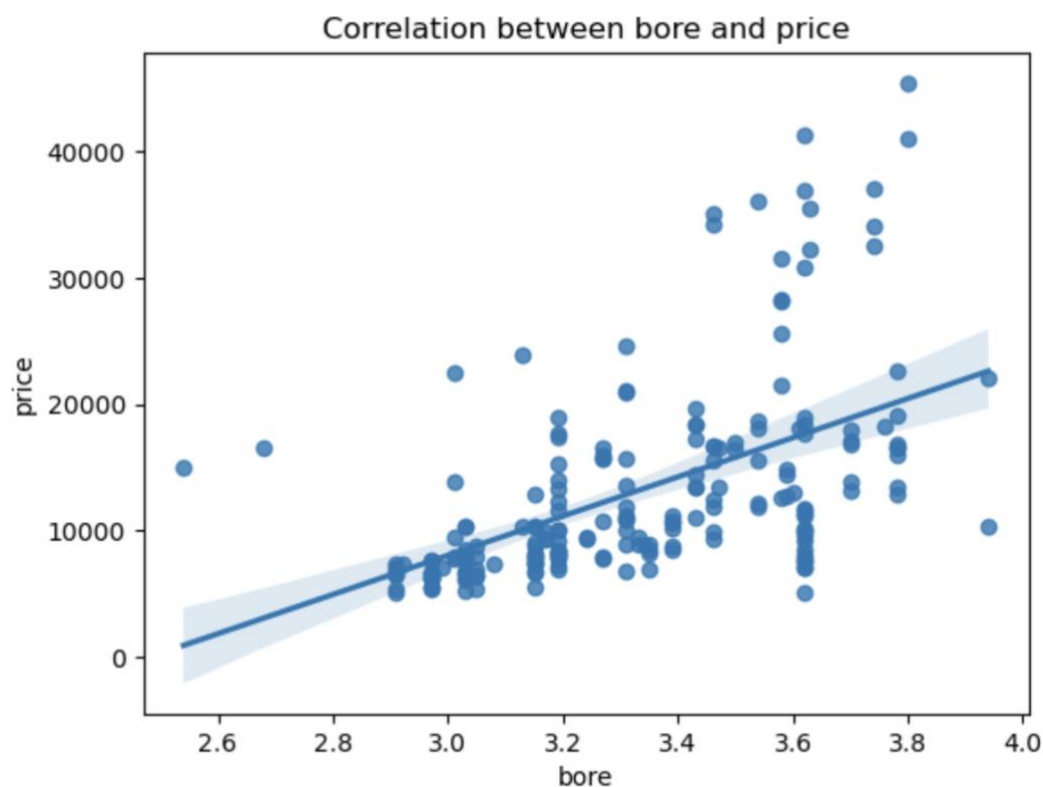
Тепер візуалізуємо кореляцію між кожним з ключових ознак та ціною.



Бачимо, що кореляція між peak rpm та ціною відсутня, що означає, що алгоритм помилково виділив цю ознаку як значущу.

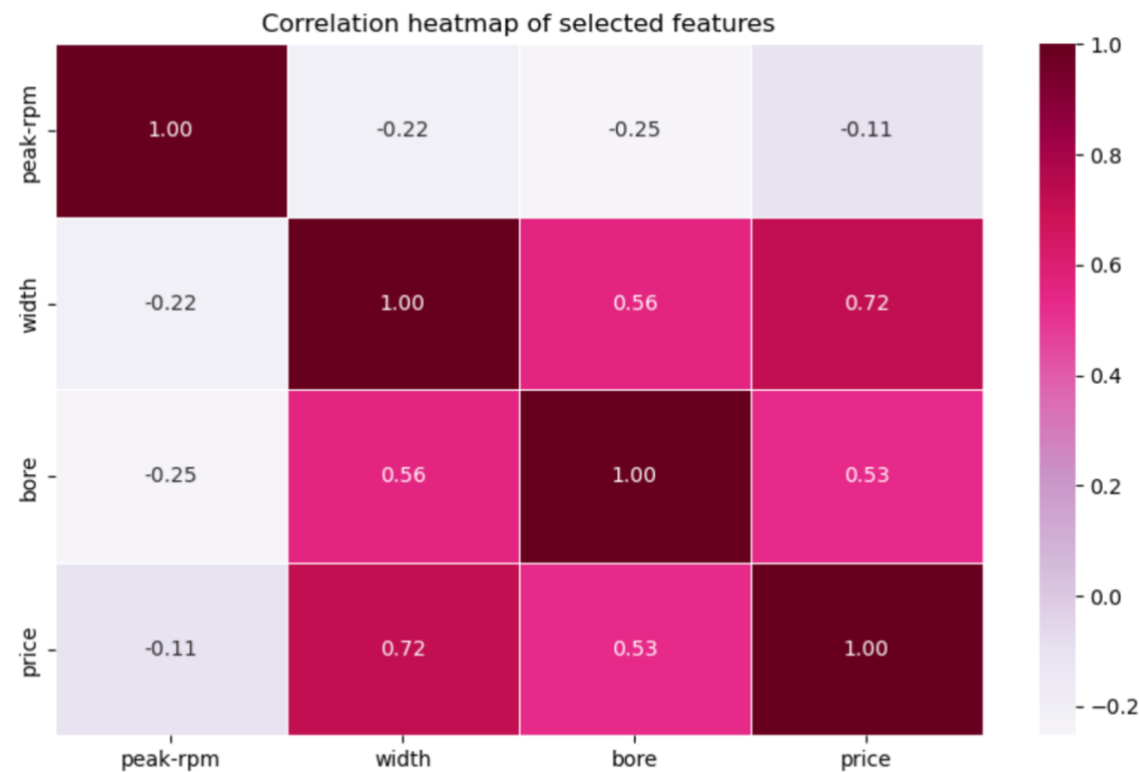


Кореляція між width та ціною загалом позитивна.



Кореляція між bore та ціною також загалом позитивна.

Побудуємо heatmap, який показує аналіз ознак на взаємозалежність:



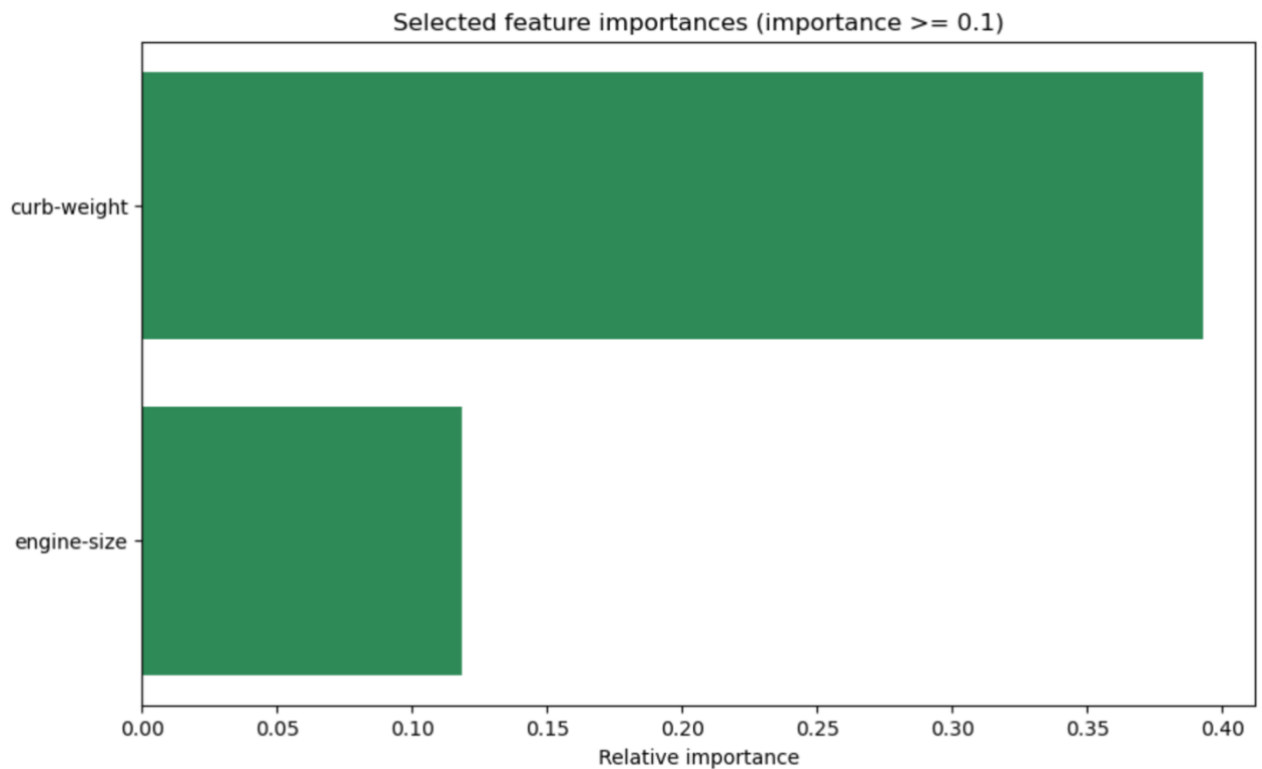
Бачимо, що кореляції між peak rpm та всіма іншими змінними немає. Між width та ціною найбільша кореляція (0.72), та між bore та ціною також позитивна кореляція (0.53).

*Висновки:* Багатофакторний вибір ознак виявив, що ключовими факторами, що впливають на ціну, є: peak-rpm, bore, та width. Після аналізу кореляції кожної ознаки з ціною було виявлено, що кореляція між peak rpm та ціною відсутня, тоді як для інших двох ознак вона позитивна.

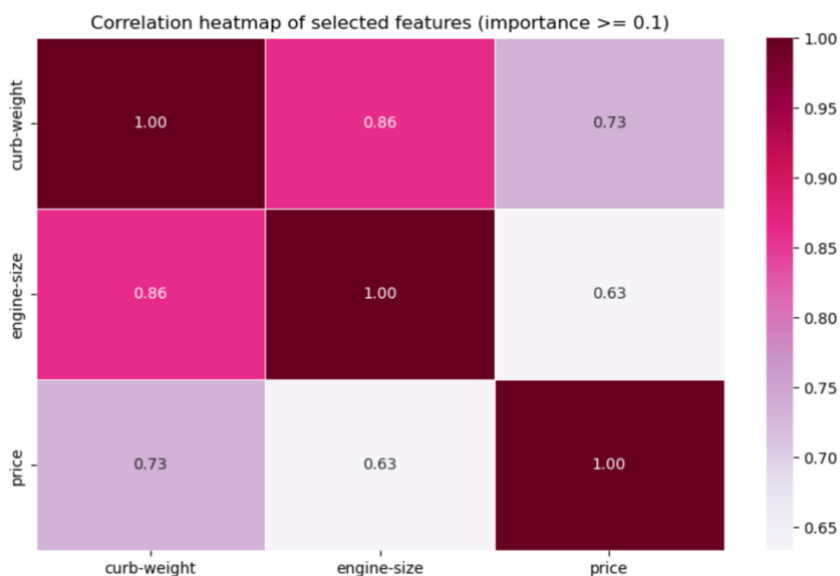


### ***Tree Based Method.***

Tree Based Method виділив наступні дві ключові ознаки: curb weight, engine size, що збігається з однофакторним методом. Візуалізуємо важливість кожної ознаки, згідно з методом:



Бачимо, що метод виявив curb weight як найбільш важливу ознаку, після якої йде engine size. Оскільки дві з ознак збігаються з однофакторним методом, перейдемо до створення heatmap:



Бачимо, що всі ознаки мають позитивну кореляцію. Найбільшу з них має curb weight (0.73), після чого йде engine size (0.63).

*Висновки:* Tree Based Method виявив, що ключовими факторами, що впливають на ціну, є: curb weight, engine size, що збігається з однофакторним методом.

Отже, два методи виявили ключові ознаки curb weight та engine size, а третій – bore та width. Тож, при виборі автомобілів на продаж слід звернути увагу на ці фактори.