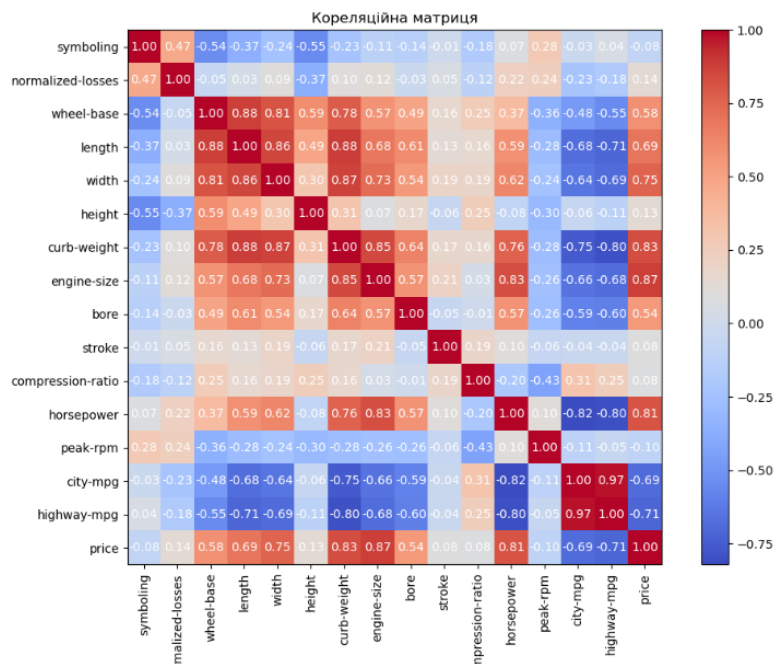


# Аналітичний звіт

## Вступ

Аналітичний звіт був зроблений під час виконання 3 лабораторної роботи. Аналіз проводився над набором даних який складається з даних з автомобільного щорічника Ward's Automotive Yearbook за 1985 рік. Метою є визначення значущих ознак за допомогою методів вибору ознак. У результаті було використані наступні методи: Pearson Correlation, SelectKBest, RandomForest.

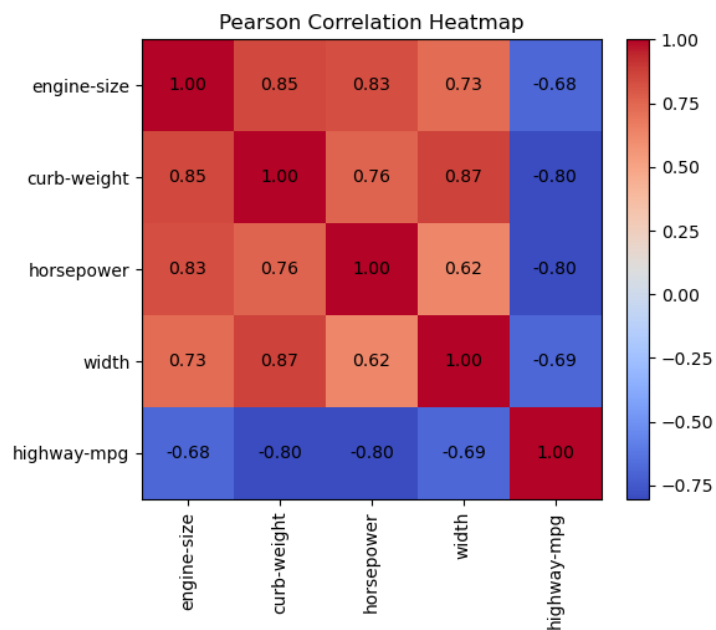
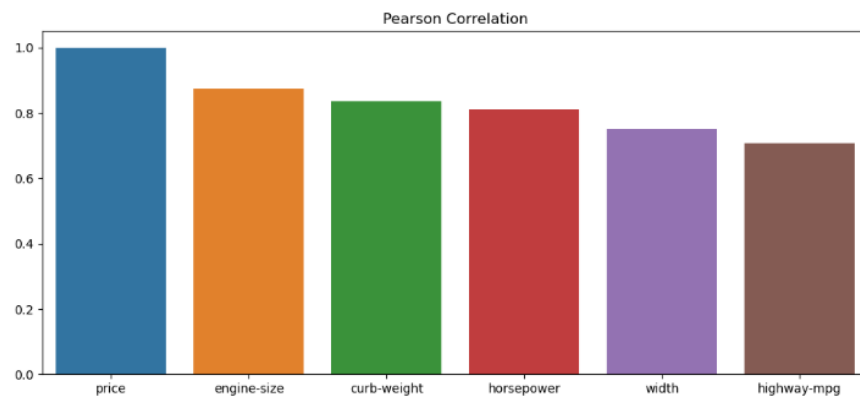
## Вибір ознак



Для початку можемо оцінити який вплив мають ознаки одне на одного. Відповідно до теплової карти можемо спостерігати, що найбільші значення коореляції мають наступні пари: price \_ curb-weight, engine-size, horsepower; wheel-base \_ length, width, curb-weight, curb-weight; length \_ width, curb-weight; width \_ curb-weight; engine-size \_ horsepower.

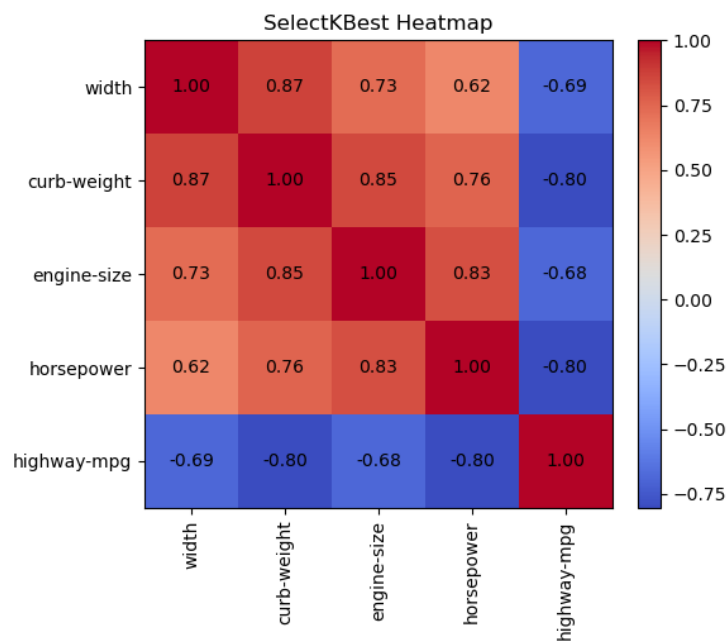
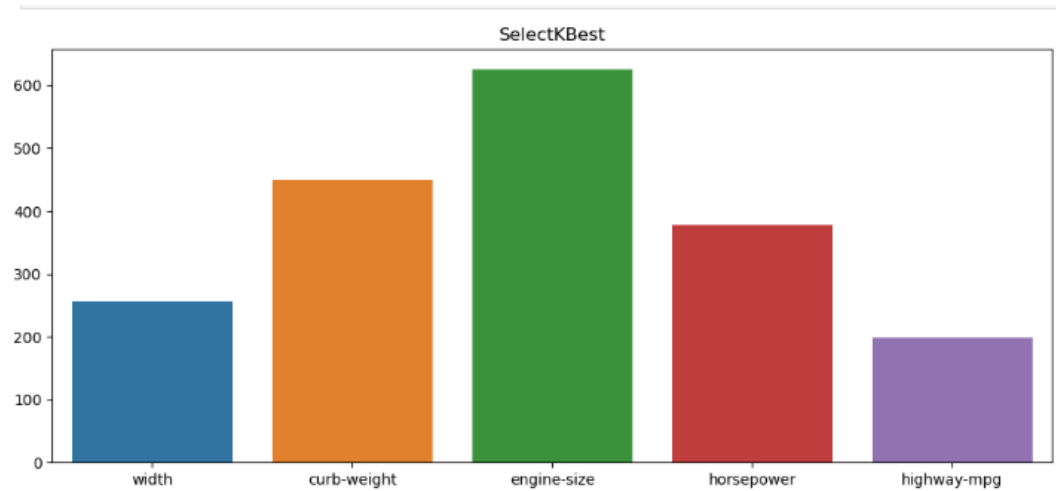
Враховуючи ці дані спробуємо зробити вибір ознак зазначеними методами та порівняти результати відбору. Зазначимо, що ціллю є ціна автомобіля.

## Pearson Correlation



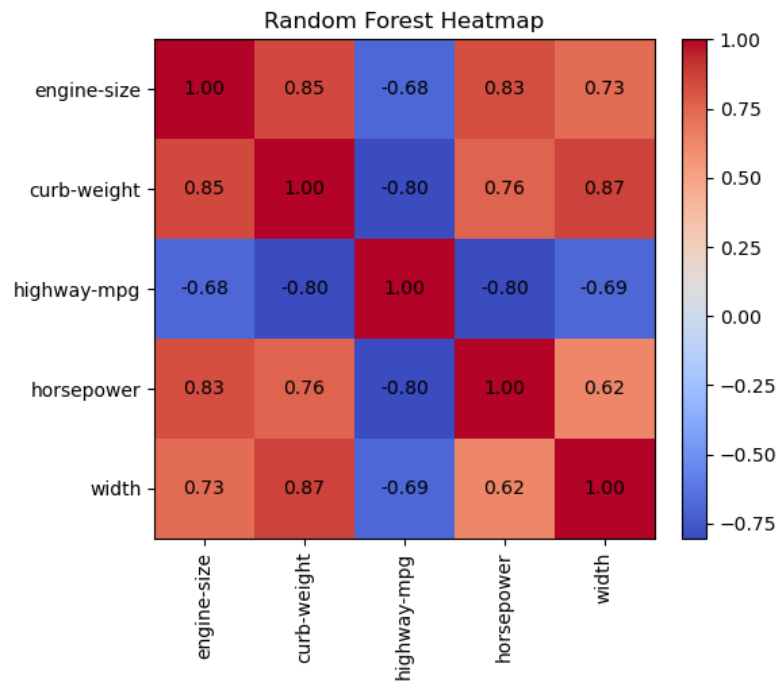
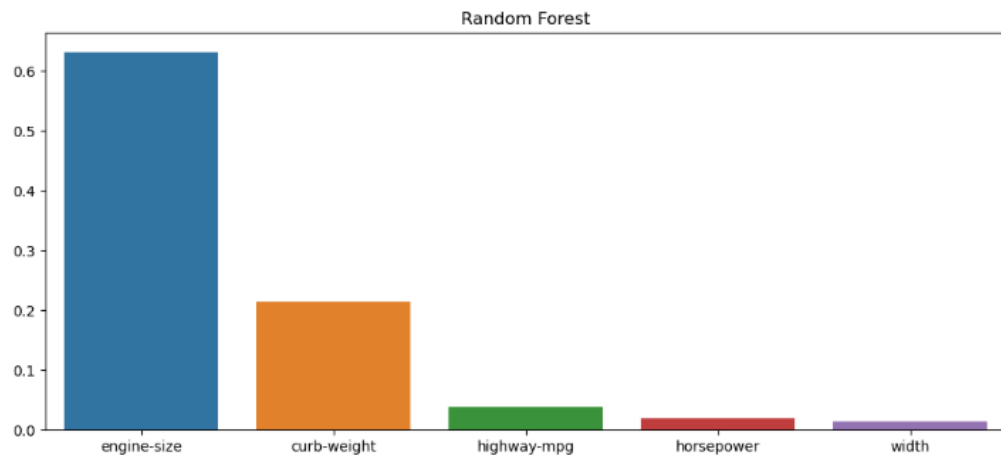
Отримуємо те, що даним набором були визначені наступні значущі ознаки: engine-size, curb-weight, horsepower, width, highway-mpg. При цьому найбільший вплив на ціну має саме розміри двигуна.

## SelectKBest



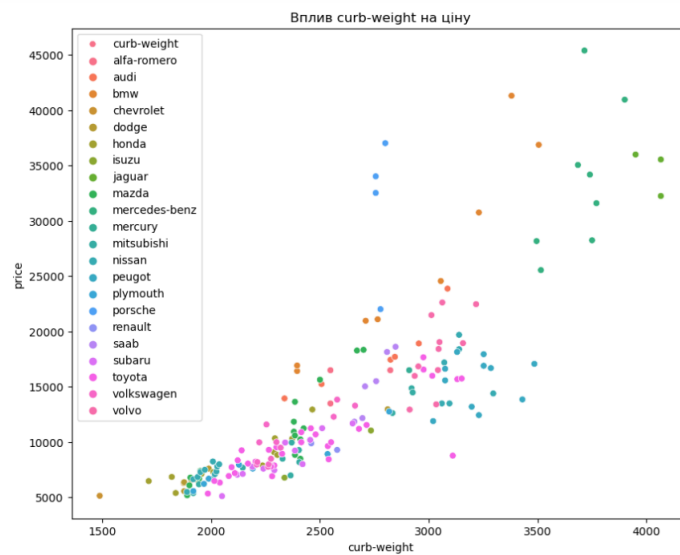
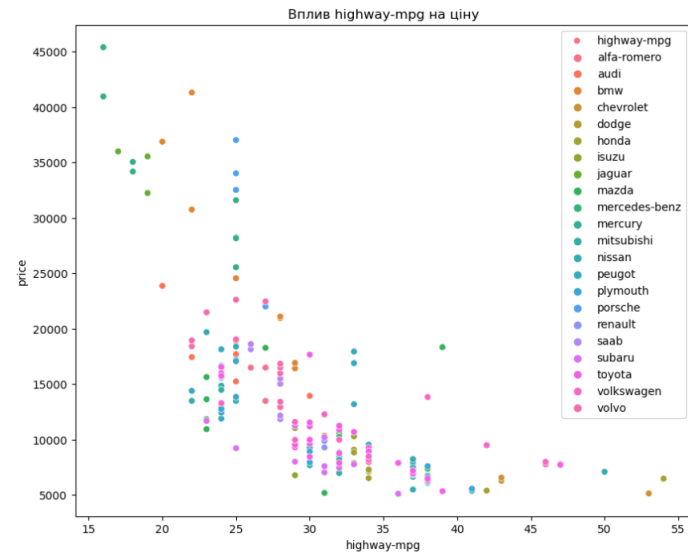
Отримуємо те, що даним набором були визначені наступні значущі ознаки: engine-size, curb-weight, horsepower, width, highway-mpg. Так само як і в першому методі найбільший вплив на ціну має саме розміри двигуна.

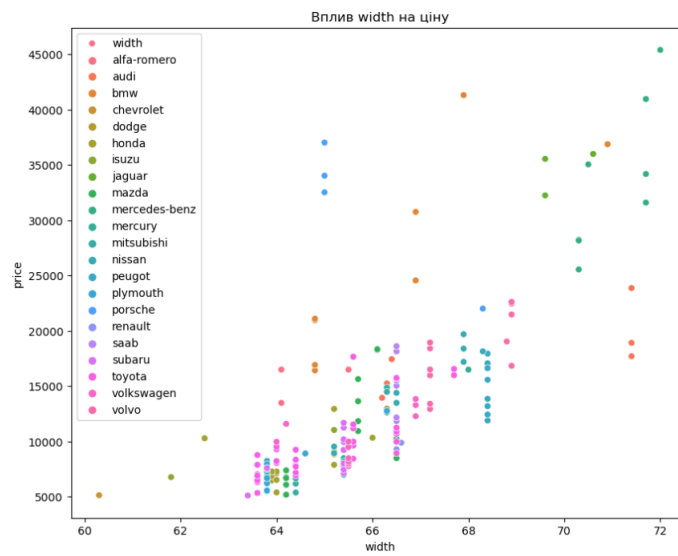
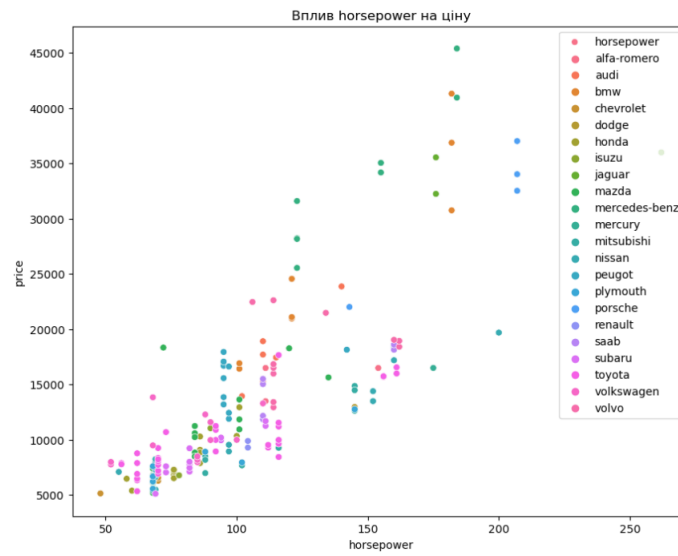
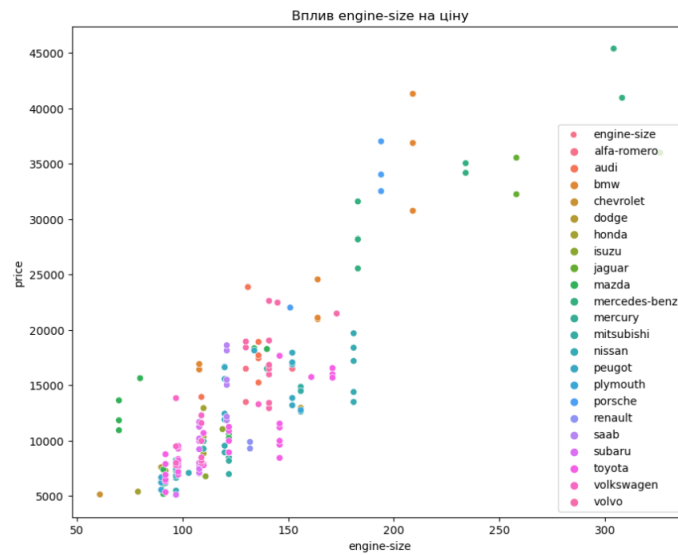
## RandomForest



Метод Random Forest так само вибирає наступні ознаки: engine-size, curb-weight, horsepower, width, highway-mpg. Так само як і в першому методі найбільший вплив на ціну має саме розміри двигуна, але на останньому місці цього разу стоїть параметр width.

# Дані як саме впливають обрані ознаки на ціну та їх виробники





## **Висновок**

У результаті дослідження було виявлено, що значущі ознаки, які впливають найбільше на ціну – це ширина автомобіля, потужність, розміри двигуна, розхід пального на шосе та curb-weight. Відповідно автомобілі у яких усі ці ознаки окрім розходу пального будуть нижчі, відповідно ціна також буде нижча. З розходом пального ситуація протилежна.