Звіт

на тему:

Вплив факторів (ознак) на ціноутворення автомобільного ринку

Підготував:

Скріпка Владислав

2023 рік

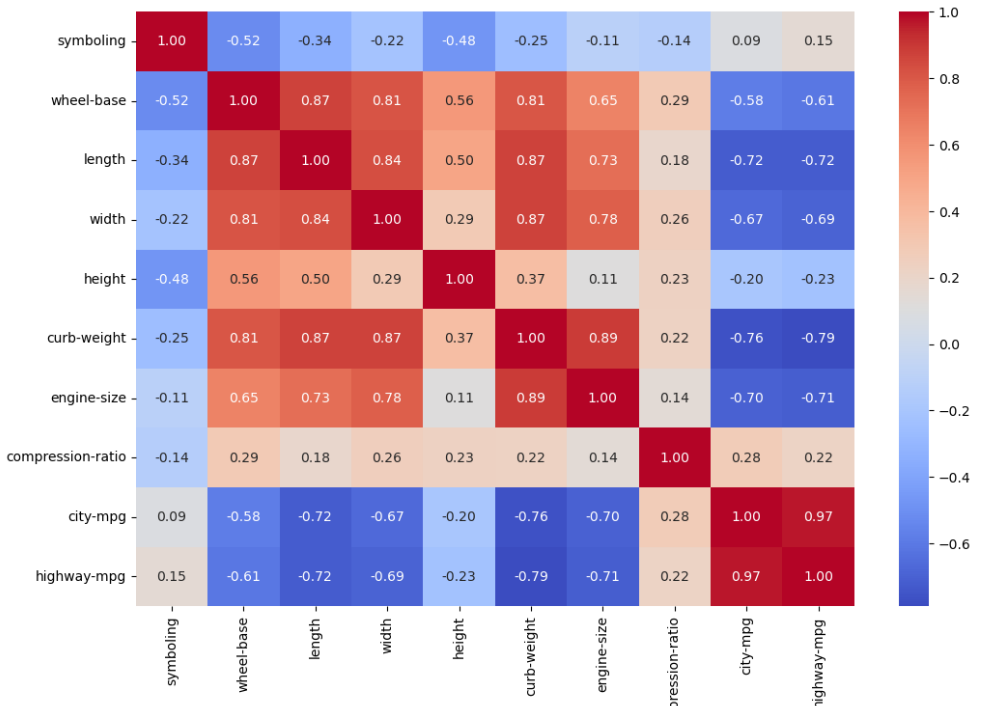
**КОНТЕКСТ**

Набір даних складається з трьох типів об’єктів: (а) специфікація автомобіля з точки зору різних характеристик, (б) присвоєний йому рейтинг страхового ризику, (в) його нормалізовані втрати під час використання порівняно з іншими автомобілями. Другий рейтинг відповідає ступеню ризику автомобіля, ніж вказує його ціна. Автомобілям спочатку присвоюється символ фактора ризику, пов'язаний з його ціною. Потім, якщо це більш ризиковано (або менше), цей символ коригується шляхом переміщення його вгору (або вниз) за шкалою. Актуарії називають цей процес «символізація». Значення +3 вказує на те, що авто є ризикованим, -3, що воно, ймовірно, досить безпечне.

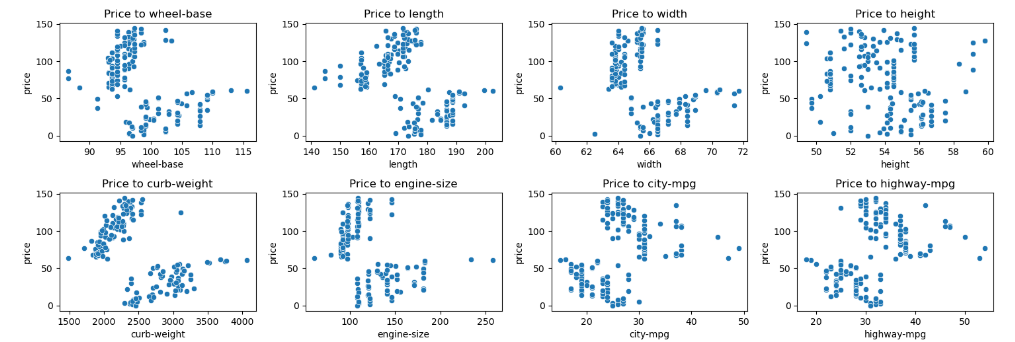
Третім фактором є відносна середня виплата збитку за рік страхування автомобіля. Це значення нормалізовано для всіх автомобілів певної класифікації розміру (дводверні маленькі, універсали, спортивні/спеціальні тощо) і являє собою середні втрати на автомобіль на рік.

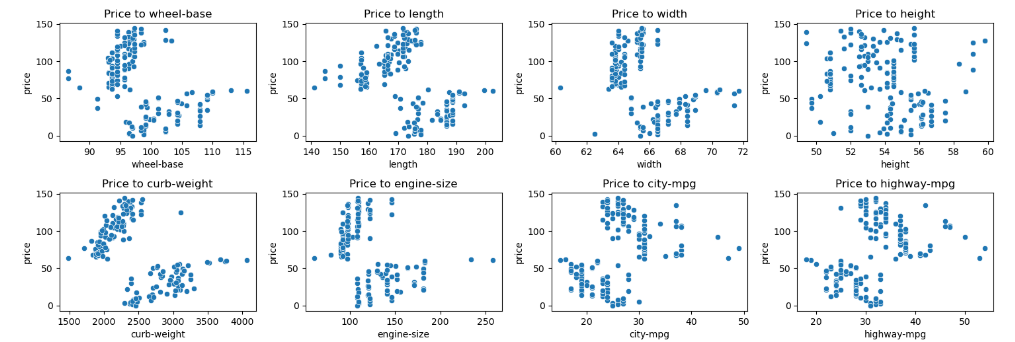
**РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ**

При дослідженні кореляції між усіма факторами можливого ціноутворення (за допомогою обчислення кореляційної матриці) та їх подальшого візуального представлення у вигляді heatmap, стало зрозуміло, що фактори symbolic, height, compression ratio мають нульову або досить незначну силу статистичного зв’язку між порівняно з іншими властивостями.



Графіки розсіювання для відношень ціни та числових властивостей:





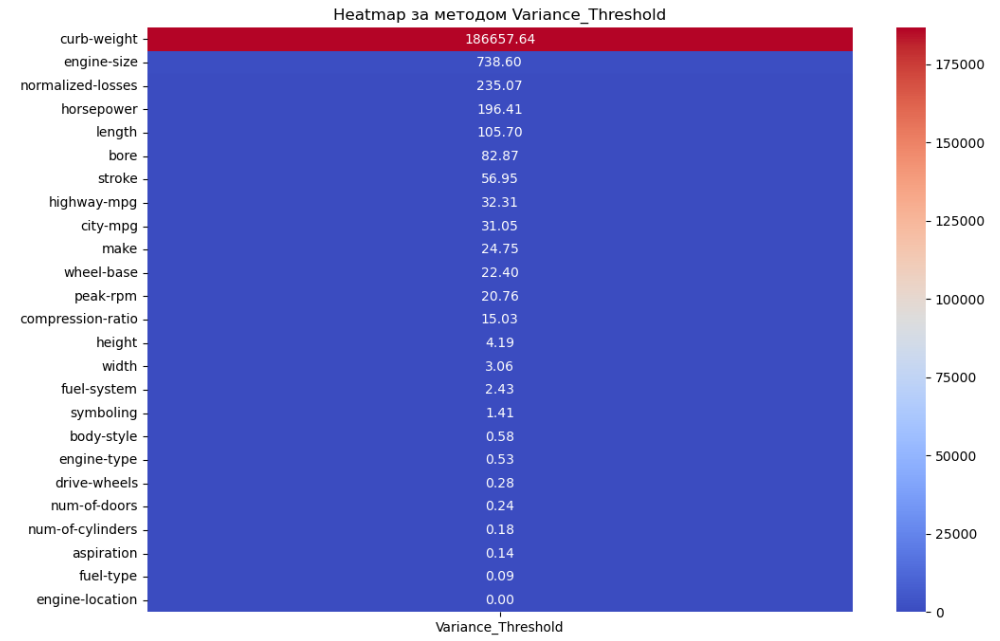
Графіки відображають відношення між ціною (price) і кожною з числових ознак (numeric\_features).

Основна ідея графіків розсіювання - вони дозволяють візуалізувати залежність між двома числовими змінними. Кожна діаграма розсіювання відповідає одній парі змінних, де ось x - це одна з ознак, а ось y - це ціна (price).

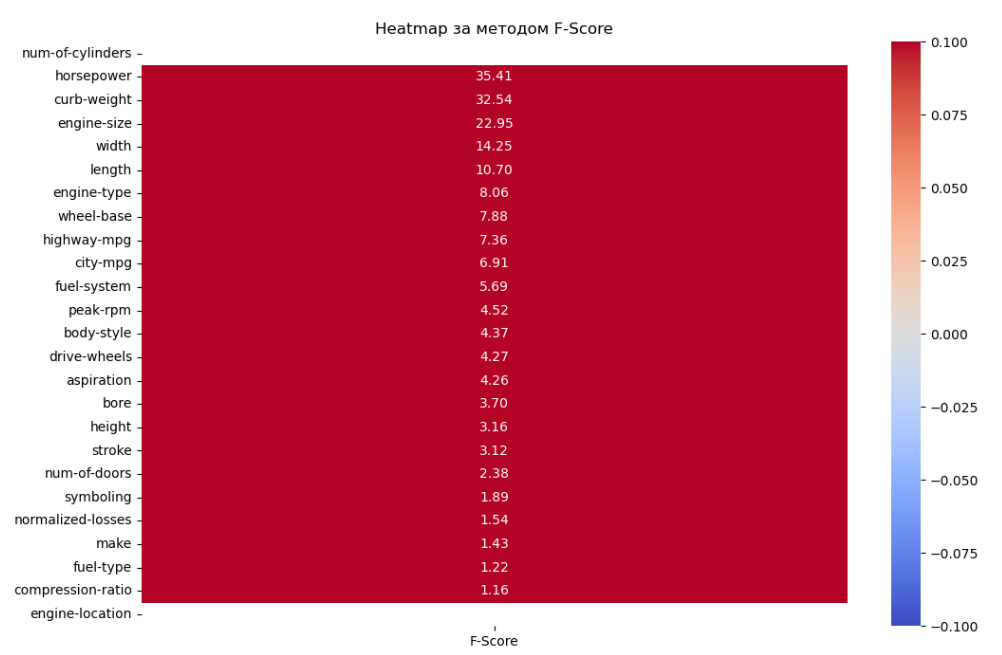
За допомогою цих графіків можемо спостерігати за тим, як змінюється ціна автомобіля в залежності від конкретної числової характеристики. Таким чином дає змогу візуалізувати можливі статистичні зв’язки між властивостями та нашою ціллю - ціною.

Проведено дослідження за методами відбору значущих ознак:

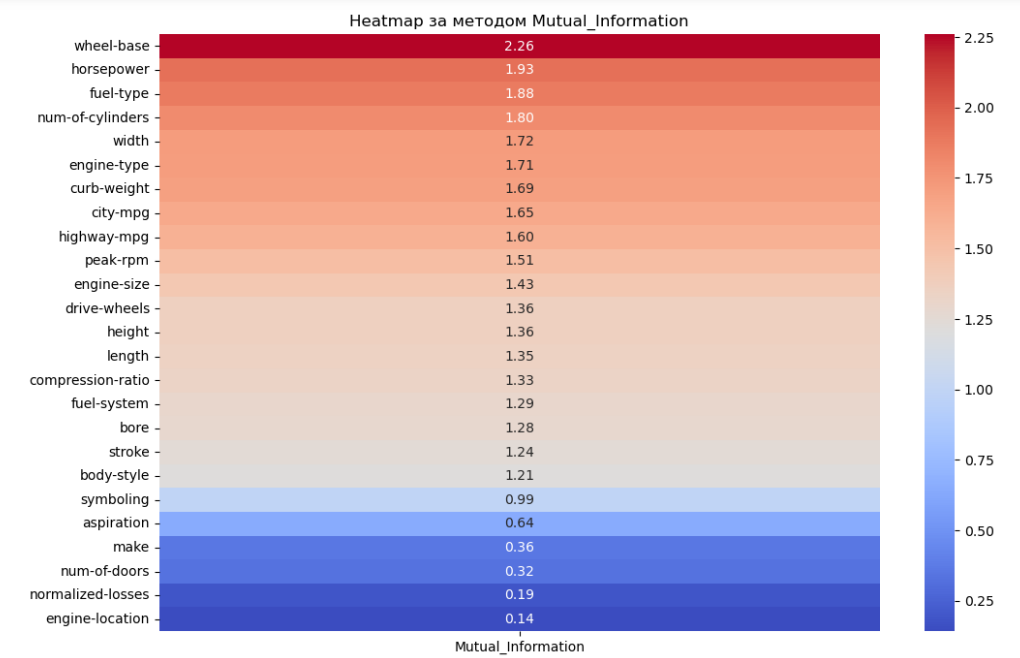
* Variance Threshold - вимірює дисперсію (варіативність) кожної ознаки, і якщо вона менше порогового значення, ознака вважається низько варіативною і може бути вилучена.



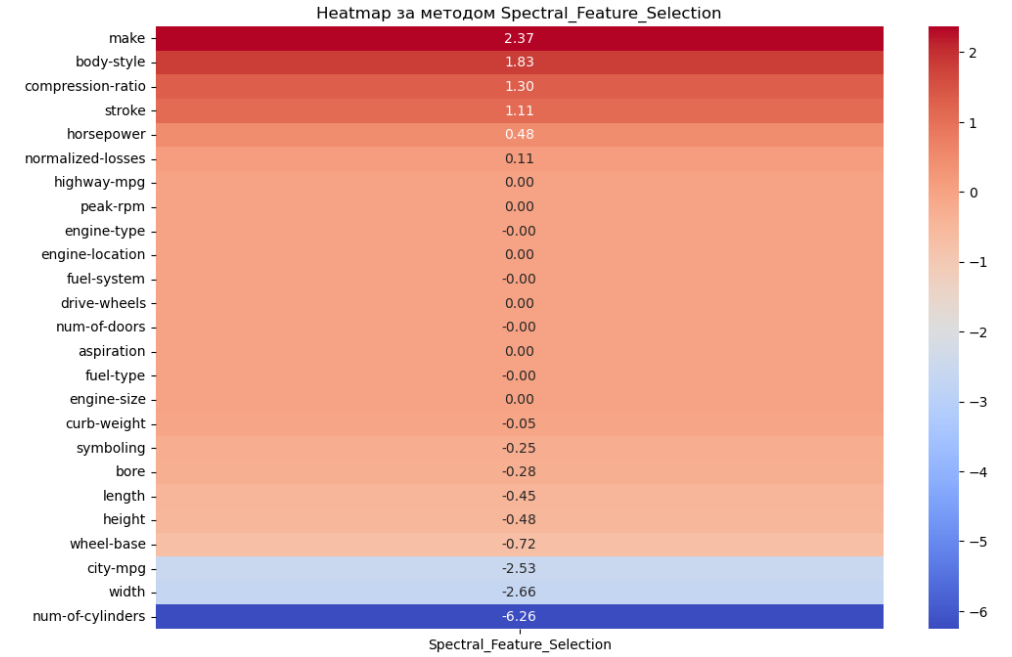
* F-Score - Кожній ознаці надається бал на основі того, наскільки добре вона розрізняє класи в залежній змінній



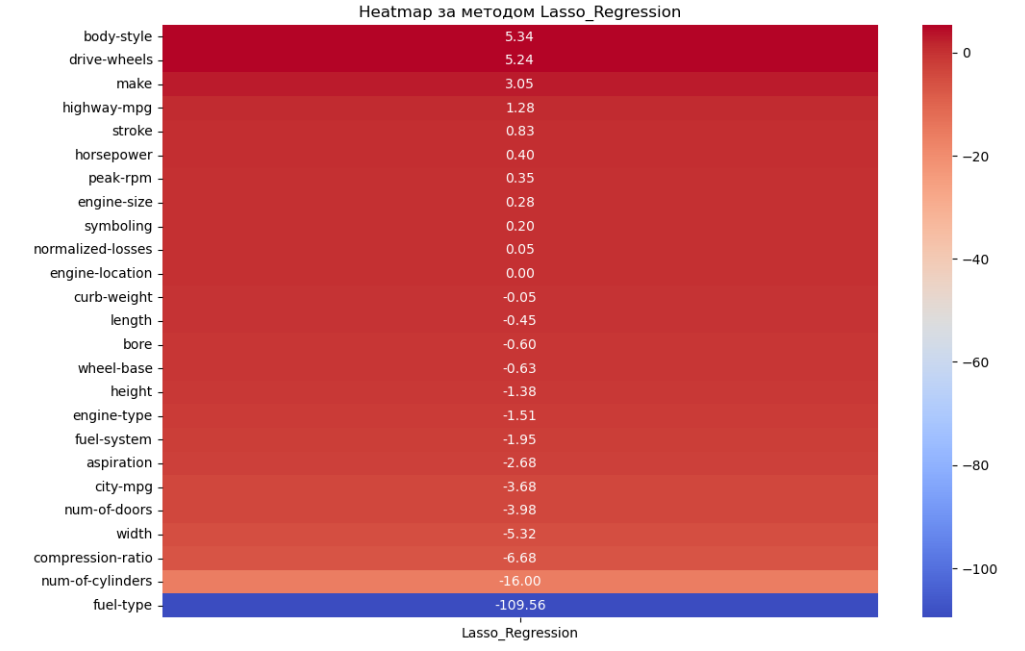
* Mutual Information – використовується теорія інформації для вимірювання того, наскільки інформативна кожна ознака для передбачення залежної змінної



* Spectral feature selection - використовується графова теорія та аналіз власних значень для визначення важливості ознак.



* Lasso regression - Мінімізує суму квадратів помилок та штрафує за абсолютні значення коефіцієнтів, що використовуються для кожної ознаки.



**ВИСНОВОК**

У результаті дослідження можемо зробити висновки щодо важливості представлених властивостей:

* make: ця ознака має велику важливість. Вона, ймовірно, важлива для моделі, але слід звернути увагу на те, що важливість може залежати від конкретного кодування марок (марки автомобіля).
* wheel-base, length, width, curb-weight, engine-size, horsepower: ці числові ознаки мають високу важливість. Зокрема, engine-size і horsepower.
* city-mpg, highway-mpg, fuel-type, aspiration, body-style, drive-wheels, engine-type, fuel-system: відповідно до досідження можна зробити висновок, що це важливі числові ознаки.
* normalized-losses, bore, stroke, compression-ratio, peak-rpm: важливість цих числових ознак середня, варто розглядати інші.
* symboling: дана ознака може вважатись не важливою для моделі, оскільки важливість відокремлена низьким значенням.
* engine-location: Має нульову важливість у двох методах. Це може означати, що ця ознака не додає значущості моделі й не впливає на ціноутворення.

Таким чином ми можемо відокремити змінні, що являються важливими у процесі ціноутворення при покупці чи продажі автомобілів.