**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Звіт до лабораторної роботи №3**

**з курсу**

**«Data Science та Big Data»**

*студента 4 курсу*

*групи ПП-41*

*спеціальності 122 «Комп’ютерні науки»*

*ОП «Прикладне програмування»*

Терновцева В.О.

*Викладач:*

*Білий Р.О.*

**Київ – 2023**

**Лабораторна робота №3**

Методи аналізу та вибору значущих ознак

(Features’ Selection Procedures)

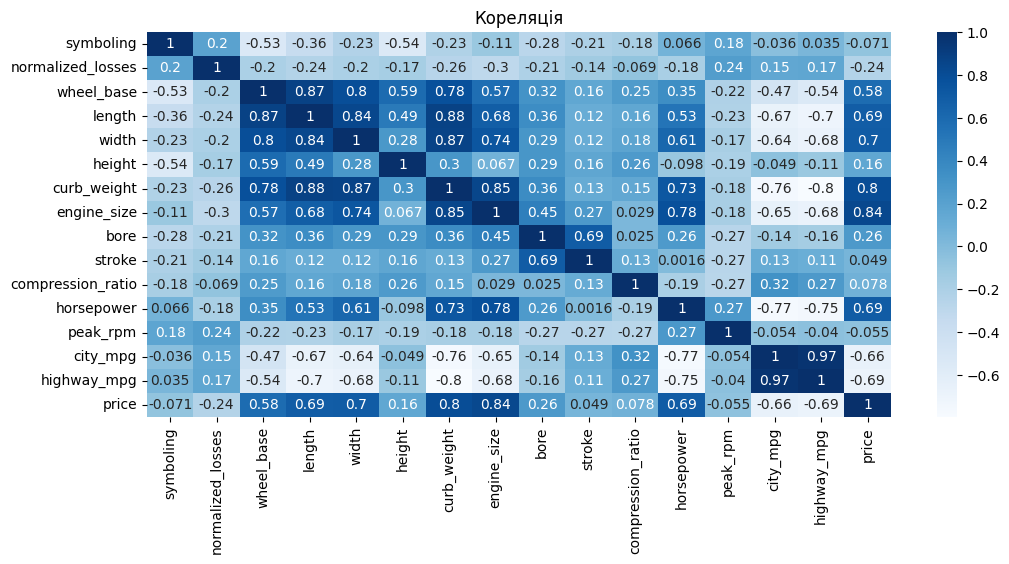
*Мета*: отримання практичних навичок аналізу та вибору значущих ознак для моделі за допомогою кореляційного аналізу, таблиць сопряжіння, аналізу багатомірні залежності та дихотомії, дисперсійного аналіз – ANOVA, критерій Хі-квадрат тощо.

**Хід виконання**

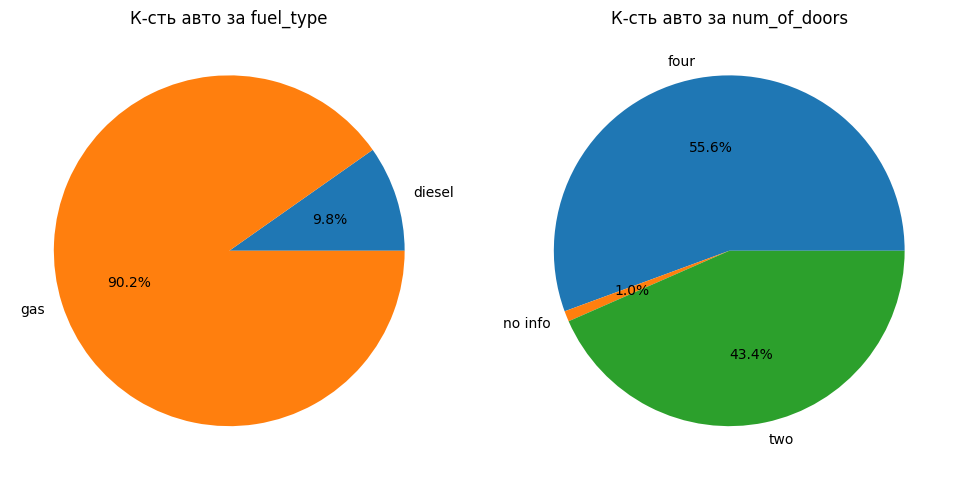
Проведемо кореляцію числових значень у датасеті. З отриманої мапи температур можна зробити кілька висновків стосовно ціни.

По-перше, на ціну автомобіля найбільший позитивний вплив мають колісна база, довжина, ширина, його споряджена маса, розміри двигуна та кінська сила.

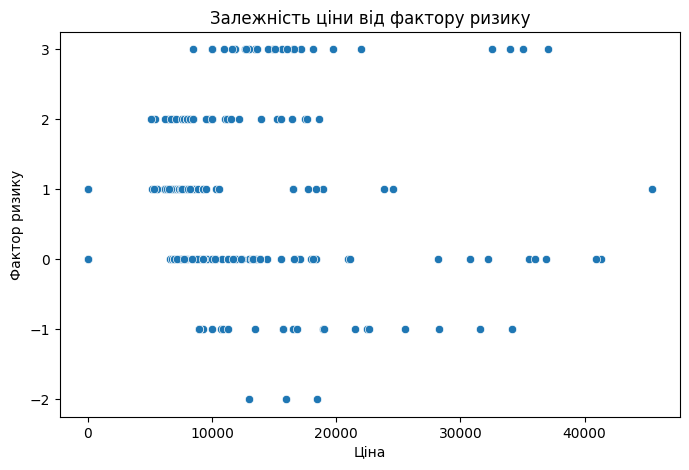
По-друге, від’ємно на ціну впливають витрати пального автомобілем. Тобто, чим вищими є витрати, тим нижчою буде ціна.



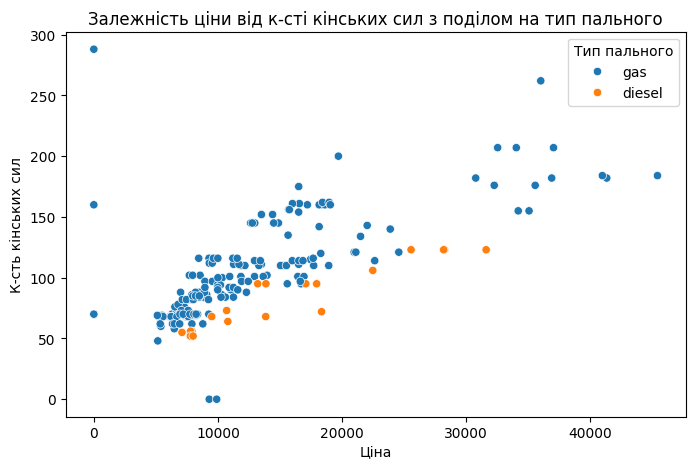
Порівняємо у відсотках кількість авто з різним пальним та кількістю дверей. Як можна побачити, 90,2% від усіх автомобілів становлять ті, які їздять на газу, в той час як авто на дизельному пальному становлять лише 9,8%. По кількості дверей автомобілі поділені майже порівну.



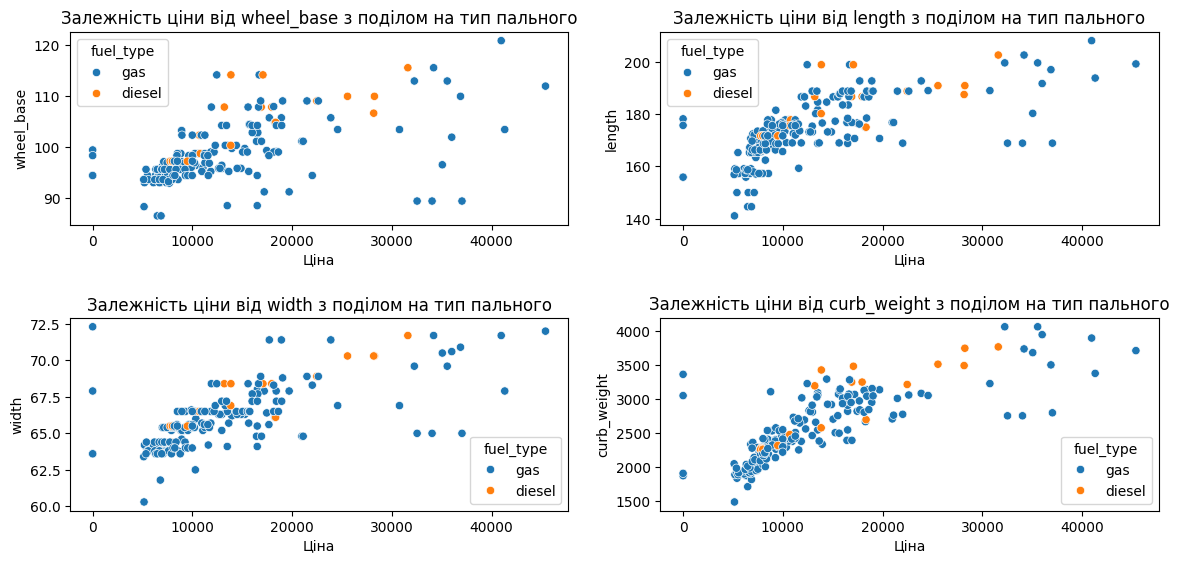
Оцінимо залежність ціни ти фактору ризику. Бачимо, що є незначний вплив, тобто при більшій безпечності авто його ціна може бути вищою, проте чіткої залежності немає.



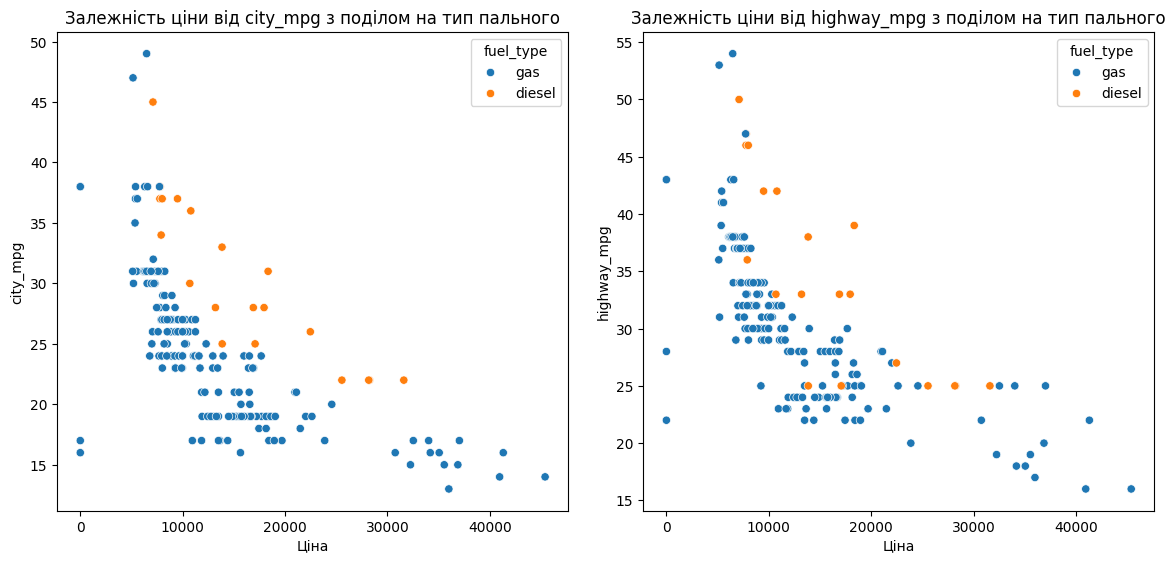
Переглянемо, наскільки на ціну впливає кількість кінських сил автомобіля. Як видно на графіку, автівки на газу при однаковій ціні мають вищу кількість кінських сил, порівняно з машинами на дизельному паливі. Загалом видно тенденцію зростання ціни зі збільшенням кінських сил.



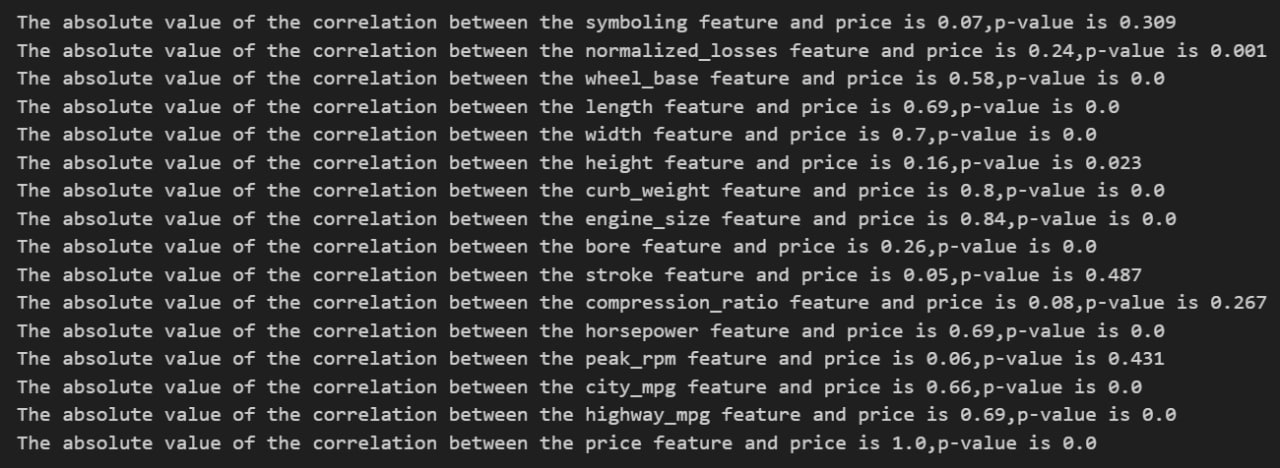
Аналогічно проведемо аналіз залежності для інших критеріїв, які були описані вище: колісна база, довжина, ширина, та споряджена маса. Для кожного з цих критеріїв показники для авто на дизельному паливі вищі, ніж середні показники для авто на газу. Колісна база має найбільший розкид значень. Для довжини та спорядженої маси пряму залежність дуже чітко видно.

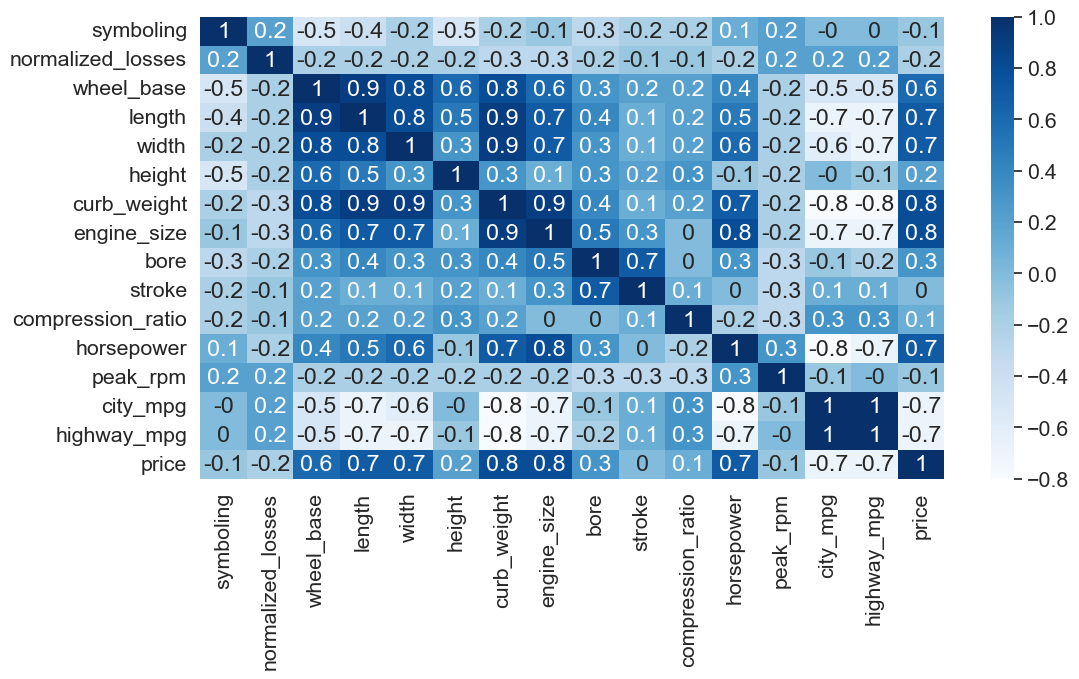


Порівняємо також, як на ціну впливають витрати пального. На графіку видно, що тип пального теж має значення, оскільки для дизельного палива при суттєво вищих витратах ціна залишається меншою. Загалом, як і було вказано вище, витрати автомобілем пального обернено пропорційні ціні.

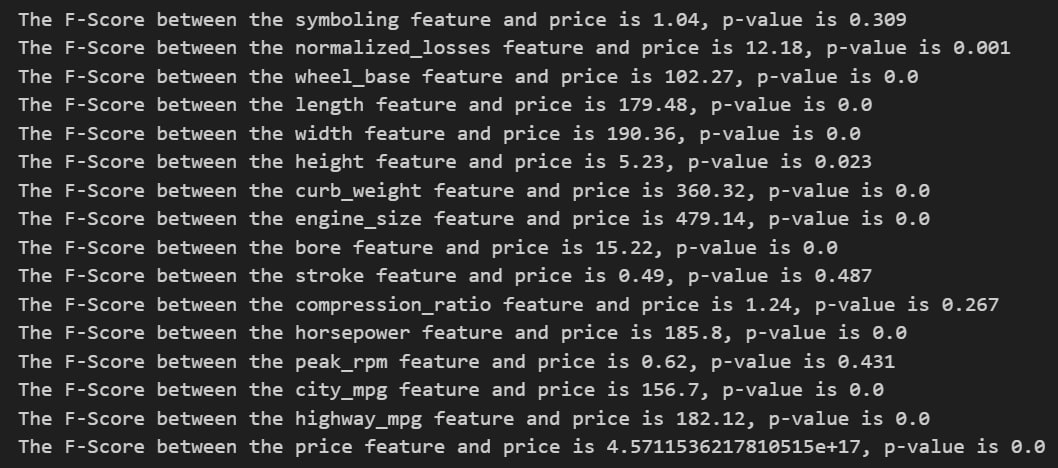


Визначимо абсолютне значення кореляції між ціною та параметрами авто та побудуємо мапу температур. Якщо абсолютне значення кореляції близьке до 1, це вказує на сильну лінійну позитивну залежність. Якщо абсолютне значення кореляції близьке до -1, це вказує обернену залежність. Коли абсолютне значення кореляції близьке до 0, це свідчить про відсутність залежності.





Визначимо F-міру між ціною та параметрами авто. Як видно з фото нижче, ложнопозитивних або ложнонегативних результатів немає.



Визначимо критерій хі-квадрат між ціною та параметрами авто.

