**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

ЗВІТ

Про виконання лабораторної роботи №3

з дисципліни: «Data Science та Big Data»

Студентки групи ПП-42

Шкаленко Ганни

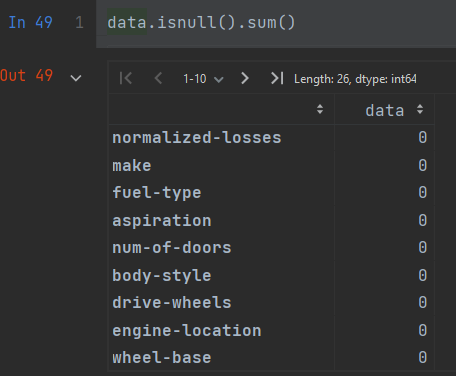
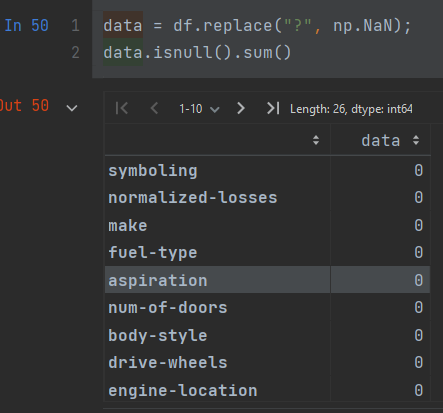
Київ – 2022

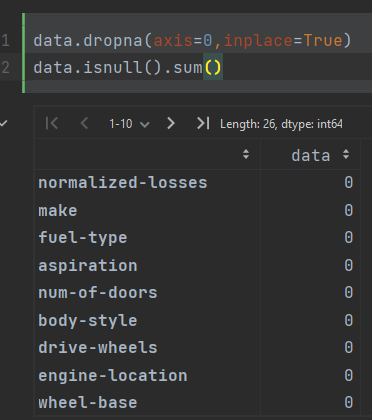
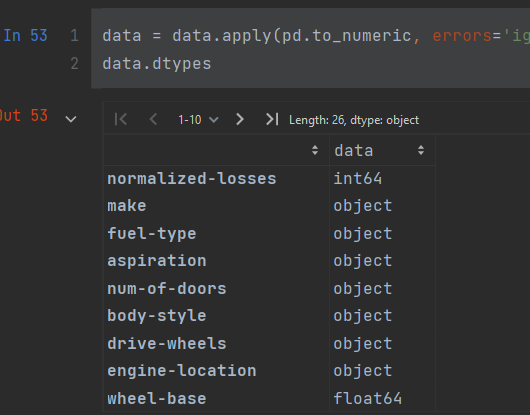
**Лабораторна робота 3.**

*Тема роботи*: Методи аналізу та вибору значущих ознак (Features’ Selection Procedures)

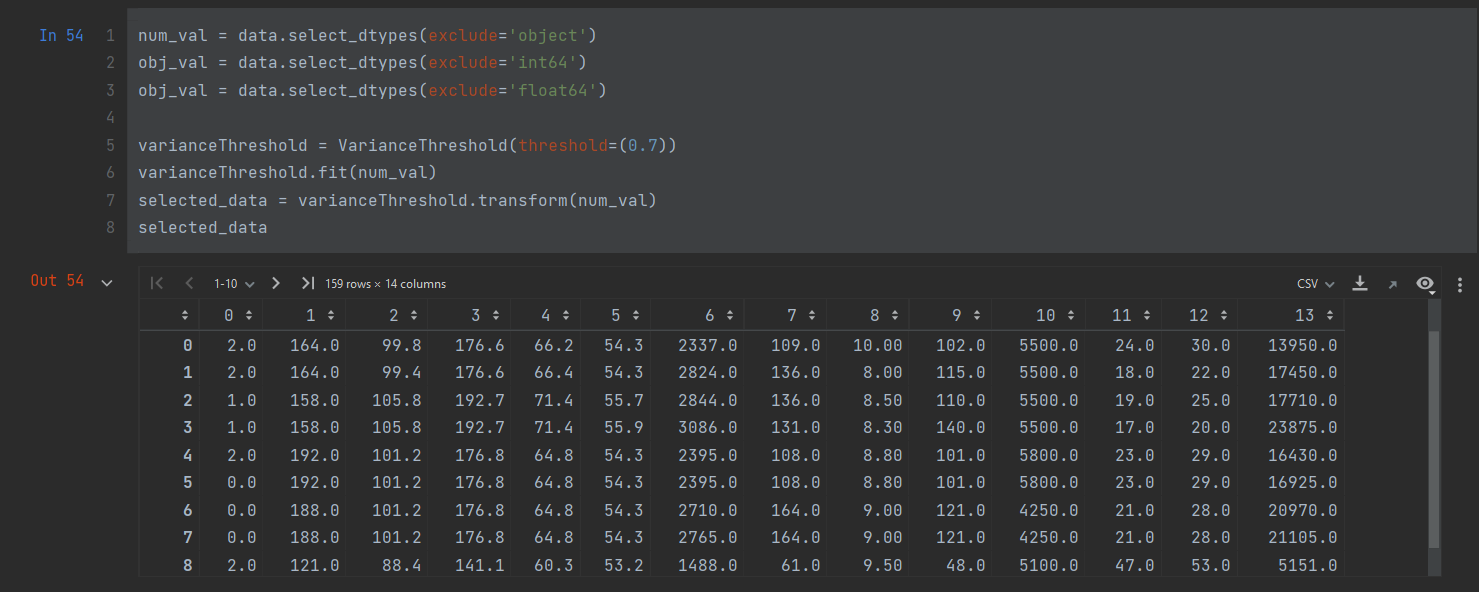
Порядок виконання

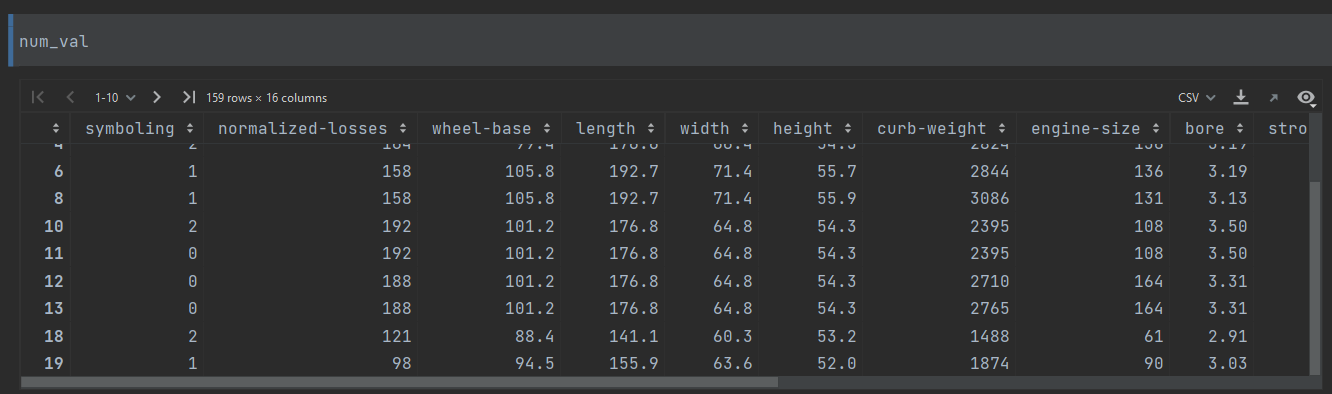
Спочатку я підготувала дані для роботи.

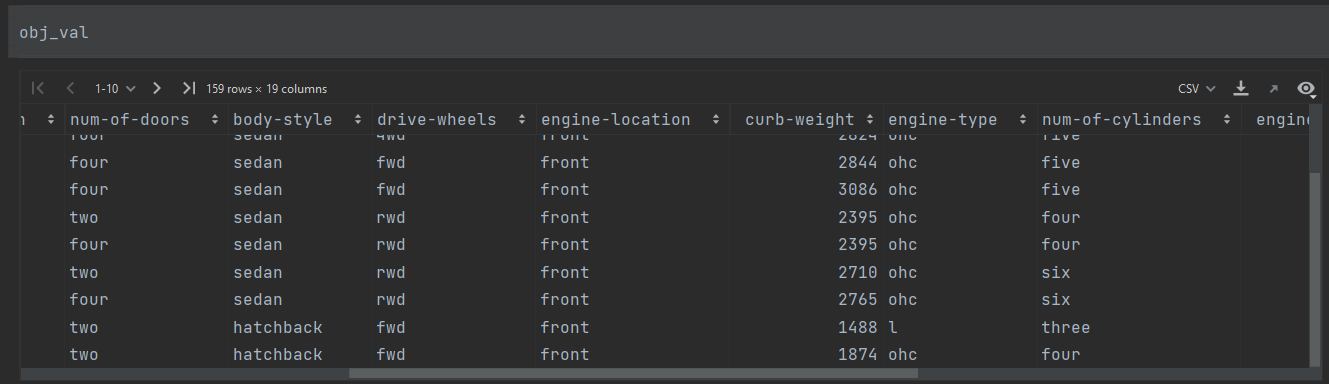
 

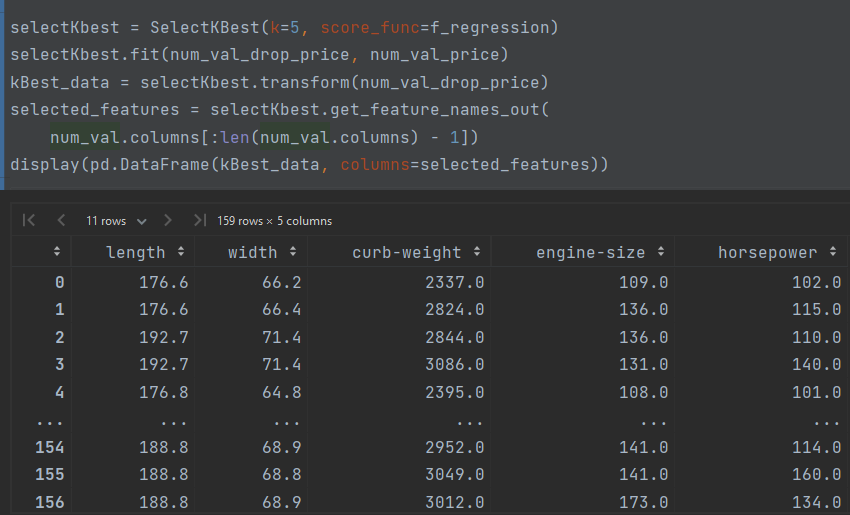
Далі використала метод для оцінки значущості факторів. Спочатку відкинула стовпці з дисперсією менше 0.7:



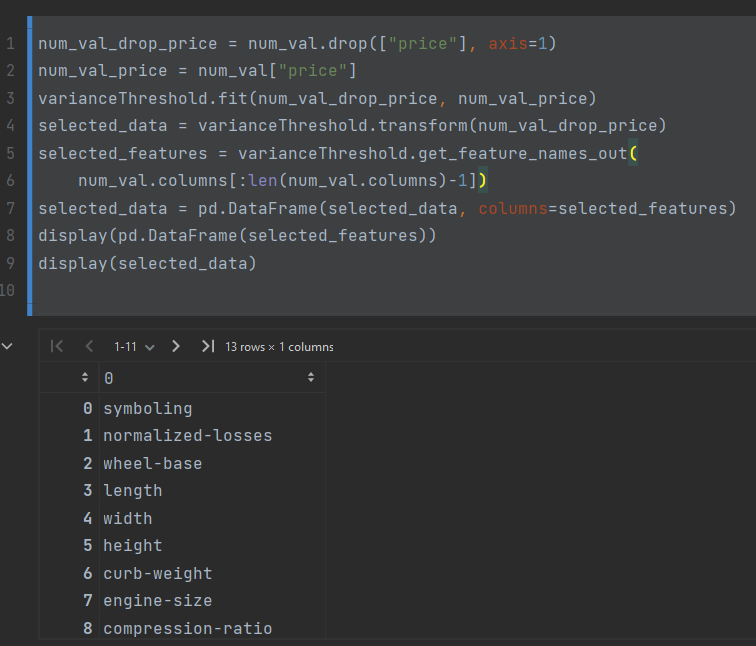




Наступний метод обирає k=5 найвпливові фактори.



Останній метод визначення виглядає так:



У підсумку з огляду цих методів можна визначити, що найбільше на ціну впливають такі показники – curb-weight, normalized-losses, peak-rpm, engine-size, horsepower.

Візуалізація кореляції:

