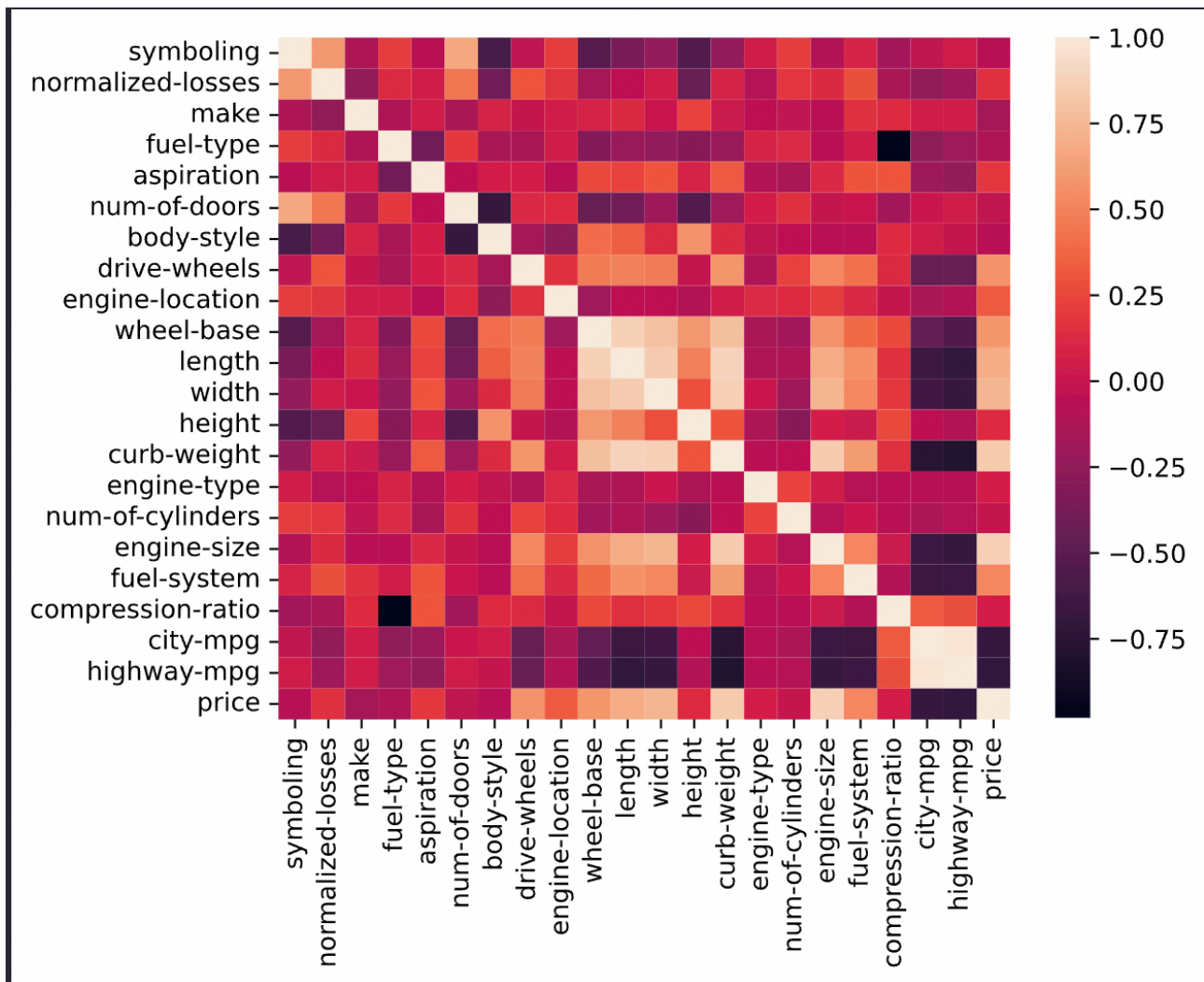


В даному аналізі було проаналізовано датасет Ward's Automotive Yearbook за 1985 рік задля визначення найважливіших характеристик, які вплинули на ціну автомобіля. Для початку було виведено загальну кореляцію метрик



10 значень з найбільшою кореляцією

curb-weight	highway-mpg	0.797465
highway-mpg	curb-weight	0.797465
price	curb-weight	0.831701
curb-weight	price	0.831701
width	length	0.841118
length	width	0.841118
engine-size	curb-weight	0.850594
curb-weight	engine-size	0.850594
width	curb-weight	0.867032
curb-weight	width	0.867032
price	engine-size	0.867435
engine-size	price	0.867435

wheel-base	length	0.874587
length	wheel-base	0.874587
length	curb-weight	0.877728
curb-weight	length	0.877728
city-mpg	highway-mpg	0.971337
highway-mpg	city-mpg	0.971337
compression-ratio	fuel-type	0.984356
fuel-type	compression-ratio	0.984356

Задля визначення найбільш важливих метрик відносно ціни було використано чотири методи.

1. Поріг дисперсії випадкової величини. За поріг було вибрано значення 0.7

	0
0	normalized-losses
1	make
2	body-style
3	wheel-base
4	length
5	width
6	height
7	curb-weight
8	engine-type
9	engine-size
10	fuel-system
11	compression-ratio
12	horsepower
13	peak-rpm
14	city-mpg
15	highway-mpg

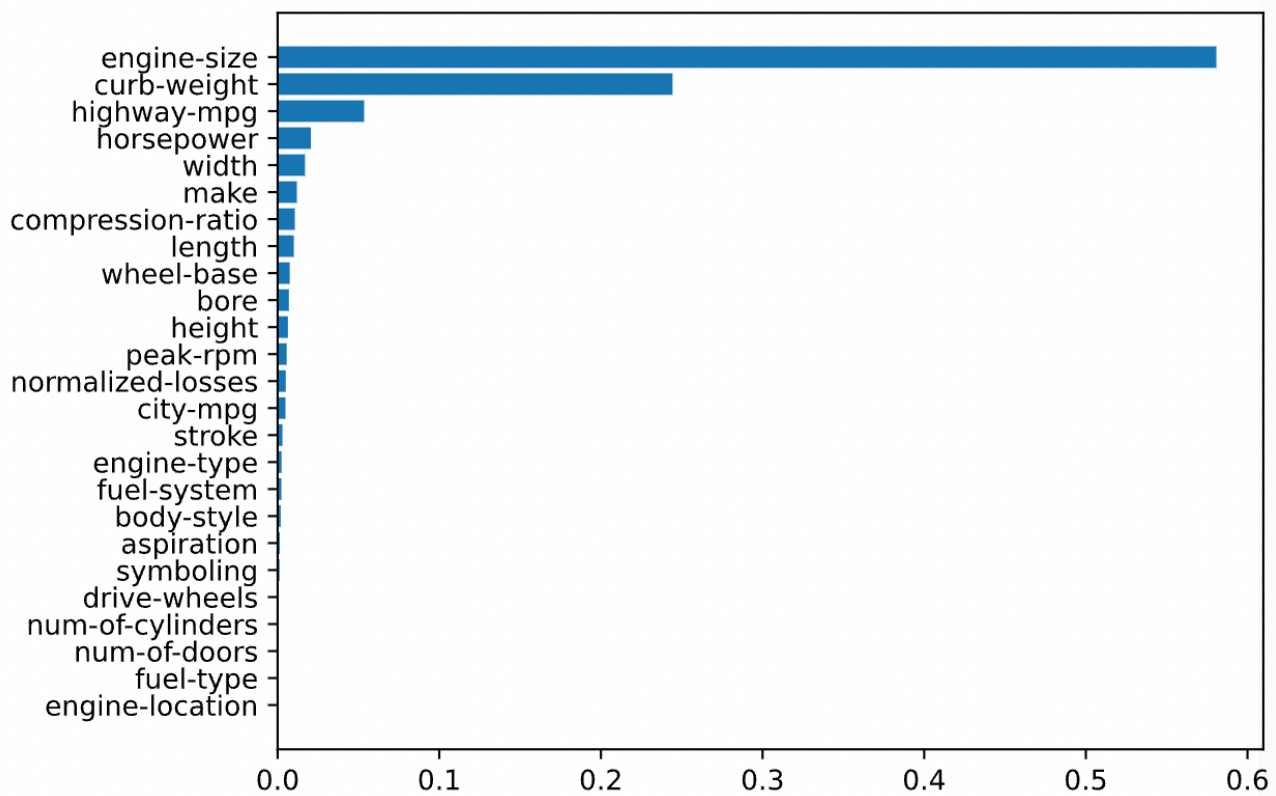
2. Вибір K найкращих з функцією f_regression, k=5

	columns	Kept
11	width	True
13	curb-weight	True
16	engine-size	True
21	horsepower	True
24	highway-mpg	True

3. Вибір K найкращих з функцією f_classif, k=5

	columns	Kept
8	engine-location	True
11	width	True
13	curb-weight	True
15	num-of-cylinders	True
16	engine-size	True

4. Визначення важливості кожної метрики з використанням RandomForestRegressor



У висновку можна сказати, що за результатом дослідження, серед всіх методів найбільш важливими можна виділити такі 4 характеристики: engine-size, curb-weight, highway-mpg, width