**Аналітичний звіт основних ознак ринку підтриманих автомобілів**

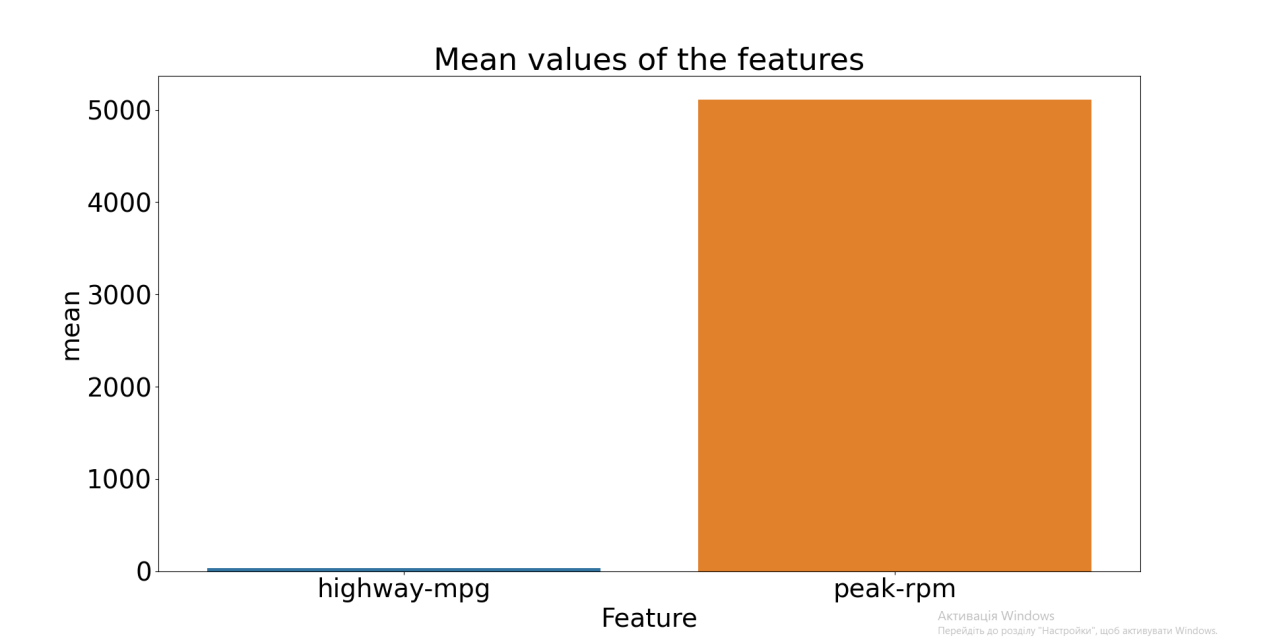
**Вступ**

### Аналітичний звіт був зроблений на основі даних з автомобільного щорічника “Ward's Automotive Yearbook” за 1985 рік. Для вибору ознак, які впливають на ціну, були вибрані такі алгоритми: однофакторний вибір ознак, рекурсивний, вимір важливості ознак. Ціль - ціна. Звіт ділиться на 3 розділи: по 1 на алгоритм. У кожному розділі міститься загальний аналіз ознак та аналіз кореляції. Також аналіз поділений на аналіз числових чисел та строкових.

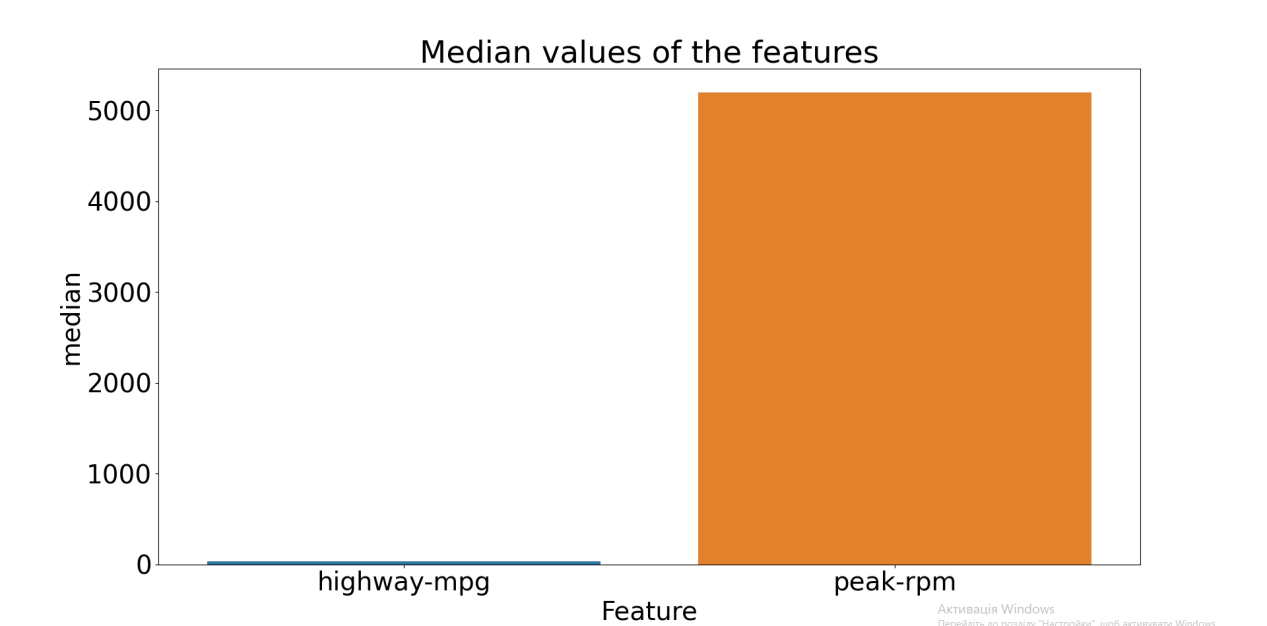
## Однофакторний вибір ознак

Алгоритм визначив важливими такі ознаки: витрати бензину на шосе, тип двигуна, пікові обороти, система заправки(від найбільш важливих до найменш). Зробимо загальний аналіз ознак. Почнемо з числових.

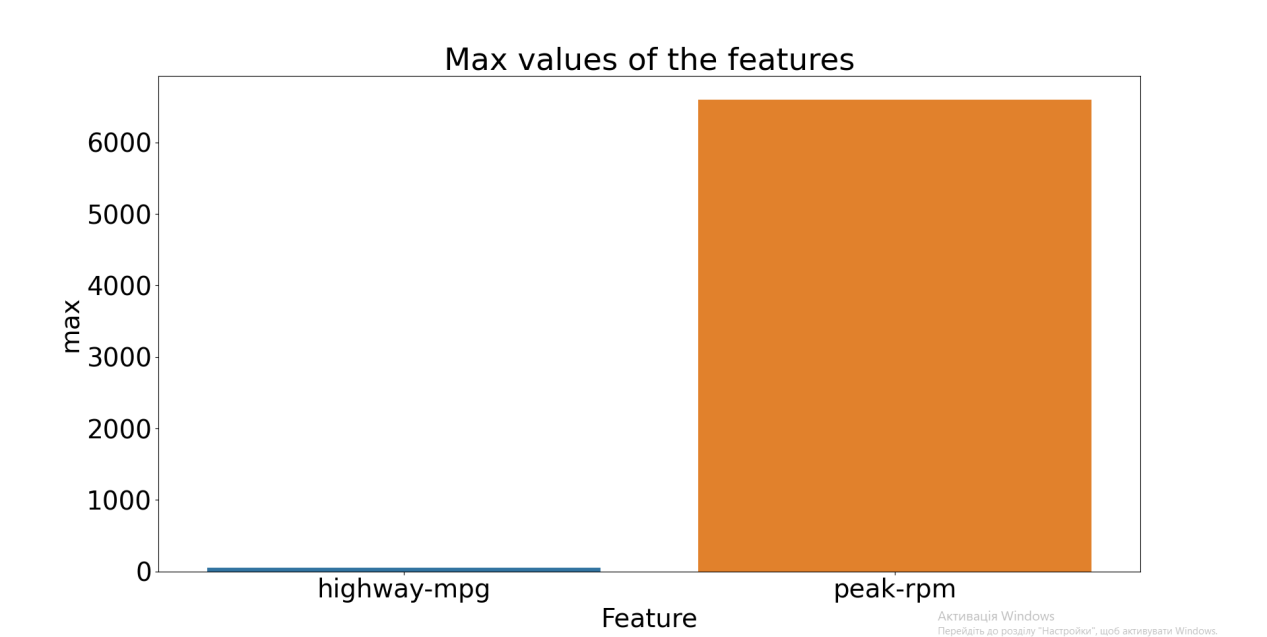
Середнє значення оборотів дорівнює близько 5000 оборотів. Значення витрати бензину дорівнює 32.



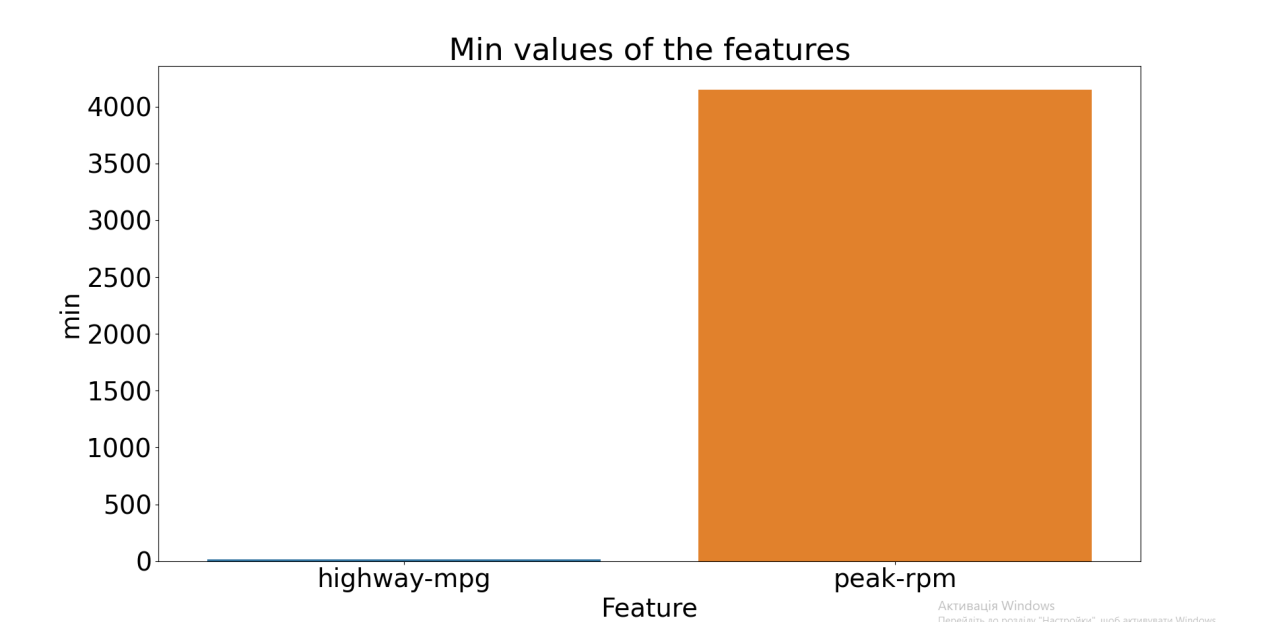
Медіана оборотів дорівнює близько 5000 оборотів. Значення витрати бензину дорівнює 32.



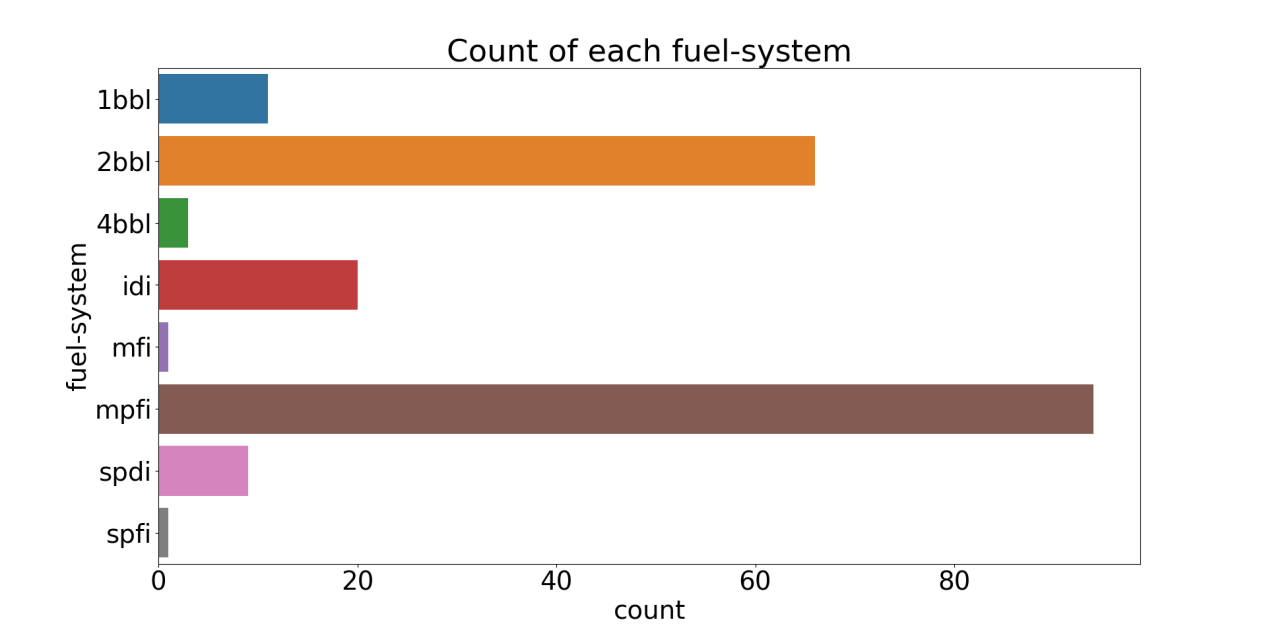
Максимальне значення оборотів дорівнює близько 6300 оборотів. Значення витрати бензину дорівнює 54.



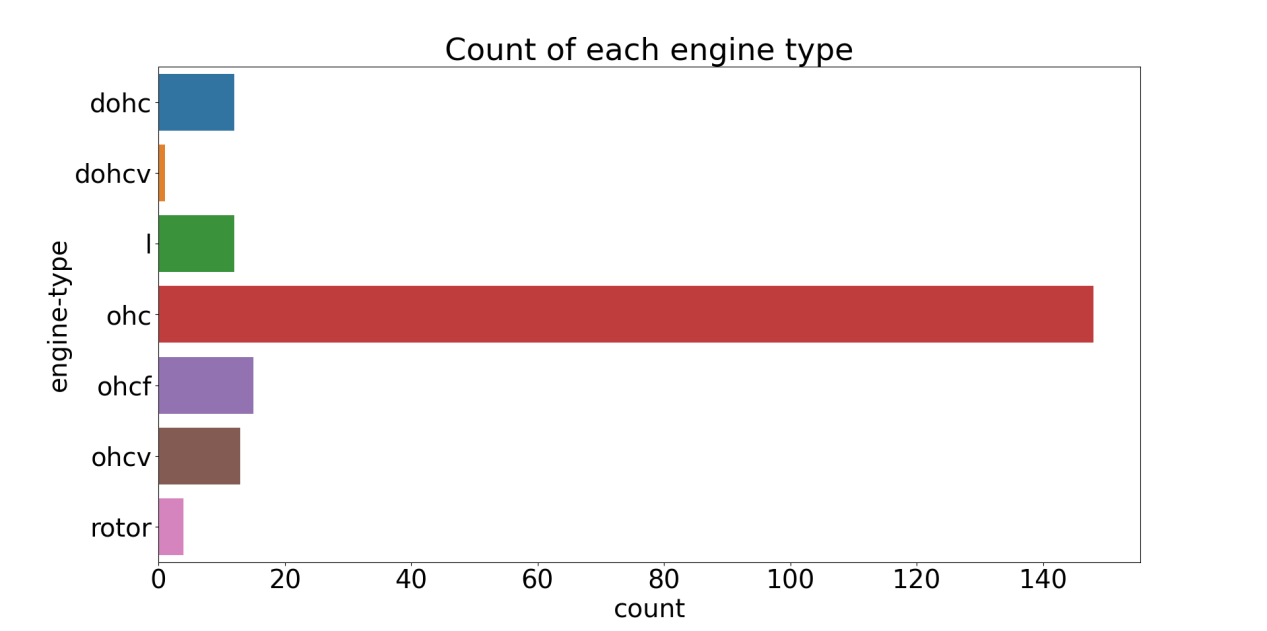
Мінімальне значення оборотів дорівнює близько 4150 оборотів. Значення витрати бензину дорівнює 18.



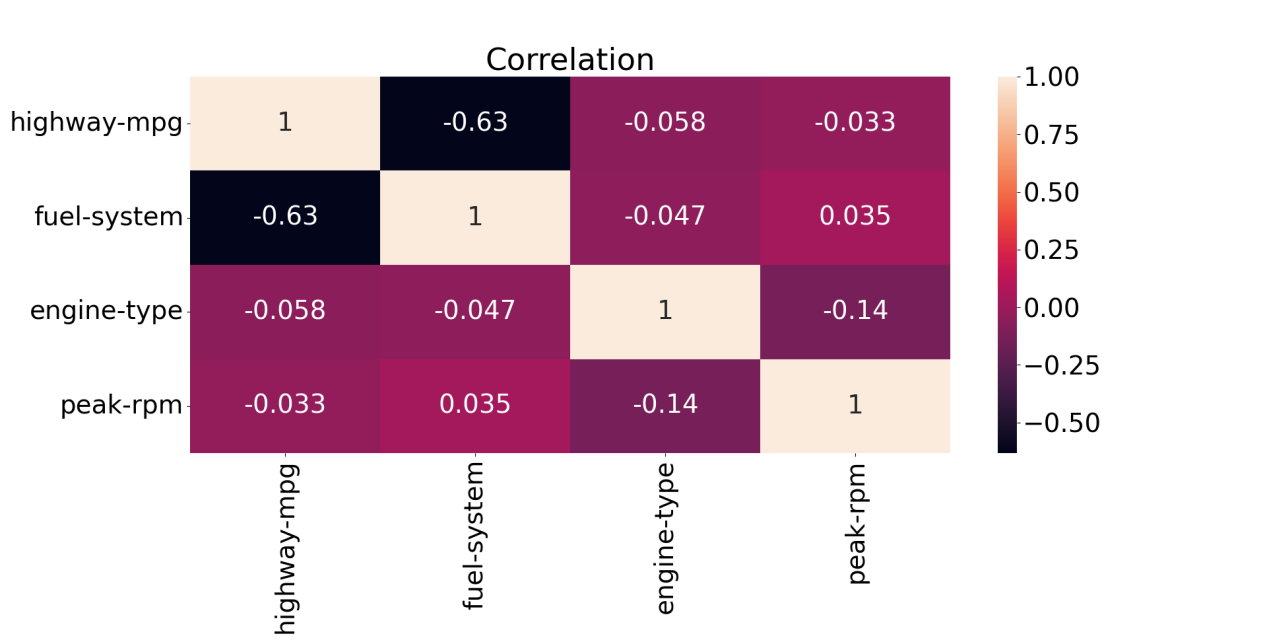
Перейдемо до аналізу строкових ознак. Можемо побачити на графіку, що найбільшою популярністю користуються системи заправки mpfi та 2bbl. Популярність інших доволі незначна.



На графіку бачимо, що надзвичайною популярністю користується тип двигуна ohc і його кількість перевищує 140 одиниць. У всіх інших кількість не доходить навіть до 20 одиниць.



Проаналізуємо графік кореляцій. Єдині ознаки, які мають позитивну кореляцію - це системи заправки та пікові оберти та мають значення 0.035. Усі інші мають негативні кореляції, але які більше за -0.1 та менші за 0, тобто доволі незначні, окрім системи заправки та витрати палива на шосу, які мають кореляцію -0.63.



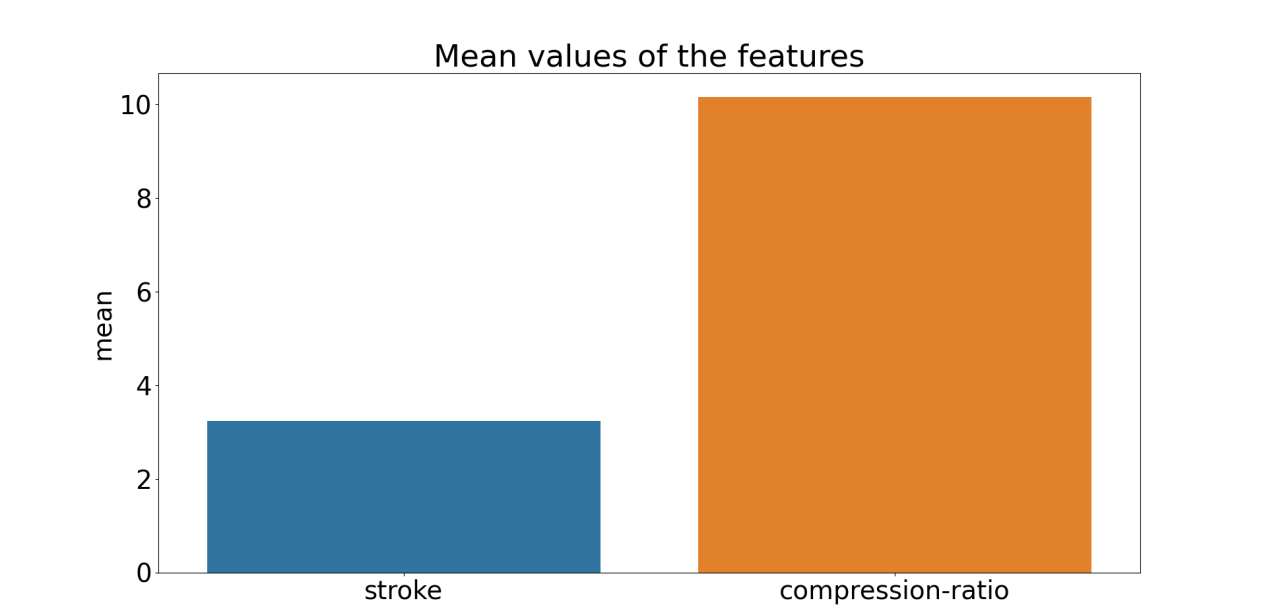
## Однофакторний вибір ознак. Висновок

За результатами методу можемо зробити висновок, що при вирішення того, які машини продавати, треба звернути увагу на такі ознаки: витрати бензину на шосе, тип двигуна, пікові обороти, система заправки. Середнє значення витрат бензину дорівнює близько 32, а пікових оборотів близько 5000. Найбільш популярні системи заправки - 2bbl та mpfi. Найбільш популярний тип двигунів - ohc. Між ознаками відсутня значуща кореляція, окрім кореляції між системою заправки та витратою бензину, які мають кореляцію -0.63.

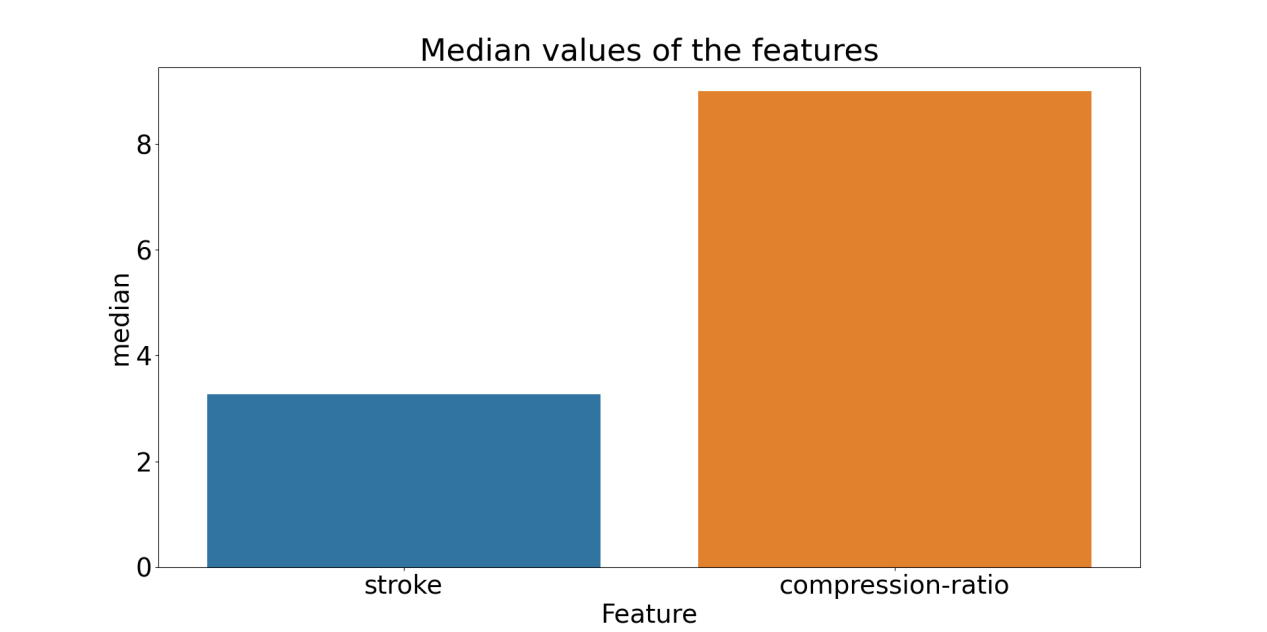
## Рекурсивний алгоритм

Алгоритм визначив важливими такі ознаки: хід двигуна, ступінь стиснення, кількість циліндрів та тип потрапляння повітря у двигун(від найбільш важливих до найменш). Зробимо загальний аналіз ознак. Почнемо з числових.

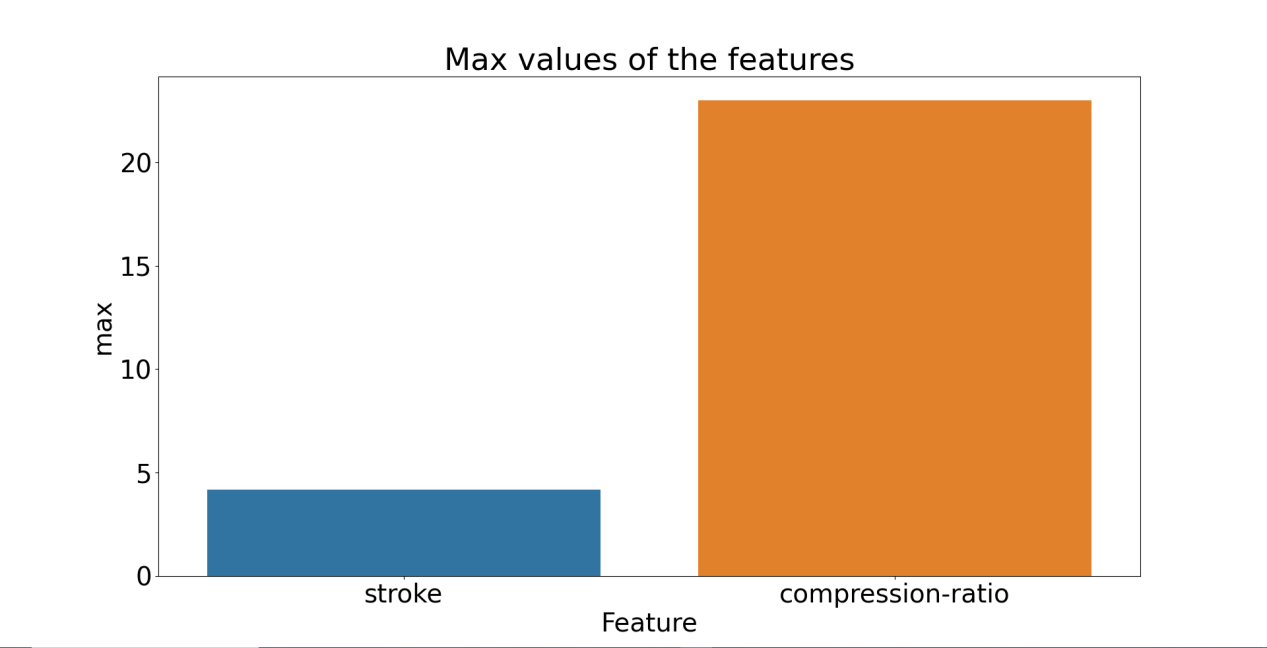
Середнє значення ступеню стистення дорівнює близько 10, а ходу двигуна близько 4.



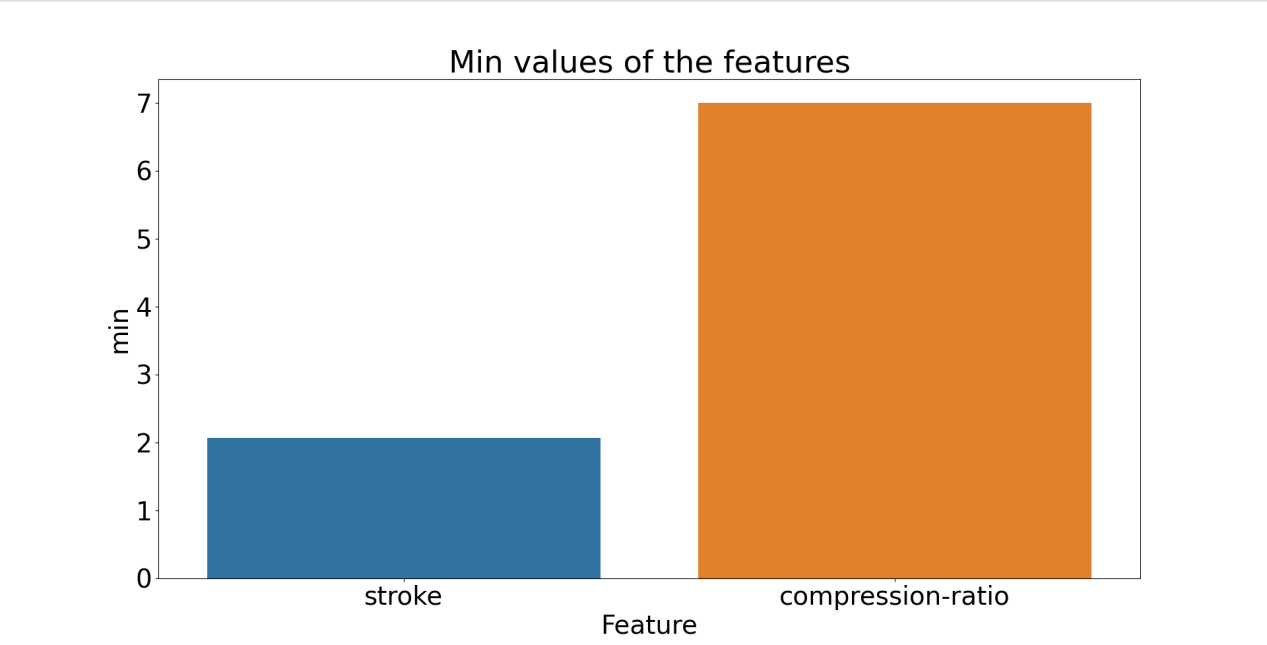
Медіана ходу двигуна дорівнює близько 4. Медіана ступеню стиснення дорівнює близько 9.



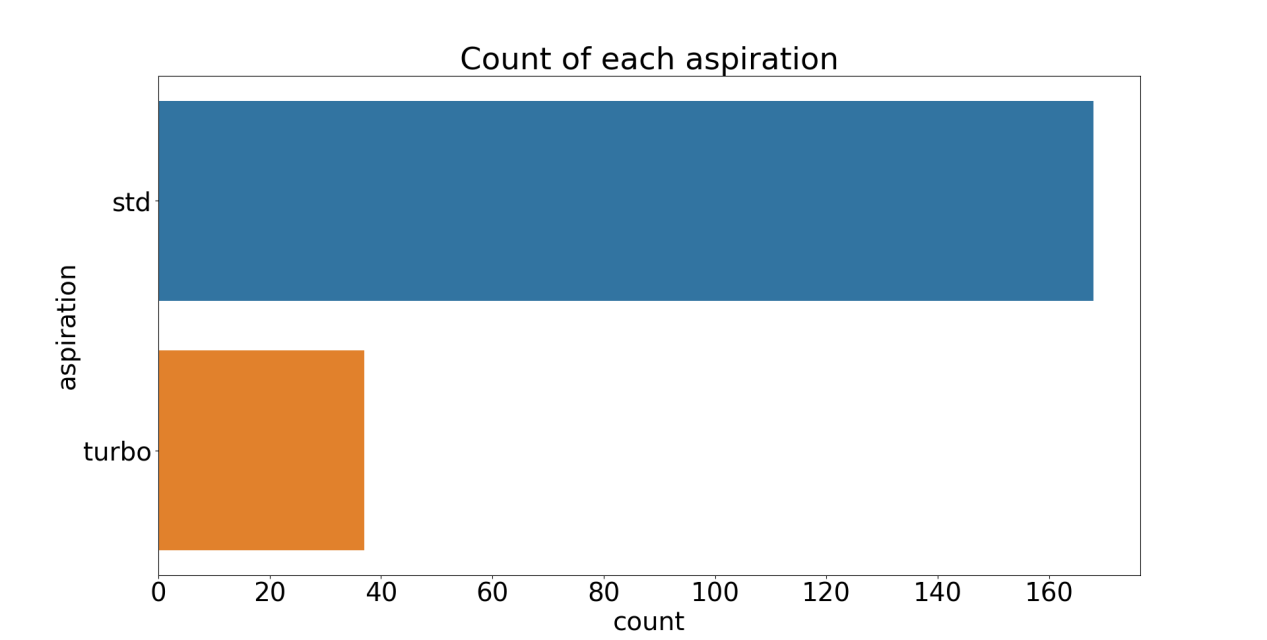
Максимальне значення cтупеню стиснення дорівнює близько 23, а ходу двигуна близько 4.



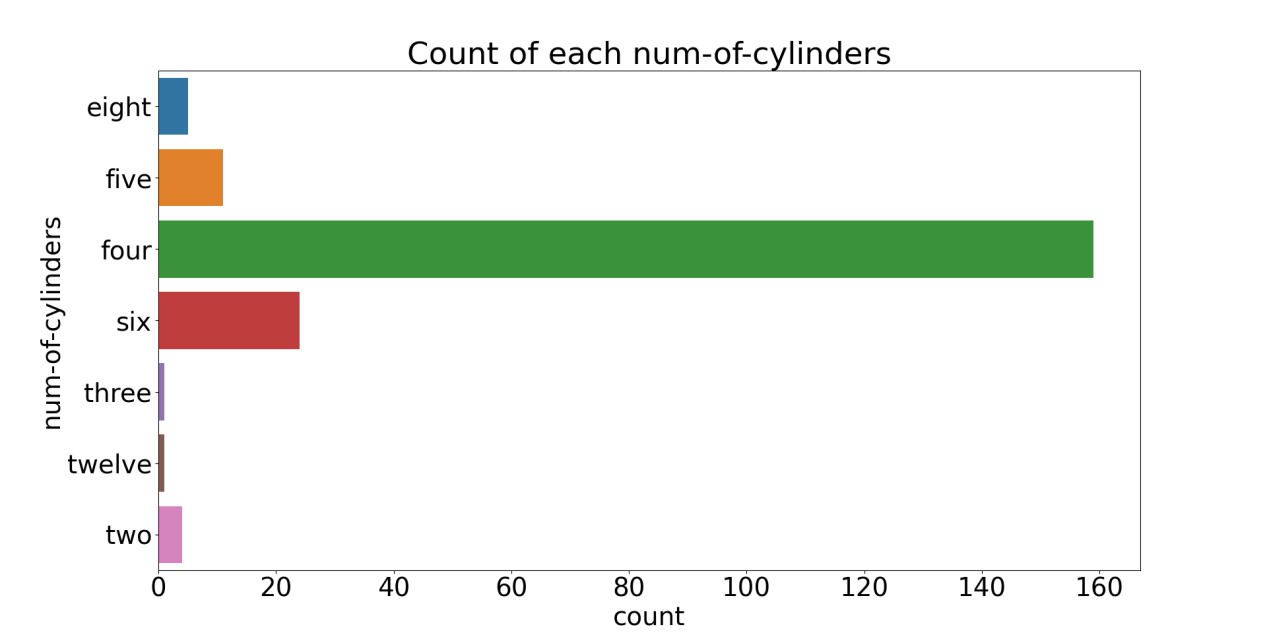
Мінімальне значення ступеню стиснення дорівнює близько 7, а ходу двигуна близько 2.



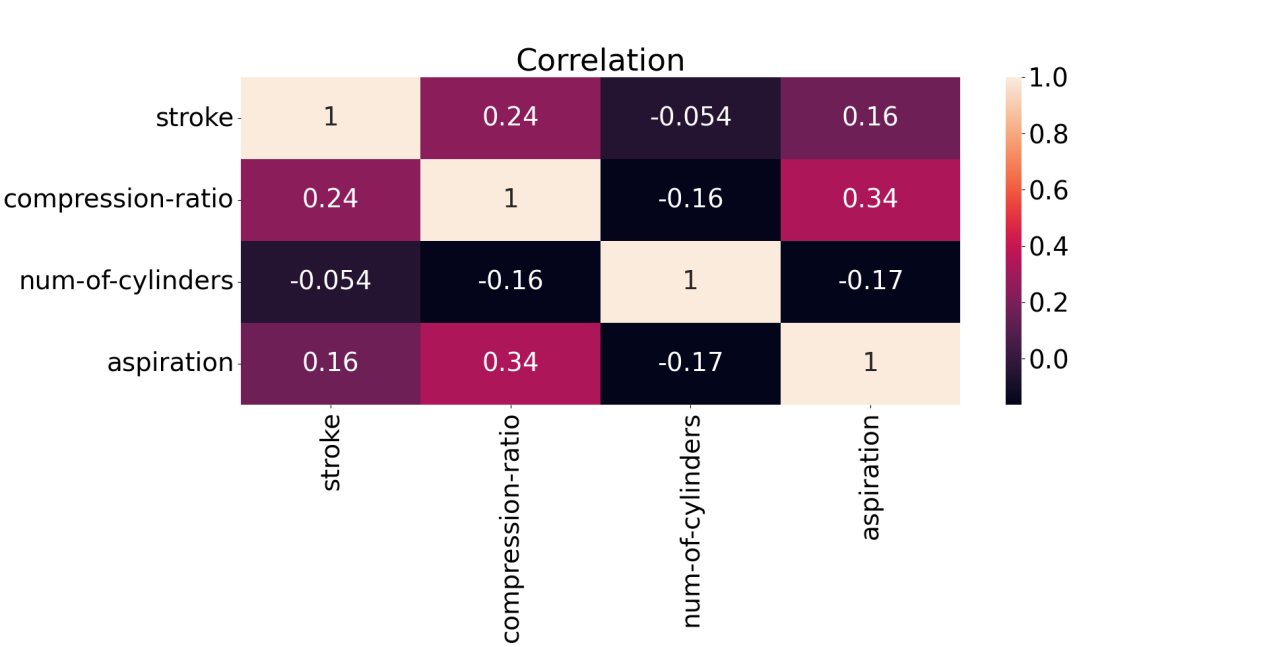
Перейдемо до аналізу строкових ознак. Можемо побачити, що найбільш популярною системою постачання повітря є std, турбо значно уступає.



На графіку бачимо, що 4 циліндри мають надзвичайну перевагу. Усі інші типи значно уступають та не користуються такою популярністю.



Проаналізуємо графік кореляцій. Є як позитивні, так і негативні коефіцієнти, а значення більшость не перевищують +-0.2. Але кореляція системи постачання повітря та ступеню стиснення дорівнює 0.34, а кореляція ступеню стиснення та ходу двигуна дорівнює 0.24. Загально кореляція не має значних значень.



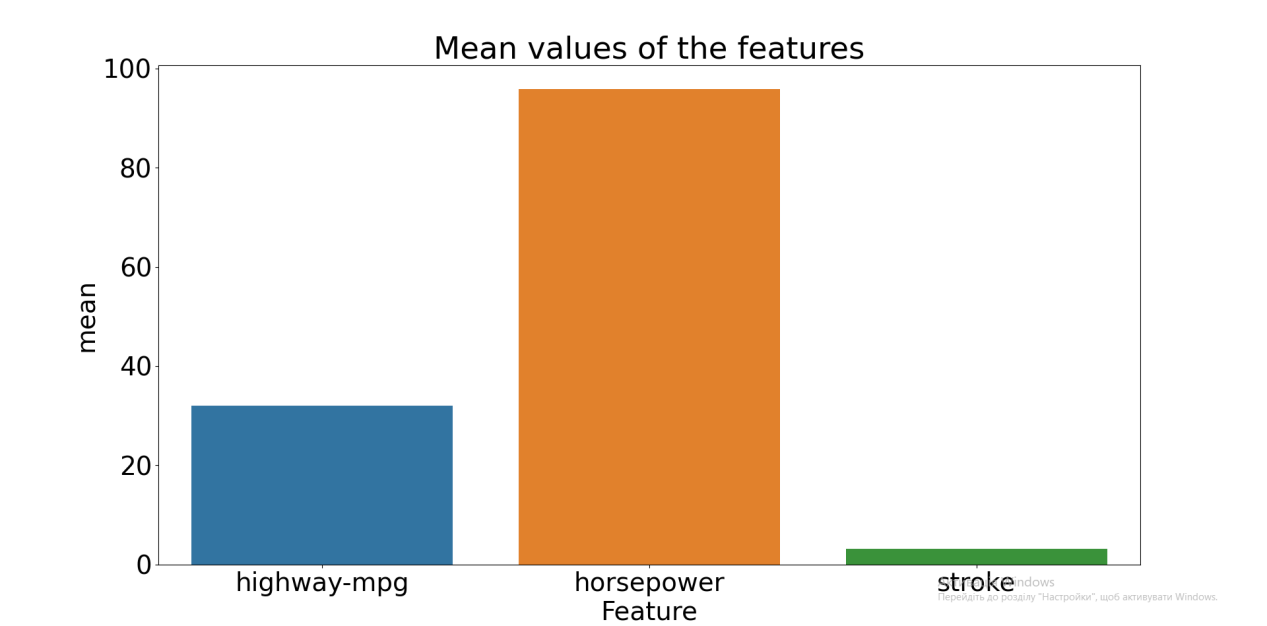
## Рекурсивний алгоритм. Висновок

За результатами методу можемо зробити висновок, що при вирішення того, які машини продавати, треба звернути увагу на такі ознаки: кількість циліндрів, система постачання повітря, ступінь стиснення та хід двигуна. Середнє значення ступеню стистення дорівнює близько 10, а ходу двигуна близько 4. Найбільш популярна кількість циліндрів - 4, а система постачання повітря - std. Між ознаками відсутня значуща кореляція, але є кореляція між системою постачання повітря та ступенем стиснення, яка дорівнює 0.34 та кореляція між ступенем стиснення та ходом двигуна, яка дорівнює 0.24.

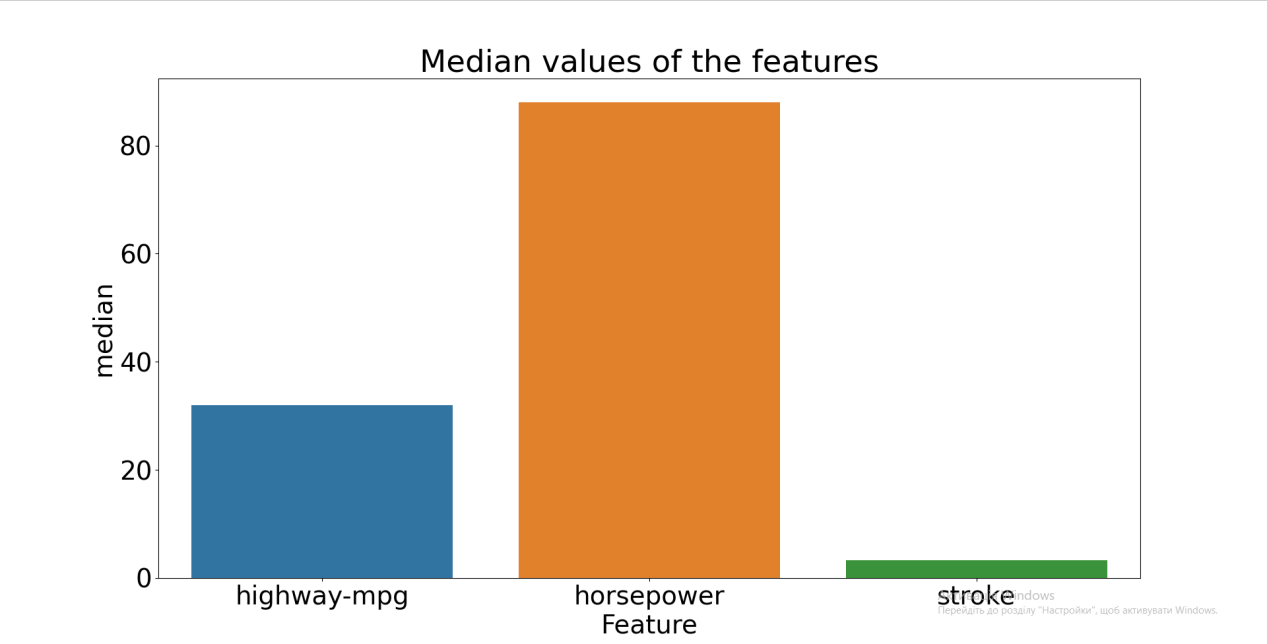
## Алгоритм відбору ознак

Алгоритм визначив важливими такі ознаки: витрати палива на шосе, кінські сили, хід двигуна та тип двигуна(від найбільш важливих до найменш). Зробимо загальний аналіз ознак. Почнемо з числових.

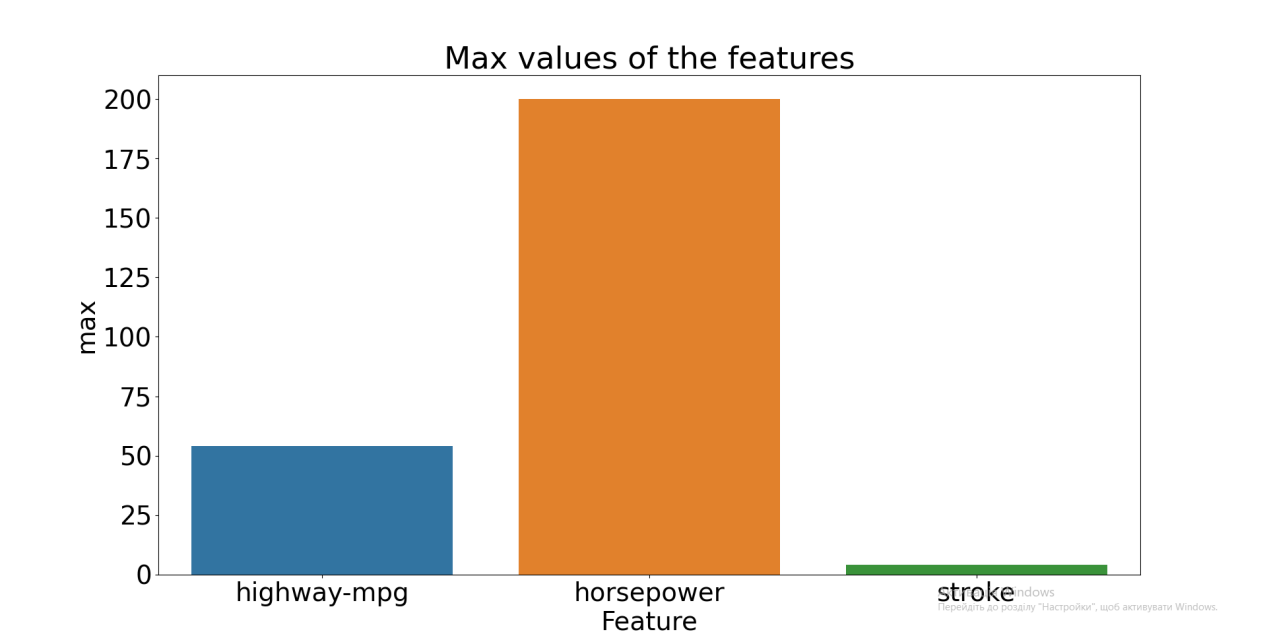
Середнє значення витрати палива на шосе становить близько 30, кінських сил близько 95, а ходу двигуна близько 3.



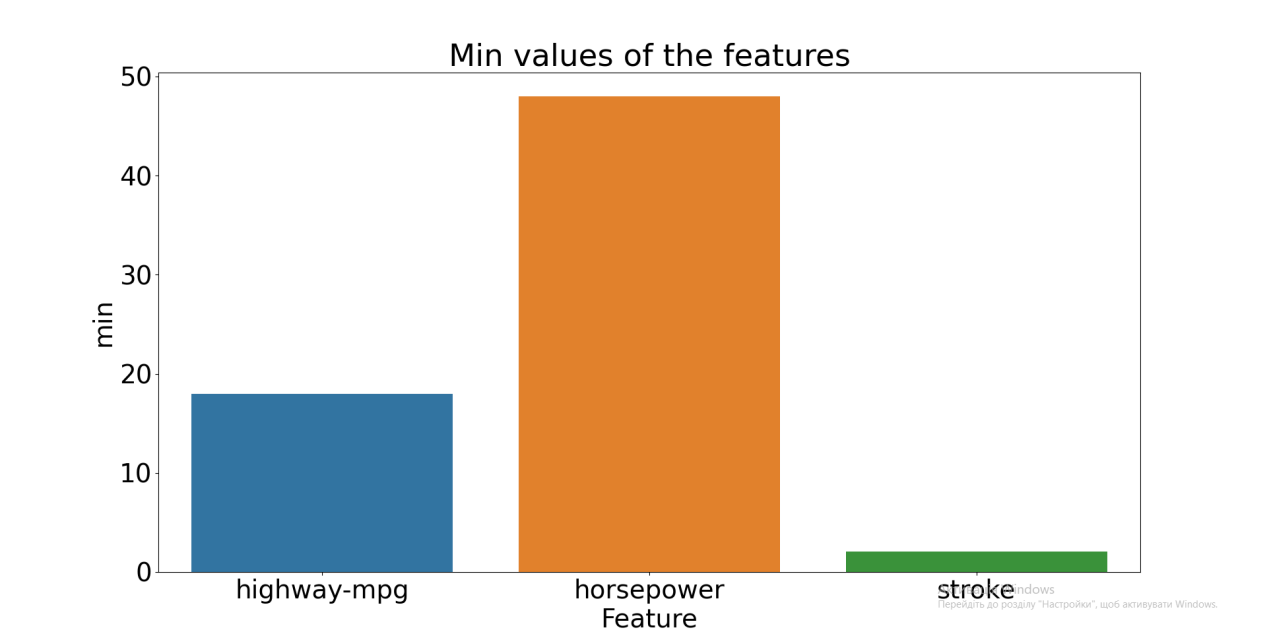
Медіана витрат палива на шосе становить близько 30, кінських сил близько 85, а ходу двигуна близько 3.



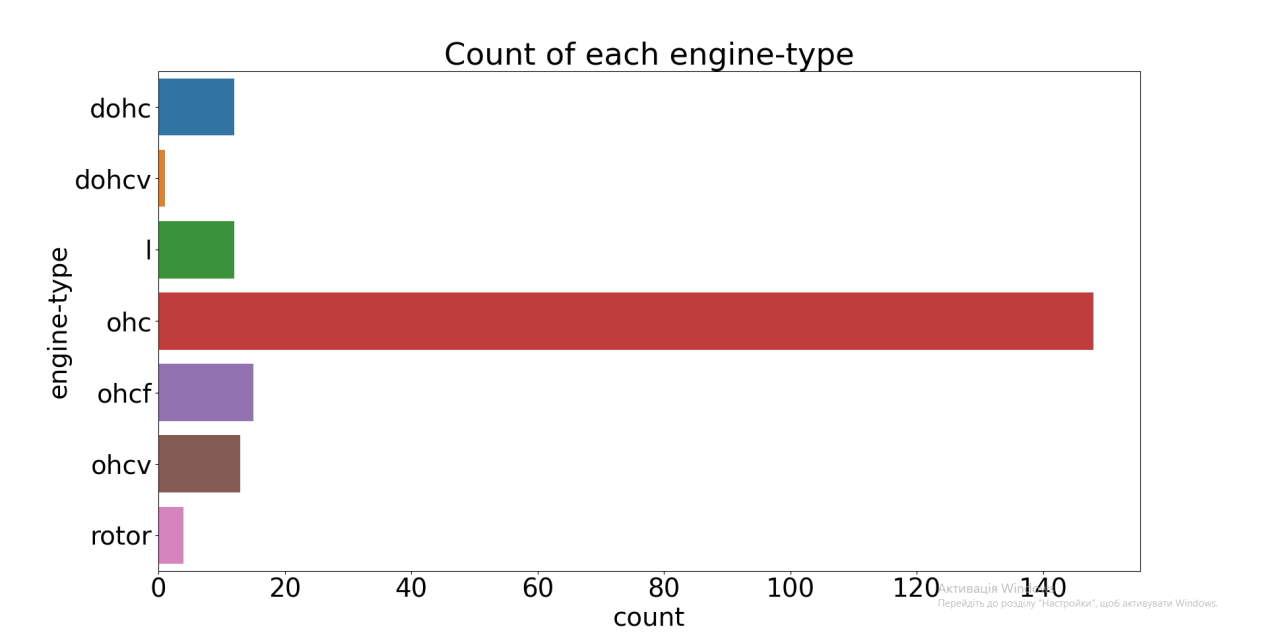
Максимальне значення витрат палива на шосе становить близько 50, кінських сил близько 200, а ходу двигуна близько 3.



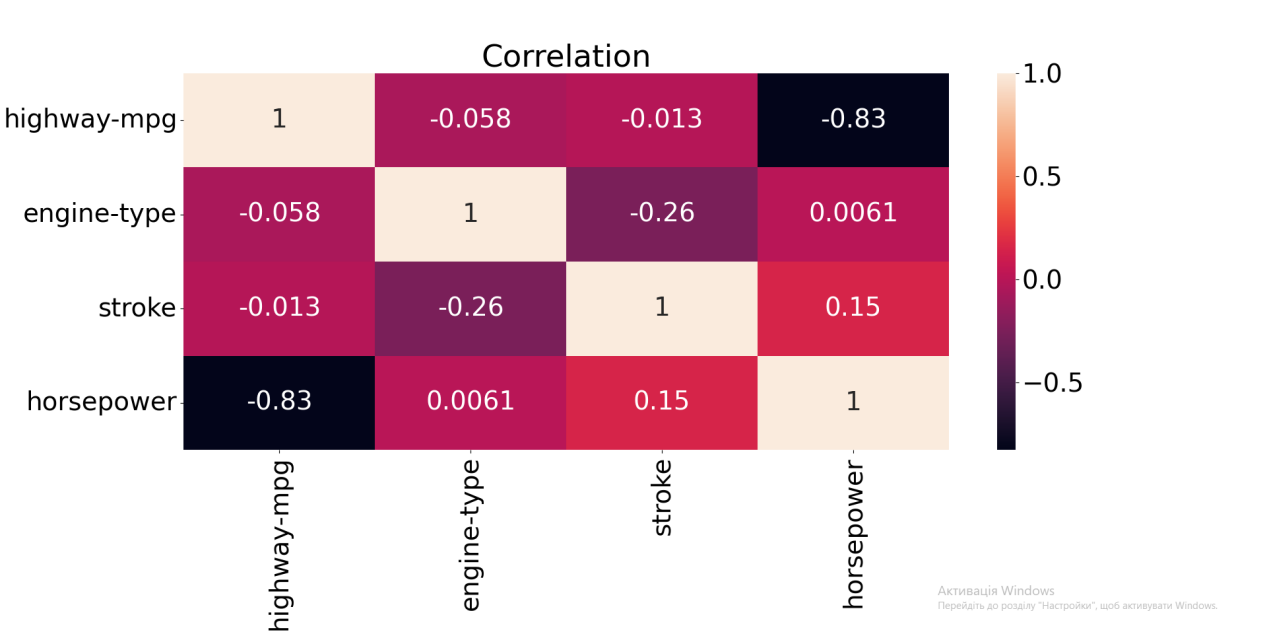
Мінімальне значення витрат палива на шосе становить близько 18, кінських сил близько 48, а ходу двигуна близько 3.



Перейдемо до аналізу строкових ознак. Можемо побачити, що найбільш популярним типом двигуна є ohc, усі інші значно уступають йому.



Проаналізуємо графік кореляцій. Є як позитивні, так і негативні коефіцієнти, а значення більшость не перевищують +-0.2. Але кореляція кінських сил та витрат бензину на шосу дорівнює -0.83, що є надзвичайно великим числом. Такоє кореляція між типом двигуна та ходом двигуна дорівнює -0.26.



## Алгоритм відбору ознак. Висновок

За результатами методу можемо зробити висновок, що при вирішення того, які машини продавати, тСокеба звернути увагу на такі ознаки: витрати палива на шосе, кінські сили, хід двигуна та тип двигуна. Середнє значення витрати палива на шосе становить близько 30, кінських сил близько 95, а ходу двигуна близько 3. Найбільш популярний тип двигуна - ohc. Між більшістю ознак відсутня значуща кореляція, але є значна кореляція кінськими силами та витратами бензину, що дорівнює -0.83.

Різні алгоритми виділяють різні ознаки, які впливають на кінцеву ціну автомобіля, але серед них є ознаки, що були обрані у 2 алгоритмах: витрата бензину на шосе, тип двигуна та хід двигуна. На них варто звернути особливу увагу при виборі автомобілів для продажу. Також була помічена значна кореляція між такими ознаками: кінські сили та витрати бензину на шосе з коефіцієнтом -0.83, система постачання повітря та ступінь стиснення з коефіцієнтом 0.34 та витрата бензину на шосе та система заправки з коефіцієнтом -0.63.