**Аналітичний звіт основних ознак ринку вживаних автомобілів**

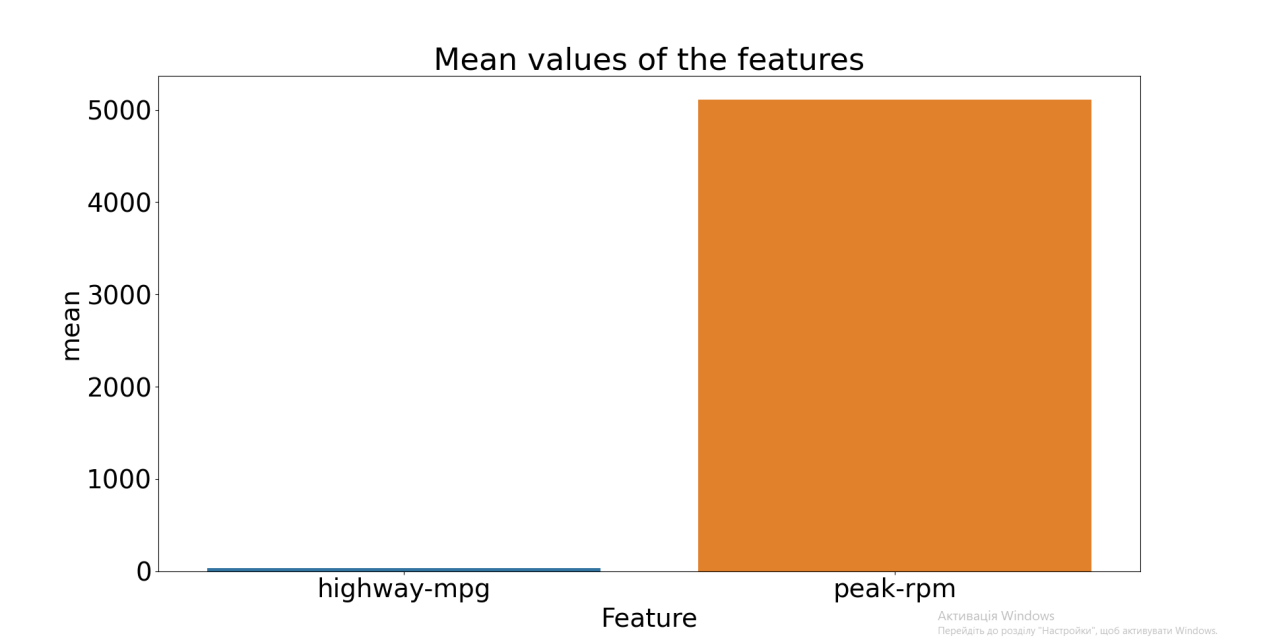
**Вступ**

### 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.

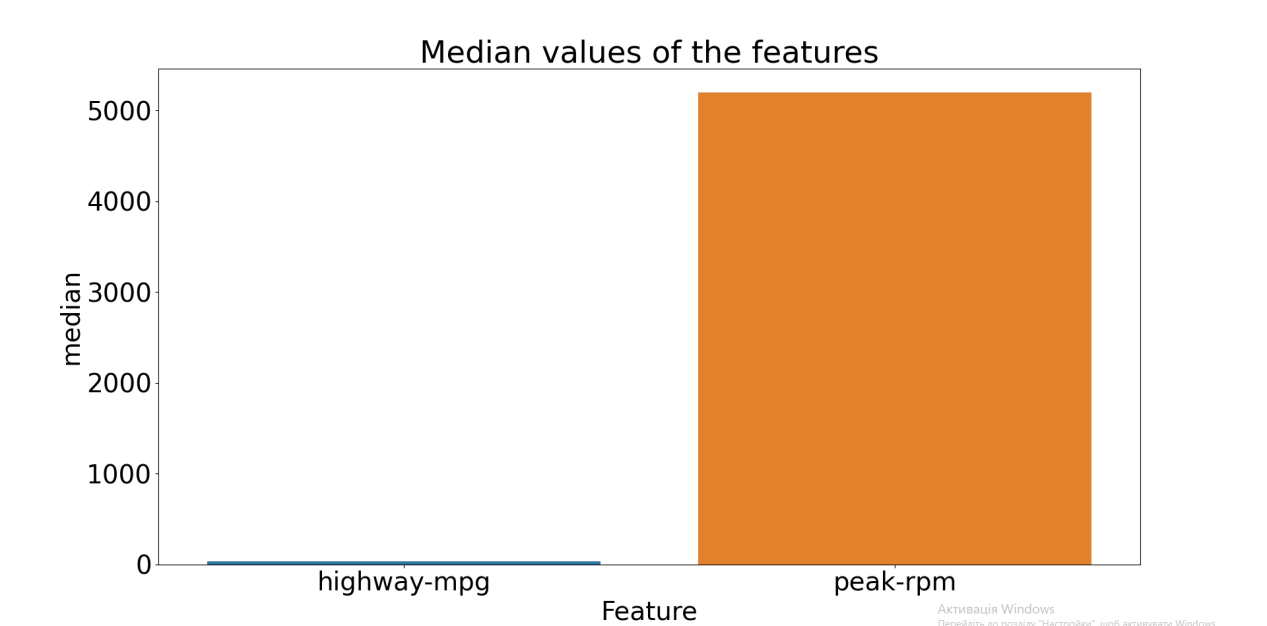
## Однофакторний вибір ознак

Алгоритм ідентифікував наступні ключові параметри автомобіля: споживання пального на шосе, тип двигуна, пікові оберти, та система заправки, розташовані від найбільш важливих до менш важливих. Давайте проведемо загальний аналіз цих характеристик, розпочавши з числових значень.

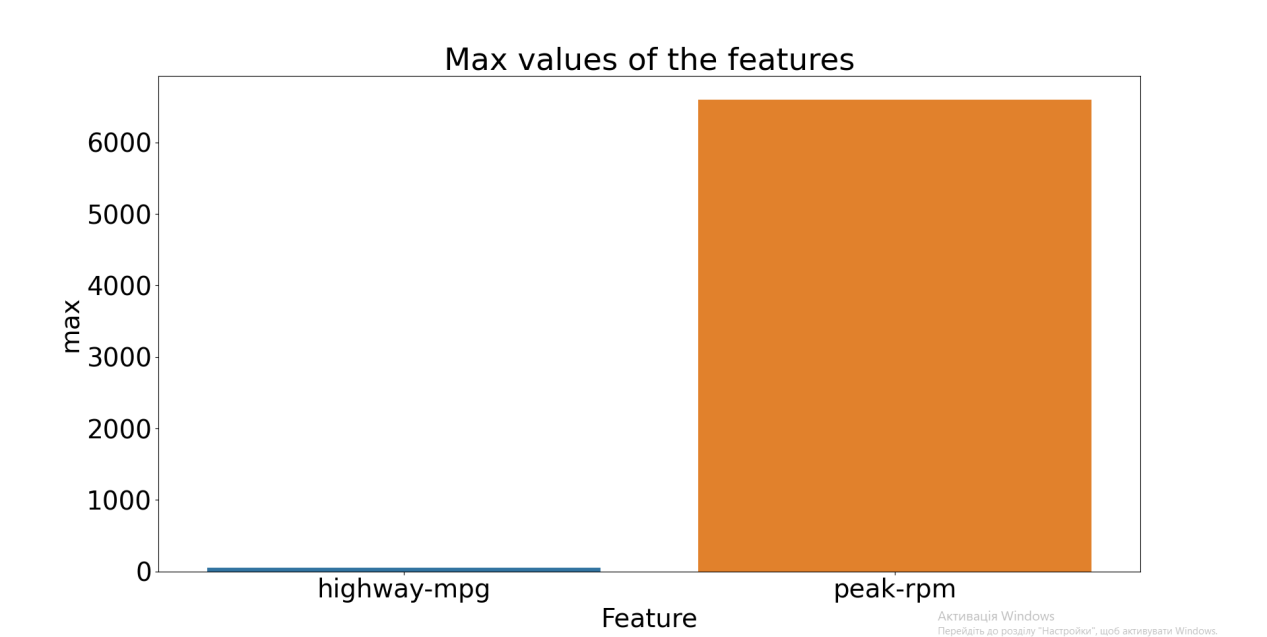
Середнє значення обертів становить приблизно 5000 обертів. Витрати пального складають 32 одиниці.



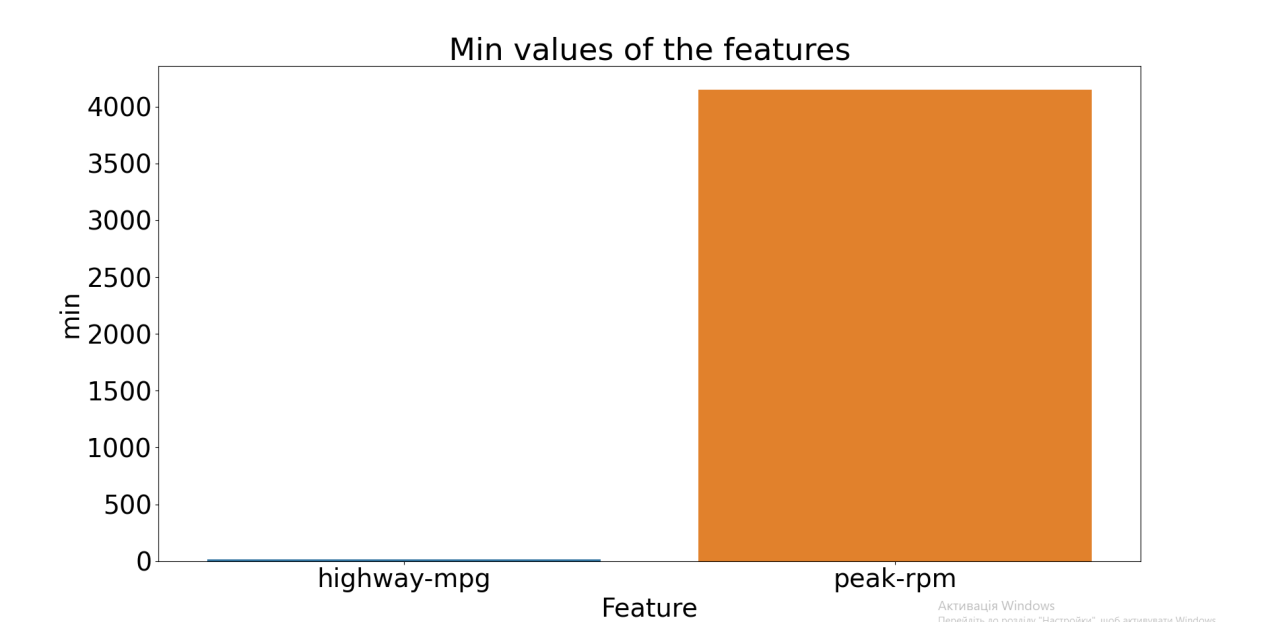
Медіана оборотів дорівнює близько 5000 оборотів. Значення витрати бензину дорівнює 32.



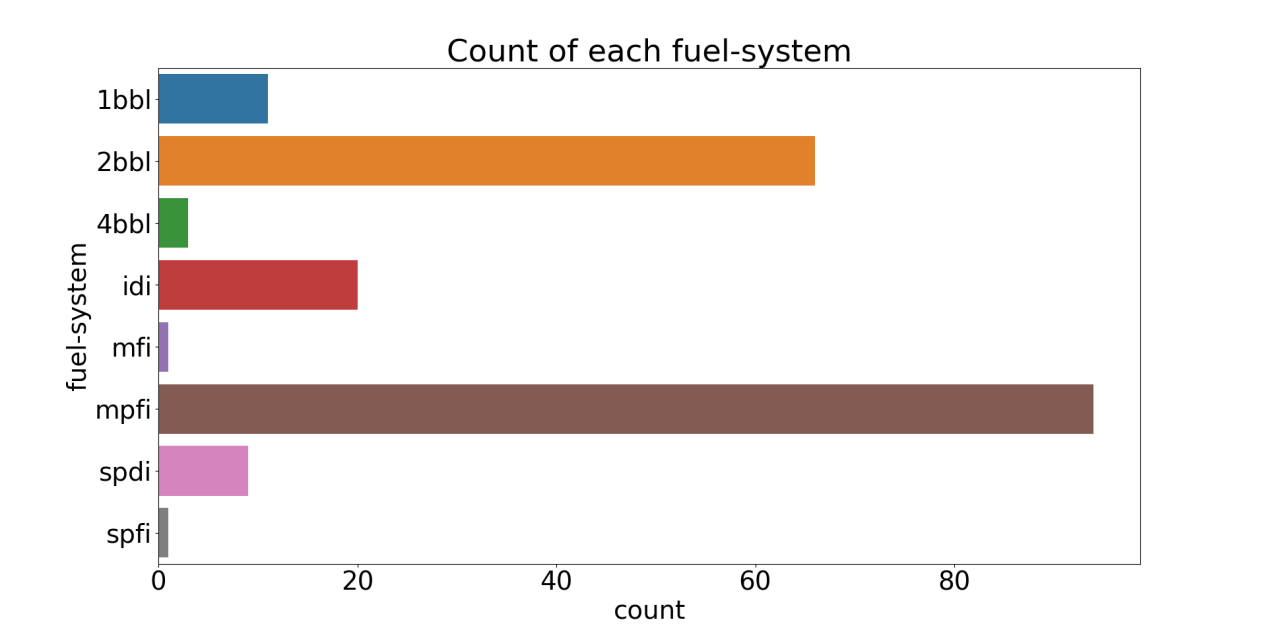
Максимальне значення оборотів дорівнює близько 6300 оборотів. Значення витрати бензину дорівнює 54.



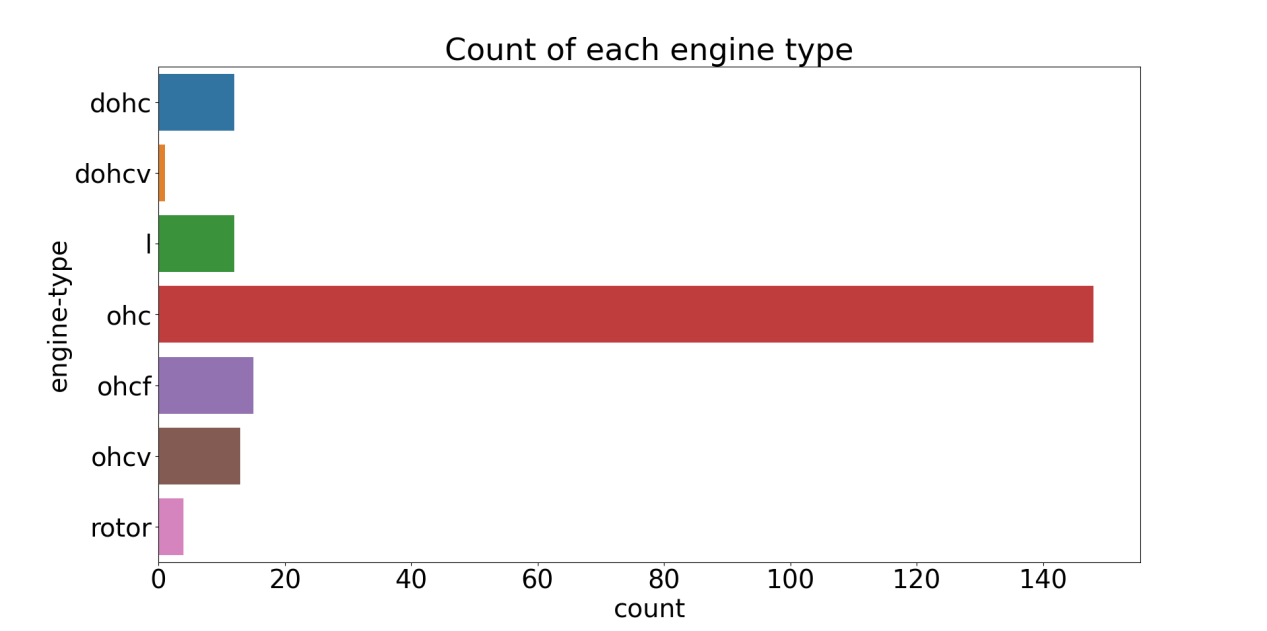
Мінімальне значення оборотів дорівнює близько 4150 оборотів. Значення витрати бензину дорівнює 18.



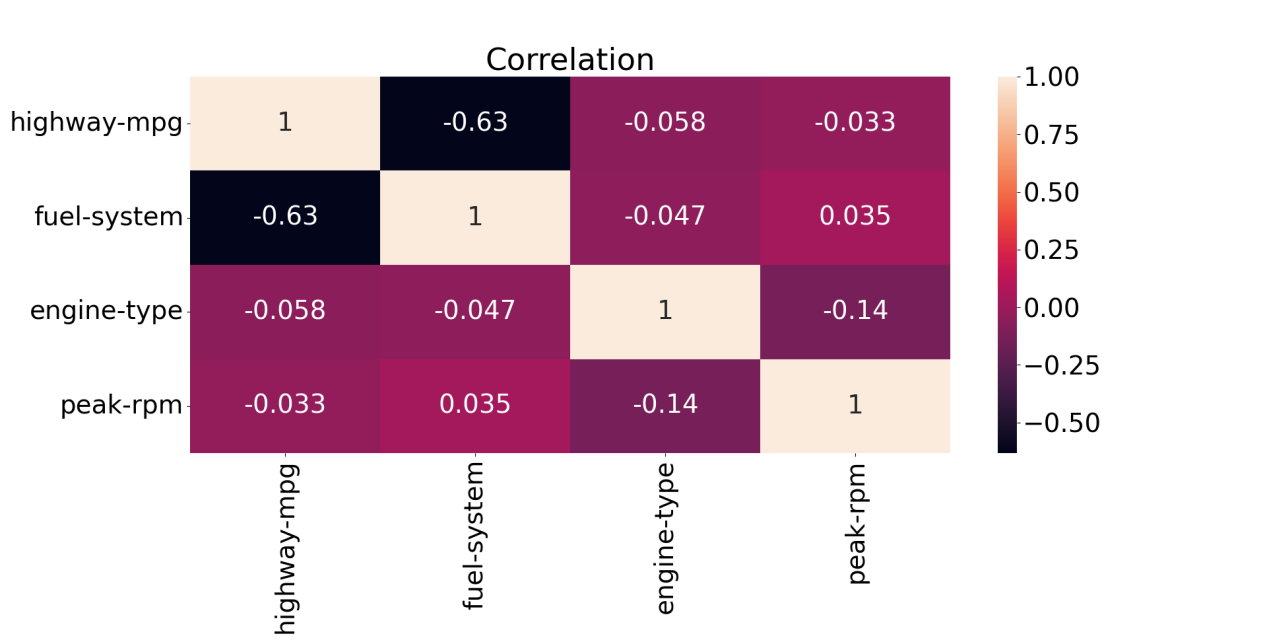
Давайте розглянемо аналіз категорійних ознак. З графіка видно, що системи заправки mpfi та 2bbl користуються найбільшою популярністю, в той час як інші системи мають значно менший рівень популярності.



За графіком видно, що тип двигуна ohc надзвичайно популярний, і його кількість перевищує 140 одиниць. У всіх інших випадках кількість не досягає навіть 20 одиниць.



Після аналізу графіку кореляцій виявлено, що єдині ознаки з позитивною кореляцією - це системи заправки та пікові оберти, і ця кореляція становить 0.035. Усі інші виявили негативні кореляції, але значення більше -0.1 та менше 0, що вказує на досить слабку взаємозв'язаність. Однак варто відзначити, що існує помітна негативна кореляція між системою заправки та витратами палива на шосе, яка становить -0.63.



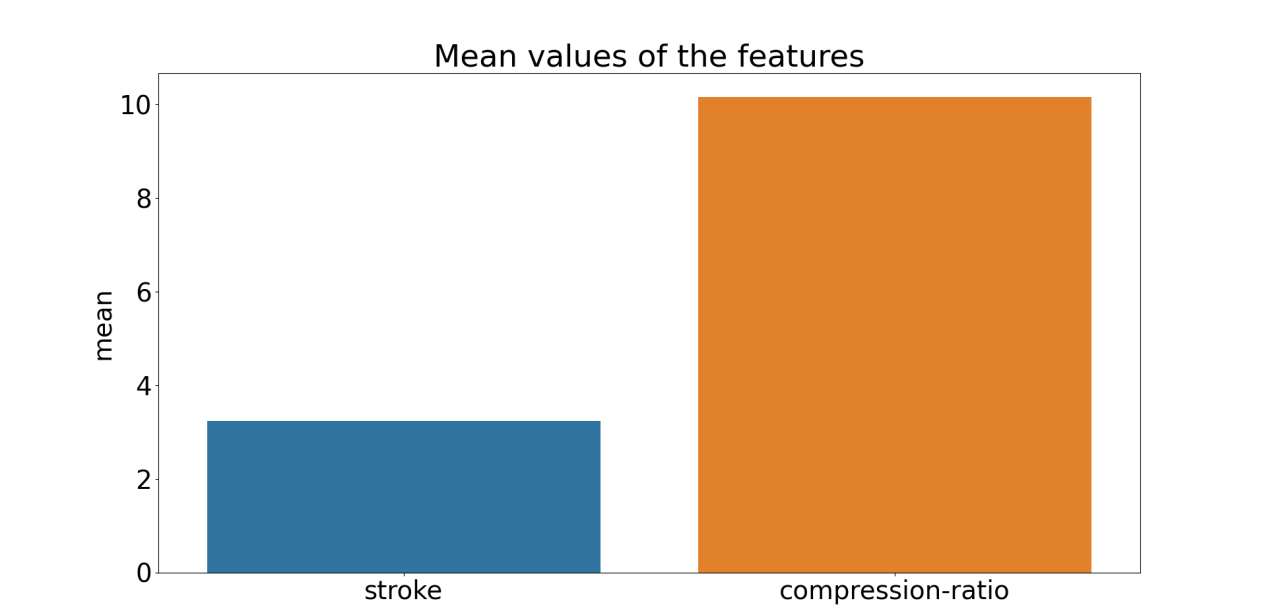
## Однофакторний вибір ознак. Висновок

За результатами аналізу можна зробити висновок, що при виборі автомобілів для продажу важливо звертати увагу на такі характеристики: витрати пального на шосе, тип двигуна, пікові оберти та система заправки. Середнє значення витрат пального складає приблизно 32, а середнє значення пікових обертів близько 5000. Системи заправки 2bbl та mpfi є найбільш популярними. Тип двигуна ohc має найвищу популярність серед інших. Важливо враховувати взаємозв'язок між системою заправки та витратами пального, оскільки вони мають помітну кореляцію -0.63.

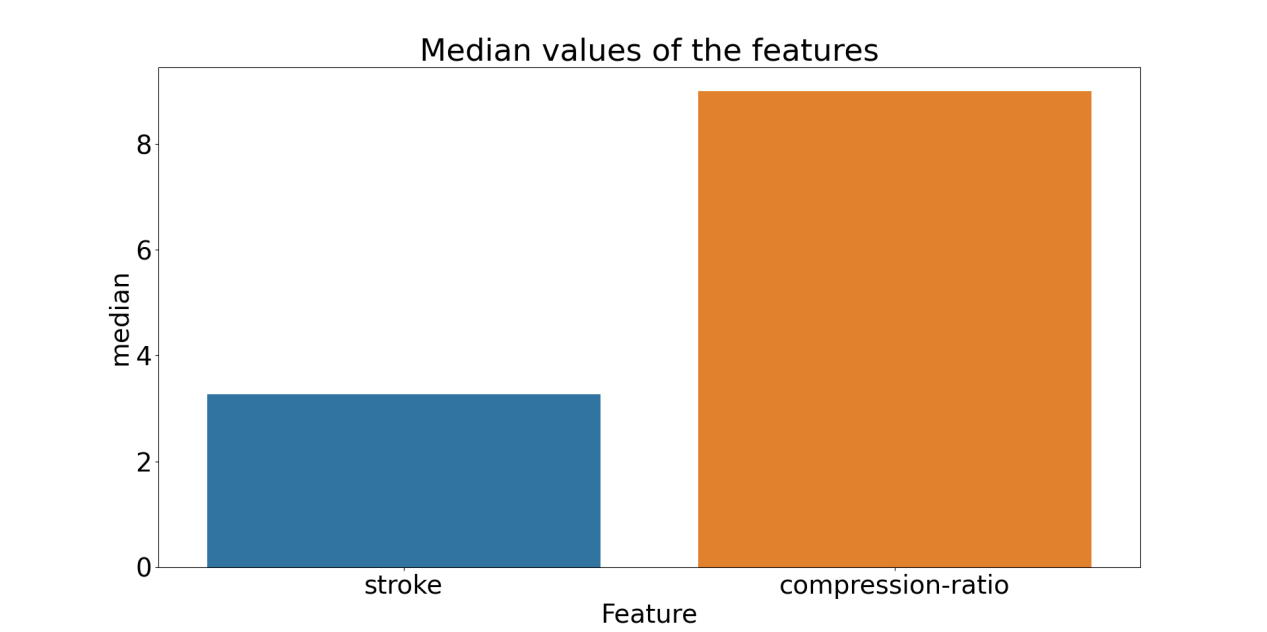
## Рекурсивний алгоритм

Алгоритм визначив наступні ключові характеристики: хід двигуна, ступінь стиснення, кількість циліндрів та тип потрапляння повітря у двигун, розташовані від найбільш важливих до менш важливих. Давайте проведемо загальний аналіз цих ознак, розпочавши з числових значень.

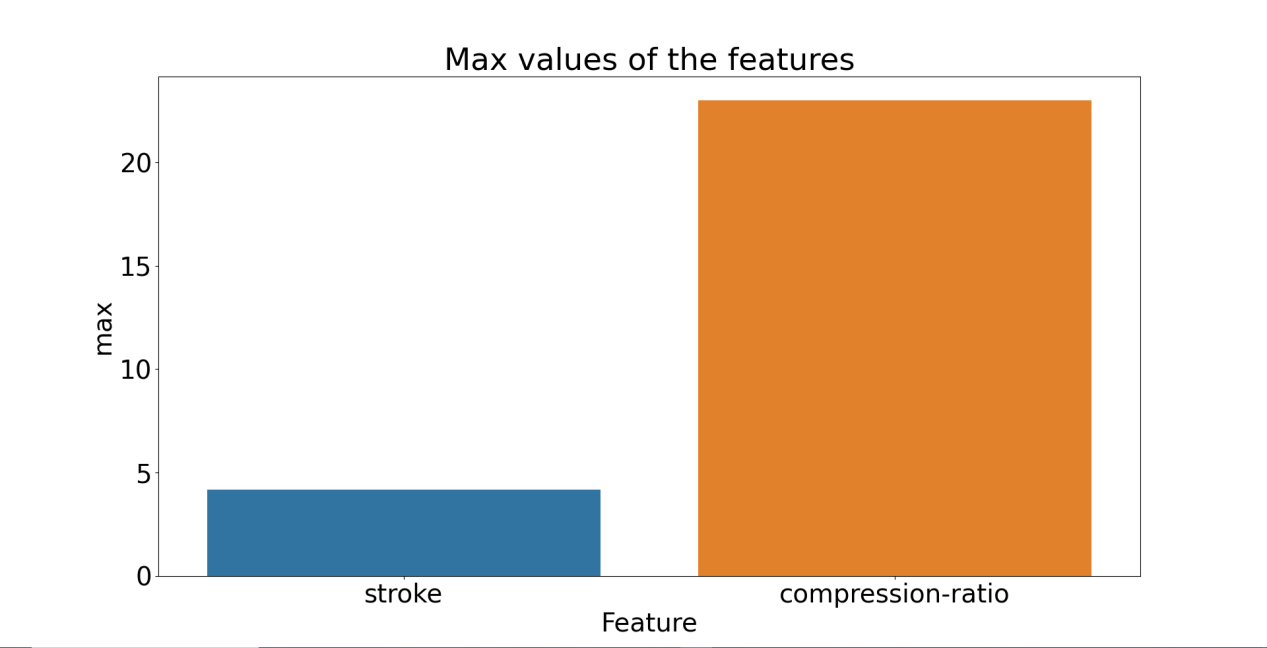
Середнє значення ступеня стиснення становить близько 10, а ходу двигуна приблизно 4.



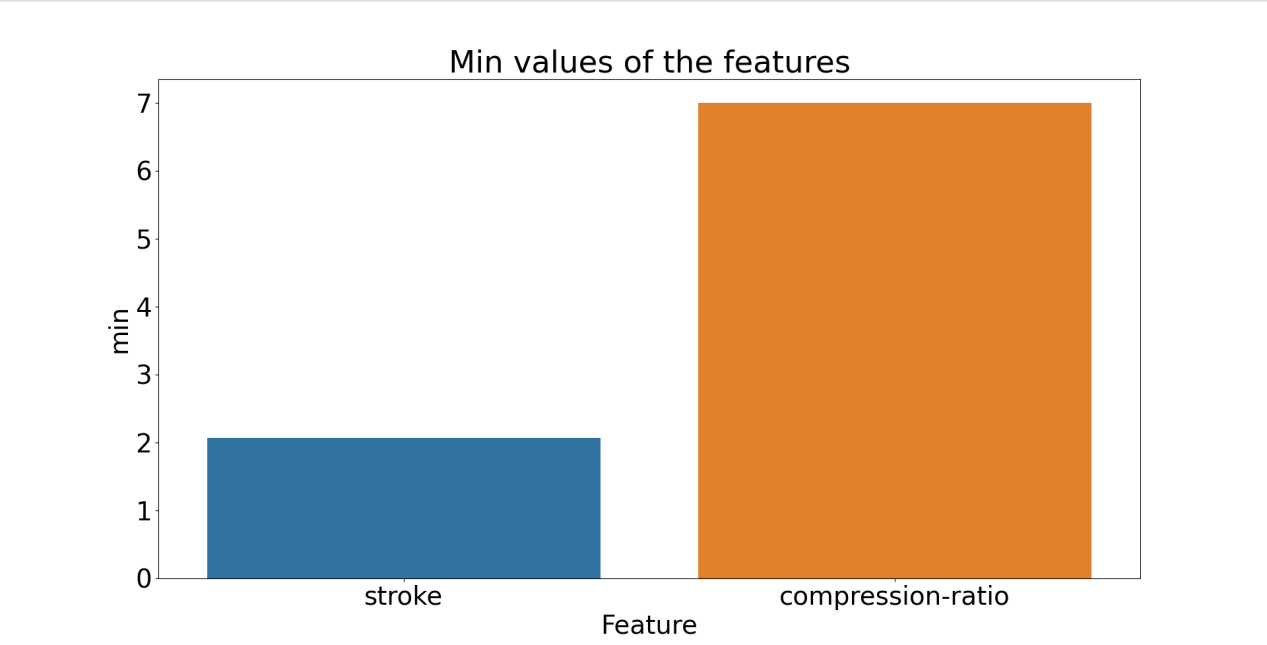
Медіана ходу двигуна дорівнює близько 4. Медіана ступеню стиснення дорівнює близько 9.



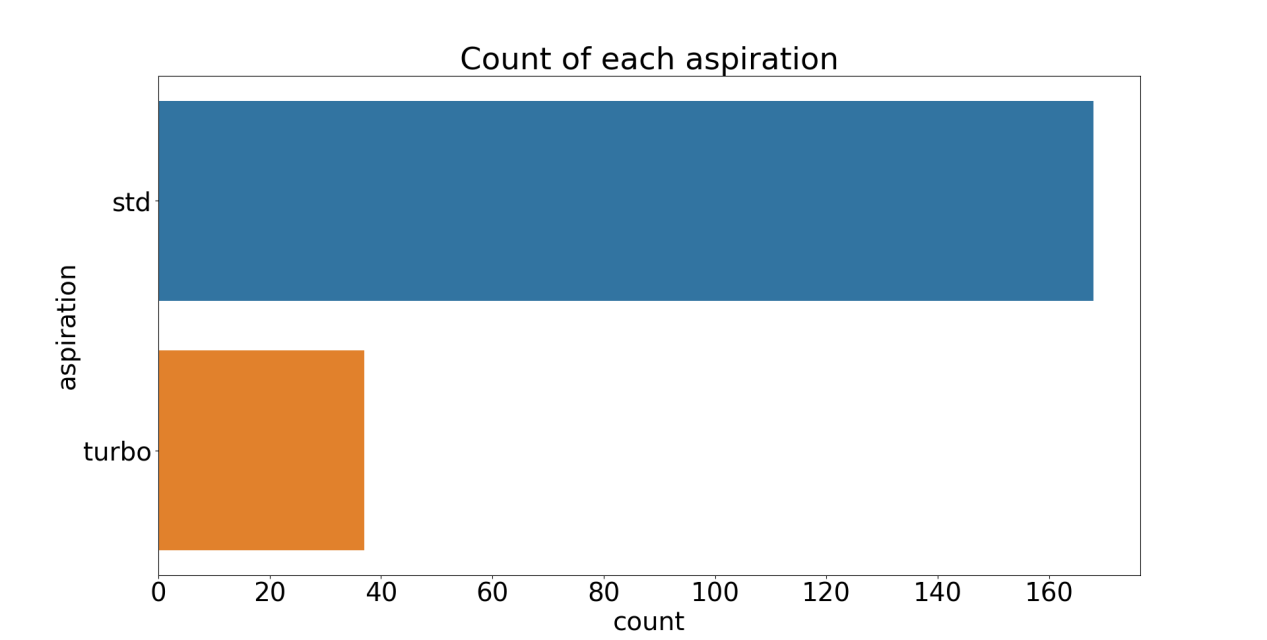
Максимальне значення cтупеню стиснення дорівнює близько 23, а ходу двигуна близько 4.



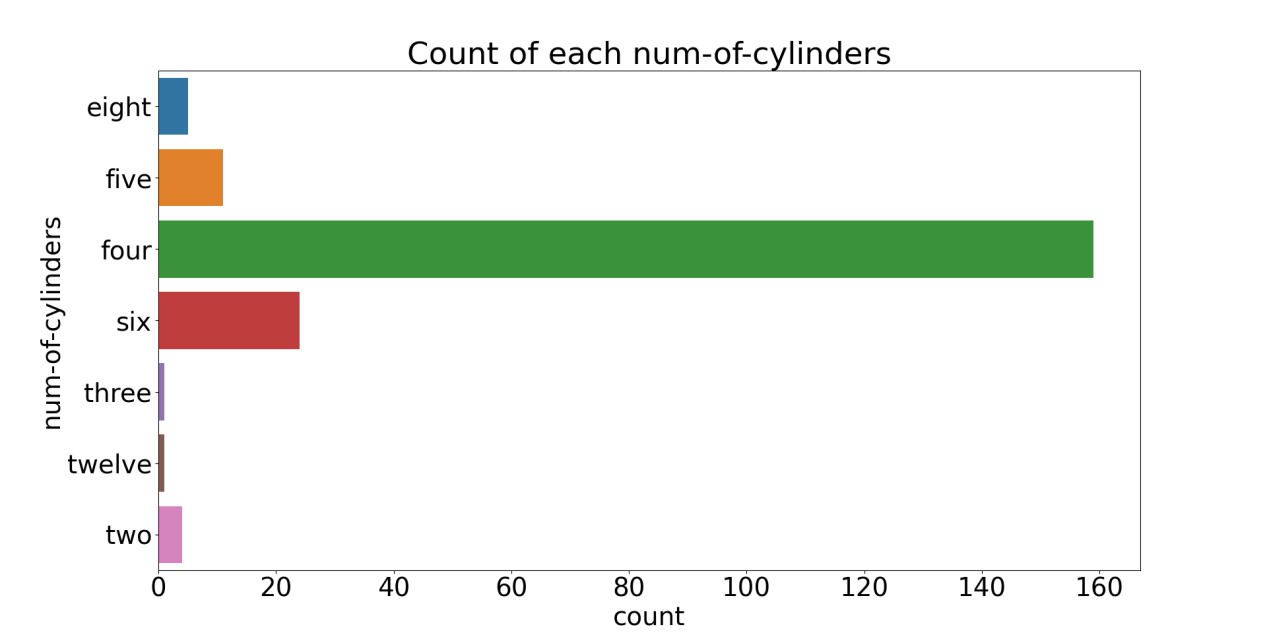
Мінімальне значення ступеню стиснення дорівнює близько 7, а ходу двигуна близько 2.

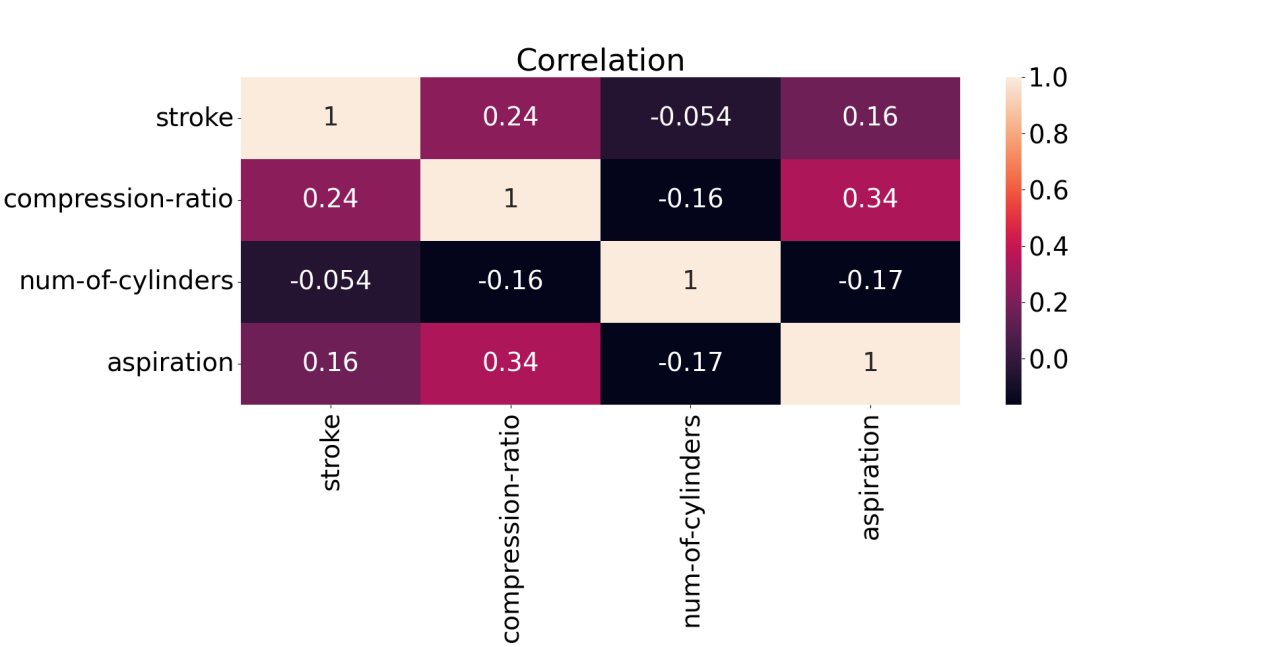


Перейдемо до аналізу строкових ознак. Можемо побачити, що найбільш популярною системою постачання повітря є std, турбо значно уступає.



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



Після аналізу графіка кореляцій виявлено, що існують як позитивні, так і негативні коефіцієнти, проте більшість значень не перевищують +-0.2. Загалом, кореляція між ознаками є досить слабкою. Зазначимо, що кореляція між системою потрапляння повітря та ступенем стиснення становить 0.34, а кореляція між ступенем стиснення та ходом двигуна - 0.24. Однак ці значення не є великими, вказуючи на відсутність значущих взаємозв'язків.

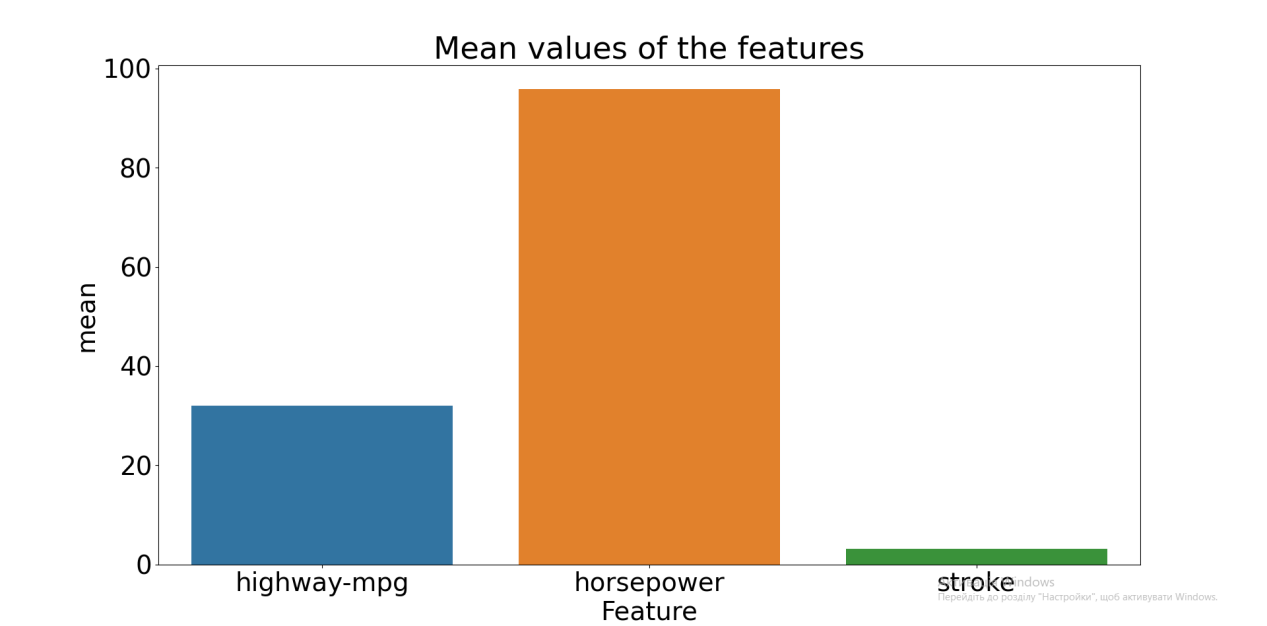
## Рекурсивний алгоритм. Висновок

За результатами аналізу можна зробити висновок, що при виборі автомобілів для продажу важливо звертати увагу на такі характеристики: кількість циліндрів, систему потрапляння повітря, ступінь стиснення та хід двигуна. Середнє значення ступеня стиснення складає приблизно 10, а середнє значення ходу двигуна близько 4. Найбільш популярна кількість циліндрів - 4, а система потрапляння повітря - std. Важливо враховувати взаємозв'язок між системою потрапляння повітря та ступенем стиснення (кореляція 0.34) та між ступенем стиснення та ходом двигуна (кореляція 0.24).

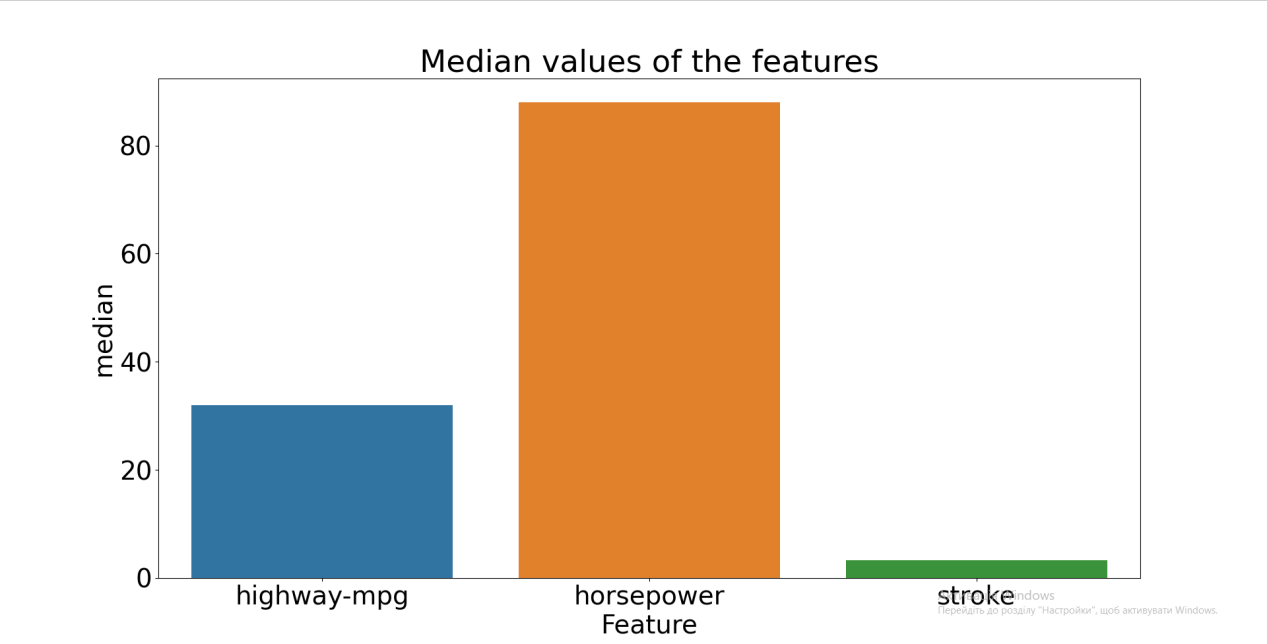
## Алгоритм відбору ознак

Алгоритм виділив наступні ключові характеристики: витрати пального на шосе, кінські сили, хід двигуна та тип двигуна, розташовані від найбільш важливих до менш важливих. Давайте проведемо загальний аналіз цих ознак, розпочавши з числових значень.

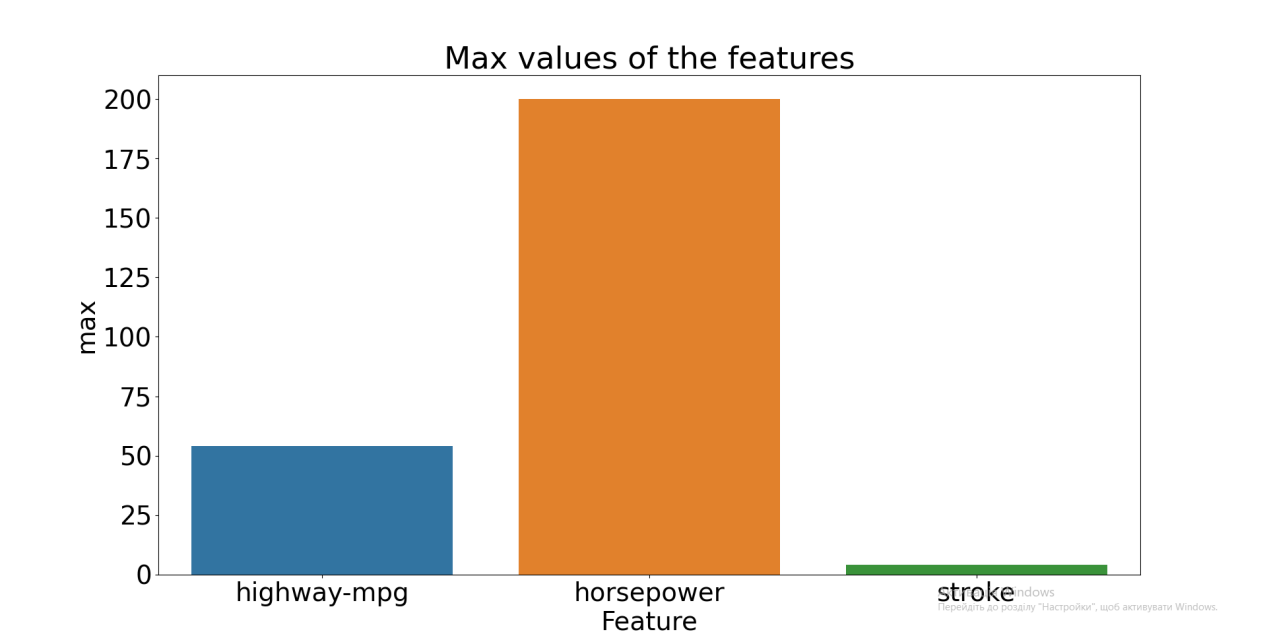
Середнє значення витрат пального на шосе становить близько 30, кінські сили приблизно 95, а середнє значення ходу двигуна дорівнює близько 3.



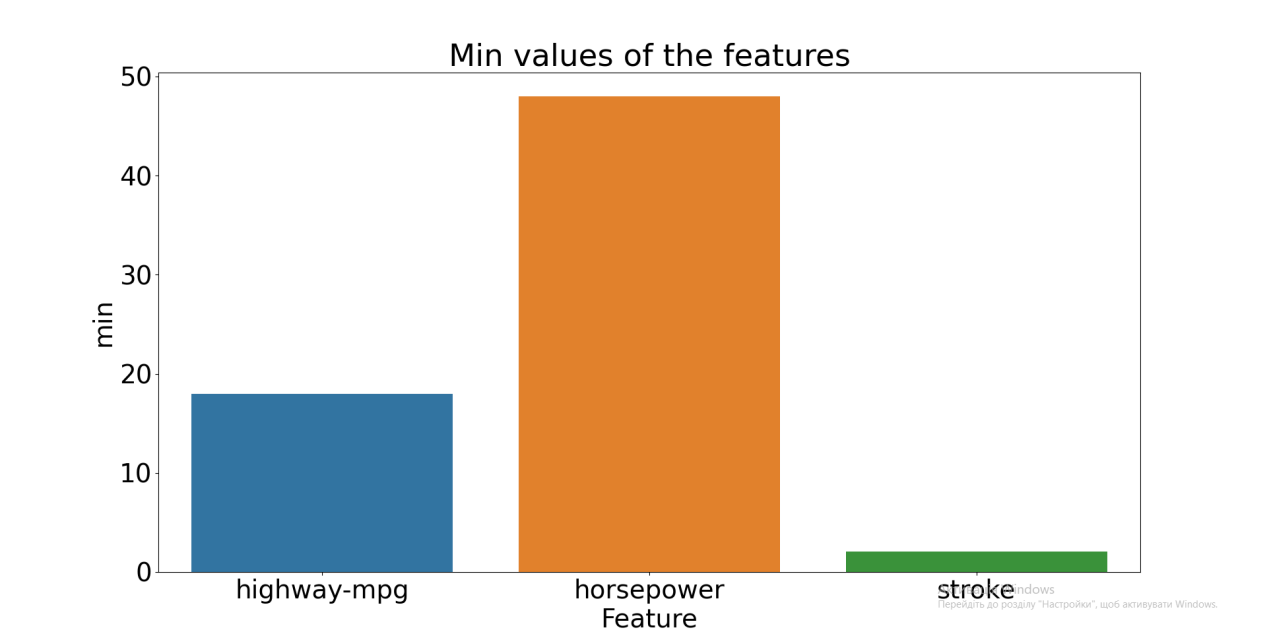
Медіана витрат палива на шосе становить близько 30, кінських сил близько 85, а ходу двигуна близько 3.



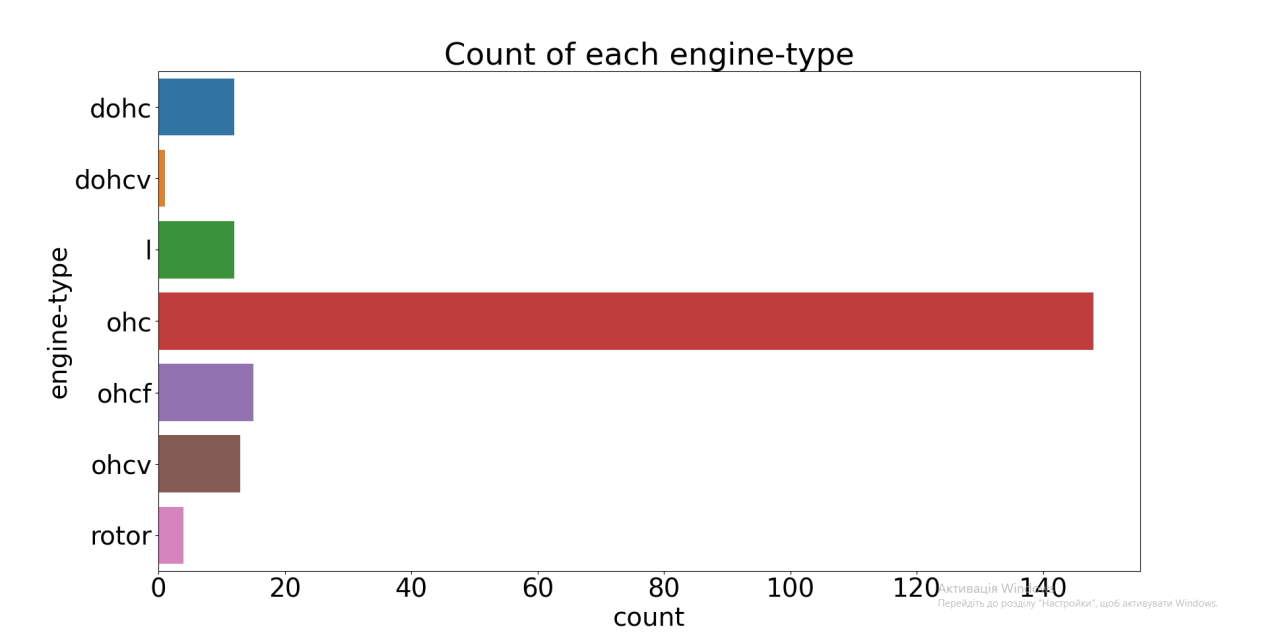
Максимальне значення витрат палива на шосе становить близько 50, кінських сил близько 200, а ходу двигуна близько 3.



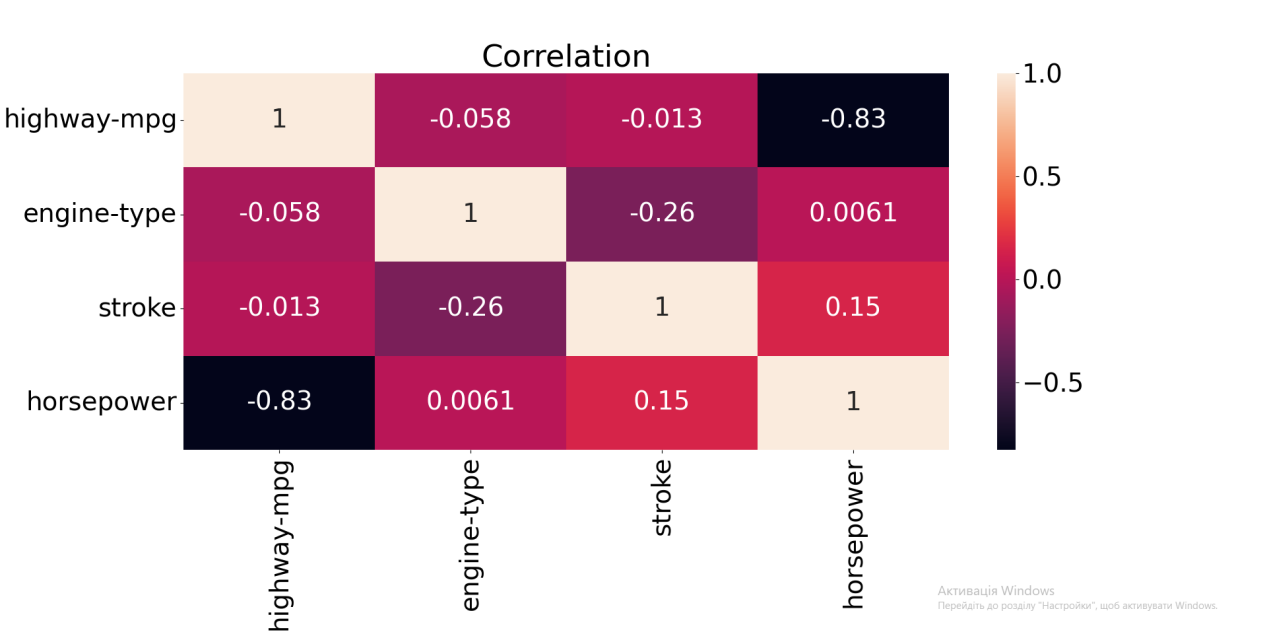
Мінімальне значення витрат палива на шосе становить близько 18, кінських сил близько 48, а ходу двигуна близько 3.



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



Після аналізу графіка кореляцій виявлено, що існують як позитивні, так і негативні коефіцієнти, але більшість значень не перевищують +-0.2. Однак важливо відзначити, що кореляція між кінськими силами та витратами бензину на шосе становить -0.83, що є дуже високим значенням. Також спостерігається помітна негативна кореляція між типом двигуна та ходом двигуна, яка дорівнює -0.26.

Ці високі значення кореляції можуть свідчити про сильну взаємозалежність між кінськими силами та витратами пального на шосе, а також між типом двигуна та ходом двигуна.

## Алгоритм відбору ознак. Висновок

За результатами аналізу можна зробити висновок, що при виборі автомобілів для продажу важливо звертати увагу на такі характеристики: витрати пального на шосе, кінські сили, хід двигуна та тип двигуна. Середнє значення витрат пального на шосе складає приблизно 30, кінські сили приблизно 95, а середнє значення ходу двигуна дорівнює близько 3. Найбільш популярний тип двигуна - ohc. Важливо враховувати значну негативну кореляцію між кінськими силами та витратами пального на шосе, яка становить -0.83.

**Загальний висновок**

Різні алгоритми визначають різноманітні ознаки, які впливають на остаточну ціну автомобіля. Проте серед них є спільні ознаки, вибрані двома алгоритмами: витрата пального на шосе, тип двигуна та хід двигуна. Важливо звертати особливу увагу на ці аспекти при виборі автомобілів для продажу. Також відзначено значну взаємозв'язок між деякими ознаками: кількість кінських сил та витрати пального на шосе з коефіцієнтом -0.83, система потрапляння повітря та ступінь стиснення з коефіцієнтом 0.34, і витрата пального на шосе та система заправки з коефіцієнтом -0.63.