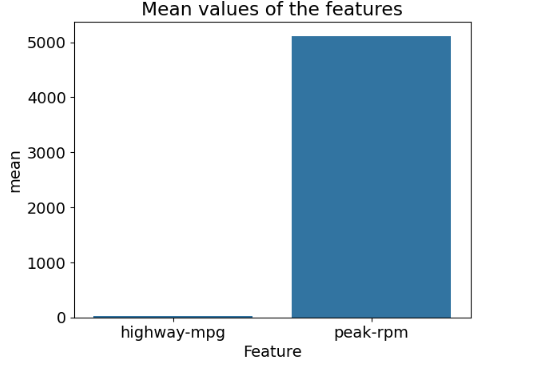
### Обрані алгоритми: однофакторний вибір ознак, рекурсивний, вимір важливості ознак. Ціль - ціна.

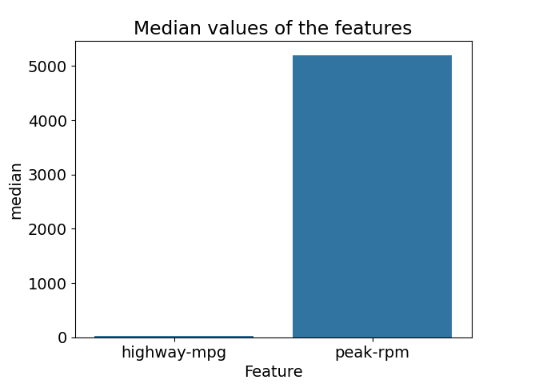
## Однофакторний вибір ознак

sklearn визначив важливими наступні ознаки: к-сть миль на галлон палива, тип двигуна, максимальні оберти, система заправки Почнемо з числових ознак:

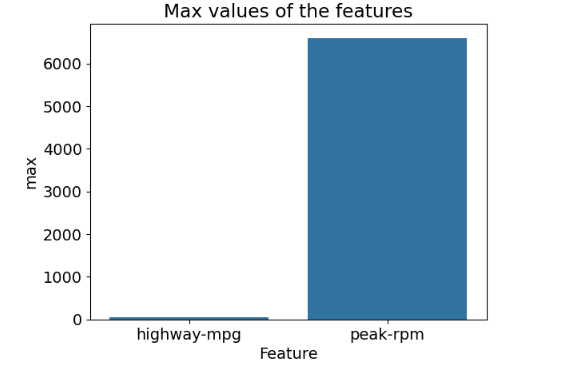
Середнє значення обертів - 5000. MPG - 32.



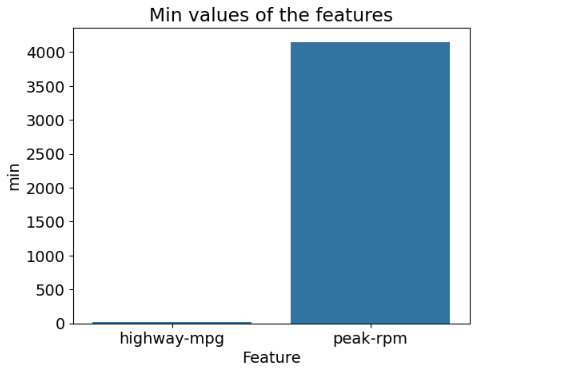
Медіана обертів - 5000. MPG - 32.



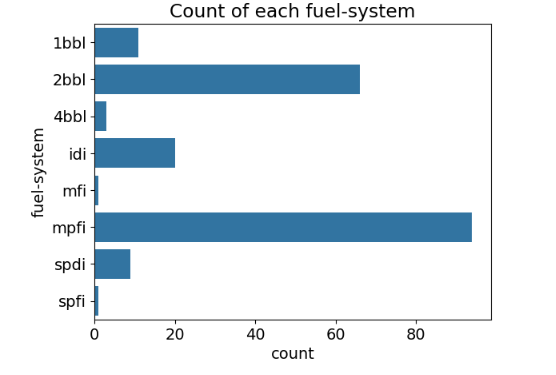
Максимальна відсічка - 6300, максимальна ефективність - 54 милі на галлон.



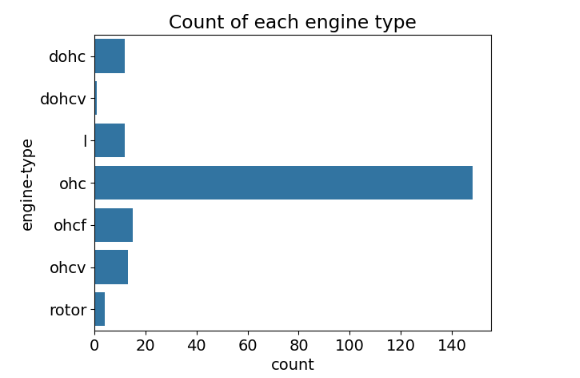
Найнижча відсічка - 4150, а найгірша ефективність - 10 миль на галлон.



Строкові ознаки:

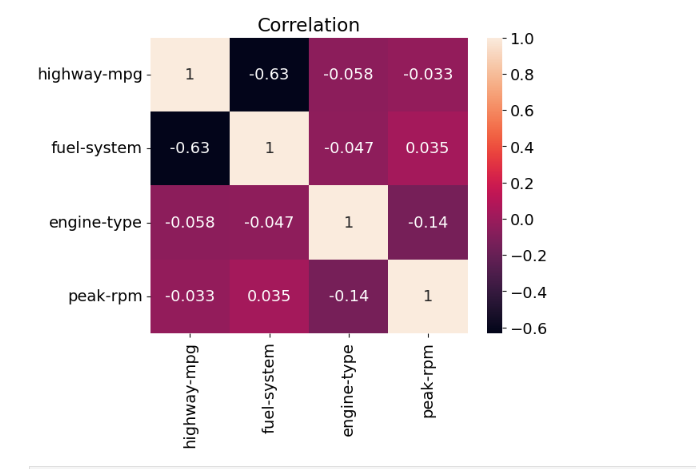


Зі строкових ознак видно, що інжекторні двигуни найбільш популярні, після них йдуть карбюраторні (2 карбюратори), після них теж карбюраторні (з одним карбюратором). Що дивно, двигуни з непрямим вприском палива більш популярні за карбюраторні.



Найбільш популярний тип двигуна - OHC (SOHC?). Двигуни з двома розподільними валами і системою зміни фаз газорозподілу майже не присутні, що також є надзвичайно дивним.

Проаналізуємо графік кореляцій. Єдині ознаки, які мають позитивну кореляцію - це системи вприску та відсічка (резонно) - кореляція 0.035. Усі інші ознаки мають негативні кореляції, але в межах -0.1 < х < 0, але система вприску та економність на трасі мають кореляцію -0.63.

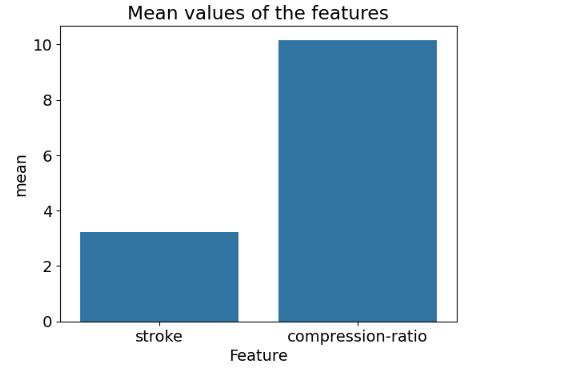


Отже, при конструюванні автомобіля необхідно звернути увагу на: витрата палива на шосе, тип двигуна (а радше кількість розподільних валів), максимальні оберти, система вприску. Середнє значення економності - близько 32 миль на галлон, а максимальних обертів - близько 5000. Найбільш популярні системи вприску - два карбюратори та інжектор. Найбільш популярний тип двигунів - SOHC.

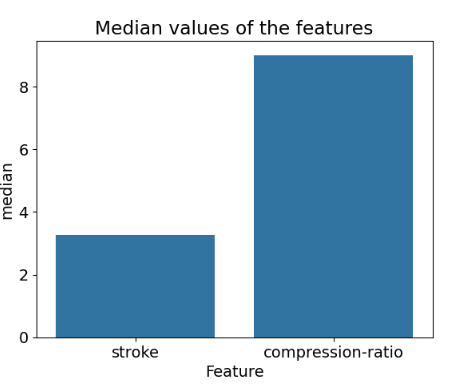
## Рекурсивний алгоритм

Ознаки: хід поршня, компресія, к-ть циліндрів та тип двигуна за отриманням повітря. Числові ознаки:

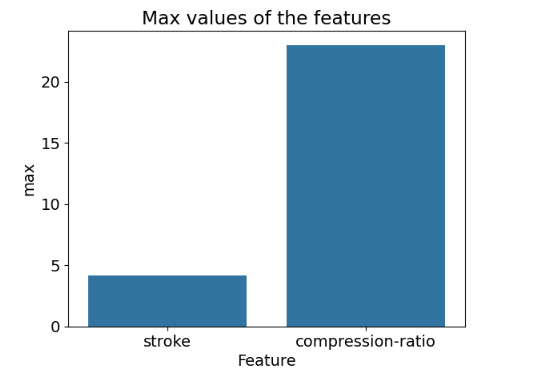
Середнє значення компресії - ~10, а ходу поршня - 3.



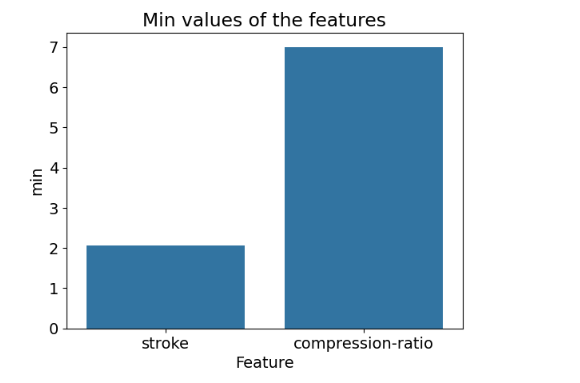
Медіана ходу поршня дорівнює ~4. Медіана компресії - близько 9.



Максимальне значення компресії дорівнює 23, а ходу поршня близько 4.

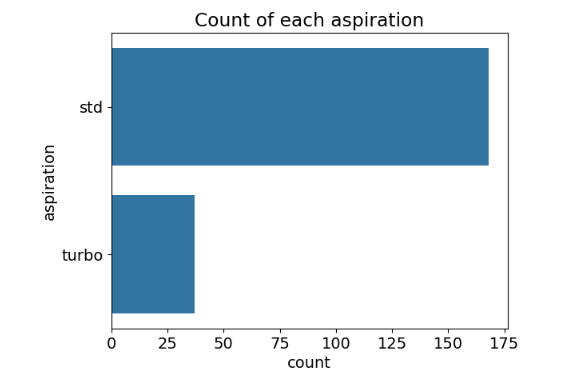


Мінімальне значення компресії дорівнює 7, а ходу поршня близько 2.

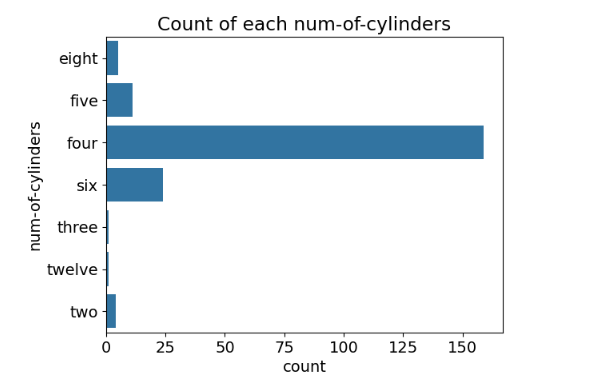


Строкові ознаки:

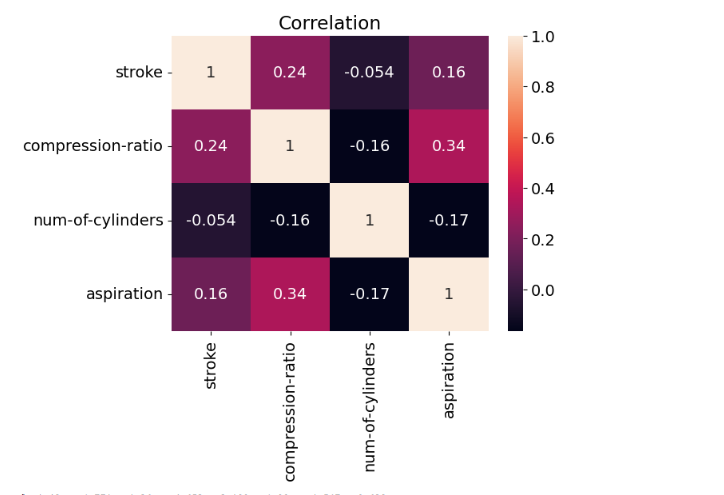
Std? Будемо вважати, що std означає стандарт (атмосферний двигун). В такому випадку атмосферних двигунів значно більше за турбові.



Бачимо, що I4 та I6/V6 двигуни найбільш популярні, що є цілком резонно. Низька популярність восьмициліндрових двиунів - це надзвичайно сумно.



З кореляцій: кореляція системи постачання повітря та компресії дорівнює 0.34 (очікувано), а кореляція компресії та ходу поршня дорівнює 0.24 (теж цілком очікувано). Загально кореляція не має значних значень.



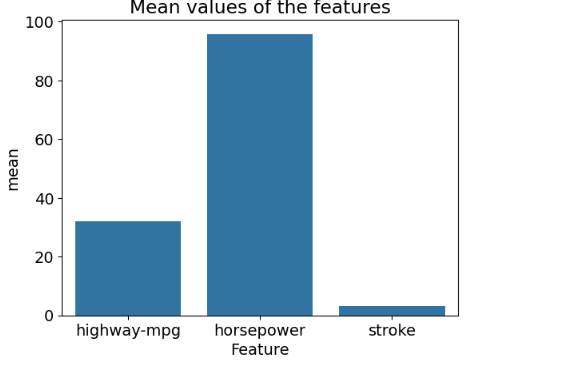
## **Висновок**

Отже, при конструюванні автомобіля необхідно звернути увагу на: кількість циліндрів, систему постачання повітря, компресію та хід поршня. Середнє значення компресії дорівнює близько 10, а ходу поршня - близько 4. Найбільш популярна кількість циліндрів - 4, а система постачання повітря - атмосферна.

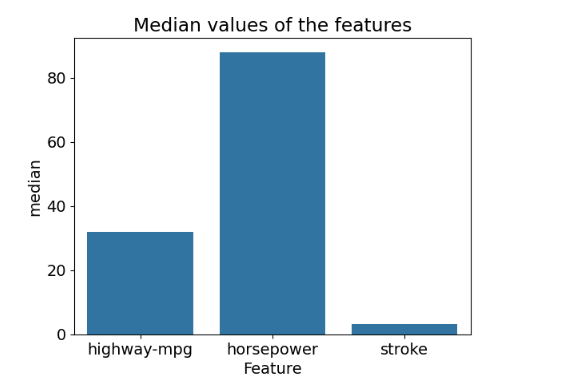
## Алгоритм відбору ознак

Ознаки: економність на шосе, потужність, хід поршня та тип двигуна. Числові ознаки:

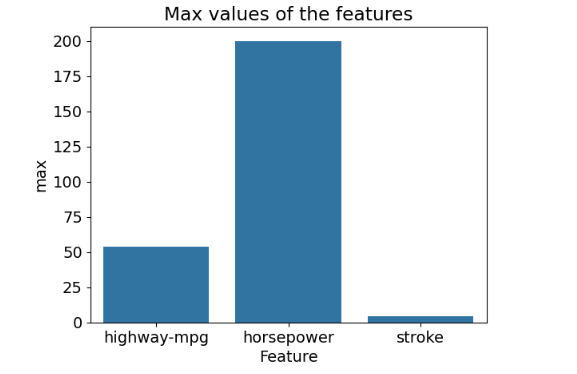
Середня к-сть миль на галлон на шосе становить близько 30, кінських сил близько 95 (?), а ходу поршня близько 3.



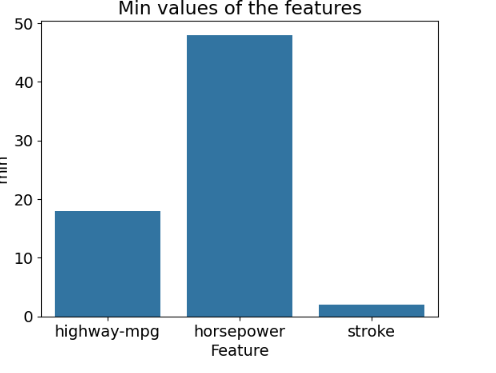
Медіана к-сті миль на галлон на шосе становить близько 30, кінських сил близько 85, а ходу поршня близько 3.



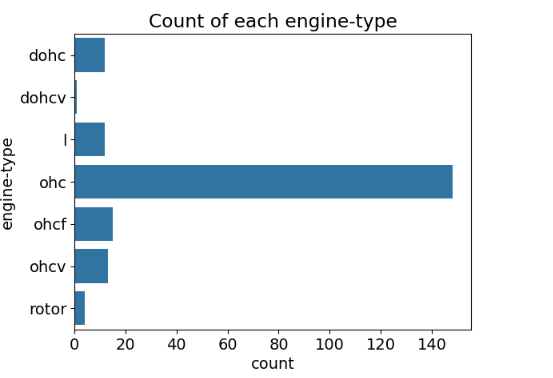
Максимальна економність на шосе становить близько 50 миль на галлон, кінських сил близько 200, а ходу поршня близько 3.



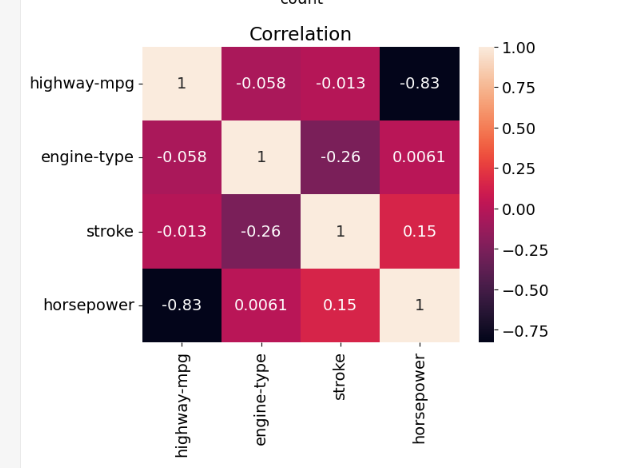
Мінімальна економність на шосе становить близько 18 миль на галлон, кінських сил близько 48 (???), а ходу поршня близько 3.



Строкові ознаки: знову двигун з одним розподільним валом найбільш популярний.



Кореляція потужності та витрати палива на шосе: -0.83. Кореляція між типом двигуна та ходом поршня: -0.26.



## **Висновок**

Отже, при конструюванні автомобіля необхідно звернути увагу на: витрати палива на шосе, потужність, хід поршня та тип двигуна. Середня к-сть миль на галлон на шосе становить близько 30, кінських сил - ~95, а ходу поршня - 3. Найбільш популярний тип двигуна - SOHC.

**Загальний висновок**

Різні алгоритми виділяють різні ознаки, що мають впливати на ціну автомобіля, але серед них є більш популярні ознаки: кількість миль на галлон палива на шосе, тип двигуна та хід поршня. На них варто звернути особливу увагу при конструюванні автомобілів. Однак варто теж звернути увагу на те, що певні ознаки можуть впливати на більш комплексні ознаки (як-от тип двигуна впливає на потужність і витрату палива).