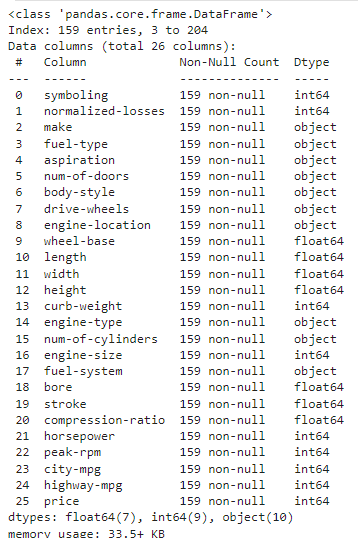
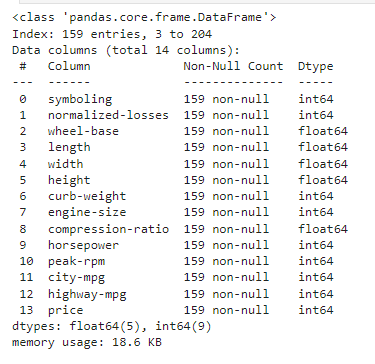
**Аналітичний звіт ознак які впливають на ціну імпортних автомобілів і вантажівок моделі 1985 року**

**Аналіз даних**



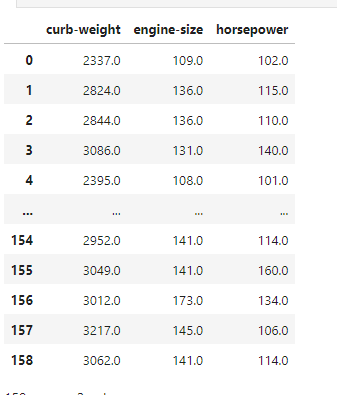
У даних виявлено 16 числових ознак та 10 строкових

**Метод дисперсії**



За результатами дисперсії було знайдено дві найменш важливі ознаки bore та stroke.

**Метод f-classif**



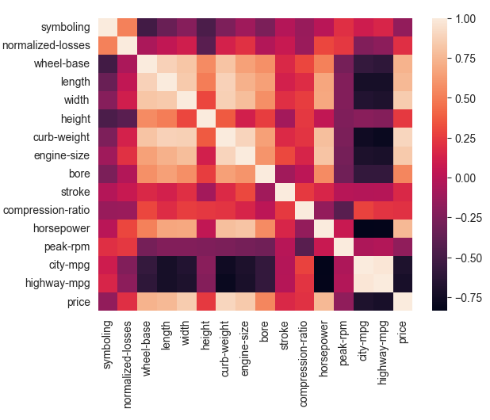
За результатами методу **f-classif** серед числових даних було відібрано три найбільш важливі ознаки що найбільше впливають на ціну є споряджена маса(curb-weight) ,об’єм двигуна(engine-size) та кількість кінських сил (horsepower).

Також було проведено кореляційний аналіз.



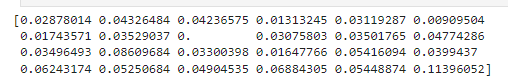
В результаті кореляція є позитивною та високою.

**Загальна кореляція**

****

Найбільше значення спостерігається між витратою палива в місті(city-mpg) та за містом(highway-mpg).Також видно що ознаки обрані методом **f-classif** мають значно вищий показник кореляції за інших.Також дані співпадають відповідно до кореляці рядка price де значення кореляції також найвищі в curb-weight ,engine-size та horsepower

**Метод ExtraTreesClassifier**

****

Цей метод є найбільш продвинутим та дає найбільш репрезентативні результати оскільки враховує також і строкові дані. Таким чином найважливішими ознаками , що найбільше впливають на ціну є:

1.Витрата палива по трасі (highway-mpg)

2.Кількість кінських сил (horsepower)

3.Тип двигуна (engine type)

4. Споряджена маса (curb-weight)

**Висновки**

За результатами проведеного дослідження найбільш важливими ознаками є кількість кінських сил та споряджена маса, які були отримані в результаті використання обох методів та аналізу кореляційного графіка. Проте виникли деякі розбіжності, які пов’язані з недосконалістю методу f-classif а саме використання лише числових ознак. Тому в ньому не було враховано ознаку тип двигуна (engine type). Але з моїх спостережень відповідно до даних кореляційного графіка випливає те що витрата палива по трасі highway-mpg дуже сильно впливає на price, але не за позитивною кореляцією а за негативною. Тобто чим менше витрата палива тим дорожче. Метод f-classif це не врахував, через що й виникла розбіжність. Найдосконалішим методом виявився ExtraTreesClassifier.