Міністерство освіти і науки України

Львівський національний університет імені Івана Франка

Факультет електроніки та комп’ютерних технологій

Кафедра системного проектування

Звіт

Про виконання лабораторної роботи №2

З курсу «Основи аналізу даних»

### **Описова статистика**

Виконав:

Студент групи ФЕІ-42

Прізвище Ім’я

Перевірив:

асистент Азаров І.В

Львів 2024

**Мета:**

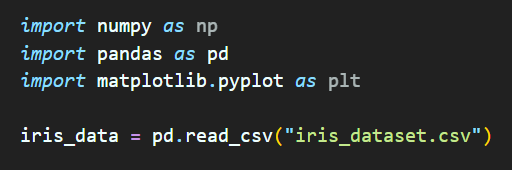
Засвоїти основні методи описової статистики для аналізу наборів даних. Навчитися обчислювати ключові статистичні показники та візуалізувати розподіл даних.

**Хід роботи:**

**Завдання**

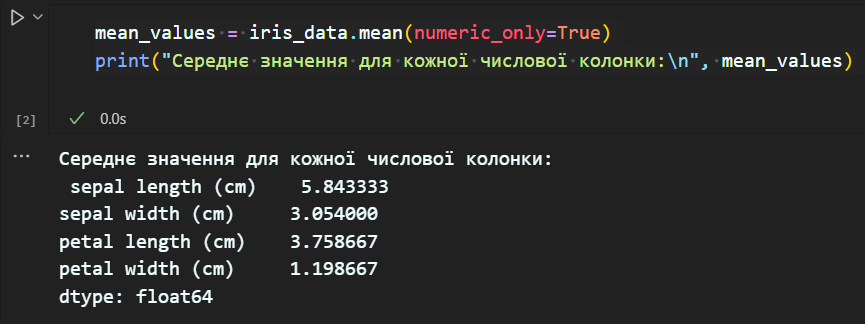
1. Використайте набір "Iris"

Підключив потрібні бібліотеки та завантажив файл:

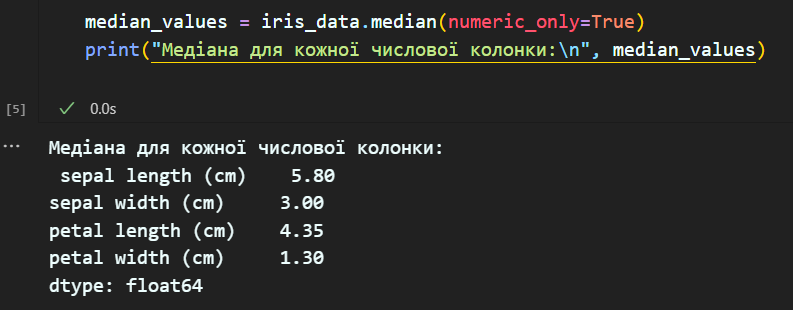


1. Обчислення середнього значення, медіани, моди:

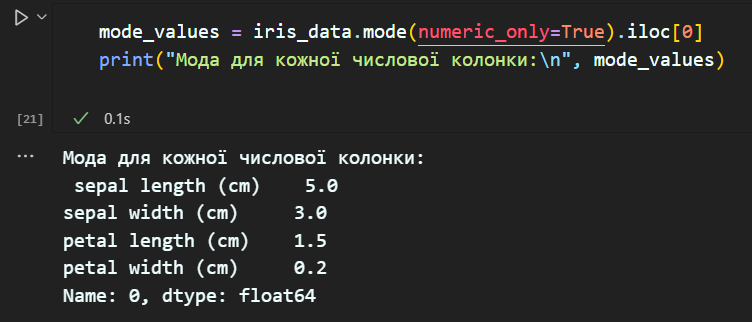
- Обчисліть середнє значення для кожної числової колонки.



- Обчисліть медіану для кожної числової колонки.

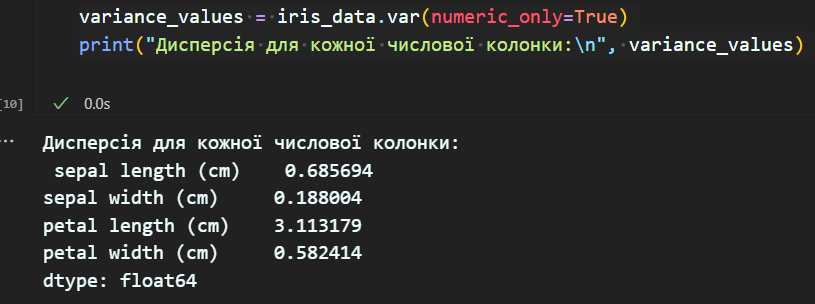


- Обчисліть моду для кожної числової колонки.

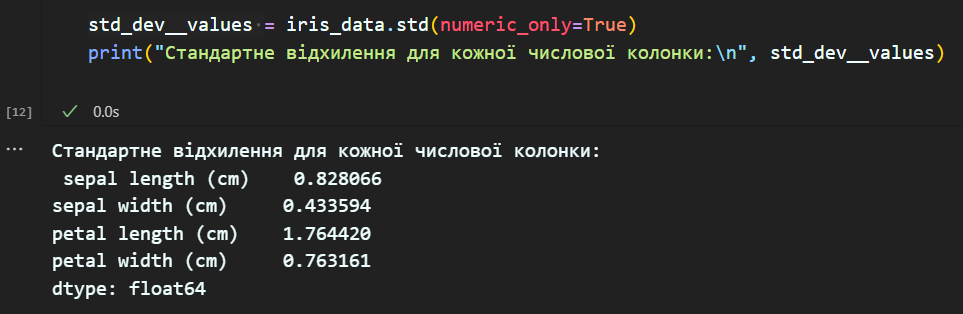


3. Обчислення дисперсії та стандартного відхилення:

- Обчисліть дисперсію для кожної числової колонки.

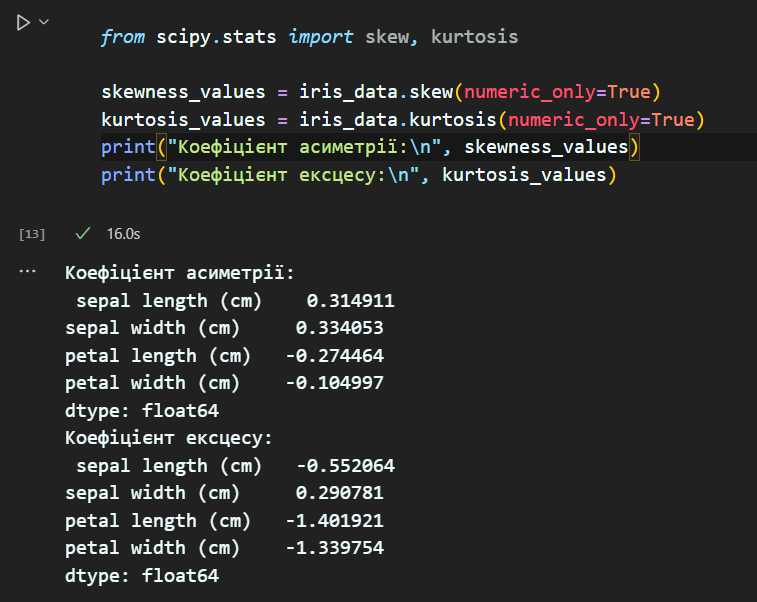


- Обчисліть стандартне відхилення для кожної числової колонки.



- Обчисліть коефіцієнт асиметрії та ексцесу для числових змінних та

проаналізуйте їх значення.

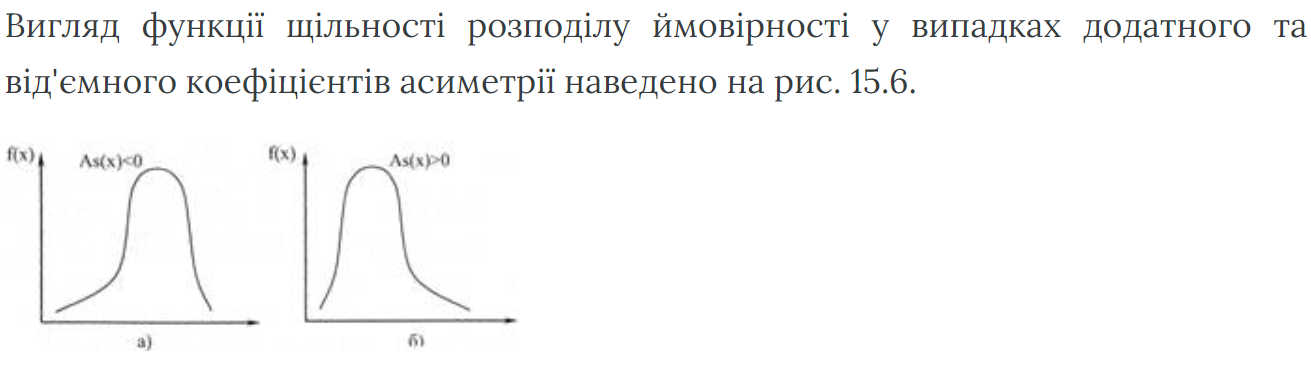


**Пояснення:**

### **Коефіцієнт асиметрії (skew)**

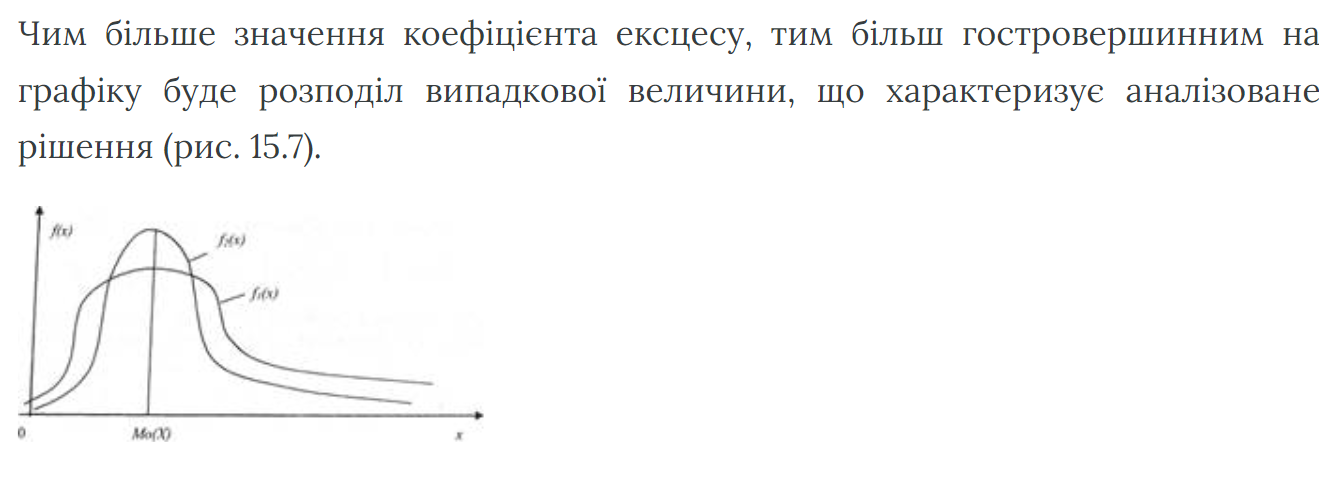
**Чим значення асиметрії ближче до 0 тим краще**·

* **0**: Розподіл симетричний (наприклад, нормальний розподіл).
* · **Додатне значення**: Розподіл має **правосторонню (позитивну) асиметрію**. Це означає, що «хвіст» розподілу тягнеться більше вправо, тобто більші значення відхиляються більше від середнього.
* · **Від'ємне значення**: Розподіл має **лівосторонню (негативну) асиметрію**. «Хвіст» розподілу довший зліва.



### **Коефіцієнт ексцесу (kurtosis)**

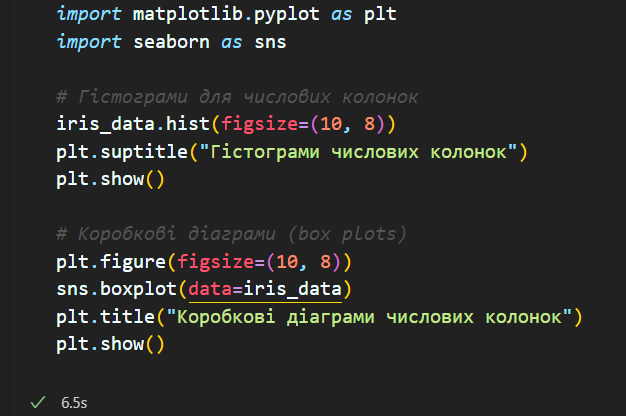
* Показує, наскільки «гострим» або «плоским» є розподіл у порівнянні з нормальним розподілом.
* **Значення:**
* **0**: Розподіл відповідає нормальному, має середню "гостроту".
* **Додатне значення**: Розподіл має **гостру вершину** (розподіл з високим піком), що означає наявність великої кількості значень близьких до середнього.
* **Від'ємне значення**: Розподіл є **плоским**, тобто має ширший пік з меншою кількістю значень біля середнього.

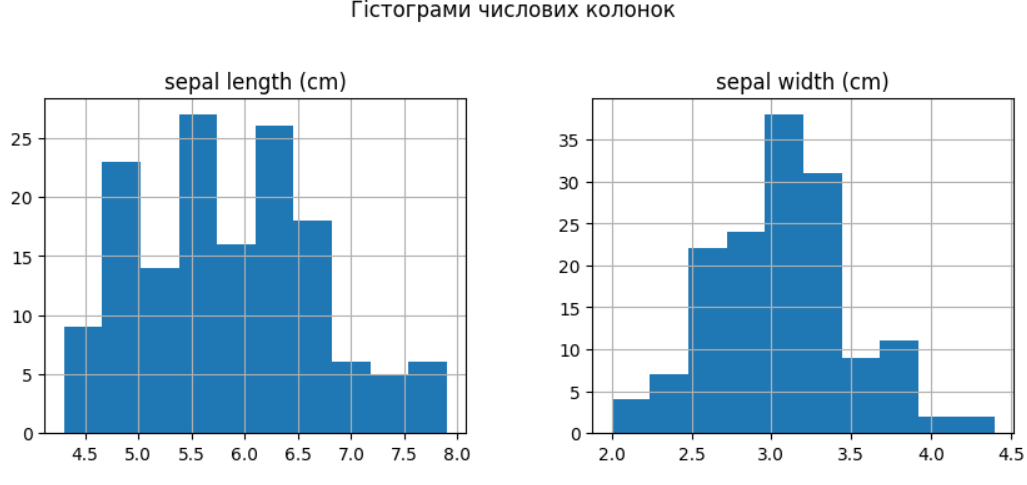


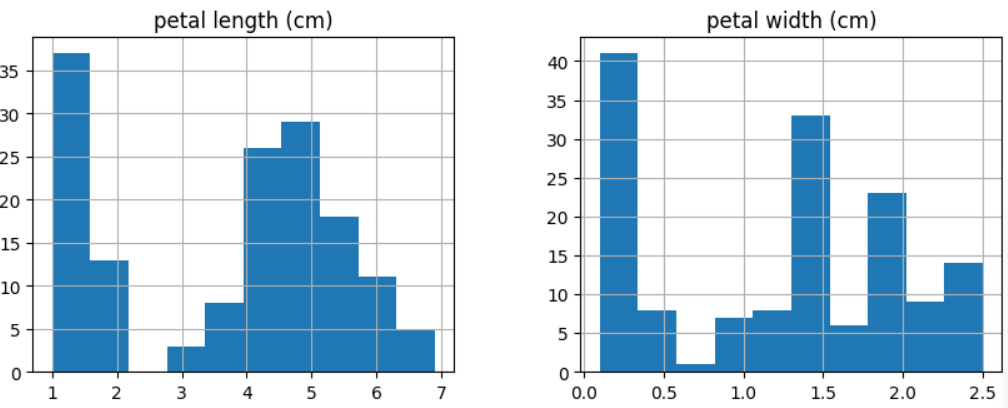
4. Візуалізація розподілу даних за допомогою гістограм та коробкових

діаграм (box plots):

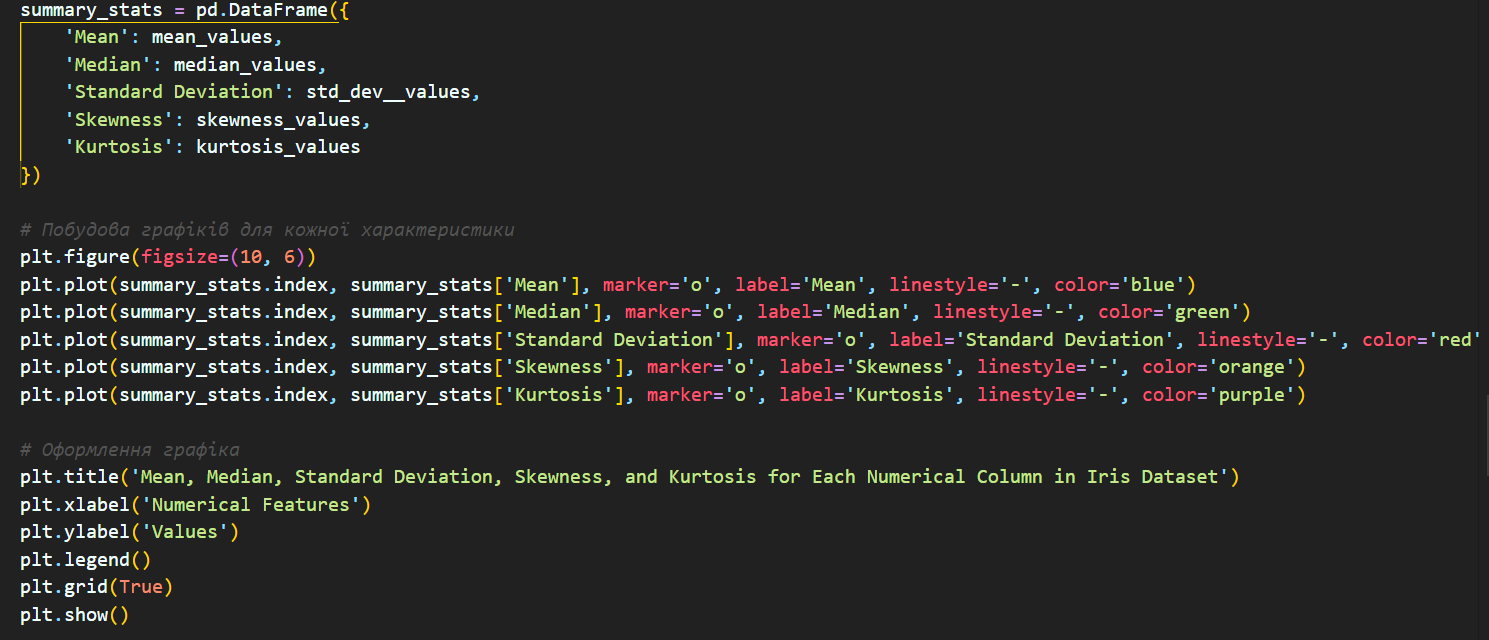
- Побудуйте гістограми для числових колонок.

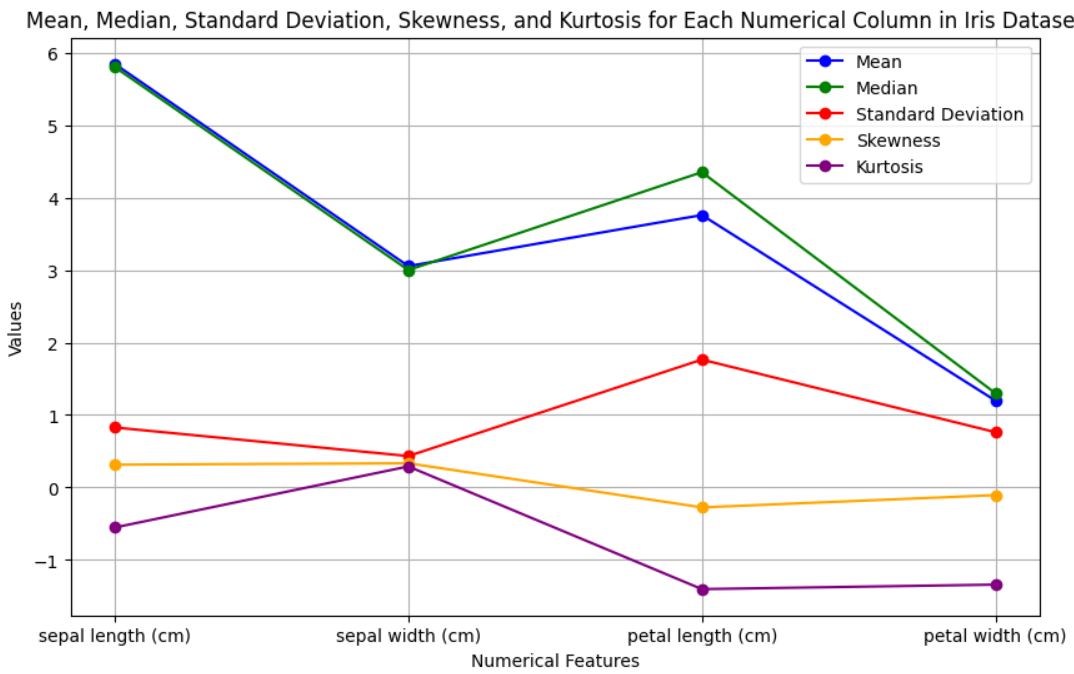






a. Побудуйте лінією три обчислені величини.

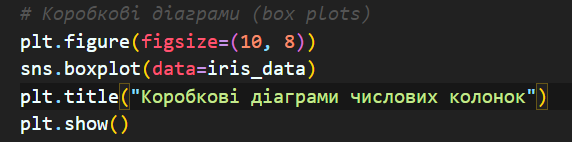


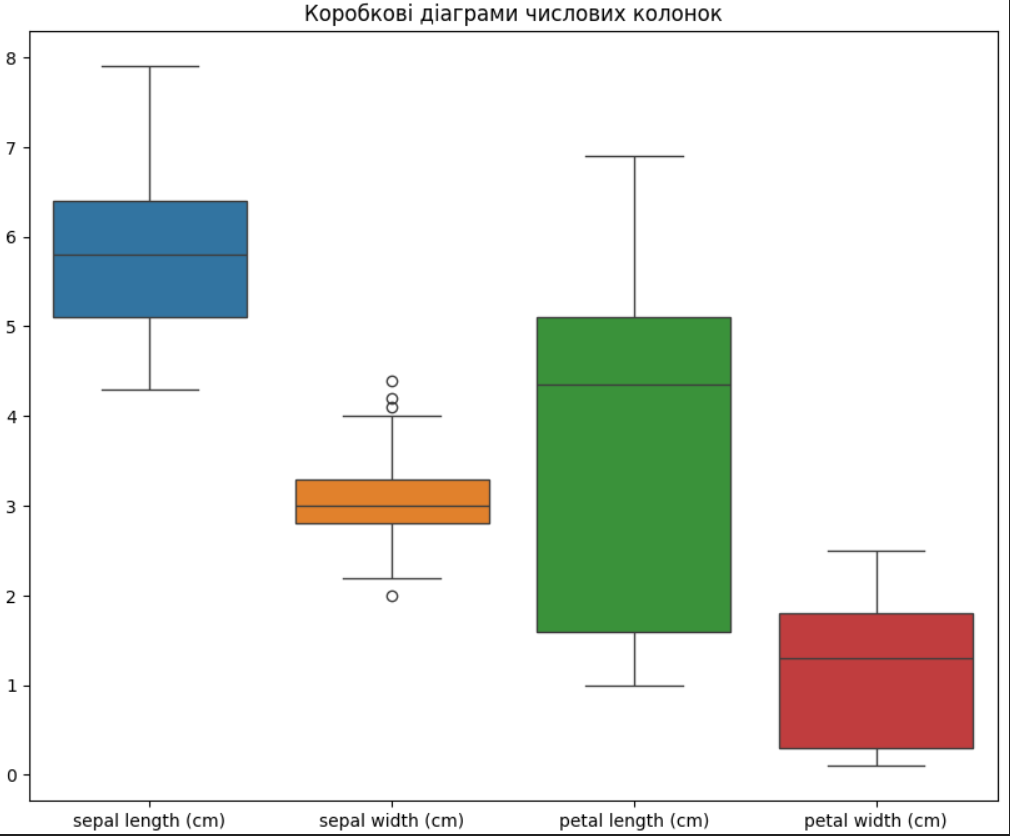


b. Вкажіть якому розподілу відповідають дані

- ТІльки для колонок зі значеннями sepal width та sepal lenght можна сказати, що дані віддалено відповідають нормальному розподілу(Гаусівському). Решта колонок (petal width, petal length), на жаль не мають явно вираженого нормального розподілу.

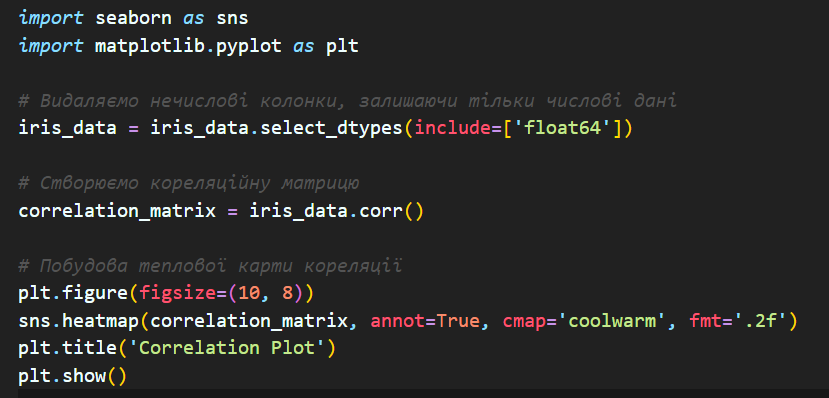
- Побудуйте коробкові діаграми (box plots) для числових колонок.

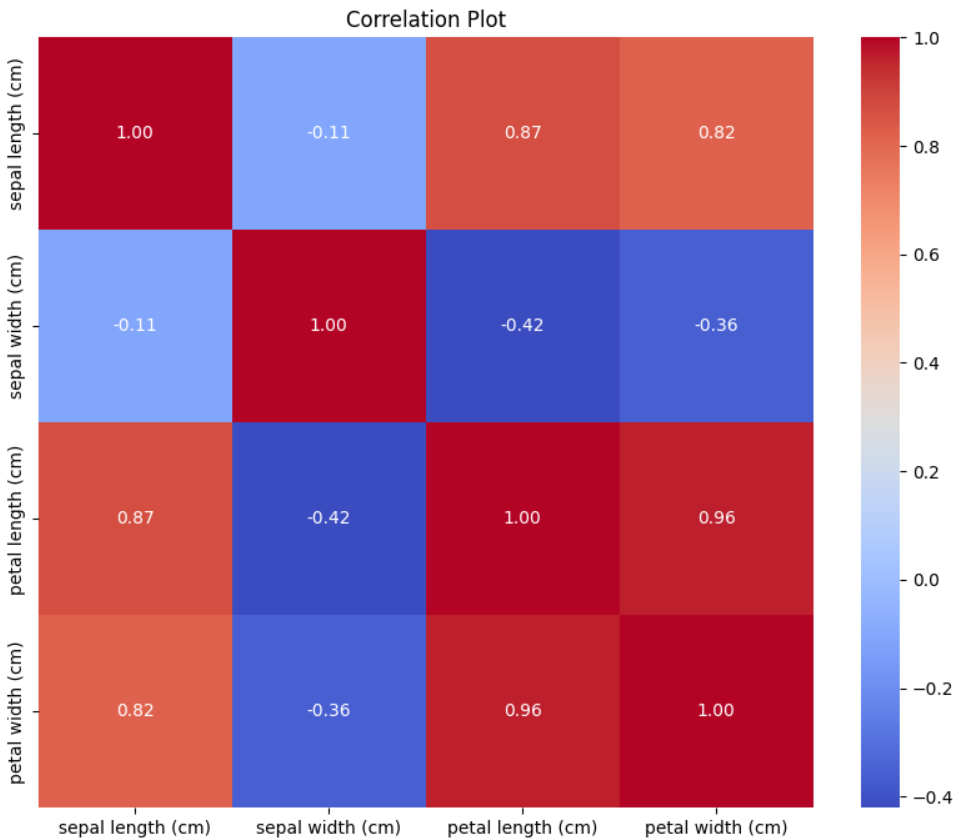




- Побудуйте графіки що дозволяють побачити взаємозв'язків між

числовими змінними. Опишіть його.





Для виявлення взяємозв’язків між числовими знінними - я побудував матрицю кореляції, яка показує на перетині двох значень число (в межах [-1,1])

Кореляція між двома величинами означає, що між ними існує певний статистичний зв'язок або залежність. Зокрема:

* **Позитивна кореляція**: Коли одна величина зростає, інша також має тенденцію зростати. Кореляційний коефіцієнт буде близьким до +1.
* **Негативна кореляція**: Коли одна величина зростає, інша зменшується. Кореляційний коефіцієнт буде близьким до -1.

З графіку видно, що майже усі значення добре корелють між собою, як позитивно(petal width I petal length), так і негативно(sepal width i petal length). На противагу, значення sepal width i sepal length погано корелють між собою, тобто залежності одного значення від іншого майже немає.

**Висновок:**

У ході виконання лабораторної роботи №2 я засвоїв основні методи описової статистики для аналізу даних та їх візуалізації. Використавши набір даних "Iris", я здійснив обчислення ключових статистичних показників, таких як середнє значення, медіана, мода, дисперсія, стандартне відхилення, а також коефіцієнти асиметрії та ексцесу для числових колонок. Також, більше дізнався про г**істограми та коробкові діаграми, які** допомогли візуалізувати розподіл числових змінних, а також виявити наявність викидів та особливості розподілу. **Кореляційний аналіз** дозволив оцінити зв'язки між змінними.