**ЛАБОРАТОРНА РОБОТА № 7**

**Мета роботи**: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити методи неконтрольованої класифікації даних у машинному навчанні.

**Завдання 1.** Кластеризація даних за допомогою методу k-середніх

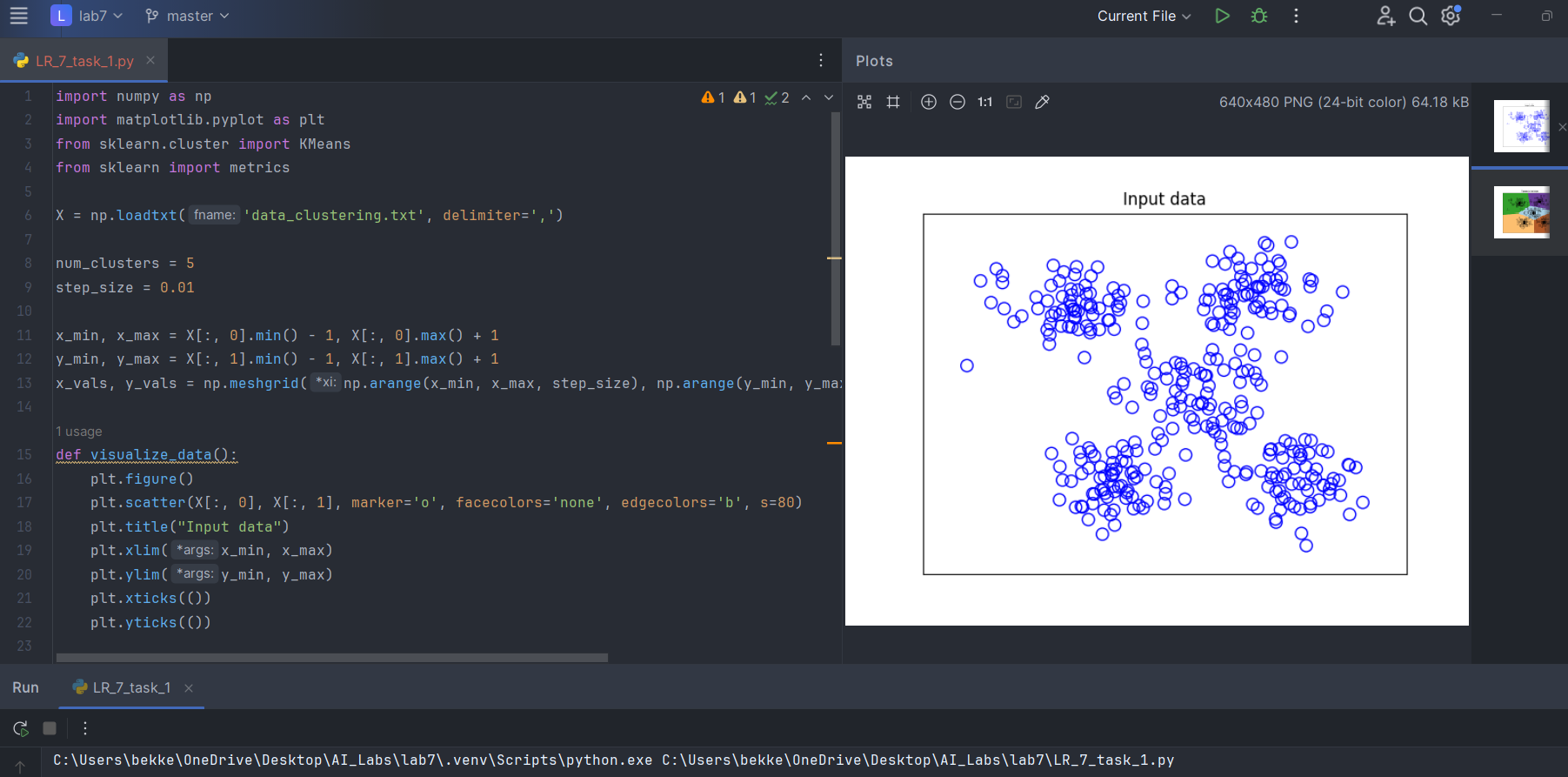


Рис 1. Візуалізація даних

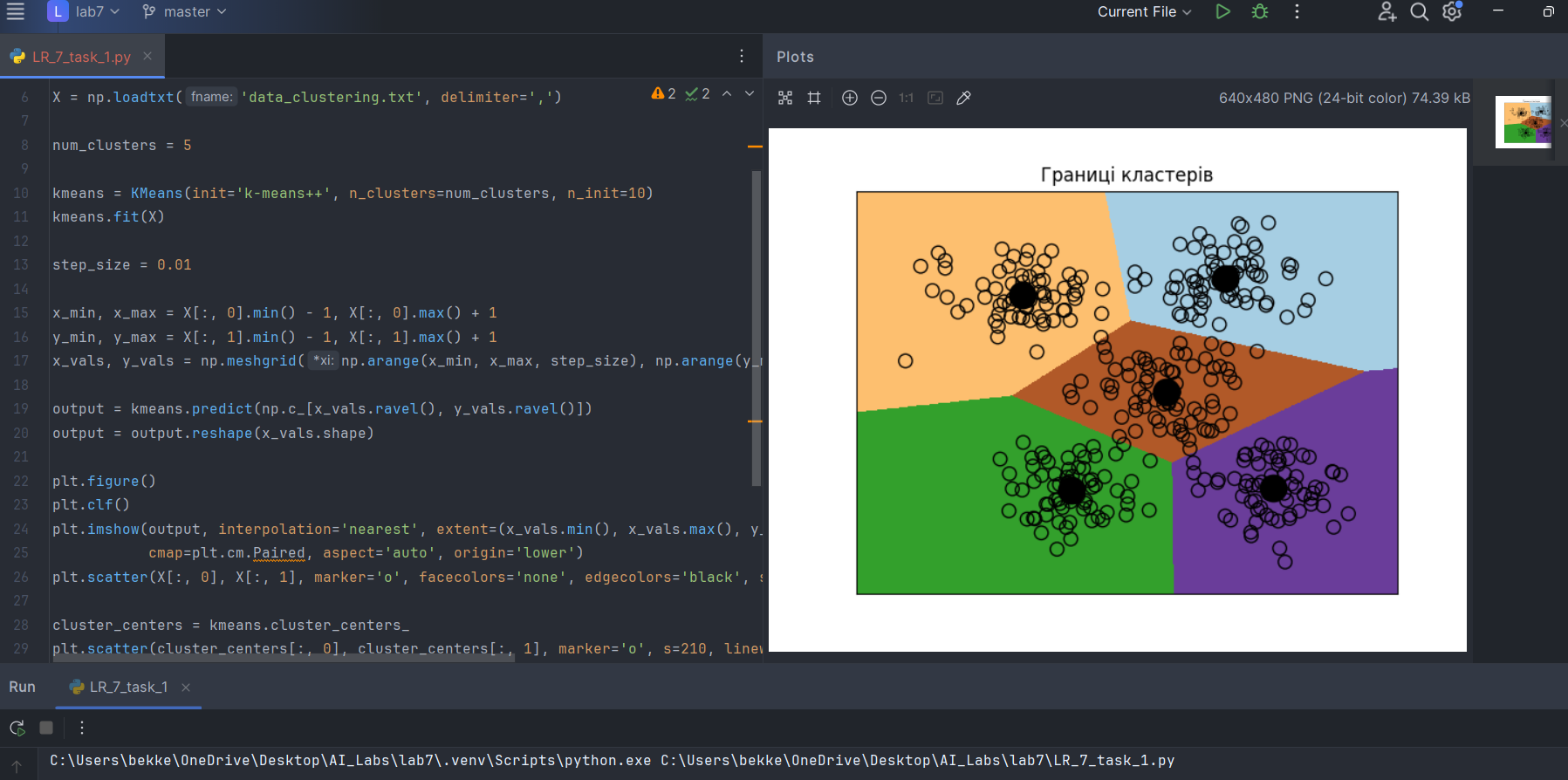


Рис 2. Візуалізація груп даних

**Висновок:** на даному етапі вдалося візуалізувати вхідні дані, а також кластери та їх межі. За допомогою цієї візуалізації можна проаналізувати розподіл даних. Також на графіку можна побачити положення центроїдів (центрів тяжіння кластеру, або головні точки).

**Завдання 2.** Кластеризація K-середніх для набору даних Iris

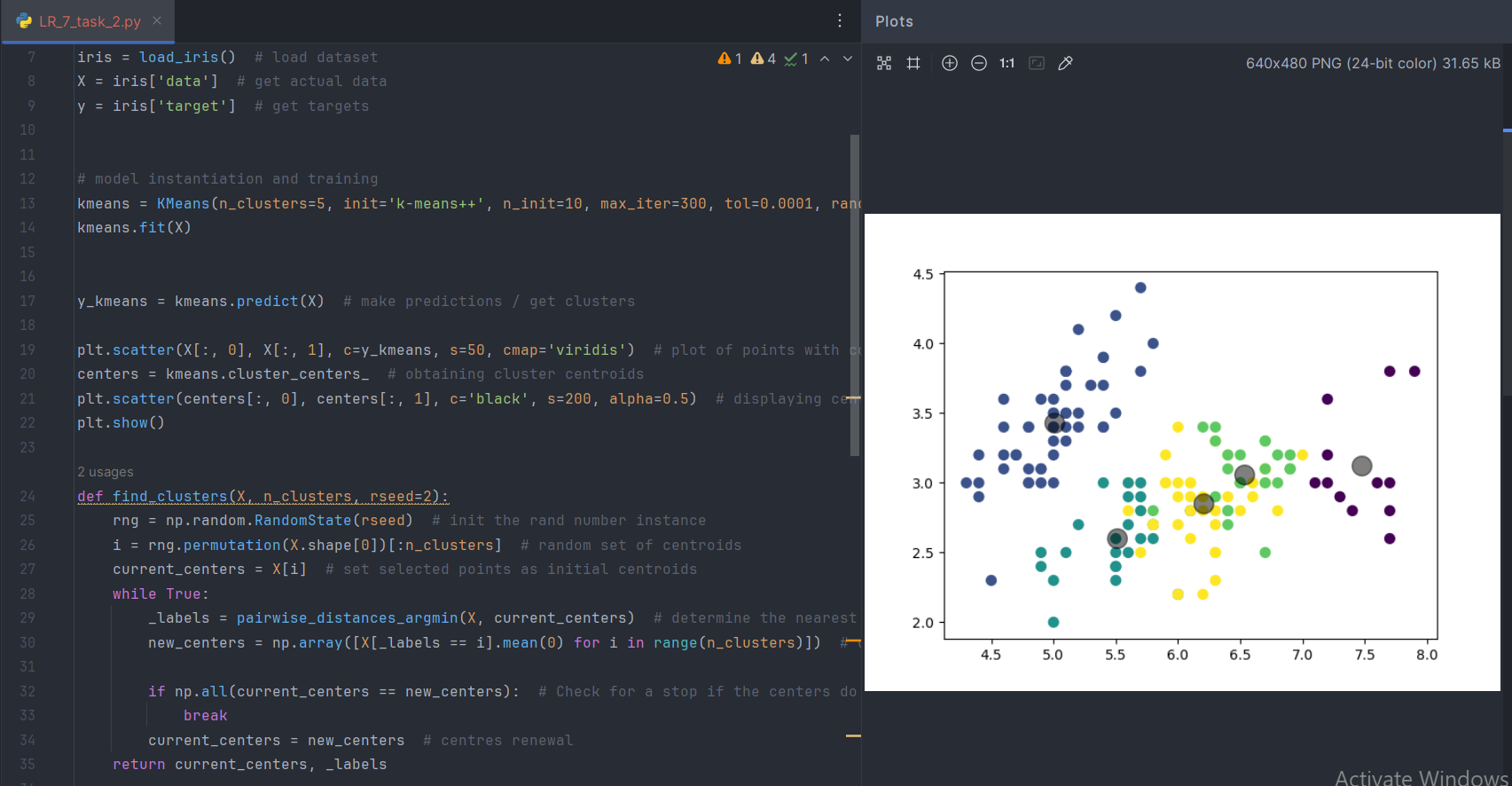
****

Рис 3. Код програми та візуалізація даних з центроїдами

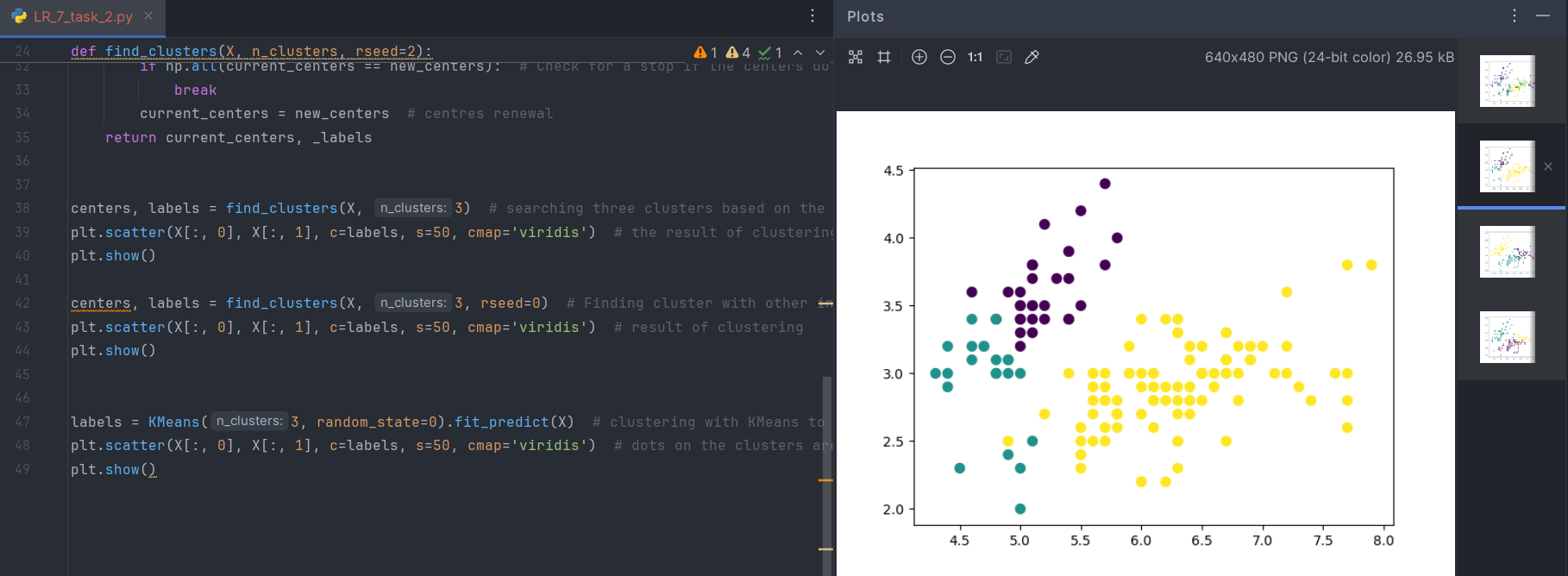


Рис 3. Код програми та візуалізація даних

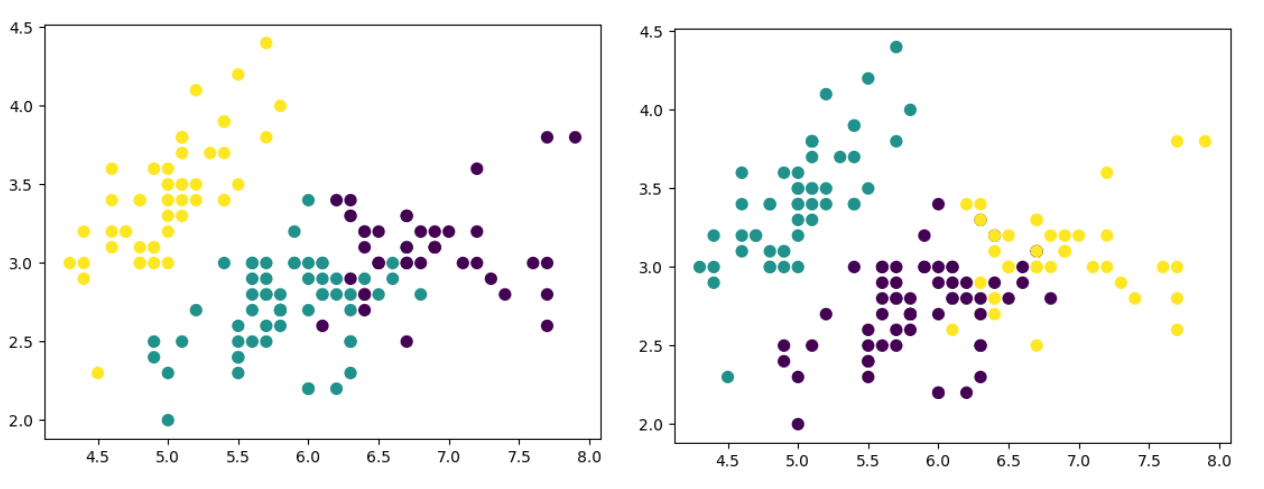


Рис 4. Візуалізація даних з початковими центрами та без

**Висновок:** на даному етапі було візуалізовано стандартну К-means кластерізацію, а також іншу, що використовує метод кластерізації з початковими центрами.

**Завдання 3.** Оцінка кількості кластерів з використанням методу зсуву середнього

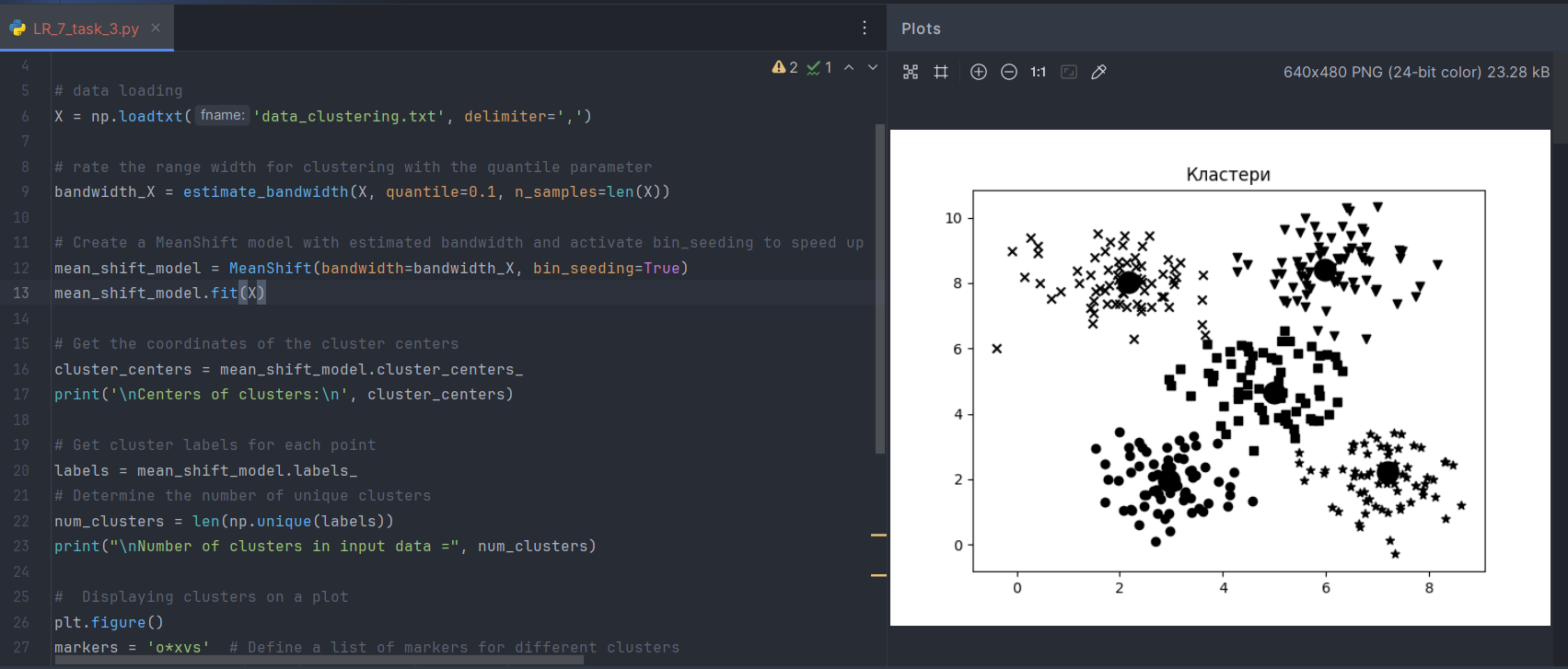


Рис 5. Візуалізація кластеризації

**Висновок:** на даному етапі вдалося виконати кластеризацію даних із використанням алгоритму **MeanShift**, який автоматично визначає кількість кластерів у даних та знаходить їх центри, що можна побачити на графіку

**Завдання 4.** Знаходження підгруп на фондовому ринку з використанням моделі поширення подібності

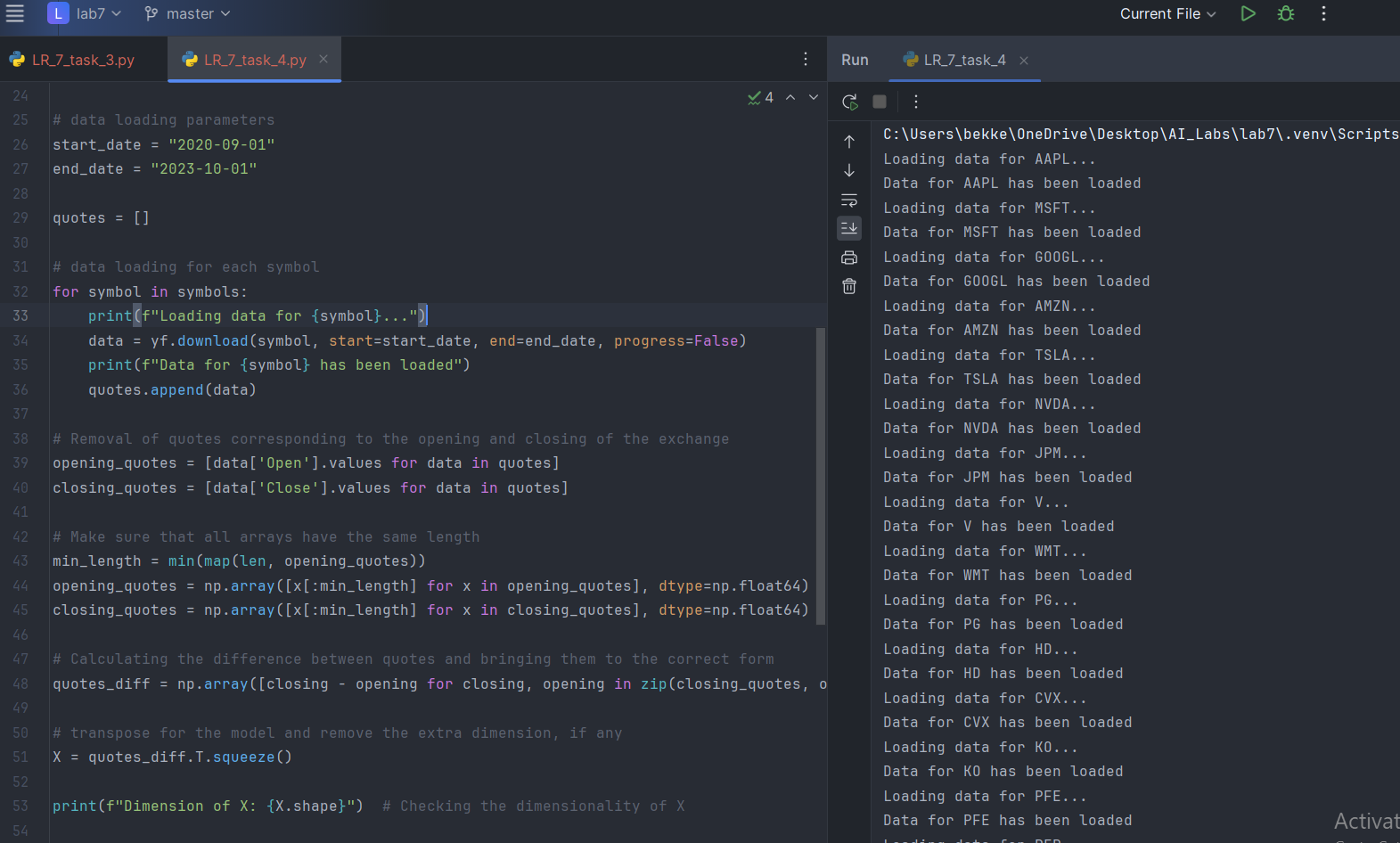


Рис 6. Завантаження даних

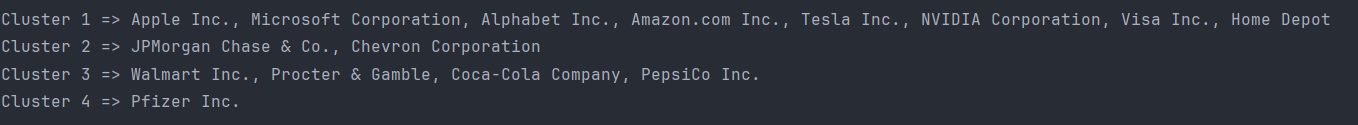


Рис 7. Результати

На даному етапі за допомогою API yfinance вдалося завантажити фінансові дані для зазначених в коді компаній, обчислити зміни між відкриттям і закриттям цін акцій, стандартизувати ці зміни, та виконати кластеризацію компаній. За допомогою метода GraphicalLassoCV вдалося побудувати ковariance-матрицю, яка допомагає зрозуміти взаємозв'язки між компаніями. Також за допомогою алгоритму кластеризації AffinityPropagation, вдалося погрупувати компанії в кластери на основі подібності їх поведінки на ринку. В результати було отримано кластеризації дані, за допомогою яких можна побачити які компанії належать до кожного кластеру.

**Висновок:** на даній лабораторній роботі я навчився використовувати кластерізацію даних за допомогою методу k-cередніх, застосував цей метод на практиці. Також навчився оцінювати кількість кластерів з використанням методу зсуву середнього, застосував ці знання на практиці. Github: [link](https://github.com/Ajoke32/AILabs)