**Міністерство освіти і науки України**

**Львівський національний університет імені Івана Франка**

Факультет електроніки та комп’ютерних технологій

**Звіт**

Про виконання лабораторної роботи №5

# **З курсу «**“Системи опрацювання даних**»**

«Кластеризація даних»

Виконав:

Студент групи ФеС-21

Шавало Андрій

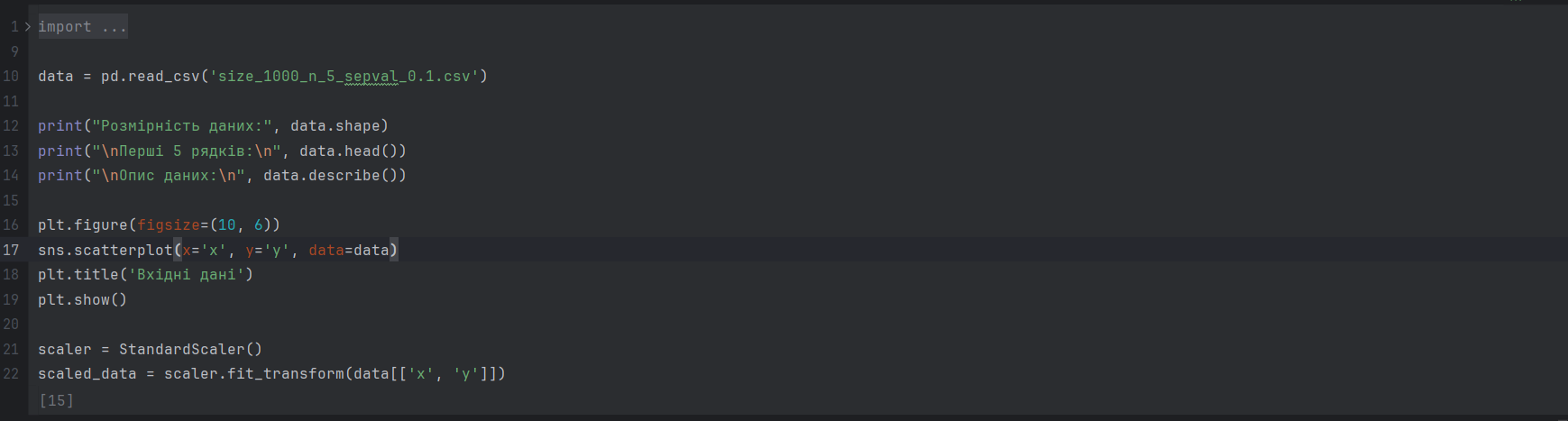
**Львів-2025**

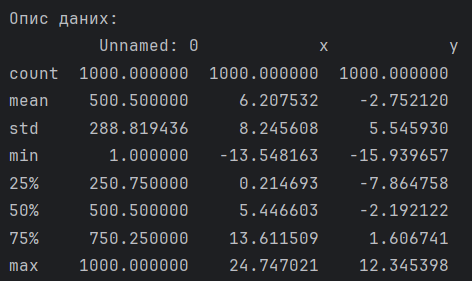
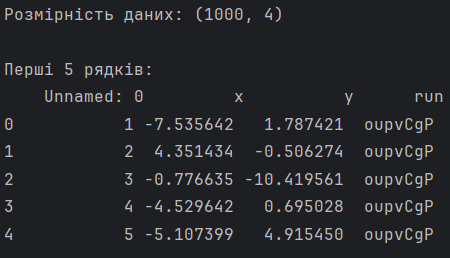
**Мета роботи:** Ознайомитися з методами кластеризації даних, реалізувати та порівняти алгоритми кластеризації (KMeans, DBSCAN, Agglomerative Clustering) на реальному датасеті, а також оцінити якість отриманих клас

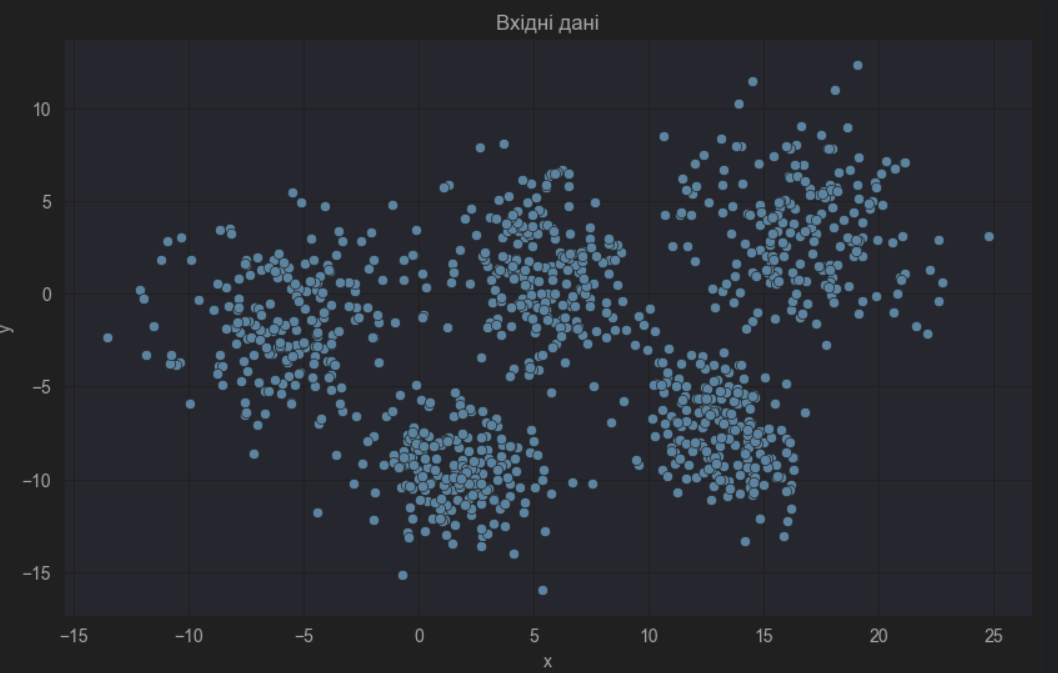
**Теоретичні відомості:** Ознайомитись з алгоритмом кластеризації к-means та прикладом його реалізації https://www.kaggle.com/kushal1996/customer-segmentation-k-meansanalysis/notebook

**Хід роботи**

**Завдання 1.** Підготовка даних o Завантажити набір даних o Виконати первинний аналіз даних: розмірність, типи змінних, наявність пропущених значень. o Провести масштабування (стандартизацію або нормалізацію) числових змінних.

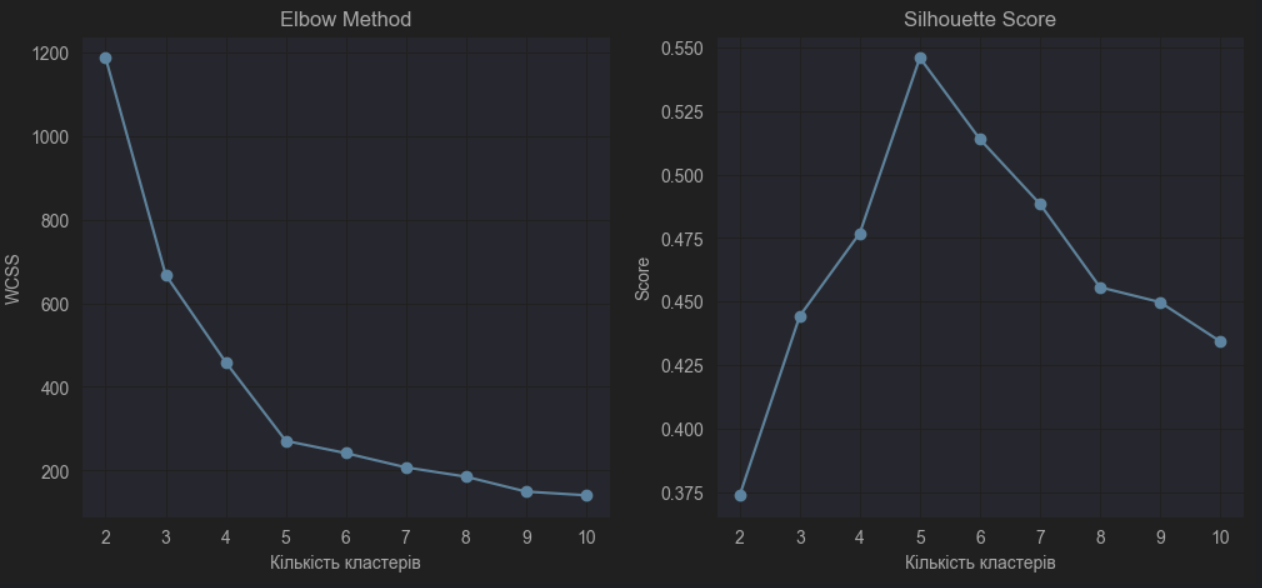




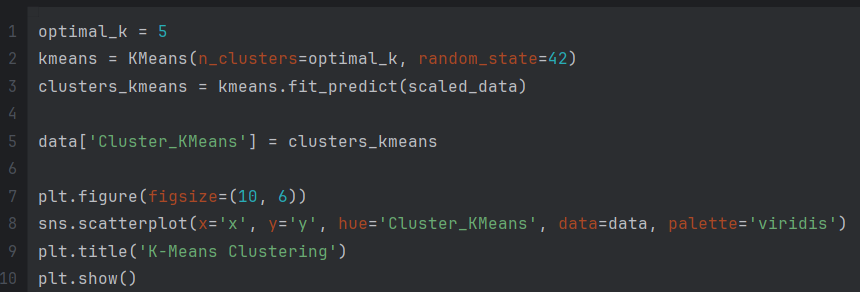


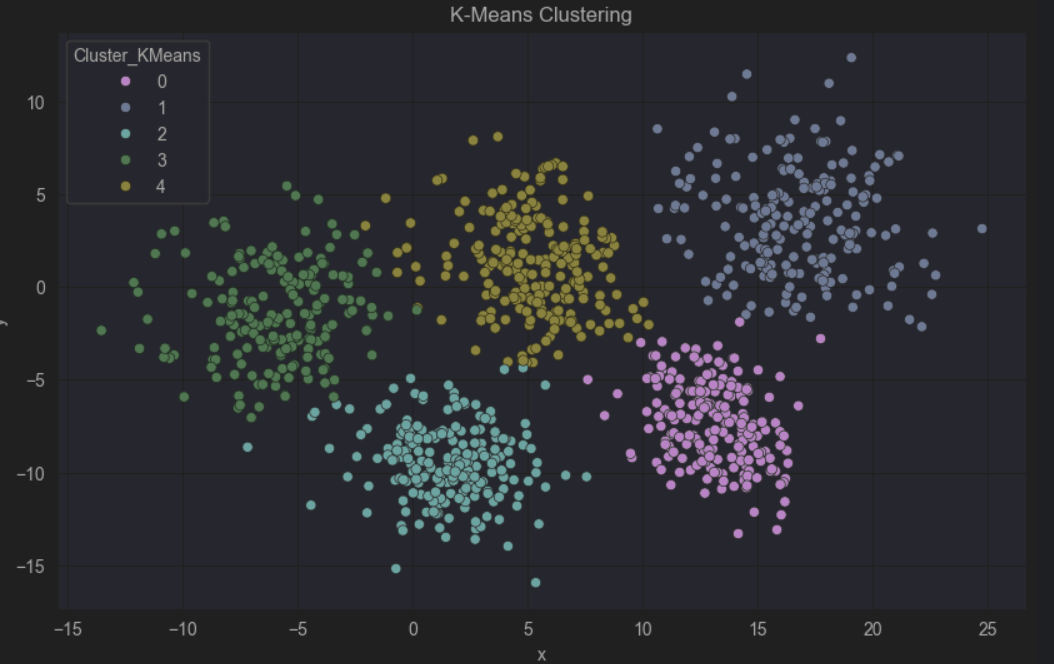
**2.** Реалізація кластеризації o Визначити оптимальну кількість кластерів за допомогою методу "Elbow Method" або Silhouette Score.



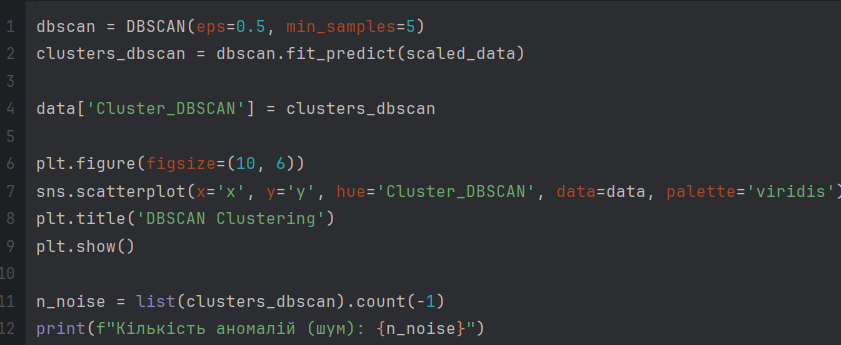


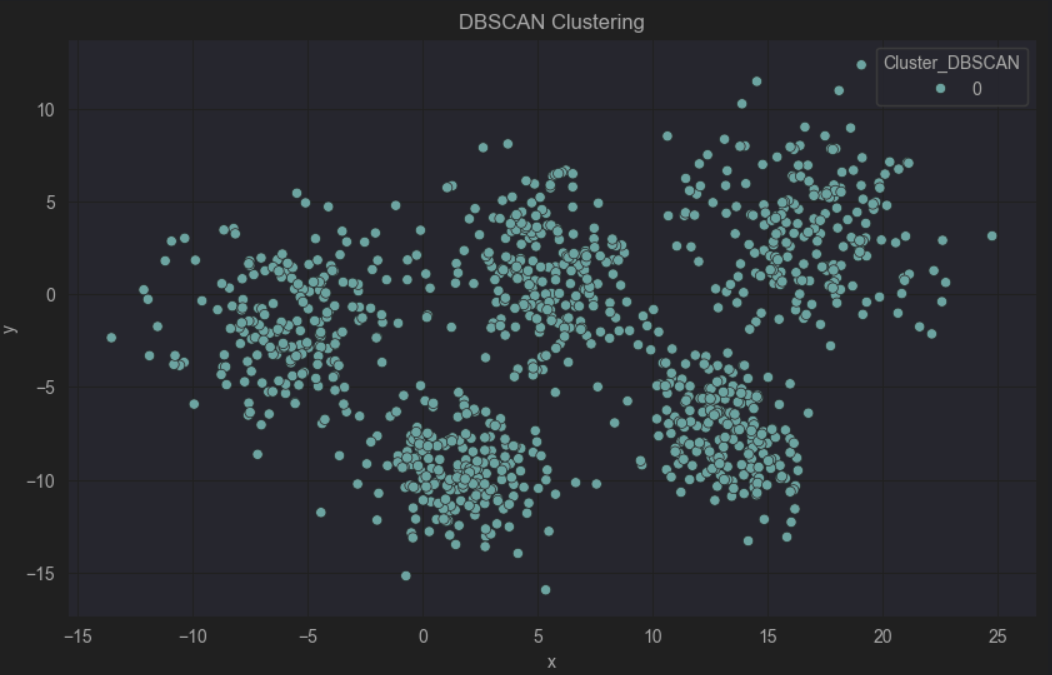
Виконати кластеризацію за допомогою K-Means.





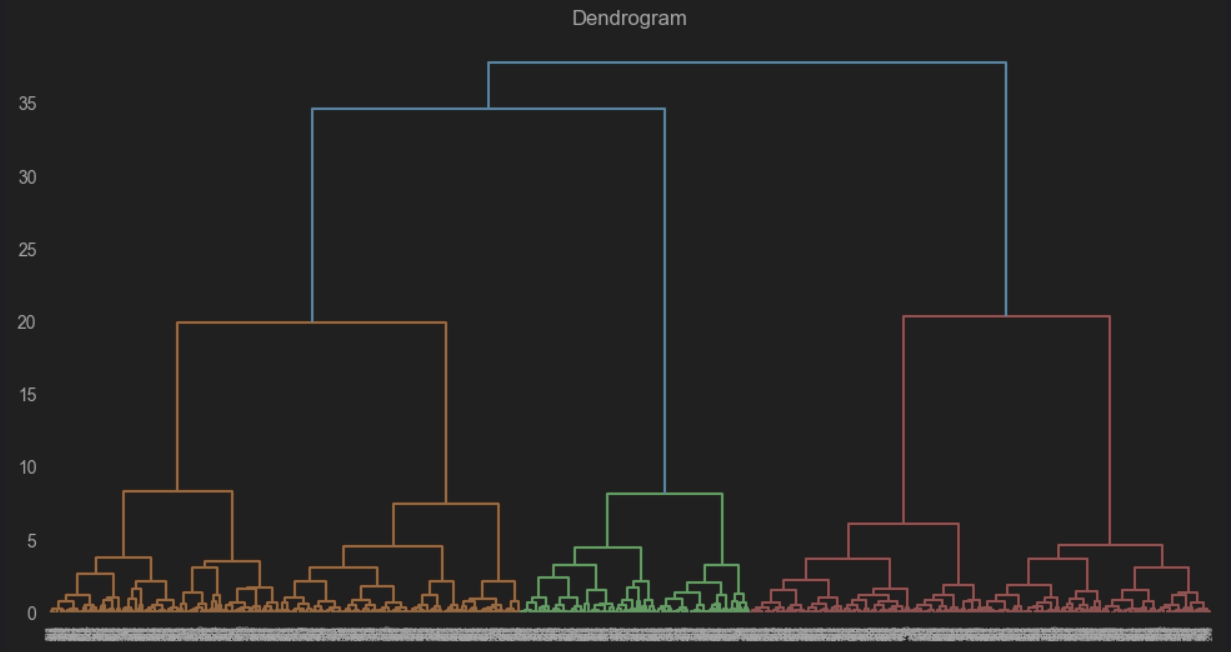
Реалізувати та застосувати DBSCAN для виявлення аномалій.

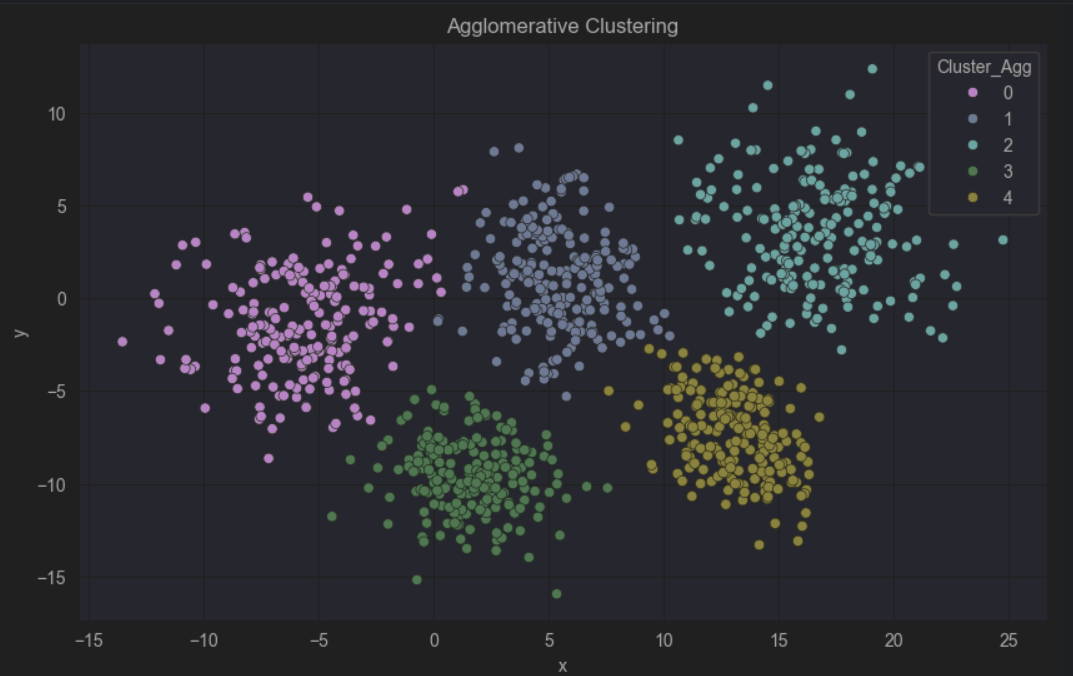




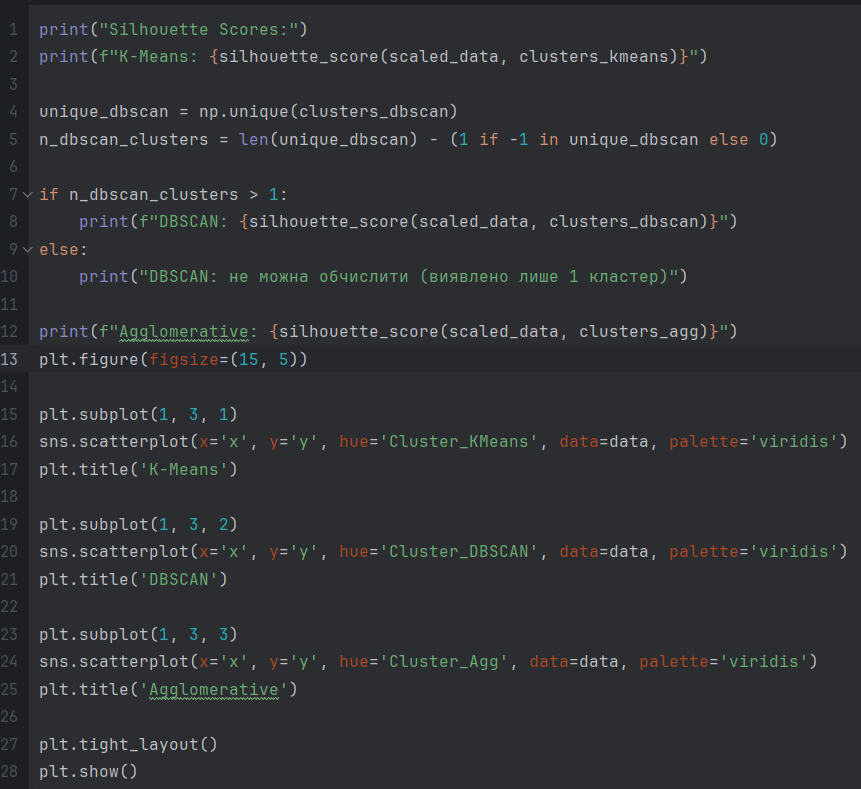
Використати Agglomerative Clustering (ієрархічну кластеризацію) та побудувати дендрограму.

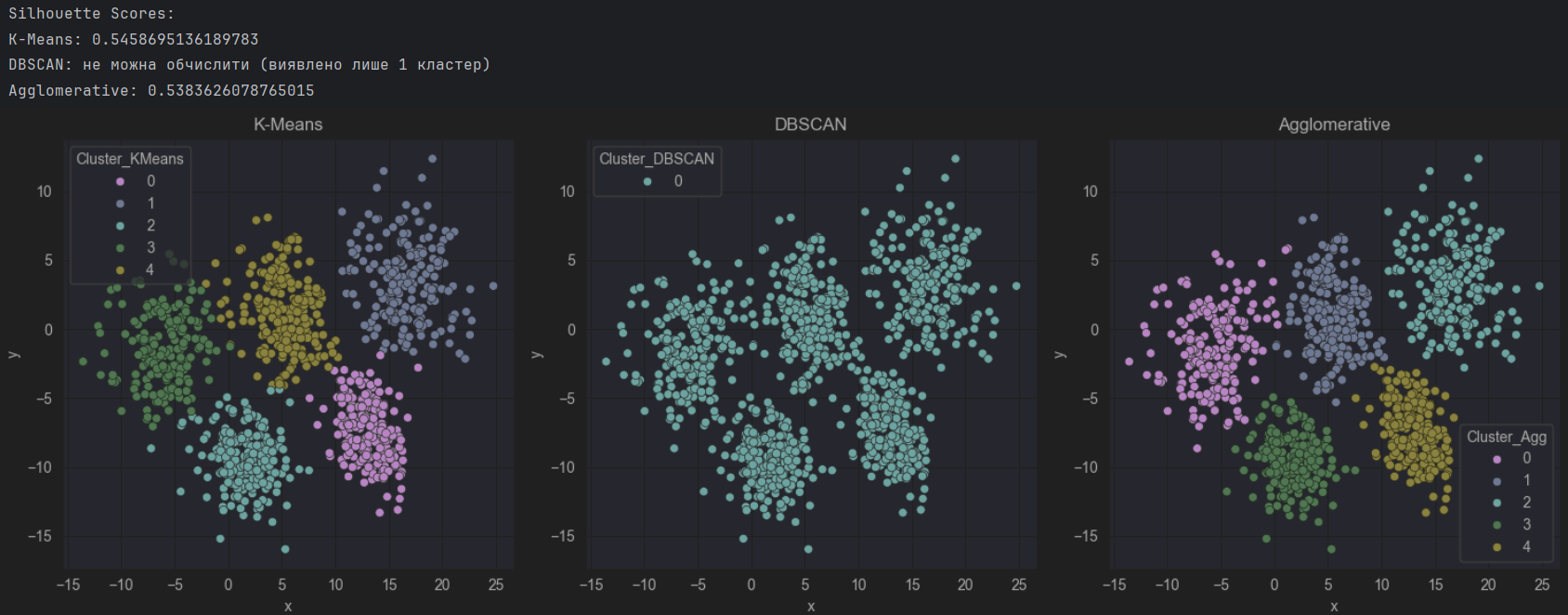






**3.** Оцінка якості кластеризації o Порівняти результати кластеризації за допомогою метрики Silhouette Score o Візуалізувати отримані кластери (наприклад, scatter plot, t-SNE, PCA). o Охарактеризувати кластери. (Наприклад, кластер тих хто часто робить недорогі покупки)





**Висновок:**

У цій лабораторній роботі я ознайомився з методами кластеризації даних та реалізував три алгоритми: K-Means, DBSCAN і Agglomerative Clustering. Я виконав попередню обробку даних, включаючи масштабування, визначив оптимальну кількість кластерів за допомогою "Elbow Method" та Silhouette Score, а також провів аналіз результатів кластеризації.

Порівняння алгоритмів показало, що K-Means добре працює на чітко відокремлених групах даних, DBSCAN виявляє аномалії, але чутливий до параметрів, а Agglomerative Clustering дає гнучкіший розподіл, проте може страждати на великому обсязі даних. Візуалізація кластерів дозволила оцінити якість отриманих розподілів.