**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

ЗВІТ

Про виконання лабораторної роботи №4

з дисципліни: «Data Science та Big Data»

Студент групи ПП-42

Черниш Дмитро

1. Київ – 2022

**Лабораторна робота №4.**

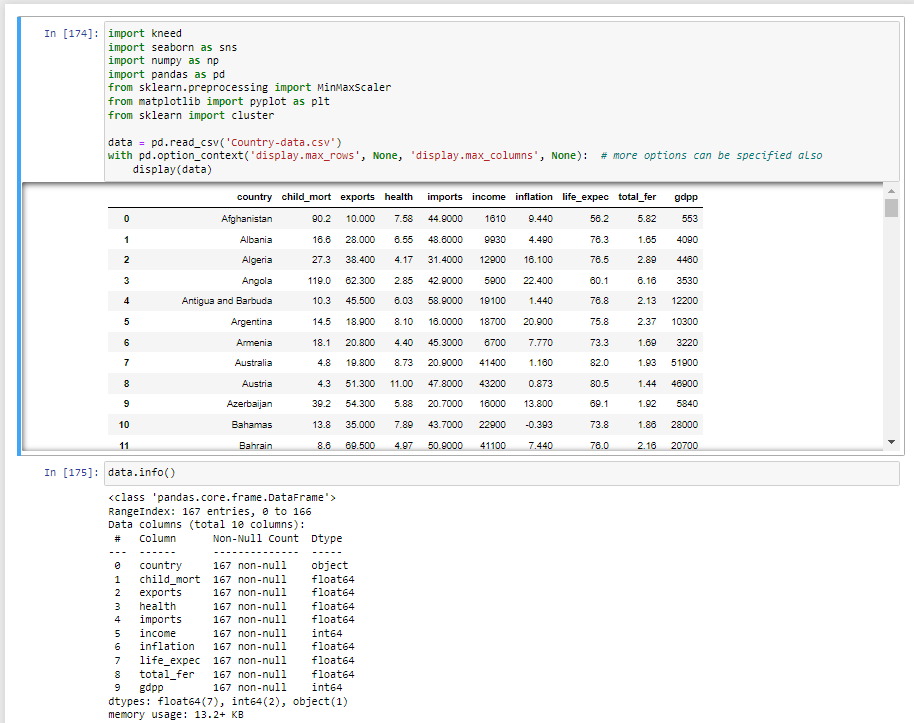
**Тема роботи:** Кластеризація та пошук оптимальних параметрів для неї

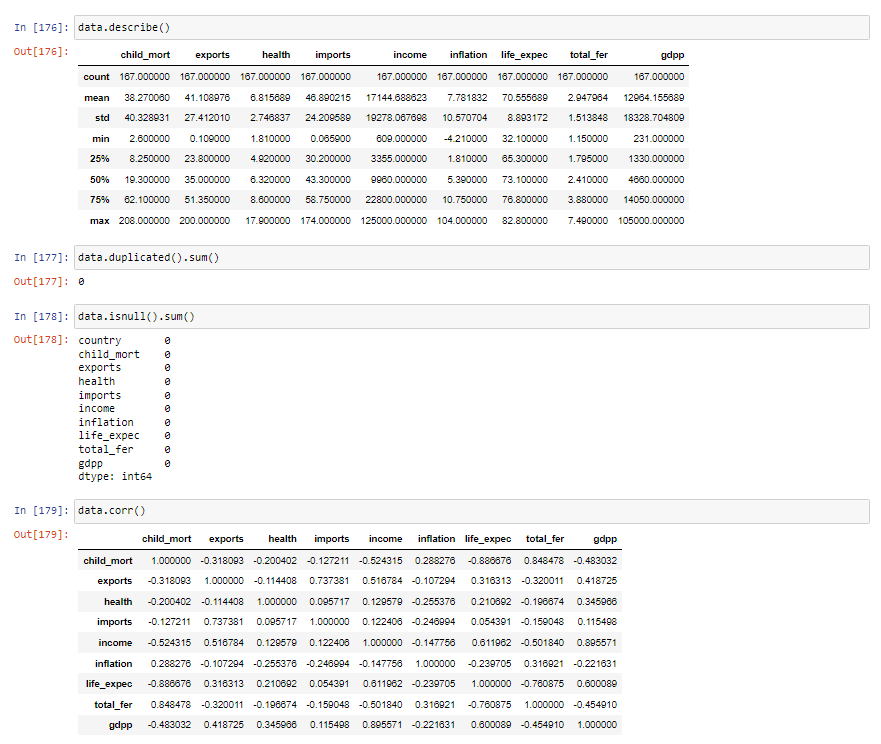
**Завдання для лабораторної роботи**

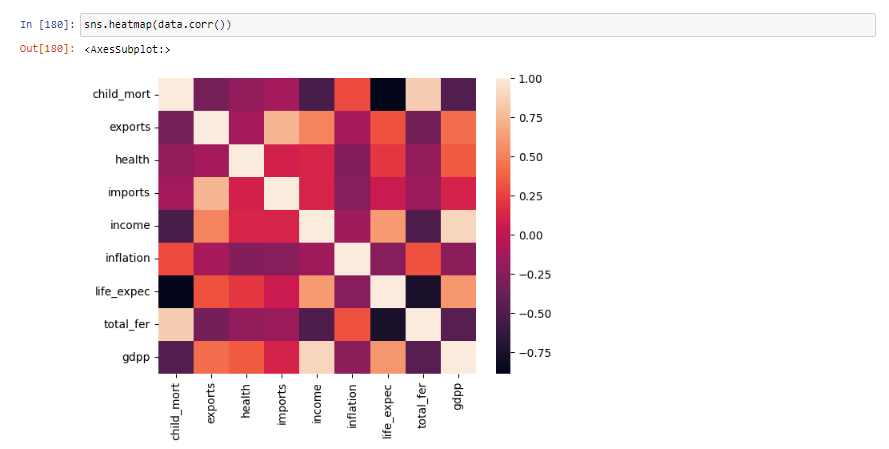
* Прочитайте надану в **Супровідних матеріалах** документацію та вивчить надані приклади в папці – Examples.
* Завантажити файл з даними із папки – data.
* Почніть із необхідної перевірки даних, очищення даних та EDA, зробіть однофакторний аналіз, двофакторний аналіз тощо. Візуалізуйте закономірності.
* Ви повинні виконати аналіз викидів для набору даних. Вибрати, залишити їх чи видалити залежно від отриманих результатів.
* Провести кластеризацію, використовуючи k-mean методологію та інші (не менше 5) з прикладів та документації.
* Спробуйте оптимізувати гіпер-параметри для кожного з обраних методів кластеризації.
* Який з протестованих вами методів дав найкращий результат?
* Візуалізуйте результати кластеризації та оптимізації гіпер-параметрів.
* Агрегуйте дані з аналізом ефективності різних методів кластеризації у результуючий dataframe.
* Підготуйте summary, використовуючи лише необхідні матеріали.
* Завантажити файл звіту та ipynb  з виконаними завданнями на git в окрему папку з відповідною назвою лабораторної роботи

**Хід роботи:**

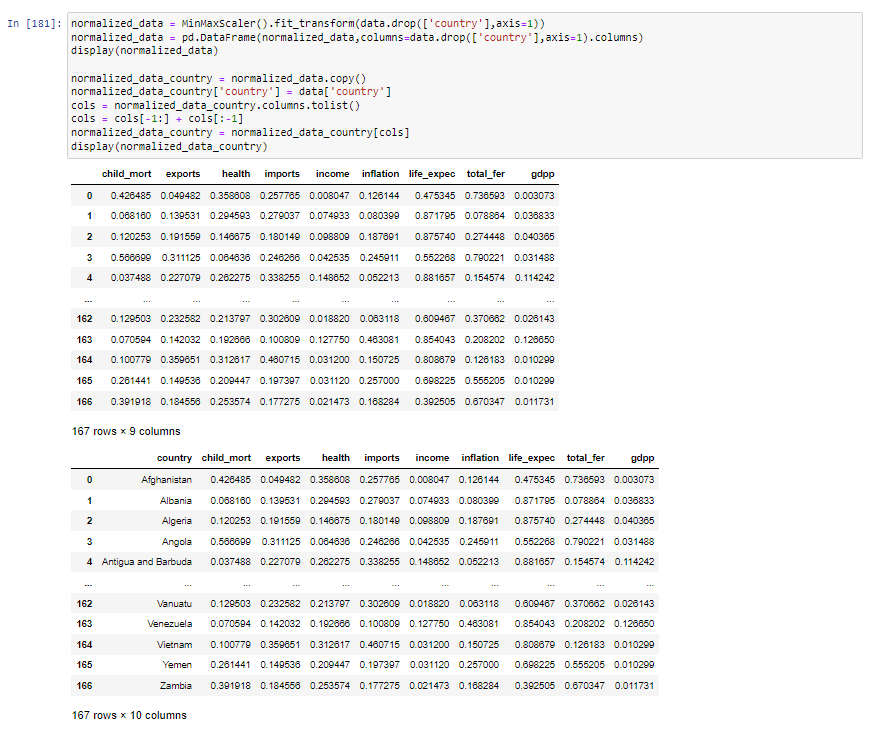
Провів EDA



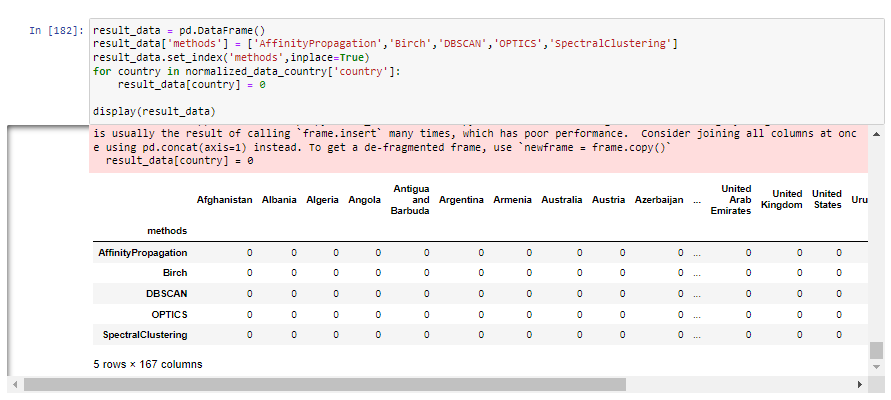




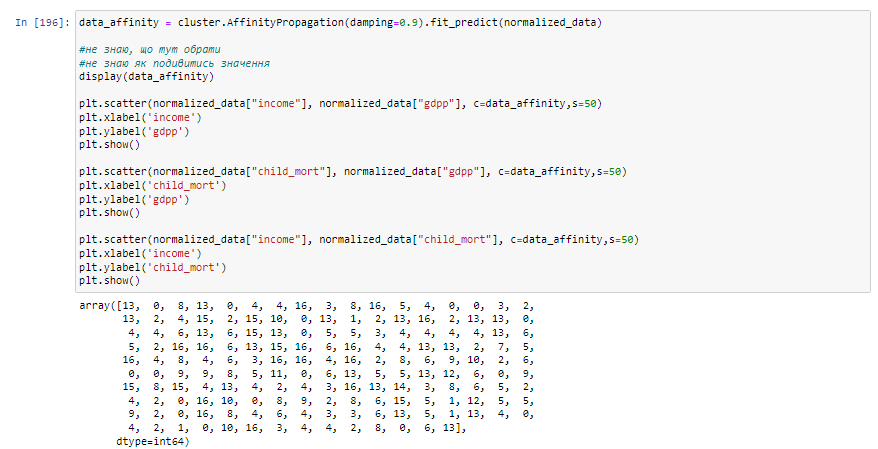
Нормалізував дані

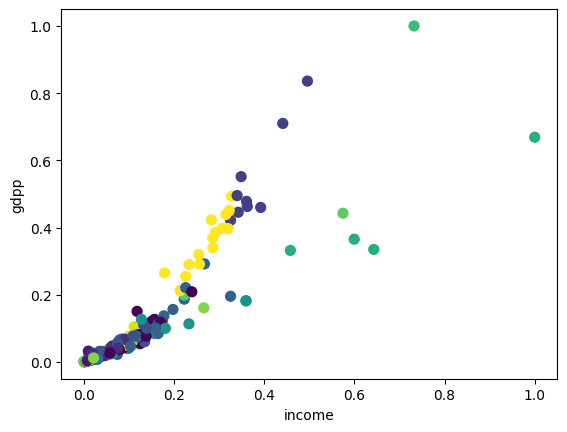


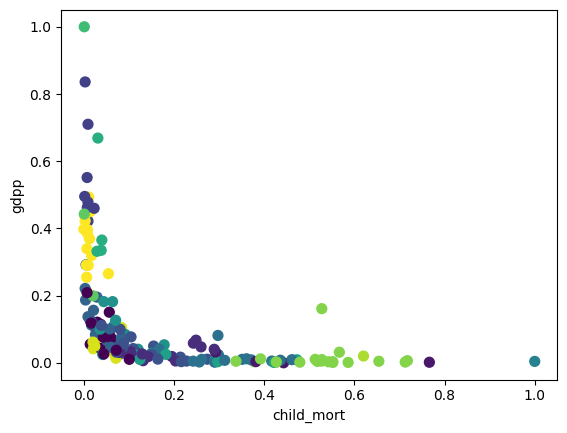
Створив результуючу таблицю

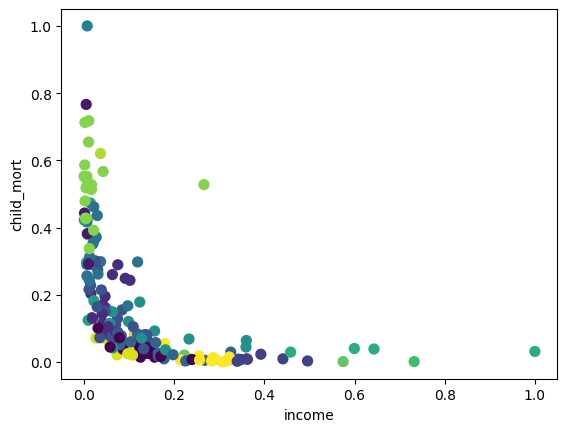


Використав алгоритм AffinityPropagation, однак не став заносити результати до таблиці, оскільки важко обрати один кластер



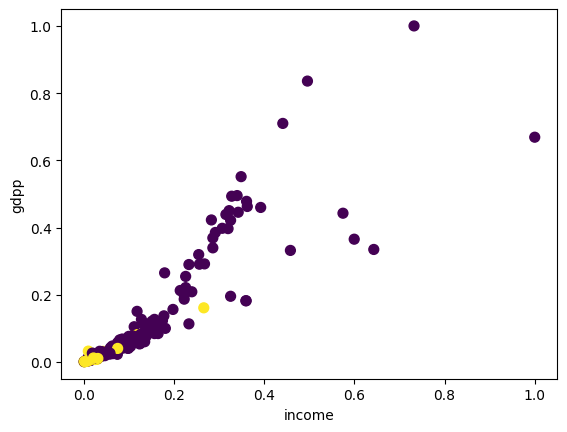


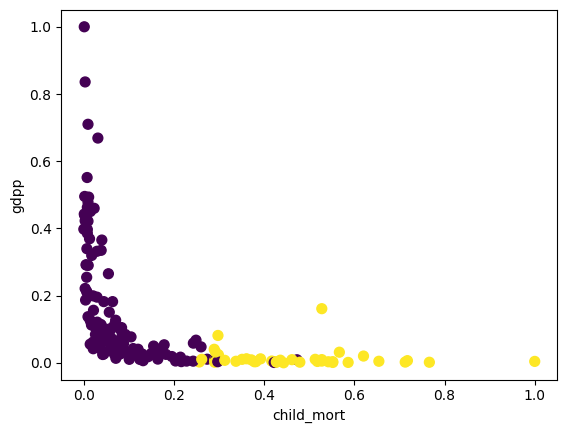


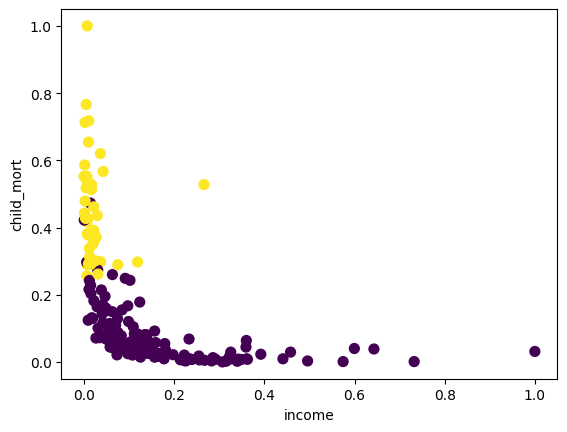


Використав алгоритм Birch



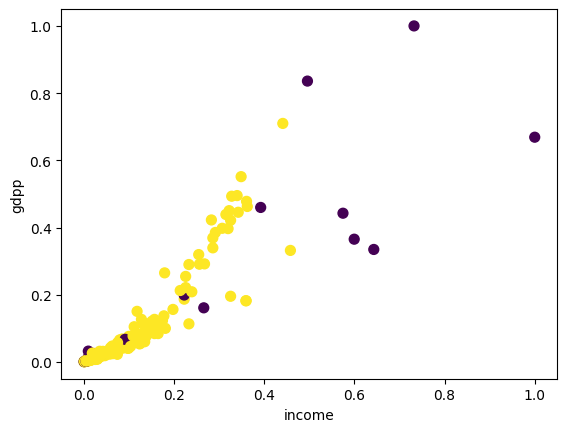


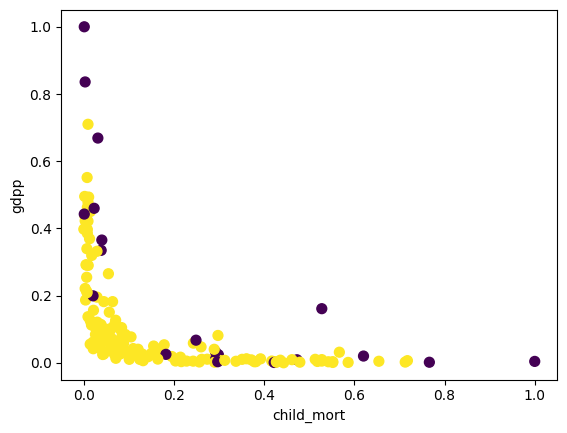


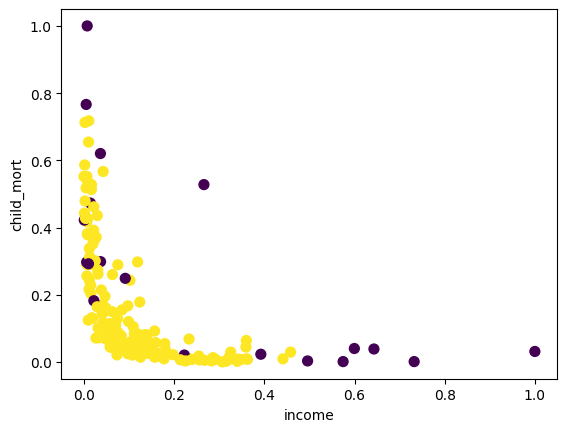


Використав алгоритм DBSCAN

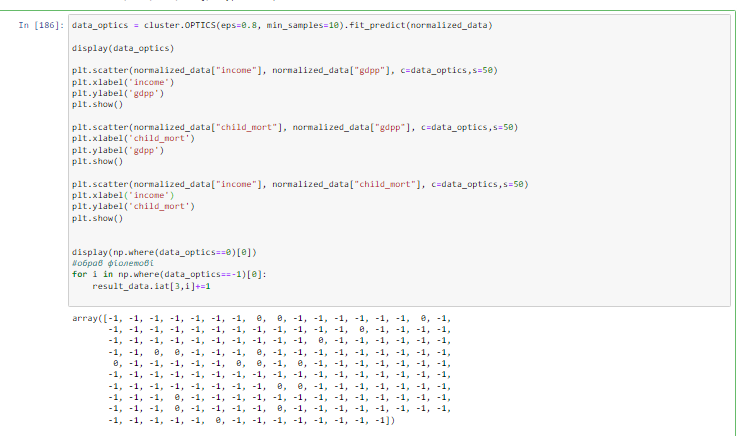


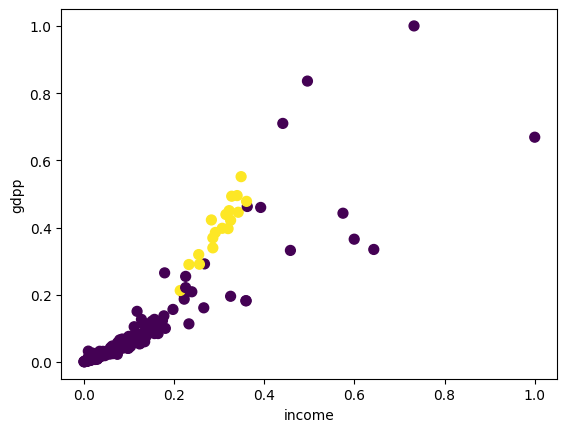


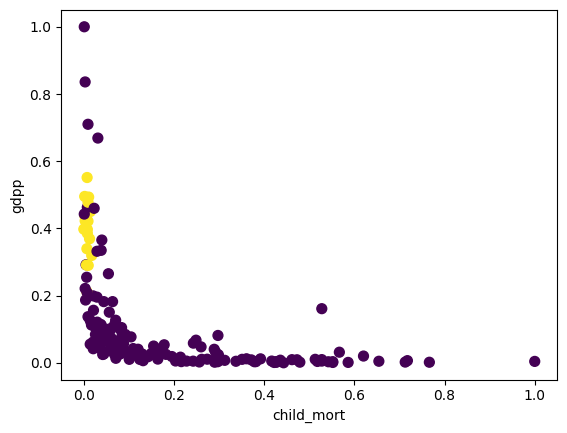


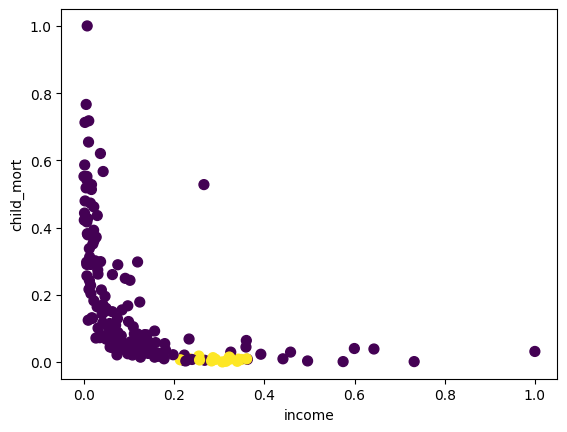


Використав алгоритм OPTICS



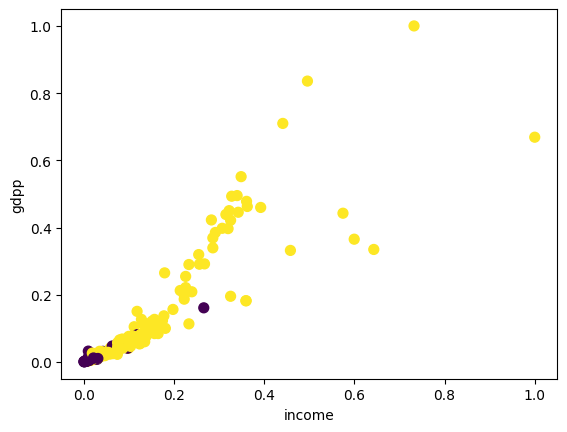


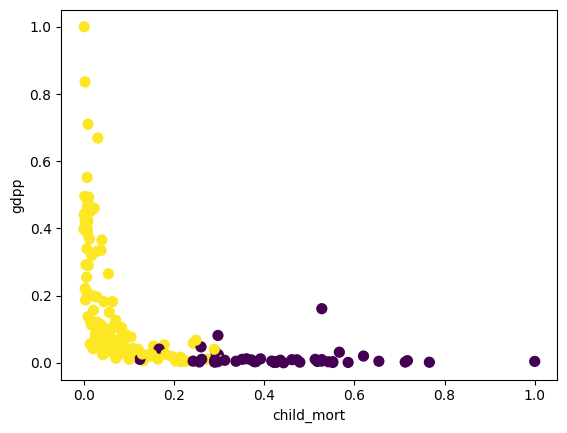


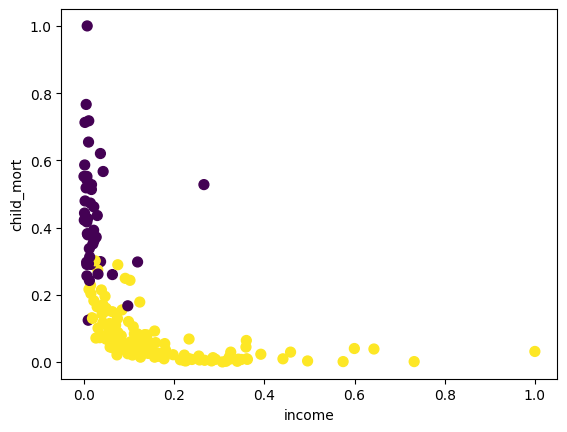


Використав алгоритм SpectralClustering

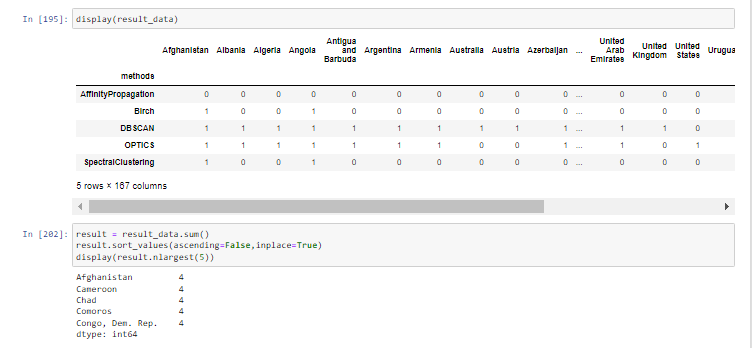








Підрахував дані та вивів 5 рекомендованих країн



**Висновки:**

У цій лабораторній я кластеризував дані та виділив країни, що потребують економічної допомоги