**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Звіт до практичної роботи №4**

**з курсу**

**«Інтелектуальний аналіз даних»**

*Студентки 3 курсу групи ПП-32*

*спеціальності 122 «Комп'ютерні науки»*

*ОП «Прикладне програмування»*

Боголій Владислави Олегівни

*Викладач:*

Білий Р.О.

**Київ – 2023**

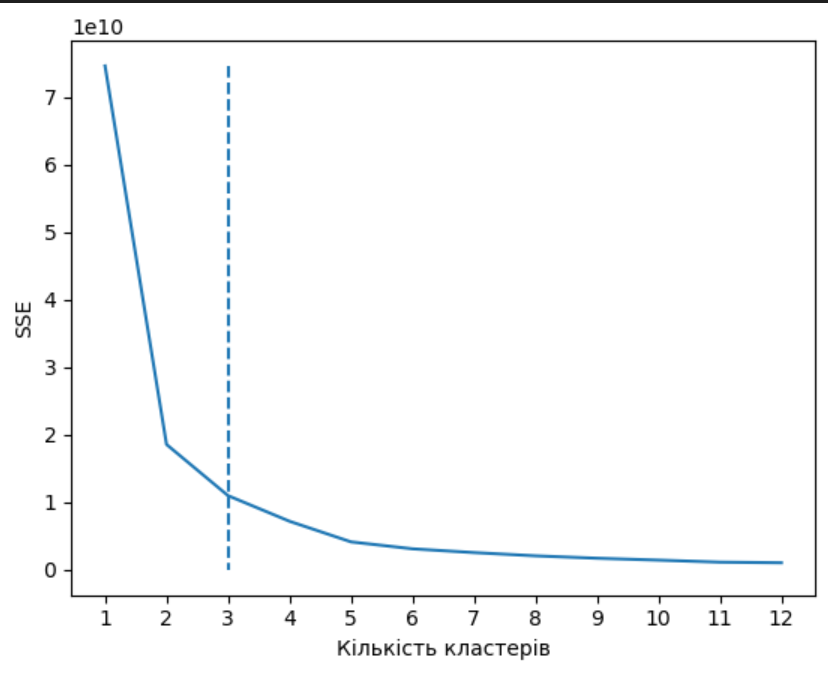
**Тема:** Кластеризація. Метод k-mean, пошук оптимальних параметрів

**Мета:** Метою лабораторної роботи є отримання практичних навичок з виконання кластеризації даних, використовуючи метод k-mean та інші з пакету sklearn.

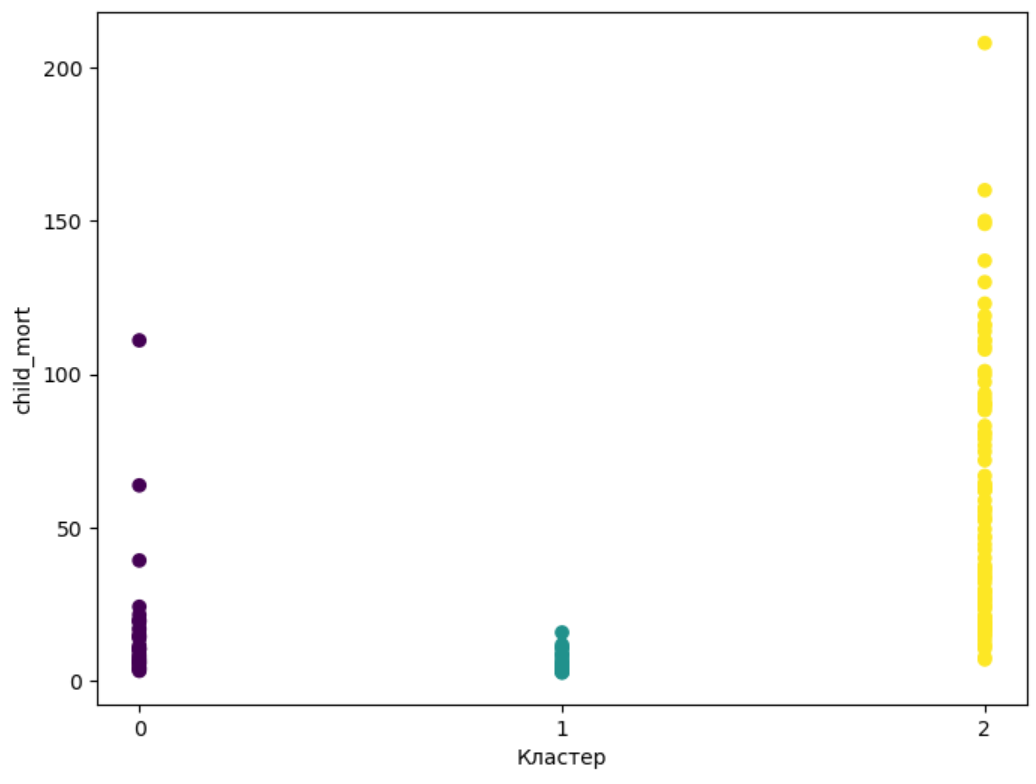
**Висновок:**

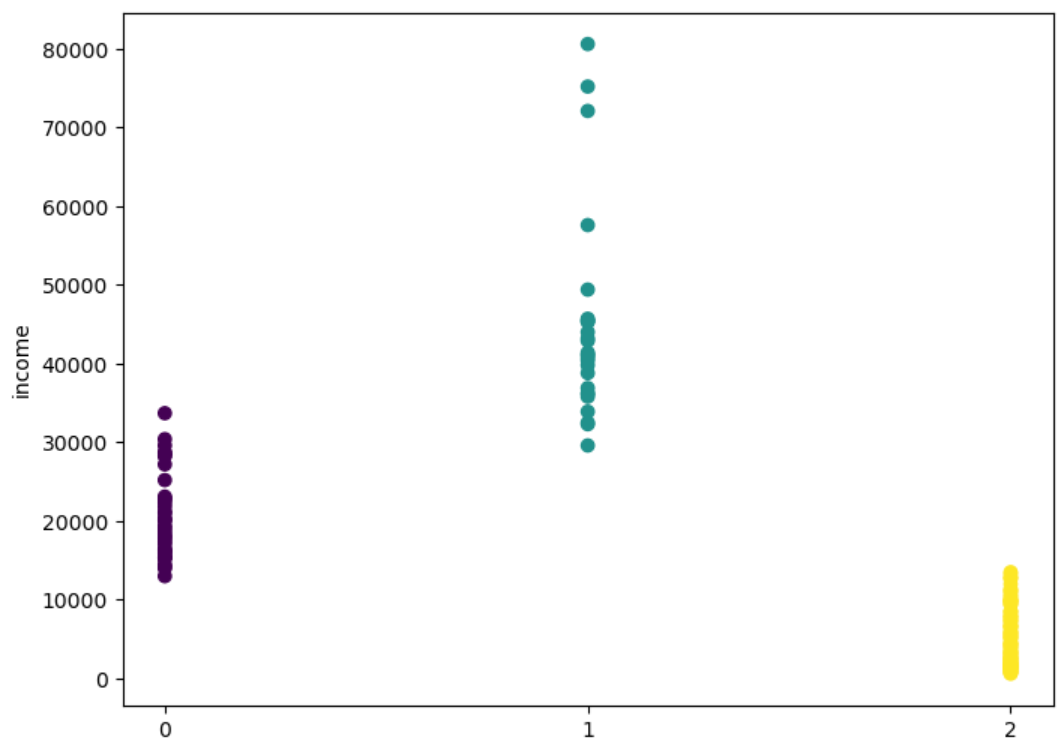
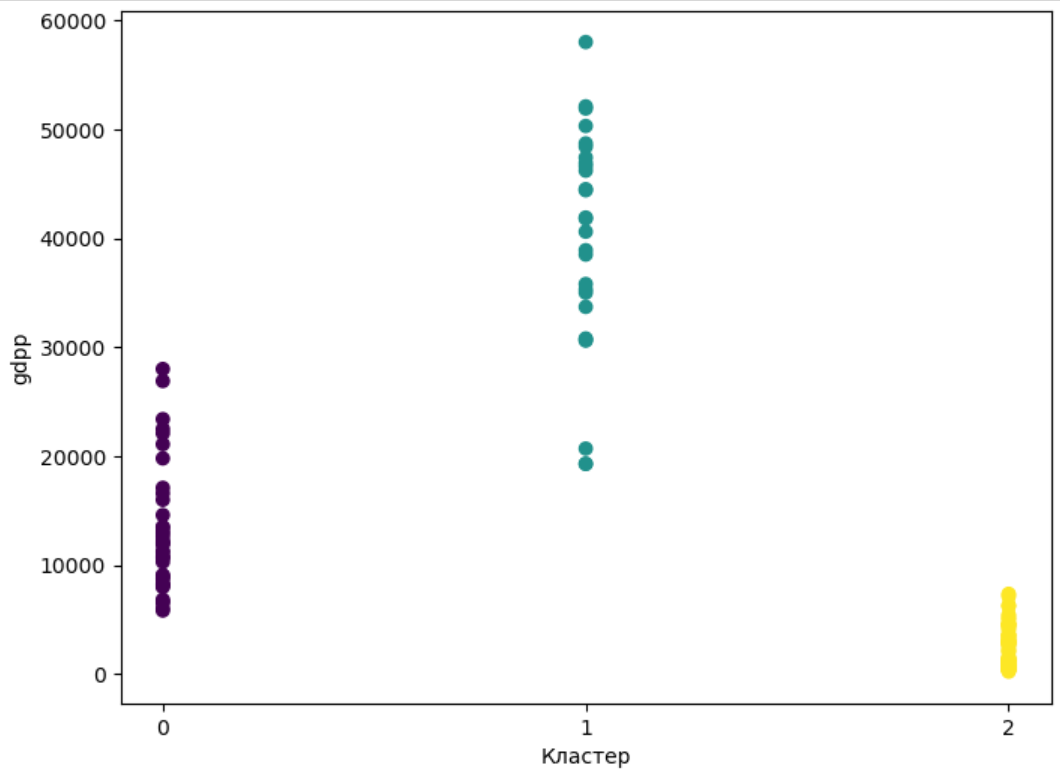
Перш за все, здійснивши аналіз викидів для даних, ми виявили кілька країн, для яких фактори доходу та ВВП мають значення достатньо високі для того, щоб не претендувати на допомогу.

Після цього методом «ліктя» було визначено оптимальну кількість кластерів для поділу країн.

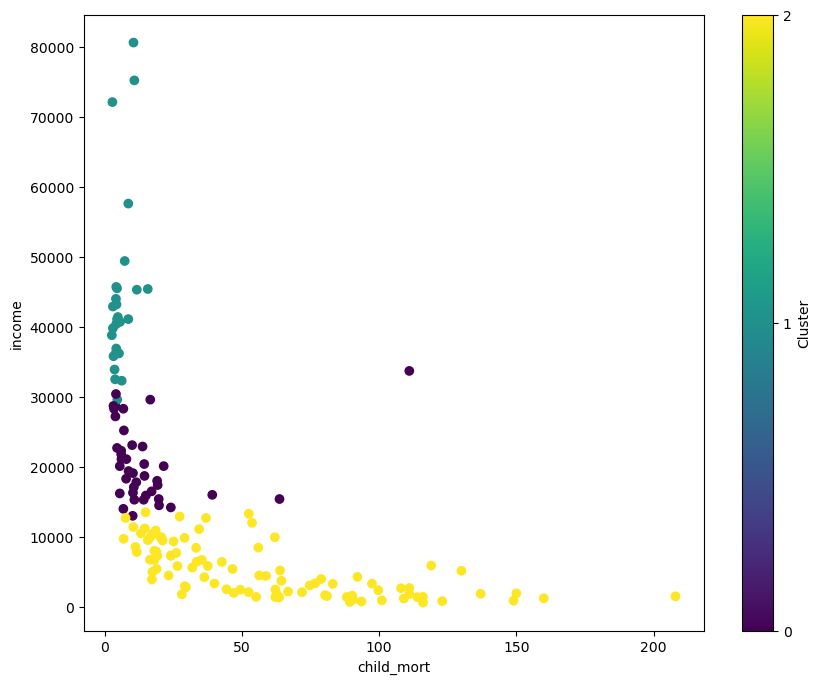


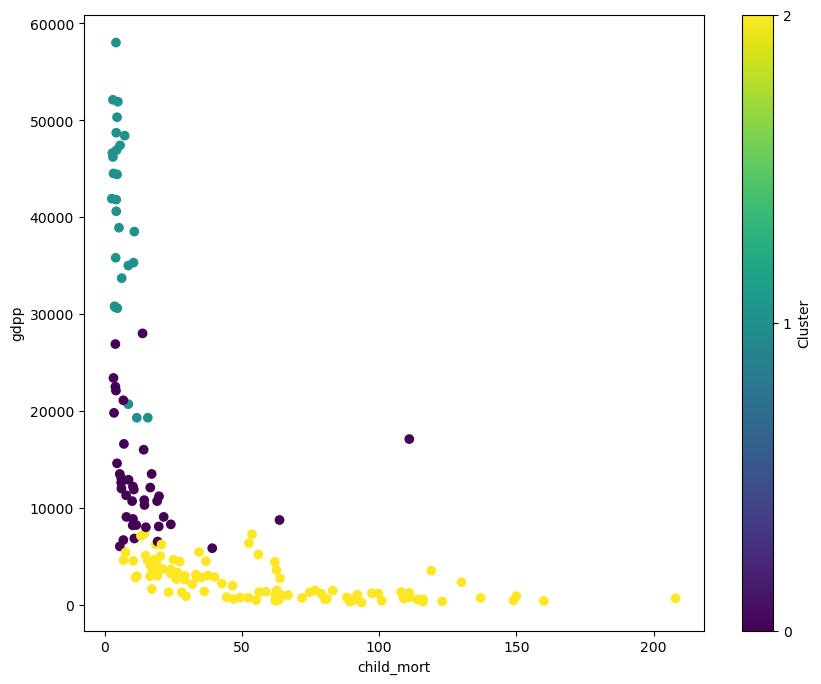
Після цього ми розподілили країни між цими кластерами та провели аналіз факторів щ урахуванням поділу країн на кластери.

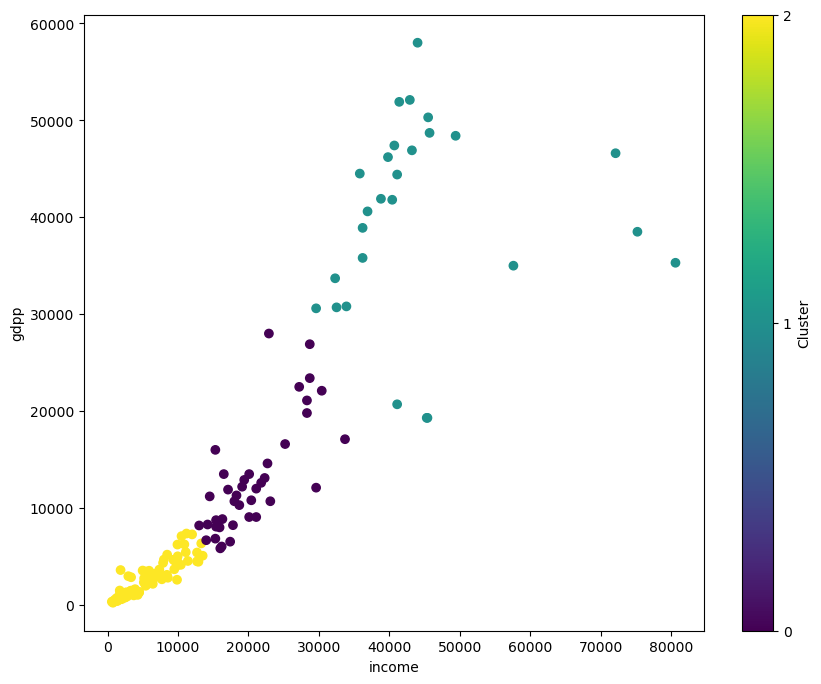


Також візуалізували залежність факторів.







У результаті обчислення середнього та візуалізації зрозуміло, що країни з кластеру 2 потребують найбільше допомоги.

Для таких країн визначимо ті, в яких найбільша дитяча смертність та найменші дохід та ВВП. Результат: 10 країн у порядку першочерговості для надання допомоги.

country child\_mort income gdpp

26 Burundi 93.6 764 231

88 Liberia 89.3 700 327

37 Congo, Dem. Rep. 116.0 609 334

111 Niger 123.0 814 348

129 Sierra Leone 160.0 1220 399

92 Madagascar 62.2 1390 413

105 Mozambique 101.0 918 419

31 Central African Republic 149.0 888 446

93 Malawi 90.5 1030 459

146 Togo 90.3 1210 488