Міністерство освіти і науки України

Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій

Кафедра прикладних інформаційних систем

Звіт

з лабораторної роботи №4

Дисципліна «Інформаційний аналіз даних»

Тема: **Кореляційний аналіз даних**

Виконав:

Афанасьєв Антон Антонович

Група ПП-31

Перевірив:

Білий Р. О.

Київ 2022

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

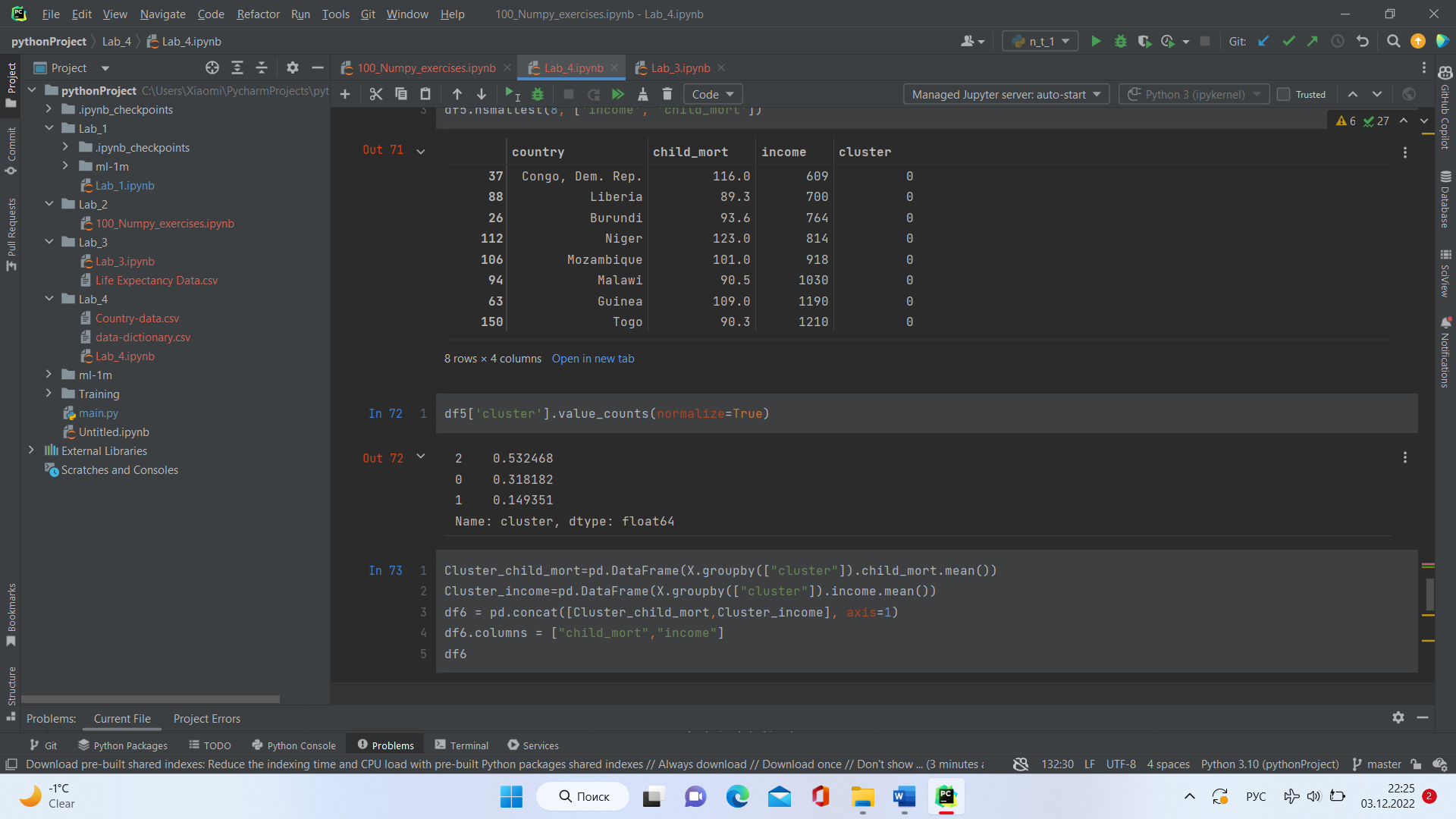
**Мета:** Метою лабораторної роботи є отримання практичних навичок з виконання кластеризації даних, використовуючи метод k-mean та інші з пакету sklearn.

***Завдання***

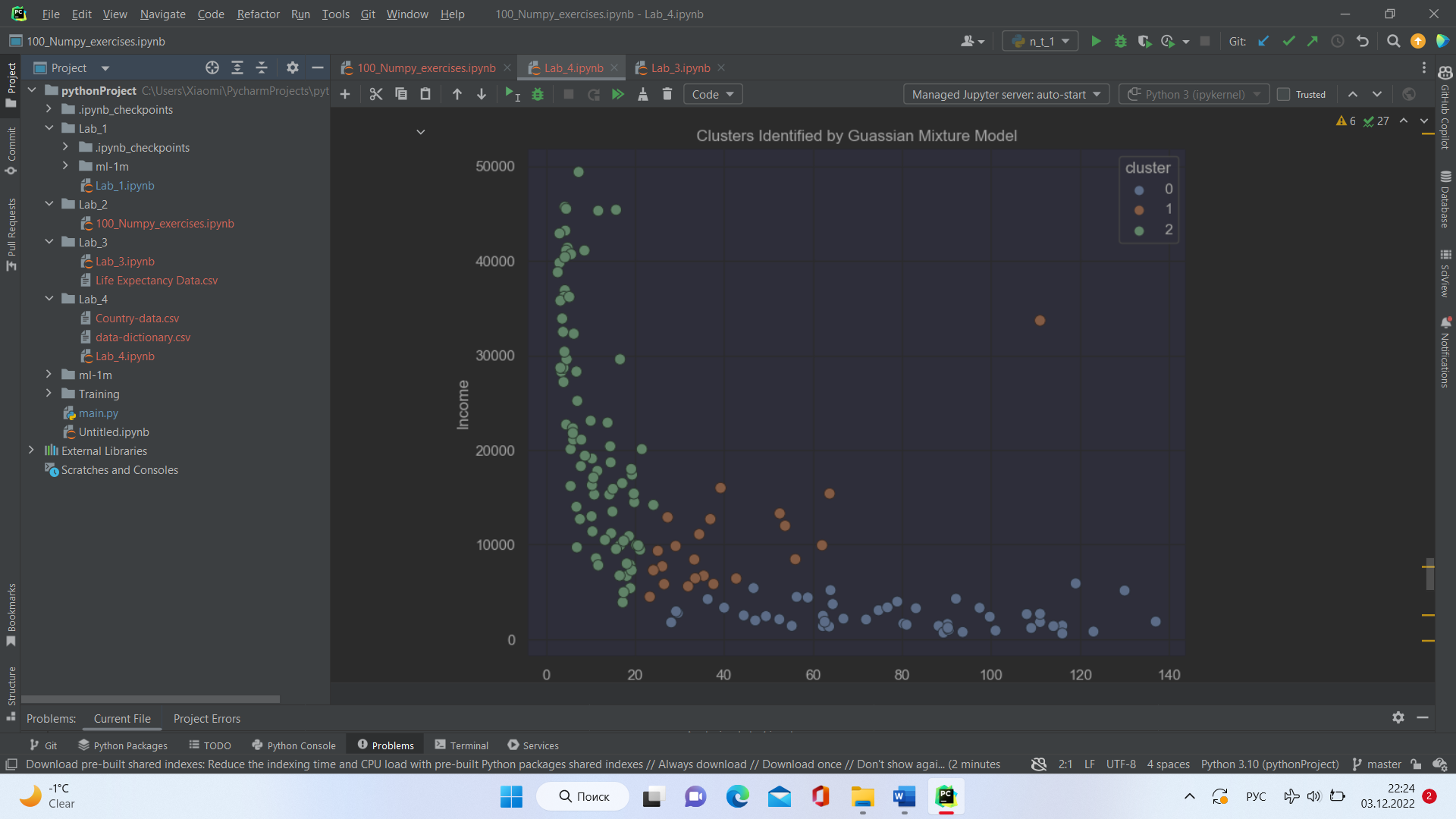
Потрібно виявити принаймні 5 країн, які найбільше потребують допомоги в результаті аналізу, який ви виконуєте. Включіть у звіт необхідні графіки та текстову інформацію.

Було підраховано відсотковий розподіл країн по кластерам.

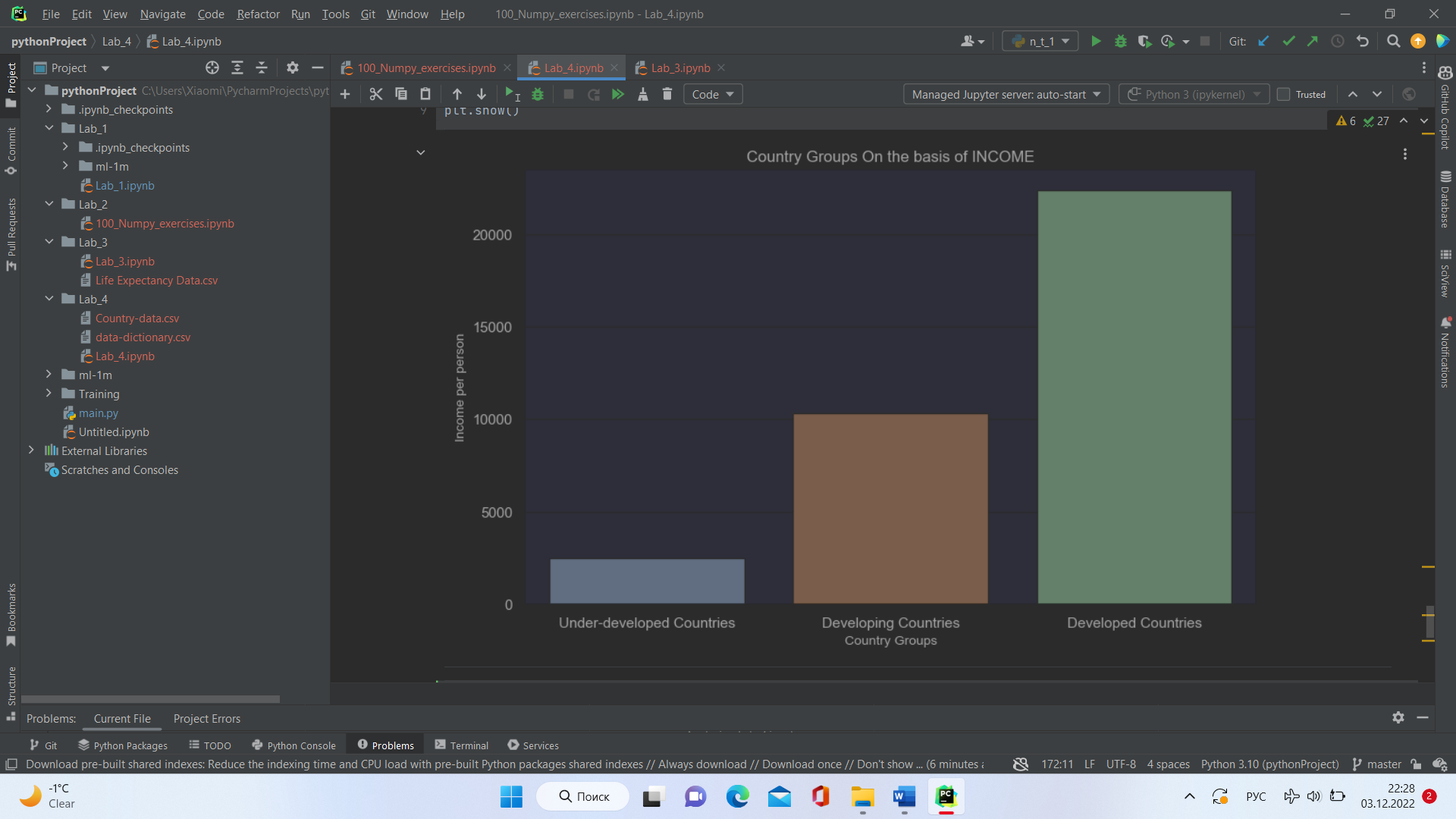
Всього країни поділені на 3 кластери: 0 – країни з низьким доходом та високою смертністю дітей, 1 – середні по доходу та по смертності, 2 – розвинені країни з високим доходом та низькою смертністю.



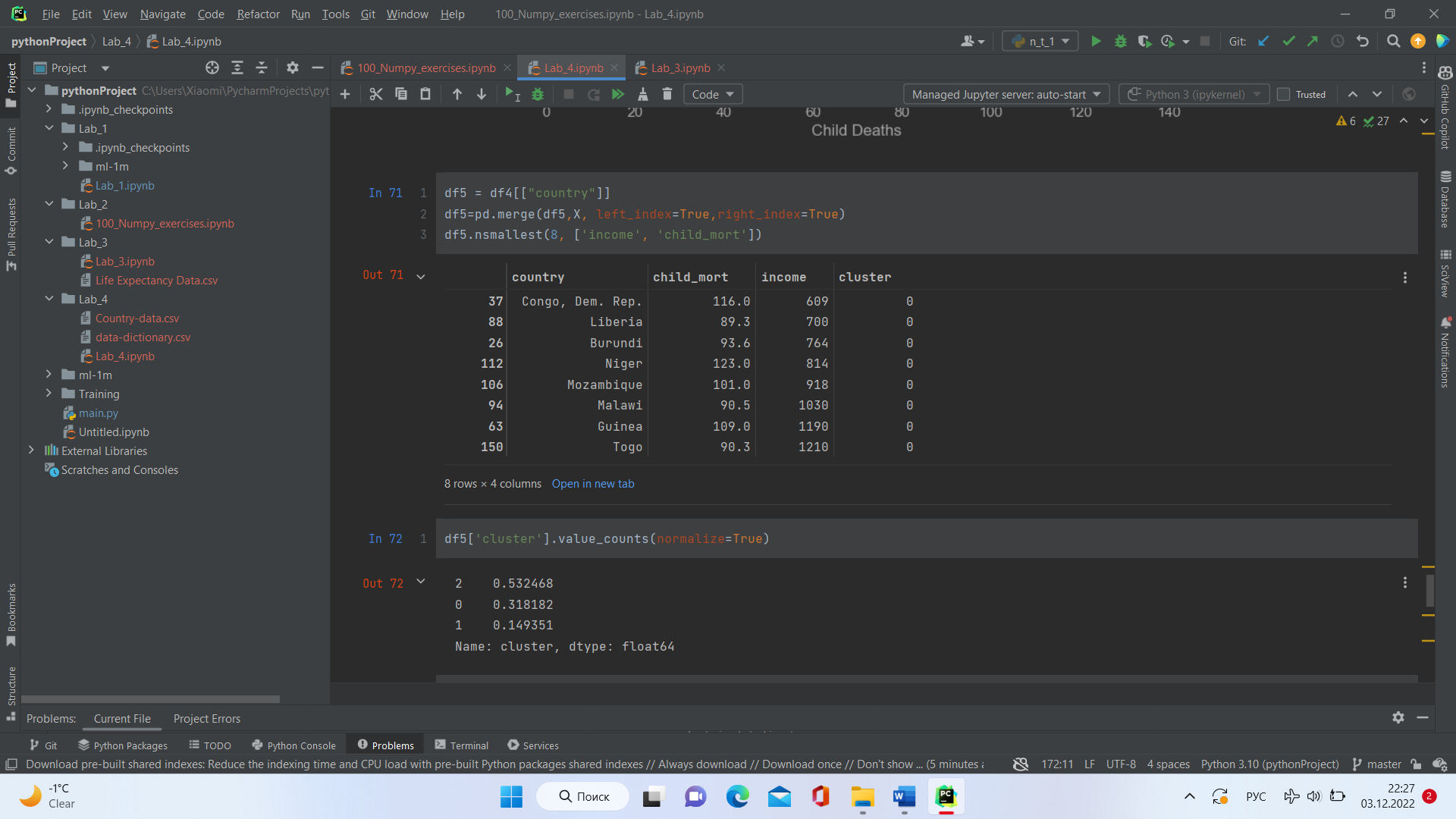
На цьому графіку можна показано розподіл середнього доходу на людину та дитячої смертності(к-сть на 1000 новонароджених). Кожна точка на графіку це окрема країна



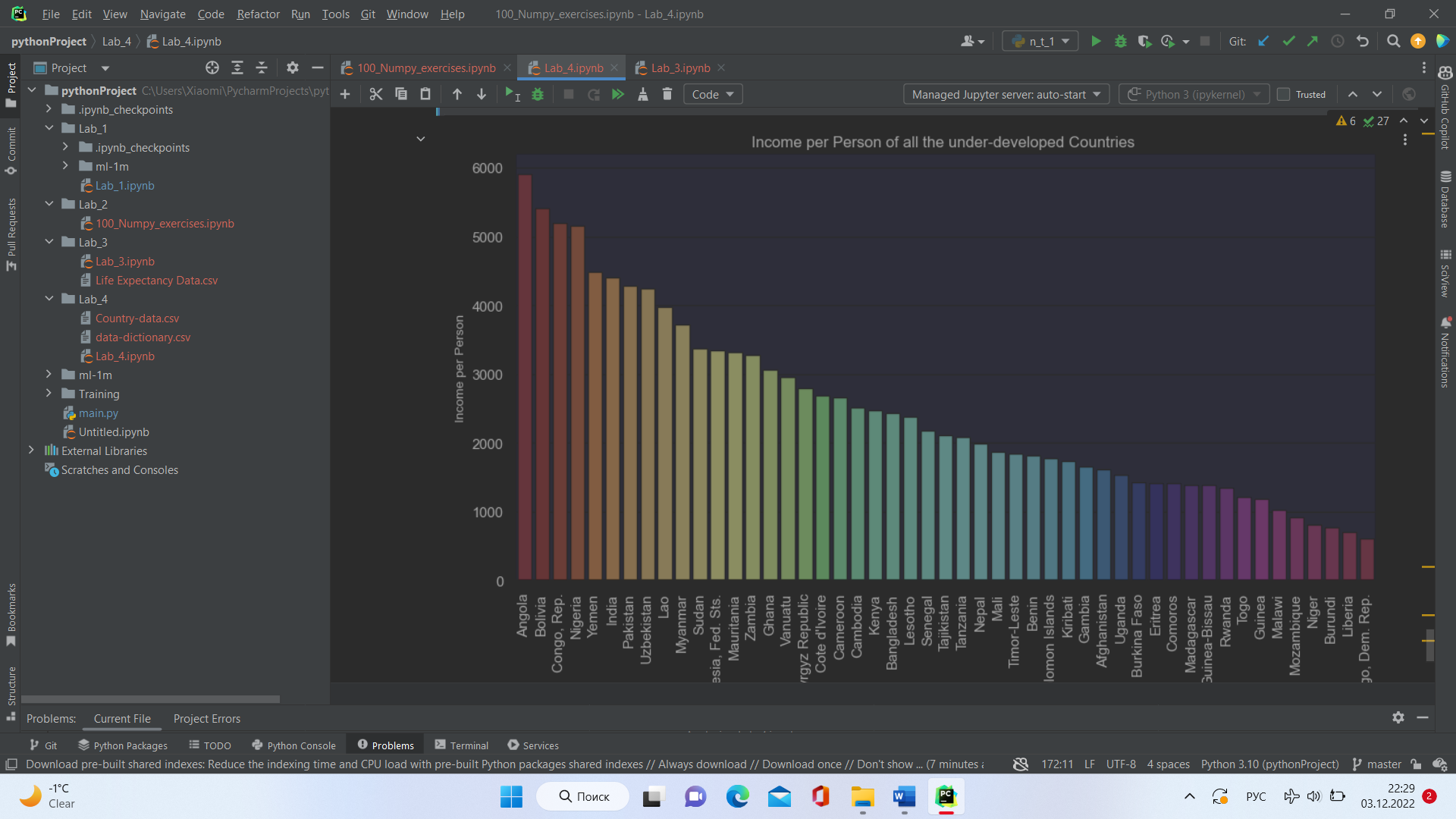
На графіку знизу показано середній дохід країни для кожного кластеру. За результатом аналізу цього графіку можна зробити висновок що середній дохід для розвинених країн(кластер 2) становить понад 25000$ на одну особу. В той час як для країн країн що розвиваються з низьким рівнем життя(кластер 0) цей показник = 2300$



Дана таблиця показує топ-8 країн з найбільшою смертністю дітей і найнижчим доходом. Всі вони належать кластеру 0 і знаходяться в Африці.



На даному графіку показані всі країни що розвиваються з низьким рівнем життя і середній дохід для їх громадян.



Після аналізу всіх даних було виявлено 6 найбідніших країн: Демократична Республіка Конго, Ліберія, Бурунді, Нігер, Мозамбік, Малаві. У всіх цих країн середній дохід на одну особу менший 1000$ в той час як у розвинених країнах він може досягати 60000$ і більше. Саме ці 6 країн на мою думку потребують допомоги.

**Висновок:** Під час виконання цієї лабораторної роботи було отримано навички з проведення кластеризації даних.