*Додаток 1*

Міністерство освіти і науки України

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського"

Факультет інформатики та обчислювальної техніки Кафедра інформатики та програмної інженерії

Звіт

з комп’ютерного практикуму №3 з дисципліни

«Аналіз даних в інформаційних системах»

на тему: «Описова статистика»

Виконав ІП-13 Ал Хадам М.Р.

(шифр, прізвище, ім'я, по батькові)

Перевірила Ліхоузова Т. А.

( прізвище, ім'я, по батькові)

Київ 2023

# **Комп’ютерний практикум 3**

Тема: описова статистика.

Мета: ознайомитись з методикою первинної обробки статистичних даних; проаналізувати вплив способу представлення даних на їх інформативність.

**Завдання**

Основне:

1. Скачати дані із файлу Data2.csv
2. Записати дані у data frame
3. Дослідити структуру даних
4. Виправити помилки в даних
5. Побудувати діаграми розмаху та гістограми
6. Додати стовпчик із щільністю населення

Додаткове:

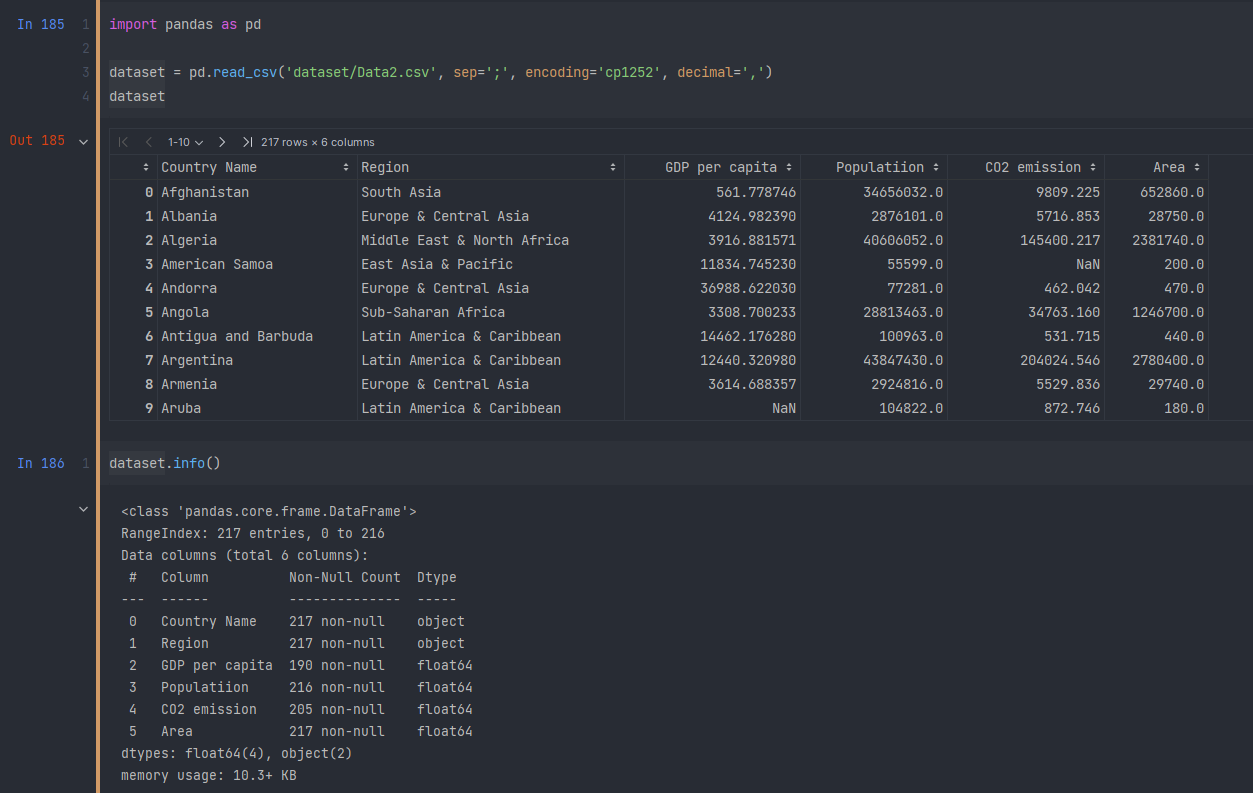
Відповісти на питання (файл Data2.csv):

1. Чи є пропущені значення? Якщо є, замінити середніми
2. Яка країна має найбільший ВВП на людину (GDP per capita)? Яка має найменшу площу?
3. В якому регіоні середня площа країни найбільша?
4. Знайдіть країну з найбільшою щільністю населення у світі? У Європі та центральній Азії?
5. Чи співпадає в якомусь регіоні середнє та медіана ВВП?
6. Вивести топ 5 країн та 5 останніх країн по ВВП та кількості СО2 на душу населення.

## **Основне завдання**

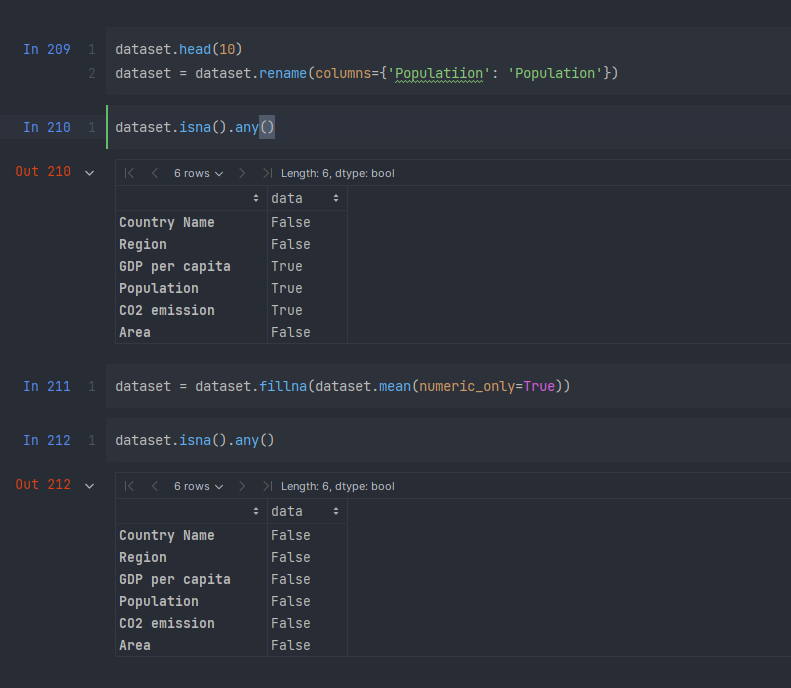
### **DataFrame та його структура**

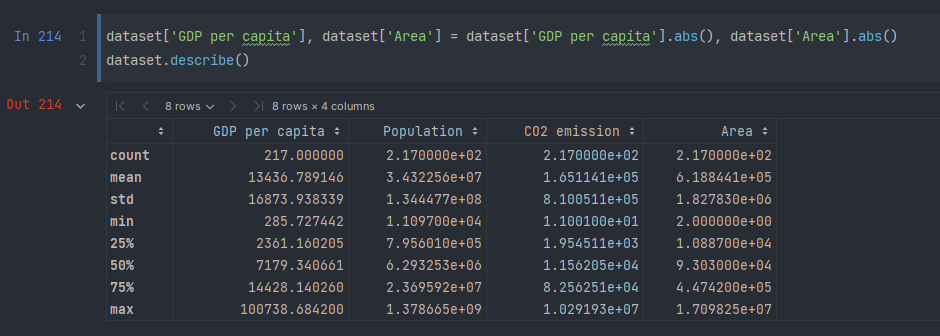
За допомогою Python бібліотеки Pandas завантажимо дані з даного csv файлу в dataframe та досліджуємо структуру даних.



### **Виправлення помилок**

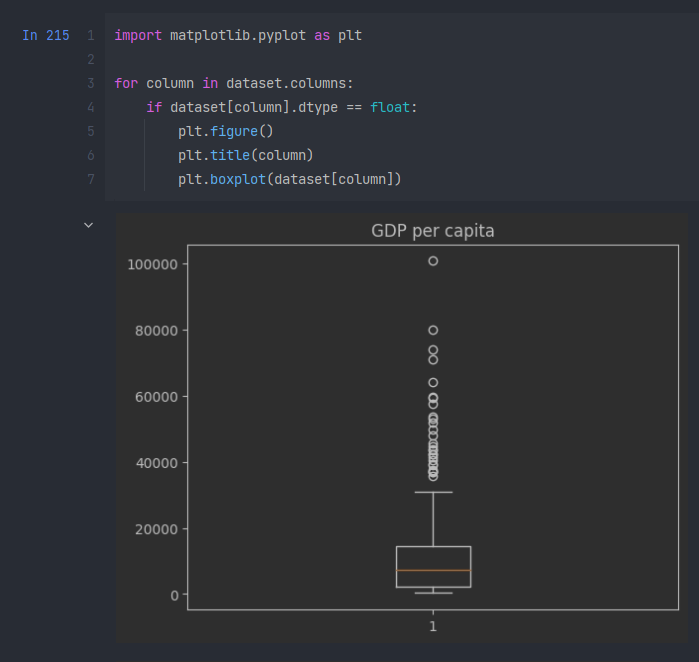
З основних помилок можемо виокремити неправильну назву одного із стовпців, відсутність числових значень в певних комірках. Замінимо відсутні значення середнім значенням по присутнім данним кожного з стовпців.

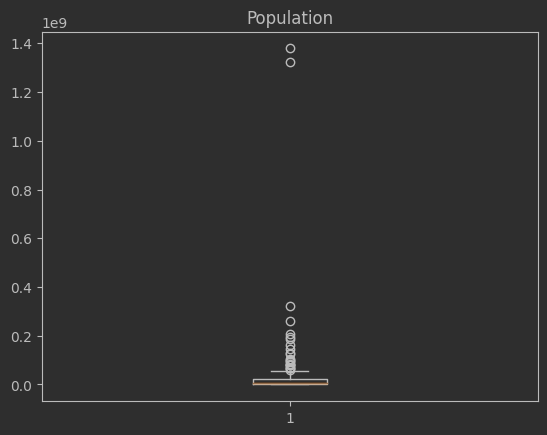




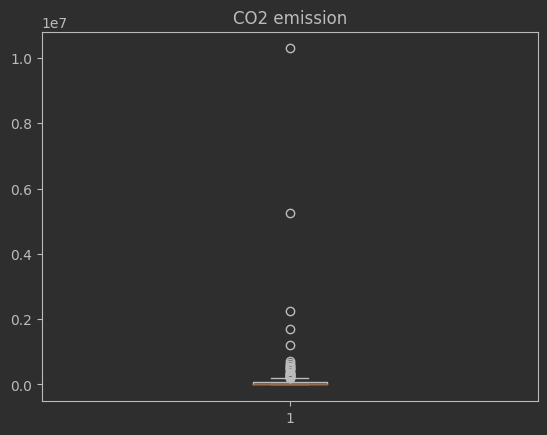
### **Діаграми розмаху та гістограми**

Побудуємо діаграми розмаху та гістограми для кожного стовпця з чисельними даними.

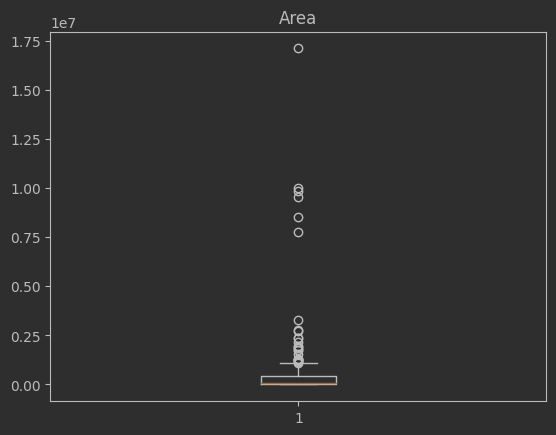


Діаграма розмаху для ВВП на душу населення

Діаграма розмаху для кількості населення

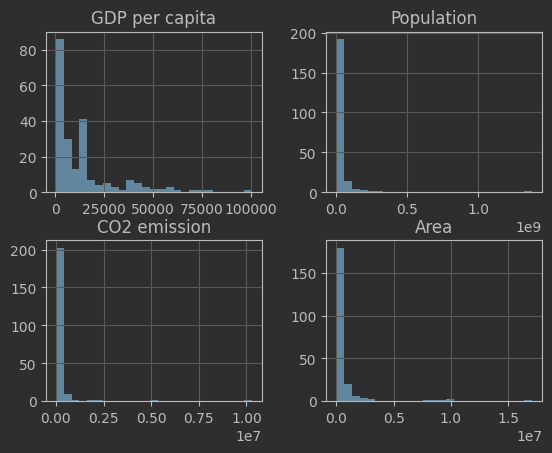


Діаграма розмаху для кількості викидів CO2



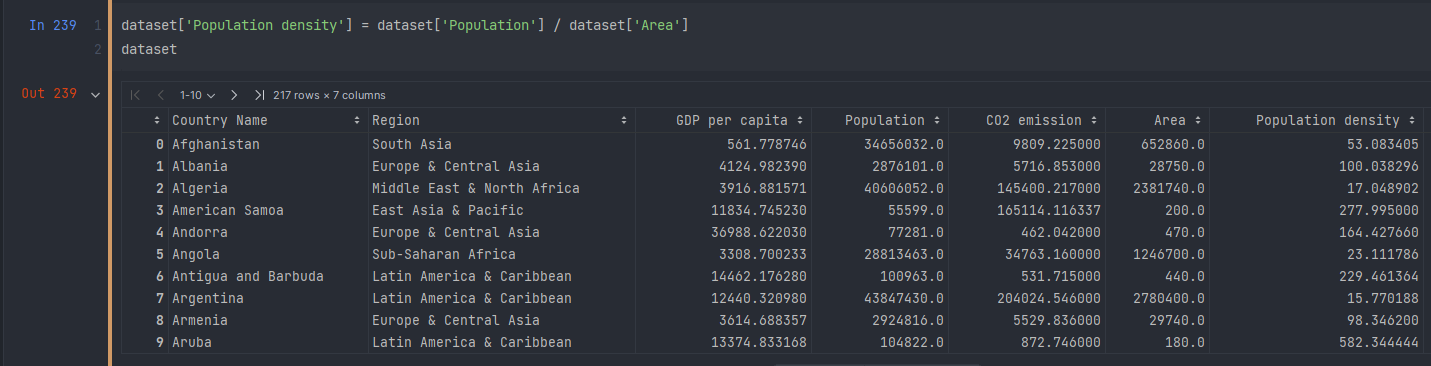
Діаграма розмаху та гістограма для площі країн

Гістограми для стовпців ВВП на душу населення, кількості населення, кількості викидів CO2, площі країн.



### **Додавання стовпчику із щільністю населення**

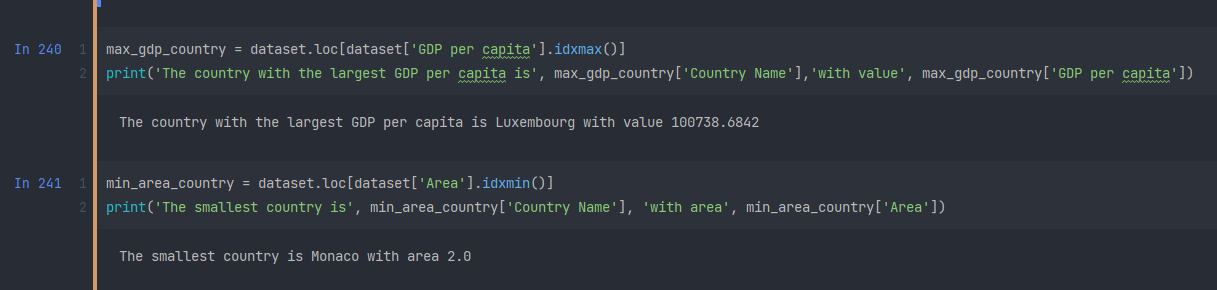
Додаємо стовпчик із щільністю населення кожної країни, який є представленням кількості населення поділеного на площу країни.



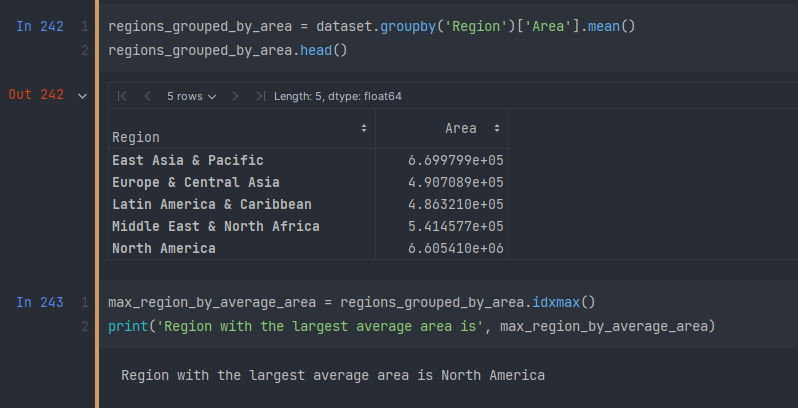
## **Додаткове завдання**

### **Країна з найбільшим ВВП на людину, з найменшою площею**

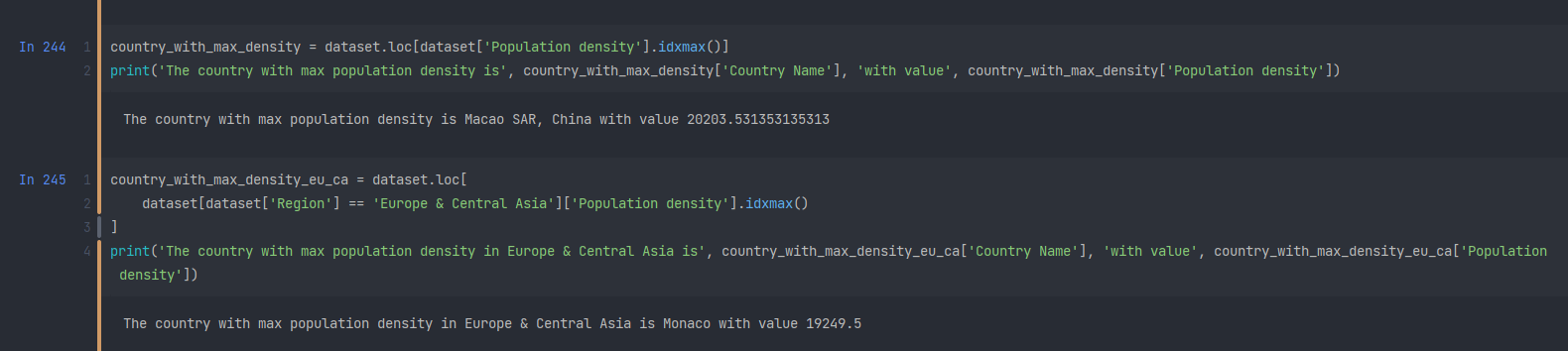
Виведемо країну з найбільшим ВВП на душу населення та країну з найменшою площею.



### **Регіон з найбільшою середньою площею країн**

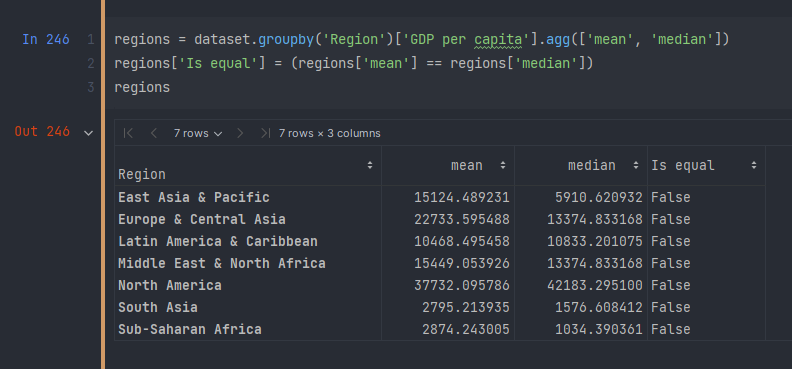


### **Країна з найбільшою щільністю населення у світі, у Європі та центральній Азії**



### **Співпадіння середнього та медіани ВВП по регіонам**

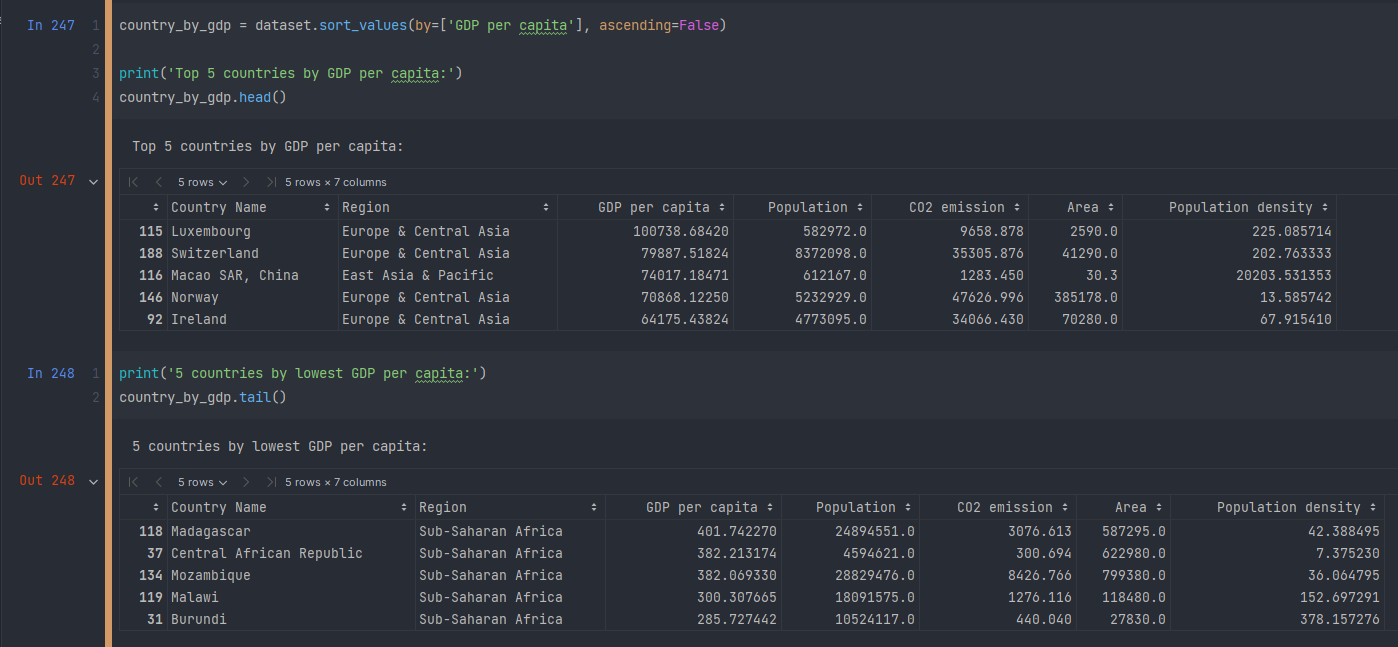
Для початку розрахуємо загальне ВВП для кожної країни та створимо окрему колонку для цих даних.



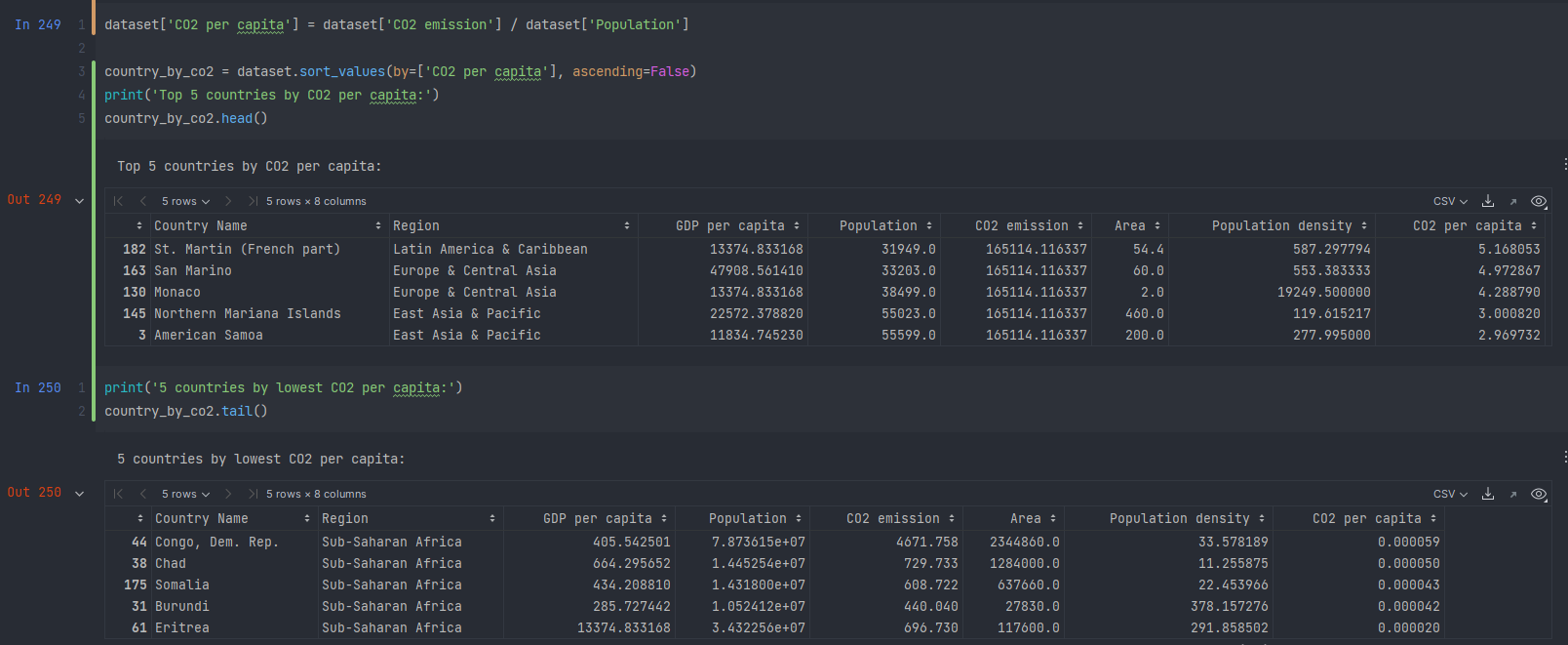
Не існує жодного регіону, де ці параметри були б рівними.

### **Топ 5 країн та 5 останніх країн по ВВП та кількості СО2 на душу населення**

Для початку розрахуємо кількість викидів СО2 на душу населення для кожної країни. Виведемо 5 країн з найбільшою кількістю ВВП на душу населення та 5 з найменшою.



Виведемо 5 країн з найбільшою кількістю викидів СО2 на душу населення та 5 з найменшою



Висновок

У цьому комп'ютерному практикумі було вивчено модуль Pandas для роботи з даними. Дані були записані в DataFrame, з виявленими помилками, тому була виконана їх очистка від від’ємних значень та заміна нульових значень на середні для більш точного аналізу. Виявлено великий розмах між даними на діаграмах розмаху, зокрема щодо кількості населення та викидів СО2. Були визначені країни з найбільшим ВВП на душу населення та з найменшою площею території, регіон з найбільшою середньою площею країн, а також країни з найбільшою густиною населення у світі та в регіоні "Європа та центральна Азія". Не було виявлено регіонів з однаковими середньою та медіаною ВВП країн. Також були визначені 5 країн з найбільшим та найменшим ВВП на душу населення та 5 з найбільшою та найменшою кількістю викидів СО2.