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

Практикум №3

з курсу «Аналіз даних в інформаційнних системах»

на тему: «Описова статистика»

|  |  |
| --- | --- |
| Викладач:  Ліхоузова Т.А. | Виконав:  студент 2 курсу  групи ІП-11 Сідак Кирил з ФІОТ |

Київ-2023

ЗМІСТ

[1. ЗАВДАННЯ 3](#_Toc131002559)

[2. ОСНОВНЕ ЗАВДАННЯ 4](#_Toc131002560)

[3. ДОДАТКОВЕ ЗАВДАННЯ 9](#_Toc131002561)

[4. ВИСНОВОК 13](#_Toc131002562)

# ЗАВДАННЯ

**Основне завдання**

Скачати дані із файлу Data2.csv

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

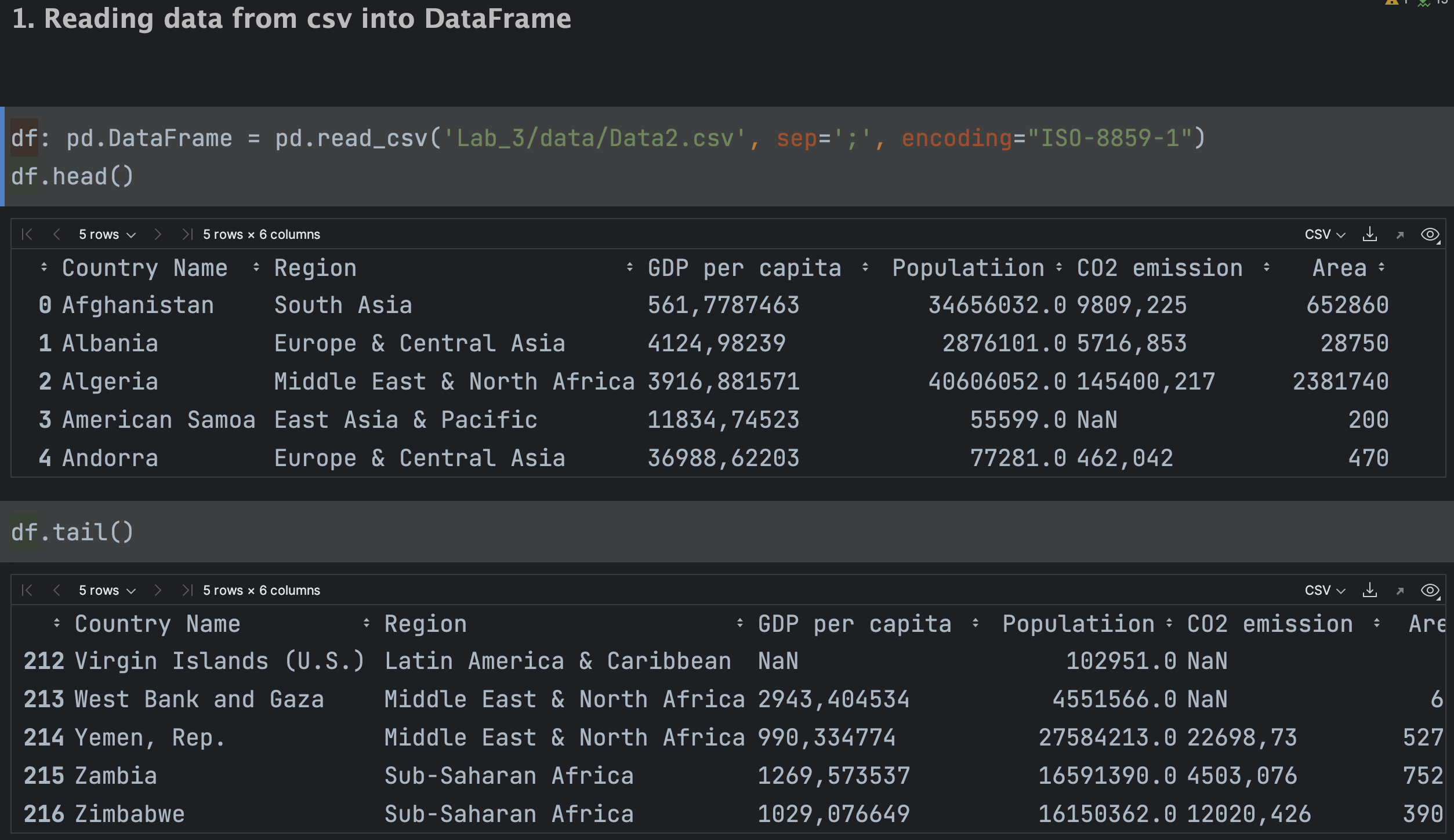
**Додаткове завдання**

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

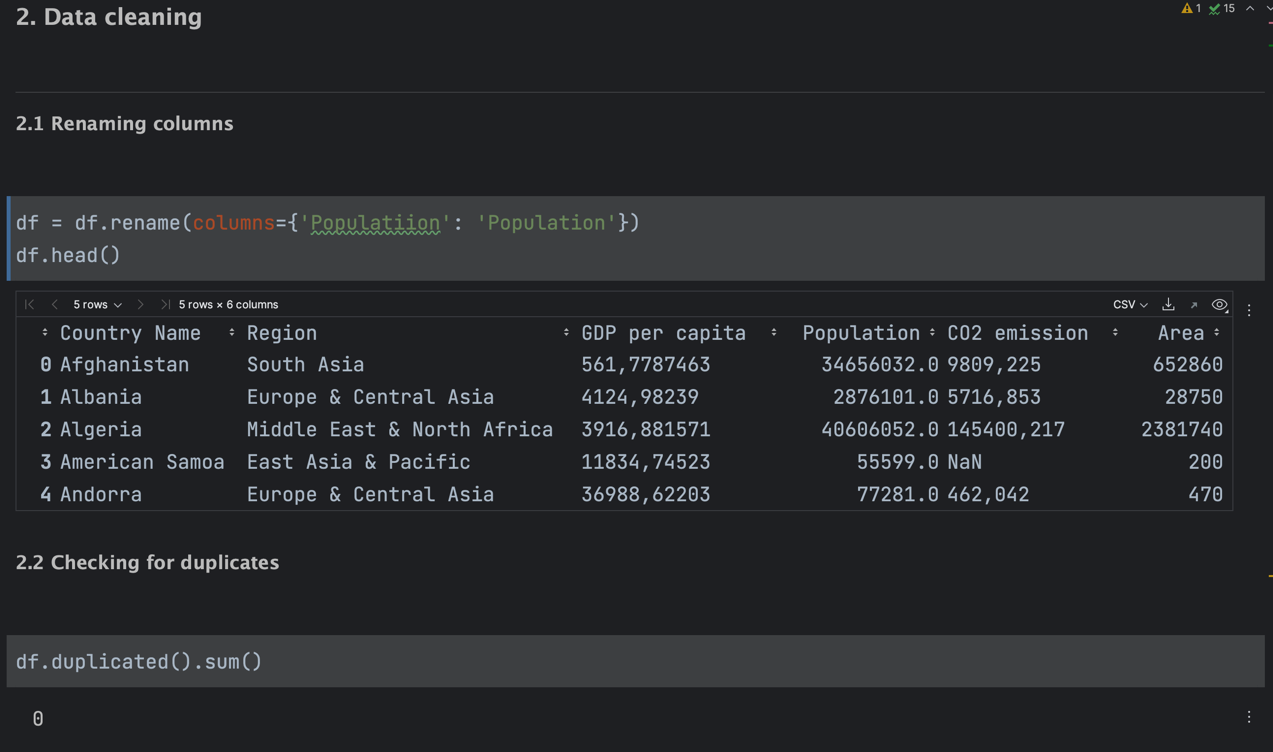
1. Записати дані у data frame

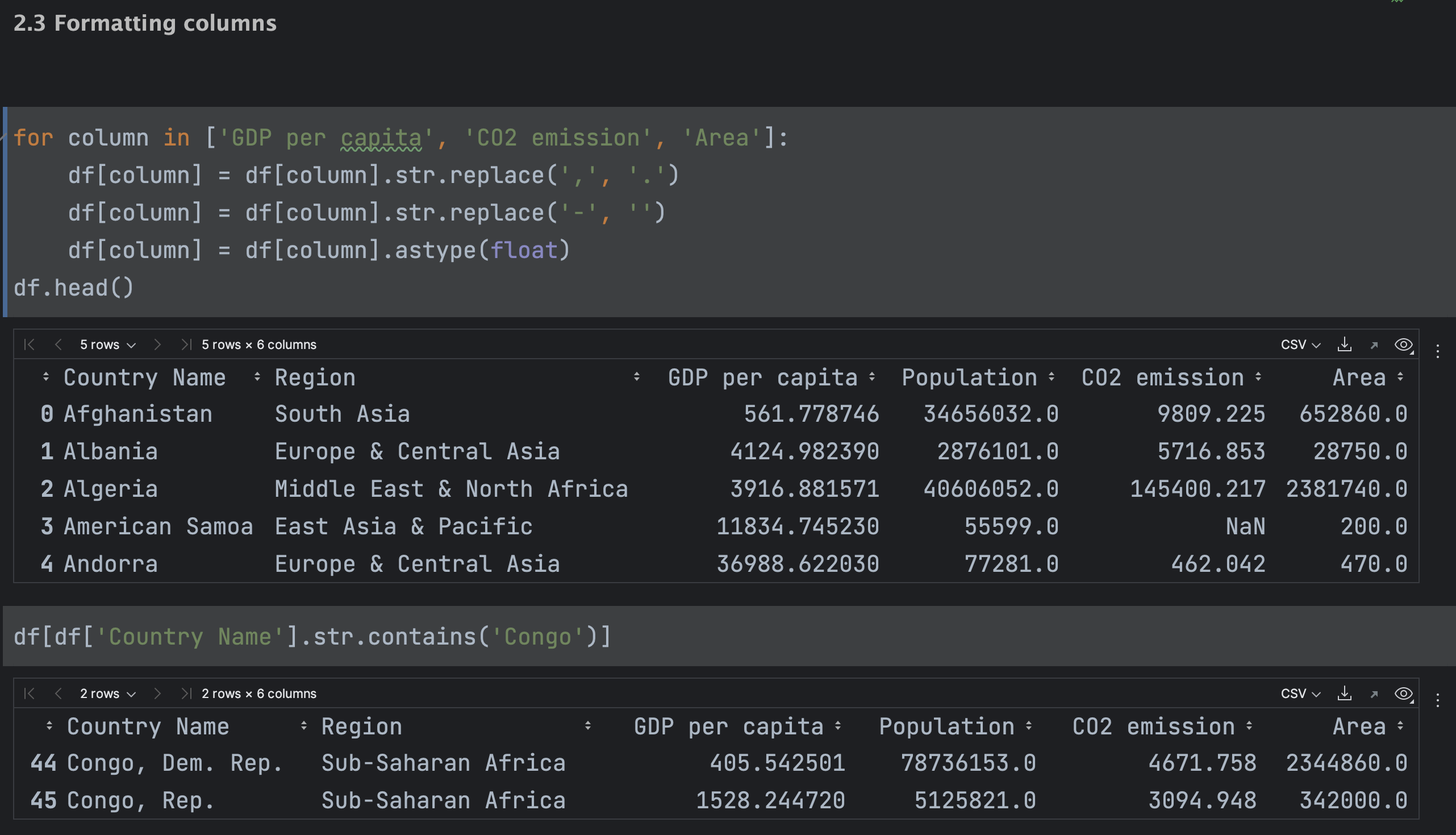


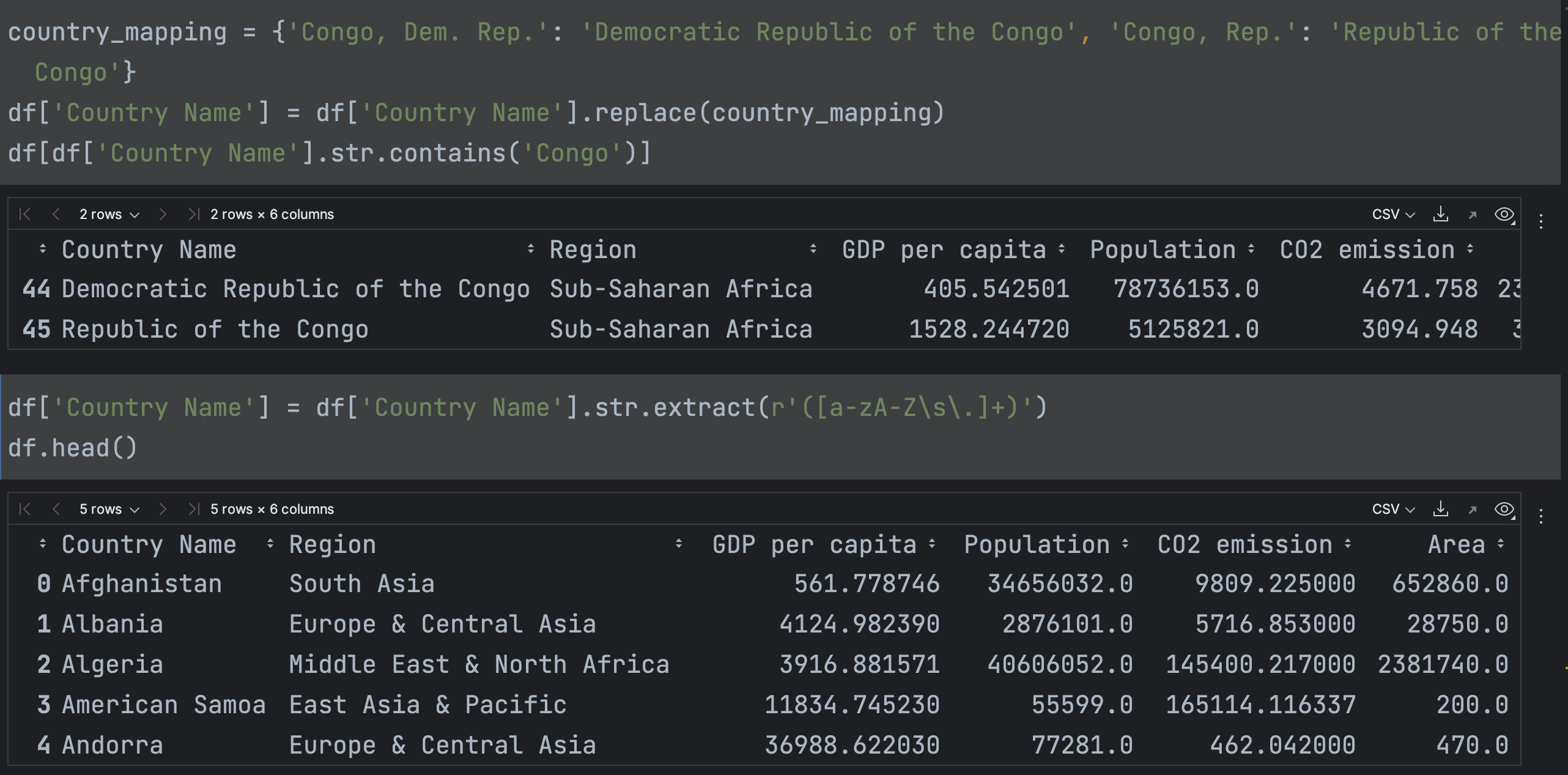
1. Дослідити структуру даних

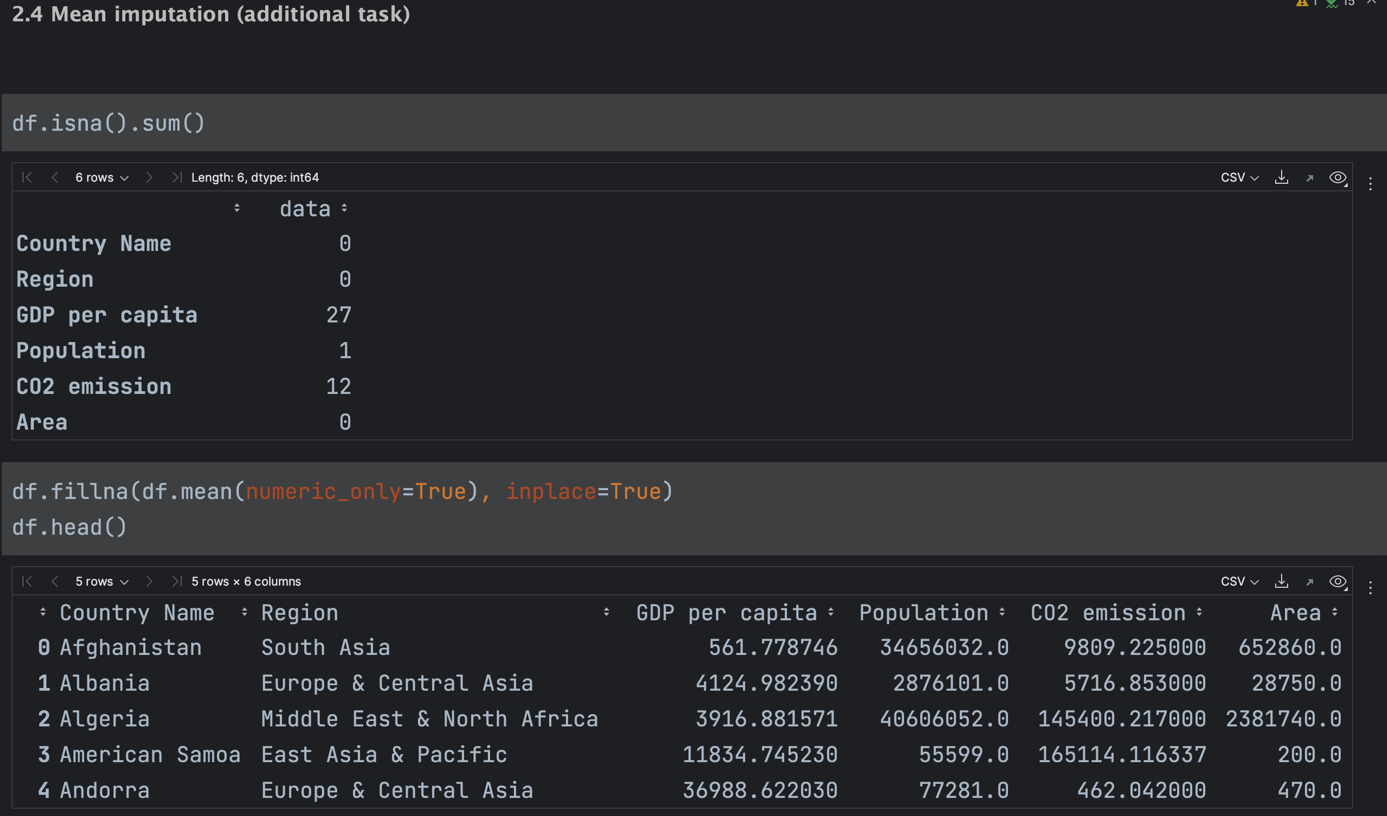
У даному датасеті присутні пропущені значення (NaN), дублікати відсутні, наявні від’ємні значення, типи даних стовпців з числовими значеннями не є числовими, замість крапки використана кома для розділення дробової частини числа, назви країн мають зайві частини.

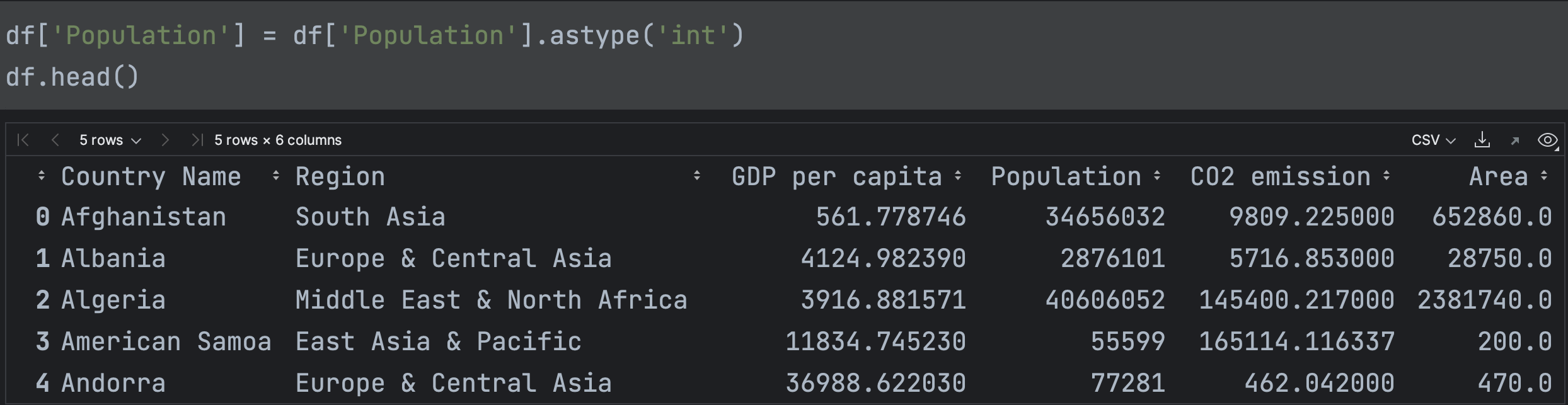
1. Виправити помилки в даних



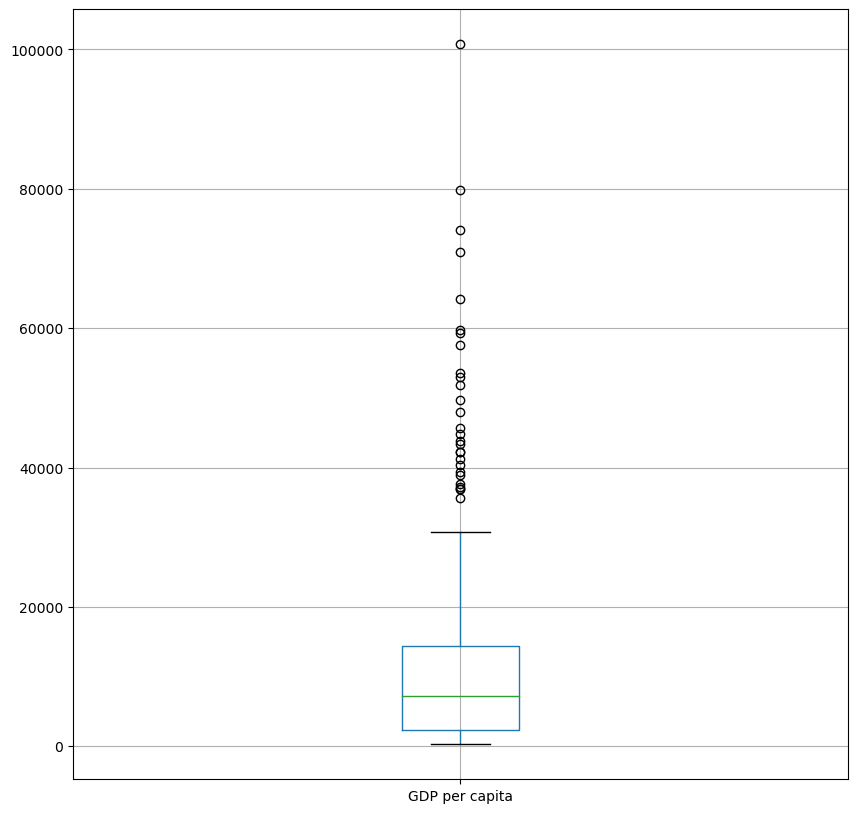








1. Побудувати діаграми розмаху та гістограми

Рисунок 4.1 – Діаграма розмаху для ВВП на душу населення

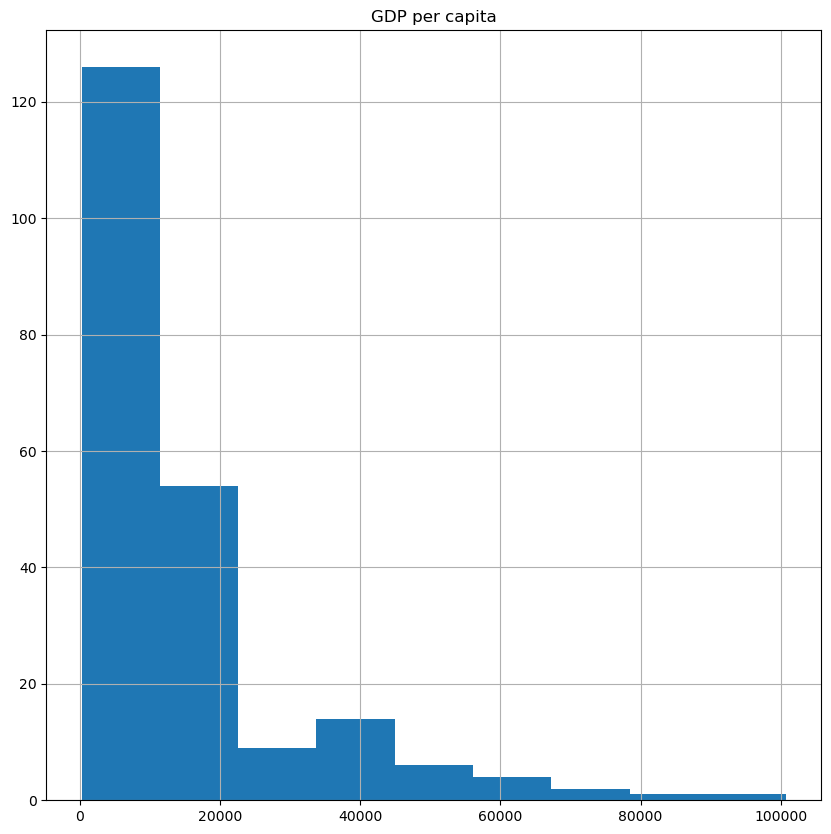
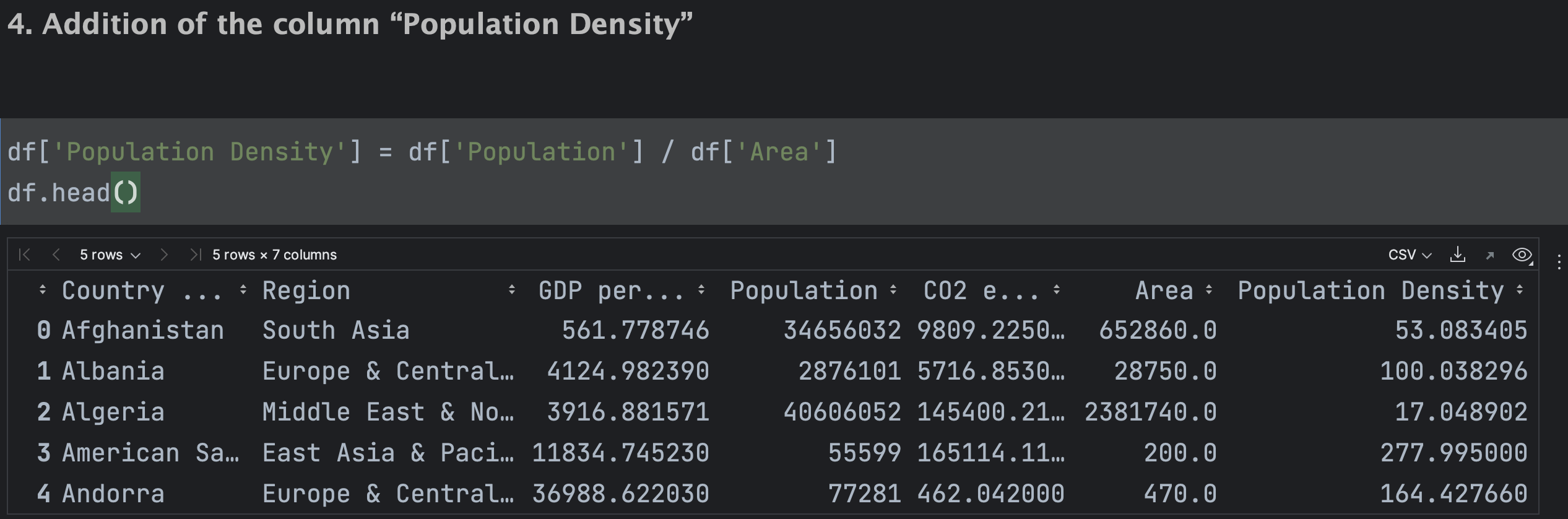


Рисунок 4.2 – Гістограма для ВВП на душу населення

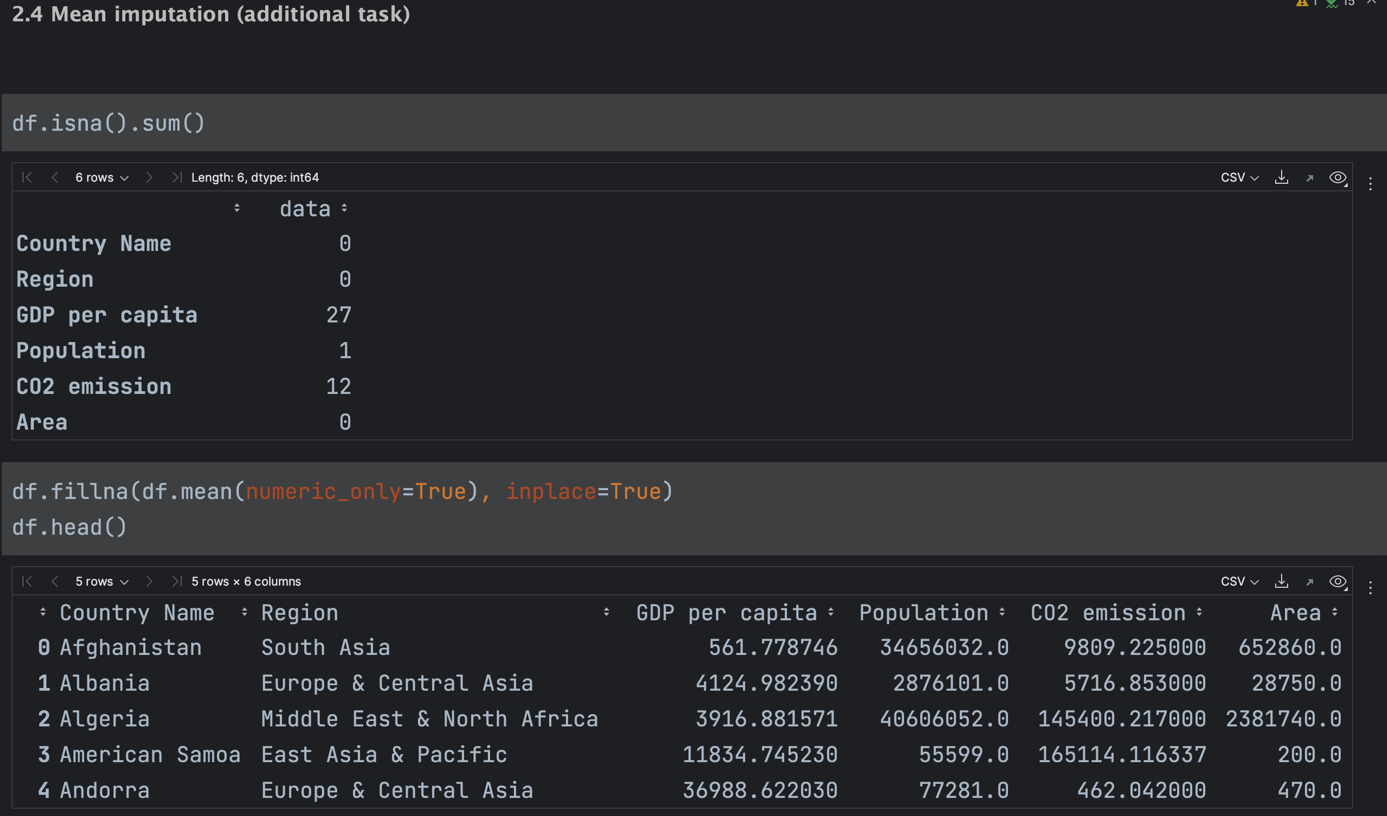
1. Додати стовпчик із щільністю населення



# ДОДАТКОВЕ ЗАВДАННЯ

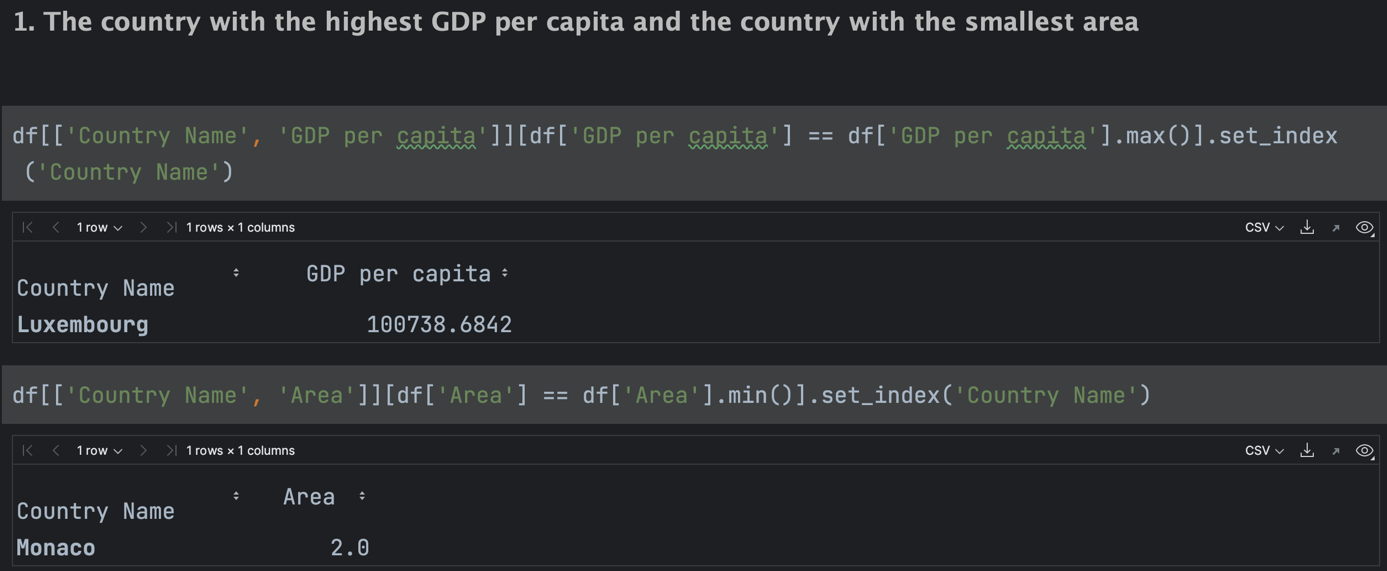
1. Чи є пропущені значення? Якщо є, замінити середніми

Пропущені значення наявні в наступних стовпцях: “GDP per capita”, “Population”, “CO2 emission”.



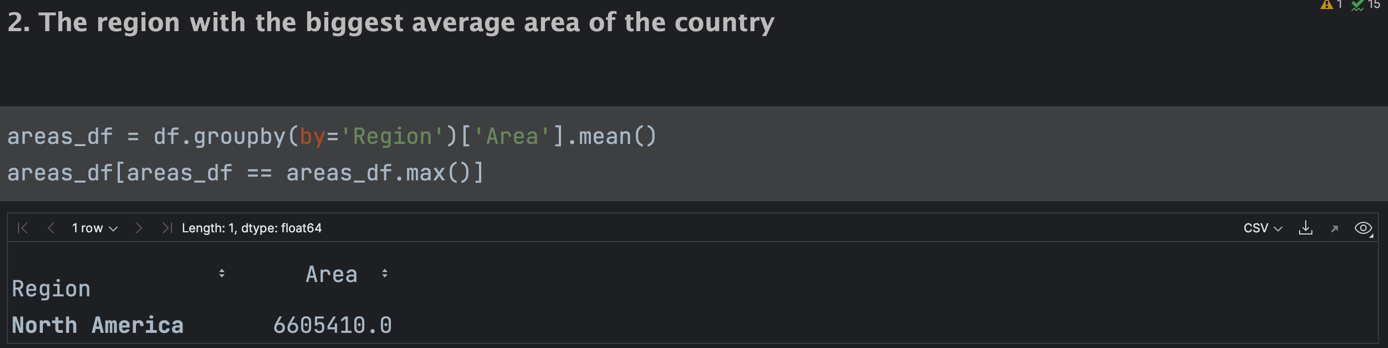
1. Яка країна має найбільший ВВП на людину (GDP per capita)? Яка має найменшу площу?

Найбільший ВВП на людину має Люксембург, а найменшу площу має Монако.



1. В якому регіоні середня площа країни найбільша?

Найбільша середня площа країни в Північній Америці.



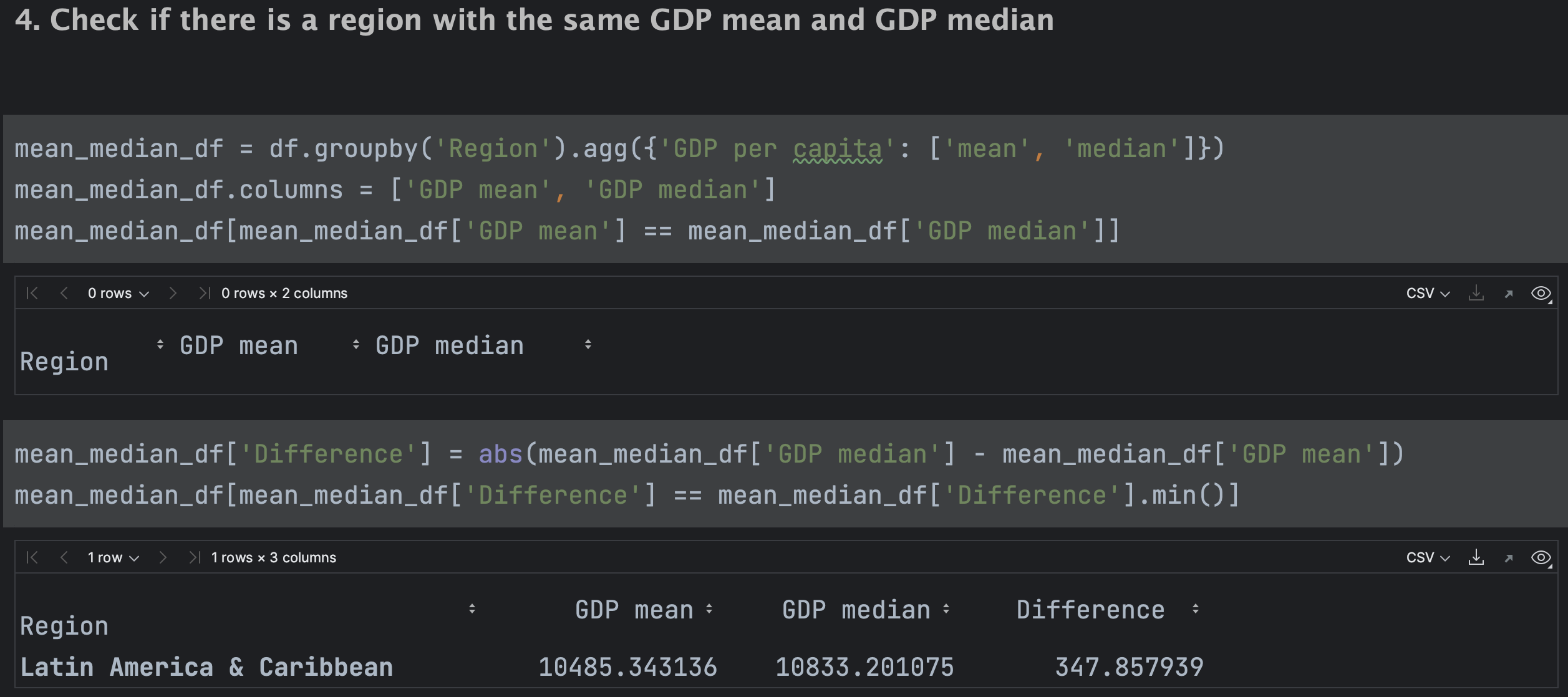
1. Знайдіть країну з найбільшою щільністю населення у світі? У Європі та центральній Азії?

Країна з найбільшою щільністю населення у світі – це Макао (особливий адміністративний район (SAR) Китаю ), а в Європі та центральній Азії – Монако.



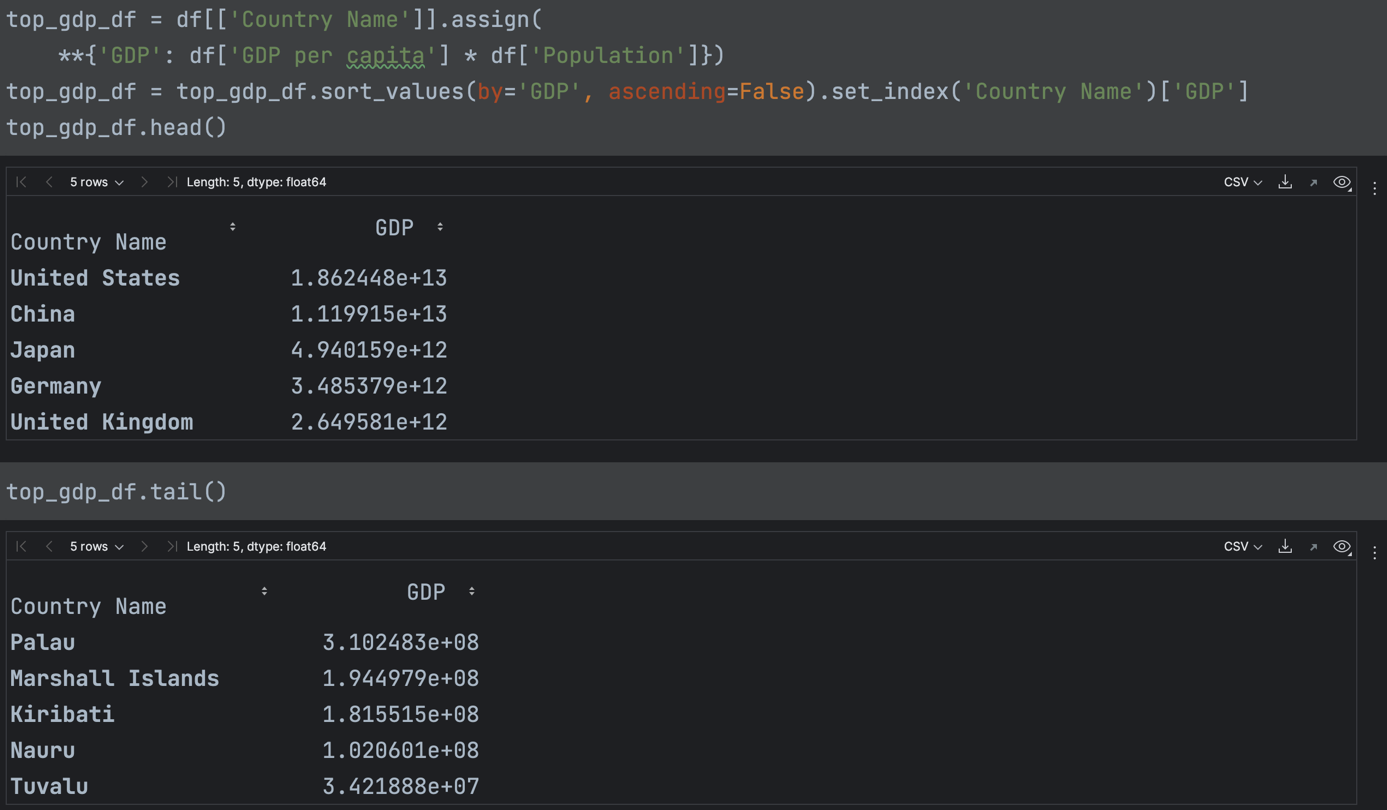
1. Чи співпадає в якомусь регіоні середнє та медіана ВВП?

У жодному регіоні середнє та медіана ВВП не співпадають. У регіоні Латинська Америка та Кариби різниця між цими значеннями є мінімальною і становить трохи більше 347.

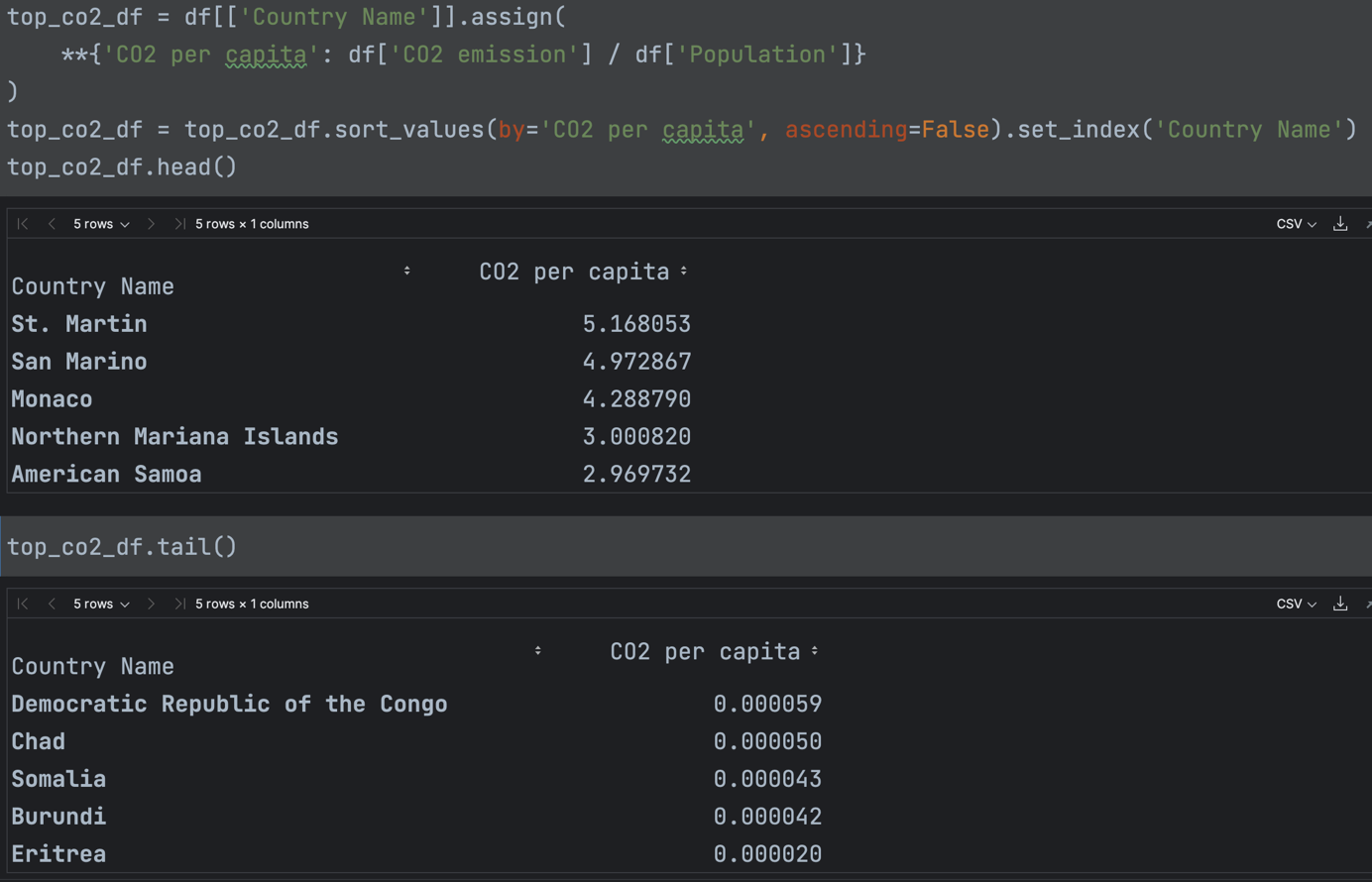


1. Вивести топ 5 країн та 5 останніх країн по ВВП та кількості СО2 на душу населення.

Топ-5 та 5 останніх країн по ВВП:



Топ-5 та 5 останніх країн по кількості СО2 на душу населення:



# ВИСНОВОК

Отже, при виконанні даної лабораторної було записано дані у DataFrame, виправлені помилки в цих даних, а саме: пропущені значення (NaN), від’ємні значення, стовпців з числовими значеннями з не числовими типами даних, крапки замість кома для розділення дробової частини числа, невідформатовані назви країн, побудовано діаграму розмаху та гістограму для числового стовпця DataFrame, додано стовпчик із щільністю населення та досліджено країни та регіони з найбільшими чи найменшими значеннями певних параметрів.