*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 студент ІП-13, Бабашев Олексій Дмитрович

(шифр, прізвище, ім'я, по батькові)

Перевірив Олійник Юрій Олександрович

( прізвище, ім'я, по батькові)

Київ 202 3

# **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:

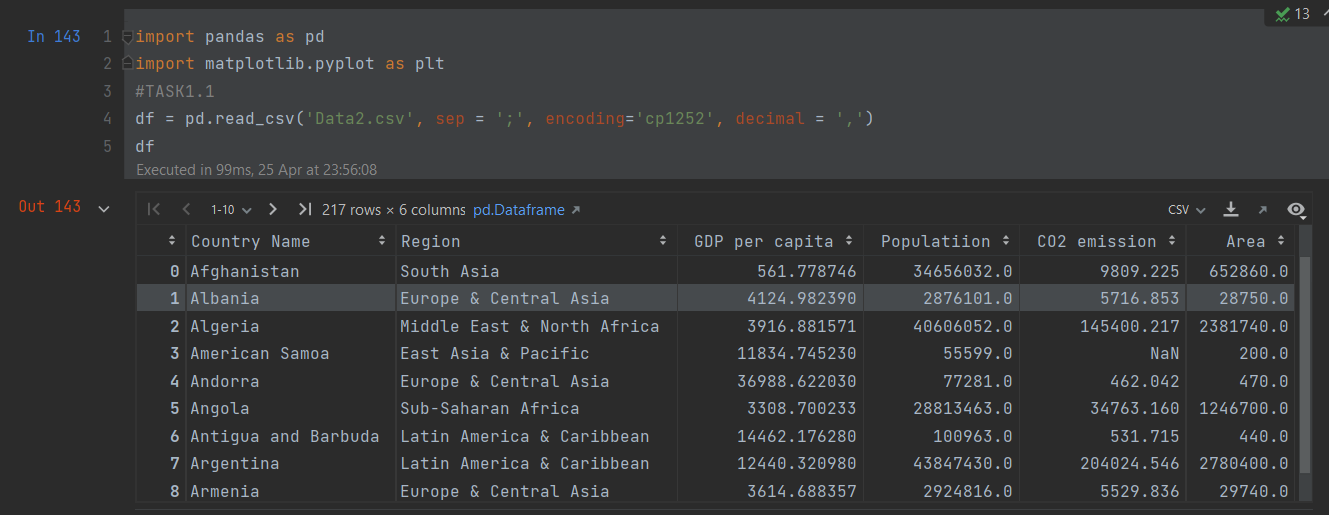
Відповісти на питання (файл Data2.csv):

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.

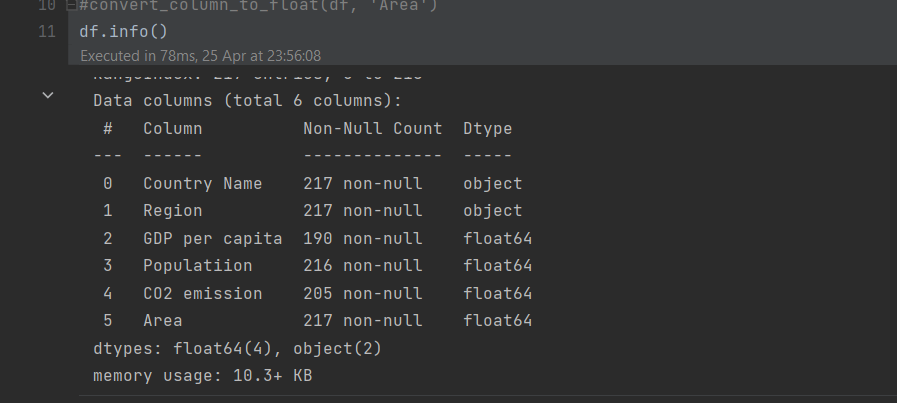
## **Основне завдання**

### **DataFrame та його структура**

За допомогою Python бібліотеки Pandas завантажимо дані з даного csv файлу в dataframe.



Досліджуємо структуру даних, використовуючи наступний код.



На рисунках вказано загальну інформацію про dataframe.

### **Виправлення помилок**

Щодо базової обробки датафрейму, змінимо назву колонки ‘Populatiion’ на ‘Population’ скриптом нижче, оскільки вона містить помилку в назві.

Graphical user interface, text

Description automatically generated

Знайдемо рядки, поля яких містять від’ємні елементи та виведемо їх.

Text

Description automatically generated

Виправимо всі існуючі від’ємні значення та виведемо їх знову, щоб перевірити внесені зміни.

Text

Description automatically generated

Також в даних наявні пусті елементи, які потрібно замінити середніми по стовпчику.

Text

Description automatically generated with low confidence

### **Діаграми розмаху та гістограми**

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

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Діаграма розмаху та гістограма для ВВП на душу населення

Chart, box and whisker chart

Description automatically generated Chart, histogram

Description automatically generated

Діаграма розмаху та гістограма для кількості населення

Chart

Description automatically generated with medium confidence Chart

Description automatically generated

Діаграма розмаху та гістограма для кількості викидів CO2

Chart, box and whisker chart

Description automatically generatedChart, histogram

Description automatically generated

Діаграма розмаху та гістограма для площі країн

Chart, box and whisker chart

Description automatically generated Chart, histogram

Description automatically generated

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

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

Graphical user interface, text, application, email

Description automatically generated

## **Додаткове завдання**

### **Заміна пропущених значень**

Демонстрація заміни пропущених значень описана в розділі «Виправлення помилок».

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

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

Text

Description automatically generated

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

Text

Description automatically generated

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

Text

Description automatically generated with low confidence

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

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

A picture containing graphical user interface

Description automatically generated

Розрахуємо середнє та медіану для кожного регіону окремо та порівняємо їх. Не існує жодного регіону, де ці параметри були б рівними.

Text

Description automatically generated

### **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

Graphical user interface, application

Description automatically generated

Виведемо 5 країн з найбільшою кількістю ВВП на душу населення та 5 з найменшою

Text

Description automatically generated

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

Graphical user interface, text, application

Description automatically generated with medium confidence

Висновок

У цьому комп’ютерному практикуму було вивчено можливості Python, а саме Pandas у роботі з даними. Вхідні дані було записано в DataFrame, структуру якого було вивчено та помічено нецілісність даних, тому я почистив дані від від’ємних значень, нульові замінив середніми для більш об’єктивної побудови гістограм та діаграм розмаху. На діаграмах розмаху було помічено великий розмах між даними. Наприклад, на діаграмі населення є дві країни з кількістю населення значно більшою за всі інші, так само і з викидами СО2, дані з ВВП на душу населення є найбільш кучними. Було визначено країну з найбільшим ВВП на душу населення у світі, з найменшою площею території, регіон з найбільшою середньою площею країн, країни з найбільшою густиною населення у світі та окремо в регіоні «Європа та центральна Азія». Регіонів з однаковими середньою та медіаною ВВП країн не виявилось, усі мають різні. Також було виведено 5 країн з найбільшим та найменшим ВВП на душу населення та 5 з найбільшою та найменшою кількістю викидів СО2.