**Міністерство освіти і науки України**

**Національний технічний університет України «КПІ» імені Ігоря Сікорського**

**Кафедра обчислювальної техніки ФІОТ**

**ЗВІТ**

**з лабораторної роботи №5**

**з навчальної дисципліни «Вступ до технології Data Science»**

**Тема:**

**ДОСЛІДЖЕННЯ ТЕХНОЛОГІЙ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ**

**Виконав:**

Студент X курсу кафедри ОТ ФІОТ,

Навчальної групи ІА-73

Петренко П.П.

**Перевірив:**

Професор кафедри ОТ ФІОТ

Писарчук О.О.

**Київ 2023**

**І. Мета:**

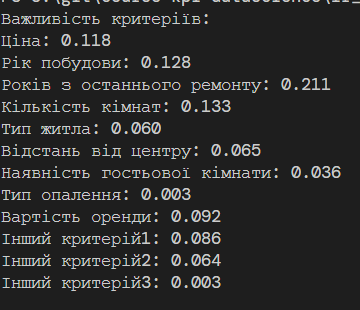
виявити дослідити та узагальнити особливості інтелектуального аналізу даних з використанням спеціалізованих пакетів мови програмування Python.

**ІІ. Завдання:**

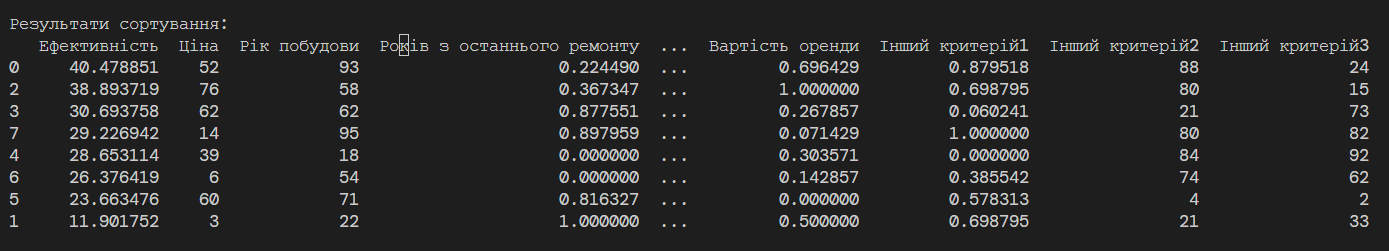
| 3, 18 | Розробити програмний скрипт, що реалізує оцінювання ефективності вибору житла. Вихідні дані містять 12 критеріїв, з яких 4 – максимізованих, решта – мінімізованих. Кількість аналогічних товарів – 8. Вхідні дані занесені у файл. Формат файлу, тип товару, зміст показників та критеріїв ефективності та їх значення обрати самостійно.  ***З використанням OLAP технологій*** провести інтелектуальний аналіз результатів розв’язку багатокритеріальної задачі, довести їх адекватність та сформувати аналітичний звіт за результатами досліджень. |
| --- | --- |

У цій лабораторній роботі я не використав OLAP з тих міркувань, що:

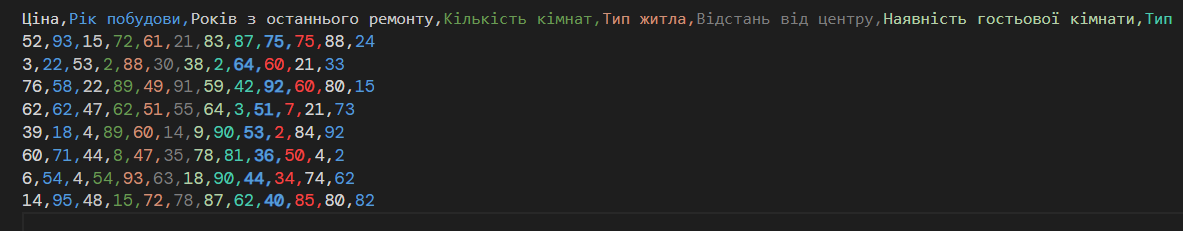
1. Це застарілий пакет 2012 року який давно не підтримується
2. Він не використовуються, а найкращою альтернативою є Pandas та Numpy
3. Для виконання цього завдання не потрібен такий інструментар, який має OLAP.

****

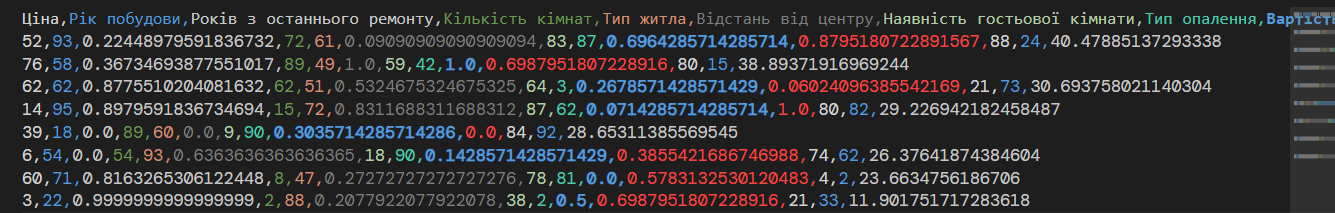
**Рисунок 1. Важливість критеріїв**

****

**Рисунок 2. Відсортований датасет за ефективністю купівлі житла**

****

**Рисунок 3. датасет**

****

**Рисунок 4. Нормалізований датасет**

**ІІІ. Довести адекватність сформованих моделей та працездатність розробленого скріпта.**

**Розроблений код повинен бути раціональним та відповідати вимогам до чистого коду.**

**import pandas as pd**

**import numpy as np**

**from sklearn.preprocessing import MinMaxScaler**

**np.random.seed(42)**

**num\_samples = 8**

**num\_criteria = 12**

**criteria\_names = ['Ціна', 'Рік побудови', 'Років з останнього ремонту', 'Кількість кімнат',**

**'Тип житла', 'Відстань від центру', 'Наявність гостьової кімнати',**

**'Тип опалення', 'Вартість оренди', 'Інший критерій1', 'Інший критерій2', 'Інший критерій3']**

**data = np.random.randint(1, 100, size=(num\_samples, num\_criteria))**

**df = pd.DataFrame(data, columns=criteria\_names)**

**df.to\_csv('variant3.csv', index=False)**

**weights = np.random.rand(num\_criteria)**

**weights /= weights.sum()**

**print("Важливість критеріїв:")**

**for i, name in enumerate(criteria\_names):**

**print(f"{name}: {weights[i]:.3f}")**

**minimize\_columns = ['Років з останнього ремонту', 'Відстань від центру', 'Вартість оренди', 'Інший критерій1']**

**scaler = MinMaxScaler()**

**df[minimize\_columns] = scaler.fit\_transform(df[minimize\_columns])**

**df['Ефективність'] = np.dot(df.values, weights)**

**df = df.sort\_values(by='Ефективність', ascending=False)**

**print("\nРезультати сортування:")**

**print(df[['Ефективність'] + criteria\_names])**

**df.to\_csv('variant3sorted.csv', index=False)**

**III. Висновок:**

виявив дослідив та узагальнив особливості інтелектуального аналізу даних з використанням спеціалізованих пакетів мови програмування Python.

Виконав: студент Петренко П.П.