

Zadania – Dane Strukturalne, Niestukturalne i Pół-Strukturalne

ZADANIE 1: Dane STRUKTURALNE — analiza sprzedaży (CSV)

Cel: Przećwiczenie pracy z danymi strukturalnymi: wczytywanie pliku CSV, tworzenie kolumn pochodnych, filtrowanie oraz agregacje.

Dane wejściowe (sales.csv):

Plik zawiera dane o sprzedaży produktów:

id, product, category, price, quantity

1, Keyboard, Electronics, 120, 3

2, Mouse, Electronics, 60, 5

3, Coffee Mug, Kitchen, 25, 10

4, Notebook, Office, 15, 20

5, Headphones, Electronics, 300, 2

6, Pen, Office, 5, 100

7, Blender, Kitchen, 200, 1

Zadania:

1. Wczytaj dane z pliku sales.csv jako DataFrame.
2. Dodaj nową kolumnę total_sale = price * quantity.
3. Oblicz łączną wartość sprzedaży dla każdej kategorii.
4. Znajdź najdroższy produkt w każdej kategorii.
5. Oblicz sumę sprzedaży wszystkich produktów.
6. (Opcjonalnie) Powtórz zadania w SQL.

ZADANIE 2: Dane NIESTRUKTURALNE — analiza logów serwera (TXT)

Cel: Nauka pracy z danymi niestukturalnymi: parsowanie tekstu, wyrażenia regularne, analiza logów.

Dane wejściowe (server.log):

[2025-01-01 10:00:00] INFO User login successful

[2025-01-01 10:00:01] ERROR Invalid password attempt

[2025-01-01 10:01:15] INFO File uploaded: report.pdf

[2025-01-01 10:02:30] WARN Memory usage high: 85%

[2025-01-01 10:05:47] ERROR Database connection timeout

[2025-01-01 10:06:10] INFO User logout

Zadania:

1. Wczytaj logi jako DataFrame.
2. Wyodrębnij timestamp, level oraz message za pomocą wyrażeń regularnych.
3. Policz wystąpienia poziomów logów (INFO/WARN/ERROR).
4. Wyświetl tylko logi typu ERROR.
5. Znajdź najczęściej pojawiające się słowa w komunikatach.

ZADANIE 3: Dane PÓŁ-STRUKTURALNE — analiza JSON

Cel: Praca z danymi pół-strukturalnymi (JSON), eksplozja tablic, praca na zagnieżdżonych polach.

Przykładowy rekord JSON:

```
{
  "customer_id": "C001",
  "name": "Jan Kowalski",
  "age": 34,
  "orders": [
    {"order_id": "O100", "value": 120, "items": 3},
    {"order_id": "O101", "value": 85, "items": 2}
  ]
}
```

Zadania:

1. Wczytaj dane JSON jako DataFrame.
2. Rozwiń tablicę orders tak, aby każdy order był osobnym rekordem.
3. Oblicz łączną wartość zamówień każdego klienta.
4. Znajdź klienta z największą liczbą zamówień.
5. Policz średnią wartość pojedynczego zamówienia.