

## Zadanie 1 — Pandas

Temat: Analiza sprzedaży produktów w sklepie internetowym.

Masz plik sales.csv z kolumnami:

order\_id, customer, category, amount, order\_date

1. Wczytaj dane do DataFrame.
2. Wyświetl tylko zamówienia z kategorii "electronics".
3. Dodaj kolumnę amount\_with\_vat (23% VAT).
4. Policz:
  - łączną wartość sprzedaży dla każdego klienta,
  - liczbę zamówień w każdej kategorii.
5. Posortuj klientów wg największej wartości sprzedaży i pokaż TOP 3.

Wskazówka: użyj groupby, agg, sort\_values.

## Zadanie 2 — NumPy

Temat: Analiza czasów biegu na 10 km.

1. Wygeneruj tablicę NumPy z 50 losowymi czasami biegu 10 km (wartości w minutach), przyjmując:
  - średnia = 50 minut,
  - odchylenie standardowe = 5 minut.
2. Oblicz:
  - średnią, medianę, min i max,
  - czasy znormalizowane (średnia = 0, odchylenie = 1),
  - indeksy 5 najlepszych czasów.
3. Wydrukuj najlepsze 5 czasów w minutach.

Wskazówka: użyj np.random.normal, np.mean, np.std, np.argsort.

## Zadanie 3 — Pandas + NumPy + Matplotlib

Temat: Analiza wyników zawodów biegowych z wizualizacją.

1. Stwórz DataFrame z danymi 10 zawodników, zawierającymi:
  - name (imię),
  - race ("10K" lub "Half"),
  - time\_min (losowo wygenerowane czasy w minutach — użyj numpy),
  - date (daty w formacie datetime).
2. Oblicz:
  - średni czas w każdej kategorii biegu (groupby),
  - najlepszy czas każdego zawodnika.
3. Zrób wykres słupkowy średnich czasów (matplotlib).

4. Dodatkowo: oznacz na wykresie kolorem słupki, w których średnia jest  $< 55$  minut dla "10K" i  $< 125$  minut dla "Half".

Wskazówka: kolorowanie możesz zrobić przez list comprehension na podstawie warunku.