KARTA PRACY UCZESTNIKA

Moduł 4 Wizualizacja danych i narracja analityczna

Sześć wykresów z uzasadnieniem i kontrolą jakości w Jupyter Notebook

Czas	Pracujesz z	Rezultat
70–90 min orientacyjnie	wynikami M3, Pandas i Matplotlib	sześcioma PNG, mapą korelacji i tabelą narracji

Cel i rezultat modułu

W M3 uzyskałeś odpowiedzi liczbowe. Teraz pokażesz je odbiorcy na wykresach i sprawdzisz, czy forma oraz opis wiernie przedstawiają dane. Przy każdym wykresie uzasadnisz wybór, porównasz go z tabelą i zapiszesz ograniczenie interpretacji.

Zanim zaczniesz

Kontynuuj projekt eksploatacja_urzadzen. Utwórz notebooks/M4_wizualizacja.ipynb, który działa samodzielnie po restarcie kernela. Zachowaj notebook M3, jego wyniki i dane processed bez zmian. Wyniki obliczaj kodem; nie przepisuj liczb ze slajdów.

Źródło w projekcie	Do czego służy
data/processed/pomiary_clean.csv	Zweryfikowane dane M2, odczytywane również w M3.
notebooks/M3_EDA.ipynb i outputs/tables	Pytania M3-P01–M3-P06, wyniki, obserwacje i ograniczenia.
docs/slownik_danych.md i README.md	Znaczenie pól, jednostki, typy oraz identyfikacja wersji danych.

Sześć wykresów w Twoim notebooku

ID i rozdział	Co przygotujesz i z czym powiążesz
M4-W01 · R2	Ranking median drgań wszystkich 8 urządzeń; M3-P03.
M4-W02 · R3	Linię dziennych median energii z pasem Q1–Q3; dzienne rozwinięcie tematu M3-P05.
M4-W03 · R3	Słupki grupowane median energii dla okresów i lokalizacji; M3-P05.
M4-W04 · R4	Histogram drgań pełnego zbioru; M3-P01.
M4-W05 · R4	Poziome wykresy pudełkowe drgań według urządzenia; M3-P01 i M3-P03.
M4-W06 · R5	Wykres punktowy temperatury i drgań z kompletnych par; M3-P06.

Workflow pracy krok po kroku

Krok 1 Skonfiguruj niezależny notebook

Uruchom notebook z katalogu notebooks i sprawdź Path.cwd(). Kod zakłada właśnie ten katalog roboczy. Jeśli plik nie istnieje, popraw punkt startu lub sprawdź wynik M2. Nie wpisuj ścieżki z nazwą dysku lub własnego konta.

```
from pathlib import Path
from hashlib import sha256
import pandas as pd
import numpy as np

print("Katalog roboczy:", Path.cwd())
PROJECT_DIR = Path.cwd().parent
TABLES_DIR = PROJECT_DIR / "outputs" / "tables"
FIGURES_DIR = PROJECT_DIR / "outputs" / "figures"
plik = PROJECT_DIR / "data" / "processed" / "pomiary_clean.csv"
assert plik.is_file(), "Sprawdź katalog roboczy i wynik M2"
hash_wejscia = sha256(plik.read_bytes()).hexdigest()
pomiary_clean = pd.read_csv(
    plik, encoding="utf-8", parse_dates=["czas", "data_instalacji"]
)
typy = dict.fromkeys([
    "rok_miesiac", "urzadzenie", "typ_urzadzenia", "lokalizacja"
], "string")
typy.update(dict.fromkeys([
    "temperatura", "drgania", "energia", "moc_nominalna_kw"
], "float64"))
typy.update({"produkcja": "Int64", "wiek_urzadzenia_dni": "Int64",
             "tryb": "category", "status": "category"})
pomiary_clean = pomiary_clean.astype(typy)
TABLES_DIR.mkdir(parents=True, exist_ok=True)
FIGURES_DIR.mkdir(parents=True, exist_ok=True)
```

W tej samej sekcji dodaj import matplotlib.pyplot as plt. Zapisz wersje Pythona, Pandas, NumPy i Matplotlib. Odtworzenie typów po odczycie CSV jest kontrolą importu; nie oznacza ponownego czyszczenia danych.

Krok 2 Potwierdź wejście przed rysowaniem

- Porównaj wszystkie 14 nazw kolumn i ich kolejność ze słownikiem M2. Sprawdź typy, kategorie oraz zgodność rok_miesiac z czasem.
- Dane kursowe mają 720 rekordów, 8 urządzeń, po 90 pomiarów na urządzenie oraz kompletny i unikalny klucz czas + urzadzenie.
- Potwierdź zakres 1 stycznia–14 lutego 2026, 45 dni, dwa braki energii i jeden brak produkcji. Temperatura i drgania nie mają braków.
- Sprawdź SHA-256 względem zapisu M2 w README.md. Rozbieżność wyjaśnij u źródła. Nie poprawiaj processed ani wyników M3, aby uzyskać oczekiwany wykres.

Krok 3 Zapisz mapę wizualizacji w rozdziale 1

Dla każdej pozycji W01–W06 zapisz odbiorcę, pytanie, wynik źródłowy, zmienne z jednostkami, poziom agregacji, typ wykresu, odrzuconą alternatywę i powód wyboru. Dodaj obserwację i ograniczenie. W02 oznacz jako nowe dzienne przeliczenie tego samego wejścia, a nie eksport istniejący już w M3.

Pytanie do namysłu: co odbiorca mógłby błędnie uznać za trend, alarm albo przyczynę? Zapisz po jednym takim ryzyku przy wybranych wykresach i wróć do niego podczas końcowego audytu.

Ranking i zmiana dzienna

Krok 4 Zbuduj ranking w rozdziale 2

Dla M4-W01 odtwórz groupby po urzadzenie, count oraz medianę drgania z pełnej ramki pomiary_clean. Porównaj wynik z M3-P03. Zachowaj wszystkie 8 urządzeń, uporządkuj mediany malejąco i wyróżnij pierwszą trójkę.

- Użyj fig, ax = plt.subplots(...), następnie ax.barh(...). Wywołaj ax.invert_yaxis(), aby pierwszy wiersz rankingu był u góry.
- Oś wartości rozpocznij od zera. Podaj miarę i jednostkę mm/s, nazwy urządzeń oraz liczebność każdej grupy.
- Dodaj etykiety wartości i wyjaśnij znaczenie wyróżnienia. Legenda jest potrzebna tylko wtedy, gdy pomaga zrozumieć kodowanie.
- Uruchamiaj całą komórkę figury od nowa. Nie nakładaj nowych słupków na obiekt ax z poprzedniego wykonania. Porównuj wartości przed zaokrągleniem.

Punkt kontrolny: tabela i wykres pokazują tę samą kolejność, 8 kategorii i po 90 pomiarów. Tytuł nie ocenia bezpieczeństwa urządzeń. W tej karcie nie otrzymujesz klucza median; wylicz je samodzielnie.

Krok 5 Przygotuj serię dzienną w rozdziale 3

M4-W02 odpowiada na pytanie o zmianę typowej wartości energii w kolejnych dniach. Jeden punkt oznacza jeden dzień i wszystkie dostępne pomiary ośmiu urządzeń w tym dniu. Nie łącz bezpośrednio 720 rekordów linią.

```
energia_dzien = (
    pomiary_clean.assign(dzien=pomiary_clean["czas"].dt.normalize())
    .groupby("dzien")["energia"]
    .agg(n="count", mediana="median",
         q1=lambda s: s.quantile(.25), q3=lambda s: s.quantile(.75))
    .reset_index().sort_values("dzien")
)
assert energia_dzien.shape == (45, 5)
assert energia_dzien["dzien"].diff().dropna().eq(pd.Timedelta(days=1)).all()
assert energia_dzien["n"].between(15, 16).all()
assert energia_dzien["n"].sum() == 718
assert energia_dzien["n"].eq(15).sum() == 2
assert (energia_dzien["q1"] <= energia_dzien["mediana"]).all()
assert (energia_dzien["mediana"] <= energia_dzien["q3"]).all()
```

- Narysuj medianę przez ax.plot(), a pas od q1 do q3 przez ax.fill_between(). Uporządkuj daty i sprawdź czytelność etykiet.
- Oś Y opisz jako medianę energii w kWh na pomiar. Pas nazwij Q1–Q3 i wyjaśnij, że pokazuje środkowe 50% wartości w dniu.
- Podaj 45 dni, 718 wartości, dzienne n = 15–16 i dwa braki bez imputacji. Pas nie oznacza przedziału ufności, błędu pomiaru ani prognozy.

Sprawdź rozumienie: dlaczego linia ma 45 punktów, a nie 720? Co traci odbiorca, gdy podpis pomija fakt, że wynik łączy osiem urządzeń? Zapisz odpowiedzi w Markdown.

Porównanie okresów i rozkładów

Krok 6 Odczytaj tabelę M3 i porównaj okresy

M4-W03 pokazuje mediany energii w trzech halach i dwóch okresach. Użyj wyniku M3-P05. Aktualny eksport ma układ wierszowy: rok_miesiac, lokalizacja, energia_mediana, rekordy i energia_n. Najpierw przekształć go do macierzy potrzebnej wykresowi.

```
nazwa_p05 = "M3_P05_energia_miesiac_lokalizacja.csv"
p05 = pd.read_csv(TABLES_DIR / nazwa_p05, encoding="utf-8")
assert p05.shape == (6, 5)
assert not p05.duplicated(["rok_miesiac", "lokalizacja"]).any()
energia_ml = p05.pivot(index="lokalizacja", columns="rok_miesiac",
                       values="energia_mediana")
energia_ml = energia_ml.reindex(
    index=["Hala A", "Hala B", "Hala C"],
    columns=["2026-01", "2026-02"])
assert energia_ml.notna().all().all()
kontrola_ml = pomiary_clean.pivot_table(
    index="lokalizacja", columns="rok_miesiac", values="energia",
    aggfunc="median").reindex_like(energia_ml)
pd.testing.assert_frame_equal(energia_ml, kontrola_ml, check_dtype=False,
                              rtol=1e-10, atol=1e-10)
```

Jeżeli Twoja tabela ma inne kolumny, sprawdź sposób eksportu w M3 i dopasuj odczyt do udokumentowanego schematu. Nie transponuj w ciemno pliku wierszowego. Nie twórz nowych wartości ani nie nadpisuj wyniku M3.

- Narysuj dwa słupki przy każdej hali. Zachowaj stałą kolejność hal, wspólną oś od zera i jednostkę kWh na pomiar.
- Serie nazwij „styczeń 2026” i „1–14 lutego 2026”. Użyj koloru i dodatkowo deseniu lub jednoznacznych etykiet. Podaj liczebności grup i braki z p05.
- Porównuj dostępne okresy. Dwa punkty miesięczne nie uzasadniają trendu, a mediany na pomiar nie oznaczają sum zużycia ani efektywności hal.

Krok 7 Pokaż rozkłady w rozdziale 4

Dla M4-W04 utwórz histogram wszystkich 720 wartości drgania. Porównaj co najmniej dwa ustawienia bins, zachowując ten sam zakres osi. Zapisz wybraną liczbę i granice przedziałów oraz uzasadnienie. Sprawdź, czy suma zliczeń odpowiada liczbie pokazanych pomiarów.

Dla M4-W05 utwórz po jednym poziomym pudełku na urządzenie. Posortuj urządzenia rosnąco według mediany i sprawdź kolejność na ekranie. Zachowaj wspólną skalę oraz po 90 pomiarów. Wyjaśnij medianę, Q1–Q3, wąsy i lokalną regułę 1,5 × IQR w każdej grupie.

W obu figurach zachowaj widoczny pełny zakres z pomiarem 4,850 mm/s. Jeśli dodajesz powiększenie, umieść je obok pełnego widoku w tym samym pliku i nazwij zakres oraz liczbę punktów poza kadrem. Różne granice histogramu wykluczają bezpośrednie porównanie wysokości słupków.

Punkt poza wąsem jest sygnałem statystycznym, nie automatycznie błędem lub awarią. W M4 zachowujesz decyzje M2, w tym status pomiaru 4,850 mm/s „do sprawdzenia”.

Zależności i komunikat dla odbiorcy

Krok 8 Sprawdź zależność w rozdziale 5

Dla M4-W06 wybierz wspólnie kolumny temperatura i drgania oraz identyfikator urządzenia. Usuwaj braki wyłącznie w dwóch polach pary. Potwierdź 720 kompletnych par i zachowaj ich przyporządkowanie do tych samych rekordów.

- Użyj ax.scatter(). Podaj jednostki °C i mm/s, zmniejsz znaczniki i zastosuj alpha, aby ujawnić zagęszczenia.
- Zachowaj maksimum 4,850 mm/s na wykresie głównym. Adnotacja może wskazywać urządzenie i wartość, lecz nie diagnozę.
- Porównaj r i n z M3-P06. Wariant bez maksimum oraz korelacje urządzeń służą ocenie wrażliwości; nie zastępują wyniku pełnego.
- Sprawdź wariant z kolorem według urządzenia. Jeżeli wprowadzasz dodatkową zmienną przez kolor lub rozmiar, wyjaśnij jej znaczenie i skalę.

Mapa korelacji jako narzędzie diagnostyczne

Wybierz jawnie cztery miary: temperatura, drgania, energia i produkcja. Oblicz macierz Pearsona i osobno liczebności kompletnych par. Nie dodawaj identyfikatorów, kodów kategorii ani wszystkich liczbowych metadanych tylko dlatego, że mają typ liczbowy.

```
miary = ["temperatura", "drgania", "energia", "produkcja"]
korelacje = pomiary_clean[miary].corr(method="pearson")
kompletne = pomiary_clean[miary].notna().astype(int)
liczebnosci = kompletne.T.dot(kompletne)
assert korelacje.shape == liczebnosci.shape == (4, 4)
assert liczebnosci.loc["energia", "produkcja"] == 717
```

Narysuj mapę przez ax.imshow() ze skalą od −1 do 1, neutralnym środkiem, wartościami wpisanymi w komórki i legendą „Korelacja Pearsona r”. Pokaż liczebności par w tabeli. Każdą zależność wykorzystaną w narracji obejrzyj również na wykresie punktowym. Niezdefiniowanego r nie zastępuj zerem.

Krok 9 Opracuj narrację w rozdziale 6

- Dla każdego PNG zapisz tytuł, podpis ze źródłem, zakresem, agregacją, jednostką, n i brakami oraz krótkie ograniczenie interpretacji.
- Oddziel wynik obliczenia, obserwację, możliwe wyjaśnienie i ograniczenie. Tytuł może być neutralny, jeśli dane nie wspierają wyraźnej obserwacji.
- Przygotuj tekst alternatywny: typ wykresu, zmienne, najważniejsza obserwacja, istotny punkt nietypowy i ograniczenie. Przy publikacji przypisz go do obiektu obrazu; sam PNG nie załatwia tej czynności.
- Sprawdź czytelność w docelowym rozmiarze i w skali szarości. Informacja nie może zależeć wyłącznie od koloru. Pozostaw dostępne tabele z dokładnymi wartościami.

Nazwy plików i zapis rezultatów

Zachowaj katalog outputs zgodny z kartami M1–M3. Zapis output na slajdach oznacza tę samą rolę, lecz w Twoim projekcie stosujesz wyłącznie outputs. Identyfikator w tekście ma zapis M4-W01, a nazwa pliku M4_W01. Nie zapisuj plików pod skróconymi nazwami M4-W01.png ani wykres_final2.png.

Plik w outputs/figures	Zawartość
M4_W01_ranking_drgan.png	Pełny ranking median drgań.
M4_W02_energia_dzien.png	Dzienne mediany i pas Q1–Q3.
M4_W03_energia_okres_lokalizacja.png	Mediany dla okresów i lokalizacji.
M4_W04_histogram_drgania.png	Rozkład drgań z pełnym zakresem.
M4_W05_boxplot_drgania_urzadzenie.png	Rozkłady ośmiu urządzeń.
M4_W06_temperatura_drgania.png	Zależność temperatury i drgań.
M4_macierz_korelacji.png	Diagnostyczna mapa, poza zestawem sześciu.

Plik w outputs/tables	Co zapisujesz
M4_energia_dzien.csv	45 wierszy: dzien, n, mediana, q1, q3.
M4_drgania_urzadzenie_podsumowanie.csv	8 wierszy: urzadzenie, n, mediana.
M4_tabela_narracji.csv	6 wierszy dokumentujących W01–W06.

Nazwy energia_dzien.csv i drgania_urzadzenie_podsumowanie.csv ze slajdów otrzymują tu prefiks M4. Dane źródłowe M3 zachowują prefiksy M3_P01, M3_P03 i M3_P05. Macierze diagnostyczne i dodatkowe kontrole mogą pozostać w notebooku.

Krok 10 Zapisz i ponownie otwórz wyniki

Po ukończeniu każdej figury wykonaj zapis, podstawiając jej nazwę z tabeli. Najpierw uzupełnij tytuł, osie i źródło; pełny podpis zachowaj również w tabeli narracji. Wykres przekazuj zawsze razem z jego opisem.

```
from IPython.display import Image, display

out = FIGURES_DIR / "M4_W01_ranking_drgan.png"
fig.savefig(out, dpi=150, bbox_inches="tight", facecolor="white")
plt.close(fig)
assert out.is_file() and out.stat().st_size > 0
display(Image(filename=str(out)))
```

Zapisuj tabele przez to_csv(..., index=False, encoding="utf-8"). Jeśli etykiety są w indeksie, przenieś je najpierw przez reset_index(). Odczytaj CSV ponownie i porównaj nazwy kolumn, klucze, braki oraz wartości; datę dzien odtwórz przez parse_dates. Nie nadpisuj plików M3. Ponowne wykonanie zapisu zastępuje wyniki M4 pod tymi samymi nazwami; potrzebną wcześniejszą wersję zarchiwizuj przed przebiegiem.

Tabela narracji i audyt wykresu

Użyj pól: id_wykresu, id_pytania, wynik_zrodlowy, plik, tytul, podpis, obserwacja, ograniczenie, tekst_alternatywny, status_qc. W plik wpisuj ścieżkę względem katalogu projektu. Stosuj pełne ID M3-P01–M3-P06; dla W02 wpisz M3-P05 i objaśnij dzienne rozwinięcie w wynik_zrodlowy. Dla W05 można wskazać M3-P01; M3-P03. Statusy: NIE_SPRAWDZONO, DO_POPRAWY, GOTOWY. Gotowość nadaj po obejrzeniu zapisanego PNG, nie automatycznie po wykonaniu kodu.

Krok 11 Sprawdź jeden wykres szczegółowo

Wybierz W03 lub inny wykres, przy którym łatwo o nadinterpretację. Zapisz co najmniej trzy wykryte ryzyka lub konkretne kontrole. Użyj tabeli audytu: problem, dowód, wpływ na odbiorcę, poprawka, wynik ponownego sprawdzenia. Zmieniaj kod i opis, a nie piksele gotowego obrazu.

Jeżeli wykonujesz ćwiczenie „przed i po” ze slajdów, zachowaj wersję wadliwą jako M4_C06_przed.png wyłącznie do dokumentacji. Poprawiona wersja W03 trafia pod swoją pełną nazwę z poprzedniej strony. W mapie wizualizacji zapisz, dlaczego nowa wersja jest wierniejsza danym.

Obszar kontroli	Warunek zakończenia
Dane	Właściwe źródło, liczebność, braki, agregacja i zgodność z tabelą.
Wizualizacja	Trafny typ, osie i jednostki, uczciwa skala, widoczne wartości skrajne.
Narracja	Tytuł i podpis zgodne z wynikiem, jawny zakres i ograniczenie.
Dostępność	Czytelny kontrast, dodatkowe rozróżnienie poza kolorem, tekst alternatywny.
Eksport	Otwarty i obejrzany PNG, brak ucięć, odtwarzalny kod i przypisany opis.

Weryfikacja opisu LLM

Jeśli korzystasz z LLM, przekaż tylko sprawdzone fakty i potrzebny kontekst bez danych poufnych. Poproś o szkic tytułu, podpisu i tekstu alternatywnego bez dopisywania przyczyn. Każde twierdzenie porównaj z tabelą i wykresem. Zapisz prompt, odpowiedź, kontrolę, werdykt i uzasadnienie w docs/rejestr_uzycia_LLM.csv, z modul = M4, albo odnotuj brak użycia LLM.

Przećwicz ocenę zdania „Hala o najniższej medianie energii jest najbardziej efektywna”. Wskaż, co można sprawdzić w tabeli, a jakich danych lub definicji brakuje do takiej oceny. Zapisz własną, ostrożniejszą wersję opisu, bez zmiany wyniku liczbowego.

Domknięcie pracy i samokontrola

Lista kontroli przed zakończeniem modułu

- □ Notebook M4 ma własne importy, ścieżki, odczyt i kontrolę wejścia.
- □ W01–W06 mają specyfikacje i powstają z wyników kodu, bez ręcznych list wartości.
- □ W02 pokazuje dni, W03 jawnie opisuje niepełny luty, a Q1–Q3 ma właściwe znaczenie.
- □ Histogram i boxplot zachowują pełny zakres; wąsy nie służą do diagnozy awarii.
- □ W06 zachowuje kompletne pary, mapa korelacji ma skalę −1 do 1 i tabelę n.
- □ Sześć wierszy narracji zawiera źródła, podpisy, ograniczenia i teksty alternatywne.
- □ Każdy finalny PNG został obejrzany w docelowym rozmiarze i ma status GOTOWY.
- □ LLM ma zapis weryfikacji albo odnotowano brak użycia; wejście i wyniki M3 są niezmienione.

Krok 12 Odtwórz cały przebieg

Zapisz notebook, wykonaj Restart Kernel and Run All Cells i popraw pierwszy ewentualny błąd. Powtórz pełny przebieg. Potwierdź niezmienność wejścia przez porównanie SHA-256 z hash_wejscia, sprawdź pliki ponownie i zapisz notebook z aktualnymi wynikami. W README.md dopisz sposób uruchomienia M4 oraz listę rezultatów.

Gdy pojawi się problem

Objaw	Najpierw sprawdź
Brak pliku lub kolumny CSV	Katalog roboczy, outputs zamiast output, prefiks M3_P05 i układ wierszowy.
Pusty PNG lub podwójne elementy	Zapis właściwego fig przed plt.close(fig); nowy fig i ax w każdym przebiegu.
Ucięte opisy lub inny wynik niż tabela	Rozmiar figury, układ etykiet, filtr, agregację i kolejność kategorii.

Dla chętnych: wykonaj ćwiczenia z prezentacji dla dziennej produkcji, temperatury w halach lub par energia–produkcja. Zachowaj nazwy M4_C03_produkcja_dzien.png, M4_C04_histogram_temperatura.png, M4_C05_boxplot_temperatura.png i M4_C05_energia_produkcja.png. Nie zastępują one sześciu wykresów W01–W06.

Na zakończenie dopisz: „Najważniejszą zmianą mojego wykresu było …, ponieważ odbiorca mógł wcześniej …”. Wskaż jedną decyzję wizualną i dowód, że poprawa pomogła zachować sens danych.