

# Lembar Praktikum

## Pertemuan 3 – Data Wrangling dengan Pandas

Mata Kuliah Data Science

September 4, 2025

### Tujuan

- Mahasiswa mampu melakukan data wrangling menggunakan Pandas.
- Mahasiswa mampu membersihkan dataset (cleaning).
- Mahasiswa mampu menangani missing values.
- Mahasiswa mampu melakukan transformasi data sederhana.

### Persiapan

1. Pastikan Python 3 dan library `pandas` sudah terinstal.
2. Siapkan dataset CSV sederhana, misalnya `data_penjualan.csv`.
3. Dataset minimal berisi kolom: `nama`, `umur`, `qty`, `price`.

### Langkah Praktikum

#### 1. Membaca Data

```
import pandas as pd

df = pd.read_csv("data_penjualan.csv")
print(df.head())
```

**Tugas:** Tampilkan 5 baris pertama dataset. Apakah ada kolom/baris yang tidak diperlukan?

#### 2. Cleaning Data

```
# Hapus kolom yang tidak diperlukan
df = df.drop(columns=["Unnamed: 0"], errors="ignore")

# Hapus data duplikat
```

```

df = df.drop_duplicates()

# Rename kolom
df = df.rename(columns={"qty": "jumlah", "price": "harga"})

```

Tugas:

- Hapus satu kolom yang tidak relevan.
- Hapus baris duplikat.
- Ubah nama minimal 2 kolom agar lebih informatif.

### 3. Menangani Missing Values

```

# Drop baris dengan NaN
df = df.dropna()

# Isi nilai kosong dengan mean
df["umur"] = df["umur"].fillna(df["umur"].mean())

```

Tugas:

- Hitung rata-rata kolom umur.
- Isi nilai kosong di kolom umur dengan rata-rata tersebut.
- Catat berapa banyak nilai kosong sebelum dan sesudah perbaikan.

### 4. Transformasi Data

```

# Replace nilai kategorikal
df["gender"] = df["gender"].replace({"M": "Male", "F": "Female"})

# Normalisasi kolom harga
df["harga_norm"] = (df["harga"] - df["harga"].min()) / \
                    (df["harga"].max() - df["harga"].min())

# Feature engineering
df["total_price"] = df["jumlah"] * df["harga"]

```

Tugas:

- Lakukan normalisasi pada kolom harga.
- Tambahkan kolom baru total\_price = jumlah × harga.
- Ganti nilai kategorikal (misalnya gender M/F menjadi Male/Female).

## **Soal Latihan Mandiri**

1. Cari tahu fungsi `interpolate()` di Pandas. Gunakan untuk mengisi nilai kosong pada kolom numerik.
2. Buat ringkasan statistik (`df.describe()`) sebelum dan sesudah cleaning. Apa perbedaannya?
3. Buat laporan singkat (1–2 halaman) tentang langkah wrangling yang Anda lakukan. Sertakan potongan kode dan hasil keluaran.

## **Pengumpulan**

Kumpulkan:

- File `.ipynb` atau `.py` berisi kode praktikum.
- Laporan singkat (PDF) tentang proses dan hasil data wrangling.