

# TUGAS MANDIRI - TOOLS DATA SCIENCE

Mata Kuliah: Data Science

Fakultas Ilmu Komputer

Universitas Sriwijaya

**Tujuan** : Mahasiswa mampu memilih tools data science yang tepat, melakukan cleaning, analisis SQL, visualisasi, dan memahami Scraping vs API.

**Dataset** : Publik ( $\geq 500$  baris,  $\geq 6$  kolom)

**Deliverables** :

- Laporan PDF (6-10 halaman).
- Dataset (CSV) + (opsional) Excel.
- SQL script (.sql) atau database SQLite (.db).
- Notebook/Script (Python/R) atau workflow (WEKA/KNIME/Orange/RapidMiner/etc).

**Langkah kerja** :

1. Tool selection: tabel tahap kerja - tools - alasan - kelebihan - keterbatasan (minimal 3 tools).
2. Cleaning: minimal 5 aksi cleaning/transformasi + bukti sebelum vs sesudah.
3. SQL: desain tabel + minimal 8 query (SELECT+WHERE, GROUP BY, ORDER BY, JOIN jika ada  $>1$  tabel, dan 1 query untuk menjawab pertanyaan analitis).
4. Visualisasi: minimal 4 grafik (distribusi, perbandingan kategori, tren waktu jika ada tanggal, dan hubungan 2 variabel) + interpretasi singkat tiap grafik.
5. Scraping vs API: jelaskan perbedaan, kapan dipakai, resiko/etika.

**Penilaian (100 poin) :**

1. Tools & penjelasan (15)
2. Cleaning (15)
3. SQL (20)
4. Visualisasi & insight (40)
5. Scraping/API & storage (20)

**Pengumpulan** :

- Upload 1 file dalam format \*.pdf, Nama file: DS-Bab\_Tools\_DS\_Nama\_NIM\_Kelas.
- Upload ke G-Form: <https://forms.gle/pxZNwKYdpUo6eQrE7>

**Deadline** : 19 Desember 2025 pukul 24.00 WIB

# TEMPLATE LAPORAN TUGAS MANDIRI

Tools Data Science, Visualisasi, dan Scraping vs API

Nama : \_\_\_\_\_

NIM : \_\_\_\_\_

Kelas : \_\_\_\_\_

Tanggal: \_\_\_\_\_

## 1. Deskripsi Dataset

Tuliskan sumber dataset (URL atau asal data), periode data, jumlah baris/kolom, dan definisi singkat tiap kolom penting.

Item	Isi
Sumber dataset	(contoh: Open Data / Kaggle / hasil pencatatan mandiri)
Ukuran data	(contoh: 1.250 baris, 9 kolom)
Tujuan analisis	(contoh: mencari pola/tren dan insight utama)

## 2. Tool Selection & Justification

Isi tabel berikut. Minimal 3 tools (mis. Excel, Python, SQL, R/Julia, KNIME/RapidMiner/Orange).

Tahap kerja	Tools	Alasan	Kelebihan	Keterbatasan
Input data	CSV/Excel			
Cleaning	Python (pandas)			
Storage	SQL (SQLite)			
Analisis	Python / R			
Visualisasi	Matplotlib/Excel			

Catatan: ganti contoh di atas sesuai pilihan Anda.

## 3. Data Cleaning & Transformasi

Jelaskan minimal 5 aktivitas cleaning/transformasi. Sertakan bukti ringkas (10 baris sebelum vs sesudah atau ringkasan statistik).

No	Masalah data	Tindakan	Bukti (before/after)
1	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)
2	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)
3	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)
4	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)
5	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)
6	(mis. missing value kolom X)	(mis. imputasi median / drop)	(mis. %missing 12% -> 0%)

## 4. SQL: Desain Tabel & Query

Sertakan: (1) desain tabel (nama kolom & tipe), (2) minimal 8 query SQL, (3) interpretasi hasil.

### 4.1 Desain tabel

Kolom	Tipe data	Keterangan
(nama_kolom)	(INTEGER/TEXT/REAL/DATE)	(arti kolom)
(nama_kolom)	(INTEGER/TEXT/REAL/DATE)	(arti kolom)
(nama_kolom)	(INTEGER/TEXT/REAL/DATE)	(arti kolom)
(nama_kolom)	(INTEGER/TEXT/REAL/DATE)	(arti kolom)
(nama_kolom)	(INTEGER/TEXT/REAL/DATE)	(arti kolom)

### 4.2 Log query SQL

No	Tujuan query	Query (ringkas)	Output
1	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
2	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
3	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
4	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
5	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
6	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
7	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)
8	(mis. Top 5 kategori)	(SELECT ... GROUP BY ... ORDER BY ...)	(tabel/grafik)

## 5. Visualisasi & Insight

Lampirkan minimal 4 grafik. Untuk tiap grafik: judul, variabel, dan 2-3 kalimat interpretasi.

No	Jenis grafik	Tujuan	Insight (2-3 kalimat)
1	(mis. Bar/Line/Scatter)	(mis. membandingkan kategori)	(apa maknanya?)
2	(mis. Bar/Line/Scatter)	(mis. membandingkan kategori)	(apa maknanya?)
3	(mis. Bar/Line/Scatter)	(mis. membandingkan kategori)	(apa maknanya?)
4	(mis. Bar/Line/Scatter)	(mis. membandingkan kategori)	(apa maknanya?)

## 6. Web Scraping vs API

Tulis 1-2 halaman: perbedaan, kapan memilih API, risiko/etika scraping.

## 7. Kesimpulan & Keterbatasan

Tuliskan 5-8 poin kesimpulan (insight utama) dan keterbatasan analisis (data/alat/metode).