

Dari Titik Titik Data ke Narasi Bermakna

Preparation Data

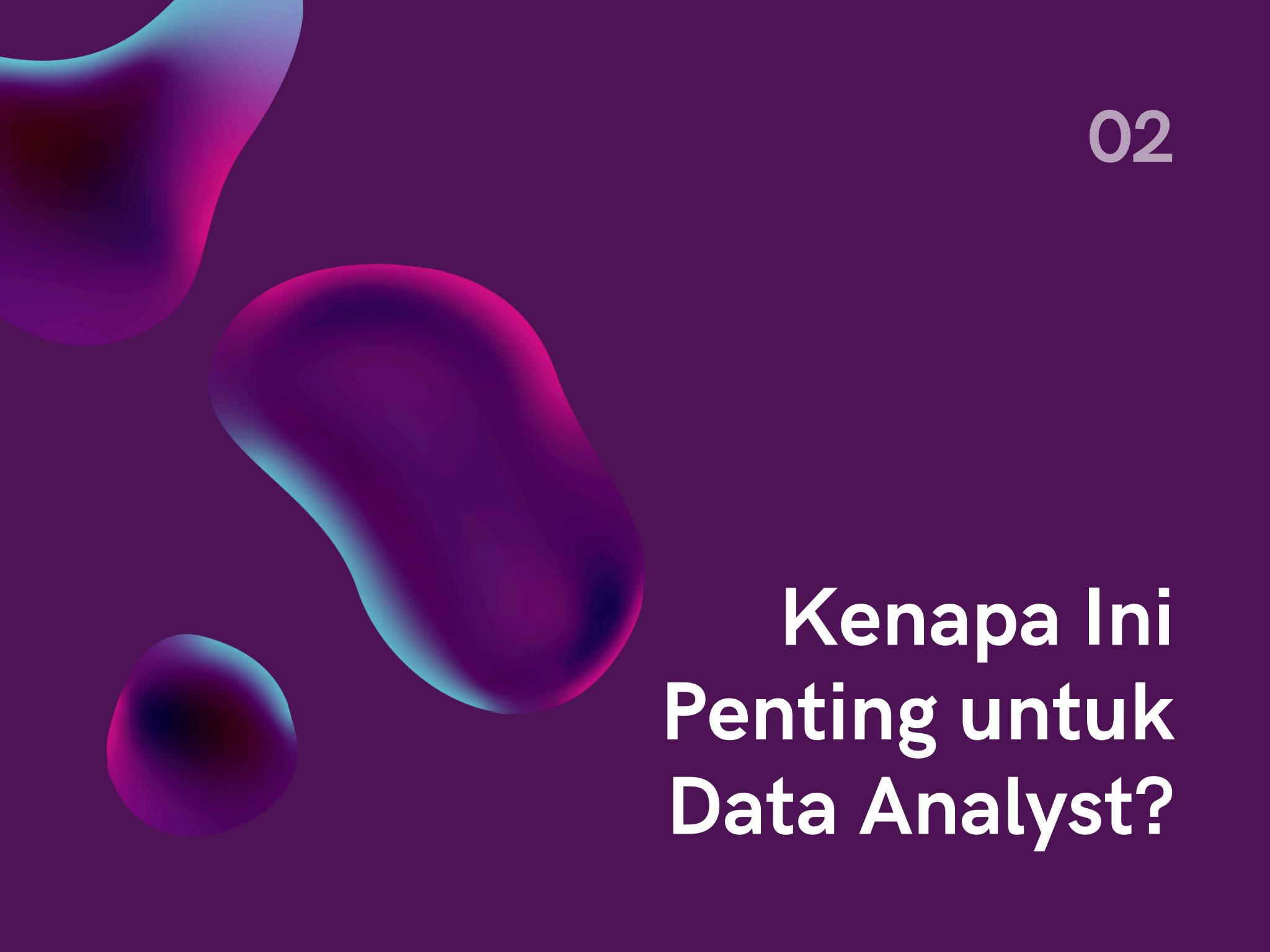
01

Melanjutkan dari Proyek Sebelumnya

- Setelah database Transjakarta dibangun...
- Data mencakup: bus, driver, rute, jadwal, transaksi.
- Tahapan selanjutnya: data preparation untuk analisis.
- Fokus: konsolidasi data dan penyusunan struktur awal.

Apa Itu Data Preparation?

- **Proses menggabungkan data dari berbagai tabel.**
- **Tujuan: membentuk dataset yang siap dianalisis.**
- **Contoh output: view yang menyatukan informasi penting.**



02

Kenapa Ini
Penting untuk
Data Analyst?

- Data analyst butuh data yang sudah digabung dan siap pakai.
- Mempermudah eksplorasi data tanpa perlu join manual.
- Menjadi dasar dalam membuat visualisasi dan laporan.

Komponen dalam Preparing Data

Function

View

Struktur inti dalam file ini adalah pembuatan view, yaitu virtual table hasil penggabungan dari beberapa tabel.

Join

Perintah di SQL yang digunakan untuk menggabungkan data dari dua atau lebih tabel, berdasarkan kolom yang saling berhubungan (biasanya kolom ID).

Select

Pemilihan data yang akan dimasukkan ke view: timestamp, nama halte, nama sopir, rute, waktu keberangkatan, dll.

```
-|SELECT
  u.user_id,
  u.name AS user_name,
  u.email,
  u.phone AS user_phone,

  c.card_id,
  c.card_number,
  c.balance,
  c.status AS card_status,
  c.issued_at,

  t.transaction_id,
  t.fare,

  s.schedule_id,
  s.departure_time,
  s.arrival_time,

  r.route_id,
  r.route_name,
  r.start_point,
  r.end_point,
  r.distance_km,

  st.stop_id,
  st.name AS stop_name,
  st.location AS stop_location,
  st.is_terminal,
```

Select Function

Query ini menyusun satu baris data yang merepresentasikan perjalanan lengkap pengguna dalam sistem transportasi. Data dimulai dari informasi pengguna (users) dan kartu yang dimiliki (cards), lalu dikaitkan dengan transaksi perjalanan (transactions), jadwal (schedules), dan rute yang ditempuh (routes). Lokasi naik atau turun diperoleh dari tabel halte (stops). Untuk memberikan konteks operasional, ditambahkan pula data bus yang digunakan (buses), sopir yang bertugas (drivers), serta tarif berdasarkan rute (ticket_prices). Gabungan ini membentuk cerita utuh perjalanan dari sisi penumpang, kendaraan, hingga rute yang dilalui.

Join Function

```
FROM users u
JOIN cards c ON u.user_id = c.user_id
JOIN transactions t ON c.card_id = t.card_id
JOIN schedules s ON t.schedule_id = s.schedule_id
JOIN routes r ON s.route_id = r.route_id
JOIN stops st ON s.stop_id = st.stop_id
JOIN buses b ON s.bus_id = b.bus_id
JOIN drivers d ON s.driver_id = d.driver_id
LEFT JOIN incidents i ON i.schedule_id = s.schedule_id;
```

Query ini menggabungkan beberapa tabel untuk menyusun informasi perjalanan pengguna secara lengkap. Dimulai dari data pengguna (users) yang dihubungkan dengan kartu transportasi mereka (cards) dan transaksi yang dilakukan (transactions). Transaksi tersebut dikaitkan dengan jadwal perjalanan (schedules), yang selanjutnya terhubung dengan rute (routes), halte (stops), bus (buses), dan sopir (drivers) yang bertugas. Selain itu, query juga menambahkan data kejadian khusus atau insiden selama perjalanan (incidents) menggunakan LEFT JOIN, agar data tetap ditampilkan meskipun tidak ada insiden. Struktur ini memungkinkan analisis menyeluruh atas satu perjalanan dari sisi pengguna hingga operasional di lapangan.

View

100 % ▾

Results Messages

| | user_id | user_name | email | user_phone | card_id | card_number | balance | card_status | issued_at | transaction_id | fare | schedule_id |
|----|---------|-----------|-----------------|-------------|---------|-------------|-----------|-------------|-------------------------|----------------|---------|-------------|
| 1 | 1 | User_1 | user1@mail.com | 08123000001 | 1 | CARD000001 | 100000.00 | active | 2025-04-19 15:13:46.640 | 1 | 3500.00 | 62 |
| 2 | 2 | User_2 | user2@mail.com | 08123000002 | 2 | CARD000002 | 100000.00 | active | 2025-04-19 15:13:46.643 | 2 | 3500.00 | 265 |
| 3 | 3 | User_3 | user3@mail.com | 08123000003 | 3 | CARD000003 | 100000.00 | active | 2025-04-19 15:13:46.643 | 3 | 3500.00 | 65 |
| 4 | 4 | User_4 | user4@mail.com | 08123000004 | 4 | CARD000004 | 100000.00 | active | 2025-04-19 15:13:46.643 | 4 | 3500.00 | 289 |
| 5 | 5 | User_5 | user5@mail.com | 08123000005 | 5 | CARD000005 | 100000.00 | active | 2025-04-19 15:13:46.643 | 5 | 3500.00 | 233 |
| 6 | 6 | User_6 | user6@mail.com | 08123000006 | 6 | CARD000006 | 100000.00 | active | 2025-04-19 15:13:46.643 | 6 | 3500.00 | 33 |
| 7 | 7 | User_7 | user7@mail.com | 08123000007 | 7 | CARD000007 | 100000.00 | active | 2025-04-19 15:13:46.643 | 7 | 3500.00 | 248 |
| 8 | 8 | User_8 | user8@mail.com | 08123000008 | 8 | CARD000008 | 100000.00 | active | 2025-04-19 15:13:46.647 | 8 | 3500.00 | 11 |
| 9 | 9 | User_9 | user9@mail.com | 08123000009 | 9 | CARD000009 | 100000.00 | active | 2025-04-19 15:13:46.647 | 9 | 3500.00 | 337 |
| 10 | 10 | User_10 | user10@mail.com | 08123000010 | 10 | CARD000010 | 100000.00 | active | 2025-04-19 15:13:46.647 | 10 | 3500.00 | 92 |

100 % ▾

Results Messages

| | departure_time | arrival_time | route_id | route_name | start_point | end_point | distance_km | stop_id | stop_name | stop_location |
|----|-------------------------|-------------------------|----------|------------|-------------|------------|-------------|---------|------------|-----------------|
| 1 | 2025-01-09 15:00:00.000 | 2025-01-09 16:00:00.000 | 2 | Koridor 2 | Pulogadung | Harmoni | 12.50 | 2 | Kota | Jakarta Barat |
| 2 | 2025-02-08 05:00:00.000 | 2025-02-08 06:00:00.000 | 3 | Koridor 3 | Kalideres | Pasar Baru | 17.20 | 7 | Dukuh Atas | Jakarta Pusat |
| 3 | 2025-01-10 05:00:00.000 | 2025-01-10 06:00:00.000 | 3 | Koridor 3 | Kalideres | Pasar Baru | 17.20 | 1 | Blok M | Jakarta Selatan |
| 4 | 2025-02-12 05:00:00.000 | 2025-02-12 06:00:00.000 | 4 | Koridor 4 | Pulogadung | Dukuh Atas | 14.10 | 5 | Kalideres | Jakarta Barat |
| 5 | 2025-02-04 05:00:00.000 | 2025-02-04 06:00:00.000 | 2 | Koridor 2 | Pulogadung | Harmoni | 12.50 | 7 | Dukuh Atas | Jakarta Pusat |
| 6 | 2025-01-06 05:00:00.000 | 2025-01-06 06:00:00.000 | 1 | Koridor 1 | Blok M | Kota | 15.00 | 7 | Dukuh Atas | Jakarta Pusat |
| 7 | 2025-02-05 19:00:00.000 | 2025-02-05 20:00:00.000 | 2 | Koridor 2 | Pulogadung | Harmoni | 12.50 | 2 | Kota | Jakarta Barat |
| 8 | 2025-01-02 09:00:00.000 | 2025-01-02 10:00:00.000 | 1 | Koridor 1 | Blok M | Kota | 15.00 | 5 | Kalideres | Jakarta Barat |
| 9 | 2025-02-19 05:00:00.000 | 2025-02-19 06:00:00.000 | 2 | Koridor 2 | Pulogadung | Harmoni | 12.50 | 4 | Harmoni | Jakarta Pusat |
| 10 | 2025-01-14 11:00:00.000 | 2025-01-14 12:00:00.000 | 1 | Koridor 1 | Blok M | Kota | 15.00 | 5 | Kalideres | Jakarta Barat |

| is_terminal | bus_id | plate_number | bus_model | capacity | bus_status | driver_id | driver_name | license_number | driver_phone | driver_status |
|-------------|--------|--------------|-----------|----------|------------|-----------|-------------|----------------|--------------|---------------|
| 1 | 2 | B2345TJ | Scania | 60 | active | 19 | Driver_19 | SIM000019 | 08213000019 | active |
| 1 | 2 | B2345TJ | Scania | 60 | active | 11 | Driver_11 | SIM000011 | 08213000011 | active |
| 1 | 7 | B7890TJ | BYD | 60 | active | 2 | Driver_2 | SIM000002 | 08213000002 | active |
| 1 | 10 | B0123TJ | Daewoo | 60 | active | 13 | Driver_13 | SIM000013 | 08213000013 | active |
| 1 | 7 | B7890TJ | BYD | 60 | active | 11 | Driver_11 | SIM000011 | 08213000011 | active |
| 1 | 7 | B7890TJ | BYD | 60 | active | 4 | Driver_4 | SIM000004 | 08213000004 | active |
| 1 | 6 | B6789TJ | Hyundai | 60 | active | 6 | Driver_6 | SIM000006 | 08213000006 | active |
| 1 | 3 | B3456TJ | Hino | 60 | active | 6 | Driver_6 | SIM000006 | 08213000006 | active |
| 0 | 2 | B2345TJ | Scania | 60 | active | 14 | Driver_14 | SIM000014 | 08213000014 | active |

Workshop Data: Gabung, Susun, Sajikan

Berfokus pada penggabungan data dari beberapa tabel menggunakan JOIN, seperti transactions, schedules, dan routes, untuk membentuk satu tampilan terpadu.

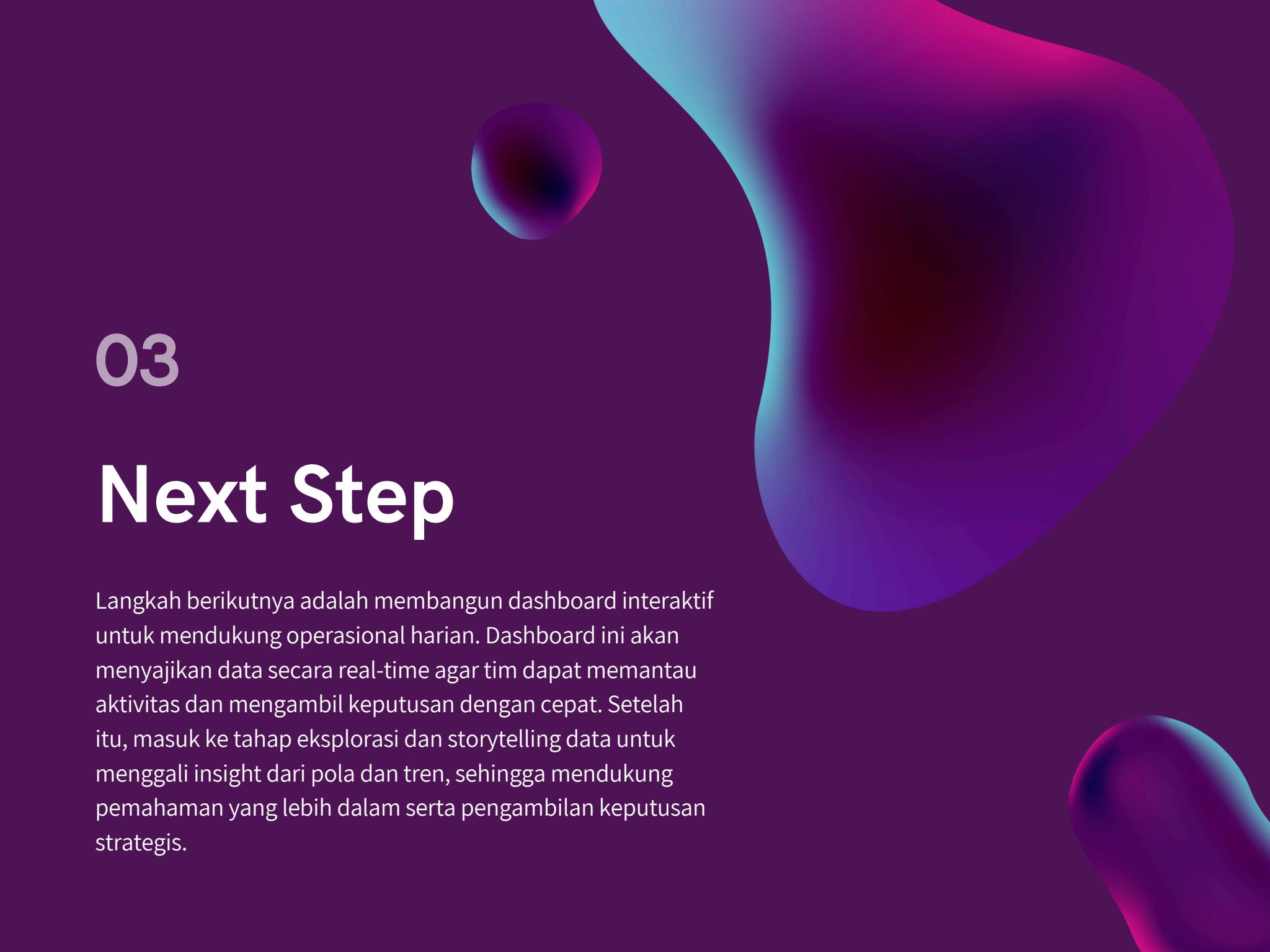
Belum dilakukan filter, agregasi, atau transformasi hanya menyusun struktur dasar dalam bentuk view sebagai fondasi analisis selanjutnya.

View sebagai Dasar Analisis

Output: view yang menyatukan informasi dari banyak tabel.

Data tetap mentah namun sudah dalam format terstruktur.

Menjadi input utama bagi tools analisis seperti Power BI atau tableau, dan dll.



03

Next Step

Langkah berikutnya adalah membangun dashboard interaktif untuk mendukung operasional harian. Dashboard ini akan menyajikan data secara real-time agar tim dapat memantau aktivitas dan mengambil keputusan dengan cepat. Setelah itu, masuk ke tahap eksplorasi dan storytelling data untuk menggali insight dari pola dan tren, sehingga mendukung pemahaman yang lebih dalam serta pengambilan keputusan strategis.

Thank You



hafidzhaqikal9@gmail.com



087814790986



haikaltech09



@hafidzhqkl