

Tugas 1

1	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
2	Customer 1	Yaris AT	MLG	3/13/2019	340.95	6	2045.7
3	Customer 1	Etios AT	SBY	4/16/2019	799.95	1	799.95
4	Customer 3	Avanza AT	MLG	6/18/2019	168.95	7	1182.65
5	Customer 3		MLG	7/24/2019	168.95	2	337.9
6	Customer 1	Avanza AT	SBY	8/21/2019	168.95	9	1520.55
7	Customer 2	Etios AT	KDR	9/7/2019	799.95	7	5599.65
8	Customer 3	Etios AT	MLG	9/7/2019	799.95	8	6399.6
9	Customer 2	Avanza AT			168.95	15	2534.25
10	Customer 1	Avanza AT	SBY	9/14/2019	168.95	15	2534.25
11	Customer 1	Etios AT	SBY	9/25/2019	799.95	2	1599.9
12	Customer 1	Avanza AT	SBY	1/21/2020	168.95	14	2365.3
13	Customer 1	Yaris AT	MLG	2/1/2020	340.95	2	681.9
14	Customer 1				340.95	7	2386.65
15	Customer 1	Etios AT	SBY	5/30/2020	799.95	9	7199.55
16	Customer 3	Avanza AT	MLG	6/7/2020	168.95	10	1689.5
17	Customer 3	Yaris AT	MLG	6/9/2020	340.95	4	1363.8
18	Customer 2		KDR	8/7/2020	799.95	14	11199.3
19	Customer 2	Avanza AT	KDR	8/8/2020	168.95	12	2027.4
20	Customer 1	Avanza AT		8/12/2020	168.95	4	675.8

Soal :

1. Berapa jumlah kolom pada data tersebut?
2. Apa arti atau isi data dari setiap kolom yang ada?
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Jawaban :

1. **Jumlah Kolom**

Dari gambar yang diberikan, terdapat **6 kolom** dalam dataset ini:

- **Customer**: Menunjukkan nama pelanggan.
- **Product**: Menunjukkan jenis produk yang dibeli.
- **Region**: Wilayah atau daerah tempat transaksi terjadi.
- **Date**: Tanggal transaksi terjadi.

- **Item unit price**: Harga per unit dari produk.
- **No.Items**: Jumlah unit produk yang dibeli.
- **Total Sale**: Total harga transaksi (dihitung dari $\text{Item unit price} \times \text{No.Items}$).

2. Apakah Ada Data Null atau Tidak Lengkap?

Dari pengamatan pada gambar, tidak terlihat adanya data yang kosong (null) atau tidak lengkap dalam tabel ini. Semua baris tampak memiliki data yang sesuai untuk setiap kolom.

3. Apakah Ada Data dengan Tipe yang Berbeda dalam Kolom yang Sama?

- **Customer**: Berisi teks (string) → Konsisten.
- **Product**: Berisi teks (string) → Konsisten.
- **Region**: Berisi teks (string) → Konsisten.
- **Date**: Berisi tanggal (string dalam format MM/DD/YYYY) → Konsisten.
- **Item unit price**: Berisi angka desimal (float) → Konsisten.
- **No.Items**: Berisi angka bulat (integer) → Konsisten.
- **Total Sale**: Berisi angka desimal (float) → Konsisten.

Tidak ditemukan data yang memiliki tipe berbeda dengan data lainnya dalam kolom yang sama.

Kesimpulannya, data tampak rapi dan tidak ada data yang kosong atau memiliki tipe data yang berbeda dari yang seharusnya.

Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
2. PDI Spoon melakukan proses extract, perhatikan pada **Tab Logging** di **Execution Results Area**, langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

Jawaban :

1. Sesuai
2. Loading Transformation → PDI Spoon membaca dan memuat transformasi yang dibuat.
Initializing Step: CSV File Input → Sistem mulai membaca file CSV yang telah dipilih.
Checking Delimiter & File Structure → Mengecek apakah delimiter sesuai dengan format file (misalnya koma , atau titik koma ;).
Reading File & Parsing Data → Data dalam CSV dibaca dan diproses sesuai dengan struktur yang telah ditentukan.
Fetching Metadata & Columns → Sistem mengambil nama kolom dengan menekan "Get Fields".
Executing Transformation → Proses ekstraksi dijalankan dan hasilnya ditampilkan pada Preview Data.
Finalizing & Logging Execution Time → PDI Spoon mencatat waktu eksekusi serta error (jika ada).
3. Pada tab **Execution Results → Logging**, terdapat beberapa huruf yang menunjukkan status proses:
 - **I (Input)** → Jumlah data yang dibaca dari sumber.
 - **O (Output)** → Jumlah data yang berhasil dikeluarkan.
 - **R (Read)** → Jumlah record (baris) yang telah dibaca.
 - **W (Written)** → Jumlah record yang ditulis ke target.
 - **U (Updated)** → Jumlah record yang diperbarui (jika ada perubahan data).
 - **E (Error)** → Jumlah error yang terjadi selama proses ekstraksi

TUGAS 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

Perbedaan isi data antara file CSV dan file Excel terletak pada format dan fitur yang didukung. File CSV hanya berisi data dalam bentuk teks dengan pemisah koma (,), sehingga tidak mendukung format seperti warna, rumus, atau multi-sheet. Sebaliknya, file Excel (XLS/XLSX) menyimpan data dalam bentuk tabel dengan sel yang dapat memiliki format khusus, warna, rumus, dan mendukung penggunaan beberapa sheet dalam satu file. Selain itu, ukuran file Excel cenderung lebih besar dibandingkan CSV karena menyimpan lebih banyak informasi tambahan, sementara CSV lebih ringan dan cocok untuk transfer data antar sistem.

2. Jelaskan apa yang terjadi pada proses **Filter rows!**
 - **Membaca data dari sumber** (misalnya file CSV).

- **Memeriksa setiap baris berdasarkan kriteria filter** (misalnya hanya menampilkan data dari tahun tertentu atau hanya produk tertentu).
 - **Menghapus baris yang tidak sesuai dengan filter.**
 - **Meneruskan data hasil filter ke langkah berikutnya.**
3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini?
Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). bandingkan dalam bentuk tabel