**LAPORAN PRAKTIKUM**

**JOBSHEET 1: DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON**

Dosen Pengampu:  
**Rudy Ariyanto, ST., M.Cs.**

****

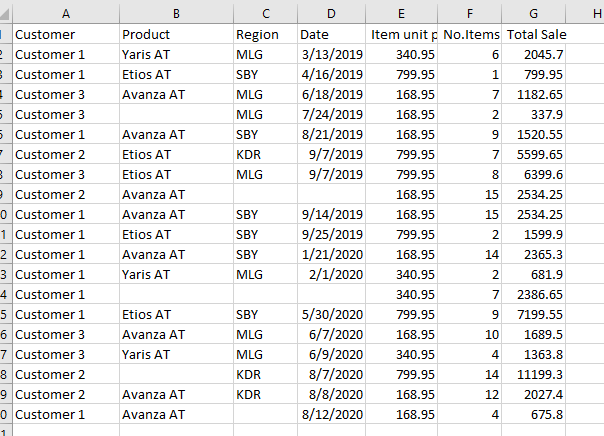
Disusun Oleh:  
**Yusra Yusuf**  
**NIM: 2341760044**

**POLITEKNIK NEGERI MALANG**  
**JURUSAN TEKNOLOGI INFORMASI**  
**PROGRAM STUDY SISTEM INFORMASI BISNIS**

**2025**

**Jobsheet 1: Data Warehouse Sederhana dan Pengenalan PDI Spoon**

**TUGAS 1**



1. Berapa jumlah kolom pada data tersebut?

**Jumlah Kolom**: Dataset memiliki **7 kolom**.

1. Apa arti atau isi data dari setiap kolom yang ada?

* **Customer**: Nama pelanggan.
* **Product**: Model mobil yang dibeli.
* **Region**: Wilayah pembelian.
* **Date**: Tanggal transaksi (format teks).
* **Item unit price**: Harga satuan mobil (angka desimal).
* **No.Items**: Jumlah unit mobil yang dibeli (bilangan bulat).
* **Total Sale**: Total penjualan (angka desimal)

1. Adakah data yang memiliki nilai null / data yang tidak lengkap?

* **Product**: 3 nilai kosong.
* **Region**: 3 nilai kosong.
* **Date**: 2 nilai kosong.

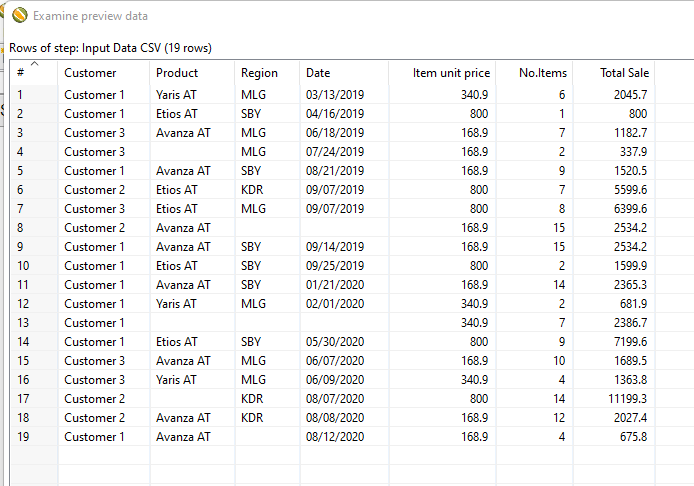
1. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

TIDAK ADA

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

SUDAH SAMA



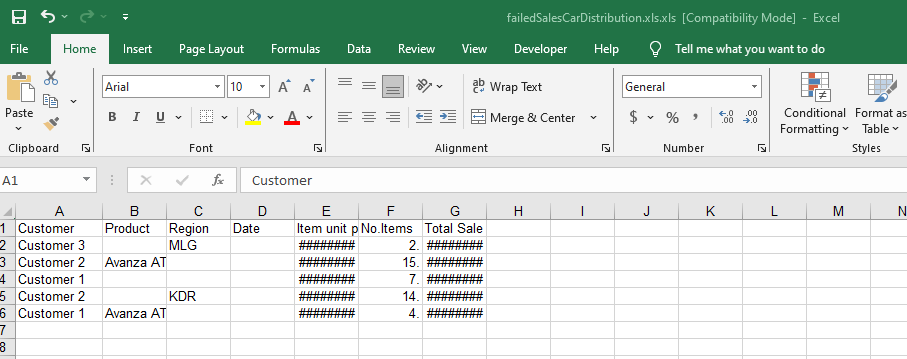
1. Saat PDI Spoon melakukan proses extract, langkah-langkahnya meliputi inisialisasi job atau transformasi, membuka koneksi ke sumber data, menjalankan query atau membaca file, serta mengambil dan memproses data. Setelah itu, Spoon melakukan parsing, validasi awal, dan transformasi jika diperlukan sebelum mengirim data ke langkah berikutnya. Sepanjang proses, Spoon mencatat informasi eksekusi di **Tab Logging**, termasuk jumlah data yang diambil, waktu eksekusi, serta potensi error seperti kegagalan koneksi atau format data yang tidak sesuai.
2. Huruf **I, O, R, W, U, E** biasanya muncul di **Tab Logging** atau indikator pada langkah-langkah transformasi dan memiliki arti sebagai berikut:

* **I (Input)** → Jumlah record yang masuk ke dalam step.
* **(Output)** → Jumlah record yang keluar dari step.
* **R (Read)** → Jumlah record yang dibaca dari sumber data.
* **W (Written)** → Jumlah record yang ditulis ke target data.
* **U (Updated)** → Jumlah record yang diperbarui dalam proses.
* **E (Error)** → Jumlah record yang mengalami error dalam proses.

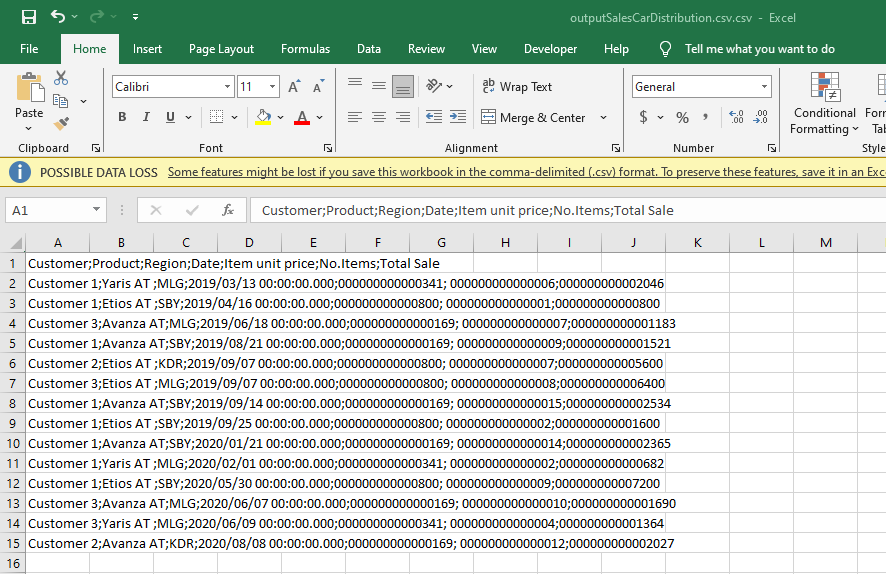
Indikator ini membantu dalam memantau aliran data dan mengidentifikasi potensi masalah dalam transformasi atau job yang dijalankan.

1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Sangat cepat
2. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

exel



csv



CSV hanya berisi data mentah, sedangkan Excel dapat menyajikan data dengan format lebih rapi dan kaya fitur.

1. **Jelaskan apa yang terjadi pada proses Filter rows!**

Filter Rows adalah langkah dalam transformasi ETL yang digunakan untuk menyaring data berdasarkan kriteria tertentu. Dalam **Pentaho Spoon**, Filter Rows bekerja dengan cara:

* Mengevaluasi setiap baris data berdasarkan kondisi yang telah ditentukan (misalnya, hanya menampilkan data dengan total penjualan > 1000).
* Memisahkan data yang memenuhi kriteria (data yang "lolos" filter) untuk diproses lebih lanjut.
* Mengabaikan atau membuang data yang tidak sesuai dengan filter.

1. **Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini. Sangat cepat**