

NIM : 2241760004

No Absen : 10

Nama Lengkap : Fitria Nur Sholikah

**Daftar Isi**

[Jobsheet 1: Data Warehouse Sederhana 3](#_Toc161414475)

[**Studi Kasus** 3](#_Toc161414476)

[**TUGAS 1** 3](#_Toc161414477)

[**A.** **Pengambilan Data (Extract)** 4](#_Toc161414478)

[**TUGAS 2** 7](#_Toc161414479)

[**B.** **Filter Data (Transform) dan Pengemasan data (Load)** 8](#_Toc161414480)

[**TUGAS 3** 14](#_Toc161414481)

[**Studi Kasus** 14](#_Toc161414482)

# **Jobsheet 1: Data Warehouse Sederhana**

## **Studi Kasus**

Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan secara berulang setiap hari setelah jam kantor maka, Pak Amir akan membuat proses untuk menarik data yang ada pada file milik supervisor tersebut secara otomatis. Namun, kadang data tersebut tidak lengkap. Sehingga Pak Amir akan mengambil data yang lengkap saja dan mengembalikan data yang tidak lengkap.

Dari studi kasus tersebut maka, akan dilakukan:

1. Cek dan Analisa data penjualan.
2. Import data dari file (Extract)
3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
4. Memindahkan data yang sudah lengkap ke file dashboard. (Load)
5. Mengumpulkan data yang belum lengkap untuk dikembalikan.

### **TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

**Jawab:** Ada 7 kolom

* Kolom Customer
* Kolom Product
* Kolom Region
* Kolom Date
* Kolom Item unit price
* Kolom No.Items
* Kolom Total Sale.

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

**Jawab:**

* Kolom 1 (Customer) : Berisi data customer.
* Kolom 2 (Product) : Berisi data nama produk.
* Kolom 3 (Region) : Berisi data alamat/kota.
* Kolom 4 (Date) : Berisi data tanggal.
* Kolom 5 (Item unit price) : Berisi data item unit price(harga barang).
* Kolom 6 (No.Items) : Berisi data nomor barang.
* Kolom 7 (Total sale) : Berisi data total penjualan.

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

**Jawab:**

Ada, pada beberapa kolom di baris tertentu.

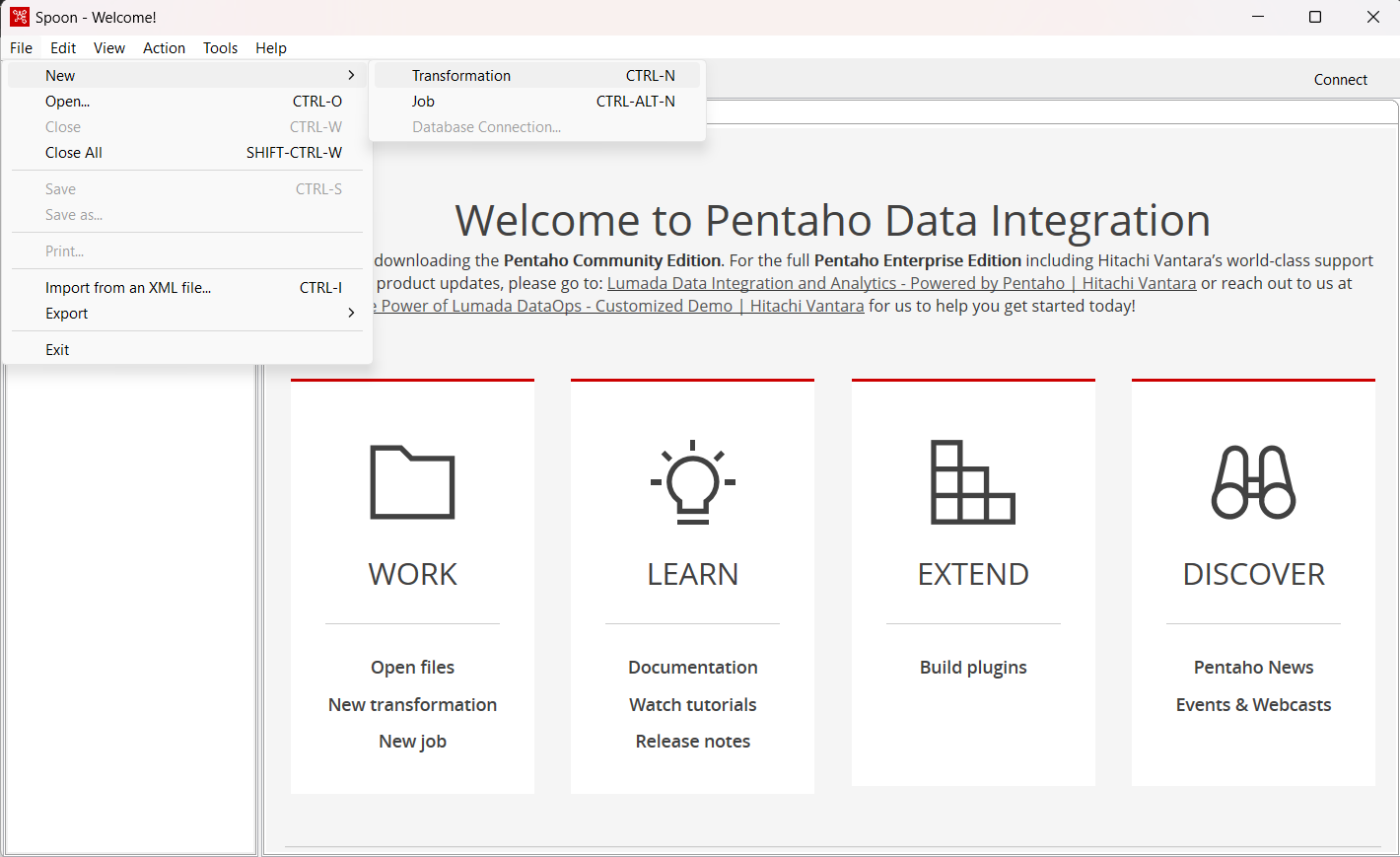
1. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

**Jawab:**

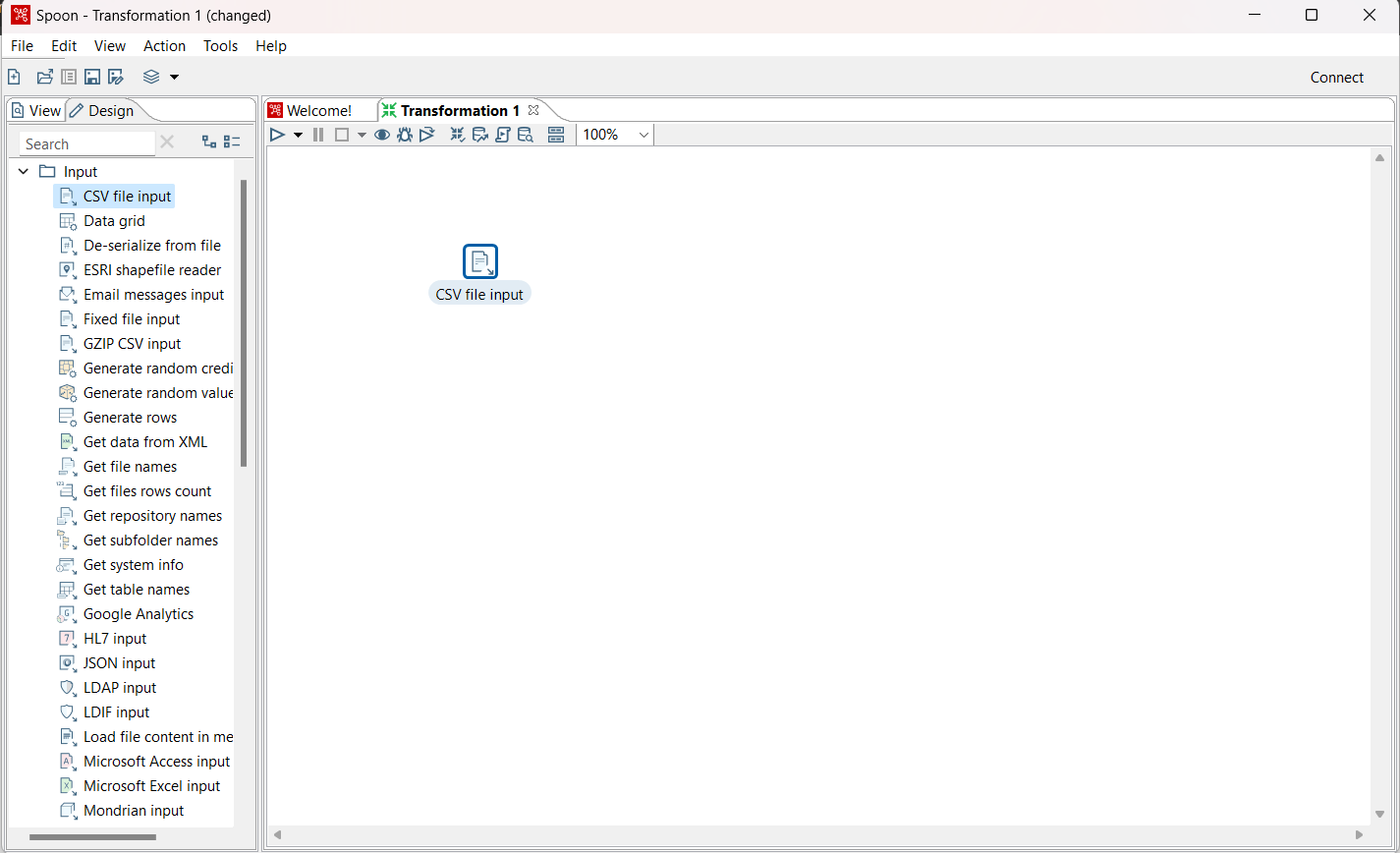
Tidak ada, karena semua data bertipe yang sama pada kolom yang sama.

## **Pengambilan Data (Extract)**

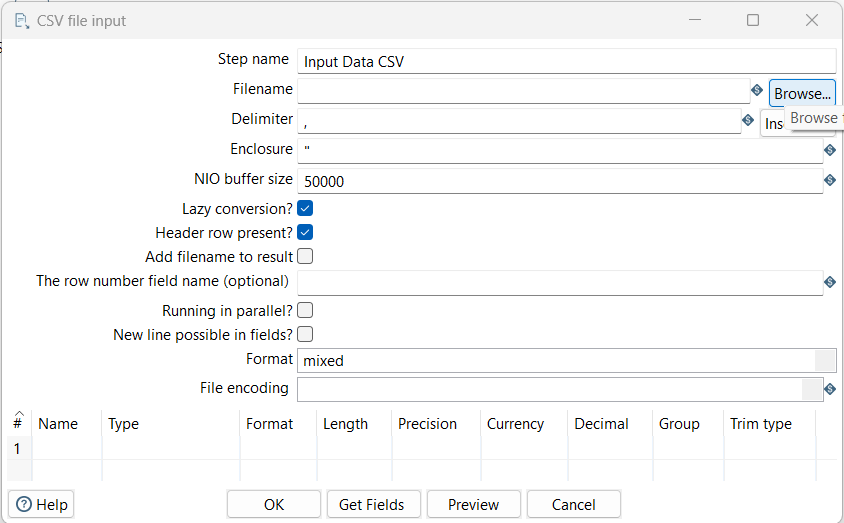
1. Buka lembar kerja **Transformation** melalui toolbar **File**.

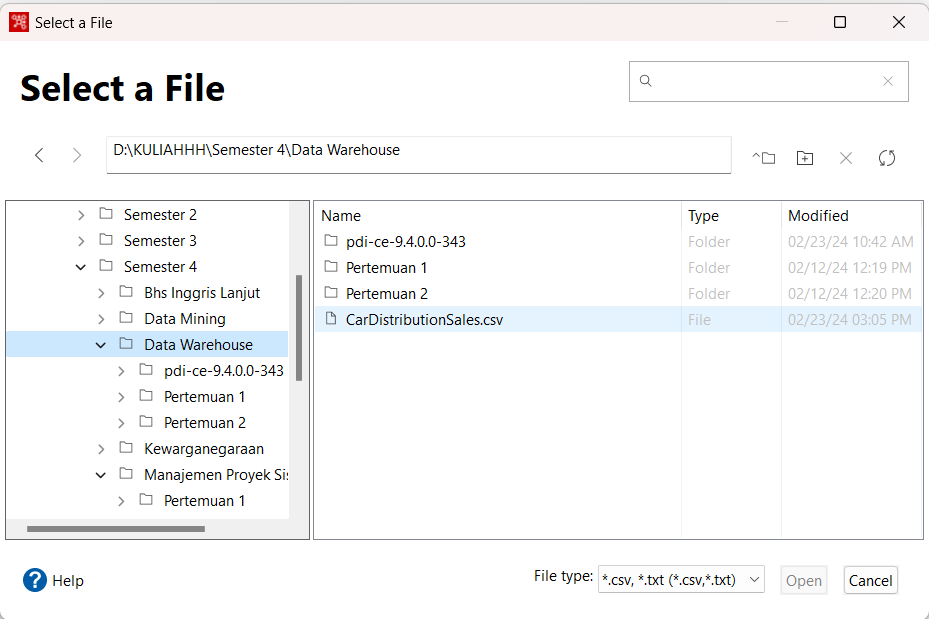


1. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area.**

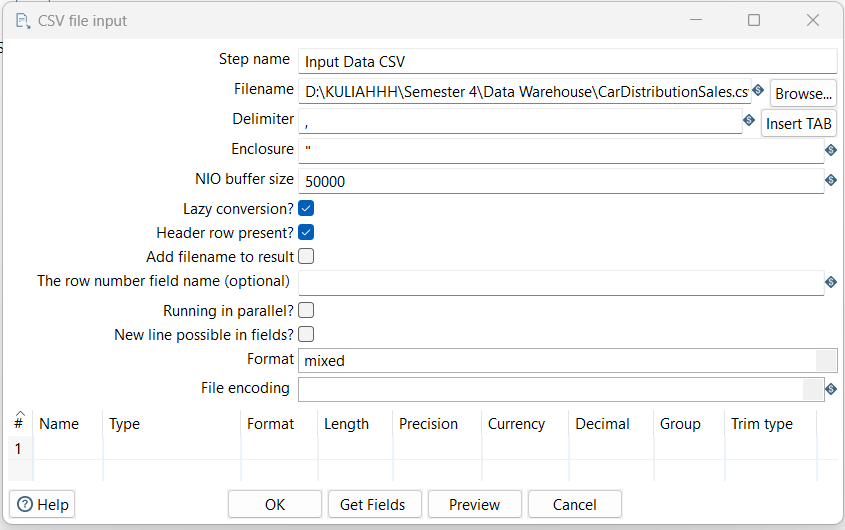


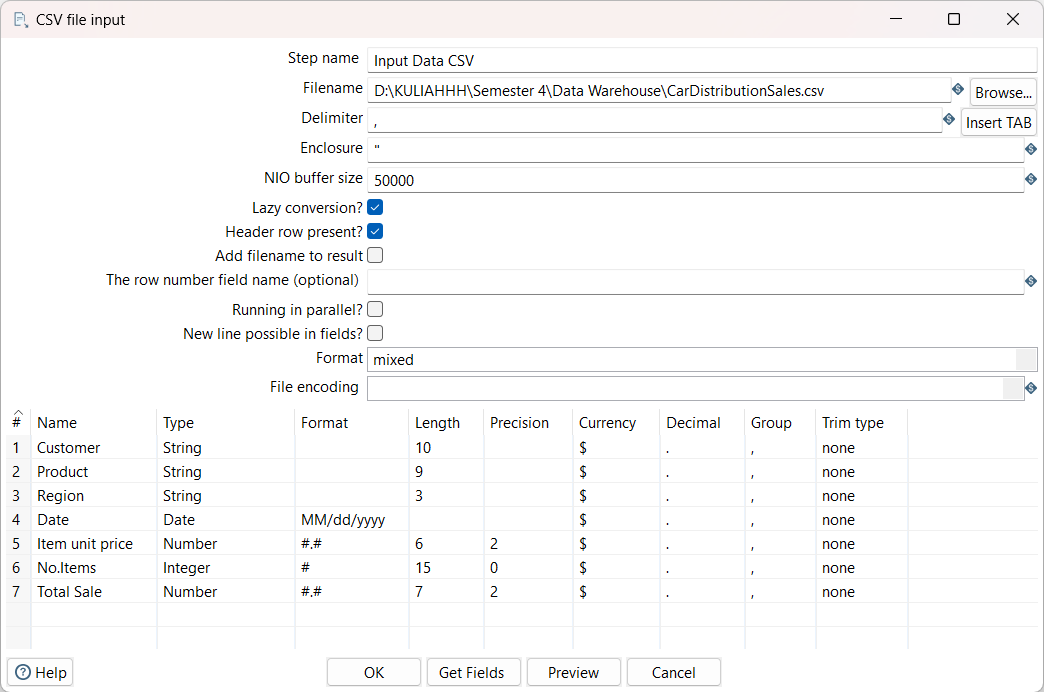
1. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasinya.
2. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
3. Pilih lokasi file **CarDistributionSales.csv** melalui **Browse** pada **filename**.



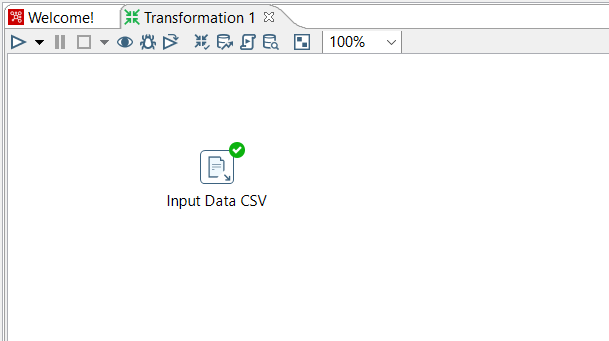


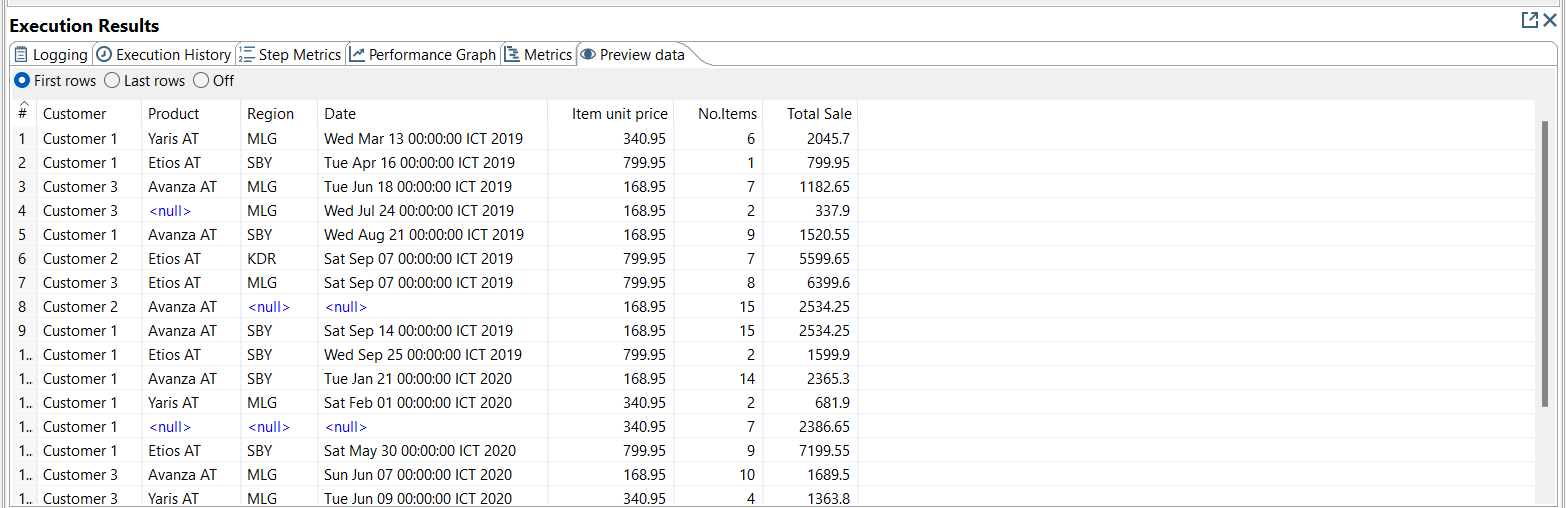
1. Jika data telah dipilih pastikan pada bagian **Delimiter** sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).
2. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai



****

1. Tutup jendela konfigurasi tersebut.
2. Tekan tombol "**Run**" pada pojok kanan atas **Area Kerja**, makan akan muncul **Execution Result Area**. Pilih tab **Prieview Data** pada **Execution Result Area**. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.

****

****

Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.

### **TUGAS 2**

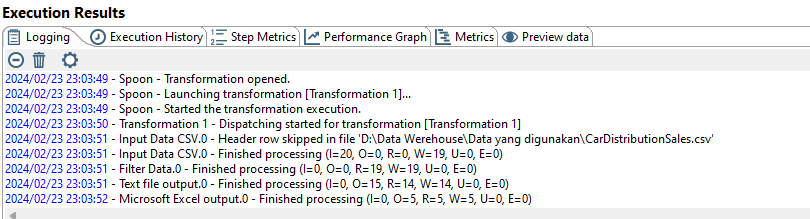
1. Apakah data hasil eksekusi sesuai dengan data aslinya?

**Jawab:**

Ya sesuai dengan data aslinya, namun terdapat data yang nilai nya null/kosong.

1. 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?

**Jawab:**



* + Spoon – Transformation opened: Proses dimulai dengan membuka transformasi dengan nama[Transformation 1]
  + Spoon - Launching transformation [Transformation 1]: PDI Spoon memulai transformasi dengan nama [Transformation 1].
  + Spoon - Started the transformation execution: Eksekusi transformasi dimulai.
  + Transformation 1 - Dispatching started for transformation [Transformation 1]: PDI Spoon memulai dispatching untuk transformasi [Transformation 1].
  + Input Data CSV.0 - Header row skipped in file 'D:\Data Werehouse\Data yang digunakan\CarDistributionSales.csv': Pada file CSV, PDI Spoon mengabaikan baris header.
  + Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0): Proses ekstraksi data dari file CSV selesai dengan input 20 baris dan output 0 baris.
  + Filter Data.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0): Proses filtering data selesai dengan input 0 baris dan output 0 baris, serta 19 baris direject.
  + Text file output.0 - Finished processing (I=0, O=15, R=14, W=14, U=0, E=0): PDI Spoon mengekspor hasil transformasi ke file teks dengan 15 baris output, 14 berhasil direkam, dan 14 berhasil ditulis.
  + Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0): PDI Spoon juga mengekspor hasil transformasi ke file Excel dengan 5 baris output, 5 berhasil direkam, dan 5 berhasil ditulis.
  + Spoon - The transformation has finished!!: Transformasi secara keseluruhan telah selesai.

1. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

**Jawab:**

https://content.gitbook.com/content/dm7Yn6WBG5LswGYao3R5/blobs/J1ClK7qle1dmasjx1VZv/image.png

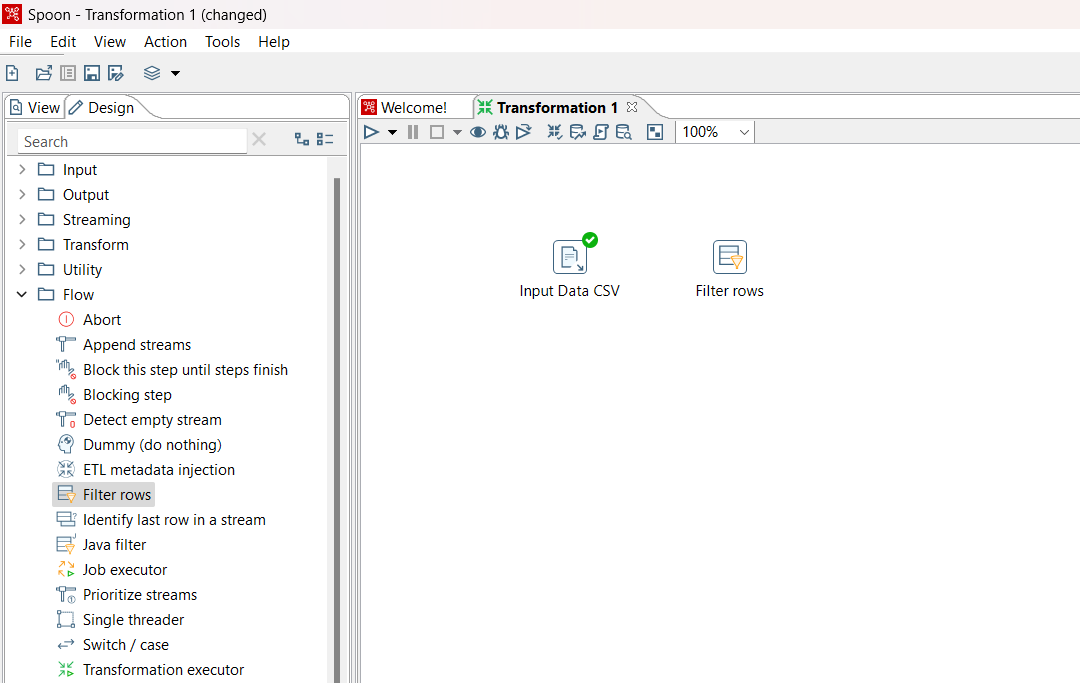
* + I (Input): Merupakan jumlah baris atau data yang masuk ke dalam suatu proses atau langkah tertentu.
  + O (Output): Merupakan jumlah baris atau data yang dihasilkan oleh suatu proses atau langkah tertentu.
  + R (Read): Merupakan jumlah baris atau data yang berhasil dibaca atau diambil dari sumber data.
  + W (Write): Merupakan jumlah baris atau data yang berhasil ditulis atau disimpan ke destinasi atau output.
  + U (Update): Merupakan jumlah baris atau data yang berhasil diperbarui dalam proses.
  + E (Error): Merupakan jumlah kesalahan atau error yang terjadi selama proses atau langkah tertentu.

1. 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.

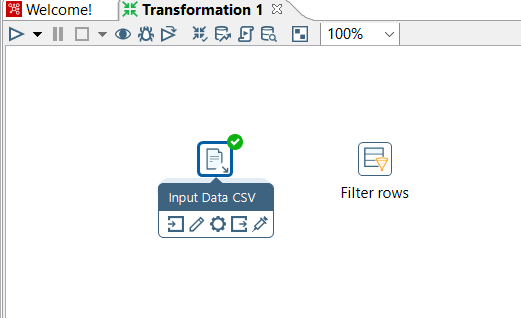
**Jawab:**

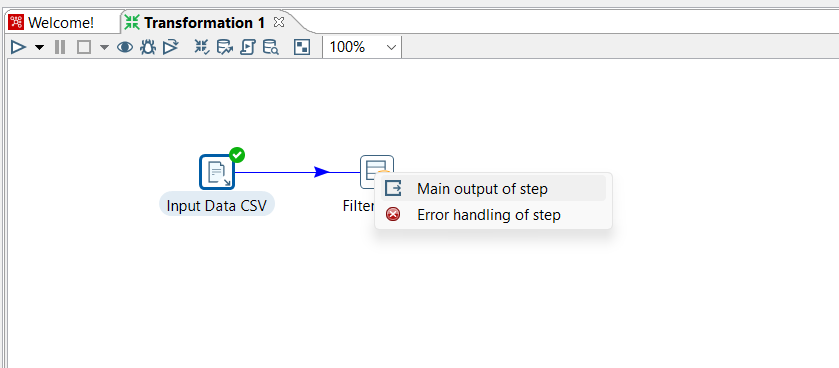
## **Filter Data (Transform) dan Pengemasan data (Load)**

* + 1. Carilah objek **Filter rows** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

****

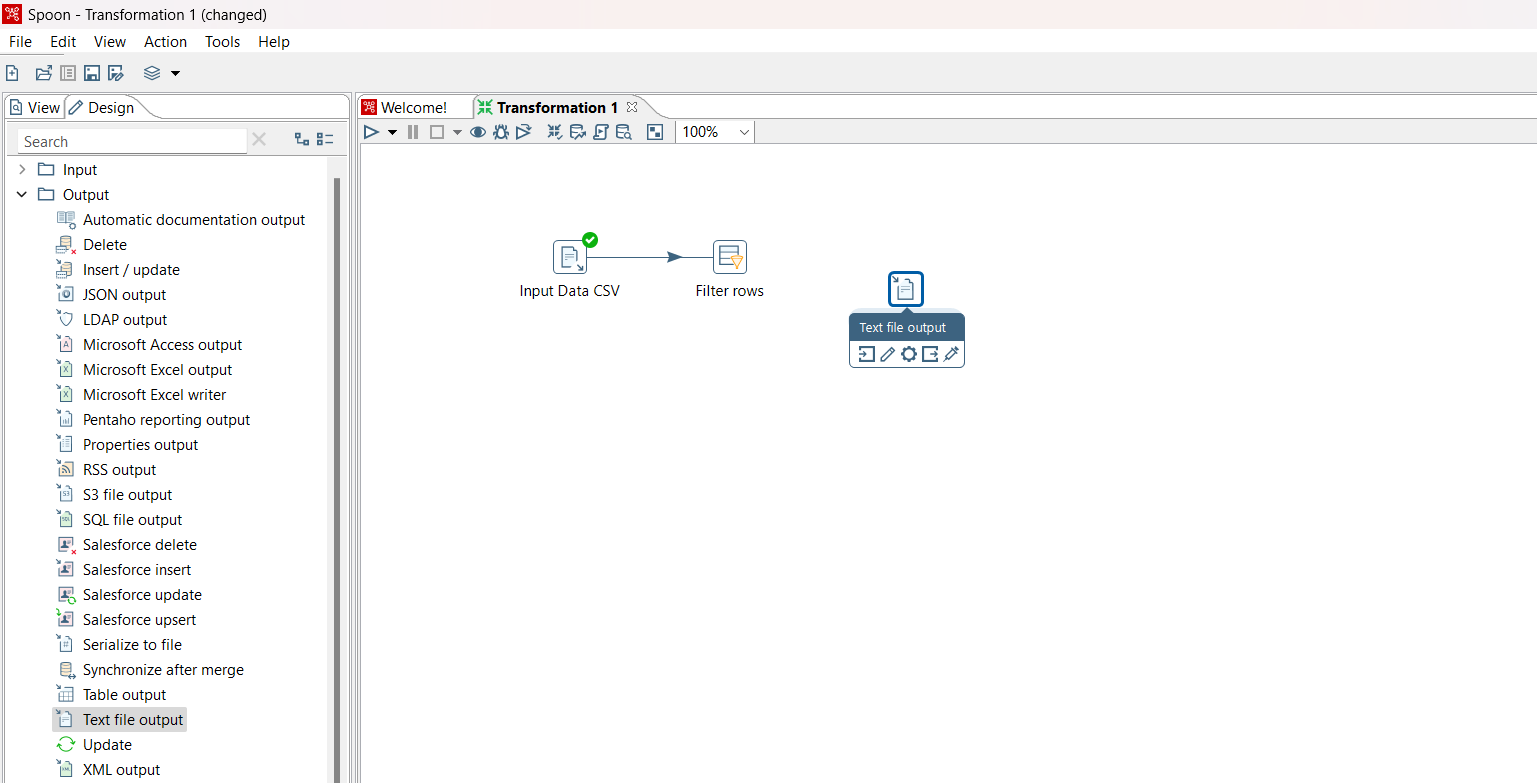
1. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.

****

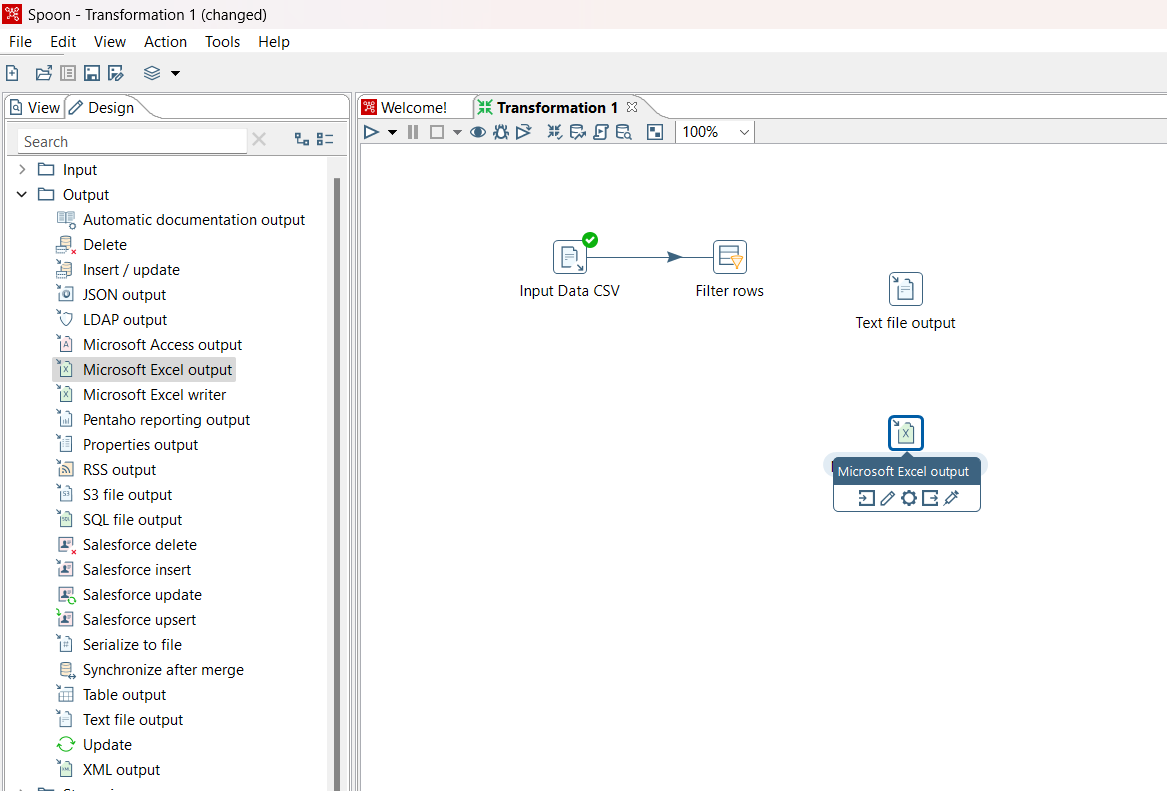
****

Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows.

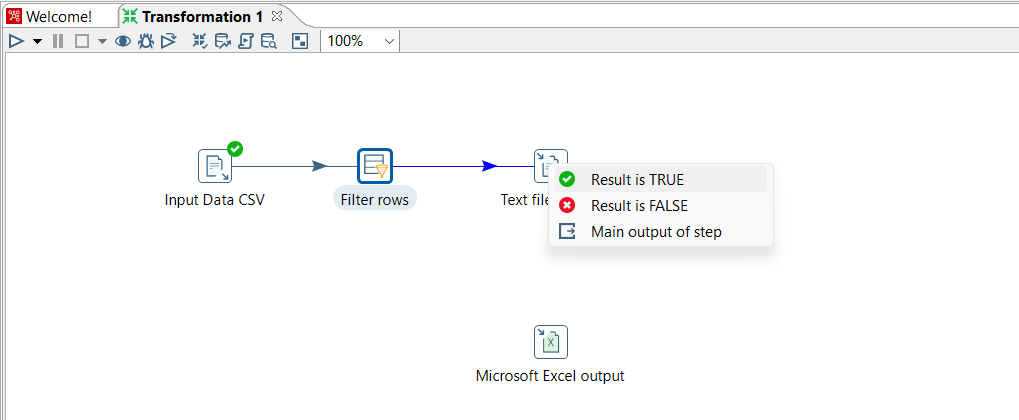
1. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

****

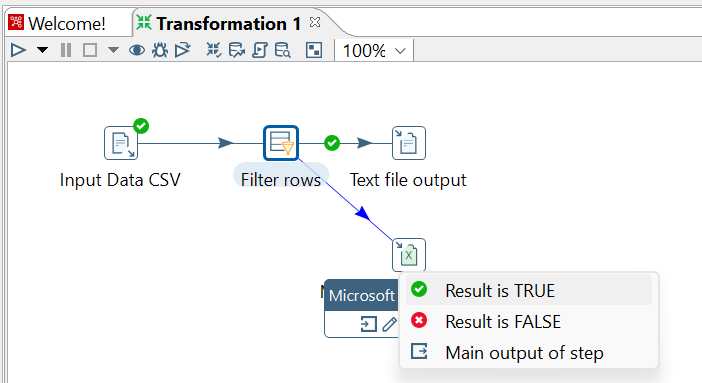
1. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

****

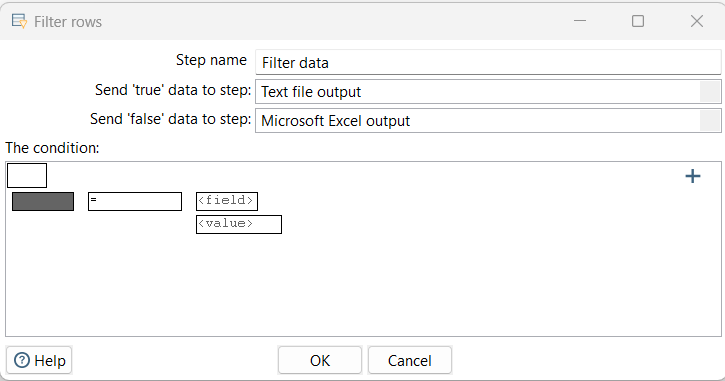
1. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagi jenis konektornya.

****

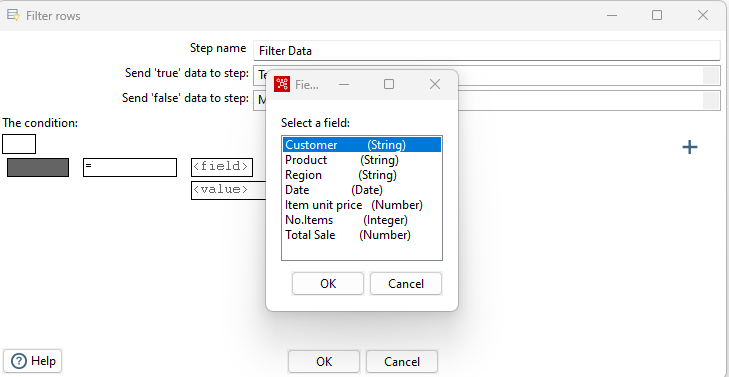
1. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagi jenis konektornya.



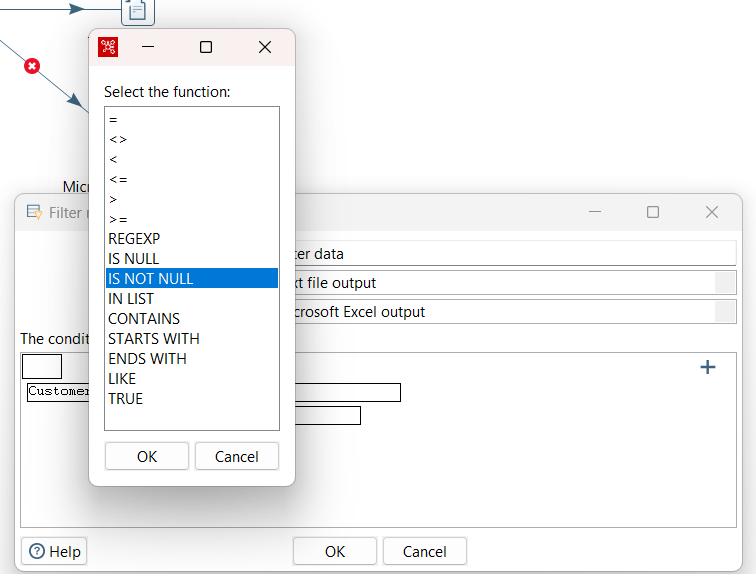
1. **Double click** pada **filter rows** hingga muncul jendela konfigurasi**.** ubah **step name** menjadi **Filter data**.
2. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
3. Klik **<Field>** pada bagian **condition**

****

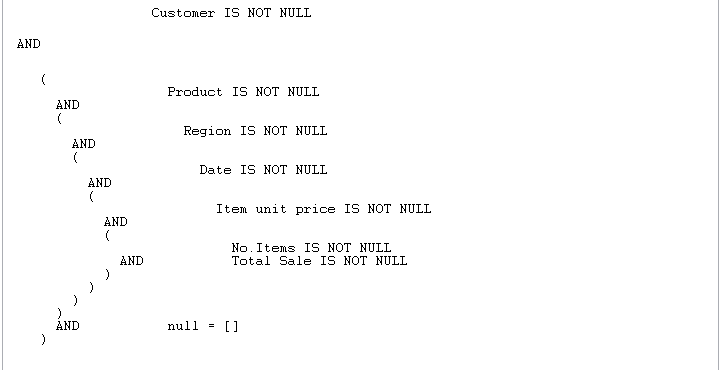
1. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer,** kemudian klik **OK.**



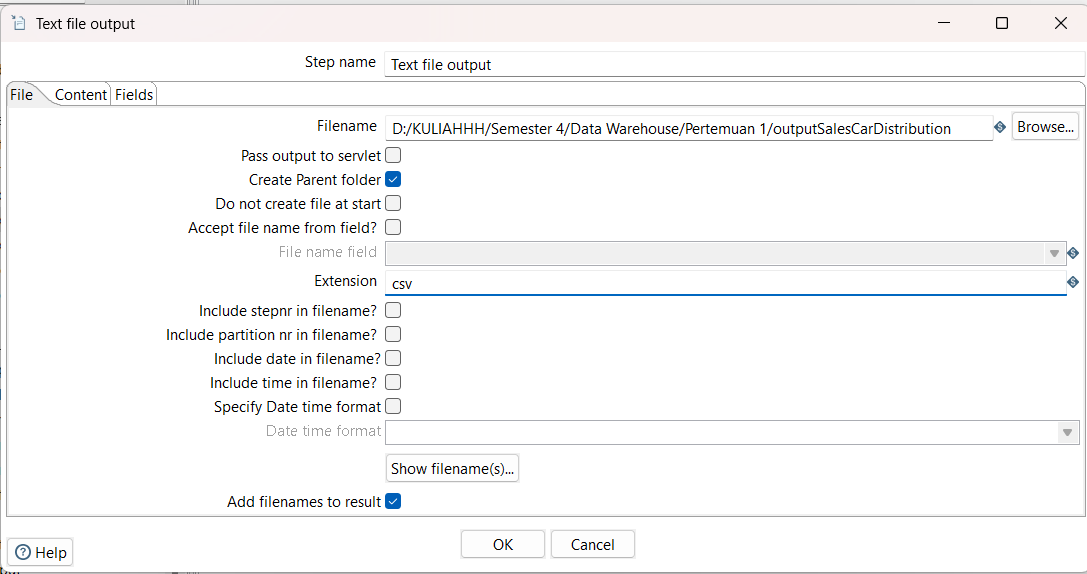
1. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null.

****

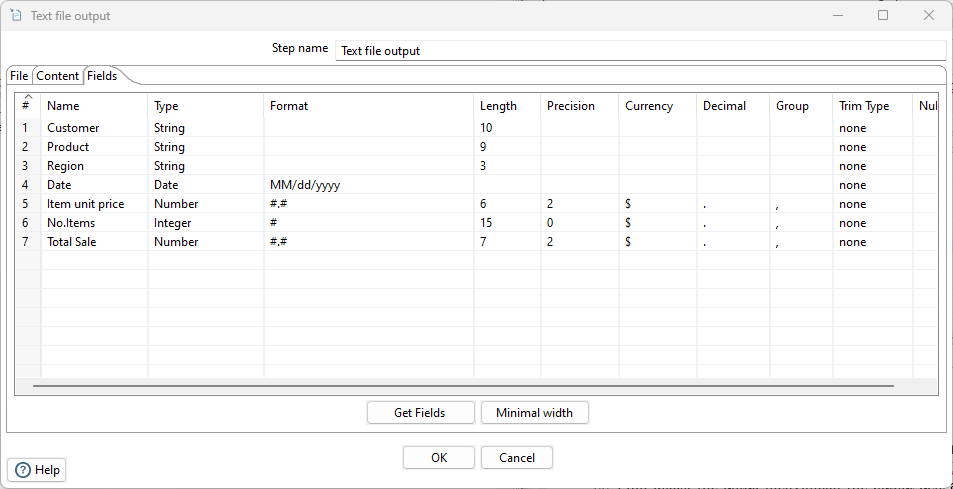
1. Tekan tombol **+** untuk menambah kondisi.
2. Pilih **statement** yang masih null untuk menambah kondisi.
3. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
4. Gunakan operator **AND** untuk menghubung
5. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.
6. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.

****

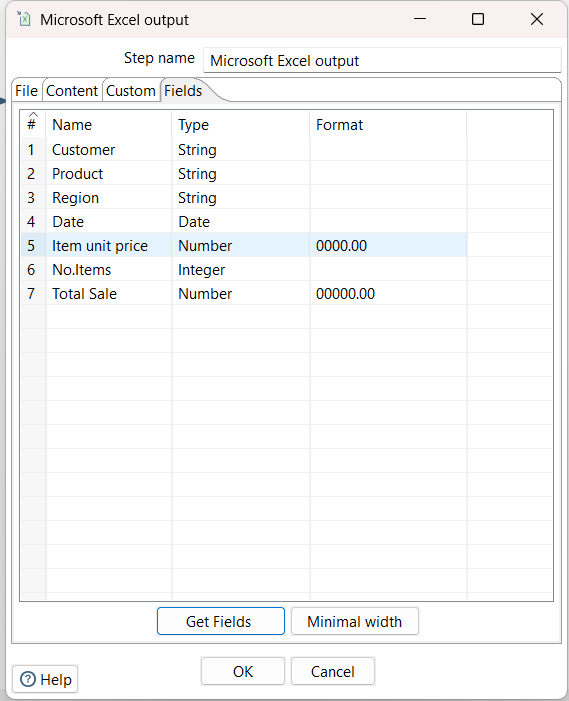
1. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**.
2. Ubah extension menjadi **csv**.

****

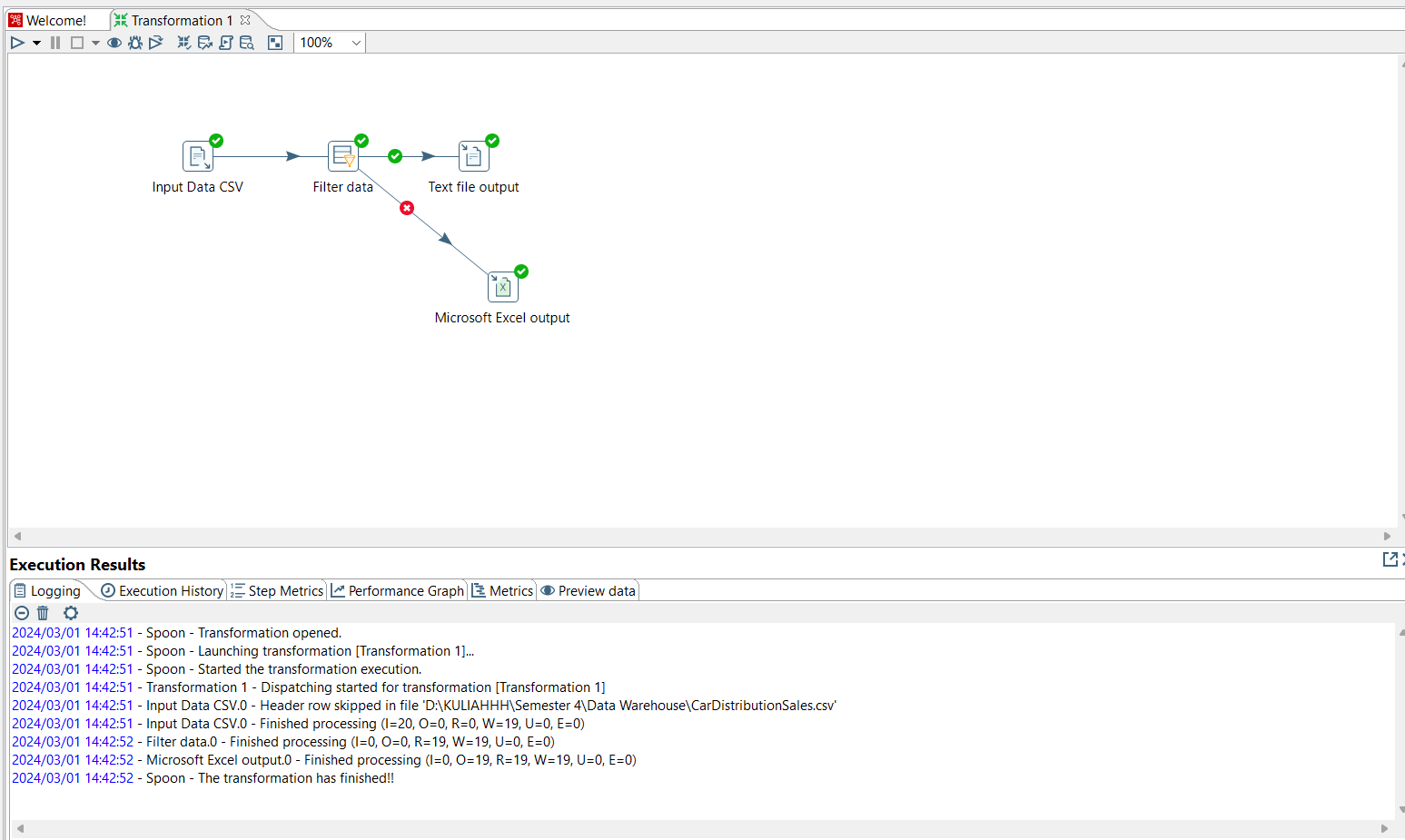
1. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.



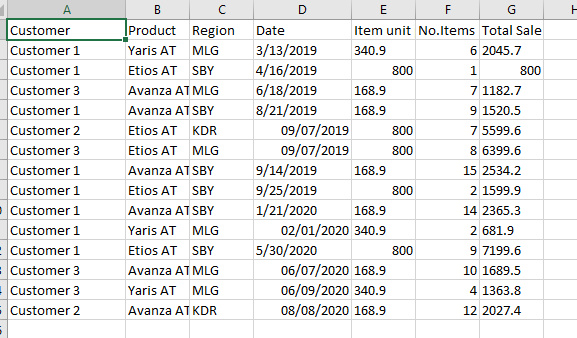
1. Tekan **ok** untuk kembali ke **work area**
2. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output, Double-click** pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.
3. Pada tab **fields** tekan tombol **Get fields** untuk mengambil field data.
4. Tekoan **ok** untuk kembali ke **work area**.

****

1. Tekan tombul **Run** pada pojok kiri atas **Work Area**. jika berhasil maka akan muncul tanda centang hijau di setiap proses.

****

Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.



Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.

### **TUGAS 3**

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

**Jawab:**

* File CSV (Comma-Separated Values) menggunakan tanda koma sebagai pemisah antar kolom. Data disimpan dalam format teks sederhana tanpa formatasi khusus.
* File Excel memiliki format yang lebih kompleks, menyertakan opsi format, formula, dan format sel. Data di Excel dapat diatur dengan lebih rapi, dan file memiliki ekstensi .xlsx atau .xls.
  + 1. Jelaskan apa yang terjadi pada proses Filter rows!

**Jawab:**

* Proses Filter Rows digunakan untuk menyaring atau memfilter baris data berdasarkan suatu kondisi tertentu.
* Pada log, terlihat bahwa proses ini selesai dengan 0 input, 0 output, 19 baris direject, dan 19 baris yang diperlakukan. Mungkin ada kondisi filter yang tidak memenuhi untuk 19 baris data tersebut.
  + 1. 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.

**Jawab:**

## **Studi Kasus**

**KASUS:**

Pak Pascanowo seorang ketua RW di Ibu Kota Nusabangsa (IKN). Terdapat 20 orang dari negara Konoha yang akan pindah dan berdomisili di IKN. Tata letak pemukiman pada IKN telah dikondisikan agar berkelompok disesuaikan dengan lokasi pekerjaannya agar akses terjangkau.

Sebagai contoh, mahasiswa akan dikumpulkan dan berdomisili di wilayah edukasi yang akses dekat dengan kampus-kampus perguruan tinggi. Sedangkan dokter akan berdomisili dekat dengan rumah sakit atau layanan kesehatan dan PNS akan berdomisili dekat dengan perkantoran dan layanan publik. Selain itu, orang yang bukan dari ketiga kategori tersebut bebas memilih domisili atau tempat tinggalnya di IKN.

Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. darai data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok,

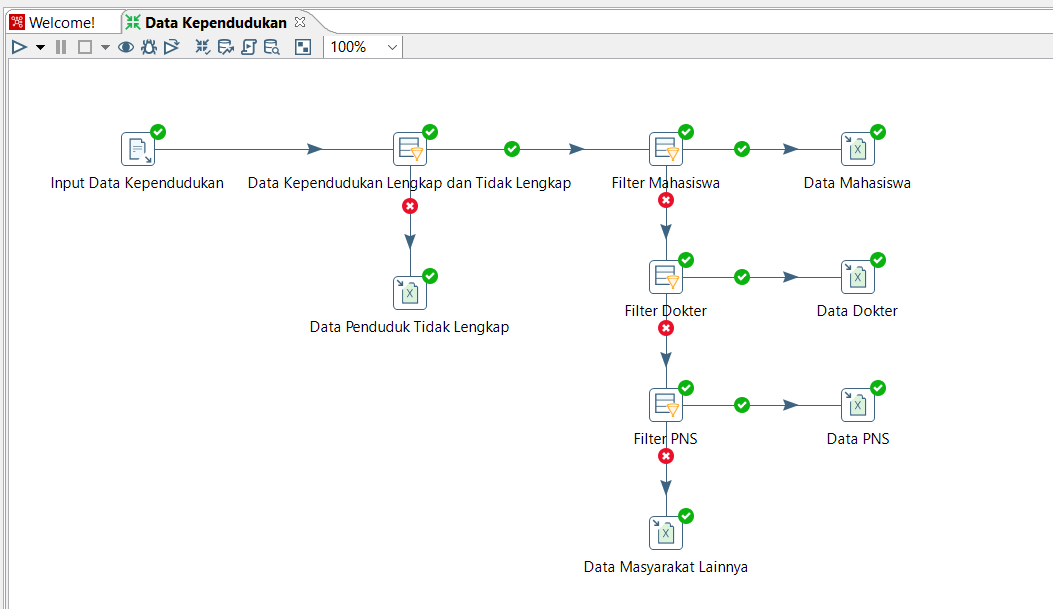
* + **Data masyarakat berprofesi mahasiswa** akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah pendidikan.
  + **Data masyarakat berprofesi PNS** akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan
  + **Data masyarakat berprofesi Dokter** akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

data-data tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

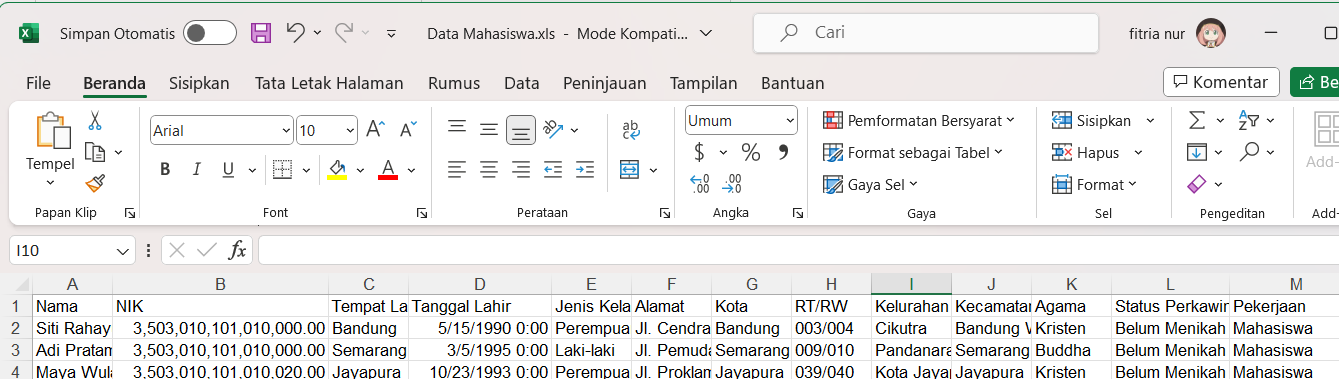
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:

* + MasyarakatMahasiswa
  + MasyarakatPNS
  + MasyarakatDokter
  + MasyarakatLainnya.

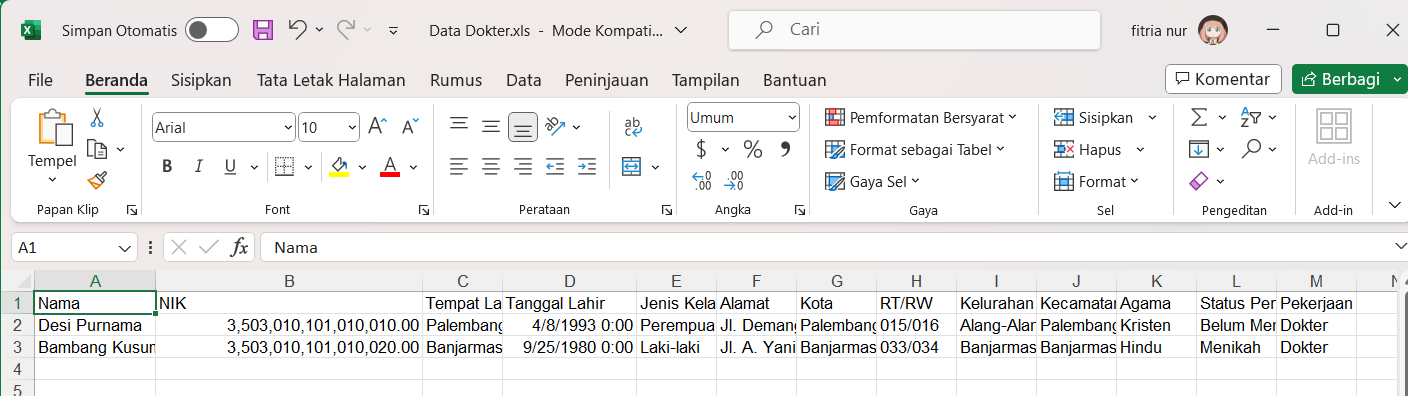
Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.

****

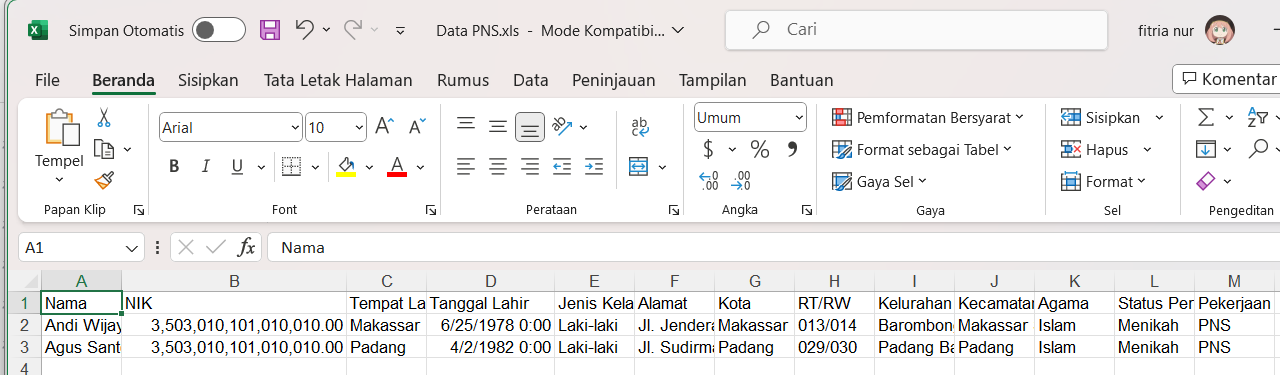
* **Data mahasiswa**



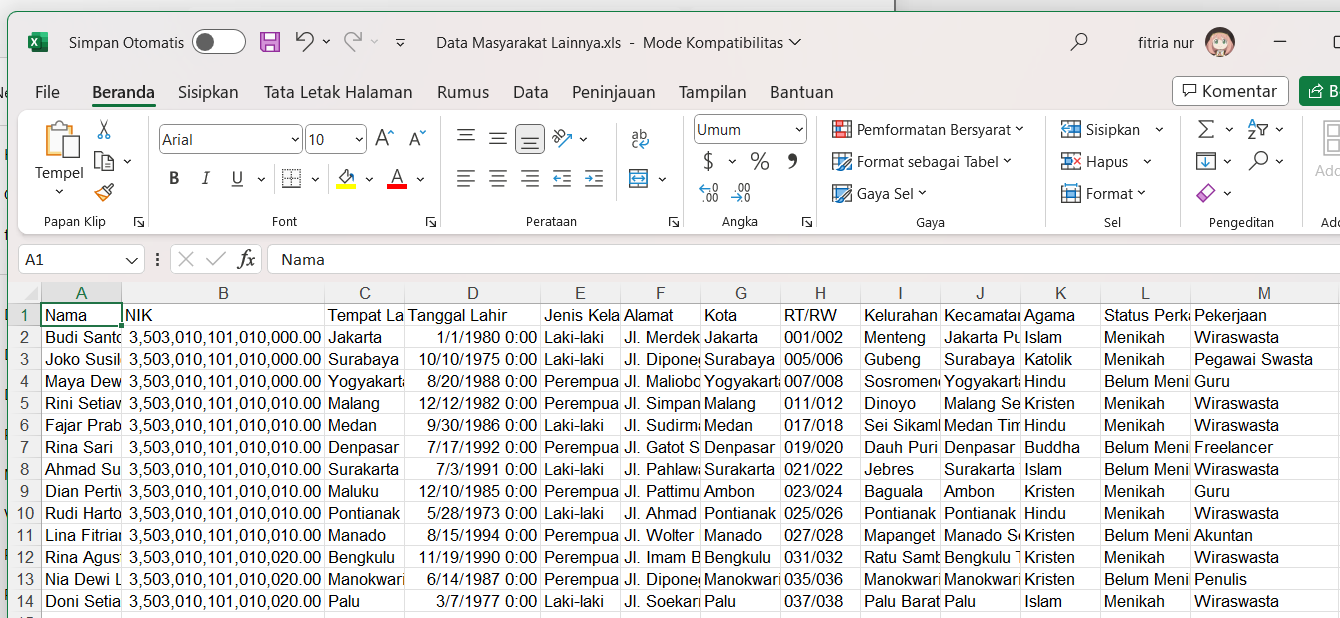
* **Data Dokter**



* **Data PNS**



* **Data Masyarakat Lainnya**

****