

2241720081

FADHLUROHMAN AL FARABI

Daftar Isi

[Pendahuluan Error! Bookmark not defined.](#_Toc98155632)

[pengambilan data Error! Bookmark not defined.](#_Toc98155632)

[filter data dan pengemasan data Error! Bookmark not defined.](#_Toc98155632)

[studi kasus Error! Bookmark not defined.](#_Toc98155632)

# PENDAHULUAN

### 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.

\*Data penjualan tersebut dapat diakses dan di download melalui link berikut :

<https://github.com/dik4rizky/datasources/blob/7e3f70f516a1b107635141d09862a8c30fbb6812/CarDistributionSales.csv>



Gambar I. Data Penjualan Mobil

### TUGAS 1

Analisa data tersebut!

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 ?

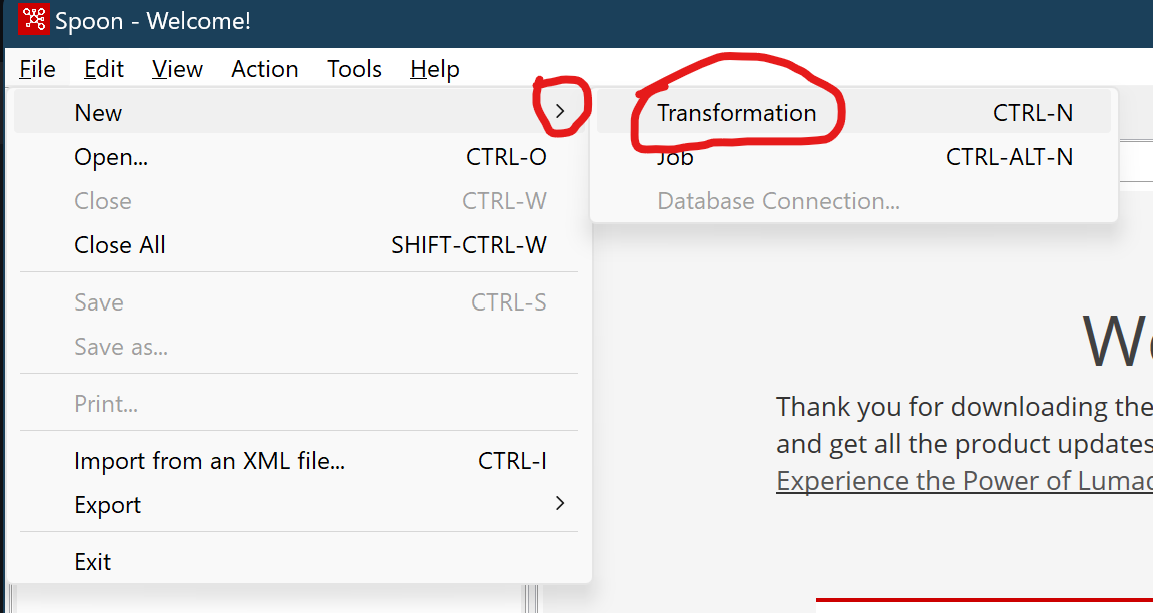
Jawaban

1. Jumlah kolom pada data diatas terdapat 7 buah. Dengan pembagian yaitu kolom customer, product, region, date, item unit price, No.items, total sale.
2. Data dari setiap kolom tersebut memiliki arti dan maksudnya masing-masing. Berikut penjelasannya
   1. Pada kolom customer mewakili daftar customer yang melakukan pembelian mobil tersebut.
   2. Pada kolom product mewakili daftar product yang dibeli oleh customer.
   3. Pada kolom region mewakili tempat dilakukannya pembelian ataupun wilayah terjadinya transaksi tersebut.
   4. Pada kolom date mewakili tanggal terjadinya pembelian mobil tersebut.
   5. Pada kolom item unit price mewakili harga barang per unit.
   6. Pada kolom No.items mewakili jumlah yang terjual pada tanggal tersebut.
   7. Pada kolom total sale mewakili keuntungan yang didapat dari hasil perhitungan antara harga per item dan jumlah yang terjual.
3. Dari data tersebut didapat beberapa data pada kolom yang tidak lengkap seperti pada kolom product, region dan date.
4. Berdasarkan data diatas tidak adanya perbedaan tipe data di kolom yang sama.

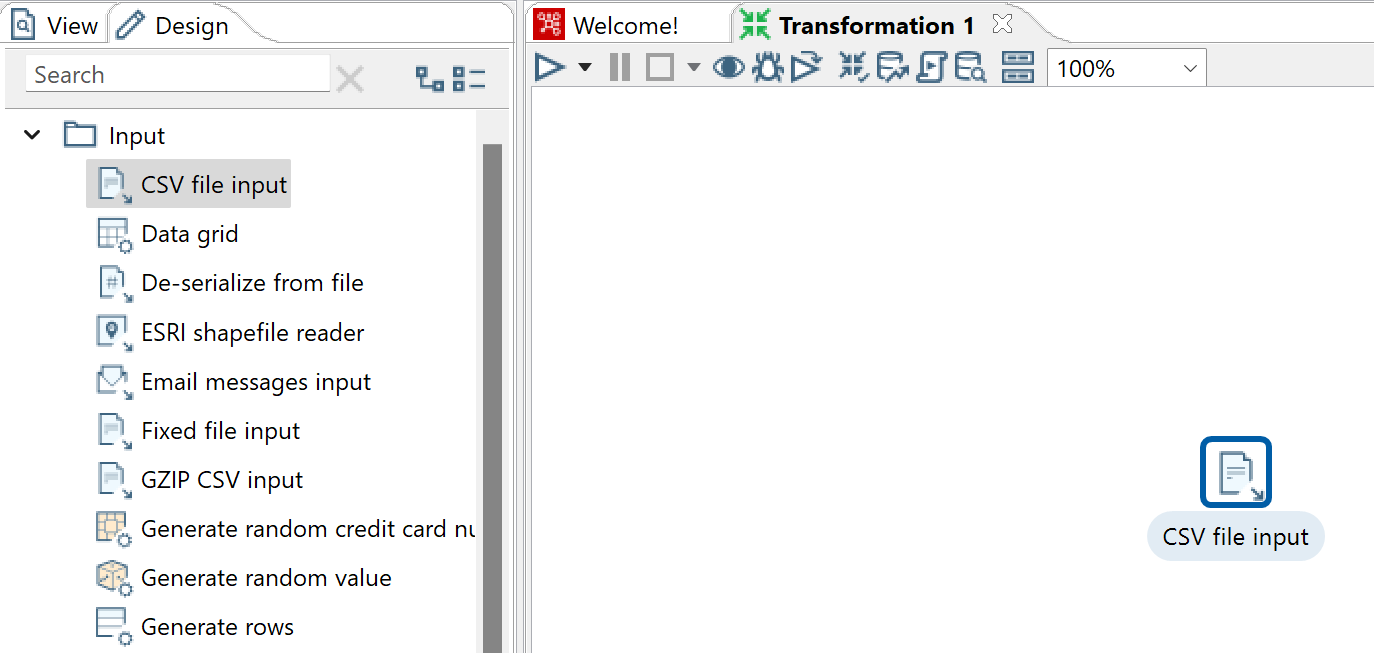
# PENGAMBILAN DATA

### Langkah-langkah

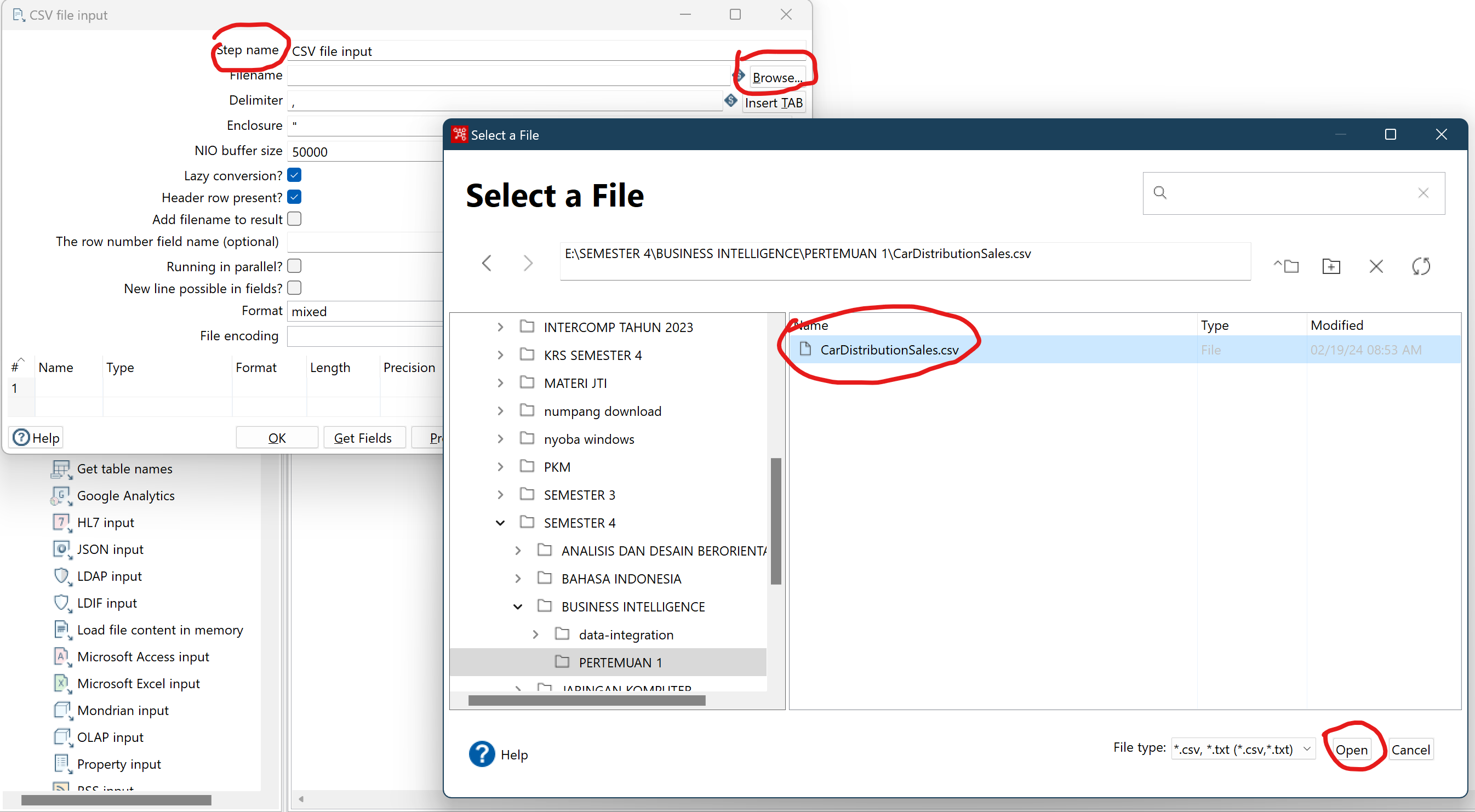
1. Buka lembar kerja **Transformation** melalui toolbar **File.**



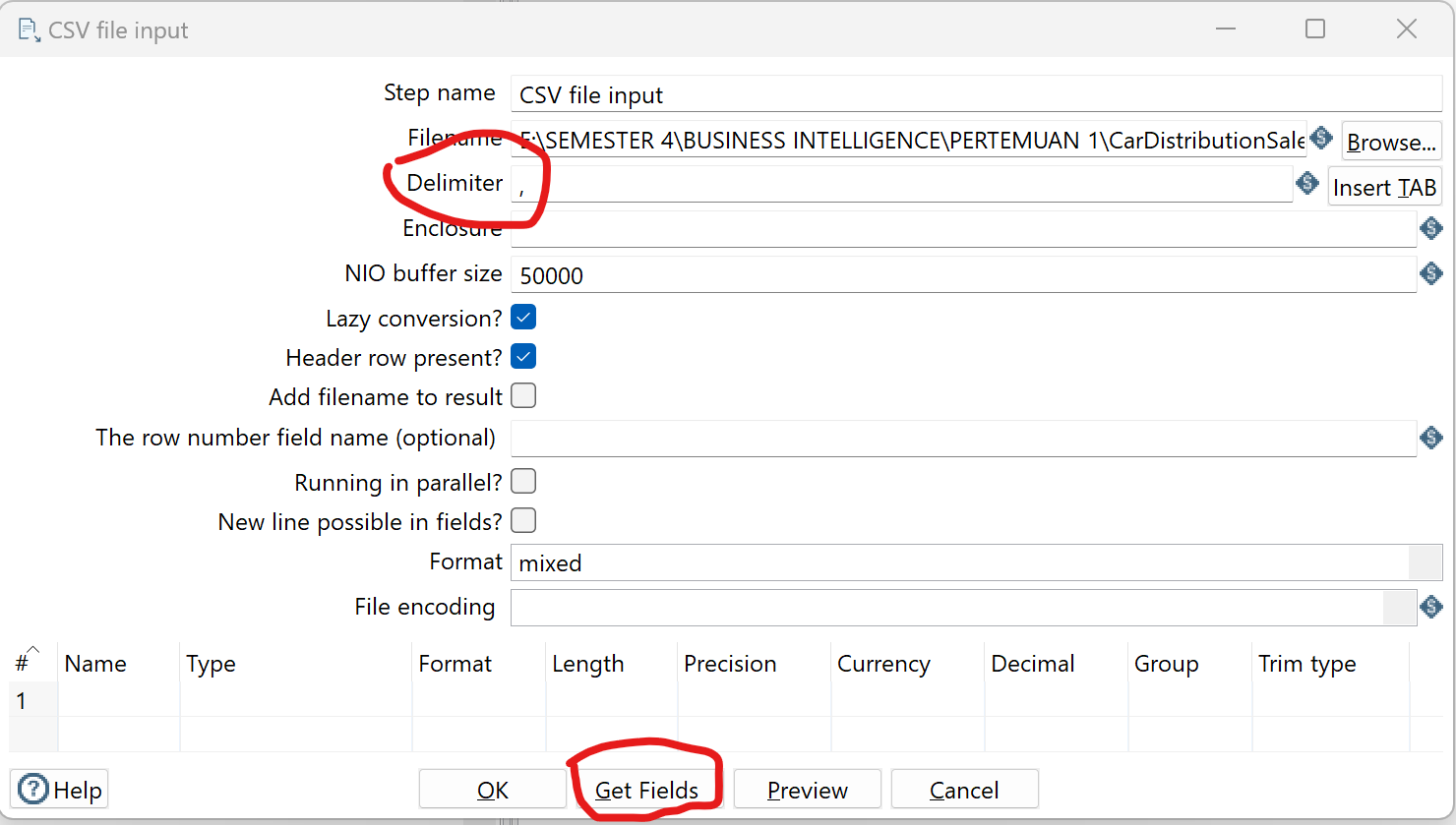
1. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area.**

****

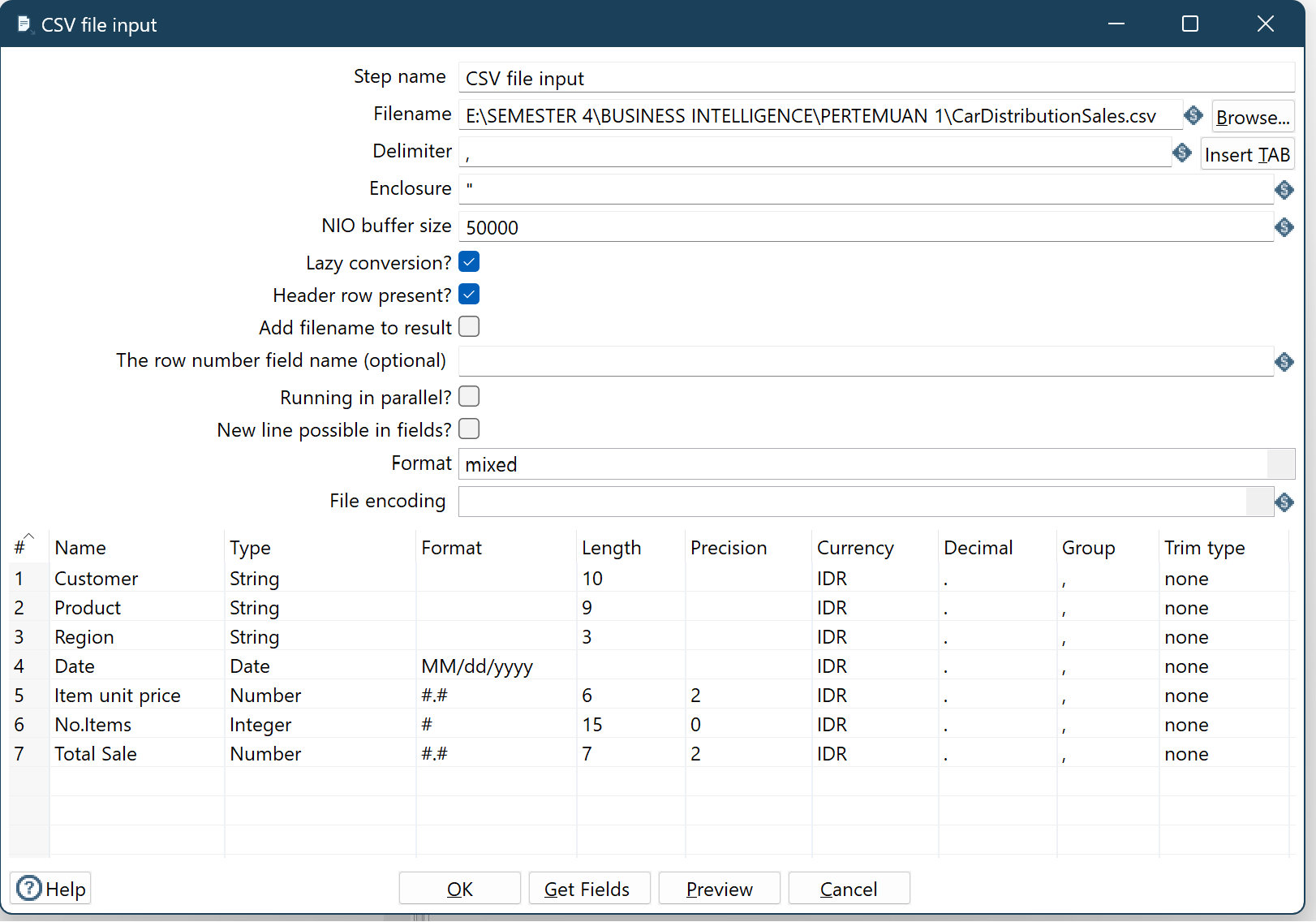
1. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasinya.
2. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
3. Pilih lokasi file **CarDistributionSales.csv** melalui **Browse** pada **filename**.



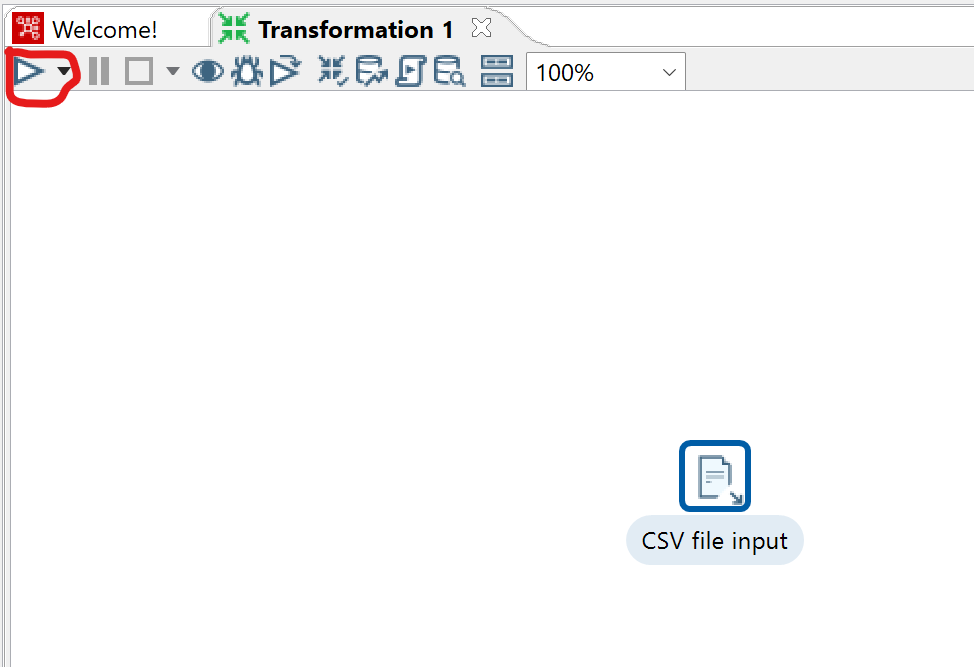
1. Jika data telah dipilih pastikan pada bagian **Delimiter** sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).

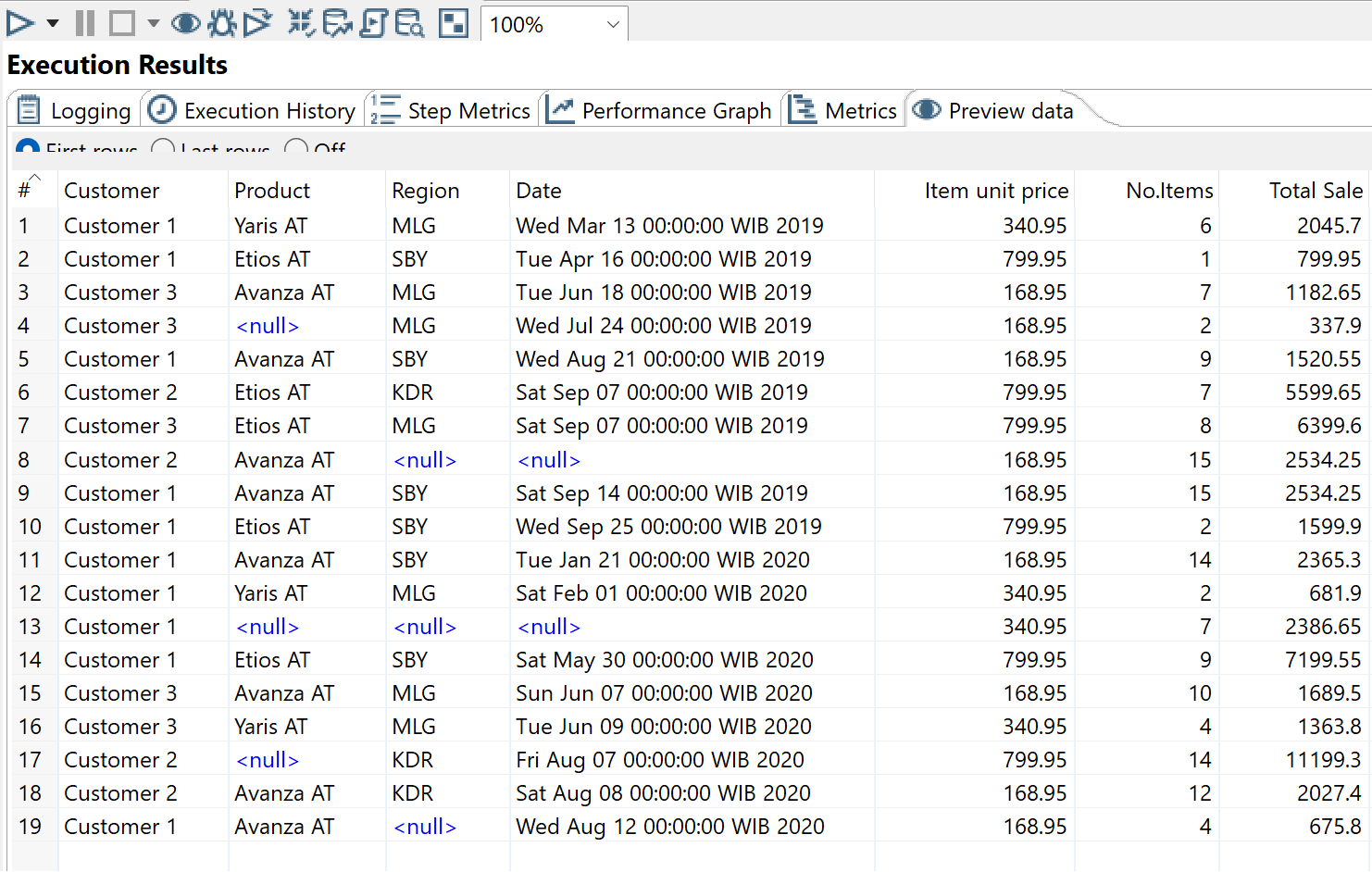


1. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai.



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.

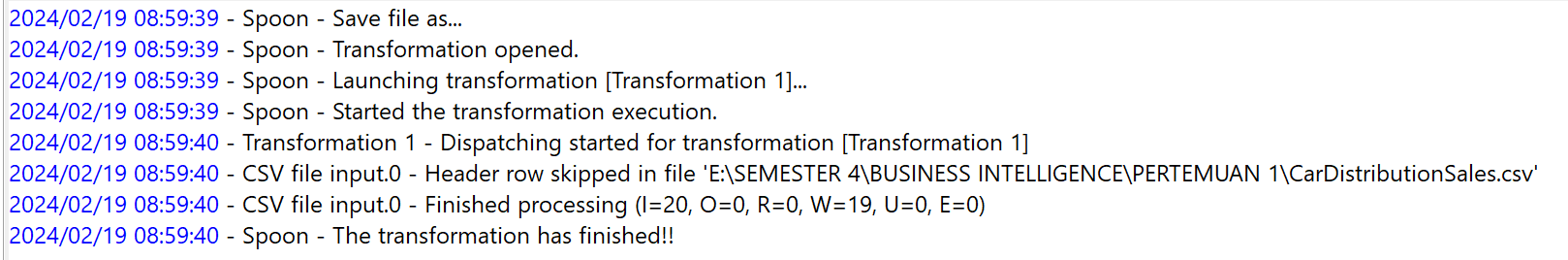




### TUGAS 2

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 ? 
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. Data hasil eksekusi sama dengan data aslinya tidak ada perbedaan.
2. 

Berdasarkan gambar diatas proses yang dilakukan adalah :

1. Proses dimulai dengan adanya penyimpanan file terlebih dahulu.
2. Setelah itu, dilanjutkan dengan **transformasi** dibuka dalam spoon.
3. Setelah itu mulai menjalankan transformasi yang disebut “Transformasi 1”.
4. Lalu, proses eksekusi transformasi dimulai.
5. Setelah itu, pengiriman dimulai untuk transformasi yang disebut “Transformasi 1”.
6. Setelah itu, informasi bahwa baris judul dilewati dalam file CSV yang diberikan berdasarkan path file nya.
7. Proses input file CSV selesai, dengan informasi sebagai berikut :
   * I (Input) : 20 baris telah dibaca dari file.
   * O (Output) : 0 baris telah dihasilkan.
   * R (Rejected) : 0 baris ditolak.
   * W (Written) : 19 baris ditulis ke dalam hasil.
   * U (Updated) : 0 baris diperbarui.
   * E (Errors) : 0 Kesalahan terjadi.
8. Transformasi telah selesai dilakukan.

Proses diatas merupakan transformasi ETL yang mengambil data dari file CSV.

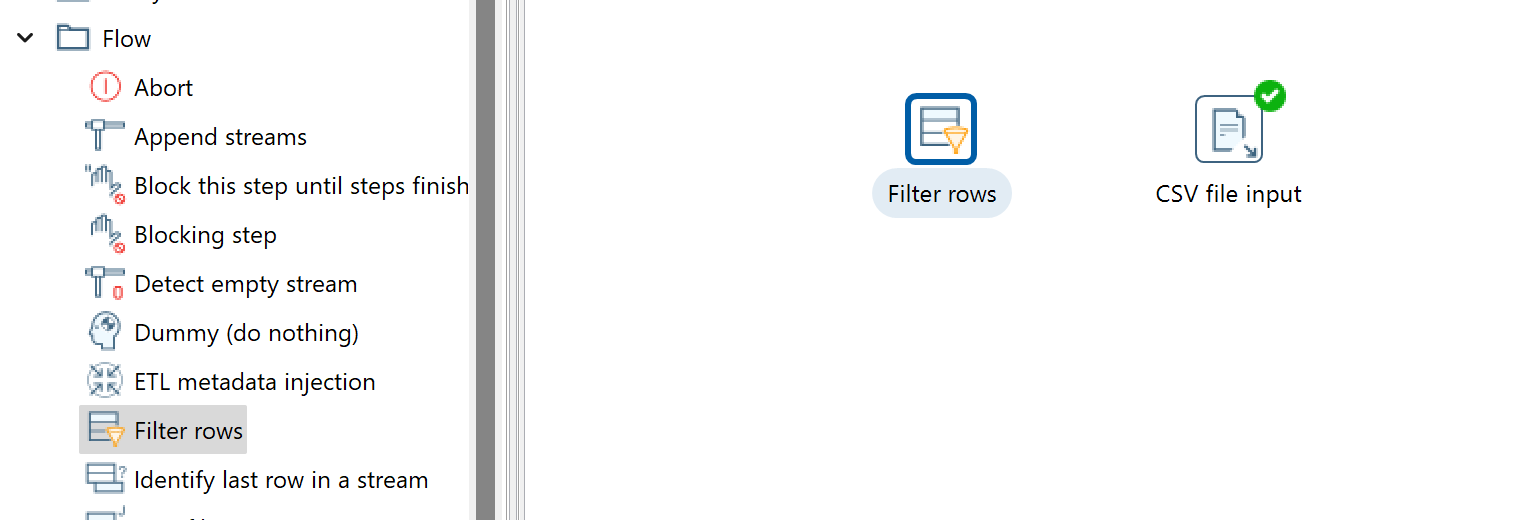
1. Informasi pada gambar tersebut merupakan informasi yang berisi tentang input (I), output (O), rejected (R), written (W), updated (U), dan error (E). Yang merupakan informasi yang didapatkan ketika telah selesai memproses transformasi ETL tersebut.
2. Berikut merupakan tabel yang menjelaskan waktu yang dibutuhkan untuk mengekstraksi data tersebut.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NO | Nama | Processor | RAM | Storage | VGA | CPU | Waktu |
| 1 | Fadhlurohman | Ryzen 7 | 16 GB | 512 GB | Radeon R8 | 16 | 0.0s |
| 2 | Aleron Tsaqif | Intel i5 | 8 GB | 512 GB | Intel Iris XE | 12 | 0.0s |
| 3 | M. Kemal | Intel i3 | 8 GB | 512 GB | Intel HD | 6 | 0.3s |
| 4 | Lalu Immaratul | Intel i7 | 16 G | 512 GB | Intel HD | 20 | 0.1s |

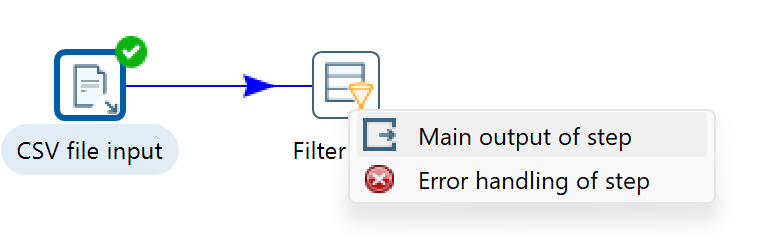
# FILTER DATA dan PENGEMASAN DATA

### Langkah-langkah

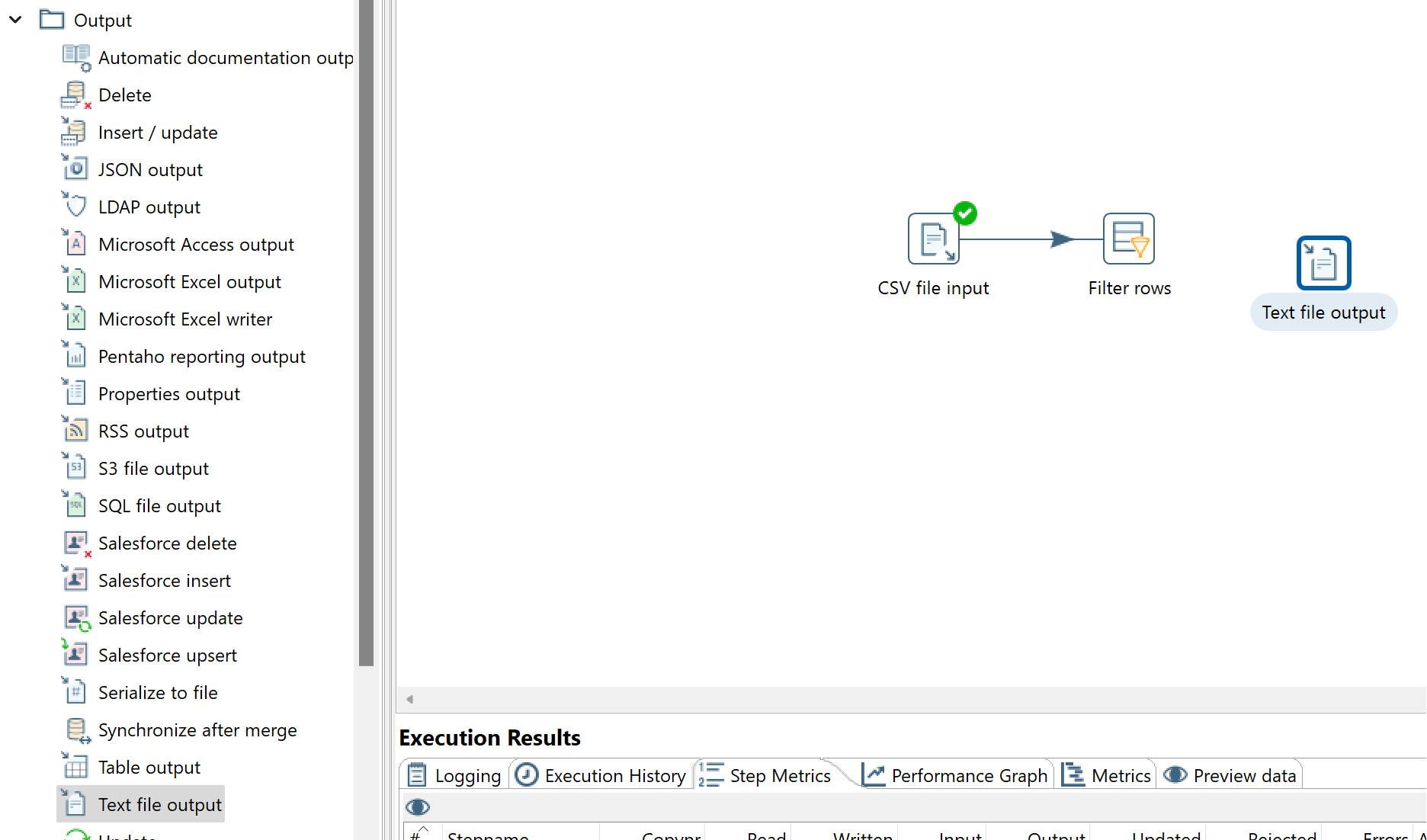
1. Carilah objek **Filter rows** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



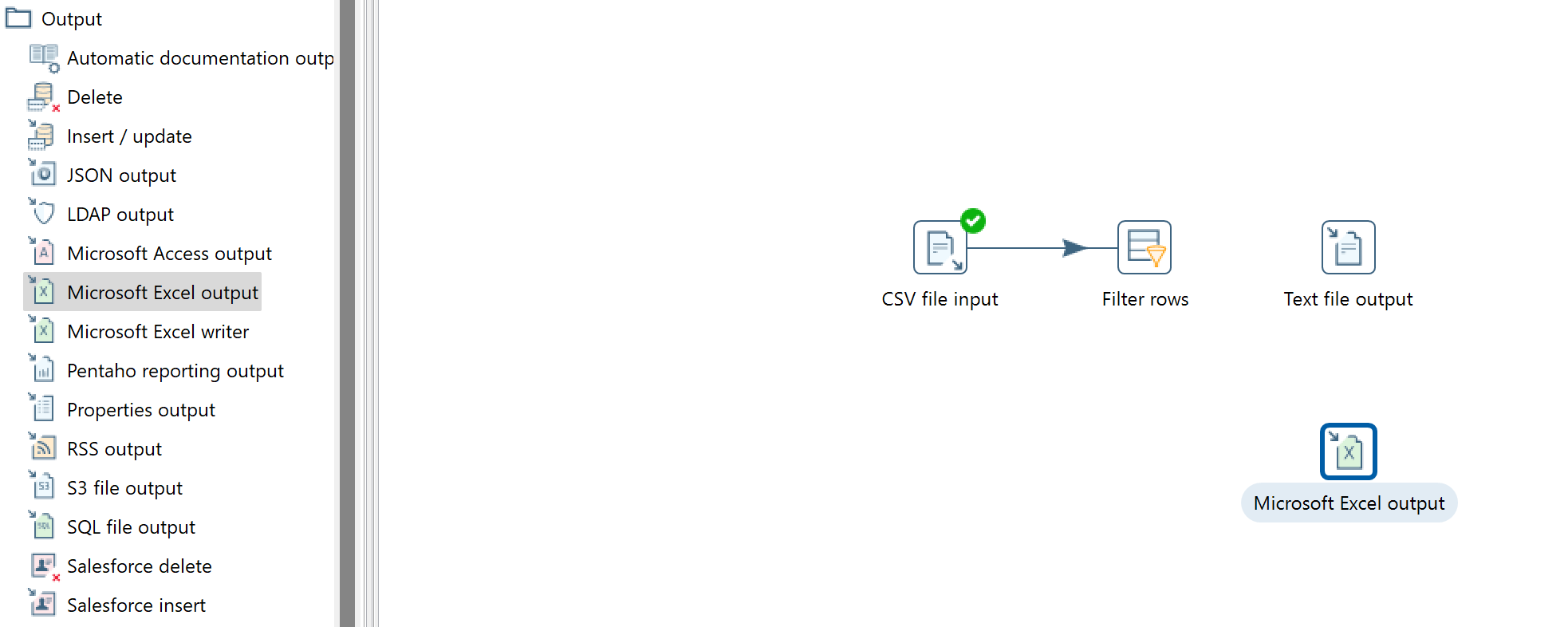
1. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.



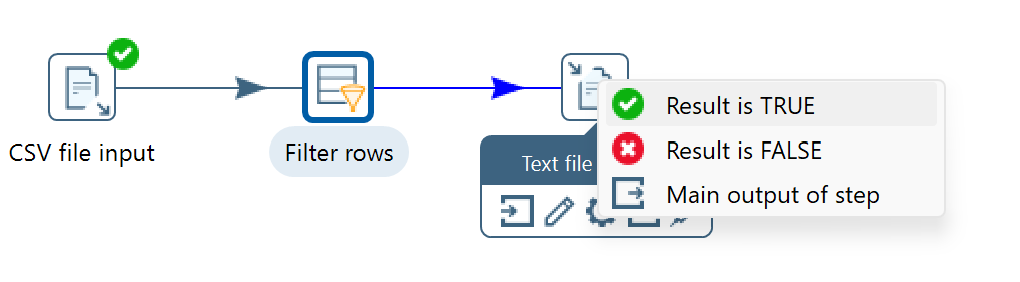
1. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



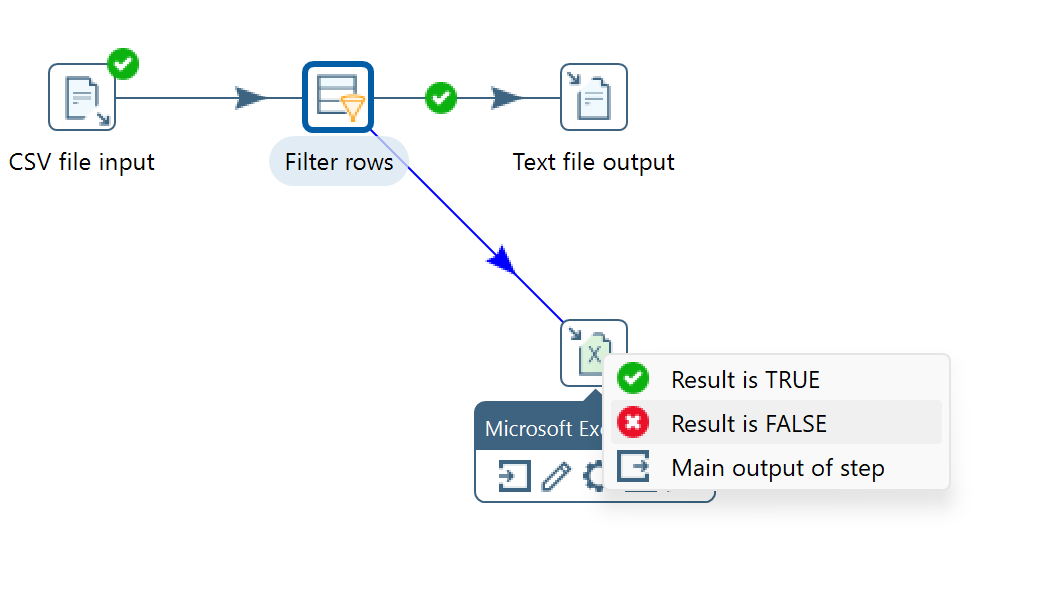
1. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



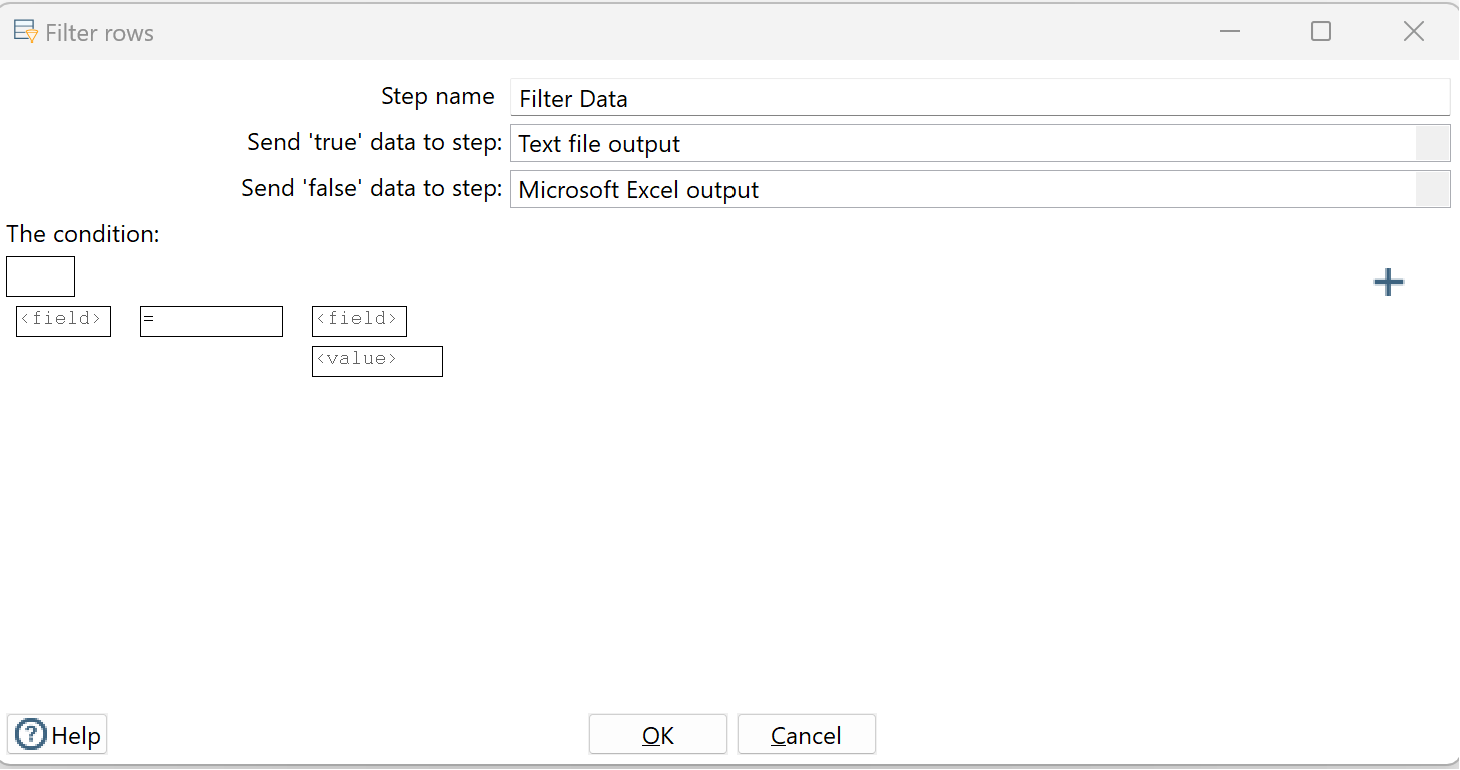
1. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagi jenis konektornya.



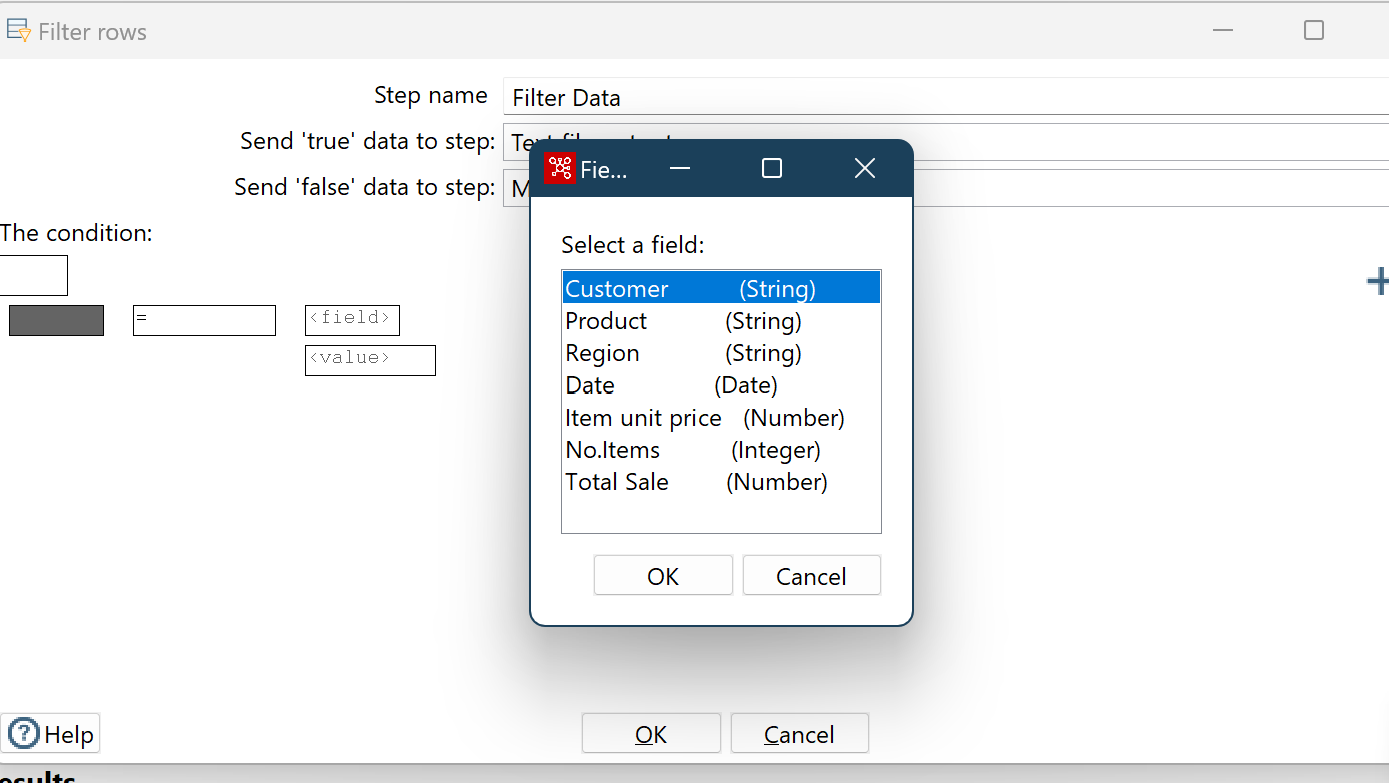
1. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagi jenis konektornya.



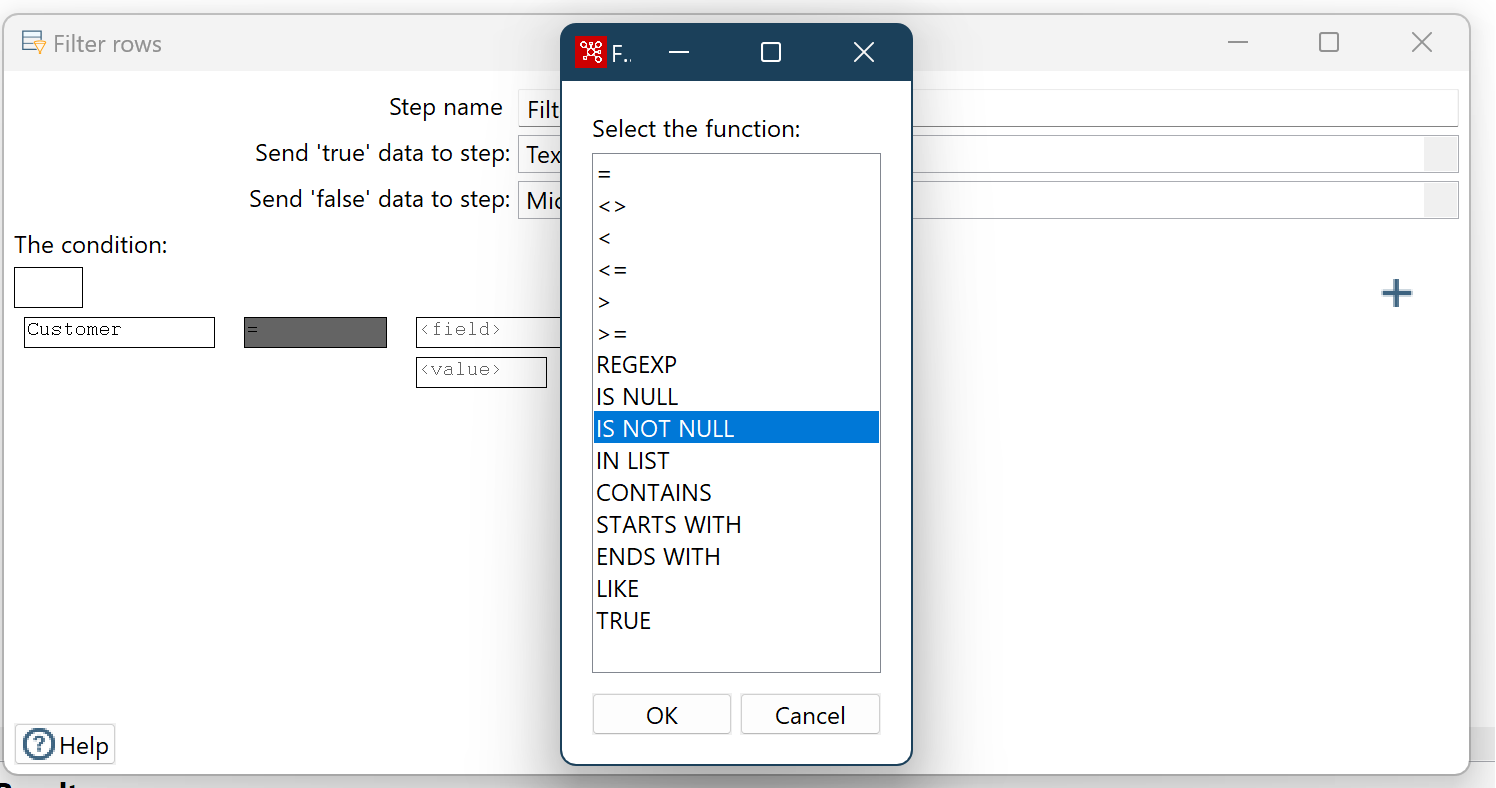
1. **Double click** pada **filter rows** hingga muncul jendela konfigurasi**.** ubah **step name** menjadi **Filter data**.
2. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
3. Klik **<Field>** pada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

****

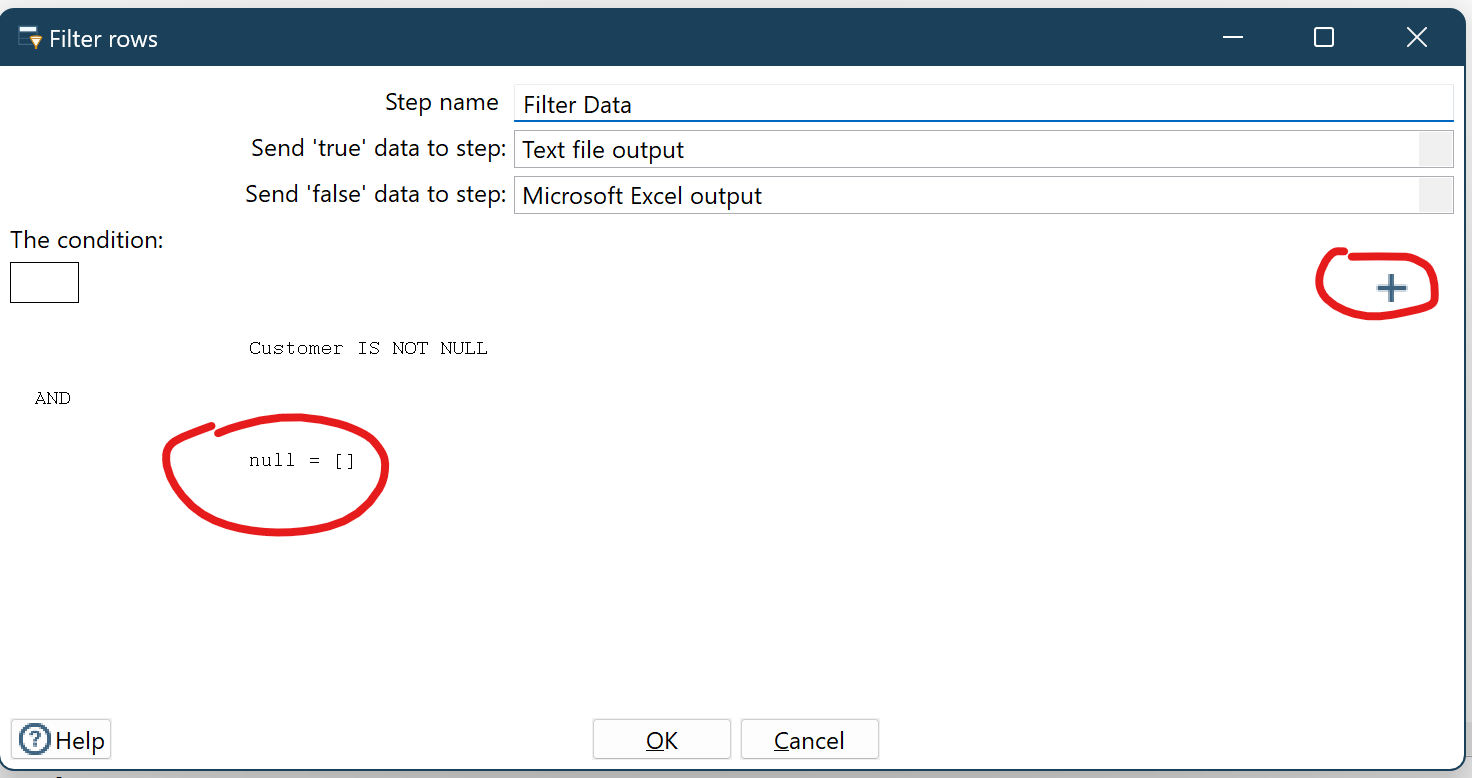
1. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer,** kemudian klik **OK.**

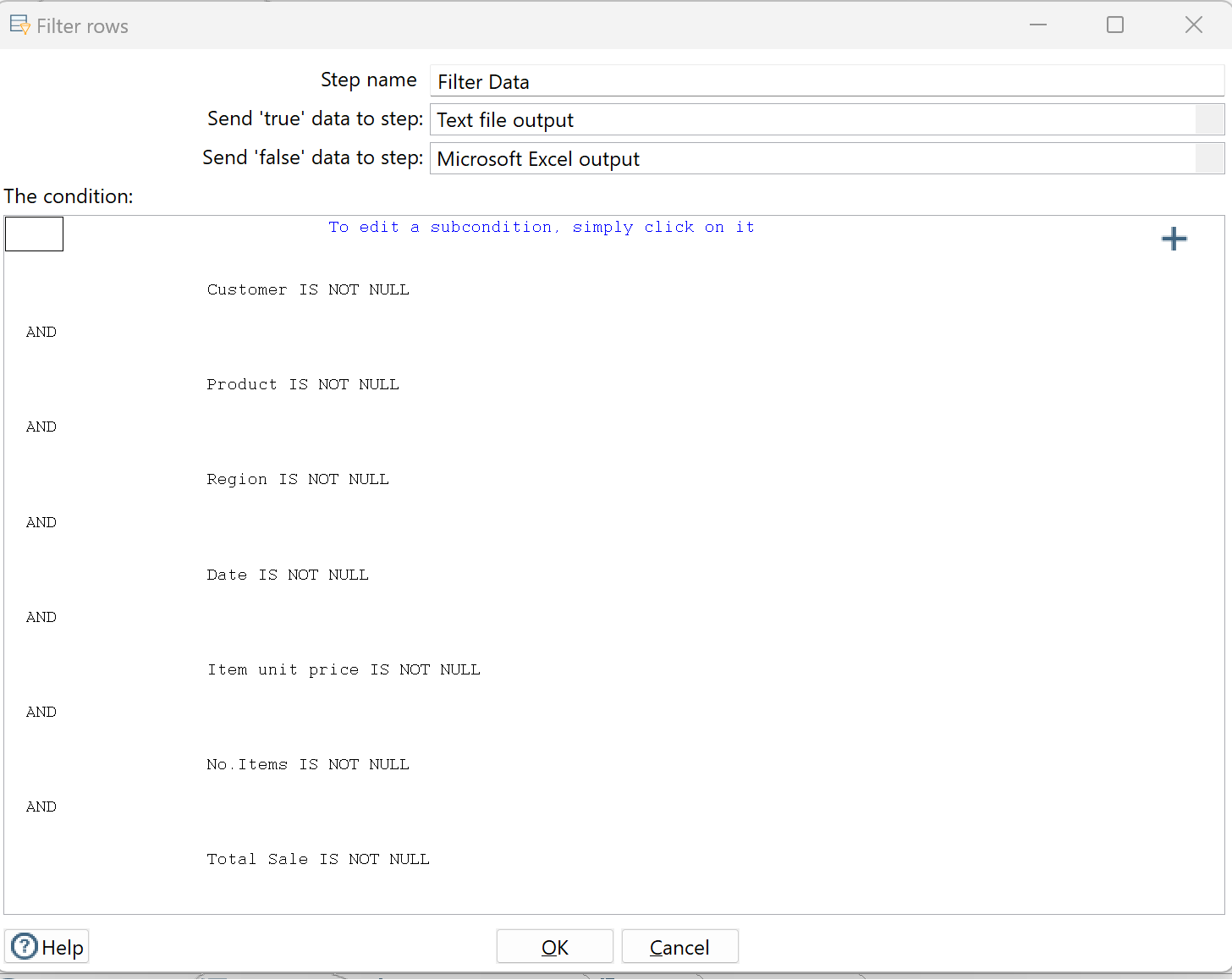
****

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

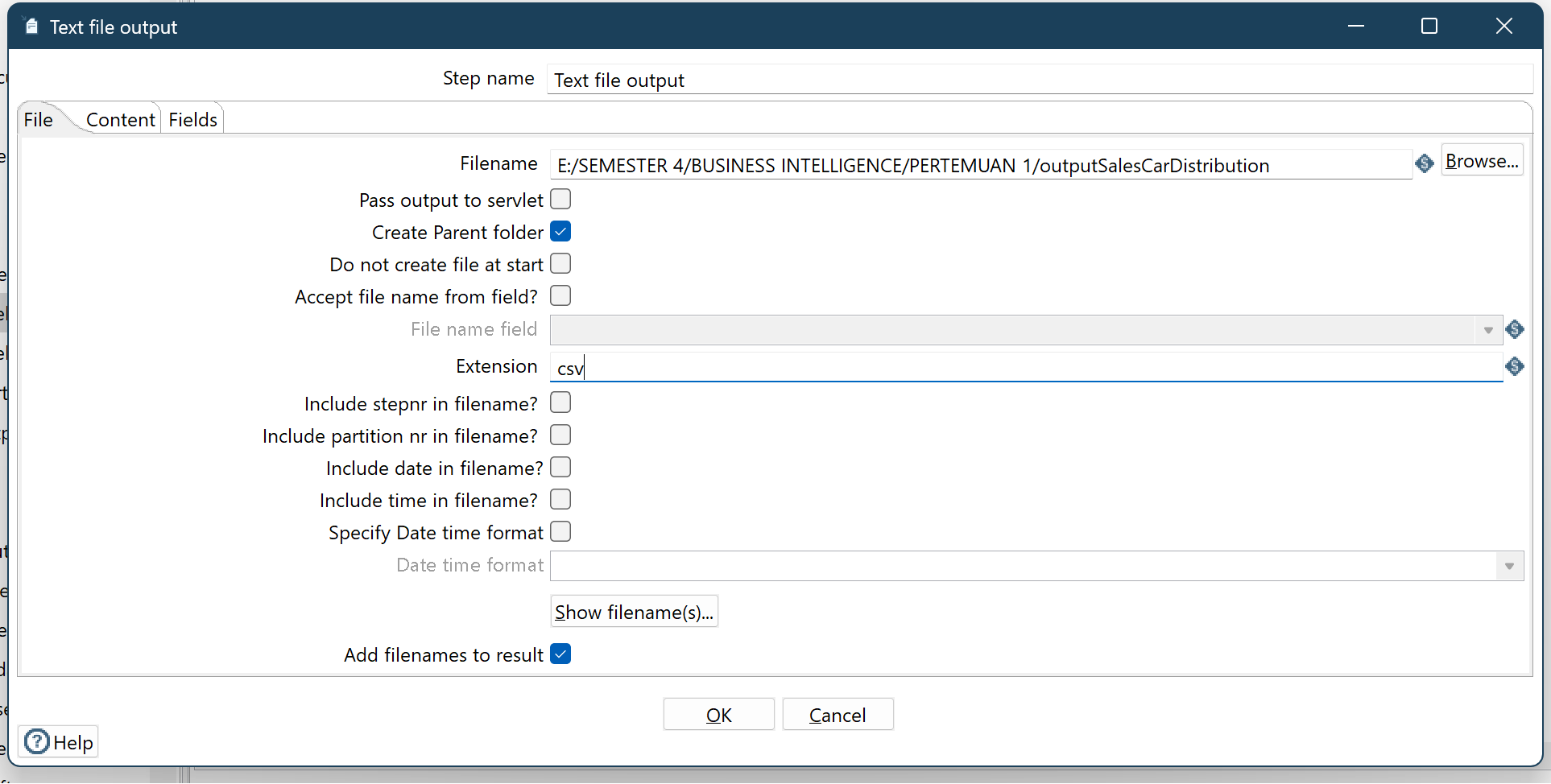


1. Tekan tombol **+** untuk menambah kondisi.
2. Pilih **statement** yang masih null untuk menambah kondisi.
3. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.

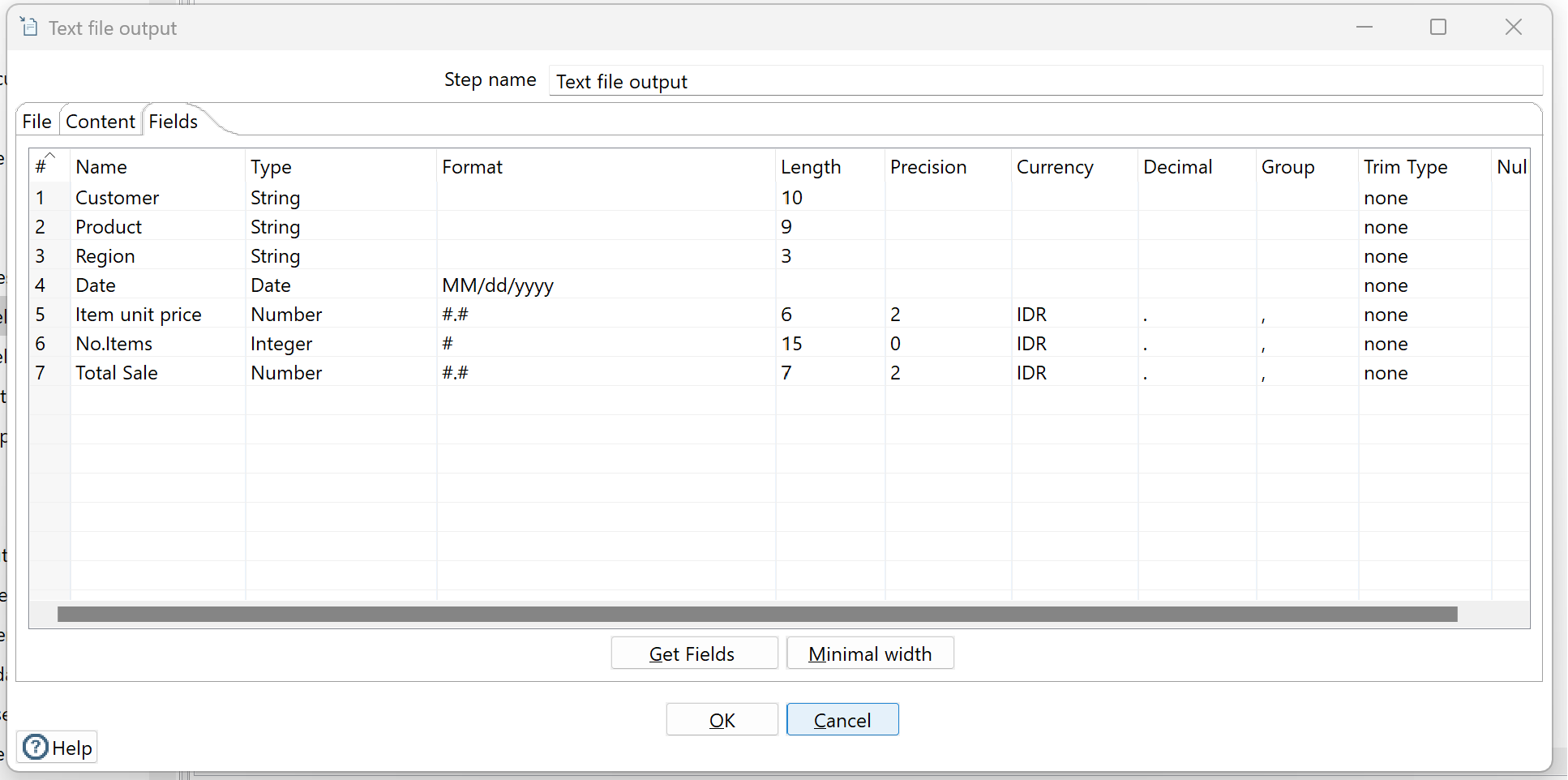




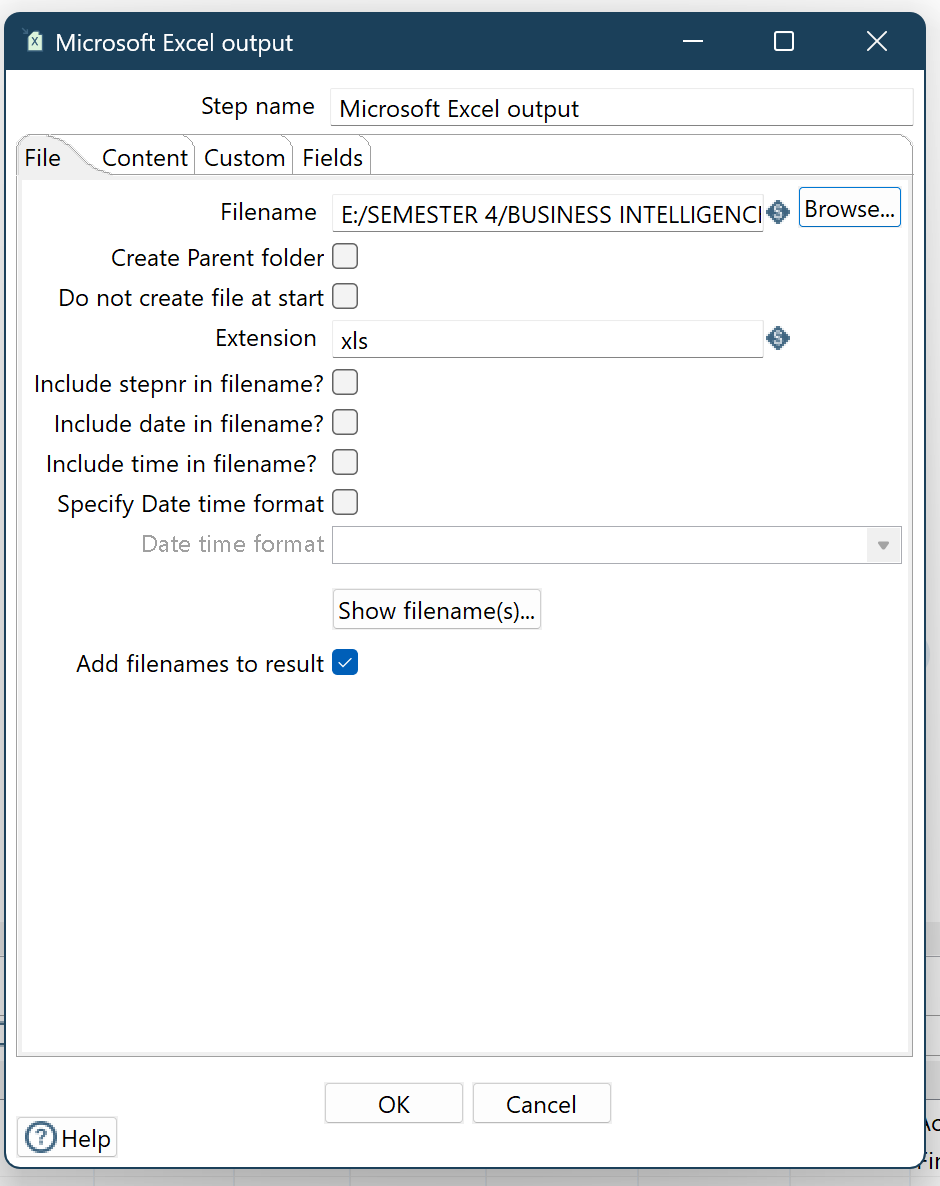
1. Gunakan operator **AND** untuk menghubung
2. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.
3. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**.
4. Ubah extension menjadi **csv.**

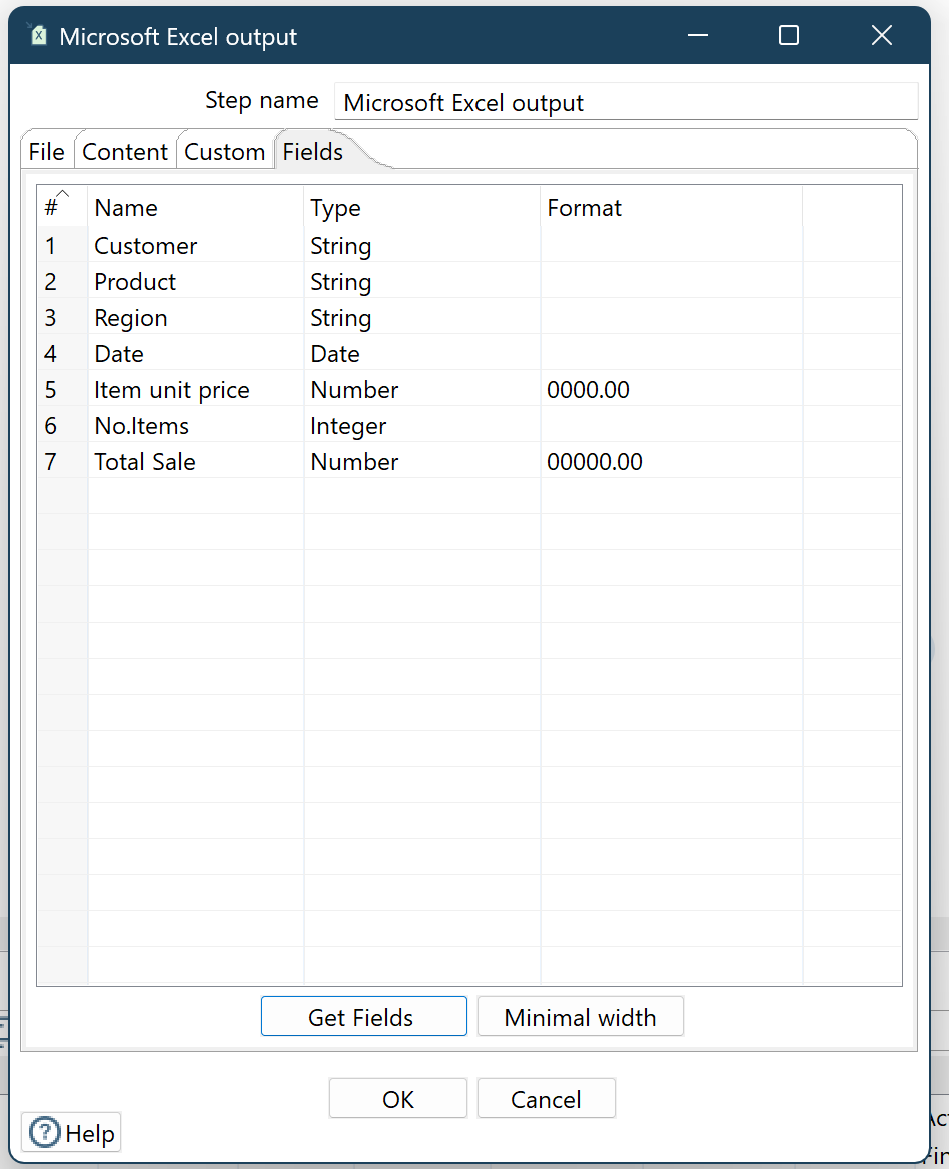
****

1. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.
2. Tekan **ok** untuk kembali ke **work area**

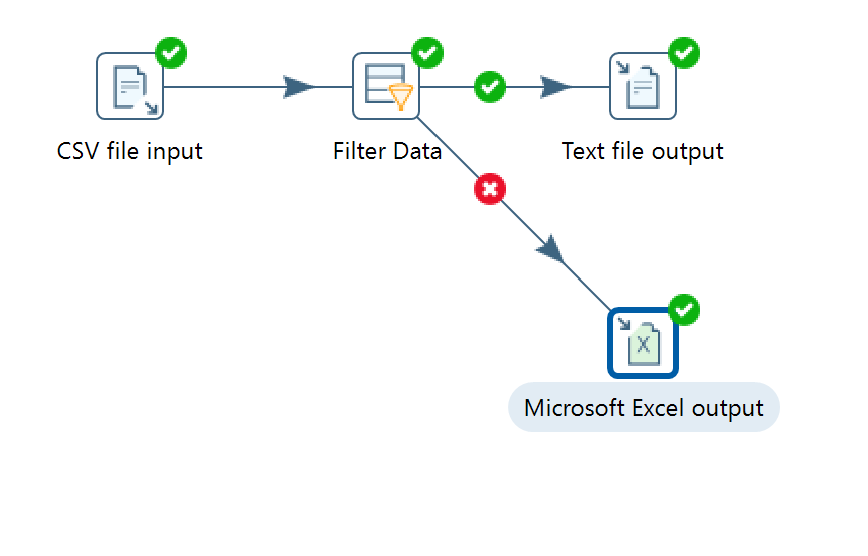
****

1. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output,** **Double-click** pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.
2. Pilih lokasi file untuk menyimpan file output dengan menekan tombol **browse**. Beri nama **failedSalesCarDistribution.** untuk extension tetap dengan extension **xls**.
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.



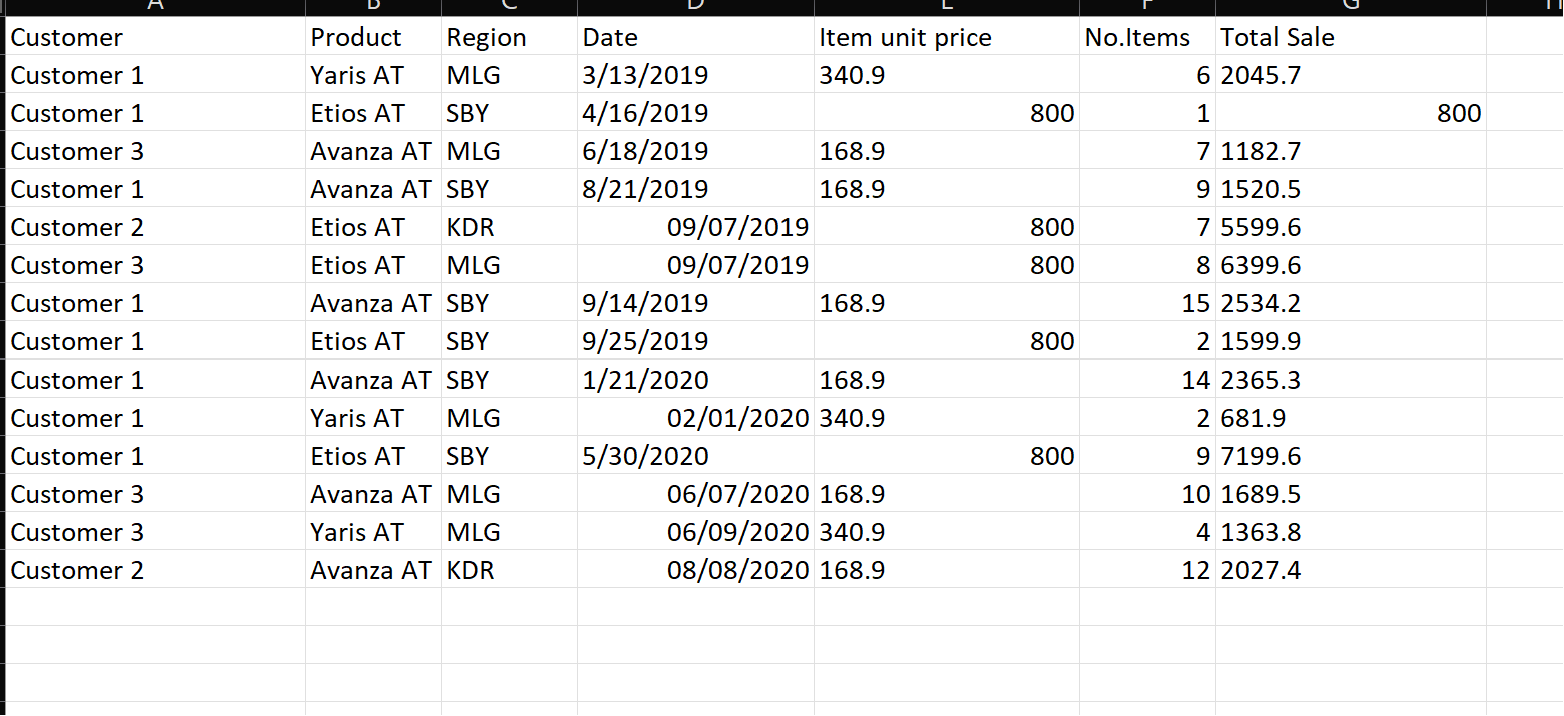
Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.

### TUGAS 3

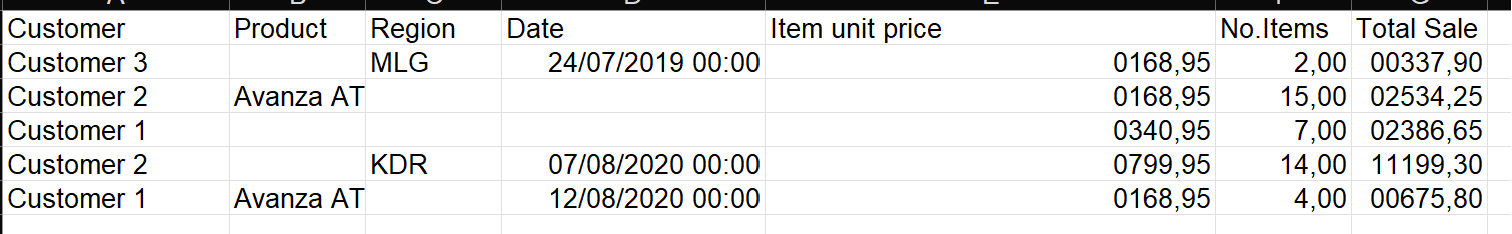
1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). bandingkan dalam bentuk tabel.

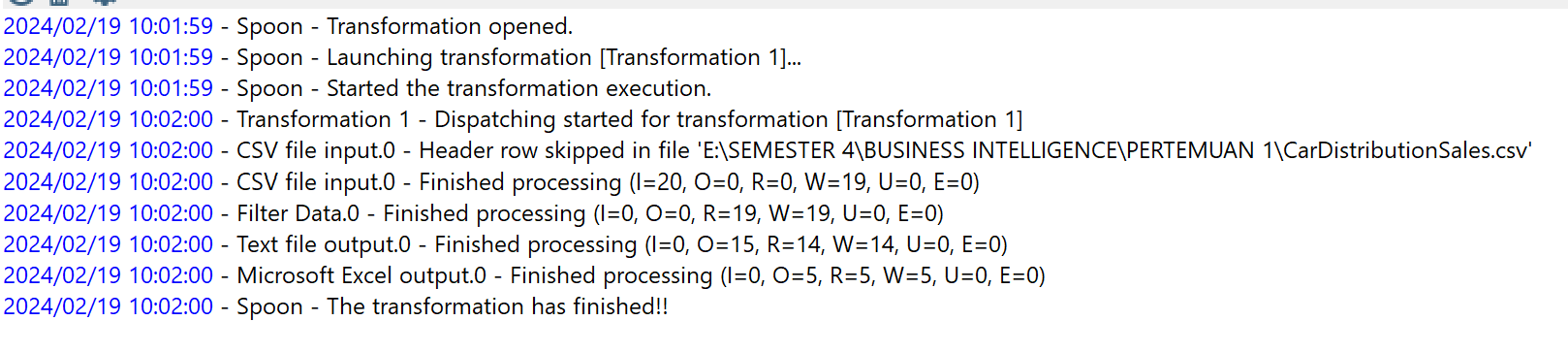
Jawaban

1. Perbedaan data output yang dihasilkan adalah
   1. Pada file csv merupakan data yang lengkap. Semua kolom terisi sehingga tidak adanya data yang bersifat null. Seperti pada gambar berikut.



* 1. Pada file excels merupakan data yang tidak lengkap. Terdapat kekosongan di dalam kolom nya. Seperti pada gambar berikut.



1. 

Berdasarkan gambar log diatas terkait prosesnya sebagai berikut

* Input data terlebih dahulu dengan 20 baris yang diinputkan.
* Lalu, ketika proses filtering 19 data dibaca dan ditulis.
* Lalu, pada filtering pertama dimana untuk memfilter data yang benar yaitu terdapat 14 data yang sesuai.
* Untuk data yang salah atau tidak lengkap yaitu sebesar 5 data.
* Ketika proses filtering selesai, maka hasil data ditulis ke dalam file yang telah ditentukan dan proses berakhir.

1. Berikut merupakan tabel yang menjelaskan tentang waktu yang diperlukan untuk melakukan eksekusi.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | Nama | Processor | RAM | Storage | VGA | CPU | Input Data CSV | Filter Data | Excel Output | Text File Output |
| 1 | Fadhlurohman | RYZEN 7 | 16 GB | 512 GB | Radeon R8 | 16 | 0.0 | 0.0 | 0.2 | 0.1 |
| 2 | M. Kemal | Intel i3 | 8 GB | 512 GB | Intel HD | 6 | 0.1 | 0.1 | 0.5 | 0.5 |
| 3 | Lalu Immaratul | Intel i7 | 16 GB | 512 GB | Intel HD | 20 | 0.0 | 0.0 | 0.2 | 0.1 |
| 4 | Aleron Tsaqif | Intel i5 | 8 GB | 512 GB | Intel Iris Xe | 12 | 0.0 | 0.0 | 0.2 | 0.1 |

# STUDI 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 membantu mengambil keputusan.

Link : <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dataDummyKependudukan.csv>

Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:

MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.