

**NAMA : SHABRINA QOTTRUNNADA**

**NIM : 2341760160**

**KELAS : SIB 2G**

**MATKUL : DATA WAREHOUSE (My Answer All)**

**JOBSHEET 1**

**DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON**

**Studi Kasus**

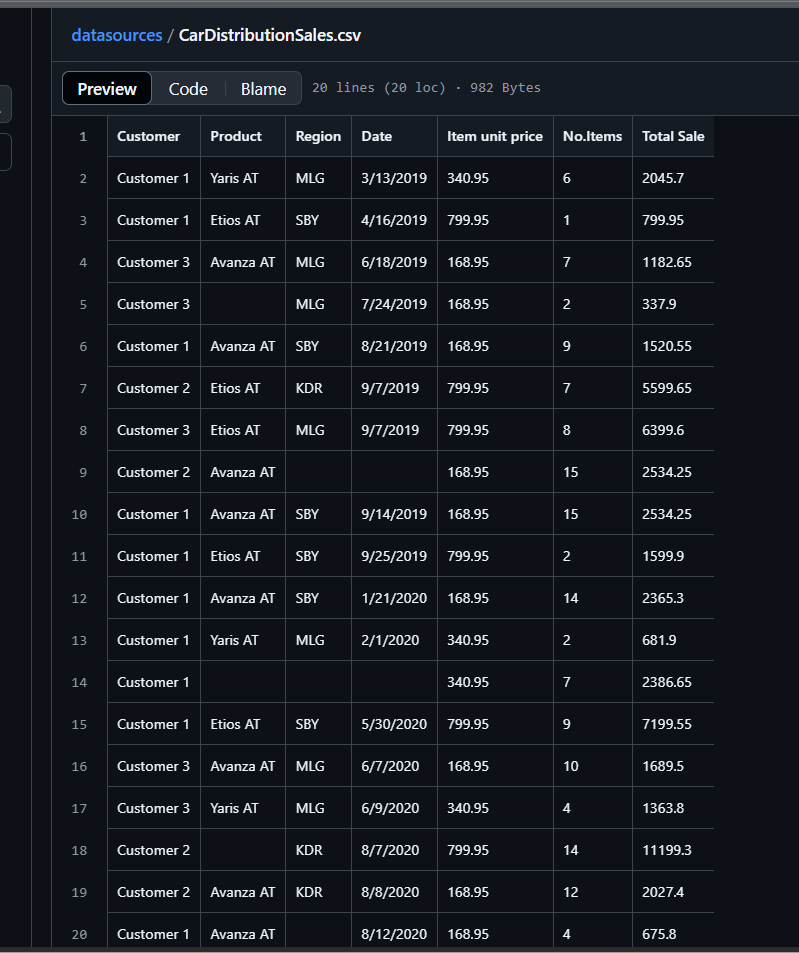
Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan 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:

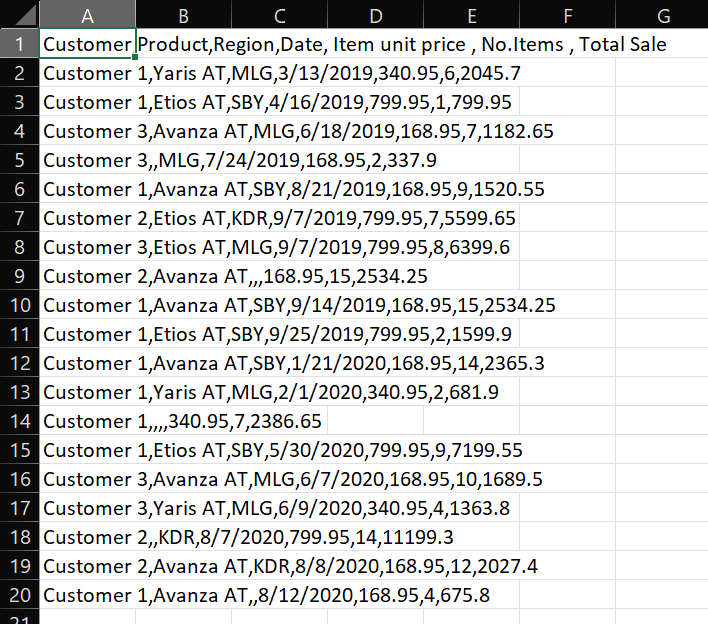
1. Cek dan Analisa data penjualan.
2. Import data dari file (Extract)
3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
4. Memindahkan data yang sudah lengkap ke file dashboard. (Load)
5. Mengumpulkan data yang belum lengkap untuk dikembalikan.

DATA

Pada Github



Pada Excel



**TUGAS 1**

**Analisa lah data tersebut!**

1. **Berapa jumlah kolom pada data tersebut?**

**My Answer**

Jumlah kolom yang ada pada data tersebut adalah 7 kolom.

1. Customer
2. Product
3. Region
4. Date
5. Item Unit Price
6. No. Items
7. Total Sale
8. **Apa arti atau isi data dari setiap kolom yang ada?**

**My Answer**

* 1. Customer : customer yang membeli barang
  2. Product : barang yang diperjual dan dibeli oleh customer
  3. Region : Lokasi customer membeli barang
  4. Date : tanggal customer melakukan transaksi
  5. Item Unit Price : harga satuan dari barang yang dibeli customer
  6. No. Items : jumlah unit yang dibeli oleh customer
  7. Total Sale : total transaksi customer dalam satu kali pembelian

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

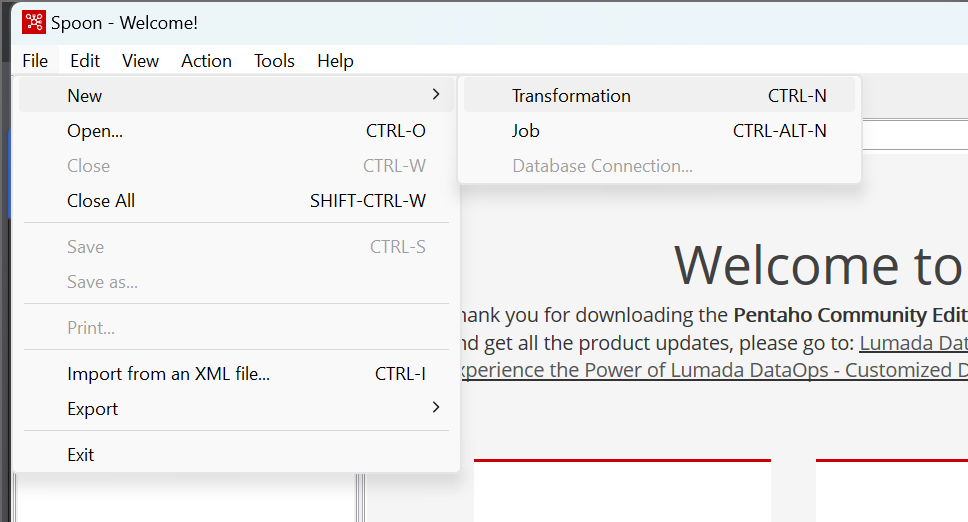
**My Answer**

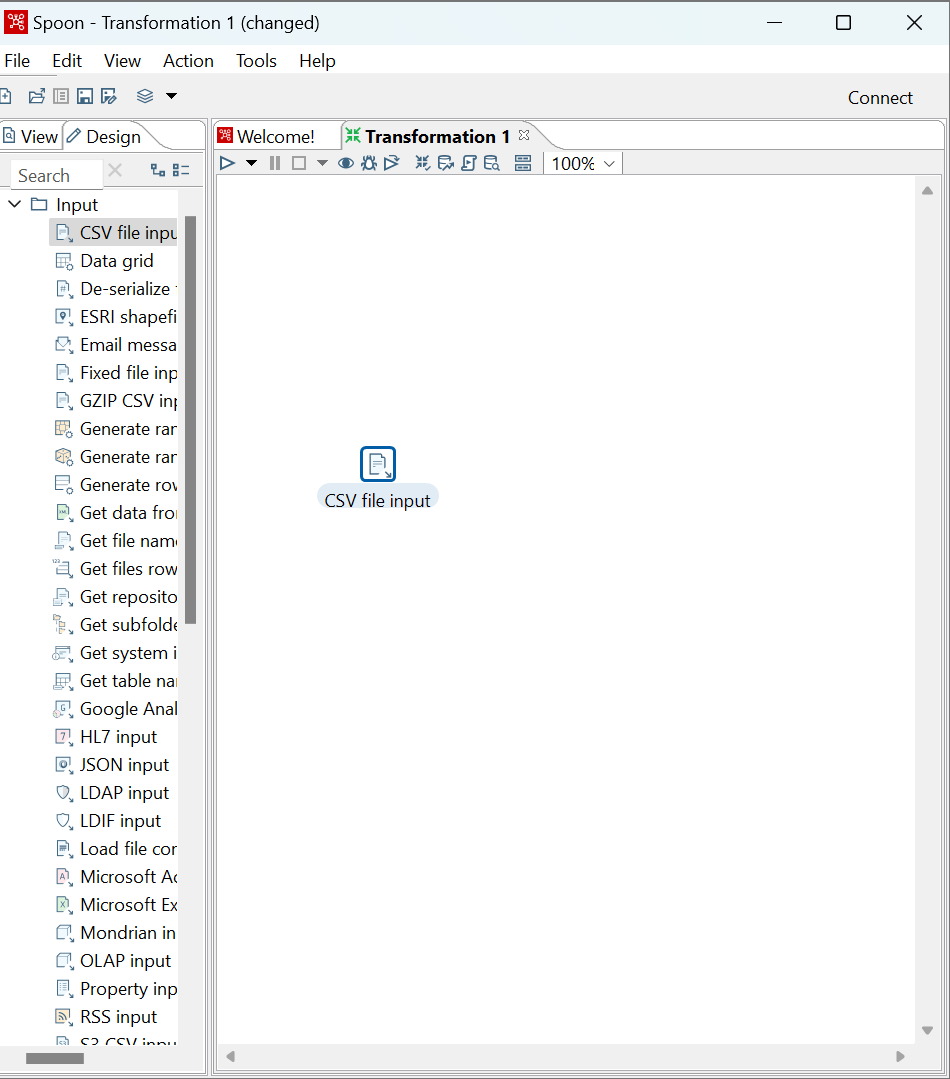
Ada, ada beberapa baris pada beberapa kolom yang memiliki nilai null atau kosong

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

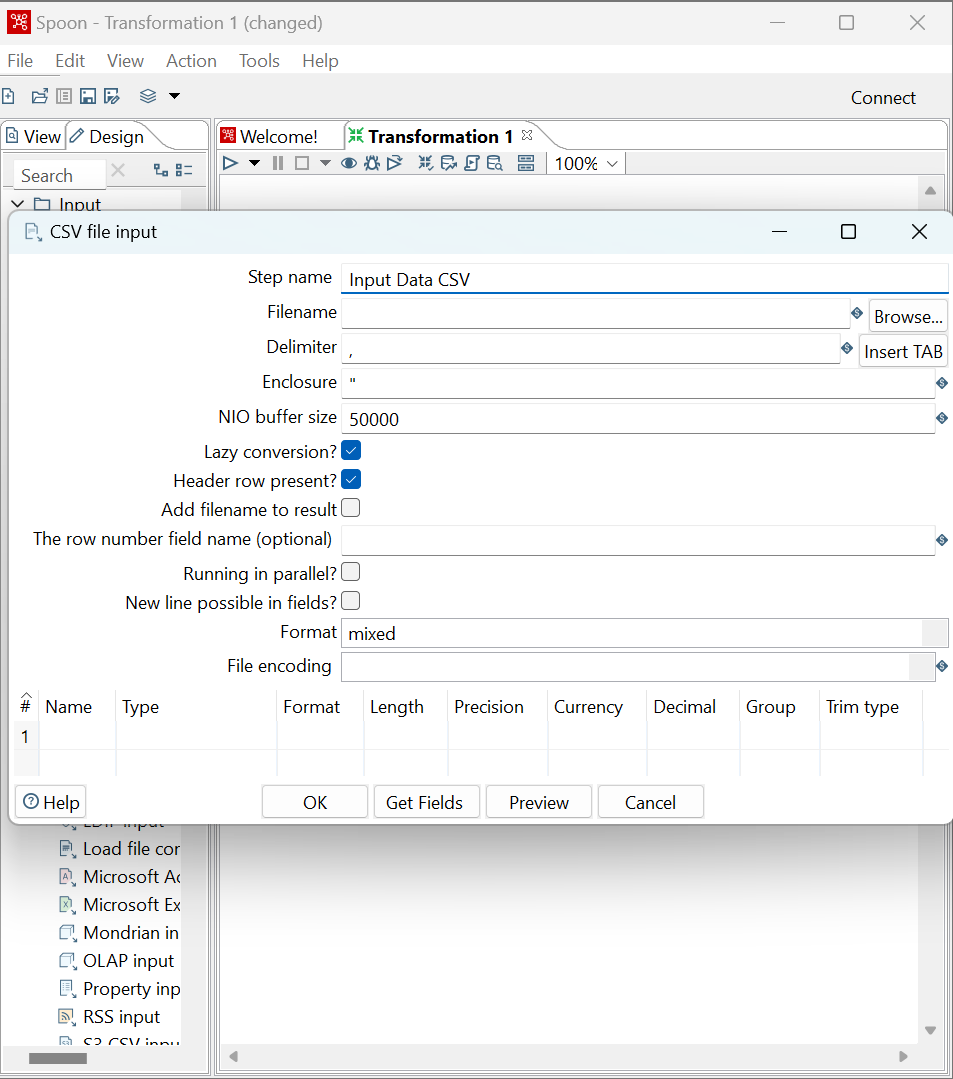
**My Answer**

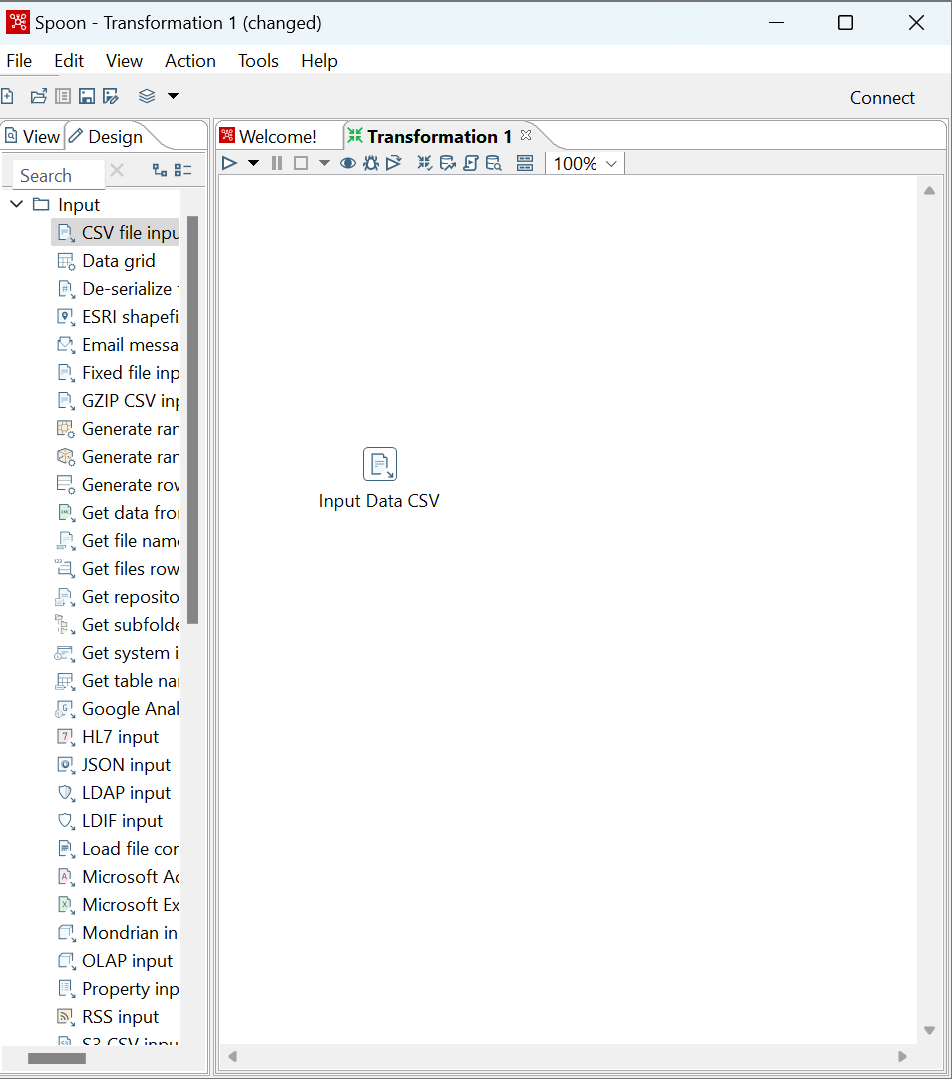
Tidak ada, karena pada data tersebut tidaka da seperti string, integer dll.

1. **PENGAMBILAN DATA (EXTRACT)**
2. **Buka lembar kerja Transformation melalui toolbar File.**
3. **Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.**

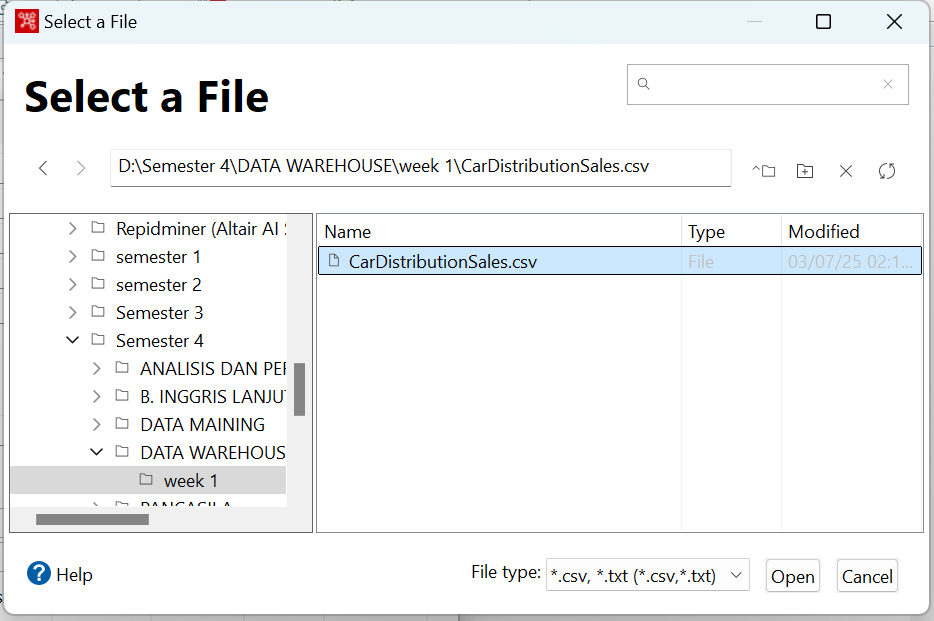


1. **Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.**
2. **Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).**

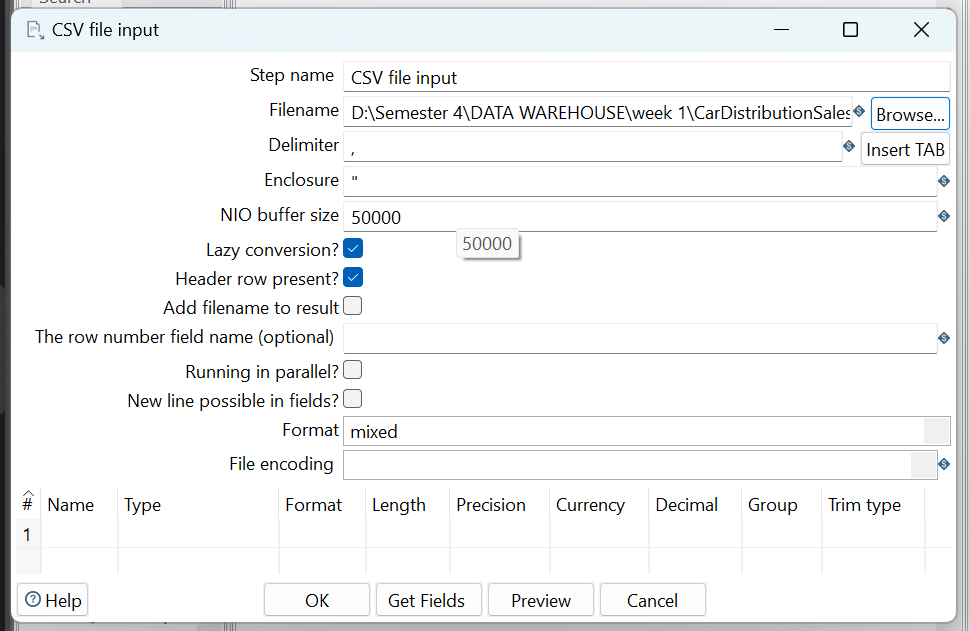


****

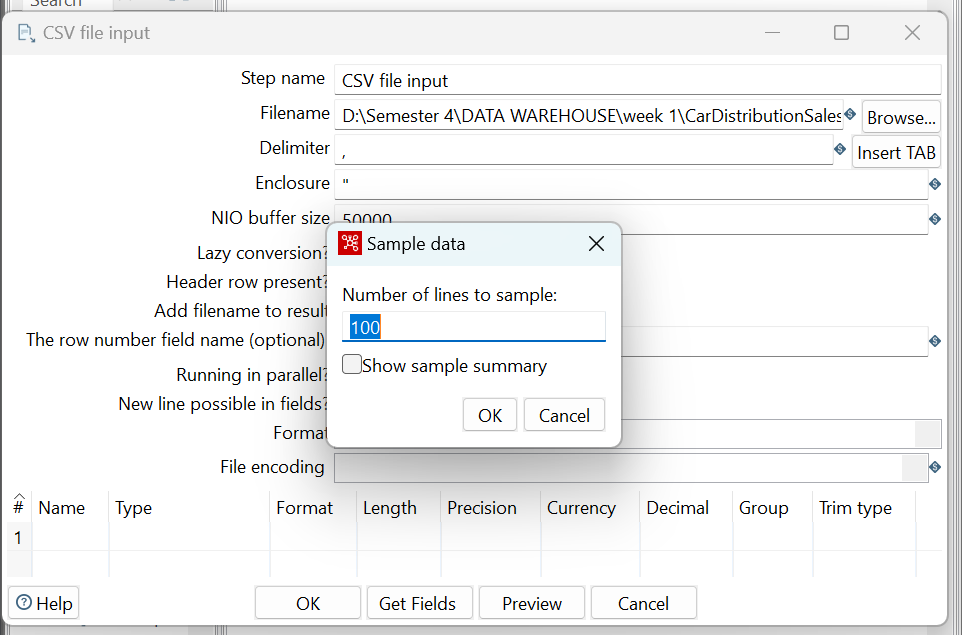
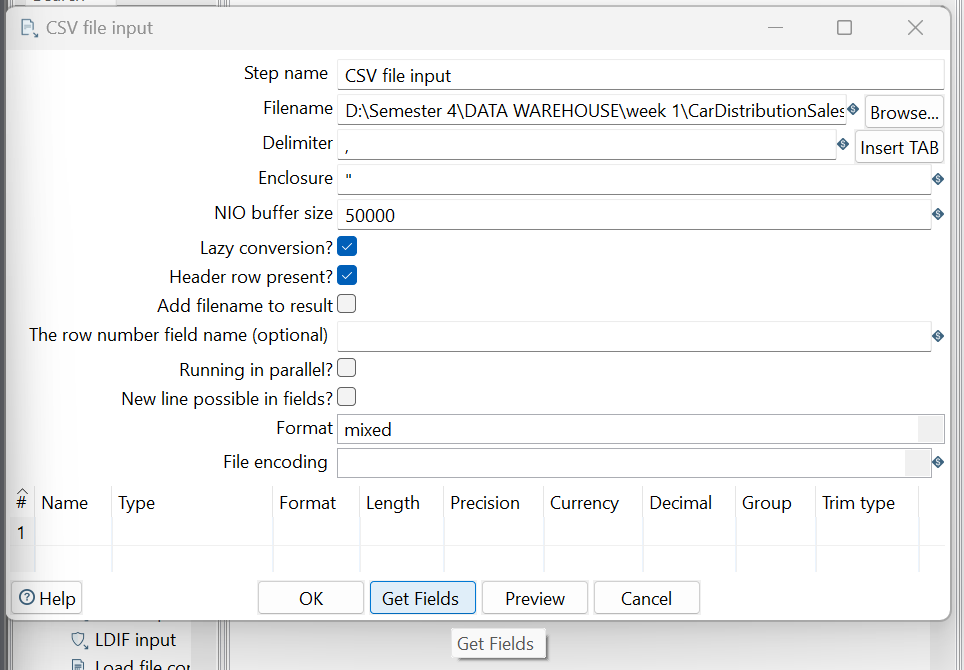
1. **Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.**

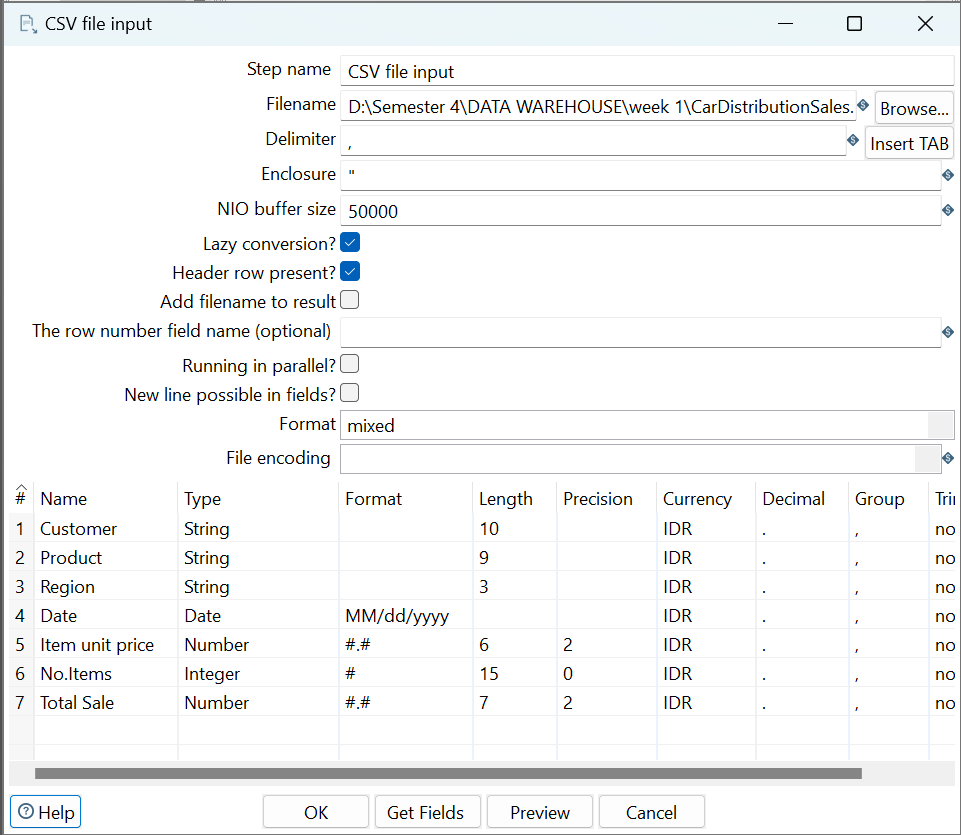
****

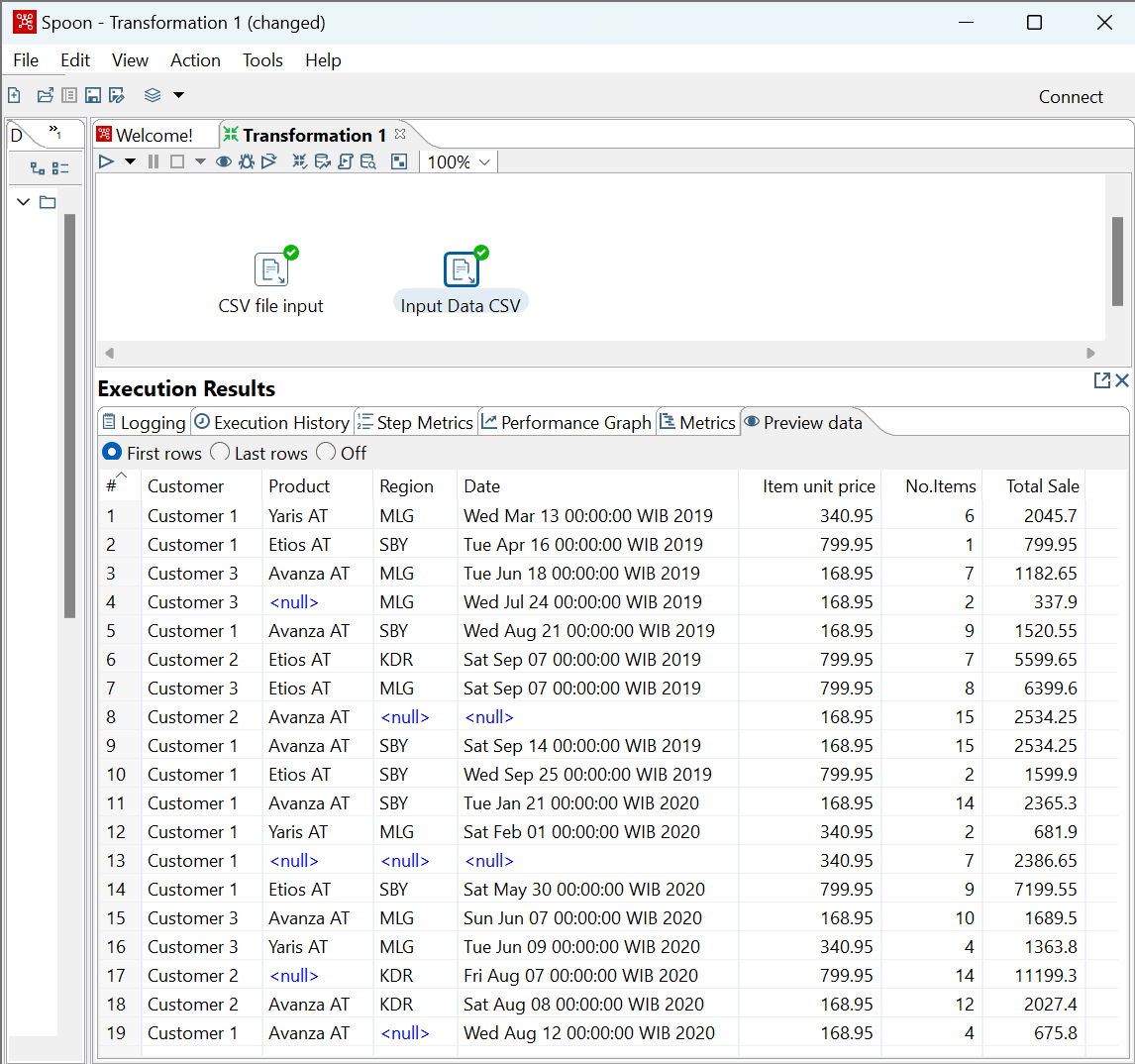
1. **Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).**

****

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

****

****

****

**TUGAS 2**

1. **Apakah data hasil eksekusi sesuai dengan data aslinya?**

**My Answer**

Iya, hasil eksekusi yang saya coba diatas hasilnya sama persis denga napa yang ada di jobsheet tersebut, sehingga bisa dibilang percobaan saya berhasil dilakukan.

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?**

**My Answer**

* 1. Yang dilakukan pertama adalah membuka PDI Spoon lembar kerja dibagian file, klik new, lalu klik transformation.
  2. Membuka koneksi sumber data, dengan cara klik CSV input data, lalu kita geser ke bagian tap pengerjaan.
  3. Lalu kita mengeksekusi Query sesuai dengan yang ada di jobsheet Langkah-langkahnya
  4. Lalu data kita rubah dengan format sesuai dan PDI akan melakukan mapping kolom.
  5. Menampilkan log eksekusi untuk melihat hasil dan mendeteksi errornya.

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

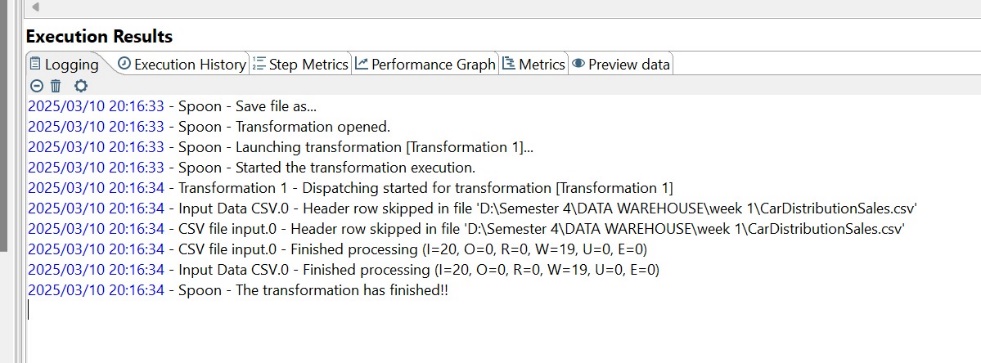
****

**My Answer**

Berdasarkan log yang ditampilkan :

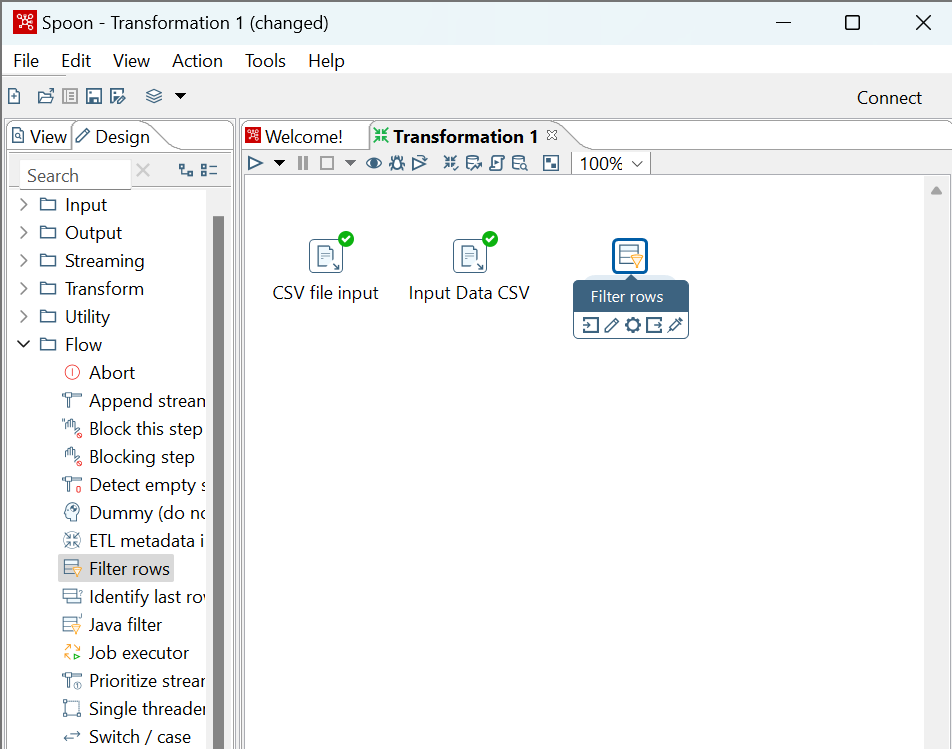
1. I (input) : jumlah record yang dibaca (sehingga I=20 merupakan 20 record yang di baca)
2. O (Output) : jumlah record yang ditulis ke tujuan akhirnya (sehingga O=0 merupakan tidak ada data yang ditulis langsung menuju ke output akhir)
3. R (Read) : Jumlah record yang dibaca ulang oleh step-step tertentu
4. W (Written) : Jumlah record yang diteruskan ke step berikutnya
5. U (Updated) : Jumlah record yang diperbarui dalam database
6. E (Error) : Jumlah error yang terjadi dalam proses yang dilakukan
7. **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.**

**My Answer**

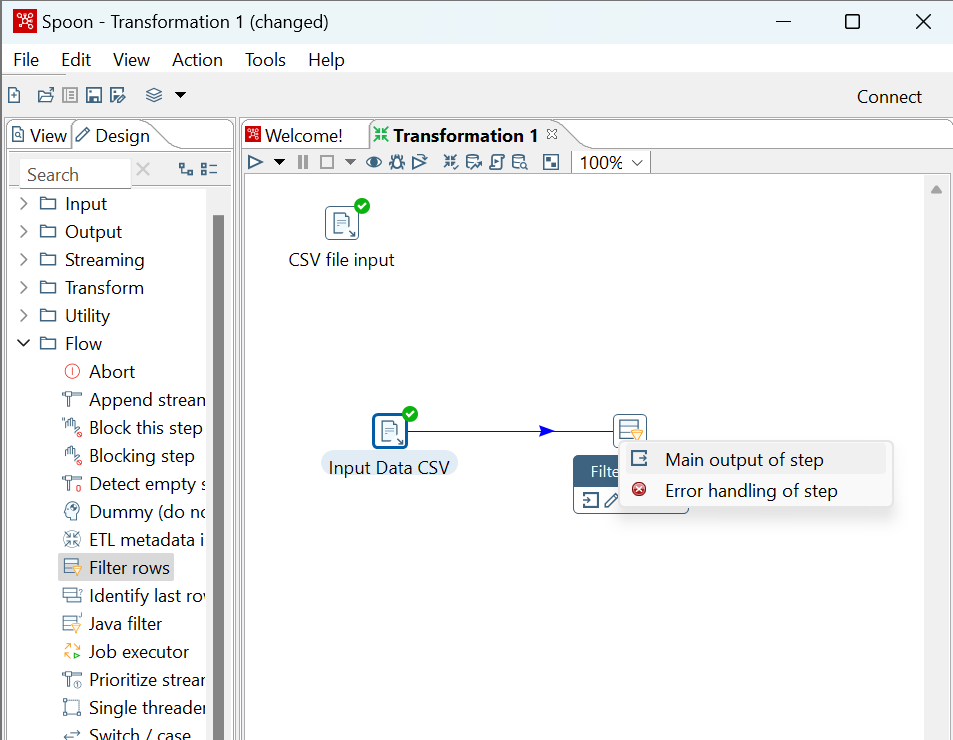


|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Processor** | **RAM** | **Storage** | **VGA** | **CPU** | **Time Of Execution** |
| **Shabrina Q** | 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz (8 CPUs), ~2.5GHz | 16384MB (16 GB) | SSD NVMe 477GB | Intel(R) Iris(R) Xe Graphics | 23% 2.25 GHz | 1 Second |
| Fakhira Z | Intel®  CoreTM i7-  13650HX  (13th Gen,  20 Cores) | 16 GB | SSD  477GB | Intel®  UHD  Graphics  +  NVIDIA  GeForce  RTX 4050 | 7% | 1 Second |
| Hanin M | Intel | 16 GB | 512 GB  SSD | GeForce  RTX  2050 | Intel(R)  Core(TM)  i7-  12650H | 0 seconds  (or less  than 1  second) |
| Sabrina R | Intel(R) Core(TM) i5-7300U CPU @ 2.60GHz (4 CPUs), ~2.7GHz | 8192MB RAM  (8 GB) |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

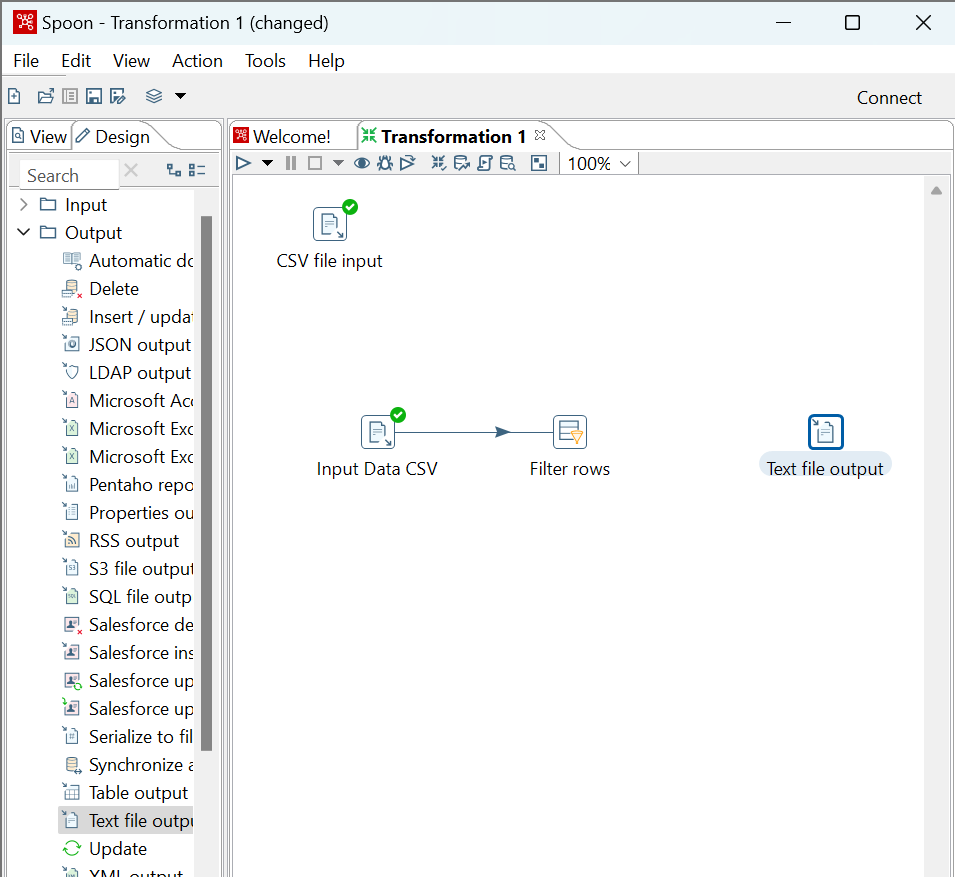
1. **FILTER DATA (TRANSFORM) DAN PENGEMASAN DATA (LOAD)**
2. **Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area.**

****

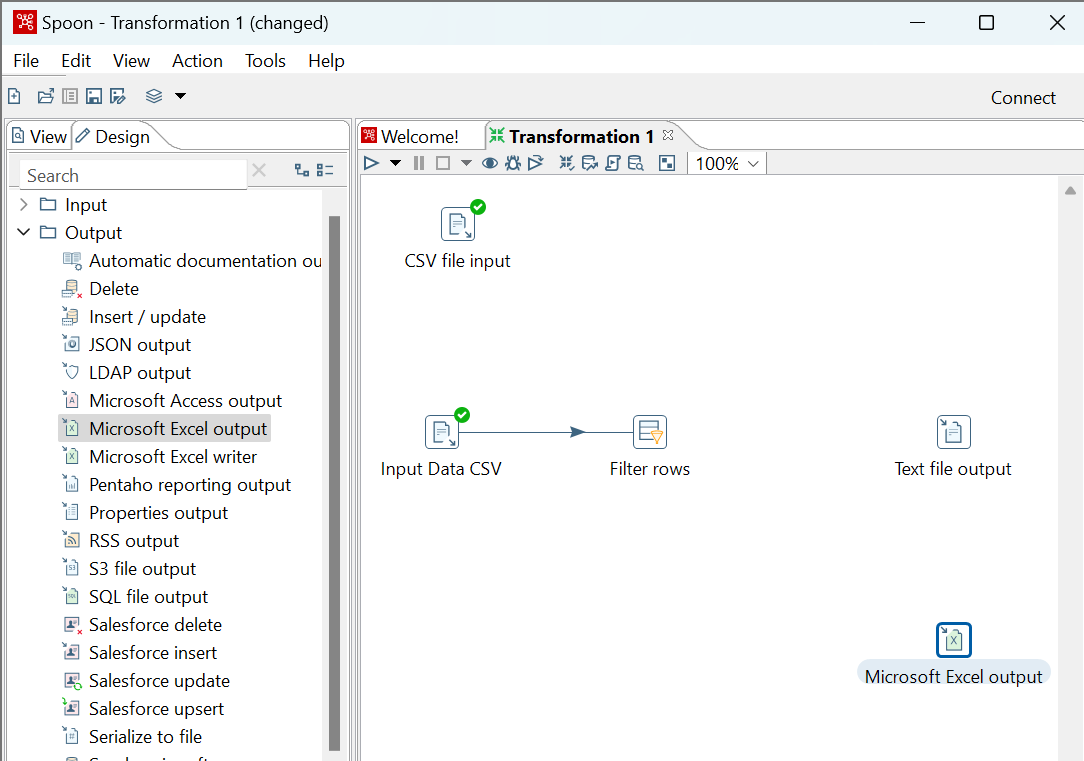
1. **Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.**

****

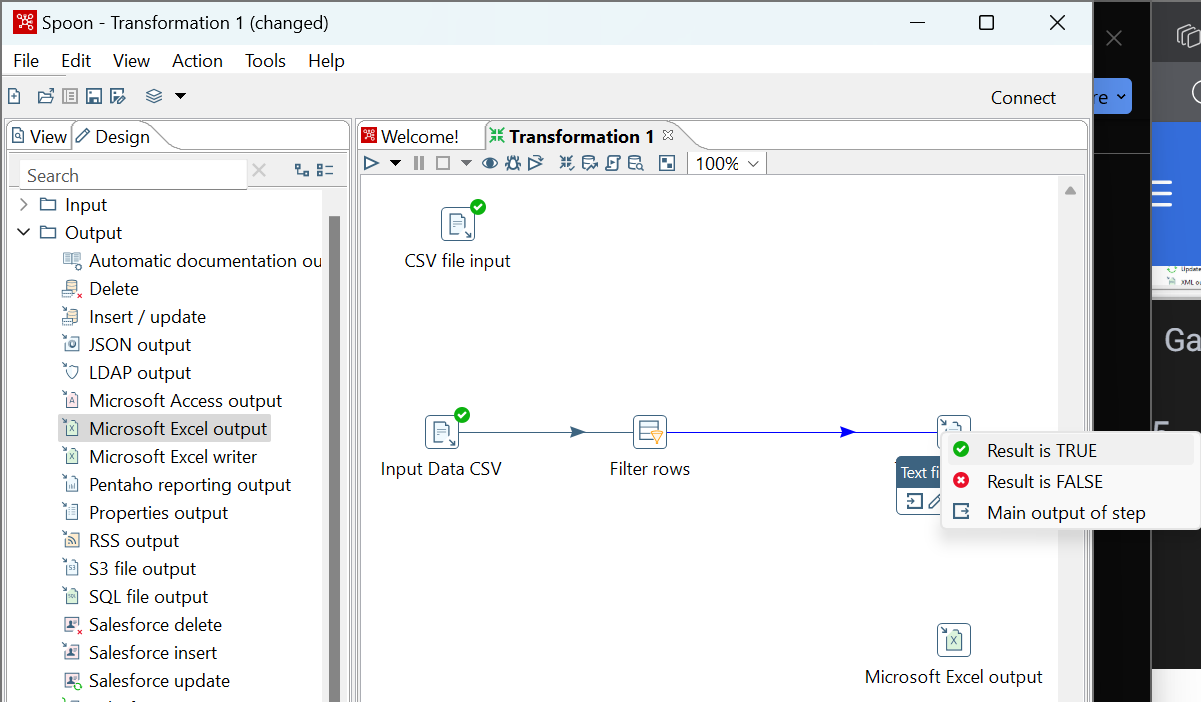
1. **Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.**

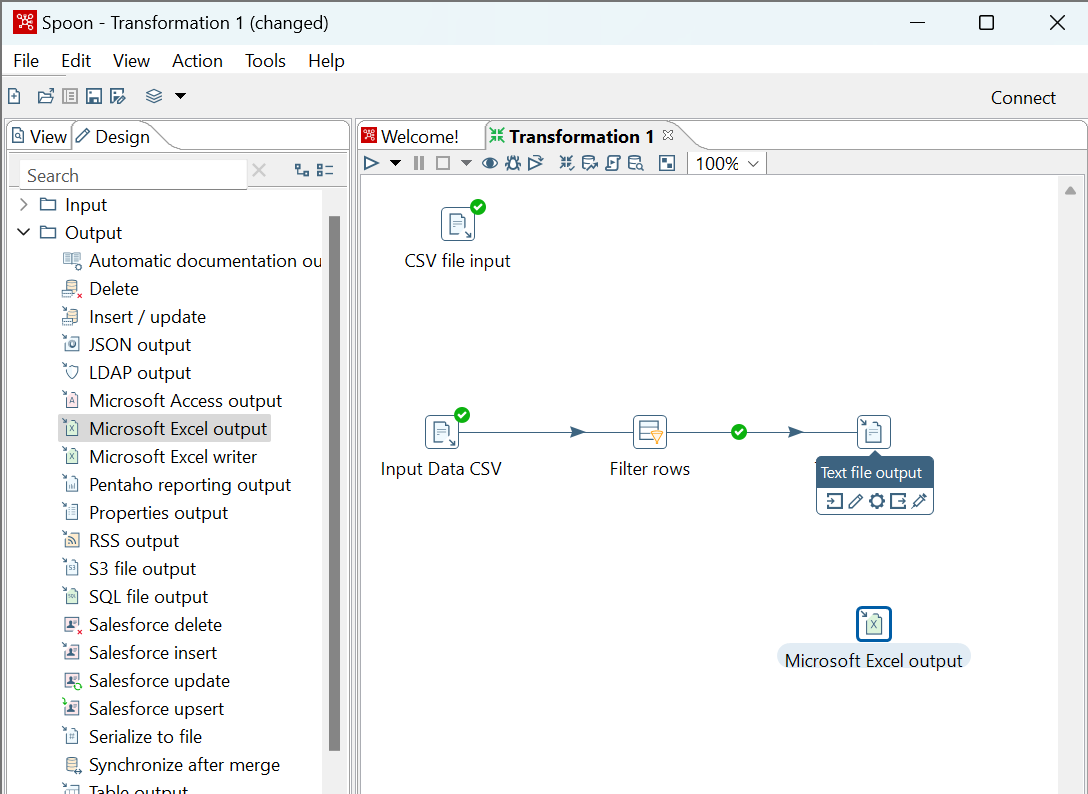
****

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

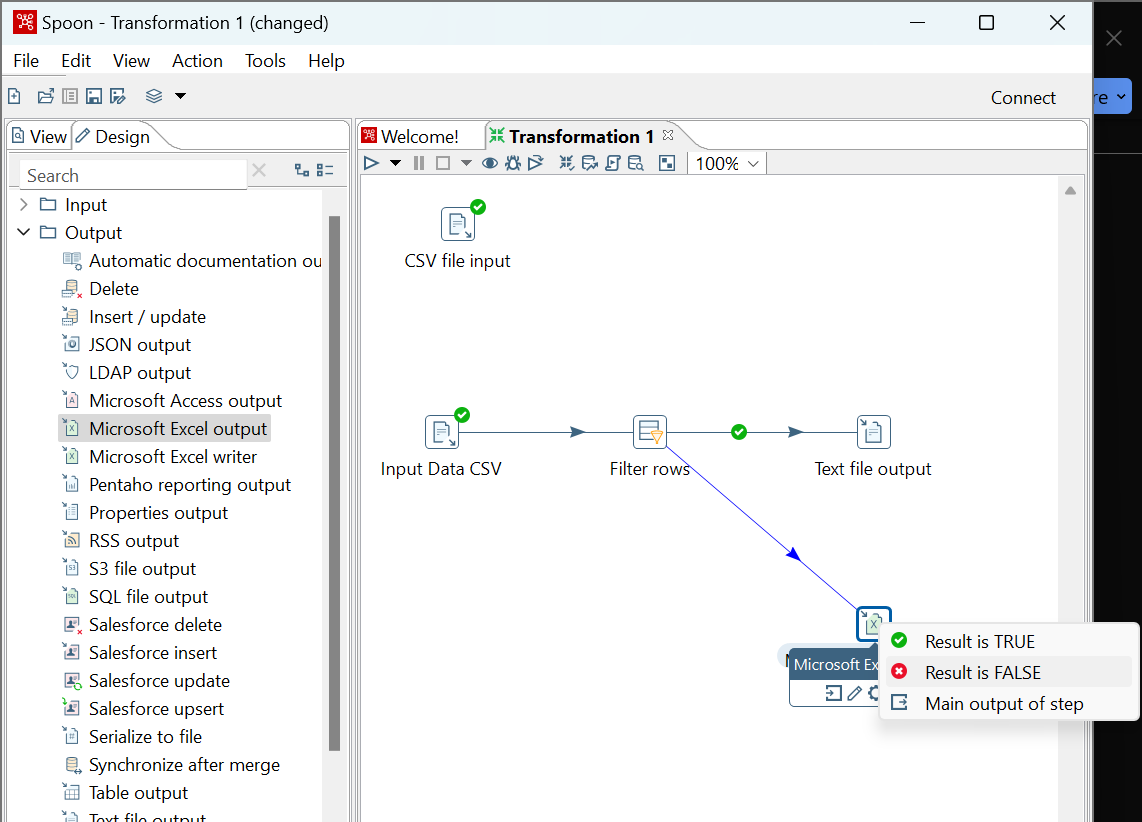
****

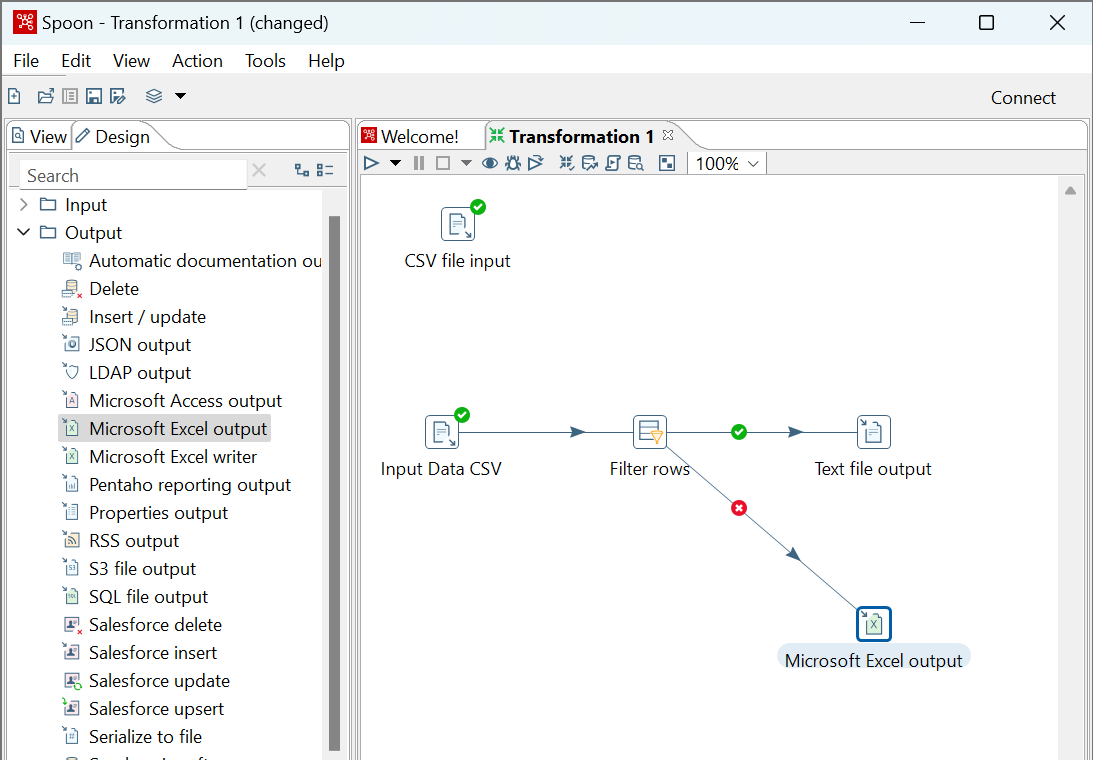
1. **Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagi jenis konektornya.**

****

****

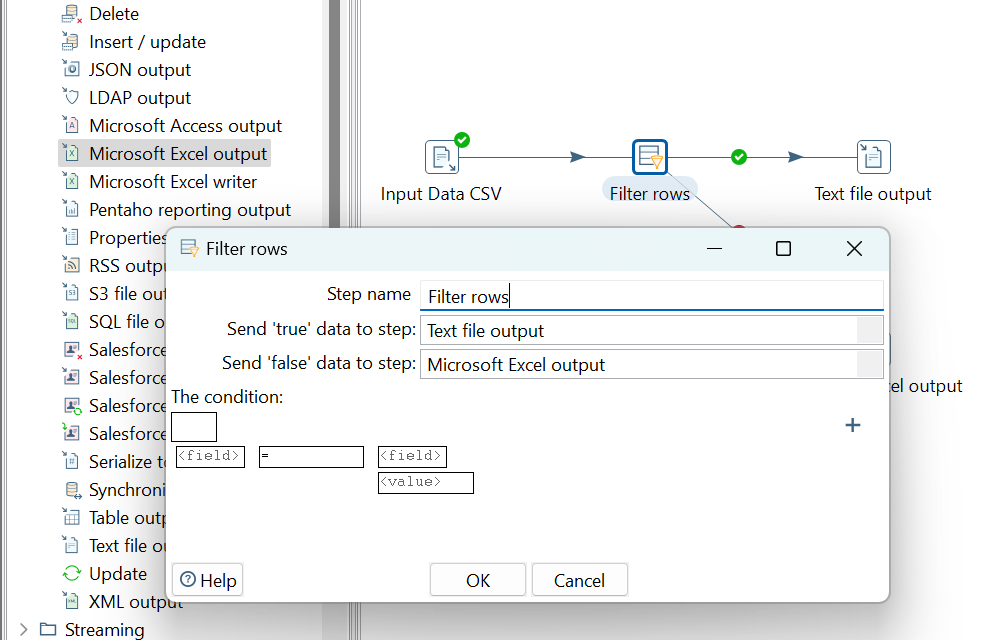
1. **Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.**

****

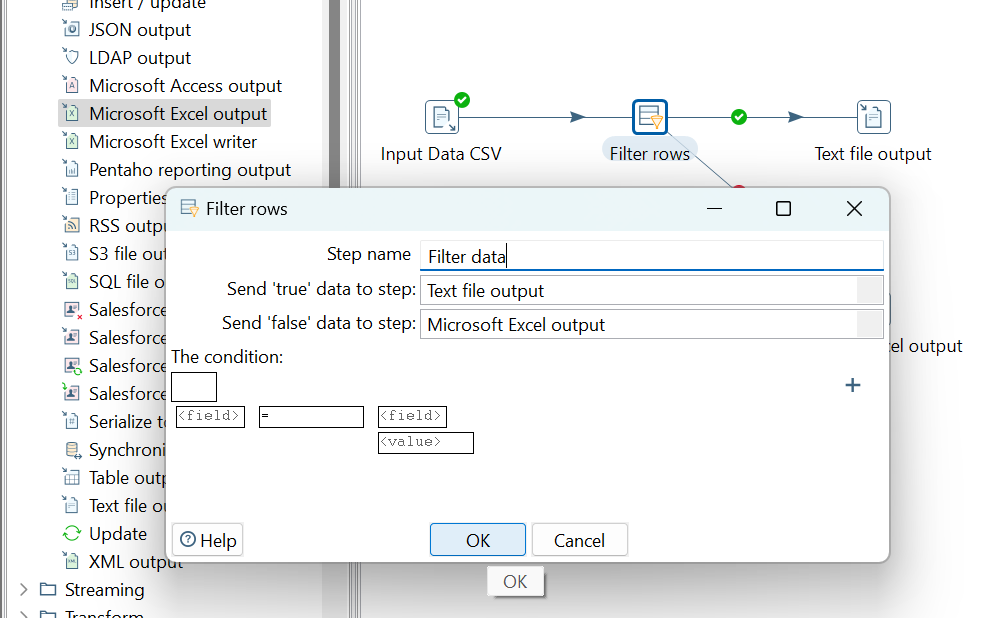
****

1. **Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.**

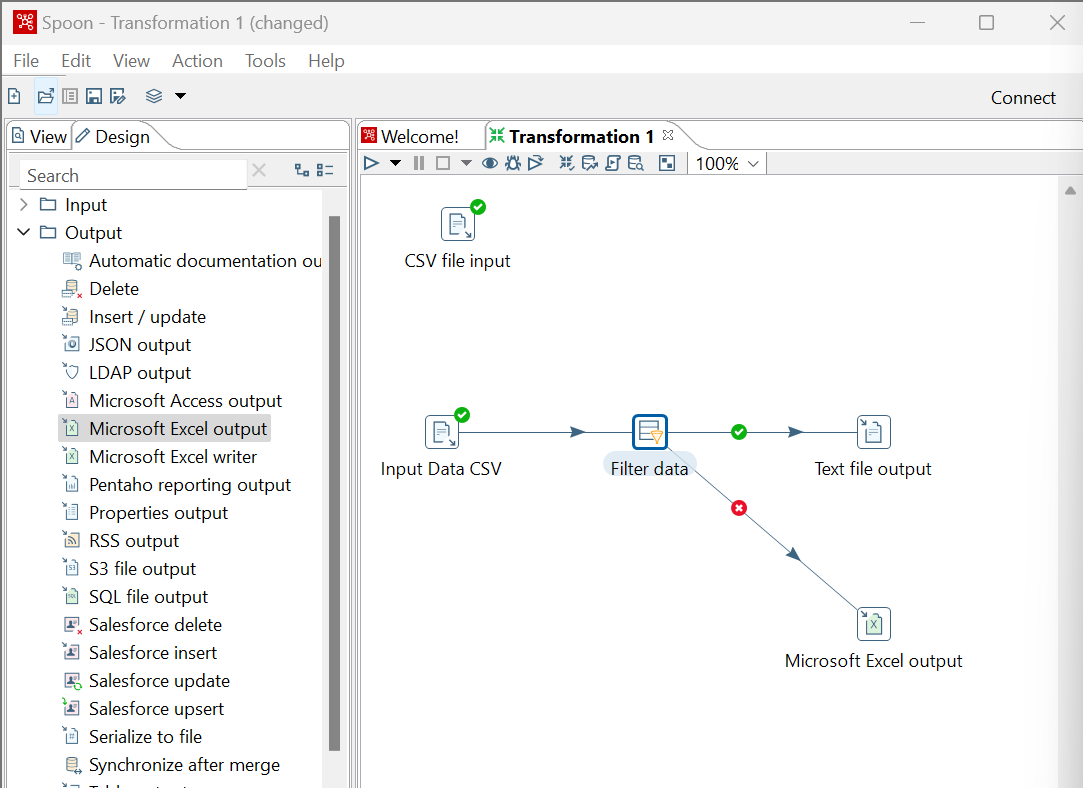
**Before :**

****

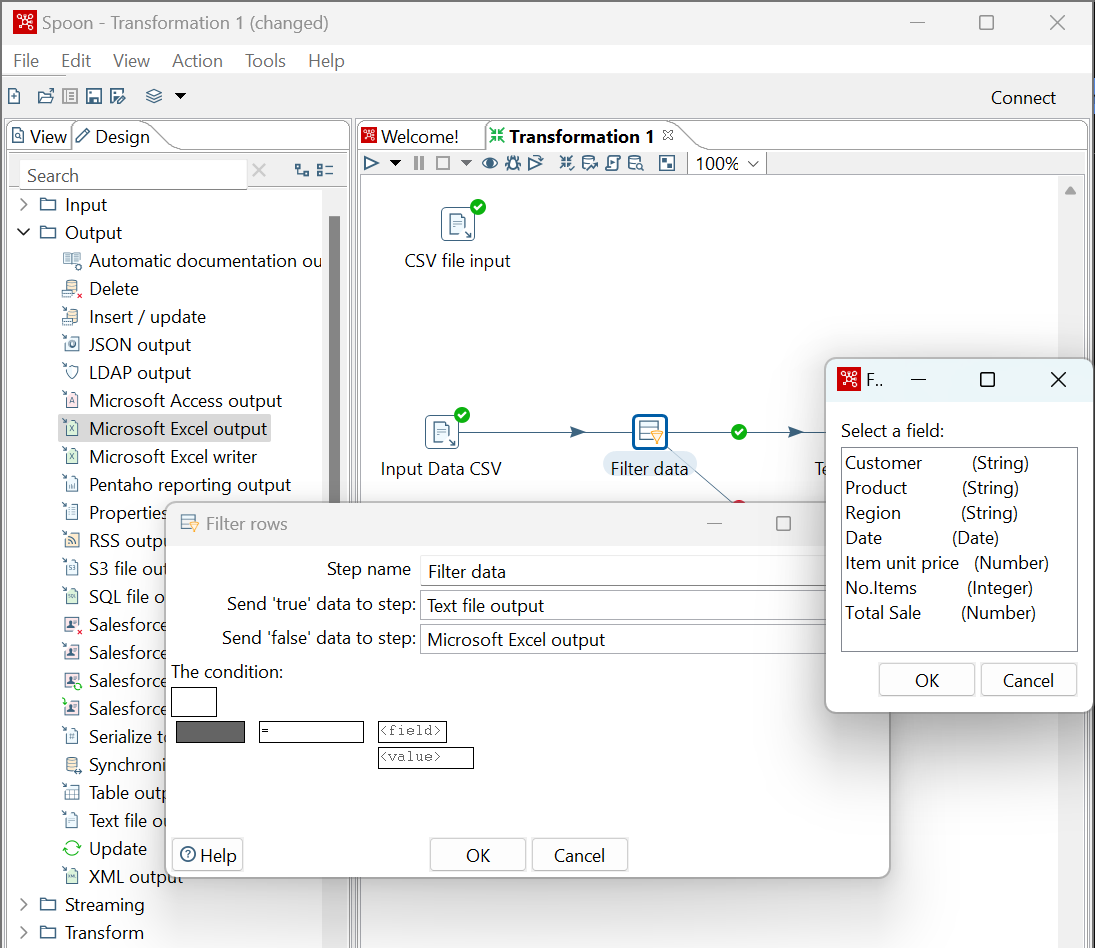
**After :**

****

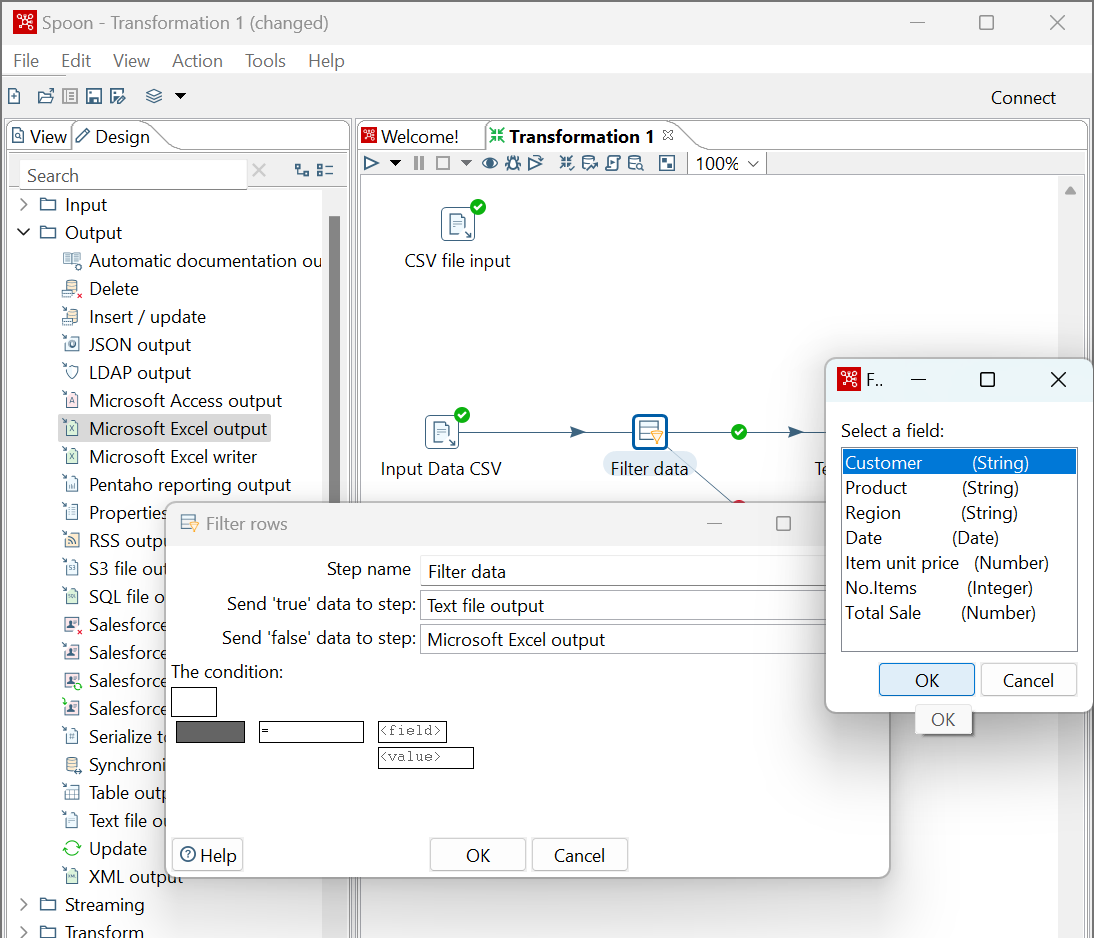
1. **Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.**

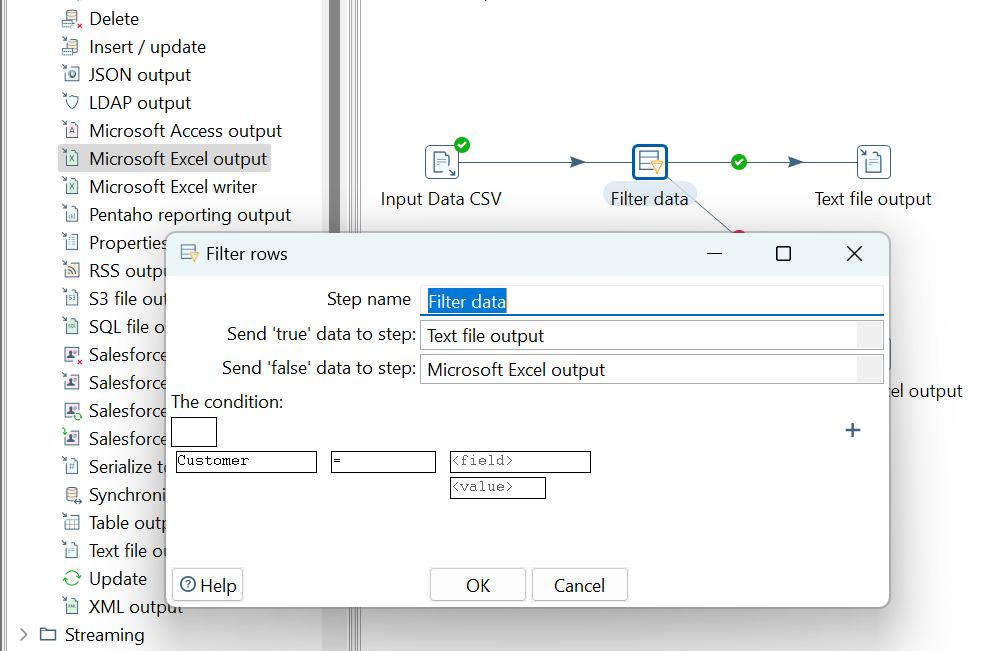
****

1. **Klik <Field> pada bagian condition**

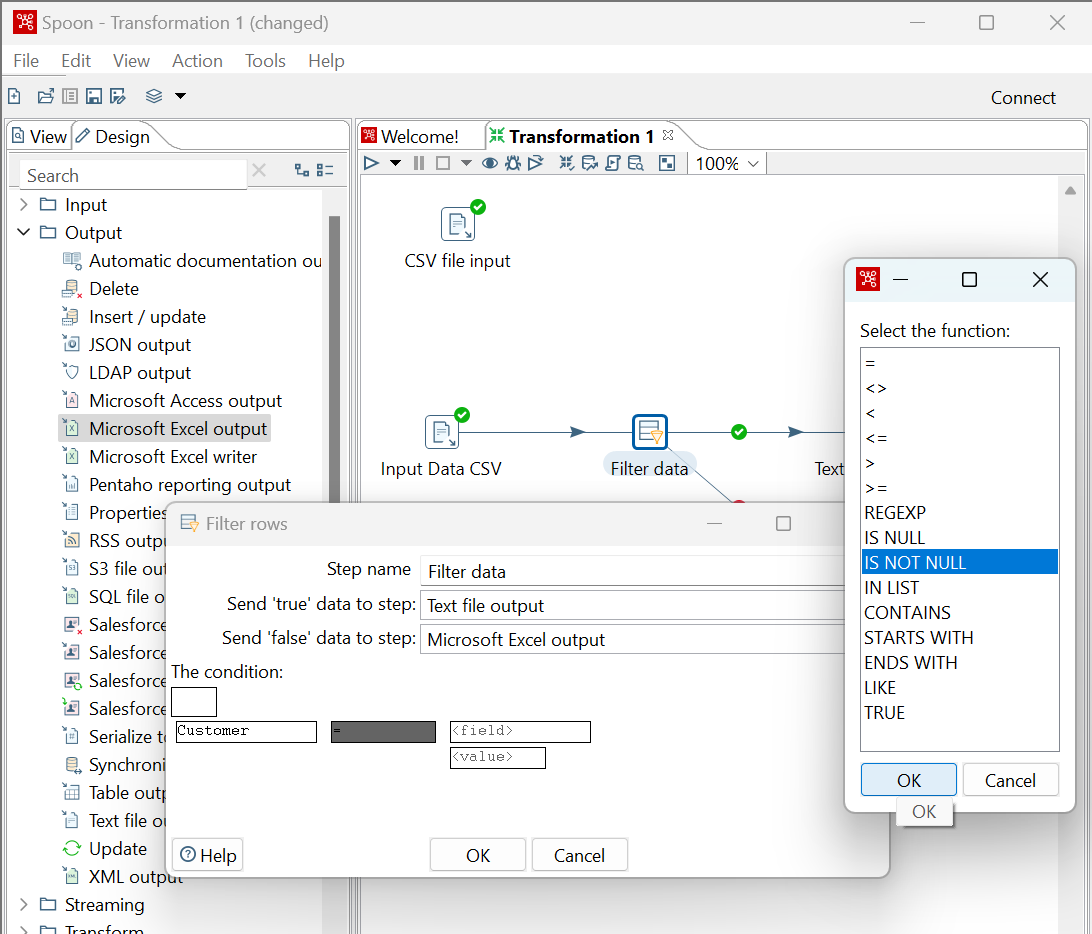
****

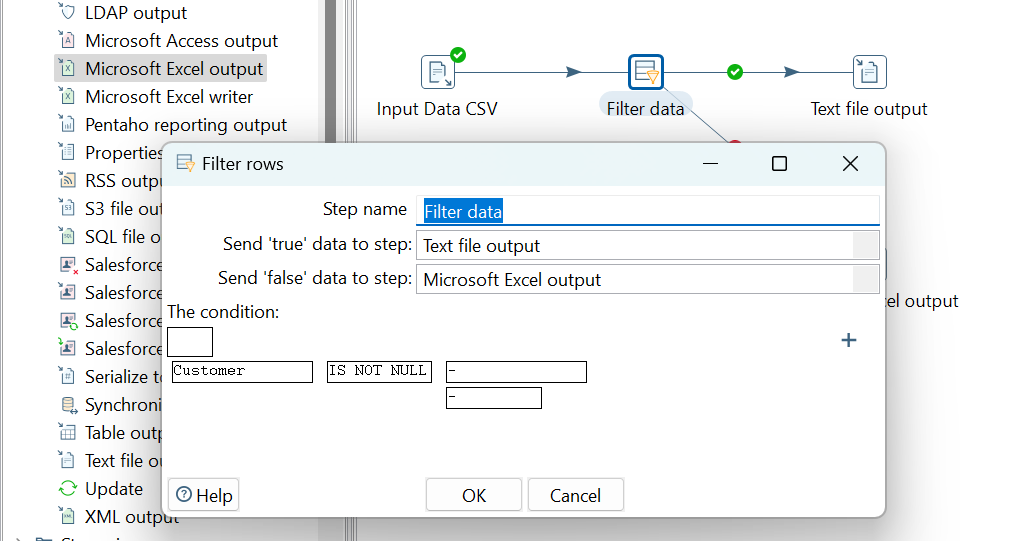
1. **Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.**

****

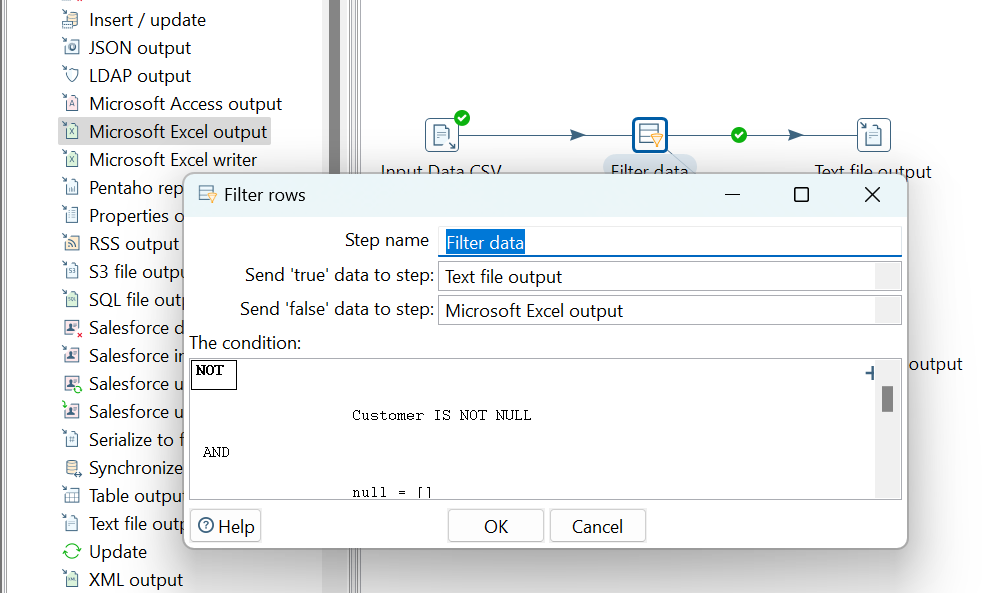
****

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

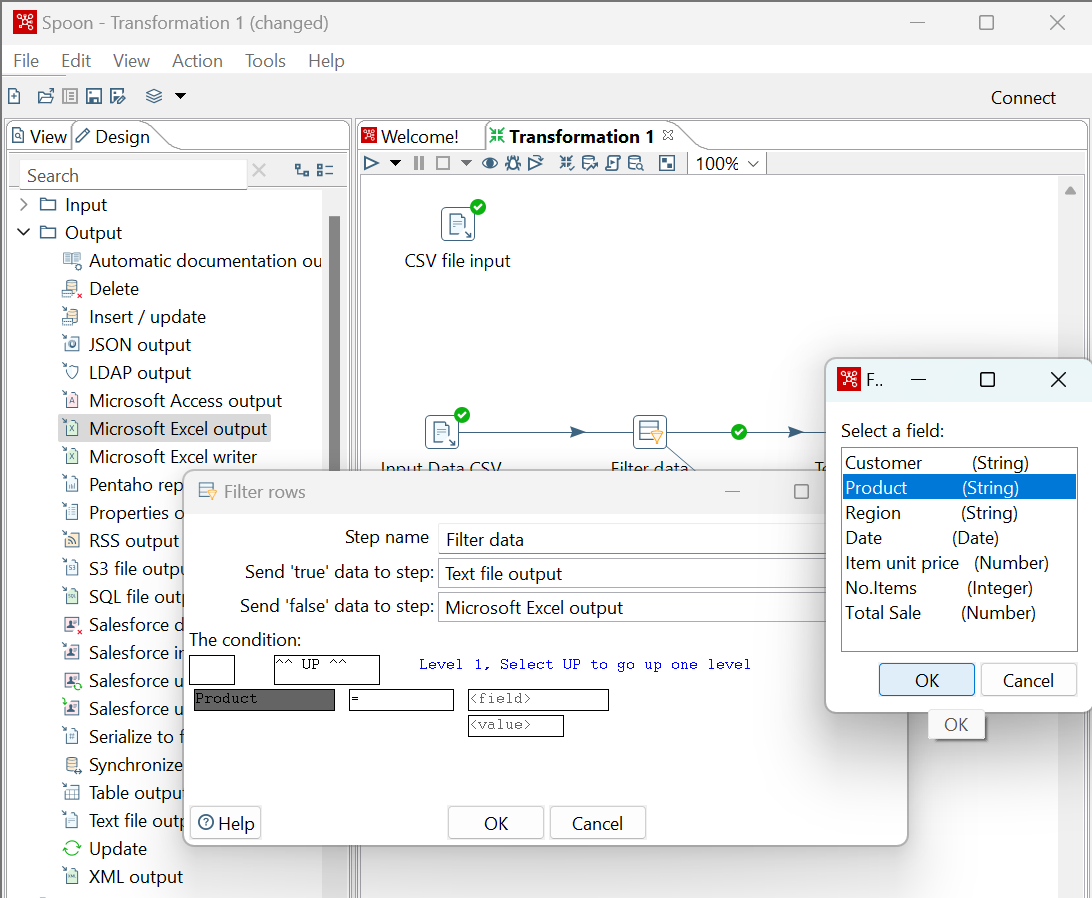
****

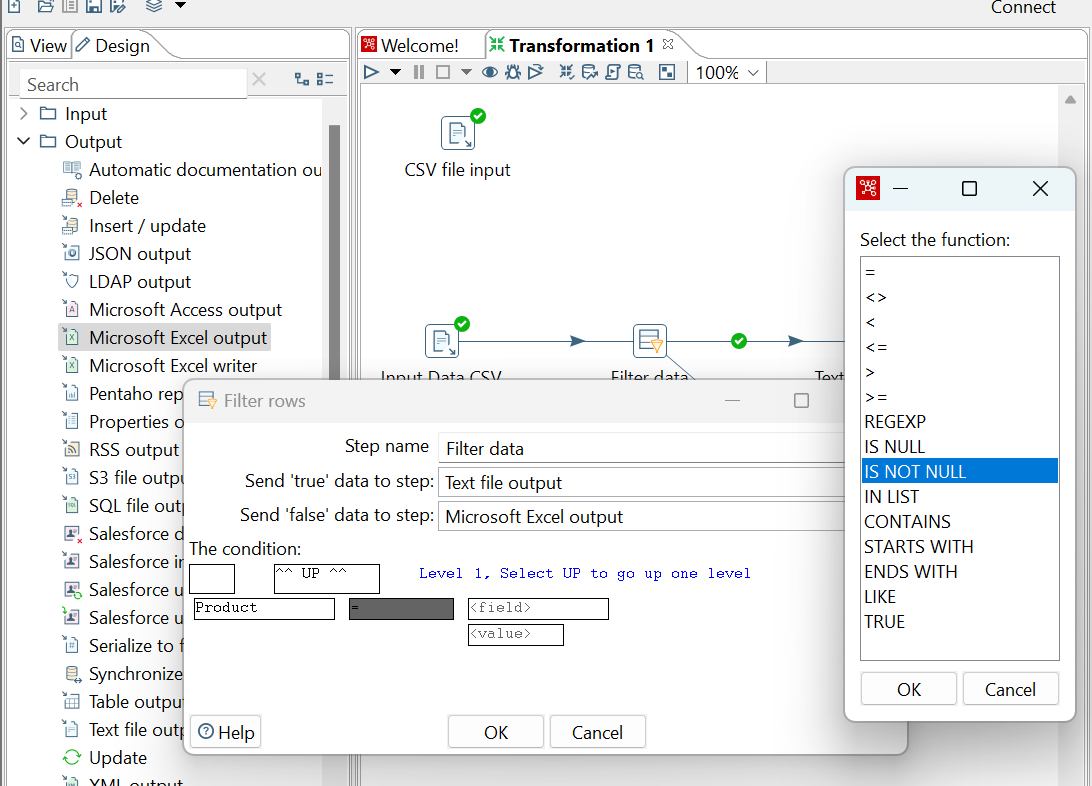
****

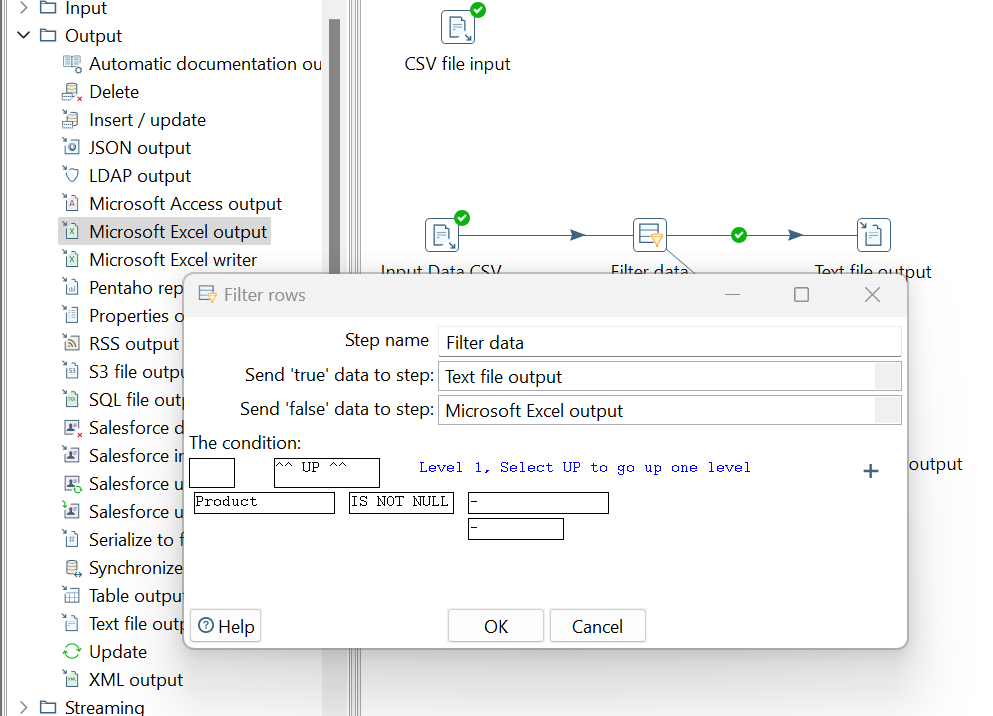
1. **Tekan tombol + untuk menambah kondisi.**

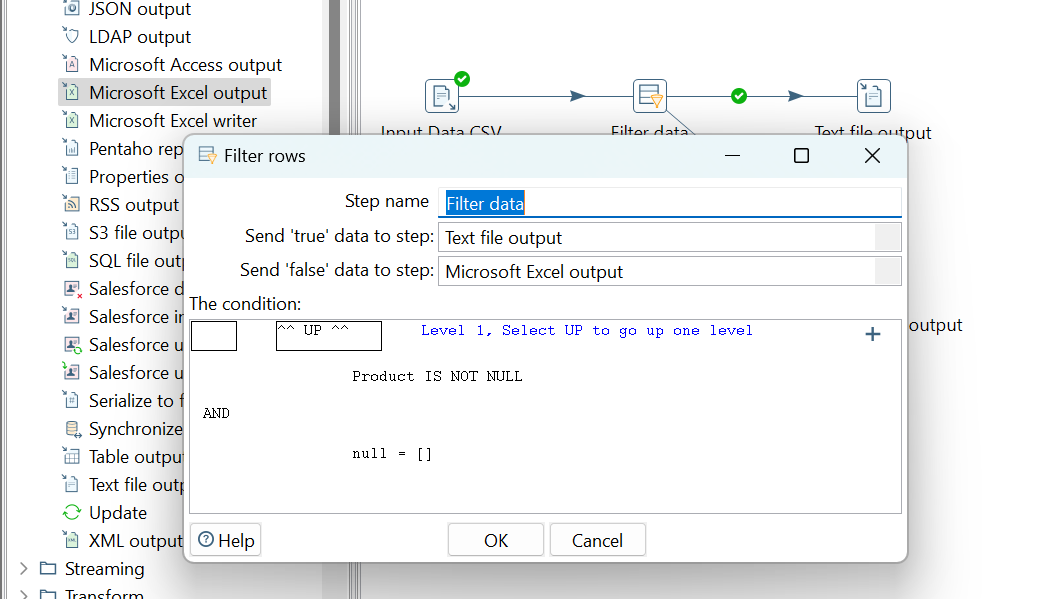
****

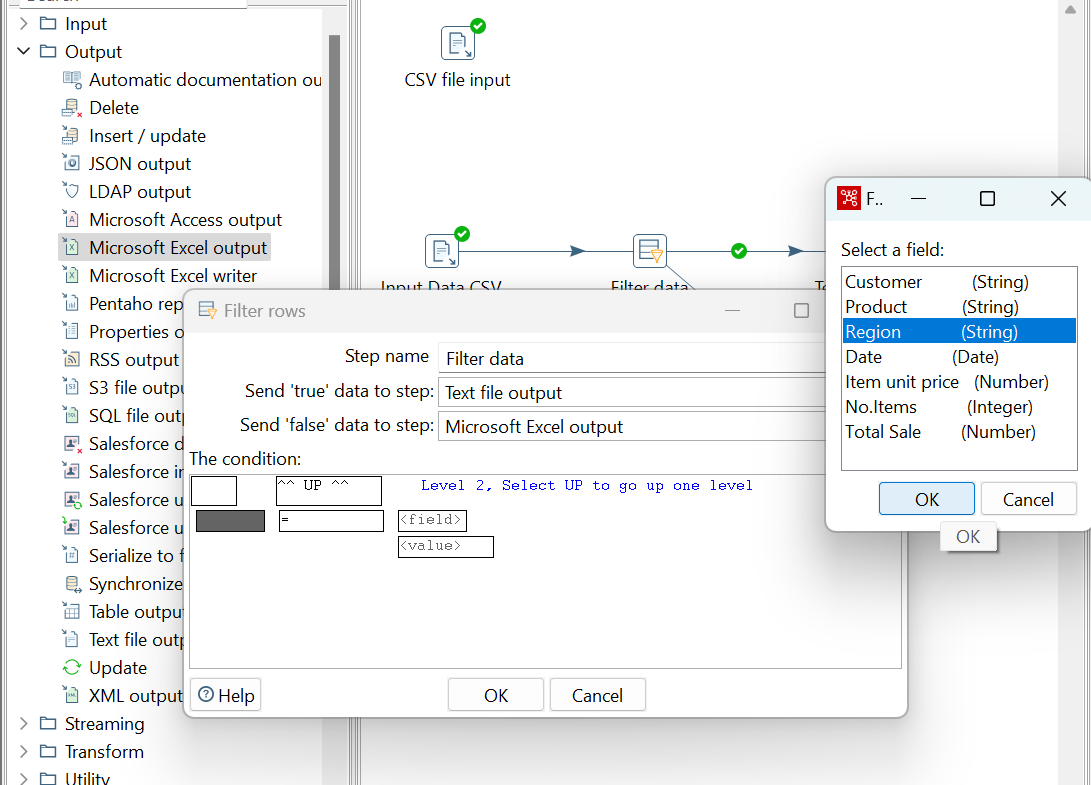
1. **Pilih statement yang masih null untuk menambah kondisi.**
2. **Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.**
3. **Gunakan operator AND untuk menghubung**

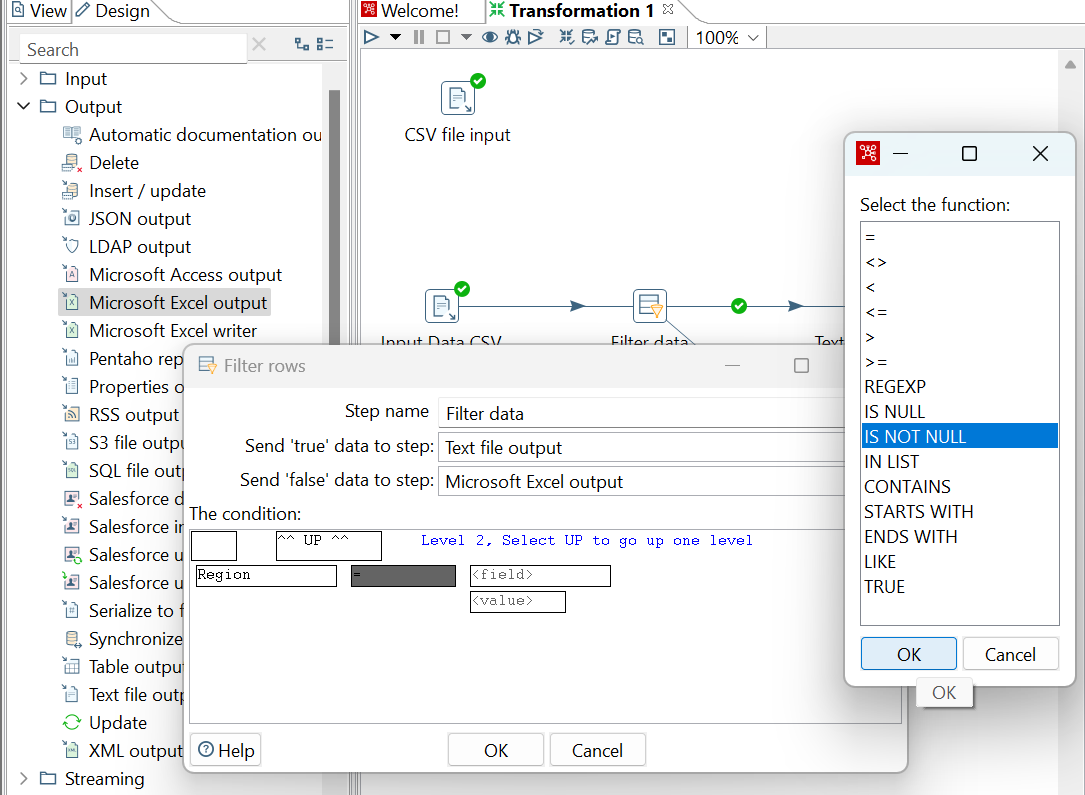
****

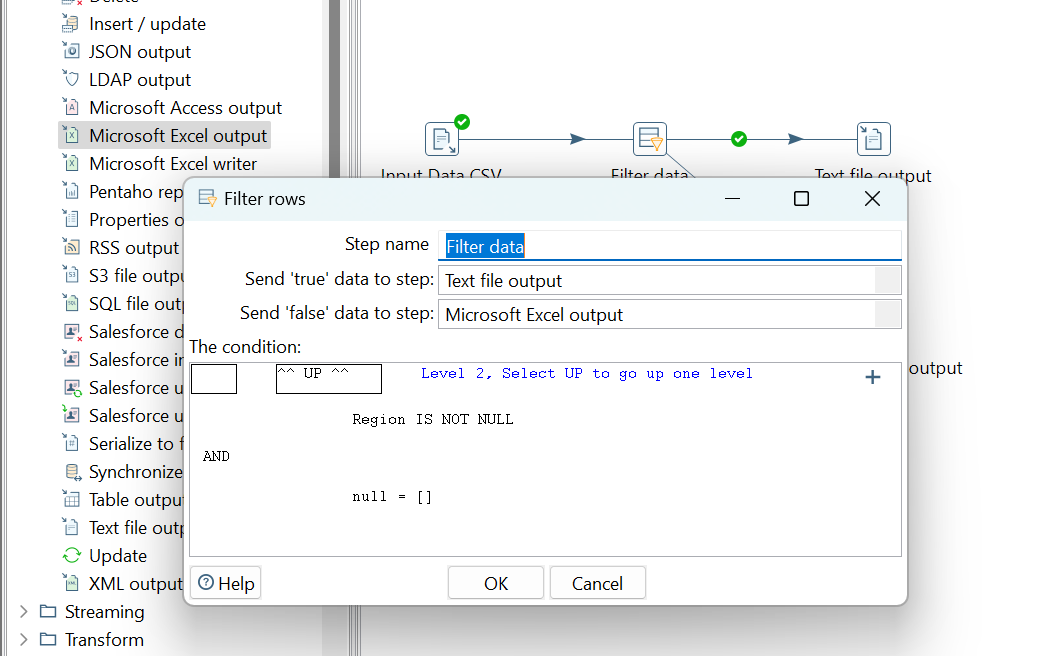
****

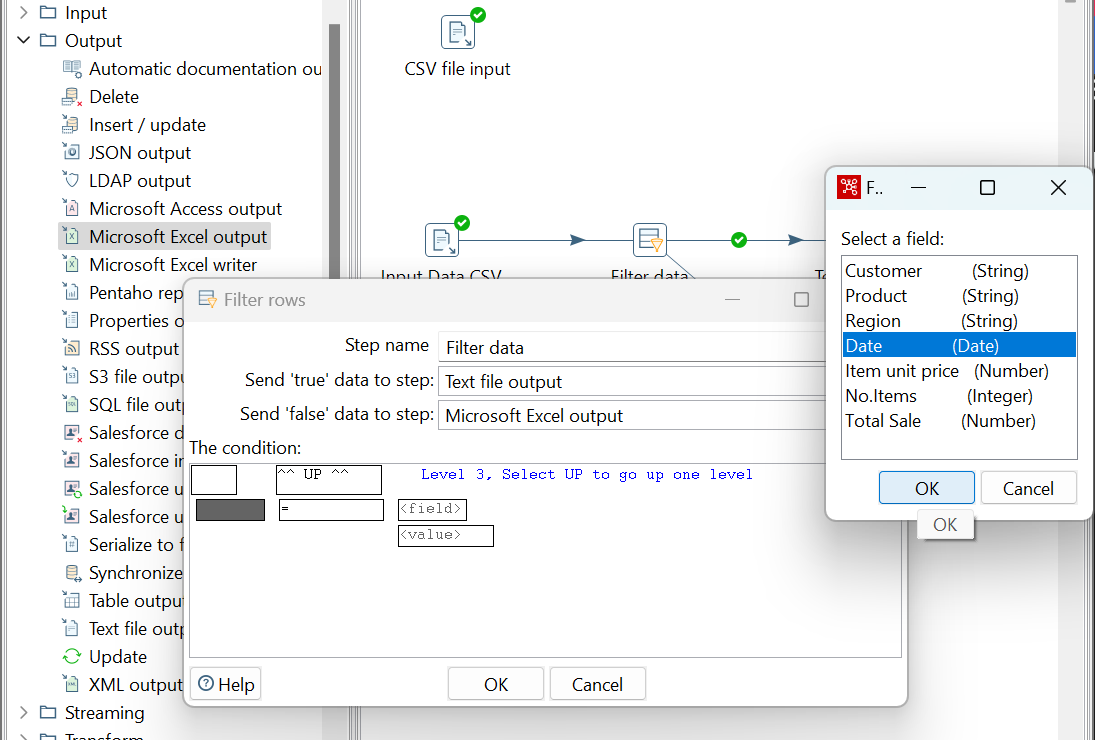
****

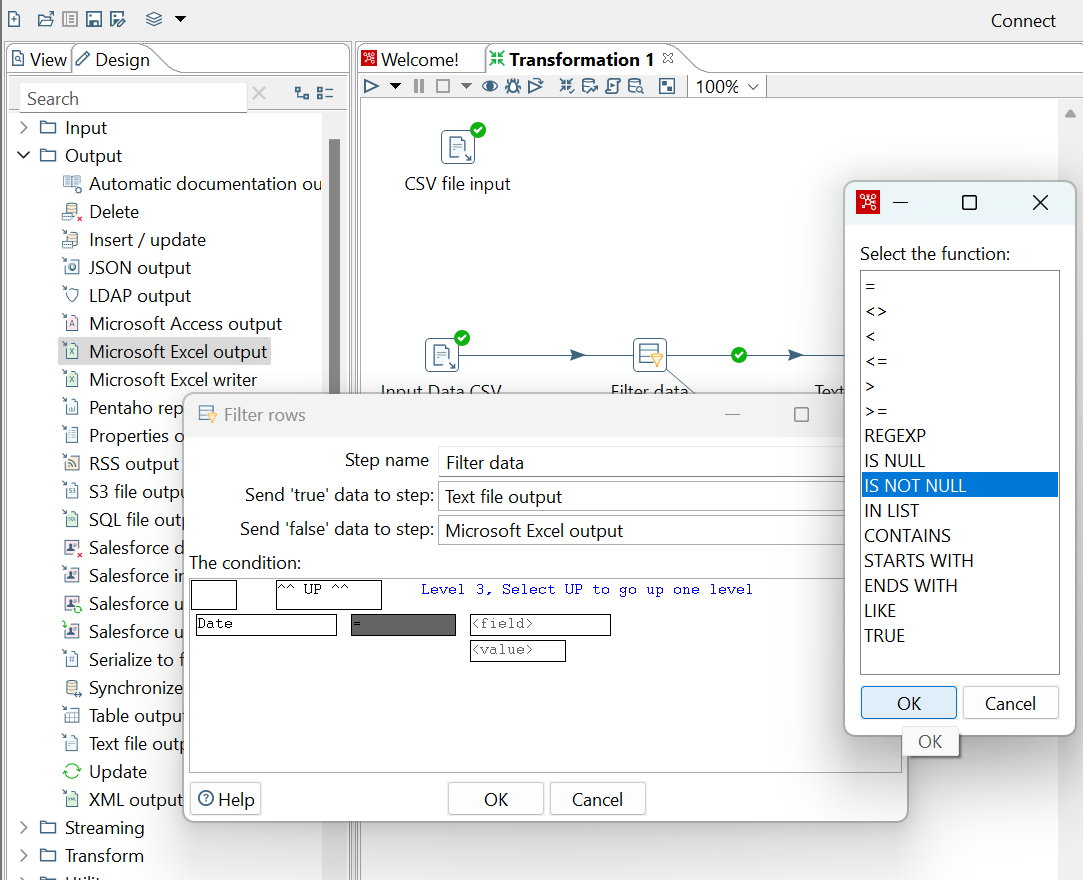
****

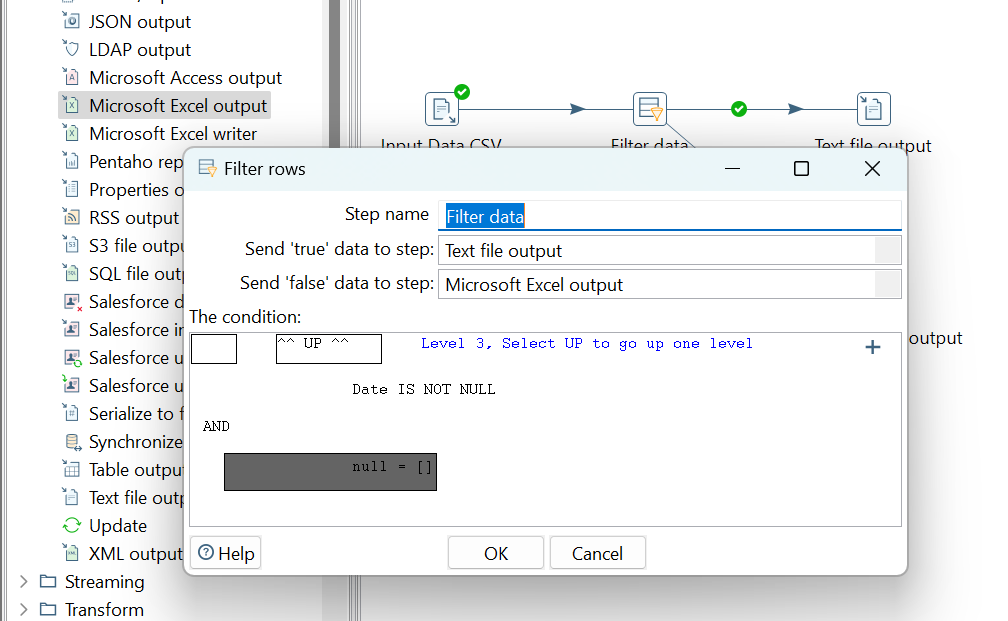
****

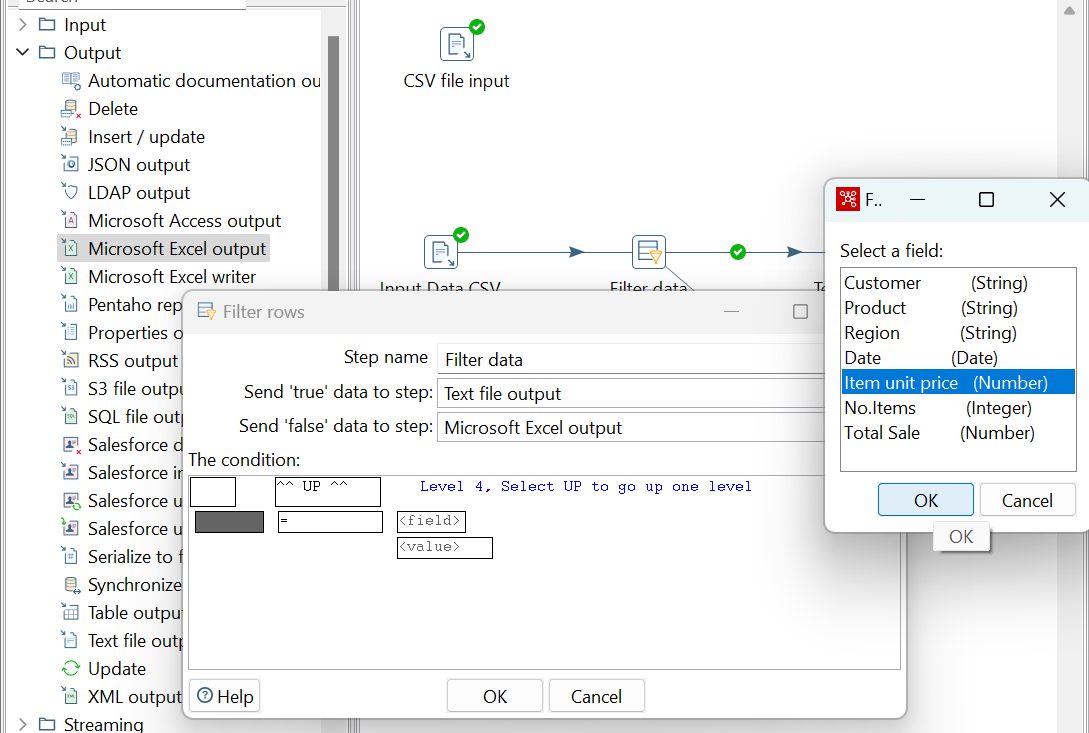
****

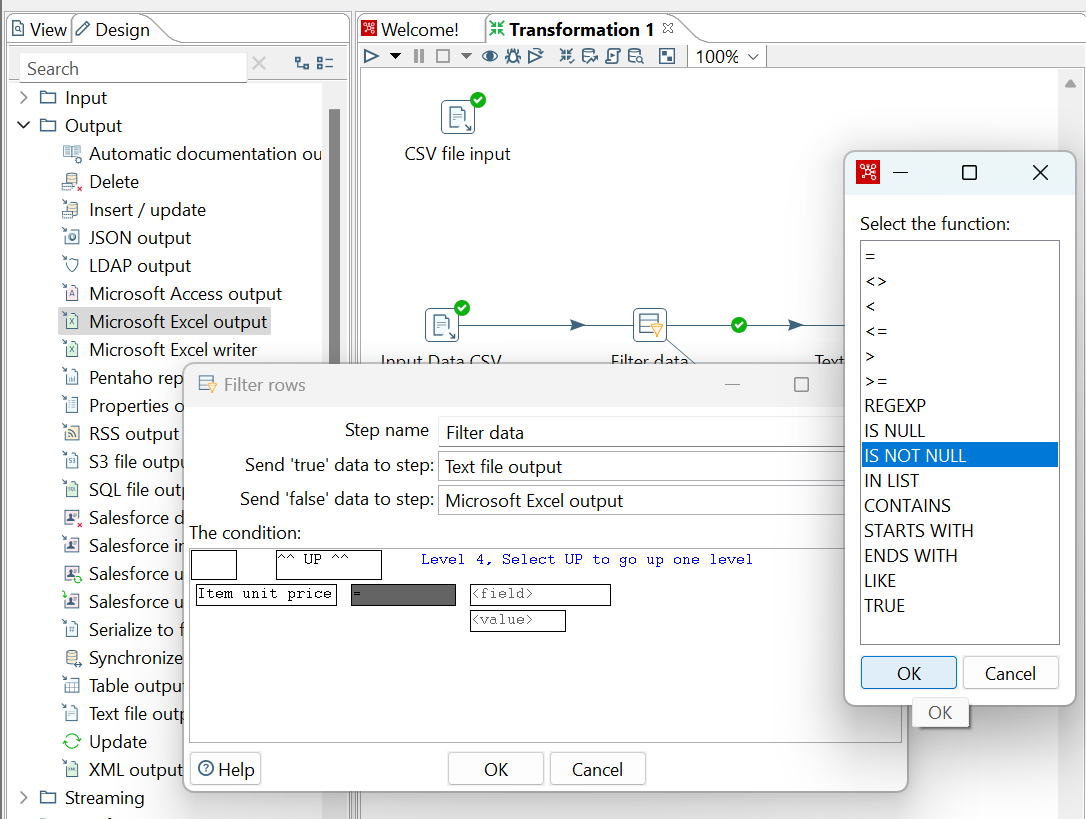
****

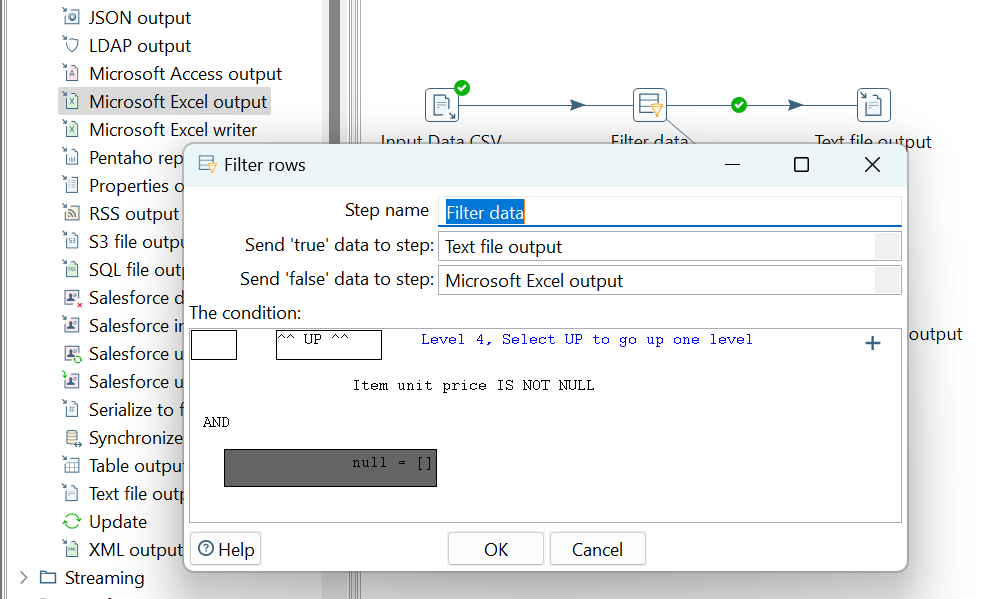
****

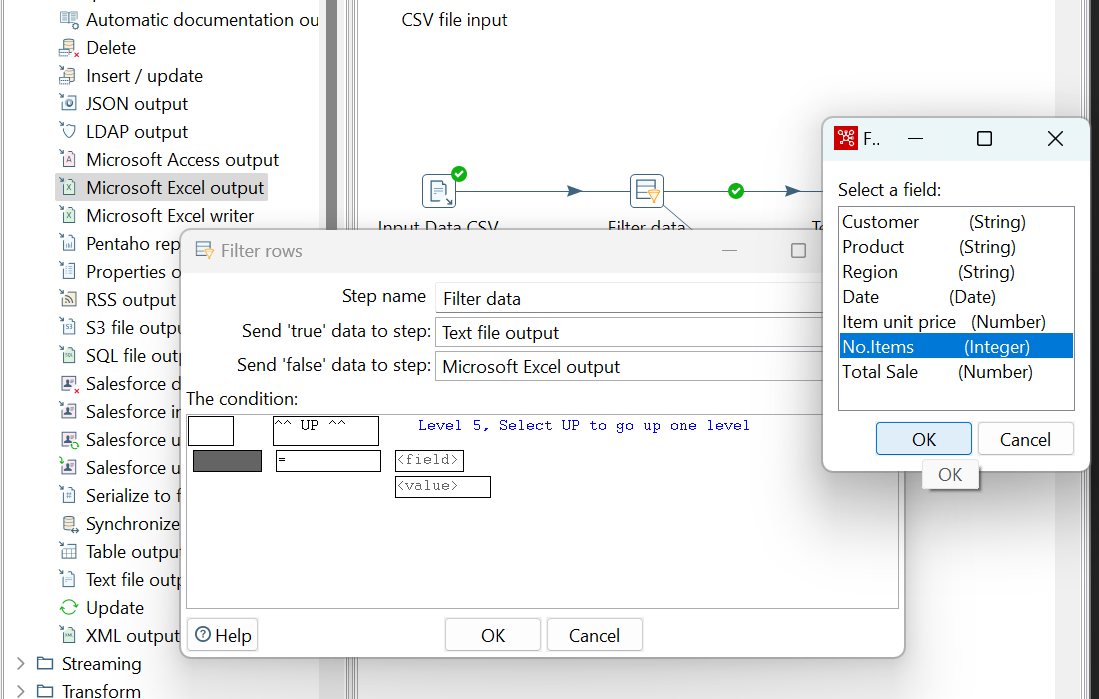
****

****

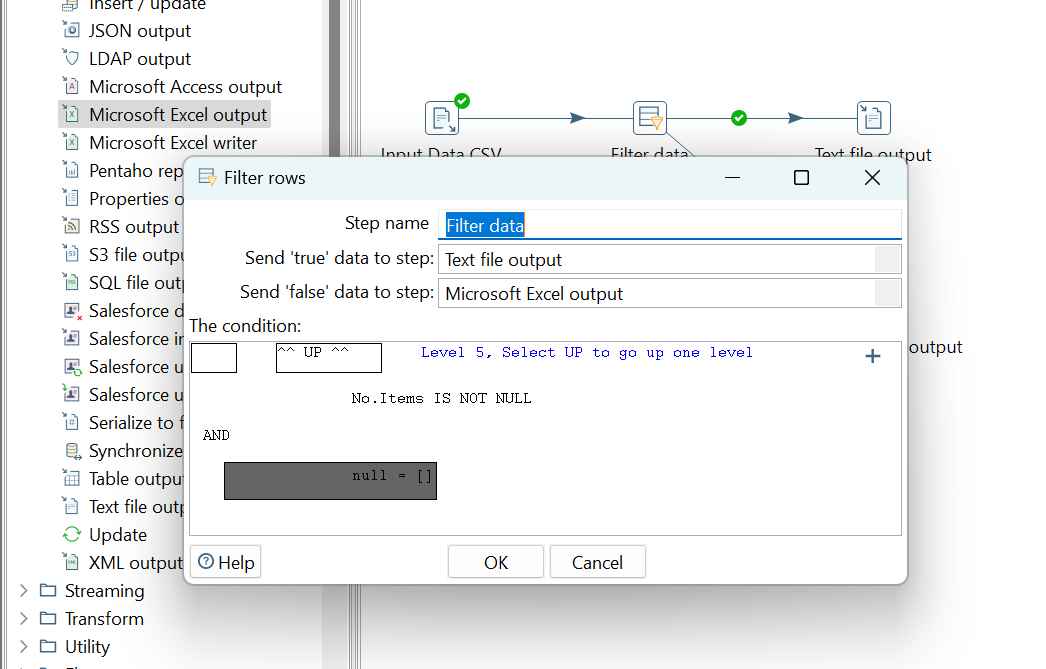
****

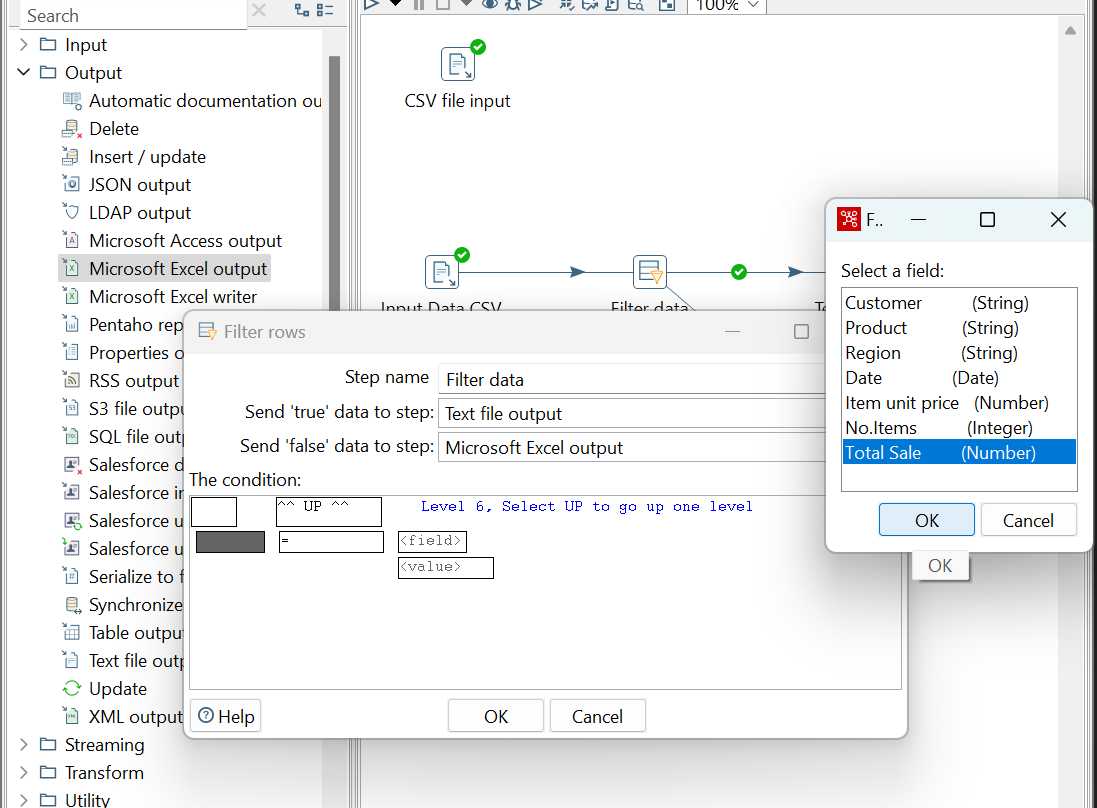
****

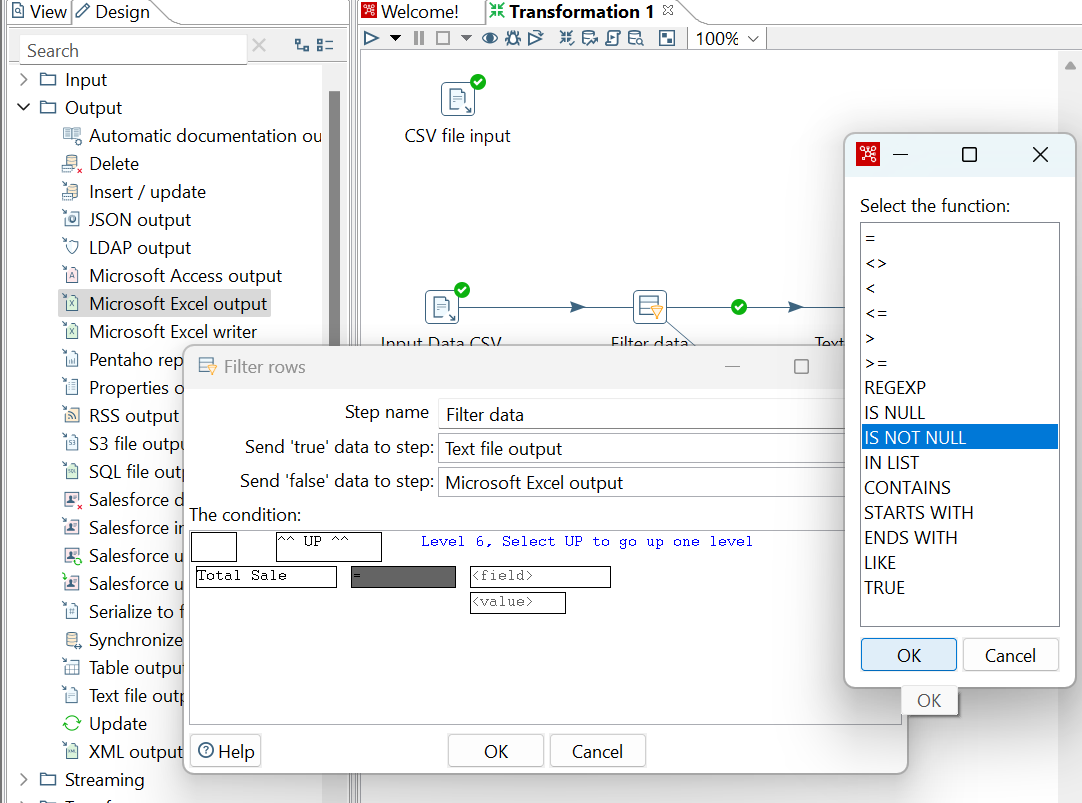
****

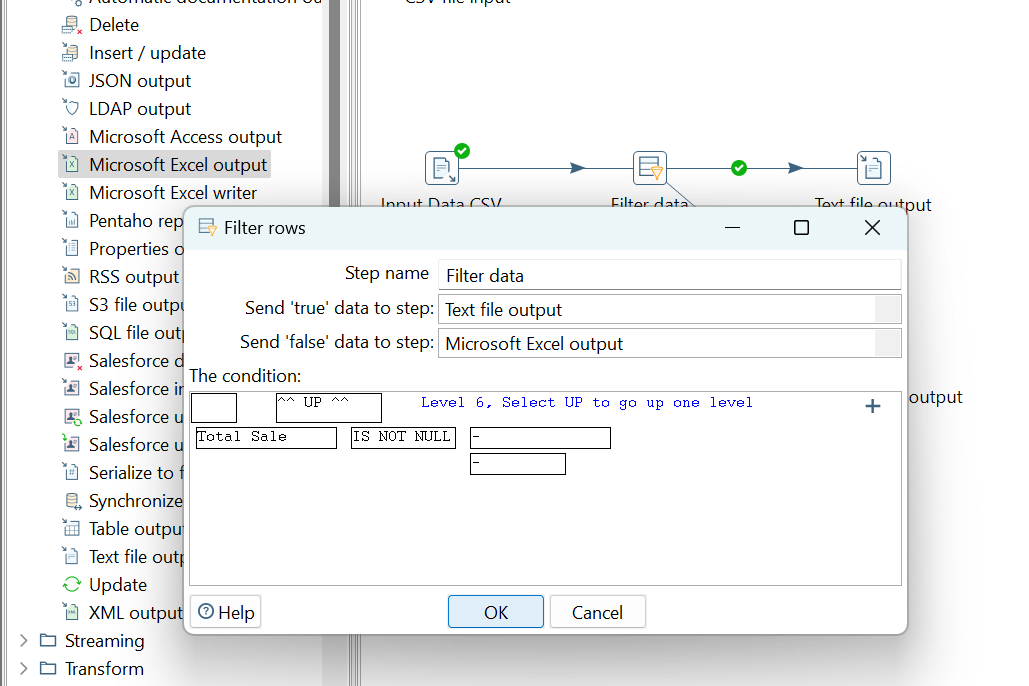
****

****

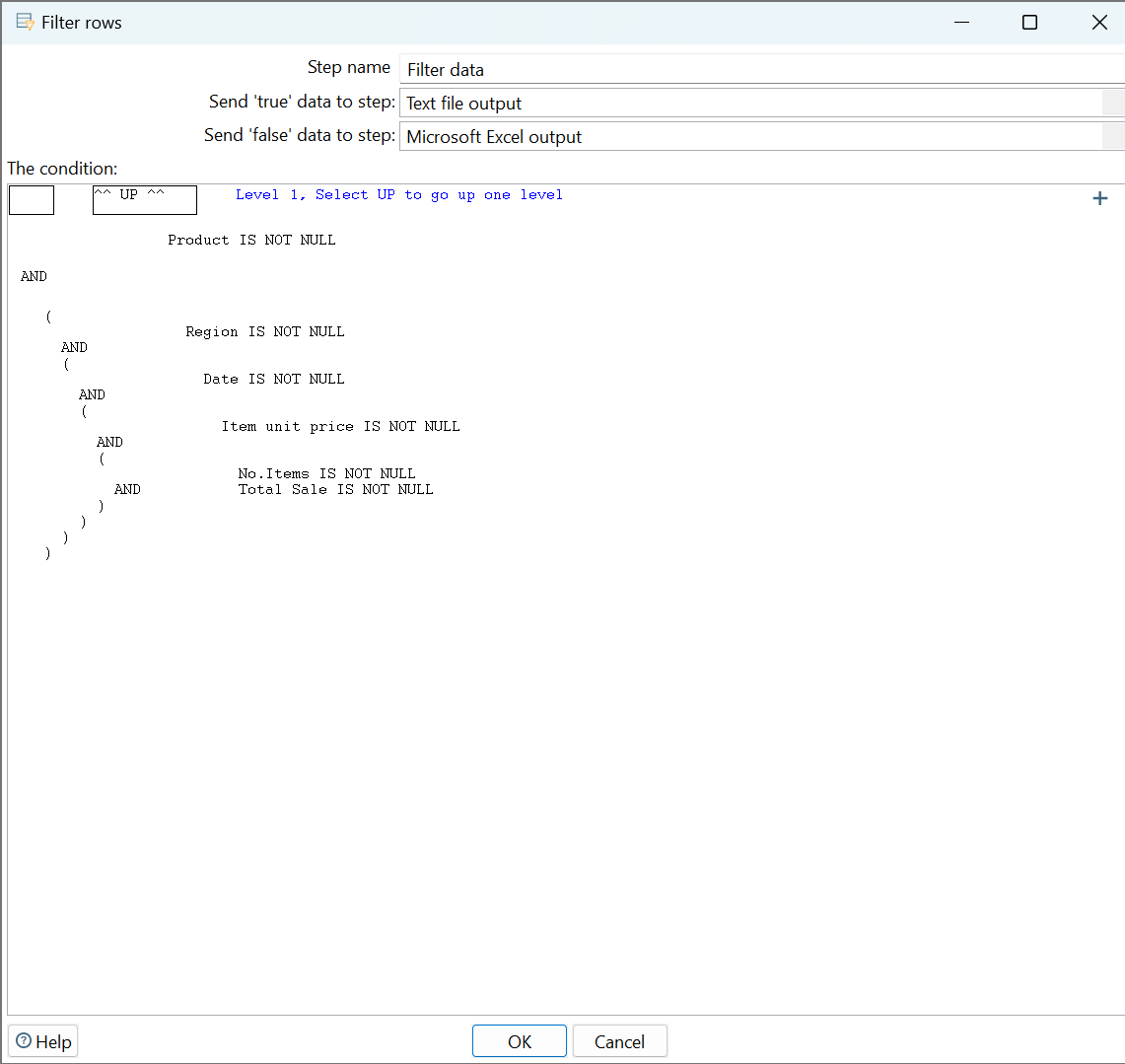
****

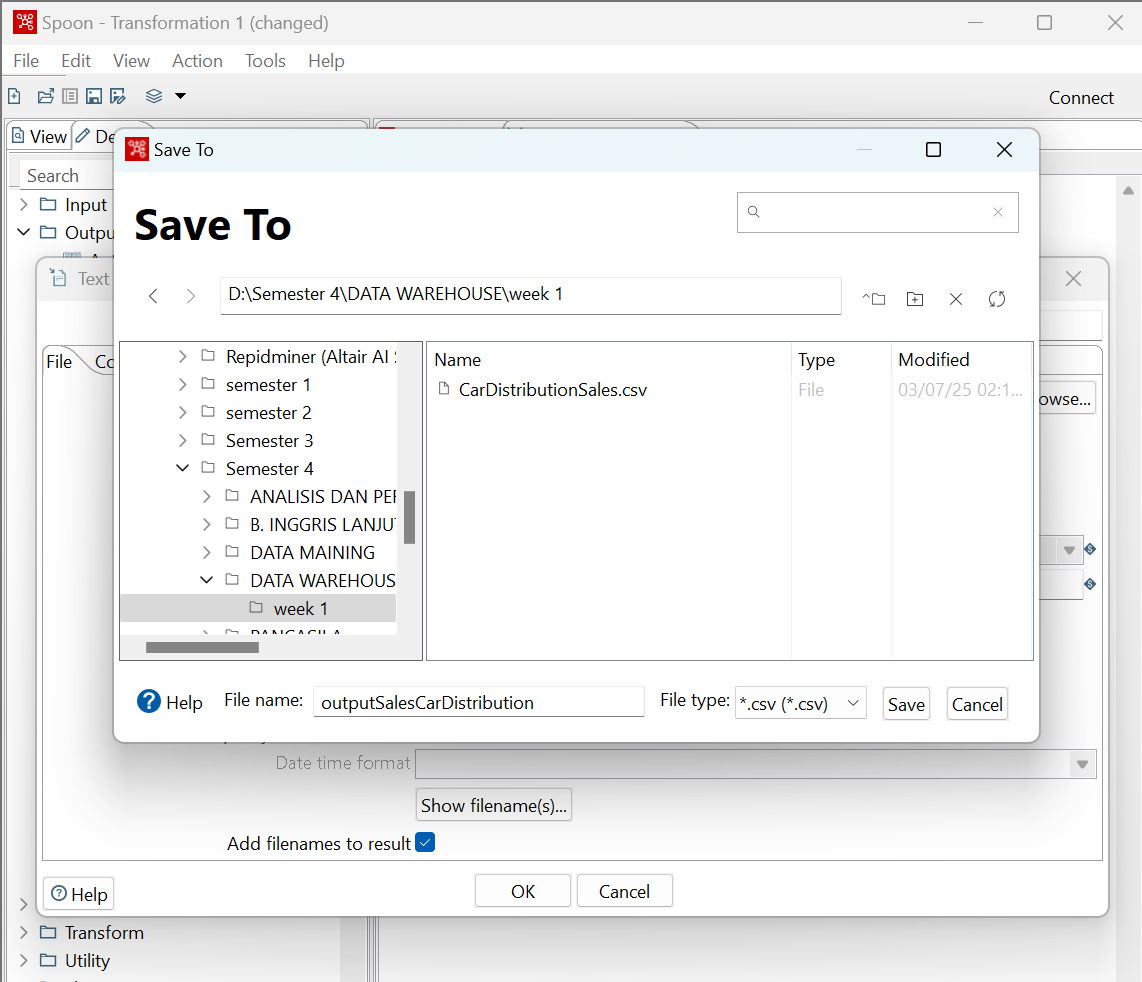
****

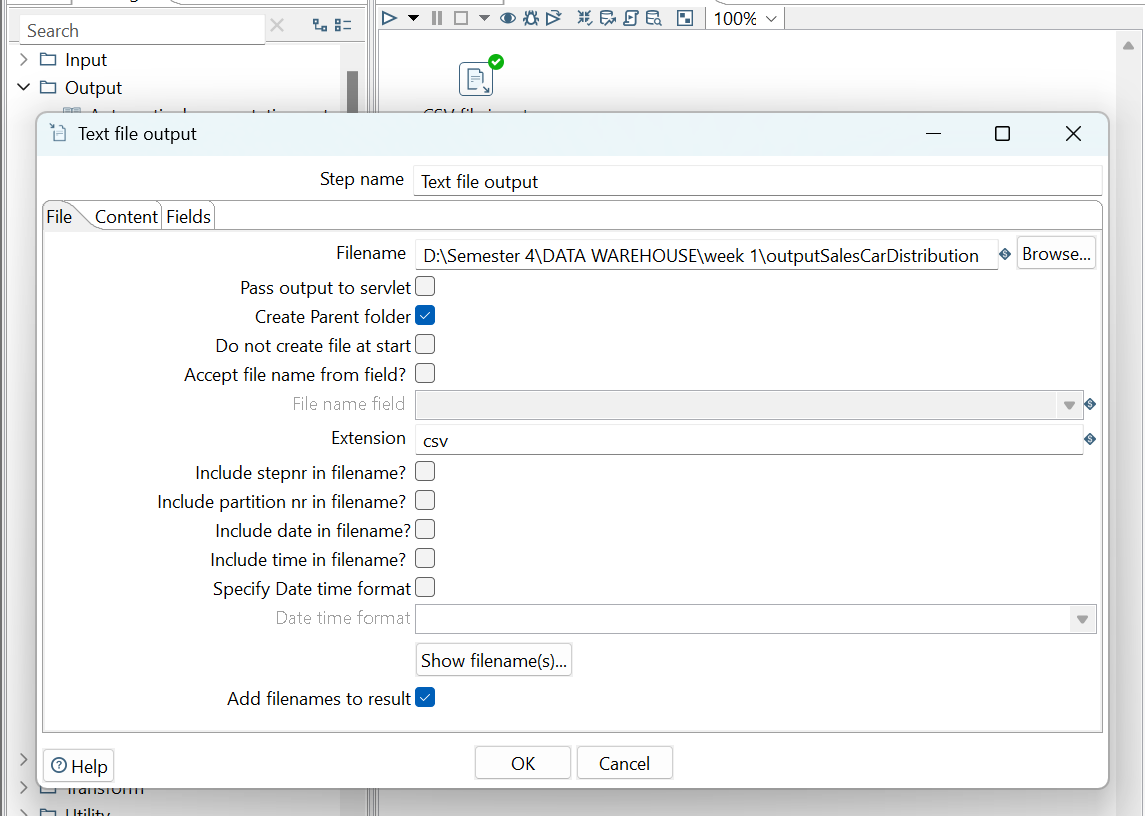
****

****

1. **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.**

****

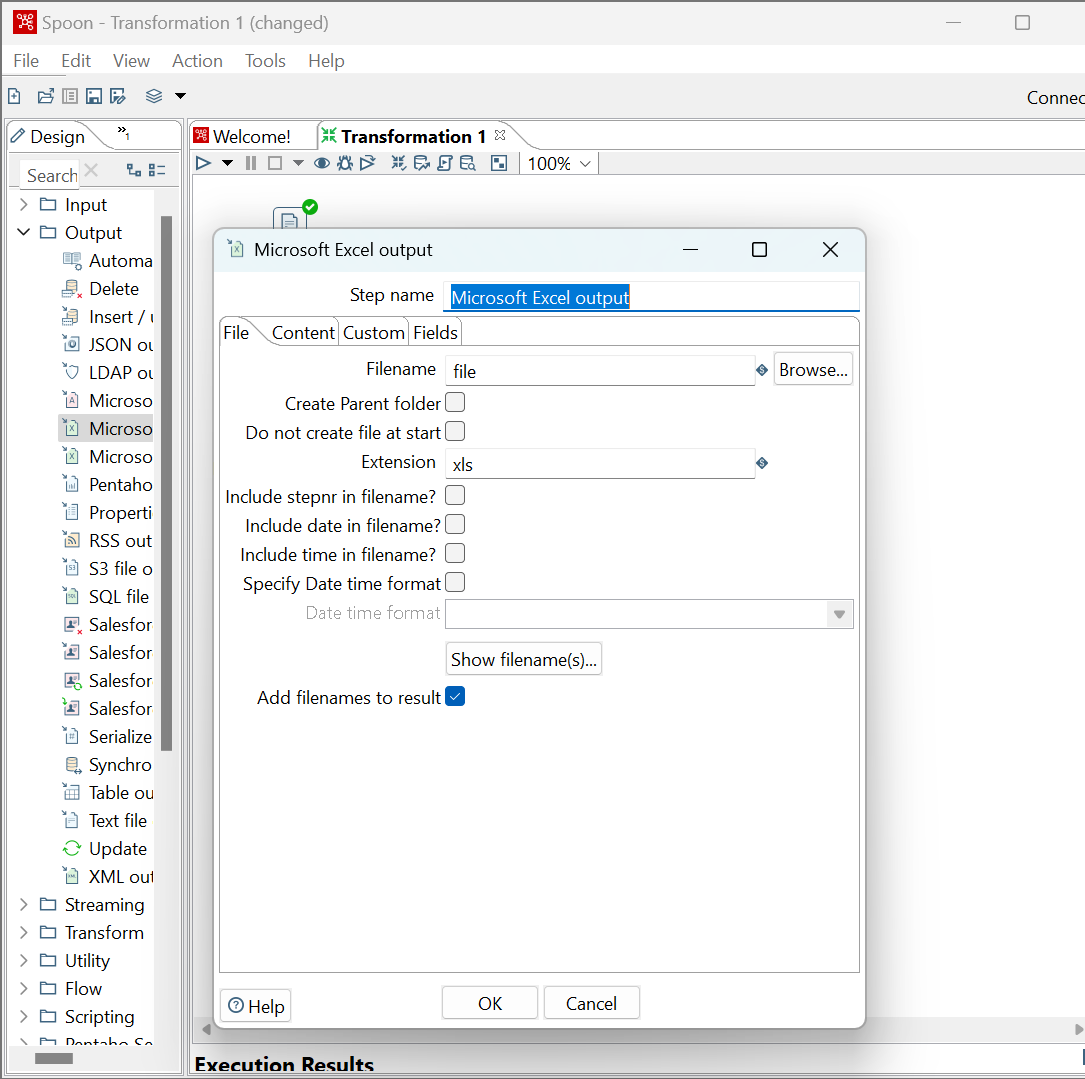
1. **Pada tab Fields, tekan tombol Get fields untuk mengambil field data.**

****

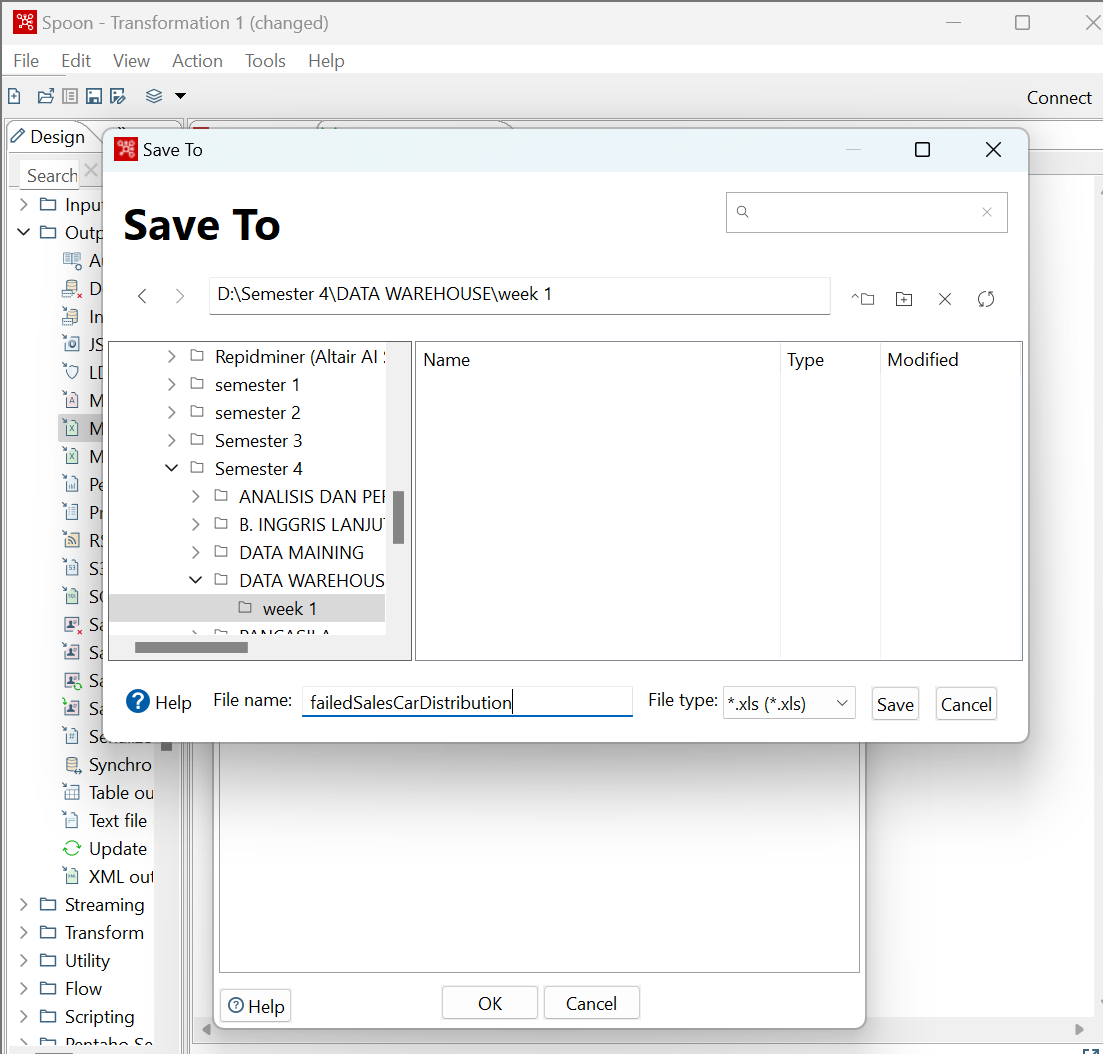
1. **Tekan ok untuk kembali ke work area**

****

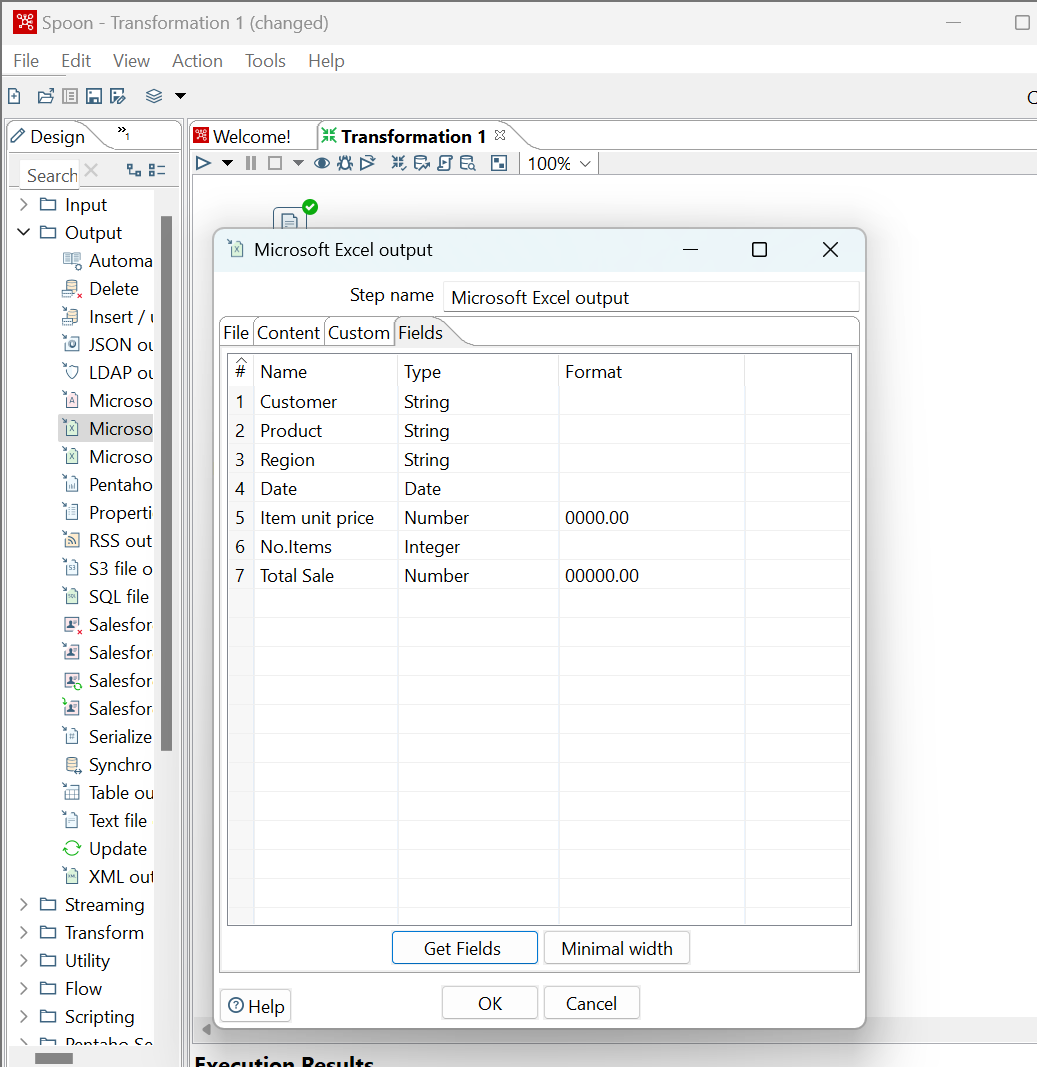
1. **Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.**

****

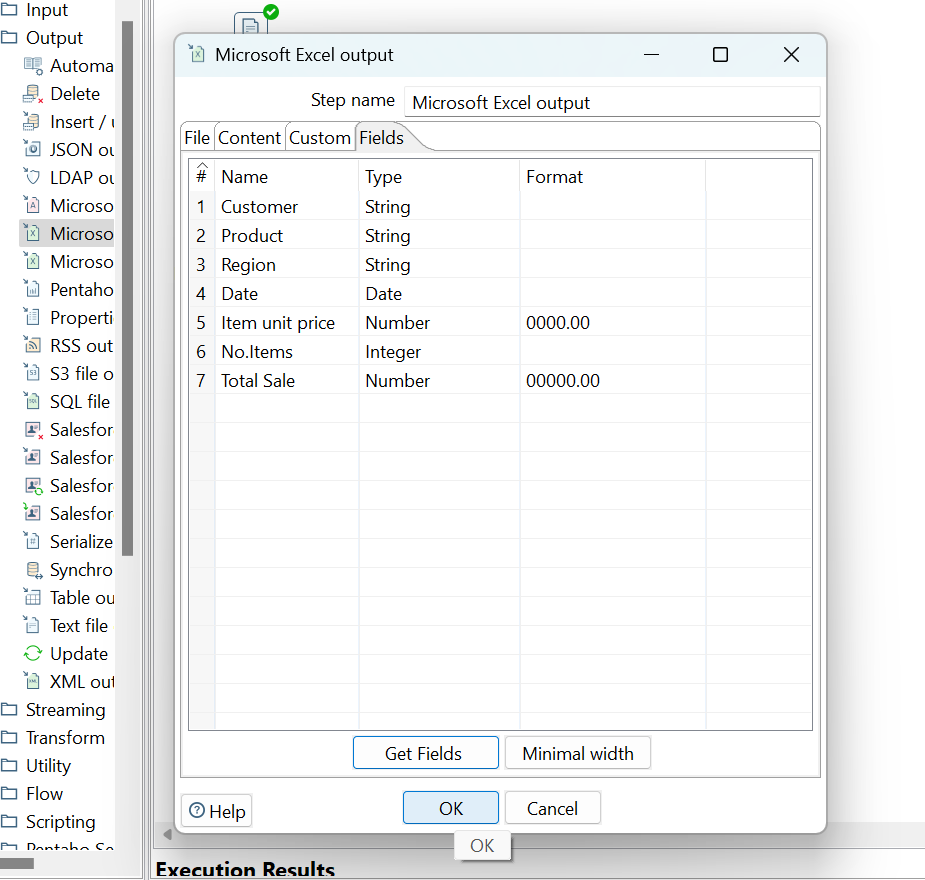
1. **Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan extension xls.**

****

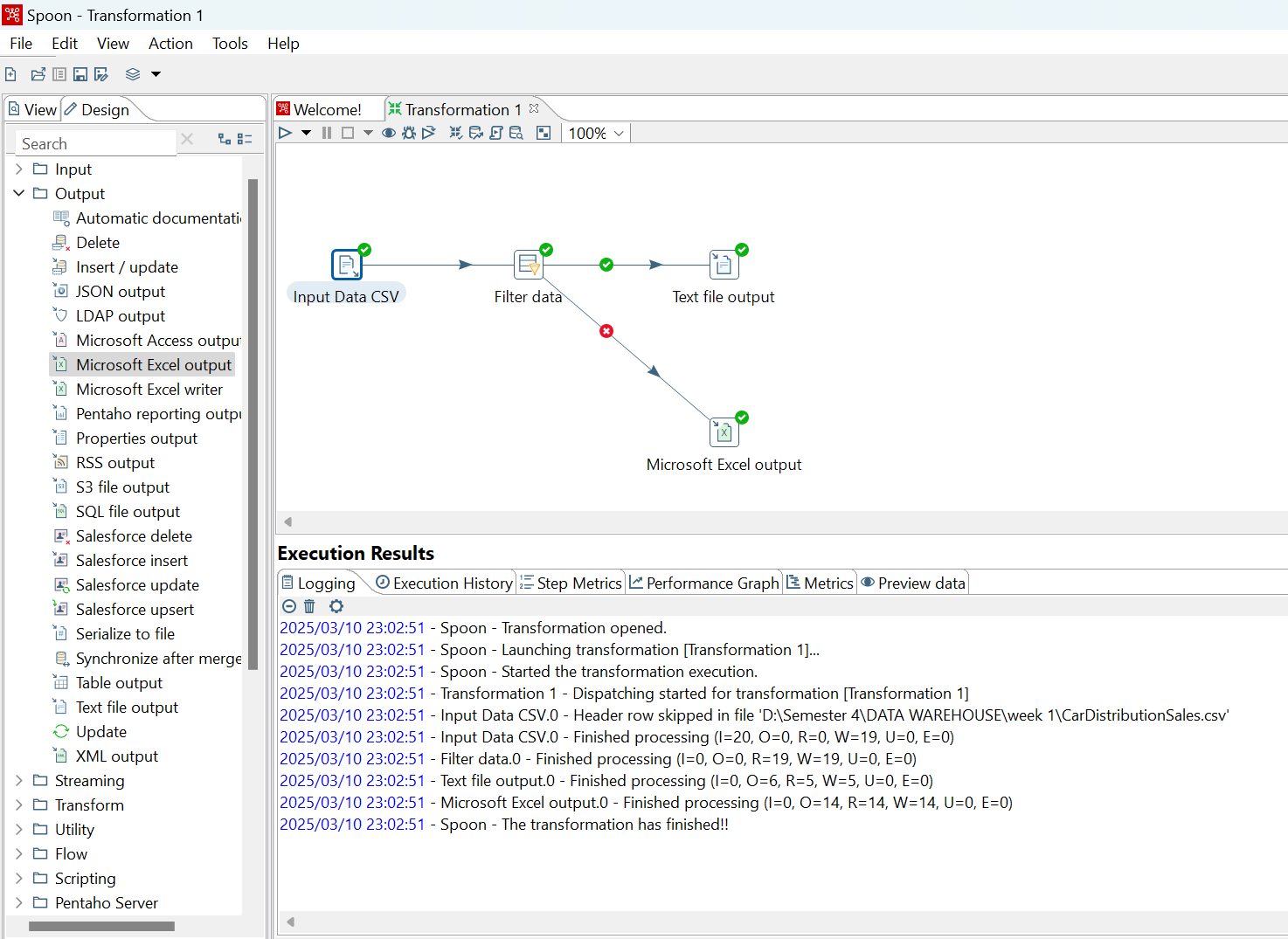
1. **Pada tab fields tekan tombol Get fields untuk mengambil field data.**

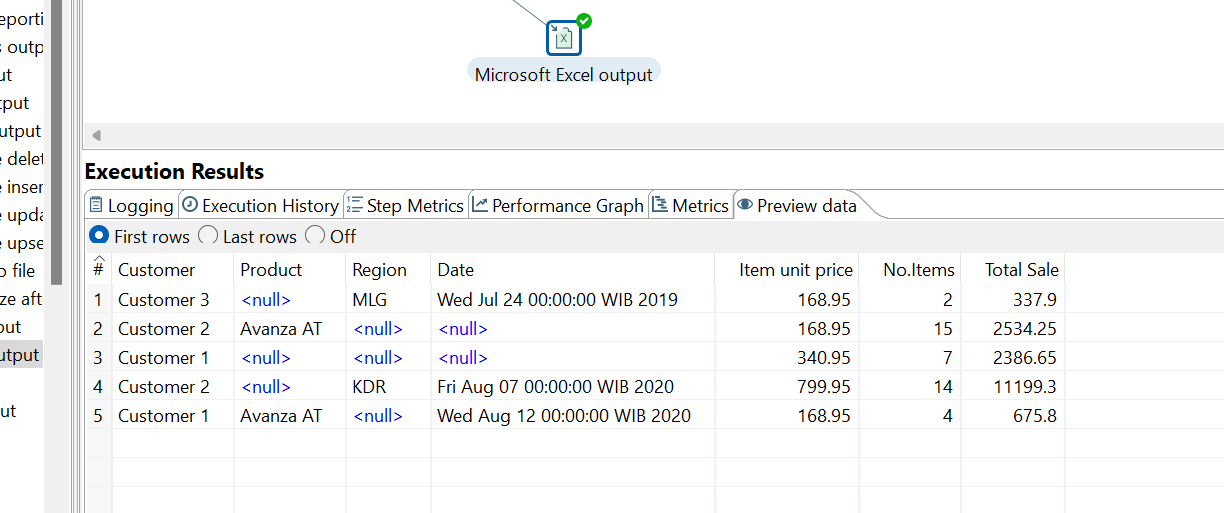
****

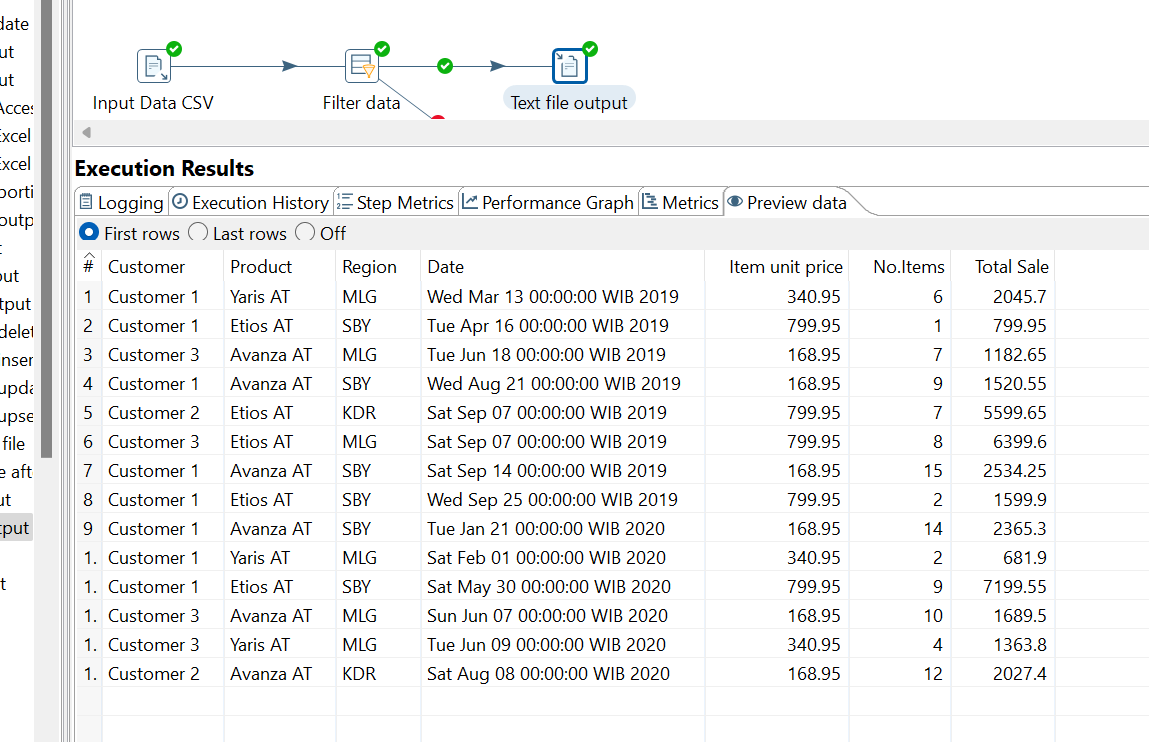
1. **Tekoan ok untuk kembali ke work area.**

****

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?**

**My Answer**

Perbedaannya bis akita lihat dari format penyimpanan datanya, yang dimana format CSV itu disimpan dalam bentuk teks biasa datanya. Sedangkan kalau dengan format XLS itu disimpan dalam bentuk format biner/XML yang kompleks.

Kita bisa juga lihat dari segi ukuran filenya karena berbeda juga dan stuktur datanya, dalam struktur data CSV itu tidak ada data yang khusus karena memakai teks biasa, sedangkan kalau XLS itu kita bisa mendukung tipe data seperti tanggal, angka, dll.

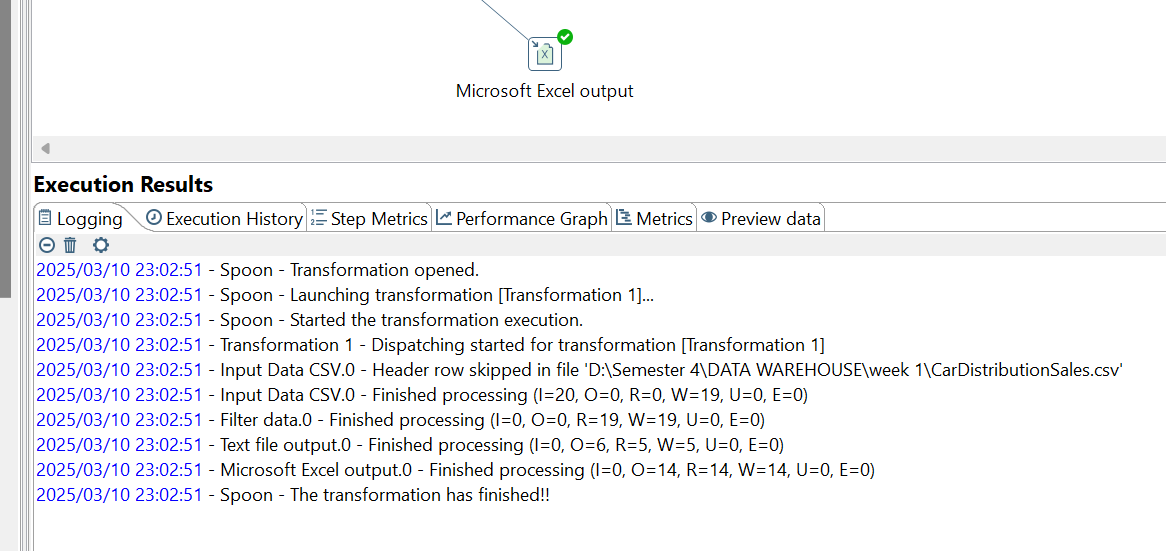
1. **Jelaskan apa yang terjadi pada proses Filter rows!**

**My Answer**

Dalam proses filter rows, CSV input data masuk ke Langkah Filter rows (filter data). Kemudian, Data yang memenuhi syarat (true) akan diteruskan ke output pertama (dalam gambar ke Text file output). Sedangkan Data yang tidak memenuhi syarat (false), akan diteruskan ke output kedua (dalam gambar ke Microsoft Excel output, tapi ada error di sana).

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**

**My Answer**

****

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Processor** | **RAM** | **Storage** | **VGA** | **CPU** | **Time Of Execution** |
| **Shabrina Q** | 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz (8 CPUs), ~2.5GHz | 16384MB (16 GB) | SSD NVMe 477GB | Intel(R) Iris(R) Xe Graphics | 23% 2.25 GHz | 0 seconds  (or less  than 1  second) |
| Fakhira Z | Intel®  CoreTM i7-  13650HX  (13th Gen,  20 Cores) | 16 GB | SSD  477GB | Intel®  UHD  Graphics  +  NVIDIA  GeForce  RTX 4050 | 7% | 1 Second |
| Hanin M | Intel | 16 GB | 512 GB  SSD | GeForce  RTX  2050 | Intel(R)  Core(TM)  i7-  12650H | 0 seconds  (or less  than 1  second) |
| Sabrina R | Intel(R) Core(TM) i5-7300U CPU @ 2.60GHz (4 CPUs), ~2.7GHz | 8192MB RAM  (8 GB) |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**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 tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Pak Pascanowo mendapatakan data penduduk yang akan pindah ke wilayahnya melalui link: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dataDummyKependudukan.csv>

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.