**LAPORAN TUGAS DATA WAREHOUSE**

**Jobsheet 1 – Pengenalan PDI Spoon**

****

Disusun Oleh:

**Evan Diantha Fafian**

**2341760163**

**D4 SISTEM INFORMASI BISNIS**

**TEKNOLOGI INFORMAS**

**POLITEKNIK NEGERI MALANG**

**2025**

**Tujuan Praktikum**

Setelah melakukan praktikum ini, mahasiswa diharapakan dapat lebih mengenal apa itu data warehouse berdasarkan uji coba membuat data warehouse sederhana

**Studi Kasus**

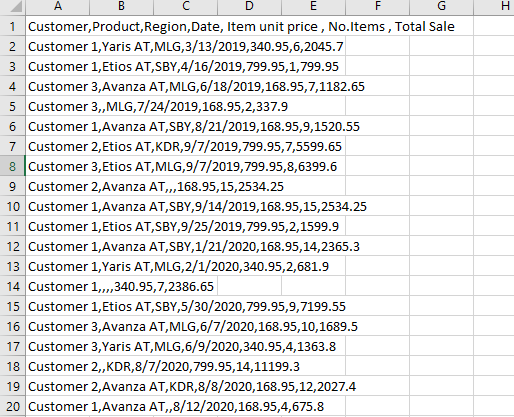
Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan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

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

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



Gambar Data Penjualan Mobil

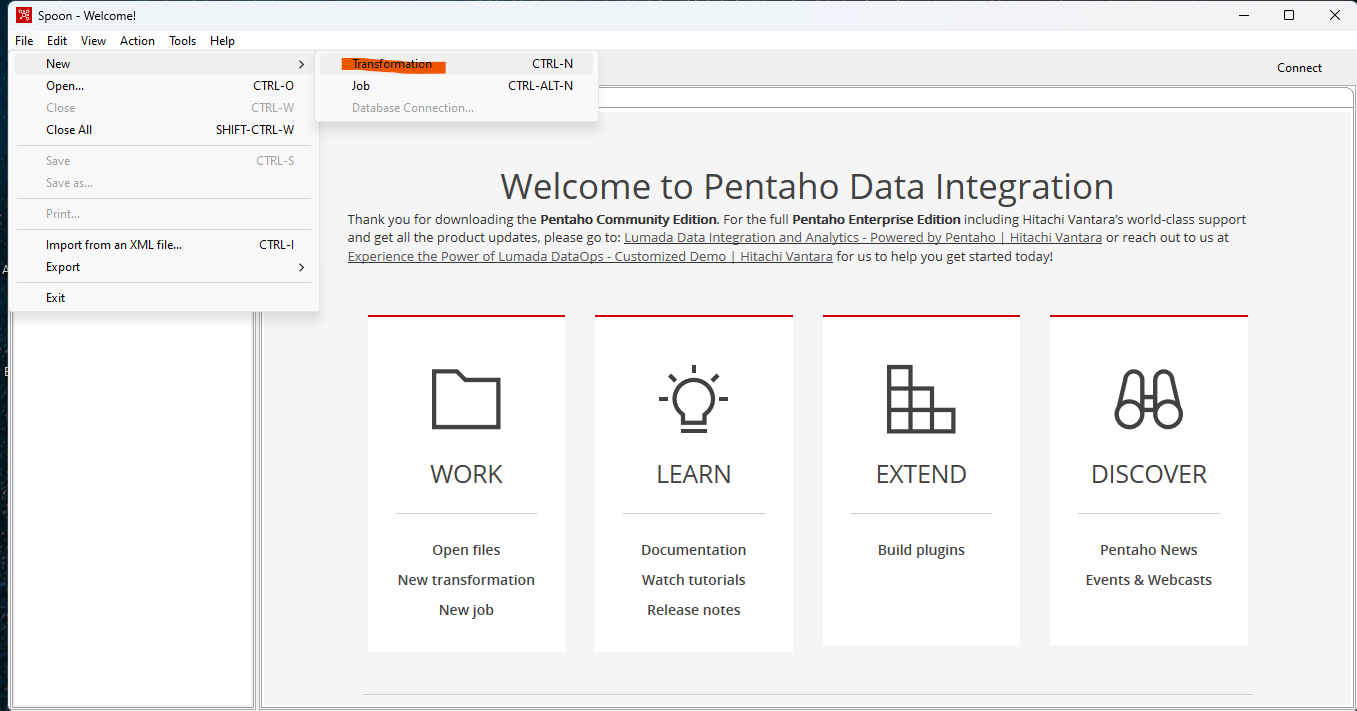
**TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?
   * Terdapat 7 kolom antara lain: Customer, Product, Region, Date, Item unit price, No. Items, Total Sales.
2. Apa arti atau isi data dari setiap kolom yang ada?
   * Customer : nama pelanggan
   * Product : jenis mobil
   * Region : kode wilayah
   * Date : tanggal transaksi (MM/DD/YYYY)
   * Item : harga tiap item
   * No. Items : jumlah unit yang terjual
   * Total Sales : total harga penjualan
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
   * Terdapat data yang kosong atau missing value, data yang kosong sebanyak 5 data. Dengan detailnya:
     1. Baris 4: Kolom Product.
     2. Baris 8: Kolom Region dan Date.
     3. Baris 13: Kolom Customer, Product, Region, dan Date.
     4. Baris 18: Kolom Product.
     5. Baris 21: Kolom Region dan Date.
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?
   * Tidak ada data yang memiliki tipe data yang berbeda pada kolom yang sama, karena semua data sudah sesuai dengan tipe data yang telah ditentukan.
     1. Customer: string
     2. Product: string
     3. Region: string
     4. Date: dat.
     5. Item unit price: double
     6. No.Items: integer
     7. Total Sale: double

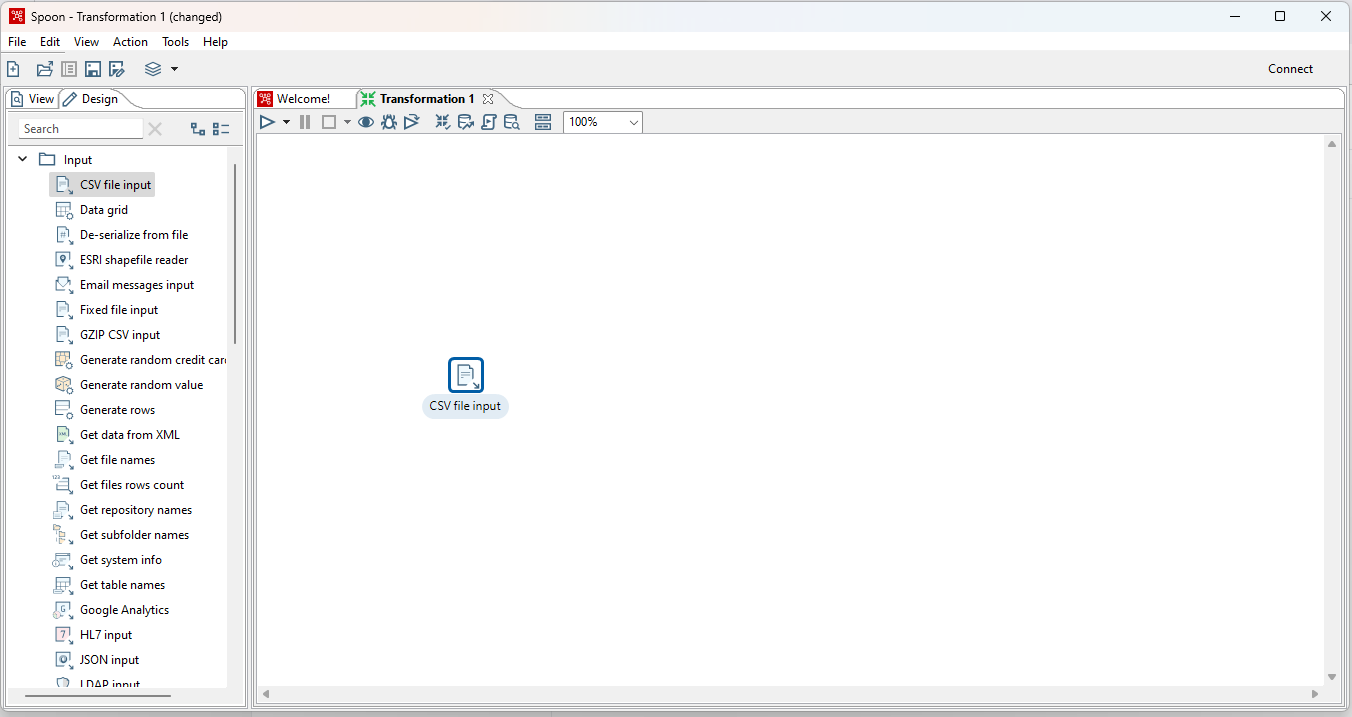
**A. Pengambilan Data (Extract)**

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



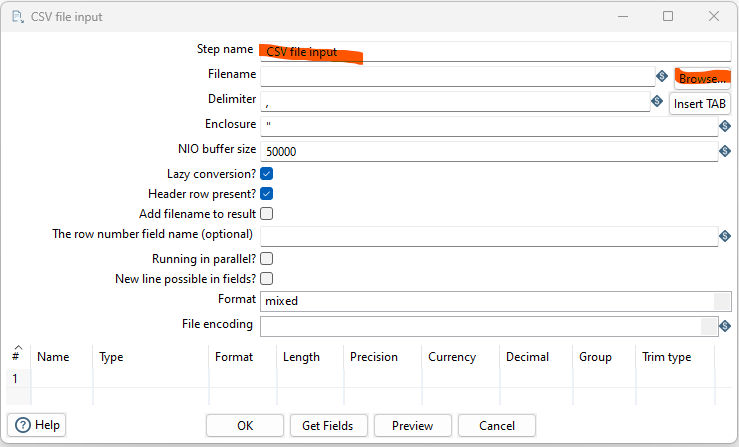
Gambar memulai membuka lembar kerja Transformation

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

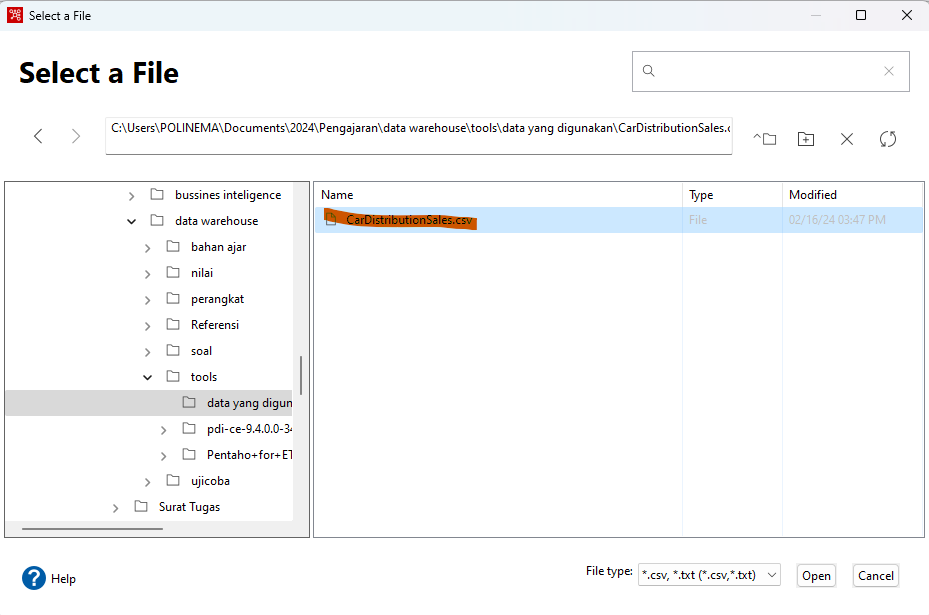


Gambar objek CSV file input pada Work Area

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**.

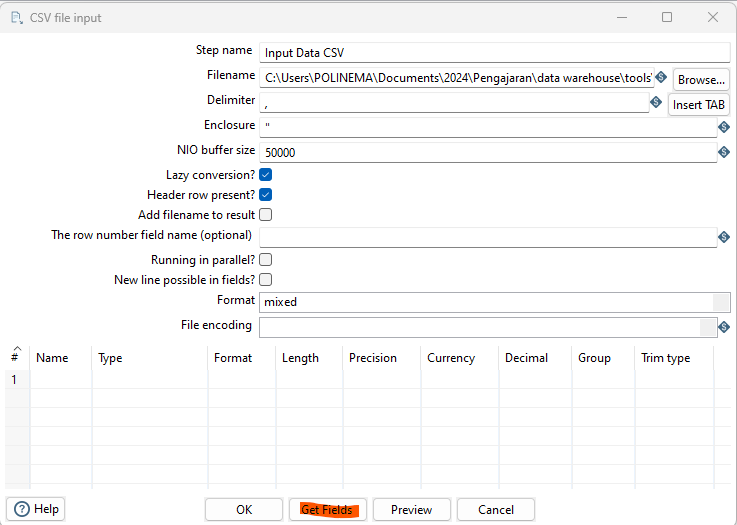


Jendela konfigurasi CSV file input

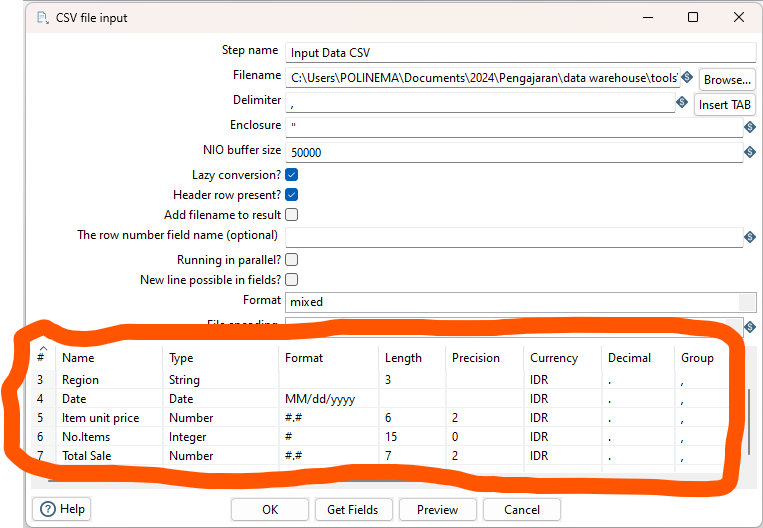


Gambar lokasi file CarDistributionSales.csv

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

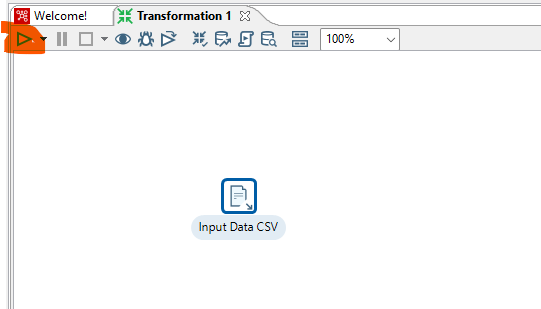


Gambar jendela konfigurasi setelah file dipilih

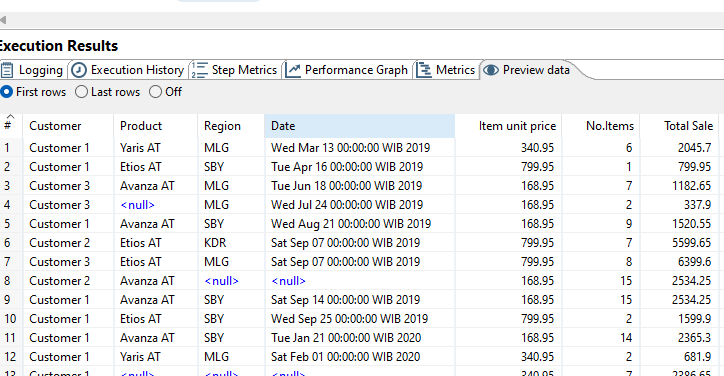


Gambar jendela konfigurasi setelah Get Fields ditekan

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.



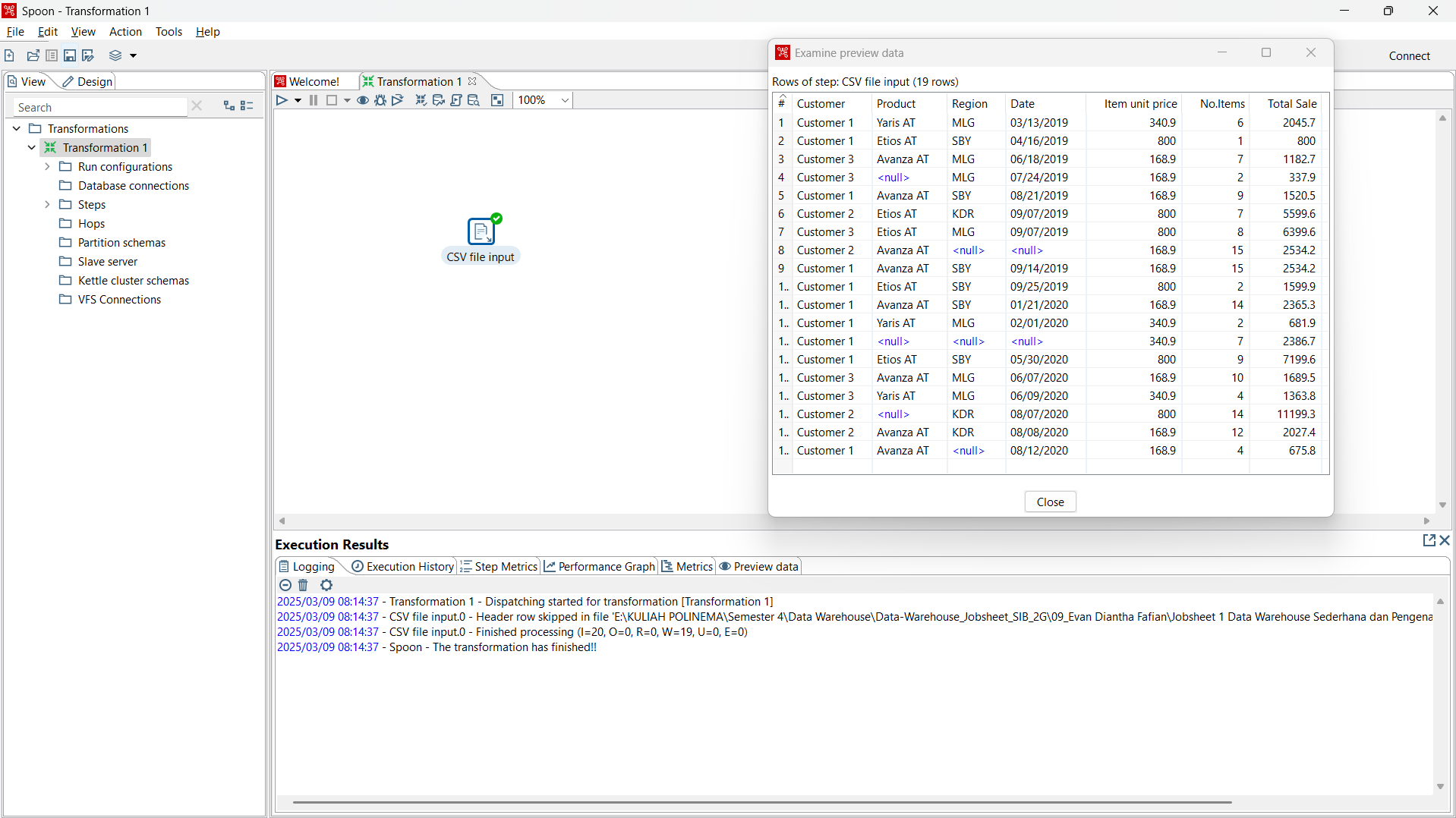
Gambar jendela transformation area



Gambar hasil dari proses

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

Hasil Extract Data



**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
   * Ya sesuai dengan data aslinya. Jika langkah-langkah dari praktikum ectract data dilakukan *step by step* nya dengan benar maka akan menampilkan data yang sama dengan data aslinya, data akan ditampilkan di executions result area, jika tampilan dari hasil datanya terdapat perbedaan mungkin bisa saja terjadi kesalahan saat melakukan konfigurasi pada saat pemilihan lokasi file .csv atau bisa delimiter yang tidak sesuai.
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?
   * Membuka koneksi ke sumber data
   * Mengeksekusi atau membaca data
   * Menjalankan transformasi
   * Transformasi
   * Mengirim data
   * Menampilkan data
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?



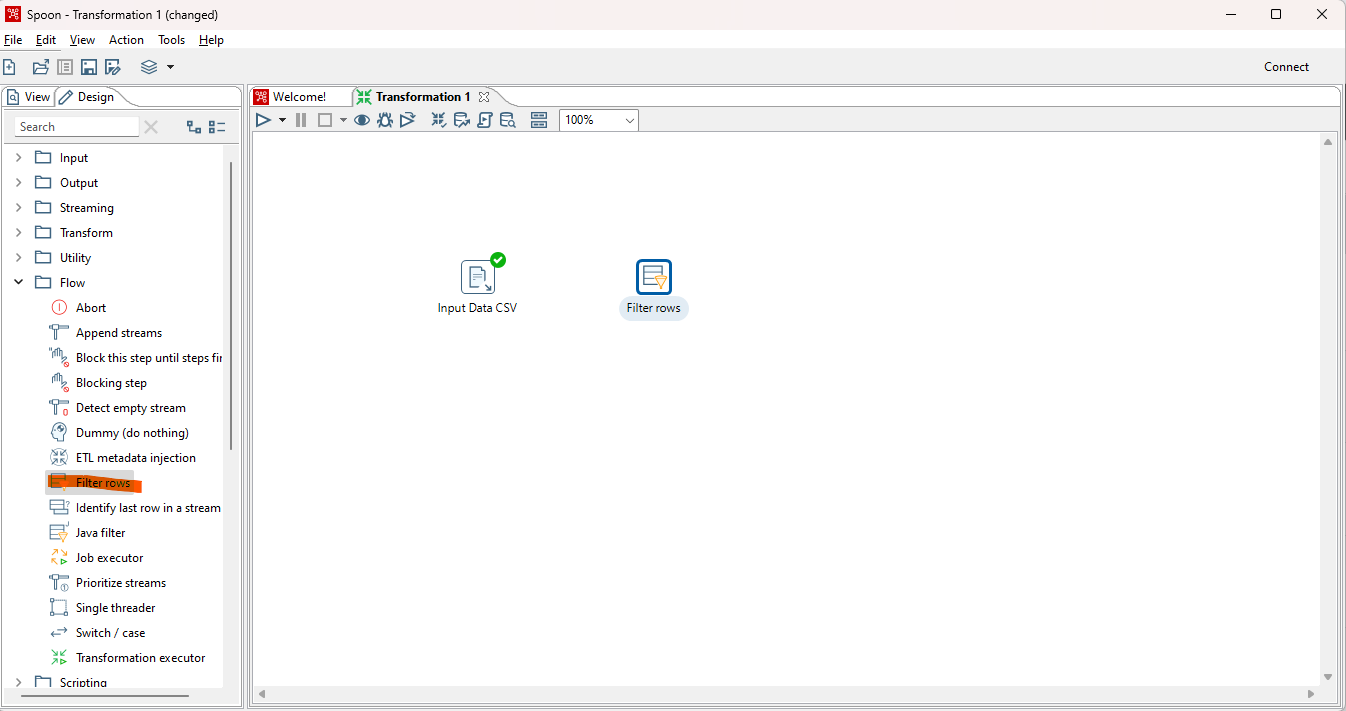
* + I (Input): Jumlah baris data yang diterima sebagai input oleh langkah tersebut.
  + O (Output): Jumlah baris data yang dikeluarkan oleh langkah tersebut ke langkah berikutnya.
  + R (Read): Jumlah baris data yang dibaca dari sumber data.
  + W (Written): Jumlah baris data yang berhasil ditulis atau diproses oleh langkah tersebut.
  + U (Updated): Jumlah baris data yang diperbarui.
  + E (Error): Jumlah baris data yang mengalami kesalahan selama pemrosesan.

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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu | Prosessor | RAM | Storage | VGA | CPU |
| 1. | Evan | 5s | Intel i7-10750H | 16 GB | SSD NVMe 512 GB | NVIDIA GTX 1650 Ti | 60% |
| 2. | Afif | 7s | AMD Ryzen 5 4600H | 8 GB | SSD SATA 256 GB | Integrated Vega 6 | 70% |
| 3. | Nopal | 6s | Intel i5-10300H | 12 GB | HDD 1 TB | NVIDIA GTX 1050 | 50% |
| 4. | Safrizal | 8s | Intel i3-10110U | 8 GB | SSD SATA 512 GB | Integrated UHD | 80% |

**B. Filter Data (Transform) dan Pengemasan data (Load)**

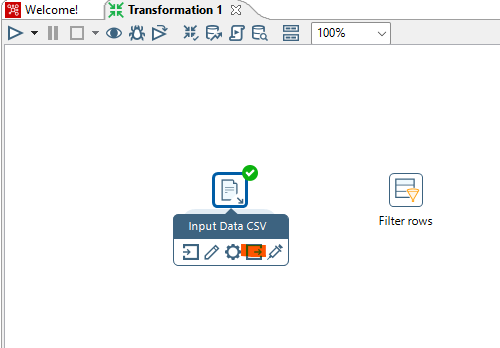
1. Carilah objek **Filter rows** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



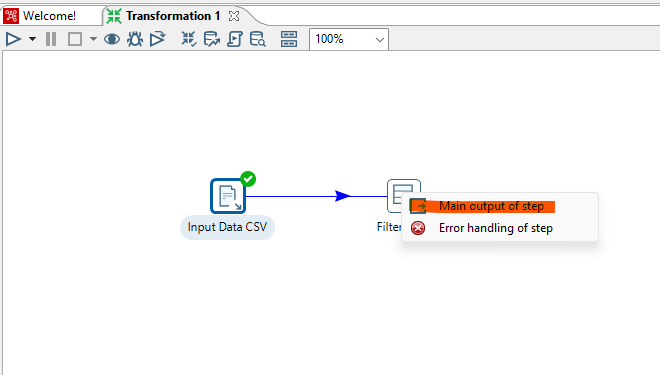
Gambar objek filter rows pada area kerja

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.

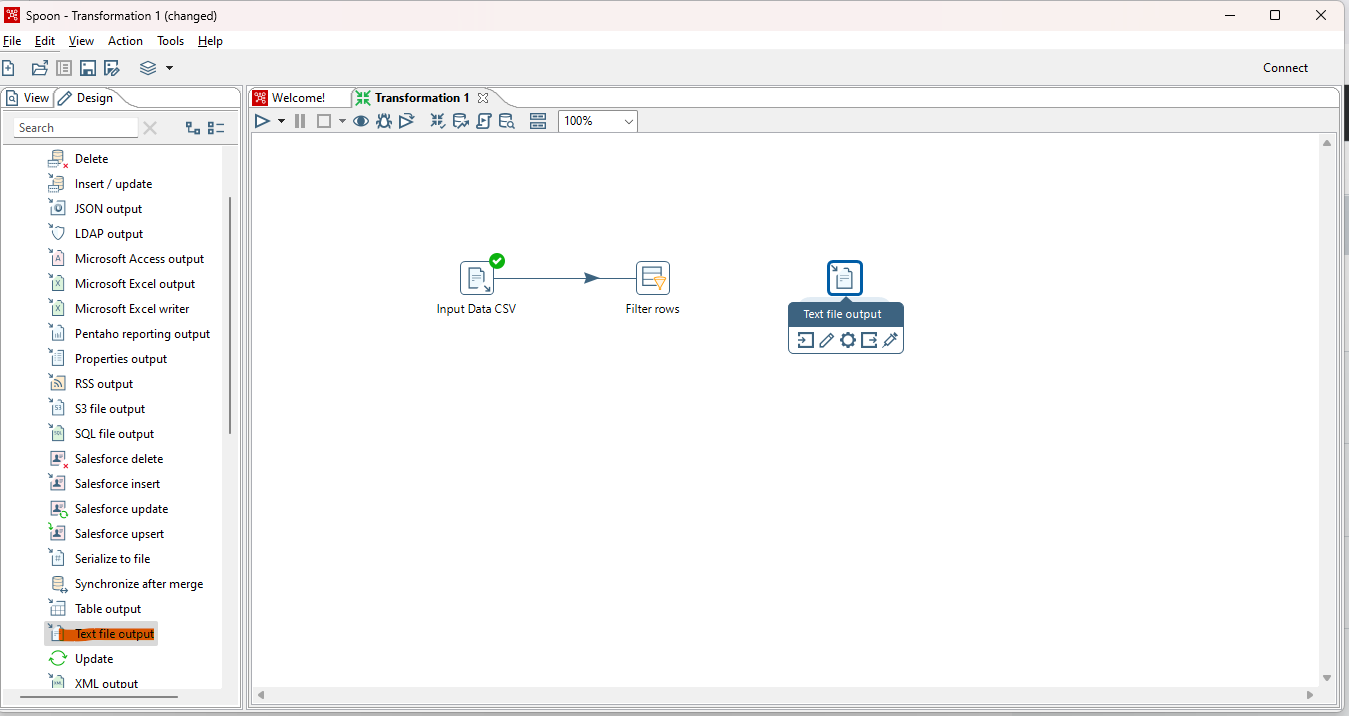


Gambar output konektor pada Input data CSV



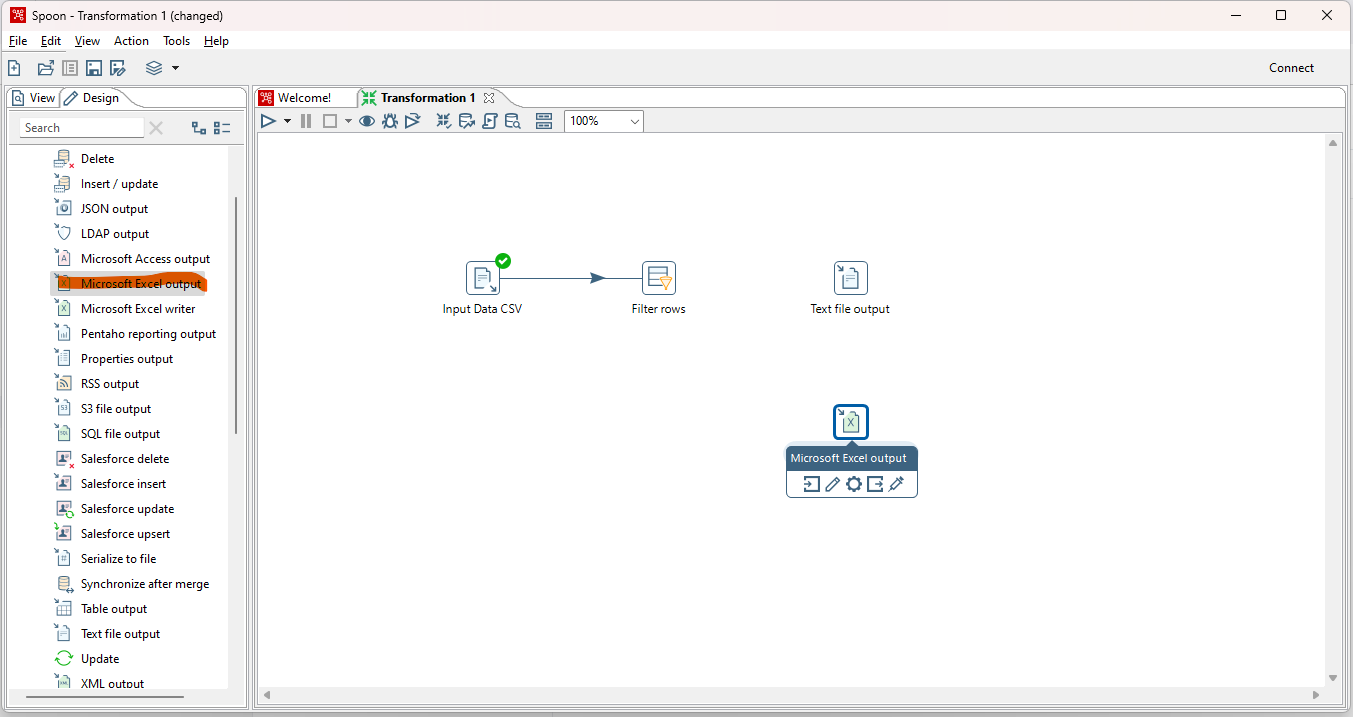
Gambar Input Data CSV terhubung dengan Filter rows

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



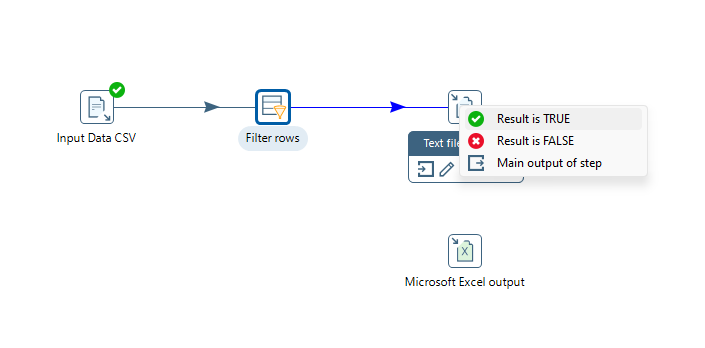
Gambar text file output pada work area

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



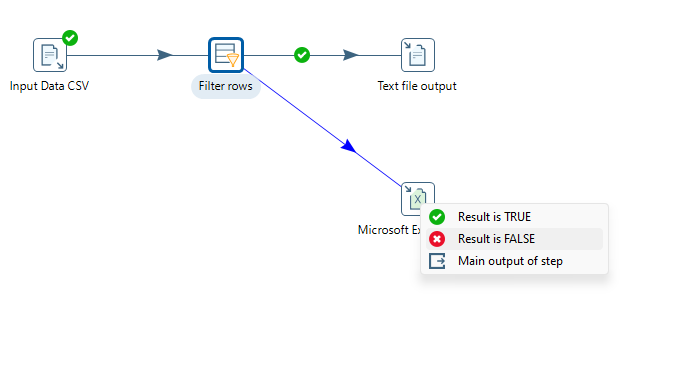
Gambar Microsofft excel output pada work area

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



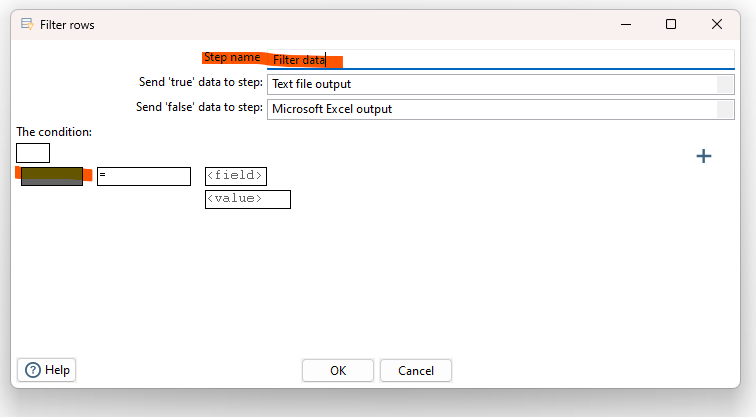
Gambar konktor text file output

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



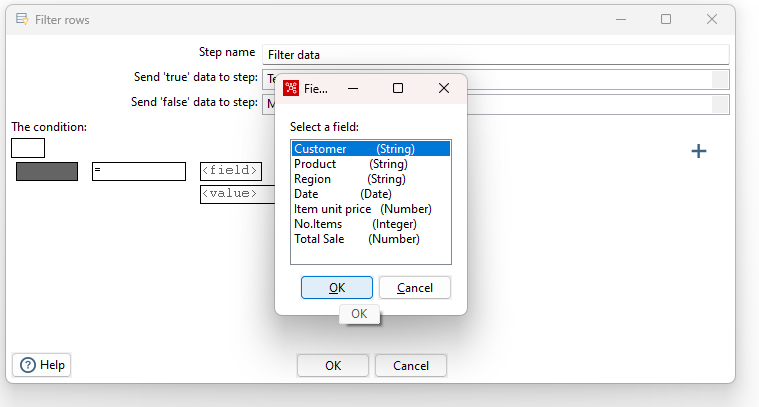
Gambar konektor microsoft excel output

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**



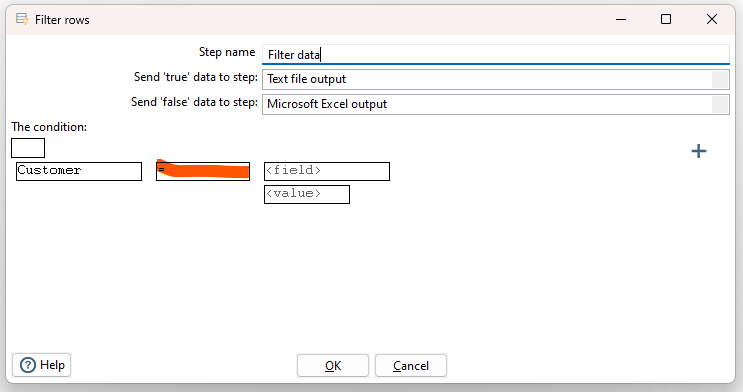
Gambar jendela konfigurasi filter rows

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

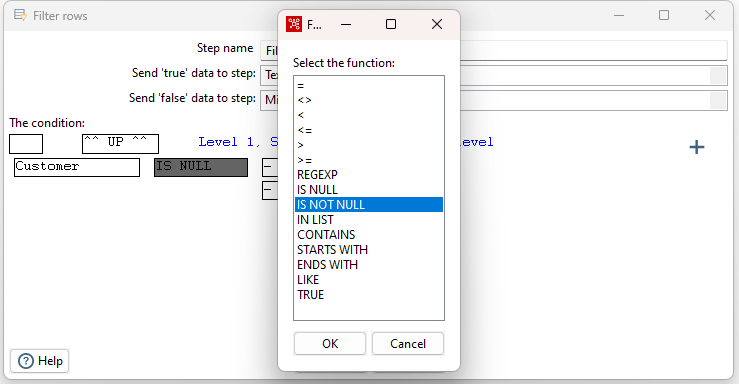


Gambara jendela pilihan field

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

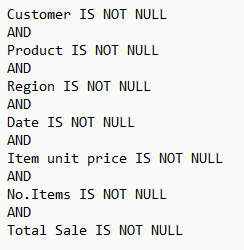


Gambar jendela filter rows bagian value

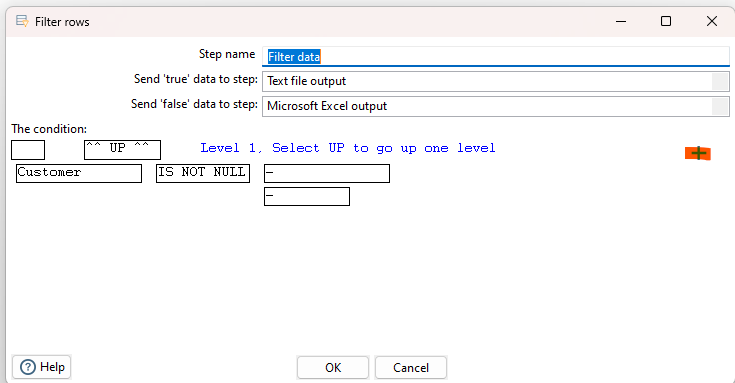


Gambar jendela filter rows pemilihan value

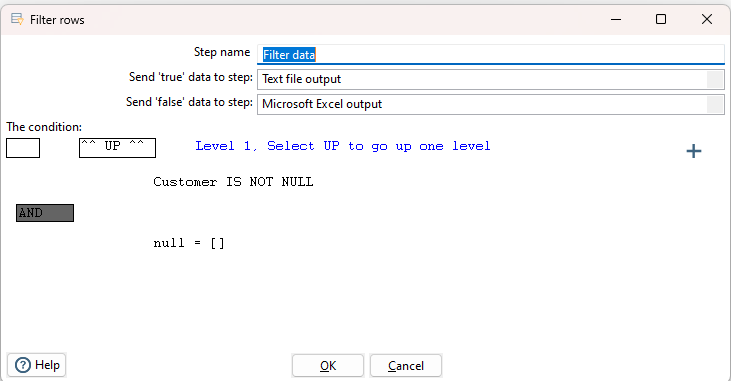
1. Tekan tombol **+** untuk menambah kondisi.
2. Pilih **statement** yang masih null untuk menambah kondisi.
3. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
4. Gunakan operator **AND** untuk menghubung



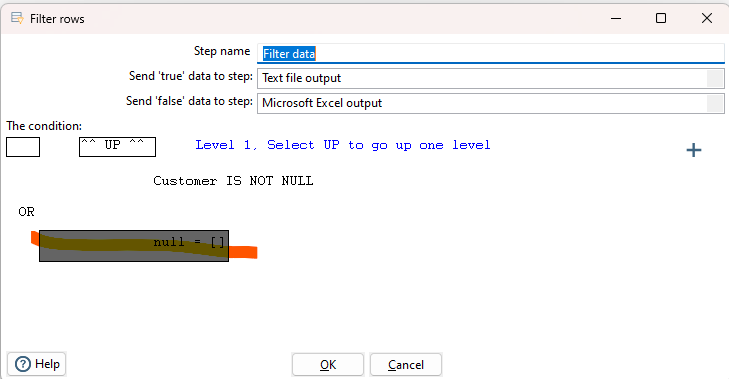
Gambar hasil statement



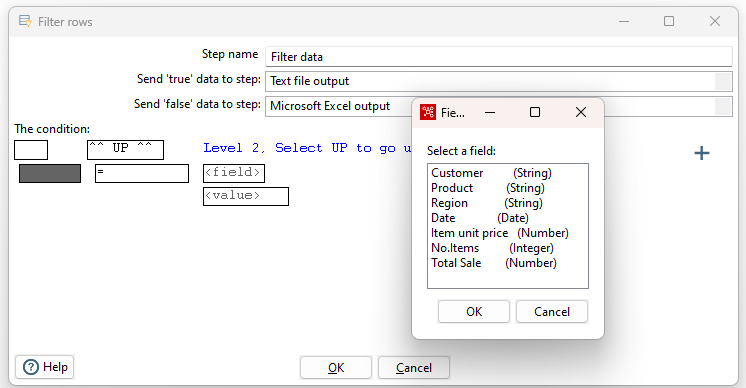
Gambar untuk menambah statement



Gambar memilih operator menjadi AND

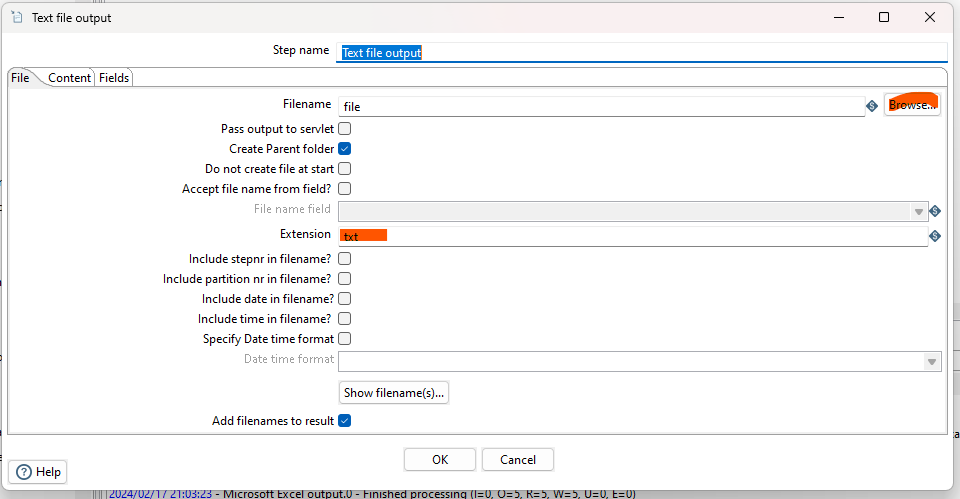


Gambar mengubah statement yang null

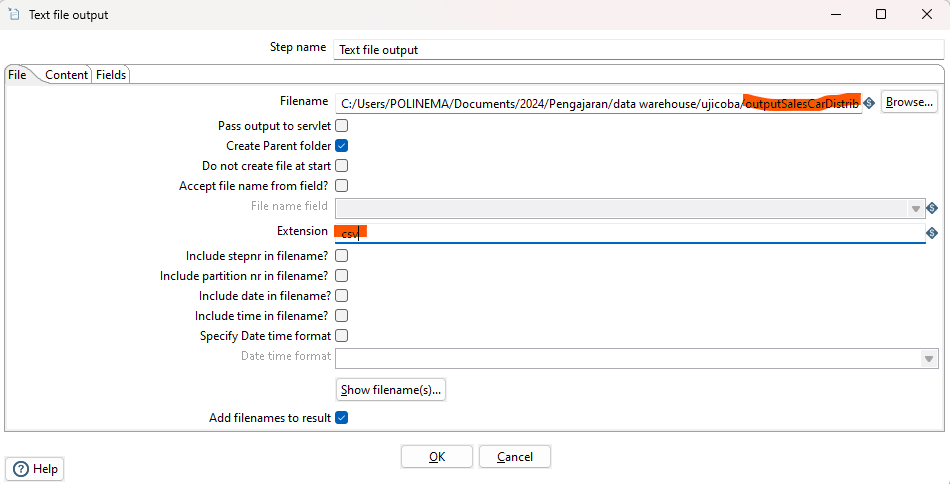


Gambar pemilihan field untuk statement

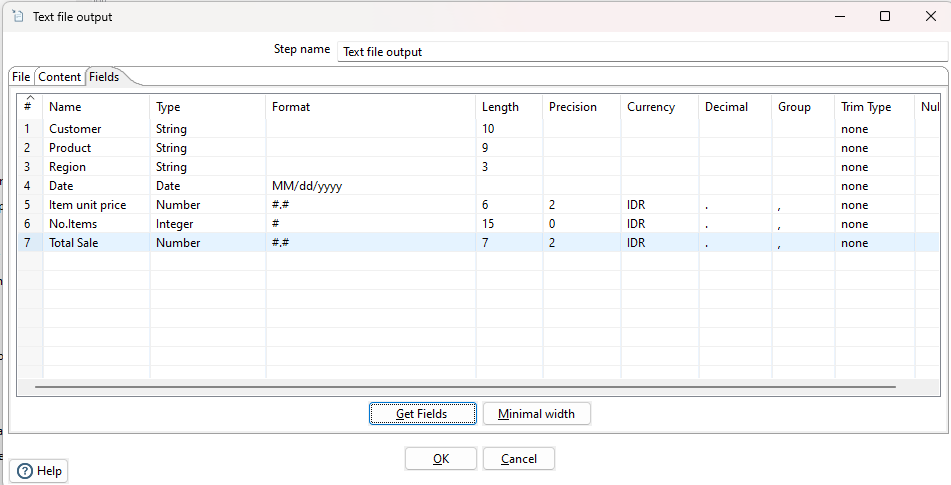
1. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.
2. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**.
3. Ubah extension menjadi **csv.**
4. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.
5. Tekan **ok** untuk kembali ke **work area**



Jendela konfigurasi text file output

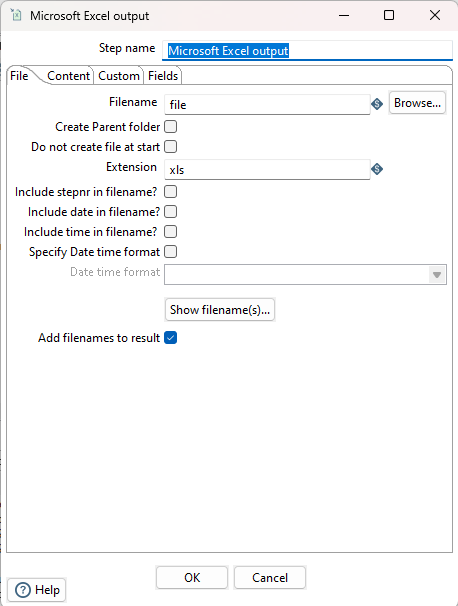


Jendela konigurasi text file input ekstensi csv

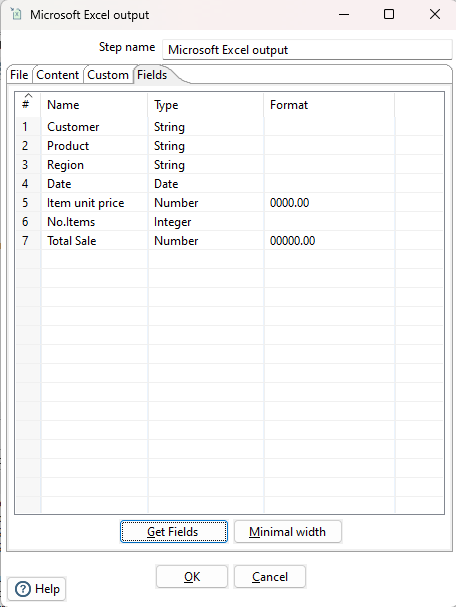


Tab get fields

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**.

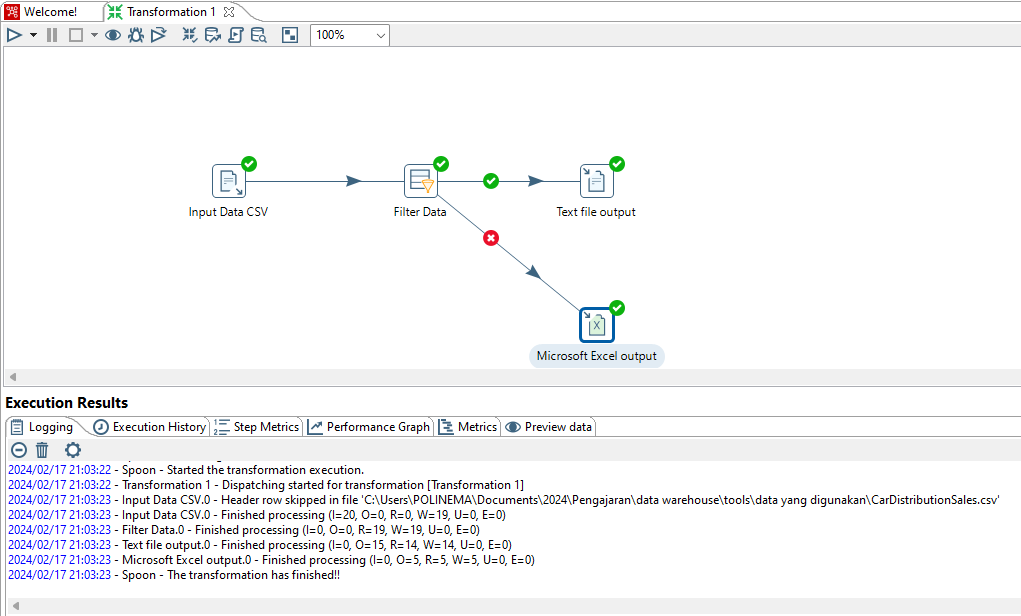


Gambar jendela konfigurasi Microsoft Excel output

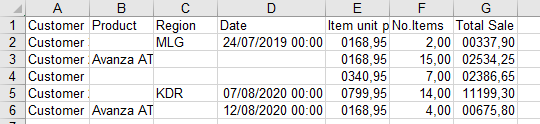


Gambar tab field microsoft excel output

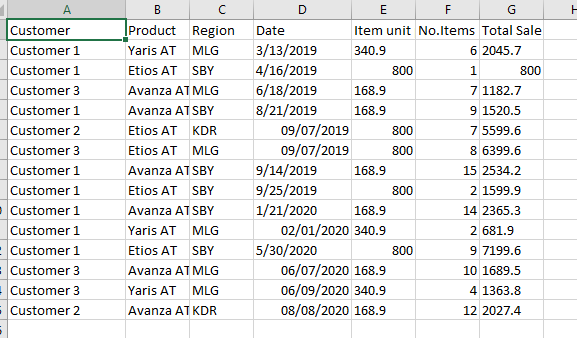
1. Tekan tombul **Run** pada pojok kiri atas **Work Area**. jika berhasil maka akan muncul tanda centang hijau di setiap proses.



Gambar proses berhasil dijalankan



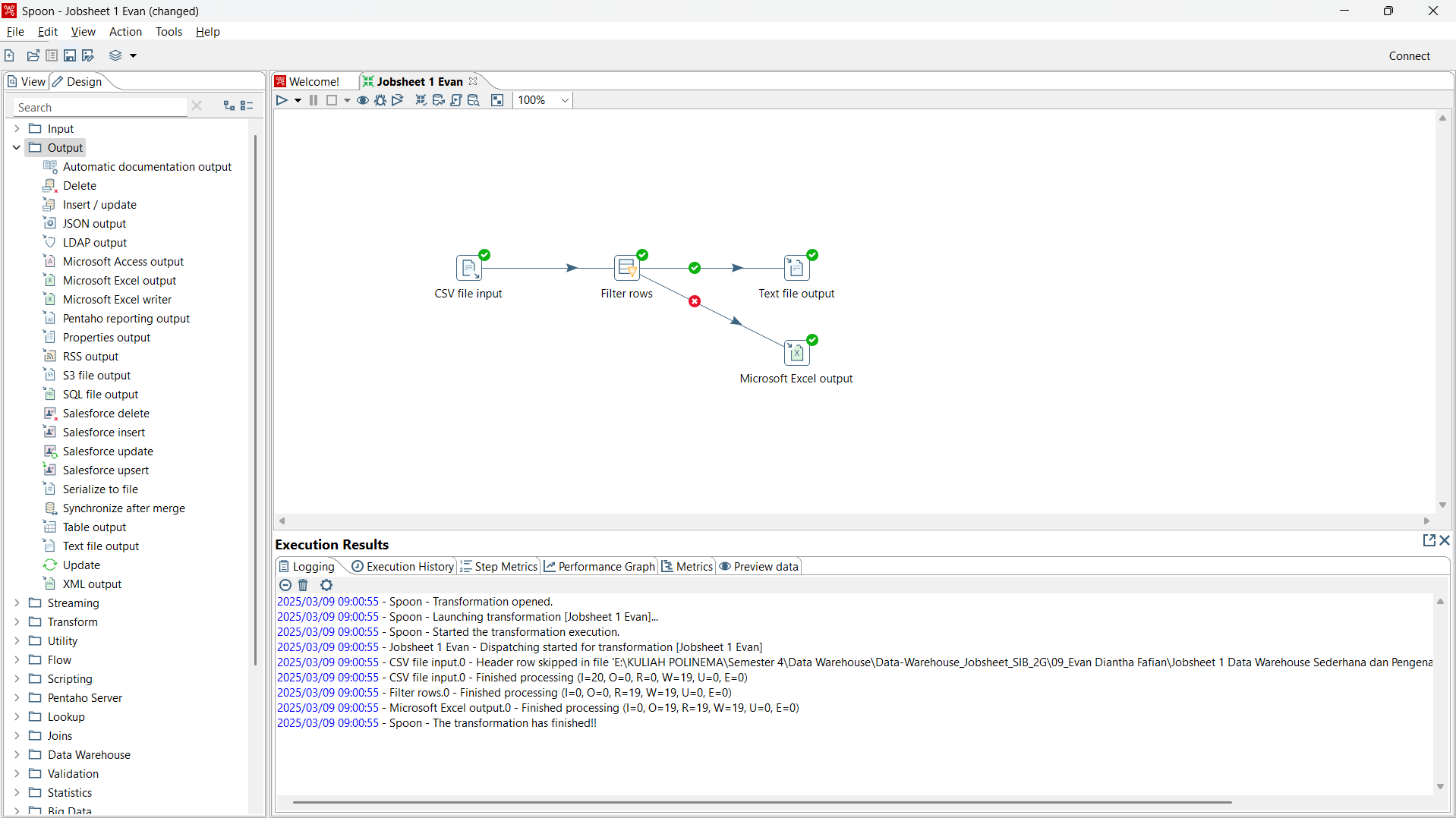
Gambar failedSalesCarDistribution



Gambar outputSalesCarDistribution

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.

Hasil Proses Filter Data dan Pengemasan Data



**TUGAS 3**

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
   * Untuk file csv atau text file itu data yang telah di filter sesuai filter datanya yaitu yang di dalam rownya terdapat nilai null akan dipisakan ke file excel
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
   * Tugas dari filter row adalah memfilter sebuah row dalam kondisi praktek kali ini memfilter row yang memiliki nilai null kita juga bisa filter dengan kondisi kondisi yang lain seperti null, cascade, dsb.
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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu | Prosessor | RAM | Storage | VGA | CPU |
| 1. | Evan | 10s | Intel i7-10750H | 16 GB | SSD NVMe 512 GB | NVIDIA GTX 1650 Ti | 60% |
| 2. | Afif | 15s | AMD Ryzen 5 4600H | 8 GB | SSD SATA 256 GB | Integrated Vega 6 | 70% |
| 3. | Nopal | 13 | Intel i5-10300H | 12 GB | HDD 1 TB | NVIDIA GTX 1050 | 50% |
| 4. | Safrizal | 12s | Intel i5-10110U | 8 GB | SSD SATA 512 GB | Integrated UHD | 65% |