**PRACTICUM DATA WAREHOUSE**

**‘Jobsheet 1 – Input Data’**

A logo of a building

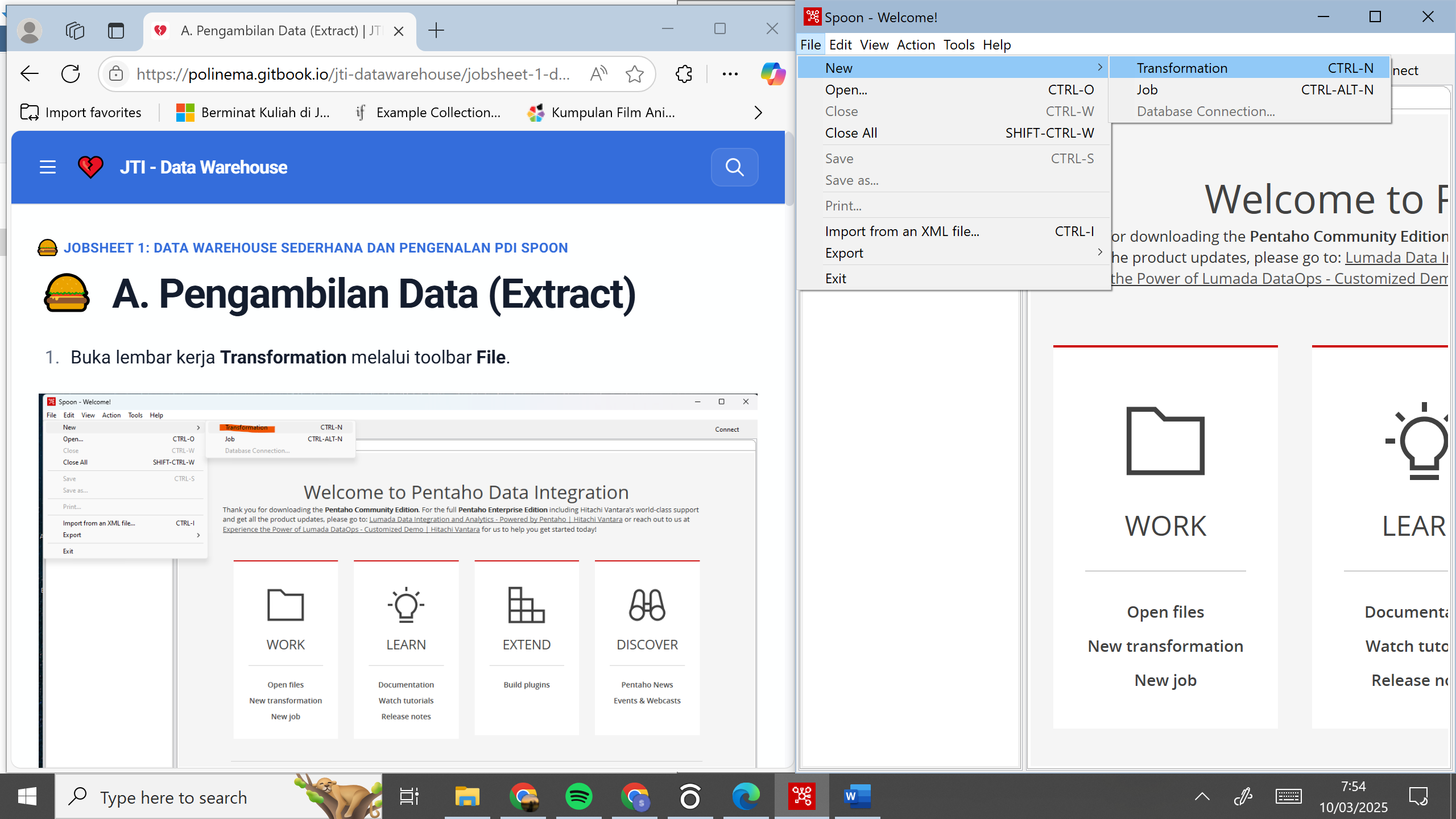
Description automatically generated

Name : Sabrina Rahmadini

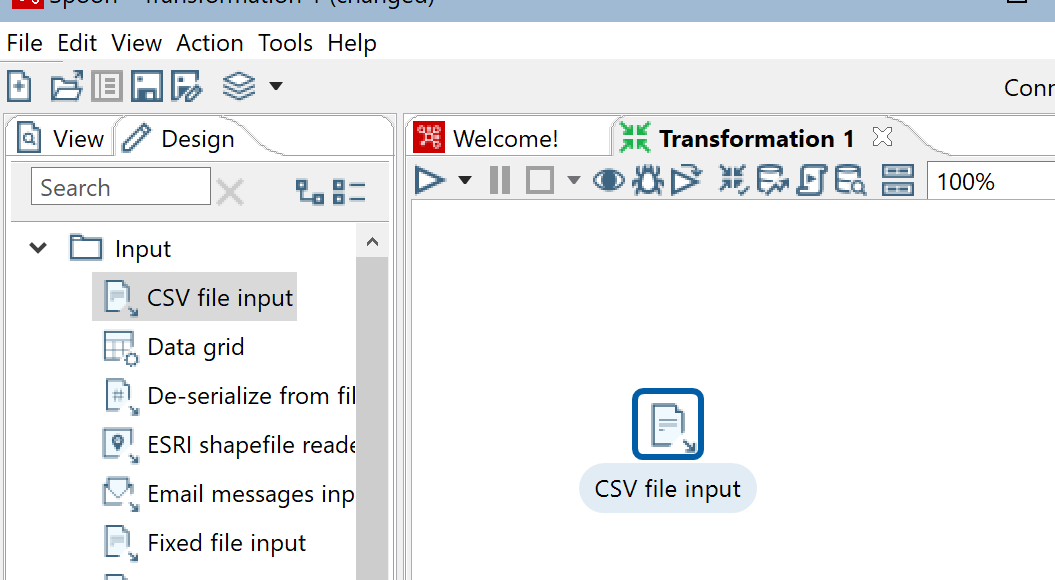
Student ID : 2341760155

Class : SIB 2G

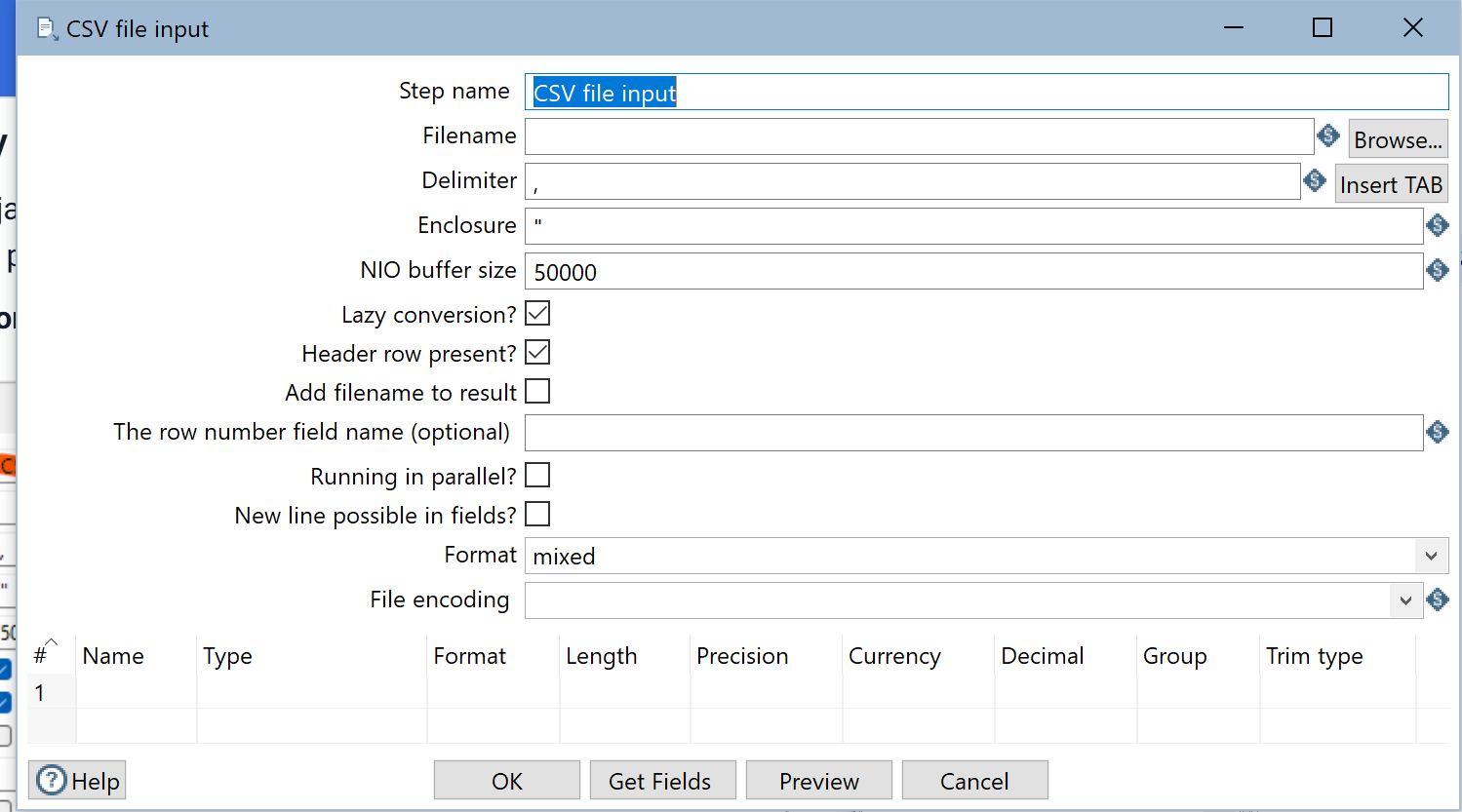
* Open the Transformation worksheet via the File toolbar.



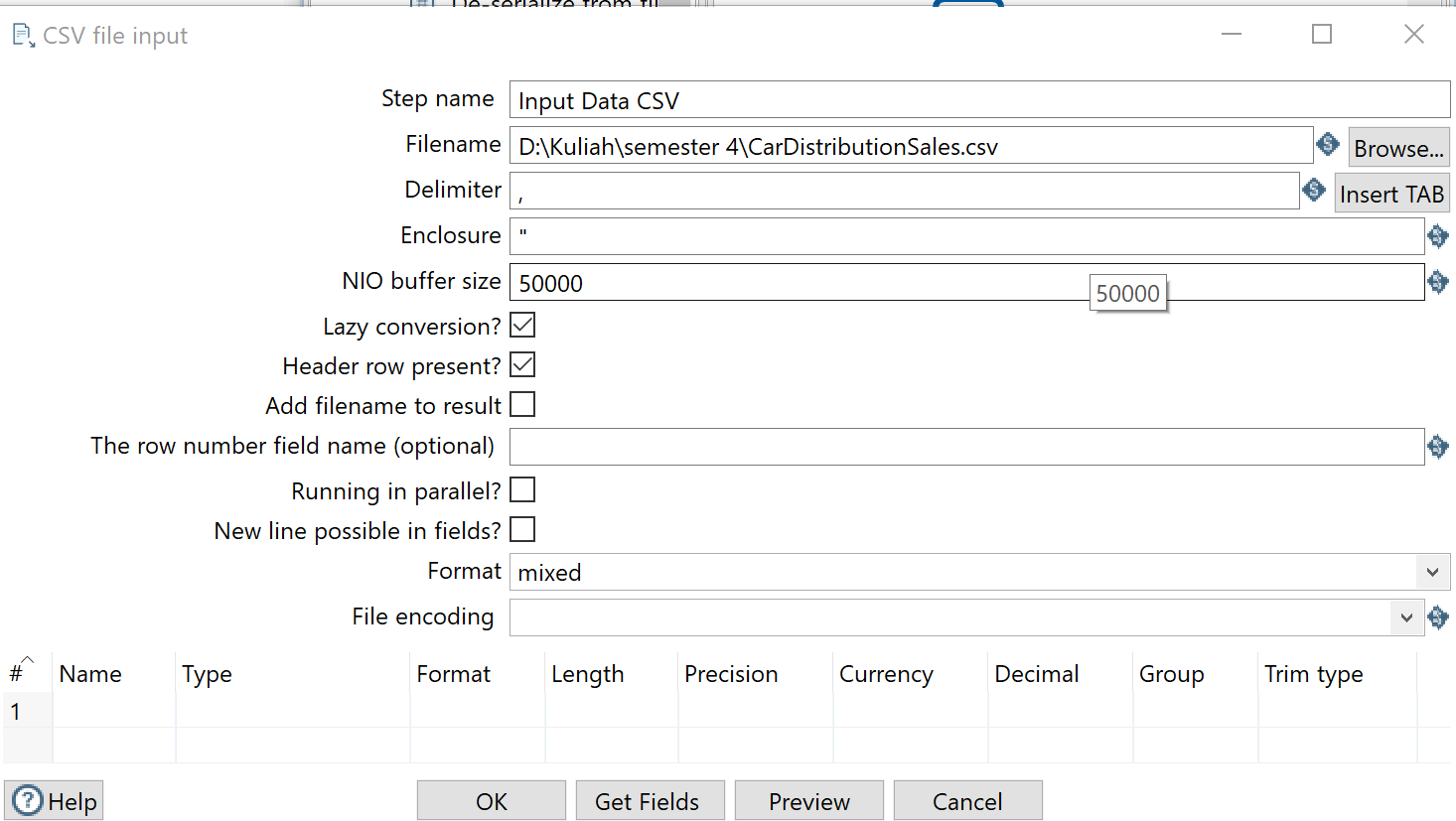
* Locate the CSV input file object in the Design Area. Drag and drop the object to the Work Area.



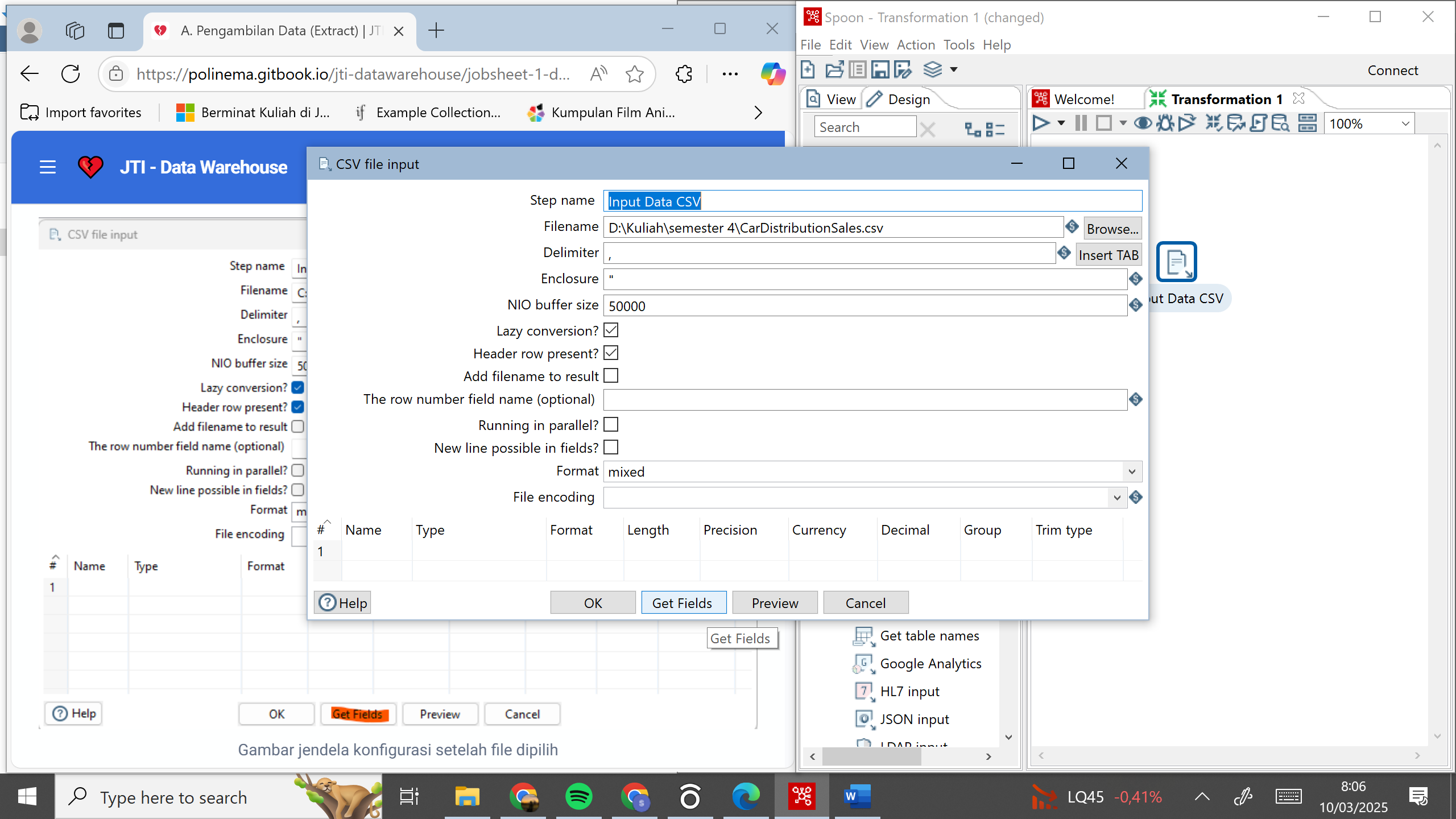
* Double-click on the CSV input file object until the configuration window appears.



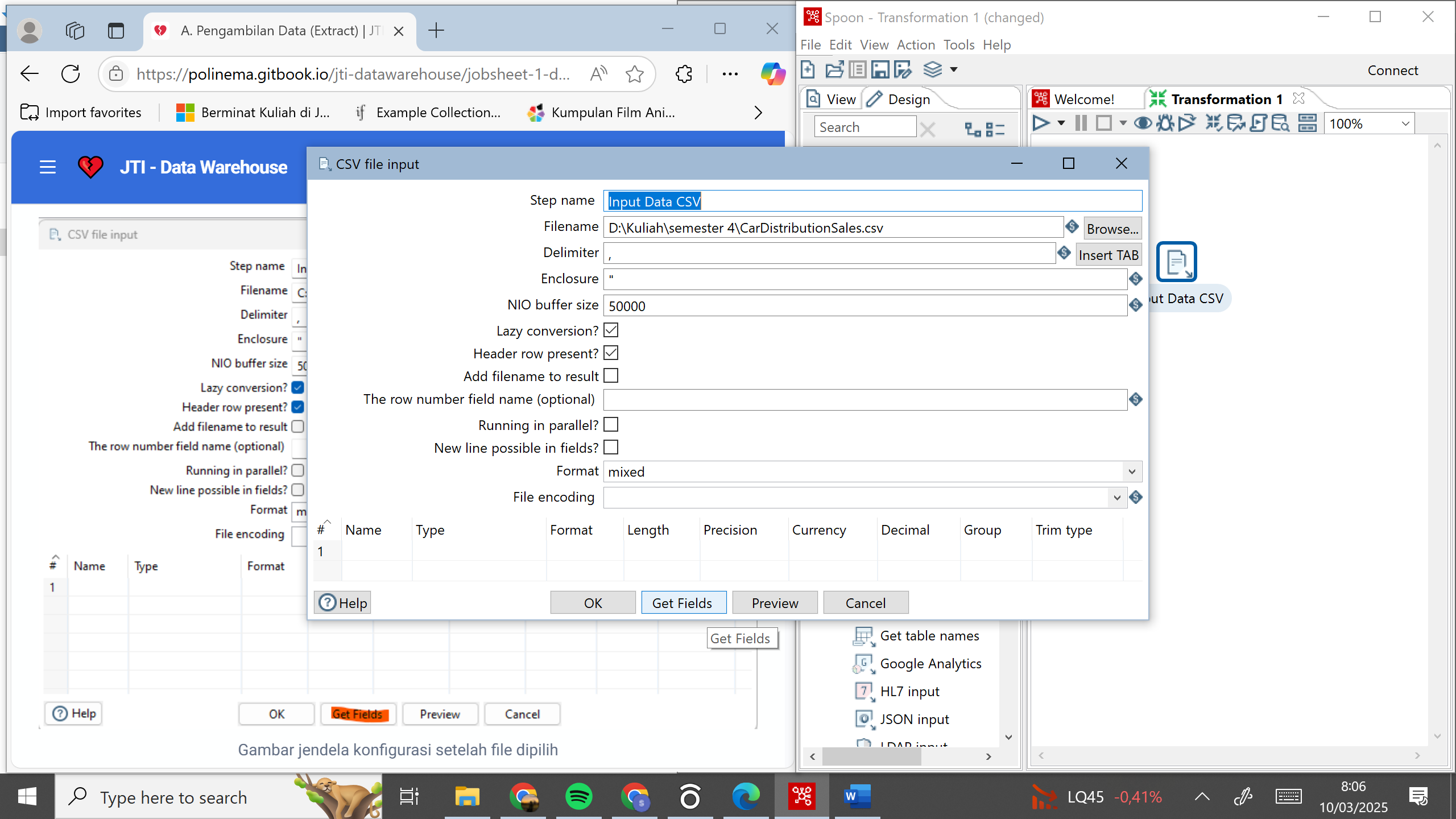
* Change the Step name to CSV Input Data, (naming is used for identification only, further naming can be adjusted).
* Select the location of the CarDistributionSales.csv file through Browse on filename.

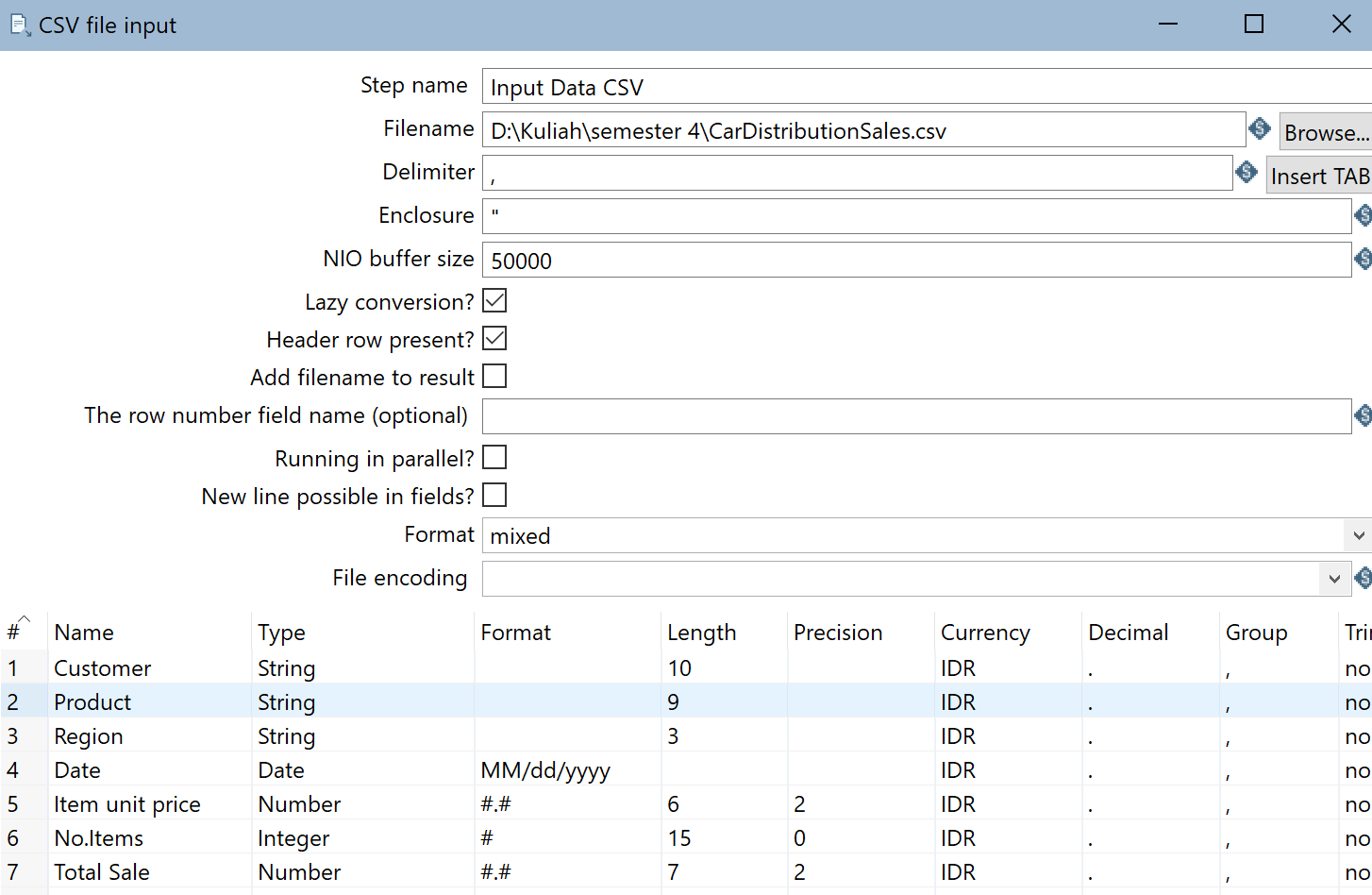


* If the data has been selected, make sure that the Delimiter section matches the data separator conditions in the selected file. (In this case it is comma “ , ” ).

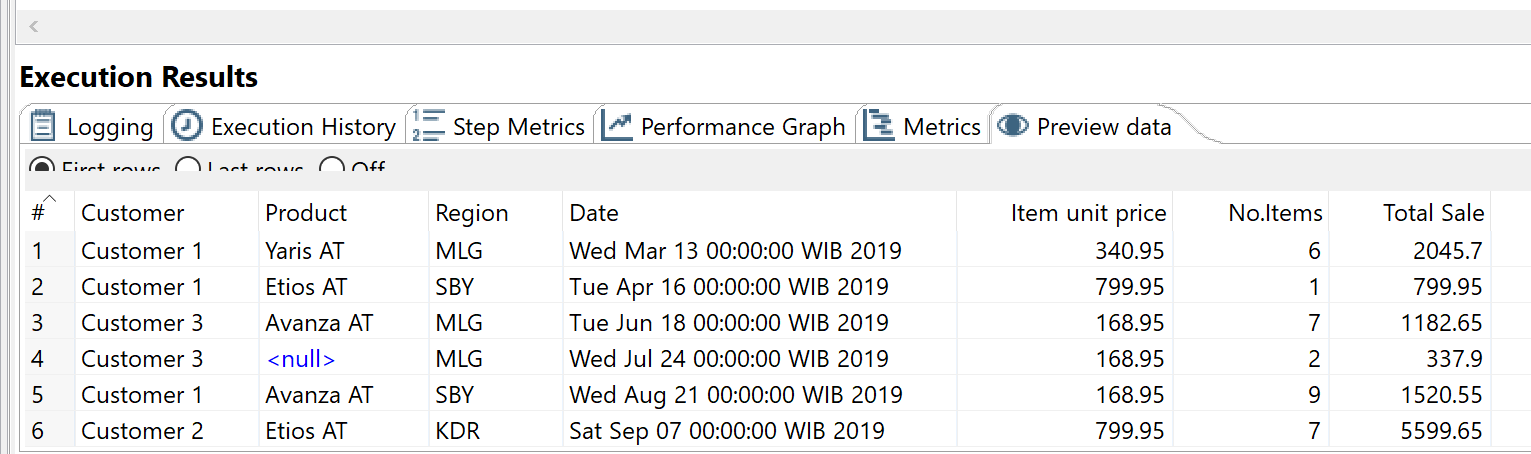


* Press Get Fields to retrieve the title of each column, make sure the column names are appropriate.





* Close the configuration window.
* Press the “Run” button on the top right corner of the Work Area, and the Execution Result Area will appear. Select the Prieview Data tab in the Execution Result Area. If the process is done correctly, the data will appear according to the original.



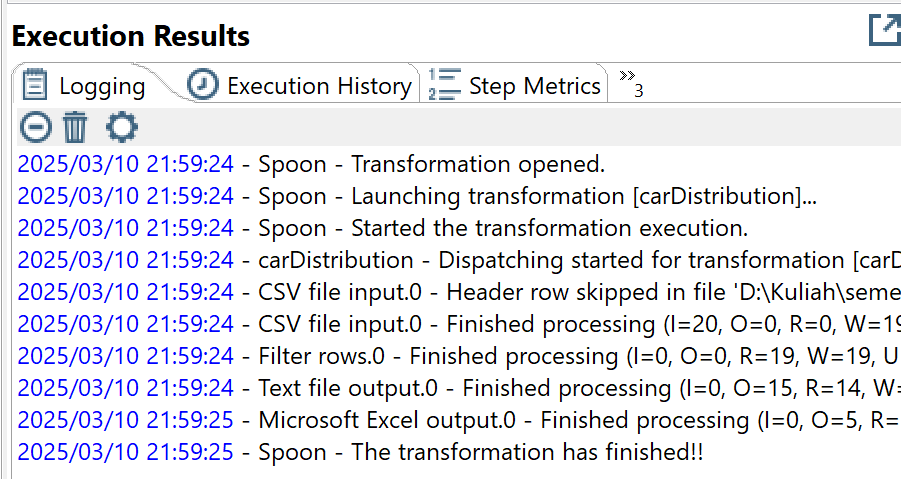
Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
   * Ya, sudah 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 ?



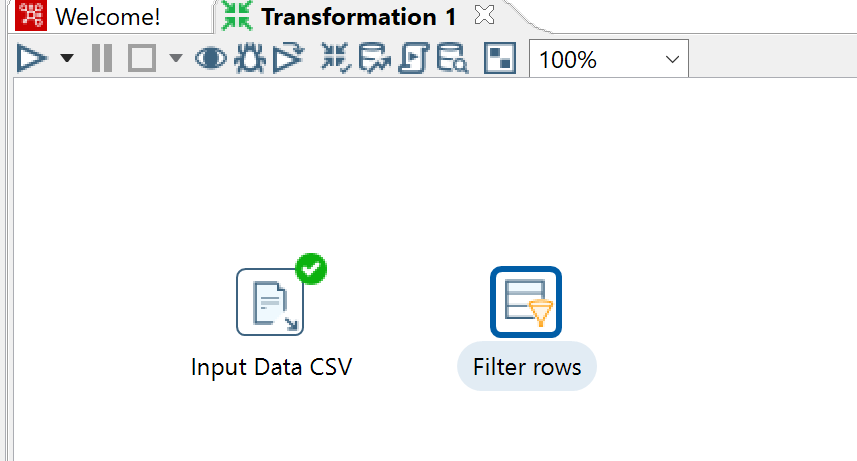
* + I (Input): Jumlah baris data yang diterima sebagai input oleh langkah tersebut.
  + O (Output): Jumlah baris data yang dikeluarkan oleh langkah tersebut ke langkah berikutnya.
  + R (Read): Jumlah baris data yang dibaca dari sumber data.
  + W (Written): Jumlah baris data yang berhasil ditulis atau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

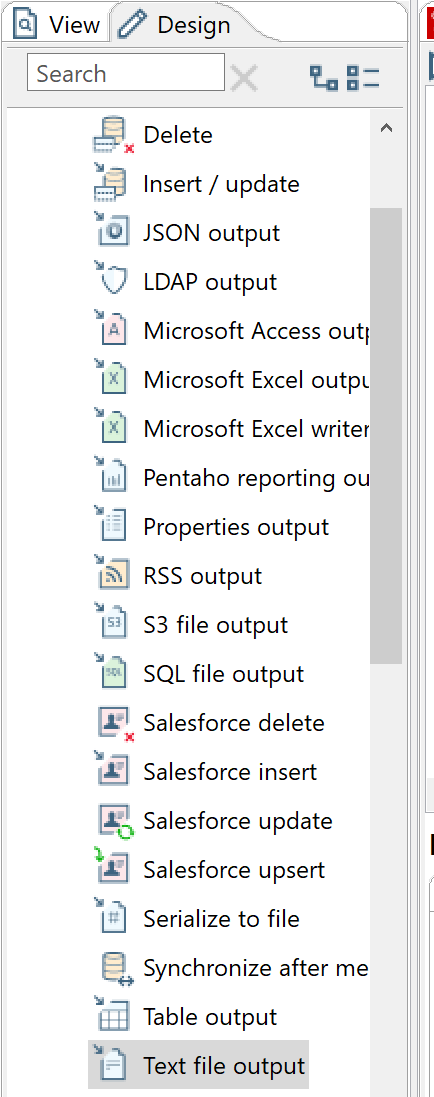
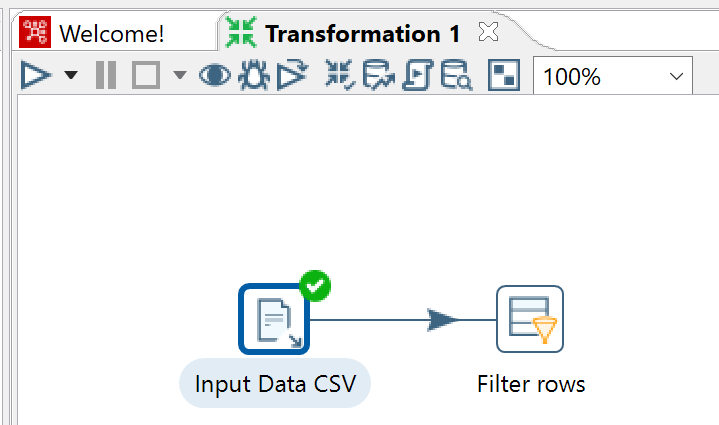
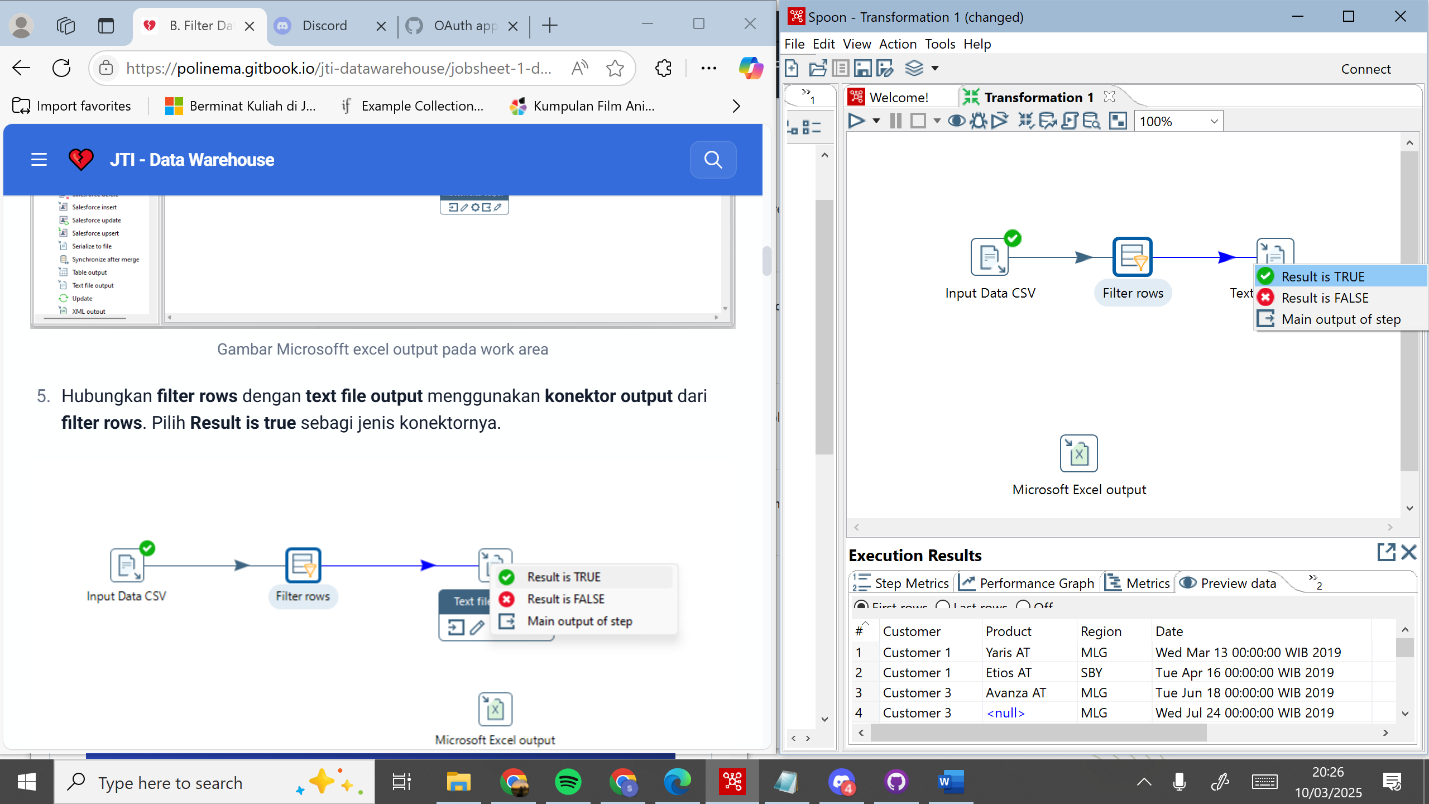
1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini?
   * 1 detik
2. Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). Bandingkan dalam bentuk tabel.

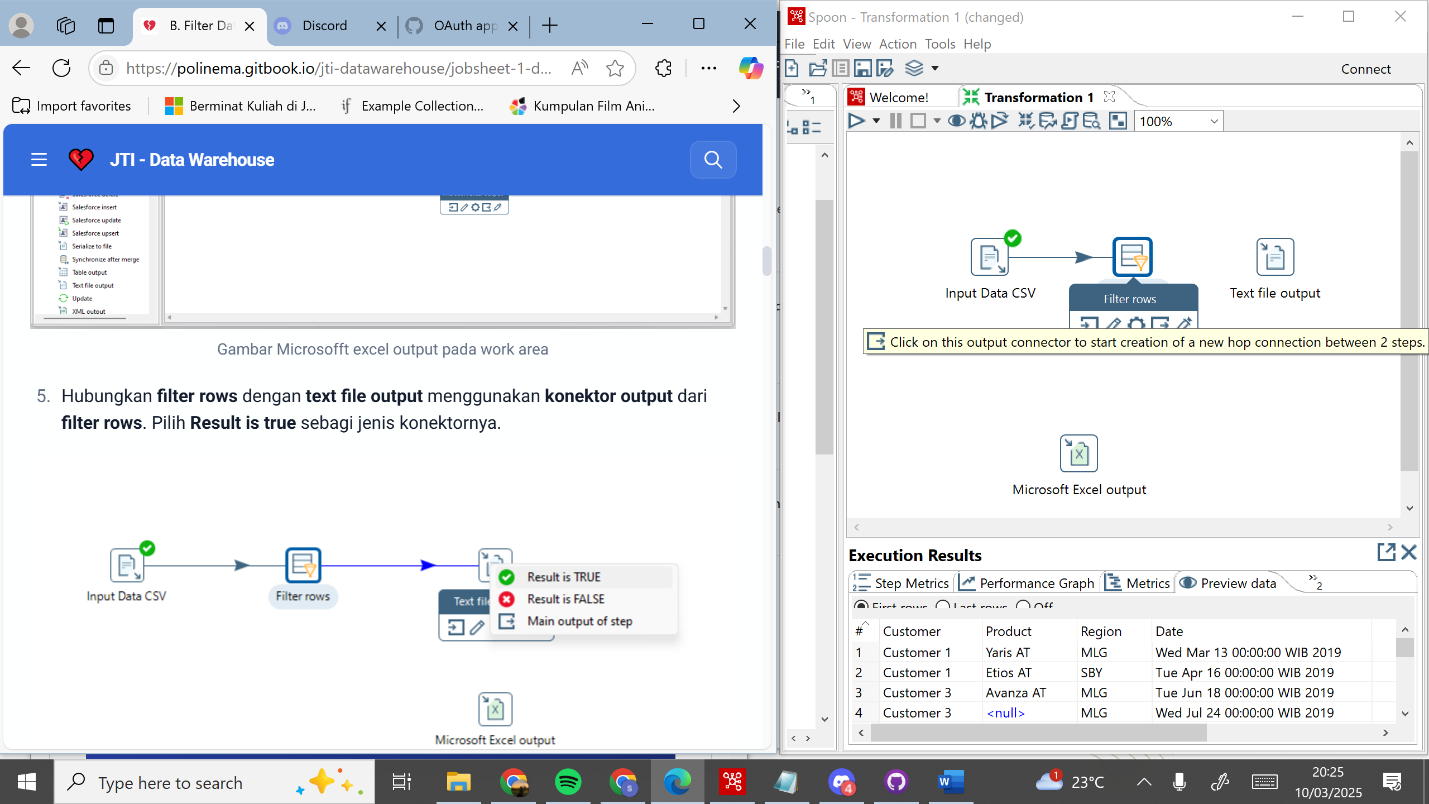
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu | Prosessor | RAM | Storage | VGA | CPU |
| 1 | Cindy | 6s | Intel i5- 10300H | 12 GB | HDD 1 TB | NVIDIA GTX 1050 | 60% |
| 2 | Sabrina | 1s | intel i5-7300U | 8192 MB(8GB) | Samsung 239 GB | Intel ® graphics 620 | 50% |
| 3 | nada | 1s | 11th gen intel®core™i5- | 16 GB | SSD NVMe 477GB | Intel®Iris ® Xe graphics | 23% |
| 4 | zahra | 1s | Intel®core ™i7-13650HX(13th Gen, 20 Cores) | 16 GB | SSD 477GB | Intel ® UHD graphics  +  NVIDIA GeForce RTX 4050 | 7% |

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

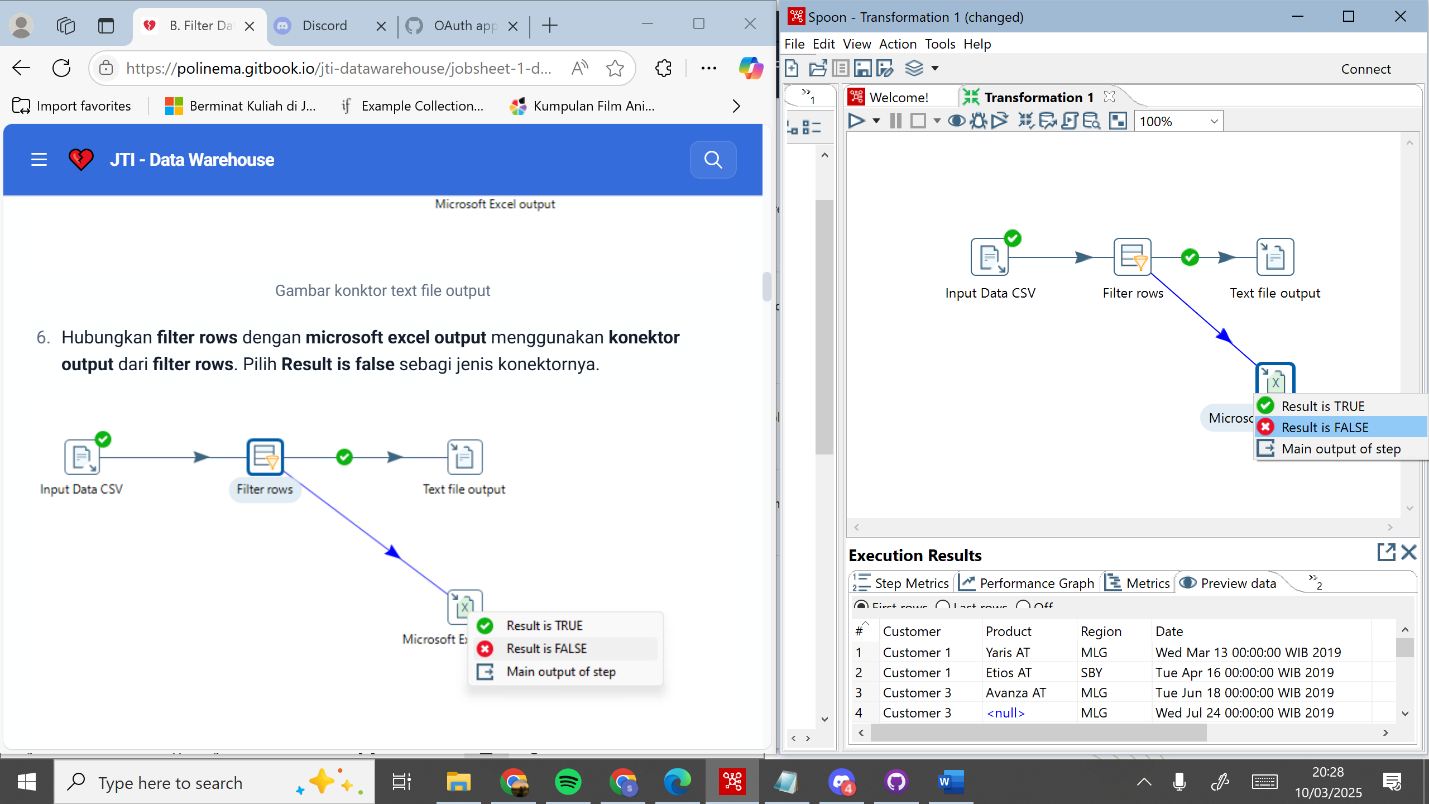
* Carilah objek **Filter rows** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

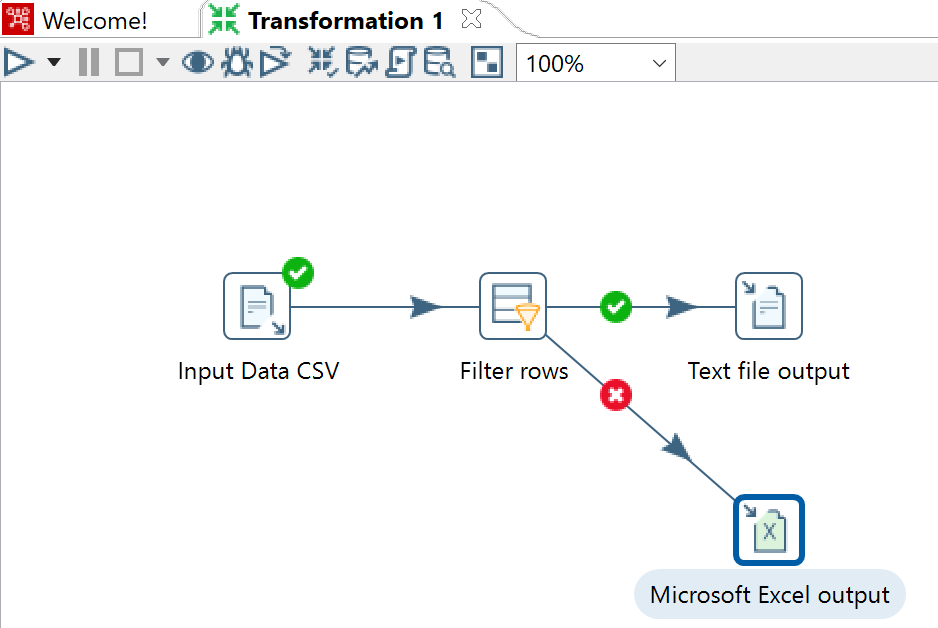


* Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.
* Carilah objek **Text file output dan Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.
* Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagi jenis konektornya.

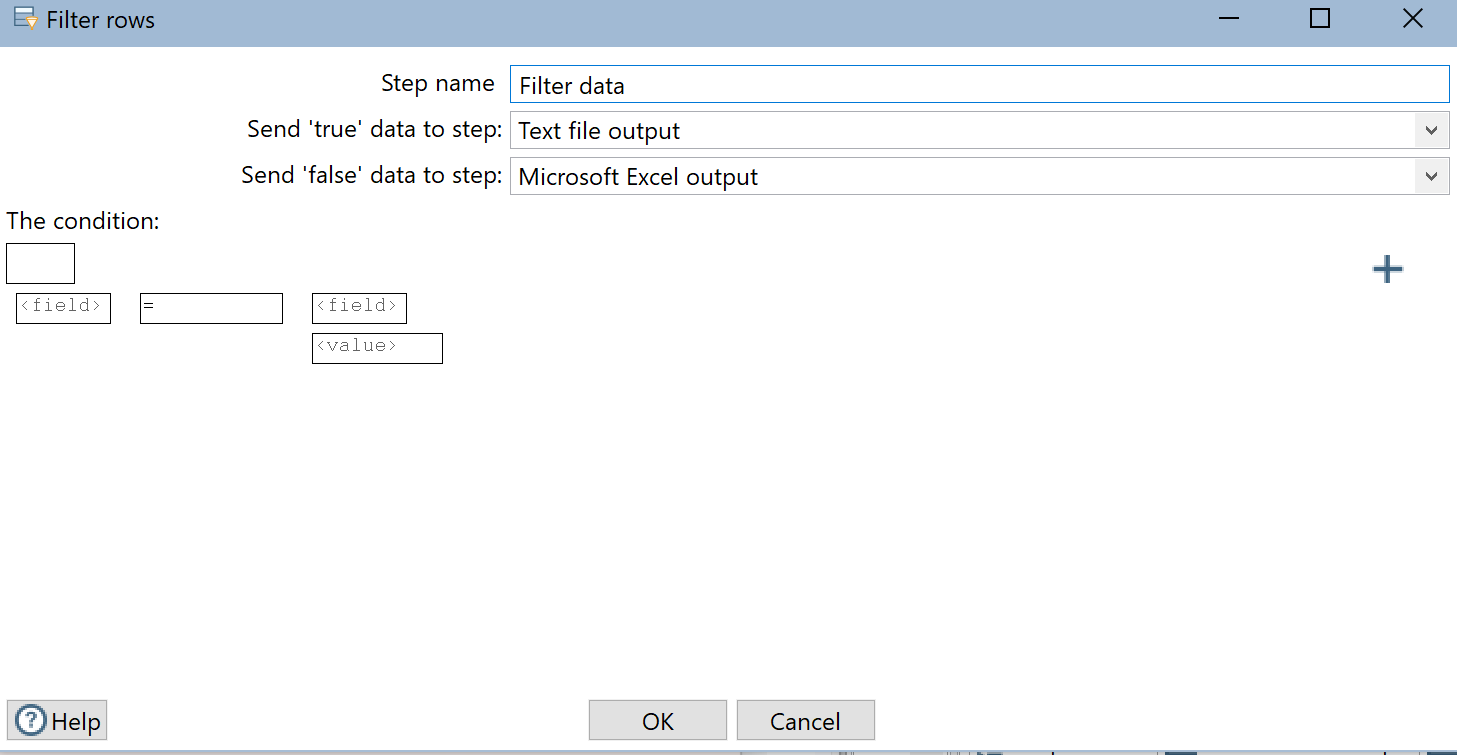
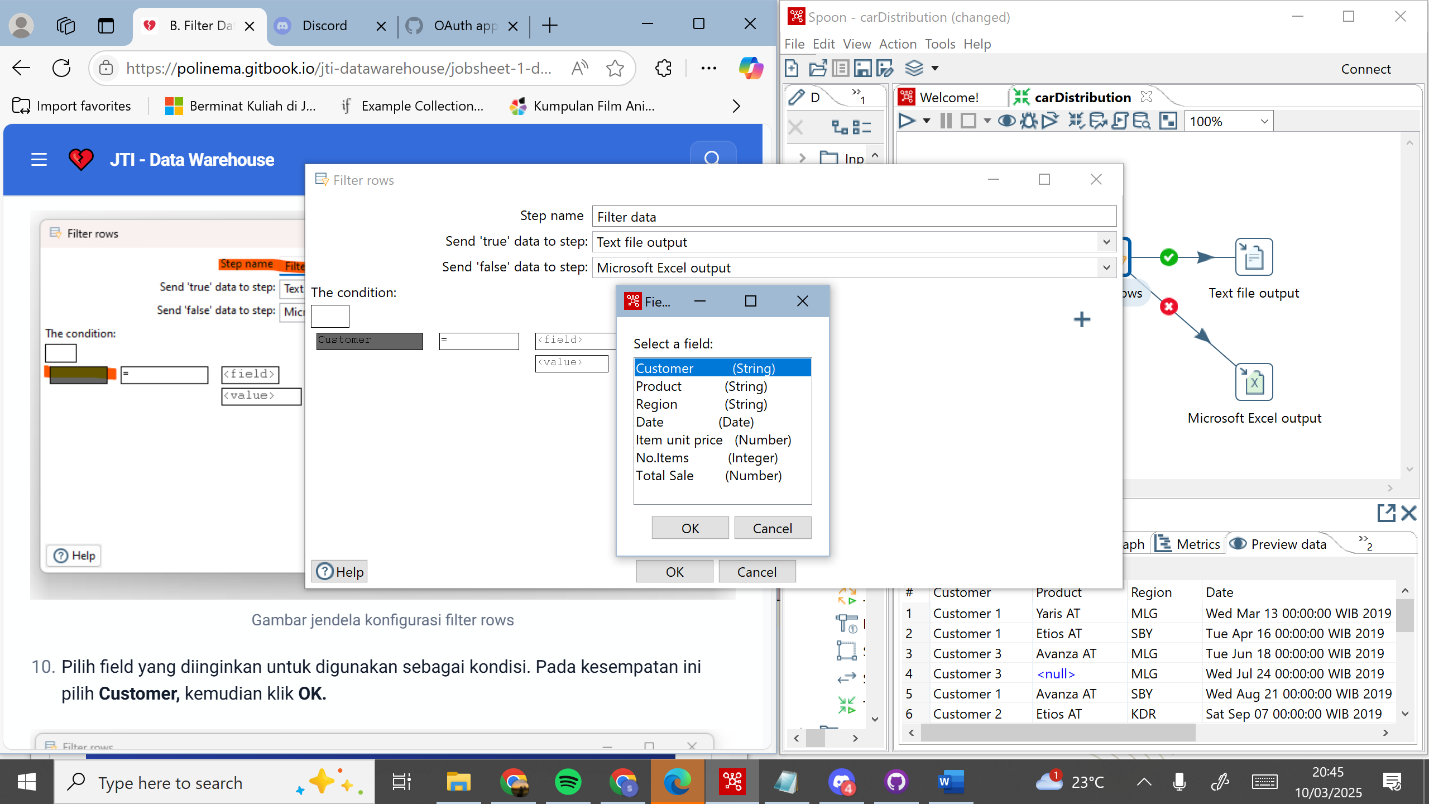


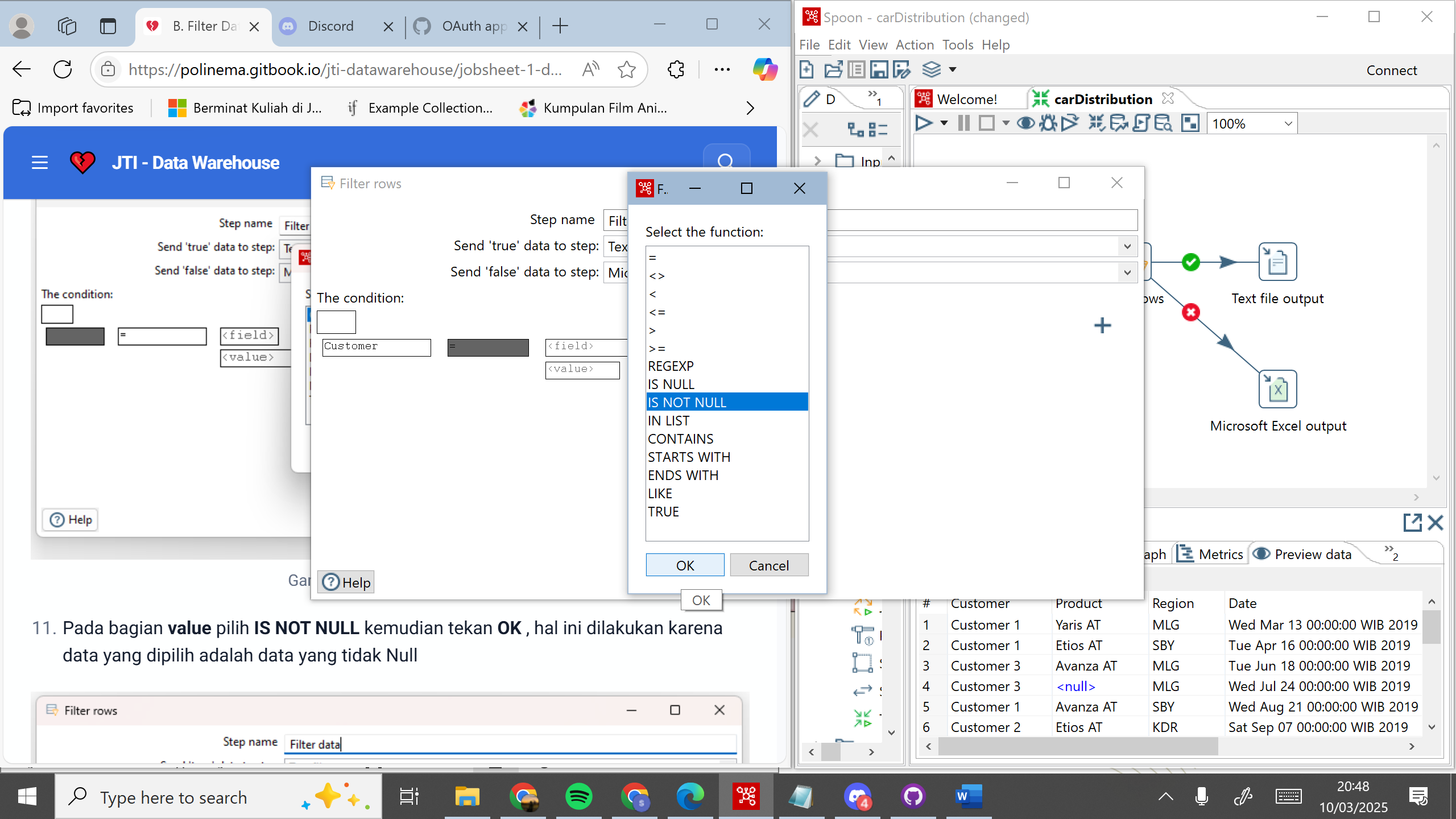
* Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagi jenis konektornya.



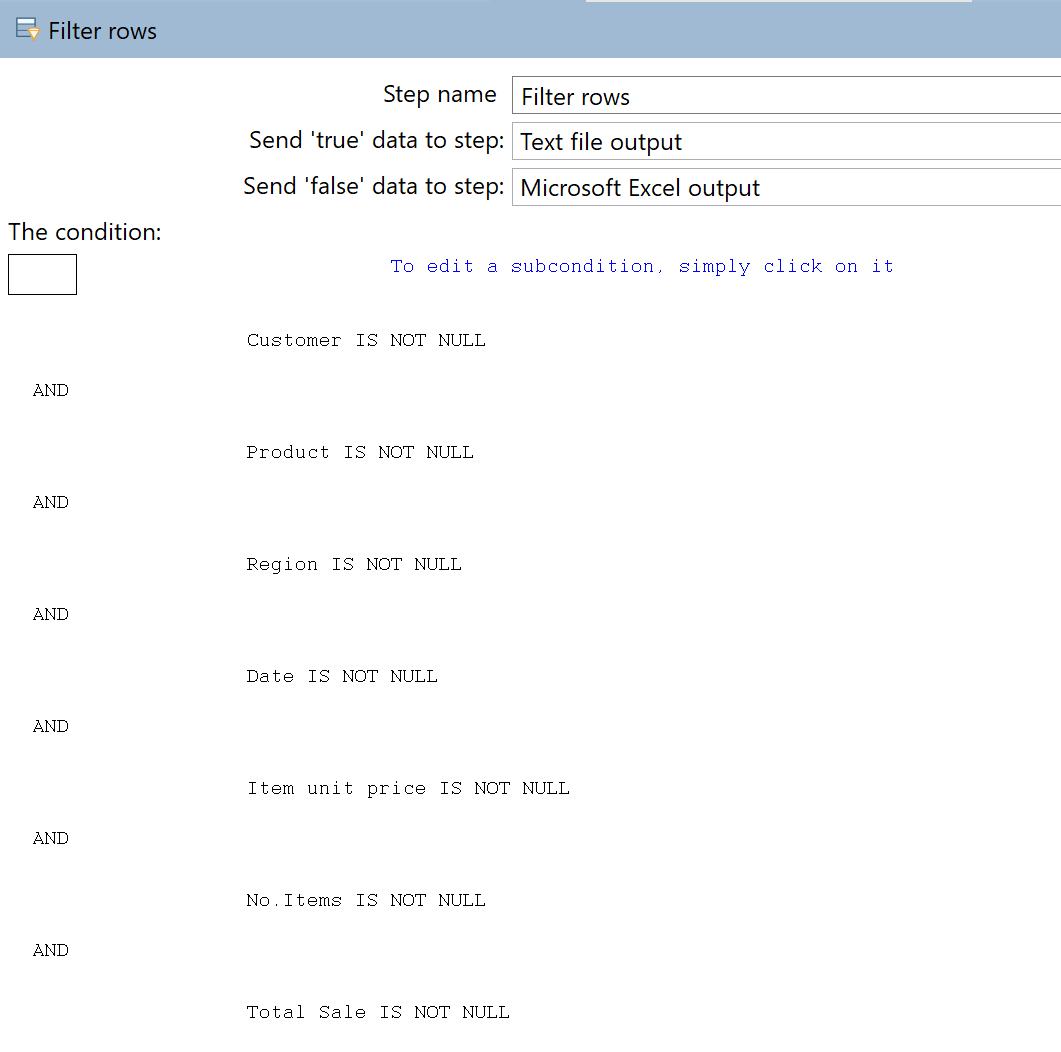


* **Double click** pada **filter rows** hingga muncul jendela konfigurasi**.** ubah **step name** menjadi **Filter data**.  
  Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
* Klik **<Field>** pada bagian **condition, Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.**

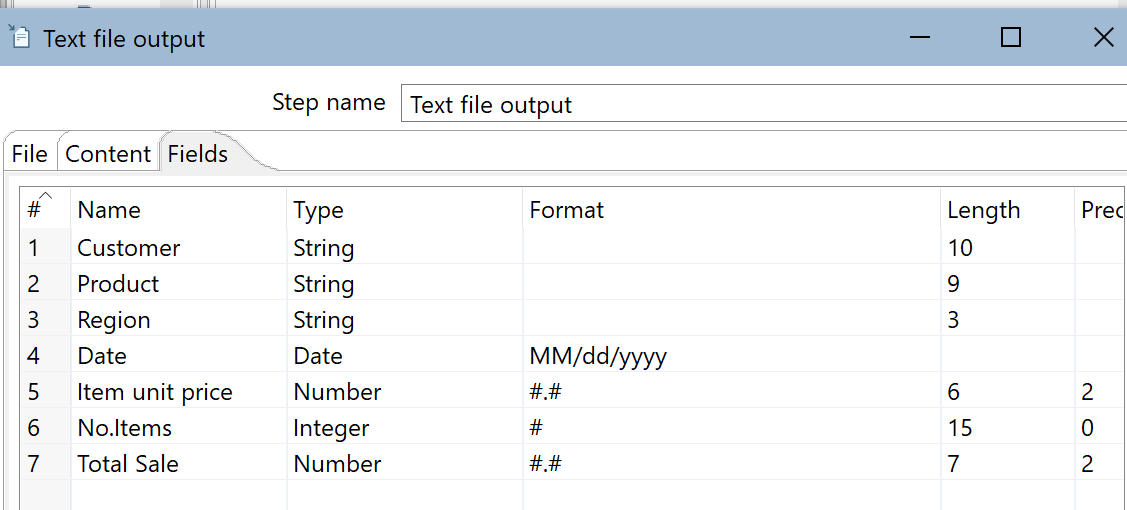


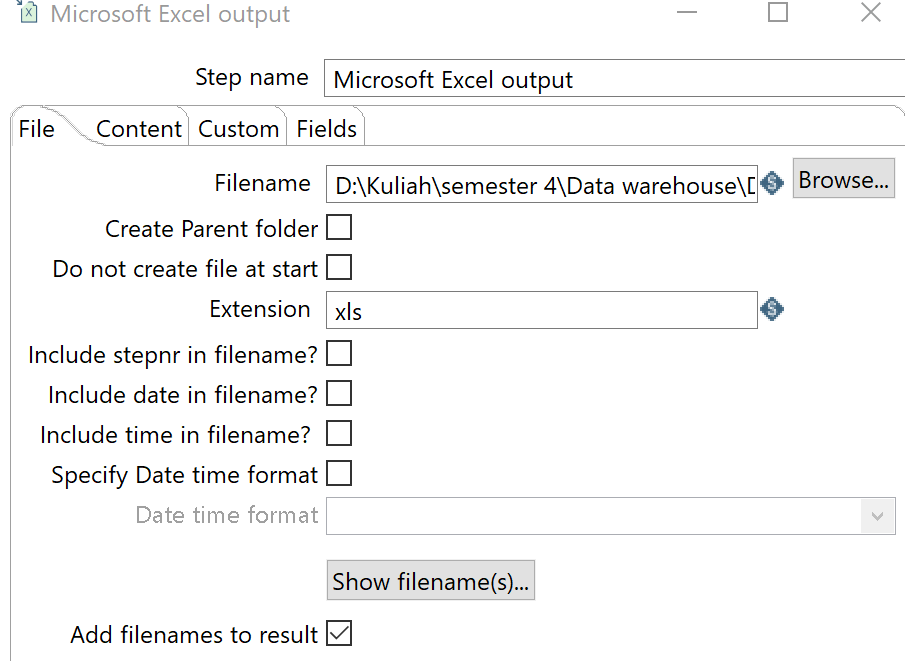


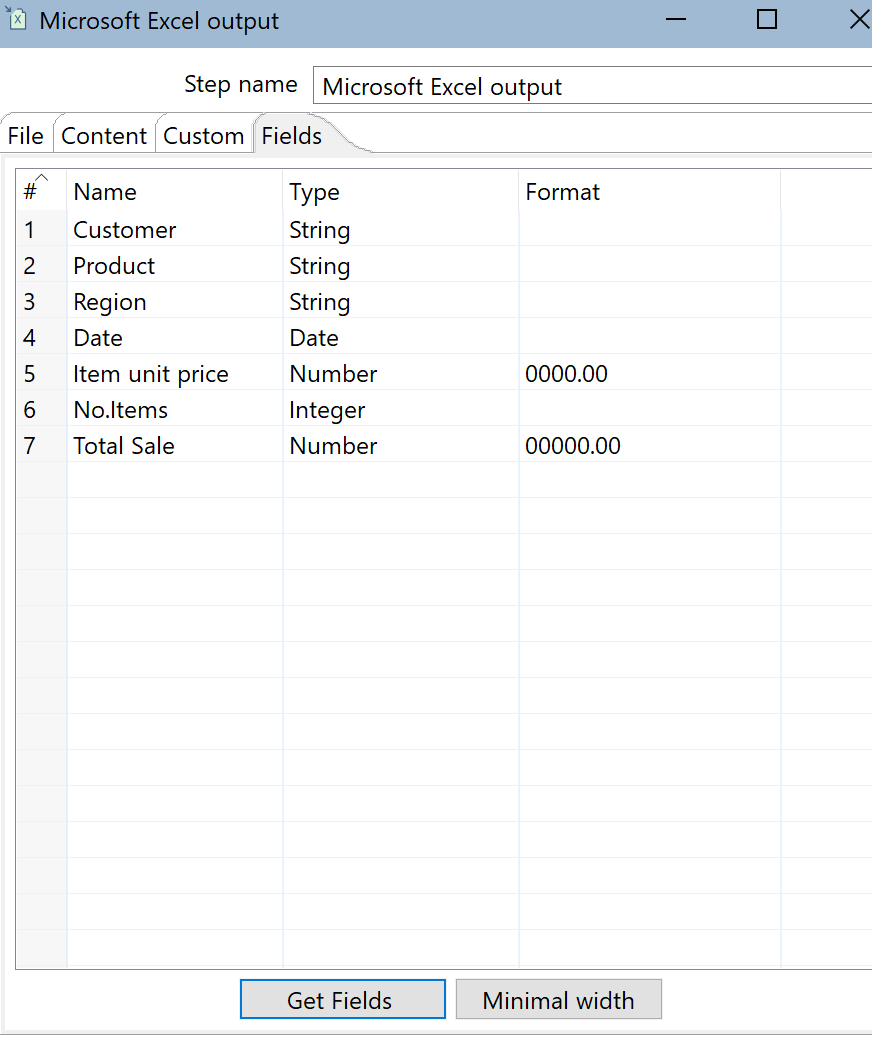
* Tekan tombol **+** untuk menambah kondisi. Pilih **statement** yang masih null untuk menambah kondisi. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
* Gunakan operator **AND** untuk menghubung



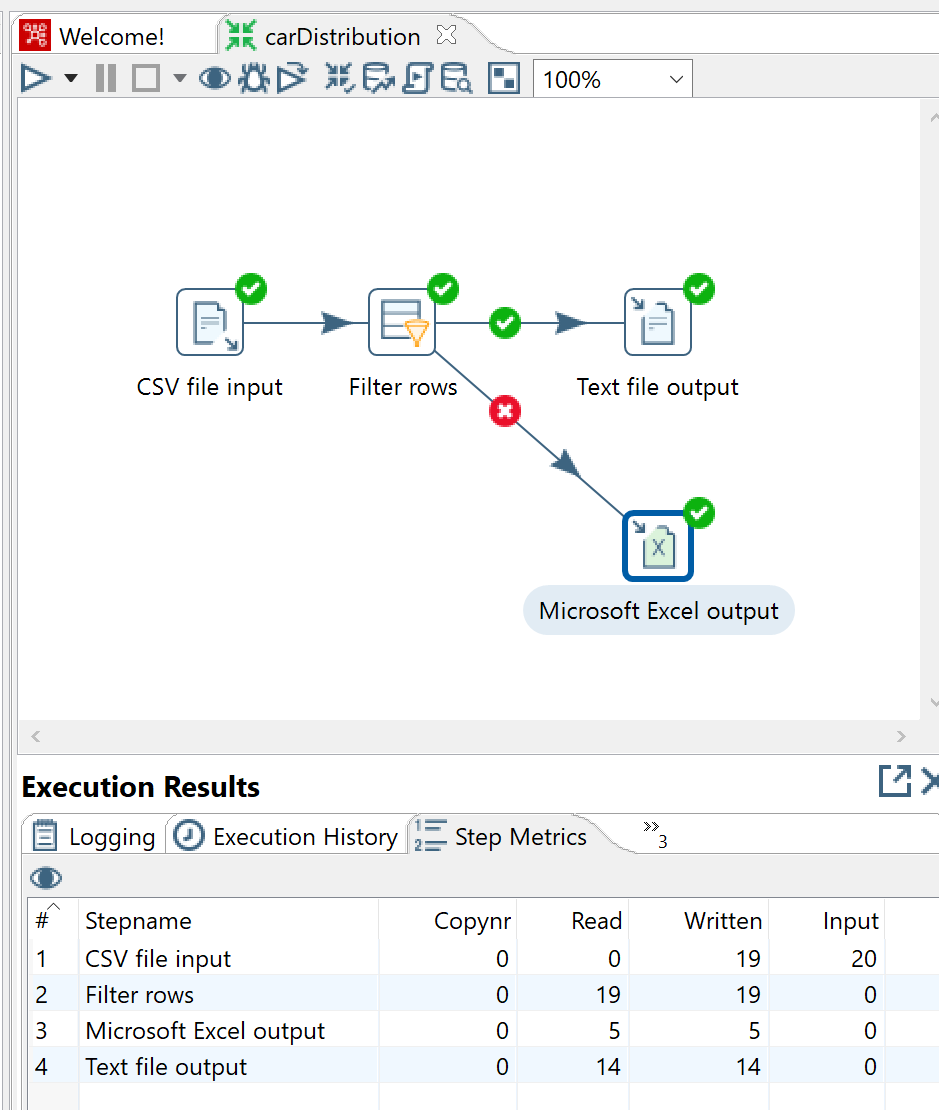
* Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.



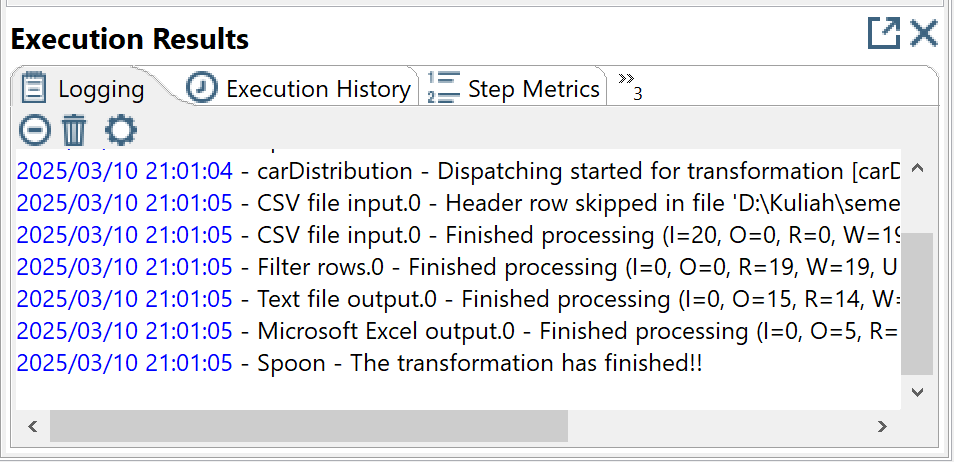




**TUGAS 3**

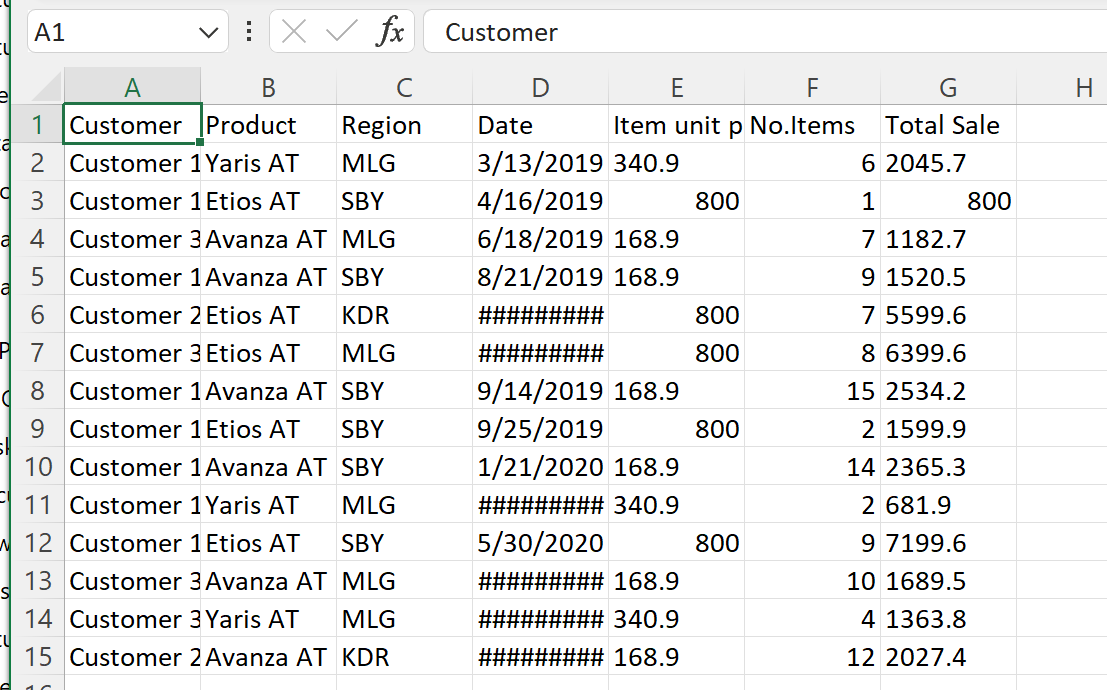


Apabila sudah berhasil di running



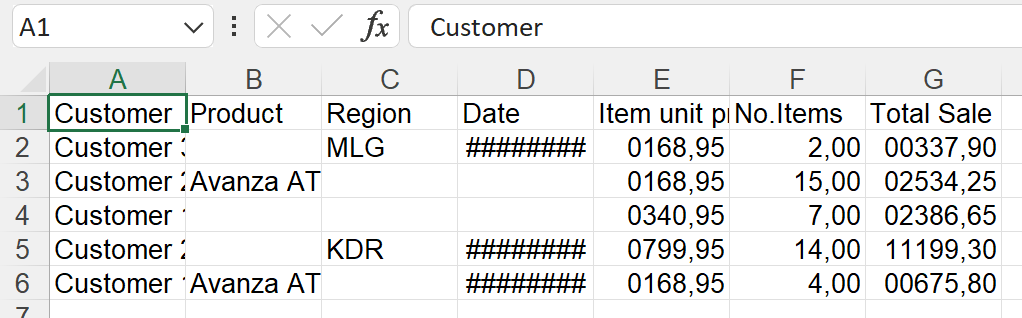
1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

🡪 **CSV:**



**Xls**

**➔** File xls berisi data yang ada nilai null atau tidak lengkap



1. Jelaskan apa yang terjadi pada proses **Filter rows\XDX**!
   * Langkah transformasi yang bertujuna untuk menyaring data berdasrkan kondisi tertentu. Jika true akan diteruskan ke langkah selanjutnya, jika false maka dapat dihapus atau diteruskan ke langkah yang lainnya.
2. 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 table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu | Prosessor | RAM | Storage | VGA | CPU |
| 1 | Cindy | 6s | Intel i5- 10300H | 12 GB | HDD 1 TB | NVIDIA GTX 1050 | 60% |
| 2 | Sabrina | 1s | intel i5-7300U | 8192 MB(8GB) | Samsung 239 GB | Intel ® graphics 620 | 50% |
| 3 | nada | 1s | 11th gen intel®core™i5- | 16 GB | SSD NVMe 477GB | Intel®Iris ® Xe graphics | 23% |
| 4 | zahra | 1s | Intel®core ™i7-13650HX(13th Gen, 20 Cores) | 16 GB | SSD 477GB | Intel ® UHD graphics  +  NVIDIA GeForce RTX 4050 | 7% |