

JOBSHEET 1
Data Warehouse Sederhana dan Pengenalan PDI Spoon
MATA KULIAH DATA WAREHOUSE



Di Susun Oleh:
Devin I'zaz Radin Dewantoro / 2341760034
SIB-2A / 10

PROGRAM STUDI D4 SISTEM INFORMASI BISNIS
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2024

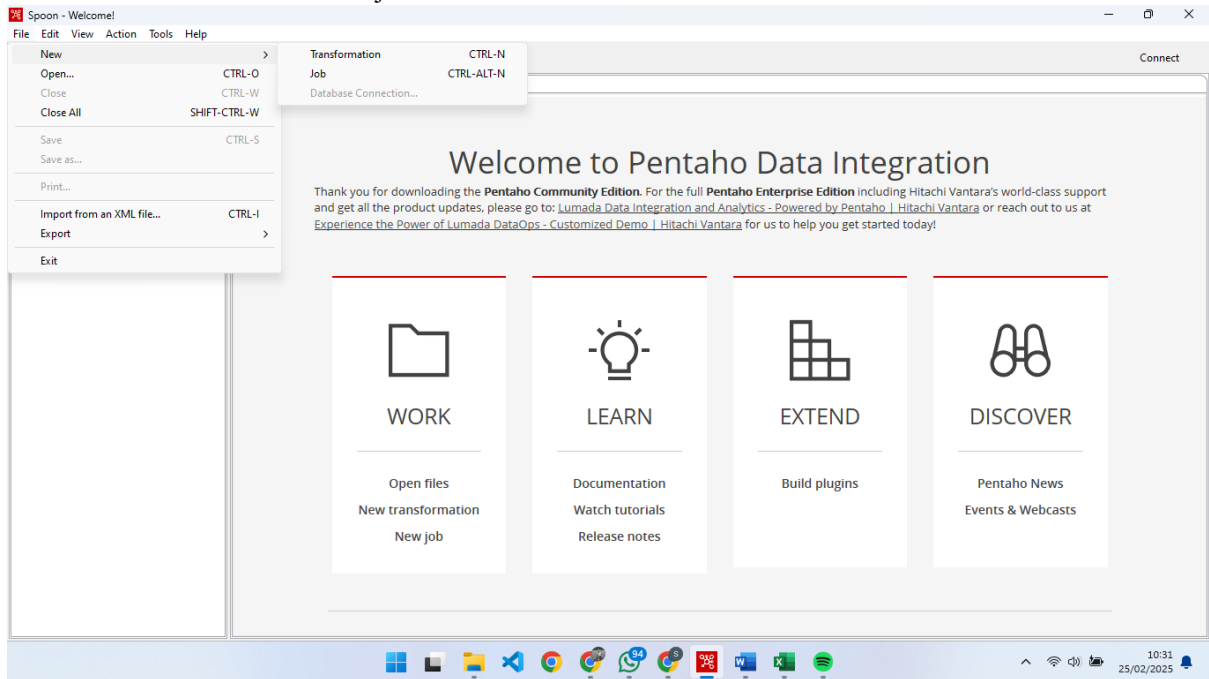
TUGAS 1

Analisa lah data tersebut!

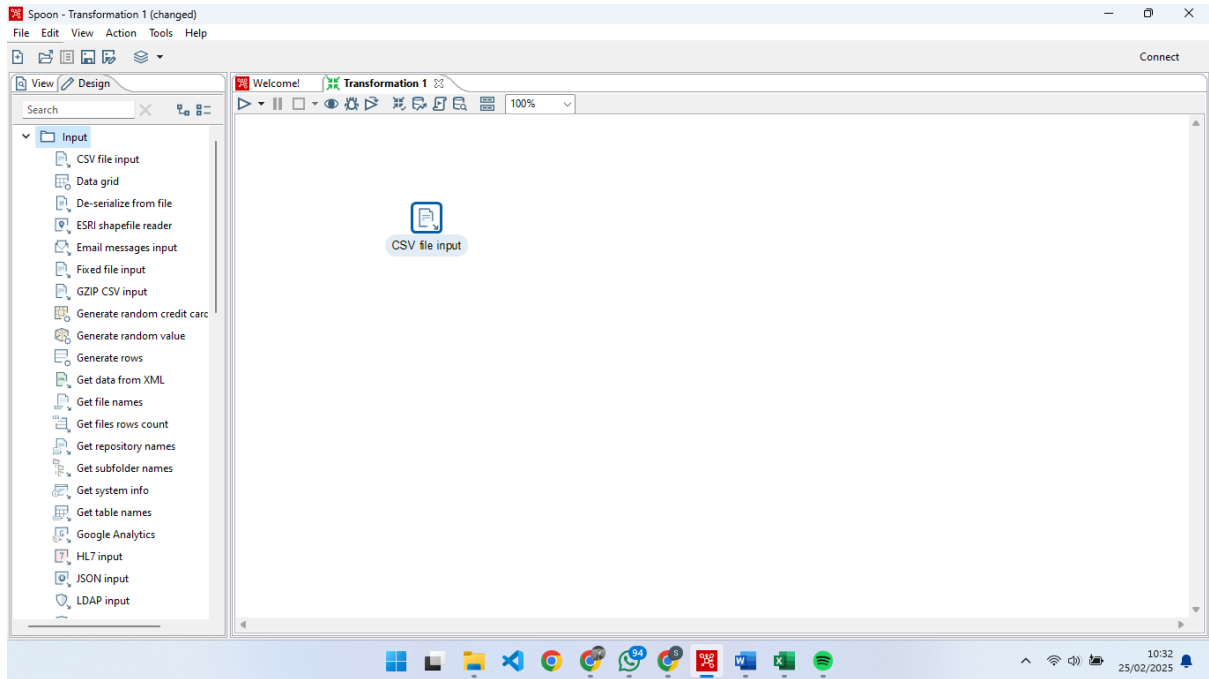
1. Berapa jumlah kolom pada data tersebut? *7 kolom.*
2. Apa arti atau isi data dari setiap kolom yang ada? *Di dalam data memiliki 7 kolom terdiri dari Customer, Product, Region, Date, Item unit price, No. items, dan Total sale dengan kolom Customer, Product, Region menggunakan ber tipe varchar kemudian kolom Date menggunakan tipe date and time, dan kolom Item unit price dan total sale menggunakan tipe data float lalu No.items ber tipe integer.*
3. Adakah data yang memiliki nilai null / data yang tidak lengkap? *ada*
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama? *Tidak ada*

A. Pengambilan Data (Extract)

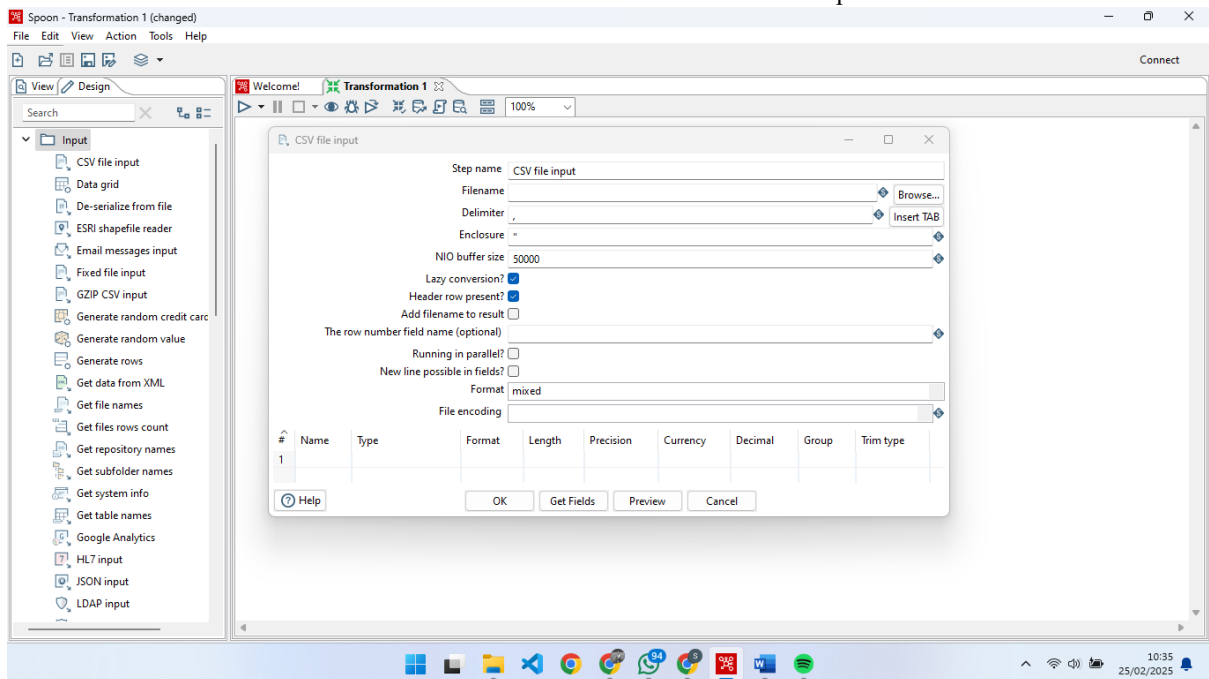
1. Buka lembar kerja Transformation melalui toolbar File.

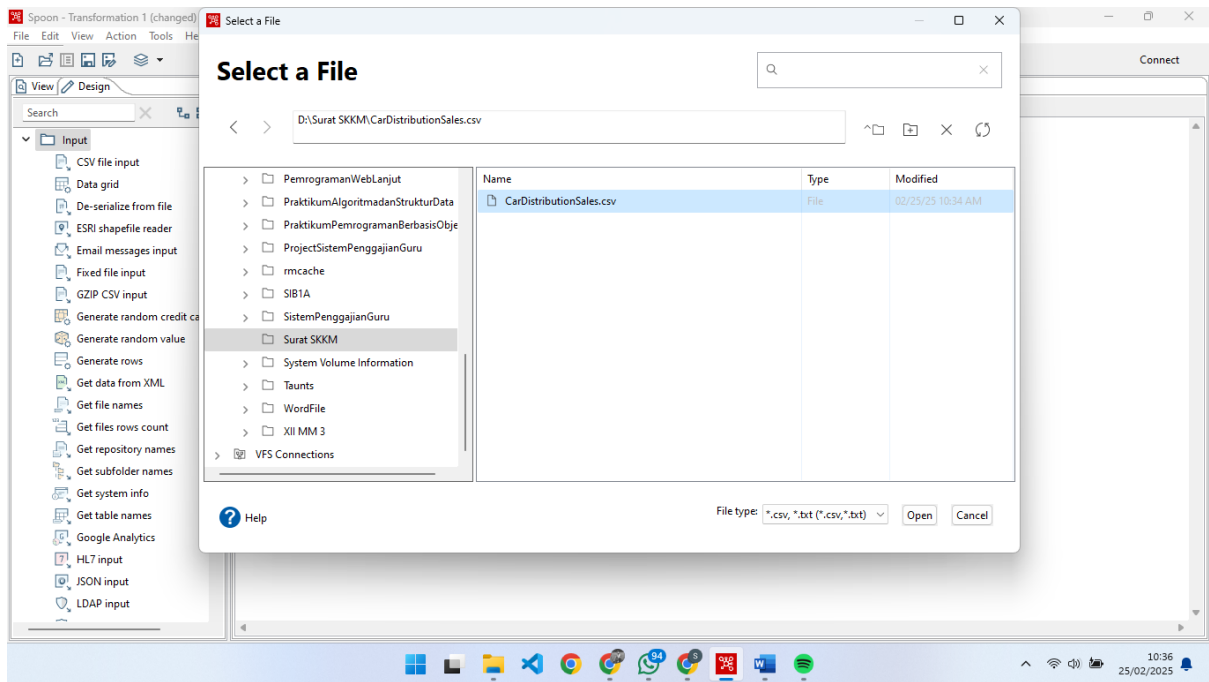


2. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.

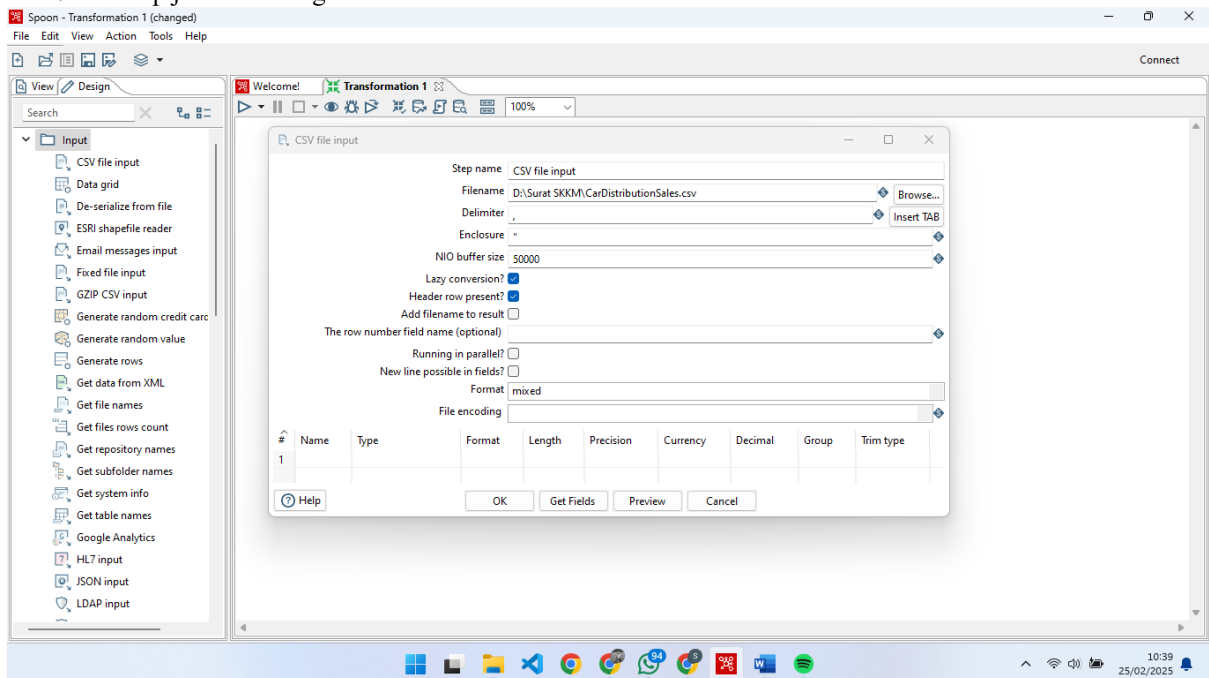


3. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasi.
4. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
5. Pilih lokasi file **CarDistributionSales.csv** melalui **Browse** pada **filename**.





5. Jika data telah dipilih pastikan pada bagian **Delimiter** sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , ").
6. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai
7. Tutup jendela konfigurasi tersebut.



CSV file input

Step name: CSV file input

Filename: D:\Surat SKKM\CarDistributionSales.csv Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result: ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

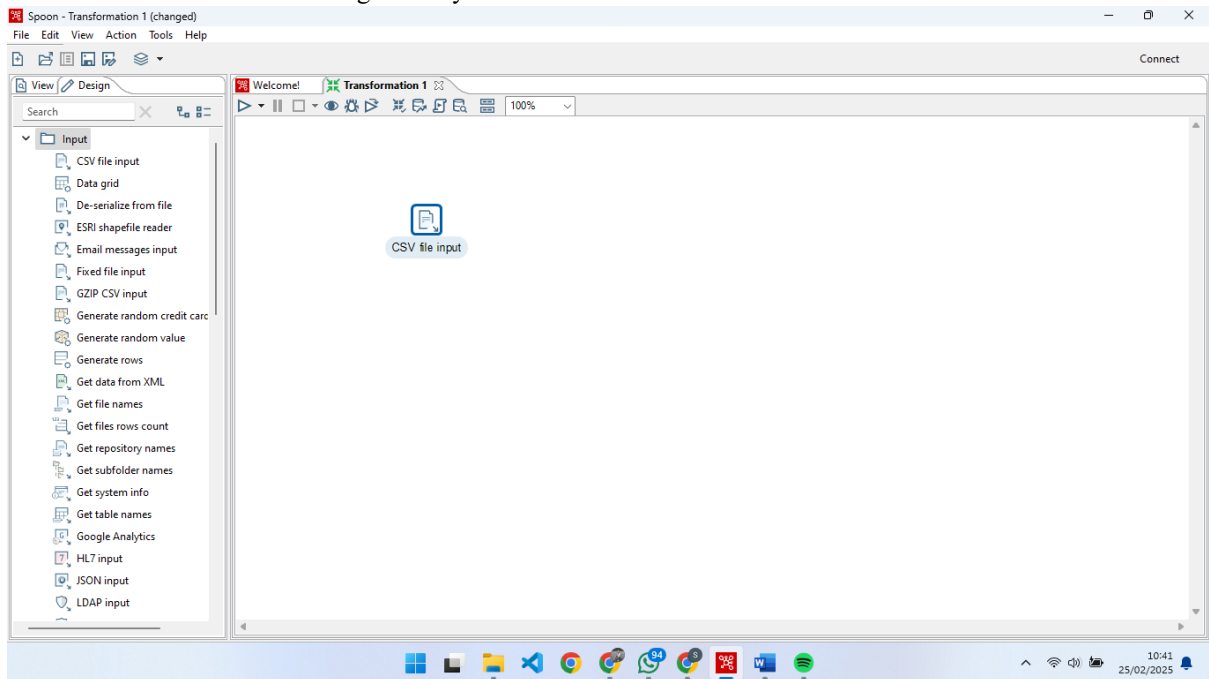
Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Customer	String		10		£	.	,	none
2	Product	String		9		£	.	,	none
3	Region	String		3		£	.	,	none
4	Date	Date	MM/dd/yyyy			£	.	,	none
5	Item unit price	Number	#,.	6	2	£	.	,	none
6	No.Items	Integer	#	15	0	£	.	,	none
7	Total Sale	Number	#,.	7	2	£	.	,	none

Help OK Get Fields Preview Cancel

8. Tekan tombol **"Run"** pada pojok kanan atas **Area Kerja**, maka akan muncul **Execution Result Area**. Pilih tab **Prview Data** pada **Execution Result Area**. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.



Execution Results

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	<null>	<null>	<null>	<null>	<null>	<null>	<null>
2	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
3	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
4	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
5	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
6	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
7	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
8	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
9	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
10	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
11	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
12	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
13	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
14	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
15	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55
16	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
17	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8
18	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
19	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 ICT 2020	168.95	12	2027.4

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

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

Execution Results

Execution History

- 2025/02/26 13:03:52 - Spoon - Transformation opened.
- 2025/02/26 13:03:52 - Spoon - Launching transformation [Transformation 1]...
- 2025/02/26 13:03:52 - Spoon - Started the transformation execution.
- 2025/02/26 13:03:52 - Transformation 1 - Dispatching started for transformation [Transformation 1]
- 2025/02/26 13:03:53 - CSV file input.0 - Header row skipped in file 'D:\Surat SKKM\CarDistributionSales.csv'
- 2025/02/26 13:03:53 - CSV file input.0 - Finished processing (I=21, O=0, R=0, W=20, U=0, E=0)
- 2025/02/26 13:03:53 - Spoon - The transformation has finished!!

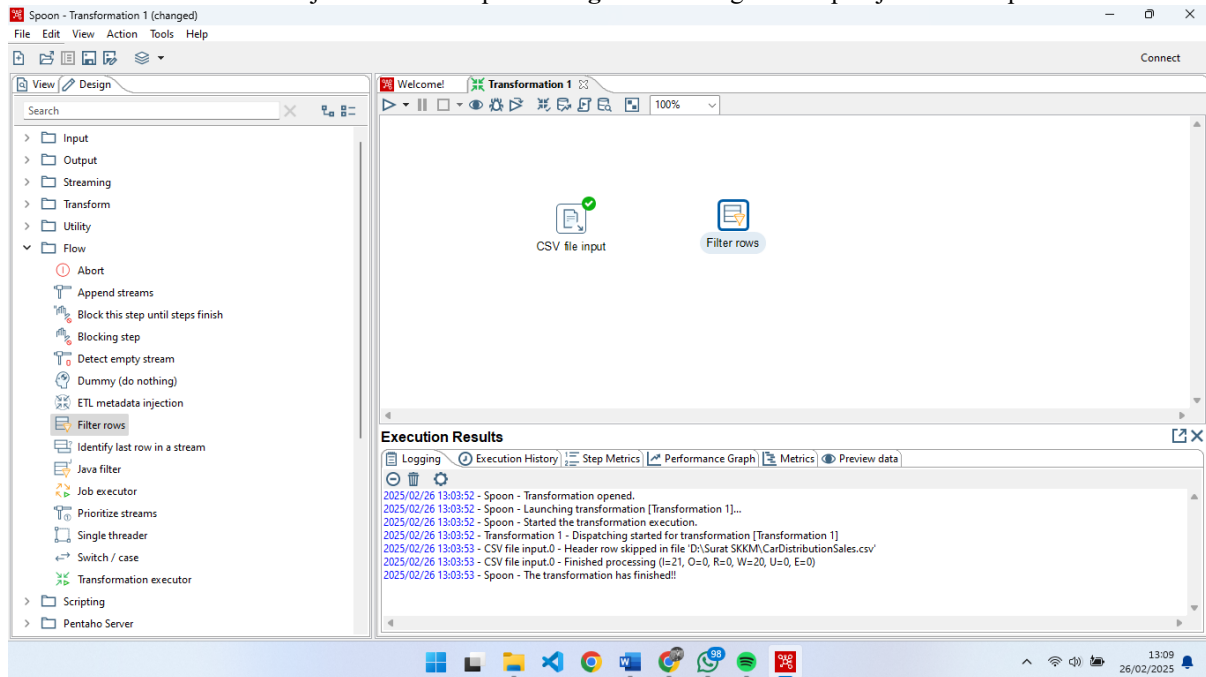
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?
2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
Menunjukkan output yang ditampilkan seperti jumlah tabel dan keluaran yang di tampilkan.

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.

No.	Nama	Processor	RAM	Storage	VGA	CPU	Waktu Transformasi
1	Devin	Intel	4 GB	128 GB	Intel HD Graphics 4000	Intel Core i5 3437U	1 detik
2	Filla	AMD	6 GB	1000 GB	Radeon Vega 8 Mobile GF	AMD Ryzen 5 3500u Processor with AMD Radeon Vega 8 Graphics	0.4 detik
3	Gegas	Intel	16 GB	512 GB	RTX 2050	Intel Core i7 12TH	0.4 detik
4	Audric	Intel	8 GB	256 GB	RTX 2050	Intel Core i5 13	0.4 detik

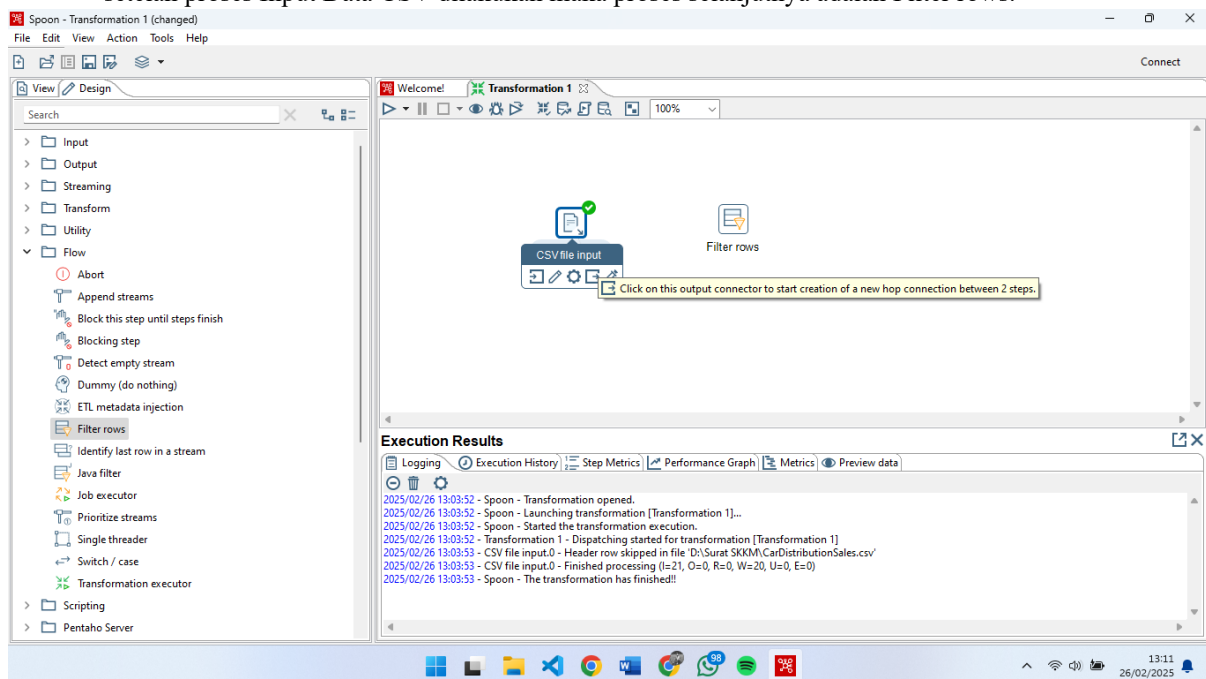
B. Filter Data (Transform) dan Pengemasan data (Load)

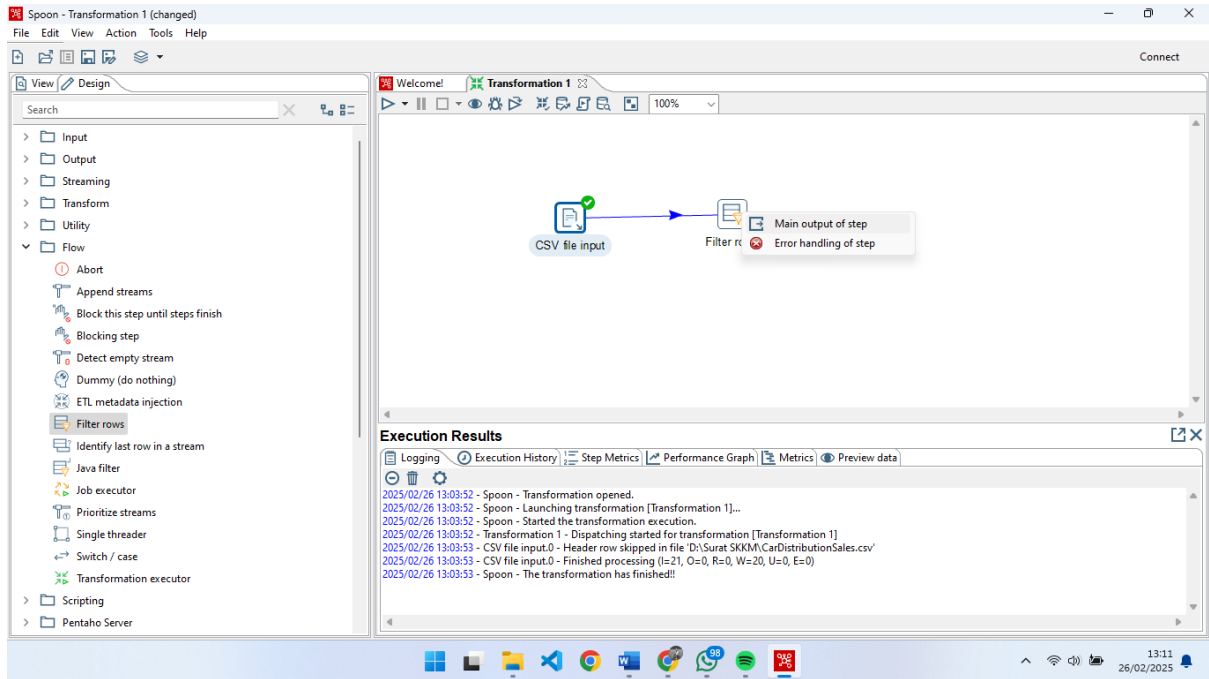
1. Carilah objek **Filter rows** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.



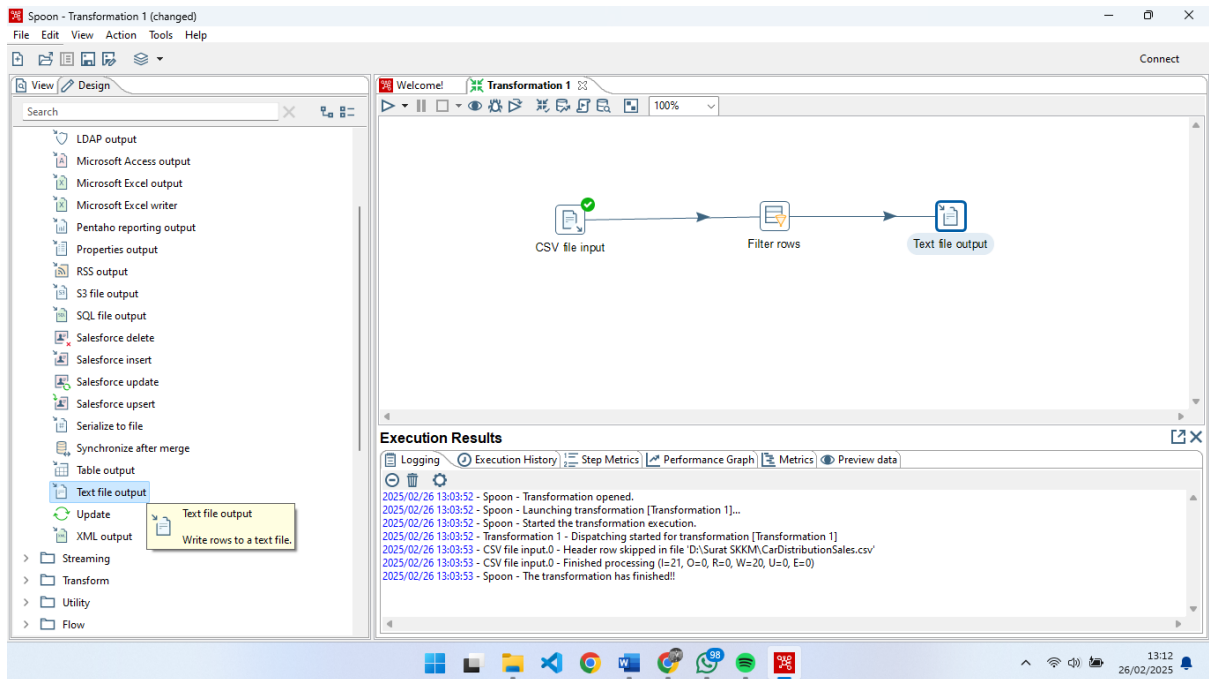
2. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.

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.

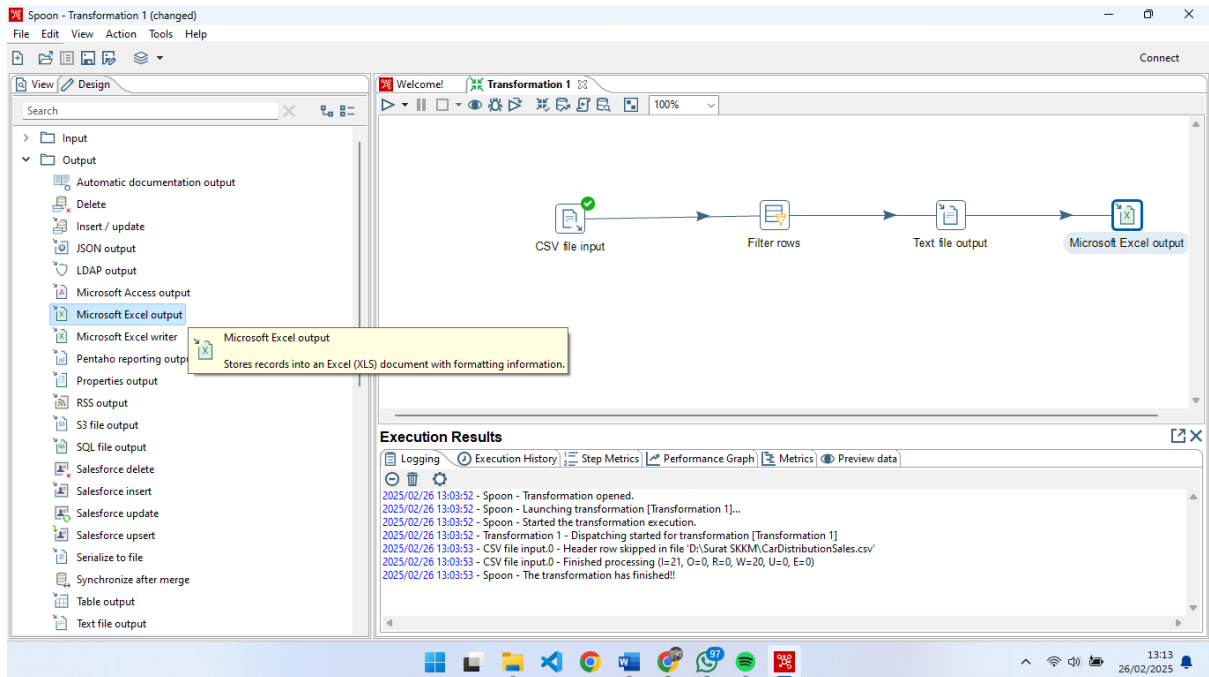




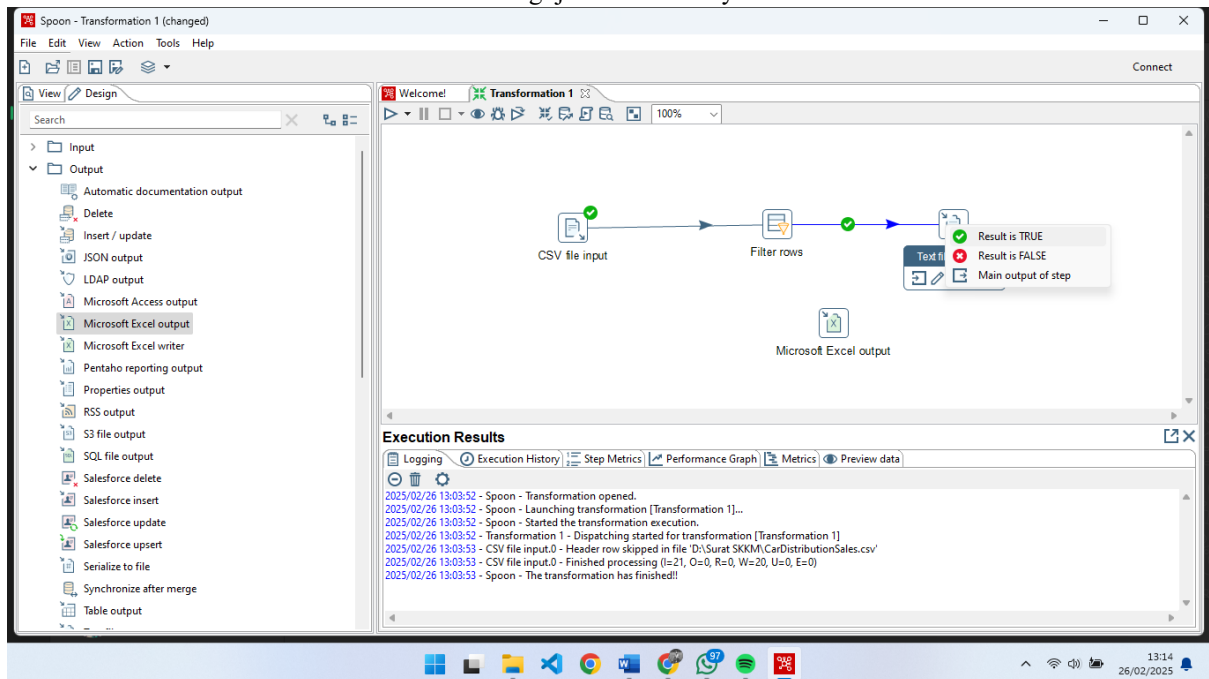
3. Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.



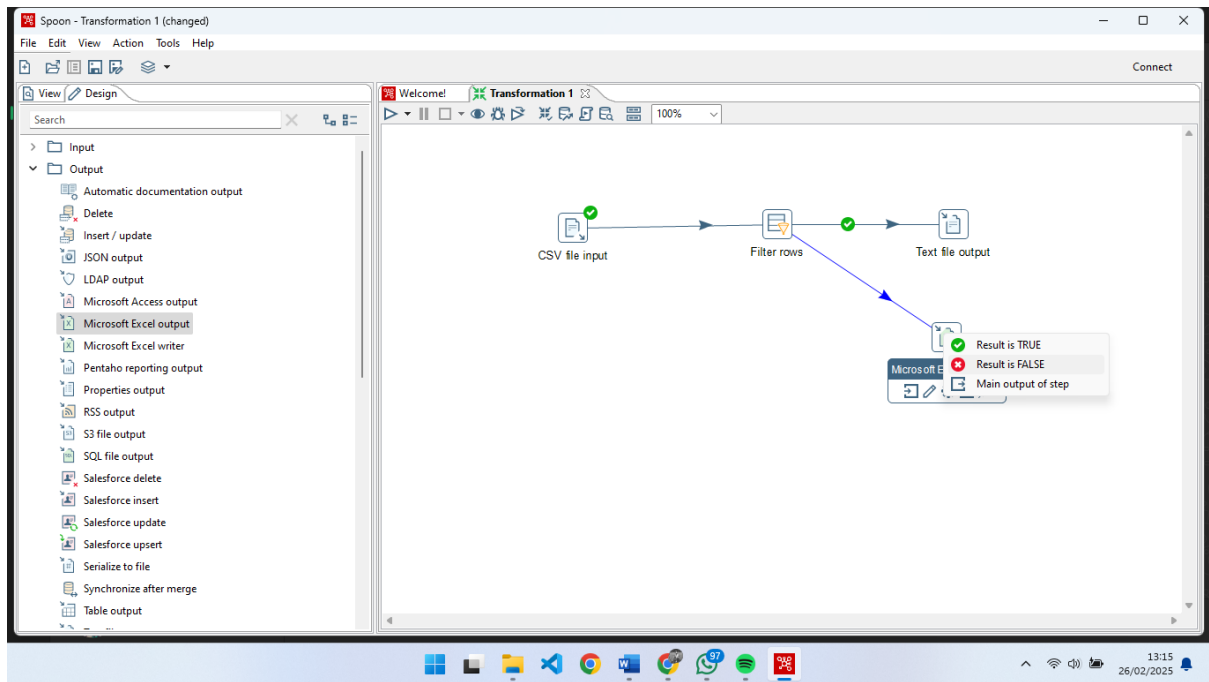
4. Carilah objek **Microsoft excel output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.



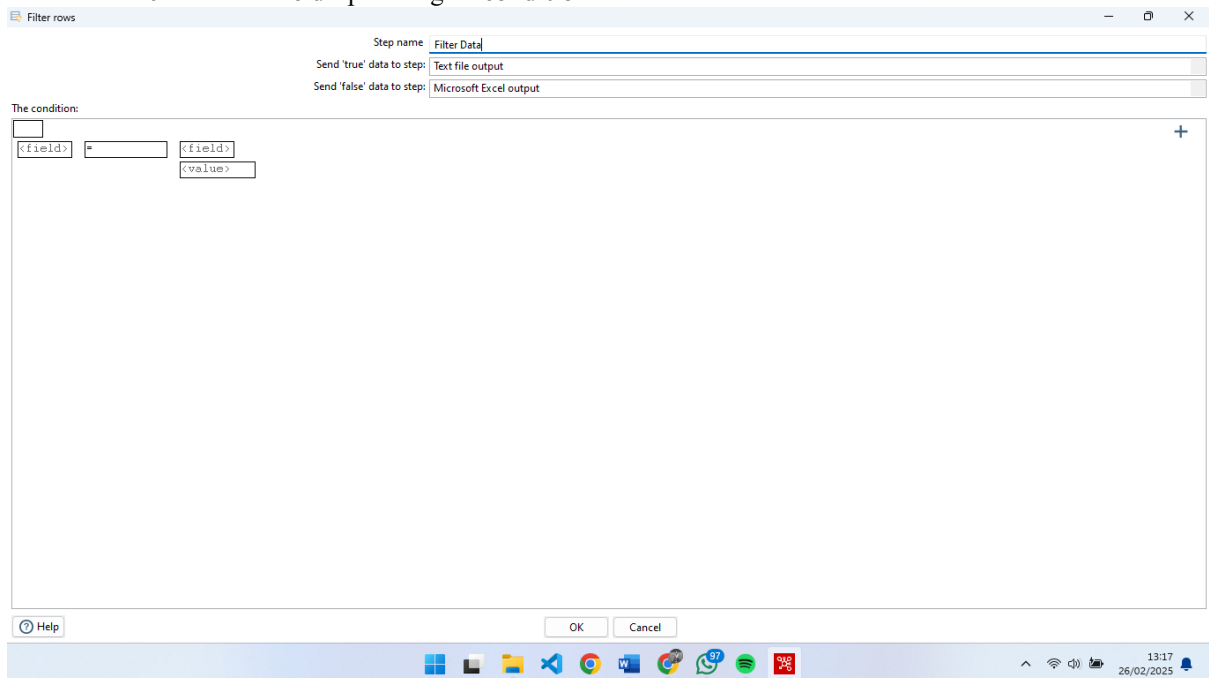
5. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagai jenis konektornya.



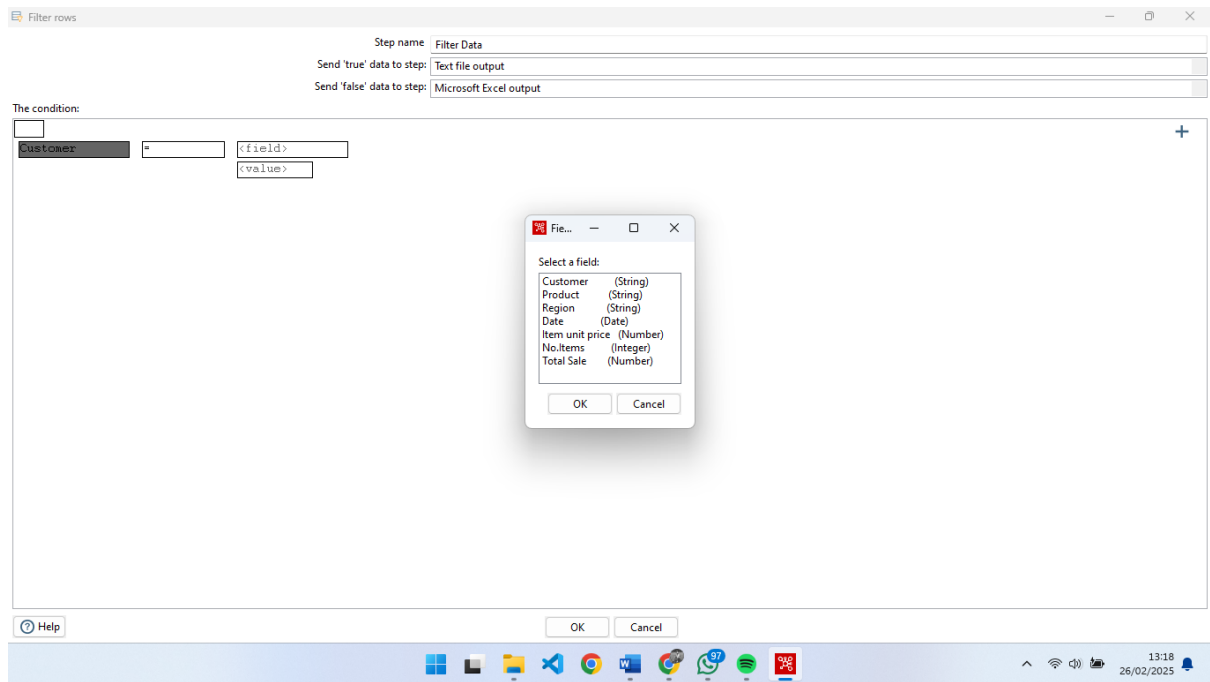
6. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagai jenis konektornya.



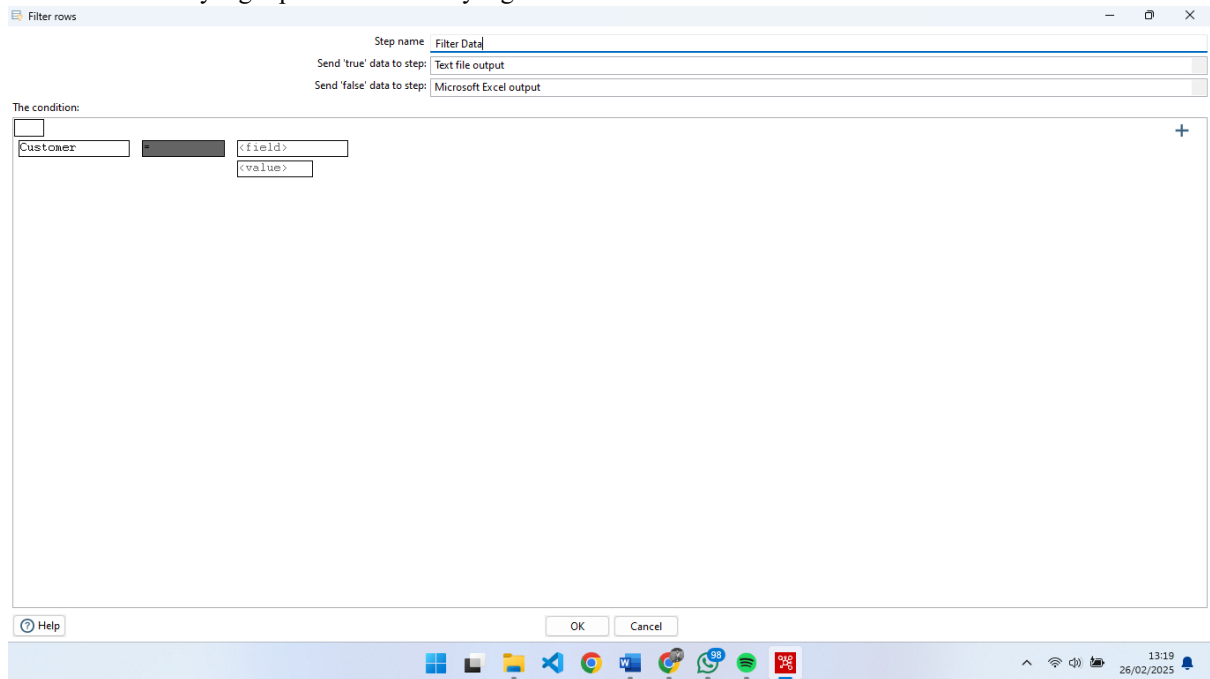
7. **Double click** pada **filter rows** hingga muncul jendela konfigurasi. ubah **step name** menjadi **Filter data**.
8. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
9. Klik **<Field>** pada bagian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

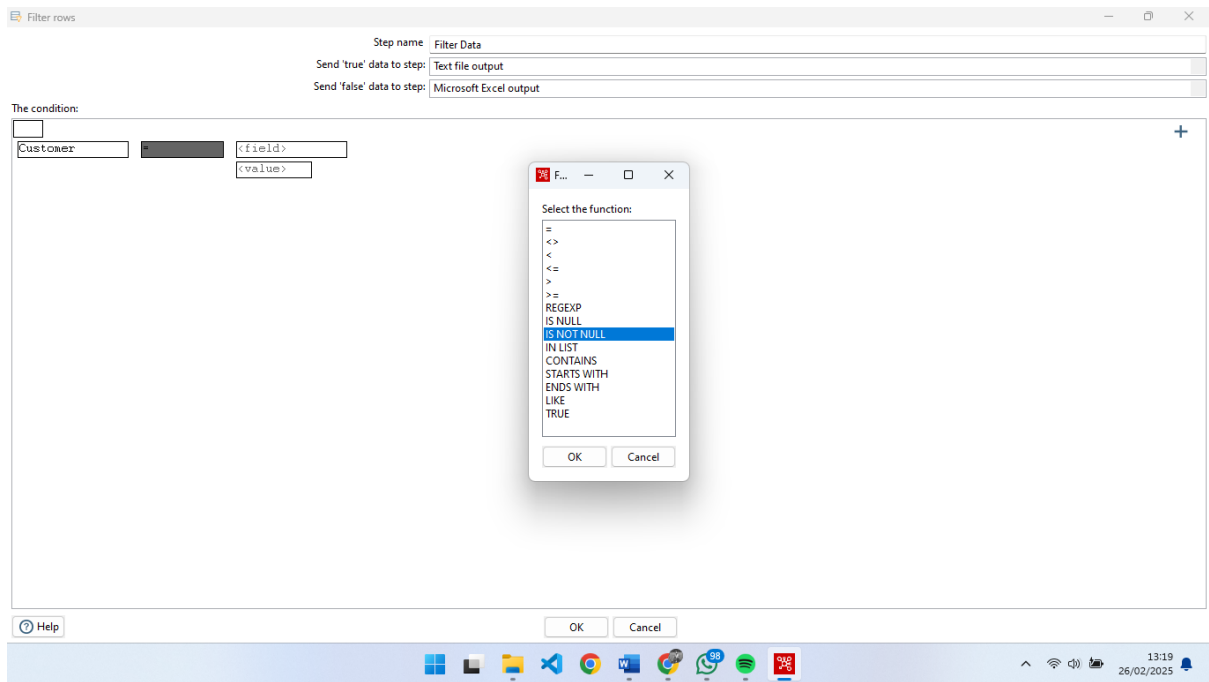


10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer**, kemudian klik **OK**.

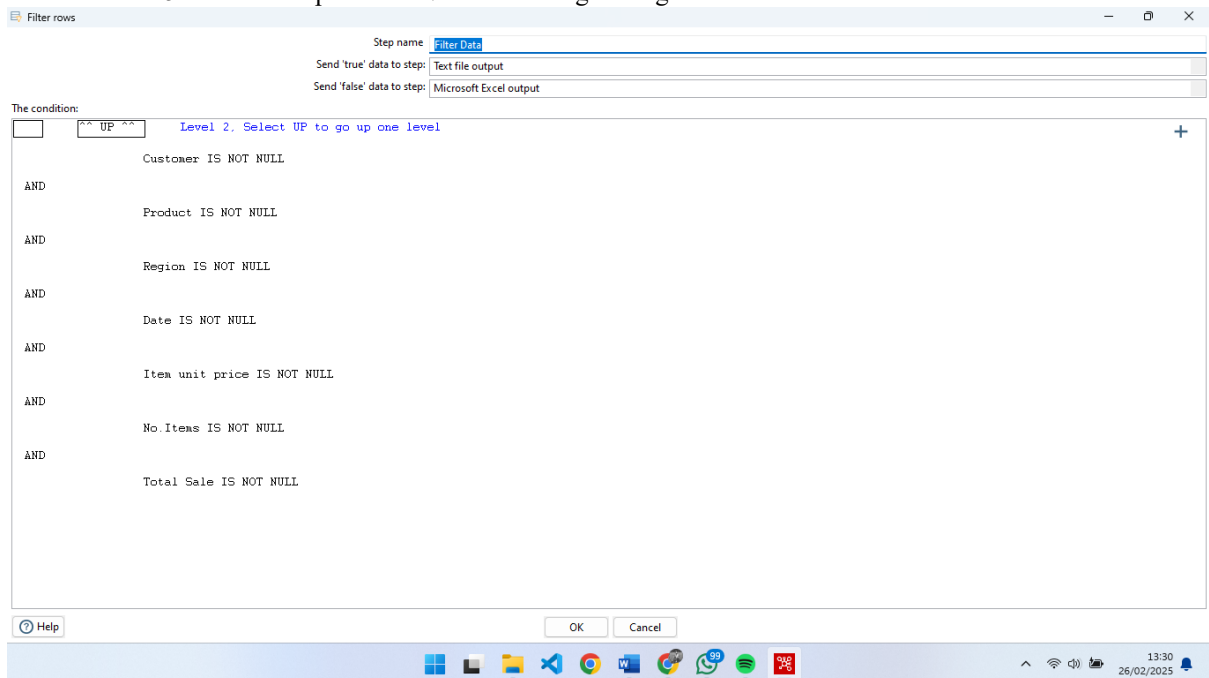


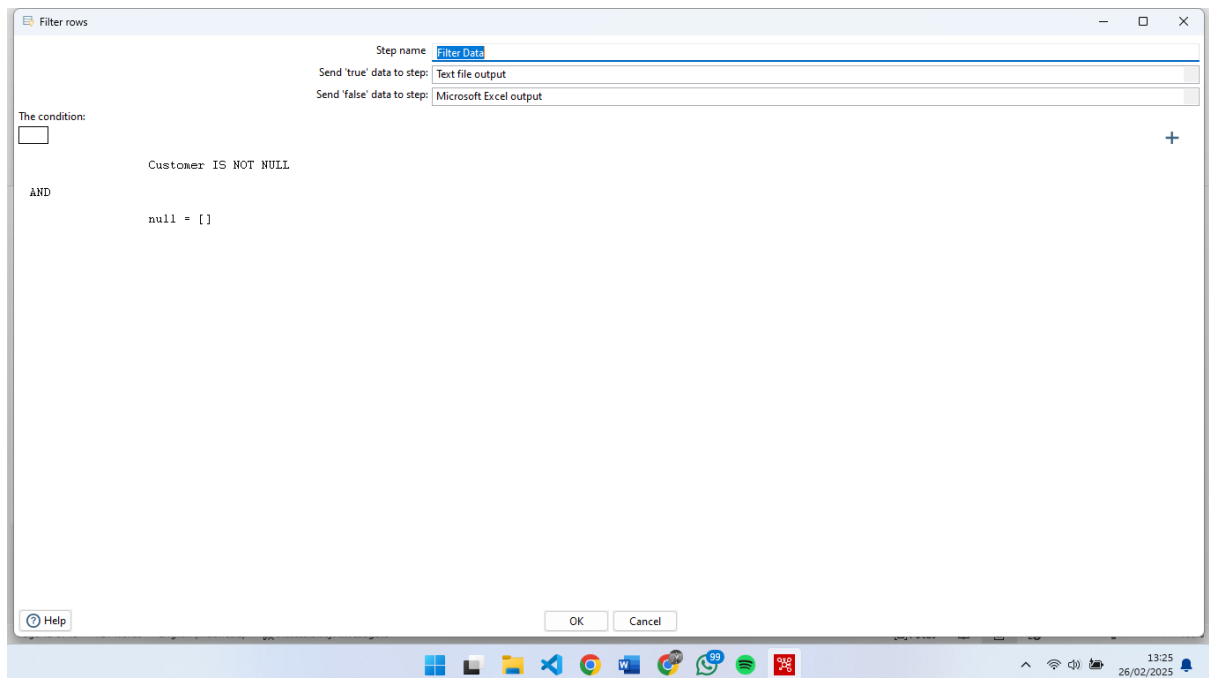
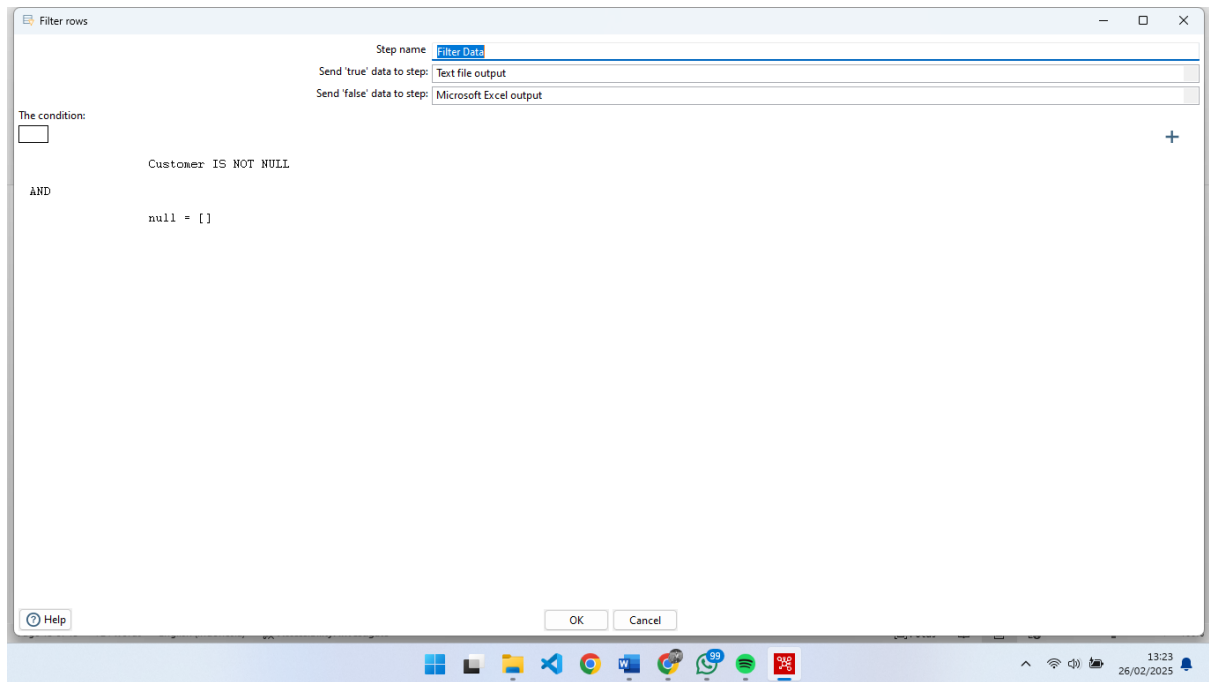
11. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null





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





16. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.
17. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**.
18. Ubah extension menjadi **csv**.
19. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.

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

Text file output

Step name: Text file output

File Content Fields

Filename: [Browse...](#)

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field ☐

File name field:

Extension: [Browse...](#)

Include stepnr in filename? ☐

Include partition nr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format:

Date time format:

Show filename(s)... [Show filename\(s\)...](#)

Add filenames to result ☒

OK Cancel

Help

13:31 26/02/2025

Text file output

Step name: Text file output

File Content Fields

Filename: [Browse...](#)

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field ☐

File name field:

Extension: [Browse...](#)

Include stepnr in filename? ☐

Include partition nr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format:

Date time format:

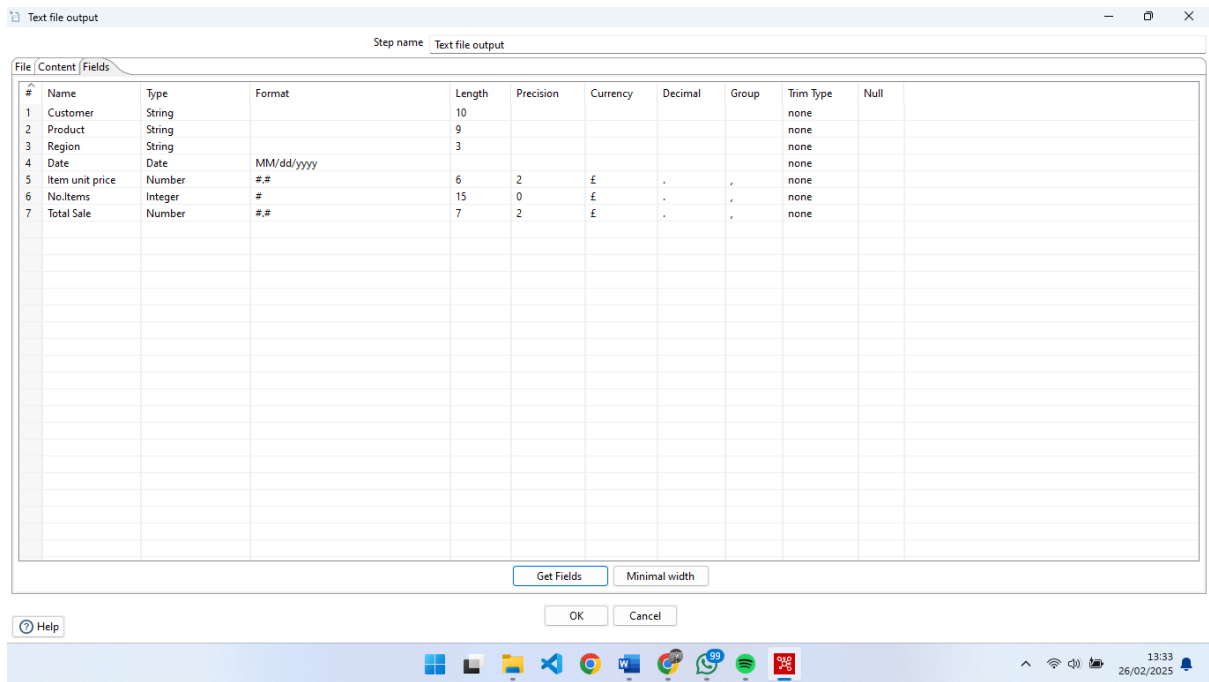
Show filename(s)... [Show filename\(s\)...](#)

Add filenames to result ☒

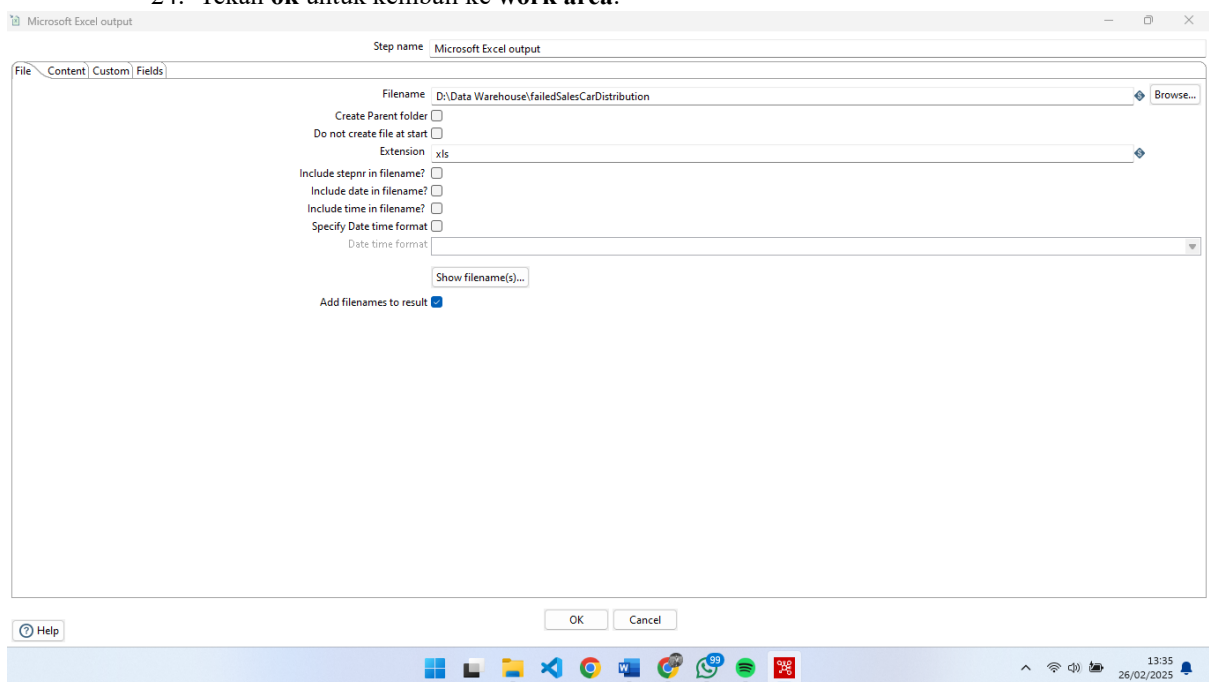
OK Cancel

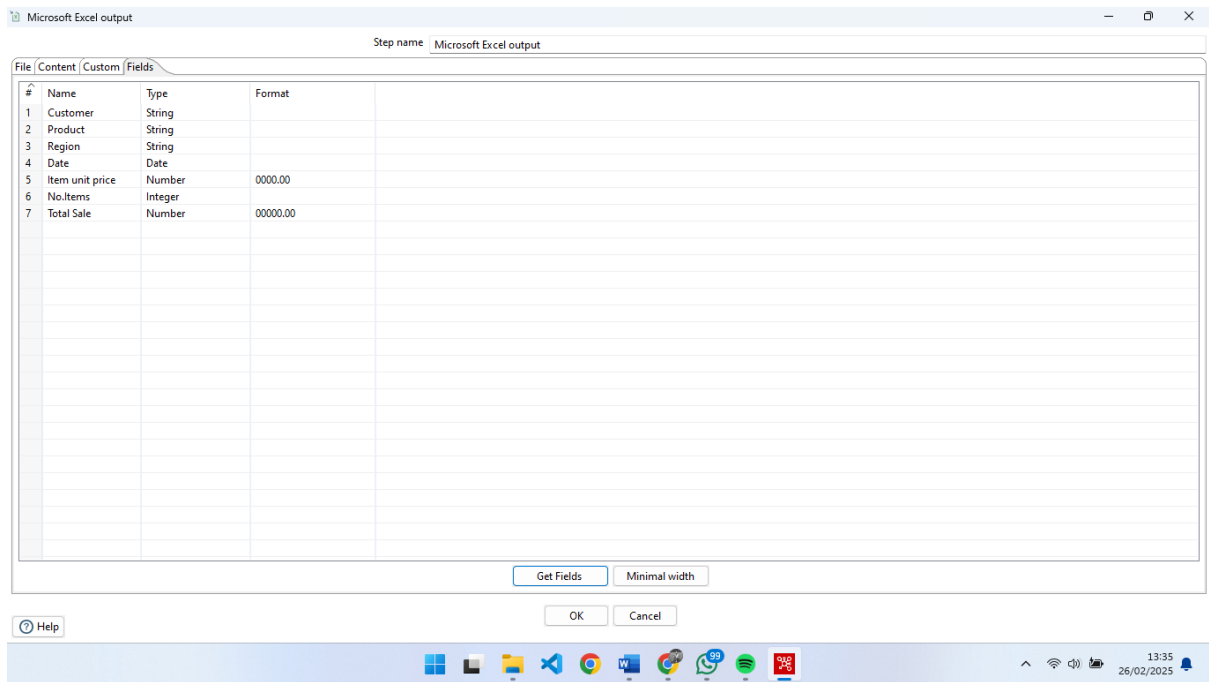
Help

13:33 26/02/2025

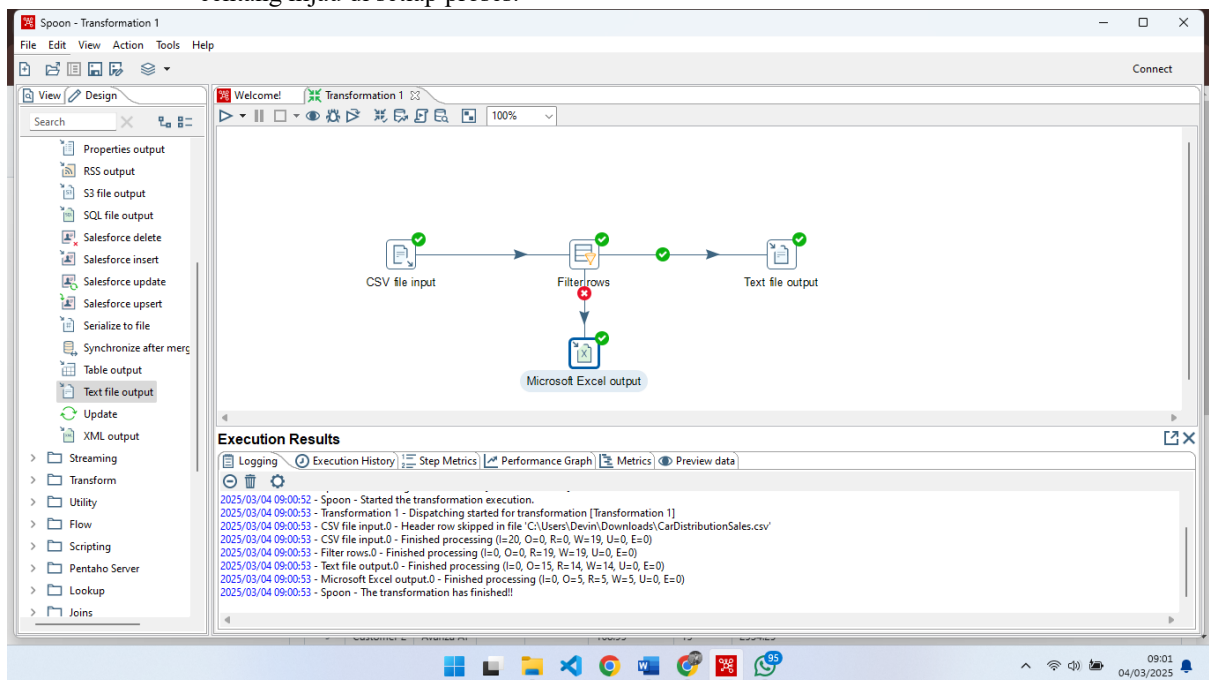


21. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output**, **Double-click** pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.
22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol **browse**. Beri nama **failedSalesCarDistribution**. untuk extension tetap dengan extension **xls**.
23. Pada tab **fields** tekan tombol **Get fields** untuk mengambil field data.
24. Tekan **ok** untuk kembali ke **work area**.





25. Tekan tombol **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?
File CSV dan file Excel memiliki cara penyimpanan data yang berbeda. File CSV menyimpan data dalam bentuk teks yang dipisahkan oleh koma, sedangkan file Excel menyimpan data dalam format yang dapat dibaca oleh aplikasi Excel.
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
Filter rows adalah proses memilih dan menampilkan hanya baris-baris data yang memenuhi kriteria tertentu.
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	Processor	RAM	Storage	VGA	CPU	Waktu Transformasi
1	Gegas	Intel	16 GB	512 GB	RTX 2050	Intel Core i7 12TH	0.4 detik
2	Devin	Intel	4 GB	128 GB	Intel HD Graphics 4000	Intel Core i5 3437U	1 detik
3	Audric	Intel	8 GB	256 GB	RTX 2050	Intel Core i5 13	0.4 detik
4	Ircham	AMD	8 GB	125 SSD 500 HDD	Radeon Vega 8 Mobile GF	AMD Ryzen 5	0.4 detik

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. dari 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.

