**LAPORAN PRAKTIKUM**

**JOBSHEET 01**

**DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON**

Disusun untuk memenuhi nilai tugas

Mata Kuliah : Data Warehouse



Oleh :

Aqueena Regita Hapsari

2341760096

SIB-2B - 03

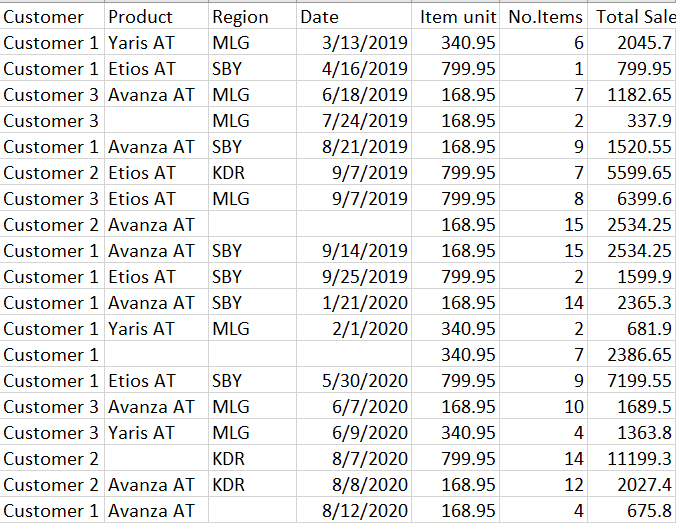
**PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS**

**POLITEKNIK NEGERI MALANG**

**TAHUN AJARAN 2024/2025**

**TUGAS 1**

* **Analisa lah data tersebut!**

****

* 1. Berapa jumlah kolom pada data tersebut?

**Jawaban :**

1. Dataset berisi 19 baris data.
2. Ada 7 kolom dengan informasi terkait penjualan mobil.
   1. Apa arti atau isi data dari setiap kolom yang ada?

**Jawaban :**

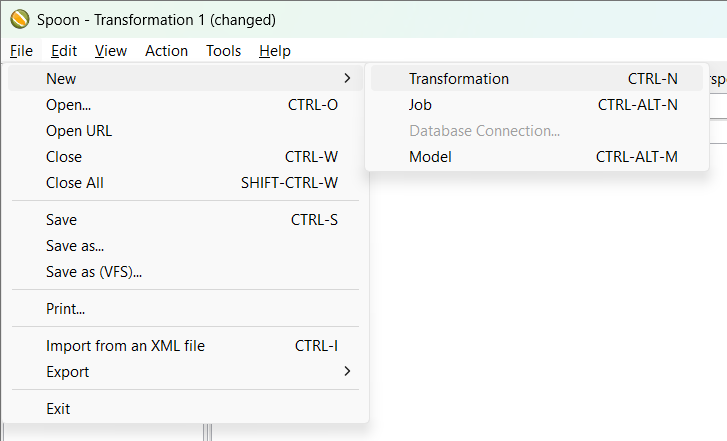
1. Customer : Nama pelanggan atau identifikasi pelanggan.
2. Product : Jenis atau model mobil yang dibeli.
3. Region : Wilayah atau lokasi pembelian.
4. Date : Tanggal transaksi pembelian.
5. Item unit price : Harga satuan per unit kendaraan.
6. No.Items : Jumlah unit kendaraan yang dibeli.
7. Total Sale : Total penjualan berdasarkan harga satuan dan jumlah unit.
   1. Adakah data yang memiliki nilai null / data yang tidak lengkap?

**Jawaban :**

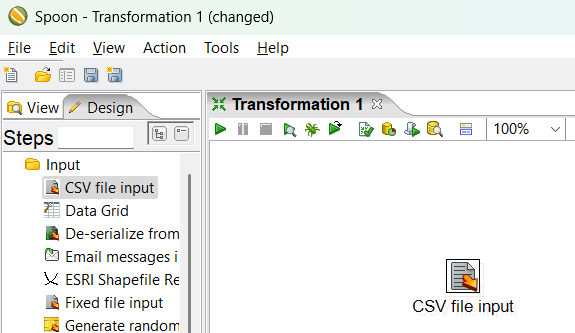
1. Kolom Product memiliki 3 nilai kosong.
2. Kolom Region memiliki 3 nilai kosong.
3. Kolom Date memiliki 2 nilai kosong.
   1. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

**Jawaban :**

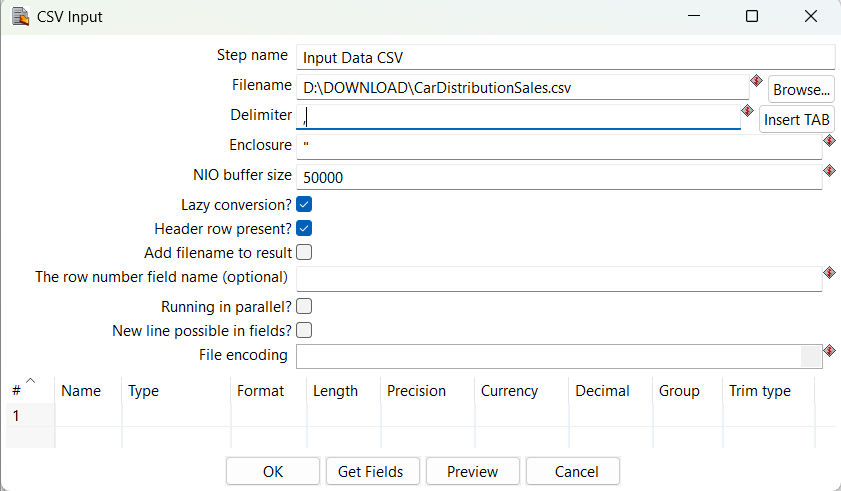
1. Kolom Date bertipe object, yang mungkin mencampur format tanggal yang berbeda atau teks non-tanggal.
2. Nama beberapa kolom memiliki spasi tambahan di awal, misalnya:
3. " Item unit price"
4. " No.Items"
5. " Total Sale"
6. Perlu dibersihkan agar lebih konsisten.
   * 1. **PENGAMBILAN DATA (EXTRACT)**
7. Buka lembar kerja Transformation melalui toolbar File.



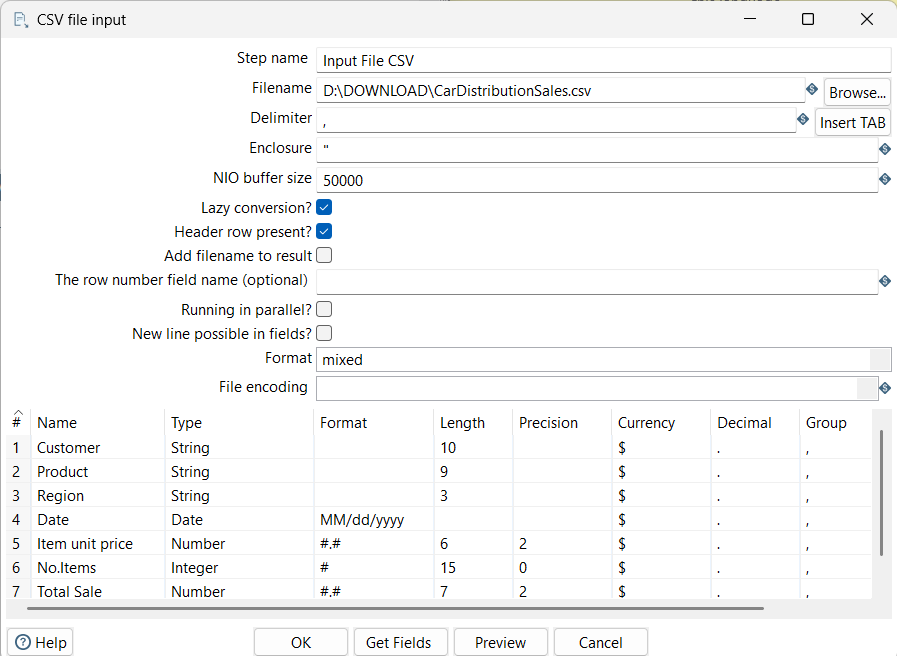
1. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.



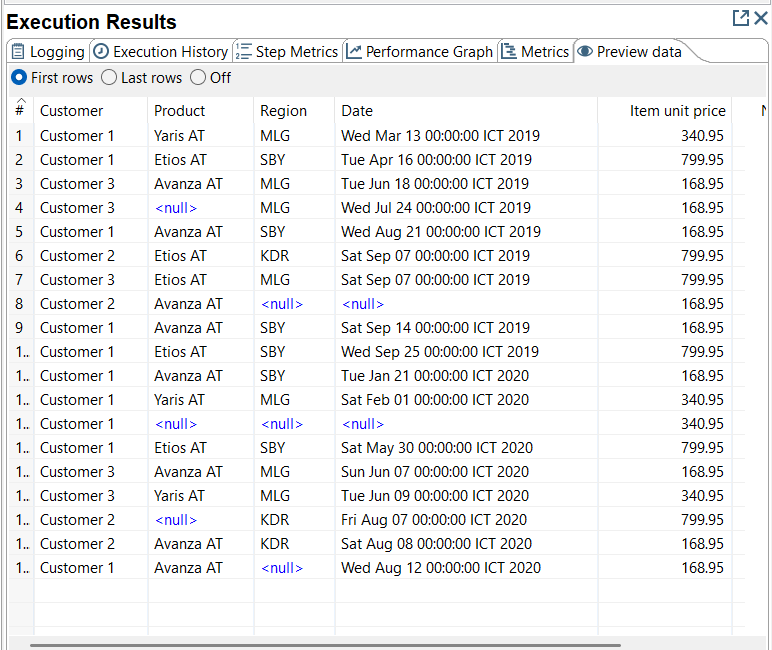
1. Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.



1. Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
2. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.
3. Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).
4. Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai

9

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.



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

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

**Jawaban :**

Iya, data hasil eksekusi sesuai dengan data aslinya

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?

**Jawaban :**

Proses ekstraksi dilakukan dengan langkah-langkah berikut di Tab Logging:

1. Initializing transformation: PDI menyiapkan proses ETL.
2. Opening CSV file: File dibuka untuk membaca isinya.
3. Parsing CSV structure: PDI membaca header kolom menggunakan Get Fields.
4. Extracting data: Data dibaca sesuai struktur.
5. Validating data: Mengecek apakah format data sesuai dengan tipe kolom.
6. Loading into memory: Data disiapkan untuk tahap selanjutnya (transformasi atau load).
7. Execution complete: Proses ekstraksi selesai.
8. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

**Jawaban :**

1. I (Input) = 20 → Jumlah total baris data yang masuk ke step ini.
2. (Output) = 0 → Tidak ada baris data yang dikirim keluar dari step ini.
3. R (Read) = 0 → Tidak ada baris data yang dibaca ulang dari sumber lain.
4. W (Written) = 19 → Sebanyak 19 baris data telah berhasil ditulis ke langkah berikutnya.
5. U (Updated) = 0 → Tidak ada data yang diperbarui.
6. E (Error) = 0 → Tidak ada error dalam proses ekstraksi.

Dapat disimpulkan bahwa proses ekstraksi data hampir bagus, dengan 19 baris data berhasil diproses dan tidak ada kesalahan (error = 0). Namun, ada 1 baris data yang tidak diproses lebih lanjut (karena Input = 20 tetapi Written = 19)

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.

**Jawaban :**

Spesifikasi laptop saya :

Processor : Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz

RAM : 8 GB

Storage : 237 GB

VGA : NVIDIA GeForce RTX 3050 Ti

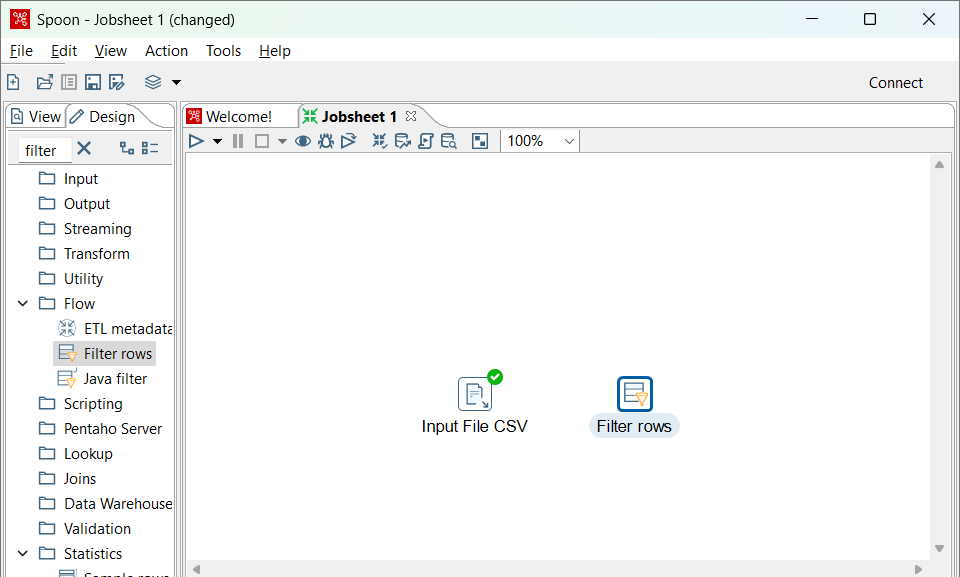
CPU : intel core i7 10th gen

Time : 0.0s

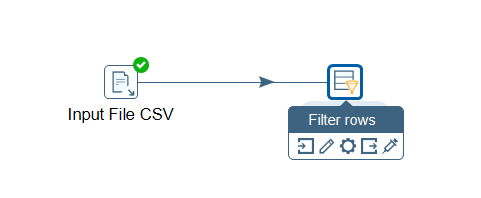
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nama | Processor | RAM | Storage | VGA | CPU | Time |
| Putera | Intel(R) Core(TM) i5-53000 CPU 2.30GHz | 16 GB | 256 GB | Intel (R) HD Graphics 5500 | Intel(R) Core(TM) i5-53000 CPU 2.30GHz | 0.1s |
| Rizky | AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz | 8 GB | 512 GB | AMD RADEON GRAPHICS | AMD RYZEN 3 7320U with Radeon Graphics | 0.1s |
| Daffa | AMD A9 9425 RADEON R5 | 8 GB | 1.2 TB | AMD RADEON (TM) R5 Graphics | AMD A9 9425 RADEON R5 | 0.3s |
| Dhevina | 12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz | 8 GB | 152 GB | Inter(R) UHD Graphics | 12th Gen Intel(R) Core(TM) i5 - 1235U | 0.0s |

* + 1. **FILTER DATA (TRANSFORM) DAN PENGEMASAN DATA (LOAD)**

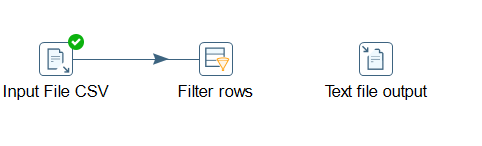
1. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area.



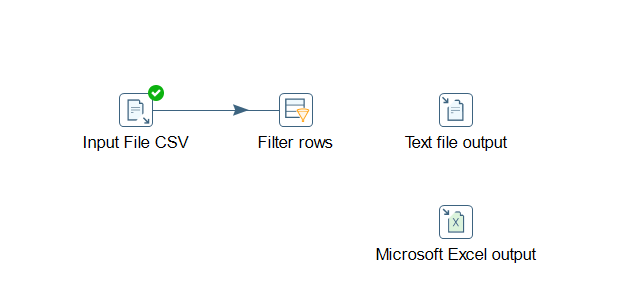
1. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.



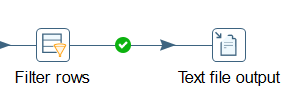
1. Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows.
2. Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.



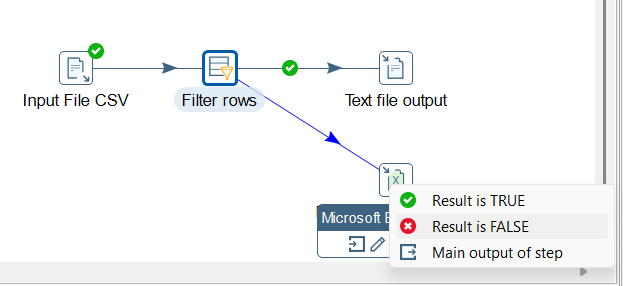
1. Carilah objek Microsoft excel output pada Design Area. Drag and drop objek tersebut pada Work Area.



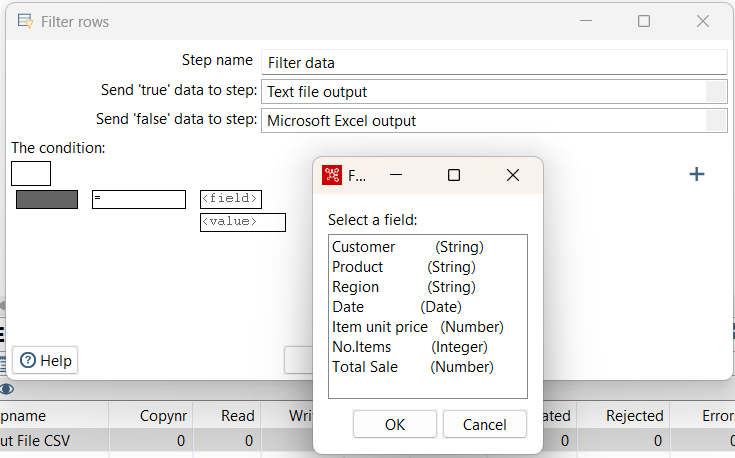
1. Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagi jenis konektornya.



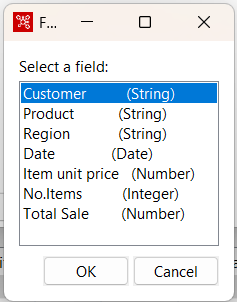
1. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.



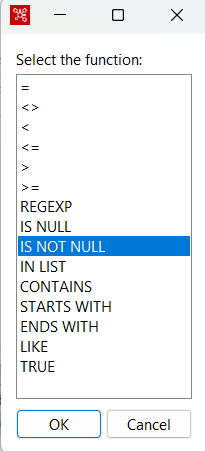
1. Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.
2. Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
3. Klik <Field> pada bagian condition



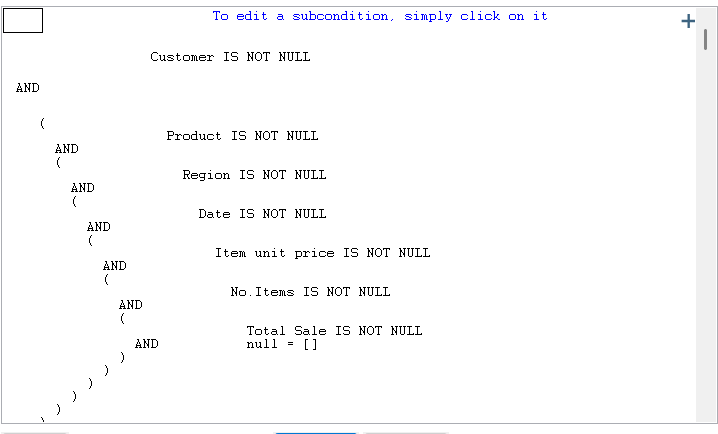
1. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.



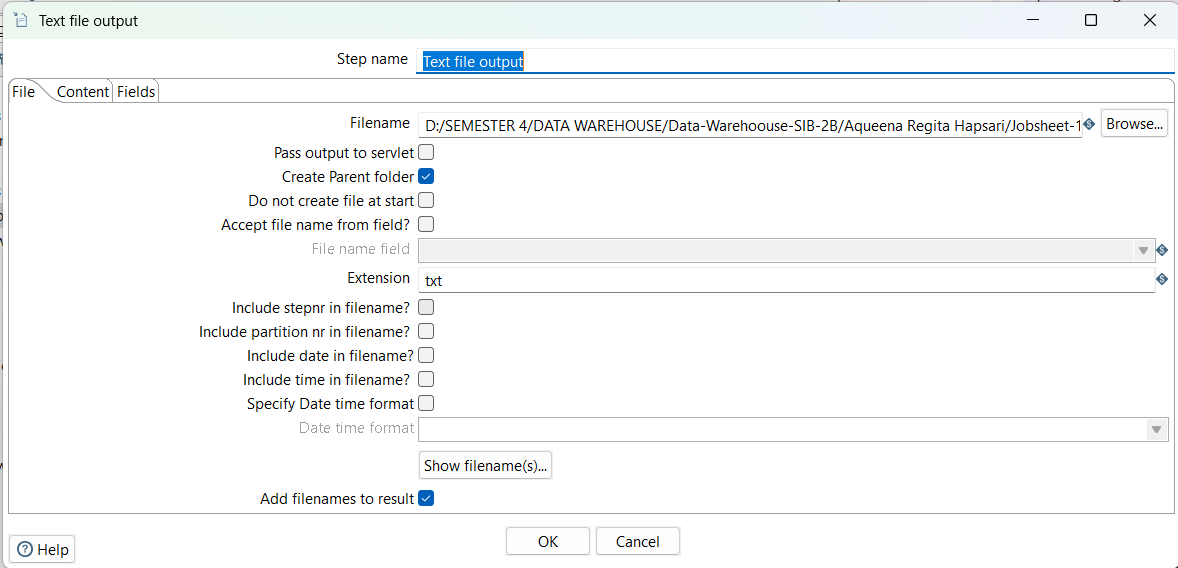
1. Pada bagian value pilih IS NOT NULL kemudian tekan OK , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



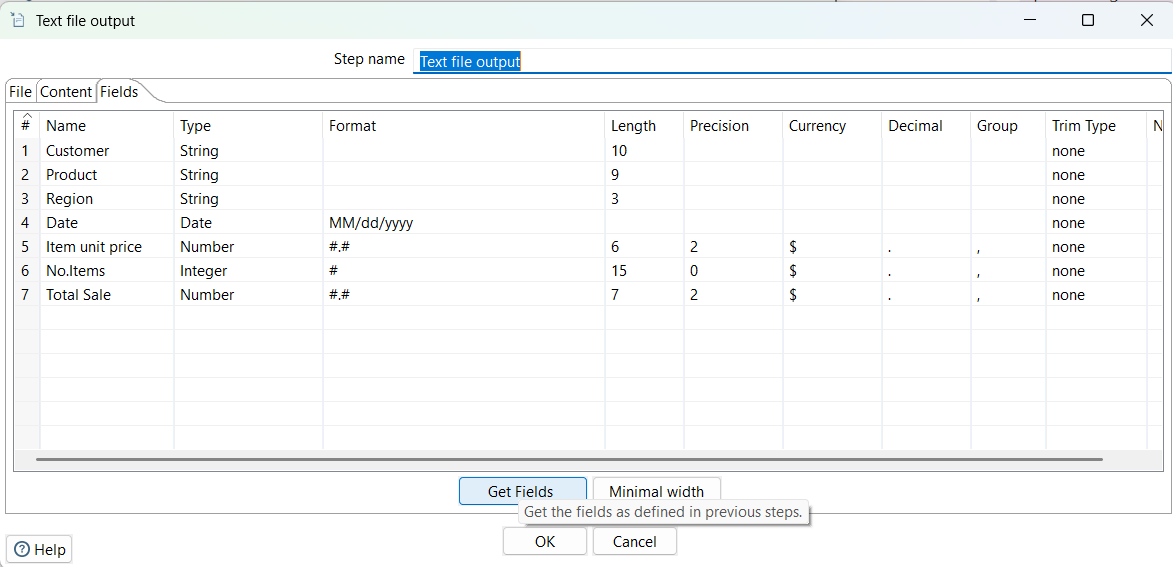
1. Tekan tombol + untuk menambah kondisi.
2. Pilih statement yang masih null untuk menambah kondisi.
3. Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
4. Gunakan operator AND untuk menghubung



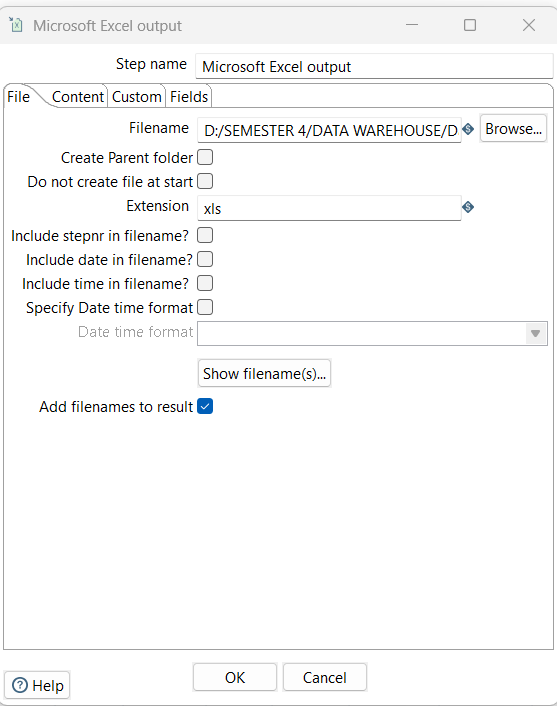
1. Jika semua field sudah masuk dalam kondisi maka tekan OK.
2. Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.
3. Ubah extension menjadi csv.



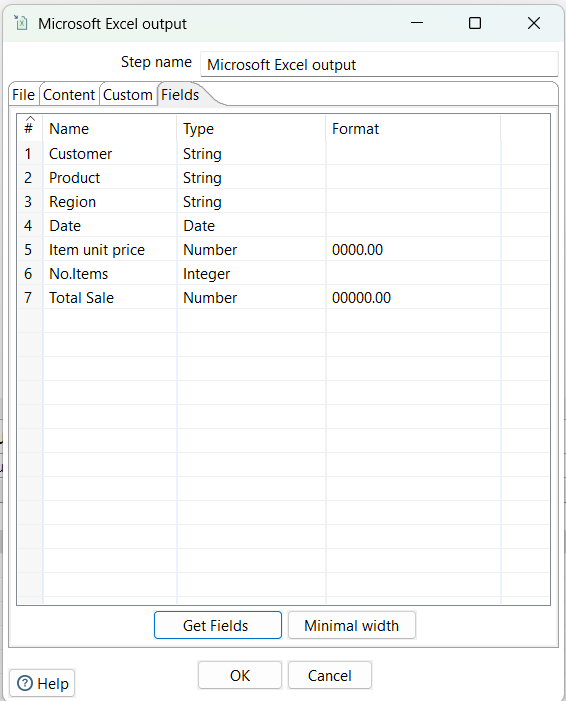
1. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.



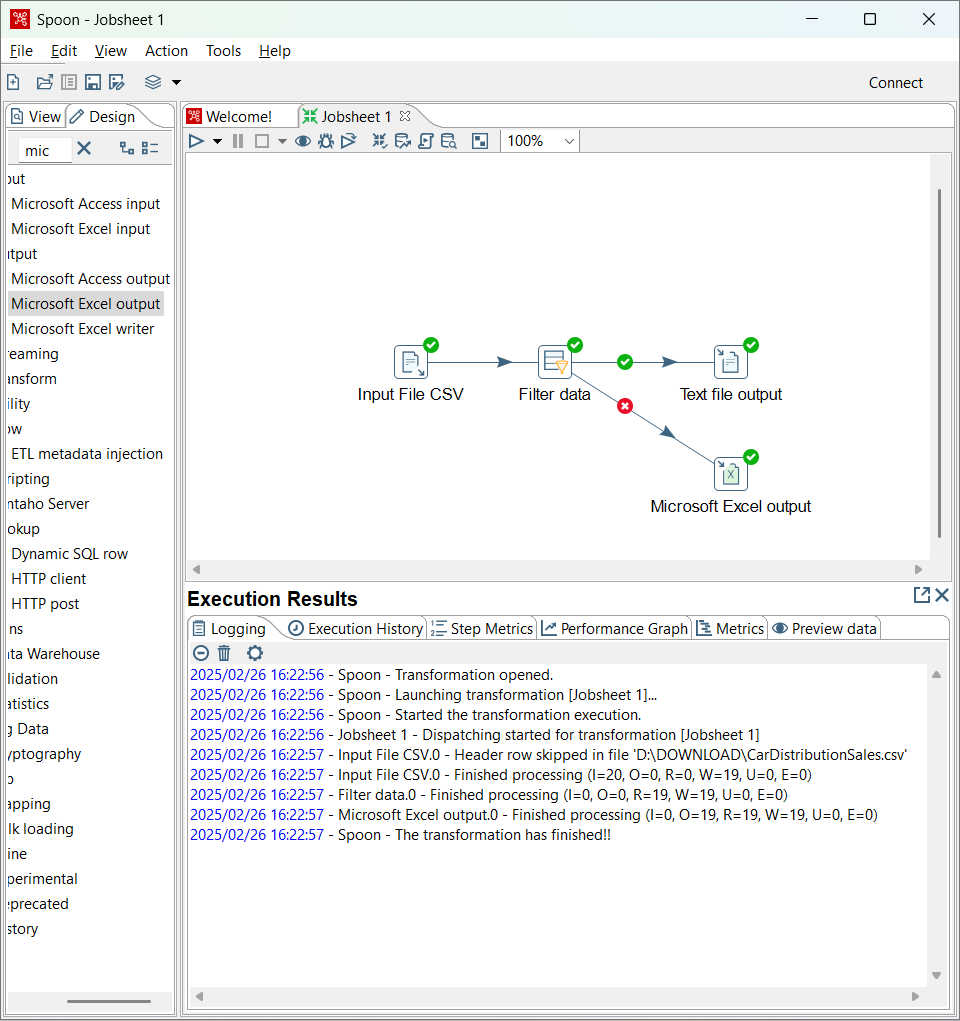
1. Tekan ok untuk kembali ke work area
2. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
3. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution.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



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



1. Tekan ok untuk kembali ke work area.
2. 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?

**Jawaban :**

1. CSV: Formatnya sederhana, hanya berisi teks dengan pemisah koma, tanpa format khusus.
2. Excel: Bisa menyimpan format tabel, warna, formula, dan berbagai jenis data yang lebih kompleks.

Perbedaan utama terjadi pada tampilan, pemformatan angka, atau encoding karakternyaa.

1. Jelaskan apa yang terjadi pada proses Filter rows!

**Jawaban :**

Yang terjadi pada proses filter rows diatas adalah :

* + - 1. Input data 🡪 data dari file di masukkan ke dalam PDI
      2. Menentukan kondisi penyaringan 🡪 filter rows digunakan untuk menentukan aturan atau kondisi yang harus dipenuhi oleh data agar bisa lolos ke tahap berikutnya
      3. Pemrosesan data 🡪 data yang sudah memenuhi kriteria akanlanjut ke langkah berikutnya dalam transformasi ETL
      4. Output data 🡪 hasil dari filter rows menghasilkan subset data yang lebih bersih dan relevan untuk dianalisis lanjutan.

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

**Jawaban :**

Spesifikasi laptop saya :

Processor : Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz

RAM : 8 GB

Storage : 237 GB

VGA : NVIDIA GeForce RTX 3050 Ti

CPU : intel core i7 10th gen

Time : 0.3s

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nama | Processor | RAM | Storage | VGA | CPU | Time |
| Putera | Intel(R) Core(TM) i5-53000 CPU 2.30GHz | 16 GB | 256 GB | Intel (R) HD Graphics 5500 | Intel(R) Core(TM) i5-53000 CPU 2.30GHz | 0.7s |
| Rizky | AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz | 8 GB | 512 GB | AMD RADEON GRAPHICS | AMD RYZEN 3 7320U with Radeon Graphics | 0.2s |
| Daffa | AMD A9 9425 RADEON R5 | 8 GB | 1.2 TB | AMD RADEON (TM) R5 Graphics | AMD A9 9425 RADEON R5 | 0.12s |
| Dhevina | 12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz | 8 GB | 152 GB | Inter(R) UHD Graphics | 12th Gen Intel(R) Core(TM) i5 - 1235U | 0.6s |

**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.

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

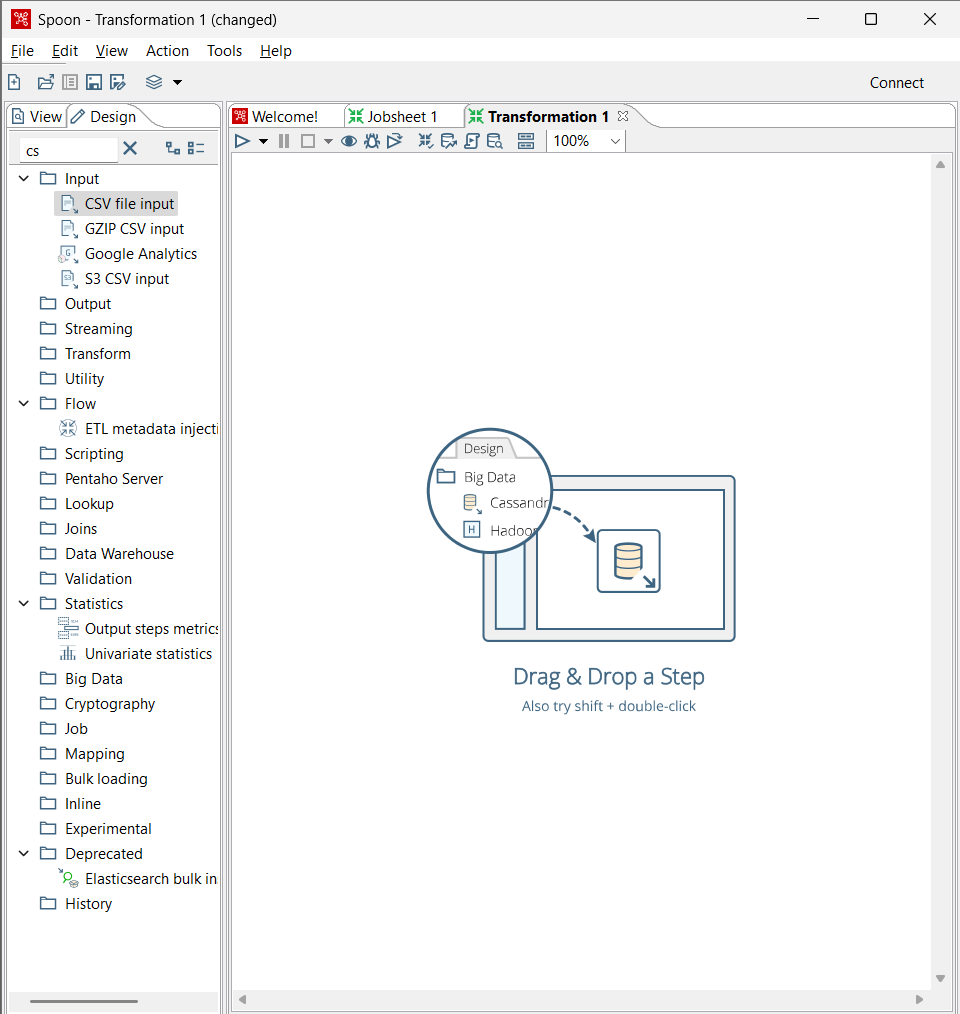
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.

**JAWABAN**

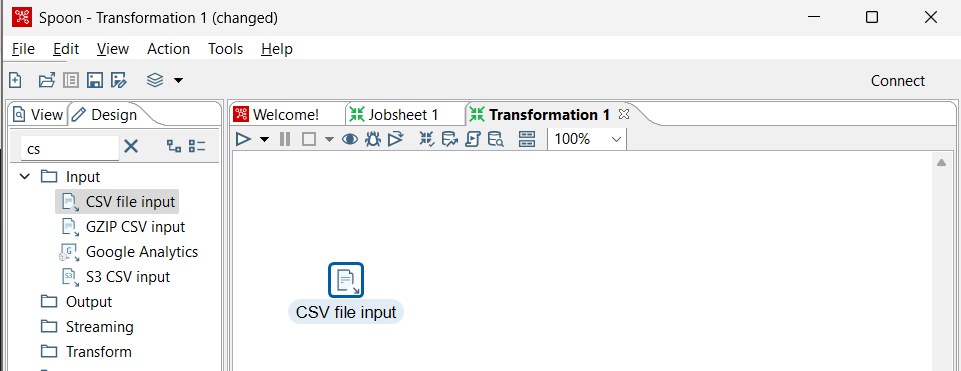
1. **Buka Spoon dan Buat Transformasi Baru**

Pertama, jalankan Spoon.bat dan tunggu sampai interface-nya terbuka. Setelah itu, buat transformasi baru dengan **klik File 🡪 New 🡪 Transformation**. Ini tempat jadi workspace utama kita buat proses data.

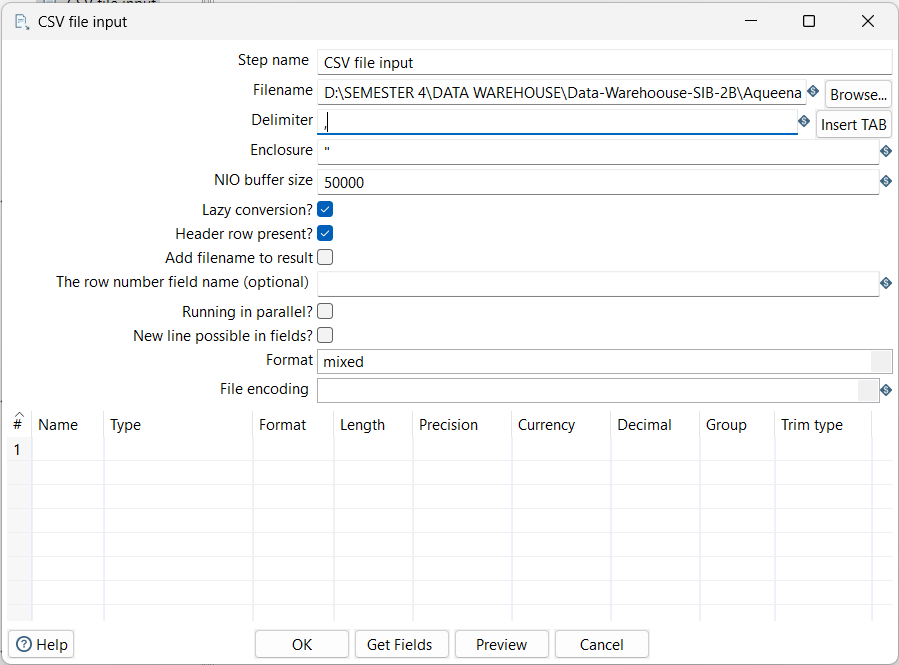


1. **Tambahkan Input Data dari CSV**

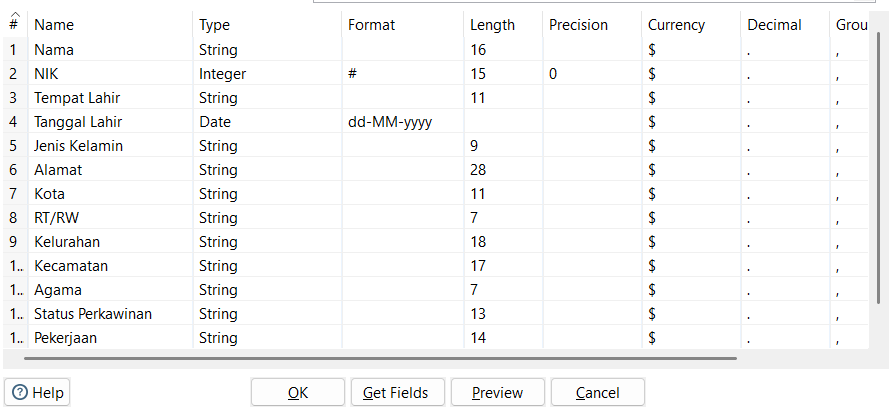
Sekarang, cari komponen CSV file input di Design Area, lalu tarik ke Work Area. Klik dua kali untuk membuka pengaturannya. Pilih file dataDummyKependudukan.csv, pastikan delimiter-nya koma (,), dan centang opsi Header supaya baris pertama dikenali sebagai nama kolom. Setelah itu, klik Get Fields supaya Spoon otomatis membaca struktur data. Kalau semua sudah sesuai, klik OK.



Gambar 1 CSV file input



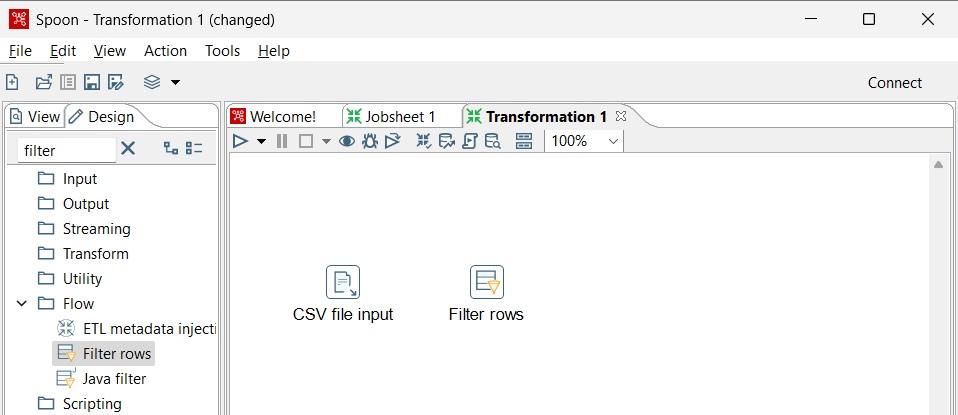
Gambar 2 Konfigurasi CSV file input



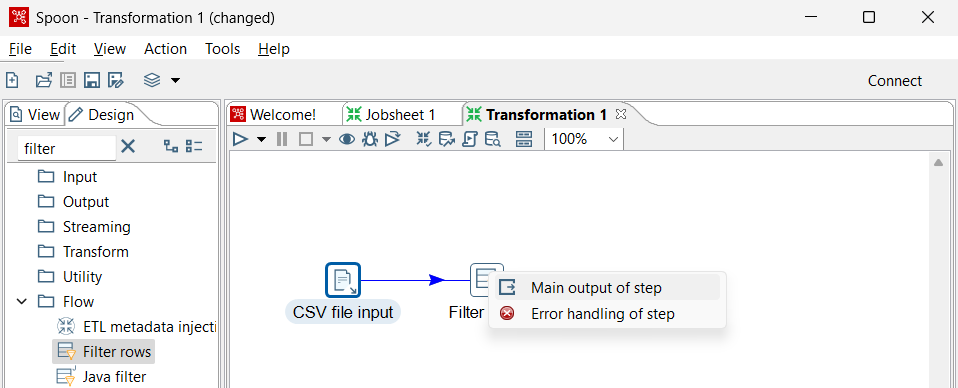
Gambar 3 Hasil

1. **Filter Data Berdasarkan Profesi**

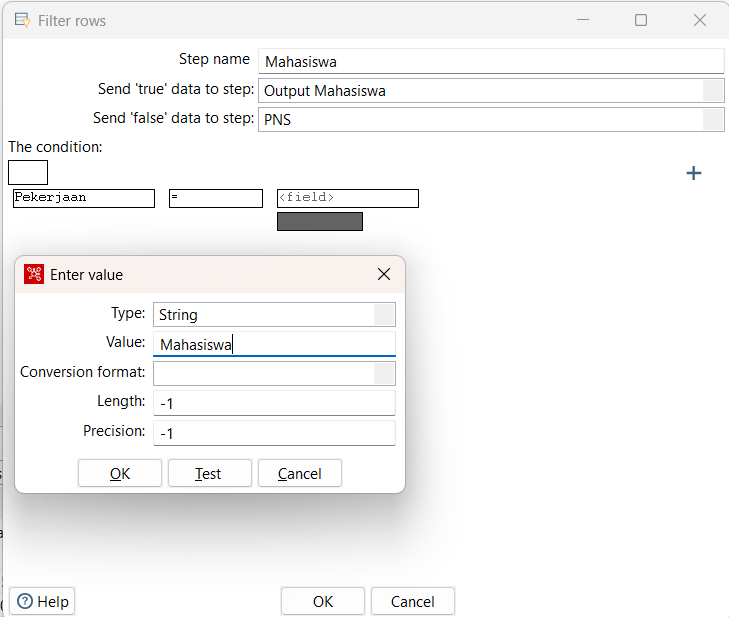
Setelah data terbaca, kita perlu memisahkannya berdasarkan profesi. Tarik Filter Rows dari panel Flow, lalu hubungkan dengan CSV file input. Klik dua kali Filter Rows, lalu atur kondisinya menjadi Profesi = "Mahasiswa". Pada bagian Send 'true' data to step, beri nama Mahasiswa Output, dan untuk Send 'false' data to step, beri nama Non-Mahasiswa Filter. Klik OK untuk menyimpan pengaturan.



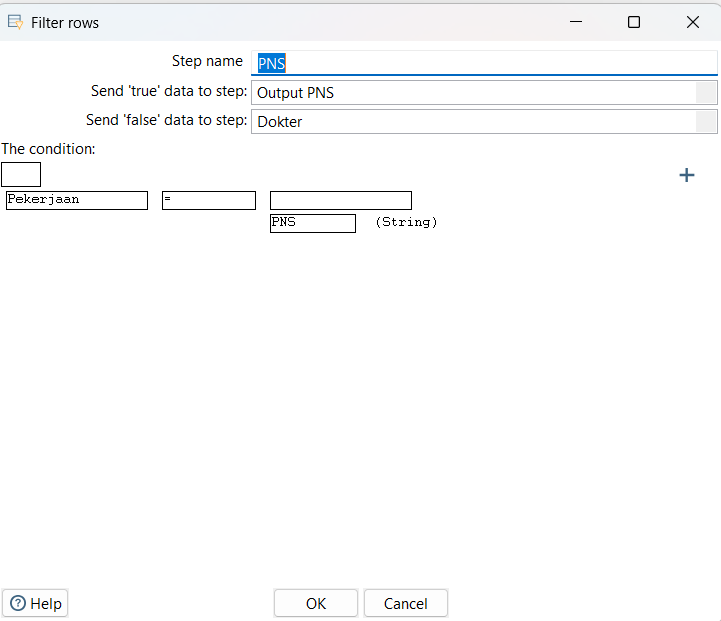
Gambar 4 Tambah Filter rows ke work space



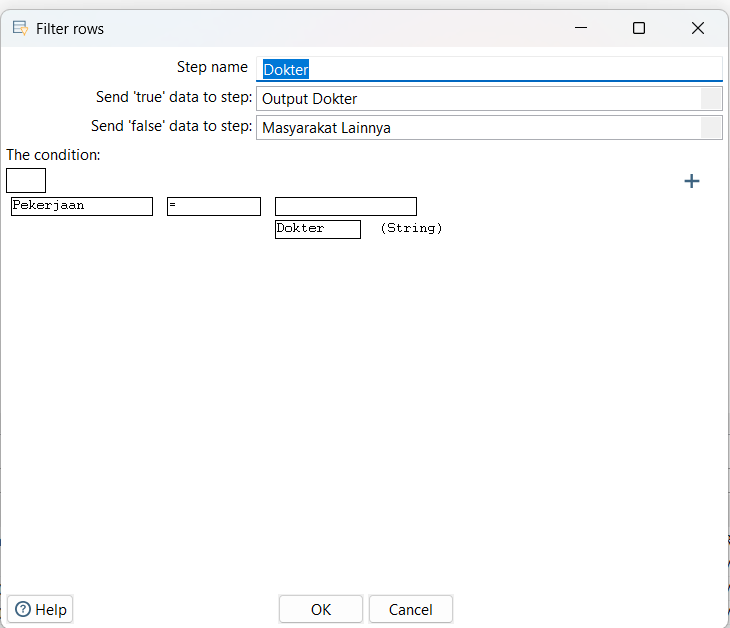
Gambar 5 Hubungkan output dari csv file input ke filter rows



Gambar 6 Filter Pekerjaan equals to "mahasiswa"



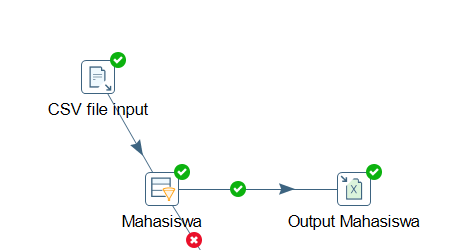
Gambar 7 Filter PNS



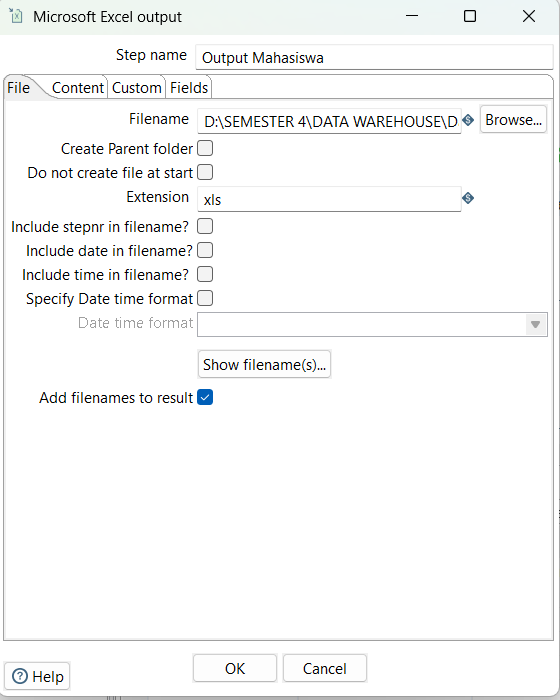
Gambar 8 Filter Dokter

1. **Simpan Data Mahasiswa ke File Terpisah**

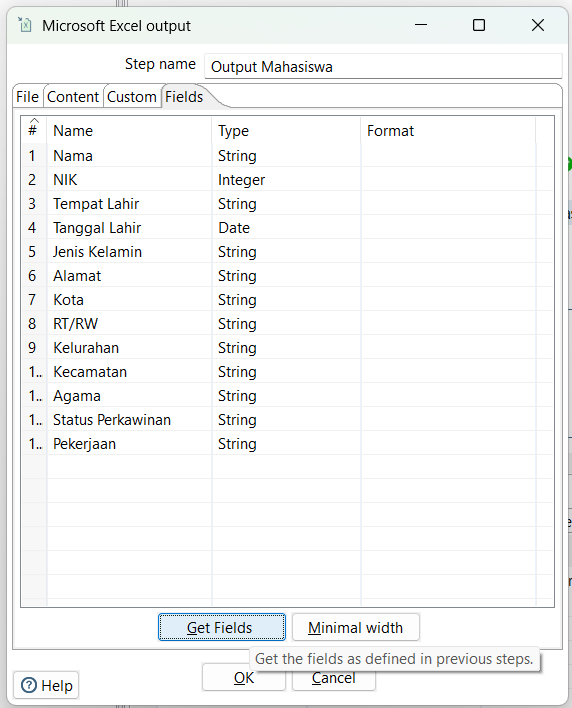
Sekarang, tarik Text file output dari Output, lalu hubungkan ke Mahasiswa Output. Klik dua kali untuk mengatur output file. Beri nama file MasyarakatMahasiswa.csv, pilih format CSV, lalu klik Get Fields supaya Spoon mengambil struktur data secara otomatis. Klik OK untuk menyimpan pengaturan ini.



Gambar 9 Drop Ms.Excel Output kemudian set hop information true



Gambar 10 Set filename untuk output



Gambar 11 Get Fields

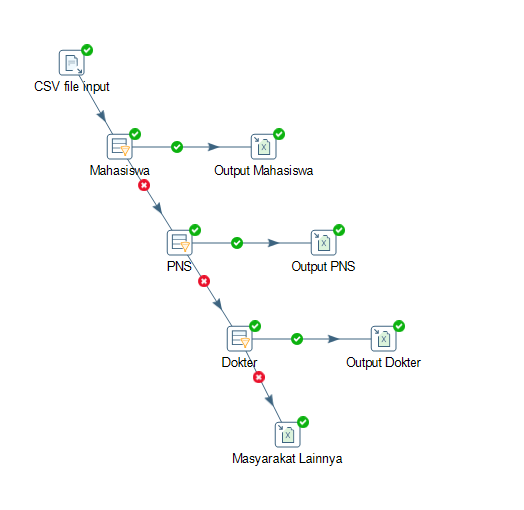
1. **Pisahkan Data PNS, Dokter, dan Lainnya**

Langkah ini mirip seperti sebelumnya, hanya saja kali ini kita memproses data yang bukan mahasiswa. Tarik lagi Filter Rows, hubungkan ke Non-Mahasiswa Filter, lalu atur kondisinya menjadi Profesi = "PNS". Beri nama output PNS Output untuk data yang cocok, dan Non-PNS Filter untuk data lainnya.

Lanjutkan dengan langkah yang sama untuk memisahkan data Dokter dengan kondisi Profesi = "Dokter", output Dokter Output, dan sisa data lainnya sebagai Lainnya Output.

1. **Simpan Data PNS, Dokter, dan Lainnya ke File Terpisah**

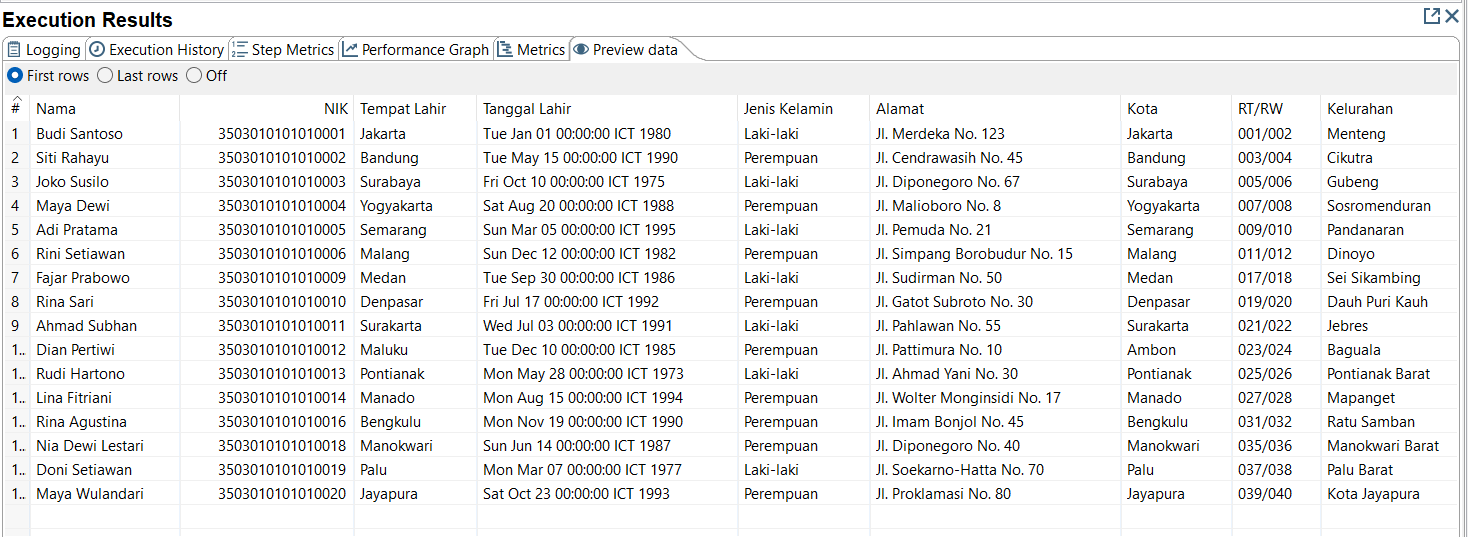
Tambahkan Text file output untuk masing-masing kelompok. Hubungkan PNS Output ke file MasyarakatPNS.csv, Dokter Output ke MasyarakatDokter.csv, dan Lainnya Output ke MasyarakatLainnya.csv. Jangan lupa klik Get Fields untuk setiap output supaya formatnya sesuai dengan input.



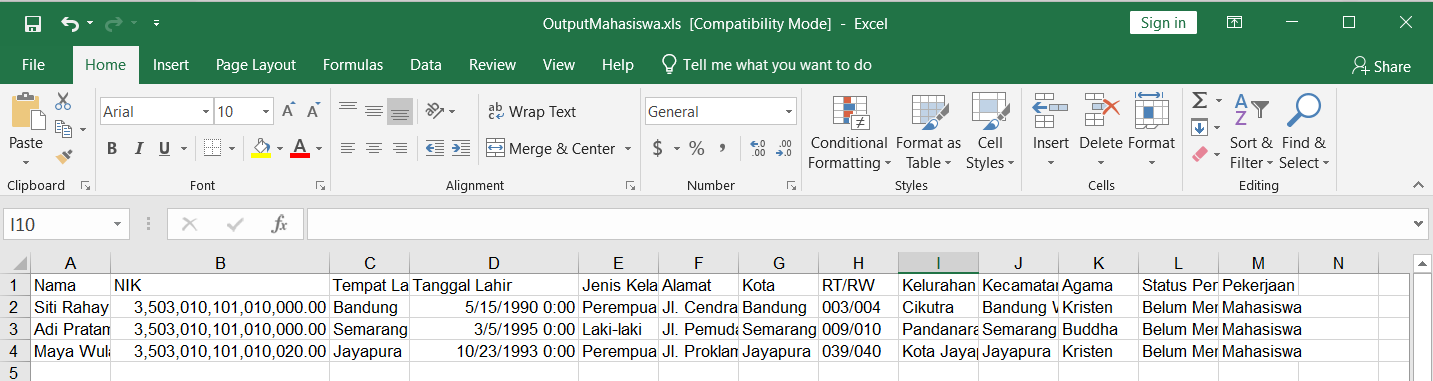
Gambar 12 Struktur Algoritma

1. **Jalankan Transformasi dan Cek Hasilnya**

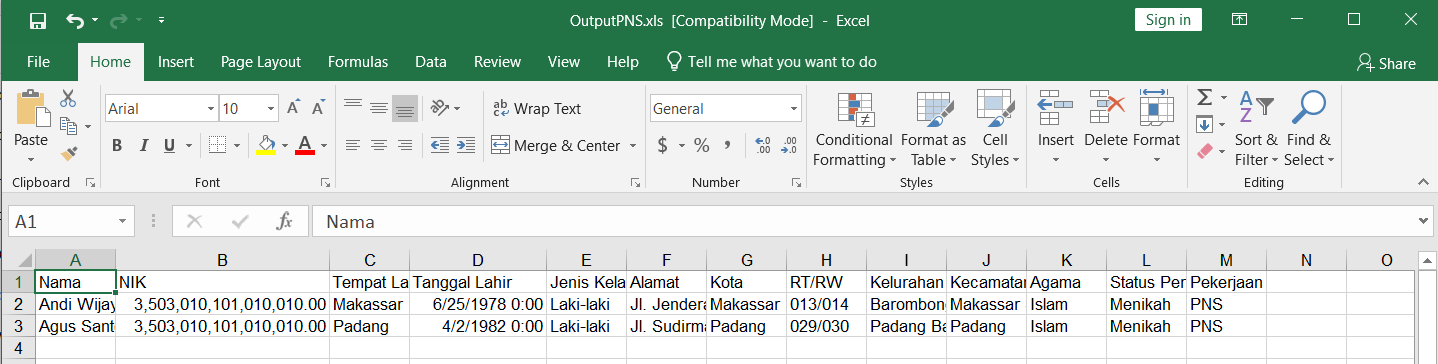
Setelah semuanya siap, simpan transformasi ini dengan klik File > Save As, beri nama misalnya PemisahanDataPenduduk.ktr, lalu klik Run (ikon hijau) untuk mengeksekusi prosesnya. Kalau semuanya benar, Spoon akan memproses data dan menyimpan hasilnya dalam empat file CSV yang berbeda sesuai dengan profesi masing-masing.



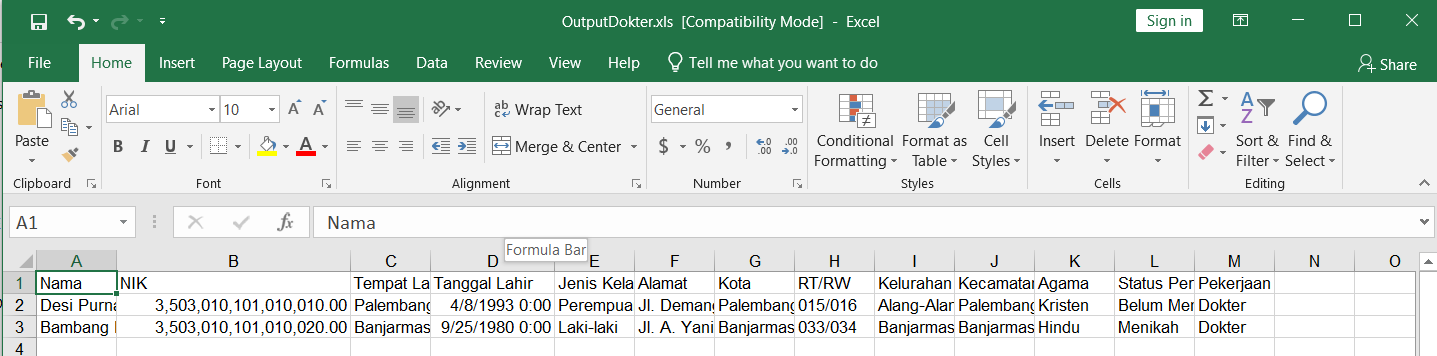
Gambar 13 Execution Result



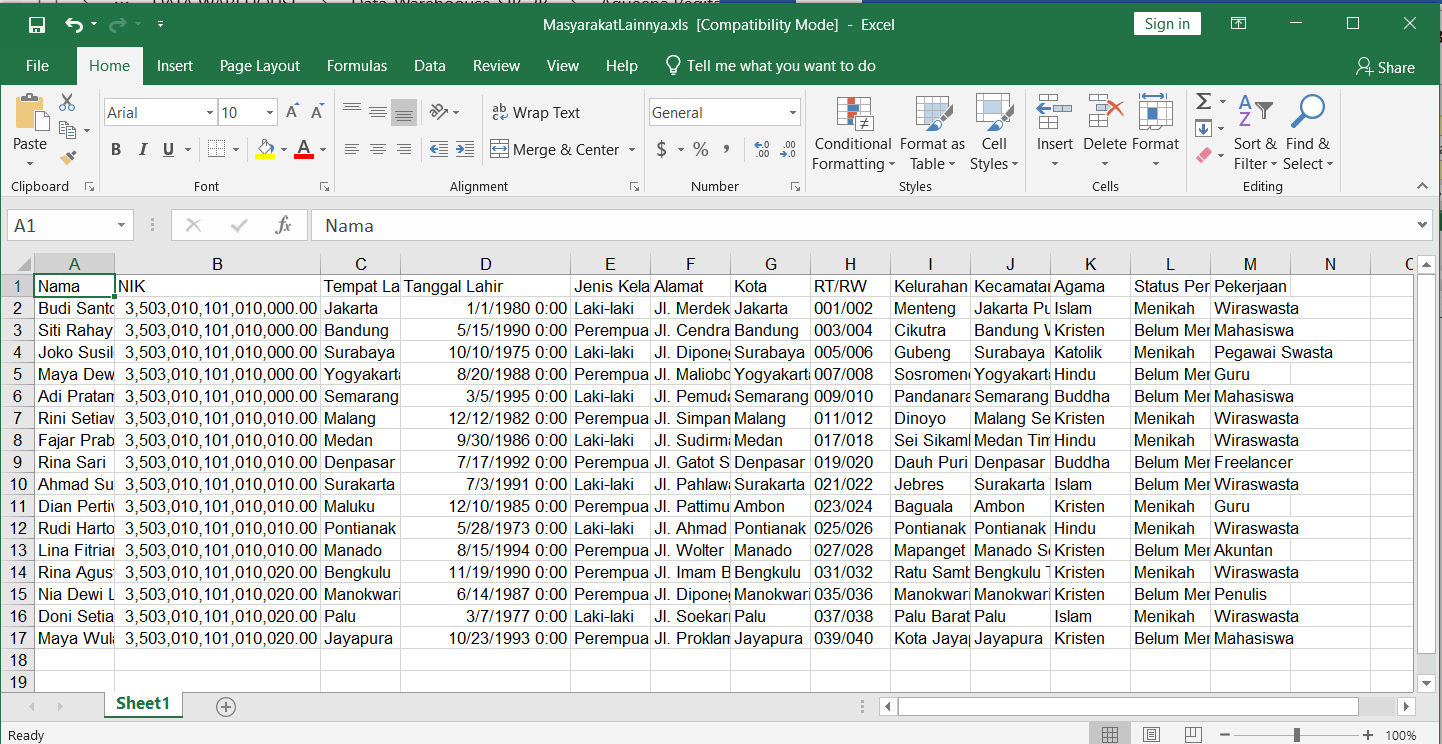
Gambar 14 Hasil Output Mahasiswa



Gambar 15 Hsil Output PNS



Gambar 16 Hasil Output Dokter



Gambar 17 Hasil Output Lainnya

Jadi, inti dari pengerjaan studi kasus ini, tentukan algoritma proses filter terlebih dahulu. Disini saya menggunakan algoritma if-else untuk proses filter data. Jadi raw data akan difilter berdasarkan profesi, if (pekerjaan != mahasiswa) maka diteruskan ke PNS, if (pekerjaan != PNS) maka di teruskan ke Dokter, if (pekerjaan != dokter) maka diteruskan ke lainnya (else if). jika kondisinya adalah if (pekerjaan = mahasiswa) maka diteruskan ke output mahasiswa (xls), begitu dan seterusnya hingga ke dokter).