**LAPORAN TUGAS DATA MINING**

**PERTEMUAN 2**

Dosen Pengampu : Vipkas Al Hadid Firdaus, S.T., M.T.



Muhammad Afiq Firdaus / 14

SIB-2B / 2341760189

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

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**2025**

**Jobsheet1** Data Warehouse sederhana dan pengenalan PDI Spoon

**Tujuan Praktikum**

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

**Studi Kasus**

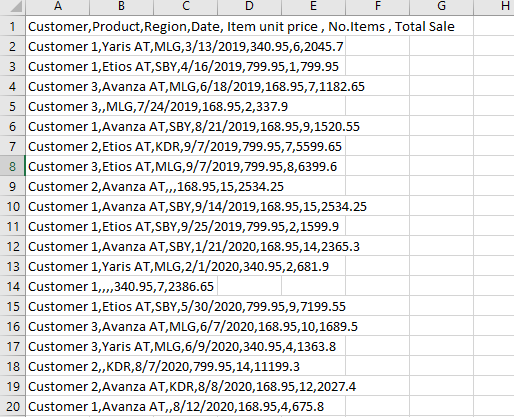
Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan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

Tugas 1

Analisa lah data tersebut!



1. Berapa jumlah kolom pada data tersebut?

**Jawaban :** 7 kolom

1. Apa arti atau isi data dari setiap kolom yang ada?

**Jawaban :**

Kolom 1 → Customer → Nomor Customer

Kolom 2 → Product → Nama Produk Mobil

Kolom 3 → Region → Nama Daerah

Kolom 4 → Date → Tanggal

Kolom 5 → Item Unit Price → Harga per unit

Kolom 6 → No Item → Nomor Item

Kolom 7 → Total Sale → Total

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

**Jawaban :**

Ada pada

• Line 5 kolom Product

• Line 9 kolom Region, Date

• Line 14 kolom Product, Region, Date

• Line 18 kolom Product,

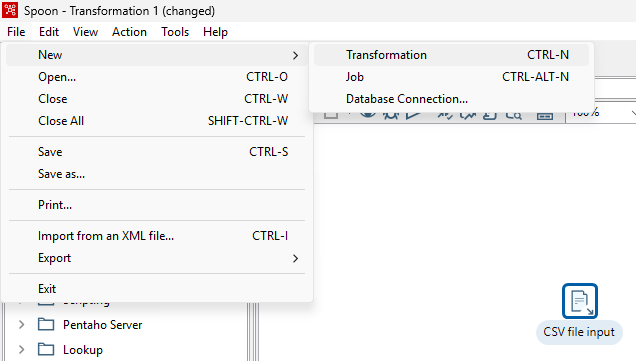
• Line 20 kolom Region

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

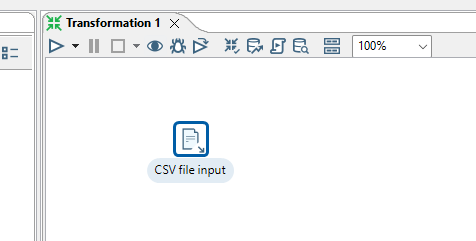
**Jawaban :**

Dalam satu kolom setiap data memiliki tipe data yang sama dan memiliki data yang kembar.

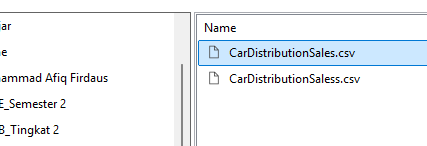
1. **Pengambilan Data (Extract)**
2. Buka lembar kerja **Transforamtion** melalui toolbar **File**



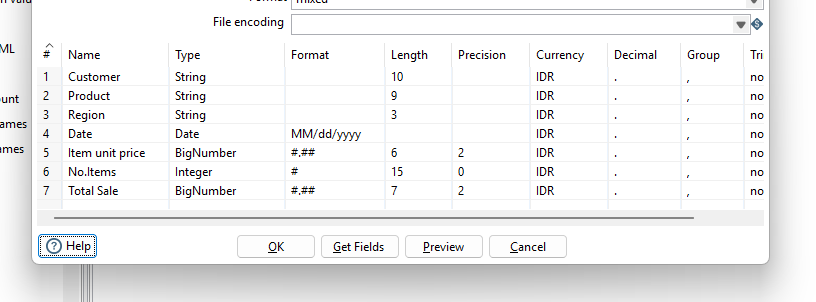
1. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.



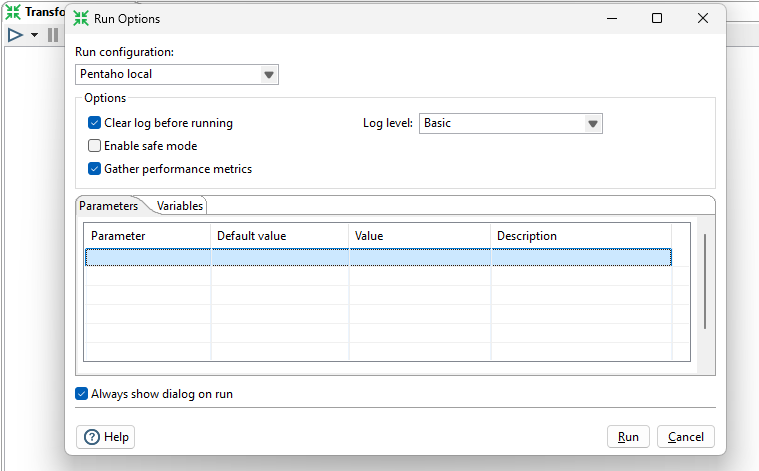
1. Memasukkan File CarDistributionSales.csv



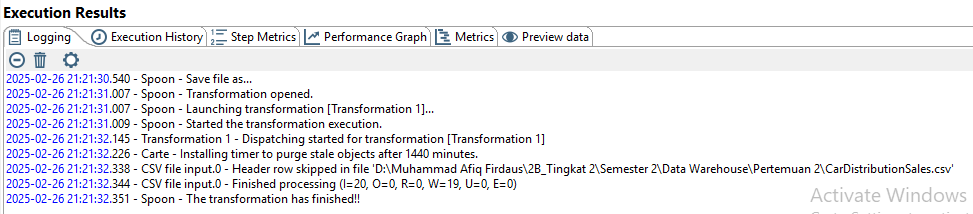
1. Klik Get Fields dan muncul nama nama kolom yang ada pada file csv tersebut.



1. Run file



1. Hasil Execute berikut



Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

**Jawaban :**

Sudah sesuai dengan file aslinya, data yang NULL tidak kosong tetapi terdapat keterangan NULL

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 :**

|  |
| --- |
| 2025-02-26 21:21:30.540 - Spoon - Save file as...  2025-02-26 21:21:31.007 - Spoon - Transformation opened.  2025-02-26 21:21:31.007 - Spoon - Launching transformation [Transformation 1]...  2025-02-26 21:21:31.009 - Spoon - Started the transformation execution.  2025-02-26 21:21:32.145 - Transformation 1 - Dispatching started for transformation [Transformation 1]  2025-02-26 21:21:32.226 - Carte - Installing timer to purge stale objects after 1440 minutes.  2025-02-26 21:21:32.338 - CSV file input.0 - Header row skipped in file 'D:\Muhammad Afiq Firdaus\2B\_Tingkat 2\Semester 2\Data Warehouse\Pertemuan 2\CarDistributionSales.csv'  2025-02-26 21:21:32.344 - CSV file input.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)  2025-02-26 21:21:32.351 - Spoon - The transformation has finished!! |

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

**Jawaban :**

I=20 → Input 20

O=0 → Output 0

R=0 → Read 0

W=19 → Written 19

U=0 → Updated 0

E=0 → Errors 0

1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini?

(Processor, RAM, Storage, VGA, CPU).

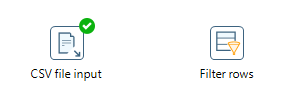
**Jawaban :**

|  |
| --- |
|  |
|  |
|  |

1. **Filter Data (Transform) dan pengemasa data (Load)**

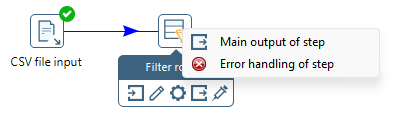
**Praktikum**

1. Masukkan Filter Rows kedalam file

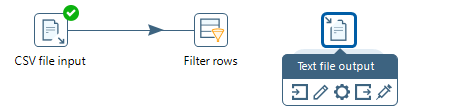


1. Sambungkan tombol output pada Input Data CSV ke Filter rows kemudian klik Main

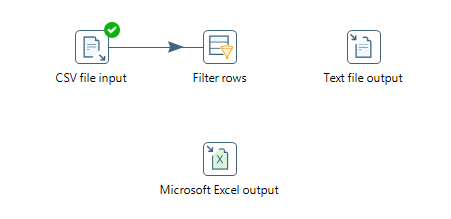
output of step



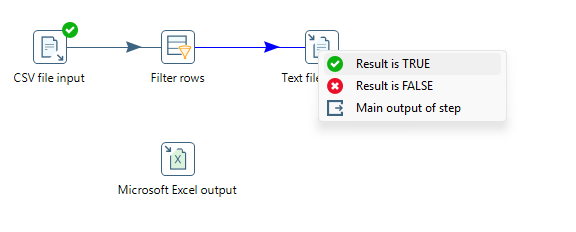
1. Tambahkan Text File Output



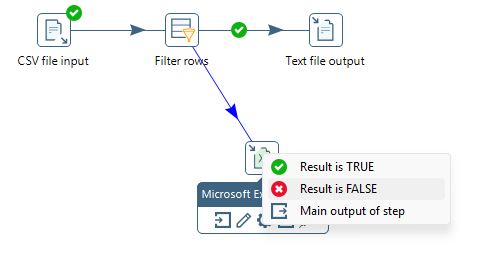
1. Tambahkan juga Microsoft Excel Output



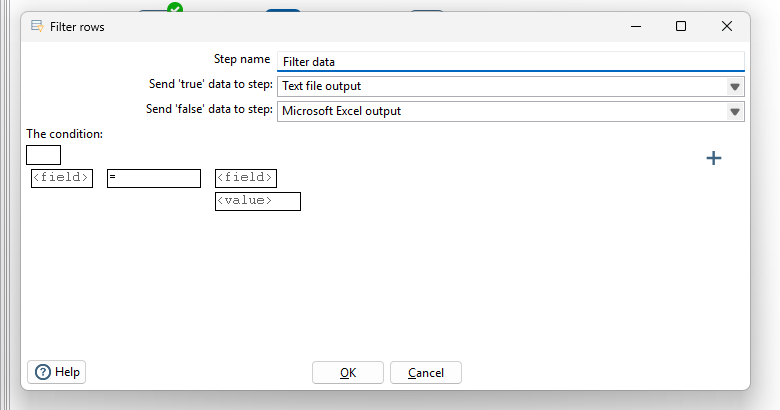
1. Sambungkan Output dari Filter Rows menuju Text File Output lalu klik TRUE



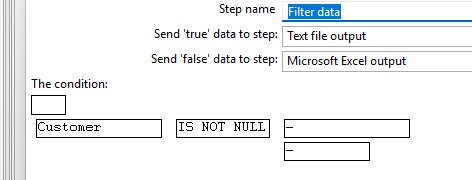
1. Sambungkan Output dari Filter Rows menuju Microsoft Excel Output lalu klik FALSE



1. Klik 2x pada Text File Output lalu ubah nama step name dan pastikan send true dan send false seperti gambar di bawah.

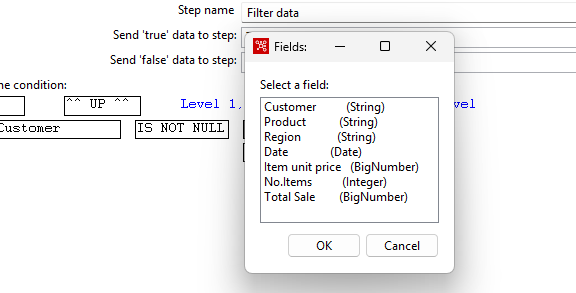


Klik <field> lalu ubah menjadi Customer kemudian kotak disebelah Customer diisi IS NOT NULL

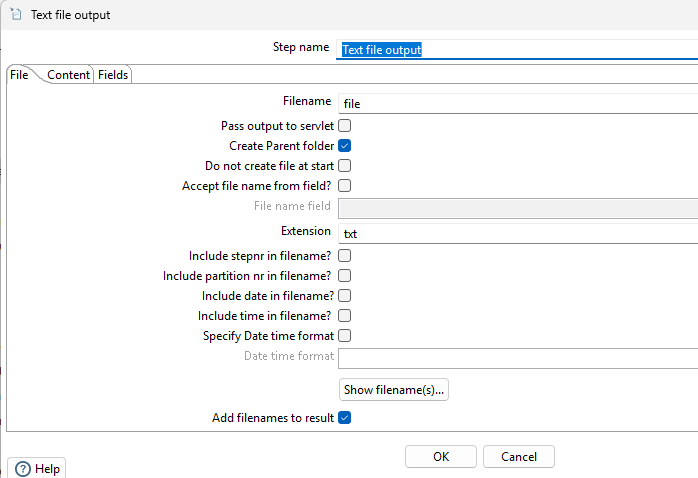


1. Klik tombol Plus untuk membuat kondisi baru, tampilan akan seperti dibawah ini. Lakukan seperti Langkah sebelumnya akan tetapi <field> diisi berurutan ke urutan selanjutnya setelah Customer.

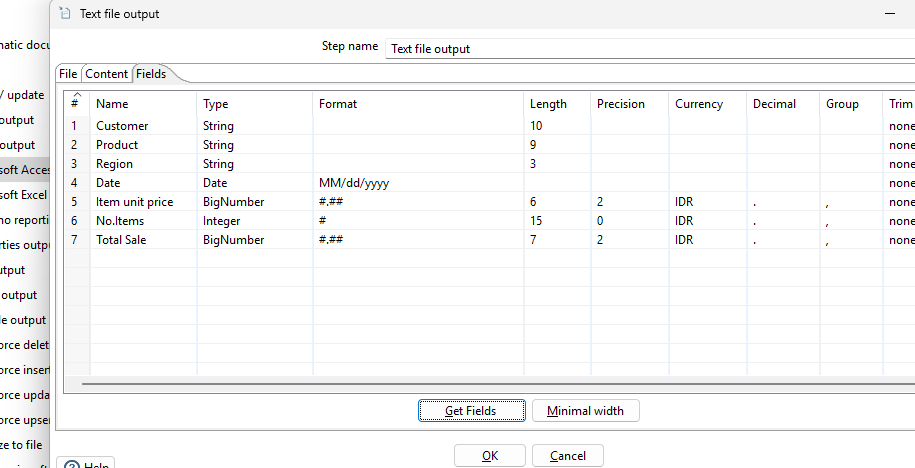
Ketika di klik OK akan terbentuk kode seperti berikut yang disambungkan dengan operator AND.



1. Klik Text File Output 2x, lalu ubah File Name seperti berikut.

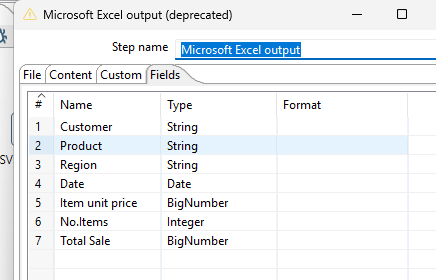


1. Klik bagian sheet Fields, lalu klik Get Fields.

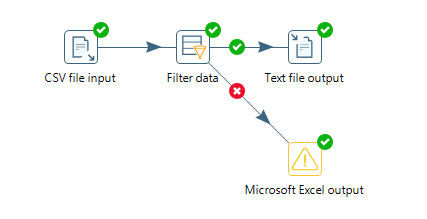


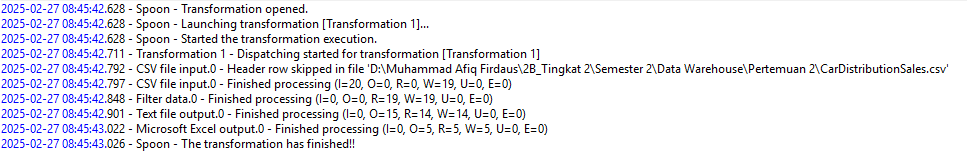
1. Klik Microsoft Excel Output kemudian ubah seperti gambar dibawah ini dan klik Get

Fields.

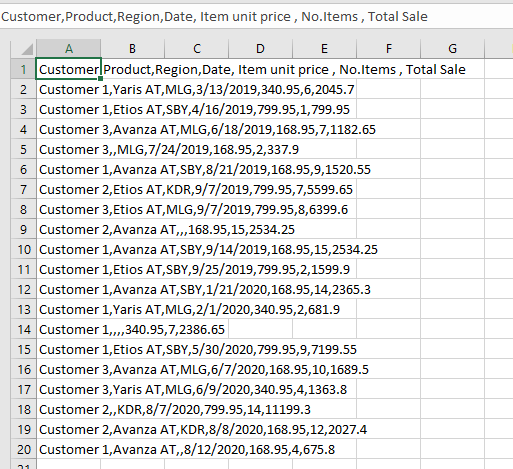


1. Klik RUN. Kemudian akan tergambar cetang hijau semua.

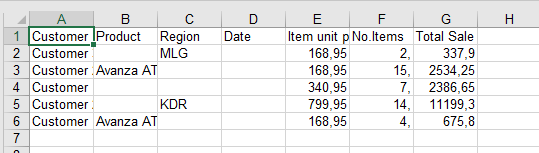




1. File outputSalesCarDistribution.csv



1. Isi file failedSalesCarDistribution.xls



**Tugas 3**

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

**Jawaban :**

* + outputSalesCarDistribution.csv

Berisi hasil dari kondisi yang benar yaitu tidak ada satupun cell yang bernilai null

* + failedSalesCarDistribution.xls

Berisi hasil dari kondisi yang salah yaitu semua row memiliki nilai null didalamnya

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

**Jawaban :**

1. Filter rows disambungkan dengan Input CSV yang bernilai file raw asli dari

CarDistribution. File rows berada di Tengah input dan output sebagai alat untuk

melakukan filterisasi hasil yang diinginkan

1. Setelah data dari input dikirimkan ke Filter Rows, filter rows melakukan tugasnya

yaitu melakukan Filterisasi Row sesuai kondisi yang di masukkan. Pada kasus

diatas kita membuat kondisi yaitu, Kondisi benilai true adalah semua row tanpa

nilai NULL(terisi semua). Sedangkan kondisi false adalah semua row yang

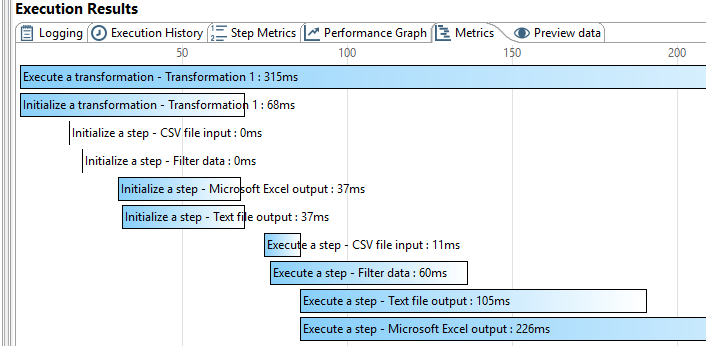
memiliki nilai NULL di dalamnya

1. Setelah melakukan proses filter, Filter row menyajikan hasil filterisasinya menjadi

2 file TRUE and FALSE

1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data

ini? identifikasi pula spesifikasi perangkat yang anda gunakan (Processor, RAM, Storage, VGA, CPU). bandingkan dalam bentuk table



**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.

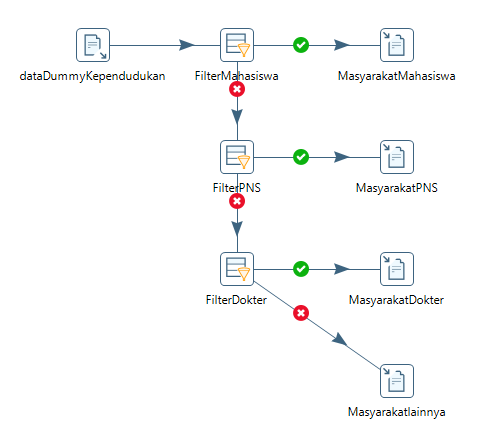
Sebagai contoh, mahasiswa akan dikumpulkan dan berdomisili di wilayah edukasi yang akses dekat dengan kampus-kampus perguruan tinggi. Sedangkan dokter akan berdomisili dekat dengan rumah sakit atau layanan kesehatan dan PNS akan berdomisili dekat dengan perkantoran dan layanan publik. Selain itu, orang yang bukan dari ketiga kategori tersebut bebas memilih domisili atau tempat tinggalnya di IKN.

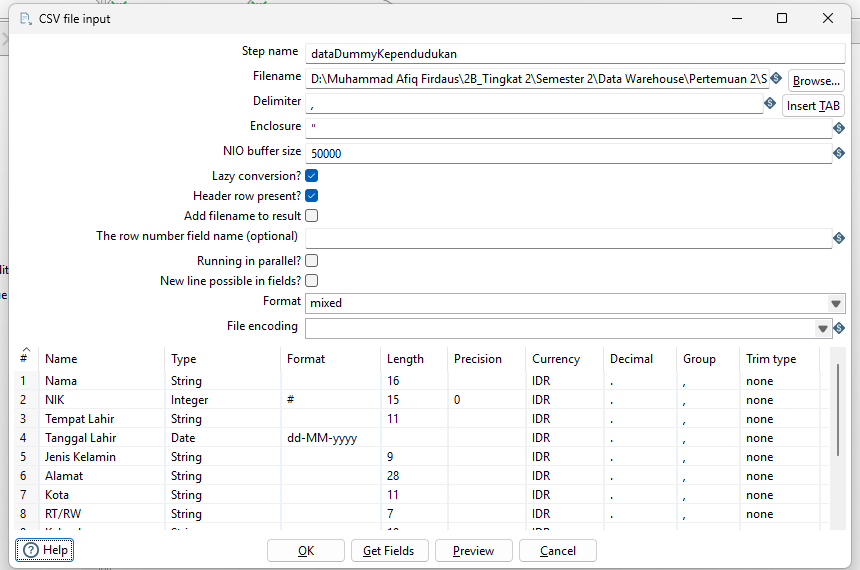
Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. darai data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah Pendidikan akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

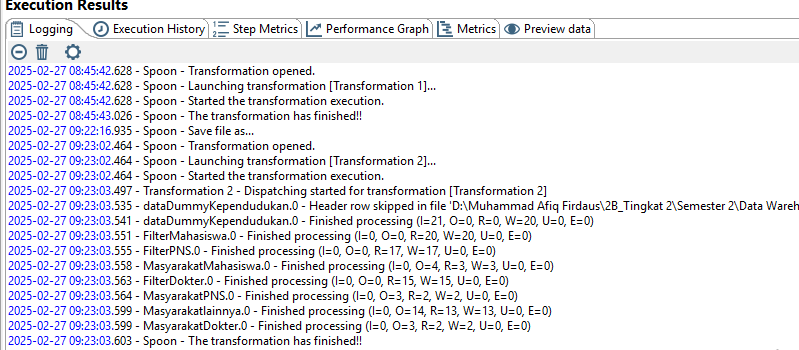
data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

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

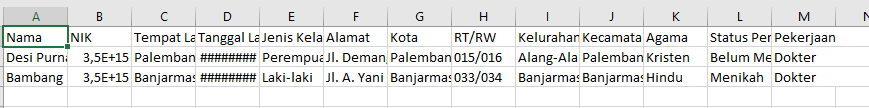
Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.



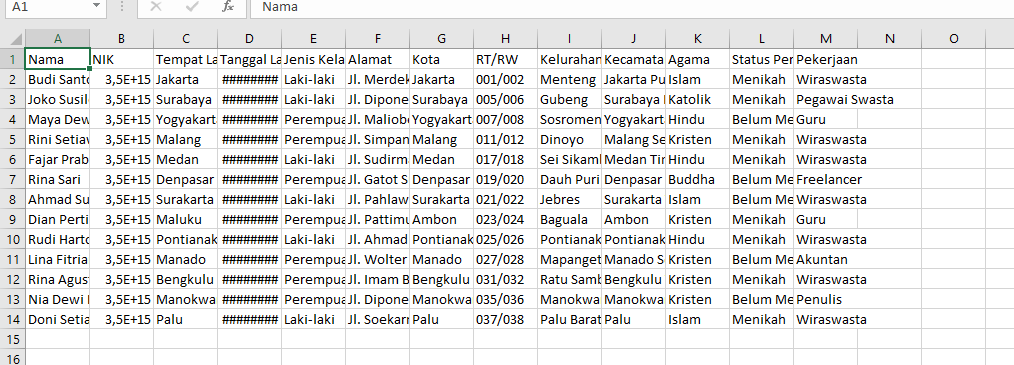




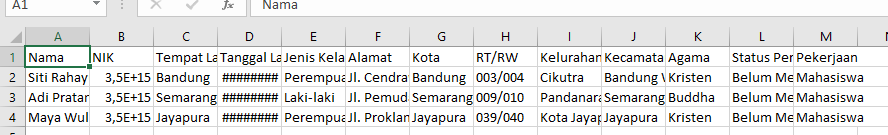
MasyarakatDokter



MasyarakatLainnya



MasyarakatMahasiswa



MasyarakatPNS

