­LAPORAN PRAKTIKUM DATA WAREHOUSE

JOBSHEET 1 : DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON



DISUSUN OLEH :

AQILA NUR AZZA (2341760022)

KELAS 2A-SIB/04

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

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

Jl. Soekarno Hatta No. 9, Jattimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur

65141

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 : Pada file CatDistributionSales.csv terdapat 7 jumlah kolom yang tersedia

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

Jawaban :

* Pada kolom A terdapat title *customer* untuk mengisi nama customers
* Pada kolom B terdapat title *product* untuk mengisi Jenis Unit
* Pada kolom C terdapat title *region* untuk mengisi Lokasi Cabang
* Pada kolom D terdapat title *date* untuk mengisi tanggal penjualan/transaksi
* Pada kolom E terdapat title *item unit price* untuk mengisi harga perunit kendaraan
* Pada kolom F terdapat title *no items* untuk mengisi jumlah unit yang terjual
* Pada kolom G terdapat title *total sale* untuk mengisi total harga penjulan (*item unit price* x *no items*)

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

Jawaban : Ada, terdapat data yang tidak lengkap pada kolom Product, Region dan Date.

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

Jawaban : Tidak ada, semua data memiliki tipe yang sama pada kolom yang sama.

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawaban : Ya, hasil eksekusi pada PDI SPOON 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 :

* Simpan file yang ditransformasikan
* File transformasi terbuka
* Menampilkan tab transformsi
* Proses eksekusi pada transformasi akan dimulai
* Proses pengiriman data untuk transformasi dimulai
* Baris pertama dalam file CSV yang berisi nama kolom dilewati dan tidak berproses
* Proses membaca file CSV selesai, dengan 20 baris data dibaca, 19 baris diproses, dan hasilnya ditulis ke tempat lain.
* Proses transformasi selesai tanpa kesalahan.

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



Jawaban :

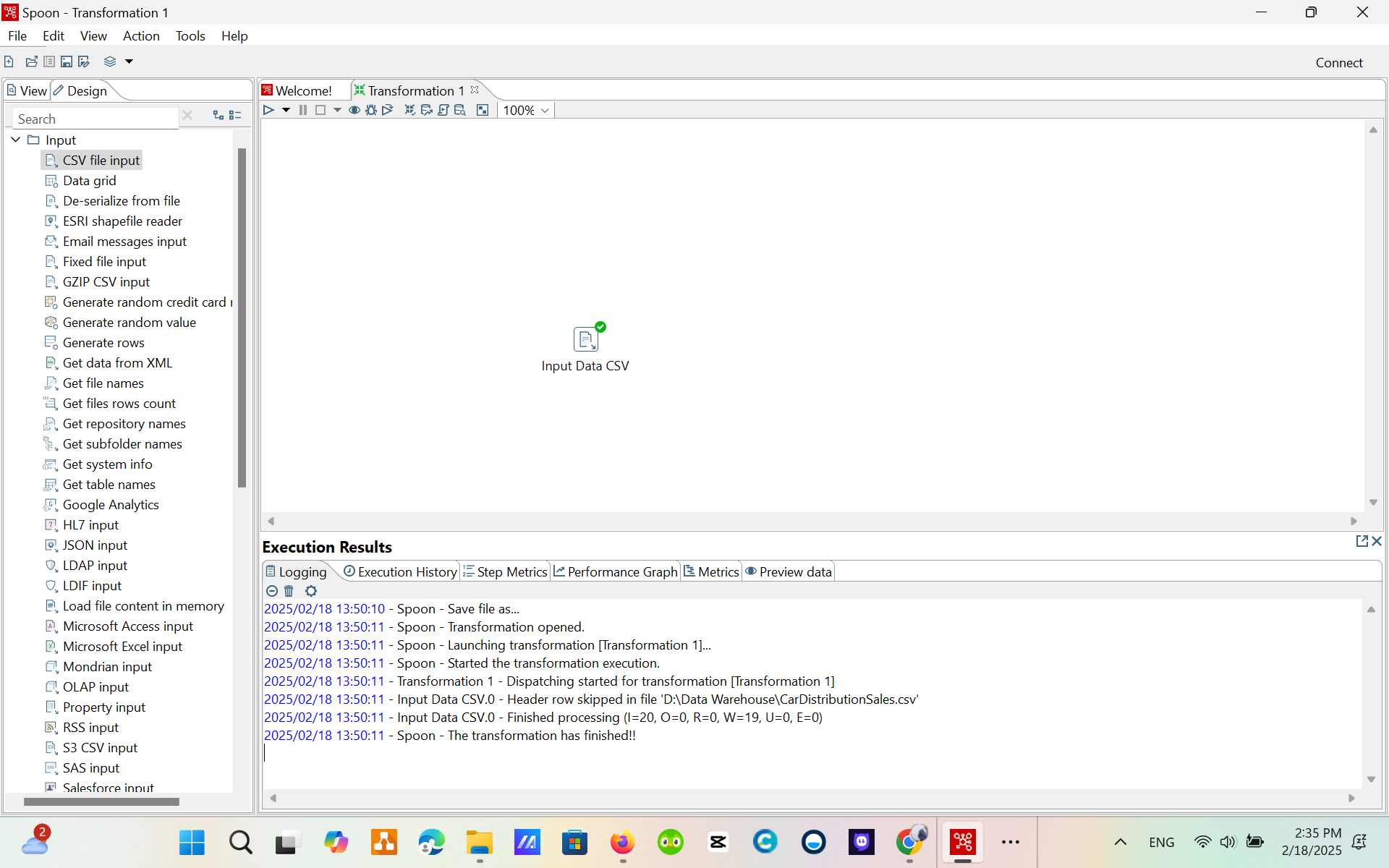
Itu merupakan proses pemrosesan data pada input CSV selesai :

* I (baris input) = 20 : 20 baris data dibaca dari file
* O (baris output)= 0 : tidak ada baris yang dikeluarkan
* R (baris yang ditulis ulang) = 0 : tidak ada baris yang dibaca ulang
* W (baris yang berhasil ditulis)= 19 : 19 baris data diproses dan ditulis ke output
* U (baris yang diperbarui)= 0 : tidak ada data yang diperbarui
* E (kesalahan (error))= 0 : tidak ada kesalahan

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 :



Untuk melakukan eksekusi Extract pada tersebut hanya membutuhkan waktu selama 1 detik saja

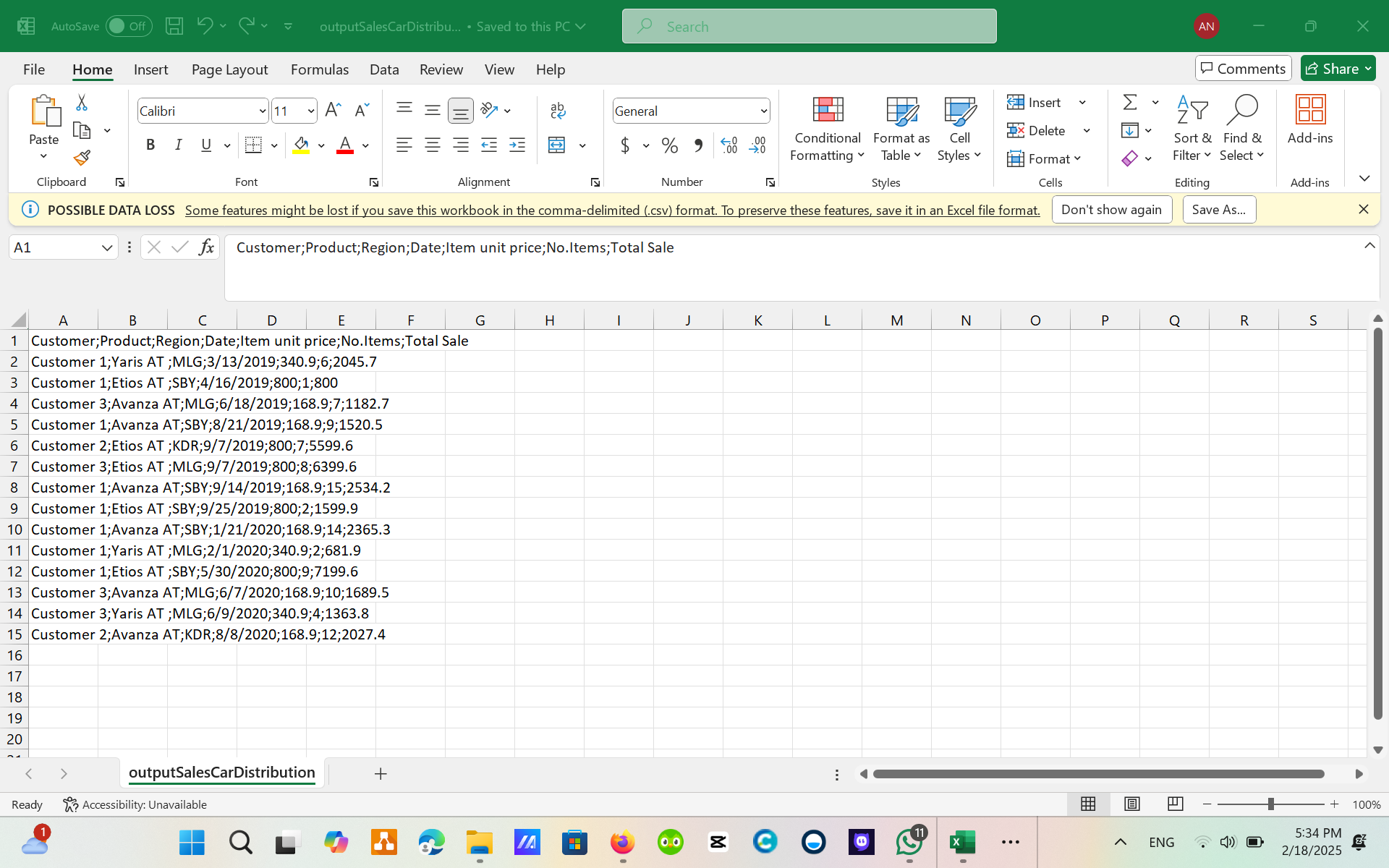
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NAME | PROCESSOR | RAM | STORAGE | VGA | CPU | TIME |
| Saya | AMD Ryzen 5 5600H | 8 GB | 512 GB | AMD Radeon(TM) Graphics | AMD Ryzen 5 5600H with Radeon Graphics | 1s |
| Yonanda | Core(TM) i5-12450H | 16 GB | 512 GB | NVIDIA GeForce RTX 3050 | 12 CPU ~2.0GHZ | 1s |
| Dahniar | Intel(R) Core(TM) i5-6300U | 8 GB | 256 GB | Intel(R) HD Graphics 520 | CPU @ 2.40GHz 2.50 GHz | 1s |
| Reishi | Inter(R) CoreTM i7-7820HQ | 16 GB | 256 GB | NVIDIA FeForce 940MX | 2.90Ghz | 1s |

**TUGAS 3**

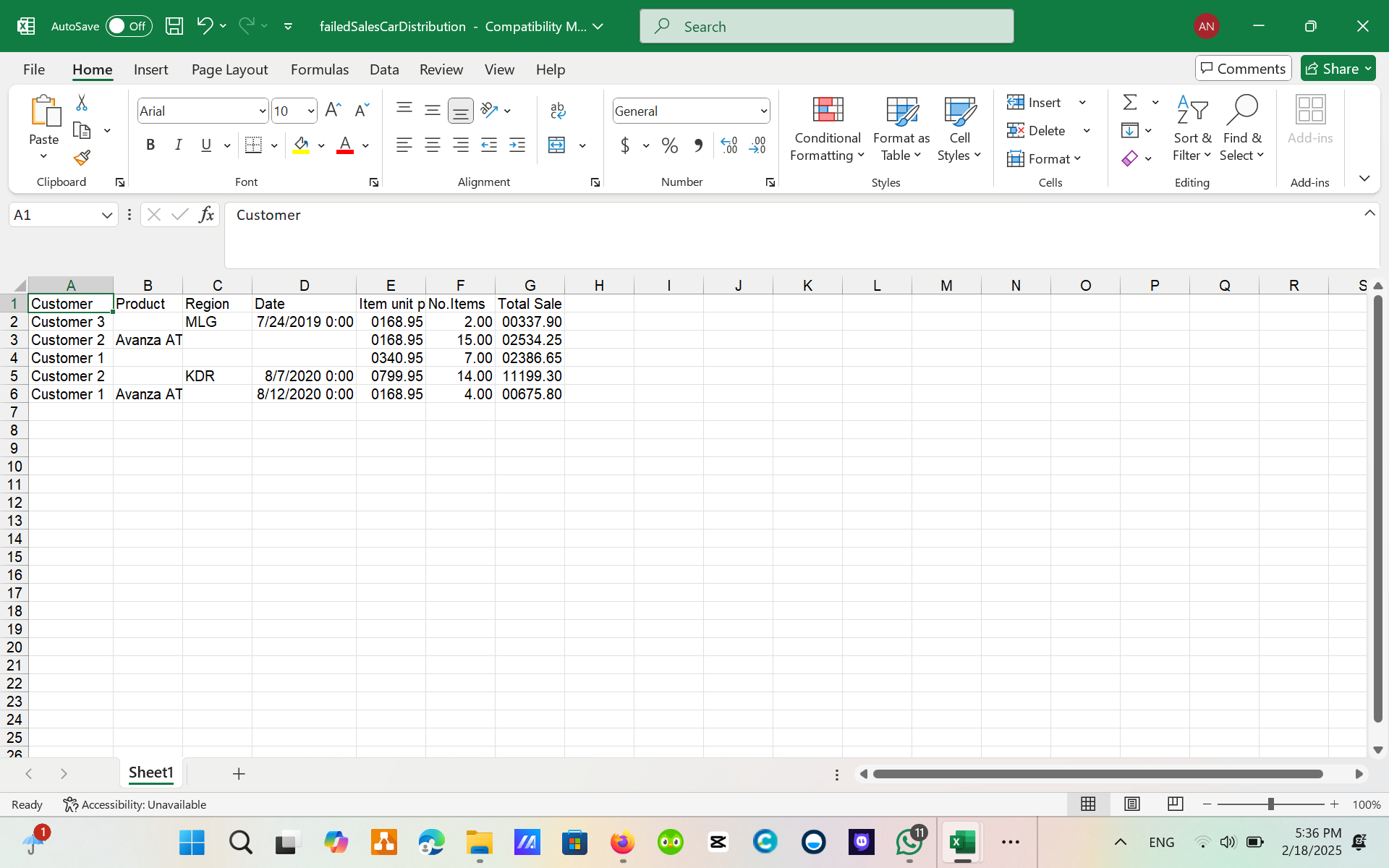
1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

Jawaban :

* + File csv



* + File excel (xlx)



|  |  |
| --- | --- |
| File CSV | File EXCEL |
| File yang berisi data dalam baris dan kolom yang setiap nilai dipisahkan dengan tanda koma | File yang berisi data dalam baris dan kolom yang setiap nilai dipisahkan dengan sel |
| Pada file tersebut telah diatur kondisi true pada Filter rows yang membuat data yang masuk pada file harus memiliki nilai | Pada file tersebut telah diatur kondisi false pada Filter rows yang membuat data yang masuk pada file terdapat data yang null atau kosong |

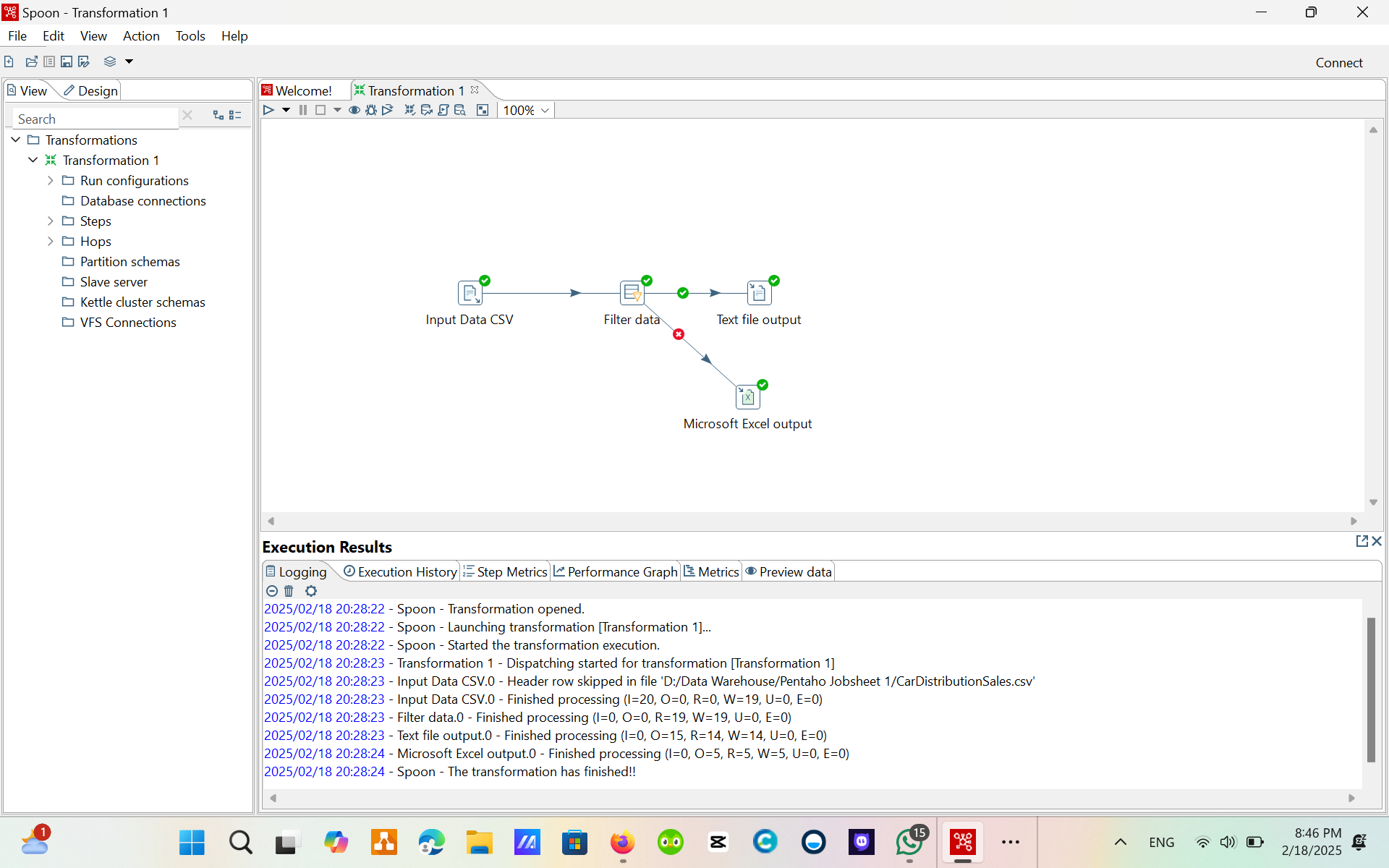
1. Jelaskan apa yang terjadi pada proses **Filter rows**!

Jawaban :

Pada proses *Filter Rows (Filter Data)* dalam proses ini berperan untuk penyaringan data dari file *Input Data CSV* sebelum data tersebut diteruskan pada output yang berbeda yaitu *Text file output* (Result is true) jika memenuhi syarat dan *Microsoft Excel Output* (Result is false) jika tidak memenuhi

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 :



Untuk melakukan eksekusi Extract Transfer Load Data dibutuhkan waktu selama 2 detik

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NAME | PROCESSOR | RAM | STORAGE | VGA | CPU | TIME |
| Saya | AMD Ryzen 5 5600H | 8 GB | 512 GB | AMD Radeon(TM) Graphics | AMD Ryzen 5 5600H with Radeon Graphics | 2s |
| Yonanda | Core(TM) i5-12450H | 16 GB | 512 GB | NVIDIA GeForce RTX 3050 | 12 CPU ~2.0GHZ | 2s |
| Dahniar | Intel(R) Core(TM) i5-6300U | 8 GB | 256 GB | Intel(R) HD Graphics 520 | CPU @ 2.40GHz 2.50 GHz | 2s |
| Reishi | Inter(R) CoreTM i7-7820HQ | 16 GB | 256 GB | NVIDIA FeForce 940MX | 2.90Ghz | 2s |

**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,

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

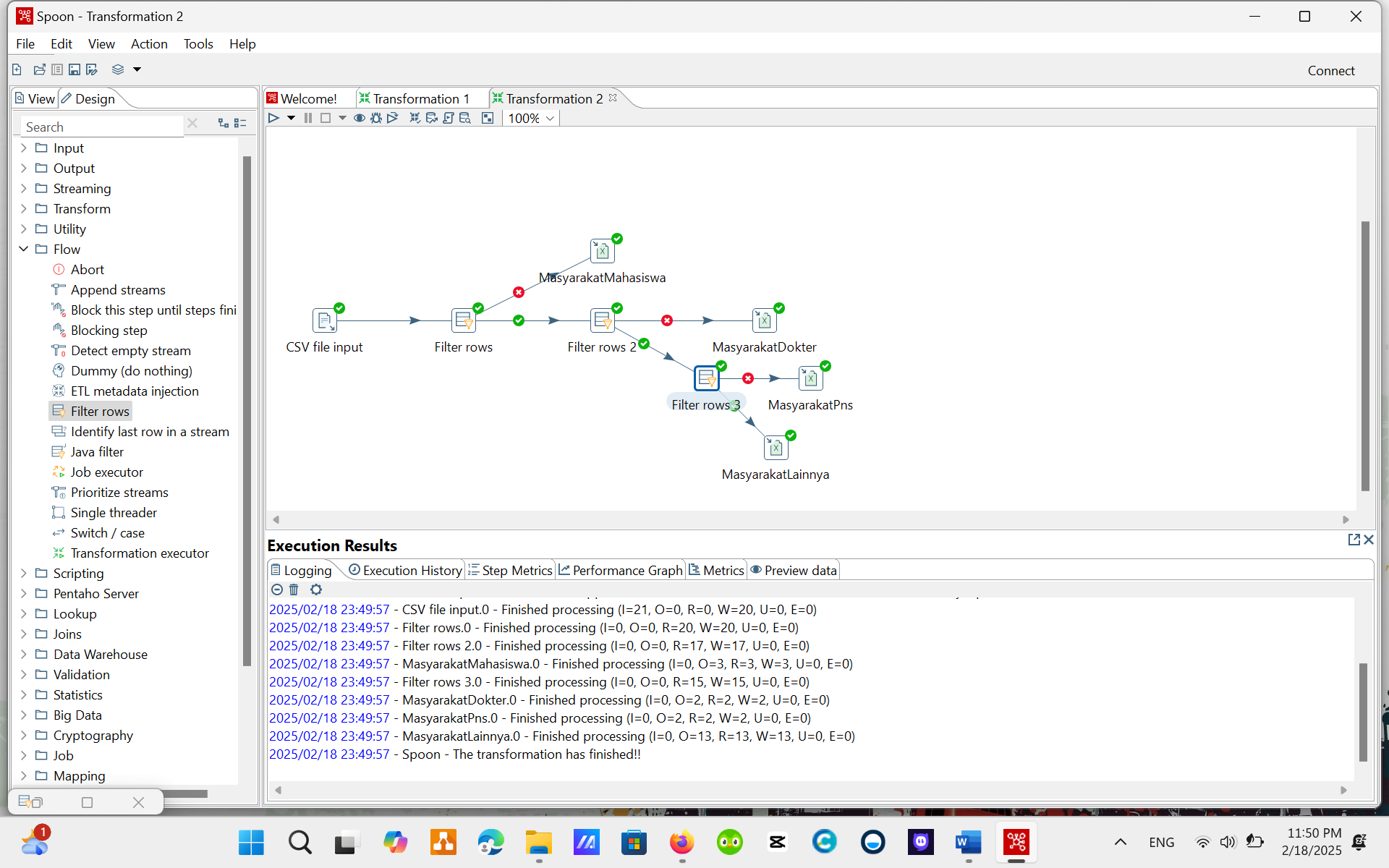
data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

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

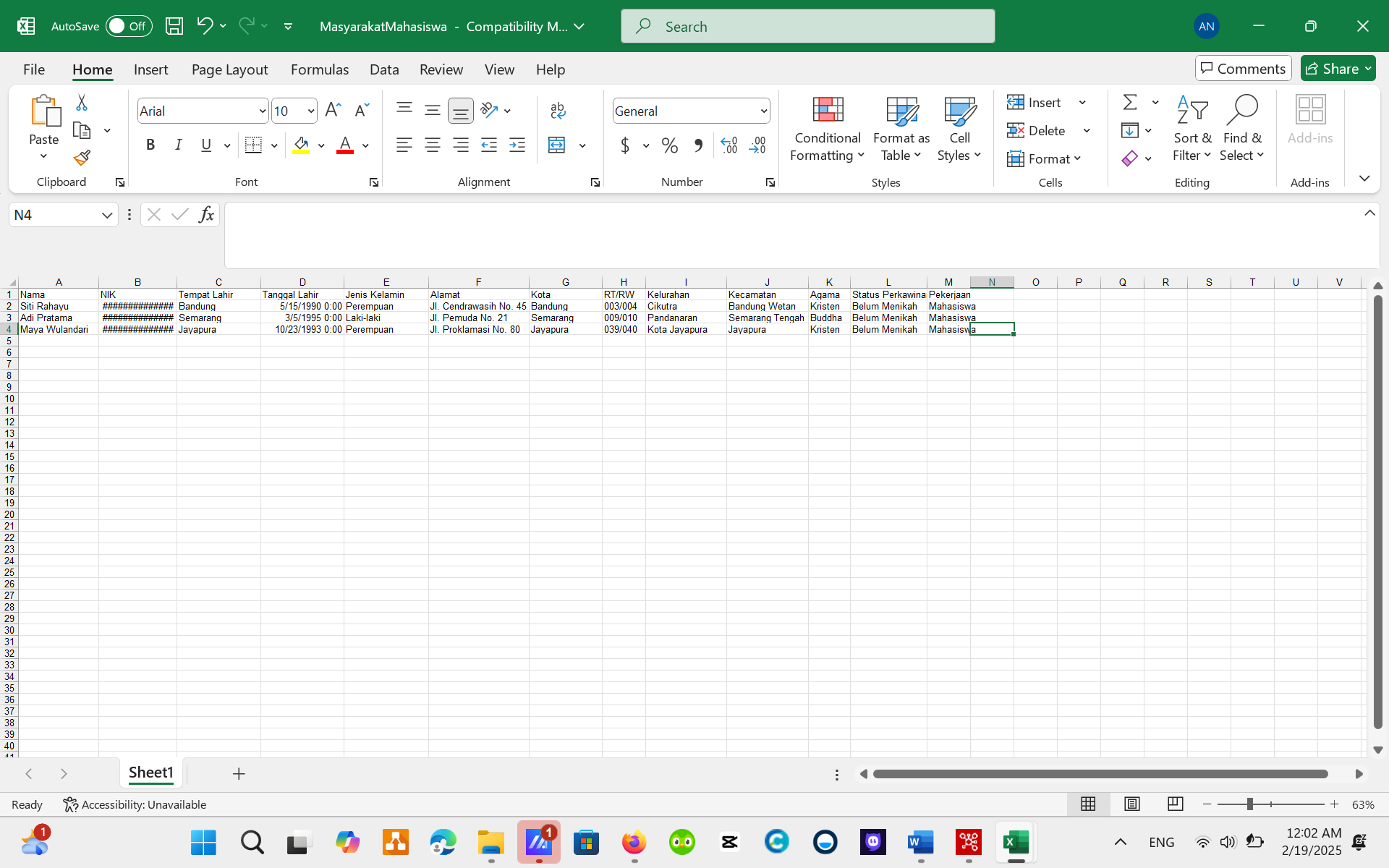
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

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

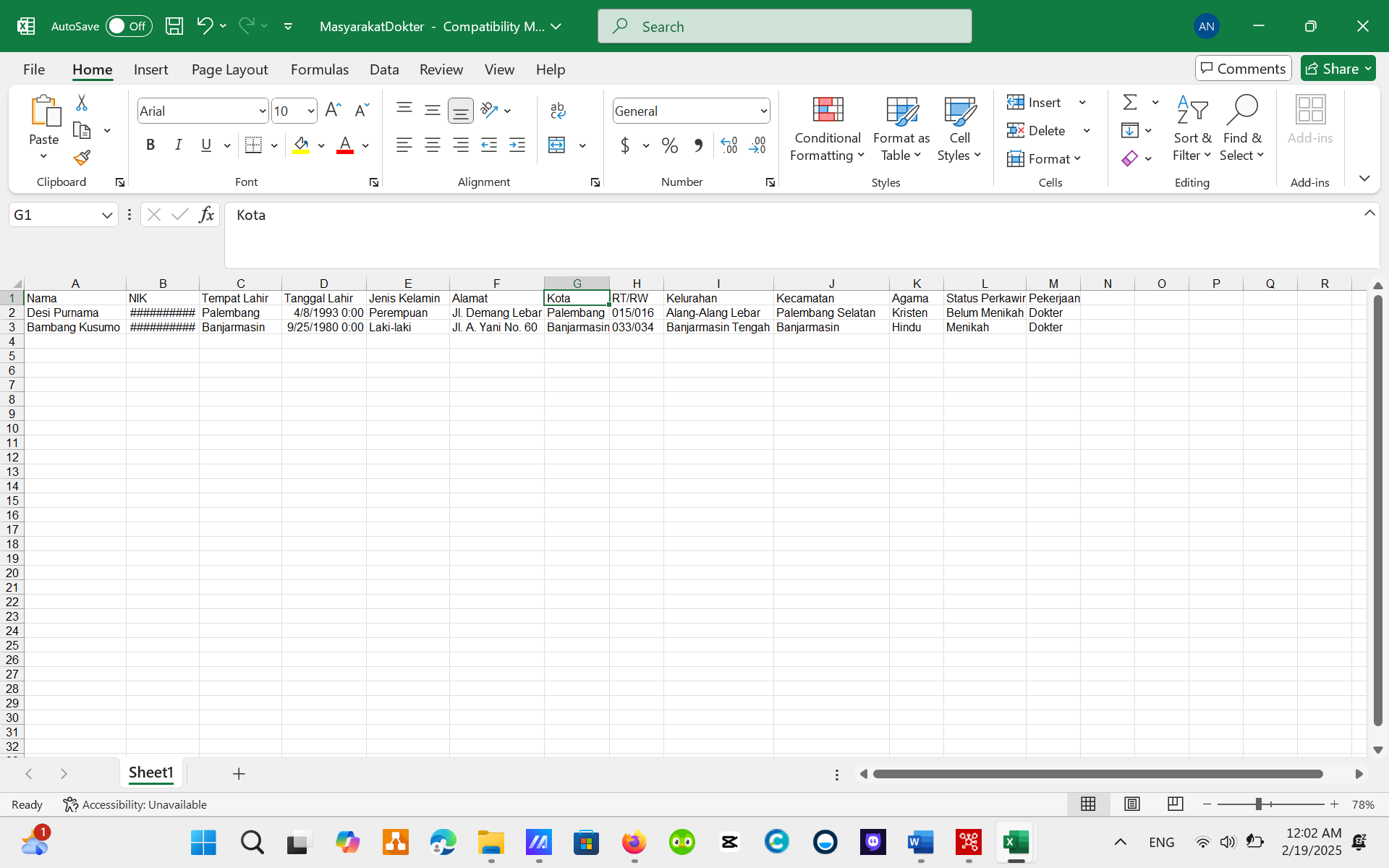
RESULT :



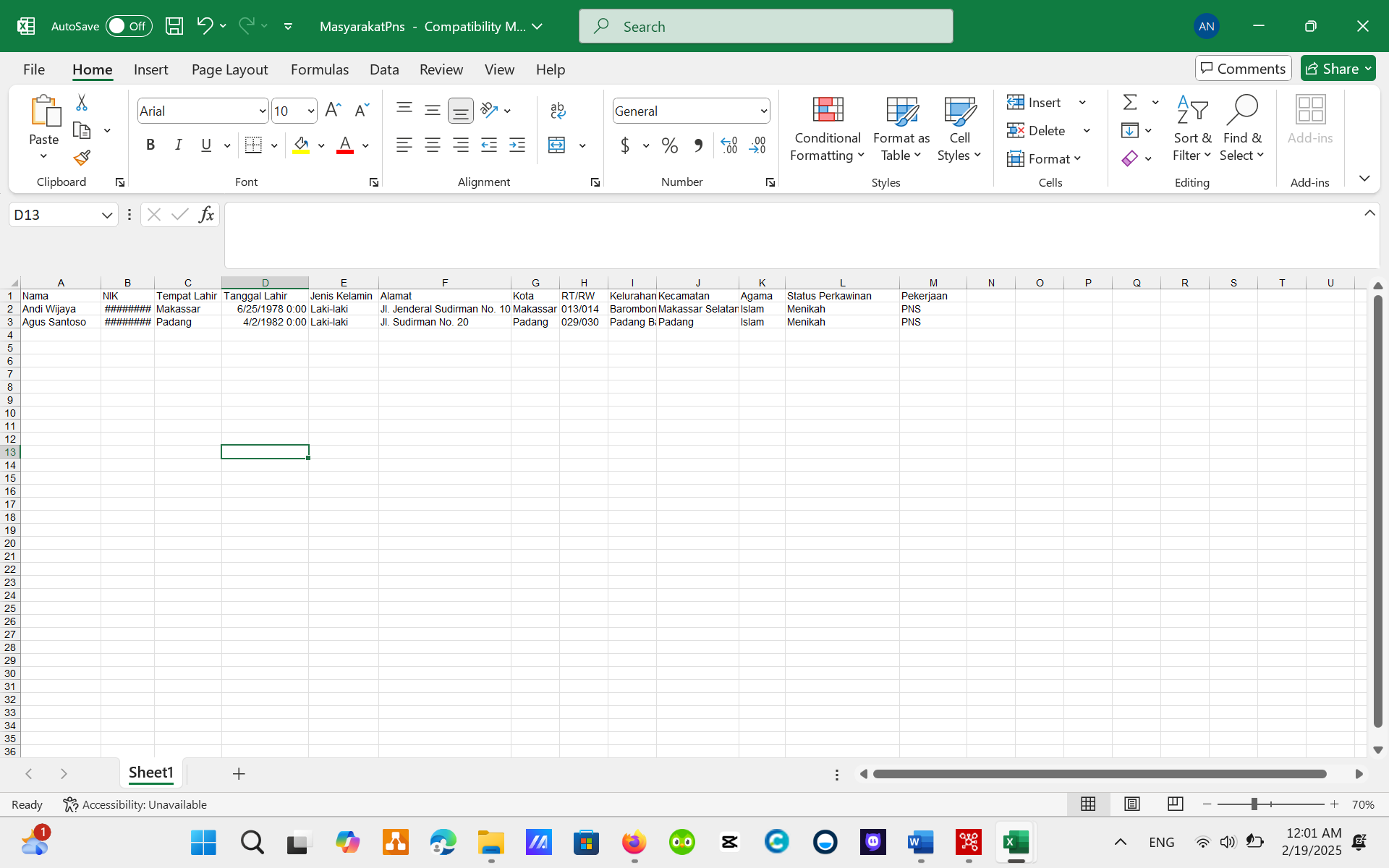
* MasyarakatMahasiswa



* MasyarakatDokter



* MasyarakatPns



* MasyarakatLainnya

