TUGAS PRAKTIKUM DATA WAREHOUSE JOBSHEET 1

MATERI DATA WAREHOUSE



Disusun Oleh:

Maisya Dhia (2341760114)

SIB 2B

SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2024

**Jobsheet 1: Data Warehouse Sederhana dan Pengenalan PDI Spoon**

**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.

**TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?
2. Apa arti atau isi data dari setiap kolom yang ada?
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

**JAWABAN**

1. Jumlah kolom tersebut ada 7
2. **Customer –** nama pelanggan yang akan melakukan pembelian

**Product –** jenis mobil yang akan dibeli

**Region –** Wilayah/ cabang dimana terjadinya transaksi

**Date -**Tanggal terjadinya transaksi pembelian

**Item unit price –** harga per unit dari mobil yang dibeli

**No.** Item – jumlah unit mobil yang dibeli

**Total sale –** total penjualan untuk transaksi tersebut ( harga per unit x jumlah unit )

1. **Ada,** pada kolom product baris ke-4 dan ke-17 kosong

**Kolom region,** baris ke-8 dan ke-19 tidak memiliki wilayah transaksi

**Kolom date,** baris ke-8 dan ke-13 tidak memiliki tanggal transaksi

1. **Ada,**

* **Kolom "Date" (Tanggal transaksi)** → Seharusnya semua berisi tanggal dalam format MM/DD/YYYY. Tapi ada beberapa baris yang kosong, sehingga akan dianggap sebagai **missing value** atau **null**.
* **Kolom "Item unit price", "No. Items", dan "Total Sale"** → Seharusnya berisi angka (float atau integer). Jika ada data kosong atau karakter lain (misalnya teks yang tidak seharusnya ada), maka akan terjadi perbedaan tipe data dalam kolom yang sama.

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
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?
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?
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.

**JAWABAN**

* 1. **Hampir sama,**
* Baris pertama(judul kolom) dilewati saat membaca file.
* Total baris cocok dengan jumlah baris dalam file(data berhasil dibaca semua)
* 19 baris di tulis ke output bisa jadi karena ada data yang tidak lengkap, bisa jadi karena terjadi *missing* data
  1. **2025-02-21 08:54:45.227 - Spoon - Transformation opened.-** PDI SPOON membuka transformasi yang dijalankan

**2025-02-21 08:54:45.228 - Spoon - Launching transformation [Transformation 1]... –** transformasi mulai dijlankan

**2025-02-21 08:54:45.228 - Spoon - Started the transformation execution. –** PDI mulai membaca dan memproses data dari file sumber

**2025-02-21 08:54:45.275 - Transformation 1 - Dispatching started for transformation [Transformation 1] –** Data dari file csv mulai diekstrak

**2025-02-21 08:54:45.324 - Input Data CSV.0 - Header row skipped in file 'file:///Applications/data-integration/Data-Warehouse/CarDistributionSales.csv' –** baris pertama ( judul kolom) tidak diproses karena hanya untuk referensi

**2025-02-21 08:54:45.327 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0) –** data selesai diproses dengan 20 input baris, 19 berhasil ditulis ke output

**2025-02-21 08:54:45.336 - Spoon - The transformation has finished!! –** proses extract selesai

* 1. **I = INPUT,** jumlah total baris yang dibaca sumber data(20)

**O = OUTPUT,** Jumlah baris yang dikirim ke output tertentu (0, tidak ada output spesifik selain W)

**R = REJECTED** , Jumlah baris yang ditolak atau gagal diproses (0, tidak ada data yang gagal)

**W = WRITTEN,** Jumlah baris yang berhasil ditulis ke target akhir (19)

**U = UPDATED, J**umlah baris yang diperbarui dalam proses ini (0)

**E = ERROR ,** Jumlah baris yang menghasilkan error selama pemrosesan (0)

* 1. **Waktu Eksekusi**

**Waktu mulai :** 2025-02-21 08:54:45.227

**Waktu selesai :** 2025-02-21 08:54:45.336

**Total waktu eksekusi:** **0.109 detik (109 milidetik)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NAMA** | **PROCESSOR** | **RAM** | **STORAGE** | **VGA** | **CPU** | **WAKTU EKSEKUSI** |
| MAISYA (saya) | Apple M1 (8-core) | 8GB | SSD 256GB / 512GB | Integrated (Apple M1 GPU, 8-core) | Apple M1 8-core | 0.109 detik |
| DHEVINA | 12th Gen Intel(R) Core(TM) i5 1235U 1.30 GHz | 8GB | 152GB | Inter(R) UHD Graphics | 12th Gen Intel(R) Core(TM) i5  1235U 1.30 GHz | 0.0 detik |
| RIZKY | AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz | 8GB | 512 gb | AMD RADEON  GRAPHICS | AMD RYZEN 3  7320U with  Radeon Graphics | 0.1 detik |
| NIMAS | AMD Ryzen  3 5300U | 16GB | 475 GB | AMD Radeon  Graphics | AMD Ryzen 3  5300U with  Radeon Graphics. | 0.1 detik |

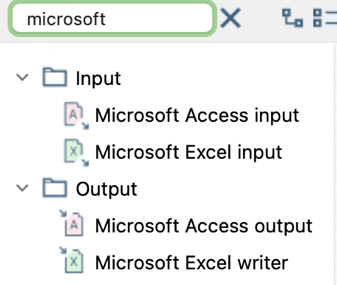
**TUGAS 3**

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
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

**JAWABAN**



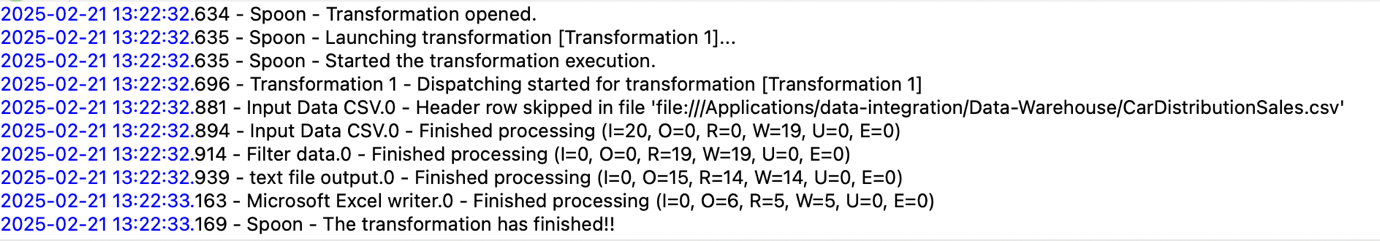
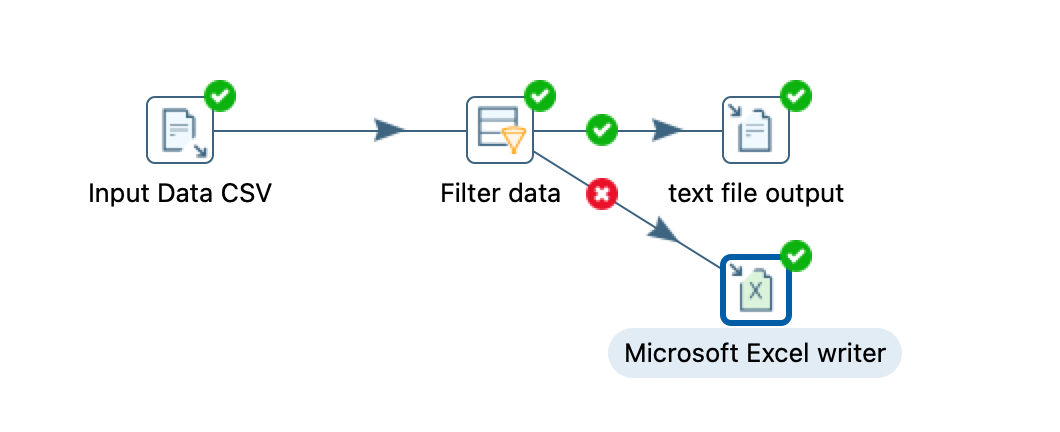
|  |  |  |
| --- | --- | --- |
| **Perbedaan** | **Csv output** | **Excel output \*excel writer** |
| **Jumlah data** | Lebih Lengkap | Hanya sebagian data |
| **Format data** | Lengkap dan sesuai | Ada data yang berubah jadi ‘null’ |
| **Data yang hilang** | Tidak ada ‘null’ | Ada beberapa baris dengan ‘null’ di beberapa kolom |
| **Tanggal** | Lengkap | Ada format tahun yang ‘null’ |
| **Produk/Region** | Lengkap | Ada beberapa baris yang ‘null’ pada Product/Region |

* 1. Filter rows dalam pentaho digunakan untuk memilih banyak data tertentu yang memnuhi syarat dan membuang data yang tidak sesuai, kemungkinan akan terjadi filter yang menyebabkan data hilang atau menjadi <null> dalam excel writer --**\*disini saya menggunakan excel writer karena ketika saya mencari tidak ada excel output--**filter bisa jadi mengambil hanya data yang tidak lengkap, sehingga hanya data dengan nilai kosong atau null yang masuk ke Excel atau bisa jadi filter salah menargetkan data yang valid, sehingga malah menyimpan data yang salah pada excel. Terakhir , format tanggal berubah dan ada yang salah karena tidak di *handle* dengan baik saat transformasi
  2. **Waktu mulai :** 2025-02-21 13:22:32.634

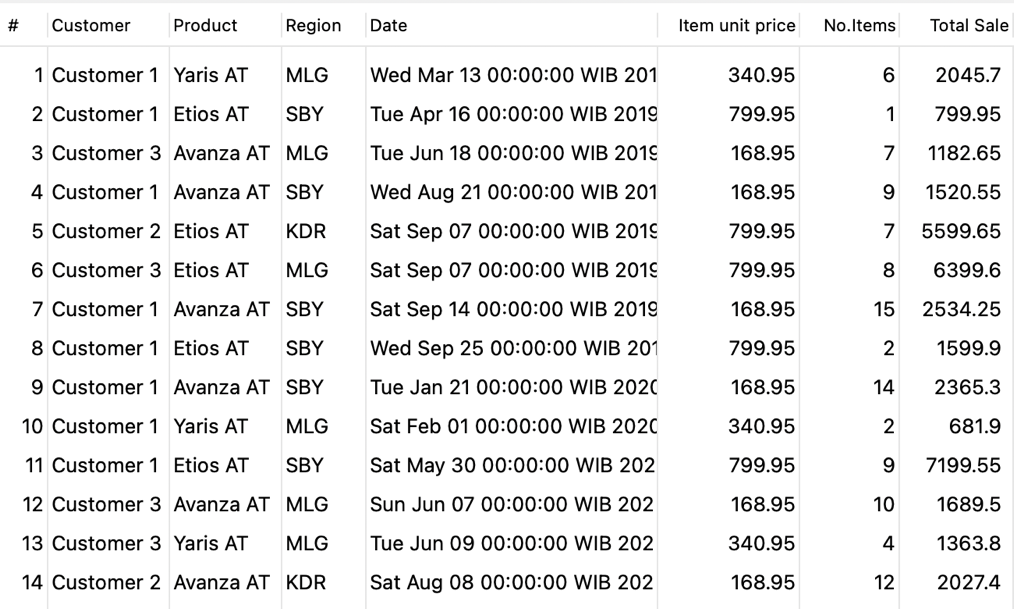
**Waktu selesai :** 2025-02-21 13:22:33.169

**Perbandingan waktu : 0.535 detik**

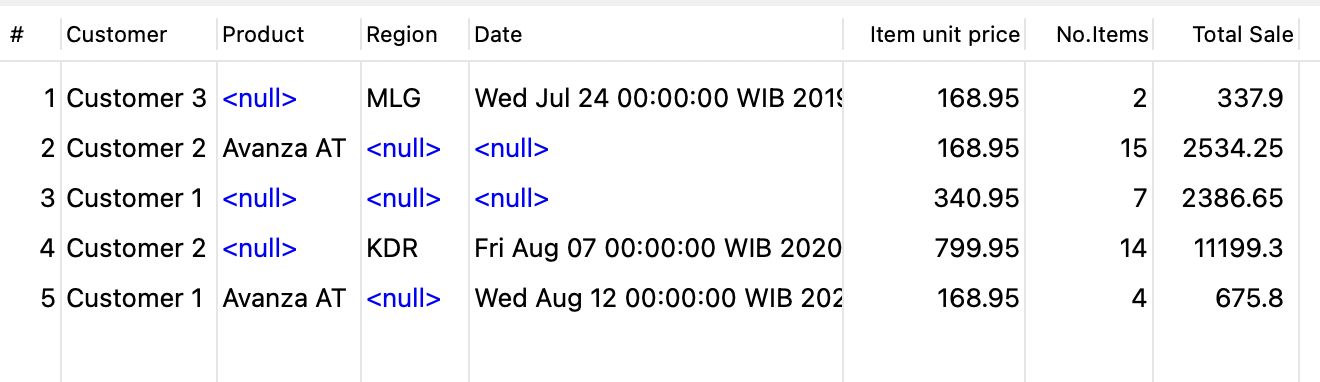
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NAMA** | **PROCESSOR** | **RAM** | **STORAGE** | **VGA** | **CPU** | **WAKTU EKSEKUSI** |
| MAISYA (saya) | Apple M1 (8-core) | 8GB | SSD 256GB | Integrated (Apple M1 GPU, 8-core) | Apple M1 8-core | 0.535 detik |
| DHEVINA | 12th Gen Intel(R) Core(TM) i5 1235U 1.30 GHz | 8GB | 152GB | Inter(R) UHD Graphics | 12th Gen Intel(R) Core(TM) i5  1235U 1.30 GHz | 0. detik |
| RIZKY | AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz | 8GB | 512 gb | AMD RADEON  GRAPHICS | AMD RYZEN 3  7320U with  Radeon Graphics | 0.2 detik |
| NIMAS | AMD Ryzen  3 5300U | 16GB | 475 GB | AMD Radeon  Graphics | AMD Ryzen 3  5300U with  Radeon Graphics. | 0.4 detik |

**HASIL JOBSHEET 1**

**HASIL DARI FILTER DATA(DATA HINGGA 19 ROWS)**

**HASIL DARI TEXT FILE OUTPUT**

**HASIL DARI MICROSOSFT WRITER**

****

# Studi Kasus

**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.

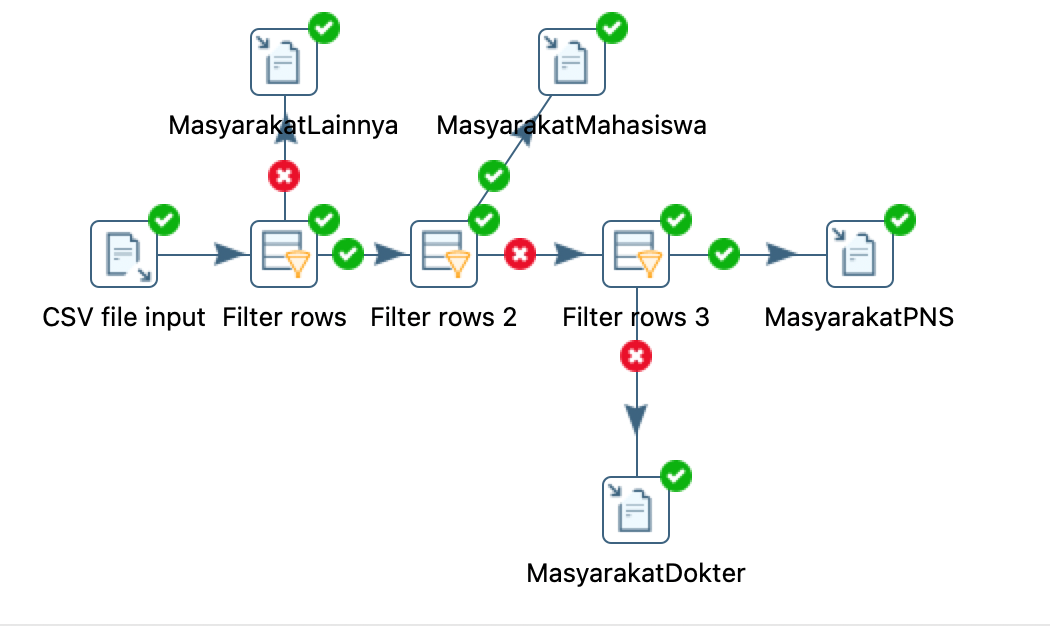
**Data masyarakat berprofesi PNS** akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan

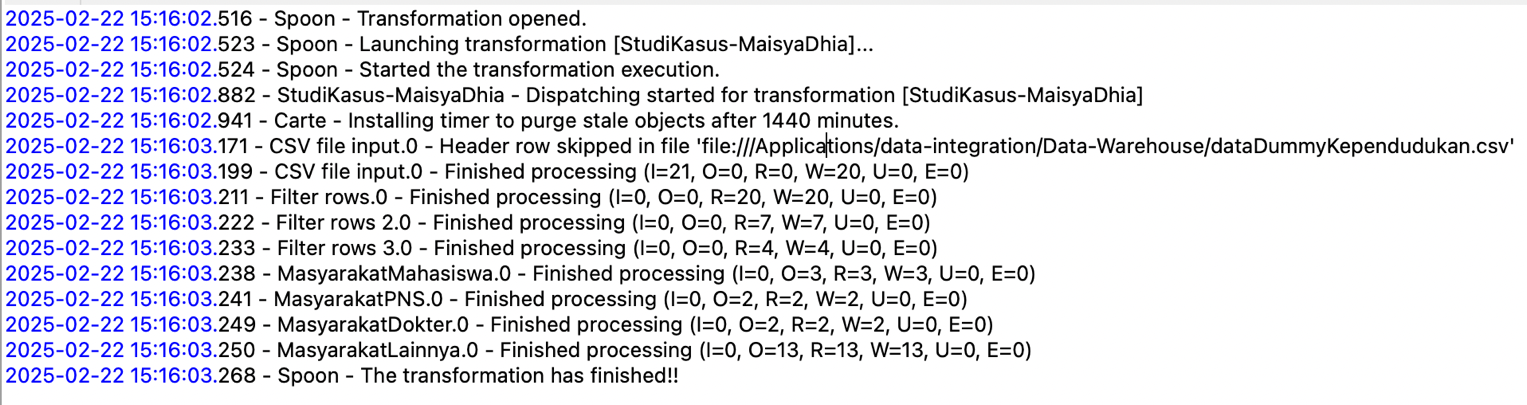
**Data masyarakat berprofesi Dokter** akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

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.





**HASIL DARI CSV FILE INPUT( 20 rows)**

****

**HASIL DARI FILTER ROWS ( 20 rows)**

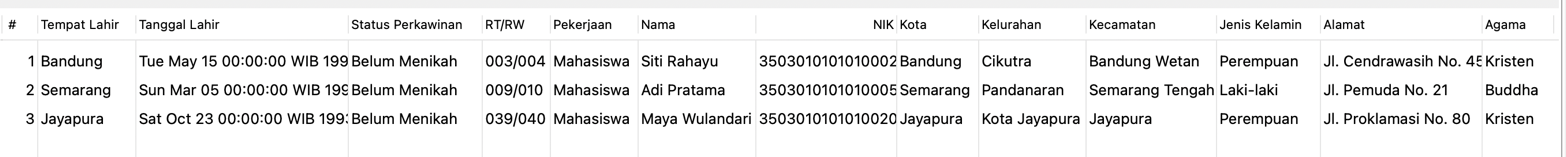
****

**TEXT FILE MILIK “MASYARAKAT LAINNYA”**

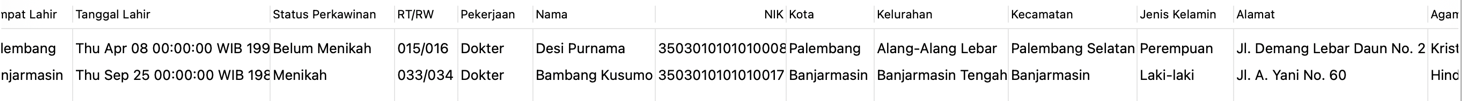
**FILTER ROWS 2**

****

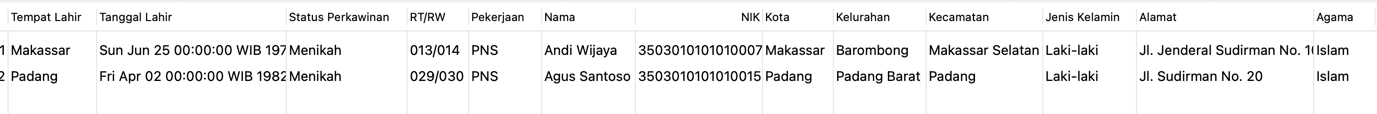
**TEXT FILE MAHASISWA**

****

**FILTER ROWS 3**

**TEXT FILE MASYARAKAT DOKTER**

**TEXT FILE MASYARAKAT PNS**

****