

DATA WAREHOUSE

JOBSHEET 1



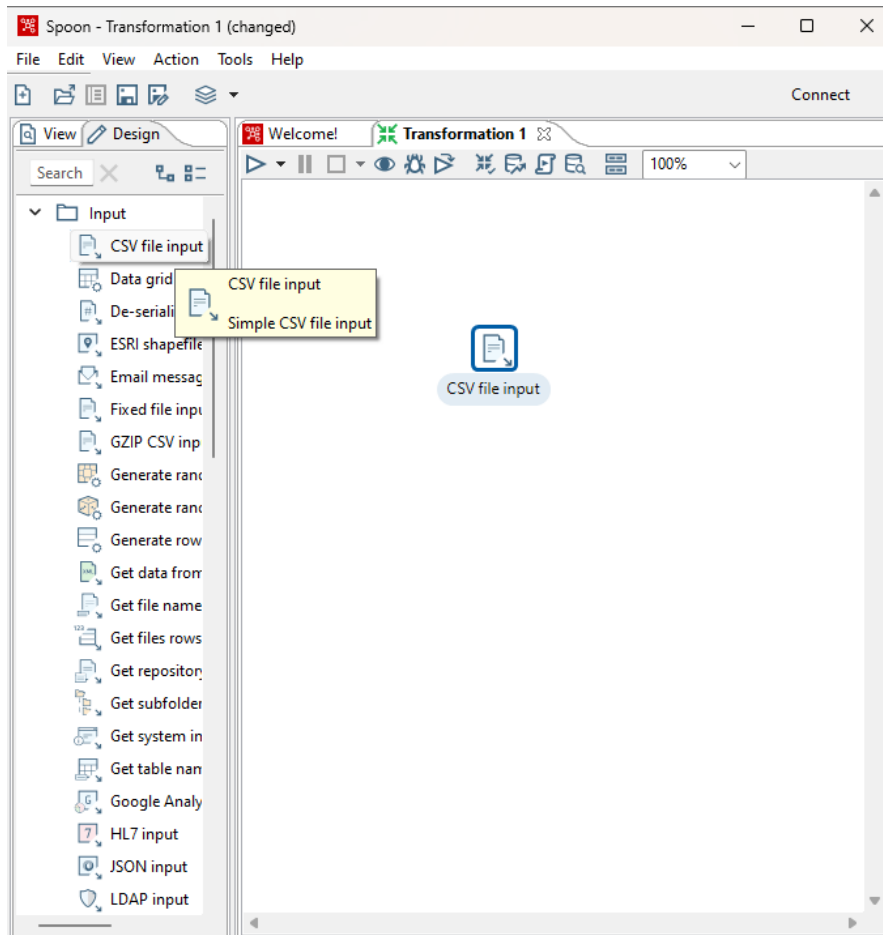
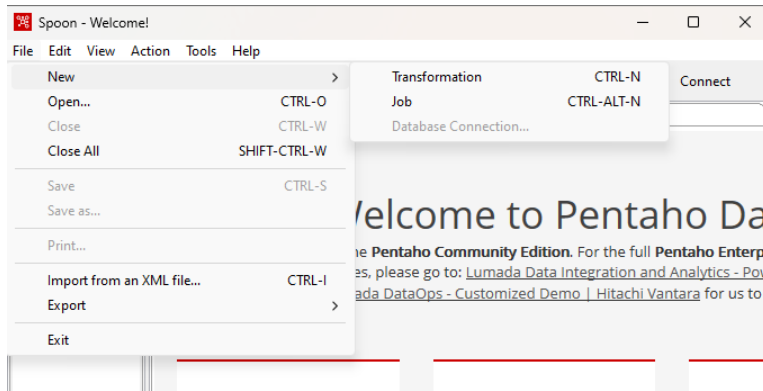
NAMA: Claudya Destine Julia Handoko

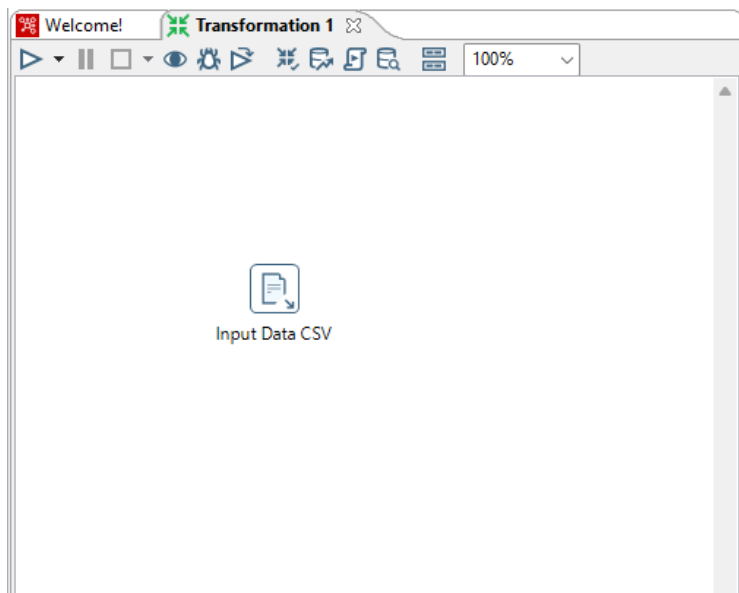
NIM: 2341760008

KELAS : 2A

PRODI: D-IV Sistem Informasi Bisnis

A. Pengambilan Data (Extract)





Run Options

Run configuration:
Pentaho local

Options

☒ Clear log before running Log level: Basic

☐ Enable safe mode

☒ Gather performance metrics

Parameters Variables

Parameter	Default value	Value	Description

☒ Always show dialog on run

[? Help](#) [Run](#) [Cancel](#)

Execution Results							
<div> Logging Execution History Step Metrics Performance Graph Metrics Preview data </div> <div> First rows Last rows Off </div>							
#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
9	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
10	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
11	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
12	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawaban : hasil eksekusi sudah sesuai dengan data aslinya. Namun, tetap perlu dilakukan pengecekan untuk memastikan tidak ada data yang hilang atau mengalami perubahan format setelah ekstraksi.

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?

Jawaban :

- a. Klik dua kali pada objek **CSV file input** hingga jendela pengaturannya muncul.
- b. Ubah **Step Name** menjadi **Input Data CSV**.
- c. Klik **Browser** pada bagian **filename** lalu pilih file **CarDistributionSales.csv**.
- d. Pastikan delimiter yang digunakan adalah koma (,).
- e. Klik **Get Fields** untuk mengambil nama kolom dari file CSV.
- f. Setelah selesai, tutup jendela konfigurasi.
- g. Tekan tombol **“Run”** di pojok kanan atas area kerja. Setelah proses berjalan, buka tab **Preview Data** di **Execution Result Area** untuk melihat hasilnya.

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

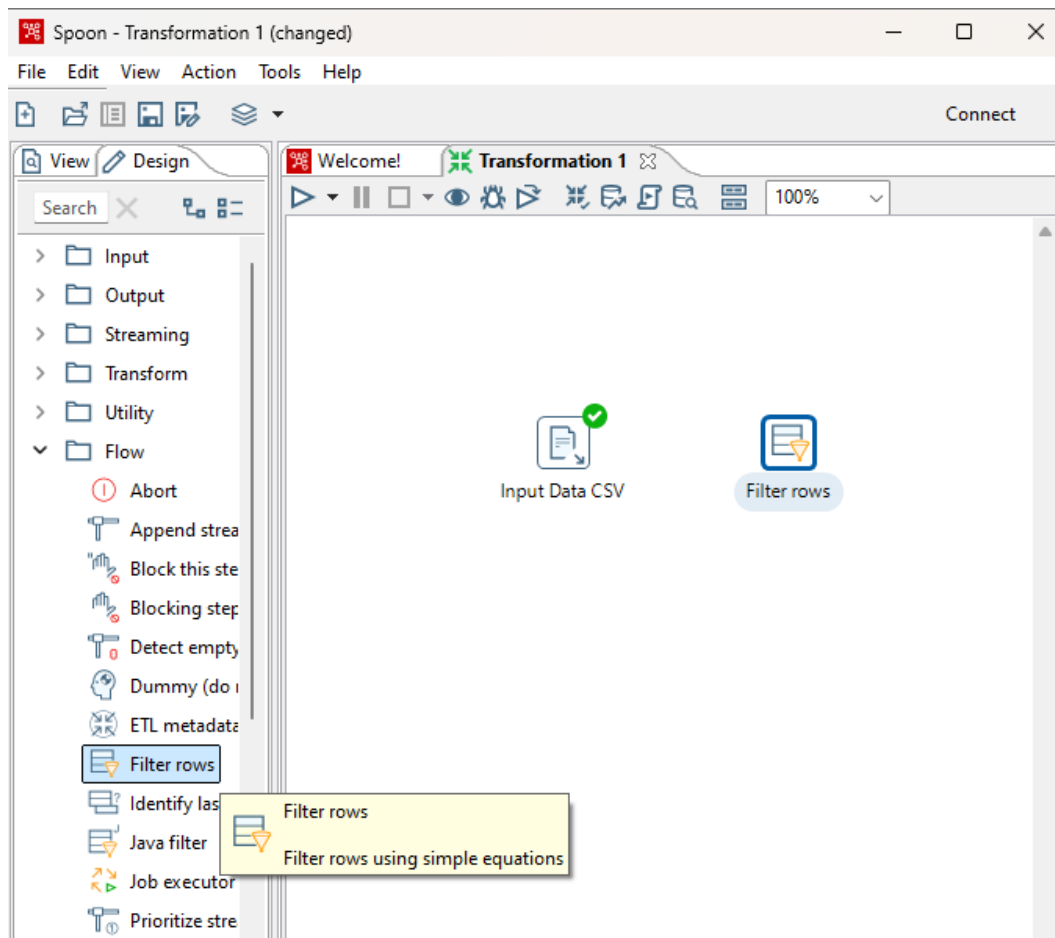
2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

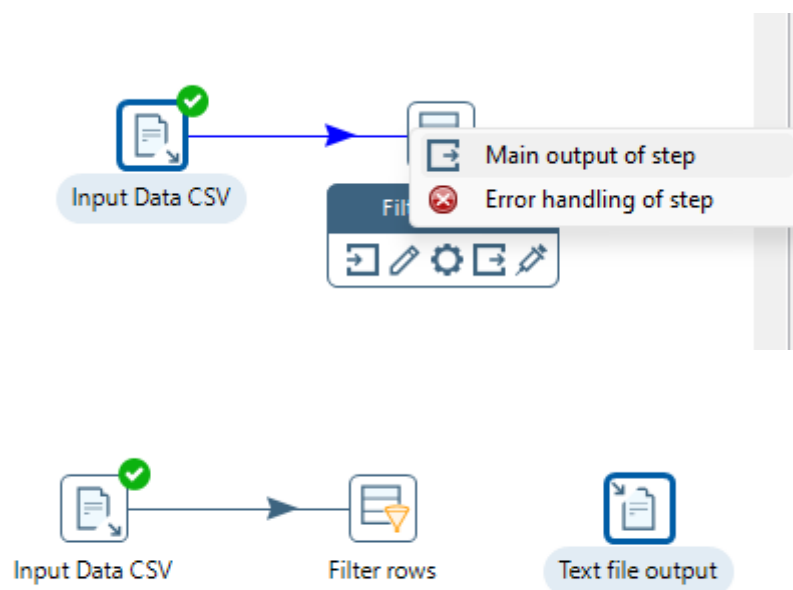
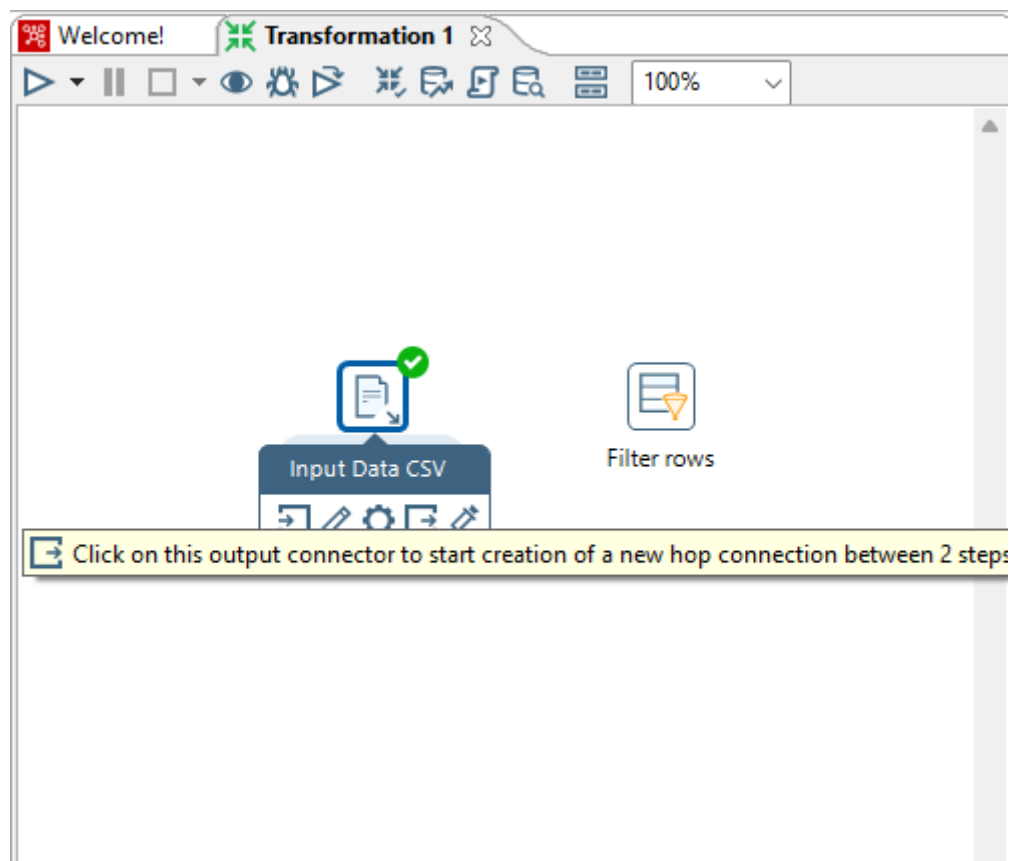
Jawaban :

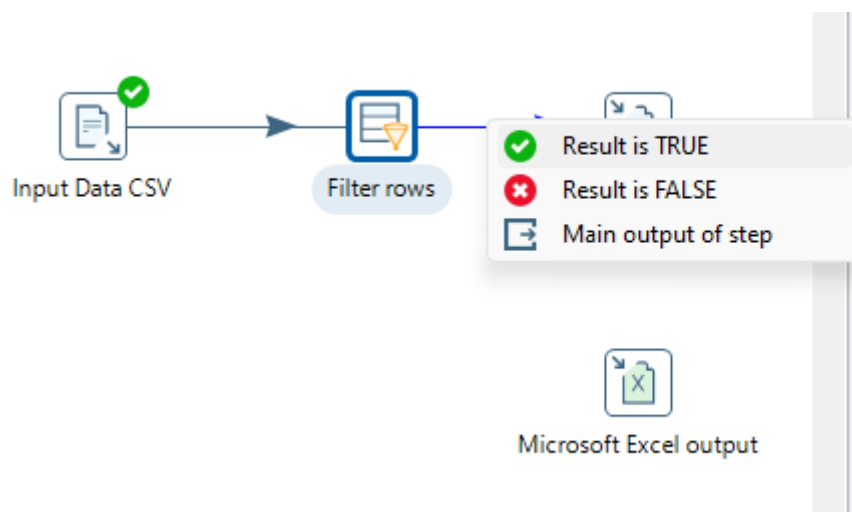
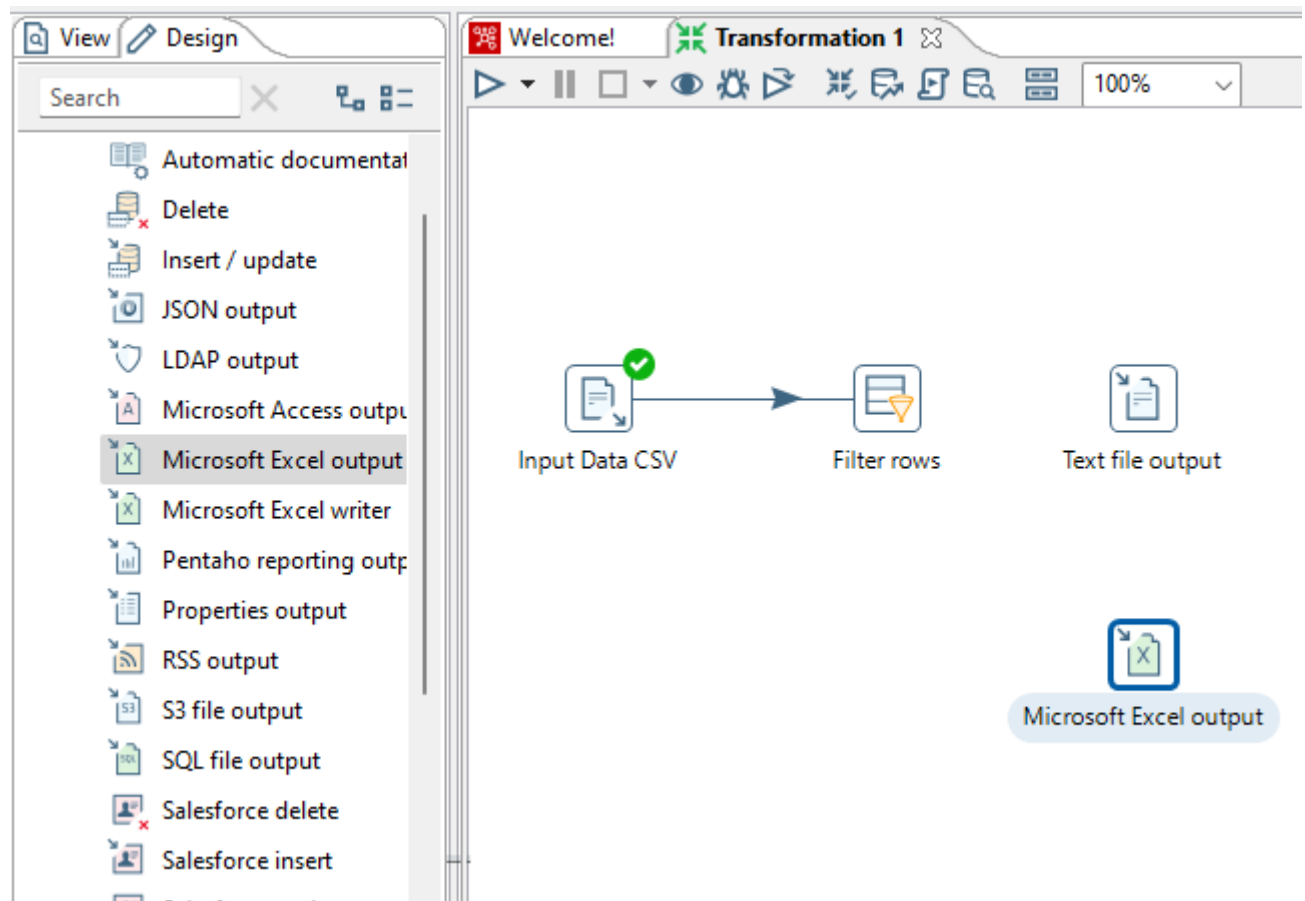
- **I (Input)** = Jumlah data yang dibaca atau dimasukkan ke dalam proses
- **O (Output)** = Jumlah data yang dikeluarkan atau ditulis ke tujuan
- **R (Read)** = Jumlah record yang dibaca dari sumber data
- **W (Written)** = Jumlah record yang berhasil ditulis ke target atau tahap berikutnya
- **U (Updated)** = Jumlah record yang diperbarui dalam database atau sistem target
- **E (Error)** = Jumlah error yang terjadi selama proses

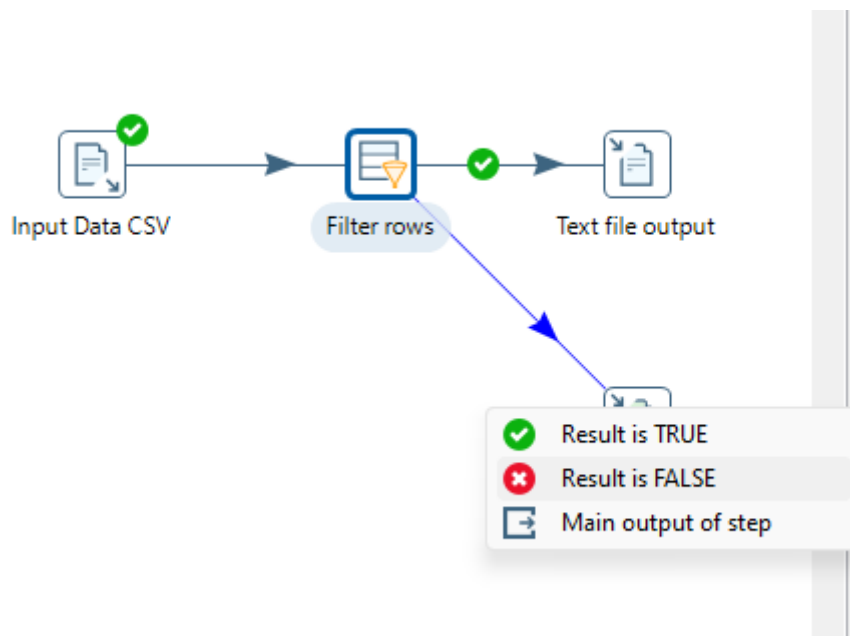
4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

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









Filter rows

Step name: Filter data

Send 'true' data to step: Text file output

Send 'false' data to step: Microsoft Excel output

The condition:

[] = [] <field>

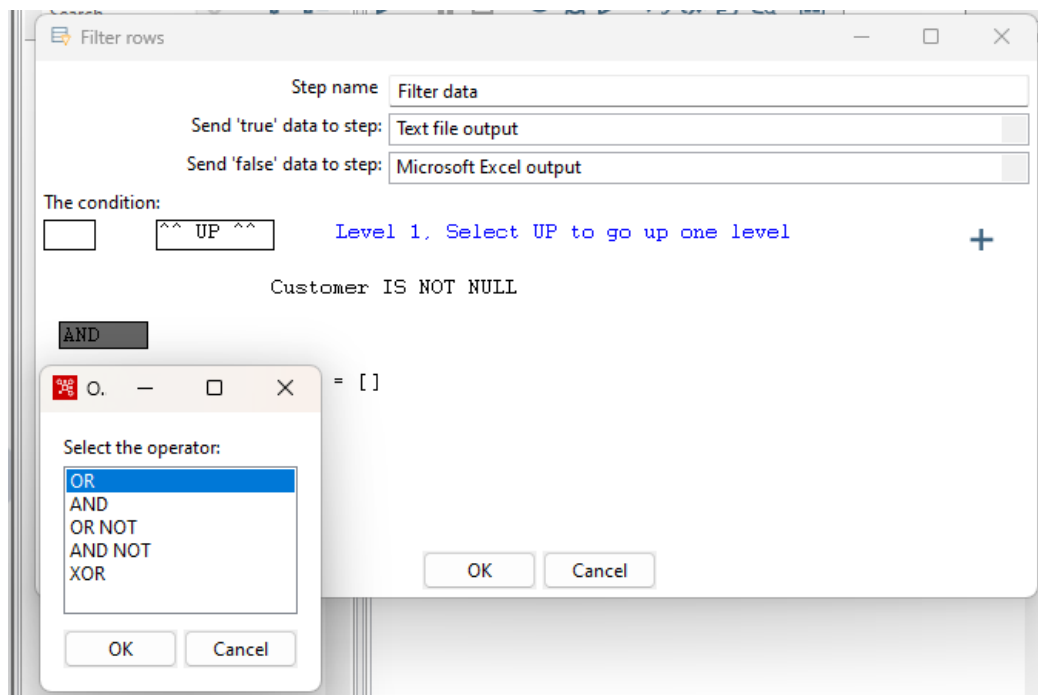
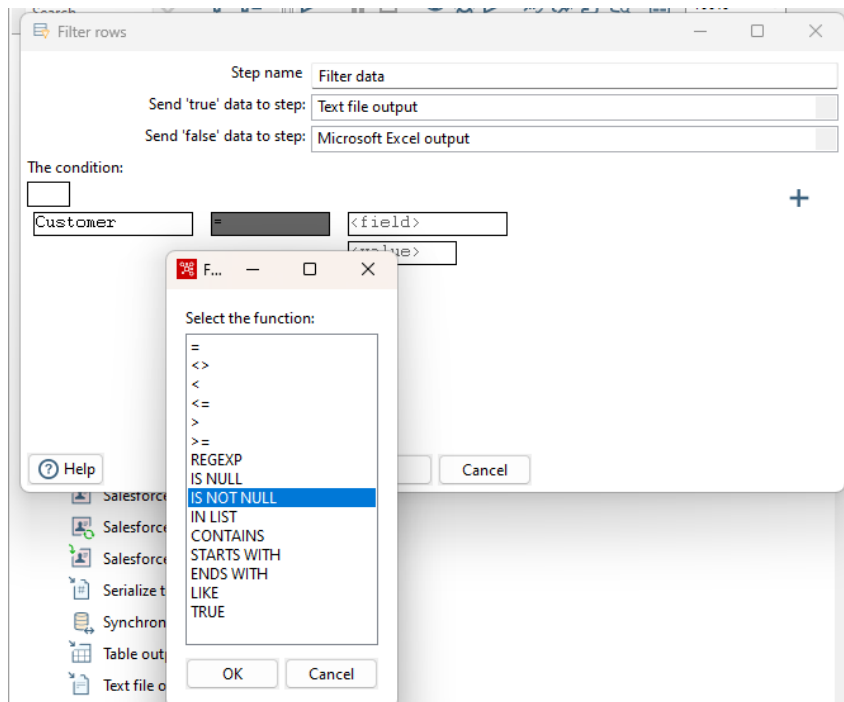
[] <value>

OK Cancel

Select a field:

Field	Type
Customer	(String)
Product	(String)
Region	(String)
Date	(Date)
Item unit price	(Number)
No.Items	(Integer)
Total Sale	(Number)

OK Cancel



Filter rows

Step name

Filter data

Send 'true' data to step:

Text file output

Send 'false' data to step:

Microsoft Excel output

The condition:

^^ UP ^^

Level 3, Select UP to go up one level

+

<field>

=

<field>

<value>

Help

OK

Cancel

Text file output

Step name

Text file output

File

Content

Fields

Filename

E:/KELAS 2/SEMESTER 4/Data Warehouse/PDI/outputSalesCarDistributionSales.csv

Browse...

Pass output to servlet

☐

Create Parent folder

☒

Do not create file at start

☐

Accept file name from field?

☐

File name field

Extension

csv

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)...

Add filenames to result

☒

Help

OK

Cancel

Microsoft Excel output

Step name: Microsoft Excel output

File Content Custom Fields

Filename: E:/KELAS 2/SEMESTER 4/Data Warehouse/PDI/failedSalesCarDistributionSales.csv

Create Parent folder: ☐

Do not create file at start: ☐

Extension: xls

Include stepnr in filename?: ☐

Include date in filename?: ☐

Include time in filename?: ☐

Specify Date time format: ☐

Date time format:

Show filename(s)...

Add filenames to result: ☒

OK Cancel

Microsoft Excel output

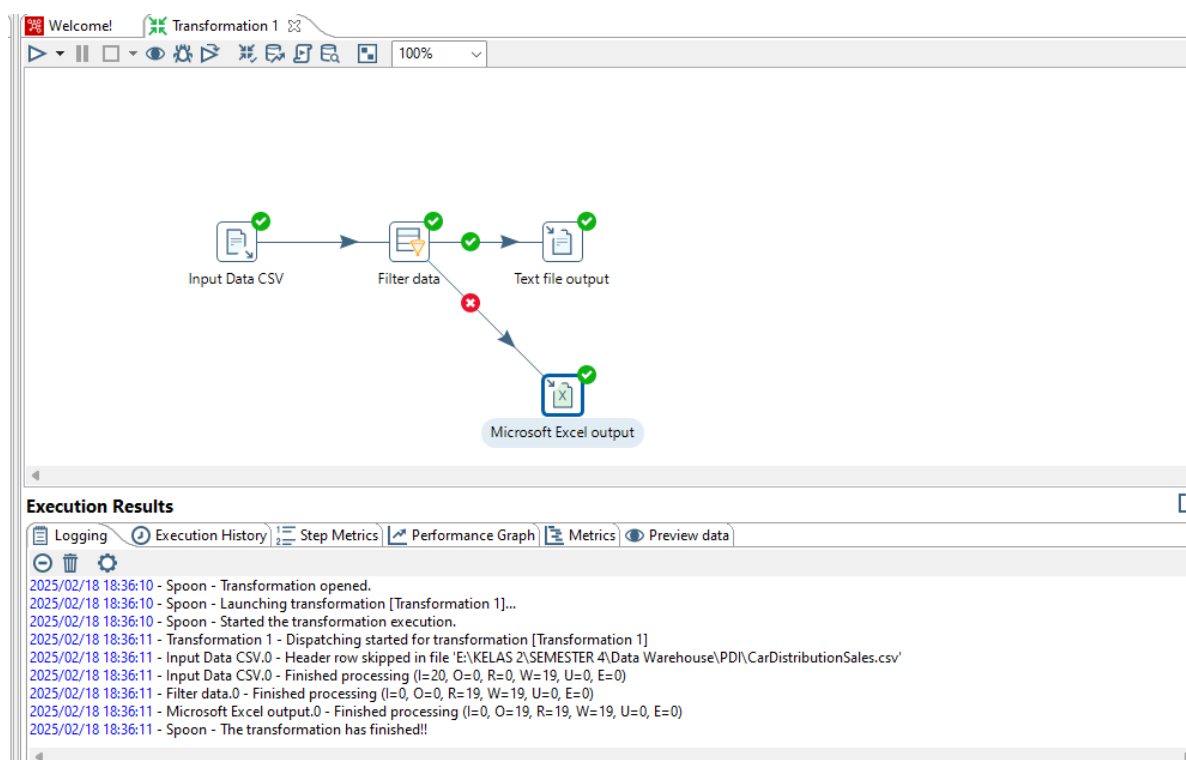
Step name: Microsoft Excel output

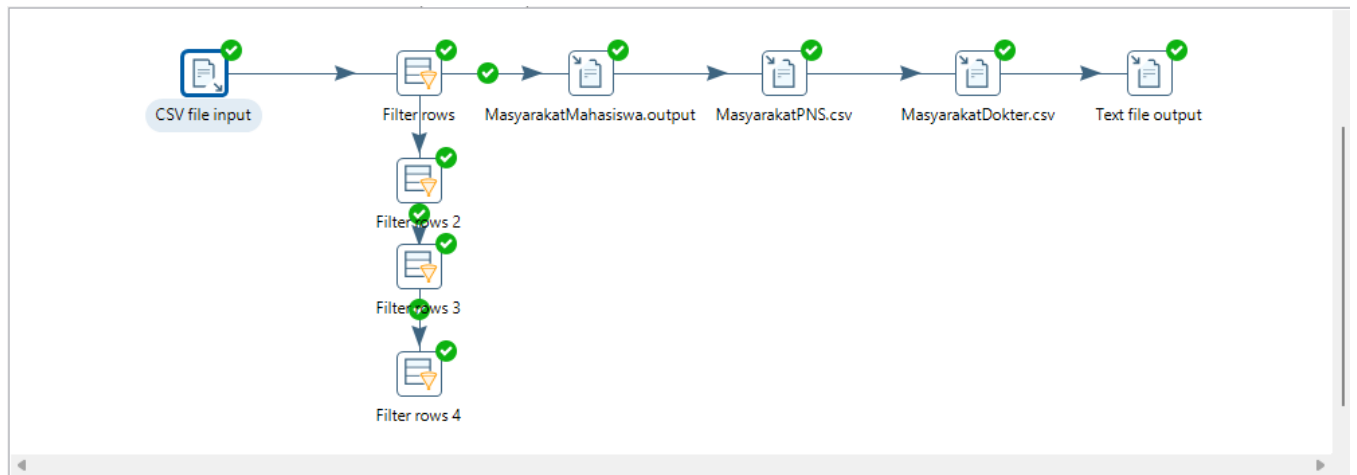
File Content Custom Fields

#	Name	Type	Format
1	Customer	String	
2	Product	String	
3	Region	String	
4	Date	Date	
5	Item unit price	Number	0000.00
6	No.Items	Integer	
7	Total Sale	Number	00000.00

Get Fields Minimal width

Help OK Cancel





Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

☒ First rows ☐ Last rows ☐ Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 ICT 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001,
2	Siti Rahayu	3503010101010002	Bandung	Tue May 15 00:00:00 ICT 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003,
3	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 ICT 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005,
4	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 ICT 1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007,
5	Adi Pratama	3503010101010005	Semarang	Sun Mar 05 00:00:00 ICT 1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009,
6	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 ICT 1982	Perempuan	Jl. Simpanan Borobudur No. 15	Malang	011,

Studi kasus

Disclaimer: Data yang tercantum di studi kasus ini adalah fiktif dan hanya digunakan untuk tujuan pembelajaran saja. Nama tempat, tanggal lahir, alamat, dan informasi lainnya yang tertera hanya tipuan belaka dan tidak memiliki kaitan dengan individu yang sebenarnya.

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

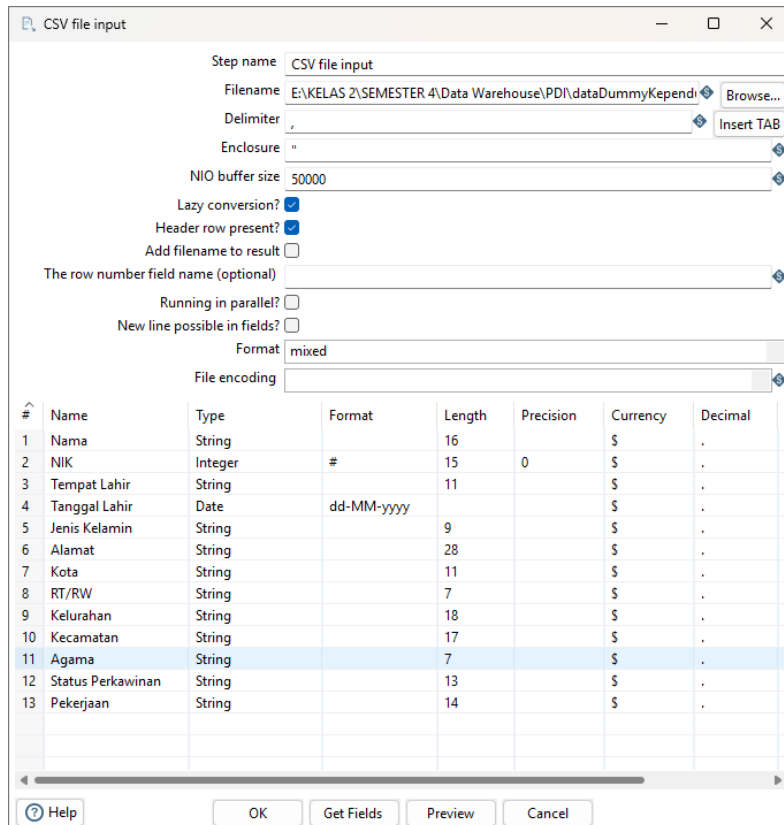
data-data tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk membantu mengambil keputusan.

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

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

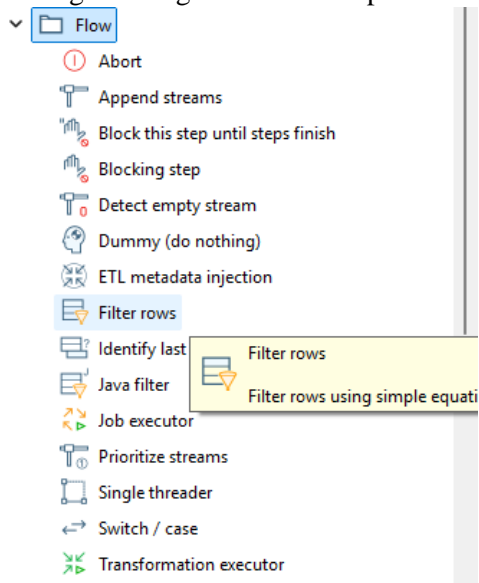
Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.

1. Menambahkan CSV File Input, masukkan file **dataDummyKependudukan.csv**, pilih **Get Fields** untuk membaca struktur kolom.



#	Name	Type	Format	Length	Precision	Currency	Decimal
1	Nama	String		16		\$.
2	NIK	Integer	#	15	0	\$.
3	Tempat Lahir	String		11		\$.
4	Tanggal Lahir	Date	dd-MM-yyyy			\$.
5	Jenis Kelamin	String		9		\$.
6	Alamat	String		28		\$.
7	Kota	String		11		\$.
8	RT/RW	String		7		\$.
9	Kelurahan	String		18		\$.
10	Kecamatan	String		17		\$.
11	Agama	String		7		\$.
12	Status Perkawinan	String		13		\$.
13	Pekerjaan	String		14		\$.

2. Menambahkan "Filter Rows", Dari panel Design, pilih **Flow** -> **Filter Rows**, Seret ke area kerja dan hubungkan dengan CSV File Input



3. Membuat **4 Filter Rows** untuk masing-masing kategori
 - **Filter Mahasiswa** (Profesi = 'Mahasiswa')

Filter rows

Step name: Filter rows

Send 'true' data to step: MasyarakatMahasiswa.output

Send 'false' data to step:

The condition:

☐ Pekerjaan = Mahasiswa (String)

Help OK Cancel

- **Filter PNS (Profesi = 'PNS')**

Filter rows

Step name: Filter rows 2

Send 'true' data to step: MasyarakatPNS.csv

Send 'false' data to step:

The condition:

☐ Pekerjaan = PNS (String)

Help OK Cancel

- **Filter Dokter (Profesi = 'Dokter')**

Filter rows

Step name: Filter rows 3

Send 'true' data to step: MasyarakatDokter.csv

Send 'false' data to step:

The condition:

☐ Pekerjaan = Dokter (String)

Help OK Cancel

- **Filter Lainnya (jika Profesi bukan Mahasiswa, PNS, atau Dokter)**

Filter rows

Step name: Filter rows 4

Send 'true' data to step: MasyarakatLainnya.csv

Send 'false' data to step:

The condition:

<field> = <field> +

<value>

Help OK Cancel

4. Menambahkan **Text File Output** untuk masing-masing Filter Rows

- **Hubungkan ke "Text File Output" bernama MasyarakatMahasiswa.csv**

Text file output

Step name: MasyarakatMahasiswa.output

File Content Fields

Filename: E:\KELAS 2\SEMESTER 4\Data Warehouse\PD\dataDummyKependudukan.csv Browse...

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field? ☐

File name field:

Extension: txt

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)...

Add filenames to result ☒

Help OK Cancel

- **Hubungkan ke "Text File Output" bernama MasyarakatPNS.csv**

Text file output

Step name: MasyarakatPNS.csv

File Content Fields

Filename: E:\KELAS 2\SEMESTER 4\Data Warehouse\PD\dataDummyKepen Browse...

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field? ☐

File name field:

Extension: txt

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)...

Add filenames to result ☒

Help OK Cancel

- Hubungkan ke **"Text File Output"** bernama MasyarakatDokter.csv

The screenshot shows the 'Text file output' configuration window with the 'Step name' set to 'MasyarakatDokter.csv'. The 'File' tab is selected, showing the following settings:

- Filename: E:\KELAS 2\SEMESTER 4\Data Warehouse\PD\dataDummyKepen (with a 'Browse...' button)
- Pass output to servlet: ☐
- Create Parent folder: ☒
- Do not create file at start: ☐
- Accept file name from field: ☐
- File name field: (empty dropdown)
- Extension: txt
- Include stepnr in filename?: ☐
- Include partition nr in filename?: ☐
- Include date in filename?: ☐
- Include time in filename?: ☐
- Specify Date time format: ☐
- Date time format: (empty dropdown)
- Show filename(s)...: (button)
- Add filenames to result: ☒

At the bottom, there are 'Help', 'OK', and 'Cancel' buttons.

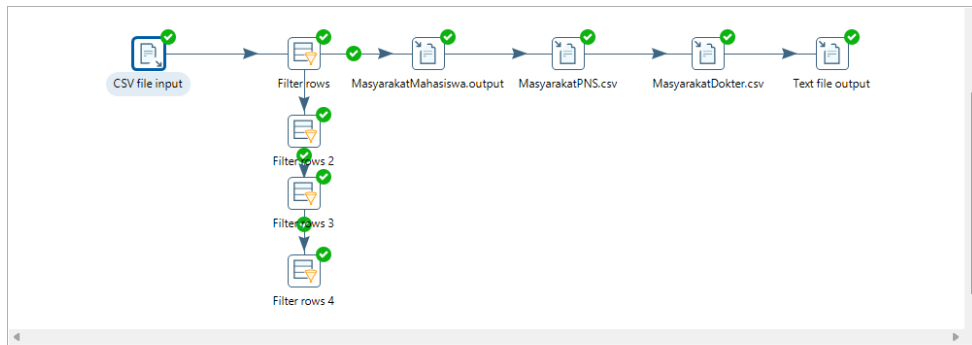
- Hubungkan ke **"Text File Output"** bernama MasyarakatLainnya.csv

The screenshot shows the 'Text file output' configuration window with the 'Step name' set to 'MasyarakatLainnya.csv'. The 'File' tab is selected, showing the following settings:

- Filename: E:\KELAS 2\SEMESTER 4\Data Warehouse\PD\dataDummyKepen (with a 'Browse...' button)
- Pass output to servlet: ☐
- Create Parent folder: ☒
- Do not create file at start: ☐
- Accept file name from field: ☐
- File name field: (empty dropdown)
- Extension: txt
- Include stepnr in filename?: ☐
- Include partition nr in filename?: ☐
- Include date in filename?: ☐
- Include time in filename?: ☐
- Specify Date time format: ☐
- Date time format: (empty dropdown)
- Show filename(s)...: (button)
- Add filenames to result: ☒

At the bottom, there are 'Help', 'OK', and 'Cancel' buttons.

5. Melakukan Run



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

☒ First rows ☐ Last rows ☐ Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/R
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 ICT 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001,
2	Siti Rahayu	3503010101010002	Bandung	Tue May 15 00:00:00 ICT 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003,
3	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 ICT 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005,
4	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 ICT 1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007,
5	Adi Pratama	3503010101010005	Semarang	Sun Mar 05 00:00:00 ICT 1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009,
6	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 ICT 1982	Perempuan	Jl. Simanung Roroburur No. 15	Malang	011,