

**LAPORAN PRAKTIKUM
DATA WAREHOUSE**

**JOB SHEET 1
DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON**



Oleh:

DHEVINA AGUSTINA

2341760065

SIB 2B

PROGRAM STUDI SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025/2026

TUGAS 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Jawaban: 7 kolom yaitu, Customer,Product,Region,Date, Item unit price , No.Items , dan Total Sale

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

Jawaban:

- a. Customer: berisi informasi customer
- b. Product: menunjukan nama produk yang dibeli oleh pelanggan
- c. Region: menunjukan wilayah atau tempat transaksi terjadi
- d. Date: menunjukan tanggal transaksi pembelian dalam format MM/DD/YYYY
- e. Item unit price: menunjukan harga satuan dari produk yang dibeli
- f. No. Items: menunjukkan no.items yang dibeli
- g. Total sale: total harga pembelian dalam transaksi tersebut

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

Jawaban: Ya, ada beberapa baris yang memiliki nilai kosong (null)

- Baris ke 4 memiliki kolom Product kosong.
- Baris ke 8 memiliki kolom Region dan Date kosong.
- Baris ke 13 memiliki beberapa kolom kosong, termasuk Product, Region, dan Date.
- Baris ke 17 memiliki kolom Product kosong
- Baris ke 19 memiliki kolom Region kosong

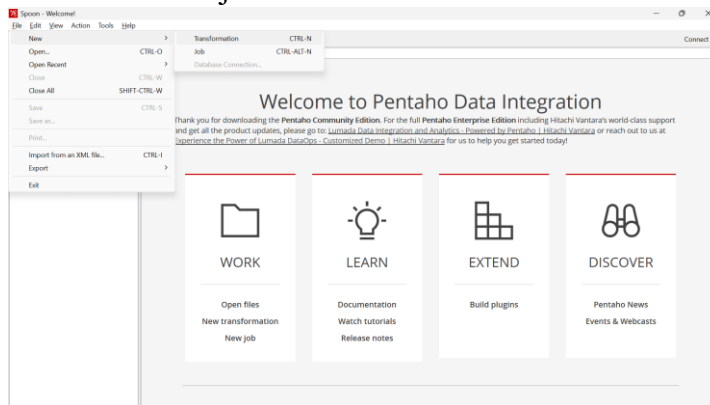
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Jawaban:

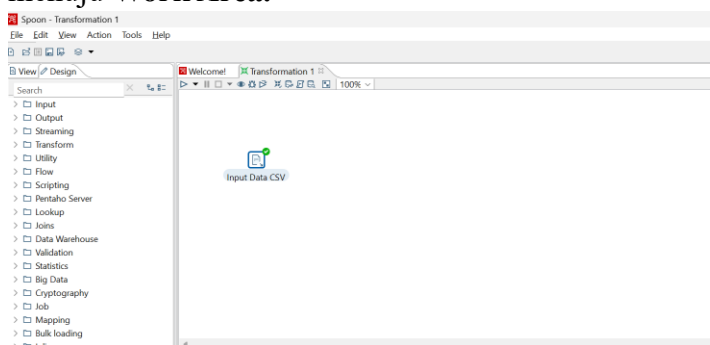
- Kolom No. Items seharusnya berisi angka, tetapi ada kemungkinan ada teks atau nilai kosong.
- Kolom Date seharusnya memiliki format tanggal (MM/DD/YYYY), tetapi beberapa baris terlihat memiliki format berbeda atau kosong.
- Kolom Total Sale juga bisa mengandung data yang salah hitung jika ada kesalahan format.

A. Pengambilan Data (Extract)

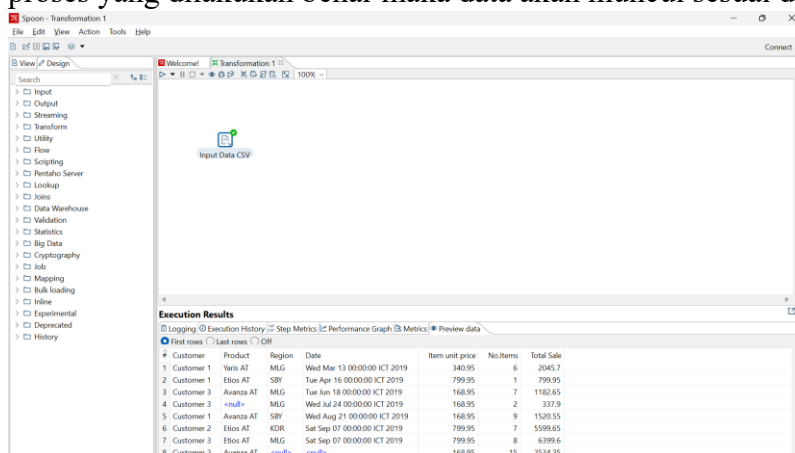
1. Buka lembar kerja Transformation melalui toolbar File.



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



3. Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.
4. Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
5. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.
6. Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , ").
7. Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai
8. Tutup jendela konfigurasi tersebut.
9. Tekan tombol "Run" pada pojok kanan atas Area Kerja, maka akan muncul Execution Result Area. Pilih tab Preview Data pada Execution Result Area. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.



TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawaban: Sesuai

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:

- Memulai Eksekusi Transformasi
- Membuka koneksi ke sumber data
- Membaca data
- Menampilkan jumlah data yang di extract

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 masuk = 20
- O (Output): jumlah data yang ditulis ke output = 0
- R (Rejected): jumlah data yang ditolak/gagal diproses = 0
- W (Written): jumlah data yang diteruskan ke Langkah berikutnya = 19
- U (Update): jumlah data yang diperbarui = 0
- E (Error): jumlah data yang gagal di proses karena error = 0

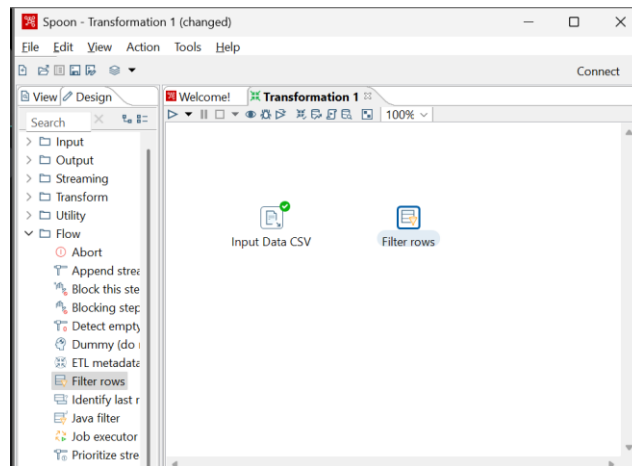
4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

Jawaban:

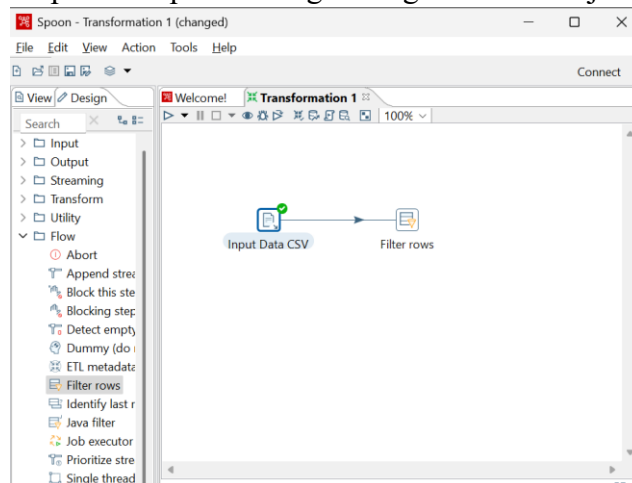
Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152 GB	Inter(R) UHD Graphics	12th Gen Intel(R) Core(TM) i5 - 1235U	0.0s
Putera	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	16,0 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0.1s
Daffa	AMD A9-9425 RADEON R5	8,0 GB	1,2 TB	AMD RADEON (TM) R5 Graphics	AMD A9-9425 RADEON R5	0.3s
Nimas	AMD Ryzen 3 5300U	16,0 GB	475 GB	AMD Radeon Graphics	AMD Ryzen 3 5300U with Radeon Graphics.	0.1s

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

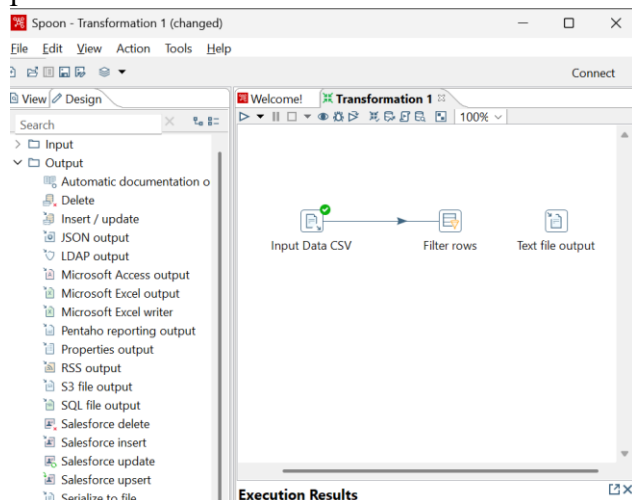
1. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area



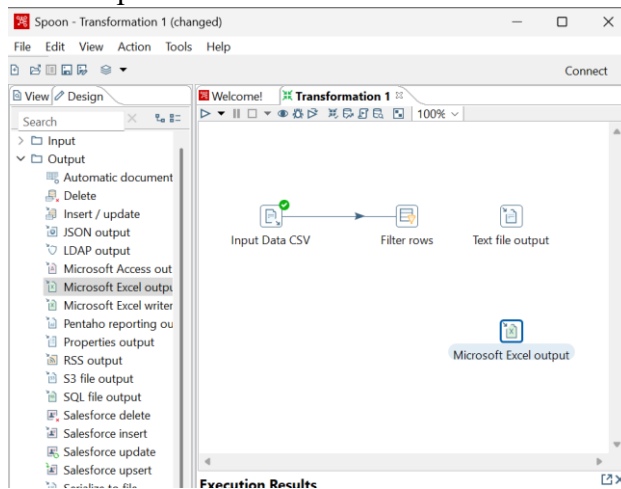
2. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.



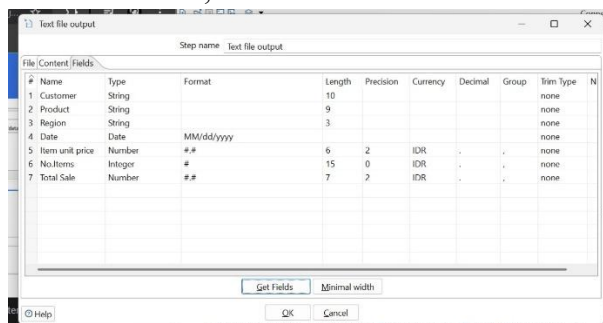
3. Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.



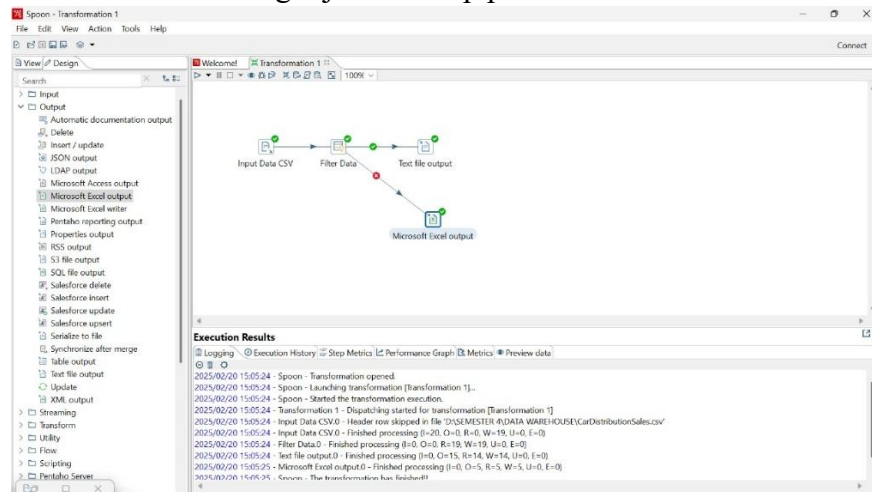
- Carilah objek Microsoft excel output pada Design Area. Drag and drop objek tersebut pada Work Area.



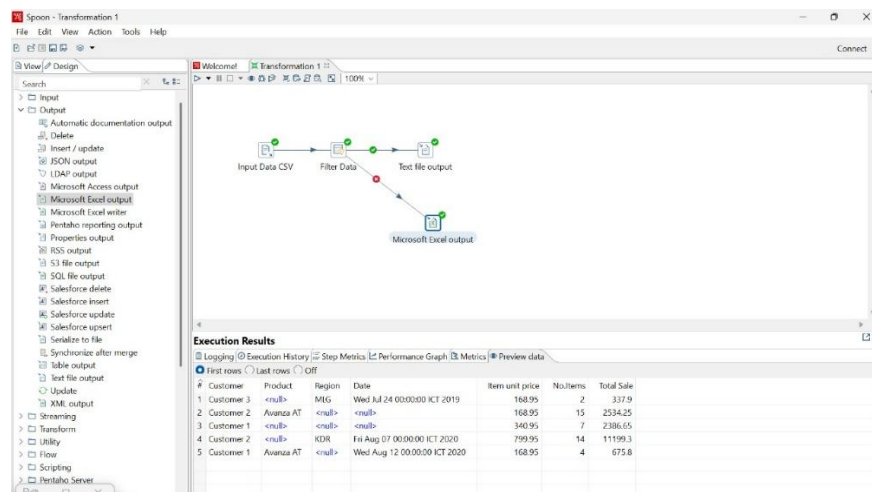
- Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagai jenis konektornya.
- Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagai jenis konektornya.
- Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.
- Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
- Klik <Field> pada bagian condition
- Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.
- Pada bagian value pilih IS NOT NULL kemudian tekan OK , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null
- Tekan tombol + untuk menambah kondisi.
- Pilih statement yang masih null untuk menambah kondisi.
- Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
- Gunakan operator AND untuk menghubungkan
- Jika semua field sudah masuk dalam kondisi maka tekan OK.
- Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.
- Ubah extension menjadi csv.
- Pada tab Fields, tekan tombol Get fields untuk mengambil field data.



20. Tekan ok untuk kembali ke work area
21. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan extension xls.
23. Pada tab fields tekan tombol Get fields untuk mengambil field data.
24. Tekoan ok untuk kembali ke work area.
25. Tekan tombol **Run** pada pojok kiri atas **Work Area**. jika berhasil maka akan muncul tanda centang hijau di setiap proses.



Proses berhasil dijalankan



failedSalesCarDistribution

Execution Results

Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
Customer 1	Yaris AT	MLG	Wed Mar 11 00:00:00 ICT 2019	340.9	6	2045.7
Customer 1	Elios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
Customer 1	Avanza AT	MLG	Tue Jan 18 00:00:00 ICT 2019	168.9	7	1182.63
Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
Customer 2	Elios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
Customer 3	Elios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.9	15	2534.25
Customer 1	Elios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9

outputSalesCarDistribution

TUGAS 3

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

Jawaban: File CSV adalah hasil extract awal sebelum dilakukan transformasi data, sedangkan file Excel sudah melewati proses pembersihan data

2. Jelaskan apa yang terjadi pada proses Filter rows!

Jawaban:

- 1) Input data: data dari sumber (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 memenuhi kriteria akan diteruskan ke langkah berikutnya dalam transformasi ETL.
 - 4) Output data: Hasil dari Filter Rows biasanya menghasilkan subset data yang lebih bersih dan relevan untuk analisis atau proses lebih lanjut.
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:

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152 GB	Inter(R) UHD Graphics	12th Gen Intel(R) Core(TM) i5 - 1235U	0.6s
Nimas	AMD Ryzen 3 5300U	16,0 GB	475 GB	AMD Radeon Graphics	AMD Ryzen 3 5300U with Radeon Graphics.	0.4s
Putera	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	16,0 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0.7s
Daffa	AMD A9-9425 RADEON R5	8,0 GB	1,2 TB	AMD RADEON (TM) R5 Graphics	AMD A9-9425 RADEON R5	0.12s

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

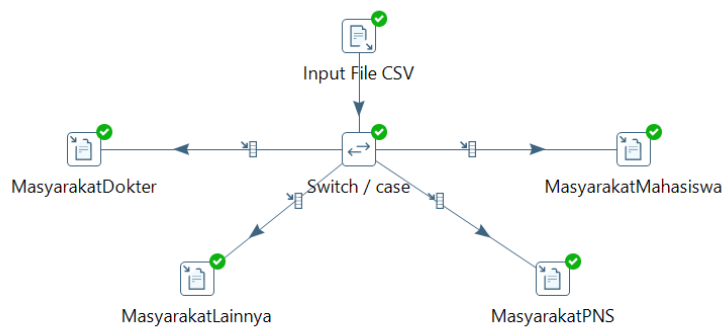
Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui link:

<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:

Berikut adalah sistem yang dibuat untuk membagi data masyarakat berdasarkan pekerjaannya.



1. Masyarakat Mahasiswa

Pada masyarakat mahasiswa terdapat 3 data

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan
1	Sei Rahayu	3503010101010002	Bandung	Tue May 15 00:00:00 ICT 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah
2	Adi Pratama	3503010101010005	Semarang	Sun May 05 00:00:00 ICT 1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandawaran	Semarang Tengah	Buddha	Belum Menikah
3	Maya Wulandari	3503010101010020	Jayapura	Sat Oct 23 00:00:00 ICT 1993	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah

2. Masyarakat PNS

Pada masyarakat PNS hanya terdapat 2 data

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan
1	Andi Wijaya	3503010101010007	Makassar	Sun Jun 25 00:00:00 ICT 1978	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar	013/014	Barombong	Makassar Selatan	Islam	Menikah
2	Agus Santoso	3503010101010015	Padang	Fri Apr 02 00:00:00 ICT 1982	Laki-laki	Jl. Sudirman No. 20	Padang	029/030	Padang Barat	Padang	Islam	Menikah

3. Masyarakat Dokter

Pada masyarakat Dokter hanya terdapat 2 data

The screenshot shows a data flow diagram with 'Input File CSV' as the source, branching into 'MasyarakatDokter', 'MasyarakatLainnya', 'MasyarakatMahasiswa', and 'MasyarakatPNS'. Below the diagram is the 'Execution Results' table.

#	Name	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama
1	Desi Purnama	3503010101010008	Palembang	Thu Apr 08 00:00:00 ICT 1993	Perempuan	Jl. Demang Lebar Daun No. 20	Palembang	015/016	Alang-Alang Lebar	Palembang Selatan	Kristen
2	Bambang Kusumo	3503010101010017	Banjarmasin	Thu Sep 25 00:00:00 ICT 1980	Laki-laki	Jl. A. Yani No. 60	Banjarmasin	033/034	Banjarmasin Tengah	Banjarmasin	Hindu

4. Masyarakat Lainnya

Pada masyarakat lainnya terdapat 13 data dengan pekerjaan Akuntan, Penulis, Wiraswasta, Pegawai Swasta, Guru dan Freelancer

The screenshot shows a data flow diagram with 'Input File CSV' as the source, branching into 'MasyarakatDokter', 'MasyarakatLainnya', 'MasyarakatMahasiswa', and 'MasyarakatPNS'. Below the diagram is the 'Execution Results' table.

#	Name	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 ICT 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng	Jakarta Pusat	Islam	Menik
2	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 ICT 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng	Surabaya Pusat	Katolik	Menik
3	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 ICT 1988	Perempuan	Jl. Maloboro No. 8	Yogyakarta	007/008	Sosromenduran	Yogyakarta Utara	Hindu	Belur
4	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 ICT 1982	Perempuan	Jl. Simpang Borobudur No. 15	Malang	011/012	Dinoyo	Malang Selatan	Kristen	Menik
5	Fajar Prabowo	3503010101010009	Medan	Tue Sep 20 00:00:00 ICT 1986	Laki-laki	Jl. Sudirman No. 30	Medan	017/018	Sat Sikambing	Medan Timur	Hindu	Menik
6	Rina Sari	3503010101010010	Dempasar	Fri Jul 17 00:00:00 ICT 1992	Perempuan	Jl. Gatot Subroto No. 30	Dempasar	019/020	Dauh Puri Kauh	Dempasar Utara	Buddha	Belur
7	Ahmad Subhan	3503010101010011	Surakarta	Wed Jul 03 00:00:00 ICT 1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	Surakarta Timur	Islam	Belur
8	Dian Pertiwi	3503010101010012	Maluku	Tue Dec 10 00:00:00 ICT 1985	Perempuan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	Ambon	Kristen	Menik
9	Rudi Hartono	3503010101010013	Pontianak	Mon May 28 00:00:00 ICT 1973	Laki-laki	Jl. Ahmad Yani No. 30	Pontianak	025/026	Pontianak Barat	Pontianak Kota	Hindu	Menik
10	Lina Fitriani	3503010101010014	Manado	Mon Aug 15 00:00:00 ICT 1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapanget	Manado Selatan	Kristen	Belur
11	Rina Agustina	3503010101010016	Bengkulu	Mon Nov 19 00:00:00 ICT 1990	Perempuan	Jl. Inam Bontol No. 45	Bengkulu	031/032	Kuta Sembilan	Bengkulu Tengah	Kristen	Menik
12	Nia Dewi Lastati	3503010101010018	Manokwari	Sun Jun 14 00:00:00 ICT 1987	Perempuan	Jl. Diponegoro No. 40	Manokwari	035/036	Manokwari Barat	Manokwari	Kristen	Belur
13	Doni Setiawan	3503010101010019	Palu	Mon Mar 07 00:00:00 ICT 1977	Laki-laki	Jl. Soekarno-Hatta No. 70	Palu	037/038	Palu Barat	Palu	Islam	Menik