

LAPORAN PRAKTIKUM
DATA WAREHOUSE

JOBSHEET 1

DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON



Oleh:

Syifa Revalina Kamila

2341760065

SIB 2B

PROGRAM STUDI SISTEM INFORMASI BISNIS
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG 2025/2026

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.

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.

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
2	Customer 1	Yaris AT	MLG	3/13/2019	340.95	6	2045.7
3	Customer 1	Etos AT	SBY	4/16/2019	799.95	1	799.95
4	Customer 3	Avanza AT	MLG	6/18/2019	168.95	7	1182.65
5	Customer 3		MLG	7/24/2019	168.95	2	337.9
6	Customer 1	Avanza AT	SBY	8/21/2019	168.95	9	1520.55
7	Customer 2	Etos AT	KDR	9/7/2019	799.95	7	5599.65
8	Customer 2	Etos AT	MLG	9/7/2019	799.95	8	6399.6
9	Customer 2	Avanza AT			168.95	15	2534.25
10	Customer 1	Avanza AT	SBY	9/14/2019	168.95	15	2534.25
11	Customer 1	Etos AT	SBY	9/23/2019	799.95	2	1599.9
12	Customer 1	Avanza AT	SBY	1/21/2020	168.95	14	2385.3
13	Customer 1	Yaris AT	MLG	2/1/2020	340.95	2	681.9
14	Customer 1				340.95	7	2386.65
15	Customer 1	Etos AT	SBY	5/30/2020	799.95	9	7199.55
16	Customer 3	Avanza AT	MLG	6/7/2020	168.95	10	1689.5
17	Customer 3	Yaris AT	MLG	6/9/2020	340.95	4	1363.8
18	Customer 2		KDR	8/7/2020	799.95	14	11199.3
19	Customer 2	Avanza AT	KDR	8/8/2020	168.95	12	2027.4
20	Customer 1	Avanza AT		8/12/2020	168.95	4	675.8

TUGAS 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Kolom pada data memiliki 7 koolom, yaitu:

- Customer – Nama pelanggan atau ID pelanggan.
- Product – Produk yang dijual (jenis mobil).
- Region – Wilayah atau cabang tempat penjualan dilakukan.
- Date – Tanggal transaksi.
- Item unit price – Harga satuan per item.
- No. Items – Jumlah unit mobil yang terjual dalam transaksi.
- Total Sale – Total hasil penjualan (dihitung sebagai Item unit price × No. Items).

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

- Customer: Mengidentifikasi pelanggan dalam bentuk teks (misalnya "Customer 1", "Customer 2").

- Product: Jenis mobil yang dibeli oleh pelanggan, seperti "Yaris AT", "Avanza AT", dll.
- Region: Cabang lokasi transaksi terjadi, seperti "MLG" (Malang), "SBY" (Surabaya), "KDR" (Kediri).
- Date: Tanggal transaksi dilakukan dalam format MM/DD/YYYY.
- Item unit price: Harga satu unit mobil dalam bentuk angka desimal.
- No. Items: Jumlah unit mobil yang dibeli dalam satu transaksi (bilangan bulat).
- Total Sale: Total nilai penjualan dalam transaksi tersebut (bilangan desimal).

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

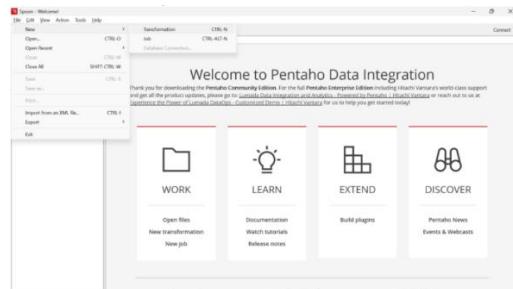
Untuk memastikan adanya nilai kosong atau data yang hilang (missing data), diperlukan pengecekan langsung pada setiap baris dalam dataset. Berdasarkan pengamatan awal terhadap gambar, tidak terlihat adanya sel kosong pada tabel. Namun, agar analisis lebih akurat, tetap diperlukan pemeriksaan lebih lanjut guna memastikan bahwa setiap baris memiliki nilai yang valid dan sesuai dengan format yang seharusnya. Pengecekan ini mencakup validasi terhadap kolom-kolom penting seperti tanggal transaksi, harga satuan, jumlah item, dan total penjualan untuk memastikan tidak ada kesalahan atau inkonsistensi dalam data.

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

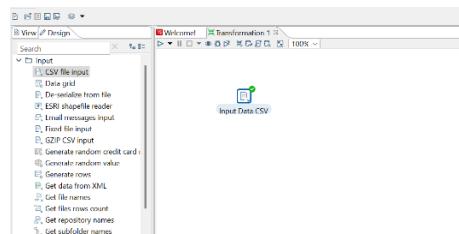
- Customer – Semua data berbentuk teks (string).
- Product – Semua data berbentuk teks (string).
- Region – Semua data berbentuk teks (string).
- Date – Harusnya berbentuk tanggal, perlu dicek apakah semua data benar dalam format yang sama.
- Item unit price – Harus berbentuk angka desimal, perlu dicek apakah ada nilai non-numerik.
- No. Items – Harus berupa bilangan bulat, perlu dicek apakah ada karakter selain angka.
- Total Sale – Harus berupa angka desimal, perlu dicek apakah ada format yang berbeda atau salah hitung.

A. Pengambilan Data (Extract)

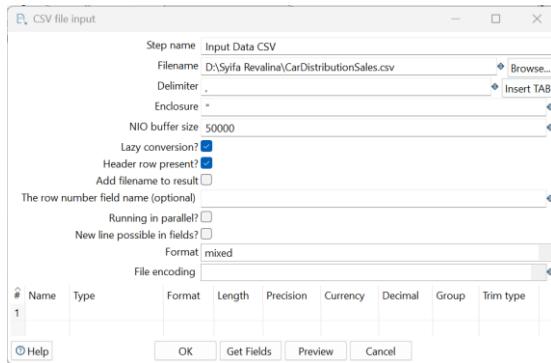
1. Buka lembar kerja Transformation melalui toolbar File.



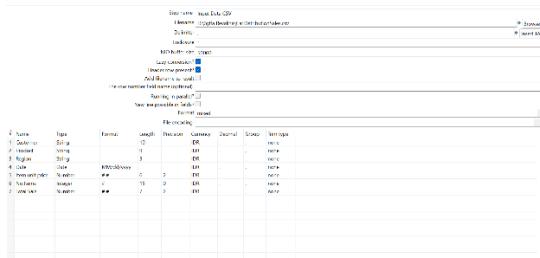
2. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.



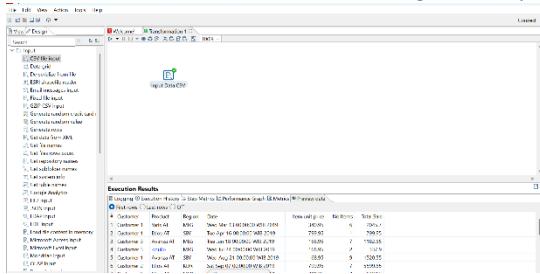
3. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasinya.
4. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
5. Pilih lokasi file **CarDistributionSales.csv** melalui **Browse** pada **filename**.



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 **Prievew 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? **Ya**
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?
 - Memulai Ekesekusi 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 ?
- I (Input): jumlah data yang masuk = 20
 - (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

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

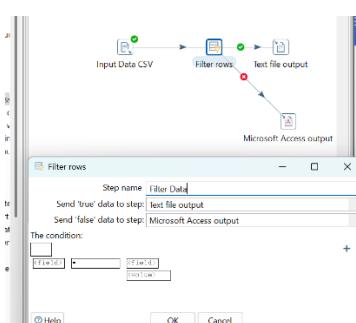
4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini?
 Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Saria	Intel(R) Core(TM) i7-12700H 2.30GHz	16,0 GB	512 GB	NVIDIA GeForce RTX 3050	Intel(R) Core(TM) i7-12700H 2.30GHz	0.0s
Alvi	Intel(R) Core(TM) i5-1135G7 2.40GHz	8,0 GB	256 GB	Intel(R) Iris Xe Graphics	Intel(R) Core(TM) i5-1135G7 2.40GHz	0.1s
Ulul	AMD Ryzen 5 5500U	8,0 GB	512 GB	AMD Radeon Vega 7 Graphics	AMD Ryzen 5 5500U	0.1s
Fauzani	AMD A6- 9220 RADEON R4	4,0 GB	1 TB	AMD RADEON (TM) R4 Graphics	AMD A6- 9220 RADEON R4	0.5s

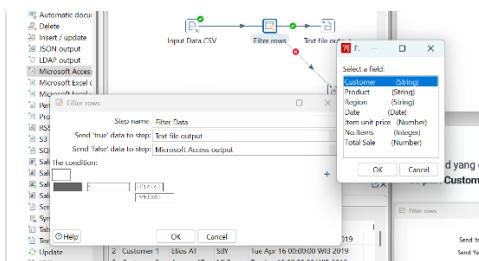
🍔 B. Filter Data (Transform) dan Pengemasan data (Load)

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. Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows.
3. Carilah objek **Text file output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.
4. Carilah objek **Microsoft excel output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.
5. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagai jenis konektornya.
6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagai jenis konektornya.

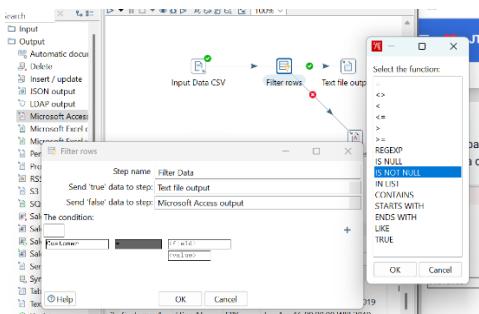
- Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.
- Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
- Klik pada bagian condition



- Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.



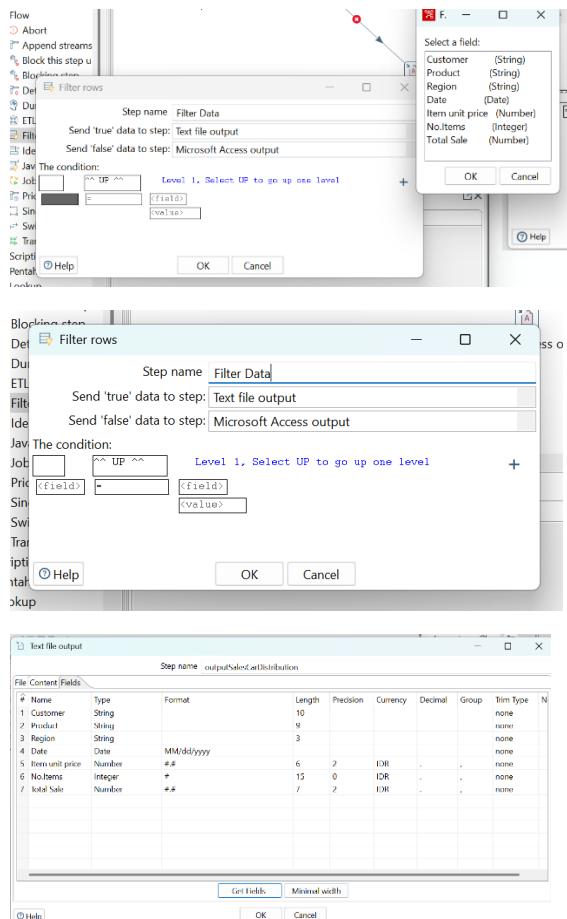
- Pada bagian value pilih IS NOT NULL kemudian tekan OK , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



- 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 menghubung
- 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.

19. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.

20. Tekan ok untuk kembali ke work area

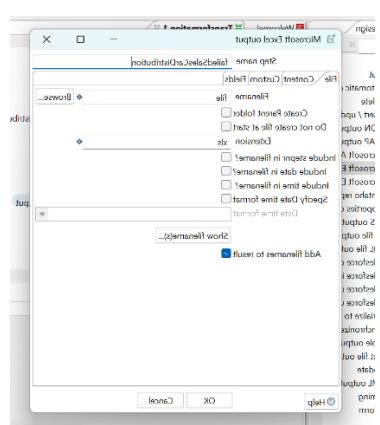


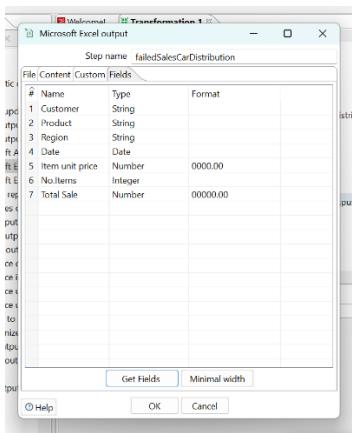
21. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output**, Double-click pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.

22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol **browse**. Beri nama **failedSalesCarDistribution**. untuk extension tetap dengan extension **xls**.

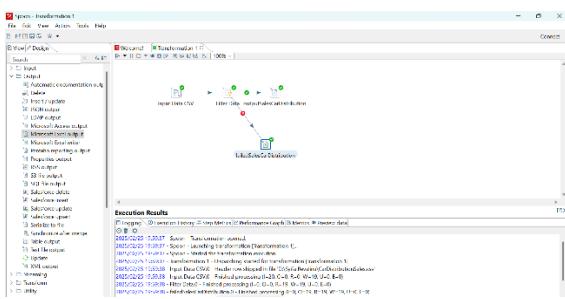
23. Pada tab **fields** tekan tombol **Get fields** untuk mengambil field data.

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.



Customer	Product	Region	Date	Item unit	No.Items	Total Sale
Customer Yaris AT	MLG	3/13/2019	340.9		6	2045.7
Customer Etios AT	SBY	4/16/2019	800		1	800
Customer Avanza AT	MLG	6/18/2019	168.9		7	1182.7
Customer Avanza AT	SBY	8/21/2019	168.9		9	1520.5
Customer Etios AT	KDR		#####	800		7 5599.6
Customer Etios AT	MLG		#####	800		8 6399.6
Customer Avanza AT	SBY	9/14/2019	168.9		15	2534.2
Customer Etios AT	SBY	9/25/2019	800		2	1599.9
Customer Avanza AT	SBY	1/21/2020	168.9		14	2365.3
Customer Yaris AT	MLG		#####	340.9		2 681.9
Customer Etios AT	SBY	5/30/2020	800		9	7199.6
Customer Avanza AT	MLG		#####	168.9		10 1689.5
Customer Yaris AT	MLG		#####	340.9		4 1363.8
Customer Avanza AT	KDR		#####	168.9		12 2027.4

TUGAS 3

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

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!**

- Input data: data dari sumber (file) di masukkan ke dalam PDI.
- Menentukan kondisi penyaringan: Filter Rows digunakan untuk menentukan aturan atau kondisi yang harus dipenuhi oleh data agar bisa lolos ke tahap berikutnya. Pemrosesan data: Data yang memenuhi kriteria akan diteruskan ke langkah berikutnya dalam transformasi ETL.
- 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

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Saria	Intel(R) Core(TM) i7-12700H 2.30GHz	16,0 GB	512 GB	NVIDIA GeForce RTX 3050	Intel(R) Core(TM) i7-12700H 2.30GHz	0.2s
Alvi	Intel(R) Core(TM) i5-1135G7 2.40GHz	8,0 GB	256 GB	Intel(R) Iris Xe Graphics	Intel(R) Core(TM) i5-1135G7 2.40GHz	0.4s
Ulul	AMD Ryzen 5 5500U	8,0 GB	512 GB	AMD Radeon Vega 7 Graphics	AMD Ryzen 5 5500U	0.3s
Fauzani	AMD A6- 9220 RADEON R4	4,0 GB	1 TB	AMD RADEON (TM) R4 Graphics	AMD A6- 9220 RADEON R4	0.8s

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.

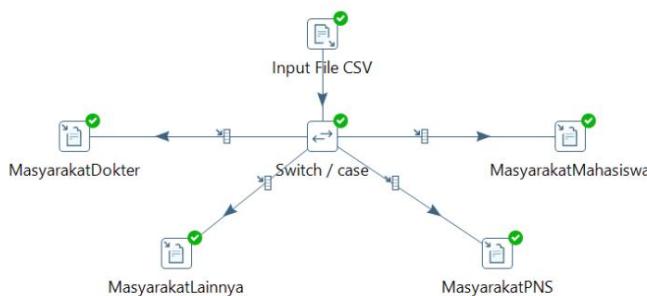
Data masyarakat berprofesi PNS akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan

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/82821b8ec341176f979956c84b79720d4ead012/dataDummyKependudukan.csv>

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



1. Masyarakat Mahasiswa

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan
1	Siti 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 Mar 05 00:00:00 ICT 1995	Laki-laki	Jl. Pernuda No. 21	Semarang	009/010	Pandanaran	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

First rows												
#	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 Santos	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

#	Nama	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	Barjarmasin	Thu Sep 25 00:00:00 ICT 1980	Laki-laki	Jl. A. Yani No. 60	Barjarmasin	033/034	Banjarmasin Tengah	Barjarmasin	Hindu

4. MasyarakatLainnya

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	Meni
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	Meni
3	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 ICT 1988	Perempuan	Jl. Malioboro 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	Meni
5	Fajar Prabowo	3503010101010009	Medan	Tue Sep 30 00:00:00 ICT 1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	Medan Timur	Hindu	Meni
6	Rina Sari	3503010101010010	Denpasar	Fri Jul 17 00:00:00 ICT 1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kauh	Denpasar 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	Jebrus	Surakarta Timur	Islam	Belur
8	Dian Pertwi	3503010101010012	Maluku	Tue Dec 10 00:00:00 ICT 1985	Perempuan	Jl. Pattimura No. 10	Amboin	023/024	Baguala	Amboin	Kristen	Meni
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	Meni
10	Lina Fitriani	3503010101010014	Manado	Mon Aug 15 00:00:00 ICT 1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapangket	Manado Selatan	Kristen	Belur
11	Rina Agustina	3503010101010016	Bengkulu	Mon Nov 19 00:00:00 ICT 1990	Perempuan	Jl. Imam Bonjol No. 45	Bengkulu	031/032	Ratu Samban	Bengkulu Tengah	Kristen	Meni
12	Nia Dewi Lestari	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	Meni