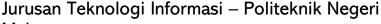
LAPORAN PRAKTIKUM DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON



Disusun Oleh:

Qusnul Diah Mawanti 2341760035

PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS JURUSAN TEKNOLOGI INFORMASI POLITEKNIK NEGERI MALANG 2025



Malang

Jobsheet 1 : Data Warehouse Sederhana dan Pengenalan

PDI Spoon

Mata Kuliah : Data Warehouse

Kelas: 2A - SIB

Nama: Qusnul Diah Mawanti

No absen: 25

NIM: 2341760035

Februari 2025

Tujuan

Setelah melakukan praktikum ini, mahasiswa diharapakan dapat lebih mengenal apa itu data warehouse berdasarkan uji coba membuat data warehouse sederhana.

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.

812/CarDistributionSales.csv

^{*}Data penjualan tersebut dapat diakses dan di download melalui link berikut: https://github.com/dik4rizky/datasources/blob/7e3f70f516a1b107635141d09862a8c30fbb6

Tugas 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Jawaban:

Data dari CarDistributionSales.csv memiliki 7 kolom.

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

Jawaban:

• Customer : Nama pelanggan

• Product: Jenis produk mobil yang dijual

• Region : Wilayah tempat penjualan dilakukan

• Date: Tanggal transaksi

• Item unit price : Harga per unit kendaraan

• No.Items: Jumlah unit kendaraan yang terjual

• Total Sale : Total nilai penjualan

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

Jawaban:

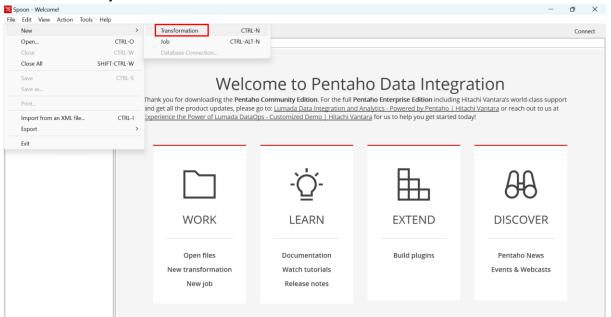
- **Product** → 3 data kosong
- **Region** → 3 data kosong
- **Date** → 2 data kosong
- 4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Jawaban:

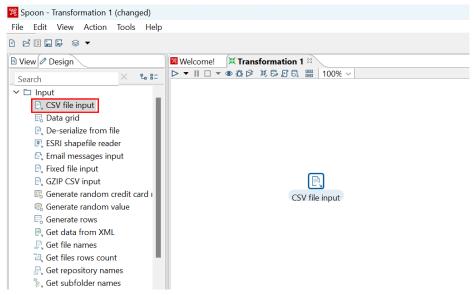
Tidak ada data yang memiliki tipe berbeda dengan data lainnya.

A. Pengenalan Data (Extract)

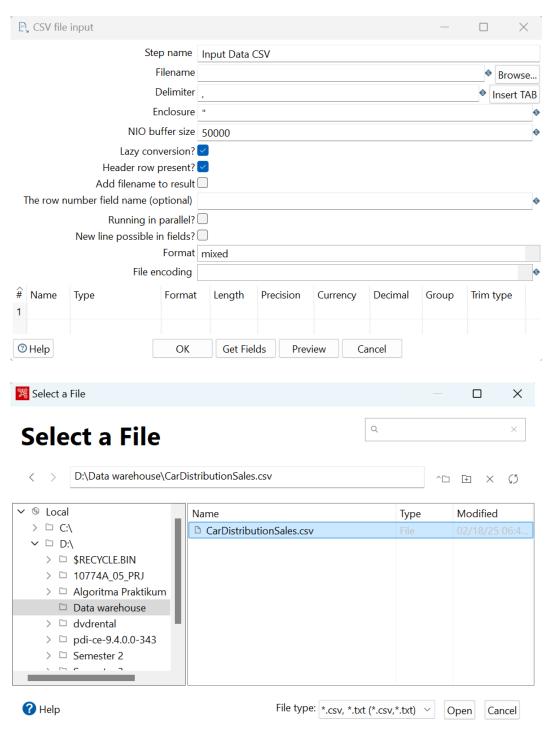
Buka lembar kerja Transformation melalui toolbar File.



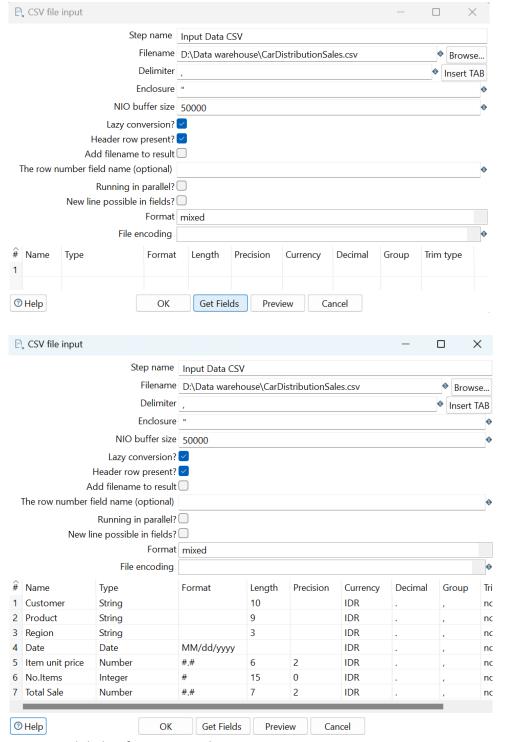
2. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area**.



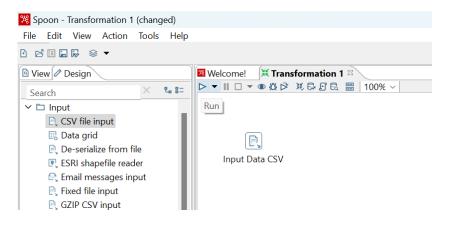
- 3. Double -click pada objek CSV file input hingga muncul jendela konfigurasinya.
- 4. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
- 5. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.



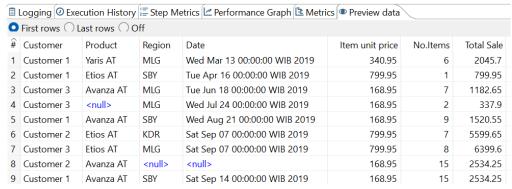
- 6. Jika data telah dipilih pastikan pada bagian **Delimiter** sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma ", ").
- 7. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai.



- 8. Tutup jendela konfigurasi tersebut.
- 9. Tekan tombol "Run" pada pojok kanan atas Area Kerja, makan akan muncul Execution Result Area. Pilih tab Prieview Data pada Execution Result Area. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.



Execution Results



Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.

Tugas 2

Analisa lah data tersebut!

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawaban:

Secara keseluruhan, isi data yang dieksekusi di Pentaho tetap sama dengan data aslinya. Namun, terdapat perbedaan dalam tampilan kolom **Date**. Pada Pentaho, kolom **Date** tidak hanya menampilkan tanggal saja, tetapi juga menyertakan keterangan waktu (timestamp).

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:

- Membuka lembar kerja **Transformation**.
- Menambahkan objek "CSV File Input" pada Design Area.
- Melakukan konfigurasi pada objek "CSV File Input".
- Menyesuaikan delimiter sesuai format file.
- Mengambil nama kolom menggunakan "Get Fields".
- Menjalankan proses extract.
- Memeriksa hasil extract pada tab "Preview Data".

3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E? 2025/02/23 10:09:36 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0) Jawaban:

Singkatan	Arti	Penjelasan
I	Input	Jumlah baris data yang masuk ke dalam proses
0	Output	Jumlah baris data yang berhasil diproses keluar
R	Read	Jumlah baris data yang telah dibaca dari sumber input
W	Write	Jumlah baris data yang telah ditulis ke dalam file output
U	Update	Jumlah baris data yang telah diperbarui
E	Error	Jumlah baris data yang mengalami kesalahan selama proses
		berlangsung

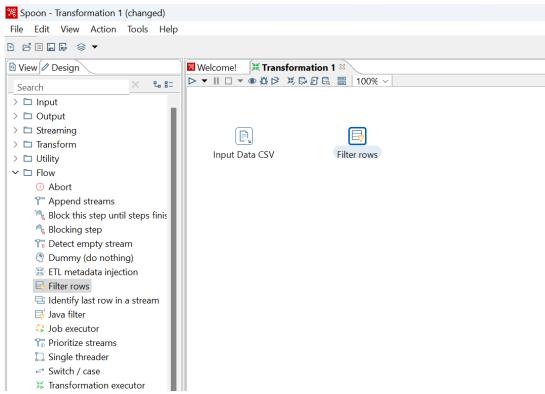
4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

Jawaban:

Nama	Waktu	Processor	RAM	Storage	VGA	СРИ
QUSNUL	1detik	AMD RYZEN 5 5625U	8 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 5 5625U with Radeon Graphics
REZA	1detik	AMD RYZEN 7 7730U	16 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 7 7730U with Radeon Graphics
DAHNIAR	1detik	Intel(R) Core(TM) i5- 6300U	8 GB	256 GB	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz
AQILA	1detik	AMD RYZEN 5 5600H	8 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 5 5600H with Radeon Graphics

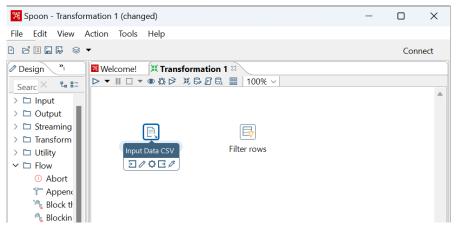
B. Filter Data (Transform) dan Pengemasan Data (Load)

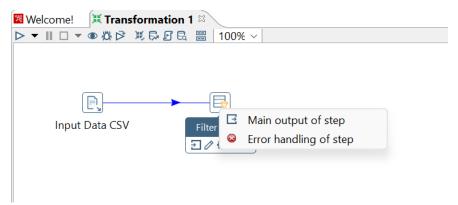
1. Carilah objek **Filter rows** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**



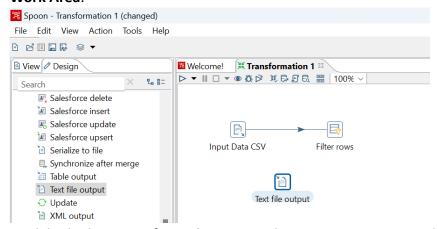
2. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.

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.

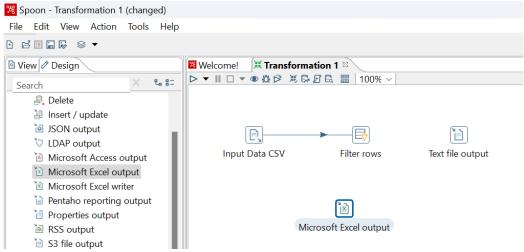




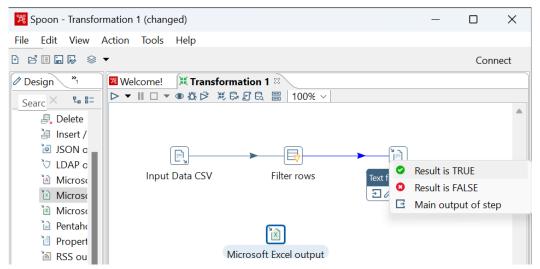
3. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



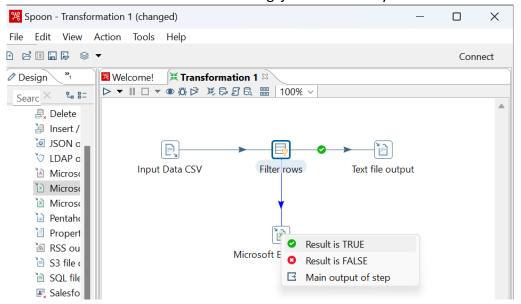
4. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



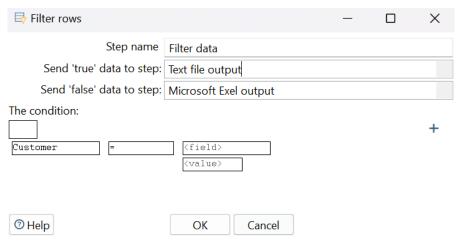
5. Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagi jenis konektornya.



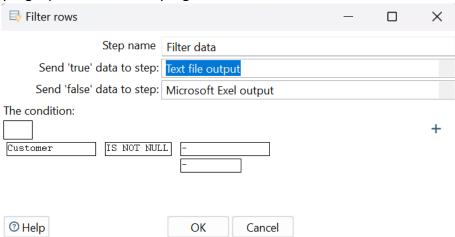
6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.



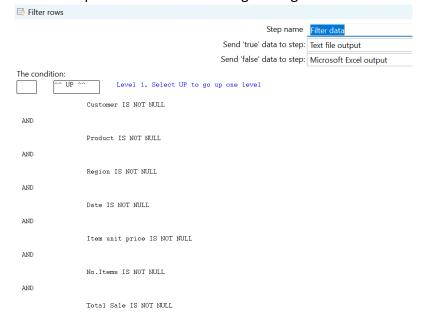
- 7. **Double click** pada **filter rows** hingga muncul jendela konfigurasi. ubah **step name** menjadi **Filter data.**
- 8. Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
- 9. Klik <Field> pada bagian condition
- 10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer,** kemudian klik **OK.**

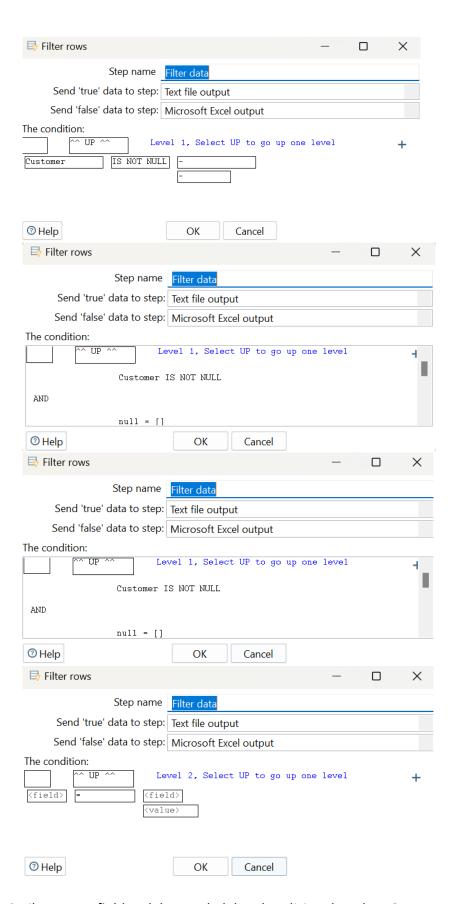


11. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



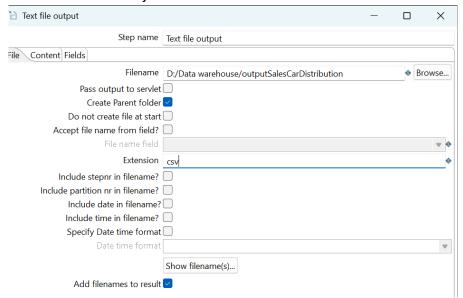
- 12. Tekan tombol + untuk menambah kondisi.
- 13. Pilih statement yang masih null untuk menambah kondisi.
- 14. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
- 15. Gunakan operator AND untuk menghubung



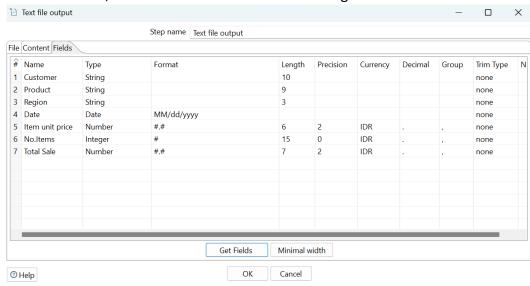


- 16. Jika semua field sudah masuk dalam kondisi maka tekan OK.
- 17. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**.

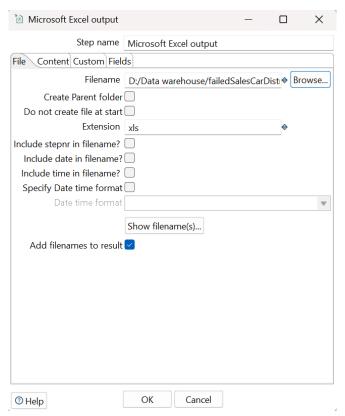
18. Ubah extension menjadi csv.



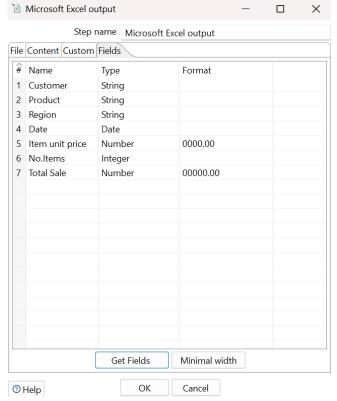
19. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.



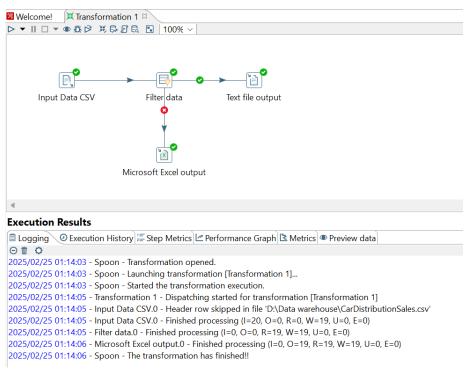
- 20. Tekan ok untuk kembali ke work area
- 21. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output, Double-click** pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.
- 22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol **browse**. Beri nama **failedSalesCarDistribution**. 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.



failedCarDistribution

	Α	В	С		D	Е	F	G
1	Customer	Product	Region	Date		Item unit p	No.Items	Total Sale
2	Customer		MLG		24/07/2019 00:00	0168,95	2,00	00337,90
3	Customer	Avanza AT				0168,95	15,00	02534,25
4	Customer					0340,95	7,00	02386,65
5	Customer		KDR		07/08/2020 00:00	0799,95	14,00	11199,30
6	Customer	Avanza AT			12/08/2020 00:00	0168,95	4,00	00675,80
7								

OutputSalesCarDistribution

1	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
2	Customer 1	Yaris AT	MLG	3/13/2019	340.9	6	2045.7
3	Customer 1	Etios AT	SBY	4/16/2019	800	1	800
4	Customer 3	Avanza AT	MLG	6/18/2019	168.9	7	1182.7
5	Customer 1	Avanza AT	SBY	8/21/2019	168.9	9	1520.5
6	Customer 2	Etios AT	KDR	09/07/2019	800	7	5599.6
7	Customer 3	Etios AT	MLG	09/07/2019	800	8	6399.6
8	Customer 1	Avanza AT	SBY	9/14/2019	168.9	15	2534.2
9	Customer 1	Etios AT	SBY	9/25/2019	800	2	1599.9
10	Customer 1	Avanza AT	SBY	1/21/2020	168.9	14	2365.3
11	Customer 1	Yaris AT	MLG	02/01/2020	340.9	2	681.9
12	Customer 1	Etios AT	SBY	5/30/2020	800	9	7199.6
13	Customer 3	Avanza AT	MLG	06/07/2020	168.9	10	1689.5
14	Customer 3	Yaris AT	MLG	06/09/2020	340.9	4	1363.8
15	Customer 2	Avanza AT	KDR	08/08/2020	168.9	12	2027.4
16							

Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.

Tugas 3

- 1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel? **Jawaban:**
 - File CSV (outputSalesCarDistribution.csv)

 Berisi data yang lengkap, yaitu data tanpa nilai kosong (null) di semua kolom

 yang digunakan dalam analisis. Data ini sudah siap untuk diproses lebih lanjut,
 seperti pembuatan dashboard atau laporan.
 - File Excel (failedSalesCarDistribution.xls)

 Berisi data yang tidak lengkap, yaitu data yang memiliki nilai kosong (null) di satu atau lebih kolom. Data ini dikumpulkan secara terpisah agar dapat diperiksa, diperbaiki, atau dilengkapi sebelum digunakan lebih lanjut.
- 2. Jelaskan apa yang terjadi pada proses **Filter rows!**

Jawaban:

Filter Rows digunakan untuk memisahkan data yang lengkap dan tidak lengkap. Data dari file CSV yang sudah diekstrak akan mengalir ke Filter Rows untuk dianalisis. Pada konfigurasi Filter Rows, kondisi yang digunakan adalah mengecek apakah suatu kolom memiliki nilai kosong (null) atau tidak.

- **Jika semua kolom memiliki nilai (IS NOT NULL)** → Data dianggap lengkap dan dikirim ke **Text File Output (CSV)**.
- Jika ada kolom yang kosong (NULL) → Data dianggap tidak lengkap dan dikirim ke Microsoft Excel Output (XLS).
- 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	Waktu	Processor	RAM	Storage	VGA	СРИ
QUSNUL	1 detik	AMD RYZEN 5 5625U	8 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 5 5625U with Radeon Graphics
REZA	1 detik	AMD RYZEN 7 7730U	16 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 7 7730U with Radeon Graphics
DAHNIAR	1 detik	Intel(R) Core(TM) i5- 6300U	8 GB	256 GB	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz

AQILA	2 detik	AMD RYZEN 5 5600H	8 GB	512 GB	AMD Radeon™ Graphics	AMD RYZEN 5 5600H with Radeon Graphics
-------	---------	----------------------	------	--------	-------------------------	----------------------------------------

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.

Jawaban:

Masyarakat berprofesi Mahasiswa

● First rows ○ Last rows ○ Off											
	NIK		Tempat Lahir Tanggal Lahir			Jenis Kelamin	Alama	t			
35030°	10101010002	Bandung Tue May 15 00:0		Tue May 15 00:00	Tue May 15 00:00:00 WIB 1990		Jl. Cen	drawasih No. 45			
35030°	10101010005	Semara	Semarang Sun Mar 0		Sun Mar 05 00:00:00 WIB 1995		Jl. Pem	uda No. 21			
ari 35030	10101010020	Jayapu	ra	Sat Oct 23 00:00:00 WIB 1993		Perempuan	Jl. Proklamasi No. 80				
RT/RW	Kelurahan		Kecan	natan	Agama	Status Perkawi	nan	Pekerjaan			
003/004	Cikutra		Bandu	ıng Wetan	Kristen	Belum Menikah		Mahasiswa			
009/010	Pandanara	n Semar		rang Tengah	Buddha	Belum Menikah		Mahasiswa			
039/040	Kota Jayap	ura	Jayapı	ura	Kristen	Belum Menikah		Mahasiswa			
	35030° 35030° ari 35030° RT/RW 003/004 009/010	NIK 350301010101010002 3503010101010005 ari 3503010101010020 RT/RW Kelurahan 003/004 Cikutra 009/010 Pandanara	NIK Tempat 3503010101010002 Bandur 3503010101010005 Semara 3503010101010020 Jayapu RT/RW Kelurahan 003/004 Cikutra 009/010 Pandanaran	NIK Tempat Lahir 3503010101010002 Bandung 3503010101010005 Semarang 3503010101010020 Jayapura RT/RW Kelurahan Kecam 003/004 Cikutra Bandu 009/010 Pandanaran Semai	NIK Tempat Lahir Tanggal Lahir 350301010101010002 Bandung Tue May 15 00:00 350301010101010005 Semarang Sun Mar 05 00:00 ari 3503010101010020 Jayapura Sat Oct 23 00:00: RT/RW Kelurahan Kecamatan 003/004 Cikutra Bandung Wetan 009/010 Pandanaran Semarang Tengah	NIK Tempat Lahir Tanggal Lahir 35030101010100002 Bandung Tue May 15 00:00:00 WIB 1990 35030101010100005 Semarang Sun Mar 05 00:00:00 WIB 1995 Sat Oct 23 00:00:00 WIB 1993 Sat Oct 23 00:00:00 WIB 1993 RT/RW Kelurahan Kecamatan Agama Cikutra Bandung Wetan Kristen O09/010 Pandanaran Semarang Tengah Buddha	NIK Tempat Lahir Tanggal Lahir Jenis Kelamin 350301010101010002 Bandung Tue May 15 00:00:00 WIB 1990 Perempuan 350301010101010005 Semarang Sun Mar 05 00:00:00 WIB 1995 Laki-laki ari 3503010101010020 Jayapura Sat Oct 23 00:00:00 WIB 1993 Perempuan RT/RW Kelurahan Kecamatan Agama Status Perkawi 003/004 Cikutra Bandung Wetan Kristen Belum Menika 009/010 Pandanaran Semarang Tengah Buddha Belum Menika	NIK Tempat Lahir Tanggal Lahir Jenis Kelamin Alama 35030101010100002 Bandung Tue May 15 00:00:00 WIB 1990 Perempuan Jl. Ceni 35030101010100005 Semarang Sun Mar 05 00:00:00 WIB 1995 Laki-laki Jl. Pemari 3503010101010020 Jayapura Sat Oct 23 00:00:00 WIB 1993 Perempuan Jl. Prokentia RT/RW Kelurahan Kecamatan Agama Status Perkawinan 003/004 Cikutra Bandung Wetan Kristen Belum Menikah 009/010 Pandanaran Semarang Tengah Buddha Belum Menikah			

• Masyarakat berprofesi PNS

First rows	Last rows O	ff							
# Nama		NIK		hir	Tanggal Lahir		Jenis Kelamin	Alamat	
1 Andi Wijaya	3503010	0101010007 Makassar			Sun Jun 25 00:00:00 WIB 1978		Laki-laki	Jl. Jenderal Sudirman No.	
2 Agus Santos	so 3503010	101010015	Padang		Fri Apr 02 00:00:00 WIB 1982		Laki-laki	Jl. Sudirman No. 20	
Kota	RT/RW	Keluraha	n Keca		camatan	Agama	Agama Status Pe		Pekerjaan
Makassar	013/014	Baromb	ong Mal		kassar Selatan	Islam	Menikah		PNS
	013/011		9		itabbai belatan				

• Masyarakat berprofesi Dokter

○ First rows ○ Last rows ○ Off										
# Nama		NIK	Tempat Lahir	Tanggal Lahir		Jenis Kelamin		Alamat		
1 Desi Purnama	3503	3503010101010008		Thu Apr 08 00:00:00 WIB 1993		Perempuan		Jl. Demang Lebar Daun No		
2 Bambang Kusu	mo 3503	3503010101010017 Banjar		Thu Sep 25 00:00:00 WIB 1980		Laki-laki		Jl. A. Yani No. 60)	
Kota	RT/RW	Kelurahan		Kecamatan	Aga	ma	Status	Perkawinan	Pekerjaan	
Palembang	015/016	Alang-Ala	ng Lebar	Palembang Selatan	Krist	ten	Belum	Menikah	Dokter	
Banjarmasin	033/034	Banjarmas	in Tengah	Banjarmasin	Hino	Hindu		ah	Dokter	