# LAPORAN PRAKTIKUM DATA WAREHOUSE

# JOBSHEET 1 DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON



Oleh: NIMAS SEPTIANDINI 2341760087

SIB 2B

PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025/2026

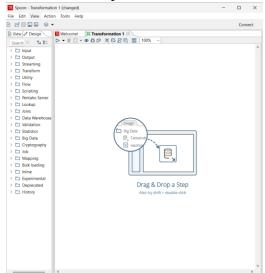
#### Tugas 1

	Α	В	С	D	E	F	G	Н
1	Customer	,Product,R	egion,Date	, Item unit	price , No	.Items , To	otal Sale	
2	Customer	1,Yaris AT,	MLG,3/13/	2019,340.9	5,6,2045.7			
3	Customer	1,Etios AT	SBY,4/16/2	019,799.95	,1,799.95			
4	Customer	3,Avanza	AT,MLG,6/1	8/2019,168	3.95,7,1182	2.65		
5	Customer	3,,MLG,7/2	24/2019,168	3.95,2,337.	9			
6	Customer	1,Avanza	AT,SBY,8/21	/2019,168	.95,9,1520.	55		
7	Customer	2,Etios AT	KDR,9/7/20	019,799.95	,7,5599.65			
8	Customer	3,Etios AT	MLG,9/7/2	019,799.95	,8,6399.6			
9	Customer	2,Avanza	AT,,,168.95,	15,2534.25	j			
10	Customer	1,Avanza	AT,SBY,9/14	/2019,168	.95,15,2534	4.25		
11	Customer	1,Etios AT	SBY,9/25/2	019,799.95	,2,1599.9			
12	Customer	1,Avanza	AT,SBY,1/21	/2020,168	.95,14,236	5.3		
13	Customer	1,Yaris AT,	MLG,2/1/2	020,340.95	,2,681.9			
14	Customer	1,,,,340.95	,7,2386.65					
15	Customer	1,Etios AT	SBY,5/30/2	020,799.95	,9,7199.55			
16	Customer	3,Avanza	AT,MLG,6/7	/2020,168.	95,10,1689	).5		
17	Customer	3,Yaris AT,	MLG,6/9/2	020,340.95	,4,1363.8			
18	Customer	2,,KDR,8/7	7/2020,799.	95,14,1119	9.3			
19	Customer	2,Avanza	AT,KDR,8/8	/2020,168.	95,12,2027	.4		
20	Customer	1,Avanza	AT,,8/12/20	20,168.95,	4,675.8			

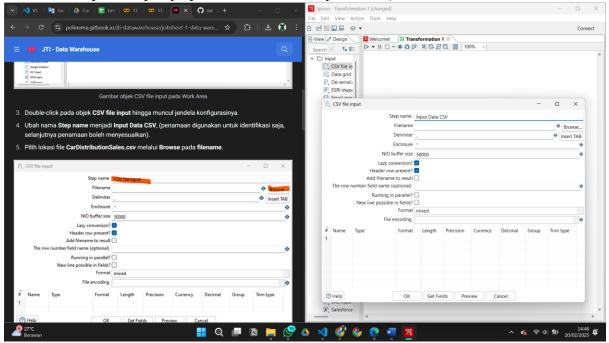
- 1. Berapa jumlah kolom pada data tersebut?
  - = Jumlah kolom pada data tersebut adalah 7
- 2. Apa arti atau isi data dari setiap kolom yang ada?
  - Customer : informasi customer
  - Product : nama produk yang dibeli pelanggan
  - Region : menunjukkan tempat pembelian produk
  - Date : menunjukkan tanggal pembelian produk
  - Item unit price: harga satuan produk
  - No.item : nomor dari item yang dibeli
  - Total Sale: total pembelian yang dilakukan
- 3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
  - = Ya, terdapat data tidak lengkap pada baris 4, 8, 13, 17, 19
- 4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?
  - 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).



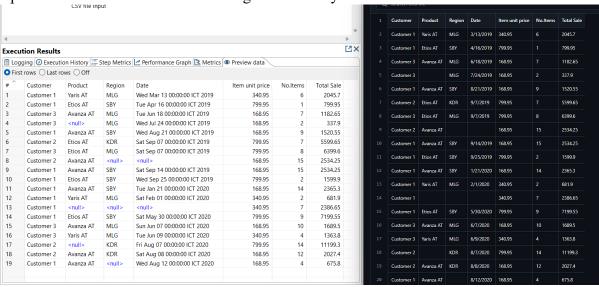
- 5. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.
- 6. Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.
- 7. Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
- 8.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

11. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.

Exe	cution Res	ults						
	Logging 🕢 Ex	ecution History	Step M	etrics 🗠 Performance Graph 🔁 Me	trics Preview data			
0	First rows O L	ast rows Of	f					
#	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></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>	<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	
1	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9	
1	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3	
1	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9	
1	Customer 1	<null></null>	<null></null>	<null></null>	340.95	7	2386.65	
1	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55	
1	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5	
1	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8	

#### Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?



- = ya, data hasil eksekusi sesuai dengan data aslilnya
- 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?
  - = Langkah Langkah yang dilakukan untuk mengekstrak data

- 1. Spoon menyimpan file transformasi
- 2. Spoon membuka file transformasi yang sudah dikonfigurasi
- 3. Spoon menjalankan transformasi
- 4. Transformasi mulai dieksekusi
- 5. Sistem melakukan distribusi tugas tugas yang diperlukan
- 6. Spoon membaca file CSV
- 7. Spoon memberikan laporan hasil pemrosesan
- 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) = input, output, read, written, updated, errors
- 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
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)

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



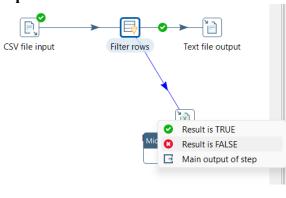
2. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.

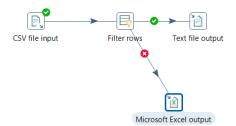


3. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

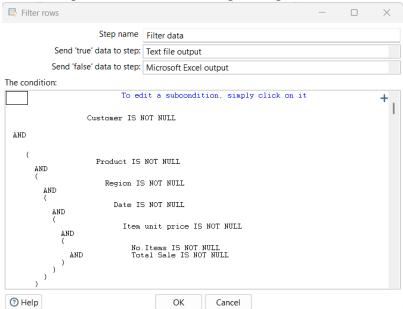


- 4. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.
- 5. Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagi jenis konektornya.
- 6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.

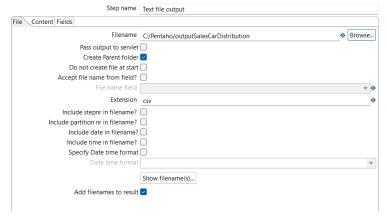


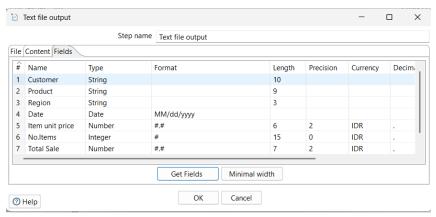


- 7. Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.
- 8. Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
- 9. Klik <Field> pada bagian condition
- 10. Pilih field 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

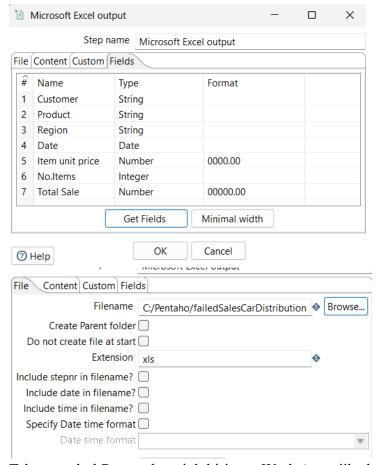


- 16. Jika semua field sudah masuk dalam kondisi maka tekan OK.
- 17. Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.
- 18. Ubah extension menjadi csv.
- 19. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.
- 20. Tekan ok untuk kembali ke work area

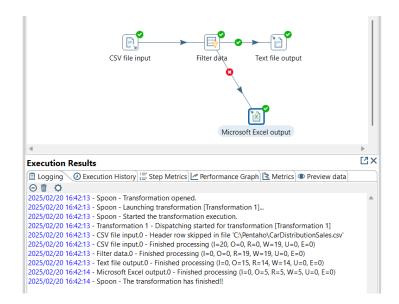




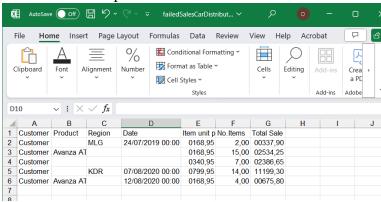
- 21. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
- 22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan extension xls.
- 23. Pada tab fields tekan tombol Get fields untuk mengambil field data.
- 24. Tekan ok untuk kembali ke work area.



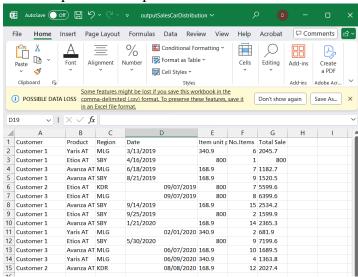
25. Tekan tombul Run pada pojok kiri atas Work Area. jika berhasil maka akan muncul tanda centang hijau di setiap proses.



Hasil failedSales pada file .csv:



Hasil outputSalesCar pada file .csv:



### Tugas 3

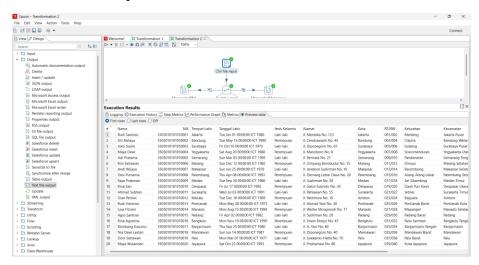
1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

- = 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**!
  - 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 table

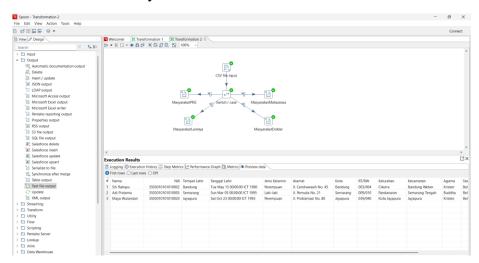
Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152 GB	Inter(R) UHD Graphics	12th Gen Intel(R) Core(TM) i5 - 1235U	0.6s
Nimas	AMD Ryzen 3 5300U	16,0 GB	475 GB	AMD Radeon Graphics	AMD Ryzen 3 5300U with Radeon Graphics.	0.4s
Putera	Intel(R) 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

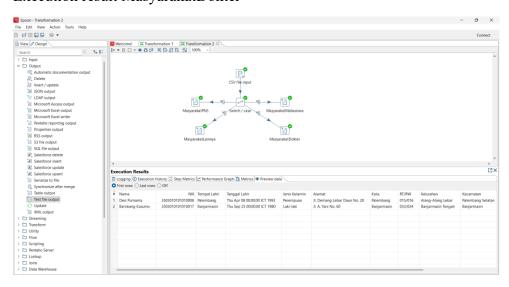
## Hasil Run data kependudukan



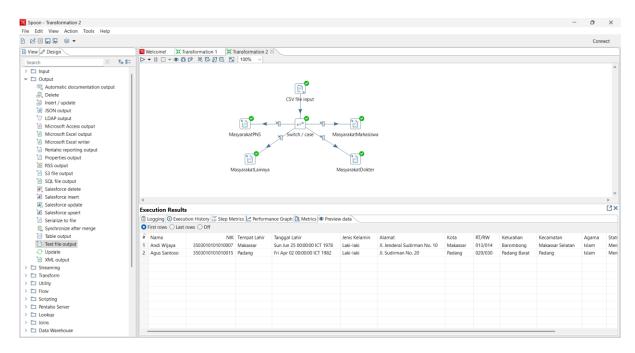
## Execution result MasyarakatMahasiswa



## Execution result MasyarakatDokter



#### Execution result MasyarakatPNS



## Execution Result MasyarakatLainnya

