

**DATA WAREHOUSE**

**JOBSHEET 1: DATA SEDERHANA DAN**

**PENGENALAN PDI SPOON**



Disusun Oleh:

KARINA IKA INDASA (2341760042)

KELAS 2A-SIB/16

**PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

Jl. Soekarno Hatta No. 9, Jattimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur

### ➤ 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.

### ➤ Tugas 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

**Jawab:** 7 kolom

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

**Jawab:**

- Customer: Nama pelanggan (misalnya ,Customer 1,Customer 2,Customer 3)
- Product: Model mobil yang dibeli (contoh: Yaris AT, Avanza AT, Etios AT)
- Region: Lokasi cabang (contoh: MLG, SBY, KDR)
- Date: Tanggal transaksi dalam format MM/DD/YYYY
- Item unit price: Harga satuan mobil.
- No.Items: Jumlah unit yang dibeli dalam satu transaksi
- Total Sale: Total nilai transaksi (hasil perkalian kolom Item unit price dan No.Items)

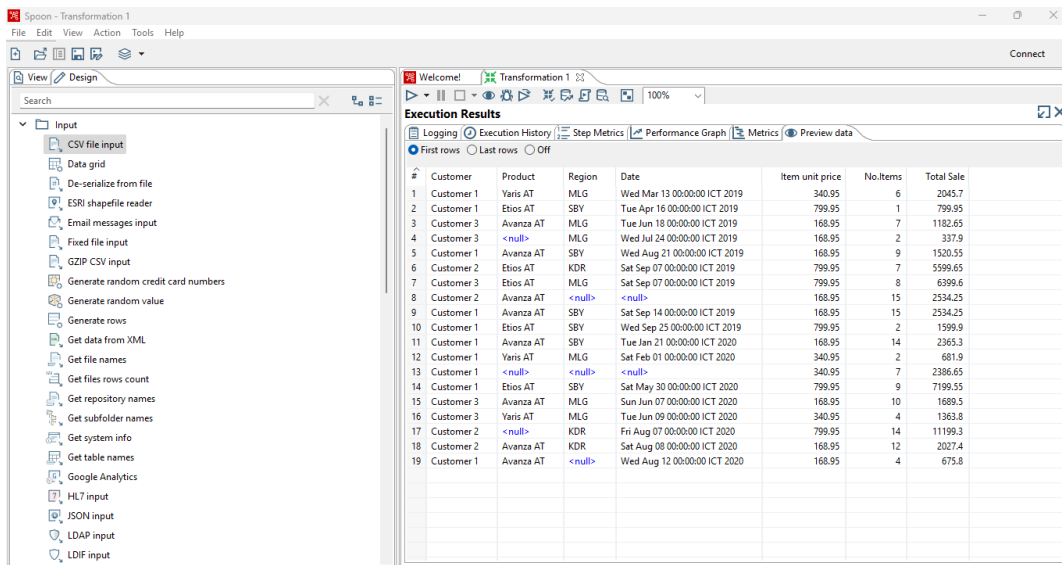
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?

**Jawab:** Ada, pada kolom product, region, date

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

**Jawab:** Tidak ada

## A. Pengambilan Data (Extract)



### TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

**Jawab:** Sesuai dengan data aslinya.

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?

```
2025/02/18 13:09:13 - Spoon - Save file as...
2025/02/18 13:09:14 - Spoon - Transformation opened.
2025/02/18 13:09:14 - Spoon - Launching transformation [Transformation 1]...
2025/02/18 13:09:14 - Spoon - Started the transformation execution.
2025/02/18 13:09:14 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2025/02/18 13:09:15 - Input Data CSV.0 - Header row skipped in file 'D:\Backup\Local Disk D\Kuliah\semester4\DW\CarDistributionSales.csv'
2025/02/18 13:09:15 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
```

**Jawab:** 2025/02/18 13:09:15 - Spoon - The transformation has finished!!

- 1) Save File → PDI Spoon menyimpan file transformasi yang akan digunakan.
  - 2) Transformation opened → Membuka transformasi yang telah dibuat sebelumnya.
  - 3) Launching transformation → Memulai proses eksekusi transformasi.
  - 4) Started the transformation execution → Transformasi mulai dijalankan.
  - 5) Dispatching transformation → Memulai distribusi proses transformasi yang telah dibuat.
  - 6) Input Data CSV- Header row skip → PDI Spoon membaca file CSV dan melewati baris header agar tidak dimasukkan ke dalam data.
  - 7) Finished processing (I=20, O=0, R=0, W=19, U=0, E=0) → Proses input data berhasil.
  - 8) The transformation has finished → Proses transformasi selesai dengan sukses.
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)

**Jawab:**

- I (Input) = 20 → PDI Spoon membaca 20 baris data dari file CSV.
- (Output) = 0 → Tidak ada data yang dikeluarkan.
- R (Read) = 0 → Tidak ada data tambahan yang dibaca dari sumber lain.

- W (Written) = 19 → Sebanyak 19 baris data berhasil ditulis ke output yang telah ditentukan.
  - U (Updated) = 0 → Tidak ada data yang diperbarui.
  - E (Errors) = 0 → Tidak ada kesalahan dalam proses transformasi.
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.

**Jawab:**

Nama	Processor	RAM	Storage	VGA	CPU	Time
Saya	Intel(R) Core(TM) i3-1005G1	4.00 GB	256 GB	Intel(R) UHD Graphics	CPU @ 1.20GHz 1.20 GHz	2s
Diajeng	Intel(R) Core(TM) i5-8250U	8.00 GB	512 GB	Intel(R) UHD Graphics 680	CPU @ 1.60GHz 1.80 GHz	1s
Aqila	AMD Ryzen 5 5600H	8.00 GB	512 GB	AMD Radeon(TM) Graphics	AMD Ryzen 5 5600H	1s
Dahniar	Intel(R) Core(TM) i5-6300U	8.00 GB	256 GB	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz	1s

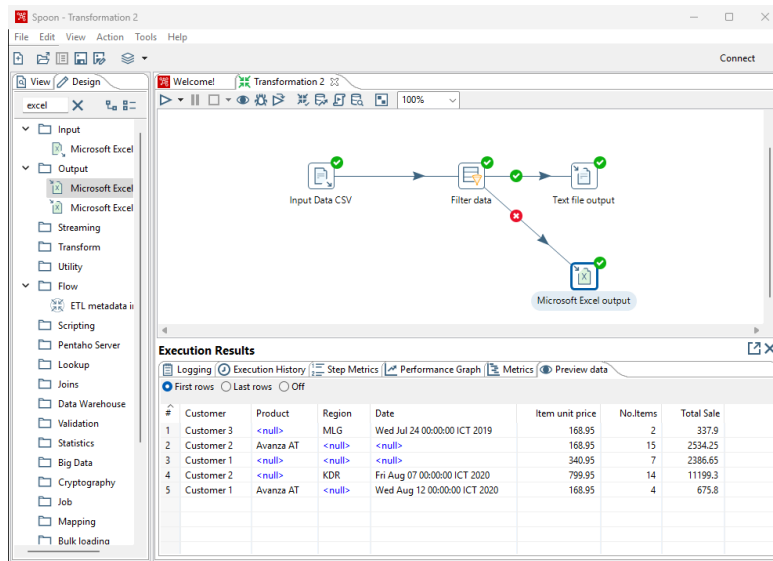
## B. Filter Data (Transform) dan Pengemasan data (Load)

- True

The screenshot shows the SAP Data Services Transformation 2 interface. The main workspace displays a data flow diagram with the following components: 'Input Data CSV', 'Filter data', 'Text file output', and 'Microsoft Excel output'. The 'Filter data' component is highlighted with a red 'X' icon, indicating a filter operation. The 'Execution Results' window is open at the bottom, showing a table of data with the following columns: #, Customer, Product, Region, Date, Item unit price, No.Items, and Total Sale. The table contains 9 rows of data.

#	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	Etiot 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 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
5	Customer 2	Etiot AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
6	Customer 3	Etiot AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
7	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
8	Customer 1	Etiot AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
9	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3

- False



### TUGAS 3

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

**Jawab:**

- File csv

**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	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 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
5	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
6	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
7	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
8	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
9	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
10	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
11	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55
12	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
13	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8
14	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 ICT 2020	168.95	12	2027.4

- File excel

outputSalesCarDistribution.csv - Excel

File Home Insert Page Layout Formulas Data Review View Help

Clipboard Font Alignment Number Styles

Conditional Formatting Table Styles Insert Delete Format Sort & Filter Find & Select Add-ins

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-separated (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	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	Etios AT	SBY	4/16/2019	800.1	800															
4	Customer 3	Avanza AT	MLG	6/18/2019	168.95	7	1182.7														
5	Customer 1	Avanza AT	SBY	8/21/2019	168.95	9	1520.5														
6	Customer 2	Etios AT	KDR	9/7/2019	800.2	7	5599.6														
7	Customer 3	Etios AT	MLG	9/7/2019	800.6	8	6399.6														
8	Customer 1	Avanza AT	SBY	9/14/2019	168.95	15	2534.2														
9	Customer 1	Etios AT	SBY	9/25/2019	800.2	2	1599.9														
10	Customer 1	Avanza AT	SBY	1/21/2020	168.95	14	2365.3														
11	Customer 1	Yaris AT	MLG	2/1/2020	340.95	2	681.9														
12	Customer 1	Etios AT	SBY	5/30/2020	800.9	9	7199.6														
13	Customer 3	Avanza AT	MLG	6/7/2020	168.95	10	1689.5														
14	Customer 3	Yaris AT	MLG	6/9/2020	340.95	4	1363.8														
15	Customer 2	Avanza AT	KDR	8/8/2020	168.95	12	2027.4														
16																					
17																					
18																					
19																					
20																					

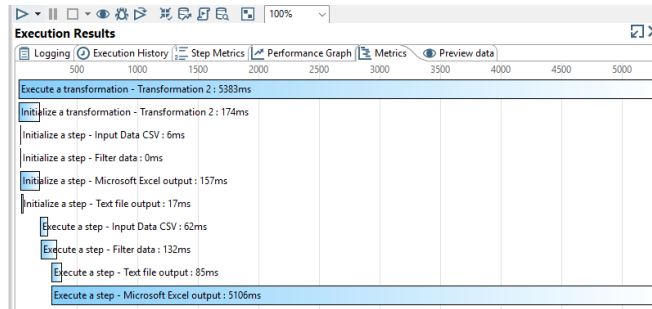
outputSalesCarDistribution

2. Jelaskan apa yang terjadi pada proses Filter rows!

**Jawab:** Proses filter data. Jika data memiliki kolom yang lengkap, maka hasilnya akan disimpan dalam file teks. Namun, jika terdapat kolom yang kosong, hasilnya akan dialihkan ke output Microsoft Excel.

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

**Jawab:**



Nama	Processor	RAM	Storage	VGA	CPU	Time
Saya	Intel(R) Core(TM) i3-1005G1	4.00 GB	256 GB	Intel(R) UHD Graphics	CPU @ 1.20GHz 1.20 GHz	2s
Diajeng	Intel(R) Core(TM) i5-8250U	8.00 GB	512 GB	Intel(R) UHD Graphics 680	CPU @ 1.60GHz 1.80 GHz	2s
Aqila	AMD Ryzen 5 5600H	8.00 GB	512 GB	AMD Radeon(TM) Graphics	AMD Ryzen 5 5600H	2s
Dahniar	Intel(R) Core(TM) i5-6300U	8.00 GB	256 GB	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz	2s

### ➤ 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. 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 emmbantu mengambil keputusan.

Pak Pascanowo mendapatakan data penduduk yang akan pindah ke wilayahnya melalui link: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aeadd012/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.

