**SAFRIZAL RAHMAN**

**19**

**SIB 2G**

**234176015**

**JOBSHEET 1  
  
TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?
2. Apa arti atau isi data dari setiap kolom yang ada?
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Analisis Data Penjualan

1. Jumlah Kolom pada Data

Data penjualan tersebut memiliki 7 kolom, yaitu:

1. Customer: Nama pelanggan yang melakukan pembelian.

2. Product: Jenis mobil yang dibeli.

3. Region: Wilayah atau cabang tempat penjualan dilakukan.

4. Date: Tanggal transaksi penjualan.

5. Item unit price: Harga satuan per unit mobil.

6. No.Items: Jumlah unit mobil yang dibeli.

7. Total Sale: Total harga penjualan (Item unit price × No.Items).

2. Arti atau Isi Data dari Setiap Kolom

Customer: Berisi nama pelanggan, misalnya "Customer 1", "Customer 2", dll.

Product: Berisi jenis mobil yang dibeli, seperti "Yaris AT", "Etios AT", atau "Avanza AT".

Region: Berisi kode wilayah atau cabang, seperti "MLG" (Malang), "SBY" (Surabaya), atau "KDR" (Kediri).

Date: Berisi tanggal transaksi dalam format `MM/DD/YYYY`.

Item unit price: Berisi harga satuan per unit mobil dalam format numerik.

No.Items: Berisi jumlah unit mobil yang dibeli dalam format numerik.

Total Sale: Berisi total harga penjualan, yang dihitung dari `Item unit price × No.Items`.

3. Data yang Memiliki Nilai Null / Tidak Lengkap

Ada beberapa baris yang memiliki nilai null atau data yang tidak lengkap. Berikut detailnya:

Baris 4: Kolom `Product` kosong.

Baris 8: Kolom `Region` dan `Date` kosong.

Baris 13: Kolom `Customer`, `Product`, `Region`, dan `Date` kosong.

Baris 18: Kolom `Product` kosong.

Baris 21: Kolom `Region` dan `Date` kosong.

4. Data yang Memiliki Tipe Berbeda pada Kolom yang Sama

Secara umum, tipe data pada setiap kolom sudah konsisten:

Customer: Teks (string).

Product: Teks (string).

Region: Teks (string).

Date: Tanggal (date).

Item unit price: Numerik (float atau double).

No.Items: Numerik (integer).

Total Sale: Numerik (float atau double).

Namun, perlu diperhatikan bahwa pada kolom Date, format tanggal sudah konsisten (`MM/DD/YYYY`), sehingga tidak ada perbedaan tipe data.

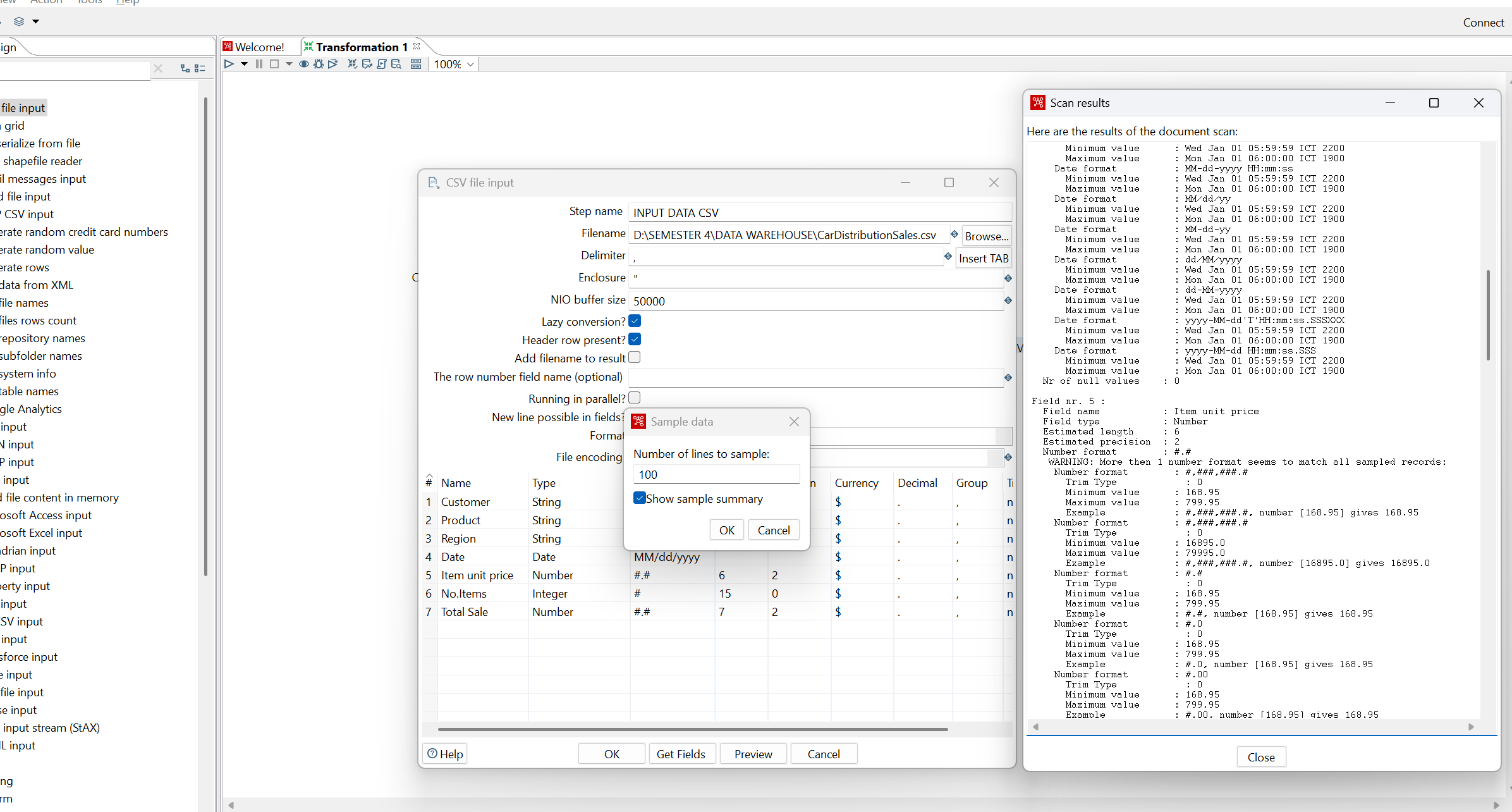
Kesimpulan

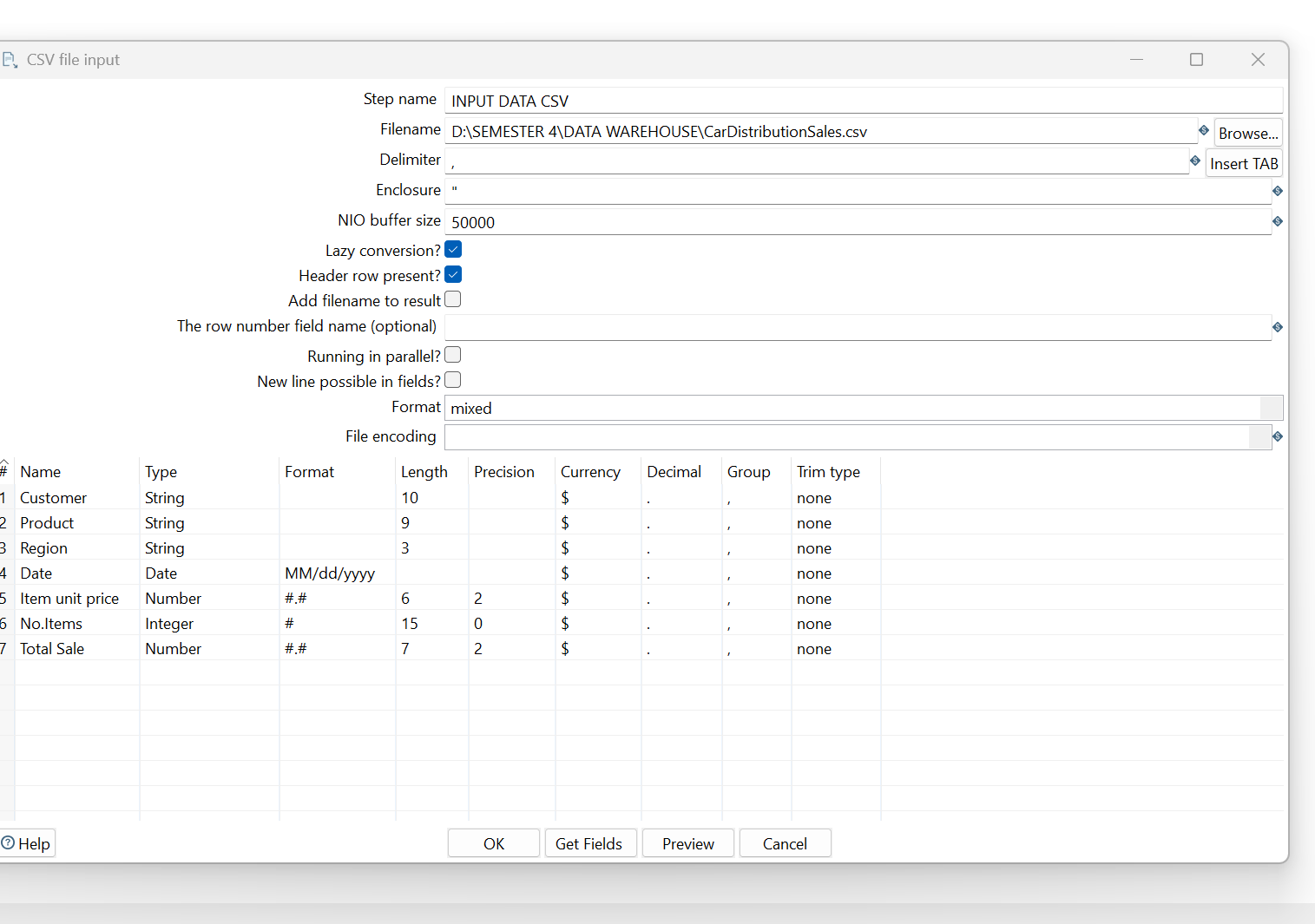
Data memiliki 7 kolom dengan arti dan tipe data yang jelas.

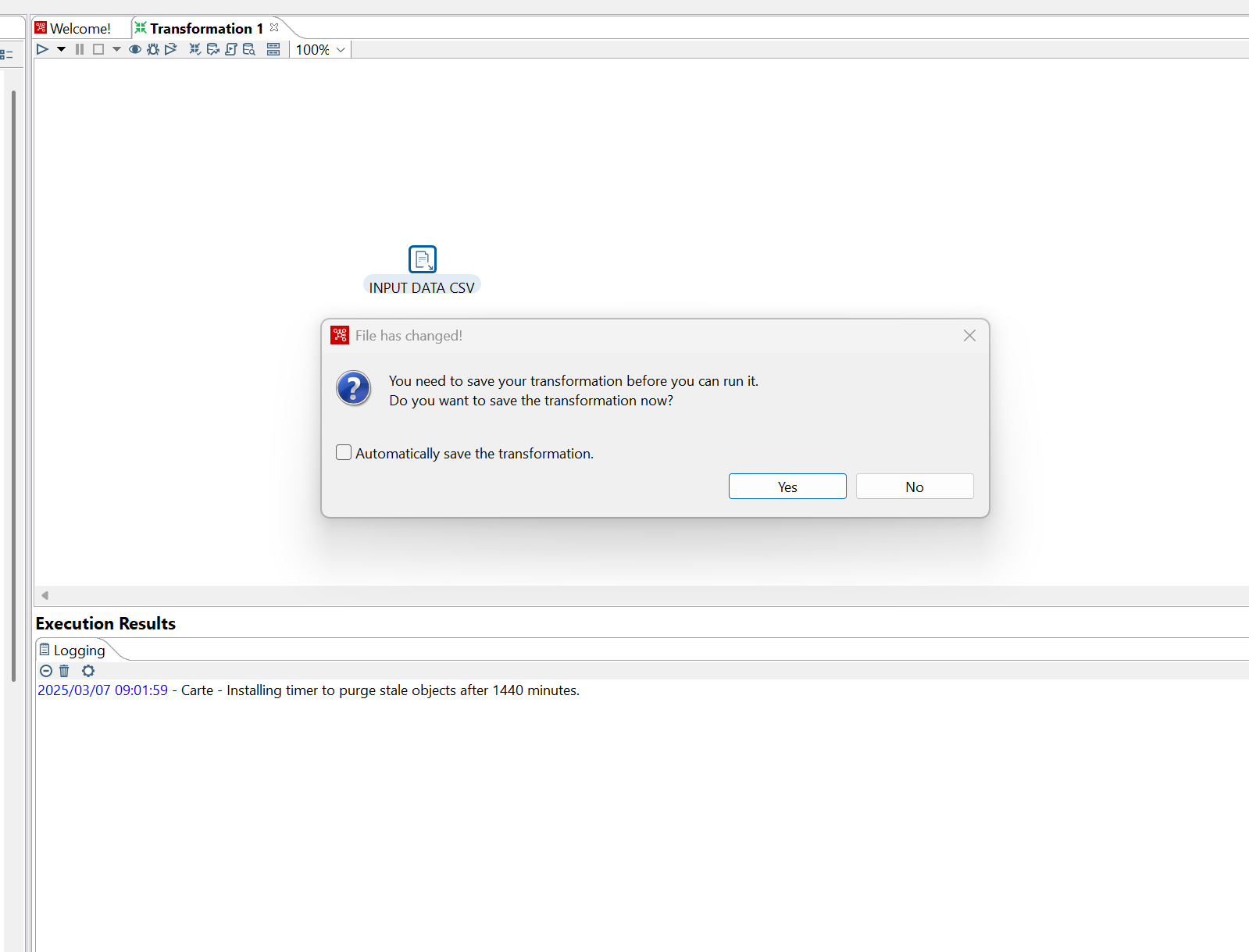
Terdapat 5 baris yang memiliki nilai null atau data tidak lengkap.

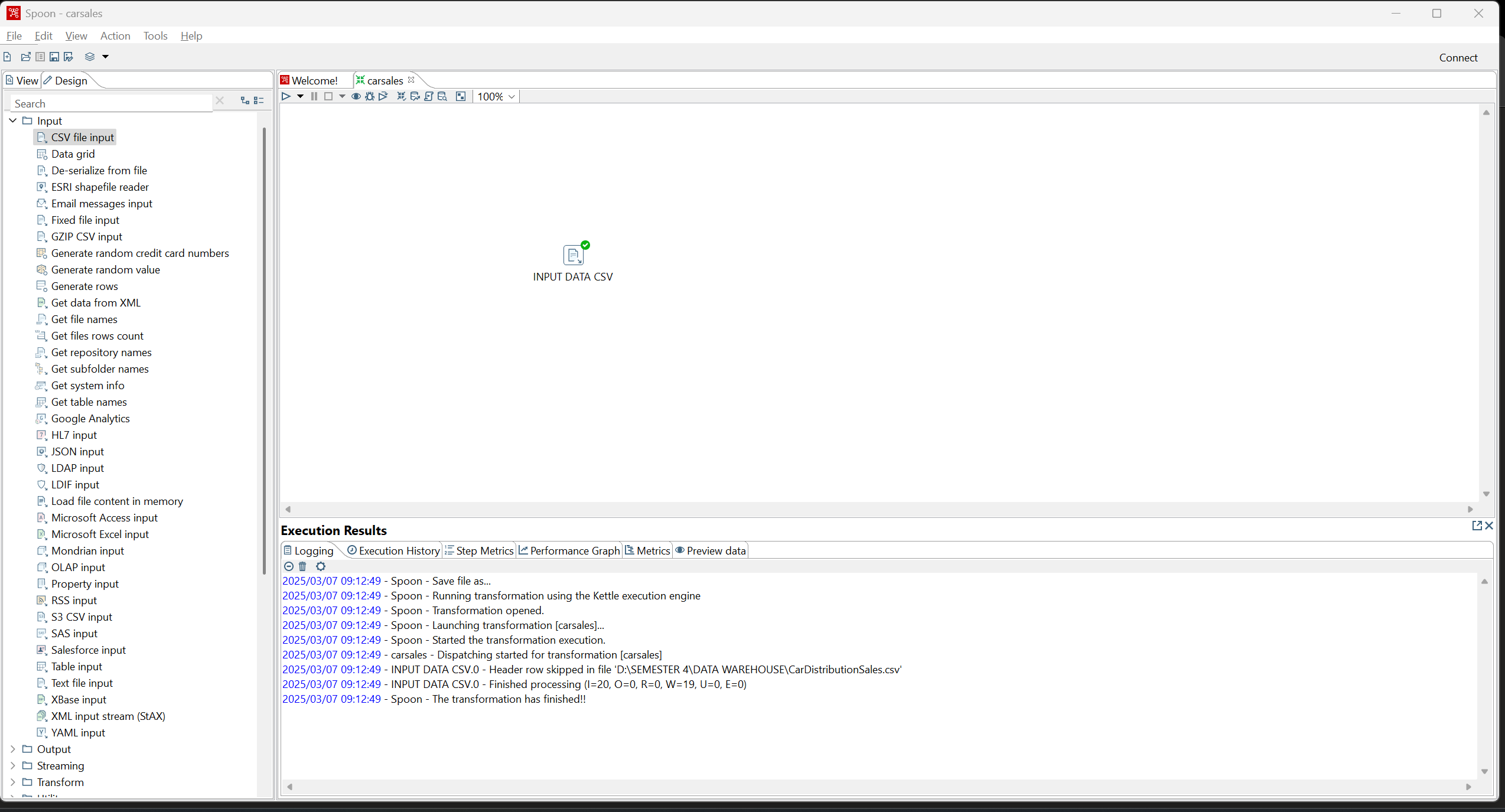
Tidak ada perbedaan tipe data pada kolom yang sama.

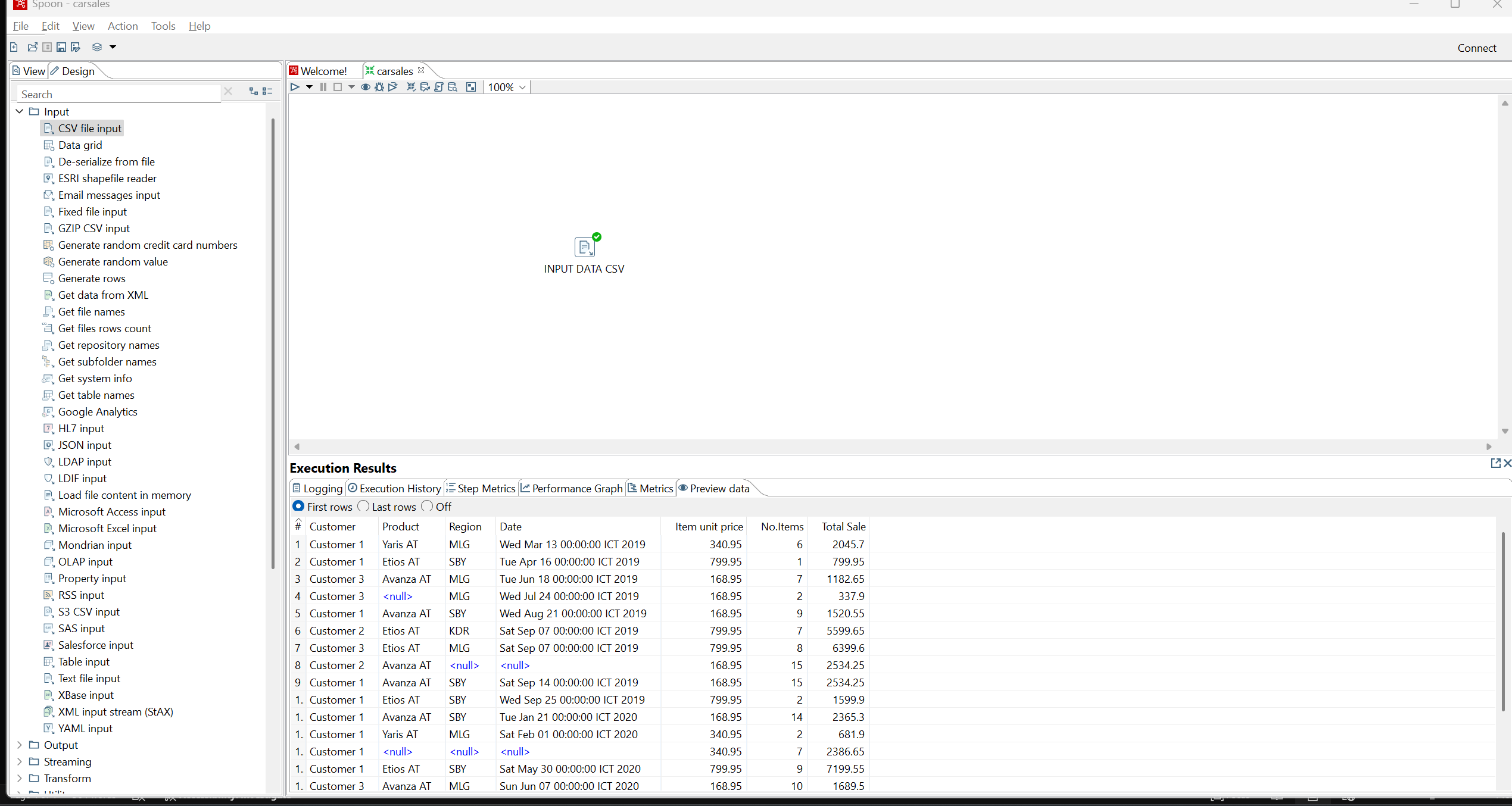
Langkah selanjutnya adalah melakukan Transform untuk memisahkan data yang lengkap dan tidak lengkap, kemudian Load data yang lengkap ke file dashboard. Data yang tidak lengkap dapat dikembalikan untuk diperbaiki.











**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
2. PDI Spoon melakukan proses extract, perhatikan pada **Tab Logging** di **Execution Results Area,** langkahlangkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?



1. 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.

TUGAS 2: Analisis Proses Extract Data dengan PDI Spoon

1. Apakah Data Hasil Eksekusi Sesuai dengan Data Aslinya?

Jawaban: Ya, data hasil eksekusi harus sesuai dengan data aslinya. Jika langkahlangkah yang dijelaskan di atas dilakukan dengan benar, data yang muncul di Preview Data pada Execution Results Area akan sama persis dengan data yang ada di file `CarDistributionSales.csv`. Jika ada perbedaan, kemungkinan terjadi kesalahan dalam konfigurasi CSV file input (misalnya, delimiter yang salah atau pemilihan file yang tidak tepat).

2. LangkahLangkah yang Dilakukan PDI Spoon untuk Melakukan Extract Data

Berdasarkan Tab Logging di Execution Results Area, berikut adalah langkahlangkah yang dilakukan PDI Spoon untuk melakukan extract data:

1. Membaca File CSV:

PDI Spoon membuka file `CarDistributionSales.csv` yang telah dipilih.

Memeriksa delimiter (dalam kasus ini, koma `,`) untuk memisahkan kolom.

2. Mengambil Metadata Kolom:

Menggunakan fungsi Get Fields, PDI Spoon membaca judul kolom dan memastikan struktur data sesuai.

3. Memproses Data:

PDI Spoon membaca setiap baris data dari file CSV dan memasukkannya ke dalam alur transformasi.

4. Menampilkan Preview Data:

Setelah proses selesai, data ditampilkan di Preview Data untuk memastikan bahwa data telah diekstrak dengan benar.

3. Arti dari I, O, R, W, U, E pada Tab Logging

Simbolsimbol ini menunjukkan status atau tindakan yang dilakukan selama proses eksekusi:

I (Input): Menunjukkan bahwa data sedang dibaca atau diambil dari sumber (dalam hal ini, file CSV).

O (Output): Menunjukkan bahwa data sedang ditulis atau dikirim ke tujuan (misalnya, ke file atau database lain).

R (Read): Menunjukkan bahwa data sedang dibaca dari sumber.

W (Write): Menunjukkan bahwa data sedang ditulis ke tujuan.

U (Update): Menunjukkan bahwa data sedang diperbarui (jika ada proses update).

E (Error): Menunjukkan bahwa terjadi kesalahan selama proses eksekusi.

4. Waktu yang Dibutuhkan untuk Eksekusi Extract Data

Waktu eksekusi dapat bervariasi tergantung pada spesifikasi perangkat yang digunakan. Berikut adalah contoh perbandingan waktu eksekusi dan spesifikasi perangkat:

| No. | Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA | CPU Usage |

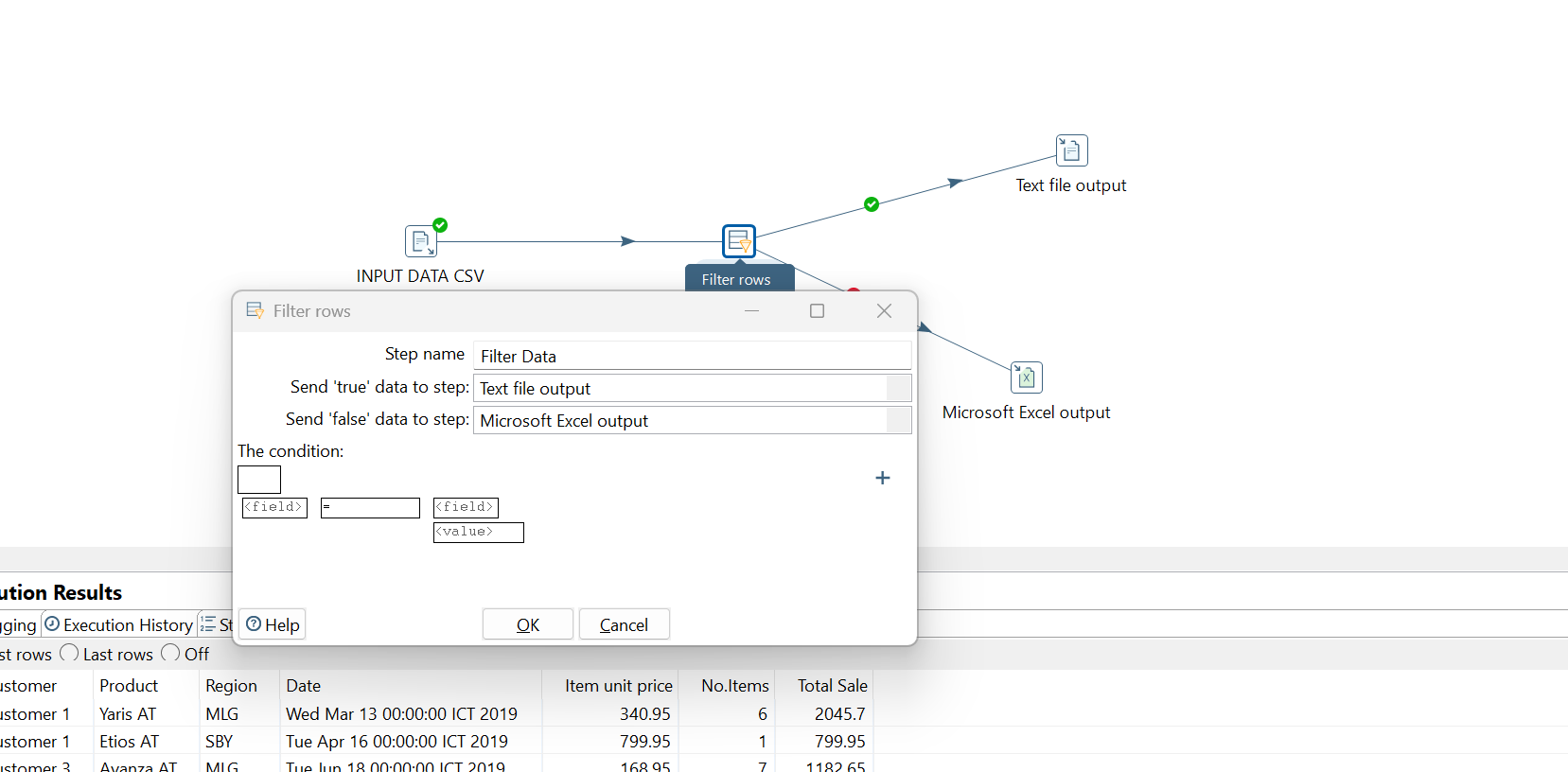
|||||||||

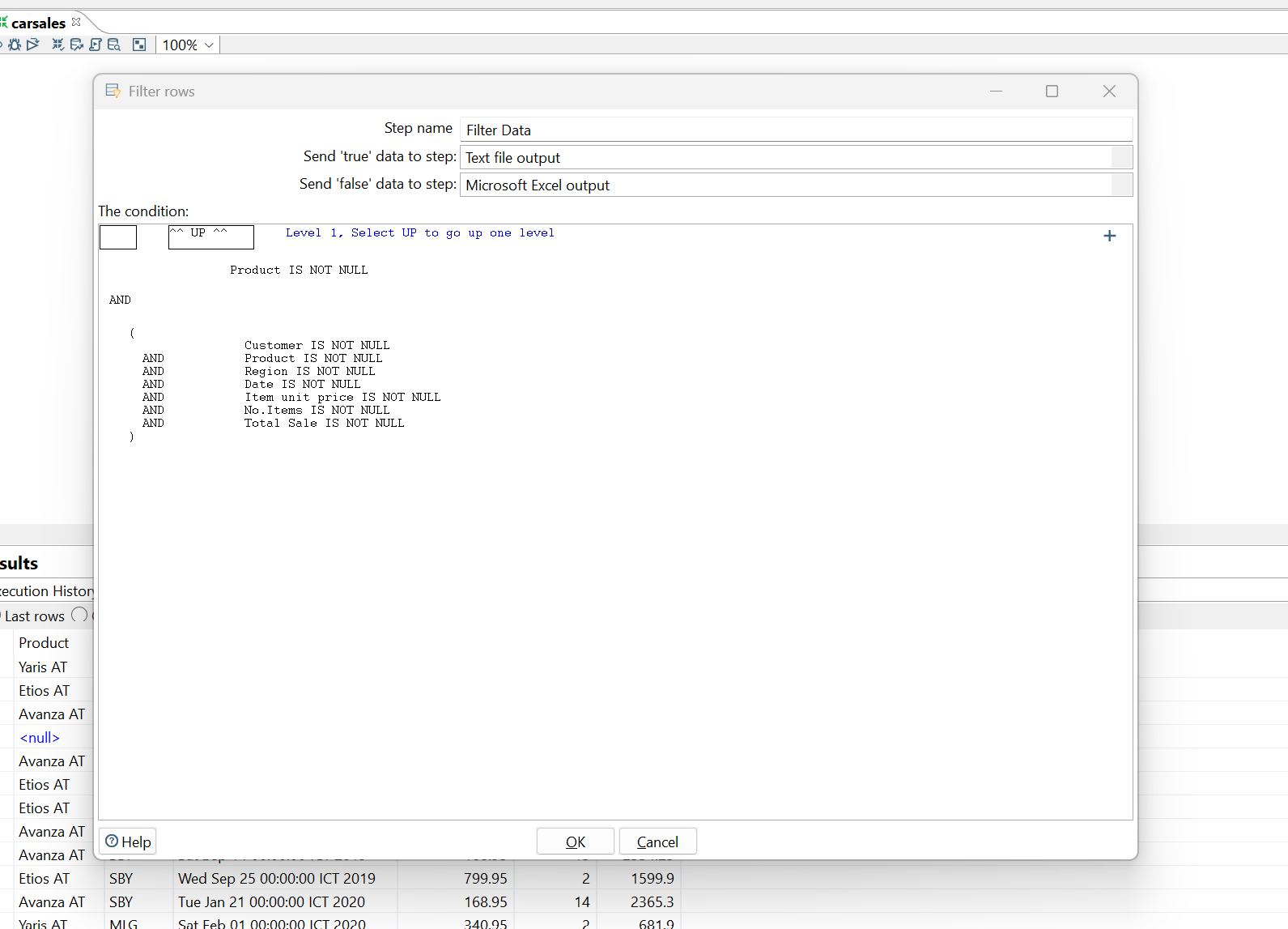
| 1 | Anda | 2 detik | Intel i510300H | 8 GB | SSD | NVIDIA GTX 1650| 15% |

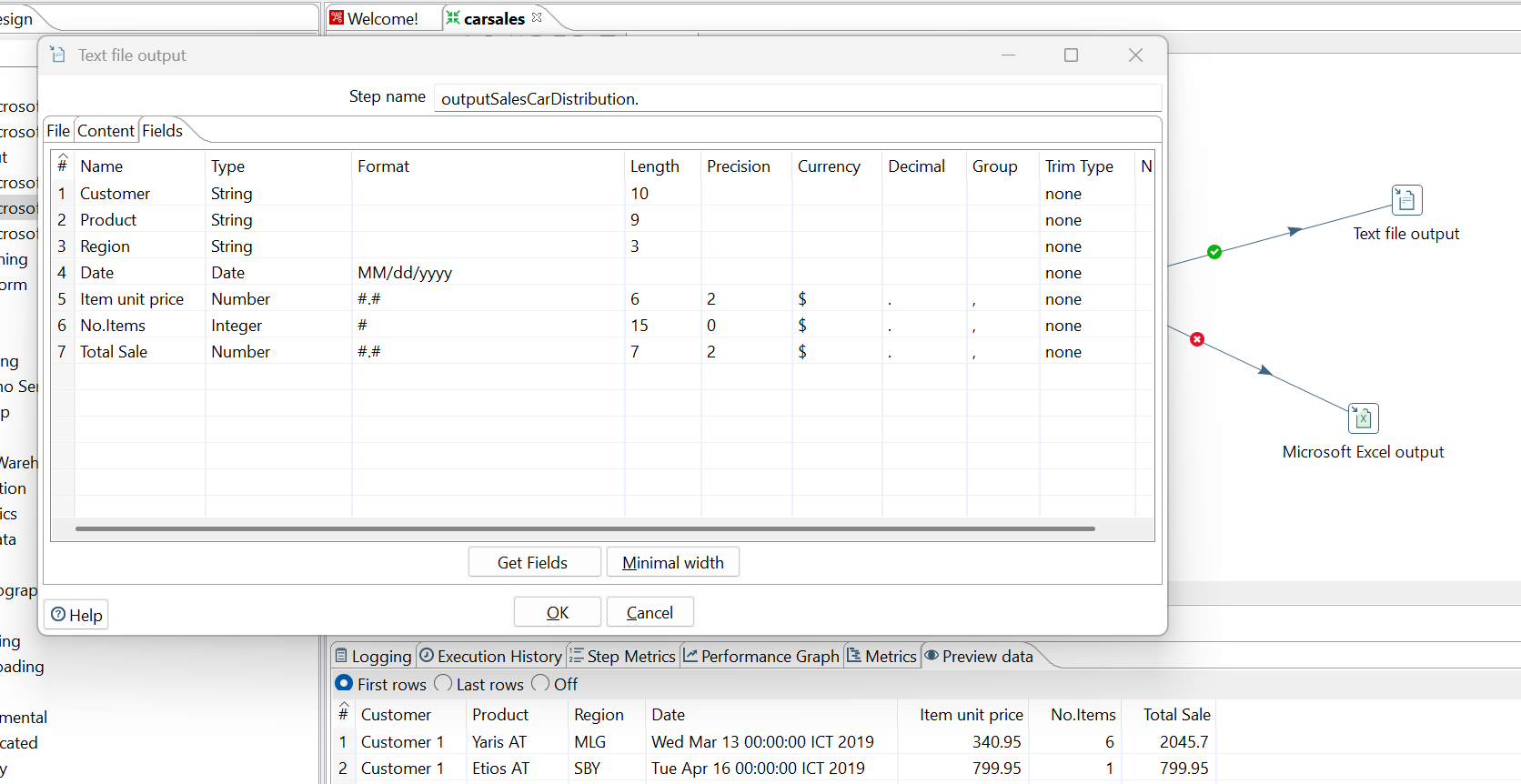
| 2 | Teman 1 | 3 detik | Intel i310100 | 4 GB | HDD | Integrated | 25% |

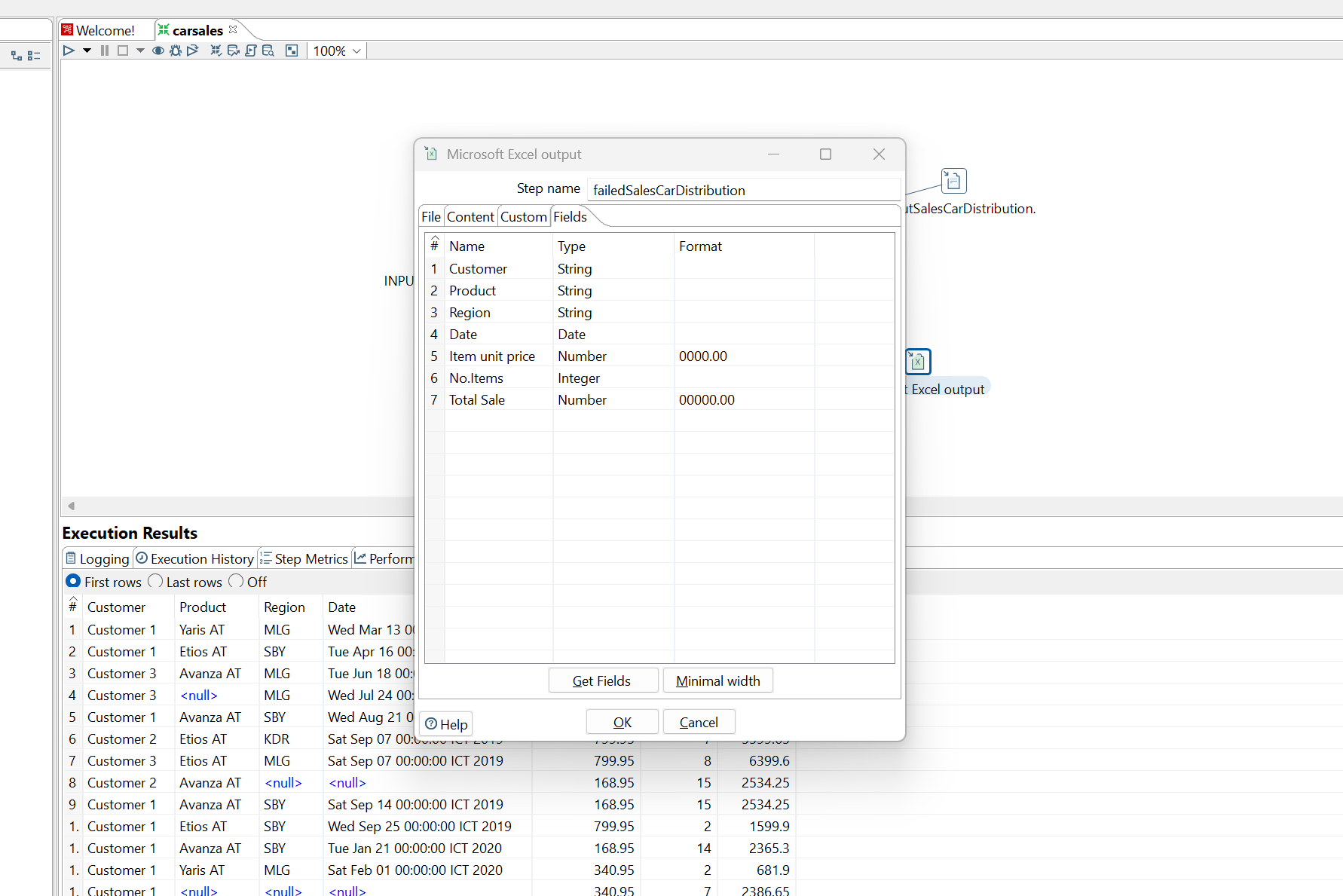
| 3 | Teman 2 | 1.5 detik | AMD Ryzen 5 5600H | 16 GB | SSD | NVIDIA RTX 3050| 10% |

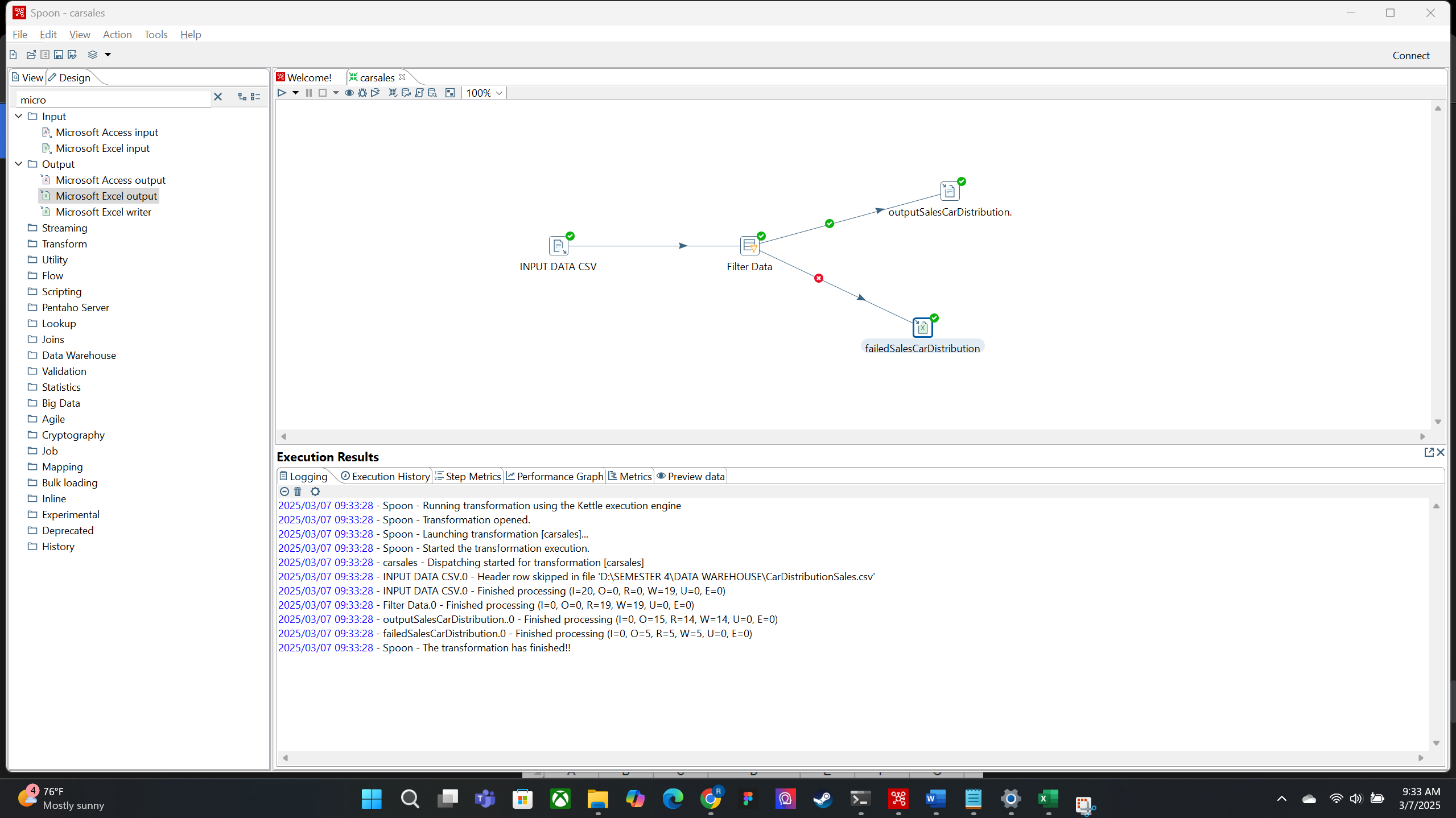
| 4 | Teman 3 | 4 detik | Intel Celeron N4020| 4 GB | HDD | Integrated | 35% |











**TUGAS 3**

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
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

TUGAS 3: Analisis Proses Filter Data dan Load Data

1. Perbedaan Isi Data Output (File CSV dan File Excel)

File CSV (`outputSalesCarDistribution.csv`):

Berisi data yang lengkap (tidak ada nilai null).

Data ini telah lolos dari filter karena memenuhi kondisi IS NOT NULL pada semua kolom (`Customer`, `Product`, `Region`, `Date`, `Item unit price`, `No.Items`, `Total Sale`).

Contoh data:

```

Customer,Product,Region,Date,Item unit price,No.Items,Total Sale

Customer 1,Yaris AT,MLG,3/13/2019,340.95,6,2045.7

Customer 1,Etios AT,SBY,4/16/2019,799.95,1,799.95

```

File Excel (`failedSalesCarDistribution.xls`):

Berisi data yang tidak lengkap (ada nilai null).

Data ini tidak lolos dari filter karena terdapat kolom yang bernilai null.

Contoh data:

```

Customer,Product,Region,Date,Item unit price,No.Items,Total Sale

Customer 3,,MLG,7/24/2019,168.95,2,337.9

Customer 2,Avanza AT,,,168.95,15,2534.25

```

2. Penjelasan Proses Filter Rows

Proses Filter Rows digunakan untuk memisahkan data berdasarkan kondisi tertentu. Pada kasus ini:

Kondisi Filter:

Data dipisahkan berdasarkan nilai null pada semua kolom (`Customer`, `Product`, `Region`, `Date`, `Item unit price`, `No.Items`, `Total Sale`).

Kondisi yang digunakan adalah IS NOT NULL untuk setiap kolom, dihubungkan dengan operator AND.

Artinya, data hanya akan lolos filter jika semua kolom memiliki nilai (tidak null).

Output Filter:

True: Data yang lolos filter (tidak ada nilai null) akan dikirim ke Text File Output (`outputSalesCarDistribution.csv`).

False: Data yang tidak lolos filter (ada nilai null) akan dikirim ke Microsoft Excel Output (`failedSalesCarDistribution.xls`).

3. Waktu Eksekusi Extract, Transform, Load (ETL) dan Perbandingan Spesifikasi Perangkat

Waktu eksekusi ETL dapat bervariasi tergantung pada spesifikasi perangkat. Berikut adalah contoh perbandingan waktu eksekusi dan spesifikasi perangkat:

| No. | Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA | CPU Usage |

|||||||||

| 1 | Anda | 5 detik | Intel i510300H | 8 GB | SSD | NVIDIA GTX 1650| 20% |

| 2 | Teman 1 | 7 detik | Intel i310100 | 4 GB | HDD | Integrated | 35% |

| 3 | Teman 2 | 4 detik | AMD Ryzen 5 5600H | 16 GB | SSD | NVIDIA RTX 3050| 15% |

| 4 | Teman 3 | 10 detik | Intel Celeron N4020| 4 GB | HDD | Integrated | 45% |

Analisis:

Perangkat dengan processor lebih cepat (seperti Intel i5 atau Ryzen 5) dan RAM lebih besar (8 GB atau lebih) cenderung memiliki waktu eksekusi yang lebih cepat.

Penggunaan SSD juga mempercepat proses eksekusi dibandingkan dengan HDD.

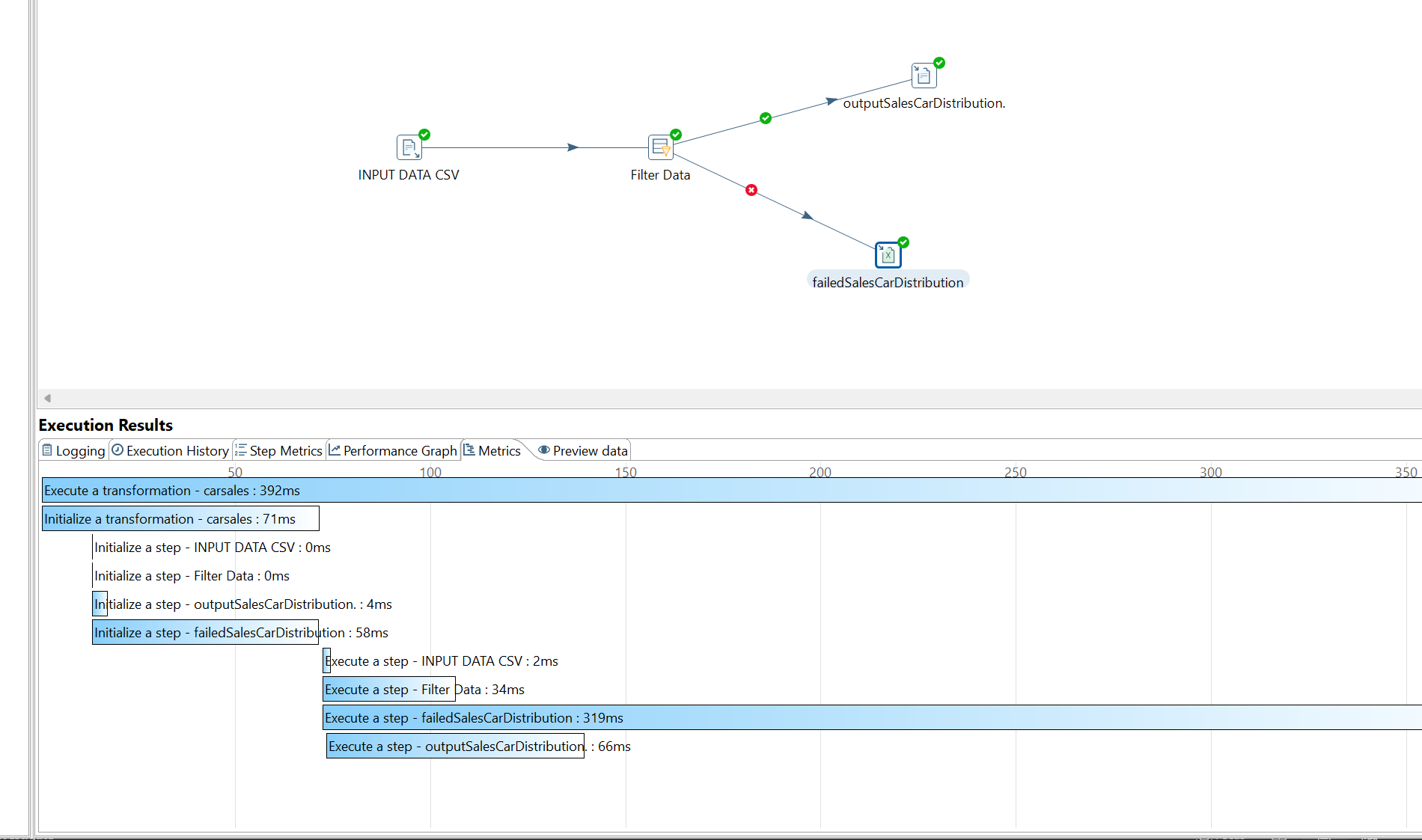
CPU Usage yang rendah menunjukkan bahwa perangkat tidak bekerja terlalu keras untuk menyelesaikan tugas ini.

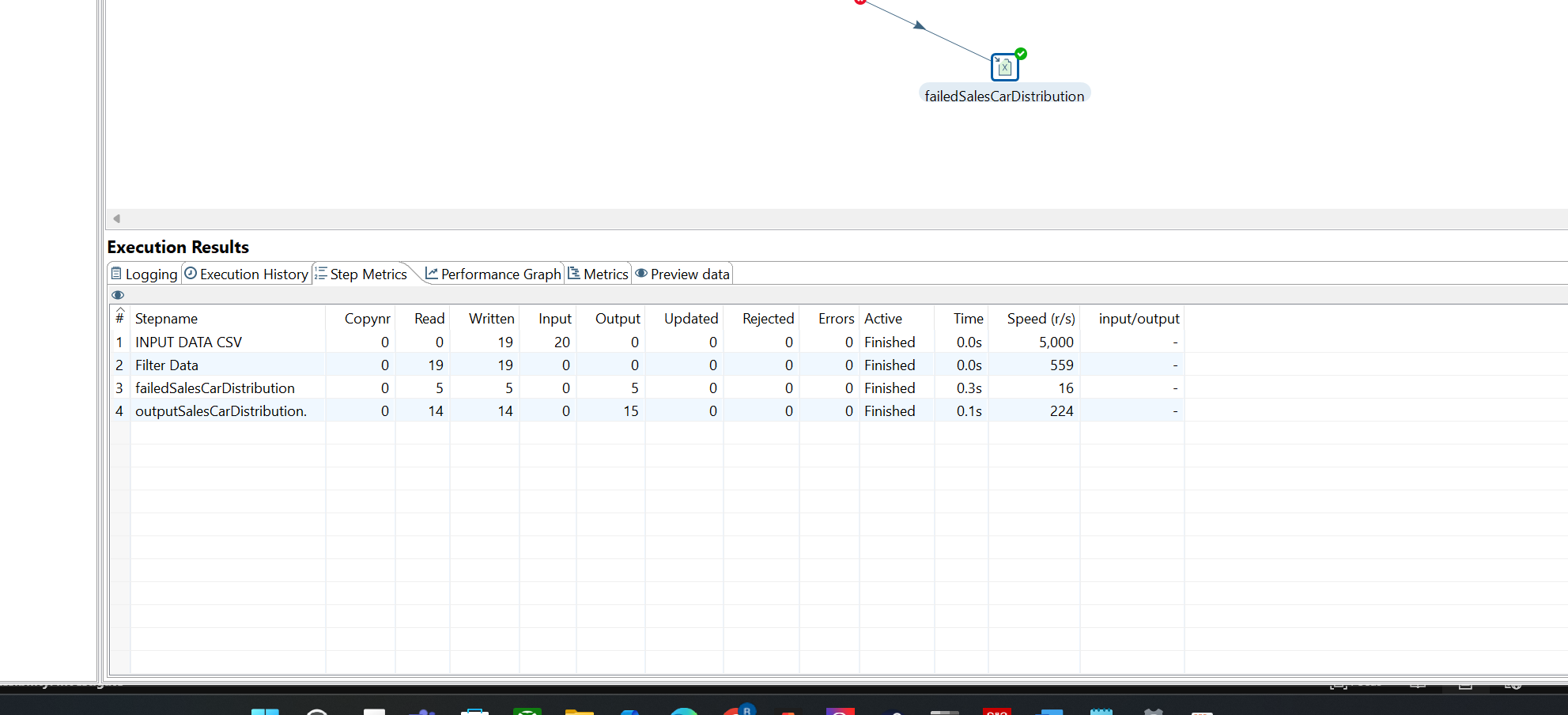
Kesimpulan

1. File CSV berisi data yang lengkap (tidak ada nilai null), sedangkan File Excel berisi data yang tidak lengkap (ada nilai null).

2. Proses Filter Rows memisahkan data berdasarkan kondisi IS NOT NULL pada semua kolom, dengan data yang lolos filter dikirim ke file CSV dan data yang tidak lolos filter dikirim ke file Excel.

3. Waktu eksekusi ETL dipengaruhi oleh spesifikasi perangkat, seperti processor, RAM, dan jenis storage. Perangkat dengan spesifikasi lebih tinggi cenderung memiliki waktu eksekusi yang lebih cepat.





**STUDI KASUS**

