


|   |  |   |
|---|--|---|
| <b>Nama:</b><br>Muhammad Rasya<br>Hadziq<br><b>NIM:</b><br>065001900033 | <br><b>Praktikum Data Warehouse</b> | <b>MODUL 11</b><br><br><b>Nama Dosen:</b><br>Ir. Teddy Siswanto, MMSi   |
| <b>Hari/Tanggal:</b><br>Sabtu, 09 07 2022                               |  | <b>Nama Asisten Labratorium:</b><br>1. Azhar Rizki Zulma<br>065001900001<br>2. Nadiya Amanda Rizkania<br>064001900003 |

## Proyek Akhir Praktikum III

### 1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

### Cleansing Data

Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:



## 1. Data Preparation

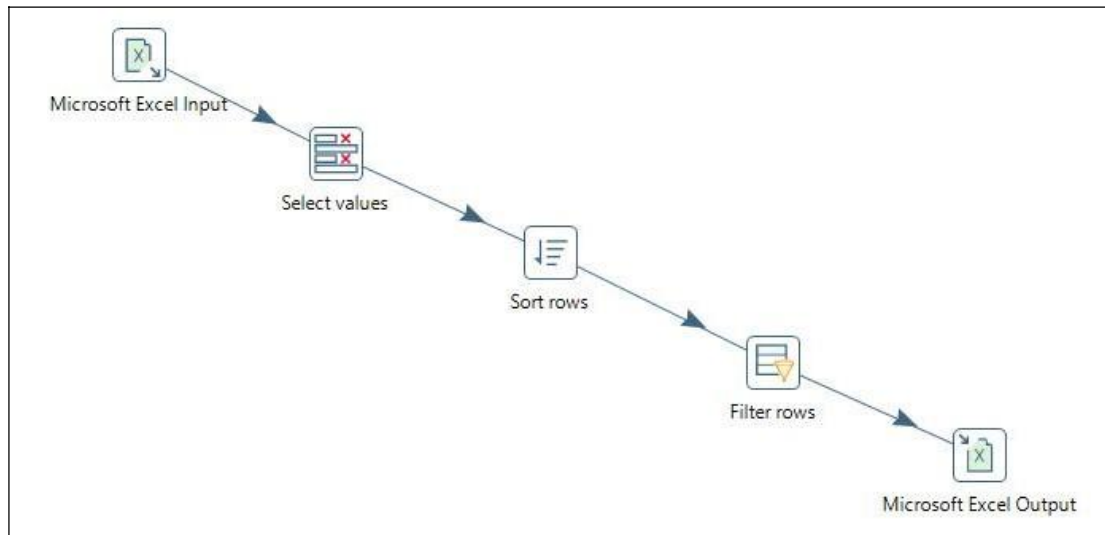
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



| Name                                     | Date modified    | Type                  | Size   |
|--|------------------|-----------------------|--------|
| Cabai Merah Keriting Desember 2018       | 14/11/2020 21:16 | Microsoft Excel C...  | 10 KB  |
| Cabai Merah Keriting November 2018       | 14/11/2020 21:14 | Microsoft Excel C...  | 10 KB  |
| Cabai Merah Keriting Oktober 2018        | 14/11/2020 21:02 | Microsoft Excel C...  | 10 KB  |
| Cabai Merah Keriting September 2018      | 14/11/2020 20:45 | Microsoft Excel C...  | 10 KB  |
| data cabai keriting merah 2018 (jan ags) | 01/05/2021 12:10 | Microsoft Excel 97... | 108 KB |

## 2. Excel Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2020.ktr***



- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

The top screenshot shows the 'Microsoft Excel input' step configuration. The 'Files' tab is active, showing 'Spread sheet type (engine)' as 'Excel 97-2003 XLS (JXL)'. Below this are fields for 'File or directory', 'Regular Expression', 'Exclude Regular Expression', and 'Password'. A 'Selected files' list contains one entry: 'C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Data Warehouse\Praktikum Video Cleanshing and Filter'. Below the list are checkboxes for 'Accept filenames from previous steps' and 'Accept filenames from previous step', along with dropdowns for 'Step to read filenames from' and 'Field in the input to use as filename'. A 'Show filename(s)...' button is at the bottom. The bottom screenshot shows the 'Sheets' tab with a table of sheet names and start positions. An 'Enter list' dialog box is open, showing 'Available items' (januari 2018, februari 2018, maret 2018, april 2018, mei 2018, juni 2018, juli 2018) and 'Your selection' (agustus 2018). The dialog has buttons for '>', '<', '<<', and '>>', and an 'Add the selected items on the left' button. The 'Get sheetname(s)...' button is at the bottom of the 'Sheets' tab.

Step name: Microsoft Excel Input

Files | Sheets | Content | Error Handling | Fields | Additional output fields

Spread sheet type (engine): Excel 97-2003 XLS (JXL)

File or directory: [Add] [Browse...]

Regular Expression: [ ]

Exclude Regular Expression: [ ]

Password: [ ]

Selected files:

| # | File/Directory   |
|---|--|
| 1 | C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Data Warehouse\Praktikum Video Cleanshing and Filter |

Accept filenames from previous steps: ☐

Accept filenames from previous step: ☐

Step to read filenames from: [ ]

Field in the input to use as filename: [ ]

Show filename(s)...

OK | Preview rows | Cancel

Step name: Microsoft Excel Input

Files | Sheets | Content | Error Handling | Fields | Additional output fields

List of sheets to read:

| # | Sheet name   | Start row | Start column |
|---|--------------|-----------|--------------|
| 1 | agustus 2018 | 0         | 0            |

Get sheetname(s)...

OK | Preview rows | Cancel

Enter list

Available items:

- januari 2018
- februari 2018
- maret 2018
- april 2018
- mei 2018
- juni 2018
- juli 2018

Your selection:

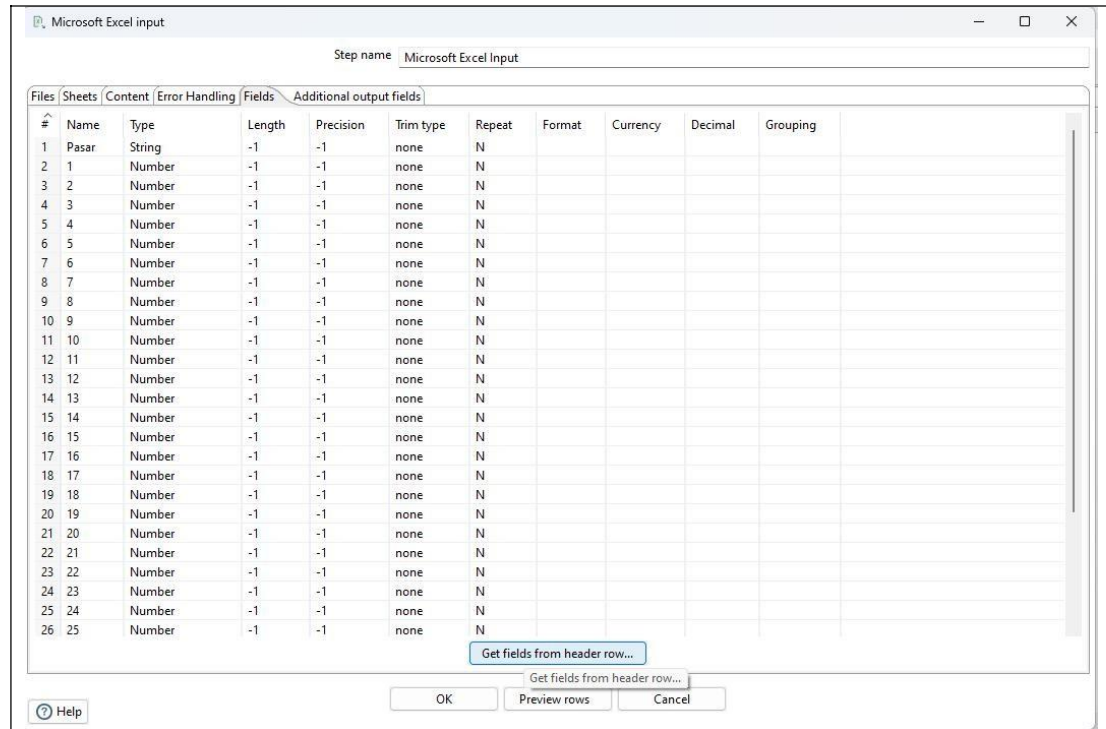
- agustus 2018

> | < | << | >>

Add the selected items on the left

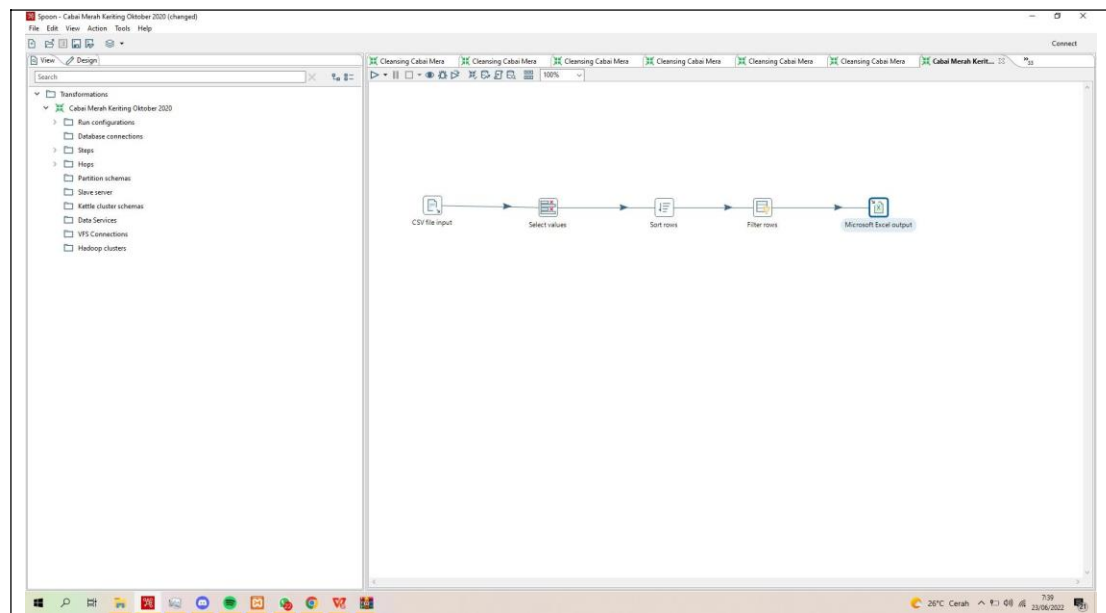
OK | Cancel





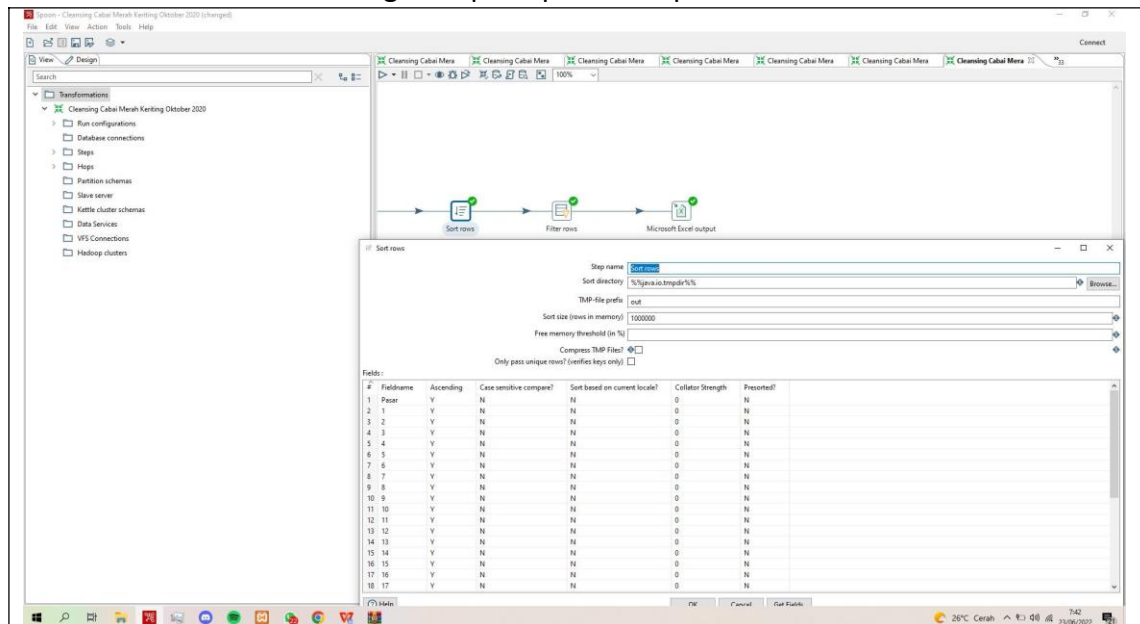
### 3. CSV Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Oktober 2020.ktr***

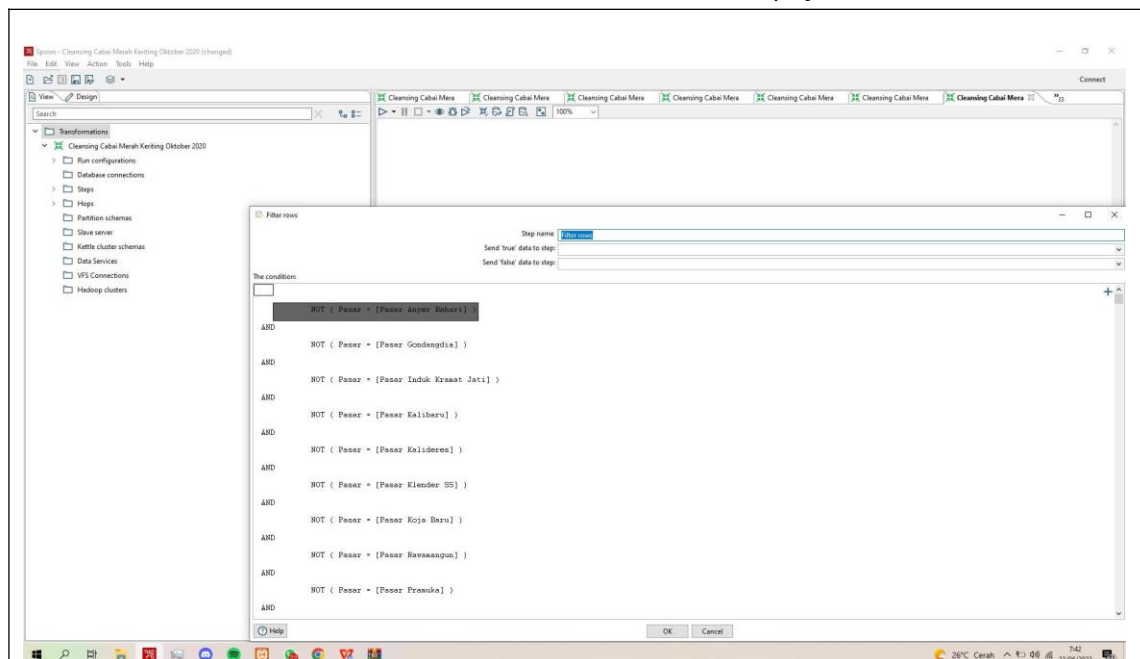


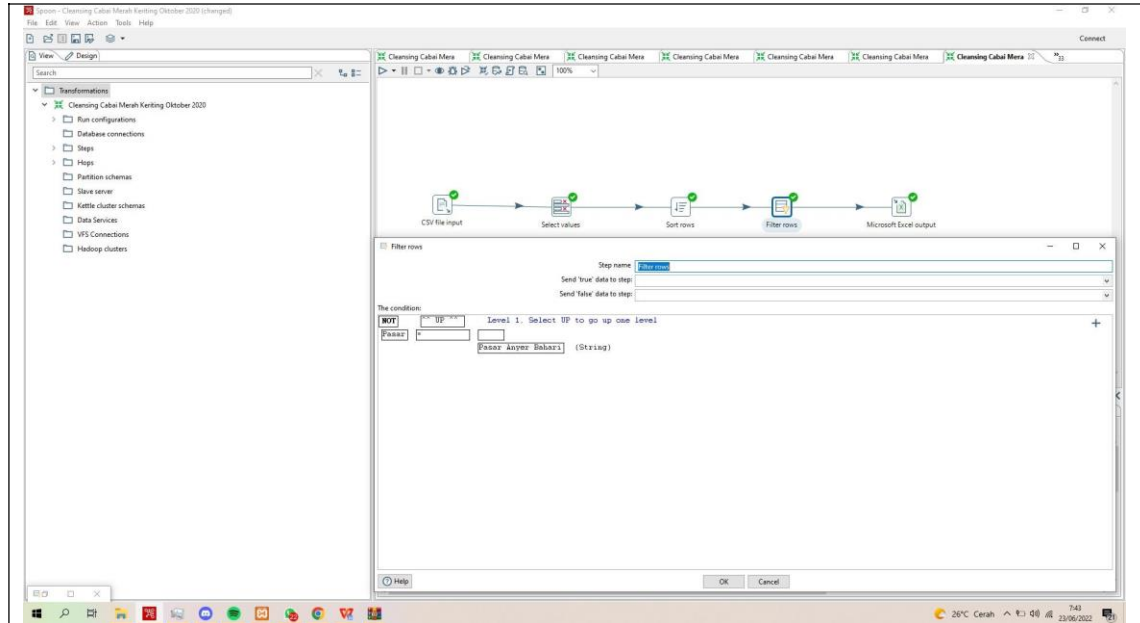


- Kemudian lakukan ascending data pada proses step Sort Rows

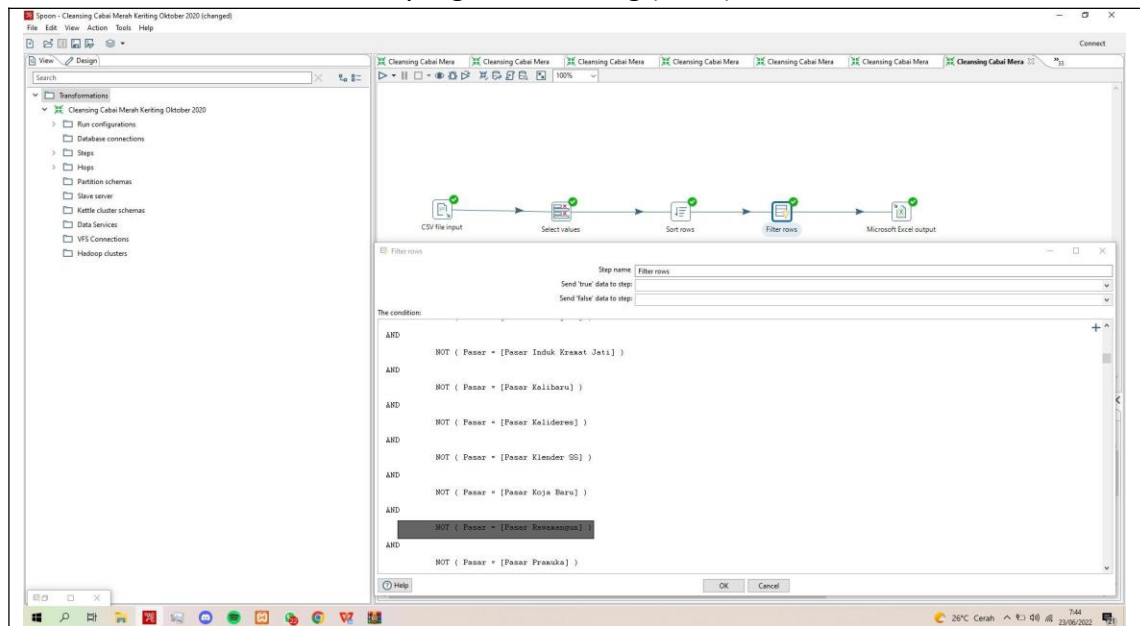


- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.





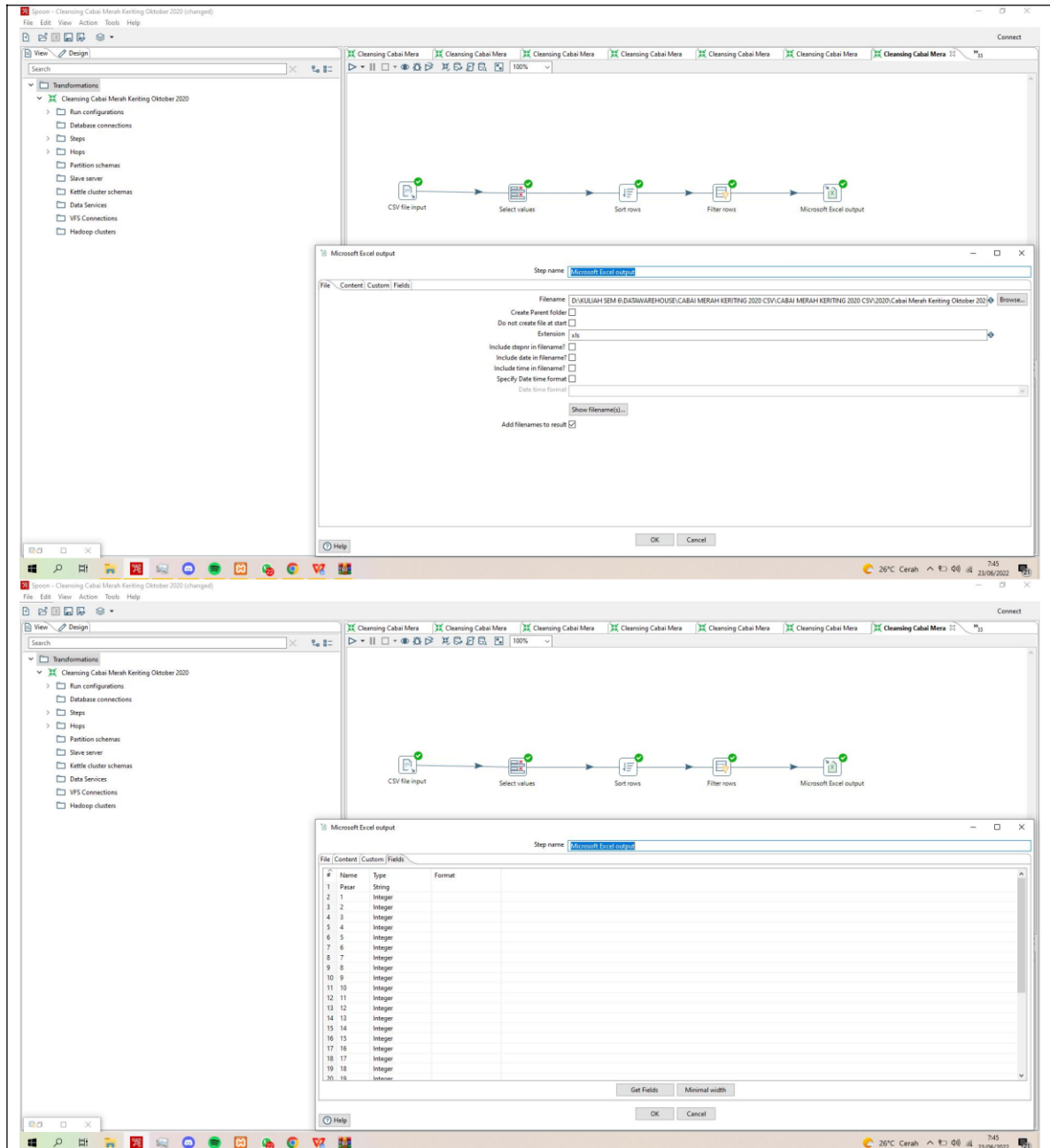
- Berikut List Data Nama Pasar yang mau dibuang (2020)



## 5. Excel Output

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama ***Cabai Merah Keriting [Bulan] [Tahun].xls*** seperti ***Cabai Merah Keriting Agustus 2020.xls*** (File & Sheet)





## 2. Alat dan Bahan

Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

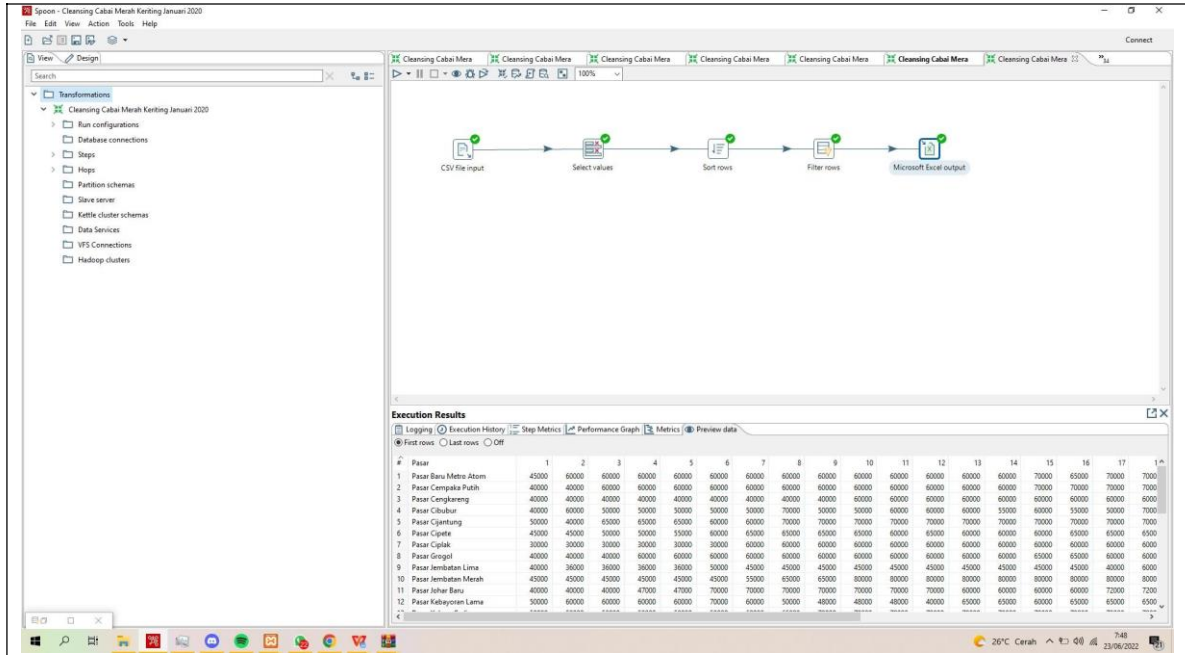
## 3. Elemen Kompetensi

a. Tugas Akhir Praktikum III – Screenshot Struktur yang sudah berhasil berjalan

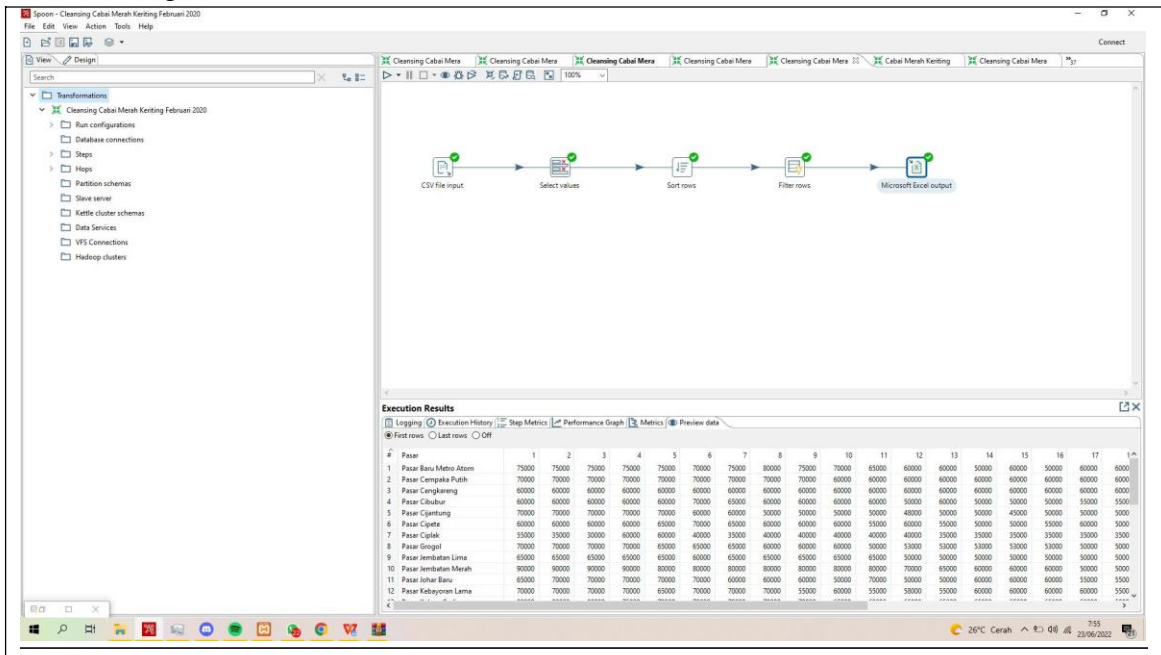
1. Cleansing Januari 2020





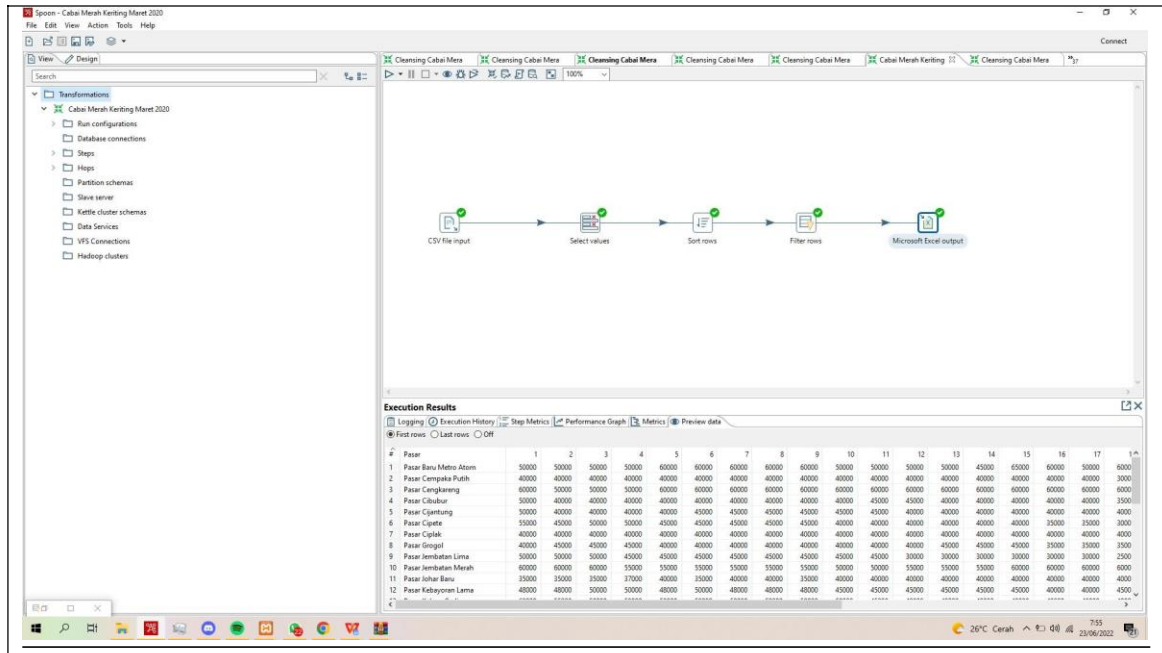


## 2. Cleansing Februari 2020

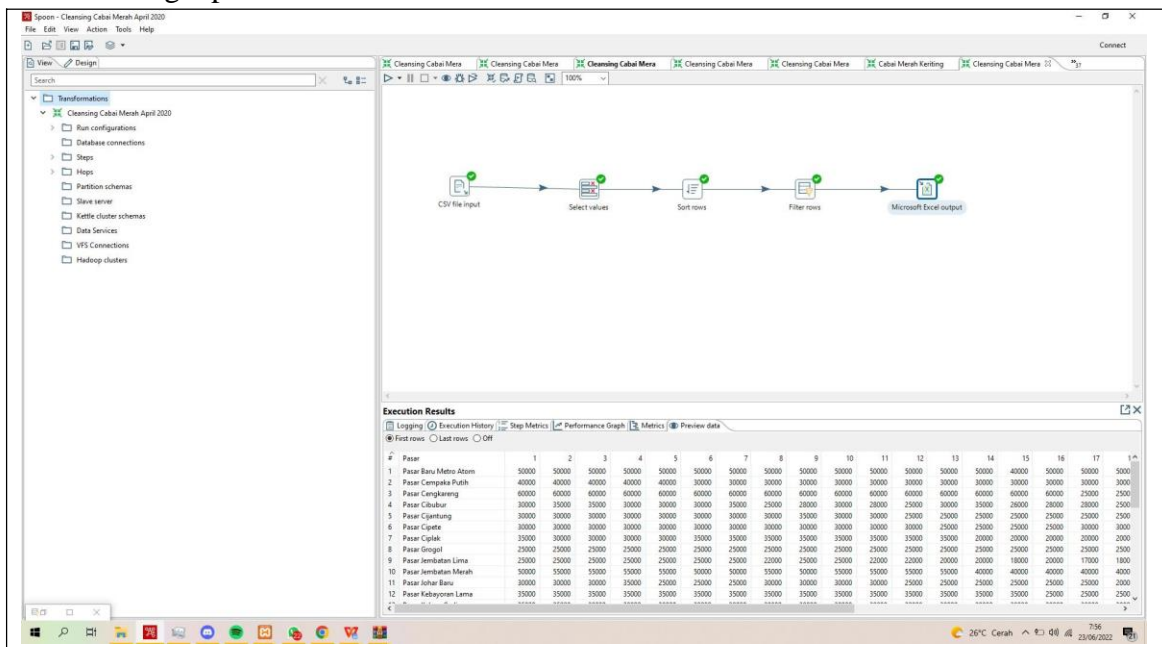


## 3. Cleansing Maret 2020



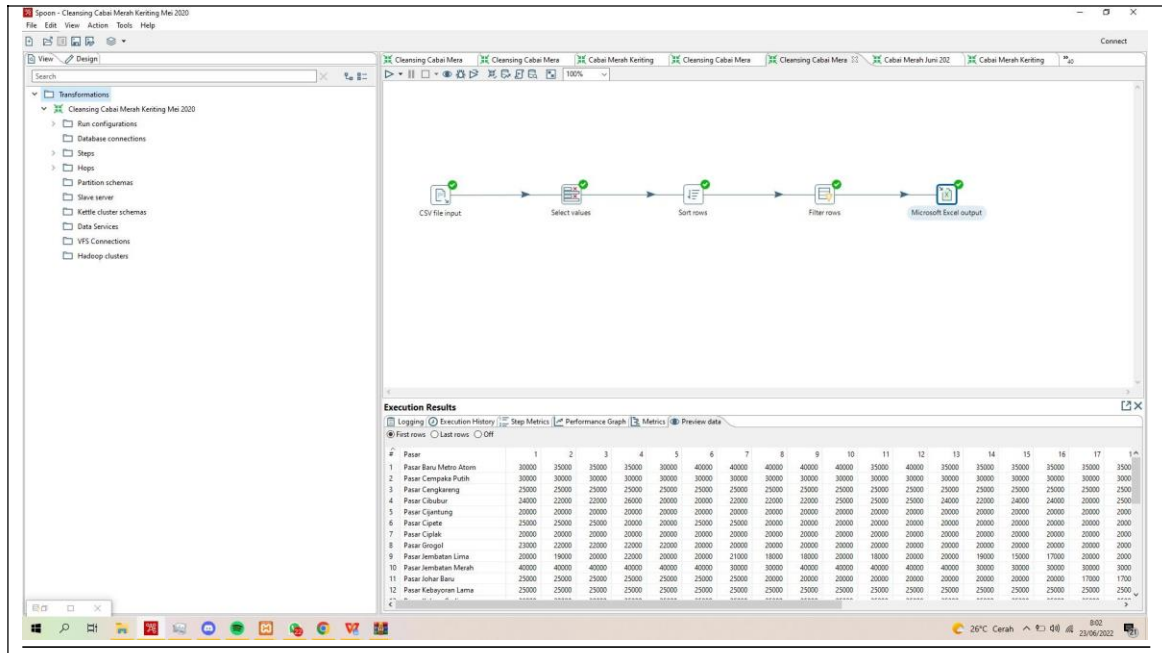


#### 4. Cleansing April 2020

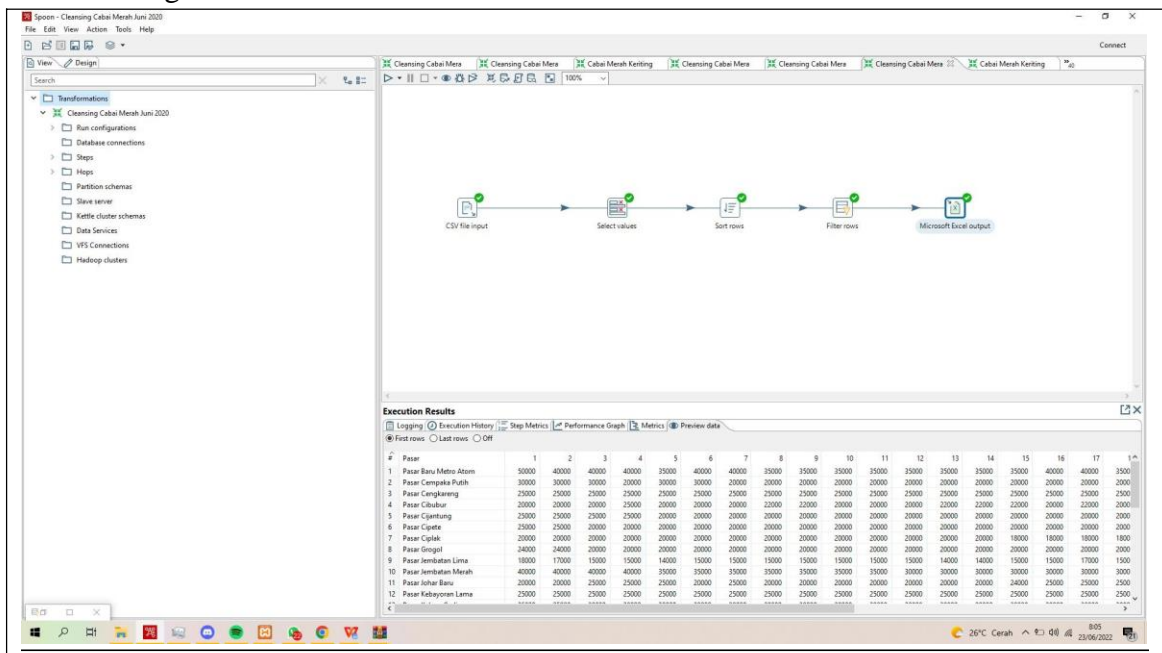


#### 5. Cleansing Mei 2020



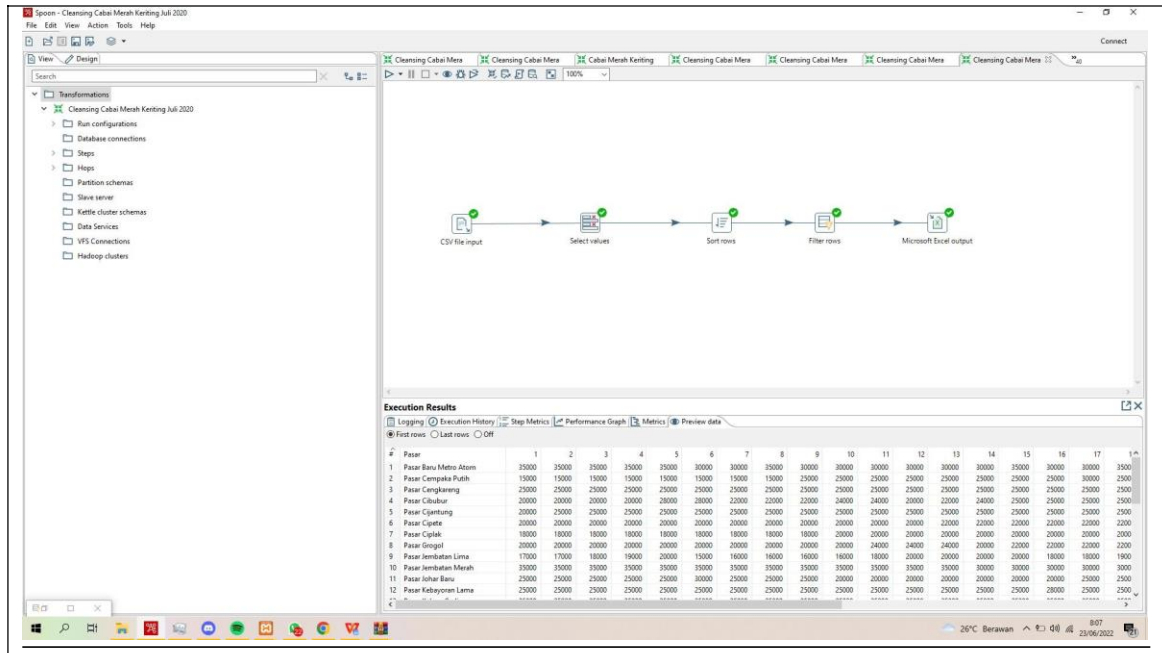


## 6. Cleansing Juni 2020

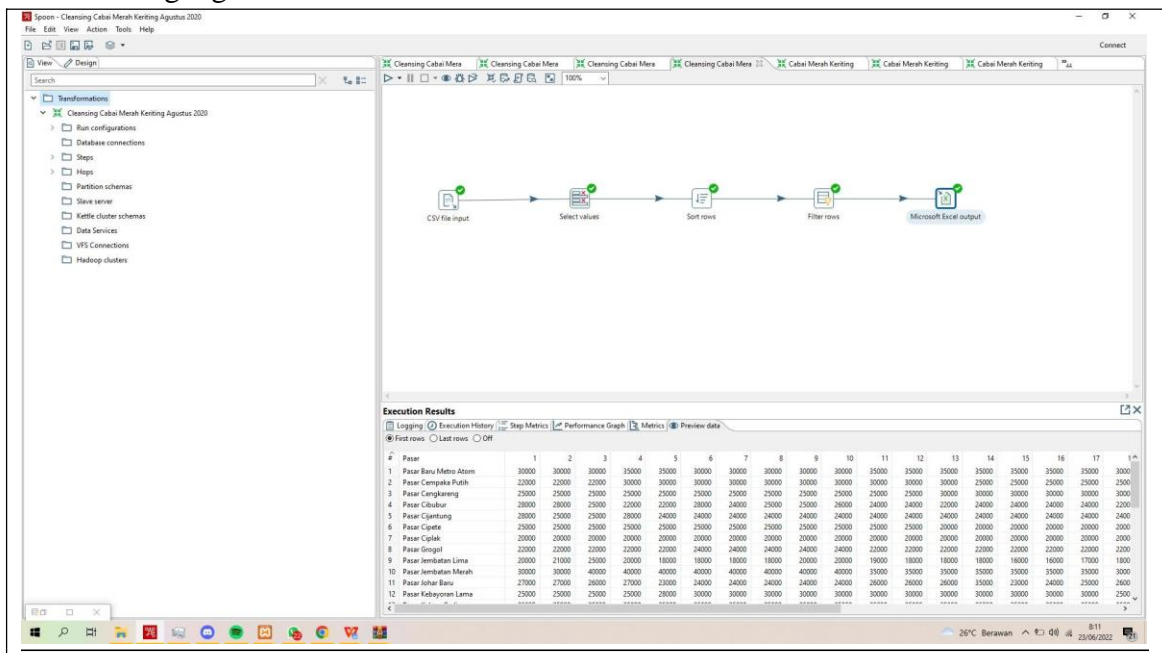


## 7. Cleansing Juli 2020



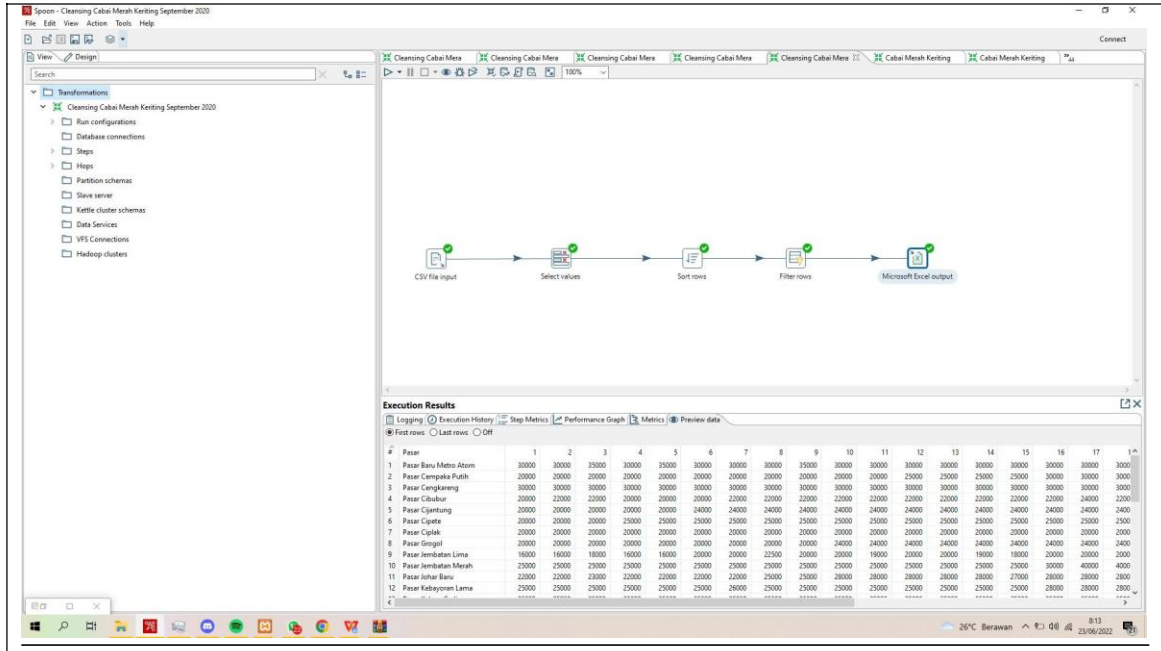


## 8. Cleansing Agustus 2020

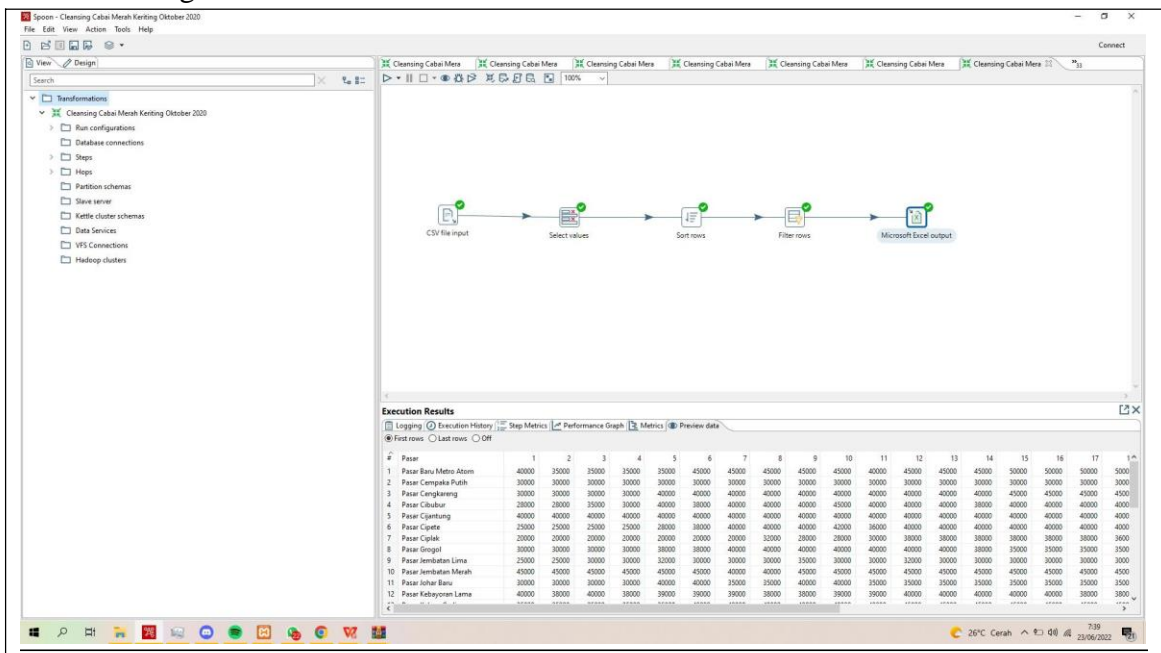


## 9. Cleansing September 2020





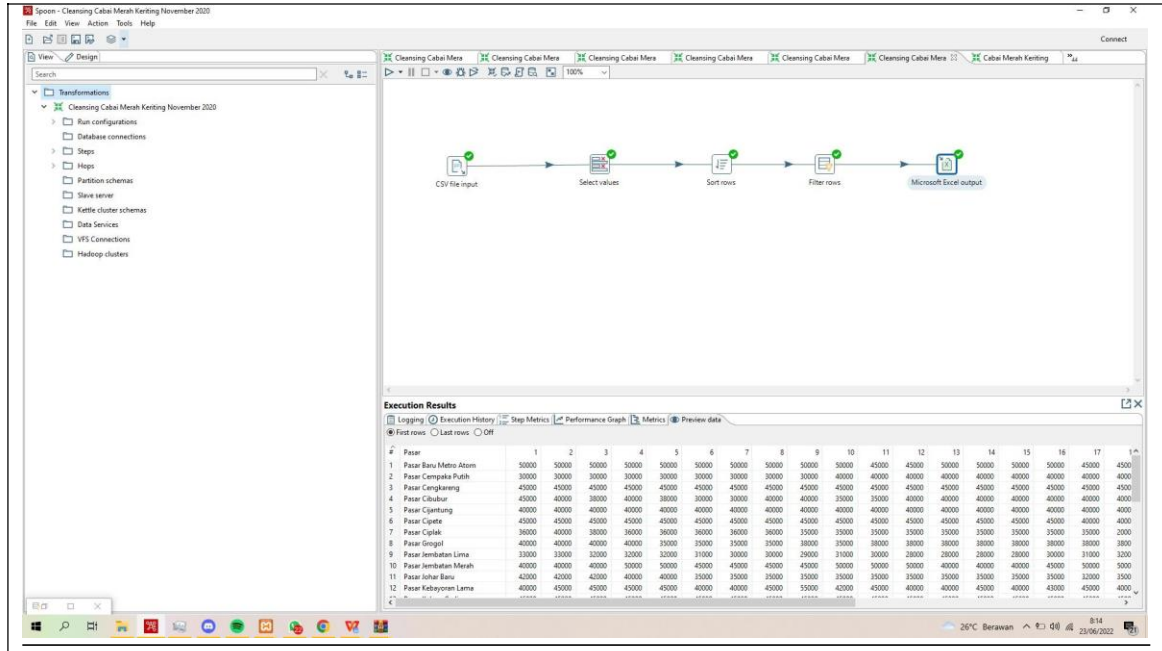
## 10. Cleansing Oktober 2020



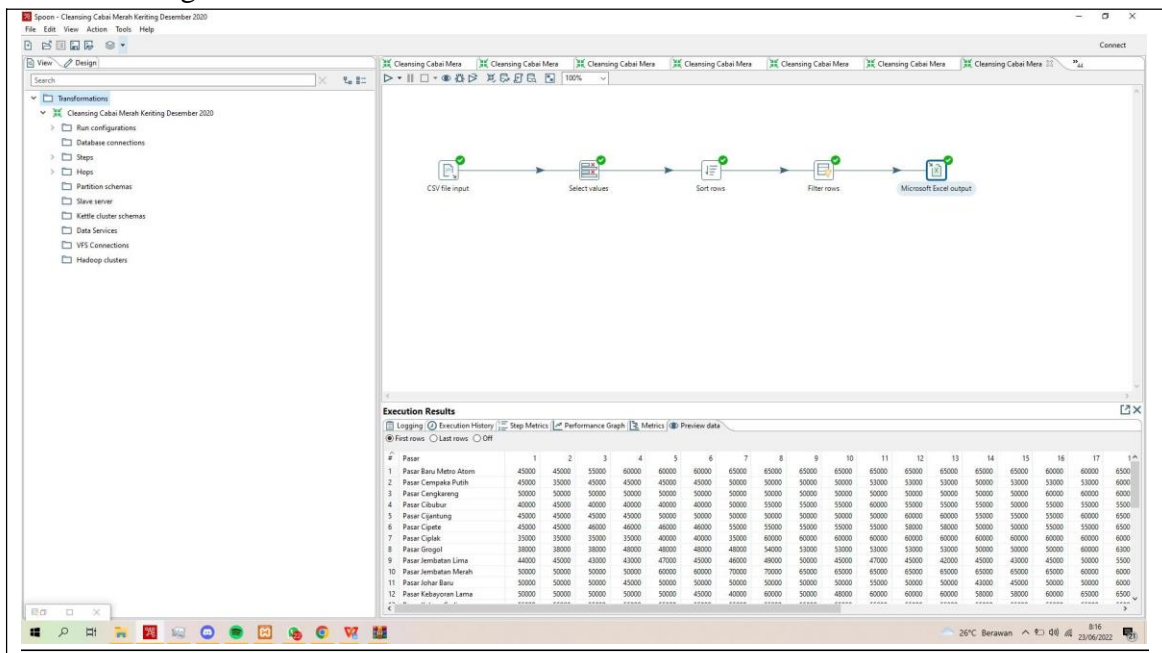
## 11. Cleansing November 2020







## 12. Cleansing Desember 2020



## 4. File Praktikum

Github Repository:

<https://github.com/rasyahdz/PraktikumDWH.git>

## 5. Soal Latihan



Jurusan Teknik Informatika & Sistem Informasi  
Fakultas Teknologi Industri – Universitas Trisakti

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:

1. Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (*missing value*).
2. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (*missing value*)

## 6. Kesimpulan

- a. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui...

## 7. Cek List (✓)

| No | Elemen Kompetensi         | Penyelesaian |               |
|----|---------------------------|--------------|---------------|
|    |                           | Selesai      | Tidak Selesai |
| 1. | Tugas Akhir Praktikum III | ✓            |               |

## 8. Formulir Umpan Balik

| No | Elemen Kompetensi         | Waktu Pengerjaan | Kriteria |
|----|---------------------------|------------------|----------|
| 1. | Tugas Akhir Praktikum III | 45 Menit         | Baik     |

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

