


<b>Nama:</b> Muhammad Rasya Hadziq <b>NIM:</b> 065001900033	 <b>Praktikum Data Warehouse</b>	<b>MODUL 9</b>  <b>Nama Dosen:</b> Ir. Teddy Siswanto, MMSi
<b>Hari/Tanggal:</b> Sabtu, 09 07 2022		<b>Nama Asisten Labratorium:</b> 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 064001900003

## Proyek Akhir Praktikum I

### 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

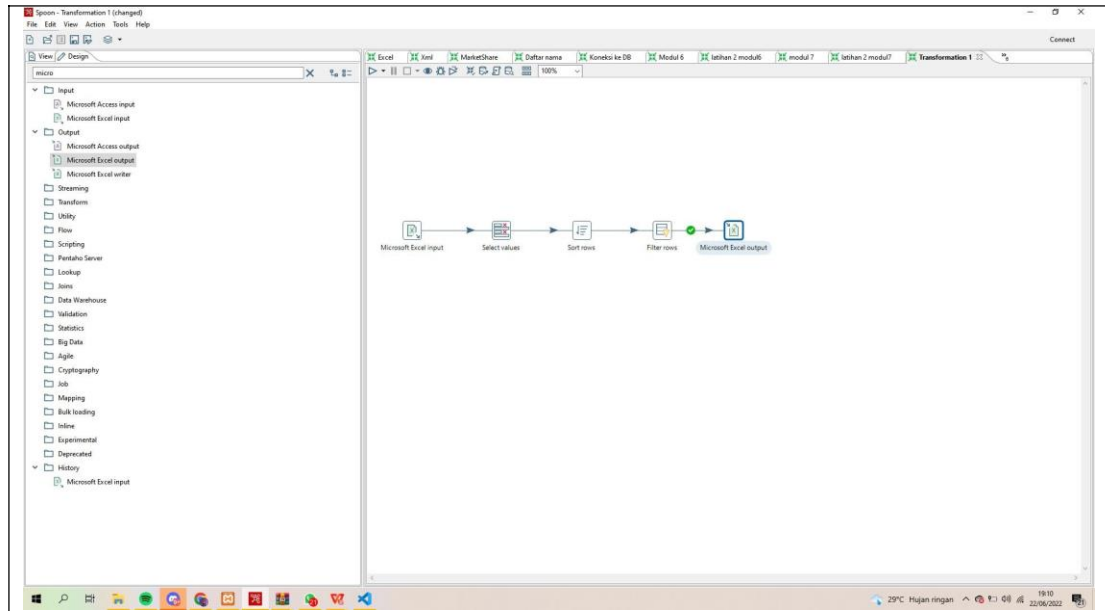
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



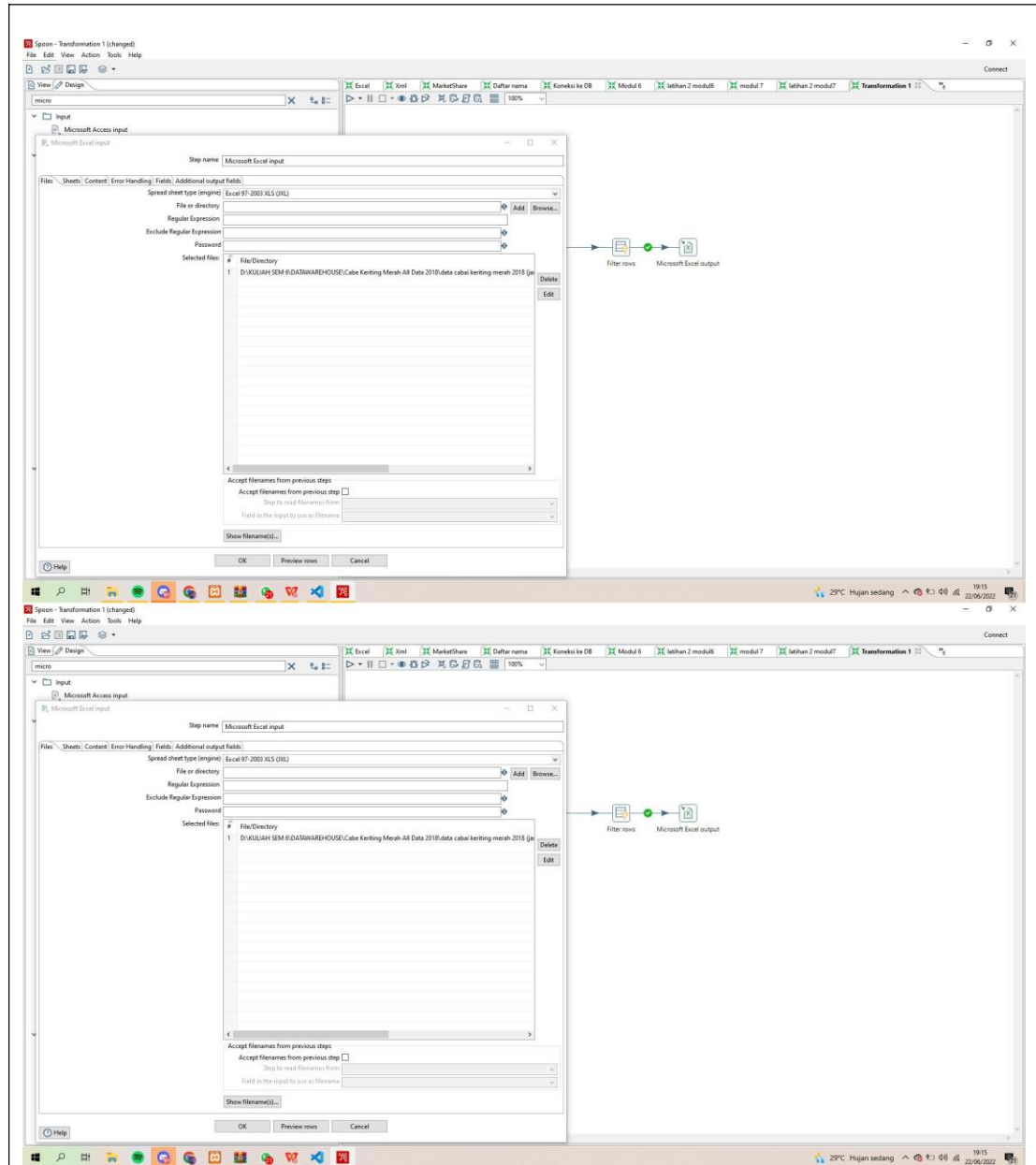
Name	Date modified	Type	Size
Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
Cabai Merah Keriting September 2018	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

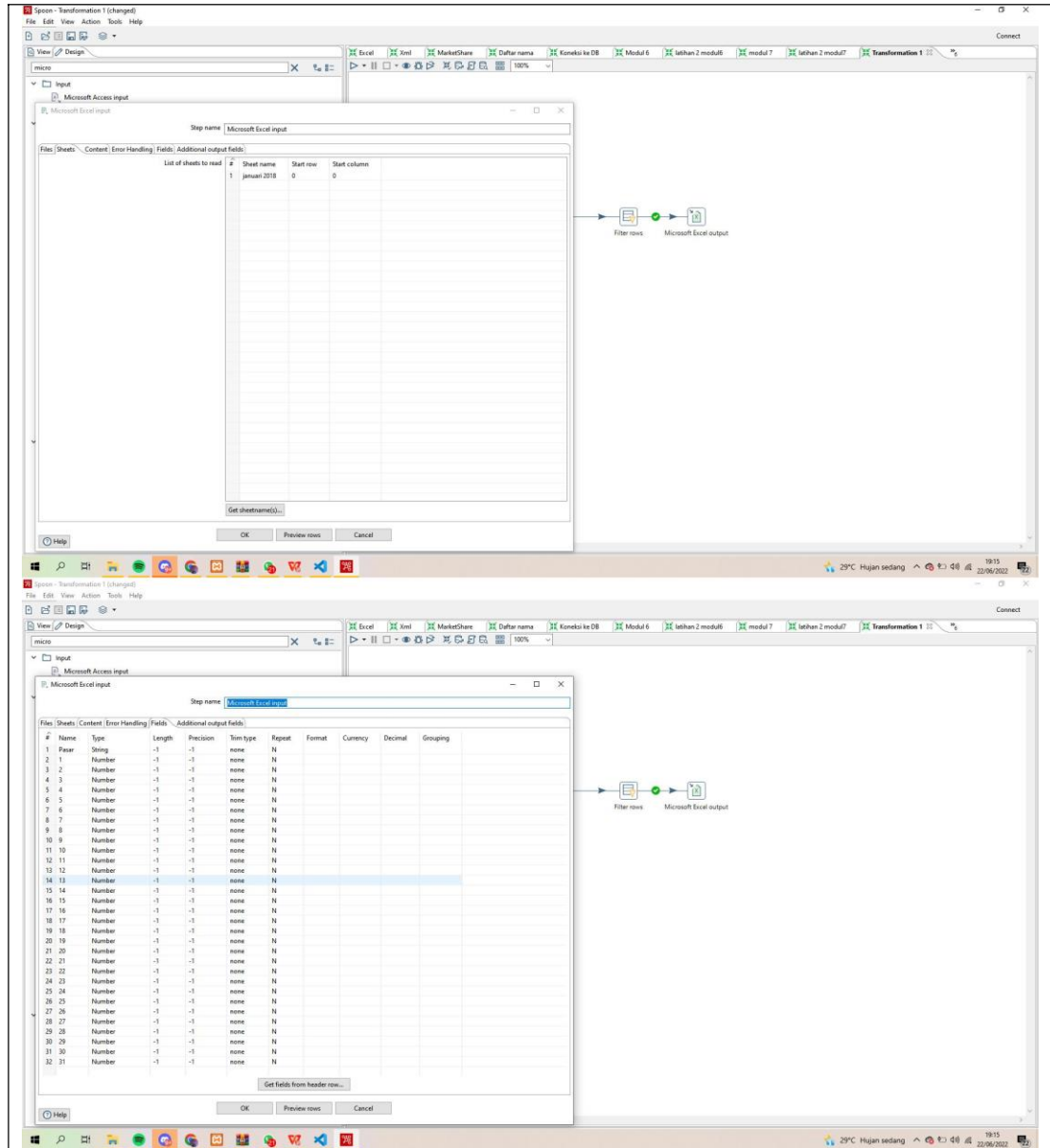
## 2. Excel Input

- 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 2018.ktr***



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

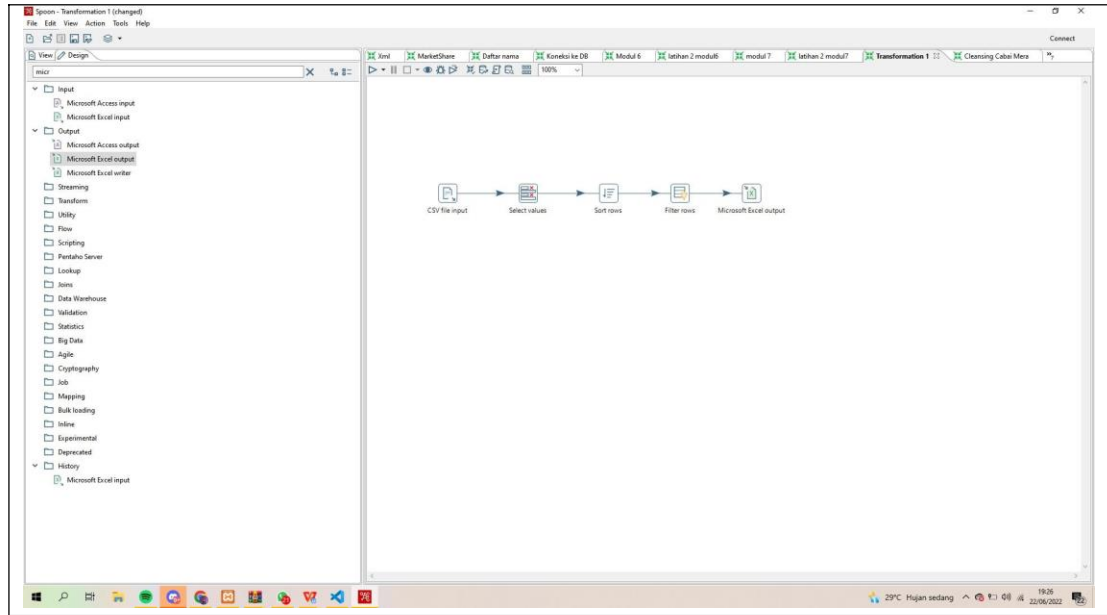




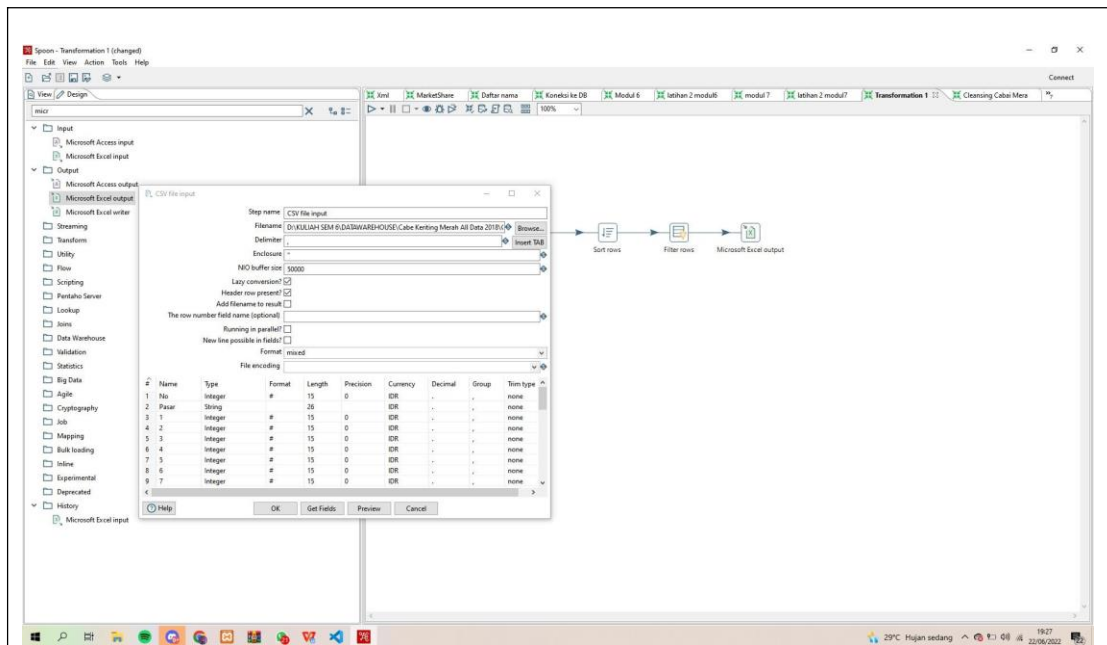
### 3. CSV Input

- 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 2018.ktr**





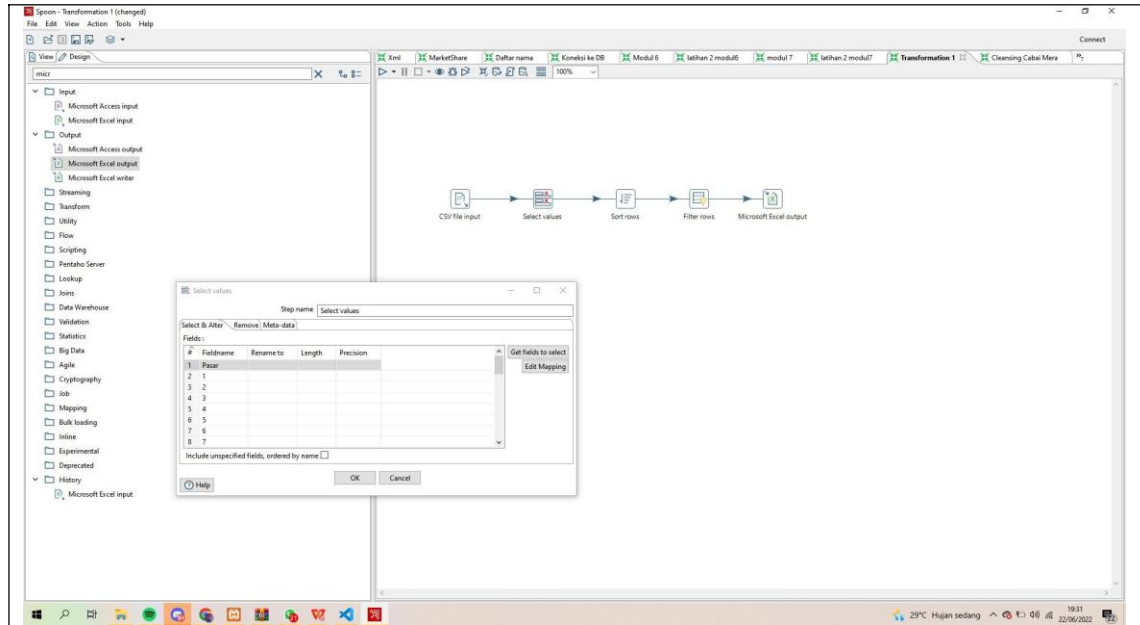
- Berikut ini adalah konfigurasi pada step CSV File Input.



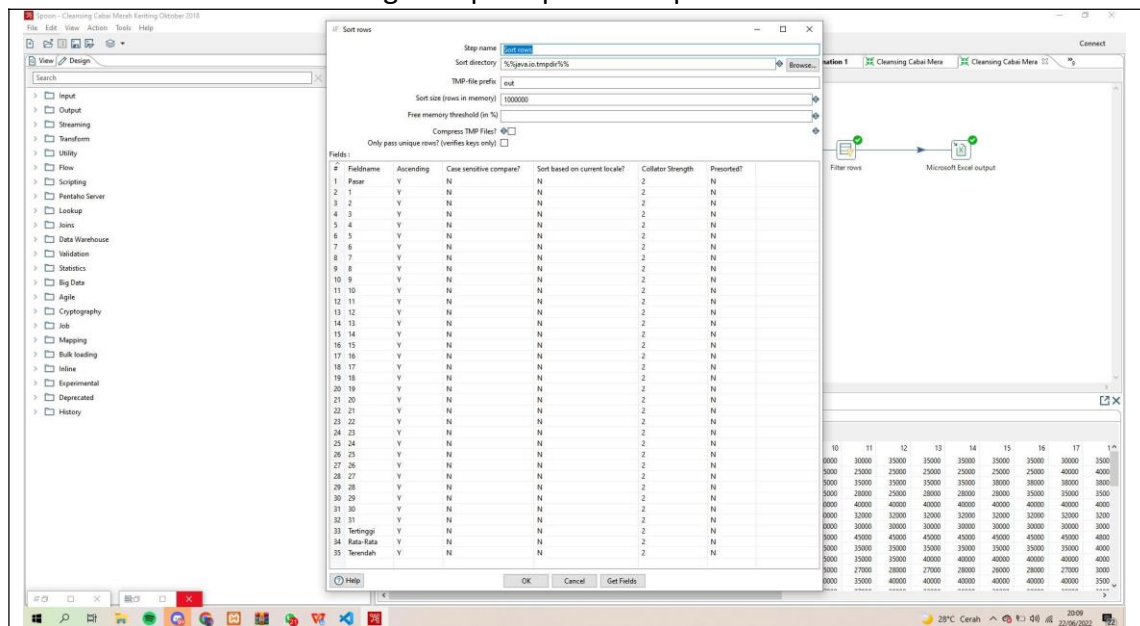
#### 4. Data Filtering

- Step selanjutnya adalah melakukan select value yang berguna untuk memilah kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.



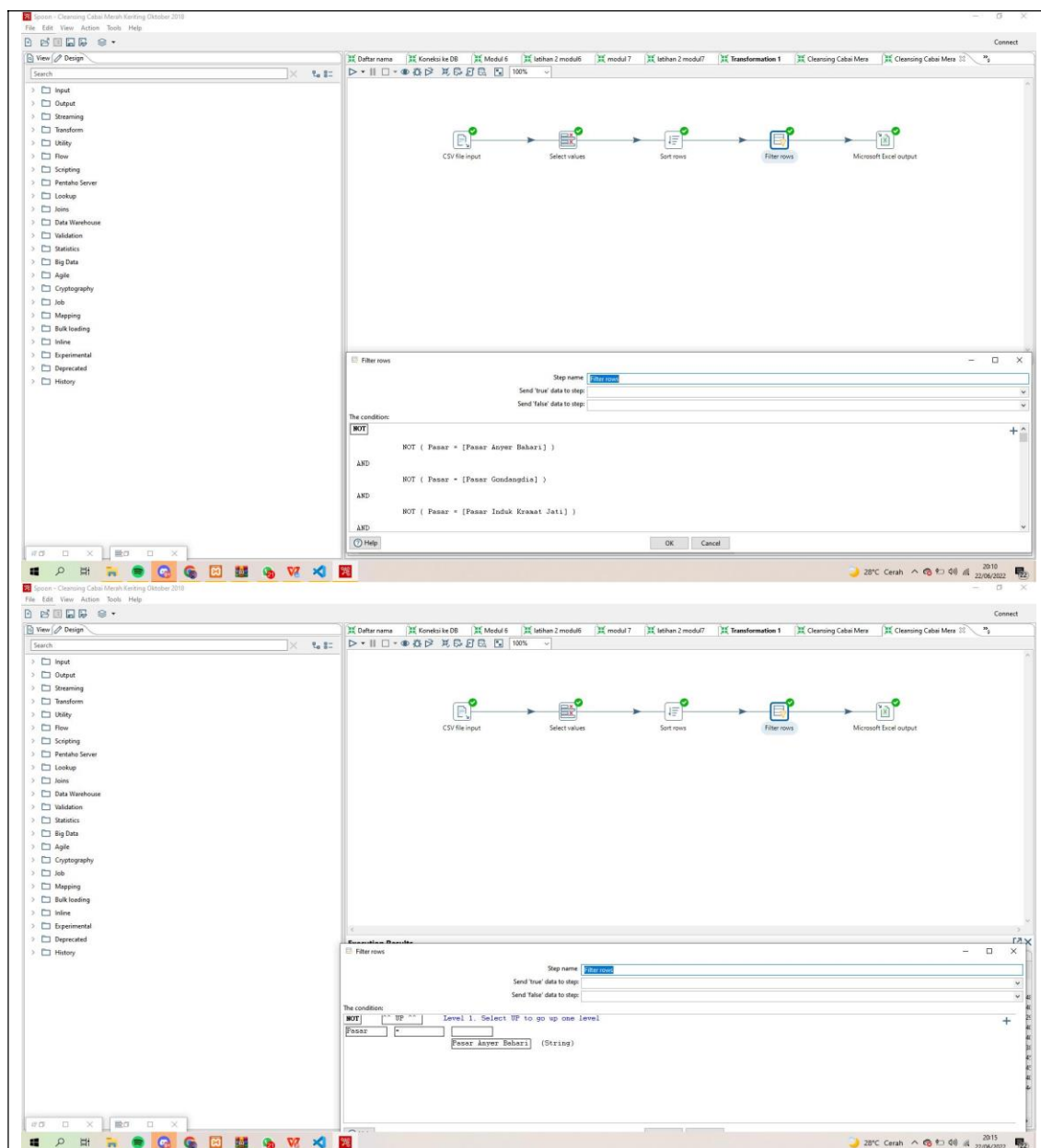


- Kemudian lakukan ascending data pada proses step Sort Rows



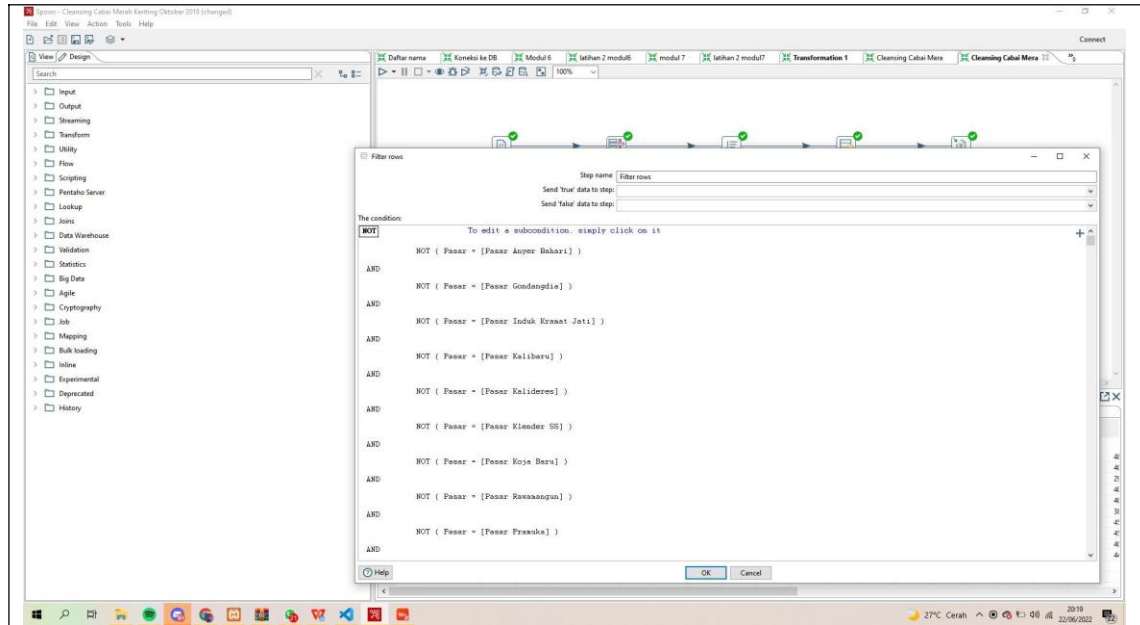
- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.





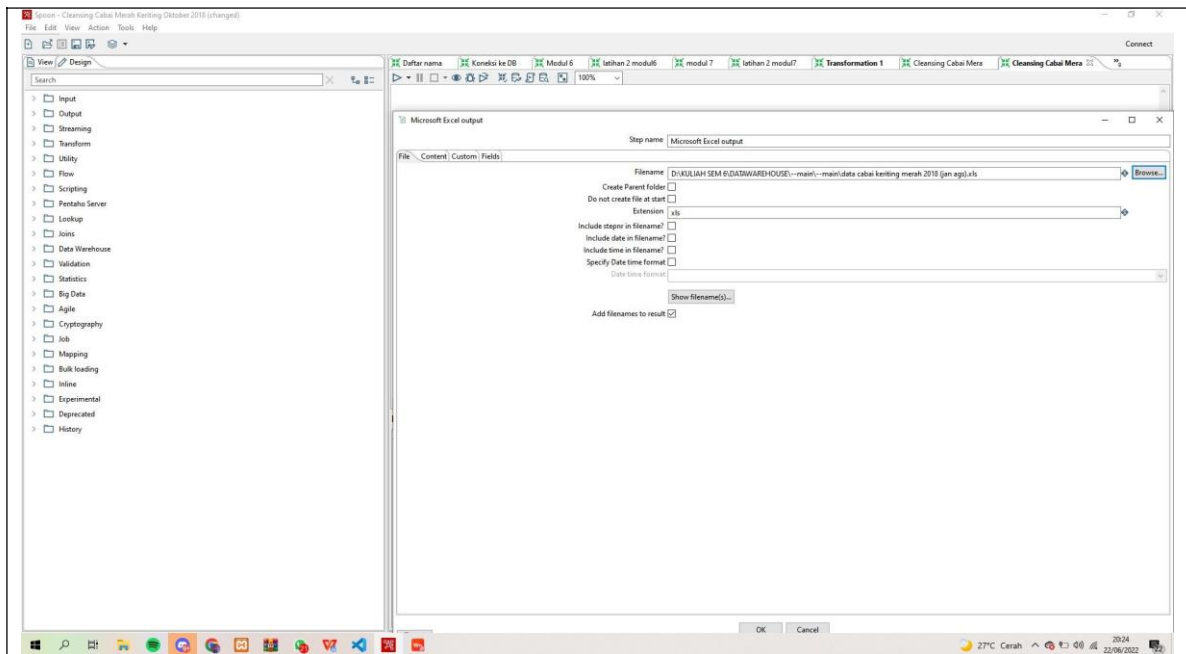
- Berikut List Data Nama Pasar yang mau dibuang (2018)



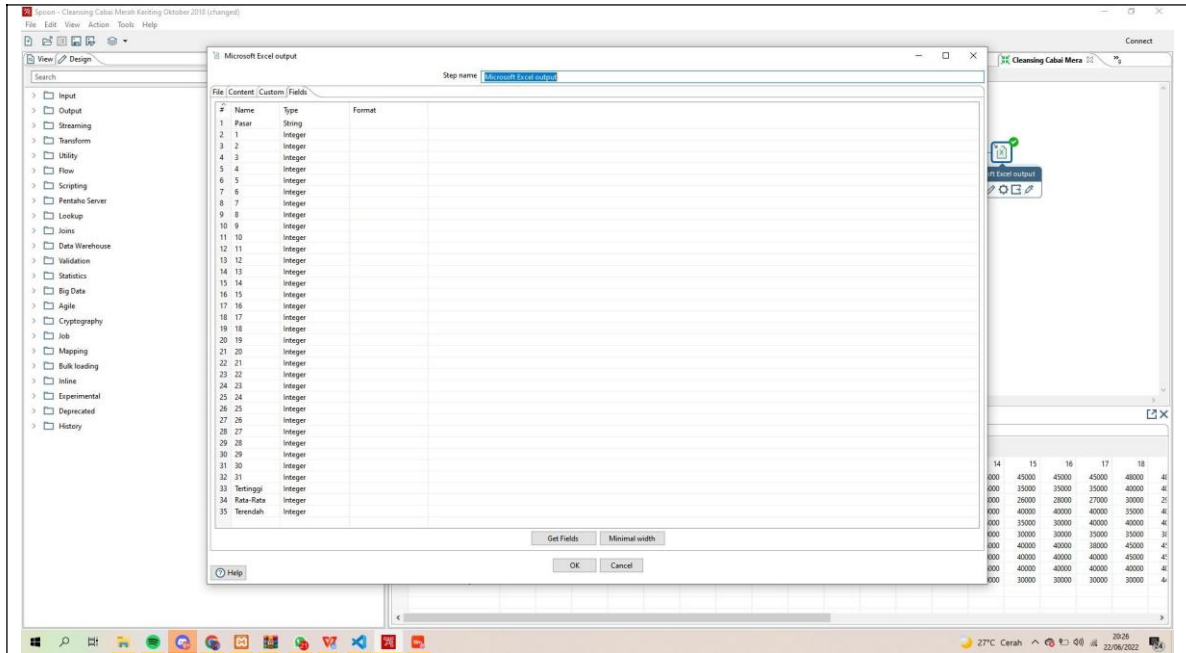


## 5. Excel Output

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 2018.xls*** (File & Sheet)







## 2. Alat dan Bahan

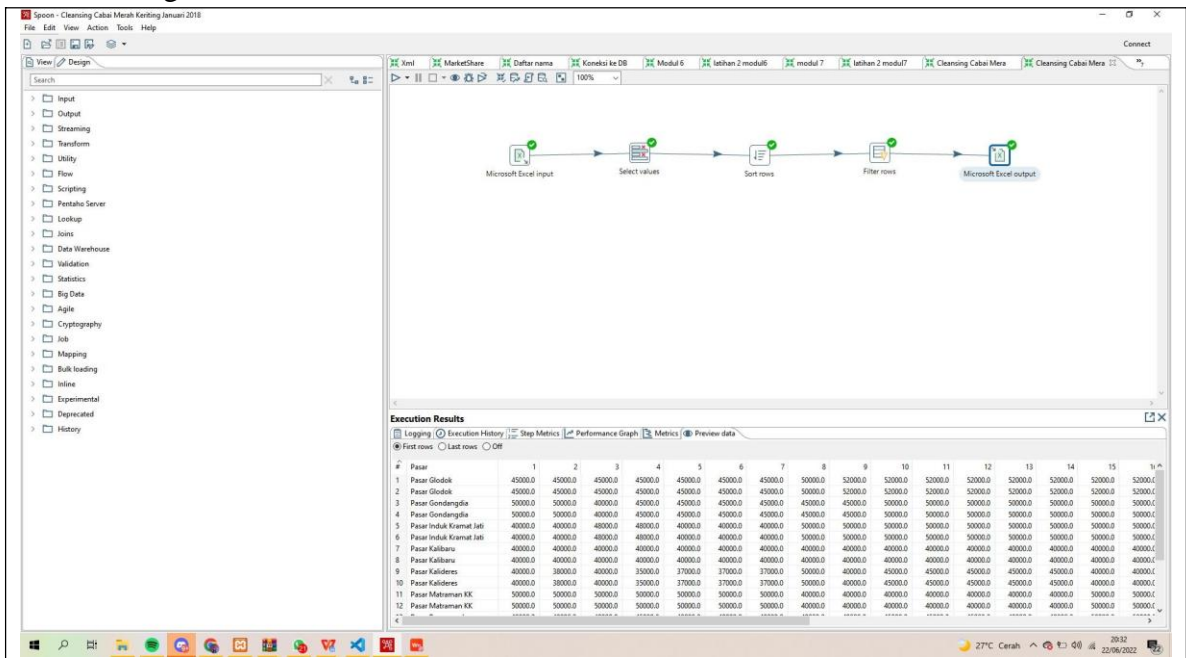
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

## 3. Elemen Kompetensi

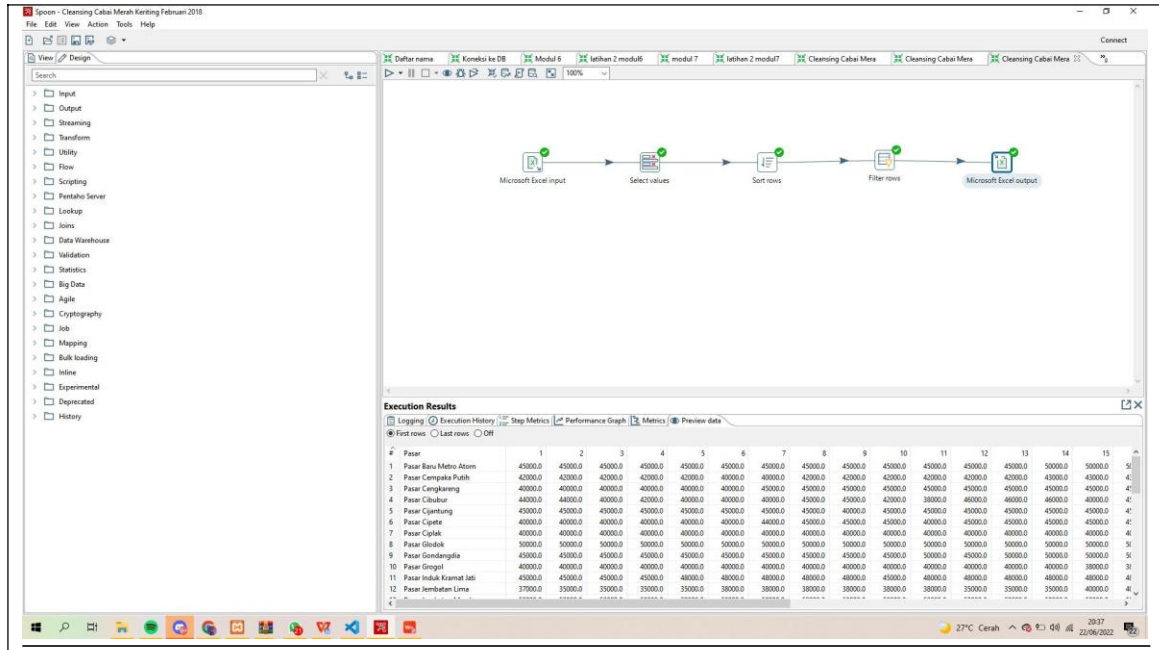
a. Tugas Akhir Praktikum I – Screenshot Struktur yang sudah berhasil berjalan

1. Cleansing Januari 2018

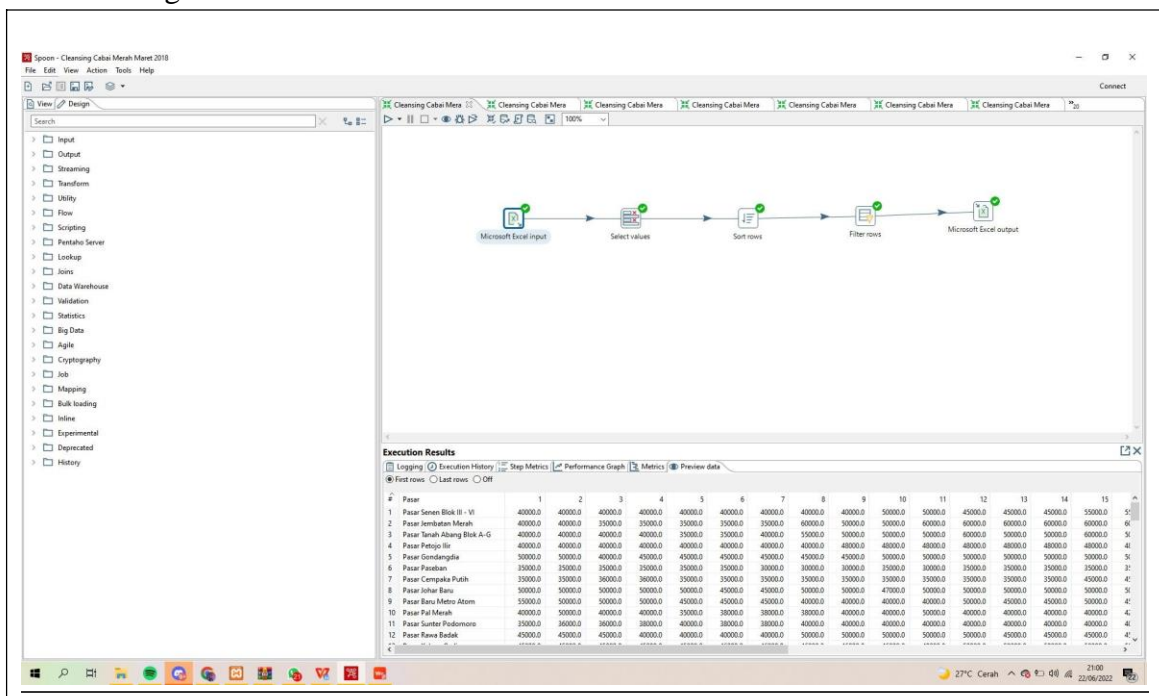


2. Cleansing Februari 2018



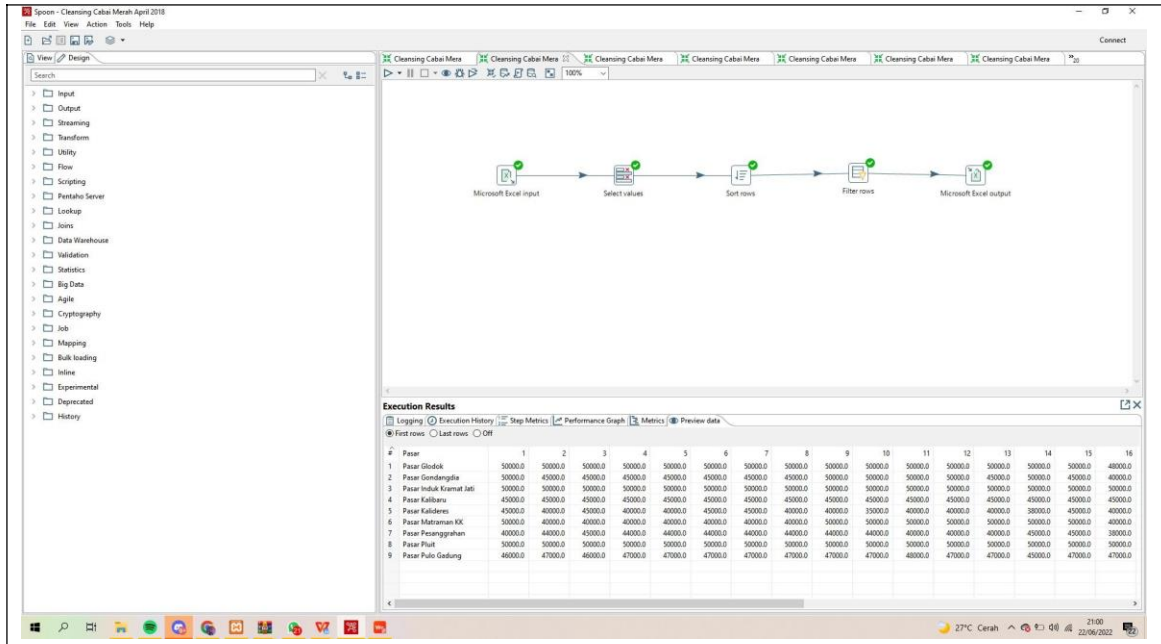


### 3. Cleansing Maret 2018

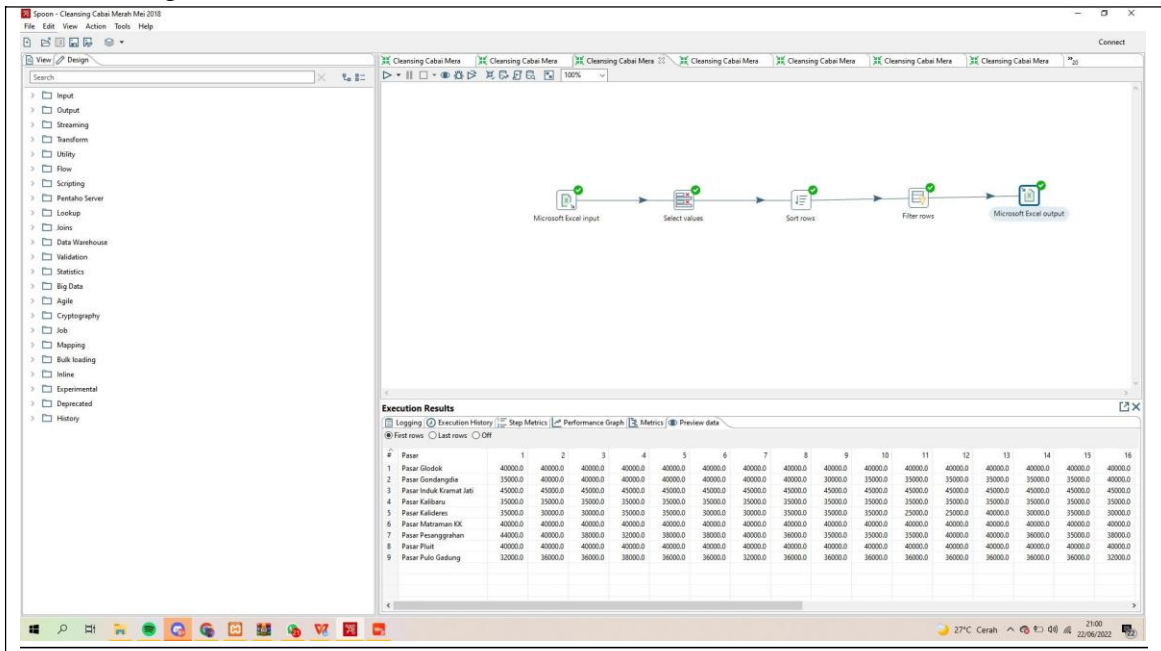


### 4. Cleansing April 2018



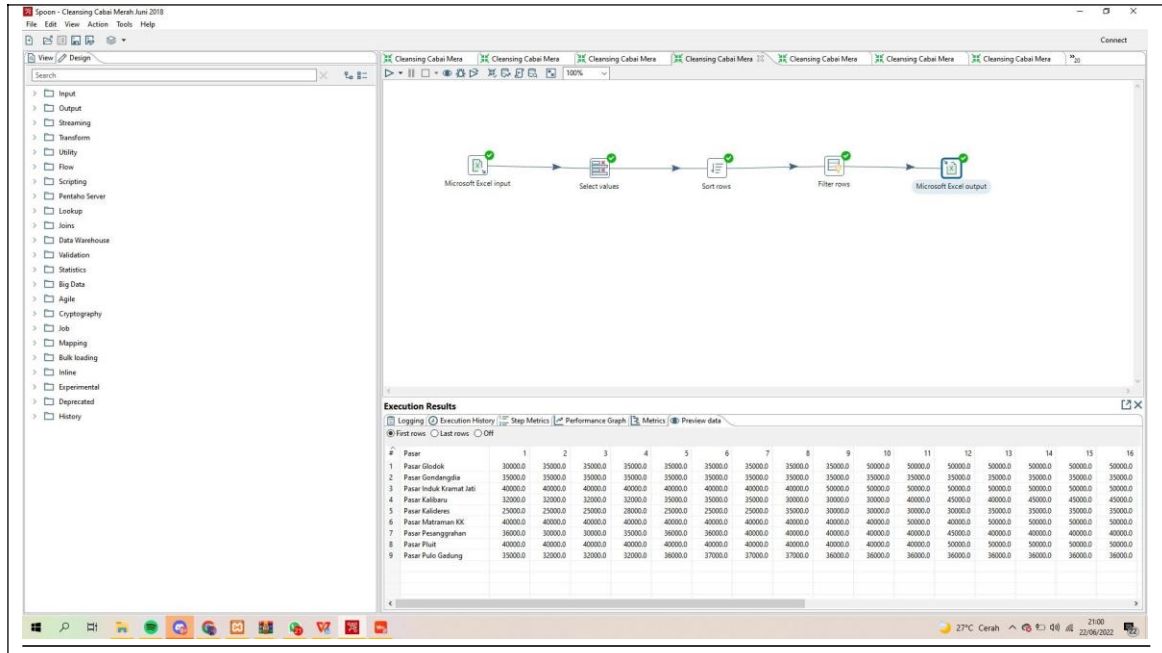


## 5. Cleansing Mei 2018

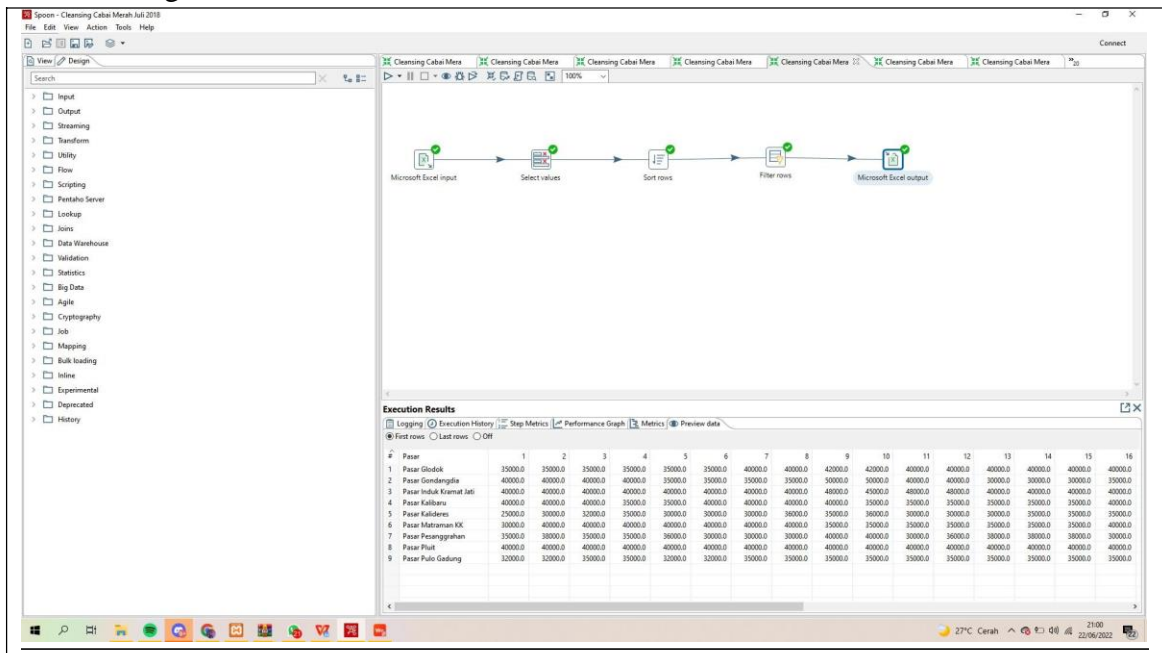


## 6. Cleansing Juni 2018



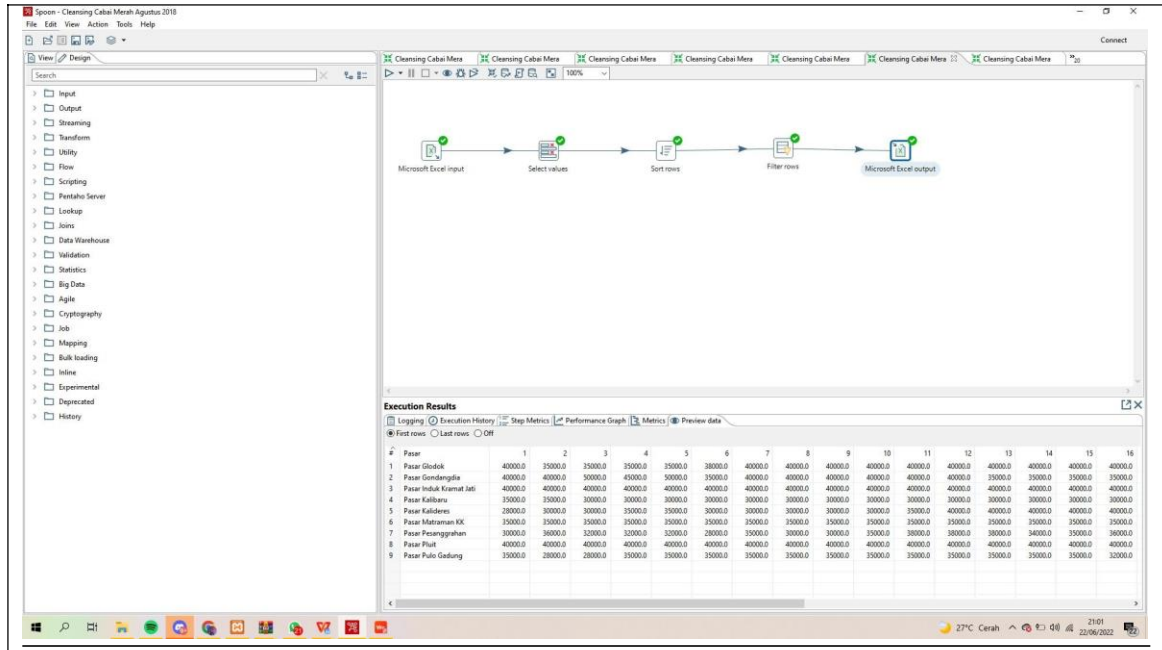


## 7. Cleansing Juli 2018

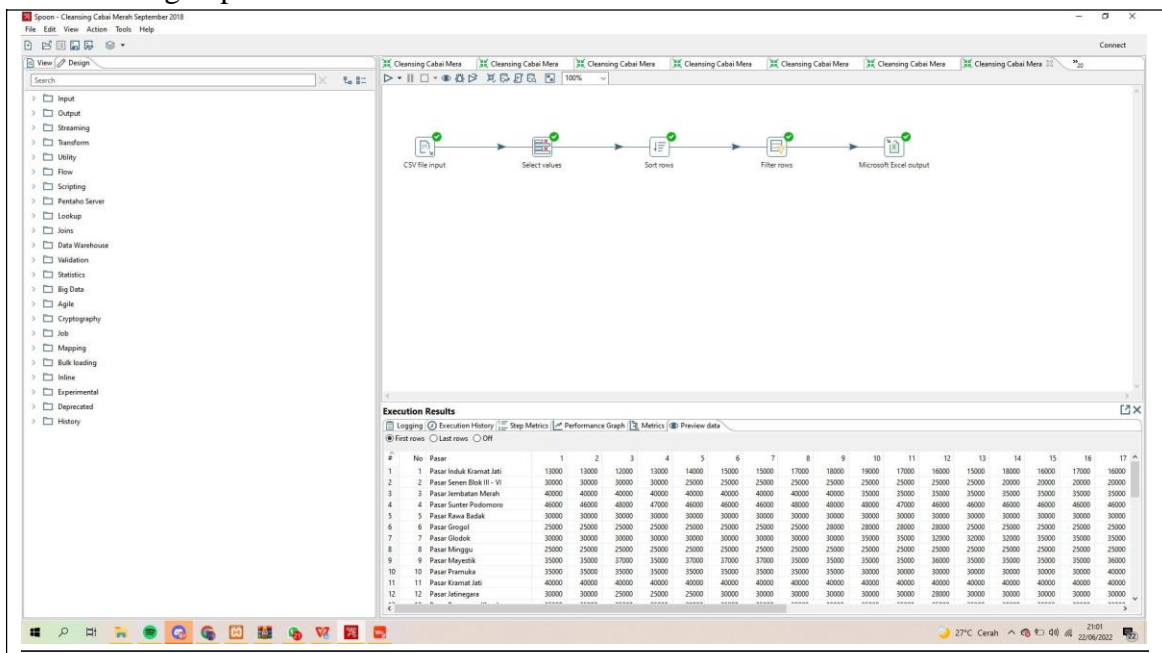


## 8. Cleansing Agustus 2018





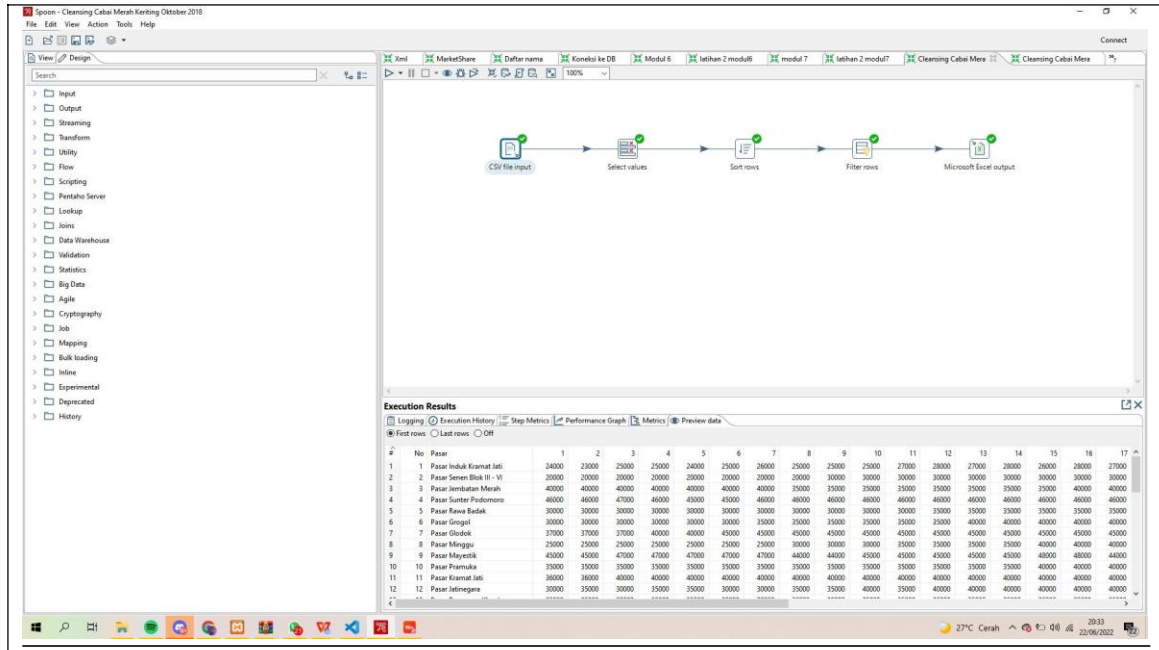
## 9. Cleansing September 2018



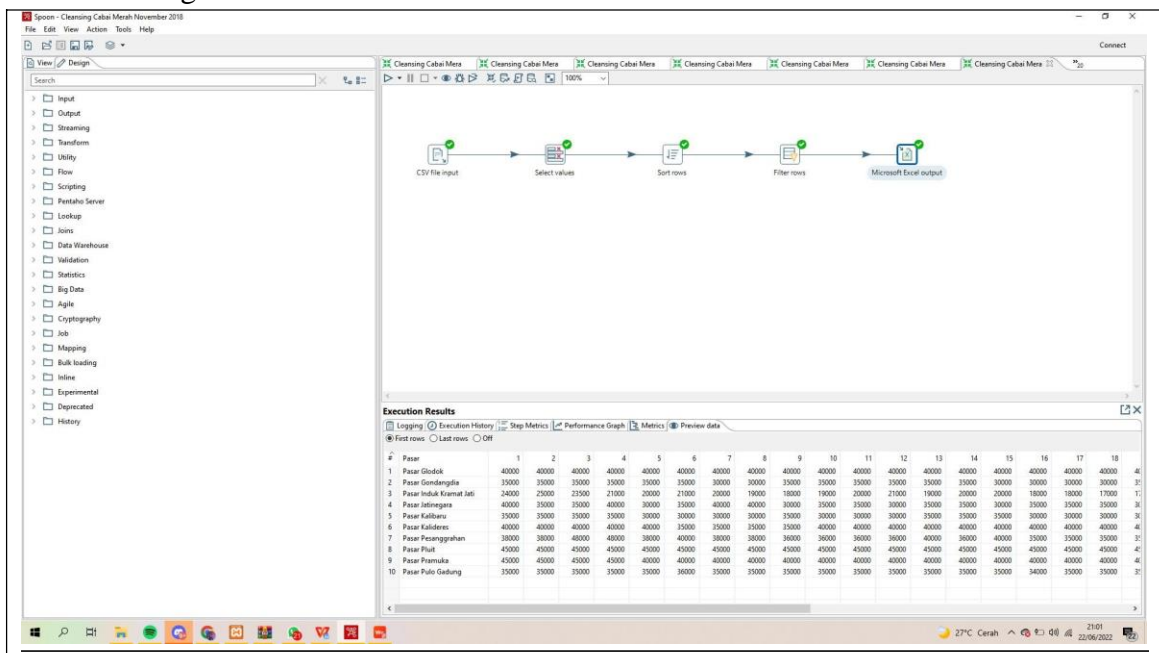
## 10. Cleansing Oktober 2018





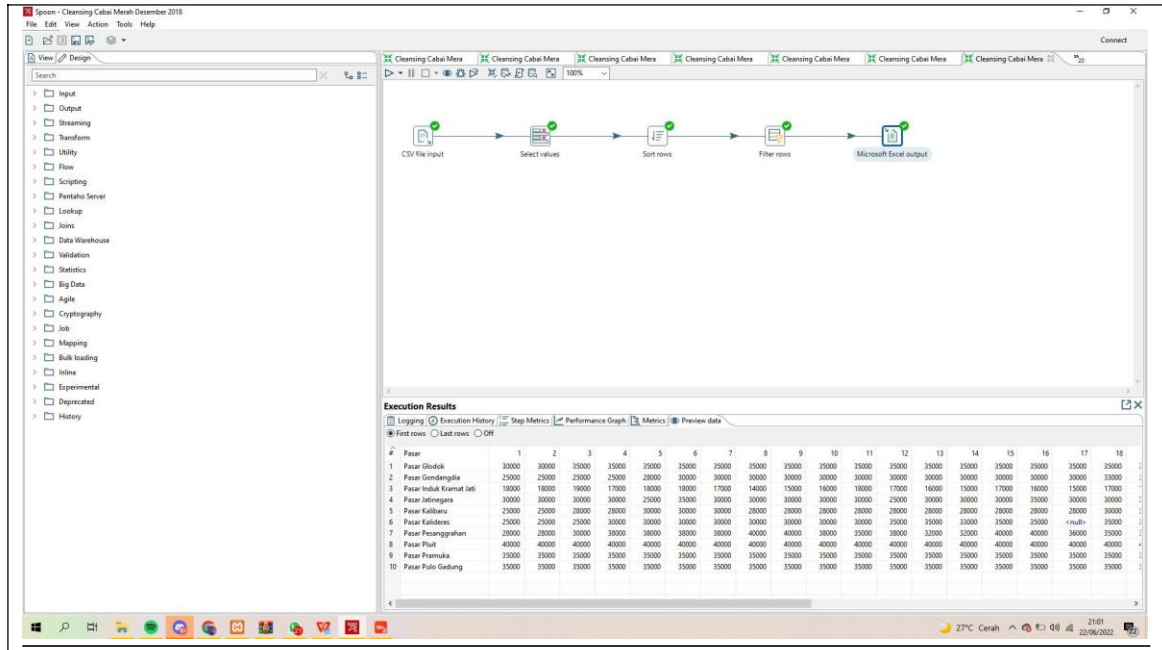


## 11. Cleansing November 2018



## 12. Cleansing Desember 2018





#### 4. File Praktikum

Github Repository:

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

#### 5. Soal Latihan

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 I	...	

**8. Formulir Umpan Balik**

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Tugas Akhir Praktikum I	... Menit	...

Keterangan:

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

