

<p>Nama: Rico Renaldy</p> <p>NIM: 065001800021</p>		<h1>MODUL 9</h1> <p>Nama Dosen: Ir. Teddy Siswanto, MMSi</p>
<p>Hari/Tanggal: Kamis, 16 Mei 2024</p>	<p><b>Praktikum Data Warehouse</b></p>	<p>Nama Asisten Laboratorium:</p> <ul style="list-style-type: none"><li>1. Nia Suhernawati – 065002100005</li><li>2. Exchell S.G Ointu - 065002100007</li></ul>

## 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

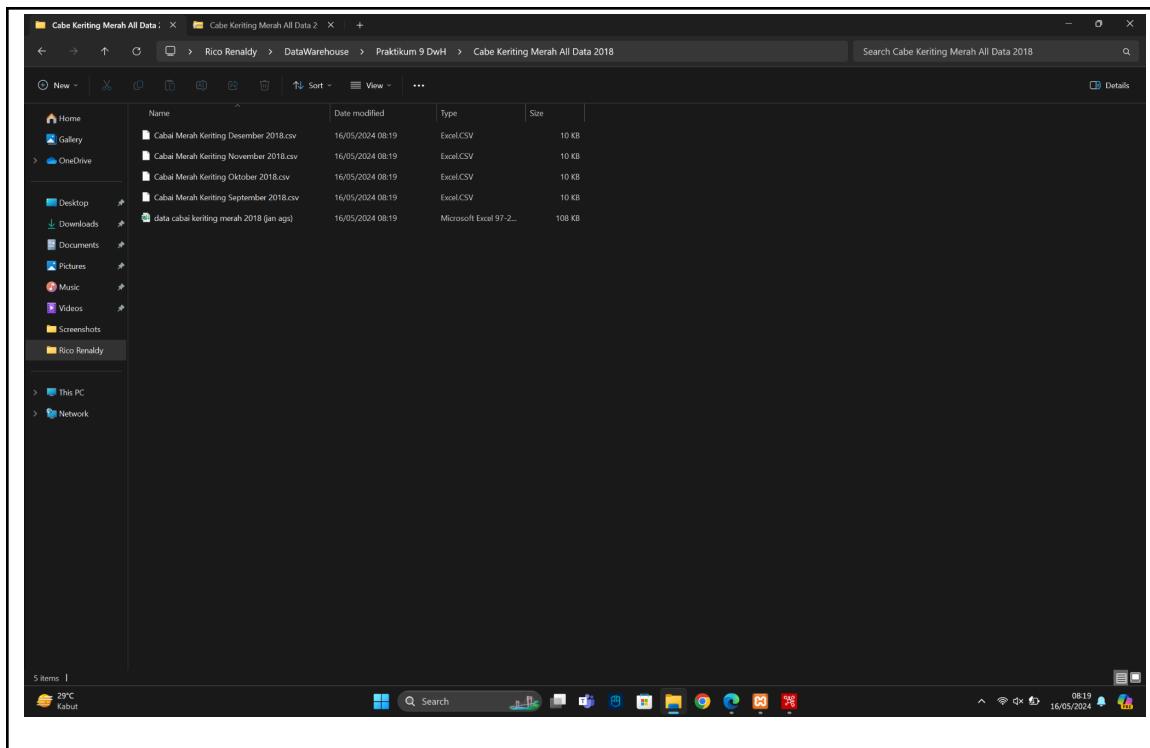


agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

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

## 1. Data Preparation

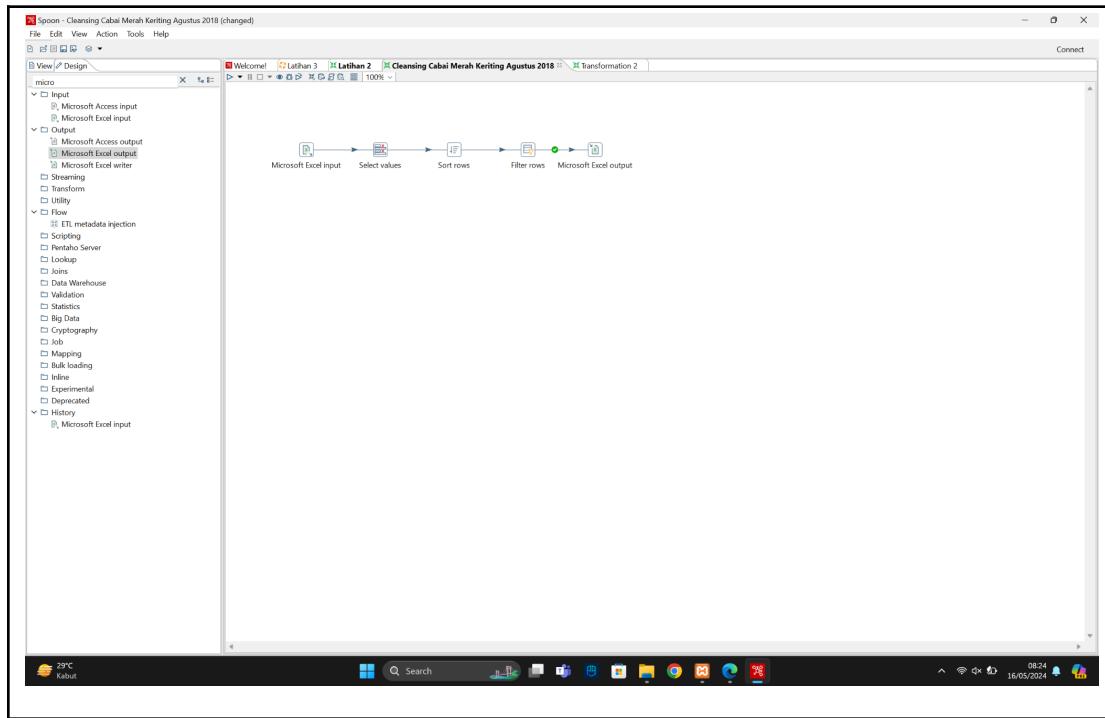
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



## 2. Excel Input

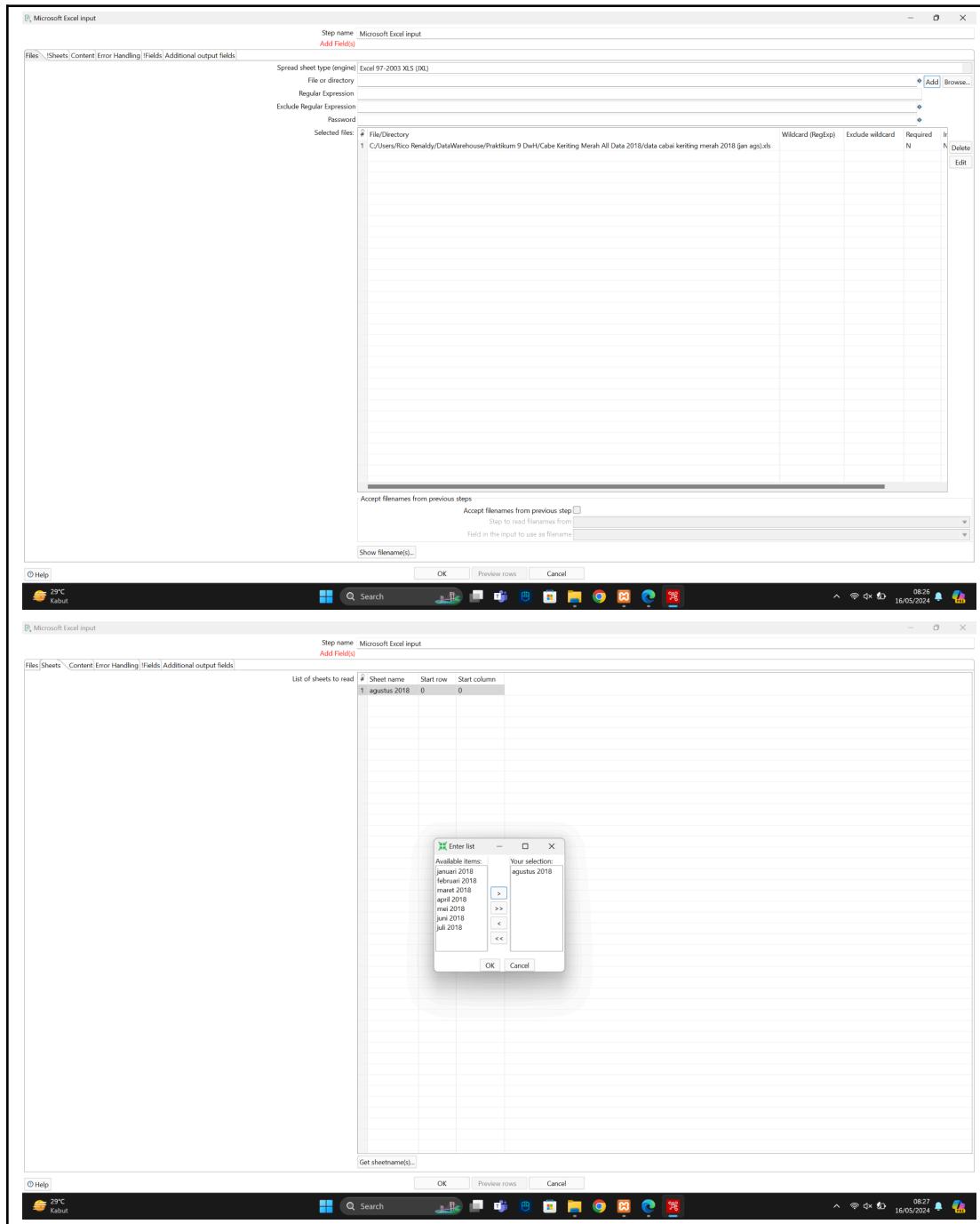
- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2018.ktr***

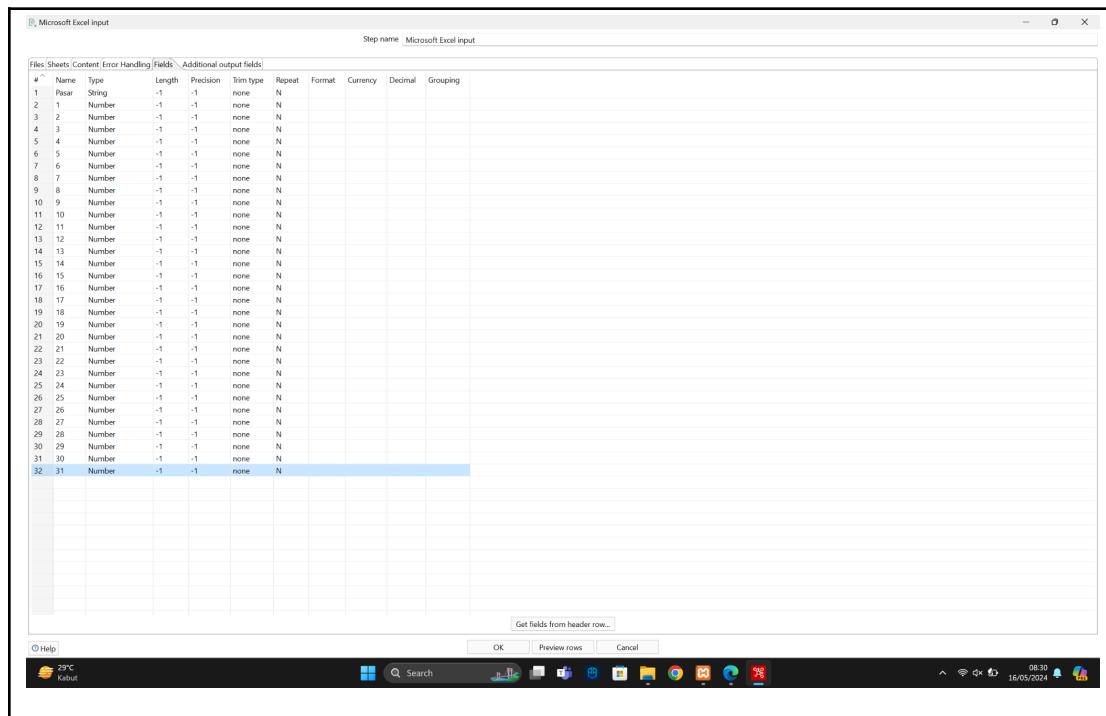




- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)



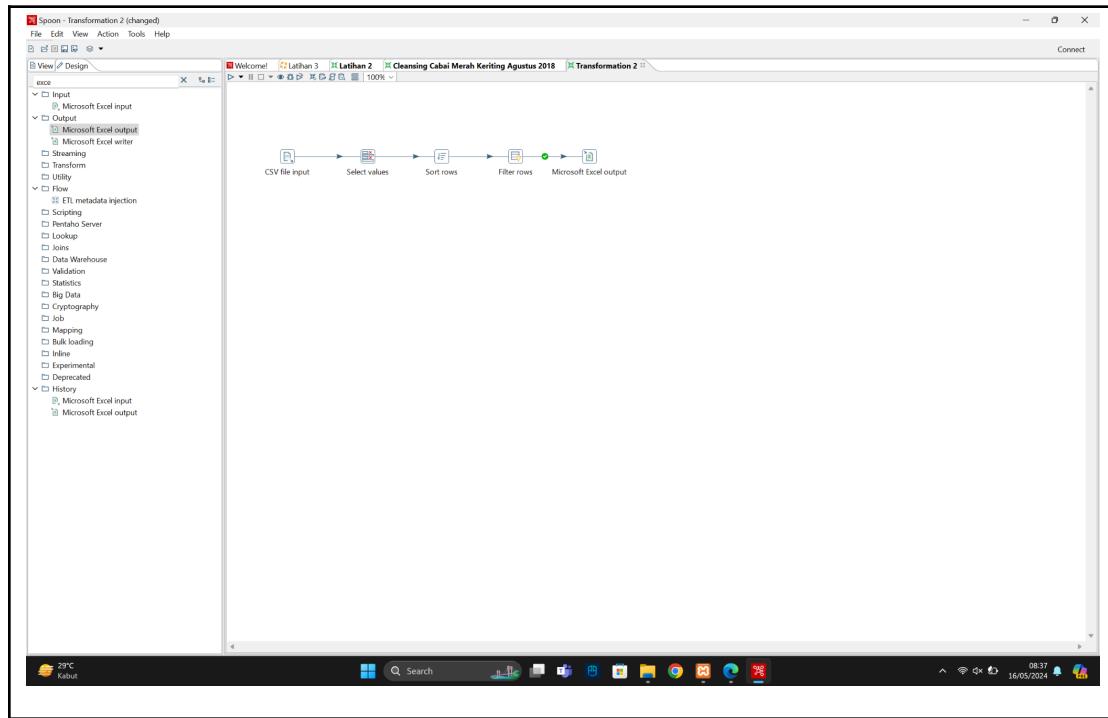




## CSV Input

- 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.



**Screenshot 1 (Top):**

Step name: CSV file input  
 Filename: C:/Users/Rico Renaldy/Desktop/Praktikum 9 Dwh/Cabai Merah Keiting Merah All Data 2018/Cabai Merah Keiting Oktober 2018.csv  
 Delimiter: ,  
 Enclosure: "  
 NIO buffer size: 50000  
 Lazy conversion?   
 Header row present?   
 Add filename to result?   
 The row number field name (optional)  
 Running in parallel?   
 New line possible in fields?   
 Format: mixed  
 File encoding:  ASCII (US-ASCII)  ISO-8859-1 (Latin 1)  UTF-8  UTF-16 Little Endian  UTF-16 Big Endian  UTF-32 Little Endian  UTF-32 Big Endian  
 # Name Type Format Length Precision Currency Decimal Group Trim type  
 1 Pasar String # 26 . IDR . . none  
 2 1 Integer # 15 0 IDR . . none  
 3 2 Integer # 15 0 IDR . . none  
 4 3 Integer # 15 0 IDR . . none  
 5 4 Integer # 15 0 IDR . . none  
 6 5 Integer # 15 0 IDR . . none  
 7 6 Integer # 15 0 IDR . . none  
 8 7 Integer # 15 0 IDR . . none  
 9 8 Integer # 15 0 IDR . . none  
 10 9 Integer # 15 0 IDR . . none  
 11 10 Integer # 15 0 IDR . . none  
 12 11 Integer # 15 0 IDR . . none  
 13 12 Integer # 15 0 IDR . . none  
 14 13 Integer # 15 0 IDR . . none  
 15 14 Integer # 15 0 IDR . . none  
 16 15 Integer # 15 0 IDR . . none  
 17 16 Integer # 15 0 IDR . . none  
 18 17 Integer # 15 0 IDR . . none  
 19 18 Integer # 15 0 IDR . . none  
 20 19 Integer # 15 0 IDR . . none  
 21 20 Integer # 15 0 IDR . . none  
 22 21 Integer # 15 0 IDR . . none  
 23 22 Integer # 15 0 IDR . . none  
 24 23 Integer # 15 0 IDR . . none  
 25 24 Integer # 15 0 IDR . . none  
 26 25 Integer # 15 0 IDR . . none  
 27 26 Integer # 15 0 IDR . . none  
 28 27 Integer # 15 0 IDR . . none  
 29 28 Integer # 15 0 IDR . . none  
 30 29 Integer # 15 0 IDR . . none  
 31 30 Integer # 15 0 IDR . . none  
 32 31 Integer # 15 0 IDR . . none  
 33 Tertinggi Integer # 15 0 IDR . . none  
 34 Rata-Rata Integer # 15 0 IDR . . none  
 35 Terendah Integer # 15 0 IDR . . none

**Screenshot 2 (Bottom):**

Step name: CSV file input  
 Filename: C:/Users/Rico Renaldy/Desktop/Praktikum 9 Dwh/Cabai Merah Keiting Merah All Data 2018/Cabai Merah Keiting Oktober 2018.csv  
 Delimiter: ,  
 Enclosure: "  
 NIO buffer size: 50000  
 Lazy conversion?   
 Header row present?   
 Add filename to result?   
 The row number field name (optional)  
 Running in parallel?   
 New line possible in fields?   
 Format: mixed  
 File encoding:  ASCII (US-ASCII)  ISO-8859-1 (Latin 1)  UTF-8  UTF-16 Little Endian  UTF-16 Big Endian  UTF-32 Little Endian  UTF-32 Big Endian  
 # Name Type Format Length Precision Currency Decimal Group Trim type  
 3 2 Integer # 15 0 IDR . . none  
 4 3 Integer # 15 0 IDR . . none  
 5 4 Integer # 15 0 IDR . . none  
 6 5 Integer # 15 0 IDR . . none  
 7 6 Integer # 15 0 IDR . . none  
 8 7 Integer # 15 0 IDR . . none  
 9 8 Integer # 15 0 IDR . . none  
 10 9 Integer # 15 0 IDR . . none  
 11 10 Integer # 15 0 IDR . . none  
 12 11 Integer # 15 0 IDR . . none  
 13 12 Integer # 15 0 IDR . . none  
 14 13 Integer # 15 0 IDR . . none  
 15 14 Integer # 15 0 IDR . . none  
 16 15 Integer # 15 0 IDR . . none  
 17 16 Integer # 15 0 IDR . . none  
 18 17 Integer # 15 0 IDR . . none  
 19 18 Integer # 15 0 IDR . . none  
 20 19 Integer # 15 0 IDR . . none  
 21 20 Integer # 15 0 IDR . . none  
 22 21 Integer # 15 0 IDR . . none  
 23 22 Integer # 15 0 IDR . . none  
 24 23 Integer # 15 0 IDR . . none  
 25 24 Integer # 15 0 IDR . . none  
 26 25 Integer # 15 0 IDR . . none  
 27 26 Integer # 15 0 IDR . . none  
 28 27 Integer # 15 0 IDR . . none  
 29 28 Integer # 15 0 IDR . . none  
 30 29 Integer # 15 0 IDR . . none  
 31 30 Integer # 15 0 IDR . . none  
 32 31 Integer # 15 0 IDR . . none  
 33 Tertinggi Integer # 15 0 IDR . . none  
 34 Rata-Rata Integer # 15 0 IDR . . none  
 35 Terendah Integer # 15 0 IDR . . none

### 3. Data Filtering



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

**Excel (Pasar – 31)**

Step name: Select values

Fields :

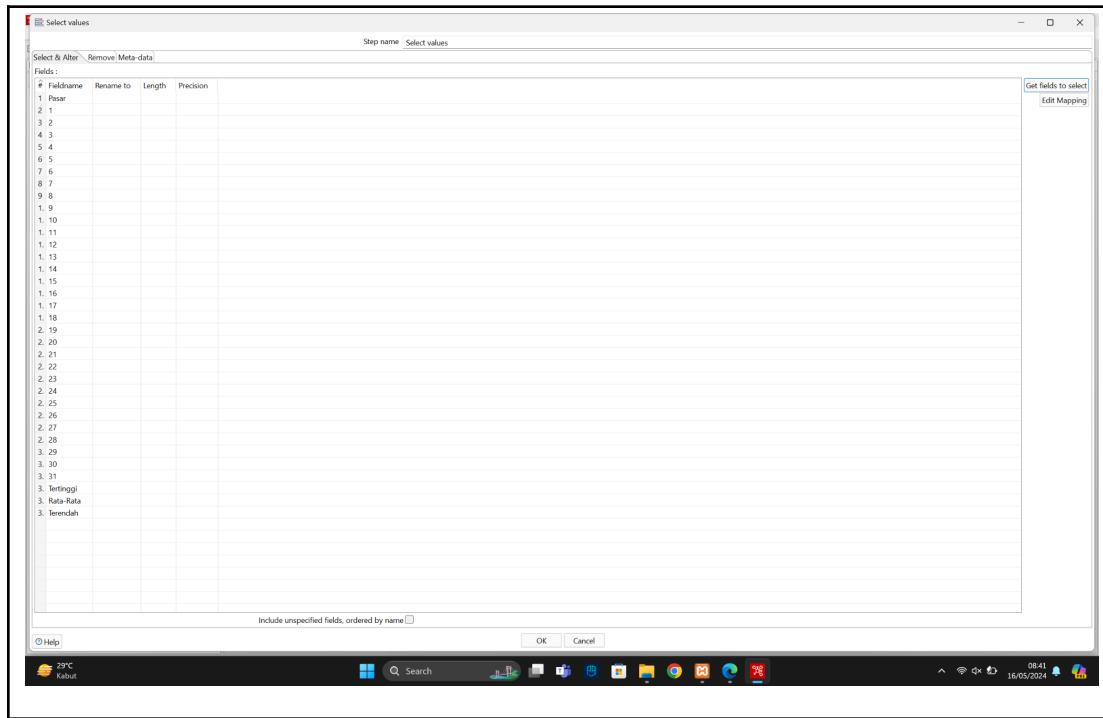
#	Fieldname	Rename to	Length	Precision
1	Pasar			
2	1			
3	2			
4	3			
5	4			
6	5			
7	6			
8	7			
9	8			
10	9			
11	10			
12	11			
13	12			
14	13			
15	14			
16	15			
17	16			
18	17			
19	18			
20	19			
21	20			
22	21			
23	22			
24	23			
25	24			
26	25			
27	26			
28	27			
29	28			
30	29			
31	30			

Include unspecified fields, ordered by name

OK Cancel

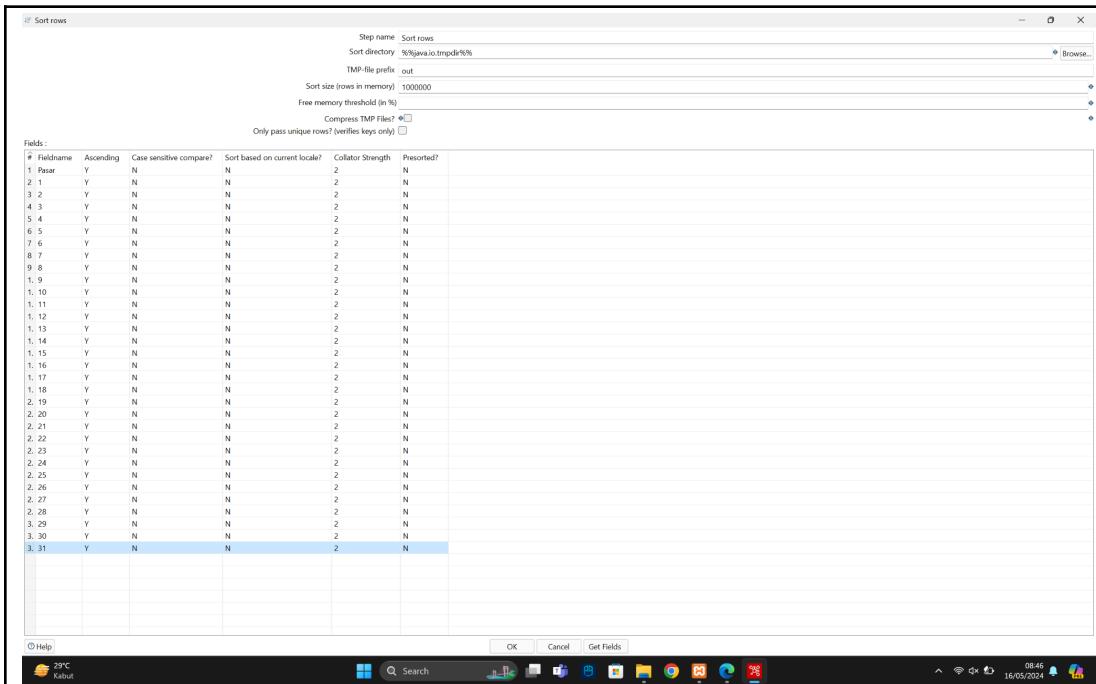
**CSV (Pasar – Terendah)**



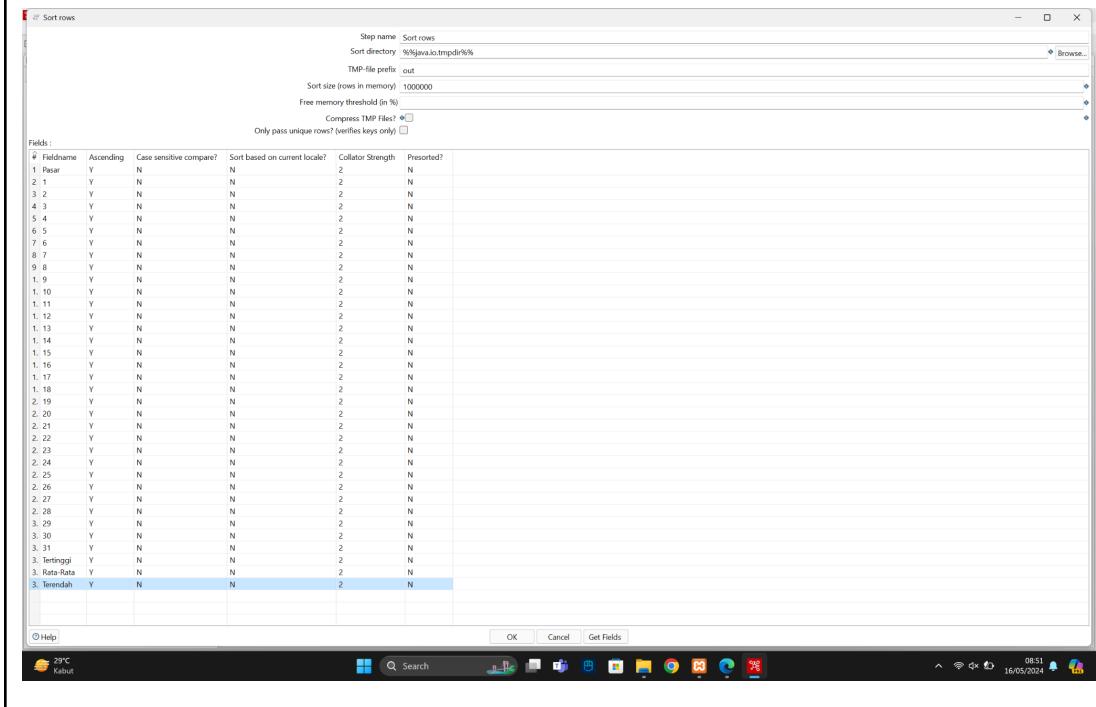


- Kemudian lakukan ascending data pada proses step Sort Rows  
**Excel (Pasar – 31)**



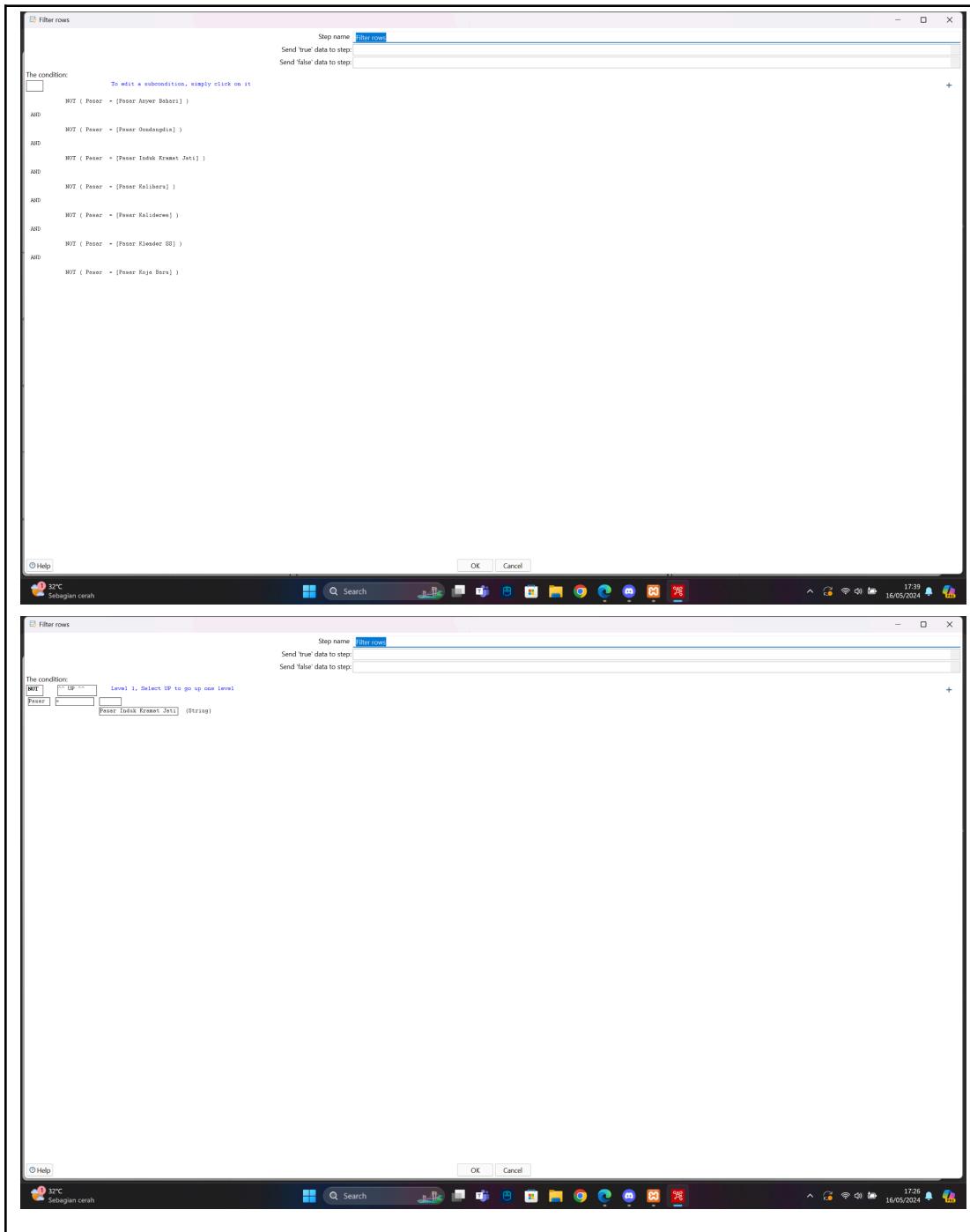


## CSV (Pasar – Terendah)



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

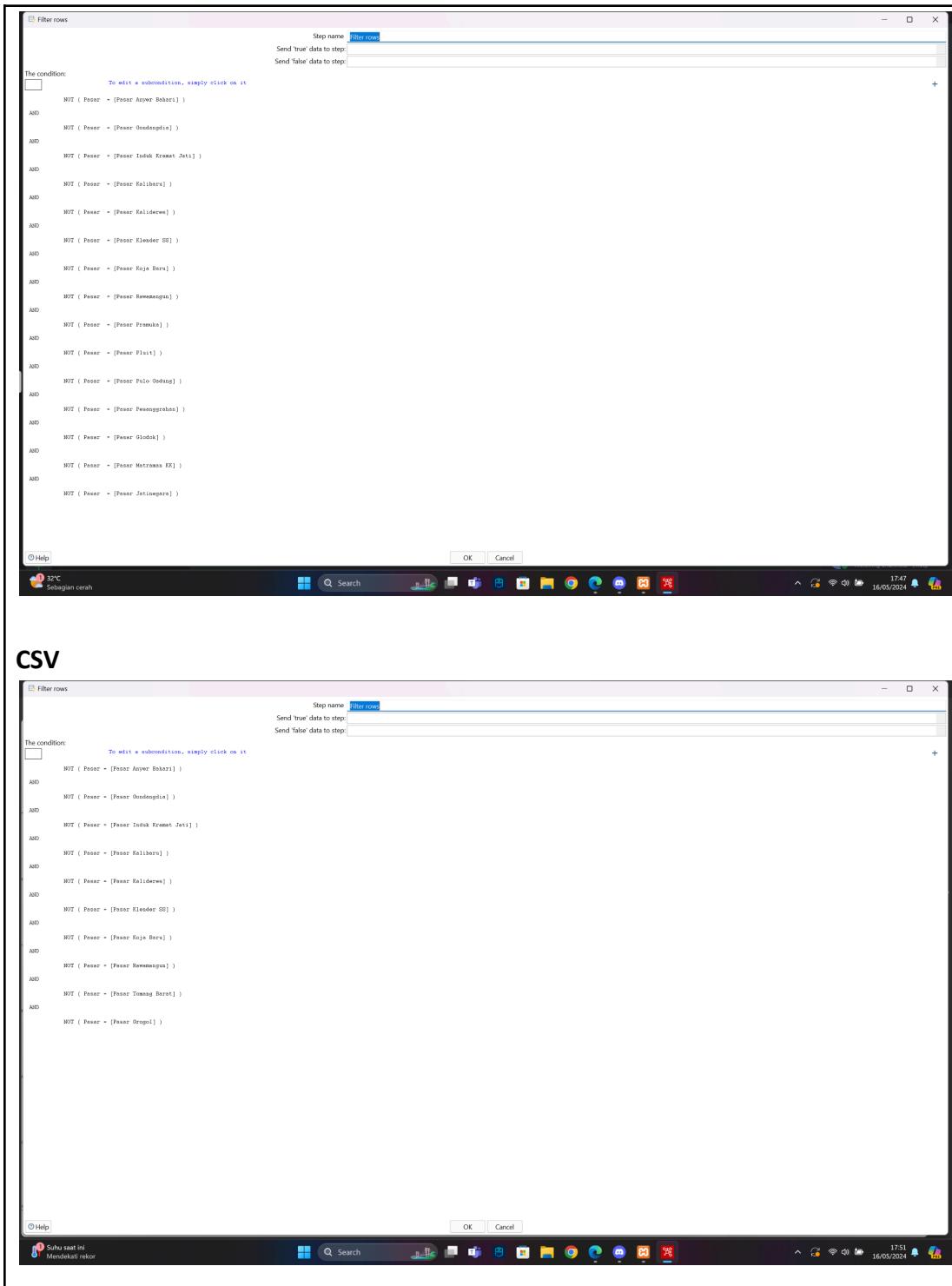




- Berikut List Data Nama Pasar yang mau dibuang (2018)

Excel

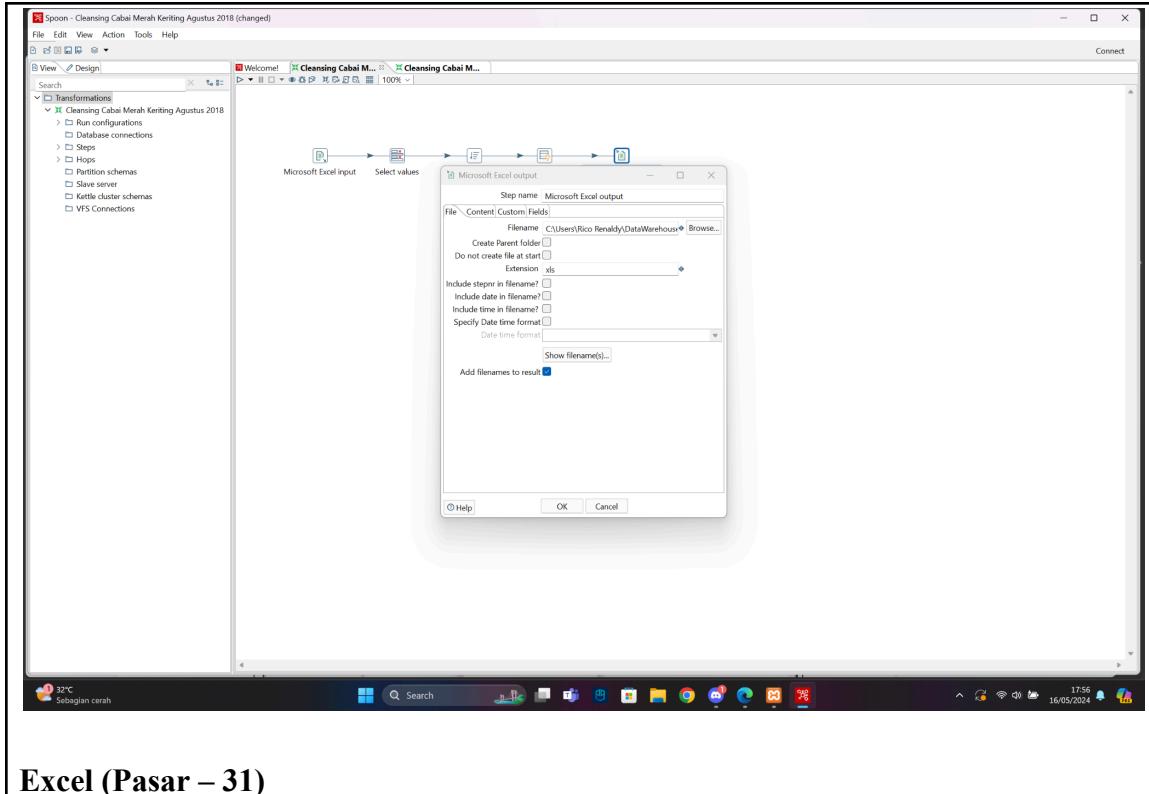




#### 4. Excel Output

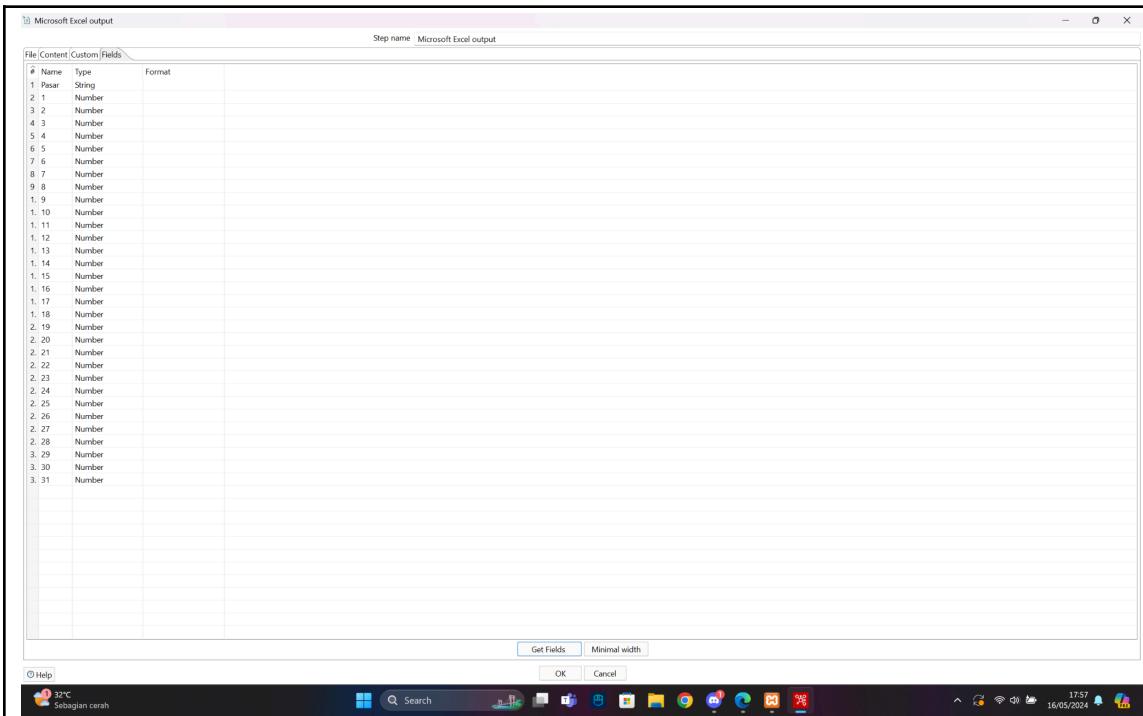


Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama **Cabai Merah Keriting [Bulan] [Tahun].xls** seperti **Cabai Merah Keriting Agustus 2018.xls** (File & Sheet)

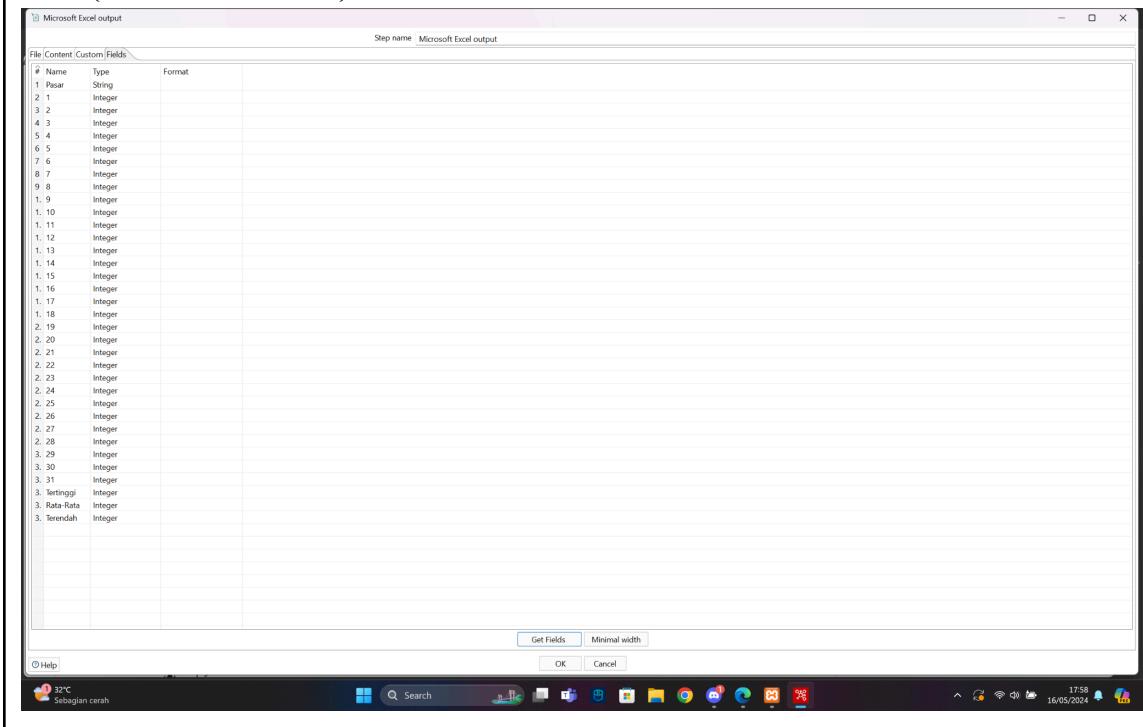


Excel (Pasar – 31)





## CSV (Pasar – Terendah)



## 2. Alat dan Bahan

Hardware : Laptop/PC



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

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



Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Pasar Baru Metro Atom	35000.0	40000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	
2. Pasar Cempaka Putih	30000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	32000.0	
3. Pasar Cengkareng	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	
4. Pasar Glodok	30000.0	30000.0	35000.0	35000.0	30000.0	35000.0	35000.0	30000.0	35000.0	32000.0	34000.0	34000.0	30000.0	28000.0	28000.0	28000.0	28000.0	
5. Pasar Jembatan Lima	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	
6. Pasar Cipeute	30000.0	25000.0	30000.0	25000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	28000.0	
7. Pasar Cipatak	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	
8. Pasar Groglol	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	28000.0	
9. Pasar Jembatan Lima	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	
1. Pasar Jembatan Merah	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	
1. Pasar Johar Baru	30000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	35000.0	
1. Pasar Kebayoran Lama	30000.0	35000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	
1. Pasar Kelapa Gading	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	30000.0	

## 9. Cleansing September 2018





10. Cleansing Oktober 2018



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

The screenshot displays two windows side-by-side. On the left is the Apache Nifi interface titled 'Spoon - Cleansing Cabai Merah Keriting Oktober 2018'. It shows a data flow with five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. The 'Transformations' sidebar is open, showing the 'Cleansing Cabai Merah Keriting Oktober 2018' run configuration. On the right is a Microsoft Excel spreadsheet titled 'Cabe Merah Keriting Oktober 2018'. The first sheet, 'Sheet1', contains a large table of data with columns labeled from 1 to 25. The data consists of 45 rows of entries, each starting with 'Pasar' followed by a name and a series of numerical values. The Excel ribbon at the top shows various tabs like File, Home, Insert, etc., and the status bar at the bottom indicates the date as 16/05/2024.

## 11. Cleansing November 2018



--

12. Cleansing Desember 2018

--

#### 4. File Praktikum

Github Repository:

--

#### 5. Soal Latihan

Soal:

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

Jawaban:

- 1.
- 2.

#### 6. Kesimpulan

- a. Dalam penggerjaan 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 Umpam Balik

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



Keterangan:

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

