


Nama: Adan Zannuba	 Praktikum Data Warehouse	MODUL 11
NIM: 065001900022		Nama Dosen: Ir. Teddy Siswanto, MMSi
Hari/Tanggal: 15 juni 2022		Nama Asisten Labratorium: 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 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

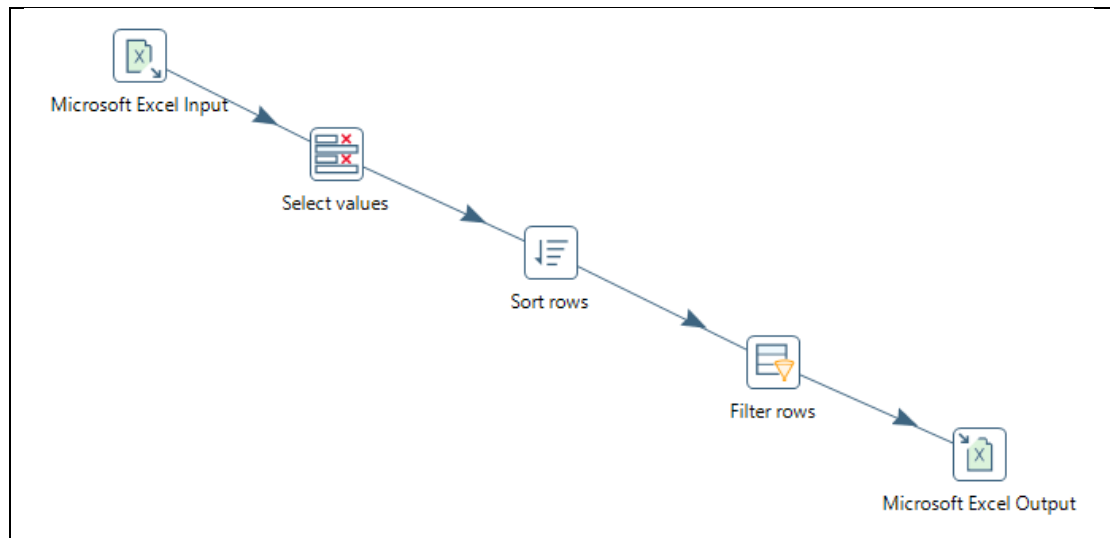
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



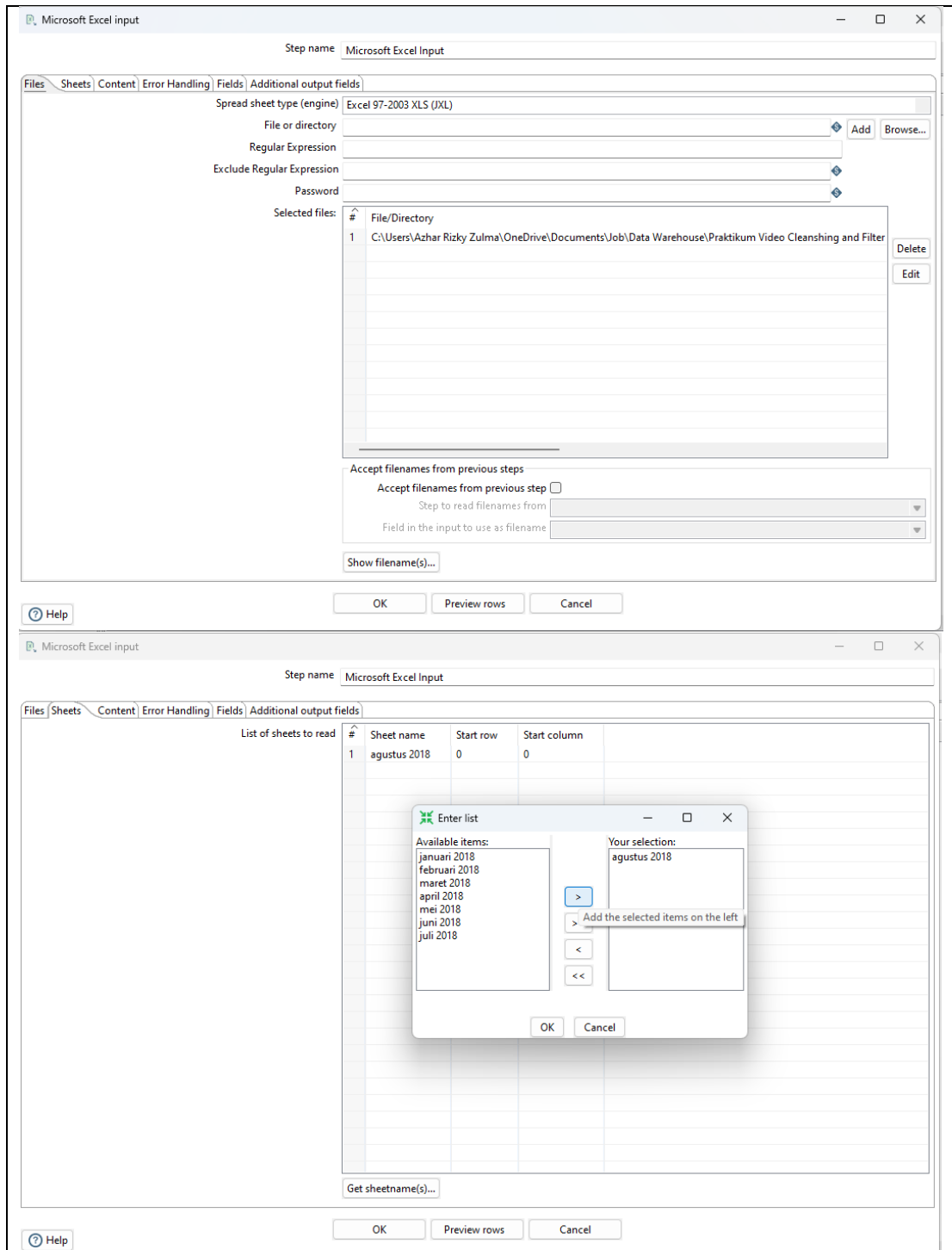
Name	Date modified	Type	Size
Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
Cabai Merah Keriting September 2018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

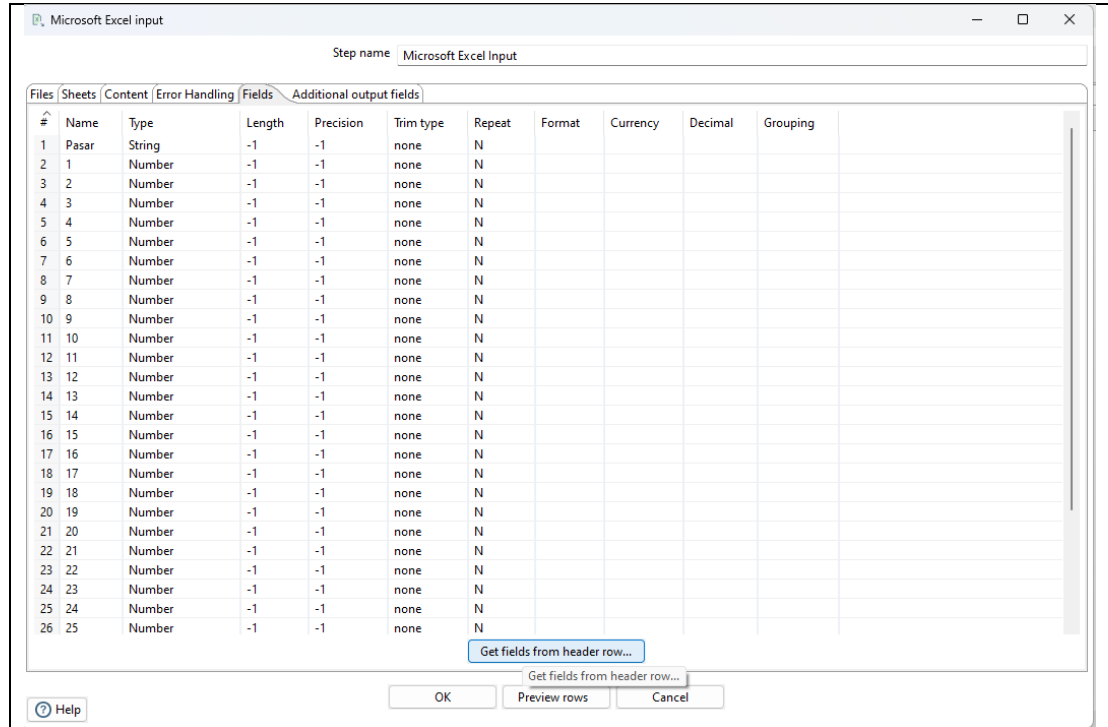
2. Excel Input

- 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***



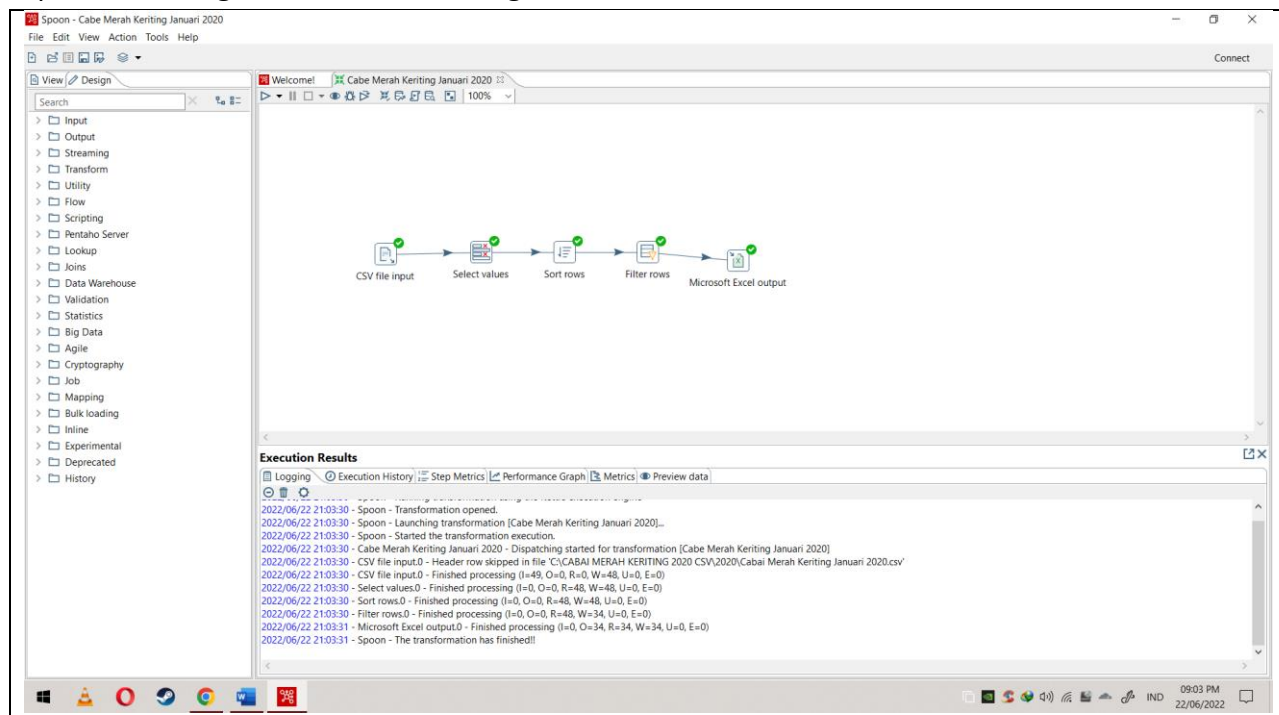
- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)



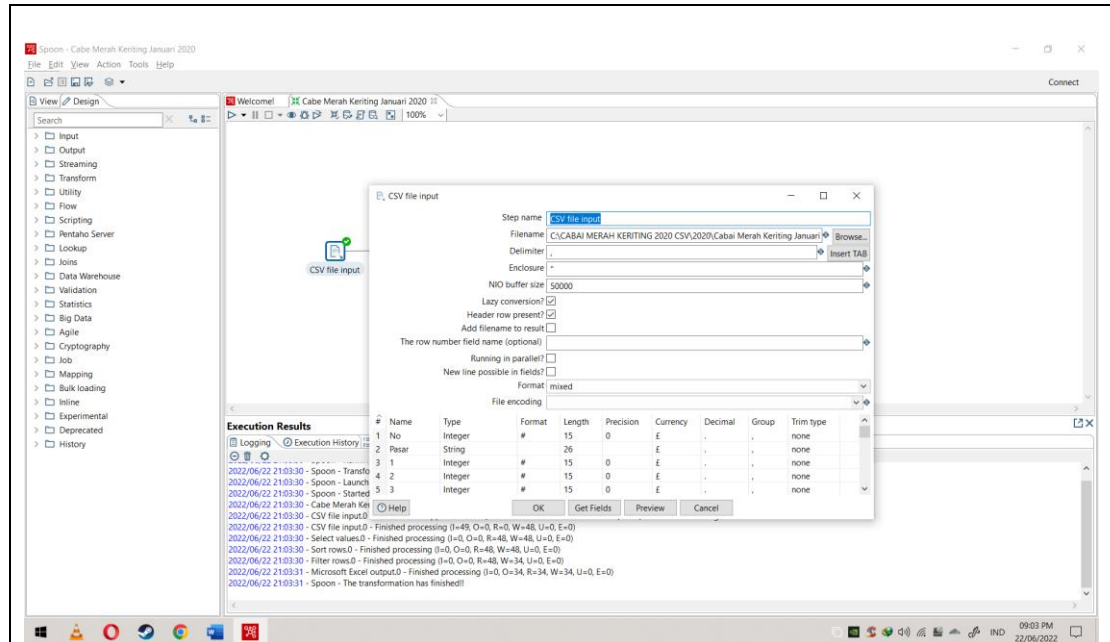


3. CSV Input

- 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**



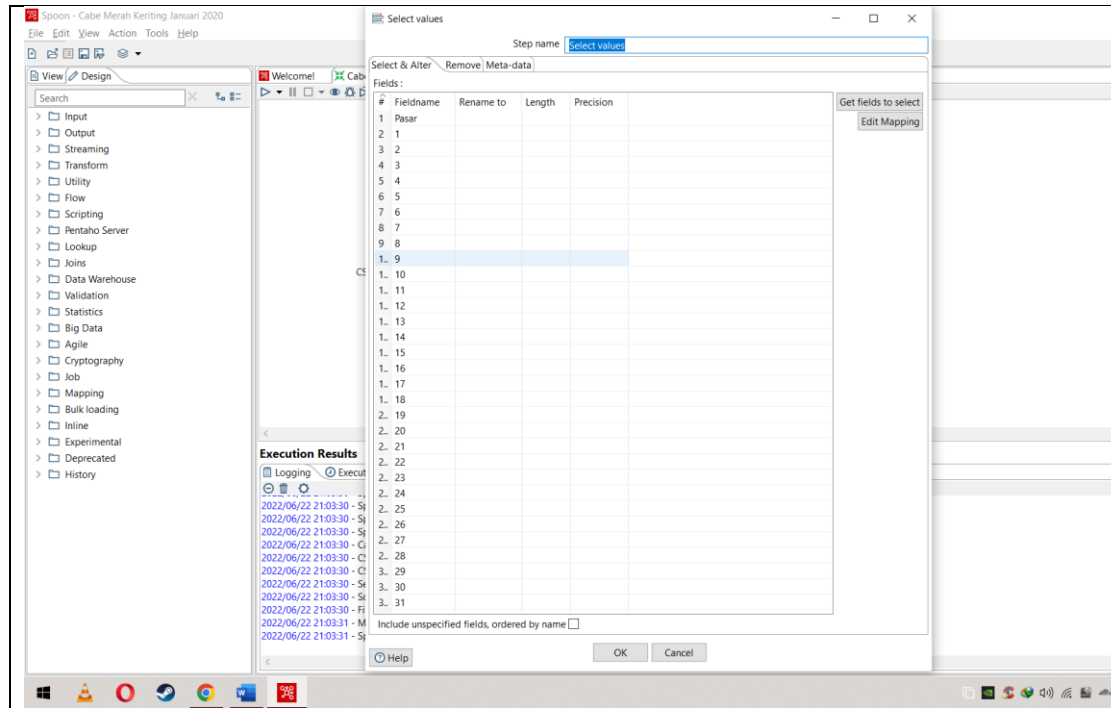
- Berikut ini adalah konfigurasi pada step CSV File Input.



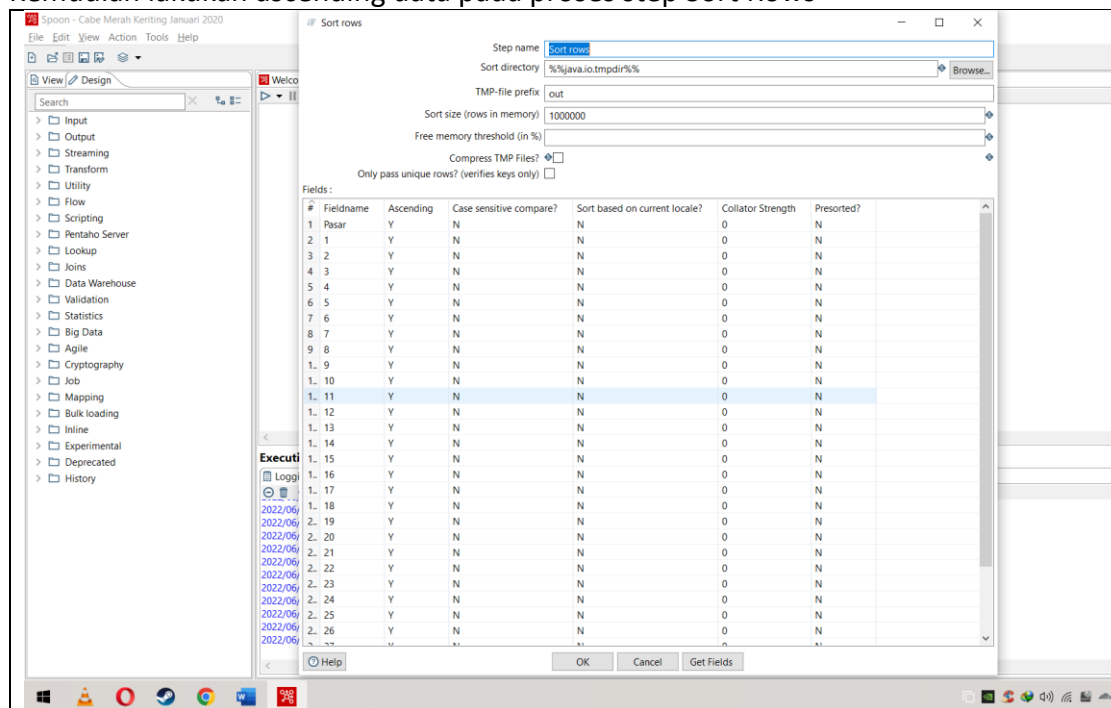
4. Data Filtering

- 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.

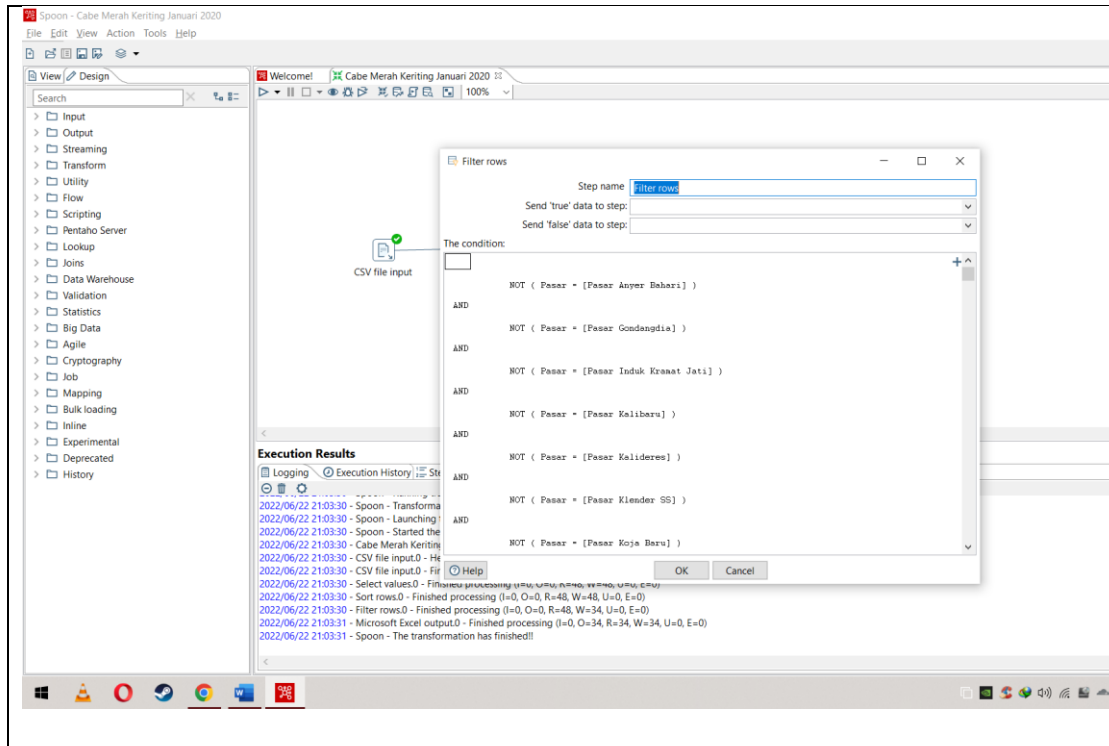




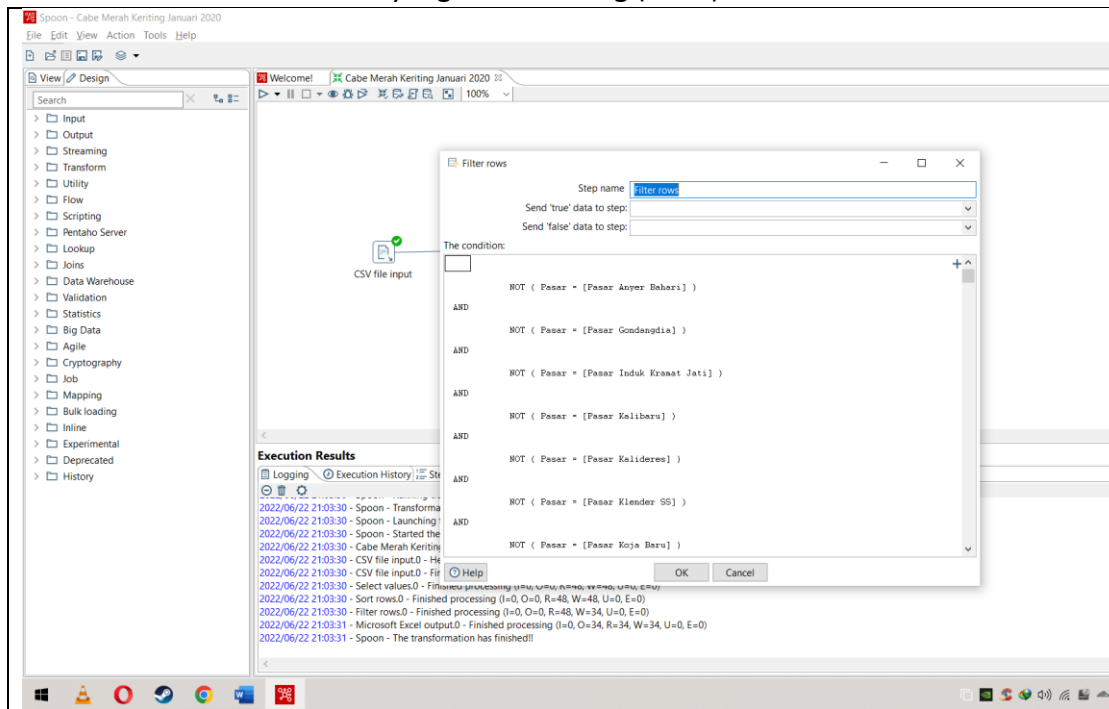
- Kemudian lakukan ascending data pada proses step Sort Rows



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



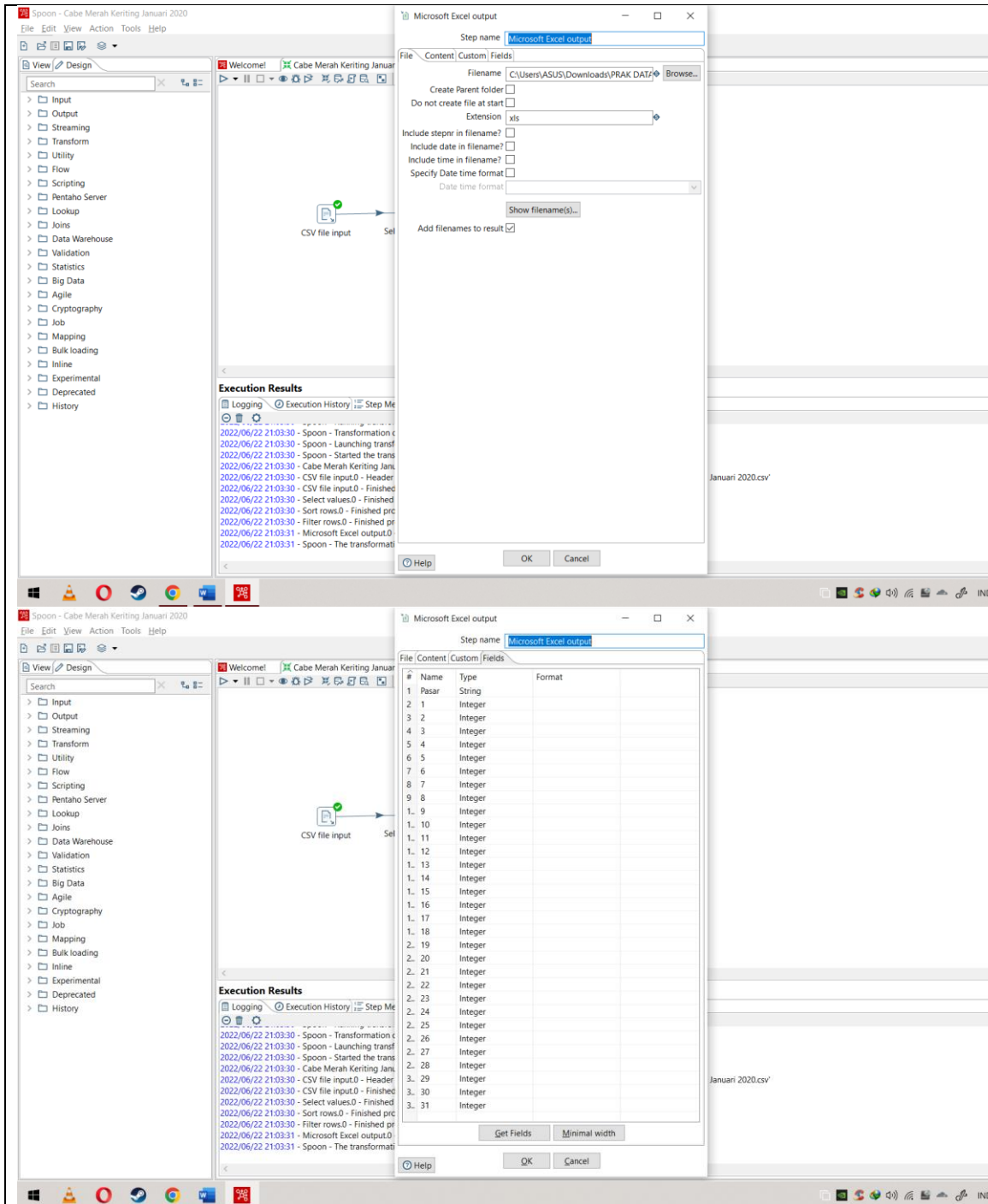
- Berikut List Data Nama Pasar yang mau dibuang (2020)



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

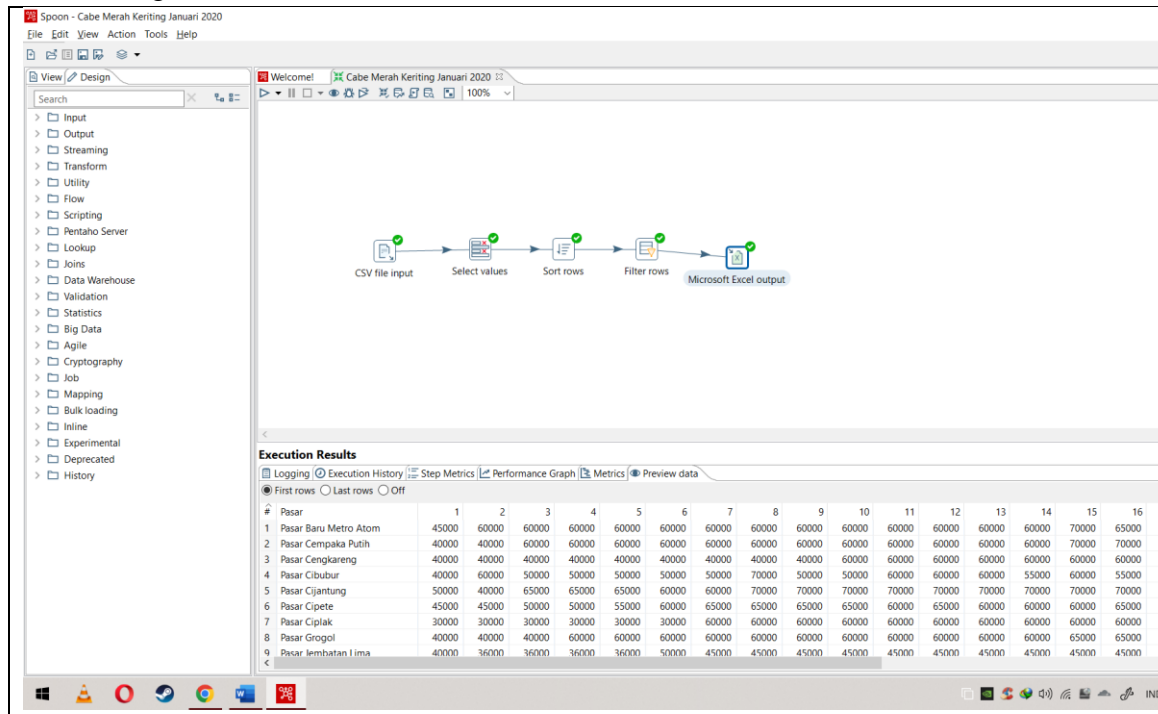
Software : Spoon Pentaho from Hitachi Vantara



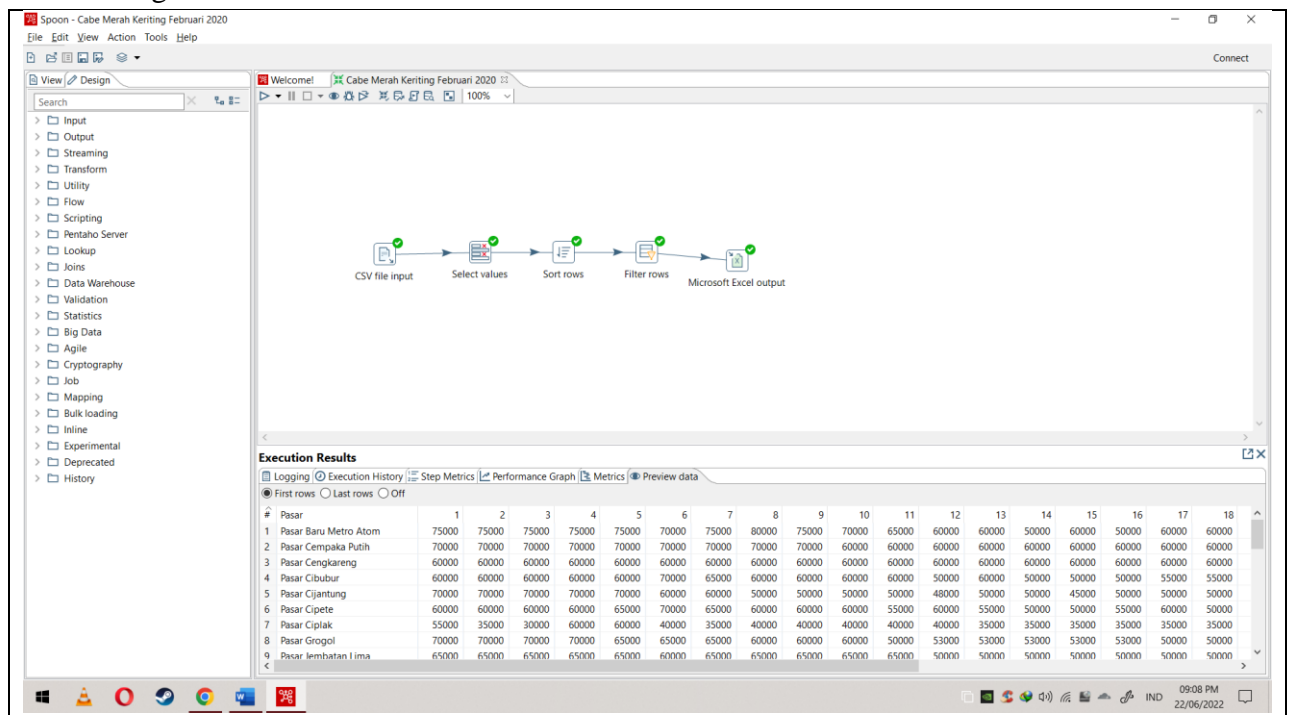
3. Elemen Kompetensi

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

1. Cleansing Januari 2020



2. Cleansing Februari 2020



3. Cleansing Maret 2020

Execution Results

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	50000	50000	50000	50000	60000	60000	60000	60000	60000	50000	50000	50000	50000	45000	65000	60000	50000	60000
2	Pasar Cempaka Putih	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	30000
3	Pasar Cengkareng	60000	50000	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
4	Pasar Cibubur	50000	40000	40000	40000	40000	40000	40000	40000	40000	40000	45000	45000	40000	40000	40000	40000	40000	35000
5	Pasar Cijantung	50000	40000	40000	40000	40000	45000	45000	45000	45000	45000	45000	40000	40000	40000	40000	40000	40000	40000
6	Pasar Cipete	55000	45000	50000	50000	45000	45000	45000	45000	45000	40000	40000	40000	40000	40000	35000	35000	30000	30000
7	Pasar Cipinang	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
8	Pasar Grogol	40000	45000	45000	45000	40000	40000	40000	40000	40000	40000	40000	40000	45000	45000	45000	35000	35000	35000
9	Pasar Jembatan Lima	50000	50000	50000	45000	45000	45000	45000	45000	45000	45000	45000	30000	30000	30000	30000	30000	30000	25000

4. Cleansing April 2020

Execution Results

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	40000	50000	50000	50000
2	Pasar Cempaka Putih	40000	40000	40000	40000	40000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	25000	25000
4	Pasar Cibubur	30000	35000	35000	30000	30000	30000	35000	25000	28000	30000	28000	25000	30000	35000	26000	28000	28000	25000
5	Pasar Cijantung	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	25000	25000	25000
6	Pasar Cipete	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	25000	30000	30000
7	Pasar Cipinang	35000	30000	30000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	20000	20000	20000	20000	20000
8	Pasar Grogol	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
9	Pasar Jembatan Lima	25000	25000	25000	25000	25000	25000	25000	22000	25000	25000	22000	20000	20000	18000	18000	17000	17000	18000

5. Cleansing Mei 2020



The screenshot shows the Spoon- Cabel Merah Kerting Mei 2020 interface. The design canvas displays a data flow: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. The Execution Results panel shows a table with 18 columns and 9 rows of data.

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	30000	35000	35000	35000	30000	40000	40000	40000	40000	40000	35000	40000	35000	35000	35000	35000	35000	35000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cibubur	24000	22000	22000	26000	20000	20000	22000	22000	22000	25000	25000	24000	22000	24000	24000	24000	24000	25000
5	Pasar Cijantung	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
6	Pasar Cipete	25000	25000	25000	20000	20000	25000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
7	Pasar Cipinang	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	23000	22000	22000	22000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
9	Pasar Jembatan Lima	20000	19000	20000	22000	20000	20000	21000	18000	18000	20000	18000	20000	20000	19000	15000	17000	20000	20000

6. Cleansing Juni 2020

The screenshot shows the Spoon- Cabel Merah Kerting Juni 2020 interface. The design canvas displays a data flow: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. The Execution Results panel shows a table with 18 columns and 9 rows of data.

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	30000	35000	35000	35000	30000	40000	40000	40000	40000	40000	35000	40000	35000	35000	35000	35000	35000	35000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cibubur	24000	22000	22000	26000	20000	20000	22000	22000	22000	25000	25000	25000	24000	22000	24000	24000	24000	25000
5	Pasar Cijantung	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
6	Pasar Cipete	25000	25000	25000	20000	20000	25000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
7	Pasar Cipinang	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	23000	22000	22000	22000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
9	Pasar Jembatan Lima	20000	19000	20000	22000	20000	20000	21000	18000	18000	20000	18000	20000	20000	19000	15000	17000	20000	20000

7. Cleansing Juli 2020



Spoon - Cabai Merah Keriting Juli 2020

File Edit View Action Tools Help

Search

Input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Agile
Cryptography
Job
Mapping
Bulk loading
Inline
Experimental
Deprecated

Welcome! Cleansing Cabai Merah Keriting Agustus 2020 Cabai Merah Keriting Juli 2020

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	35000	35000	35000	35000	35000	30000	30000	35000	30000	30000	30000	30000	30000
2	Pasar Cempaka Putih	15000	15000	15000	15000	15000	15000	15000	15000	25000	25000	25000	25000	25000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cibubur	20000	20000	20000	20000	28000	28000	22000	22000	22000	24000	24000	20000	22000
5	Pasar Cijantung	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
6	Pasar Cipete	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000

8. Cleansing Agustus 2020

Spoon - Cleansing Cabai Merah Keriting Agustus 2020 (changed)

File Edit View Action Tools Help

Search

Transformations
Cleansing Cabai Merah Keriting
Run configurations
Database connections
Steps
Hops
Partition schemas
Slave server
Kettle cluster schemas
Data Services
VFS Connections
Hadoop clusters

Welcome! Cleansing Cabai Merah Keriting Agustus 2020

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	30000	30000	30000	35000	35000	30000	30000	30000	30000	30000	35000	35000	35000
2	Pasar Cempaka Putih	22000	22000	22000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	30000
4	Pasar Cibubur	28000	28000	25000	22000	28000	24000	25000	25000	25000	26000	24000	24000	22000
5	Pasar Cijantung	28000	25000	25000	28000	24000	24000	24000	24000	24000	24000	24000	24000	24000
6	Pasar Cipete	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	20000

9. Cleansing September 2020



Spoon - Cabai Merah Keriting September 2020

File Edit View Action Tools Help

View Design

Search

- > Input
- > Output
- > Streaming
- > Transform
- > Utility
- > Flow
- > Scripting
- > Pentaho Server
- > Lookup
- > Joins
- > Data Warehouse
- > Validation
- > Statistics
- > Big Data
- > Agile
- > Cryptography
- > Job
- > Mapping
- > Bulk loading
- > Inline
- > Experimental
- > Deprecated
- > Utilities

Welcome! | Cleansing Cabai Merah Keriting Agustus 2020 | Cabai Merah Keriting September 2020

100%

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	30000	30000	35000	30000	35000	30000	30000	30000	35000	30000	30000	30000	30000
2	Pasar Cempaka Putih	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	25000	25000
3	Pasar Cengkareng	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
4	Pasar Cibubur	20000	22000	22000	20000	20000	20000	22000	22000	22000	22000	22000	22000	22000
5	Pasar Cijantung	20000	20000	20000	20000	20000	24000	24000	24000	24000	24000	24000	24000	24000
6	Pasar Cipete	20000	20000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000

10. Cleansing Oktober 2020

Spoon - Cabai Merah Keriting Oktober 2020

File Edit View Action Tools Help

View Design

Search

- > Input
- > Output
- > Streaming
- > Transform
- > Utility
- > Flow
- > Scripting
- > Pentaho Server
- > Lookup
- > Joins
- > Data Warehouse
- > Validation
- > Statistics
- > Big Data
- > Agile
- > Cryptography
- > Job
- > Mapping
- > Bulk loading
- > Inline
- > Experimental
- > Deprecated
- > Utilities

Welcome! | Cleansing Cabai Merah Keriting Agustus 2020 | Cabai Merah Keriting Oktober 2020

100%

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	40000	35000	35000	35000	35000	45000	45000	45000	45000	45000	40000	45000	45000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	30000	30000	30000	30000	40000	40000	40000	40000	40000	40000	40000	40000	40000
4	Pasar Cibubur	28000	28000	35000	30000	40000	38000	40000	40000	40000	45000	40000	40000	40000
5	Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
6	Pasar Cipete	25000	25000	25000	25000	28000	38000	40000	40000	40000	42000	36000	40000	40000

11. Cleansing November 2020



Spoon - Cabai Merah Keriting November 2020

File Edit View Action Tools Help

View Design

Search

- > Input
- > Output
- > Streaming
- > Transform
- > Utility
- > Flow
- > Scripting
- > Pentaho Server
- > Lookup
- > Joins
- > Data Warehouse
- > Validation
- > Statistics
- > Big Data
- > Agile
- > Cryptography
- > Job
- > Mapping
- > Bulk loading
- > Inline
- > Experimental
- > Deprecated
- > Utility

Welcome! | Cleansing Cabai Merah Keriting Agustus 2020 | Cabai Merah Keriting November 2020

100%

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	45000	45000	50000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	40000	40000	40000	40000
3	Pasar Cengkareng	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
4	Pasar Cibubur	45000	40000	38000	40000	38000	30000	30000	40000	40000	35000	35000	40000	40000
5	Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
6	Pasar Cipete	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000

12. Cleansing Desember 2020

Spoon - Cabai Merah Keriting Desember 2020

File Edit View Action Tools Help

View Design

Search

- > Input
- > Output
- > Streaming
- > Transform
- > Utility
- > Flow
- > Scripting
- > Pentaho Server
- > Lookup
- > Joins
- > Data Warehouse
- > Validation
- > Statistics
- > Big Data
- > Agile
- > Cryptography
- > Job
- > Mapping
- > Bulk loading
- > Inline
- > Experimental
- > Deprecated
- > History

Welcome! | Cleansing Cabai Merah Keriting Agustus 2020 | Cabai Merah Keriting Desember 2020

100%

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pasar Baru Metro Atom	45000	45000	55000	60000	60000	60000	65000	65000	65000	65000	65000	65000	65000
2	Pasar Cempaka Putih	45000	35000	45000	45000	45000	45000	50000	50000	50000	50000	53000	53000	53000
3	Pasar Cengkareng	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000
4	Pasar Cibubur	40000	45000	40000	40000	40000	40000	50000	55000	55000	55000	60000	55000	55000
5	Pasar Cijantung	45000	45000	45000	45000	50000	50000	50000	50000	50000	50000	50000	60000	60000
6	Pasar Cipete	45000	45000	46000	46000	46000	46000	55000	55000	55000	55000	55000	58000	58000

4. File Praktikum

Github Repository:

<https://github.com/adamzannuba11/PraktikumDWH>



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

5. Soal Latihan

Soal:

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

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	... Menit	...

Keterangan:

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

