


Nama: (Isi Nama Anda)	 Praktikum Data Warehouse	MODUL 10
NIM: (Isi NIM Anda)		Nama Dosen: Ir. Teddy Siswanto, MMSi
Hari/Tanggal: Hari, Tanggal Bulan 2022		Nama Asisten Labratorium: 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 064001900003

Proyek Akhir Praktikum II

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

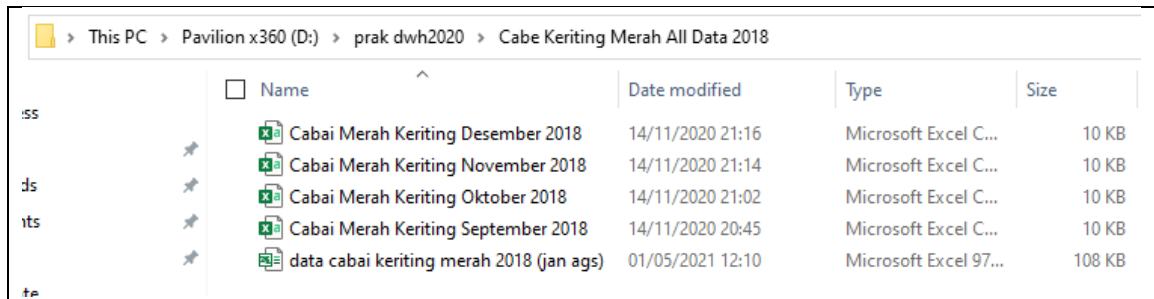
Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses 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

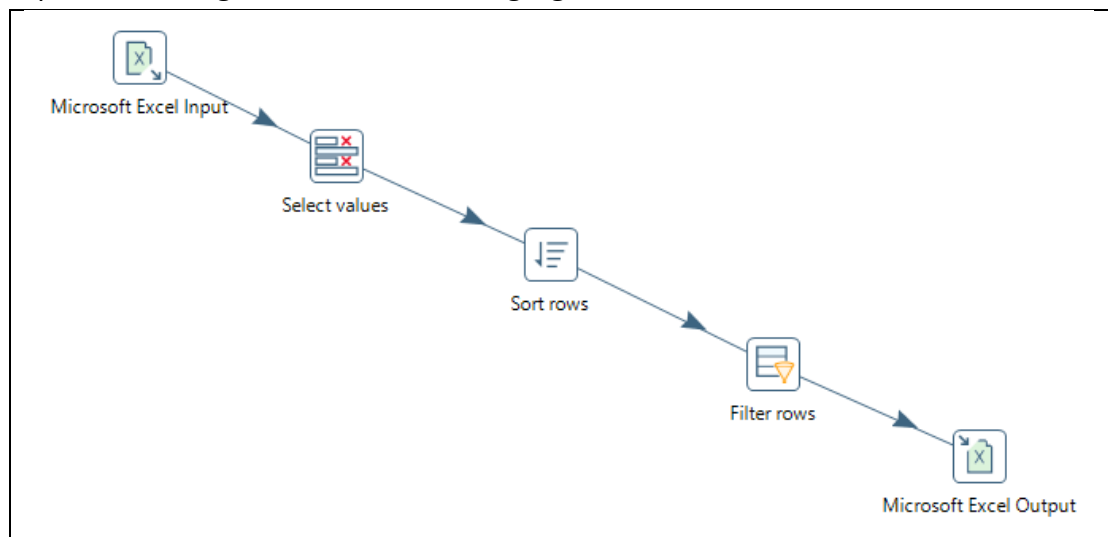
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



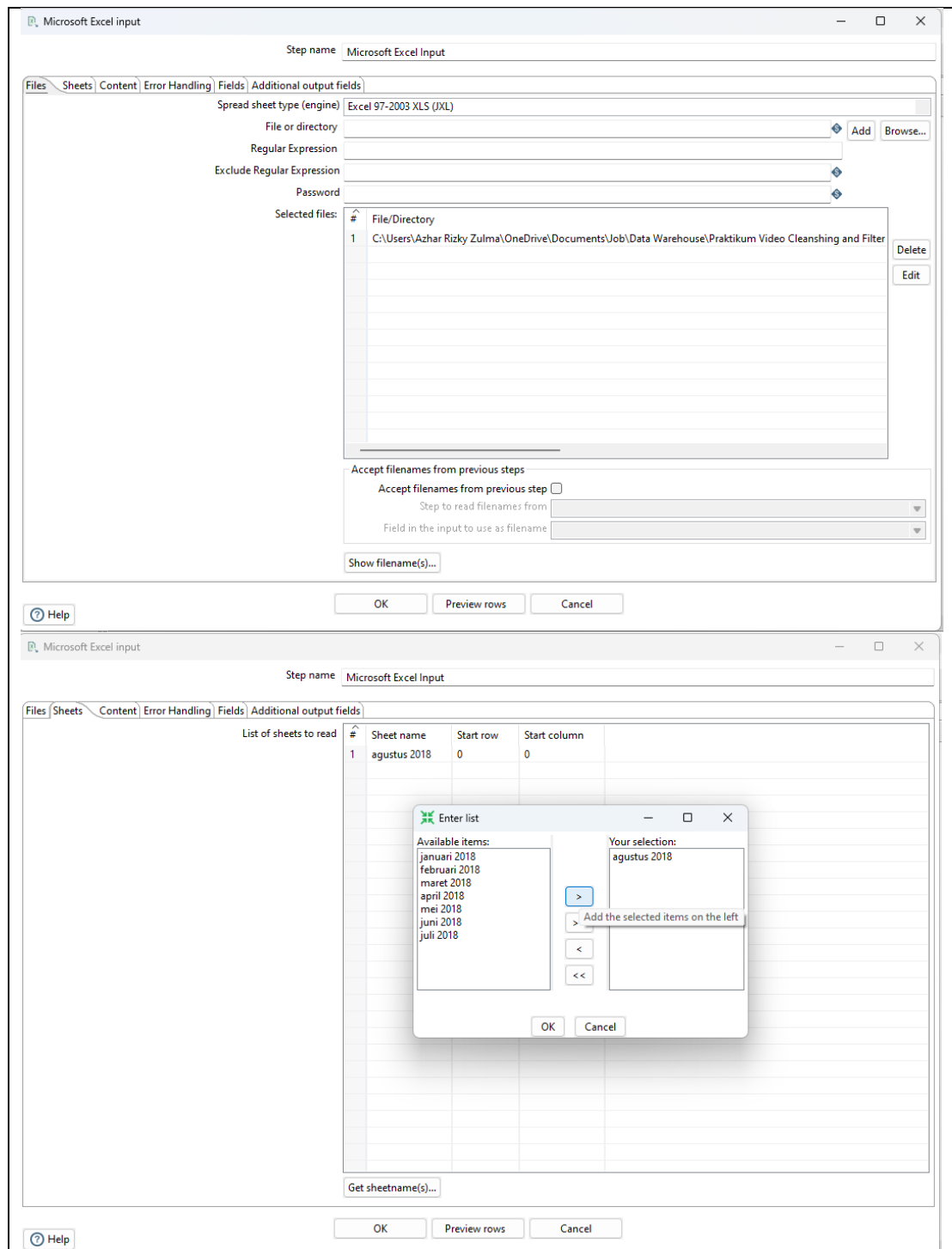
Name	Date modified	Type	Size
Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
Cabai Merah Keriting September 2018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

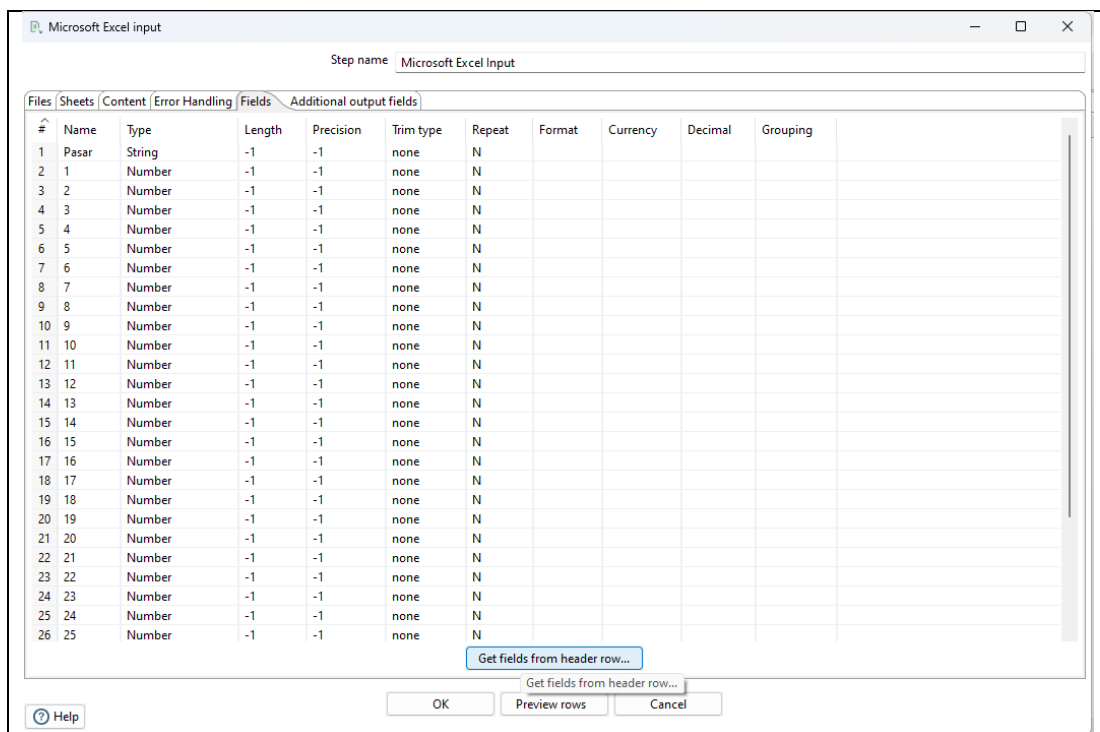
2. Excel Input

- 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 2019.ktr***



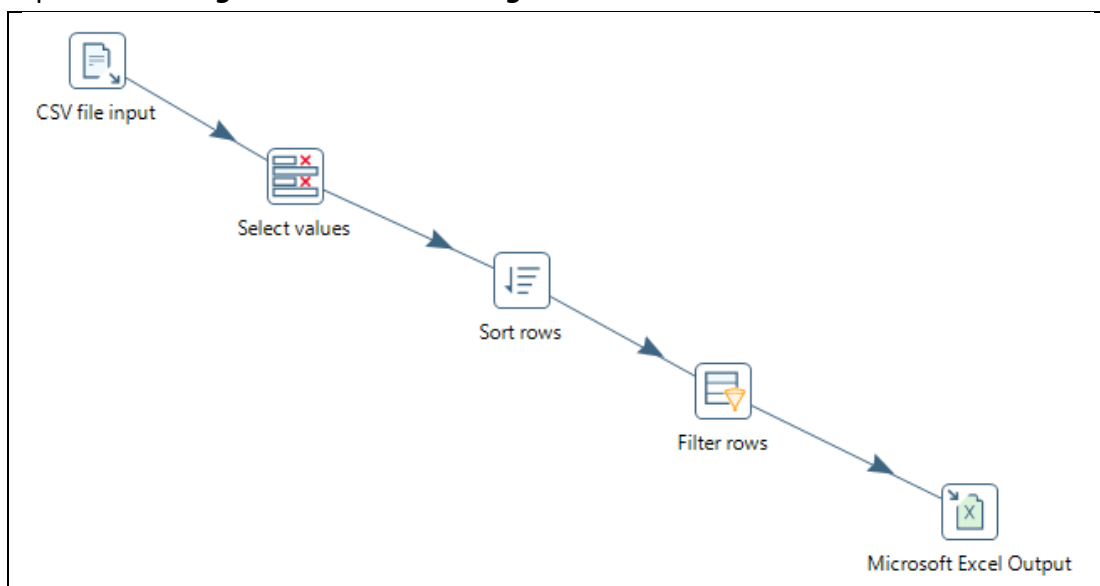
- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)



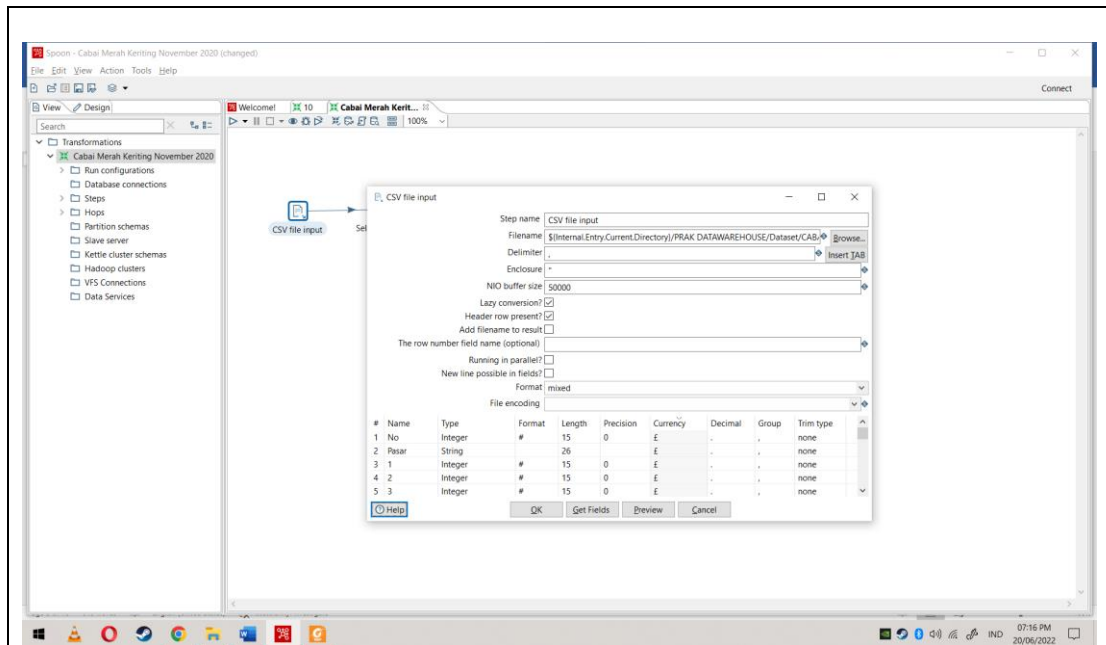


3. CSV Input

- 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 2019.ktr***

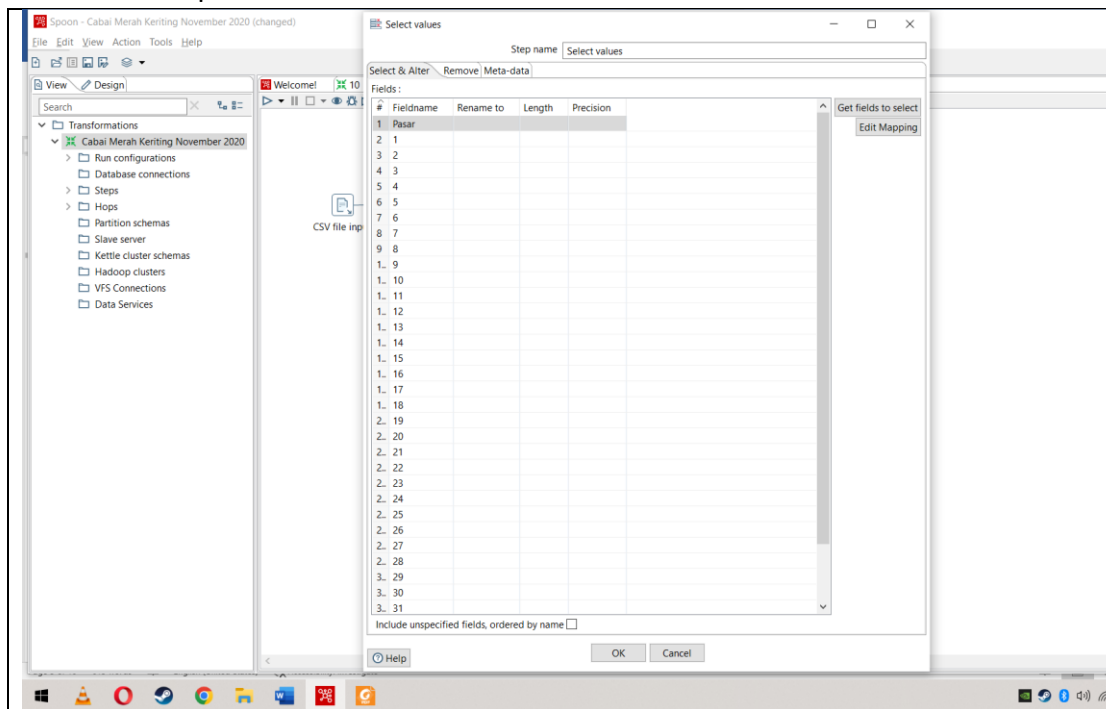


- Berikut ini adalah konfigurasi pada step CSV File Input.



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

Sort rows

Step name: Sort rows

Sort directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	Pasar	Y	N	N	2	N
2	1	Y	N	N	2	N
3	2	Y	N	N	2	N
4	3	Y	N	N	2	N
5	4	Y	N	N	2	N

Help OK Cancel Get Fields

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

Filter rows

Step name: Filter rows

Send 'true' data to step:

Send 'false' data to step:

The condition:

+

NOT (Pasar = [Pasar Anyer Bahari])

AND

NOT (Pasar = [Pasar Gondangdia])

AND

NOT (Pasar = [Pasar Induk Kramat Jati])

AND

NOT (Pasar = [Pasar Kalibaru])

AND

NOT (Pasar = [Pasar Kalideres])

Help OK Cancel



Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

Level 1. Select UP to go up one level

(String)

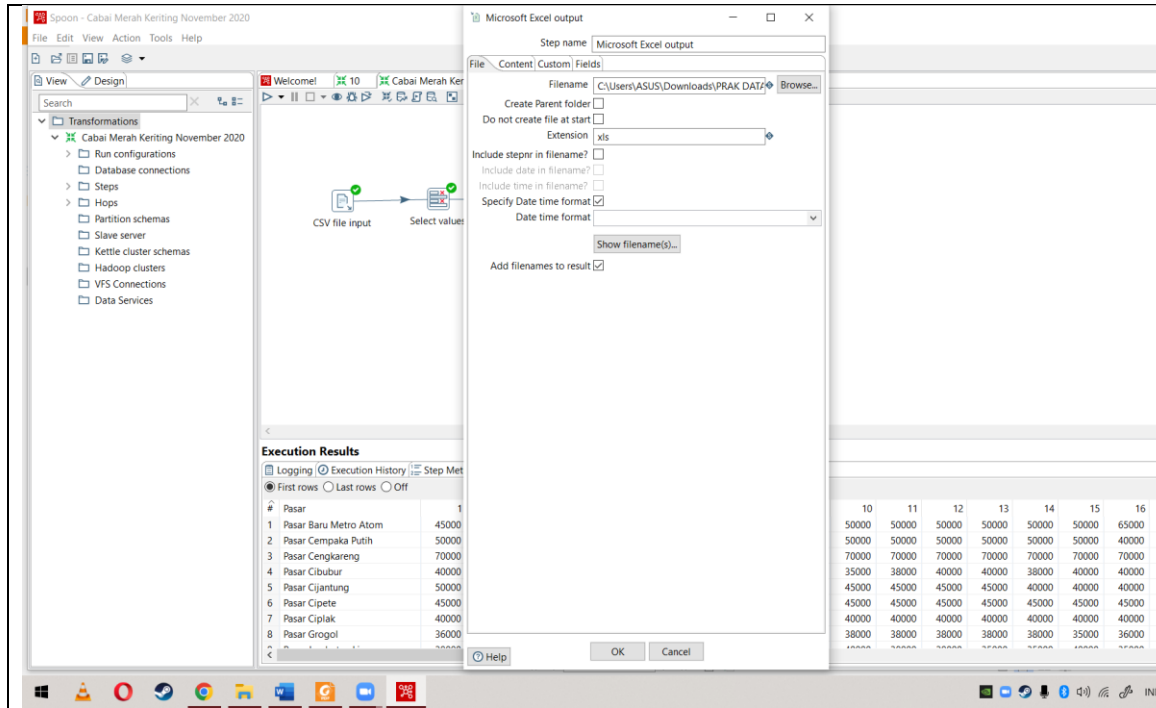
- Berikut List Data Nama Pasar yang mau dibuang (2019)

	NOT (Pasar = [Pasar Anyer Bahari])
AND	
	NOT (Pasar = [Pasar Gondangdia])
AND	
	NOT (Pasar = [Pasar Induk Kramat Jati])
AND	
	NOT (Pasar = [Pasar Kalibaru])
AND	
	NOT (Pasar = [Pasar Kalideres])
AND	
	NOT (Pasar = [Pasar Klender SS])
AND	
	NOT (Pasar = [Pasar Koja Baru])
AND	
	NOT (Pasar = [Pasar Rawa Mangun])
AND	
	NOT (Pasar = [Pasar Pramuka])
AND	
	NOT (Pasar = [Pasar Pluit])
AND	
	NOT (Pasar = [Pasar Pulo Gadung])
AND	
	NOT (Pasar = [Pasar Pesanggrahan])
AND	
	NOT (Pasar = [Pasar Glodok])
AND	
	NOT (Pasar = [Pasar Matraman KK])
AND	
	NOT (Pasar = [Pasar Jatinegara])



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 2019.xls*** (File & Sheet)



2. Alat dan Bahan

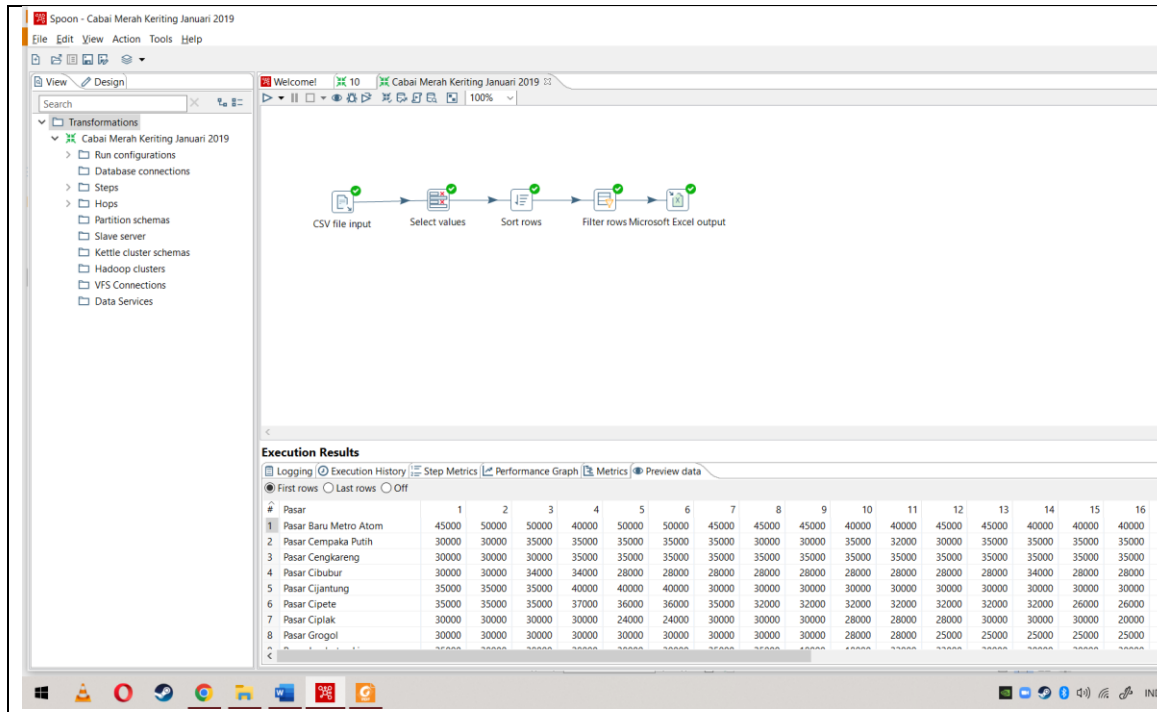
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

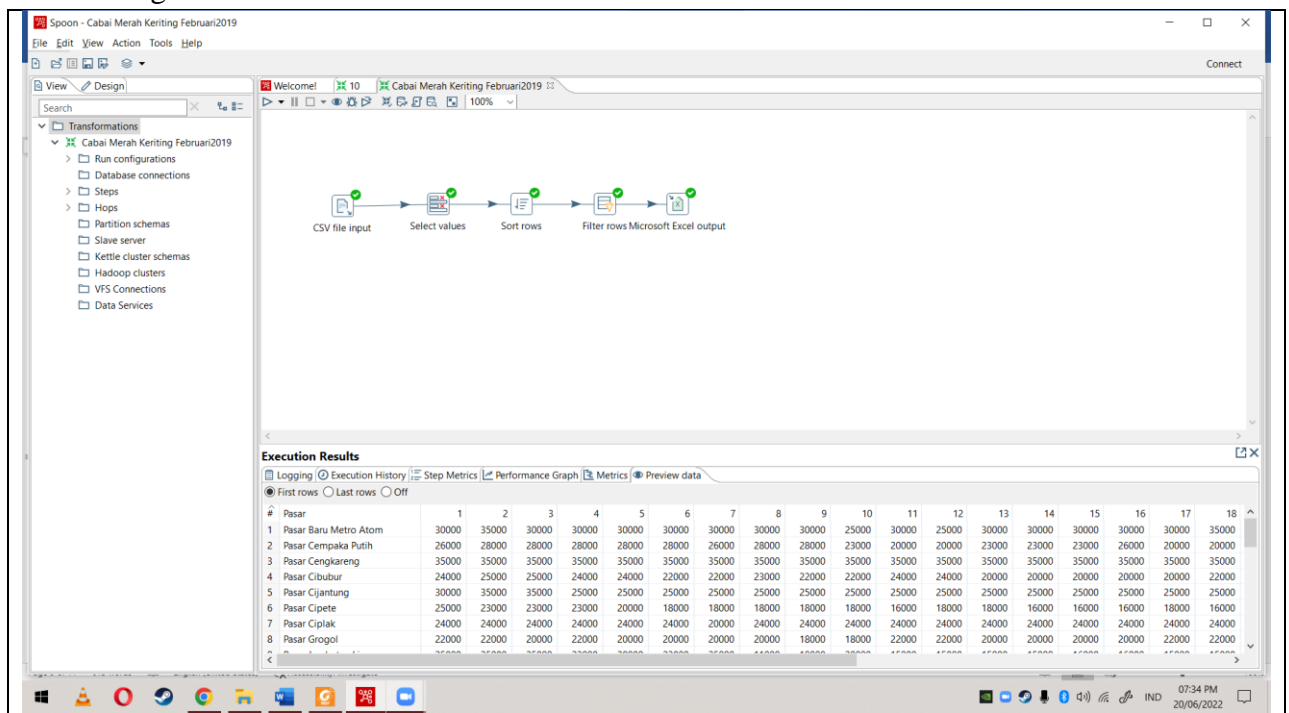
3. Elemen Kompetensi

- Tugas Akhir Praktikum II – Screenshot Struktur yang sudah berhasil berjalan
 - Cleansing Januari 2019



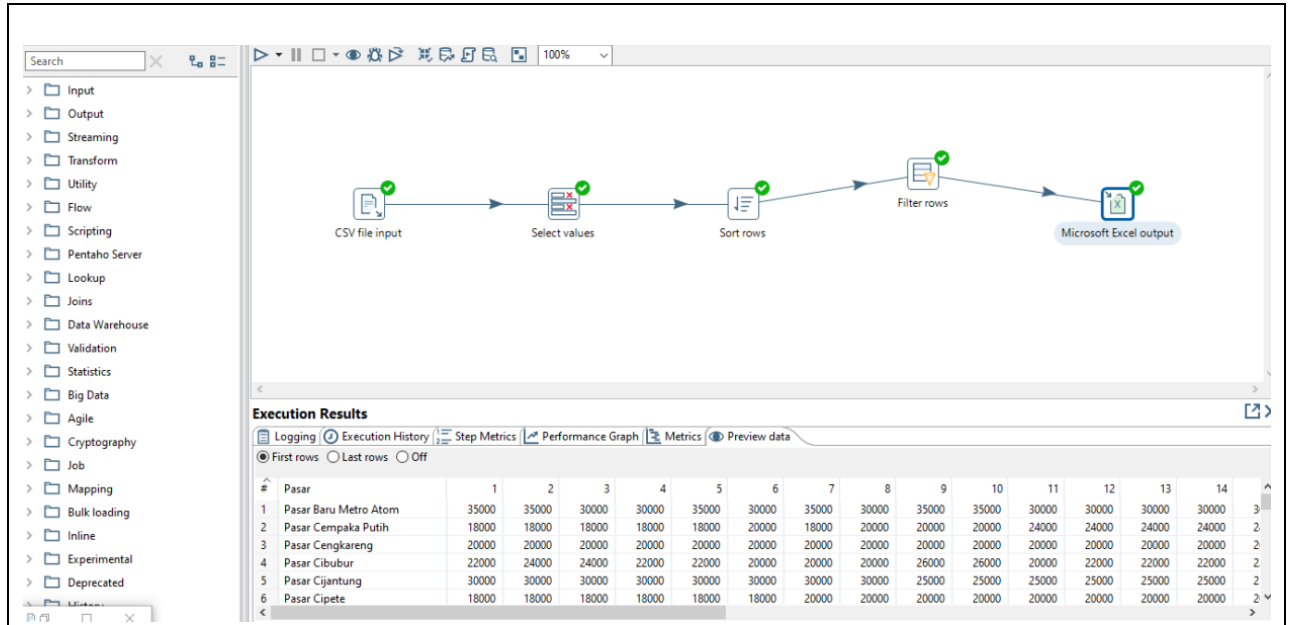


2. Cleansing Februari 2019

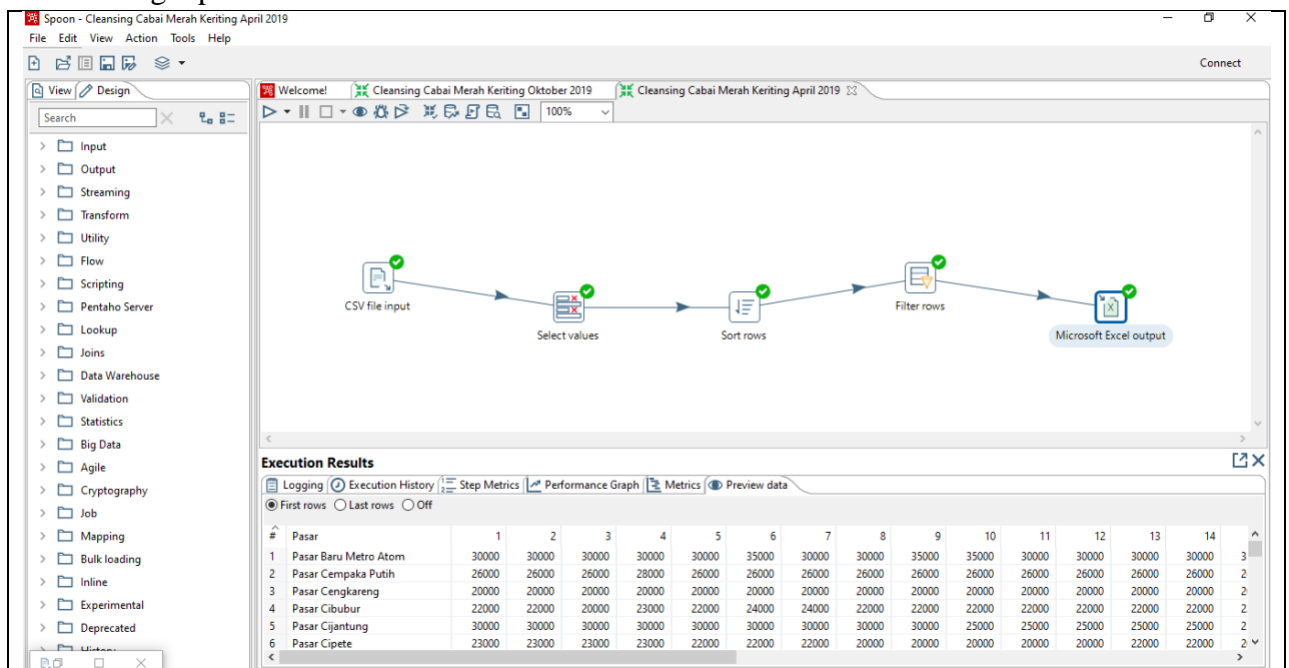


3. Cleansing Maret 2019



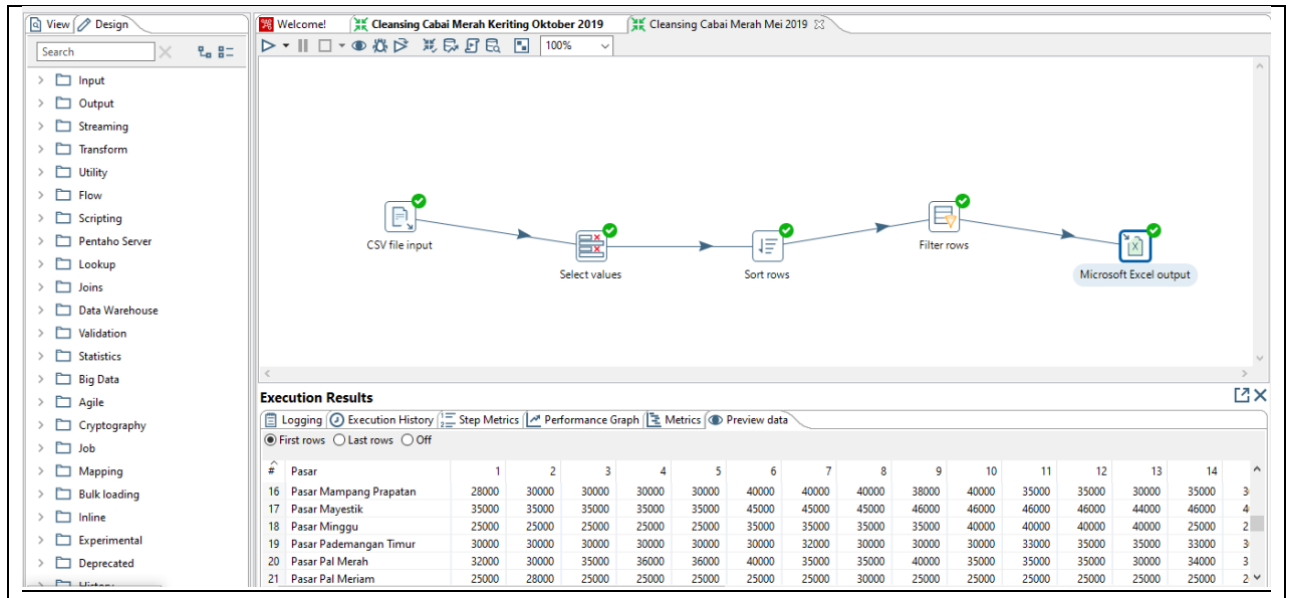


4. Cleansing April 2019

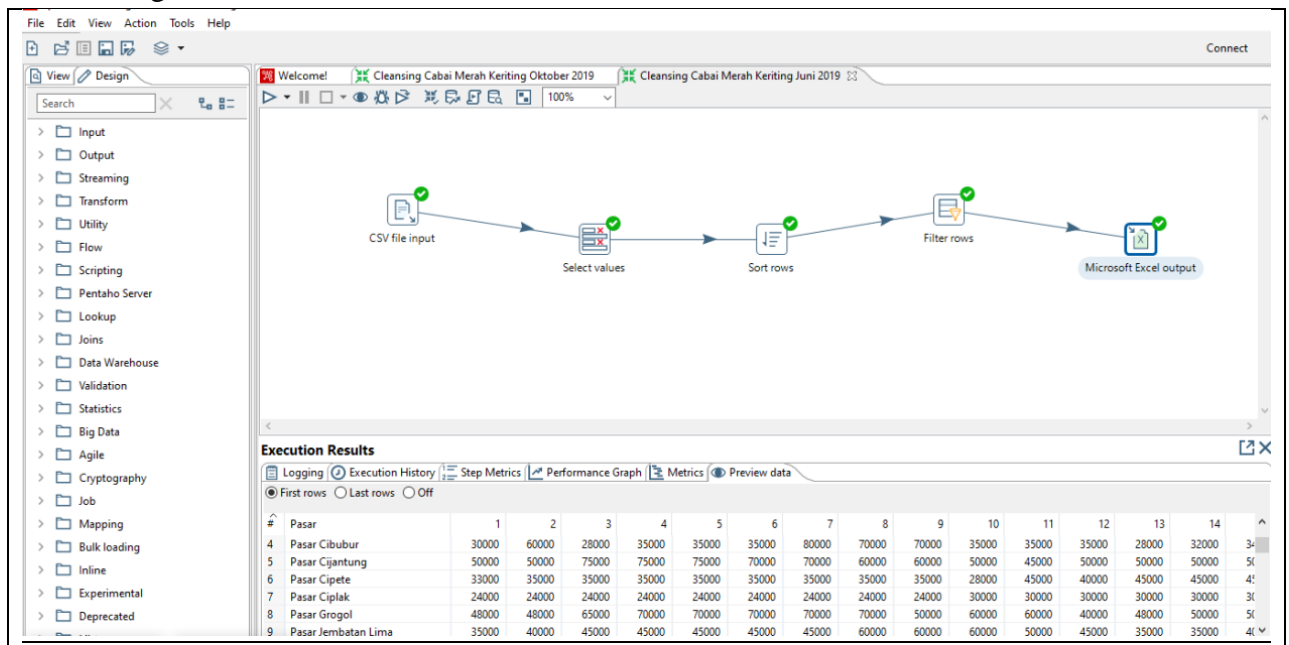


5. Cleansing Mei 2019





6. Cleansing Juni 2019



7. Cleansing Juli 2019



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

Welcome! Cleansing Cabai Merah Keriting Oktober 2019 Cleansing Cabai Merah Keriting Juli 2019

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	14	
1	Pasar Baru Metro Atom	60000	70000	70000	70000	75000	70000	70000	65000	70000	70000	70000	70000	70000	70000	7
2	Pasar Cempaka Putih	50000	50000	50000	50000	50000	50000	50000	60000	60000	60000	60000	60000	60000	60000	6
3	Pasar Cengkareng	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	70000	7
4	Pasar Cibubur	50000	55000	28000	45000	55000	55000	55000	55000	55000	65000	65000	60000	60000	60000	6
5	Pasar Cijantung	65000	65000	65000	70000	70000	70000	65000	70000	70000	70000	70000	70000	70000	68000	7
6	Pasar Cipete	55000	55000	60000	60000	60000	60000	65000	65000	65000	65000	60000	60000	60000	60000	6

8. Cleansing Agustus 2019

Spoon - Cleansing Cabai Merah Keriting Agustus 2019

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

Welcome! Cleansing Cabai Merah Keriting Oktober 2019 Cleansing Cabai Merah Keriting Agustus 2019

100%

CSV file input

Select values

Sort rows

Filter rows

Microsoft Excel output

Execution Results

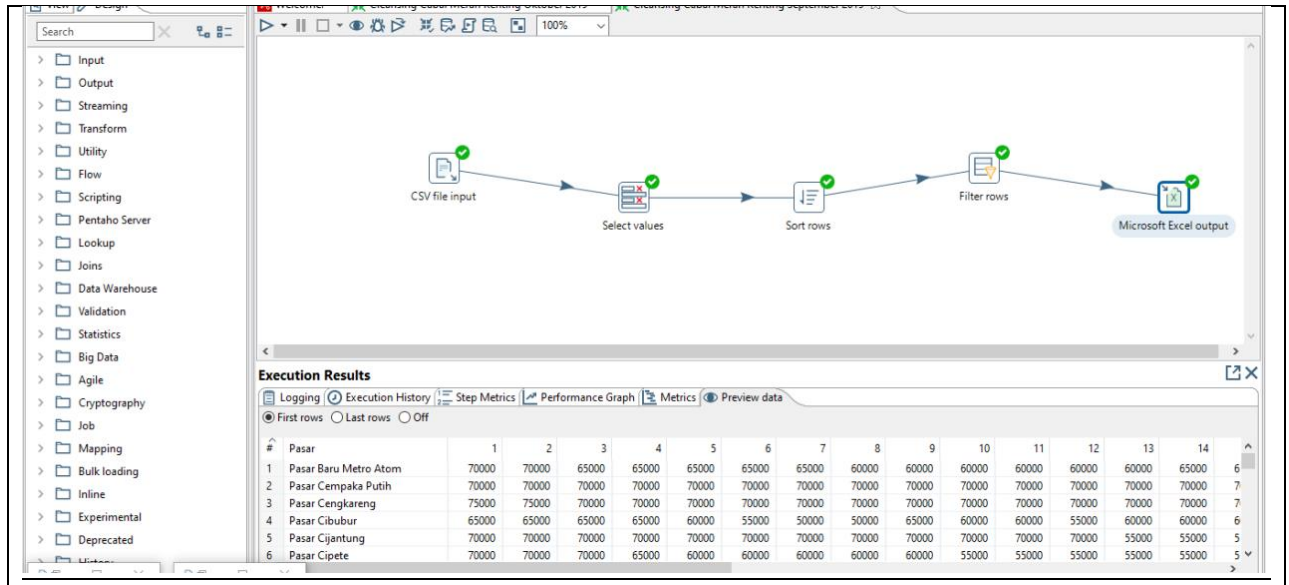
Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

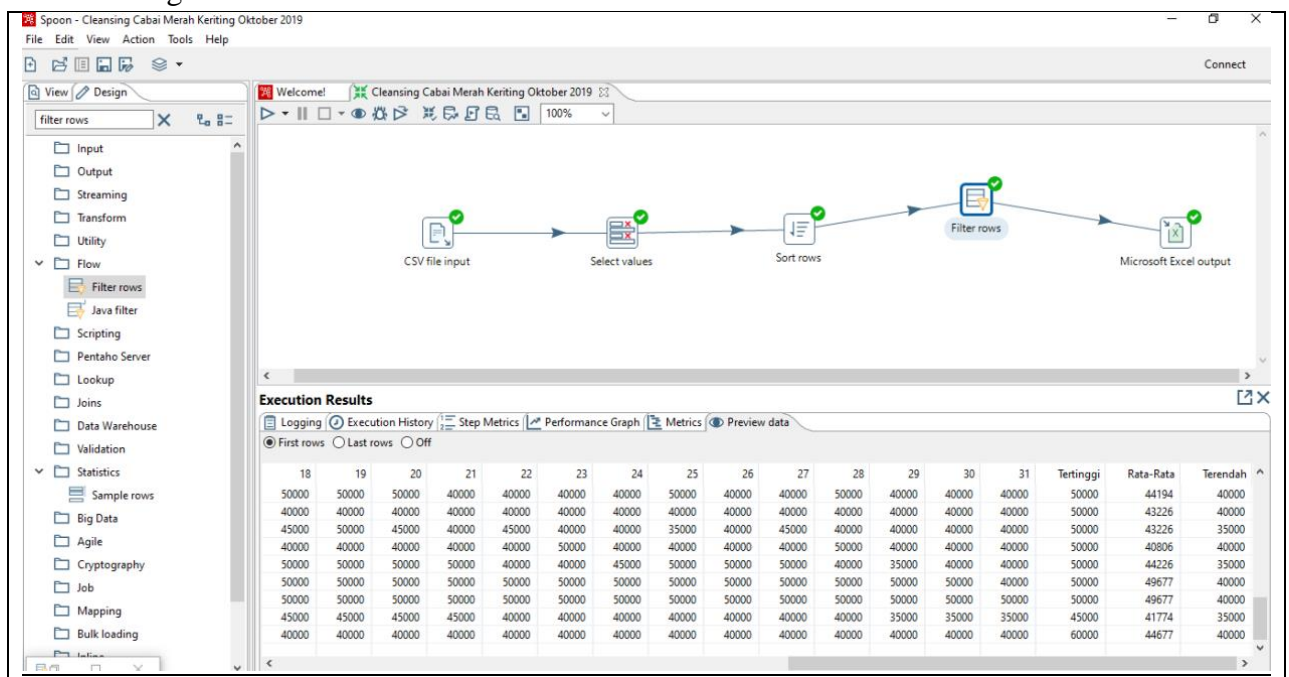
#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	Pasar Baru Metro Atom	80000	75000	75000	75000	70000	75000	75000	75000	75000	75000	75000	75000	75000	75000	7500
2	Pasar Cempaka Putih	60000	60000	72000	72000	72000	72000	72000	72000	72000	72000	72000	72000	85000	85000	8500
3	Pasar Cengkareng	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	7000
4	Pasar Cibubur	70000	65000	65000	65000	65000	65000	65000	70000	70000	60000	60000	60000	65000	7500	
5	Pasar Cijantung	70000	70000	70000	70000	70000	70000	70000	70000	90000	90000	90000	80000	85000	8500	
6	Pasar Cipete	60000	60000	60000	60000	65000	70000	70000	70000	75000	75000	75000	70000	80000	8500	

9. Cleansing September 2019





10. Cleansing Oktober 2019



11. Cleansing November 2019



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	45000	45000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	65000	60000	50000
2	Pasar Cempaka Putih	50000	50000	50000	50000	40000	40000	40000	40000	40000	50000	50000	50000	50000	50000	50000	40000	40000	40000
3	Pasar Cengkareng	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000
4	Pasar Cibubur	40000	40000	40000	40000	40000	38000	38000	35000	35000	38000	40000	40000	38000	40000	40000	40000	40000	40000
5	Pasar Cijantung	50000	50000	50000	45000	45000	45000	50000	50000	50000	45000	45000	45000	45000	40000	40000	40000	45000	45000
6	Pasar Cipete	45000	45000	45000	45000	45000	40000	40000	45000	45000	45000	45000	45000	45000	45000	45000	46000	46000	46000
7	Pasar Ciplak	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
8	Pasar Grogol	36000	35000	35000	38000	38000	38000	38000	38000	38000	38000	38000	38000	38000	35000	36000	36000	36000	36000

12. Cleansing Desember 2019

Execution Results

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Pasar Baru Metro Atom	40000	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	40000
3	Pasar Cengkareng	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
4	Pasar Cibubur	35000	35000	36000	34000	32000	34000	35000	35000	30000	32000	34000	34000	40000	40000
5	Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
6	Pasar Cipete	35000	35000	35000	35000	36000	35000	35000	35000	35000	40000	40000	40000	40000	35000

4. File Praktikum

Github Repository:

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

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.
- 2.

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 II	...	

8. Formulir Umpan Balik

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

Keterangan:

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

