# PRAKTIKUM DATA WAREHOUSING DAN DATA MINING MODUL 2

# DASAR ETL (EXTRACT TRANSFORM LOAD)



### Disusun oleh:

Adinda Aulia Hapsari

L200220037

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA
TAHUN 2024

Setelah kegiatan selesai, lembar kerja ini dicetak (di-print) dan dikumpulkan ke

asisten.

(Diisi oleh Asisten)

Nilai Praktek:

NIM : L200220037

Nama : Adinda Aulia Hapsari

Nama Asisten : Diva Halimah

Tanggal Praktikum: 27 September 2024

Tanda Tangan:

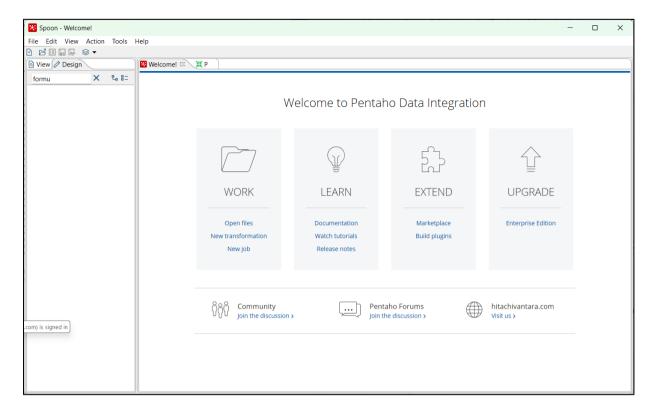
# **KEGIATAN PRAKTIKUM**

#### 2.4.1 Ekstraksi Data Excel ke Pentaho

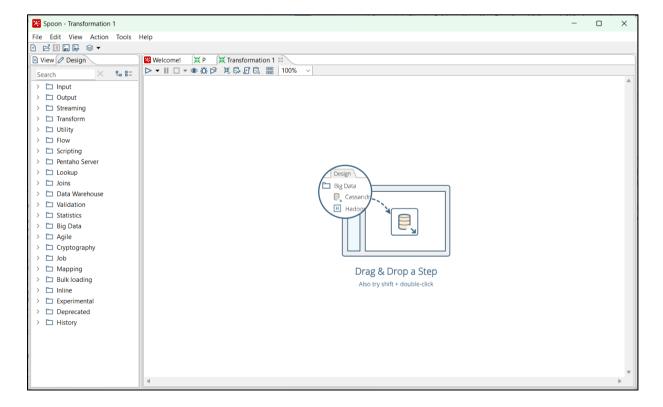
1. Buka program aplikasi Ms. Excel dan siapkan data seperti pada Tabel Target Penjualan di bawah ini. Anda juga bisa unduh pada URL berikut: <a href="https://gitea.ums.ac.id/yusufsn/Praktikum\_DWDM/src/branch/master/Data/ETL/Target%20Penjualan%20New.xlsx">https://gitea.ums.ac.id/yusufsn/Praktikum\_DWDM/src/branch/master/Data/ETL/Target%20Penjualan%20New.xlsx</a>.

Kode Cabang	Kode Kategori	Kode Produk	Tahun	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CABANG-555	PROD-0000001	PROD-000001	2008	3264	2965	5538	4248	3050	4823	3375	3922	3783	4154	3173	4058
CABANG-039	PROD-0000003	PROD-0000003	2009	1906	1580	2033	1789	2405	2435	2084	2767	2828	2469	2528	2235
CABANG-039	PROD-0000004	PROD-0000004	2009	3554	2800	2506	2864	3340	3302	3172	2120	2054	2321	3200	3449
CABANG-039	PROD-0000005	PROD-0000005	2008	2667	2002	2021	3439	2169	3051	3159	2040	2238	2249	1932	1822
CABANG-039	PROD-0000006	PROD-0000006	2008	3940	3600	3282	2407	3773	2341	3742	2636	2859	3091	2317	4374
CABANG-039	PROD-0000007	PROD-0000007	2008	1670	1124	1763	1567	1609	1605	2026	1277	1173	1514	1725	1867
CABANG-039	PROD-0000008	PROD-0000008	2008	452	375	380	282	287	369	490	343	293	348	345	458
CABANG-039	PROD-0000009	PROD-0000009	2008	1353	1791	1996	2008	1466	1881	1939	2293	1220	1377	1262	1546
CABANG-039	PROD-0000010	PROD-0000010	2008	308	455	487	521	506	441	520	359	431	536	321	389
CABANG-039	PROD-0000011	PROD-0000011	2008	521	380	570	377	292	516	347	429	516	392	427	491
CABANG-039	PROD-0000012	PROD-0000012	2008	1005	1047	1628	1071	1467	1183	903	1351	1199	1305	1319	1363
CABANG-039	PROD-0000013	PROD-0000013	2008	3777	5230	3465	5721	5330	5617	4402	4350	5907	4372	5754	5280
CABANG-039	PROD-0000014	PROD-0000014	2008	2976	4216	4554	3105	4049	4243	3622	4030	5725	5603	3527	4656
CABANG-039	PROD-0000015	PROD-0000015	2008	9223	8017	5016	7549	5629	8474	6421	7097	7395	7864	7092	4807
CABANG-039	PROD-0000016	PROD-0000016	2008	####	####	####	####	####	####	####	####	####	####	####	####
CABANG-039	PROD-0000017	PROD-0000017	2008	533	343	465	528	379	524	411	445	382	294	594	629
CABANG-039	PROD-0000018	PROD-0000018	2008	4045	2525	3722	3236	2930	2784	4338	3413	4343	3666	4440	4364
CABANG-039	PROD-0000019	PROD-0000019	2008	1674	2384	2176	1653	2734	1863	1714	1753	1912	2324	3077	1936
CABANG-039	PROD-0000020	PROD-0000020	2008	1881	1698	1722	1330	2007	2147	2042	1480	1545	2332	2684	2329
CABANG-039	PROD-0000021	PROD-0000021	2008	4430	4889	3743	3536	4923	3172	5059	4787	3855	3888	3235	3597
CABANG-039	PROD-0000022	PROD-0000022	2008	####	####	####	####	####	####	####	####	####	####	####	####
CABANG-039	PROD-0000023	PROD-0000023	2008	5151	5229	4500	3906	4795	4036	6414	5240	4303	6353	5265	4573
CABANG-039	PROD-0000024	PROD-0000024	2008	5457	4905	4278	3950	4902	4126	6493	5523	5208	5761	6775	4910
CABANG-039	PROD-0000025	PROD-0000025	2008	2838	2865	2924	3629	3481	3931	3350	3541	2751	2776	2826	3657
CABANG-039	PROD-0000026	PROD-0000026	2008	266	255	511	276	248	229	232	371	284	233	316	361
CABANG-039	PROD-0000027	PROD-0000027	2008	4493	3328	4594	3387	3443	4697	4157	4334	4433	4044	4179	5182

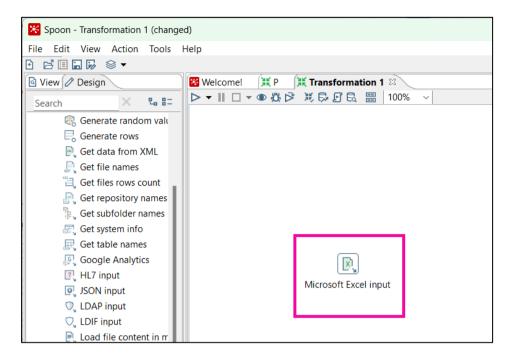
2. Jalankan aplikasi Pentaho Data Integration dengan mengaktifkan file Spoon.bat (klik kanan >> Run as Administrator), maka akan muncul tampilan awal seperti gambar di bawah.



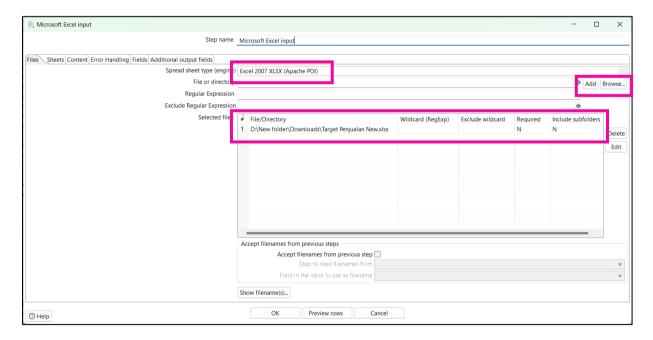
3. Buat transformasi baru dengan File >> New >> Transformation (CTRL+N).



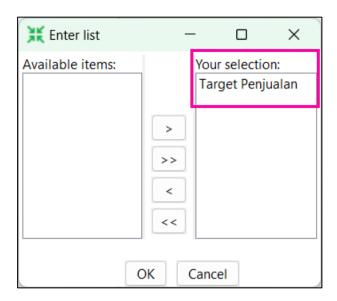
4. Pada sidebar sebelah kiri pilih tab Design, ketikkan Microsoft Excel input pada kolom Search dan tarik ke canvas.



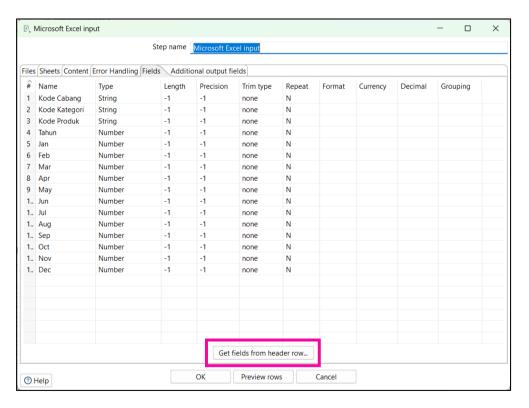
5. Lakukan konfigurasi pada step Microsoft Excel input dengan cara klik kanan >> Edit. Pada tab Files ganti Spread sheet type menjadi Excel 2007, lalu pilih file excel yang ingin di-ekstrak pada File or directory dan tekan tombol Add. Pastikan file excel tersebut ada di tabel Selected Files.



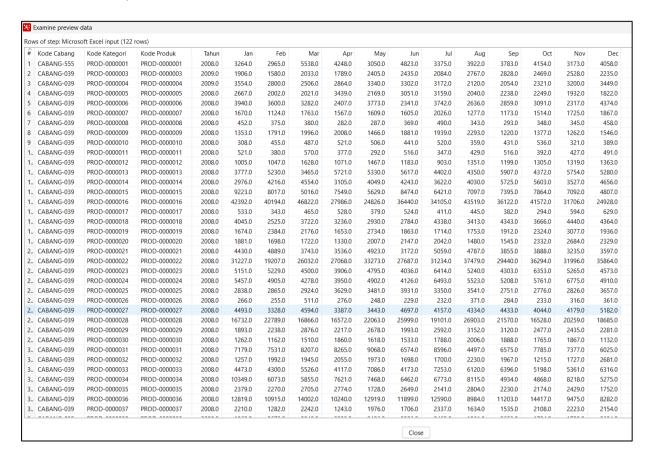
6. Pada tab !Sheets, tekan tombol Get sheetname(s)... untuk memilih sheet pada file excel yang ingin di-ekstrak datanya. Pilih sheet pada Available items dan pindahkan ke kotak Your selection. Tujuannya agar sheet yang kita pilih sesuai, umumnya hal ini akan berpengaruh apabila pada file excel tersebut terdapat lebih dari satu sheet.



7. Pada tab !Fields, tekan tombol Get fields from header row... untuk mendapatkan semua kolom yang ada pada sheet excel yang kita pilih.



8. Tekan tombol Preview rows untuk memastikan apakah data yang akan di-ekstrak sudah benar. Jika sudah benar, maka tekan tombol OK.

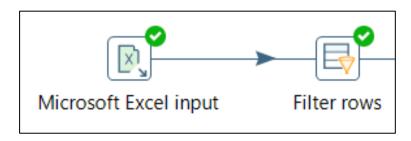


9. Proses ekstraksi selesai.

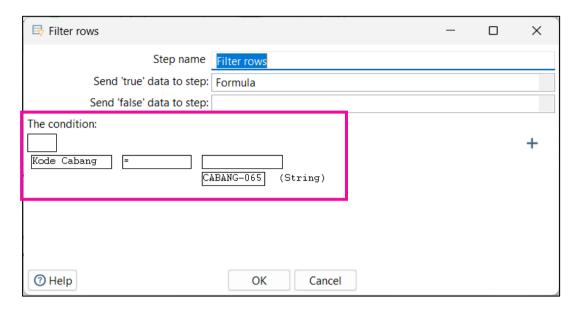
#### 2.4.2 Transformasi Data

Pada kegiatan ini, akan dilakukan proses transformasi data yaitu filtering (filter data sehingga hanya menyisakan Kode Cabang = 'CABANG-065'), dilanjutkan melakukan penghitungan Total\_Target\_Q1 dengan menambahkan data target pada bulan di Kuartal pertama (Jan, Feb, Mar). Tahapan praktikum sebagai berikut:

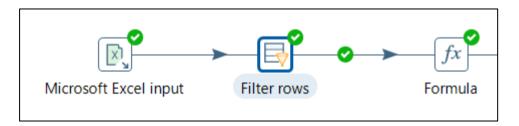
1. Pada sidebar Design ketik Filter rows dan lakukan drag and drop ke canvas. Hubungkan step 1 (Microsoft Excel input) dengan step 2 (Filter rows) dengan cara tekan tombol Shift lalu klik kiri dan tarik dari step 1 hingga terhubung ke step 2.

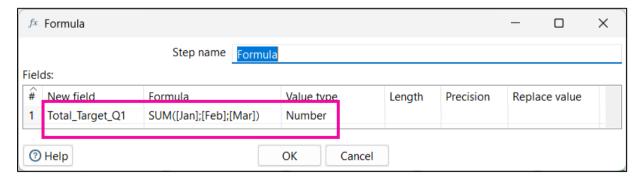


2. Konfigurasi filtering (filter data sehingga hanya menyisakan Kode Cabang = 'CABANG-065'). Klik kanan pada Filter rows pilih Edit, tambahkan kondisi seperti berikut lalu klik OK.

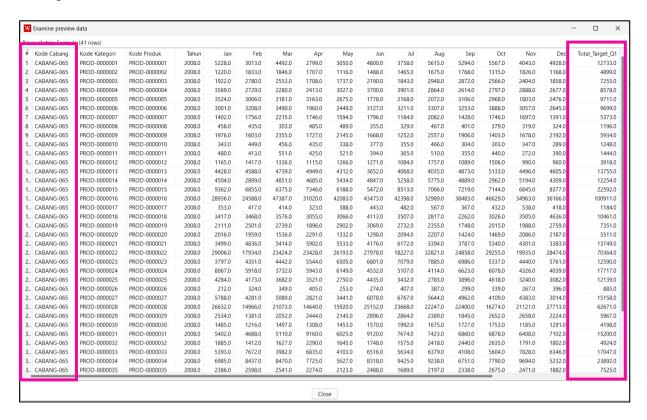


- 3. Jalankan dengan menekan tombol Run dan simpan.
- 4. Langkah selanjutnya yaitu melakukan penghitungan Total\_Target\_ Q1. Pada sidebar Design, ketik Formula lalu drag and drop ke canvas, hubungkan dengan step sebelumnya (pilih Result is TRUE) lalu klik kanan dan Edit. Isikan seperti berikut.

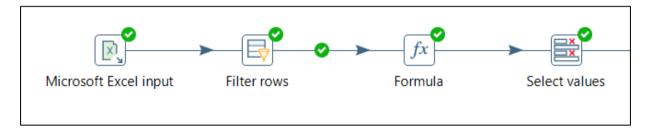




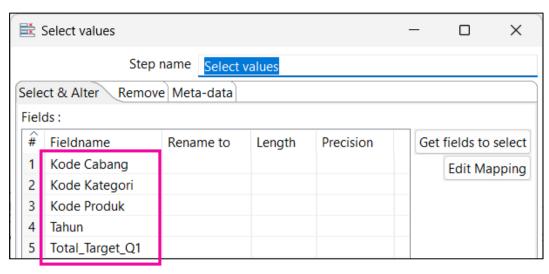
5. Jalankan dengan menekan tombol Run. Untuk melihat apakah hasil yang diharapkan sudah sesuai atau belum, klik kanan pada step Formula dan pilih Preview >> Quick Launch. Jika sudah sesuai maka proses transformasi data sudah selesai.

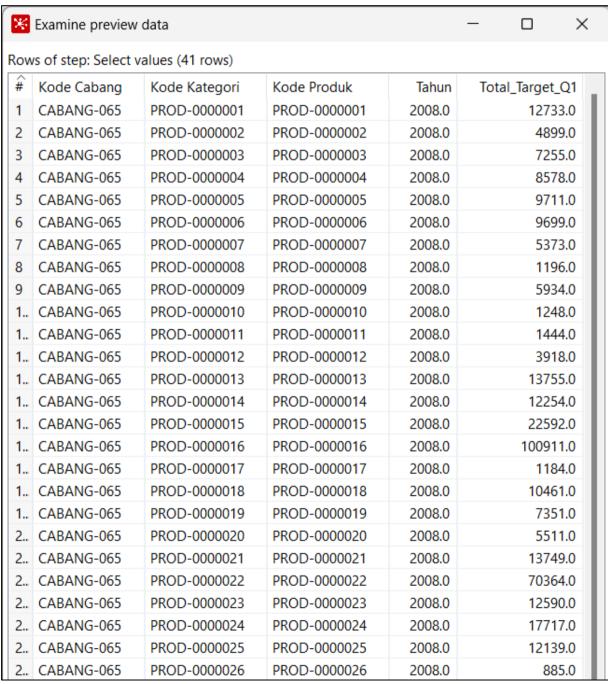


6. Selanjutnya kita dapat menyimpan data hasil transformasi data tersebut pada komputer dalam format tertentu, pada praktikum ini kita akan menyimpan dalam format excel. Pada sidebar Design pilih Select values lalu drag and drop ke canvas, hubungkan dengan step sebelumnya lalu klik kanan >> Edit seperti berikut.

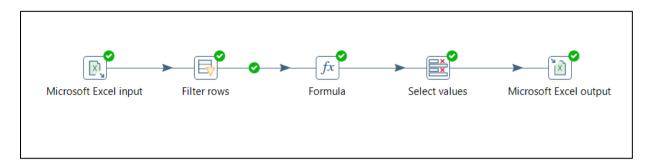


Pada praktikum ini kolom yang akan ditampilkan hanya lima saja, yaitu : Kode Cabang, Kode Kategori, Kode Produk, Tahun dan Total\_Target\_Q1. Anda juga bisa menambahkan kolom bulan Jan-Des jika diperlukan. Periksa hasilnya dengan klik kanan Preview >> Quick Launch.

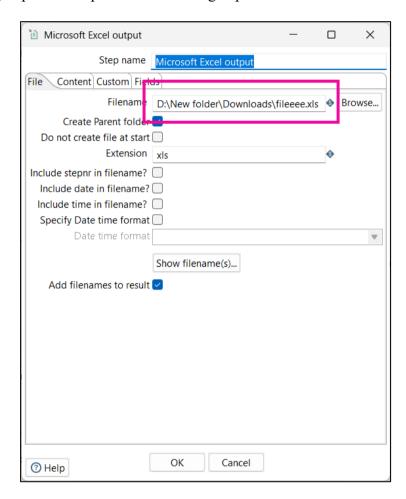




7. Selanjutnya pada sidebar Design, ketik Microsoft Excel output lalu drag and drop ke canvas, hubungkan dengan step terakhir (Select values) seperti berikut ini.



8. Atur lokasi penyimpanan dan penamaan file dengan pilih klik kanan >> Edit.



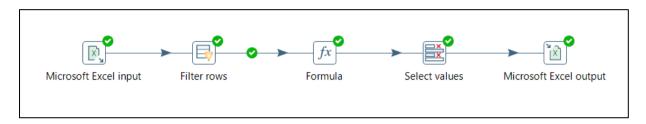
9. Jalankan dengan menekan tombol Run. Untuk melihat apakah isi file sudah sesuai atau belum dapat dilihat pada direktori lokasi file yang dipilih. Berikut adalah hasilnya.

A1	. ~	$[\times \checkmark fx]$	Kode Cabang		
4	Α	В	С	D	Е
1	Kode Cabang	Kode Kategori	Kode Produk	Tahun	Total_Target_Q1
2	CABANG-065	PROD-0000001	PROD-0000001	2.008,00	12.733,00
3	CABANG-065	PROD-0000002	PROD-0000002	2.008,00	4.899,00
4	CABANG-065	PROD-0000003	PROD-0000003	2.008,00	7.255,00
5	CABANG-065	PROD-0000004	PROD-0000004	2.008,00	8.578,00
6	CABANG-065	PROD-0000005	PROD-0000005	2.008,00	9.711,00
7	CABANG-065	PROD-0000006	PROD-0000006	2.008,00	9.699,00
8	CABANG-065	PROD-0000007	PROD-0000007	2.008,00	5.373,00
9	CABANG-065	PROD-0000008	PROD-0000008	2.008,00	1.196,00
10	CABANG-065	PROD-0000009	PROD-0000009	2.008,00	5.934,00
11	CABANG-065	PROD-0000010	PROD-0000010	2.008,00	1.248,00
12	CABANG-065	PROD-0000011	PROD-0000011	2.008,00	1.444,00
13	CABANG-065	PROD-0000012	PROD-0000012	2.008,00	3.918,00
14	CABANG-065	PROD-0000013	PROD-0000013	2.008,00	13.755,00
15	CABANG-065	PROD-0000014	PROD-0000014	2.008,00	12.254,00
16	CABANG-065	PROD-0000015	PROD-0000015	2.008,00	22.592,00
17	CABANG-065	PROD-0000016	PROD-0000016	2.008,00	100.911,00
18	CABANG-065	PROD-0000017	PROD-0000017	2.008,00	1.184,00
19	CABANG-065	PROD-0000018	PROD-0000018	2.008,00	10.461,00
20	CABANG-065	PROD-0000019	PROD-0000019	2.008,00	7.351,00
21	CABANG-065	PROD-0000020	PROD-0000020	2.008,00	5.511,00
22	CABANG-065	PROD-0000021	PROD-0000021	2.008,00	13.749,00
23	CABANG-065	PROD-0000022	PROD-0000022	2.008,00	70.364,00
24	CABANG-065	PROD-0000023	PROD-0000023	2.008,00	12.590,00
25	CABANG-065	PROD-0000024	PROD-0000024	2.008,00	17.717,00
26	CABANG-065	PROD-0000025	PROD-0000025	2.008,00	12.139,00
27	CABANG-065	PROD-0000026	PROD-0000026	2.008,00	885,00
28	CABANG-065	PROD-0000027	PROD-0000027	2.008,00	15.158,00
29	CABANG-065	PROD-0000028	PROD-0000028	2.008,00	62.671,00
30	CABANG-065	PROD-0000029	PROD-0000029	2.008,00	5.967,00
	< >	Sheet1	+		

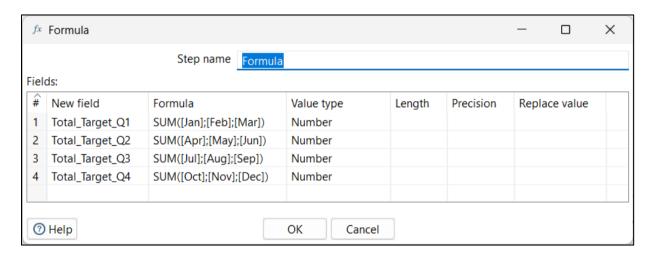
#### **TUGAS**

Dengan menggunakan file excel "Target Penjualan New.xlsx", tambahkan penghitungan Total Target pada Quartal 2 (Q2), Total Target pada Quartal 3 (Q3) serta Total Target pada Quartal 4 (Q4)!

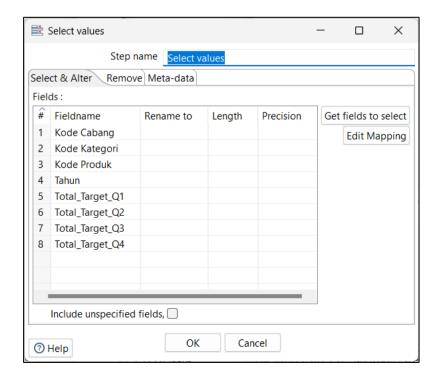
a. Menggunakan susunan design yang sama dengan kegiatan praktikum sebelumnya.



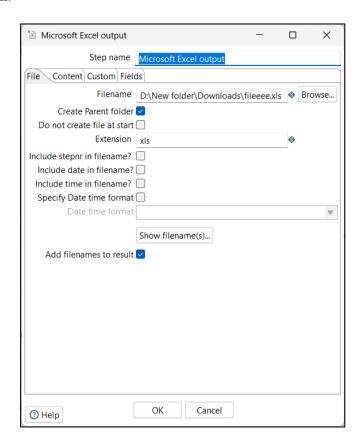
b. Menambahkan field Target Q2 hingga Q4 pada Formula.



c. Menambahkan Total Target Q2 hingga Q4 pada Select Values.



#### d. Simpan file, lalu run.



## e. Hasil pada file excel.

