

# DATA ANALYSIS AND INTEGRATION

## Project Report

---

GONÇALO GALANTE, 96216

SIMÃO FONSECA, 96911

Grupo 12

# PRESENTATION TOPICS

- 0 BRIEF INTRODUCTION**
- 1 TASK 1:** NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023
- 2 TASK 2:** SCATTER PLOT BETWEEN THE NUMBER OF CONTRACTS AND THE ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023
- 3 TASK 3:** CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B
- 4 TASK 4:** DATA WAREHOUSE SCHEMA CREATION WITH SQL

# PRESENTATION TOPICS

- 5 **TASK 5:** TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES
- 6 **TASK 6:** TRANSFORMATION TO POPULATE THE FACT TABLE
- 7 **TASK 7:** OLAP CUBE DEFINITION BASED ON THE PREVIOUS DATA WAREHOUSE
- 8 **TASK 8:** DATA CUBE ANALYSIS
- 9 **END OF THE PRESENTATION**

# BRIEF INTRODUCTION

In this project, two real-world datasets about the energy distribution network in Portugal were used.



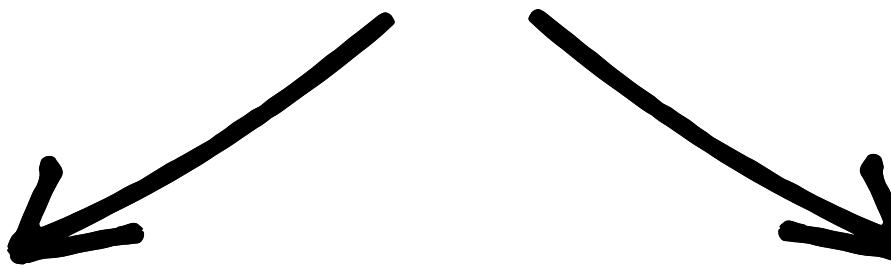
**Number of active energy  
contracts by meter type**

DATASET A



**Monthly consumption  
by municipality**

DATASET B



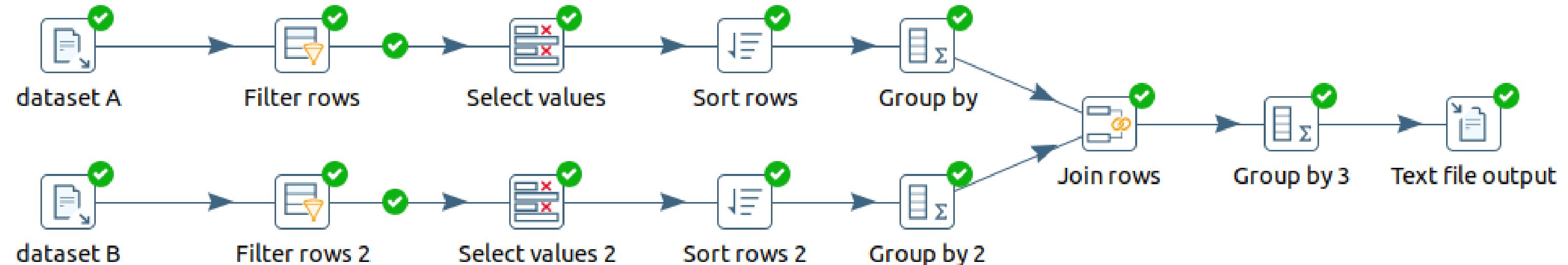
# TASK 1

NUMBER OF CONTRACTS AND ENERGY  
CONSUMPTION PER MUNICIPALITY IN JUNE 2023

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



To get the number of contracts and energy consumption per municipality in June 2023, a transformation in Pentaho was made:

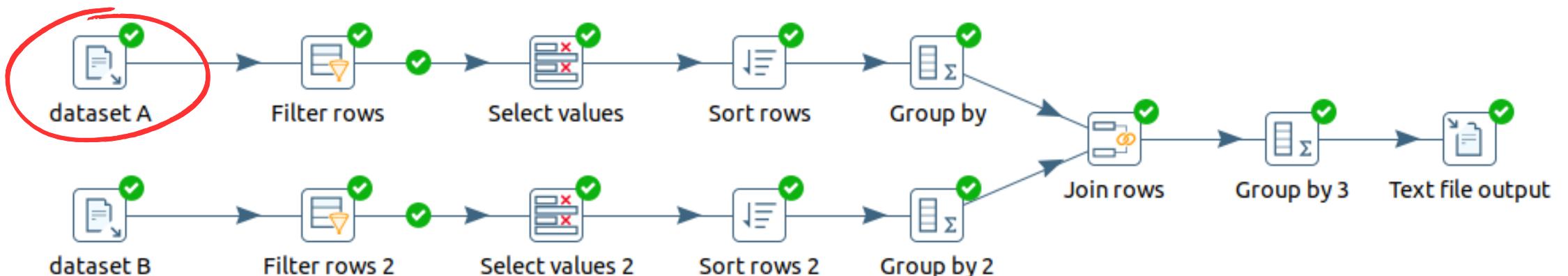


The Configuration Windows and Previews for each step, as well as the final transformation's output are shown in the next slides.

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Entire Transformation



CSV File input (dataset A) Configuration Window

**CSV File input**

Step name: dataset A

Filename: /home/aid/Desktop/21-contadores-de-energia.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	String		7		\$	.	,	none
4 District	String		16		\$	.	,	none
5 Municipality	String		27		\$	.	,	none
6 Parish	String		138		\$	.	,	none
7 incl. Smart Meter	String		3		\$	.	,	none
8 Number of CPE's	Integer	#	15	0	\$	.	,	none
9 DistrictCode	Integer	#	15	0	\$	.	,	none
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	,	none
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

CSV File input (dataset A) Preview

Rows of step: dataset A (10000 rows)

Examine preview data

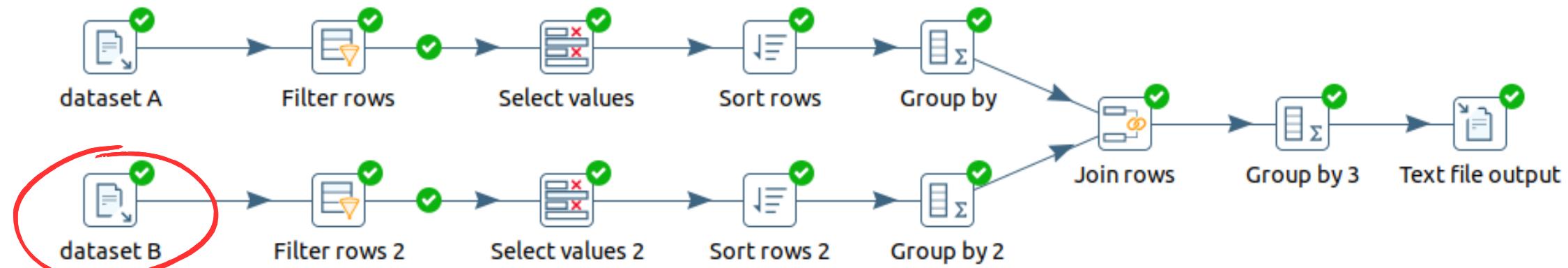
	Year	Month	Date	District	Municipality	Parish	incl. Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2023	8	2023-08	Braga	Vila Nova de Famalicão	União das freguesias de Carreira e Bente	NÃO	120	3	312	031253
2	2022	7	2022-07	Faro	Loulé (São Sebastião)	Loulé (São Sebastião)	NÃO	1714	8	808	080809
3	2022	11	2022-11	Coimbra	Condeixa-a-Nova	Ega	SIM	658	6	604	060406
4	2022	6	2022-06	Lisboa	Cadaval	União das freguesias de Painho e Figueiros	NÃO	947	11	1104	110413
5	2022	9	2022-09	Aveiro	Anadia	São Lourenço do Bairro	SIM	396	1	103	010310
6	2022	3	2022-03	Guarda	Vila Nova de Foz Côa	Vila Nova de Foz Côa	NÃO	1515	9	914	091419
7	2022	3	2022-03	Coimbra	Penacova	Sazes do Lorvão	NÃO	299	6	613	061310
8	2022	3	2022-03	Porto	Lousada	União das freguesias de Silvares, Pias, Nogueira e Alvarenga	NÃO	2445	13	1305	130532
9	2022	8	2022-08	Viana do Castelo	Viana do Castelo	Amonde	SIM	43	16	1609	160903
10	2022	8	2022-08	Porto	Póvoa de Varzim	União das freguesias de Aguçadoura e Navais	NÃO	2390	13	1313	131314
11	2022	10	2022-10	Vila Real	Valpaços	Santa Valha	NÃO	239	17	1712	171220
12	2022	10	2022-10	Bragança	Mirandela	União das freguesias de Avantos e Romeu	NÃO	184	4	407	040738
13	2022	2	2022-02	Bragança	Mirandela	Vale de Asnes	SIM	68	4	407	040731
14	2022	7	2022-07	Castelo Branco	Castelo Branco	Santo André das Tojeiras	SIM	201	5	502	050221
15	2022	7	2022-07	Viana do Castelo	Caminha	União das freguesias de Venade e Azevedo	SIM	172	16	1602	160225
16	2023	6	2023-06	Braga	Vila Nova de Famalicão	União das freguesias de Lemenhe, Mouquim e Jesufrei	SIM	10	3	312	031256
17	2022	5	2022-05	Viana do Castelo	Monção	Trute	SIM	44	16	1604	160431
18	2022	3	2022-03	Guarda	Guarda	Pêra do Moço	NÃO	399	9	907	090730
19	2022	3	2022-03	Santarém	Cartaxo	Valada	SIM	175	14	1406	140605
20	2023	1	2023-01	Santarém	Abrantes	Rio de Moinhos	SIM	167	14	1401	140108
21	2023	7	2023-07	Santarém	Rio Maior	São Sebastião	SIM	117	14	1414	141411
22	2023	4	2023-04	Aveiro	Sever do Vouga	Pessegueiro do Vouga	SIM	919	1	117	011704
23	2023	4	2023-04	Coimbra	Arganil	Plódão	SIM	177	6	601	060111
24	2022	2	2022-02	Aveiro	Arouca	União das freguesias de Covelo de Paivó e Janarde	NÃO	138	1	104	010424
25	2022	12	2022-12	Évora	Mourão	Mourão	NÃO	232	7	708	070803
26	2022	10	2022-10	Bragança	Macedo de Cavaleiros	Lamalonga	NÃO	238	4	405	040517
27	2023	2	2023-02	Braga	Vieira do Minho	Eira Vedra	NÃO	216	3	311	031107
28	2022	7	2022-07	Guarda	Guarda	Pega	NÃO	75	9	907	090729
29	2023	8	2023-08	Bragança	Vinhais	União das freguesias de Sobreiro de Baixo e Alvaredos	NÃO	59	4	412	041240
30	2022	7	2022-07	Guarda	Pinhel	Manigoto	NÃO	126	9	910	091014
31	2023	6	2023-06	Leiria	Porto de Mós	Porto de Mós - São João Baptista e São Pedro	SIM	3815	10	1016	101614
32	2023	5	2023-05	Leiria	Porto de Mós	São Bento	SIM	272	10	1016	101610
33	2022	5	2022-05	Castelo Branco	Idanha-a-Nova	Proença-a-Velha	NÃO	266	5	505	050511
34	2022	5	2022-05	Viseu	São Pedro do Sul	Sul	NÃO	790	18	1816	181616

Close Stop Get more rows

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Entire Transformation



CSV File input 2 (dataset B) Configuration Window

**CSV file input**

Step name: dataset B

Filename: /home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	String		7		\$	.	,	none
4 District	String		16		\$	.	,	none
5 Municipality	String		27		\$	.	,	none
6 parish	String		30		\$	.	,	none
7 Voltage level	String		32		\$	.	,	none
8 Active Energy (kWh)	Number	#####.###	12	3	\$	.	,	none
9 DistrictCode	Integer	#	15	0	\$	.	,	none
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	,	none
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

CSV File input 2 (dataset B) Preview

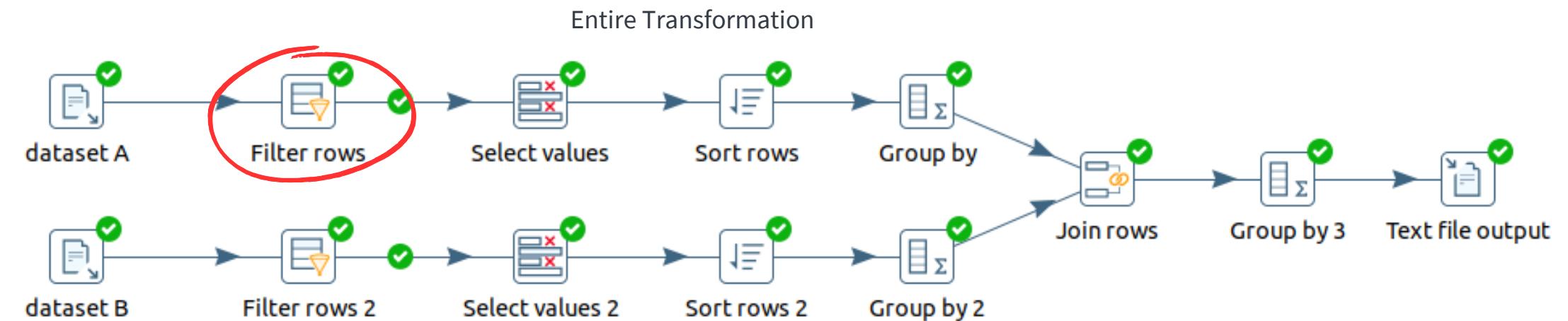
Rows of step: dataset B (10000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRA	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133

Examine preview data

Close Stop Get more rows

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Filter rows Configuration Window

**Filter rows**

Step name: Filter rows

Send 'true' data to step: Select values

Send 'false' data to step:

The condition:

- Year = [2023]
- AND
- Month = [6]
- AND
- Date = [2023-06]

**OK** **Cancel**

Filter rows Preview

Examine preview data

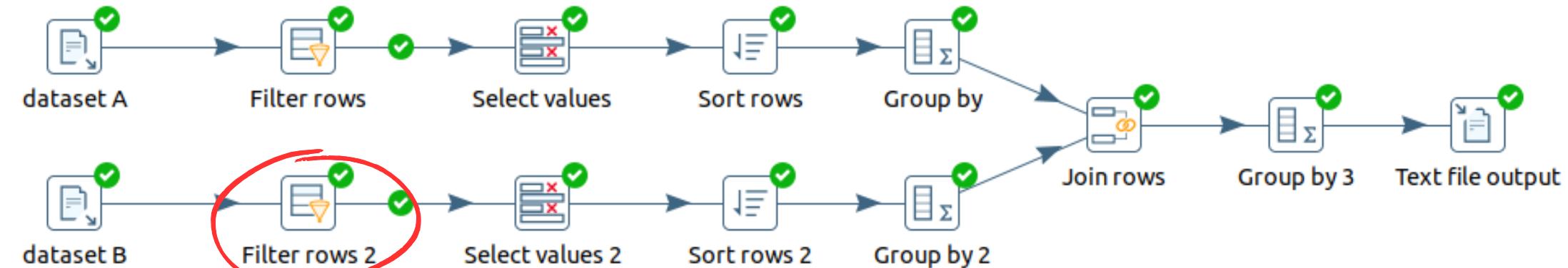
Rows of step: Filter rows (5759 rows)

#	Year	Month	Date	District	Municipality	Parish	incl. Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2023	6	2023-06	Braga	Vila Nova de Famalicão	União das freguesias de Lemenhe, Mouquim e Jesufréi	SIM	10	3	312	031256
2	2023	6	2023-06	Leiria	Porto de Mós	Porto de Mós - São João Baptista e São Pedro	SIM	3815	10	1016	101614
3	2023	6	2023-06	Viseu	Sátão	São Miguel de Vila Boa	SIM	366	18	1817	181709
4	2023	6	2023-06	Viseu	São João da Pesqueira	Ervedosa do Douro	NÃO	134	18	1815	181502
5	2023	6	2023-06	Santarém	Coruche	Couço	NÃO	1380	14	1409	140902
6	2023	6	2023-06	Braga	Barcelos	União das freguesias de Creixomil e Mariz	SIM	287	3	302	030297
7	2023	6	2023-06	Porto	Maia	Vila Nova da Telha	SIM	1879	13	1306	130616
8	2023	6	2023-06	Braga	Celorico de Basto	União das freguesias de Caçarilhe e Infesta	NÃO	92	3	305	030524
9	2023	6	2023-06	Lisboa	Cadaval	Alguber	SIM	190	11	1104	110401
10	2023	6	2023-06	Faro	São Brás de Alportel	São Brás de Alportel	NÃO	2513	8	812	081201
11	2023	6	2023-06	Bragança	Vila Flor	Rios	NÃO	12	4	410	041009
12	2023	6	2023-06	Viana do Castelo	Viana do Castelo	Perre	NÃO	118	16	1609	160926
13	2023	6	2023-06	Guarda	Guarda	Arrifana	NÃO	53	9	907	090706
14	2023	6	2023-06	Bragança	Bragança	Serapicos	NÃO	154	4	402	040247
15	2023	6	2023-06	Portalegre	Crato	Gáfete	NÃO	34	12	1206	120604
16	2023	6	2023-06	Braga	Guimarães	Lordelo	SIM	760	3	308	030828
17	2023	6	2023-06	Viseu	Tondela	Ferreirós do Dão	NÃO	55	18	1821	182107
18	2023	6	2023-06	Setúbal	Grândola	Carvalhal	SIM	3027	15	1505	150505
19	2023	6	2023-06	Coimbra	Cantanhede	Ourentã	NÃO	241	6	602	060209
20	2023	6	2023-06	Viana do Castelo	Caminha	Lanhelas	NÃO	88	16	1602	160211
21	2023	6	2023-06	Castelo Branco	Sertã	União das freguesias de Ermida e Figueiredo	SIM	127	5	509	050917
22	2023	6	2023-06	Porto	Vila do Conde	Modivas	NÃO	363	13	1316	131617
23	2023	6	2023-06	Lisboa	Cadaval	Peral	SIM	297	11	1104	110407
24	2023	6	2023-06	Santarém	Benavente	Santo Estêvão	SIM	629	14	1405	140503
25	2023	6	2023-06	Viana do Castelo	Paredes de Coura	Romarigões	SIM	66	16	1605	160519
26	2023	6	2023-06	Setúbal	Moita	Moita	SIM	10262	15	1506	150603
27	2023	6	2023-06	Guarda	Guarda	Pega	NÃO	57	9	907	090729
28	2023	6	2023-06	Beja	Ourique	União das freguesias de Garvão e Santa Luzia	SIM	211	2	212	021207
29	2023	6	2023-06	Bragança	Miranda do Douro	São Martinho de Angueira	NÃO	86	4	406	040613
30	2023	6	2023-06	Santarém	Cartaxo	União das freguesias do Cartaxo e Vale da Pinta	NÃO	711	14	1406	140609
31	2023	6	2023-06	Viana do Castelo	Valença	União das freguesias de Valença, Cristelo Covo e Arão	NÃO	763	16	1608	160820
32	2023	6	2023-06	Porto	Felgueiras	Refontoura	NÃO	475	13	1303	130317
33	2023	6	2023-06	Viseu	Castro daire	Cabril	SIM	85	18	1803	180303
34	2023	6	2023-06	Braga	Vila Verde	Lanhais	SIM	156	3	313	031324
35	2023	6	2023-06	Aveiro	Espinho	Espinho	NÃO	509	1	107	010702

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Entire Transformation



Filter rows 2 Configuration Window

**Filter rows**

Step name: **Filter rows 2**

Send 'true' data to step: **Select values 2**

Send 'false' data to step:

The condition:

Year = [2023]

AND

Month = [6]

AND

Date = [2023-06]

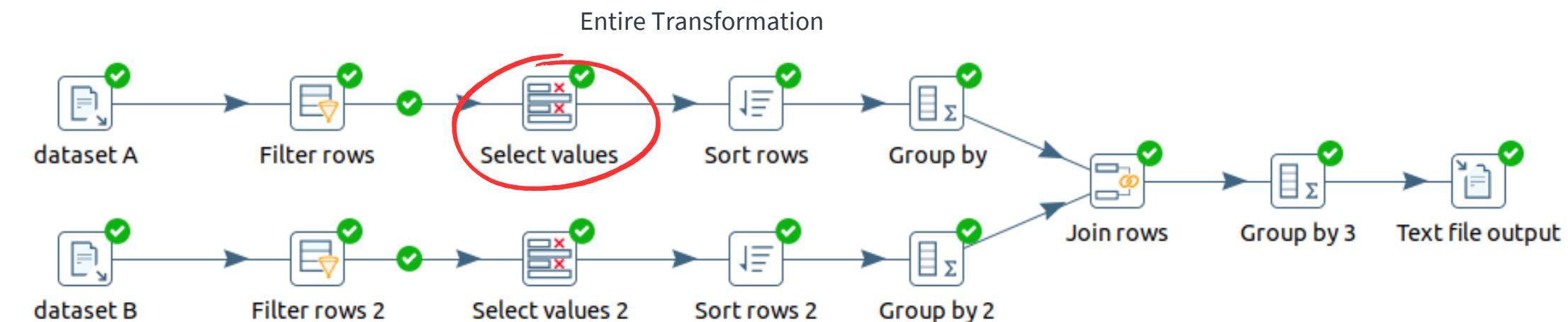
**OK**   **Cancel**

Filter rows 2 Preview

Rows of step: Filter rows 2 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2023	6	2023-06	BRAGA	Fafe	UF CEPAES E FAREJA	Baixa Tensão	260821.127	3	307	030741
2	2023	6	2023-06	BRAGA	Vieira do Minho	TABUACAS	Baixa Tensão	141120.732	3	311	031118
3	2023	6	2023-06	BRAGANCA	Bragança	UF CASTRELOS E CARRAZEDO	Baixa Tensão	26446.06	4	402	040251
4	2023	6	2023-06	VILA REAL	Ribeira de Pena	SANTA MARINHA	Baixa Tensão	62337.971	17	1709	170906
5	2023	6	2023-06	BEJA	Almodôvar	ALDEIA DOS FERNANDES	Baixa Tensão	72344.789	2	202	020208
6	2023	6	2023-06	BEJA	Moura	UF SAFARA S ALEIXO RESTAURACAO	Muito Alta, Alta e Média Tensões	24839.872	2	210	021010
7	2023	6	2023-06	BRAGA	Barcelos	FRAGOSO	Baixa Tensão	245897.125	3	302	030235
8	2023	6	2023-06	BRAGA	Braga	PRISCOS	Baixa Tensão	190551.117	3	303	030336
9	2023	6	2023-06	BRAGA	Celorico de Basto	REGO	Muito Alta, Alta e Média Tensões	386792.869	3	305	030517
10	2023	6	2023-06	CASTELO BRANCO	Castelo Branco	SANTO ANDRE DAS TOJEIRAS	Baixa Tensão	100302.887	5	502	050221
11	2023	6	2023-06	CASTELO BRANCO	Covilhã	PAUL	Baixa Tensão	204203.678	5	503	050314
12	2023	6	2023-06	LEIRIA	Óbidos	OLHO MARINHO	Muito Alta, Alta e Média Tensões	58738	10	1012	101203
13	2023	6	2023-06	PORTALEGRE	Sousel	CANO	Baixa Tensão	302824.501	12	1215	121501
14	2023	6	2023-06	PORTO	Lousada	UF FIGUEIRAS E COVAS	Muito Alta, Alta e Média Tensões	83619.162	13	1305	130529
15	2023	6	2023-06	VIANA DO CASTELO	Ponte da Barca	SAMPRIZ	Baixa Tensão	24810.18	16	1606	160619
16	2023	6	2023-06	VIANA DO CASTELO	Ponte da Barca	VADE (SAO PEDRO)	Baixa Tensão	30428.374	16	1606	160623
17	2023	6	2023-06	VIDEU	Armamar	SANTA CRUZ	Muito Alta, Alta e Média Tensões	50214.065	18	1801	180111
18	2023	6	2023-06	CASTELO BRANCO	Proenca-a-Nova	MONTES DA SENHORA	Baixa Tensão	116234.296	5	508	050802
19	2023	6	2023-06	LEIRIA	Óbidos	S MARIA S PEDRO E SOBRAL LAGOA	Muito Alta, Alta e Média Tensões	414970	10	1012	101210
20	2023	6	2023-06	LISBOA	Cascais	UF CASCAIS E ESTORIL	Baixa Tensão	16141248.413	11	1105	110508
21	2023	6	2023-06	PORTO	Gondomar	UF MELRES E MEDAS	Baixa Tensão	703621.47	13	1304	130416
22	2023	6	2023-06	BRAGA	Vieira do Minho	UF CANICADA E SOENGAS	Baixa Tensão	100579.504	3	311	031124
23	2023	6	2023-06	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão	27060.313	9	907	090718
24	2023	6	2023-06	SANTAREM	Santarém	UF AZOIA DE CIMA E TREMES	Baixa Tensão	376478.883	14	1416	141630
25	2023	6	2023-06	VILA REAL	Chaves	REDONDELO	Muito Alta, Alta e Média Tensões	19065.52	17	1703	170324
26	2023	6	2023-06	VILA REAL	Vila Pouca de Aguiar	ALVAO	Baixa Tensão	94804.088	17	1713	171319
27	2023	6	2023-06	VIDEU	Viseu	BODIOSA	Baixa Tensão	355461.733	18	1823	182304
28	2023	6	2023-06	BRAGA	Guimarães	PINHEIRO	Baixa Tensão	165941.058	3	308	030836
29	2023	6	2023-06	BRAGA	Póvoa de Lanhoso	FERREIROS	Baixa Tensão	44778.917	3	309	030908
30	2023	6	2023-06	BRAGANCA	Mirandela	COBRO	Baixa Tensão	15859.244	4	407	040713
31	2023	6	2023-06	CASTELO BRANCO	Belmonte	INGUIAS	Baixa Tensão	80495.606	5	501	050104
32	2023	6	2023-06	EVORA	Reguengos de Monsaraz	MONSARAZ	Muito Alta, Alta e Média Tensões	112681.74	7	711	071103
33	2023	6	2023-06	GUARDA	Guarda	CASTANHEIRA	Baixa Tensão	37087.186	9	907	090712
34	2023	6	2023-06	LISBOA	Odivelas	UF PONTINHA E FAMOES	Muito Alta, Alta e Média Tensões	528311.858	11	1116	111608

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Select values Configuration Window

Select values

Step name

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 Municipality			
2 Number of CPE's			

Include unspecified fields, ordered

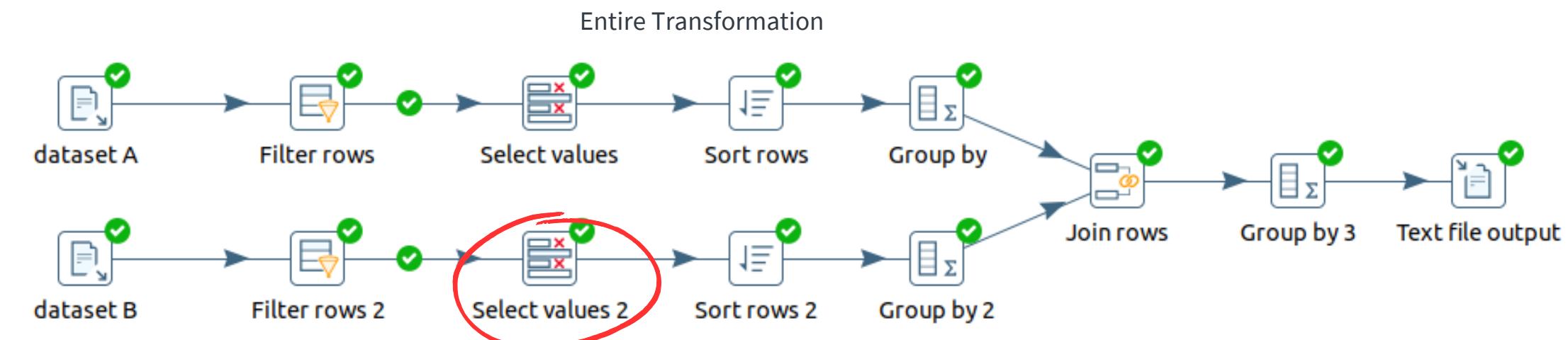
Select values Preview

Examine preview data

Rows of step: Select values (5759 rows)

Municipality	Number of CPE's
Vila Nova de Famalicão	10
Porto de Mós	3815
Sátão	366
São João da Pesqueira	134
Coruche	1380
Barcelos	287
Maia	1879
Celorico de Basto	92
Cadaval	190
São Brás de Alportel	2513
Vila Flor	12
Viana do Castelo	118
Guarda	53
Bragança	154
Crato	34
Guimarães	760
Tondela	55
Grândola	3027
Cantanhede	241
Caminha	88
Sertã	127
Vila do Conde	363
Cadaval	297
Benavente	629
Paredes de Coura	66
Moita	10262
Guarda	57
Ouriue	211
Miranda do douro	86
Cartaxo	711
Valenca	763
Felgueiras	475
Castro daire	85
Vila Verde	156
Espinho	509

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Select values 2 Configuration Window

Select values

Step name **Select values 2**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 Municipality	Municipality_2		
2 Active Energy (kwh)			

Get fields to select Edit Mapping

Include unspecified fields, ordered

OK Cancel

Help

Select values 2 Preview

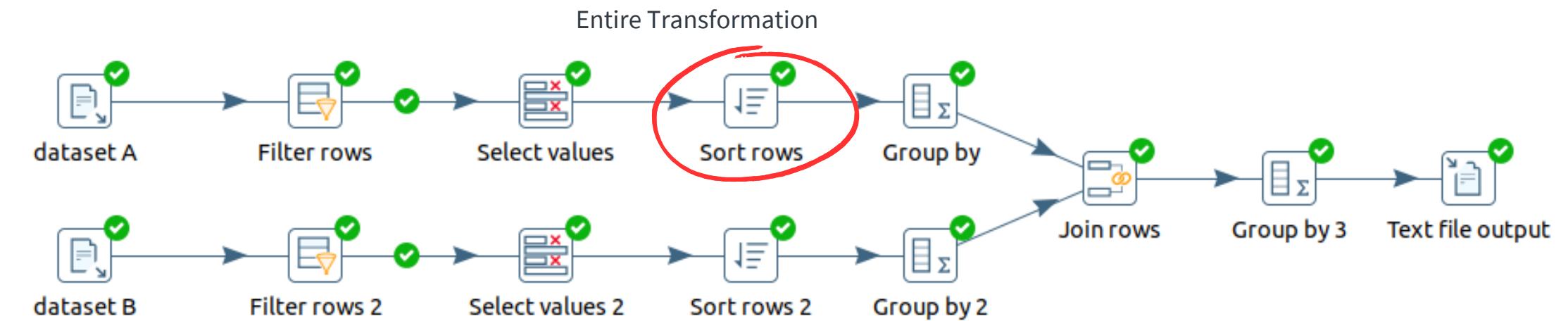
Rows of step: Select values 2 (1000 rows)

Examine preview data

Municipality_2	Active Energy (kwh)
1 Fafe	260821.127
2 Vieira do Minho	141120.732
3 Bragança	26446.06
4 Ribeira de Pena	62337.971
5 Almodôvar	72344.789
6 Moura	24839.872
7 Barcelos	245897.125
8 Braga	190551.117
9 Celorico de Basto	386792.869
10 Castelo Branco	100302.887
11 Covilhã	204203.678
12 Óbidos	58738
13 Sousel	302824.501
14 Lousada	83619.162
15 Ponte da Barca	24810.18
16 Ponte da Barca	30428.374
17 Armamar	50214.065
18 Proença-a-Nova	116234.296
19 Óbidos	414970
20 Cascais	16141248.413
21 Gondomar	703621.47
22 Vieira do Minho	100579.504
23 Guarda	27060.313
24 Santarém	376478.883
25 Chaves	19065.52
26 Vila Pouca de Aguiar	94804.088
27 Viseu	355461.733
28 Guimarães	165941.058
29 Póvoa de Lanhoso	44778.917
30 Mirandela	15859.244
31 Belmonte	80495.606
32 Reguengos de Monsaraz	112681.74
33 Guarda	37087.186
34 Odivelas	528311.858

Close Stop Get more rows

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Sort rows Configuration Window

Sort rows

Step name: Sort rows

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

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 Municipality	Y	N	N	0	N

OK Cancel Get Fields

Sort rows Preview

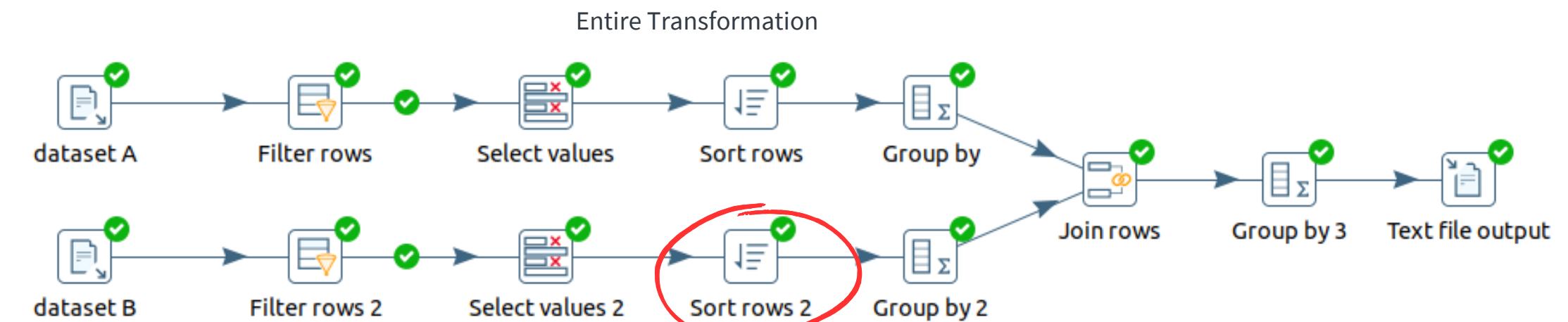
Rows of step: Sort rows (5759 rows)

Municipality	Number of CPE's
Abrantes	751
Abrantes	348
Abrantes	380
Abrantes	1391
Abrantes	660
Abrantes	9840
Abrantes	903
Abrantes	1437
Abrantes	1598
Abrantes	263
Abrantes	260
Abrantes	616
Abrantes	743
Abrantes	167
Abrantes	471
Abrantes	378
Abrantes	504
Abrantes	310
Abrantes	725
Abrantes	934
Abrantes	307
Abrantes	517
Abrantes	714
Abrantes	220
Abrantes	243
Abrantes	500
Abrantes	40
Abrantes	158
Abrantes	139
Abrantes	167
Abrantes	300
Abrantes	19
Abrantes	337
Abrantes	317
Abrantes	337

Examine preview data

Close

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Sort rows 2 Configuration Window

Sort rows

Step name: Sort rows 2

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

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	Presort
Municipality_2	Y	N	N	0	N

OK Cancel Get Fields

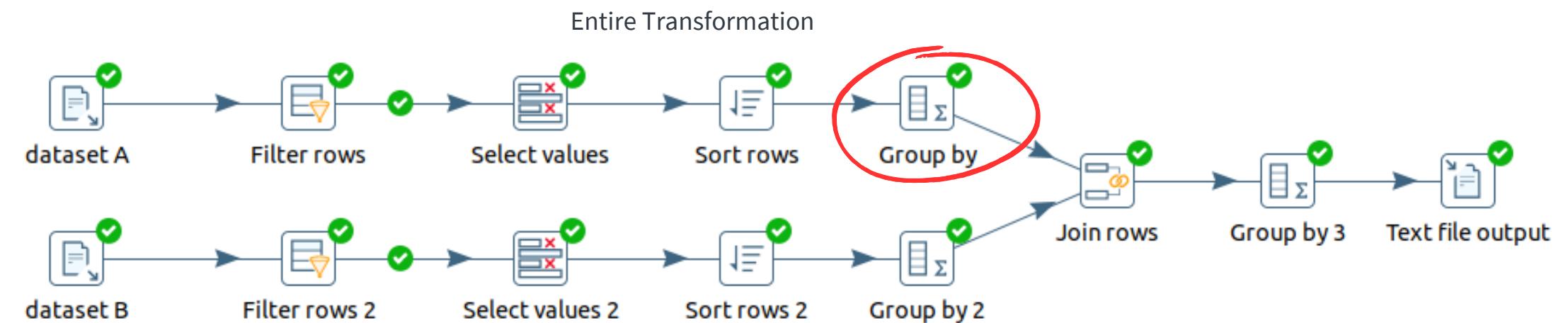
Sort rows 2 Preview

Examine preview data

Municipality_2	Active Energy (kWh)
Abrantes	2589572.393
Abrantes	609789.647
Abrantes	192437.608
Abrantes	150692.446
Abrantes	250548.476
Abrantes	563054.365
Abrantes	156245.515
Abrantes	38562.774
Abrantes	177322.669
Abrantes	168351.892
Abrantes	81690.02
Abrantes	666703.809
Abrantes	534904.641
Abrantes	2429622.729
Abrantes	288245.288
Abrantes	257655.992
Abrantes	461791.879
Abrantes	85106.295
Abrantes	79647.267
Abrantes	145162.008
Abrantes	640829.128
Abrantes	29078.754
Aguiar da Beira	30596.258
Aguiar da Beira	390008.292
Aguiar da Beira	40977.393
Aguiar da Beira	78953.736
Aguiar da Beira	100316.568
Aguiar da Beira	171164.614
Aguiar da Beira	42543.906
Aguiar da Beira	83922.628
Aguiar da Beira	99840.593
Aguiar da Beira	68207.182
Alandroal	355948.552
Alandroal	63751.087

Close Stop Get more rows

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Group by Configuration Window

**Group by**

Step name **Group by**

Include all rows?

Temporary files directory `%java.io.tmpdir%`

TMP-file prefix `grp`

Add line number, restart in each group

Line number field name

Always give back a result row

The fields that make up the group:

**Group field**

1   Municipality	<input type="button" value="Get Fields"/>
------------------	---

Aggregates:

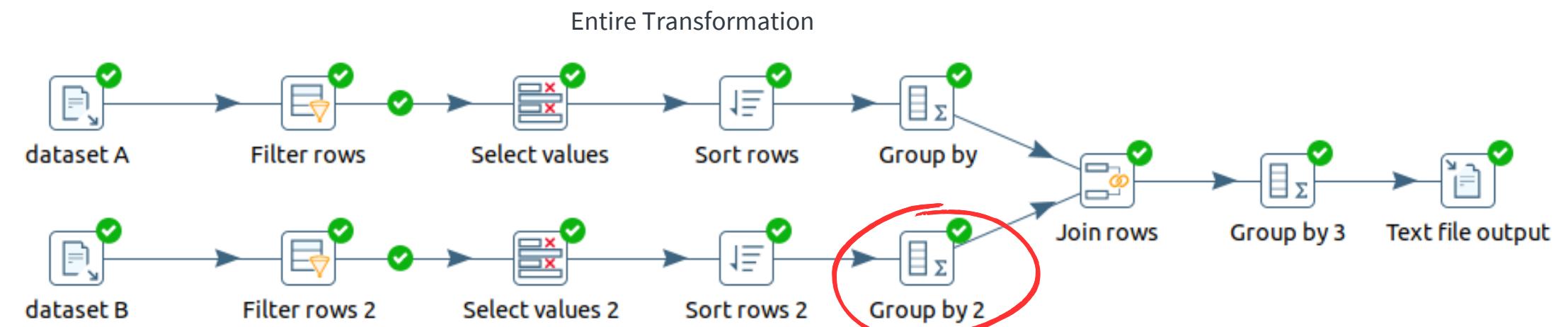
Name	Subject	Type	Value	<input type="button" value="Get lookup fields"/>
1   CPEs	Number of CPE's	Sum		

Group by Preview

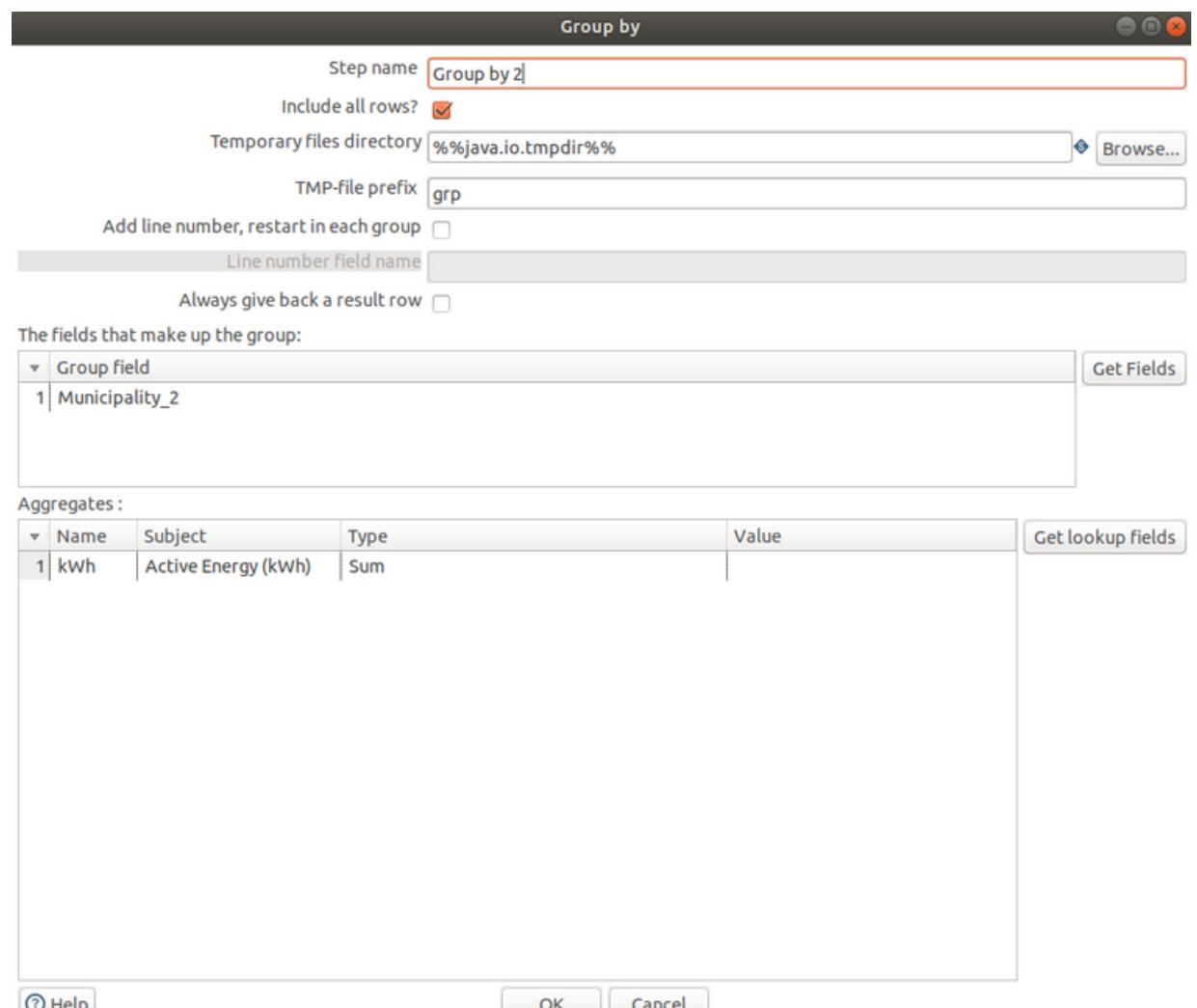
Rows of step: Group by (278 rows)

Municipality	CPEs
1 Abrantes	25180
2 Aguiar da Beira	4933
3 Alandroal	4133
4 Albergaria-a-Velha	14241
5 Albufeira	54157
6 Alcanena	8265
7 Alcobaça	39245
8 Alcochete	11819
9 Alcoutim	3703
10 Alcácer do Sal	8480
11 Alenquer	24858
12 Alfândega da Fé	4167
13 Alijó	9146
14 Aljezur	6523
15 Aljustrel	6058
16 Almada	116447
17 Almeida	7014
18 Almeirim	14117
19 Almodôvar	5693
20 Alpiarça	4630
21 Alter do Chão	2892
22 Alvaiázere	5897
23 Alvito	1725
24 Amadora	98733
25 Amarante	31797
26 Amares	11329
27 Anadia	17295
28 Ansião	9048
29 Arcos de Valdevez	19112
30 Arganil	11140
31 Armamar	5147
32 Arouca	12670
33 Arraiolos	4874
34 Arronches	2319
35 Arruda dos Vinhos	7678

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Group by 2 Configuration Window



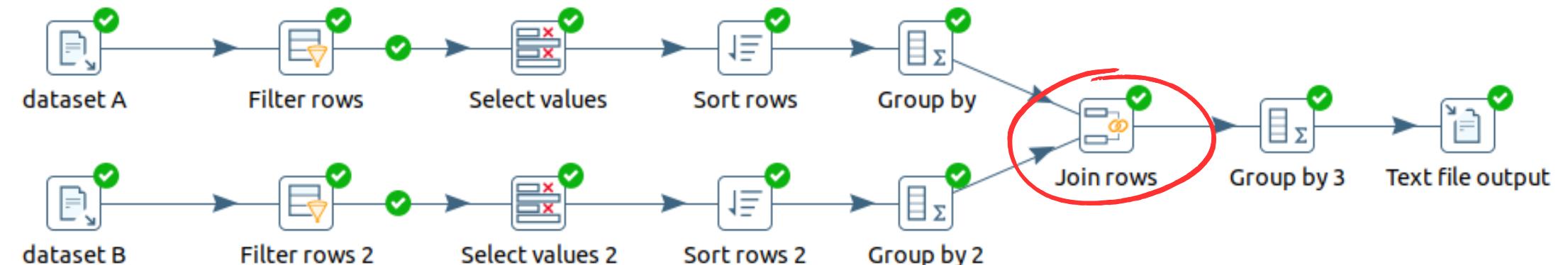
Group by 2 Preview

Rows of step: Group by 2 (1000 rows)		Examine preview data
Municipality_2	Active Energy (kwh)	kwh
1 Abrantes	2589572.393	10567936.841
2 Abrantes	609789.647	10567936.841
3 Abrantes	192437.608	10567936.841
4 Abrantes	150692.446	10567936.841
5 Abrantes	250548.476	10567936.841
6 Abrantes	563054.365	10567936.841
7 Abrantes	156245.515	10567936.841
8 Abrantes	38562.774	10567936.841
9 Abrantes	177322.669	10567936.841
10 Abrantes	168351.892	10567936.841
11 Abrantes	81690.02	10567936.841
12 Abrantes	666703.809	10567936.841
13 Abrantes	534904.641	10567936.841
14 Abrantes	2429622.729	10567936.841
15 Abrantes	288245.288	10567936.841
16 Abrantes	257655.992	10567936.841
17 Abrantes	461791.879	10567936.841
18 Abrantes	85106.295	10567936.841
19 Abrantes	79647.267	10567936.841
20 Abrantes	145162.008	10567936.841
21 Abrantes	640829.128	10567936.841
22 Aguiar da Beira	29078.754	1135609.924
23 Aguiar da Beira	30596.258	1135609.924
24 Aguiar da Beira	390008.292	1135609.924
25 Aguiar da Beira	40977.393	1135609.924
26 Aguiar da Beira	78953.736	1135609.924
27 Aguiar da Beira	100316.568	1135609.924
28 Aguiar da Beira	171164.614	1135609.924
29 Aguiar da Beira	42543.906	1135609.924
30 Aguiar da Beira	83922.628	1135609.924
31 Aguiar da Beira	99840.593	1135609.924
32 Aguiar da Beira	68207.182	1135609.924
33 Alandroal	355948.552	1838245.246
34 Alandroal	63751.087	1838245.246

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Entire Transformation



Join rows Configuration Window

Join rows (cartesian product)

Step name:

Temp directory:

TMP-file prefix:

Max. cache size (in rows):

Main step to read from:

The condition:

Municipality = Municipality\_2

Join rows Preview

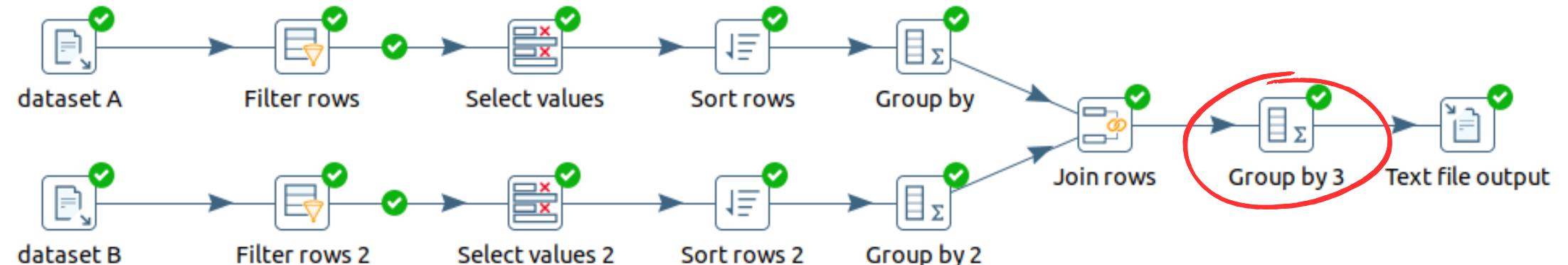
Rows of step: Join rows (1000 rows)

	Municipality_2	Active Energy (kWh)	kWh	Municipality	CPEs
1	Abrantes	2589572.393	10567936.841	Abrantes	25180
2	Abrantes	609789.647	10567936.841	Abrantes	25180
3	Abrantes	192437.608	10567936.841	Abrantes	25180
4	Abrantes	150692.446	10567936.841	Abrantes	25180
5	Abrantes	250548.476	10567936.841	Abrantes	25180
6	Abrantes	563054.365	10567936.841	Abrantes	25180
7	Abrantes	156245.515	10567936.841	Abrantes	25180
8	Abrantes	38562.774	10567936.841	Abrantes	25180
9	Abrantes	177322.669	10567936.841	Abrantes	25180
10	Abrantes	168351.892	10567936.841	Abrantes	25180
11	Abrantes	81690.02	10567936.841	Abrantes	25180
12	Abrantes	666703.809	10567936.841	Abrantes	25180
13	Abrantes	534904.641	10567936.841	Abrantes	25180
14	Abrantes	2429622.729	10567936.841	Abrantes	25180
15	Abrantes	288245.288	10567936.841	Abrantes	25180
16	Abrantes	257655.992	10567936.841	Abrantes	25180
17	Abrantes	461791.879	10567936.841	Abrantes	25180
18	Abrantes	85106.295	10567936.841	Abrantes	25180
19	Abrantes	79647.267	10567936.841	Abrantes	25180
20	Abrantes	145162.008	10567936.841	Abrantes	25180
21	Abrantes	640829.128	10567936.841	Abrantes	25180
22	Aguiar da Beira	29078.754	1135609.924	Aguiar da Beira	4933
23	Aguiar da Beira	30596.258	1135609.924	Aguiar da Beira	4933
24	Aguiar da Beira	390008.292	1135609.924	Aguiar da Beira	4933
25	Aguiar da Beira	40977.393	1135609.924	Aguiar da Beira	4933
26	Aguiar da Beira	78953.736	1135609.924	Aguiar da Beira	4933
27	Aguiar da Beira	100316.568	1135609.924	Aguiar da Beira	4933
28	Aguiar da Beira	171164.614	1135609.924	Aguiar da Beira	4933
29	Aguiar da Beira	42543.906	1135609.924	Aguiar da Beira	4933
30	Aguiar da Beira	83922.628	1135609.924	Aguiar da Beira	4933
31	Aguiar da Beira	99840.593	1135609.924	Aguiar da Beira	4933
32	Aguiar da Beira	68207.182	1135609.924	Aguiar da Beira	4933
33	Alandroal	355948.552	1838245.246	Alandroal	4133
34	Alandroal	63751.087	1838245.246	Alandroal	4133

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Entire Transformation



Group by 3 Configuration Window

Group by

Step name **Group by 3**

Include all rows?

Temporary files directory `%%java.io.tmpdir%%`

TMP-file prefix `grp`

Add line number, restart in each group

Line number field name

Always give back a result row

The fields that make up the group:

Group field

1 Municipality  
2 CPEs  
3 kWh

Aggregates:

Name	Subject	Type	Value
1			

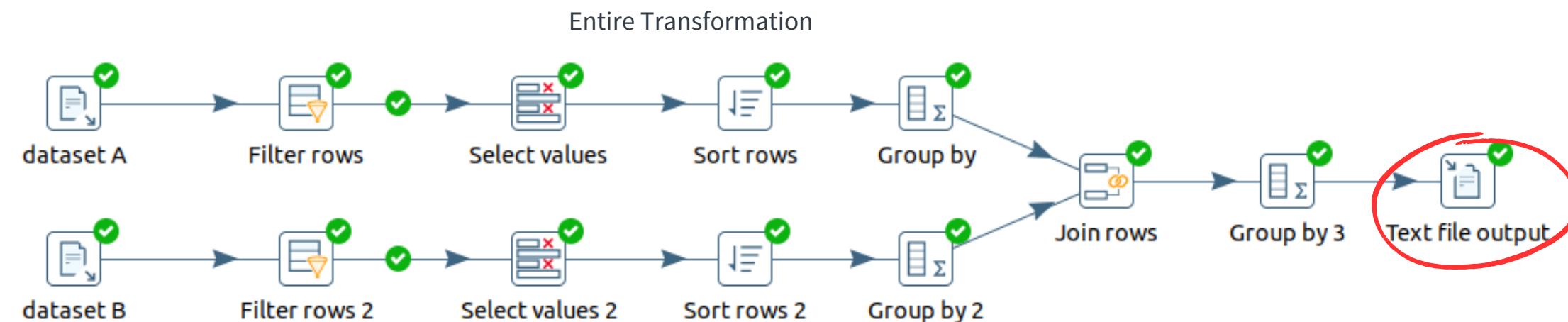
Group by 3 Preview

Rows of step: Group by 3 (278 rows)

Examine preview data

Municipality	CPEs	kWh
1 Abrantes	25180	10567936.841
2 Aguiar da Beira	4933	1135609.924
3 Aaldoal	4133	1838245.246
4 Albergaria-a-Velha	14241	14765120.434
5 Albufeira	54157	33437993.117
6 Alcanena	8265	5643207.701
7 Alcobaça	39245	19699225.738
8 Alcochete	11819	6670715.332
9 Alcoutim	3703	611524.557
10 Alcácer do Sal	8480	6333071.731
11 Alenquer	24858	18889521.167
12 Alfândega da Fé	4167	758239.718
13 Alijo	9146	2051978.29
14 Aljezur	6523	1880390.444
15 Aljustrel	6058	15121881.473
16 Almada	116447	39254590.524
17 Almeida	7014	1219835.386
18 Almeirim	14117	6655643.254
19 Almodôvar	5693	1455934.812
20 Alpiarça	4630	3980930.421
21 Alter do Chão	2892	858120.347
22 Alvaiázere	5897	1248745.845
23 Alvito	1725	1220644.494
24 Amadora	98733	38592131.582
25 Amarante	31797	9305899.397
26 Amares	11329	3512297.001
27 Anadia	17295	9902342.348
28 Ansião	9048	2611803.207
29 Arcos de Valdevez	19112	7737799.646
30 Arganil	11140	3583461.19
31 Armamar	5147	1484130.569
32 Arouca	12670	4066420.363
33 Arraiolos	4874	2092141.816
34 Aronches	2319	798947.881

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Text file output Configuration Window - File

**Text file output**

Step name **Text file output**

**File**

- Filename  **\${Internal.Entry.Current.Directory}/cpe\_kwh\_municipality**
- Pass output to servlet
- Create Parent folder
- Do not create file at start
- Accept file name from field?
- File name field
- Extension **csv**
- Include stepnr in filename?
- Include partition nr in filename?
- Include date in filename?
- Include time in filename?
- Specify Date time format
- Date time format
- Show filename(s)
- Add filenames to result

**Content**

**Fields**

OK Cancel

Text file output Configuration Window - Content

**Text file output**

Step name **Text file output**

**File**

**Content**

**Fields**

- Append
- Separator **;**
- Enclosure **"**
- Force the enclosure around fields?
- Disable the enclosure fix?
- Header
- Footer
- Format **CR+LF terminated (Windows, DOS)**
- Compression **None**
- Encoding **UTF-8**
- Right pad fields
- Fast data dump (no formatting)
- Split every ... rows
- Add Ending line of file

**OK** **Cancel**

Text file output Configuration Window - Fields

**Text file output**

Step name **Text file output**

**File**

**Content**

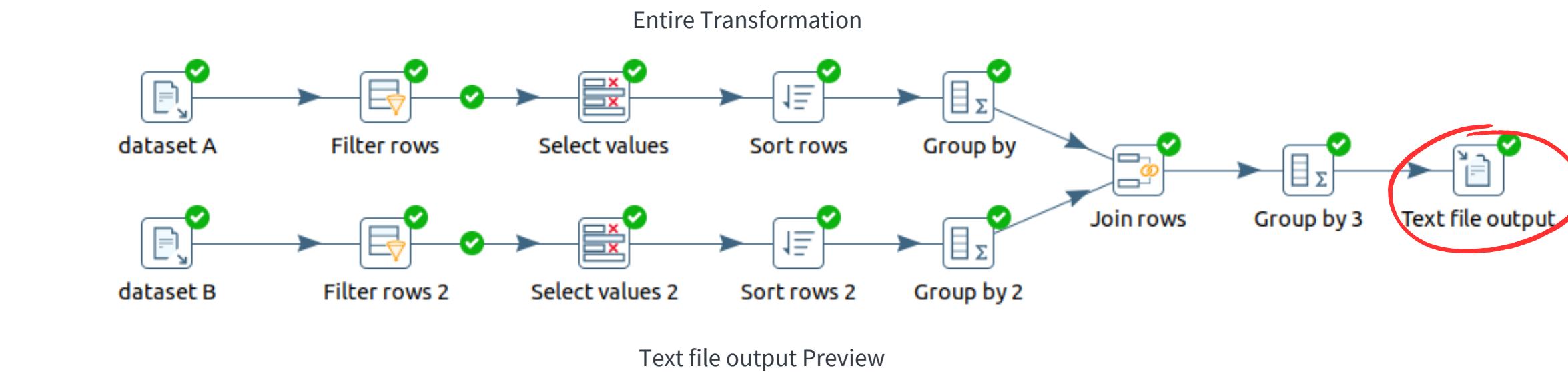
**Fields**

Name	Type	Format	Length	Precision	Currency	Deci
1 Municipality	String				.	.
2 CPEs	Integer	0				
3 kwh	Number	0.#####				

Get Fields Minimal width

**OK** **Cancel**

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Examine preview data

Municipality	CPES	kwh
1 Abrantes	25180	10567936.841
2 Aguiar da Beira	4933	1135609.924
3 Alandroal	4133	1838245.246
4 Albergaria-a-Velha	14241	14765120.434
5 Albufeira	54157	33437993.117
6 Alcanena	8265	5643207.701
7 Alcobaça	39245	19699225.738
8 Alcochete	11819	6670715.332
9 Alcoutim	3703	611524.557
10 Alcácer do Sal	8480	6333071.731
11 Alenquer	24858	18889521.167
12 Alfândega da Fé	4167	758239.718
13 Alijó	9146	2051978.29
14 Aljezur	6523	1880390.444
15 Aljustrel	6058	15121881.473
16 Almada	116447	39254590.524
17 Almeida	7014	1219835.386
18 Almeirim	14117	6655643.254
19 Almodôvar	5693	1455934.812
20 Alpiarça	4630	3980930.421
21 Alter do Chão	2892	858120.347
22 Alvaiázere	5897	1248745.845
23 Alvito	1725	1220644.494
24 Amadora	98733	38592131.582
25 Amarante	31797	9305899.397
26 Amares	11329	3512297.001
27 Anadia	17295	9902342.348
28 Ansião	9048	2611803.207
29 Arcos de Valdevez	19112	7737799.646
30 Arganil	11140	3583461.19
31 Armamar	5147	1484130.569
32 Arouca	12670	4066420.363
33 Arraiolos	4874	2092141.816
34 Arronches	2319	798947.881

# 1. NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



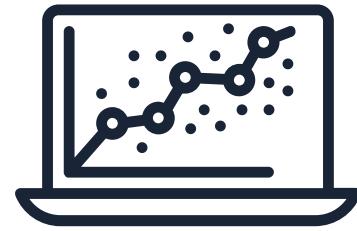
Final output file:

	A	B	C
1	Municipality	CPEs	kWh
2	Abrantes	25180	10567936.841
3	Aguiar da Beira	4933	1135609.924
4	Alandroal	4133	1838245.246
5	Albergaria-a-Velha	14241	14765120.434
6	Albufeira	54157	33437993.117
7	Alcanena	8265	5643207.701
8	Alcobaça	39245	19699225.738
9	Alcochete	11819	6670715.332
10	Alcoutim	3703	611524.557
11	Alcácer do Sal	8480	6333071.731
12	Alenquer	24858	18889521.167
13	Alfândega da Fé	4167	758239.718
14	Alijó	9146	2051978.29
15	Aljezur	6523	1880390.444
16	Aljustrel	6058	15121881.473
17	Almada	116447	39254590.524
18	Almeida	7014	1219835.386
19	Almeirim	14117	6655643.254
20	Almodôvar	5693	1455934.812
21	Alpiarça	4630	3980930.421
22	Alter do Chão	2892	858120.347
23	Alvalázere	5897	1248745.845
24	Alvito	1725	1220644.494
25	Amadora	98733	38592131.582
26	Amarante	31797	9305899.397
27	Amares	11329	3512297.001
28	Anadia	17295	9902342.348
29	Ansião	9048	2611803.207
30	Arcos de Valdevez	19112	7737799.646
31	Arganil	11140	3583461.19
32	Armamar	5147	1484130.569
33	Arouca	12670	4066420.363
34	Arraiolos	4874	2092141.816
35	Arronches	2319	798947.881
36	Arruda dos Vinhos	7678	2988192.087
37	Aveiro	51443	59927297.078
38	Avis	3182	4463850.29
39	Azambuja	12414	12774466.277
40	Baião	11943	2663459.128
41	Barcelos	57289	33547153.266
42	Barrancos	1239	401405.503
43	Barreiro	49322	16597481.08
44	Batalha	9851	6249562.411
45	Beja	23054	45508349.423
46	Belmonte	5120	1113175.786
47	Benavente	17696	12162078.692
48	Bombarral	8893	3151119.435
49	Borba	4292	1760942.903
50	Boticas	4850	1384722.293
51	Braga	106285	57788696.454

# TASK 2

SCATTER PLOT BETWEEN THE NUMBER OF  
CONTRACTS AND THE ENERGY CONSUMPTION  
PER MUNICIPALITY IN JUNE 2023

## 2. SCATTER PLOT BETWEEN THE NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



To study the correlation between the number of contracts and energy consumption, a scatter plot was made in DataCleaner.

Analysis Job

cpe\_kwh\_municipality

Scatter plot

Scatter plot Configuration Window

Scatter plot | DataCleaner

Scatter plot

Documentation Rename

Input columns

Variable1: CPEs

Variable2: kWh

Optional properties (1)

Group column: - none -

Table cpe\_kwh\_municipality Configuration Window

cpe\_kwh\_municipality | DataCleaner

cpe\_kwh\_municipality

project.cpe\_kwh\_municipality

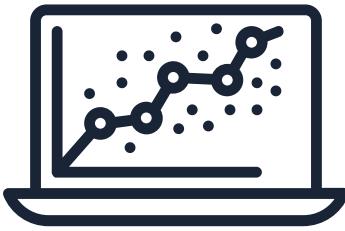
Name Type

Name	Type
Municipality	String
CPEs	Integer
kWh	Double

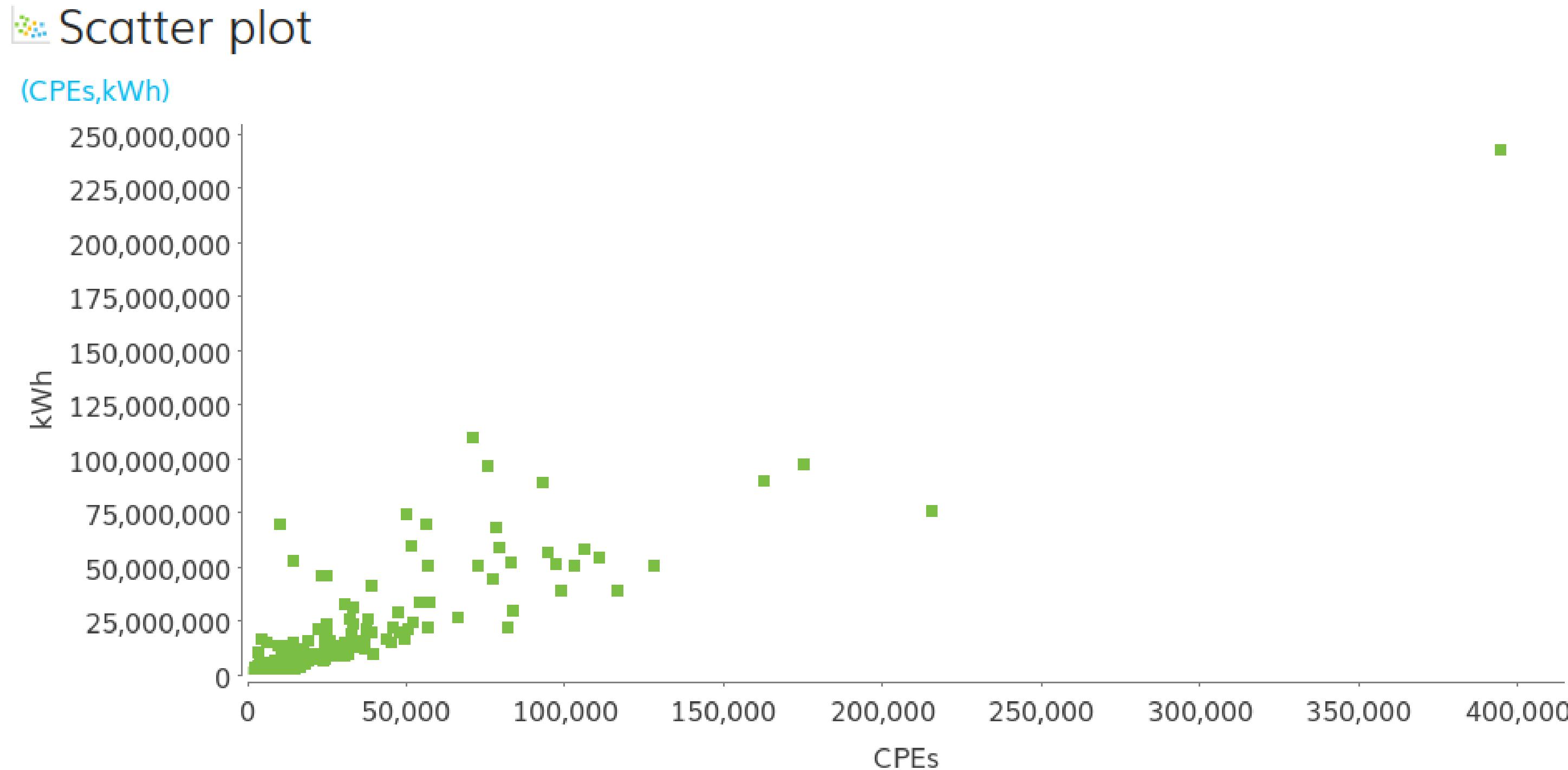
X Close

X Close

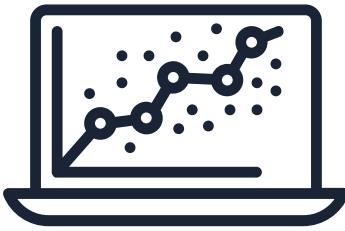
## 2. SCATTER PLOT BETWEEN THE NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



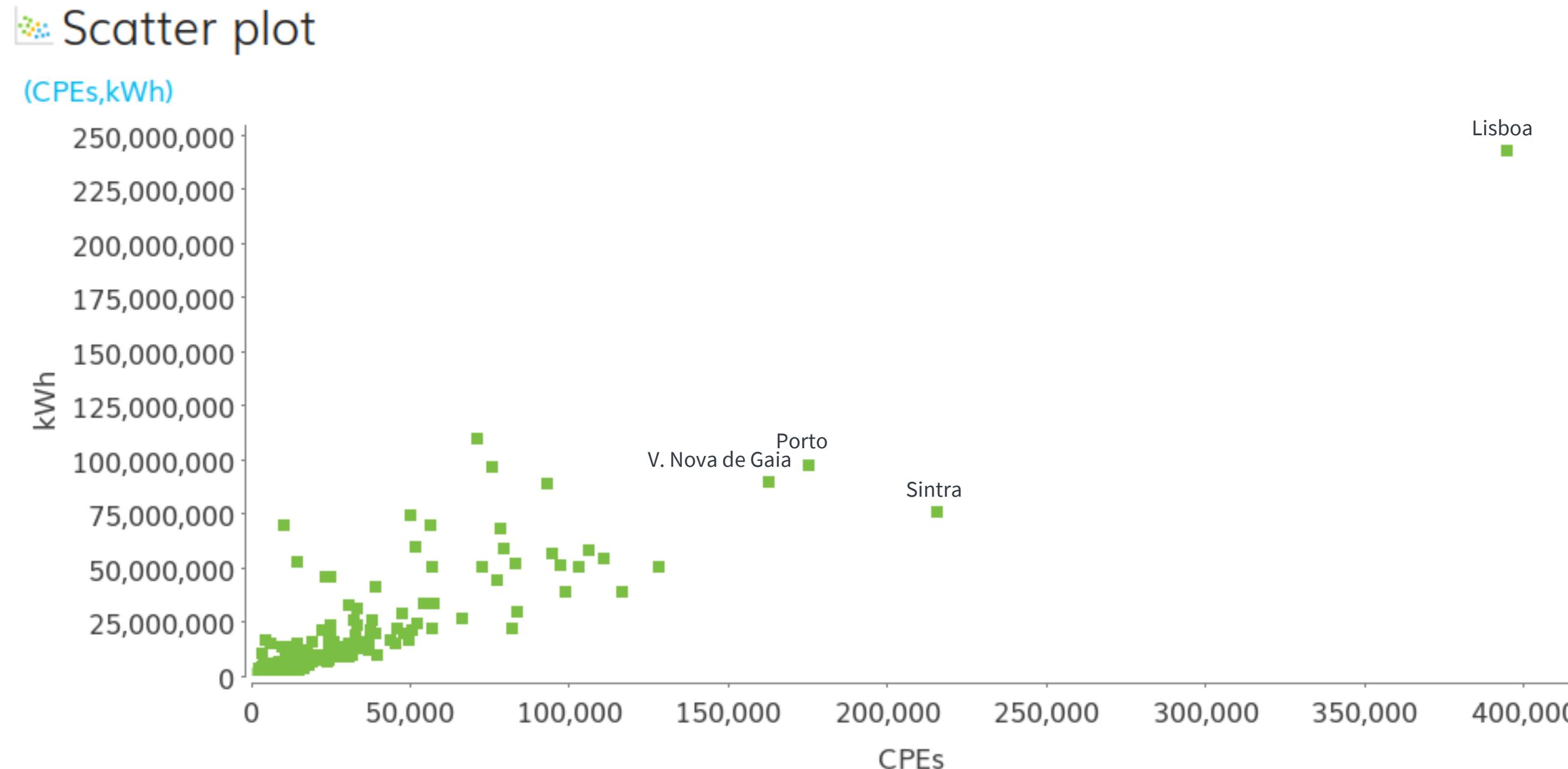
Scatter plot obtained between the number of contracts, CPE, and the energy consumption, kWh:



## 2. SCATTER PLOT BETWEEN THE NUMBER OF CONTRACTS AND ENERGY CONSUMPTION PER MUNICIPALITY IN JUNE 2023



Top-4 municipalities with the most number of contracts, CPEs, and/or consumption, kWh



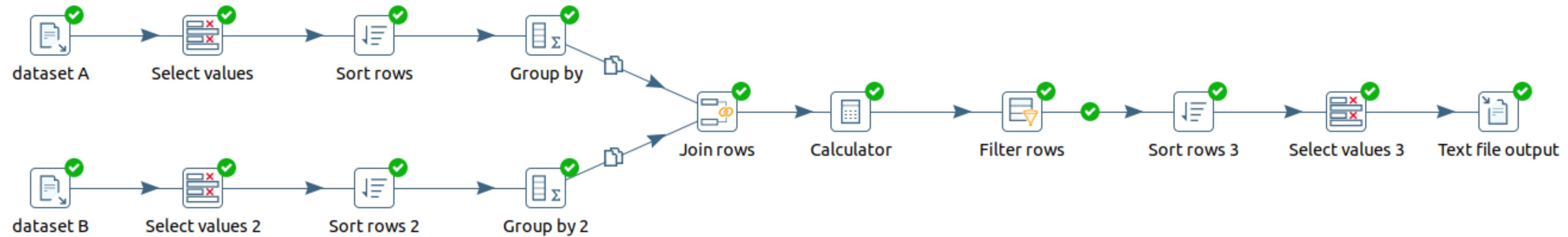
# TASK 3

CORRECT MAPPING OF DISTRICTS IN DATASET A  
TO DISTRICTS IN DATASET B

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



A duplicate detection approach using a string-matching measure and a threshold was employed in order to correctly map the districts in both datasets. The following transformation was made in Pentaho:

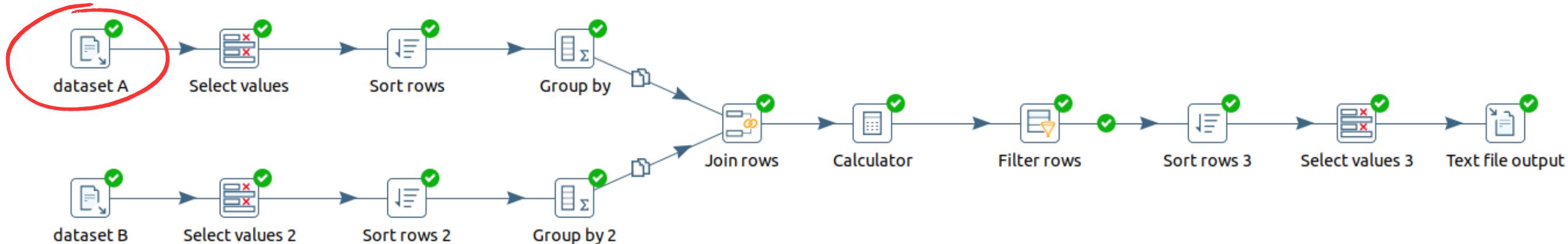


The Configuration Windows and Previews for each step, as well as the final transformation's output are shown in the next slides.

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Entire Transformation



CSV File input (dataset A) Configuration Window

**CSV File input**

Step name: dataset A

Filename: /home/aid/Desktop/21-contadores-de-energia.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	String		7		\$	.	,	none
4 District	String		16		\$	.	,	none
5 Municipality	String		27		\$	.	,	none
6 Parish	String		138		\$	.	,	none
7 incl. Smart Meter	String		3		\$	.	,	none
8 Number of CPE's	Integer	#	15	0	\$	.	,	none
9 DistrictCode	Integer	#	15	0	\$	.	,	none
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	,	none
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

CSV File input (dataset A) Preview

Rows of step: dataset A (10000 rows)

Examine preview data

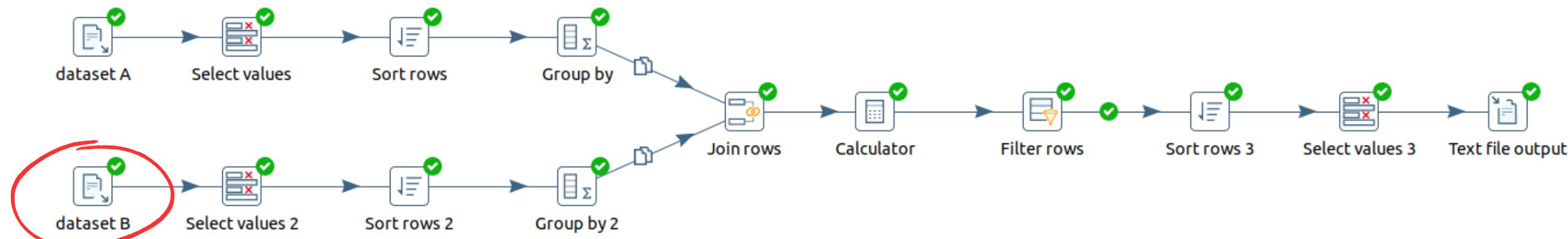
	Year	Month	Date	District	Municipality	Parish	incl. Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2023	8	2023-08	Braga	Vila Nova de Famalicão	União das freguesias de Carreira e Bente	NÃO	120	3	312	031253
2	2022	7	2022-07	Faro	Loulé (São Sebastião)	Loulé (São Sebastião)	NÃO	1714	8	808	080809
3	2022	11	2022-11	Coimbra	Condeixa-a-Nova	Ega	SIM	658	6	604	060406
4	2022	6	2022-06	Lisboa	Cadaval	União das freguesias de Painho e Figueiros	NÃO	947	11	1104	110413
5	2022	9	2022-09	Aveiro	Anadia	São Lourenço do Bairro	SIM	396	1	103	010310
6	2022	3	2022-03	Guarda	Vila Nova de Foz Côa	Vila Nova de Foz Côa	NÃO	1515	9	914	091419
7	2022	3	2022-03	Coimbra	Penacova	Sazes do Lorvão	NÃO	299	6	613	061310
8	2022	3	2022-03	Porto	Lousada	União das freguesias de Silvares, Pias, Nogueira e Alvarenga	NÃO	2445	13	1305	130532
9	2022	8	2022-08	Viana do Castelo	Viana do Castelo	Amonde	SIM	43	16	1609	160903
10	2022	8	2022-08	Porto	Póvoa de Varzim	União das freguesias de Aguçadoura e Navais	NÃO	2390	13	1313	131314
11	2022	10	2022-10	Vila Real	Valpaços	Santa Valha	NÃO	239	17	1712	171220
12	2022	10	2022-10	Bragança	Mirandela	União das freguesias de Avantos e Romeu	NÃO	184	4	407	040738
13	2022	2	2022-02	Bragança	Mirandela	Vale de Asnes	SIM	68	4	407	040731
14	2022	7	2022-07	Castelo Branco	Castelo Branco	Santo André das Tojeiras	SIM	201	5	502	050221
15	2022	7	2022-07	Viana do Castelo	Caminha	União das freguesias de Venade e Azevedo	SIM	172	16	1602	160225
16	2023	6	2023-06	Braga	Vila Nova de Famalicão	União das freguesias de Lemenhe, Mouquim e Jesufrei	SIM	10	3	312	031256
17	2022	5	2022-05	Viana do Castelo	Monção	Trute	SIM	44	16	1604	160431
18	2022	3	2022-03	Guarda	Guarda	Pêra do Moço	NÃO	399	9	907	090730
19	2022	3	2022-03	Santarém	Cartaxo	Valada	SIM	175	14	1406	140605
20	2023	1	2023-01	Santarém	Abrantes	Rio de Moinhos	SIM	167	14	1401	140108
21	2023	7	2023-07	Santarém	Rio Maior	São Sebastião	SIM	117	14	1414	141411
22	2023	4	2023-04	Aveiro	Sever do Vouga	Pessegueiro do Vouga	SIM	919	1	117	011704
23	2023	4	2023-04	Coimbra	Arganil	Plódão	SIM	177	6	601	060111
24	2022	2	2022-02	Aveiro	Arouca	União das freguesias de Covelo de Paivó e Janarde	NÃO	138	1	104	010424
25	2022	12	2022-12	Évora	Mourão	Mourão	NÃO	232	7	708	070803
26	2022	10	2022-10	Bragança	Macedo de Cavaleiros	Lamalonga	NÃO	238	4	405	040517
27	2023	2	2023-02	Braga	Vieira do Minho	Eira Vedra	NÃO	216	3	311	031107
28	2022	7	2022-07	Guarda	Guarda	Pega	NÃO	75	9	907	090729
29	2023	8	2023-08	Bragança	Vinhais	União das freguesias de Sobreiro de Baixo e Alvaredos	NÃO	59	4	412	041240
30	2022	7	2022-07	Guarda	Pinhel	Manigoto	NÃO	126	9	910	091014
31	2023	6	2023-06	Leiria	Porto de Mós	Porto de Mós - São João Baptista e São Pedro	SIM	3815	10	1016	101614
32	2023	5	2023-05	Leiria	Porto de Mós	São Bento	SIM	272	10	1016	101610
33	2022	5	2022-05	Castelo Branco	Idanha-a-Nova	Proença-a-Velha	NÃO	266	5	505	050511
34	2022	5	2022-05	Viseu	São Pedro do Sul	Sul	NÃO	790	18	1816	181616

Close Stop Get more rows

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Entire Transformation



CSV File input (dataset B) Configuration Window

**CSV file input**

Step name: dataset B

Filename: /home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	String		7		\$	.	,	none
4 District	String		16		\$	.	,	none
5 Municipality	String		27		\$	.	,	none
6 parish	String		30		\$	.	,	none
7 Voltage level	String		32		\$	.	,	none
8 Active Energy (kWh)	Number	#####.###	12	3	\$	.	,	none
9 DistrictCode	Integer	#	15	0	\$	.	,	none
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	,	none
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

CSV File input (dataset B) Preview

Rows of step: dataset B (10000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VIDEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VIDEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VIDEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VIDEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VIDEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133

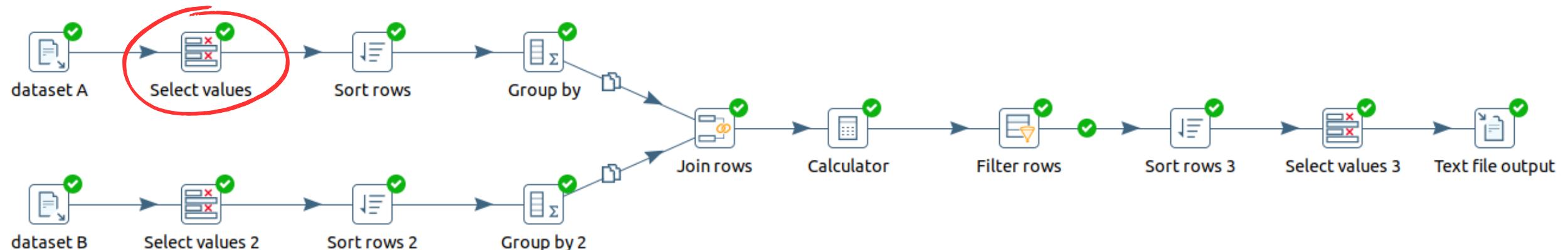
Examine preview data

Close Stop Get more rows

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Entire Transformation



Select Values Configuration Window

Select values

Step name

Select & Alter Remove Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision
1	District	District_A		

Include unspecified fields, ordered by name

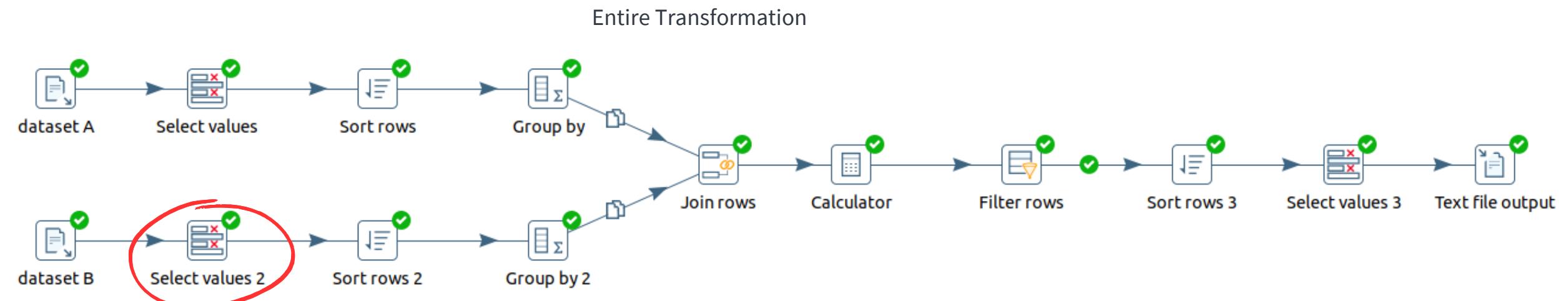
Select Values Preview

Rows of step: Select values (1000 rows)

Examine preview data

Row	District_A
1	Braga
2	Faro
3	Coimbra
4	Lisboa
5	Aveiro
6	Guarda
7	Coimbra
8	Porto
9	Viana do Castelo
10	Porto
11	Vila Real
12	Braganca
13	Braganca
14	Castelo Branco
15	Viana do Castelo
16	Braga
17	Viana do Castelo
18	Guarda
19	Santarém
20	Santarém
21	Santarém
22	Aveiro
23	Coimbra
24	Aveiro
25	Évora
26	Braganca
27	Braga
28	Guarda
29	Braganca
30	Guarda
31	Leiria
32	Leiria
33	Castelo Branco
34	Viseu

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Select Values 2 Configuration Window

Select values

Step name: Select values 2

Select & Alter Remove Meta-data

Fieldname	Rename to	Length	Precision
1 District	District_B		

Include unspecified fields, ordered by name

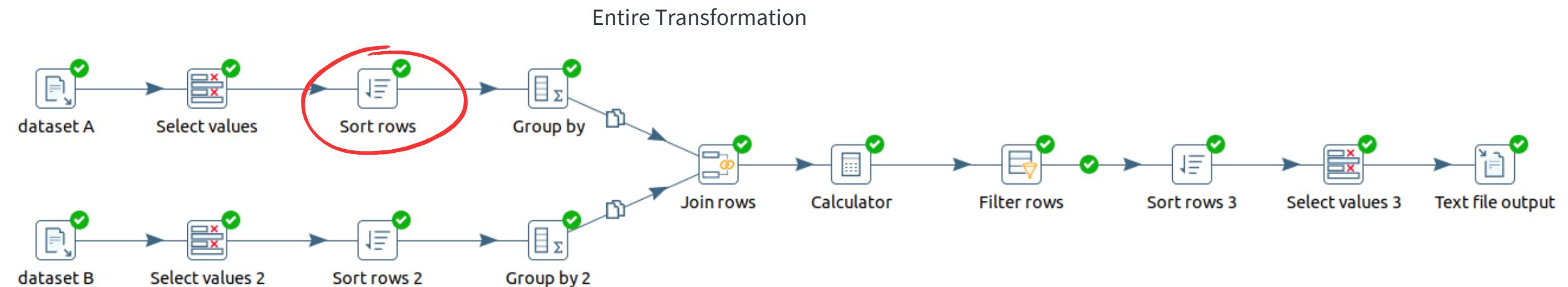
Select Values 2 Preview

Examine preview data

Rows of step: Select values 2 (1000 rows)

District_B
1 AVEIRO
2 BRAGA
3 FARO
4 GUARDA
5 PORTO
6 VIANA DO CASTELO
7 GUARDA
8 LEIRIA
9 LEIRIA
10 AVEIRO
11 AVEIRO
12 EVORA
13 EVORA
14 LISBOA
15 PORTO
16 SETUBAL
17 VIANA DO CASTELO
18 LEIRIA
19 LEIRIA
20 SANTAREM
21 VIANA DO CASTELO
22 VILA REAL
23 VISEU
24 VISEU
25 VISEU
26 LEIRIA
27 PORTO
28 SANTAREM
29 VISEU
30 VISEU
31 AVEIRO
32 BRAGANCA
33 GUARDA
34 PORTO

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Sort rows Configuration Window

Sort rows

Step name	Sort rows				
Sort directory	%java.io.tmpdir%				
TMP-file prefix	out				
Sort size (rows in memory)	1000000				
Free memory threshold (in %)					
Compress TMP Files?	<input type="checkbox"/>				
Only pass unique rows? (verifies keys only)	<input type="checkbox"/>				
Fields :					
Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1 District_A	Y	N	N	0	N

Help      OK      Cancel      Get Fields

Sort rows Preview

Rows of step: Sort rows (1000 rows)

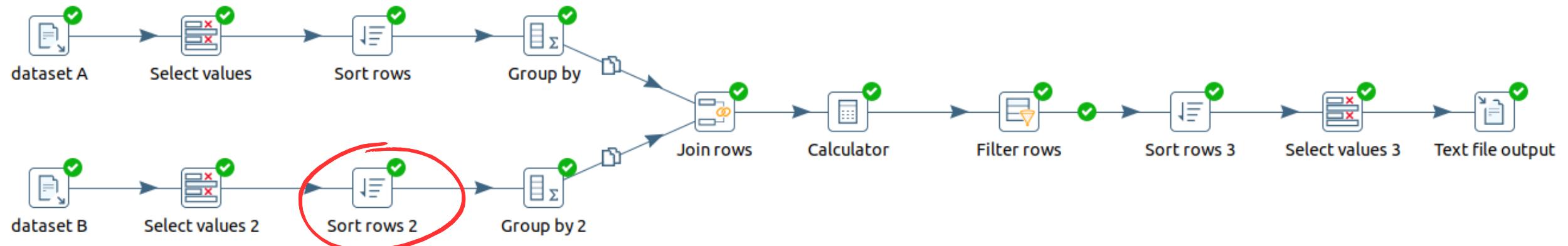
District_A	
1	Aveiro
2	Aveiro
3	Aveiro
4	Aveiro
5	Aveiro
6	Aveiro
7	Aveiro
8	Aveiro
9	Aveiro
10	Aveiro
11	Aveiro
12	Aveiro
13	Aveiro
14	Aveiro
15	Aveiro
16	Aveiro
17	Aveiro
18	Aveiro
19	Aveiro
20	Aveiro
21	Aveiro
22	Aveiro
23	Aveiro
24	Aveiro
25	Aveiro
26	Aveiro
27	Aveiro
28	Aveiro
29	Aveiro
30	Aveiro
31	Aveiro
32	Aveiro
33	Aveiro
34	Aveiro

Examine preview data      Close      Stop      Get more rows

## **3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B**



## Entire Transformation



## Sort rows 2 Configuration Window

**Sort rows**

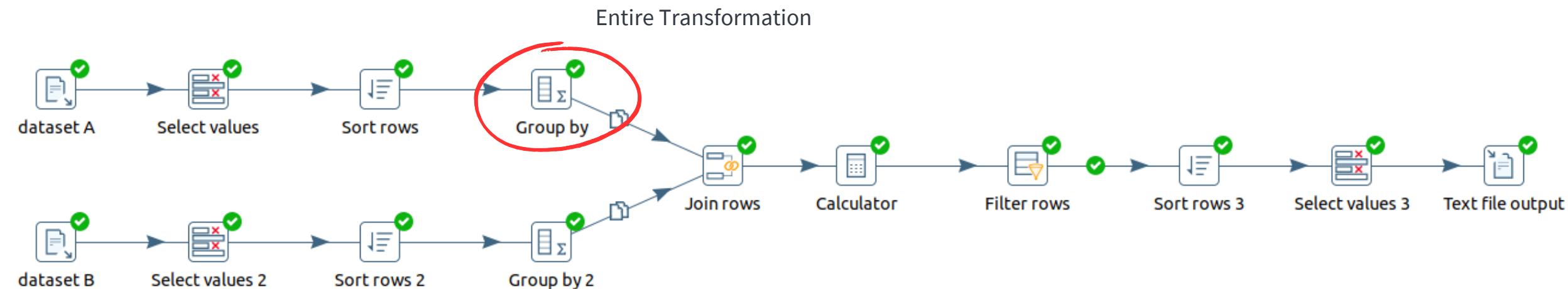
Step name	Sort rows 2					
Sort directory	%%java.io.tmpdir%% <a href="#">Browse</a>					
TMP-file prefix	out					
Sort size (rows in memory)	1000000					
Free memory threshold (in %)						
Compress TMP Files?	<input type="checkbox"/>					
Only pass unique rows? (verifies keys only)	<input type="checkbox"/>					
<b>Fields :</b>						
▼	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	District_B	Y	N	N	0	N

## Sort rows 2 Preview

Rows of step: Sort rows 2 (1000 rows)

	District_B
1	AVEIRO
2	AVEIRO
3	AVEIRO
4	AVEIRO
5	AVEIRO
6	AVEIRO
7	AVEIRO
8	AVEIRO
9	AVEIRO
10	AVEIRO
11	AVEIRO
12	AVEIRO
13	AVEIRO
14	AVEIRO
15	AVEIRO
16	AVEIRO
17	AVEIRO
18	AVEIRO
19	AVEIRO
20	AVEIRO
21	AVEIRO
22	AVEIRO
23	AVEIRO
24	AVEIRO
25	AVEIRO
26	AVEIRO
27	AVEIRO
28	AVEIRO
29	AVEIRO
30	AVEIRO
31	AVEIRO
32	AVEIRO
33	AVEIRO
34	AVEIRO

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Group by Configuration Window

**Group by**

Step name

Include all rows?

Temporary files directory

TMP-file prefix

Add line number, restart in each group

Line number field name

Always give back a result row

The fields that make up the group:

Group field

1	District_A
---	------------

Aggregates :

Name	Subject	Type	Value
1			

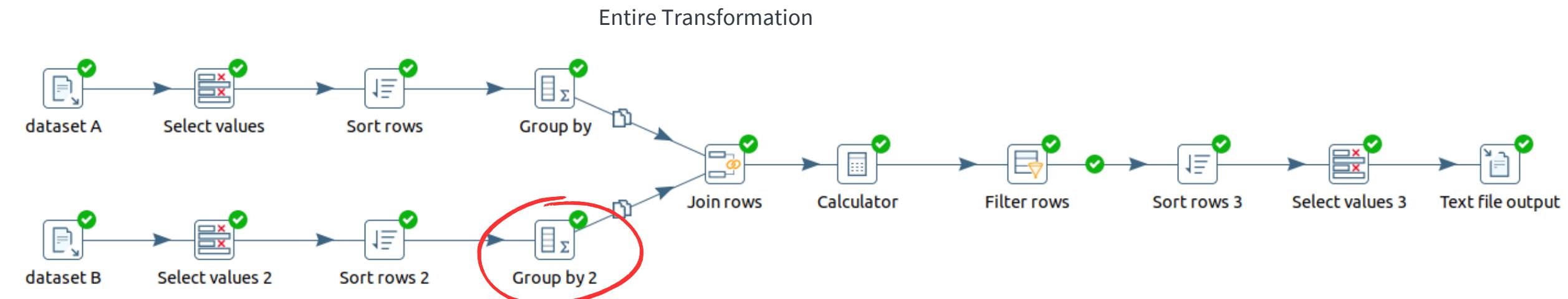
Group by Preview

**Examine preview data**

Rows of step: Group by (18 rows)

	District_A
1	Aveiro
2	Beja
3	Braga
4	Bragança
5	Castelo Branco
6	Coimbra
7	Faro
8	Guarda
9	Leiria
10	Lisboa
11	Portalegre
12	Porto
13	Santarém
14	Setúbal
15	Viana do Castelo
16	Vila Real
17	Viseu
18	Évora

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Group by 2 Configuration Window

Group by

Step name

Include all rows?

Temporary files directory

TMP-file prefix

Add line number, restart in each group

Line number field name

Always give back a result row

The fields that make up the group:

Group field

1 District_B
--------------

Aggregates :

Name	Subject	Type	Value	Get lookup fields
1				

Group by 2 Preview

Examine preview data

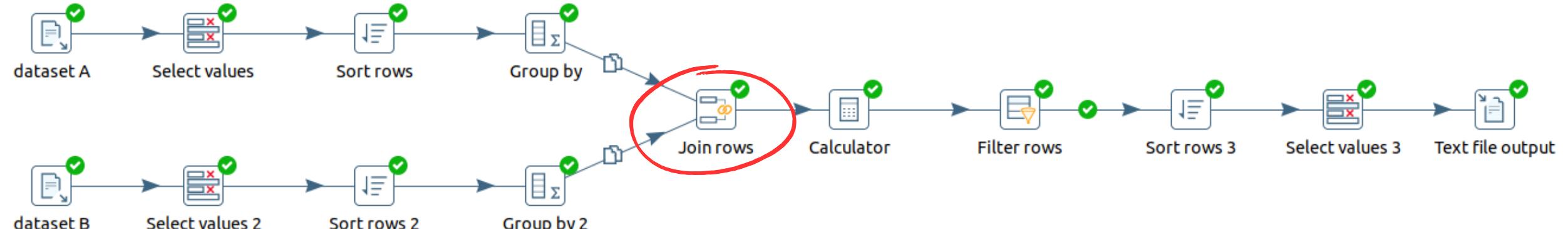
Rows of step: Group by 2 (18 rows)

District_B
1 AVEIRO
2 BEJA
3 BRAGA
4 BRAGANCA
5 CASTELO BRANCO
6 COIMBRA
7 EVORA
8 FARO
9 GUARDA
10 LEIRIA
11 LISBOA
12 PORTALEGRE
13 PORTO
14 SANTAREM
15 SETUBAL
16 VIANA DO CASTELO
17 VILA REAL
18 VISEU

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Entire Transformation



Get all possible combinations

Join rows Configuration Window

**Join rows (cartesian product)**

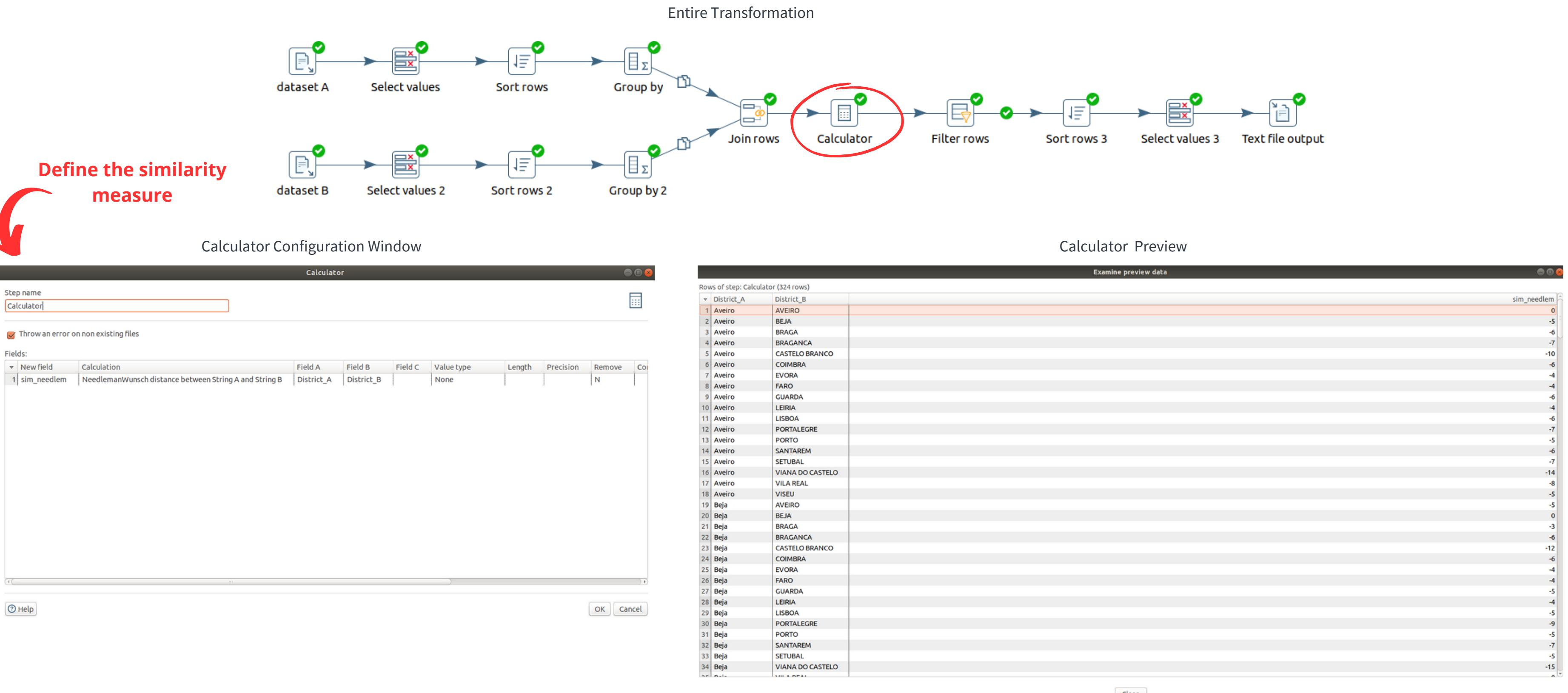
Step name	Join rows
Temp directory	%%java.io.tmpdir%%
TMP-file prefix	out
Max. cache size (in rows)	500
Main step to read from	<input type="button" value="▼"/>
The condition:	
<input type="text"/> <field> = <input type="text"/> <field> <input type="text"/> <value>	
<input type="button" value="Help"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/>	

Join rows Preview

Rows of step: Join rows (324 rows)

District_A	District_B
Aveiro	AVEIRO
Aveiro	BEJA
Aveiro	BRAGA
Aveiro	BRAGANCA
Aveiro	CASTELO BRANCO
Aveiro	COIMBRA
Aveiro	EVORA
Aveiro	FARO
Aveiro	GUARDA
Aveiro	LEIRIA
Aveiro	LISBOA
Aveiro	PORTALEGRE
Aveiro	PORTO
Aveiro	SANTAREM
Aveiro	SETUBAL
Aveiro	VIANA DO CASTELO
Aveiro	VILA REAL
Aveiro	VISEU
Beja	AVEIRO
Beja	BEJA
Beja	BRAGA
Beja	BRAGANCA
Beja	CASTELO BRANCO
Beja	COIMBRA
Beja	EVORA
Beja	FARO
Beja	GUARDA
Beja	LEIRIA
Beja	LISBOA
Beja	PORTALEGRE
Beja	PORTO
Beja	SANTAREM
Beja	SETUBAL
Beja	VIANA DO CASTELO

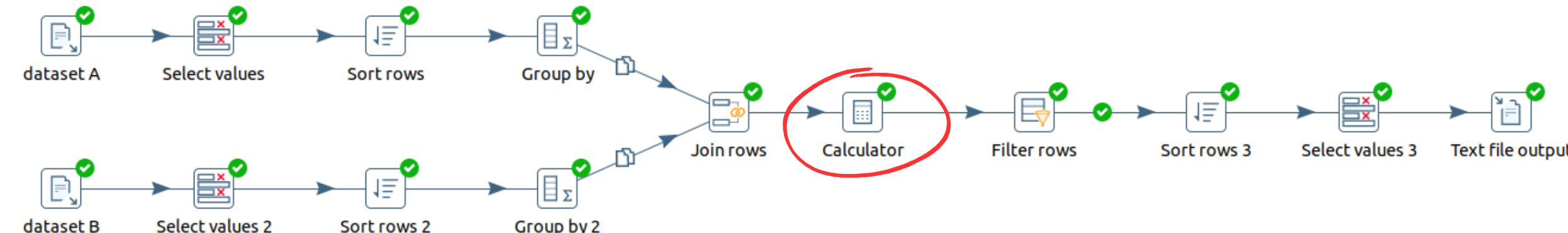
### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Entire Transformation



Calculator Preview (rows of interest)

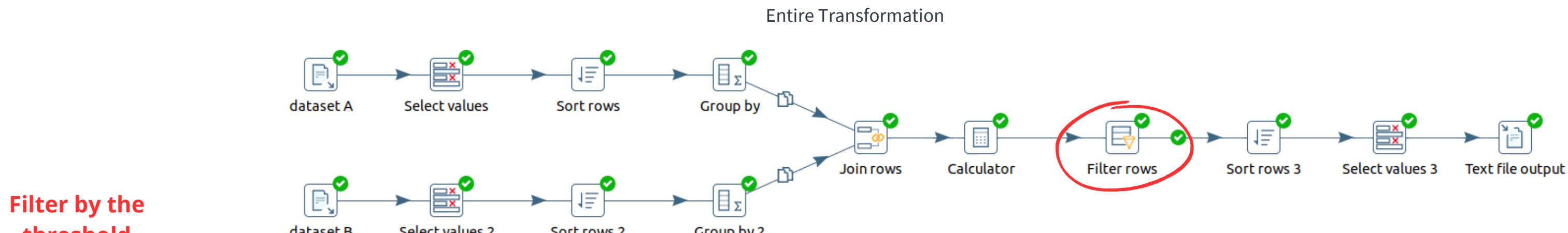
319	Évora	PORTO	-4
21	Beja	BRAGA	-3
38	Braga	BEJA	-3
40	Braga	BRAGANCA	-3
57	Bragança	BRAGA	-3
121	Faro	PORTO	-3
106	Porto	FARO	-3
58	Bragança	BRAGANCA	-1
30	Santarém	SANTAREM	-1
49	Setúbal	SETUBAL	-1
313	Évora	EVORA	-1
1	Aveiro	AVEIRO	0
20	Beja	BEJA	0
39	Braga	BRAGA	0
77	Castelo Branco	CASTELO BRANCO	0
96	Coimbra	COIMBRA	0
116	Faro	FARO	0
35	Guarda	GUARDA	0
154	Leiria	LEIRIA	0
173	Lisboa	LISBOA	0
192	Portalegre	PORTALEGRE	0
111	Porto	PORTO	0
168	Viana do Castelo	VIANA DO CASTELO	0
187	Vila Real	VILA REAL	0
306	Viseu	VISEU	0

Highest jump!

Threshold

Needleman-Wunsch measure:  
"The higher the score, the better"

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Filter by the threshold

Filter rows Window

Filter rows

Step name: Filter rows

Send 'true' data to step: Sort rows 3

Send 'false' data to step:

The condition:

sim\_needlem >=  -1 (Integer)

OK Cancel

Filter rows Preview

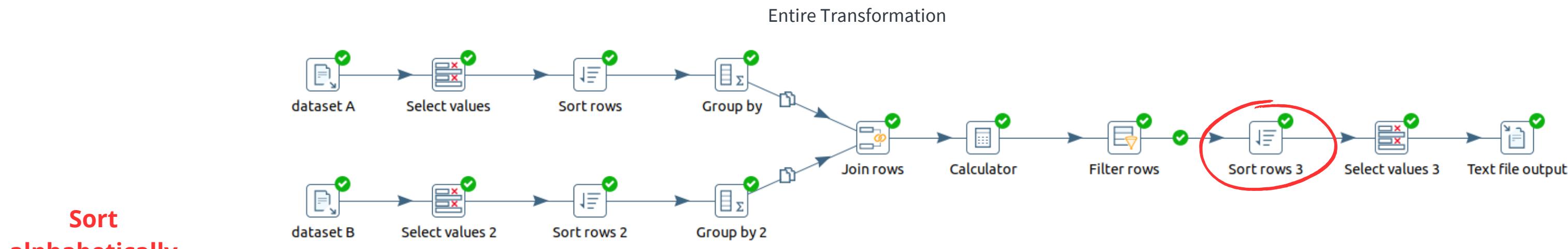
Rows of step: Filter rows (18 rows)

District_A	District_B	sim_needlem
1 Aveiro	AVEIRO	0
2 Beja	BEJA	0
3 Braga	BRAGA	0
4 Bragança	BRAGANCA	-1
5 Castelo Branco	CASTELO BRANCO	0
6 Coimbra	COIMBRA	0
7 Faro	FARO	0
8 Guarda	GUARDA	0
9 Leiria	LEIRIA	0
10 Lisboa	LISBOA	0
11 Portalegre	PORTALEGRE	0
12 Porto	PORTO	0
13 Santarém	SANTAREM	-1
14 Setúbal	SETUBAL	-1
15 Viana do Castelo	VIANA DO CASTELO	0
16 Vila Real	VILA REAL	0
17 Viseu	VISEU	0
18 Évora	EVORA	-1

Examine preview data

Close

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



**Sort alphabetically**



Sort rows 3 Configuration Window

Sort rows

Step name: Sort rows 3

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

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  District_B	Y	N	N	0	N

OK Cancel Get Fields

Sort rows 3 Preview

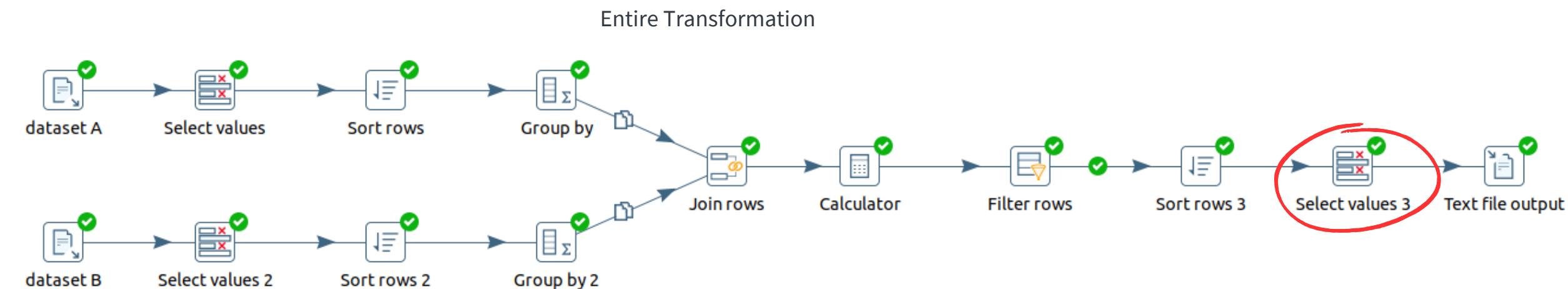
Rows of step: Sort rows 3 (18 rows)

Examine preview data

District_A	District_B	sim_needlem
1  Aveiro	AVEIRO	0
2  Beja	BEJA	0
3  Braga	BRAGA	0
4  Bragança	BRAGANCA	-1
5  Castelo Branco	CASTELO BRANCO	0
6  Coimbra	COIMBRA	0
7  Évora	EVORA	0
8  Faro	FARO	0
9  Guarda	GUARDA	0
10  Leiria	LEIRIA	0
11  Lisboa	LISBOA	0
12  Portalegre	PORTALEGRE	0
13  Porto	PORTO	0
14  Santarém	SANTAREM	-1
15  Setúbal	SETUBAL	0
16  Viana do Castelo	VIANA DO CASTELO	0
17  Vila Real	VILA REAL	0
18  Viseu	VISEU	0

Close

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Select values 3 Configuration Window

Select values

Step name **Select values 3**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 District_A			
2 District_B			

**Get fields to select** **Edit Mapping**

Include unspecified fields, ordered by name

OK Cancel

Select values 3 Preview

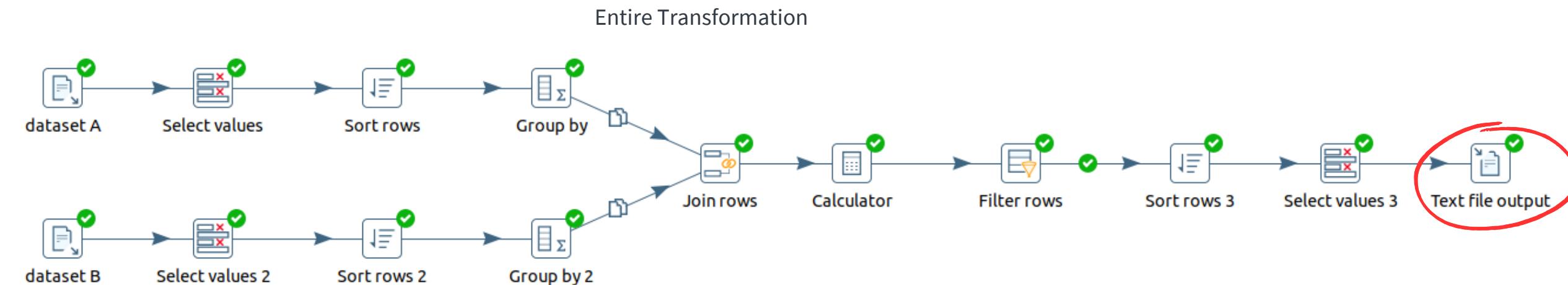
Rows of step: Select values 3 (18 rows)

District_A	District_B
1 Aveiro	AVEIRO
2 Beja	BEJA
3 Braga	BRAGA
4 Bragança	BRAGANCA
5 Castelo Branco	CASTELO BRANCO
6 Coimbra	COIMBRA
7 Évora	EVORA
8 Faro	FARO
9 Guarda	GUARDA
10 Leiria	LEIRIA
11 Lisboa	LISBOA
12 Portalegre	PORTALEGRE
13 Porto	PORTO
14 Santarém	SANTAREM
15 Setúbal	SETUBAL
16 Viana do Castelo	VIANA DO CASTELO
17 Vila Real	VILA REAL
18 Viseu	VISEU

Examine preview data

Close

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Text file output Configuration Window - File

Text file output

Step name: Text file output

**File**

- Filename: \${Internal.Entry.Current.Directory}/districts\_A\_B
- Pass output to servlet:
- Create Parent folder:
- Do not create file at start:
- Accept file name from field?:
- File name field:
- Extension: csv
- Include stepnr in filename?
- Include partition nr in filename?
- Include date in filename?
- Include time in filename?
- Specify Date time format?
- Date time format:
- Show filename(s)...
- Add filenames to result:

**OK** **Cancel**

Text file output Configuration Window - Content

Text file output

Step name: Text file output

**Content**

- Append:
- Separator: ;
- Enclosure: "
- Force the enclosure around fields?
- Disable the enclosure fix?
- Header:
- Footer:
- Format: CR+LF terminated (Windows, DOS)
- Compression: None
- Encoding: UTF-8
- Right pad fields:
- Fast data dump (no formatting)
- Split every ... rows:
- Add Ending line of file:

**OK** **Cancel**

Text file output Configuration Window - Fields

Text file output

Step name: Text file output

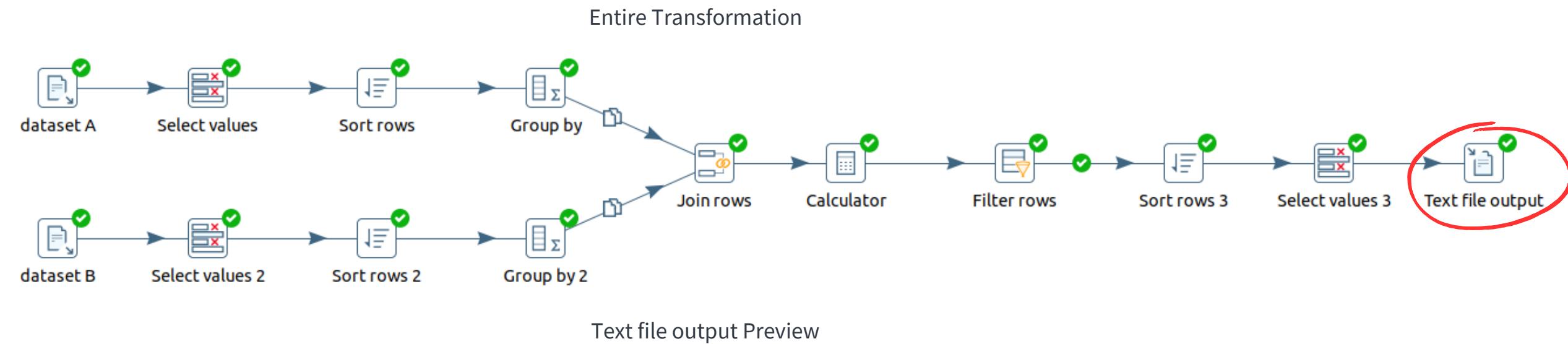
**Fields**

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1 District_A	String		16						none
2 District_B	String		16						none

**Get Fields** **Minimal width**

**OK** **Cancel**

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Text file output Preview

Examine preview data

Rows of step: Text file output (18 rows)	
District_A	District_B
1 Aveiro	AVEIRO
2 Beja	BEJA
3 Braga	BRAGA
4 Bragança	BRAGANCA
5 Castelo Branco	CASTELO BRANCO
6 Coimbra	COIMBRA
7 Évora	EVORA
8 Faro	FARO
9 Guarda	GUARDA
10 Leiria	LEIRIA
11 Lisboa	LISBOA
12 Portalegre	PORTALEGRE
13 Porto	PORTO
14 Santarém	SANTAREM
15 Setúbal	SETUBAL
16 Viana do Castelo	VIANA DO CASTELO
17 Vila Real	VILA REAL
18 Viseu	VISEU

Close

### 3. CORRECT MAPPING OF DISTRICTS IN DATASET A TO DISTRICTS IN DATASET B



Final output file:

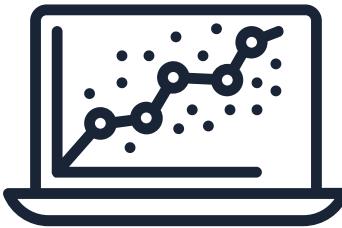
Liberation Sans 10 District\_A

	A	B	C
1	District_A	District_B	
2	Aveiro	AVEIRO	
3	Beja	BEJA	
4	Braga	BRAGA	
5	Bragança	BRAGANCA	
6	Castelo Branco	CASTELO BRANCO	
7	Coimbra	COIMBRA	
8	Évora	EVORA	
9	Faro	FARO	
10	Guarda	GUARDA	
11	Leiria	LEIRIA	
12	Lisboa	LISBOA	
13	Portalegre	PORTALEGRE	
14	Porto	PORTO	
15	Santarém	SANTAREM	
16	Setúbal	SETUBAL	
17	Viana do Castelo	VIANA DO CASTELO	
18	Vila Real	VILA REAL	
19	Viseu	VISEU	
20			
21			

# TASK 4

DATA WAREHOUSE SCHEMA CREATION WITH SQL

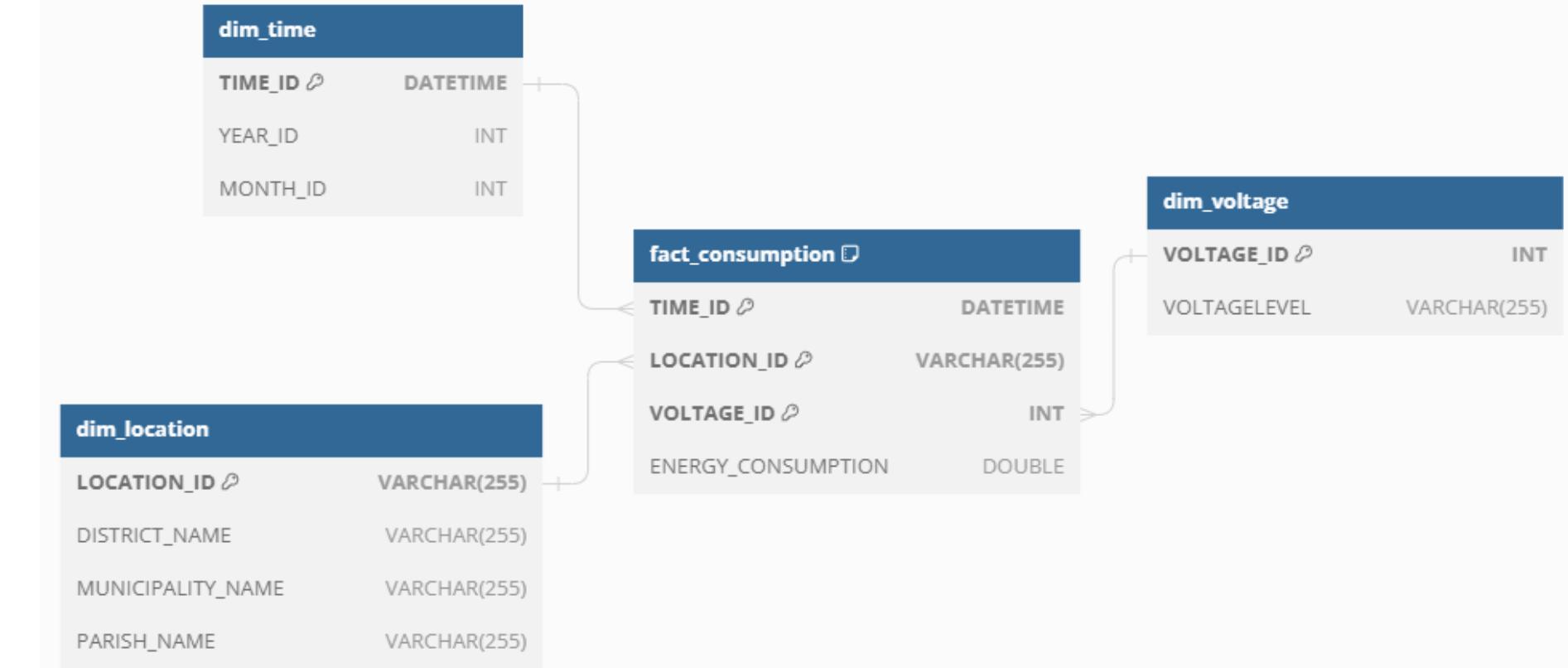
# 4. DATA WAREHOUSE SCHEMA CREATION WITH SQL



In order to create the data warehouse schema, this SQL script was made:

```
DROP DATABASE IF EXISTS datasetb_dw;  
  
CREATE DATABASE datasetb_dw;  
  
USE datasetb_dw;  
  
CREATE TABLE dim_voltage (  
    VOLTAGE_ID INT,  
    VOLTAGELEVEL VARCHAR(255),  
    PRIMARY KEY (VOLTAGE_ID)  
);  
  
CREATE TABLE dim_location (  
    LOCATION_ID VARCHAR(255),  
    DISTRICT_NAME VARCHAR(255),  
    MUNICIPALITY_NAME VARCHAR(255),  
    PARISH_NAME VARCHAR(255),  
    PRIMARY KEY (LOCATION_ID)  
);  
  
CREATE TABLE dim_time (  
    TIME_ID DATETIME,  
    YEAR_ID INT,  
    MONTH_ID INT,  
    PRIMARY KEY (TIME_ID)  
);  
  
CREATE TABLE fact_consumption (  
    TIME_ID DATETIME,  
    LOCATION_ID VARCHAR(255),  
    VOLTAGE_ID INT,  
    ENERGY_CONSUMPTION DOUBLE,  
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),  
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),  
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),  
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)  
);
```

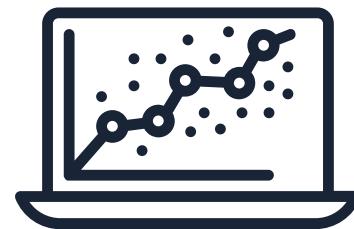
Data Warehouse Schema Design



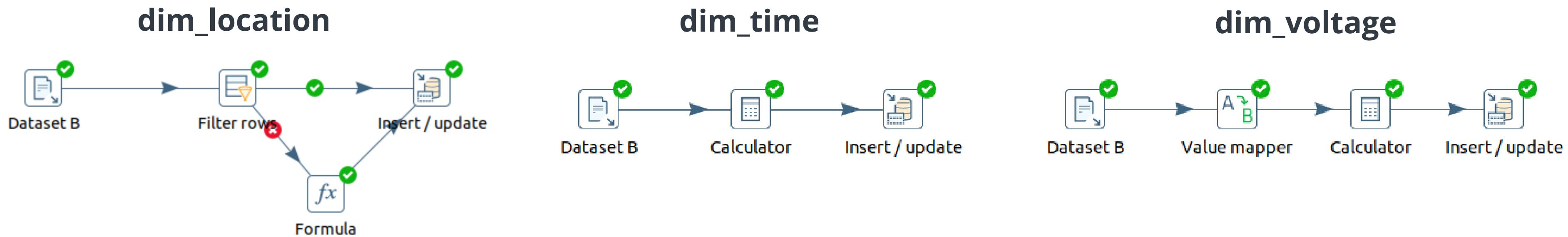
# **TASK 5**

## TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES

## 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES

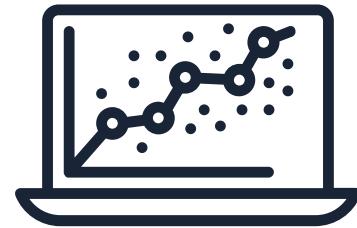


The following transformations were created to populate each dimension table:



The Configuration Windows, Previews for each step and for each dimension, as well as their output, are shown in the next slides.

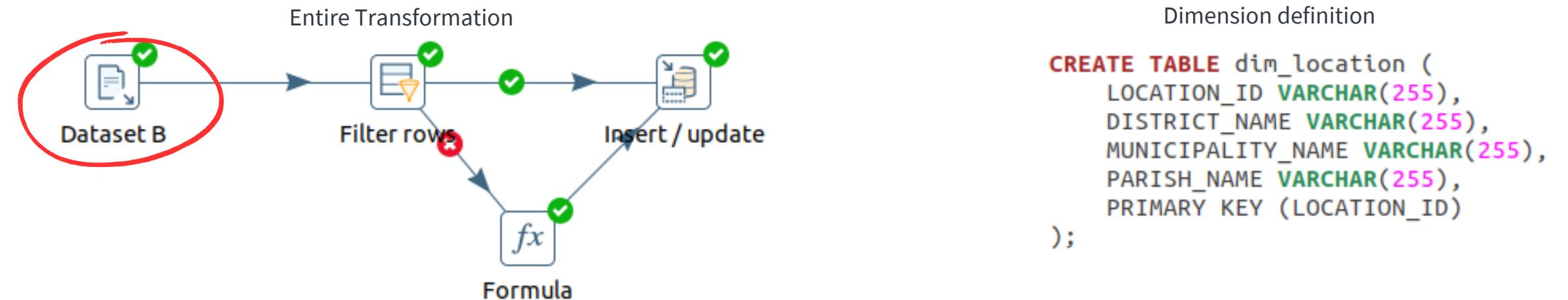
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



**dim\_location**

**dim\_time**

**dim\_voltage**



Dimension definition

```

CREATE TABLE dim_location (
    LOCATION_ID VARCHAR(255),
    DISTRICT_NAME VARCHAR(255),
    MUNICIPALITY_NAME VARCHAR(255),
    PARISH_NAME VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
    
```

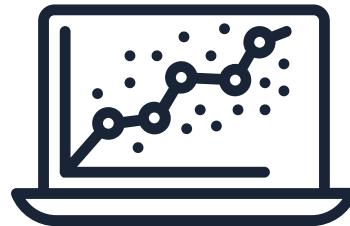
CSV File input (Dataset B) Configuration Window

CSV file input																																																																																																																					
Step name <b>Dataset B</b>																																																																																																																					
Filename <input type="text" value="/home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.csv"/> <input type="button" value="Browse..."/>																																																																																																																					
Delimiter <input type="text" value;"=""/> <input type="button" value="Insert TAB"/>																																																																																																																					
Enclosure <input type="text" value="#"/> <input type="button" value=""/>																																																																																																																					
NIO buffer size <input type="text" value="50000"/> <input type="button" value=""/>																																																																																																																					
Lazy conversion? <input type="checkbox"/>																																																																																																																					
Header row present? <input checked="" type="checkbox"/>																																																																																																																					
Add filename to result? <input type="checkbox"/>																																																																																																																					
The row number field name (optional) <input type="text"/>																																																																																																																					
Running in parallel? <input type="checkbox"/>																																																																																																																					
New line possible in fields? <input type="checkbox"/>																																																																																																																					
Format <input type="text" value="mixed"/>																																																																																																																					
File encoding <input type="text" value="UTF-8"/>																																																																																																																					
<table border="1"> <thead> <tr> <th>Name</th> <th>Type</th> <th>Format</th> <th>Length</th> <th>Precision</th> <th>Currency</th> <th>Decimal</th> <th>Group</th> <th>Trim type</th> </tr> </thead> <tbody> <tr><td>1 Year</td><td>Number</td><td>#.##</td><td>15</td><td>0</td><td>\$</td><td>.</td><td>,</td><td>both</td></tr> <tr><td>2 Month</td><td>Number</td><td>#.##</td><td>15</td><td>0</td><td>\$</td><td>.</td><td>,</td><td>both</td></tr> <tr><td>3 Date</td><td>String</td><td></td><td>8</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> <tr><td>4 District</td><td>String</td><td></td><td>17</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> <tr><td>5 Municipality</td><td>String</td><td></td><td>28</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> <tr><td>6 parish</td><td>String</td><td></td><td>31</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> <tr><td>7 Voltage level</td><td>String</td><td></td><td>33</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> <tr><td>8 Active Energy (kwh)</td><td>Number</td><td>#####.###</td><td>20</td><td>3</td><td>\$</td><td>.</td><td>,</td><td>both</td></tr> <tr><td>9 DistrictCode</td><td>Number</td><td>#.##</td><td>15</td><td>0</td><td>\$</td><td>.</td><td>,</td><td>both</td></tr> <tr><td>10 DistrictMunicipalityCode</td><td>Number</td><td>#.##</td><td>15</td><td>0</td><td>\$</td><td>.</td><td>,</td><td>both</td></tr> <tr><td>11 DistrictMunicipalityParishCode</td><td>String</td><td></td><td>6</td><td></td><td>\$</td><td>.</td><td>,</td><td>none</td></tr> </tbody> </table>										Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type	1 Year	Number	#.##	15	0	\$	.	,	both	2 Month	Number	#.##	15	0	\$	.	,	both	3 Date	String		8		\$	.	,	none	4 District	String		17		\$	.	,	none	5 Municipality	String		28		\$	.	,	none	6 parish	String		31		\$	.	,	none	7 Voltage level	String		33		\$	.	,	none	8 Active Energy (kwh)	Number	#####.###	20	3	\$	.	,	both	9 DistrictCode	Number	#.##	15	0	\$	.	,	both	10 DistrictMunicipalityCode	Number	#.##	15	0	\$	.	,	both	11 DistrictMunicipalityParishCode	String		6		\$	.	,	none
Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type																																																																																																													
1 Year	Number	#.##	15	0	\$	.	,	both																																																																																																													
2 Month	Number	#.##	15	0	\$	.	,	both																																																																																																													
3 Date	String		8		\$	.	,	none																																																																																																													
4 District	String		17		\$	.	,	none																																																																																																													
5 Municipality	String		28		\$	.	,	none																																																																																																													
6 parish	String		31		\$	.	,	none																																																																																																													
7 Voltage level	String		33		\$	.	,	none																																																																																																													
8 Active Energy (kwh)	Number	#####.###	20	3	\$	.	,	both																																																																																																													
9 DistrictCode	Number	#.##	15	0	\$	.	,	both																																																																																																													
10 DistrictMunicipalityCode	Number	#.##	15	0	\$	.	,	both																																																																																																													
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none																																																																																																													
<input type="button" value="Help"/> <input type="button" value="OK"/> <input type="button" value="Get Fields"/> <input type="button" value="Preview"/> <input type="button" value="Cancel"/>																																																																																																																					

CSV File input (Dataset B) Preview

Rows of step: Dataset B (10000 rows)												Examine preview data		
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode			
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809			
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120			
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503			
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144			
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612			
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629			
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315			
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102			
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909			
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204			
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610			
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526			
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302			
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115			
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001			
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804			
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156			
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304			
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801			
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909			
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737			
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102			
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525			
24	2022	6	2022-06	VISEU	Resende	CARQUEIRE	Baixa Tensão	80697.018	18	1813	181303			
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111			
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102			
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731			
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416			
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519			
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307			
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507			
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802			
33	2022	8	2022-08	GUARDA	Sabugal	MALCAT	Baixa Tensão	64069.193	9	911	091118			
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133			

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES

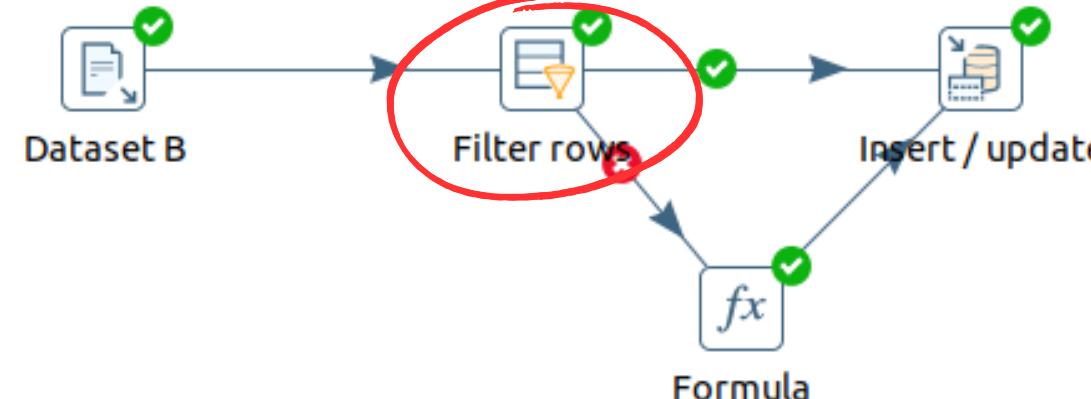


**dim\_location**

**dim\_time**

**dim\_voltage**

Entire Transformation



Dimension definition

```
CREATE TABLE dim_location (
    LOCATION_ID VARCHAR(255),
    DISTRICT_NAME VARCHAR(255),
    MUNICIPALITY_NAME VARCHAR(255),
    PARISH_NAME VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
```

Filter rows Configuration Window

Filter rows

Step name: Filter rows

Send 'true' data to step: Insert / update

Send 'false' data to step: Formula

The condition:

DistrictMunicipalityParishCode IS NOT NULL

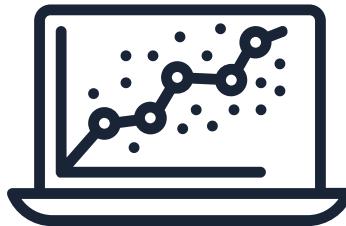
OK Cancel

Filter rows Preview

Row	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRA	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIJA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansiao	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133
35	2022	9	2022-09	BRAGA	Fafe	ARMIL	Baixa Tensão	73635.213	3	307	030705
36	2022	9	2022-09	COIMBRA	Coimbra	SAO SILVESTRE	Baixa Tensão	217019.542	6	602	060224

Separate the rows with non NULL "DistrictMunicipalityParishCode" values to get them ready to insert in our dim\_location

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



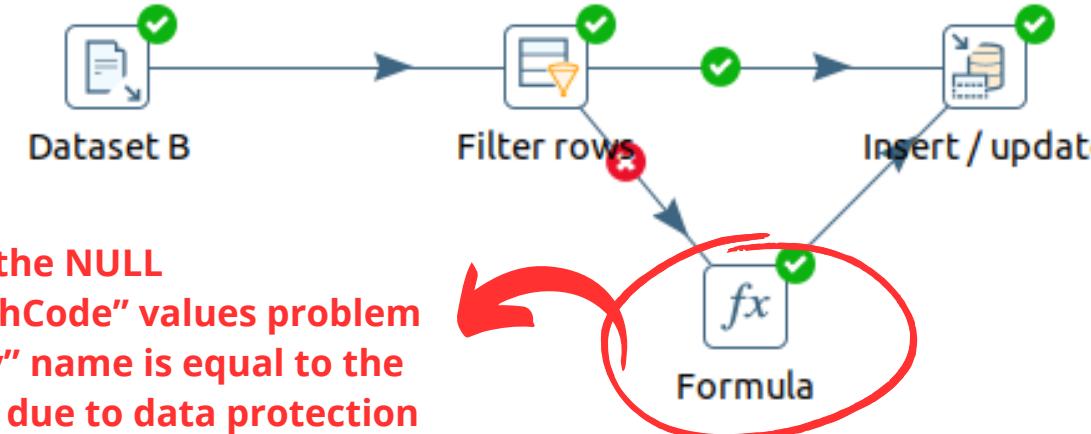
**dim\_location**

**dim\_time**

**dim\_voltage**

To handle the NULL  
"DistrictMunicipalityParishCode" values problem  
where the "Municipality" name is equal to the  
"parish" for each district, due to data protection

Entire Transformation



Dimension definition

```
CREATE TABLE dim_location (
    LOCATION_ID VARCHAR(255),
    DISTRICT_NAME VARCHAR(255),
    MUNICIPALITY_NAME VARCHAR(255),
    PARISH_NAME VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
```

Formula Configuration Window

Fields:	Step name	Formula	Value type	Length	Precision	Replace value
1	DistrictMunicipalityParishCode	CONCATENATE("DP-";[Municipality])	String	255	0	DistrictMunicipalityParishCode

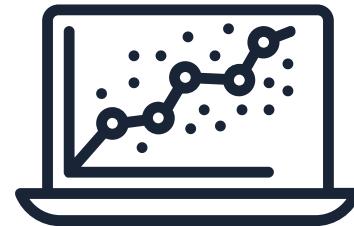
- "DistrictMunicipalityParishCode" values are our primary key **LOCATION\_ID**
- Replace NULL "DistrictMunicipalityParishCode" values with values in format : **DP-Municipality**, where DP = Data Protection

We didn't  
want to  
lose data!

Formula Preview

Rows of step: Formula (502 rows)											
1	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	10	2022-10	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	26063215.115	2	205	DP-Beja
2	2023	7	2023-07	SANTAREM	Santarém	SANTAREM	Muito Alta, Alta e Média Tensões	309404	14	1416	DP-Santarém
3	2022	4	2022-04	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1362003.362	6	603	DP-Coimbra
4	2021	8	2021-08	BRAGA	Braga	BRAGA	Baixa Tensão	8475.189	3	303	DP-Braga
5	2022	9	2022-09	EVORA	Évora	EVORA	Muito Alta, Alta e Média Tensões	155809.452	7	705	DP-Évora
6	2022	2	2022-02	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	1165957.062	1	105	DP-Aveiro
7	2022	9	2022-09	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	141537.433	11	1106	DP-Lisboa
8	2021	2	2021-02	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	21384335.482	2	205	DP-Beja
9	2021	4	2021-04	BRAGA	Braga	BRAGA	Muito Alta, Alta e Média Tensões	4547294.017	3	303	DP-Braga
10	2022	1	2022-01	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1823917.481	13	1312	DP-Porto
11	2023	5	2023-05	SANTAREM	Santarém	SANTAREM	Muito Alta, Alta e Média Tensões	566345.445	14	1416	DP-Santarém
12	2021	12	2021-12	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1687356.898	13	1312	DP-Porto
13	2021	4	2021-04	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1407654.581	6	603	DP-Coimbra
14	2021	2	2021-02	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	557907.697	1	105	DP-Aveiro
15	2021	11	2021-11	PORTALEGRE	Portalegre	PORTALEGRE	Muito Alta, Alta e Média Tensões	125079.866	12	1214	DP-Portalegre
16	2022	10	2022-10	BRAGA	Braga	BRAGA	Baixa Tensão	9232.9	3	303	DP-Braga
17	2023	4	2023-04	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	27644395.719	2	205	DP-Beja
18	2023	1	2023-01	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1555889.967	6	603	DP-Coimbra
19	2020	11	2020-11	SETUBAL	Setúbal	SETUBAL	Muito Alta, Alta e Média Tensões	1794380.955	15	1512	DP-Setúbal
20	2021	8	2021-08	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	155161.606	11	1106	DP-Lisboa
21	2022	12	2022-12	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	145345.5	11	1106	DP-Lisboa
22	2021	12	2021-12	PORTALEGRE	Portalegre	PORTALEGRE	Muito Alta, Alta e Média Tensões	157235.368	12	1214	DP-Portalegre
23	2023	7	2023-07	BRAGA	Braga	BRAGA	Baixa Tensão	3644	3	303	DP-Braga
24	2022	4	2022-04	BRAGA	Braga	BRAGA	Baixa Tensão	8989.848	3	303	DP-Braga
25	2021	2	2021-02	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1694402.287	13	1312	DP-Porto
26	2022	7	2022-07	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	1344615.147	1	105	DP-Aveiro
27	2020	11	2020-11	LEIRIA	Leiria	LEIRIA	Muito Alta, Alta e Média Tensões	443450.57	10	1009	DP-Leiria
28	2021	8	2021-08	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	426451.09	2	205	DP-Beja
29	2022	5	2022-05	FARO	Faro	FARO	Muito Alta, Alta e Média Tensões	195960.189	8	805	DP-Faro
30	2023	7	2023-07	SETUBAL	Setúbal	SETUBAL	Muito Alta, Alta e Média Tensões	1399553	15	1512	DP-Setúbal
31	2022	4	2022-04	BRAGANCA	Bragança	BRAGANCA	Muito Alta, Alta e Média Tensões	1680728.238	4	402	DP-Bragança
32	2021	9	2021-09	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	22895422.523	2	205	DP-Beja
33	2022	7	2022-07	BRAGA	Braga	BRAGA	Muito Alta, Alta e Média Tensões	4221354.331	3	303	DP-Braga
34	2020	12	2020-12	EVORA	Évora	EVORA	Muito Alta, Alta e Média Tensões	64812.502	7	705	DP-Évora
35	2023	4	2023-04	BRAGANCA	Bragança	BRAGANCA	Muito Alta, Alta e Média Tensões	1491579.71	4	402	DP-Bragança

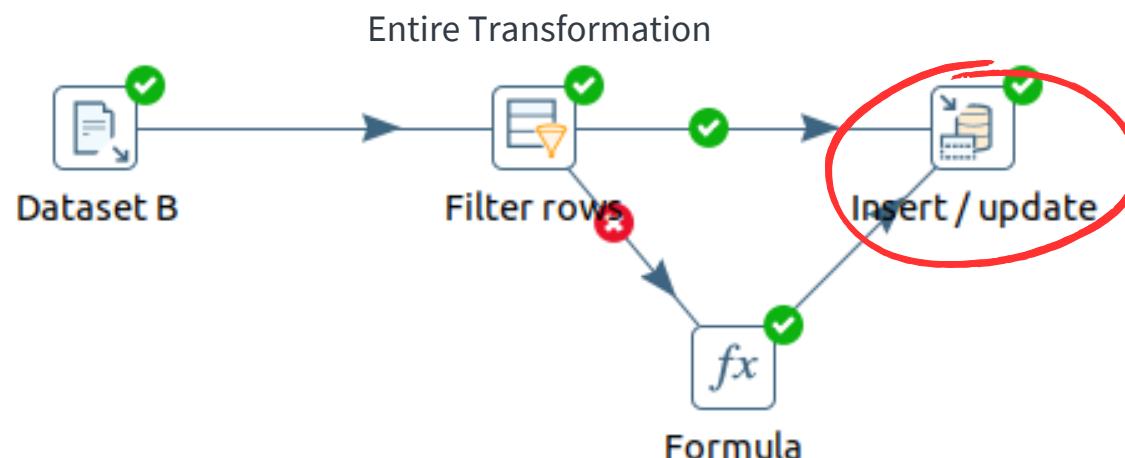
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



**dim\_location**

**dim\_time**

**dim\_voltage**



Dimension definition

```
CREATE TABLE dim_location (
    LOCATION_ID VARCHAR(255),
    DISTRICT_NAME VARCHAR(255),
    MUNICIPALITY_NAME VARCHAR(255),
    PARISH_NAME VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
```

Insert / update Configuration Window

Insert / update

Step name: **Insert / update**

Connection: datasetb\_dw

Target schema: datasetb\_dw

Target table: dim\_location

Commit size: 100

Don't perform any updates:

The key(s) to look up the value(s):

Table field	Comparator	Stream field1	Stream field2
1 LOCATION_ID	=	DistrictMunicipalityParishCode	

Update fields:

Table field	Stream field	Update
1 DISTRICT_NAME	District	Y
2 MUNICIPALITY_NAME	Municipality	Y
3 PARISH_NAME	parish	Y
4 LOCATION_ID	DistrictMunicipalityParishCode	Y

Get update fields | Edit mapping

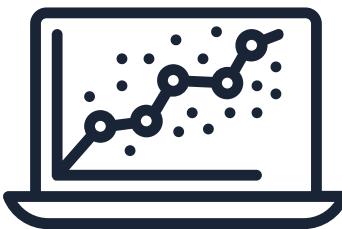
OK Cancel SQL

Insert / update Preview

Rows of step: Insert / update (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRA	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS E VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VIDEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VIDEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VIDEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VIDEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VIDEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133
35	2022	9	2022-09	BRAGA	Fafe	ARMIL	Baixa Tensão	73635.213	3	307	030705

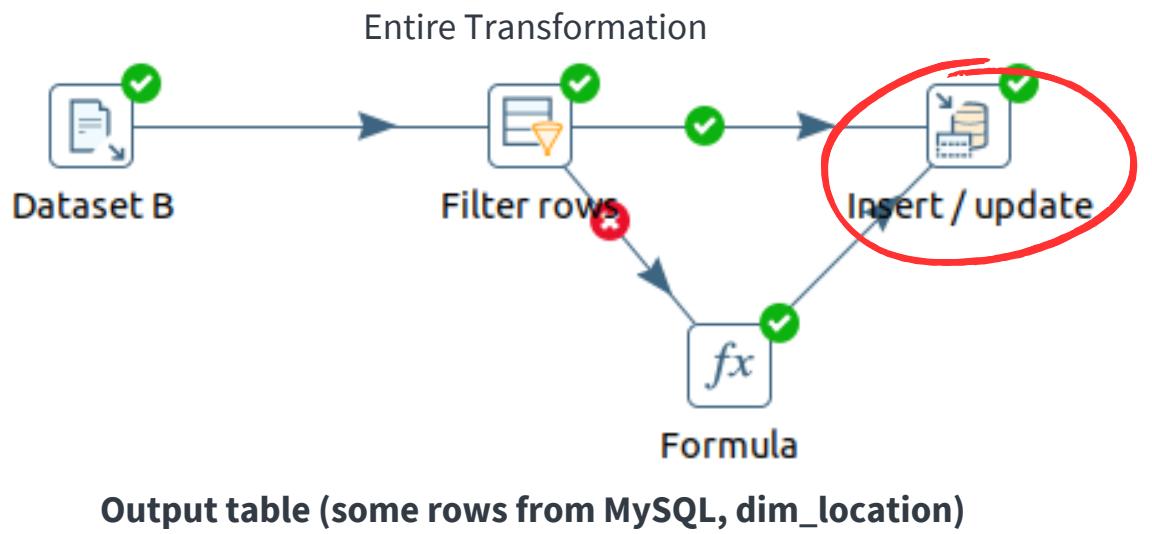
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage

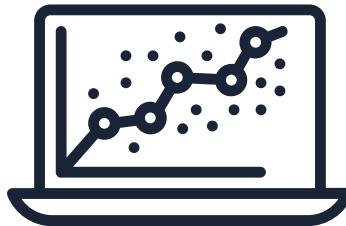


#	LOCATION_ID	DISTRICT_NAME	MUNICIPALITY_NAME	PARISH_NAME
1	010103	AVEIRO	Águeda	AGUADA DE CIMA
2	010109	AVEIRO	Águeda	FERMENTELOS
3	010112	AVEIRO	Águeda	MACINHATA DO VOUGA
4	010119	AVEIRO	Águeda	VALONGO DO VOUGA
5	010121	AVEIRO	Águeda	UF AGUEDA E BORRALHA
6	010122	AVEIRO	Águeda	UF BARRO E AGUADA DE BAIXO
7	010123	AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA A...
8	010124	AVEIRO	Águeda	UF RECARDAES E ESPINHEL
9	010125	AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA
10	010126	AVEIRO	Águeda	UF TROFA SEGADAS LAMAS V...
11	010127	AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCoba
12	010202	AVEIRO	Albergaria-a-Velha	ALQUERUBIM
13	010203	AVEIRO	Albergaria-a-Velha	ANGEJA
14	010204	AVEIRO	Albergaria-a-Velha	BRANCA
15	010206	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS
16	010209	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALM...
17	010210	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSS...
18	010304	AVEIRO	Anadia	AVELAS DE CAMINHO
19	010305	AVEIRO	Anadia	AVELAS DE CIMA
20	010307	AVEIRO	Anadia	MOITA
21	010309	AVEIRO	Anadia	SANGALHOS
22	010310	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO
23	010312	AVEIRO	Anadia	VILA NOVA DE MONSARROS
24	010313	AVEIRO	Anadia	VILARINHO DO BAIRRO
25	010316	AVEIRO	Anadia	UF DA GANDARA BAIRRO E A...
26	010317	AVEIRO	Anadia	UF ARCOES E MOGOFORES
27	010318	AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIR...
28	010402	AVEIRO	Arouca	ALVARENGA
29	010407	AVEIRO	Arouca	CHAVE
30	010409	AVEIRO	Arouca	ESCARIZ
31	010411	AVEIRO	Arouca	FERMEDO
32	010413	AVEIRO	Arouca	MANSORES
33	010414	AVEIRO	Arouca	MOLDES
34	010415	AVEIRO	Arouca	ROSSAS
35	010416	AVEIRO	Arouca	SANTA EULALIA
36	010417	AVEIRO	Arouca	SAO MIGUEL DO MATO
37	010418	AVEIRO	Arouca	TROPECO
38	010419	AVEIRO	Arouca	URRO

Dimension definition

```
CREATE TABLE dim_location (
    LOCATION_ID VARCHAR(255),
    DISTRICT_NAME VARCHAR(255),
    MUNICIPALITY_NAME VARCHAR(255),
    PARISH_NAME VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
```

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES

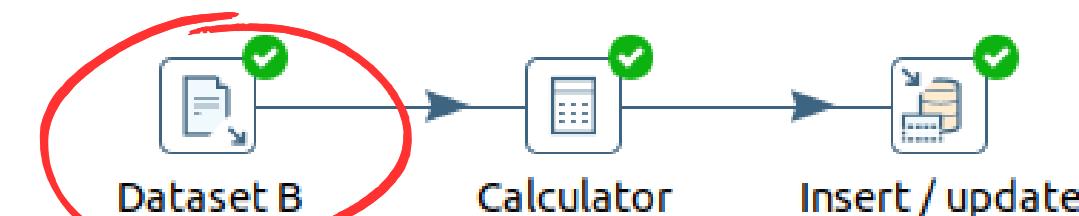


dim\_location

dim\_time

dim\_voltage

Entire Transformation



Dimension definition

```

CREATE TABLE dim_time (
    TIME_ID DATETIME,
    YEAR_ID INT,
    MONTH_ID INT,
    PRIMARY KEY (TIME_ID)
);
    
```

CSV File input (Dataset B) Configuration Window

**CSV file input**

Step name: Dataset B

Filename: /home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.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: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Number	#.##	15	0	\$	.	,	both
2 Month	Number	#.##	15	0	\$	.	,	both
3 Date	String		8		\$	.	,	none
4 District	String		17		\$	.	,	none
5 Municipality	String		28		\$	.	,	none
6 parish	String		31		\$	.	,	none
7 Voltage level	String		33		\$	.	,	none
8 Active Energy (kwh)	Number	#####.###	20	3	\$	.	,	both
9 DistrictCode	Number	#.##	15	0	\$	.	,	both
10 DistrictMunicipalityCode	Number	#.##	15	0	\$	.	,	both
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

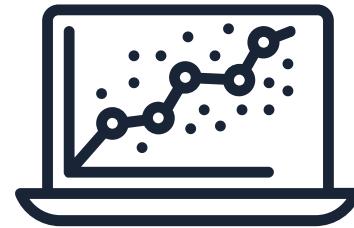
OK Get Fields Preview Cancel

CSV File input (Dataset B) Preview

Rows of step: Dataset B (10000 rows)												Examine preview data		
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode			
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809			
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120			
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503			
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144			
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612			
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629			
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315			
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102			
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909			
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204			
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610			
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526			
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302			
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115			
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001			
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804			
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156			
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304			
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801			
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909			
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737			
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102			
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525			
24	2022	6	2022-06	VISEU	Resende	CARQUEJE	Baixa Tensão	80697.018	18	1813	181303			
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111			
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102			
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731			
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416			
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519			
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307			
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507			
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802			
33	2022	8	2022-08	GUARDA	Sabugal	MALCAT	Baixa Tensão	64069.193	9	911	091118			
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133			

Close Stop Get more rows

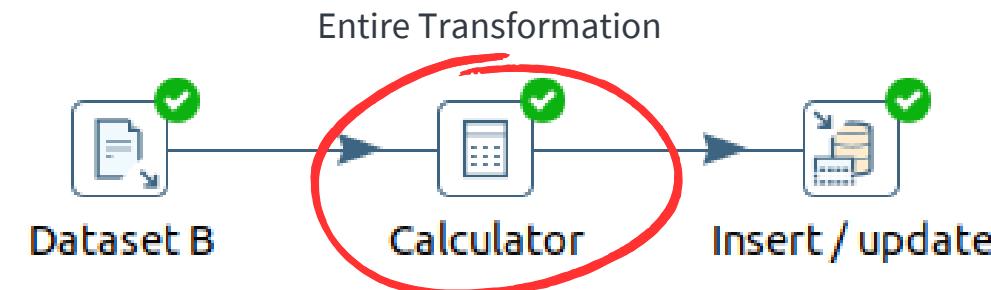
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage



Dimension definition

```

CREATE TABLE dim_time (
    TIME_ID DATETIME,
    YEAR_ID INT,
    MONTH_ID INT,
    PRIMARY KEY (TIME_ID)
);
  
```

Calculator Configuration Window

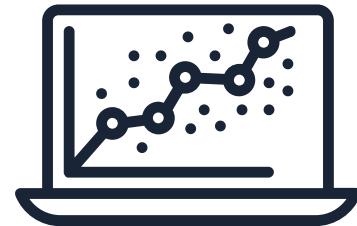
New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol	Currency symbol
dateConverted	Create a copy of field A	Date			Date		N		YYYY-MM			

Calculator Preview

Rows of step: Calculator (1000 rows)													Examine preview data		
#	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	dateConverted			
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	2022-03			
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	2022-03			
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2022-03			
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	2022-03			
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2022-03			
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	2022-03			
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	2022-04			
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2022-04			
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2022-04			
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2022-05			
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIWA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	2022-05			
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2022-05			
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	2022-05			
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	2022-05			
15	2022	5	2022-05	PORTO	Paredes	AGUARIA DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	2022-05			
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	2022-05			
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	2022-05			
18	2022	6	2022-06	LEIRIA	Ansiao	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2022-06			
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	2022-06			
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2022-06			
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	2022-06			
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2022-06			
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2022-06			
24	2022	6	2022-06	VISEU	Resende	CARQUEIRE	Baixa Tensão	806970.018	18	1813	181303	2022-06			
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	2022-06			
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2022-07			
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2022-07			
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELERA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416	2022-07			
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	2022-07			
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	2022-07			
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2022-08			
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2022-08			
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118	2022-08			
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	2022-08			

Convert the Date field (string) to date type with the YYYY-MM mask

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES

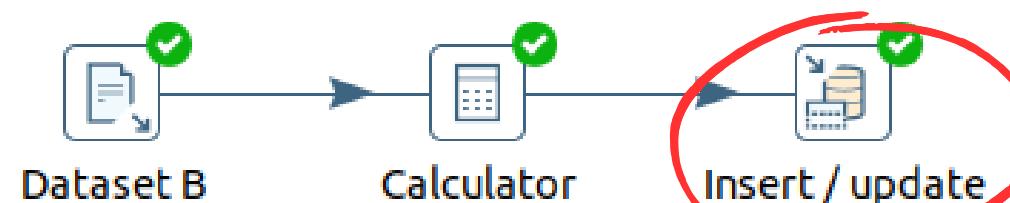


dim\_location

dim\_time

dim\_voltage

Entire Transformation



Dimension definition

```

CREATE TABLE dim_time (
    TIME_ID DATETIME,
    YEAR_ID INT,
    MONTH_ID INT,
    PRIMARY KEY (TIME_ID)
);
    
```

Insert / update Configuration Window

Table field	Comparator	Stream field1	Stream field2	Get fields
1 TIME_ID	=		dateConverted	

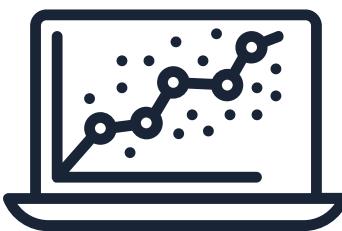
  

Table field	Stream field	Update	Get update fields
1 YEAR_ID	Year	Y	
2 MONTH_ID	Month	Y	
3 TIME_ID	dateConverted	Y	

Insert / update Preview

Rows of step: Insert / update (1000 rows)												Examine preview data		
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	dateConverted		
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	2022-03		
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	2022-03		
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2022-03		
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	2022-03		
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2022-03		
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	2022-03		
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	2022-04		
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2022-04		
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2022-04		
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2022-05		
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	2022-05		
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2022-05		
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	2022-05		
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CUMBRA	Baixa Tensão	1254374.575	11	1111	111115	2022-05		
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	2022-05		
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	2022-05		
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	2022-05		
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2022-06		
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	2022-06		
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2022-06		
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	2022-06		
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2022-06		
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2022-06		
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303	2022-06		
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	2022-06		
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2022-07		
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2022-07		
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416	2022-07		
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	2022-07		
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	2022-07		
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2022-08		
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2022-08		
33	2022	8	2022-08	GUARDA	Sabugal	MALCATIA	Baixa Tensão	64069.193	9	911	091118	2022-08		
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	2022-08		

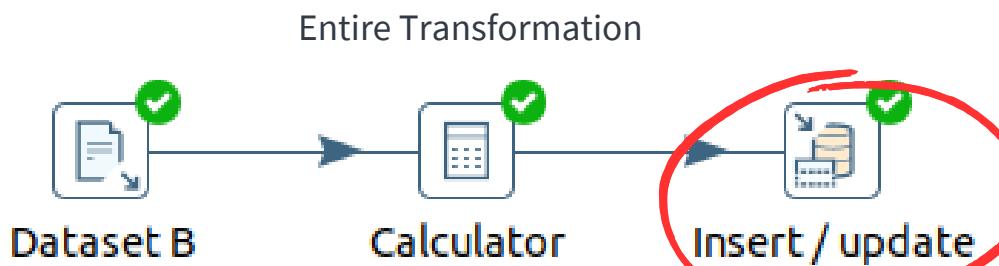
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage



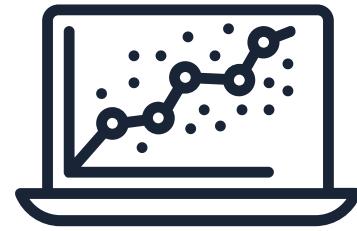
Dimension definition

```
CREATE TABLE dim_time (
    TIME_ID DATETIME,
    YEAR_ID INT,
    MONTH_ID INT,
    PRIMARY KEY (TIME_ID)
);
```

Output table (some rows from MySQL, dim\_time)

#	TIME_ID	YEAR_ID	MONTH_ID
1	2020-11-01 00:00:00	2020	11
2	2020-12-01 00:00:00	2020	12
3	2021-01-01 00:00:00	2021	1
4	2021-02-01 00:00:00	2021	2
5	2021-03-01 00:00:00	2021	3
6	2021-04-01 00:00:00	2021	4
7	2021-05-01 00:00:00	2021	5
8	2021-06-01 00:00:00	2021	6
9	2021-07-01 00:00:00	2021	7
10	2021-08-01 00:00:00	2021	8
11	2021-09-01 00:00:00	2021	9
12	2021-10-01 00:00:00	2021	10
13	2021-11-01 00:00:00	2021	11
14	2021-12-01 00:00:00	2021	12
15	2022-01-01 00:00:00	2022	1
16	2022-02-01 00:00:00	2022	2
17	2022-03-01 00:00:00	2022	3
18	2022-04-01 00:00:00	2022	4
19	2022-05-01 00:00:00	2022	5
20	2022-06-01 00:00:00	2022	6
21	2022-07-01 00:00:00	2022	7
22	2022-08-01 00:00:00	2022	8
23	2022-09-01 00:00:00	2022	9
24	2022-10-01 00:00:00	2022	10
25	2022-11-01 00:00:00	2022	11
26	2022-12-01 00:00:00	2022	12
27	2023-01-01 00:00:00	2023	1
28	2023-02-01 00:00:00	2023	2

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage



CSV File input (Dataset B) Configuration Window

CSV file input

Step name **Dataset B**

Filename **/home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.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 **UTF-8**

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Number	#.##	15	0	\$	.	,	both
2 Month	Number	#.##	15	0	\$	.	,	both
3 Date	String		8		\$	.	,	none
4 District	String		17		\$	.	,	none
5 Municipality	String		28		\$	.	,	none
6 parish	String		31		\$	.	,	none
7 Voltage level	String		33		\$	.	,	none
8 Active Energy (kwh)	Number	#####.###	20	3	\$	.	,	both
9 DistrictCode	Number	#.##	15	0	\$	.	,	both
10 DistrictMunicipalityCode	Number	#.##	15	0	\$	.	,	both
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

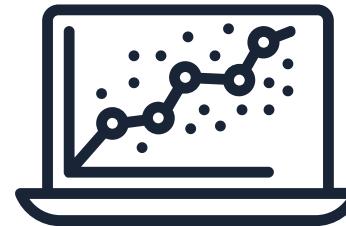
CSV File input (Dataset B) Preview

Rows of step: Dataset B (10000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VISEU	Resende	CARQUEIRE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCAT	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133

Close Stop Get more rows

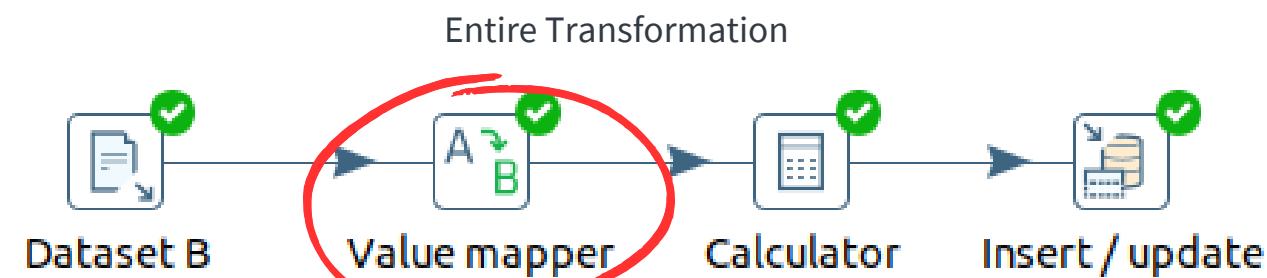
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage



Dimension definition

```

CREATE TABLE dim_voltage (
    VOLTAGE_ID INT,
    VOLTAGELEVEL VARCHAR(255),
    PRIMARY KEY (VOLTAGE_ID)
);
    
```

Value mapper Configuration Window

**Value mapper**

Step name: **Value mapper**

Fieldname to use: **Voltage level**

Target field name (empty=overwrite): **VOLTAGE\_ID**

Default upon non-matching:

Source value	Target value
Baixa Tensão	1
Muito Alta, Alta e Média Tensões	2

Field values:

Source value	Target value
Baixa Tensão	1
Muito Alta, Alta e Média Tensões	2

OK Cancel

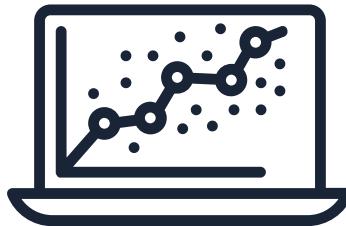
Value mapper Preview

Rows of step: Value mapper (1000 rows)

Examine preview data

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	VOLTAGE_ID
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	1
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	1
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	1
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	1
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	1
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	1
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIDA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	1
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	1
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	1
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	1
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	1
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	1
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	1
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguiã	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303	1
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	1
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Baixa Tensão	77169.225	14	1414	141416	2
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	1
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	1
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118	1
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	1

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage

Convert the VOLTAGE\_ID string to int

Calculator Configuration Window

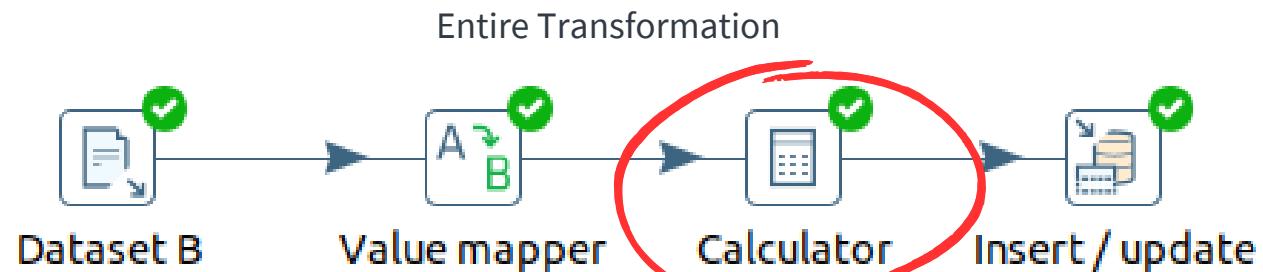
**Calculator**

Step name: Calculator

Fields:

New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Def.
1	VOLTAGE_ID	Create a copy of field A	VOLTAGE_ID		Integer		N			

Throw an error on non existing files



Dimension definition

```

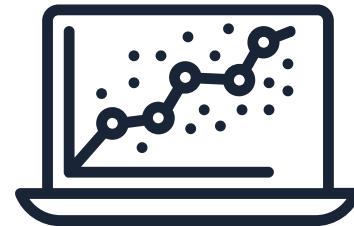
CREATE TABLE dim_voltage (
    VOLTAGE_ID INT,
    VOLTAGELEVEL VARCHAR(255),
    PRIMARY KEY (VOLTAGE_ID)
);
    
```

Calculator Preview

Rows of step: Calculator (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipality	DistrictMunicipalityParishCc	VOLTAGE_ID	VOLTAGE_ID_1
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	1	1
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	1	1
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2	2
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	1	1
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2	2
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	1	1
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	1	1
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2	2
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRA	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2	2
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2	2
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	1	1
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2	2
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	1	1
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	1	1
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	1	1
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	1	1
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	1	1
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2	2
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	1	1
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2	2
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	1	1
22	2022	6	2022-06	VILA REAL	Santa Marta de Penag	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2	2
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2	2
24	2022	6	2022-06	VISEU	Resende	CARQUEIRA	Baixa Tensão	80697.018	18	1813	181303	1	1
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	1	1
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2	2
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2	2
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416	2	2
29	2022	7	2022-07	VISEU	Lamego	SAMODEAS	Baixa Tensão	15388.178	18	1805	180519	1	1
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	1	1
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2	2
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2	2
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118	1	1
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	1	1

# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage

Entire Transformation



Dimension definition

```

CREATE TABLE dim_voltage (
    VOLTAGE_ID INT,
    VOLTAGELEVEL VARCHAR(255),
    PRIMARY KEY (VOLTAGE_ID)
);
  
```

Insert / update Configuration Window

**Insert / update**

Step name: **Insert / update**

Connection: **datasetb\_dw**

Target schema: **datasetb\_dw**

Target table: **dim\_voltage**

Commit size: **100**

Don't perform any updates:

The key(s) to look up the value(s):

Table field	Comparator	Stream field1	Stream field2
VOLTAGE_ID	=	VOLTAGE_ID	

Update fields:

Table field	Stream field	Update
VOLTAGELEVEL	Voltage level	Y
VOLTAGE_ID	VOLTAGE_ID	Y

**Get fields**, **Get update fields**, **Edit mapping**

OK, Cancel, SQL, Help

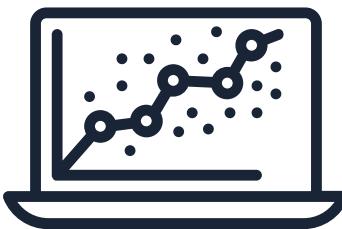
Insert / update Preview

Rows of step: Insert / update (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kW)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	VOLTAGE_ID	VOLTAGE_ID_1
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	1	1
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	1	1
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2	2
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	1	1
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2	2
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	1	1
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	1	1
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2	2
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2	2
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2	2
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	1	1
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2	2
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	1	1
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	1	1
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	1	1
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	1	1
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	1	1
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2	2
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	1	1
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2	2
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	1	1
22	2022	6	2022-06	VILA REAL	Santa Marta de Penagui	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2	2
23	2022	6	2022-06	VIDEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2	2
24	2022	6	2022-06	VIDEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303	1	1
25	2022	6	2022-06	VIDEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	1	1
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2	2
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2	2
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416	2	2
29	2022	7	2022-07	VIDEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	1	1
30	2022	7	2022-07	VIDEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	1	1
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2	2
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2	2
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118	1	1
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	1	1

Close, Stop, Get more rows

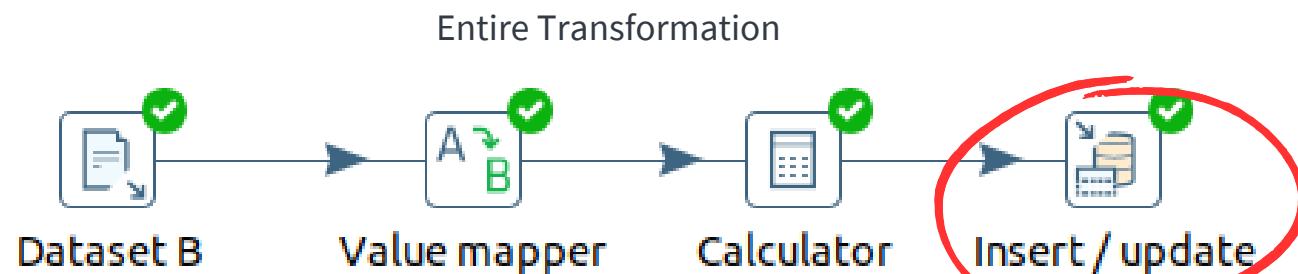
# 5. TRANSFORMATIONS TO POPULATE THE DIMENSION TABLES



dim\_location

dim\_time

dim\_voltage



Dimension definition

```
CREATE TABLE dim_voltage (
    VOLTAGE_ID INT,
    VOLTAGELEVEL VARCHAR(255),
    PRIMARY KEY (VOLTAGE_ID)
);
```

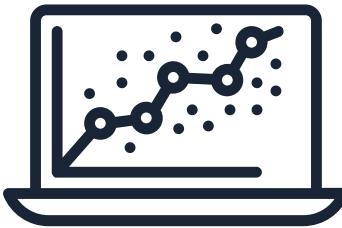
Output table (from MySQL, dim\_voltage)

#	VOLTAGE_ID	VOLTAGELEVEL
1	1	Baixa Tensão
2	2	Muito Alta, Alta e Média Tensões
*	NULL	NULL

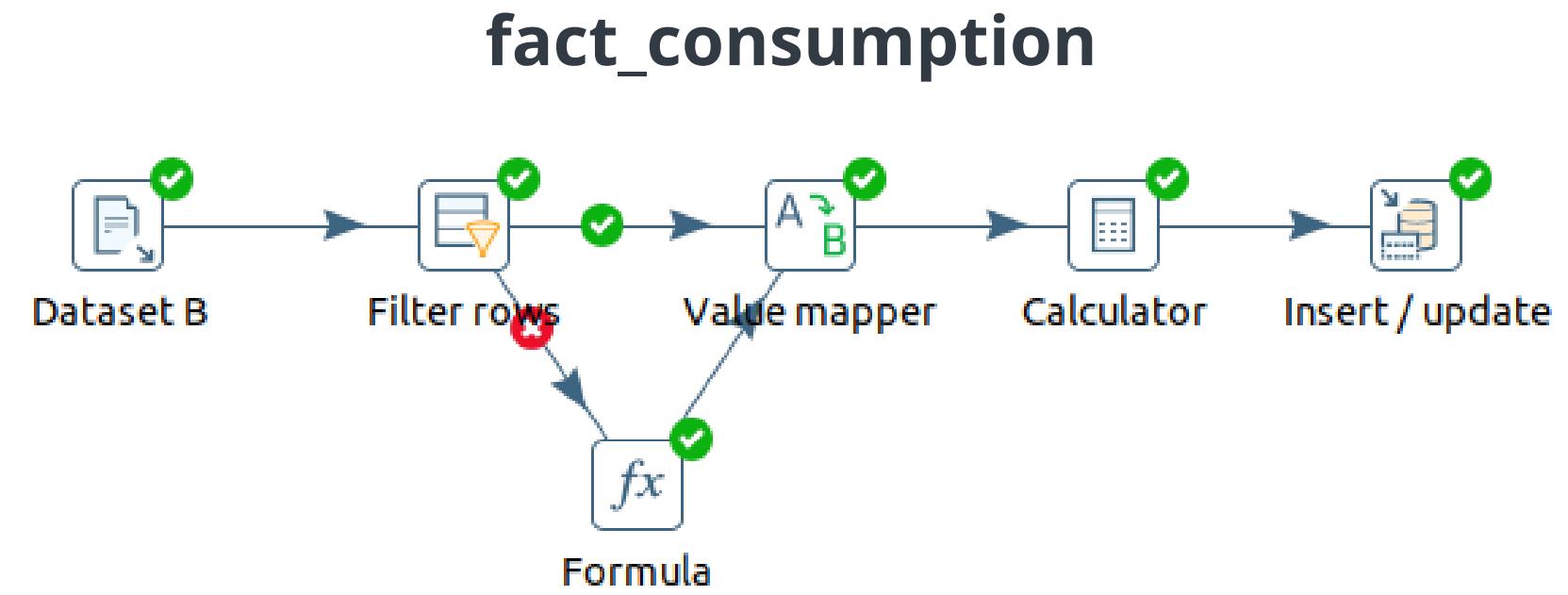
# TASK 6

TRANSFORMATION TO POPULATE THE FACT TABLE

## 6. TRANSFORMATION TO POPULATE THE FACT TABLE

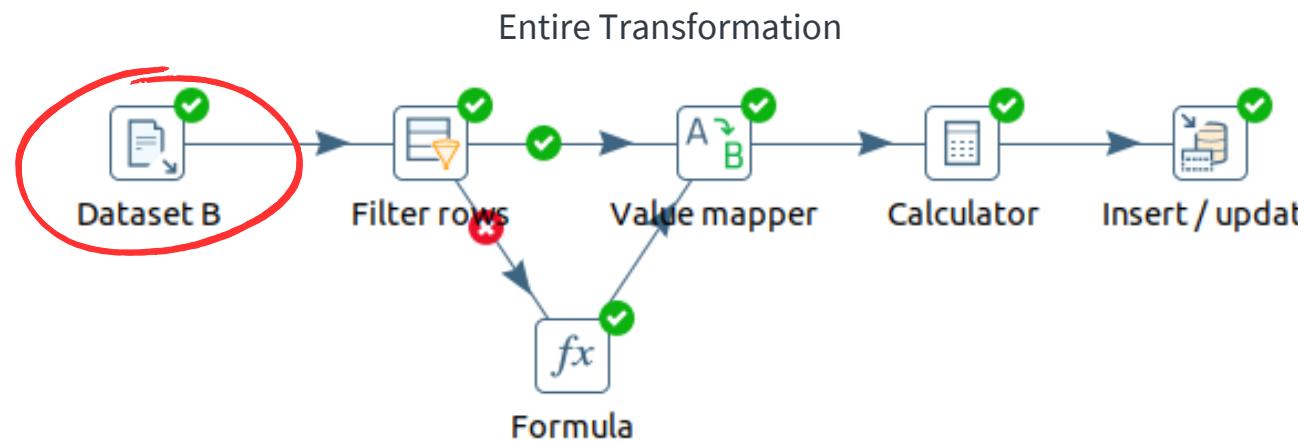
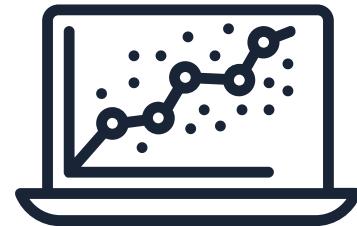


The following transformation was created to populate the fact table:



The Configuration Windows, Previews for each step, as well as the output, are shown in the next slides.

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Dimension definition

```

CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
  
```

CSV File input (Dataset B) Configuration Window

CSV file input

Step name **Dataset B**

Filename **/home/aid/Desktop/3-consumos-faturados-por-municipio-ultimos-10-anos.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 **UTF-8**

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Number	#.##	15	0	\$	.	,	both
2 Month	Number	#.##	15	0	\$	.	,	both
3 Date	String		8		\$	.	,	none
4 District	String		17		\$	.	,	none
5 Municipality	String		28		\$	.	,	none
6 parish	String		31		\$	.	,	none
7 Voltage level	String		33		\$	.	,	none
8 Active Energy (kwh)	Number	#####.###	20	3	\$	.	,	both
9 DistrictCode	Number	#.##	15	0	\$	.	,	both
10 DistrictMunicipalityCode	Number	#.##	15	0	\$	.	,	both
11 DistrictMunicipalityParishCode	String		6		\$	.	,	none

OK Get Fields Preview Cancel

CSV File input (Dataset B) Preview

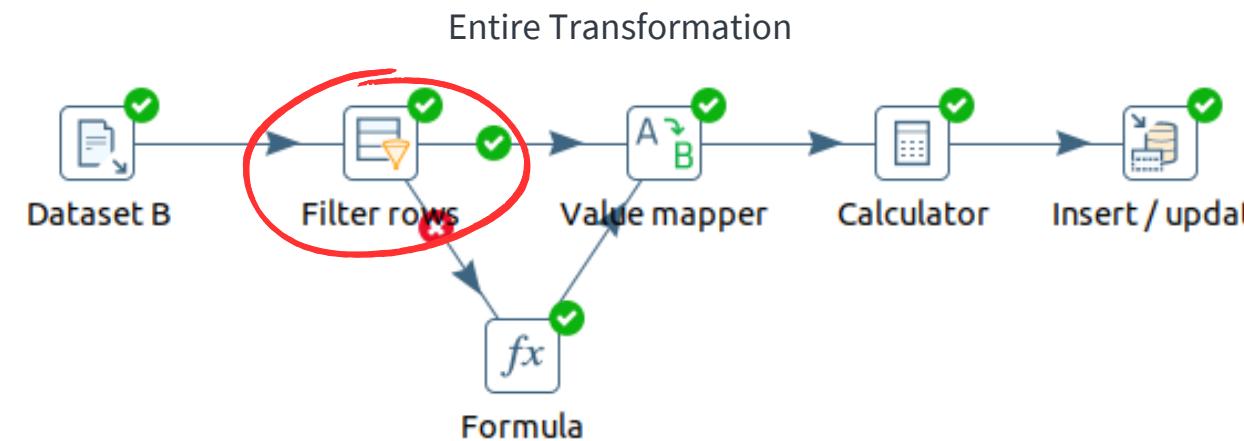
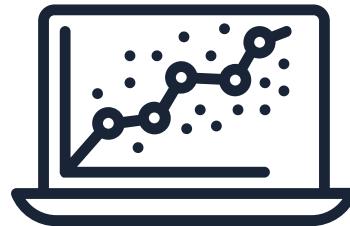
Examine preview data

Rows of step: Dataset B (10000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VISEU	Resende	CARQUEIRA	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATIA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133

Close Stop Get more rows

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Dimension definition

```

CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
  
```

Filter rows Configuration Window

**Filter rows**

Step name: **Filter rows**

Send 'true' data to step: **Insert / update**

Send 'false' data to step: **Formula**

The condition:

**DistrictMunicipalityParishCode** IS NOT NULL

**OK** **Cancel**

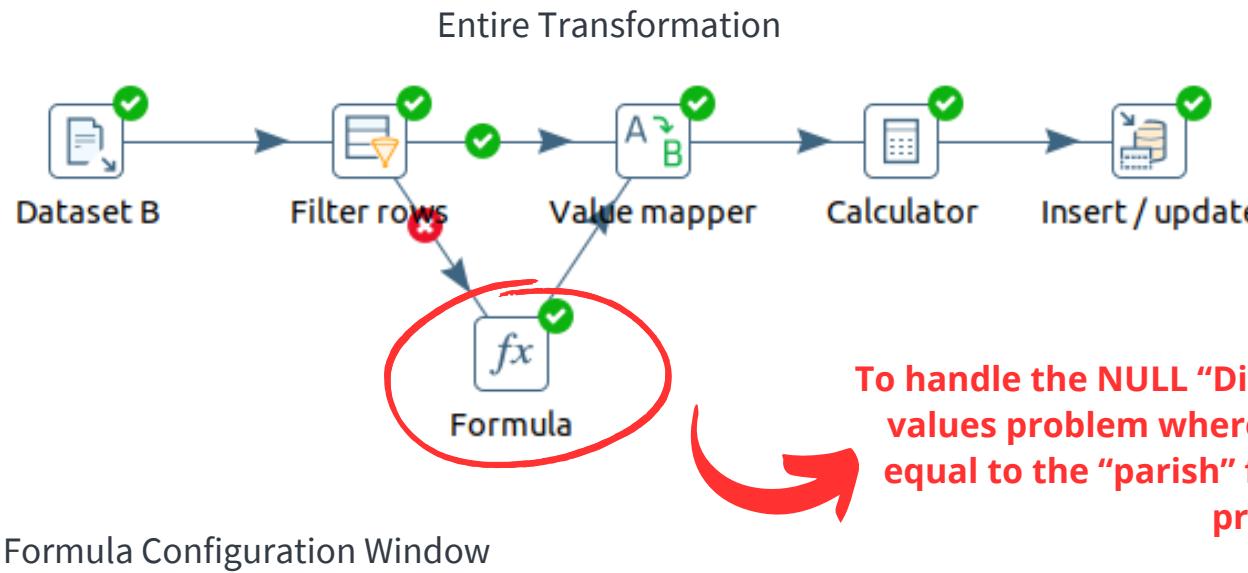
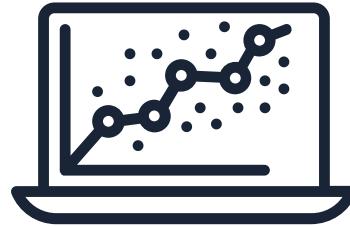
Filter rows Preview

Rows of step: Filter rows (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRA	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIWA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156
18	2022	6	2022-06	LEIRIA	Ansiao	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133
35	2022	9	2022-09	BRAGA	Fafe	ARMIL	Baixa Tensão	73635.213	3	307	030705
36	2022	9	2022-09	COIMBRA	Coimbra	SAO SILVESTRE	Baixa Tensão	217019.542	6	602	060224

Separate the rows with non NULL "DistrictMunicipalityParishCode" values to get them ready to insert in our dim\_location

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Formula Configuration Window

Step name **Formula**

Fields:	New field	Formula	Value type	Length	Precision	Replace value
	DistrictMunicipalityParishCode	CONCATENATE("DP-";[Municipality])	String			DistrictMunicipalityParishCode

OK Cancel

- "DistrictMunicipalityParishCode" values are our primary key **LOCATION\_ID**
- Replace NULL "DistrictMunicipalityParishCode" values with values in format : **DP-Municipality**, where DP = Data Protection

Dimension definition

```

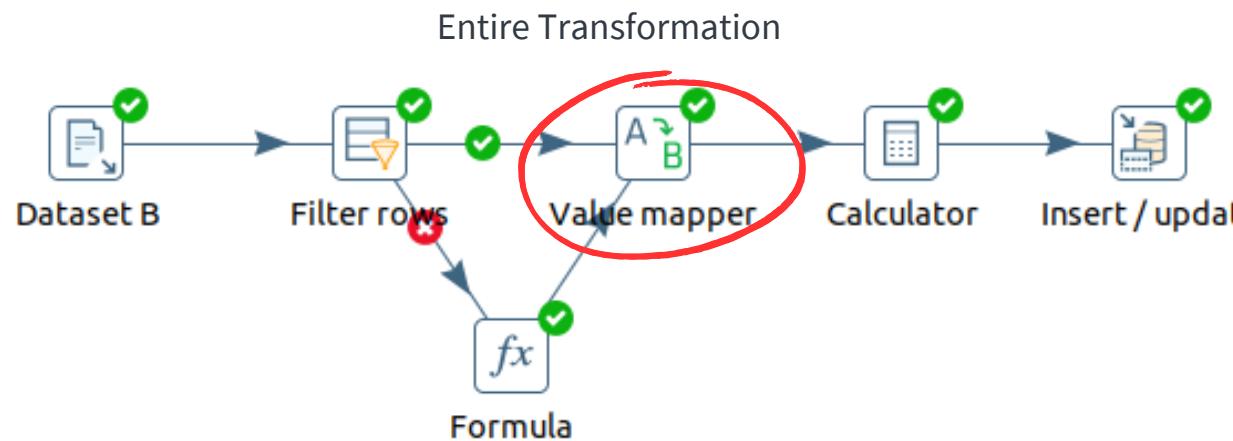
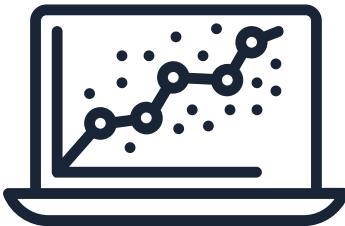
CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
    
```

Formula Preview

Rows of step: Formula (502 rows)											
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2022	10	2022-10	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	26063215.115	2	205	DP-Beja
2	2023	7	2023-07	SANTAREM	Santarém	SANTAREM	Muito Alta, Alta e Média Tensões	309404	14	1416	DP-Santarém
3	2022	4	2022-04	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1362003.362	6	603	DP-Coimbra
4	2021	8	2021-08	BRAGA	Braga	BRAGA	Baixa Tensão	8475.189	3	303	DP-Braga
5	2022	9	2022-09	EVORA	Évora	EVORA	Muito Alta, Alta e Média Tensões	155809.452	7	705	DP-Évora
6	2022	2	2022-02	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	1165957.062	1	105	DP-Aveiro
7	2022	9	2022-09	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	141537.433	11	1106	DP-Lisboa
8	2021	2	2021-02	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	21384335.482	2	205	DP-Beja
9	2021	4	2021-04	BRAGA	Braga	BRAGA	Muito Alta, Alta e Média Tensões	4547294.017	3	303	DP-Braga
10	2022	1	2022-01	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1823917.481	13	1312	DP-Porto
11	2023	5	2023-05	SANTAREM	Santarém	SANTAREM	Muito Alta, Alta e Média Tensões	566345.445	14	1416	DP-Santarém
12	2021	12	2021-12	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1687356.898	13	1312	DP-Porto
13	2021	4	2021-04	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1407654.581	6	603	DP-Coimbra
14	2021	2	2021-02	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	557907.697	1	105	DP-Aveiro
15	2021	11	2021-11	PORTALEGRE	Portalegre	PORTALEGRE	Muito Alta, Alta e Média Tensões	125079.866	12	1214	DP-Portalegre
16	2022	10	2022-10	BRAGA	Braga	BRAGA	Baixa Tensão	9232.9	3	303	DP-Braga
17	2023	4	2023-04	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	27644395.719	2	205	DP-Beja
18	2023	1	2023-01	COIMBRA	Coimbra	COIMBRA	Muito Alta, Alta e Média Tensões	1555889.967	6	603	DP-Coimbra
19	2020	11	2020-11	SETUBAL	Setúbal	SETUBAL	Muito Alta, Alta e Média Tensões	1794380.955	15	1512	DP-Setúbal
20	2021	8	2021-08	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	155161.606	11	1106	DP-Lisboa
21	2022	12	2022-12	LISBOA	Lisboa	LISBOA	Muito Alta, Alta e Média Tensões	145345.5	11	1106	DP-Lisboa
22	2021	12	2021-12	PORTALEGRE	Portalegre	PORTALEGRE	Muito Alta, Alta e Média Tensões	157235.368	12	1214	DP-Portalegre
23	2023	7	2023-07	BRAGA	Braga	BRAGA	Baixa Tensão	3644	3	303	DP-Braga
24	2022	4	2022-04	BRAGA	Braga	BRAGA	Baixa Tensão	8989.848	3	303	DP-Braga
25	2021	2	2021-02	PORTO	Porto	PORTO	Muito Alta, Alta e Média Tensões	1694402.287	13	1312	DP-Porto
26	2022	7	2022-07	AVEIRO	Aveiro	AVEIRO	Muito Alta, Alta e Média Tensões	1344615.147	1	105	DP-Aveiro
27	2020	11	2020-11	LEIRIA	Leiria	LEIRIA	Muito Alta, Alta e Média Tensões	443450.57	10	1009	DP-Leiria
28	2021	8	2021-08	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	426451.09	2	205	DP-Beja
29	2022	5	2022-05	FARO	Faro	FARO	Muito Alta, Alta e Média Tensões	195960.189	8	805	DP-Faro
30	2023	7	2023-07	SETUBAL	Setúbal	SETUBAL	Muito Alta, Alta e Média Tensões	1399553	15	1512	DP-Setúbal
31	2022	4	2022-04	BRAGANCA	Bragança	BRAGANCA	Muito Alta, Alta e Média Tensões	1680728.238	4	402	DP-Bragança
32	2021	9	2021-09	BEJA	Beja	BEJA	Muito Alta, Alta e Média Tensões	22895422.523	2	205	DP-Beja
33	2022	7	2022-07	BRAGA	Braga	BRAGA	Muito Alta, Alta e Média Tensões	4221354.331	3	303	DP-Braga
34	2020	12	2020-12	EVORA	Évora	EVORA	Muito Alta, Alta e Média Tensões	64812.502	7	705	DP-Évora
35	2023	4	2023-04	BRAGANCA	Braganca	BRAGANCA	Muito Alta, Alta e Média Tensões	1491579.71	4	402	DP-Bragança

We didn't want to lose data!

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Dimension definition

```
CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
```

Value mapper Configuration Window

**Value mapper**

Step name: Value mapper

Fieldname to use: Voltage level

Target field name (empty=overwrite): VOLTAGE\_ID

Default upon non-matching:

Source value	Target value
Baixa Tensão	1
Muito Alta, Alta e Média Tensões	2

OK Cancel

Value mapper Preview

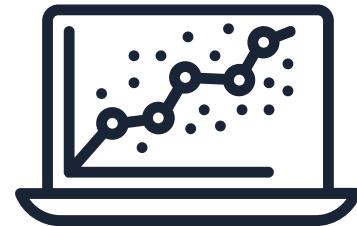
Rows of step: Value mapper (1000 rows)

Examine preview data

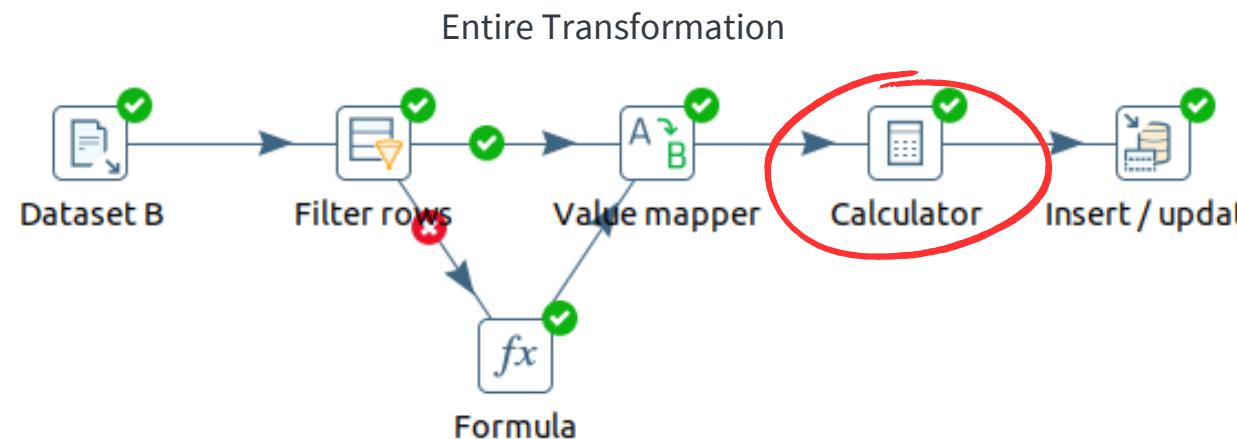
	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	VOLTAGE_ID
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	1
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	1
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEIXE	Muito Alta, Alta e Média Tensões	114819.863	8	805	080503	2
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	Baixa Tensão	34461.911	9	911	091144	1
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Tensões	666877.841	13	1316	131612	2
6	2022	3	2022-03	VIANA DO CASTELO	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	1
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	1
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Tensões	471518.525	10	1001	100102	2
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Tensões	422717.641	10	1009	100909	2
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Tensões	903593.147	1	102	010204	2
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	1
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	Muito Alta, Alta e Média Tensões	1057184.011	7	705	070526	2
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	1
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	1
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	1
16	2022	5	2022-05	SETUBAL	Palmeira	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	1
17	2022	5	2022-05	VIANA DO CASTELO	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	1
18	2022	6	2022-06	LEIRIA	Ansiao	CHAO DE COUCE	Muito Alta, Alta e Média Tensões	591322.42	10	1003	100304	2
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	1
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Tensões	4341502.722	14	1409	140909	2
21	2022	6	2022-06	VIANA DO CASTELO	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	1
22	2022	6	2022-06	VILA REAL	Santa Marta de Penaguião	CUMIEIRA	Muito Alta, Alta e Média Tensões	40879.84	17	1711	171102	2
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Tensões	828976.474	18	1805	180525	2
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303	1
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	1
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Tensões	1088142.387	10	1011	101102	2
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Tensões	1680417.491	13	1317	131731	2
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Tensões	77169.225	14	1414	141416	2
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	1
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	1
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Tensões	67492.802	1	115	011507	2
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Tensões	141715.613	4	408	040802	2
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	Baixa Tensão	64069.193	9	911	091118	1
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	1

Close Stop Get more rows

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Convert the Date field  
(string) to date type  
and the VOLTAGE\_ID  
to int



Dimension definition

```

CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
  
```

Calculator Configuration Window

Calculator

Step name	Calculator								
<input checked="" type="checkbox"/> Throw an error on non existing files									
Fields:									
New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask
1 VOLTAGE_ID_INT	Create a copy of field A	VOLTAGE_ID			Integer		N		
2 dateConverted	Create a copy of field A	Date			Date		N		yyyy-MM

OK Cancel

Calculator Preview

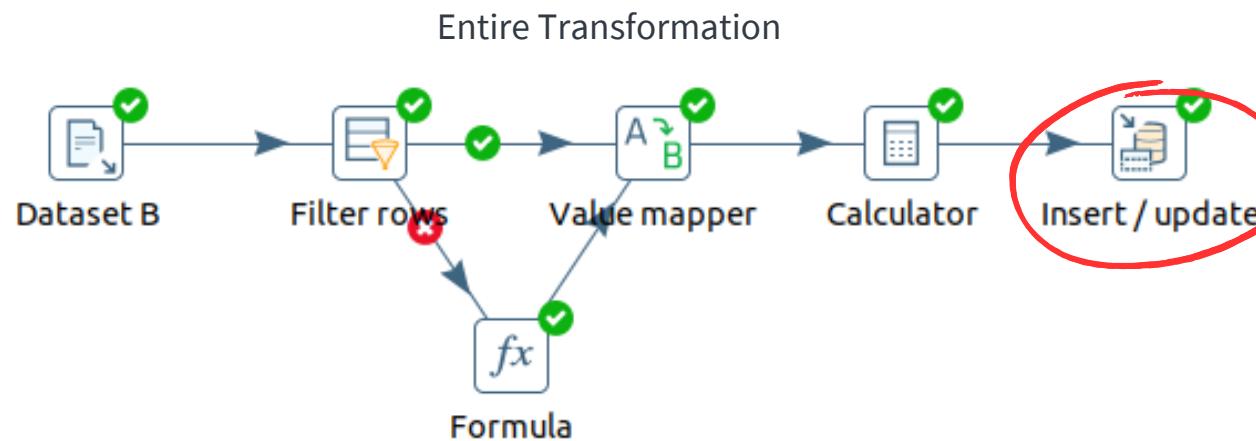
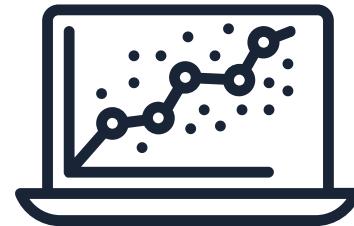
Rows of step: Calculator (1000 rows)

Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParish	VOLTAGE_ID	VOLTAGE_ID_INT	dateConverted
1	2022	3	2022-03	AVEIRO	Estarreja	Baixa Tensão	389752.454	1	108	010809	1	1	2022-03
2	2022	3	2022-03	BRAGA	Vieira do Minho	Baixa Tensão	583541.604	3	311	031120	1	1	2022-03
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	114819.863	8	805	080503	2	2	2022-03
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUAS	34461.911	9	911	091144	1	1	2022-03
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	666877.841	13	1316	131612	2	2	2022-03
6	2022	3	2022-03	VIANA DO CASTE	Ponte da Barca	UF TOUVEDO	50632.677	16	1606	160629	1	1	2022-03
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	21363.331	9	913	091315	1	1	2022-04
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEBIZERAO	471518.525	10	1001	100102	2	2	2022-04
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	422717.641	10	1009	100909	2	2	2022-04
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	903593.147	1	102	010204	2	2	2022-05
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIJA PEDORIDO E PARAISO	520199.687	1	106	010610	1	1	2022-05
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIRO	1057184.011	7	705	070526	2	2	2022-05
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	516207.244	7	713	071302	1	1	2022-05
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	1254374.575	11	1111	111115	1	1	2022-05
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	204595.462	13	1310	131001	1	1	2022-05
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	2759588.721	15	1508	150804	1	1	2022-05
17	2022	5	2022-05	VIANA DO CASTE	Arcos de Valdevez	UF GRADE E CARRALCOVA	53644.213	16	1601	160156	1	1	2022-05
18	2022	6	2022-06	LEIRIA	Ansiao	CHAO DE COUCE	591322.42	10	1003	100304	2	2	2022-06
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	134028.104	10	1008	100801	1	1	2022-06
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	4341502.722	14	1409	140909	2	2	2022-06
21	2022	6	2022-06	VIANA DO CASTE	Ponte de Lima	REFOIOS DO LIMA	269518.952	16	1607	160737	1	1	2022-06
22	2022	6	2022-06	VILA REAL	Santa Marta de Penag	CUMIEIRA	40879.84	17	1711	171102	2	2	2022-06
23	2022	6	2022-06	VIDEU	Lamego	LAMEGO (ALMACAVE E SE)	828976.474	18	1805	180525	2	2	2022-06
24	2022	6	2022-06	VIDEU	Resende	CARQUERE	80697.018	18	1813	181303	1	1	2022-06
25	2022	6	2022-06	VIDEU	Tondela	MOLELOS	299236.255	18	1821	182111	1	1	2022-06
26	2022	7	2022-07	LEIRIA	Nazare	NAZARE	1088142.387	10	1011	101102	2	2	2022-07
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	1680417.491	13	1317	131731	2	2	2022-07
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	77169.225	14	1414	141416	2	2	2022-07
29	2022	7	2022-07	VIDEU	Lamego	SAMODAES	15388.178	18	1805	180519	1	1	2022-07
30	2022	7	2022-07	VIDEU	Viseu	CAVERNAES	209277.865	18	1823	182307	1	1	2022-07
31	2022	8	2022-08	AVEIRO	Ovar	VALEG	67492.802	1	115	011507	2	2	2022-08
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	141715.613	4	408	040802	2	2	2022-08
33	2022	8	2022-08	GUARDA	Sabugal	MALCATA	64069.193	9	911	091118	1	1	2022-08
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	255718.426	13	1311	131133	1	1	2022-08

Examine preview data

Close Stop Get more rows

# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



Dimension definition

```

CREATE TABLE fact_consumption (
    TIME_ID DATETIME,
    LOCATION_ID VARCHAR(255),
    VOLTAGE_ID INT,
    ENERGY_CONSUMPTION DOUBLE,
    PRIMARY KEY (VOLTAGE_ID, LOCATION_ID, TIME_ID),
    FOREIGN KEY (VOLTAGE_ID) REFERENCES dim_voltage (VOLTAGE_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID)
);
    
```

Insert / update Configuration Window

Insert / update

Step name: **Insert / update**

Connection: datasetb\_dw

Target schema: datasetb\_dw

Target table: fact\_consumption

Commit size: 100

Don't perform any updates:

The key(s) to look up the value(s):

Table field	Comparator	Stream field1	Stream field2
1 LOCATION_ID	=	DistrictMunicipalityParishCode	
2 VOLTAGE_ID	=	VOLTAGE_ID_INT	
3 TIME_ID	=	dateConverted	

Update fields:

Table field	Stream field	Update
1 ENERGY_CONSUMPTION	Active Energy (kWh)	Y
2 LOCATION_ID	DistrictMunicipalityParishCode	Y
3 VOLTAGE_ID	VOLTAGE_ID_INT	Y
4 TIME_ID	dateConverted	Y

Get fields | Get update fields | Edit mapping | OK | Cancel | SQL | ? Help | Close | Stop | Get more rows

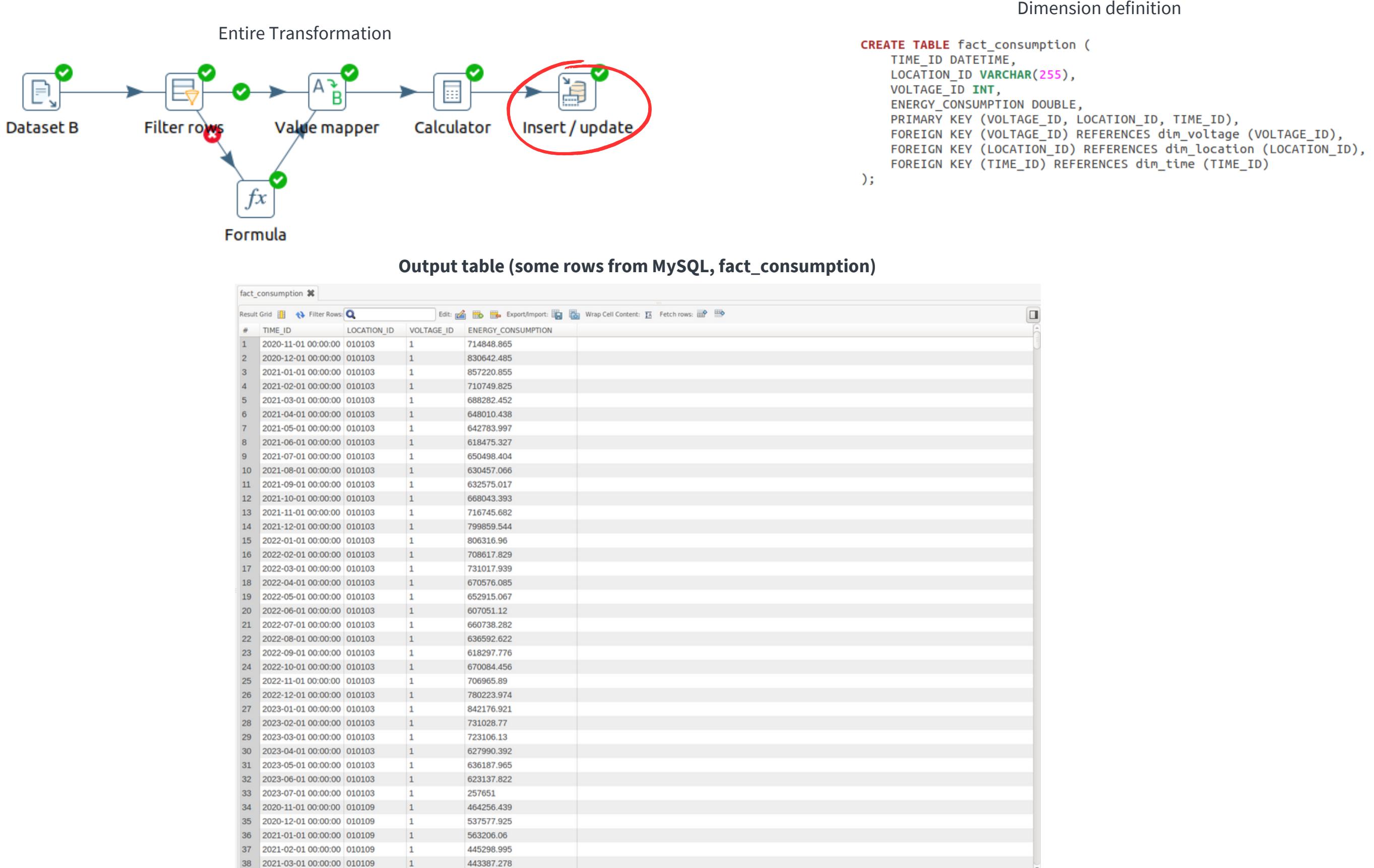
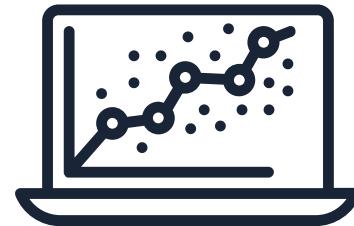
Insert / update Preview

Rows of step: Insert / update (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kW)	DistrictCode	DistrictMunicipalityCo	DistrictMunicipalityParishCo	VOLTAGE_ID	VOLTAGE_ID_IN	dateConverted
1	2022	3	2022-03	AVEIRO	Estarreja	UF CANELAS E FERMELA	Baixa Tensão	389752.454	1	108	010809	1	1	2022-03
2	2022	3	2022-03	BRAGA	Vieira do Minho	VIEIRA DO MINHO	Baixa Tensão	583541.604	3	311	031120	1	1	2022-03
3	2022	3	2022-03	FARO	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Ten	114819.863	8	805	080503	2	2	2022-03
4	2022	3	2022-03	GUARDA	Sabugal	UF RUVINA RUIVOS VALE EGUA	Baixa Tensão	34461.911	9	911	091144	1	1	2022-03
5	2022	3	2022-03	PORTO	Vila do Conde	JUNQUEIRA	Muito Alta, Alta e Média Ten	666877.841	13	1316	131612	2	2	2022-03
6	2022	3	2022-03	VIANA DO CASTE	Ponte da Barca	UF TOUVEDO	Baixa Tensão	50632.677	16	1606	160629	1	1	2022-03
7	2022	4	2022-04	GUARDA	Trancoso	REBOLEIRO	Baixa Tensão	21363.331	9	913	091315	1	1	2022-04
8	2022	4	2022-04	LEIRIA	Alcobaça	ALFEIZERAO	Muito Alta, Alta e Média Ten	471518.525	10	1001	100102	2	2	2022-04
9	2022	4	2022-04	LEIRIA	Leiria	COIMBRAO	Muito Alta, Alta e Média Ten	422717.641	10	1009	100909	2	2	2022-04
10	2022	5	2022-05	AVEIRO	Albergaria-a-Velha	BRANCA	Muito Alta, Alta e Média Ten	903593.147	1	102	010204	2	2	2022-05
11	2022	5	2022-05	AVEIRO	Castelo de Paiva	UF RAIVA PEDORIDO E PARAISO	Baixa Tensão	520199.687	1	106	010610	1	1	2022-05
12	2022	5	2022-05	EVORA	Évora	UF S MANCOS S VICENTE PIGEIR	Muito Alta, Alta e Média Ten	1057184.011	7	705	070526	2	2	2022-05
13	2022	5	2022-05	EVORA	Viana do Alentejo	VIANA DO ALENTEJO	Baixa Tensão	516207.244	7	713	071302	1	1	2022-05
14	2022	5	2022-05	LISBOA	Sintra	CASAL DE CAMBRA	Baixa Tensão	1254374.575	11	1111	111115	1	1	2022-05
15	2022	5	2022-05	PORTO	Paredes	AGUIAR DE SOUSA	Baixa Tensão	204595.462	13	1310	131001	1	1	2022-05
16	2022	5	2022-05	SETUBAL	Palmela	QUINTA DO ANJO	Baixa Tensão	2759588.721	15	1508	150804	1	1	2022-05
17	2022	5	2022-05	VIANA DO CASTE	Arcos de Valdevez	UF GRADE E CARRALCOVA	Baixa Tensão	53644.213	16	1601	160156	1	1	2022-05
18	2022	6	2022-06	LEIRIA	Ansião	CHAO DE COUCE	Muito Alta, Alta e Média Ten	591322.42	10	1003	100304	2	2	2022-06
19	2022	6	2022-06	LEIRIA	Figueiró dos Vinhos	AGUDA	Baixa Tensão	134028.104	10	1008	100801	1	1	2022-06
20	2022	6	2022-06	SANTAREM	Coruche	UF CORUCHE FAJARDA EERRA	Muito Alta, Alta e Média Ten	4341502.722	14	1409	140909	2	2	2022-06
21	2022	6	2022-06	VIANA DO CASTE	Ponte de Lima	REFOIOS DO LIMA	Baixa Tensão	269518.952	16	1607	160737	1	1	2022-06
22	2022	6	2022-06	VILA REAL	Santa Marta de Penagui	CUMIEIRA	Muito Alta, Alta e Média Ten	40879.84	17	1711	171102	2	2	2022-06
23	2022	6	2022-06	VISEU	Lamego	LAMEGO (ALMACAVE E SE)	Muito Alta, Alta e Média Ten	828976.474	18	1805	180525	2	2	2022-06
24	2022	6	2022-06	VISEU	Resende	CARQUERE	Baixa Tensão	80697.018	18	1813	181303	1	1	2022-06
25	2022	6	2022-06	VISEU	Tondela	MOLELOS	Baixa Tensão	299236.255	18	1821	182111	1	1	2022-06
26	2022	7	2022-07	LEIRIA	Nazaré	NAZARE	Muito Alta, Alta e Média Ten	1088142.387	10	1011	101102	2	2	2022-07
27	2022	7	2022-07	PORTO	Vila Nova de Gaia	UF SERZEDO E PEROSINHO	Muito Alta, Alta e Média Ten	1680417.491	13	1317	131731	2	2	2022-07
28	2022	7	2022-07	SANTAREM	Rio Maior	UF MARMELEIRA E ASSENTIZ	Muito Alta, Alta e Média Ten	77169.225	14	1414	141416	2	2	2022-07
29	2022	7	2022-07	VISEU	Lamego	SAMODAES	Baixa Tensão	15388.178	18	1805	180519	1	1	2022-07
30	2022	7	2022-07	VISEU	Viseu	CAVERNAES	Baixa Tensão	209277.865	18	1823	182307	1	1	2022-07
31	2022	8	2022-08	AVEIRO	Ovar	VALEGA	Muito Alta, Alta e Média Ten	67492.802	1	115	011507	2	2	2022-08
32	2022	8	2022-08	BRAGANCA	Mogadouro	BEMPOSTA	Muito Alta, Alta e Média Ten	141715.613	4	408	040802	2	2	2022-08
33	2022	8	2022-08	GUARDA	Sabugal	MALCATTA	Baixa Tensão	64069.193	9	911	091118	1	1	2022-08
34	2022	8	2022-08	PORTO	Penafiel	RECEZINHOS (SAO MARTINHO)	Baixa Tensão	255718.426	13	1311	131133	1	1	2022-08

Examine preview data | Close | Stop | Get more rows

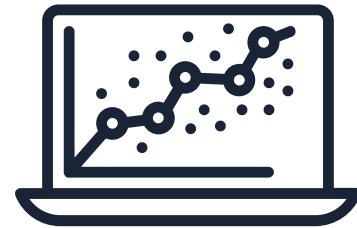
# 6. TRANSFORMATION TO POPULATE THE FACT TABLE



# TASK 7

OLAP CUBE DEFINITION BASED ON THE PREVIOUS  
DATA WAREHOUSE

# 7. OLAP CUBE DEFINITION BASED ON THE PREVIOUS DATA WAREHOUSE



The OLAP Cube definition as well as the corresponding XML file, created with the Pentaho Schema Workbench, are shown below:

OLAP Cube schema

The screenshot shows the Pentaho Schema Workbench interface with the title bar "Schema - datasetb\_dw (datasetb.xml)". The toolbar contains icons for various schema elements. The main pane displays a hierarchical tree of schema components:

- Schema**: A folder icon.
- Consumptions**: A cube icon.
  - Table: fact\_consumption**: A grid icon.
  - Location**: A location pin icon.
    - Location Hierarchy**: A hierarchy icon.
      - District**: A building icon.
      - Municipality**: A building icon.
      - Parish**: A building icon.
    - Table: dim\_location**: A grid icon.
  - Voltage**: A voltage icon.
    - Voltage Hierarchy**: A hierarchy icon.
      - Voltage Level**: A building icon.
      - Table: dim\_voltage**: A grid icon.
  - Time**: A time icon.
    - Time Hierarchy**: A hierarchy icon.
      - Year**: A building icon.
      - Month**: A building icon.
    - Table: dim\_time**: A grid icon.
  - Energy Consumption**: A chart icon.

XML file

```
<Schema name="datasetb_dw">
  <Cube name="Consumptions" visible="true" cache="true" enabled="true">
    <Table name="fact_consumption"> </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="LOCATION_ID" name="Location">
      <Hierarchy name="Location Hierarchy" visible="true" hasAll="true" allMemberName="All Locations" primaryKey="LOCATION_ID">
        <Table name="dim_location"> </Table>
        <Level name="District" visible="true" column="DISTRICT_NAME" type="String" uniqueMembers="false" levelType="Regular"> </Level>
        <Level name="Municipality" visible="true" column="MUNICIPALITY_NAME" type="String" uniqueMembers="false" levelType="Regular"> </Level>
        <Level name="Parish" visible="true" column="PARISH_NAME" type="String" uniqueMembers="false" levelType="Regular"> </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="VOLTAGE_ID" name="Voltage">
      <Hierarchy name="Voltage Hierarchy" visible="true" hasAll="true" allMemberName="All Voltages" primaryKey="VOLTAGE_ID">
        <Table name="dim_voltage"> </Table>
        <Level name="Voltage Level" visible="true" column="VOLTAGELEVEL" type="String" uniqueMembers="false" levelType="Regular"> </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="TimeDimension" visible="true" foreignKey="TIME_ID" name="Time">
      <Hierarchy name="Time Hierarchy" visible="true" hasAll="true" allMemberName="All Years" primaryKey="TIME_ID">
        <Table name="dim_time"> </Table>
        <Level name="Year" visible="true" column="YEAR_ID" type="Integer" uniqueMembers="false" levelType="TimeYears"> </Level>
        <Level name="Month" visible="true" column="MONTH_ID" type="Integer" uniqueMembers="false" levelType="TimeMonths"> </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="Energy Consumption" column="ENERGY_CONSUMPTION" datatype="Numeric" aggregator="sum" visible="true"> </Measure>
  </Cube>
</Schema>
```

# TASK 8

DATA CUBE ANALYSIS

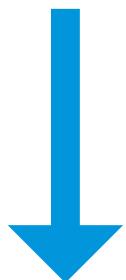
## 8. DATA CUBE ANALYSIS



Analysis:

with Saiku Analytics

**Energy consumption by district and year, with complete yearly data**



Energy consumption variation per district  
from one year to the next



Relation between the energy consumption  
variation and the voltage level

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



- 1) Check the years with complete data (from January to December)

Query

```
1 SELECT
2     [Measures].[Energy Consumption] ON COLUMNS,
3     Hierarchize({[Time.Time Hierarchy].[Year].Members, [Time.Time Hierarchy].[Month].Members}) ON ROWS
4 FROM [Consumptions]
5
```

- This query shows the energy consumption by years and respective months, which make it possible to conclude which years are complete.
- The query **result** is shown in the next slide.

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



Result		
Year	Month	Energy Consumption
2020		7,737,650,934.718
	11	3,784,273,853.729
	12	3,953,377,080.989
2021		44,587,075,708.57
	1	4,162,390,715.635
	2	3,524,177,902.769
	3	3,665,941,542.609
	4	3,439,014,139.886
	5	3,625,989,446.056
	6	3,564,464,319.096
	7	3,751,698,512.408
	8	3,648,692,938.926
	9	3,688,731,004.21
	10	3,753,118,919.856
	11	3,833,051,400.197
	12	3,929,804,866.922
2022		45,557,235,463.158
	1	4,062,405,055.189
	2	3,674,352,071.875
	3	3,927,421,094.605
	4	3,643,103,818.613
	5	3,761,850,586.832
	6	3,686,539,550.318
	7	3,994,944,076.363
	8	3,715,173,383.658
	9	3,663,044,396.387
	10	3,783,139,293.508
	11	3,776,152,593.106
	12	3,869,109,542.704
2023		23,864,616,531.698
	1	4,177,060,870.284
	2	3,740,556,142.862
	3	3,902,594,793.458
	4	3,519,701,021.643
	5	3,656,531,121.926
	6	3,663,127,952.525
	7	1,205,044,629

} 2 months

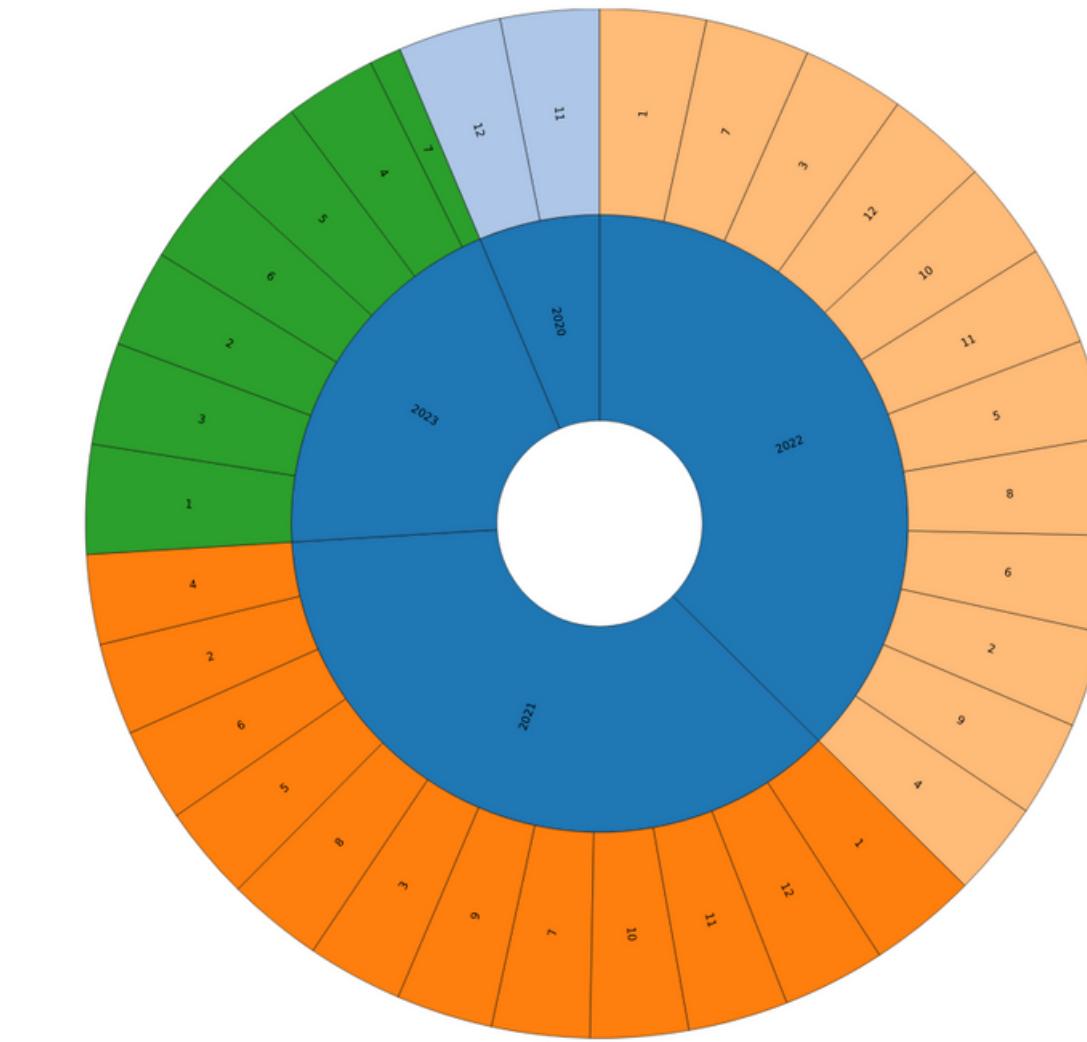
} 12 months

} 12 months

} 7 months

2021  
2022

Also, with Saiku sunburst chart:



- Years with complete data: **2021, 2022**

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



### 2) Energy consumption variation per district from one year to the next

#### Query

```
1 SELECT
2   CROSSJOIN({[Time.Time Hierarchy].[Year].[2021], [Time.Time Hierarchy].[Year].[2022]}, Measures.[Energy Consumption]) ON COLUMNS
3   [Location.Location Hierarchy].[District].Members ON ROWS
4 FROM [Consumptions]
5
```

- This query shows the energy consumption by district and complete years found in the previous analysis, 2021 and 2022.
- Query **result** is in the next slide, as well as the **energy consumption variation** studied with Saiku Spark Line on the query result.

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



With the **spark line** applied to the **query result** it is possible to study the variation from one year to the other.

Result			
District	2021	2022	
	Energy Consumption	Energy Consumption	
AVEIRO	4,381,127,458	4,351,723,261.454	/
BEJA	1,087,511,836	1,136,176,040	/
BRAGA	3,407,041,095	3,454,781,865.22	/
BRAGANCA	379,473,577	388,817,972	/
CASTELO BRANCO	771,822,945.85	816,148,419.15	/
COIMBRA	2,689,109,688	2,612,783,627	/
EVORA	712,514,243	780,338,813	/
FARO	2,166,312,114	2,367,894,146.31	/
GUARDA	467,885,614	482,750,861	/
LEIRIA	2,517,152,013	2,553,090,826.97	/
LISBOA	8,270,141,703	8,670,500,520.346	/
PORCALEGRE	468,289,813	482,877,968.406	/
PORTO	7,314,575,940.72	7,386,737,067.874	/
SANTAREM	2,161,810,637	2,158,709,002.642	/
SETUBAL	4,776,806,313	4,853,113,930.288	/
VIANA DO CASTELO	1,083,665,399	1,098,441,193.29	/
VILA REAL	547,451,452	572,536,374.612	/
VISEU	1,384,383,867	1,389,813,573.596	/

### Variation per District

Energy consumption **DECREASED** from 2021 to 2022:

- Aveiro
- Coimbra
- Santarém



Energy consumption **INCREASED** from 2021 to 2022:

- |                  |              |                    |
|------------------|--------------|--------------------|
| • Beja           | • Guarda     | • Viana do Castelo |
| • Braga          | • Leiria     | • Vila Real        |
| • Bragança       | • Lisboa     | • Viseu            |
| • Castelo Branco | • Portalegre |                    |
| • Évora          | • Porto      |                    |
| • Faro           | • Setúbal    |                    |



## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



- 3) Energy consumption variation for each voltage level by district and from one year to the next

### Query

```
1 SELECT
2   CROSSJOIN({[Time.Time Hierarchy].[Year].[2021], [Time.Time Hierarchy].[Year].[2022]}, Measures.[Energy Consumption]) ON COLUMNS,
3   CROSSJOIN([Location.Location Hierarchy].[District].Members, [Voltage.Voltage Hierarchy].[Voltage Level].Members) ON ROWS
4 FROM [Consumptions]
5 |
```

- This query shows the energy consumption by voltage level, district, and complete years, 2021 and 2022.
- The query **result** is show in the next slide, as well as the **energy consumption variation by voltage level** studied with Saiku Spark Line on the query result.

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



With the **spark line** applied to the **query result** it is possible to study the variation by voltage level from one year to the other.

Result		2021	2022	
District	Voltage Level	Energy Consumption	Energy Consumption	
AVEIRO	Baixa Tensão	1,369,144,179	1,376,191,990.334	/
	Muito Alta, Alta e Média Tensões	3,011,983,279	2,975,531,271.12	\
BEJA	Baixa Tensão	350,164,592	352,540,358	/
	Muito Alta, Alta e Média Tensões	737,347,244	783,635,682	/
BRAGA	Baixa Tensão	1,597,686,548	1,637,806,521.22	/
	Muito Alta, Alta e Média Tensões	1,809,354,547	1,816,975,344	/
BRAGANCA	Baixa Tensão	294,537,617	301,887,725	/
	Muito Alta, Alta e Média Tensões	84,935,960	86,930,247	/
CASTELO BRANCO	Baixa Tensão	417,897,687.85	433,234,808.15	/
	Muito Alta, Alta e Média Tensões	353,925,258	382,913,611	/
COIMBRA	Baixa Tensão	860,819,448	871,775,371	/
	Muito Alta, Alta e Média Tensões	1,828,290,240	1,741,008,256	\
EVORA	Baixa Tensão	389,105,397	395,550,362	/
	Muito Alta, Alta e Média Tensões	323,408,846	384,788,451	/
FARO	Baixa Tensão	1,593,012,633	1,714,950,018.31	/
	Muito Alta, Alta e Média Tensões	573,299,481	652,944,128	/
GUARDA	Baixa Tensão	310,149,733	315,599,290	/
	Muito Alta, Alta e Média Tensões	157,735,881	167,151,571	/
LEIRIA	Baixa Tensão	1,057,089,736	1,077,604,176.97	/
	Muito Alta, Alta e Média Tensões	1,460,062,277	1,475,486,650	/
LISBOA	Baixa Tensão	4,658,049,308	4,809,788,973.346	/
	Muito Alta, Alta e Média Tensões	3,612,092,395	3,860,711,547	/
PORTALEGRE	Baixa Tensão	269,608,128	271,239,807.406	/
	Muito Alta, Alta e Média Tensões	198,681,685	211,638,161	/
PORTO	Baixa Tensão	3,784,636,658.72	3,825,691,624.874	/
	Muito Alta, Alta e Média Tensões	3,529,939,282	3,561,045,443	/
SANTAREM	Baixa Tensão	941,982,822	962,584,317.642	/
	Muito Alta, Alta e Média Tensões	1,219,827,815	1,196,124,685	\
SETUBAL	Baixa Tensão	1,649,722,208	1,663,463,155.288	/
	Muito Alta, Alta e Média Tensões	3,127,084,105	3,189,650,775	/
VIANA DO CASTELO	Baixa Tensão	496,557,035	513,075,020.29	/
	Muito Alta, Alta e Média Tensões	587,108,364	585,366,173	\
VILA REAL	Baixa Tensão	380,659,744	392,108,689.612	/
	Muito Alta, Alta e Média Tensões	166,791,708	180,427,685	/
VISEU	Baixa Tensão	688,262,055	708,253,657.596	/
	Muito Alta, Alta e Média Tensões	696,121,812	681,559,916	\

**NEXT STEP:** compare the result of this analysis with the results of the previous one, in 2), to find whether such increase/decrease is due to an increase/decrease at a low or at a high voltage level.

# 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



Result from 3)

District	Voltage Level	2021	2022	
		Energy Consumption	Energy Consumption	
AVEIRO	Baixa Tensão	1,369,144,179	1,376,191,990.334	/
	Muito Alta, Alta e Média Tensões	3,011,983,279	2,975,531,271.12	\
BEJA	Baixa Tensão	350,164,592	352,540,358	/
	Muito Alta, Alta e Média Tensões	737,347,244	783,635,682	/
BRAGA	Baixa Tensão	1,597,686,548	1,637,806,521.22	/
	Muito Alta, Alta e Média Tensões	1,809,354,547	1,816,975,344	/
BRAGANCA	Baixa Tensão	294,537,617	301,887,725	/
	Muito Alta, Alta e Média Tensões	84,935,960	86,930,247	/
CASTELO BRANCO	Baixa Tensão	417,897,687.85	433,234,808.15	/
	Muito Alta, Alta e Média Tensões	353,925,258	382,913,611	/
COIMBRA	Baixa Tensão	860,819,448	871,775,371	/
	Muito Alta, Alta e Média Tensões	1,828,290,240	1,741,008,256	\
EVORA	Baixa Tensão	389,105,397	395,550,362	/
	Muito Alta, Alta e Média Tensões	323,408,846	384,788,451	/
FARO	Baixa Tensão	1,593,012,633	1,714,950,018.31	/
	Muito Alta, Alta e Média Tensões	573,299,481	652,944,128	/
GUARDA	Baixa Tensão	310,149,733	315,599,290	/
	Muito Alta, Alta e Média Tensões	157,735,881	167,151,571	/
LEIRIA	Baixa Tensão	1,057,089,736	1,077,604,176.97	/
	Muito Alta, Alta e Média Tensões	1,460,062,277	1,475,486,650	/
LISBOA	Baixa Tensão	4,658,049,308	4,809,788,973.346	/
	Muito Alta, Alta e Média Tensões	3,612,092,395	3,860,711,547	/
PORTALEGRE	Baixa Tensão	269,608,128	271,239,807.406	/
	Muito Alta, Alta e Média Tensões	198,681,685	211,638,161	/
PORTO	Baixa Tensão	3,784,636,658.72	3,825,691,624.874	/
	Muito Alta, Alta e Média Tensões	3,529,939,282	3,561,045,443	/
SANTAREM	Baixa Tensão	941,982,822	962,584,317.642	/
	Muito Alta, Alta e Média Tensões	1,219,827,815	1,196,124,685	\
SETUBAL	Baixa Tensão	1,649,722,208	1,663,463,155.288	/
	Muito Alta, Alta e Média Tensões	3,127,084,105	3,189,650,775	/
VIANA DO CASTELO	Baixa Tensão	496,557,035	513,075,020.29	/
	Muito Alta, Alta e Média Tensões	587,108,364	585,366,173	\
VILA REAL	Baixa Tensão	380,659,744	392,108,689.612	/
	Muito Alta, Alta e Média Tensões	166,791,708	180,427,685	/
VISEU	Baixa Tensão	688,262,055	708,253,657.596	/
	Muito Alta, Alta e Média Tensões	696,121,812	681,559,916	\

Result from 2)

District	2021	2022	
	Energy Consumption	Energy Consumption	
AVEIRO	4,381,127,458	4,351,723,261.454	\
BEJA	1,087,511,836	1,136,176,040	/
BRAGA	3,407,041,095	3,454,781,865.22	/
BRAGANCA	379,473,577	388,817,972	/
CASTELO BRANCO	771,822,945.85	816,148,419.15	/
COIMBRA	2,689,109,688	2,612,783,627	\
EVORA	712,514,243	780,338,813	/
FARO	2,166,312,114	2,367,894,146.31	/
GUARDA	467,885,614	482,750,861	/
LEIRIA	2,517,152,013	2,553,090,826.97	/
LISBOA	8,270,141,703	8,670,500,520.346	/
PORTALEGRE	468,289,813	482,877,968.406	/
PORTO	7,314,575,940.72	7,386,737,067.874	/
SANTAREM	2,161,810,637	2,158,709,002.642	\
SETUBAL	4,776,806,313	4,853,113,930.288	/
VIANA DO CASTELO	1,083,665,399	1,098,441,193.29	/
VILA REAL	547,451,452	572,536,374.612	/
VISEU	1,384,383,867	1,389,813,573.596	/

Compare



Method and conclusions  
in the next slides

# 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



## Method

Examples for some districts:

AVEIRO	Baixa Tensão	1,369,144,179	1,376,191,990.334	/
	Muito Alta, Alta e Média Tensões	3,011,983,279	2,975,531,271.12	\

AVEIRO		4,381,127,458	4,351,723,261.454	\
--------	--	---------------	-------------------	---

} Low voltage level increased  
**Decrease** is due to the **High voltage level decrease**

BRAGANCA	Baixa Tensão	294,537,617	301,887,725	/
	Muito Alta, Alta e Média Tensões	84,935,960	86,930,247	/

BRAGANCA		379,473,577	388,817,972	/
----------	--	-------------	-------------	---

} Both Low and High voltage level increased  
**Both** Low and High voltage contributed to the **increase**

SANTAREM	Baixa Tensão	941,982,822	962,584,317.642	/
	Muito Alta, Alta e Média Tensões	1,219,827,815	1,196,124,685	\

SANTAREM		2,161,810,637	2,158,709,002.642	\
----------	--	---------------	-------------------	---

} Low voltage level increased  
**Decrease** is due to the **High voltage level decrease**

VIANA DO CASTELO	Baixa Tensão	496,557,035	513,075,020.29	/
	Muito Alta, Alta e Média Tensões	587,108,364	585,366,173	\

VIANA DO CASTELO		1,083,665,399	1,098,441,193.29	/
------------------	--	---------------	------------------	---

} High voltage decreased  
**Increase** is due to the **Low voltage level increase**

- The conclusions are shown in the next slide.

## 8. DATA CUBE ANALYSIS

Energy consumption by district and year, with complete yearly data



### Analysis conclusions: variations by district and voltage level

Energy consumption **DECREASED** from 2021 to 2022:

Due to **High voltage level** decrease:

- Aveiro
- Coimbra
- Santarém

Due to **Low voltage level** decrease:

Due to both **Low and High voltage levels** decrease:

Energy consumption **INCREASED** from 2021 to 2022:

Due to **Low voltage level** increase:

- Viana do Castelo
- Viseu

Due to both **Low and High voltage levels** increase:

- |                  |              |             |
|------------------|--------------|-------------|
| • Beja           | • Faro       | • Porto     |
| • Braga          | • Guarda     | • Setúbal   |
| • Bragança       | • Leiria     | • Vila Real |
| • Castelo Branco | • Lisboa     |             |
| • Évora          | • Portalegre |             |

Due to **High voltage level** increase:

# THANK YOU

---

GONÇALO GALANTE, 96216

SIMÃO FONSECA, 96911