

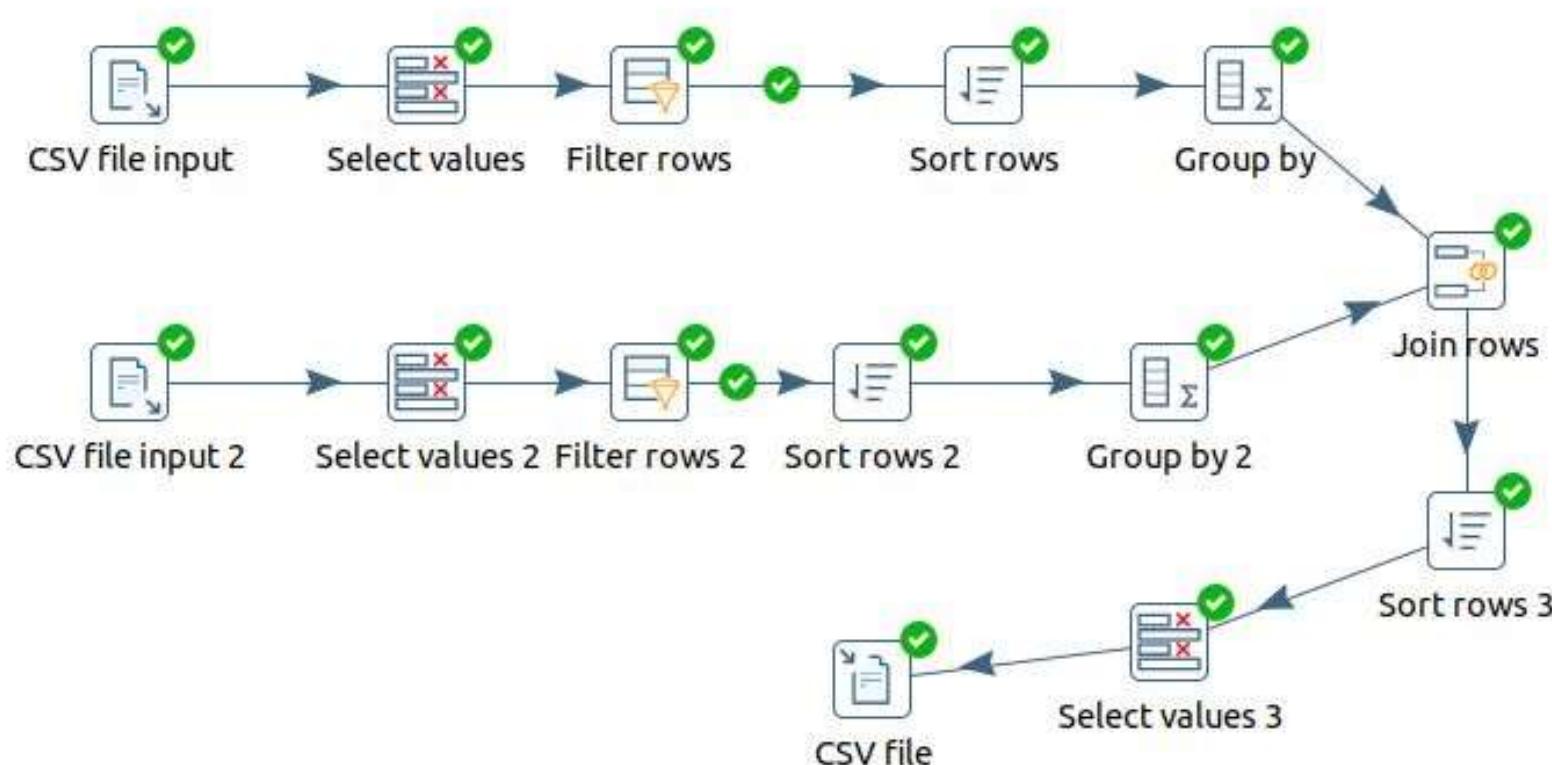
2023/2024

# Data Analysis and Integration Project Presentation

Diogo Clemente dos Santos 99200  
Ana Carolina Baptista 95529

01

# Transformation



# CSV file input

CSV file input

Step name: CSV file input

Filename: \$[Internal Entry:Current Directory]/Downloads/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 by
1	Date	String		7		\$	,	,	none
2	Municipality	String		27		\$	,	,	none
3	Number of CPE's	Integer	#	15	0	\$	,	,	none
4	DistrictMunicipalityCode	Integer	#	15	0	\$	,	,	none

Help OK Get Fields Preview Cancel

# CSV file input

Examine preview data

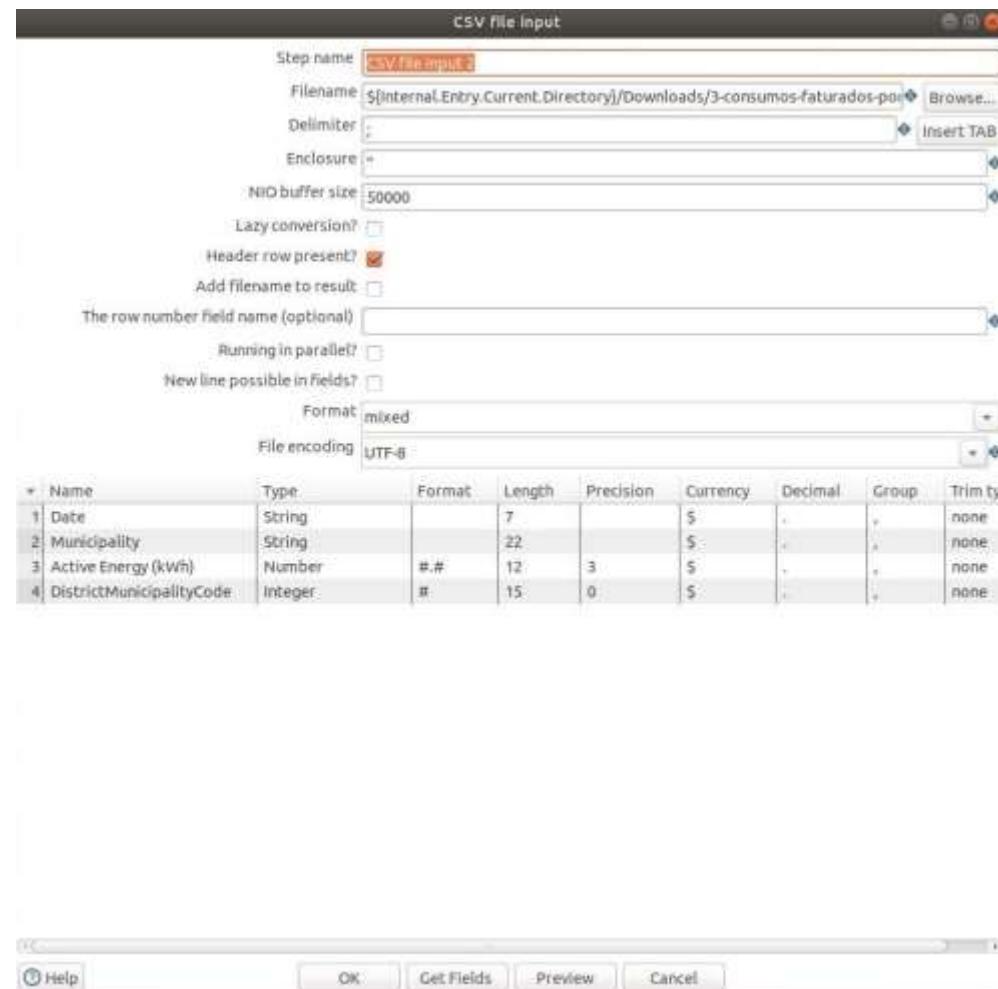
Rows of step: CSV file input (1000 rows)

	Date	Municipality	Number of CPE's	DistrictMunicipalityCode
1	2022-06	Espinho	453	107
2	2022-06	Guimarães	134	308
3	2022-09	Portel	221	709
4	2022-09	Mogadouro	42	408
5	2023-03	Braga	333	303
6	2023-03	Vila Flor	93	410
7	2022-07	Cuba	163	207
8	2023-07	Seia	323	912
9	2023-09	Guimarães	503	308
10	2023-09	Baião	293	1302
11	2022-05	Fafe	545	307
12	2022-05	Caldas da Rainha	612	1006
13	2022-02	Celorico da Beira	101	903
14	2022-01	Montalegre	62	1706
15	2022-12	Ponte da Barca	542	1606
16	2022-12	Felgueiras	575	1303
17	2022-04	Castelo Branco	566	502
18	2023-06	Celorico da Beira	163	903
19	2023-05	Borba	315	703
20	2022-09	Moimenta da Beira	185	1807
21	2022-09	Alandroal	340	701
22	2023-03	Cinfães	1056	1804
23	2022-11	Guarda	238	907
24	2022-07	Lousada	486	1305

[Close](#)

[Show Log](#)

# CSV file input 2



# CSV file input 2

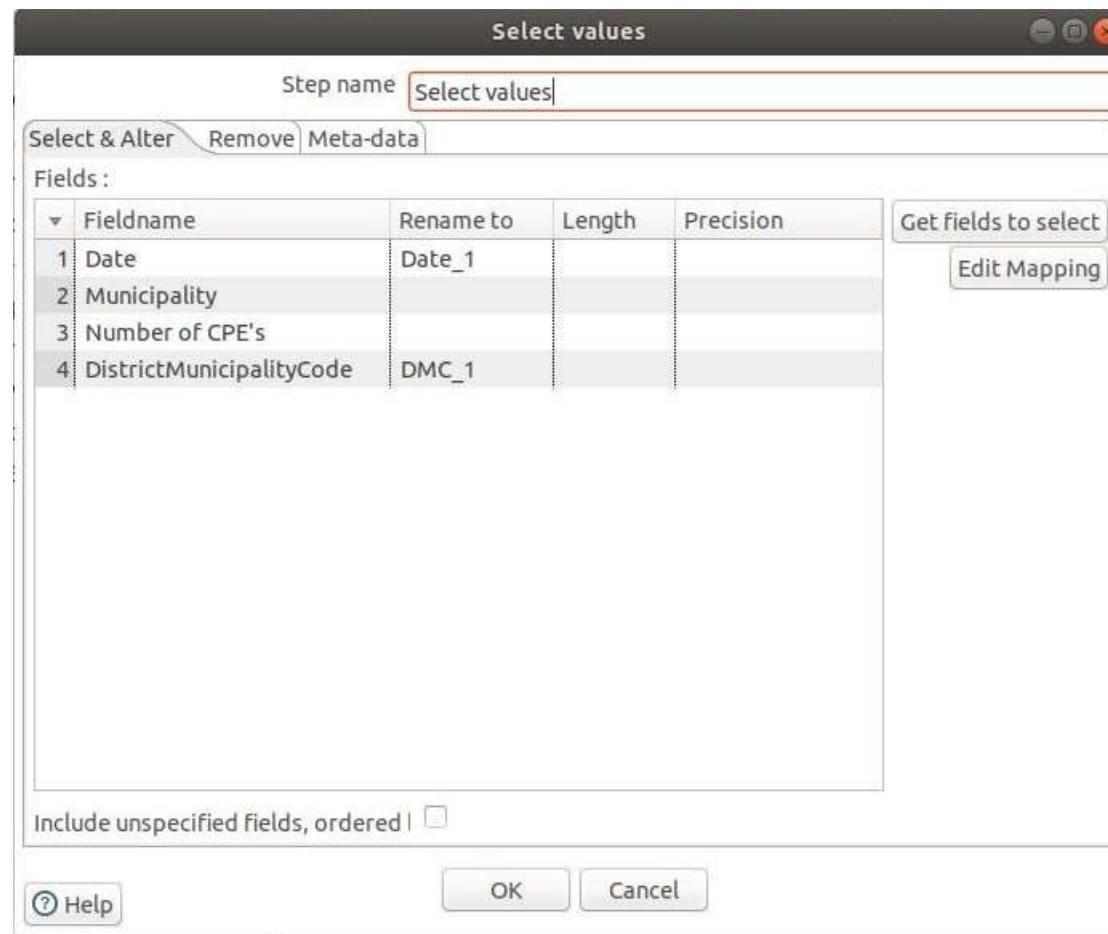
Examine preview data

Rows of step: CSV File input 2 (1000 rows)

	Date	Municipality	Active Energy (kWh)	DistrictMunicipalityCode
1	2022-07	Amadora	3821309.7	1115
2	2022-07	Vila do Conde	757616.4	1316
3	2022-08	Espinho	2244683.1	107
4	2022-08	Penamacor	46110.4	507
5	2022-08	Guarda	29907.3	907
6	2022-08	Sabugal	65586.5	911
7	2022-08	Óbidos	500908.2	1012
8	2022-08	Porto de Mós	142244.8	1016
9	2022-08	Amadora	3685384.5	1115
10	2022-08	Caminha	65482.1	1602
11	2022-08	Caminha	599125.9	1602
12	2022-08	Tondela	256920.4	1821
13	2022-09	Aveiro	615474.5	105
14	2022-09	Ilhavo	1922566	110
15	2022-09	Oliveira do Bairro	1191897.7	114
16	2022-09	Fafe	266406.9	307
17	2022-09	Portimão	2662235.5	811
18	2022-09	Guarda	29762.4	907
19	2022-09	Lourinhã	582614.3	1108
20	2022-09	Santo Tirso	1039783.1	1314
21	2022-09	Maçao	71593	1413
22	2022-09	Vouzela	178751.3	1824
23	2022-10	Arganil	2319197	601
24	2022-10	Montemor-o-Velho	286168.2	610

[Close](#) [Show Log](#)

# Select values



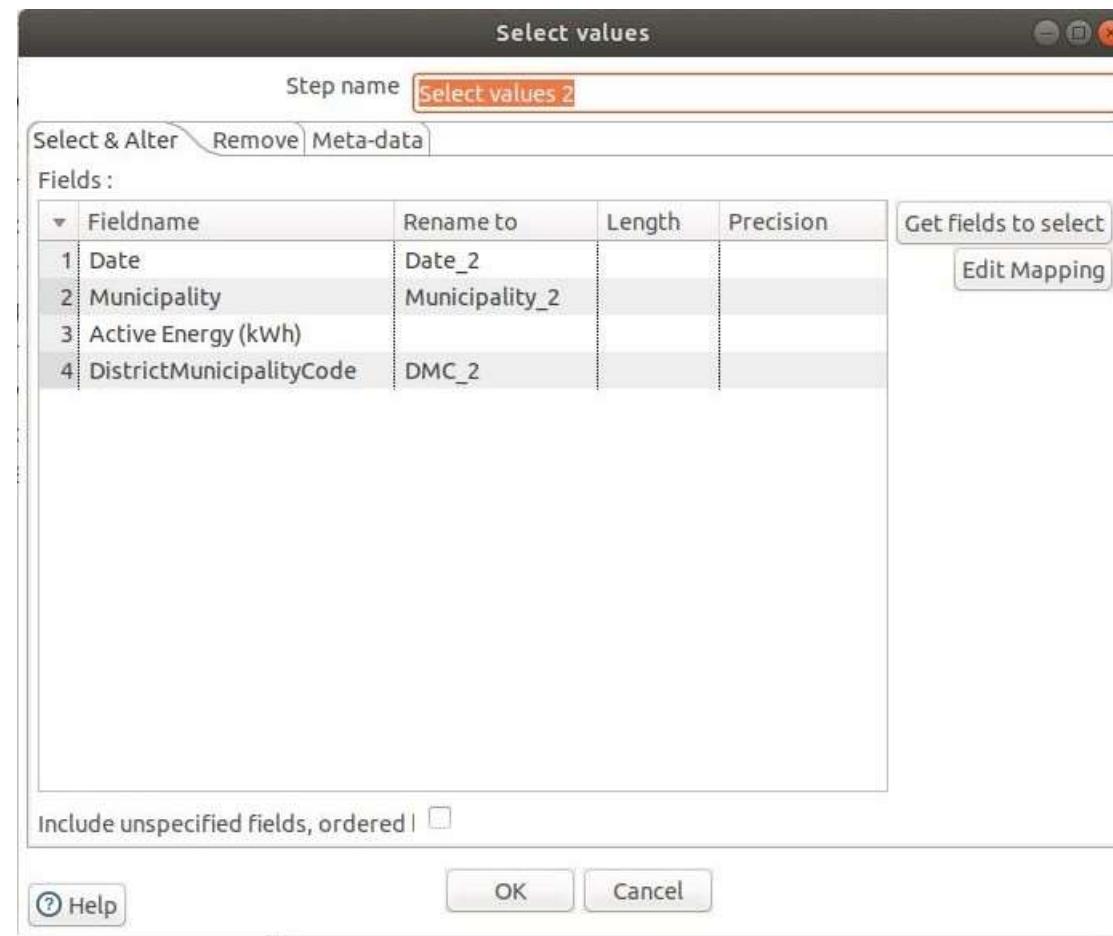
# Select values

Examine preview data

Rows of step: Select values (1000 rows)

	Date_1	Municipality	Number of CPE's	DMC_1
1	2022-06	Espinho	453	107
2	2022-06	Guimarães	134	308
3	2022-09	Portel	221	709
4	2022-09	Mogadouro	42	408
5	2023-03	Braga	333	303
6	2023-03	Vila Flor	93	410
7	2022-07	Cuba	163	207
8	2023-07	Seia	323	912
9	2023-09	Guimarães	503	308
10	2023-09	Baião	293	1302
11	2022-05	Fafe	545	307
12	2022-05	Caldas da Rainha	612	1006
13	2022-02	Celorico da Beira	101	903
14	2022-01	Montalegre	62	1706
15	2022-12	Ponte da Barca	542	1606
16	2022-12	Felgueiras	575	1303
17	2022-04	Castelo Branco	566	502
18	2023-06	Celorico da Beira	163	903
19	2023-05	Borba	315	703
20	2022-09	Moimenta da Beira	185	1807
21	2022-09	Alandroal	340	701
22	2023-03	Cinfães	1056	1804
23	2022-11	Guarda	238	907
24	2022-07	Lousada	486	1305

# Select values 2



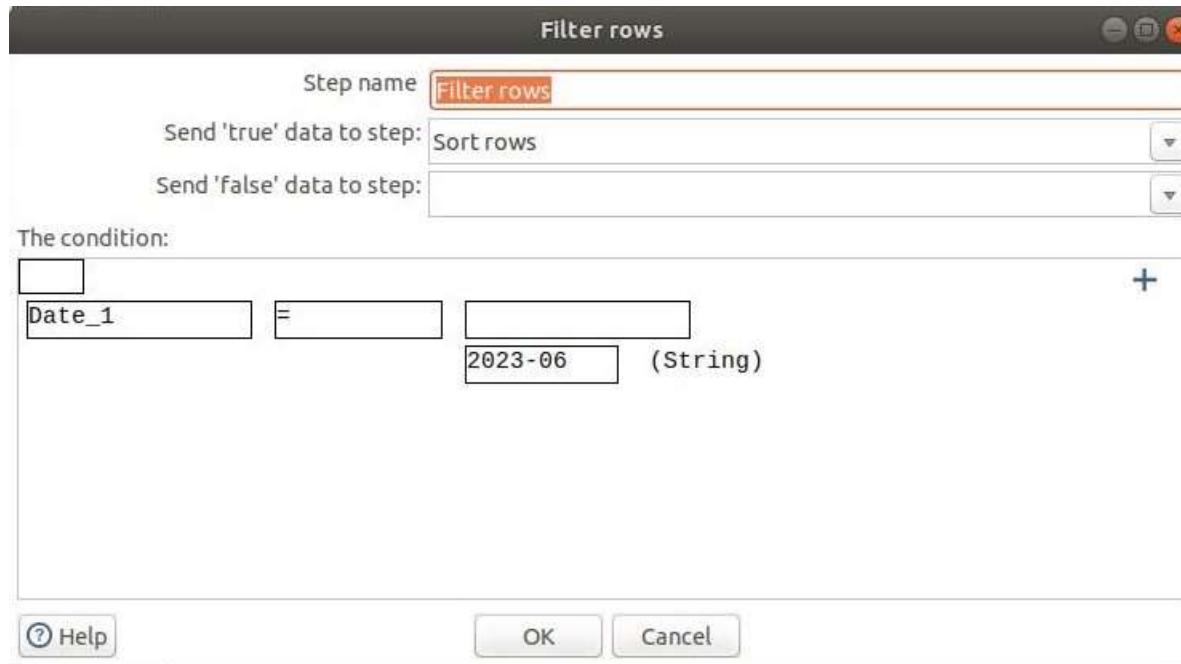
# Select values 2

Examine preview data

Rows of step: Select values 2 (1000 rows)

	Date_2	Municipality_2	Active Energy (kWh)	DMC_2
1	2022-07	Amadora	3821309.7	1115
2	2022-07	Vila do Conde	757616.4	1316
3	2022-08	Espinho	2244683.1	107
4	2022-08	Penamacor	46110.4	507
5	2022-08	Guarda	29907.3	907
6	2022-08	Sabugal	65586.5	911
7	2022-08	Óbidos	500908.2	1012
8	2022-08	Porto de Mós	142244.8	1016
9	2022-08	Amadora	3685384.5	1115
10	2022-08	Caminha	65482.1	1602
11	2022-08	Caminha	599125.9	1602
12	2022-08	Tondela	256920.4	1821
13	2022-09	Aveiro	615474.5	105
14	2022-09	Ilhavo	1922566	110
15	2022-09	Oliveira do Bairro	1191897.7	114
16	2022-09	Fafe	266406.9	307
17	2022-09	Portimão	2662235.5	811
18	2022-09	Guarda	29762.4	907
19	2022-09	Lourinhã	582614.3	1108
20	2022-09	Santo Tirso	1039783.1	1314
21	2022-09	Maçao	71593	1413
22	2022-09	Vouzela	178751.3	1824
23	2022-10	Arganil	2319197	601
24	2022-10	Montemor-o-Velho	286168.2	610

# Filter rows



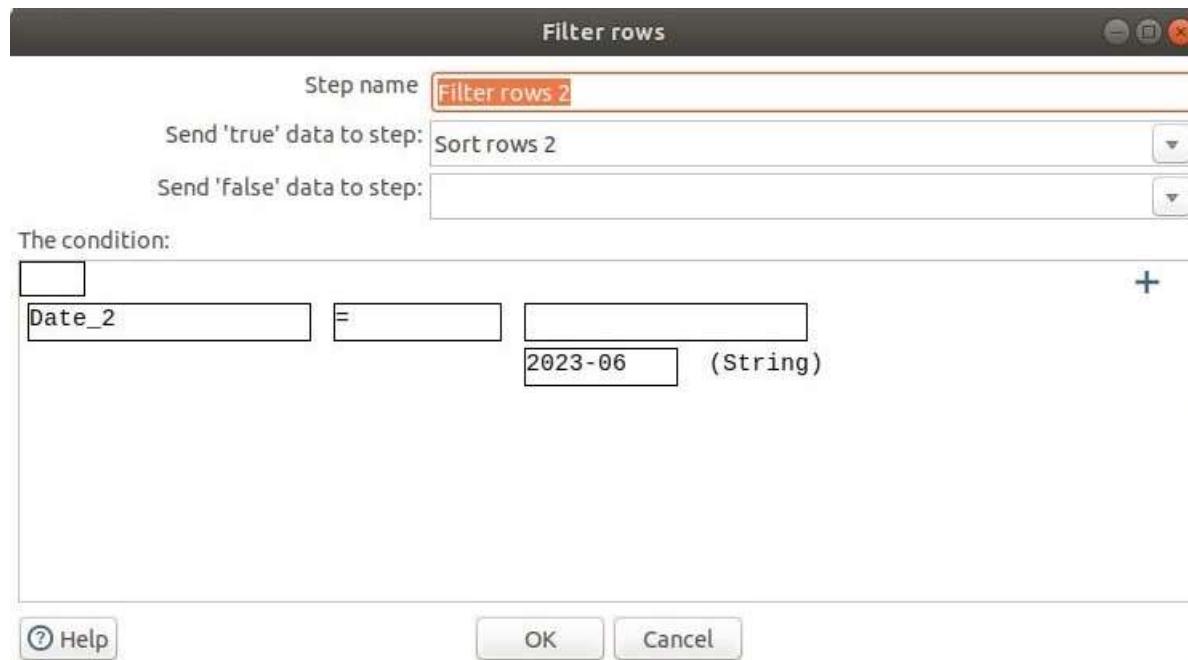
# Filter rows

Examine preview data

Rows of step: Filter rows (1000 rows)

	Date_1	Municipality	Number of CPE's	DMC_1
1	2023-06	Celorico da Beira	163	903
2	2023-06	Braga	362	303
3	2023-06	Idanha-a-Nova	39	505
4	2023-06	Albergaria-a-Velha	469	102
5	2023-06	Resende	607	1813
6	2023-06	Cascais	4172	1105
7	2023-06	Marco de Canaveses	550	1307
8	2023-06	Penela	173	614
9	2023-06	Barcelos	530	302
10	2023-06	Vila do Conde	354	1316
11	2023-06	Melgaço	88	1603
12	2023-06	Lagoa	3301	806
13	2023-06	Covilhã	398	503
14	2023-06	Fornos de Algodres	113	905
15	2023-06	Tarouca	243	1820
16	2023-06	Ribeira de Pena	147	1709
17	2023-06	Alijó	214	1701
18	2023-06	Vagos	3523	118
19	2023-06	Alijó	322	1701
20	2023-06	Caminha	193	1602
21	2023-06	Ponte de Lima	506	1607
22	2023-06	Paredes	493	1310
23	2023-06	Odemira	490	211
24	2023-06	Sabugal	138	911

# Filter rows 2



# Filter rows 2

Examine preview data

Rows of step: Filter rows 2 (1000 rows)

	Date_2	Municipality_2	Active Energy (kWh)	DMC_2
1	2023-06	Miranda do Corvo	426337.4	609
2	2023-06	Portimão	1029578.7	811
3	2023-06	Porto de Mós	1378504.1	1016
4	2023-06	Felgueiras	136161.8	1303
5	2023-06	Arcos de Valdevez	31831.3	1601
6	2023-06	Vila Real	67637.4	1714
7	2023-06	Almodôvar	90220.6	202
8	2023-06	Odemira	671764.2	211
9	2023-06	Torres Vedras	79163.1	1113
10	2023-06	Fronteira	105237.8	1208
11	2023-06	Vila do Conde	398136.6	1316
12	2023-06	Coruche	163823.8	1409
13	2023-06	Viseu	126909.2	1823
14	2023-06	Ferreira do Alentejo	2152018.5	208
15	2023-06	Braga	767437.8	303
16	2023-06	Póvoa de Lanhoso	93535.9	309
17	2023-06	Covilhã	2508581.7	503
18	2023-06	Arganil	201297	601
19	2023-06	São Pedro do Sul	84389.1	1816
20	2023-06	Vouzela	48901.4	1824
21	2023-06	Vagos	320913.5	118
22	2023-06	Mirandela	38065.5	407
23	2023-06	Alandroal	623162.5	701
24	2023-06	Alenquer	128920.6	1101

# Sort rows

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)

This is the free me  
maximum JVM me  
We will increase th

Fields :

▼	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted
1	DMC_1	Y	N	N	0	N

① Help

# Sort rows

Examine preview data

	Date_1	Municipality	Number of CPE's	DMC_1
1	2023-06	Águeda	156	101
2	2023-06	Águeda	1073	101
3	2023-06	Águeda	300	101
4	2023-06	Águeda	1395	101
5	2023-06	Águeda	323	101
6	2023-06	Águeda	705	101
7	2023-06	Águeda	2876	101
8	2023-06	Águeda	7673	101
9	2023-06	Águeda	354	101
10	2023-06	Águeda	147	101
11	2023-06	Águeda	1608	101
12	2023-06	Águeda	476	101
13	2023-06	Águeda	764	101
14	2023-06	Águeda	1698	101
15	2023-06	Águeda	2150	101
16	2023-06	Águeda	169	101
17	2023-06	Águeda	168	101
18	2023-06	Águeda	56	101
19	2023-06	Águeda	1946	101
20	2023-06	Águeda	325	101
21	2023-06	Águeda	143	101
22	2023-06	Águeda	297	101
23	2023-06	Albergaria-a-Velha	469	102
24	2023-06	Albergaria-a-Velha	371	102

# Sort rows 2

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 %) `0`

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presort
1 DMC_2	Y	N	N	0	N

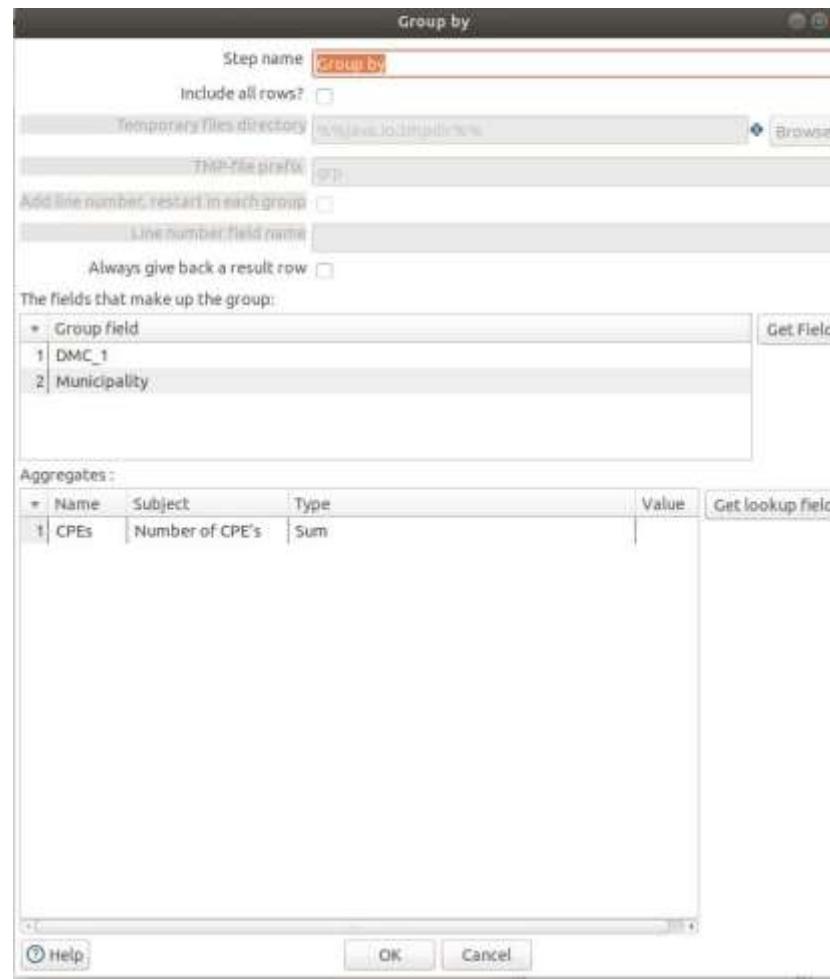
# Sort rows 2

Examine preview data

Rows of step: Sort rows 2 (1000 rows)

	Date_2	Municipality_2	Active Energy (kWh)	DMC_2
1	2023-06	Águeda	49163.6	101
2	2023-06	Águeda	7042.1	101
3	2023-06	Águeda	587642	101
4	2023-06	Águeda	723851.5	101
5	2023-06	Águeda	120308.5	101
6	2023-06	Águeda	401124.8	101
7	2023-06	Águeda	2920985.3	101
8	2023-06	Águeda	748527.1	101
9	2023-06	Águeda	623137.8	101
10	2023-06	Águeda	404615.9	101
11	2023-06	Águeda	1206462.7	101
12	2023-06	Águeda	202800.3	101
13	2023-06	Águeda	323120.7	101
14	2023-06	Águeda	753311.3	101
15	2023-06	Águeda	111918.9	101
16	2023-06	Águeda	2153746	101
17	2023-06	Águeda	198295.9	101
18	2023-06	Águeda	1131063.2	101
19	2023-06	Águeda	6499536.4	101
20	2023-06	Águeda	1822944.9	101
21	2023-06	Águeda	1602205.5	101
22	2023-06	Águeda	767498.1	101
23	2023-06	Albergaria-a-Velha	350953.7	102
24	2023-06	Albergaria-a-Velha	799719	102

# Group by



# Group by

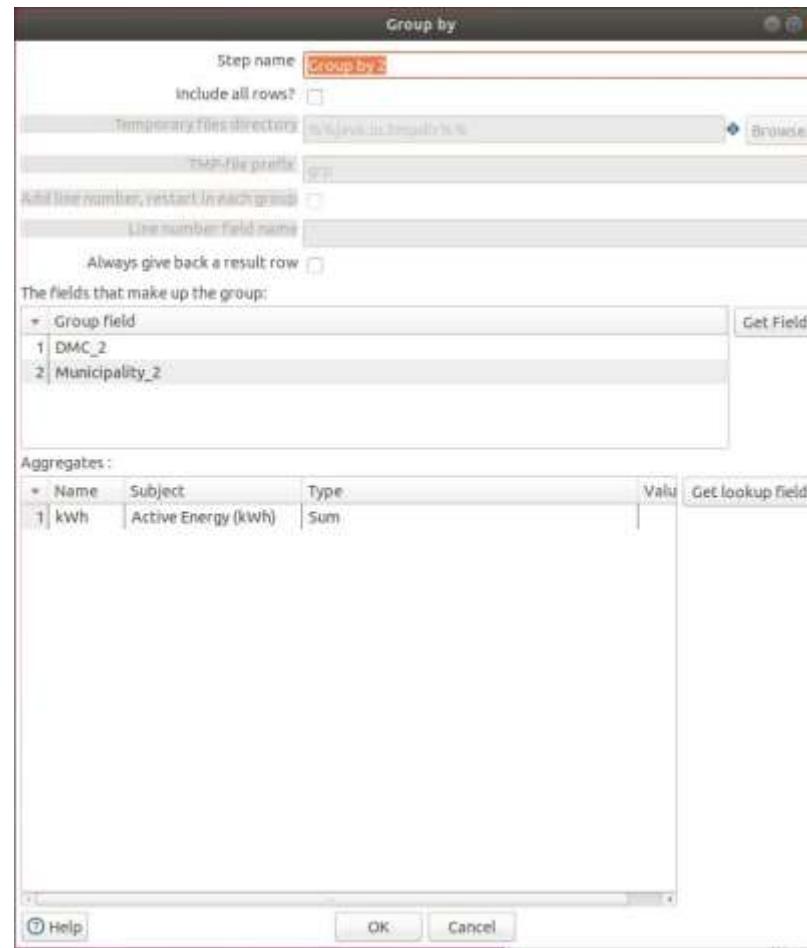
Examine preview data

Rows of step: Group by (278 rows)

	DMC_1	Municipality	CPEs
1	101	Águeda	24802
2	102	Albergaria-a-Velha	14241
3	103	Anadia	17295
4	104	Arouca	12670
5	105	Aveiro	51443
6	106	Castelo de Paiva	8082
7	107	Espinho	18639
8	108	Estarreja	14469
9	109	Santa Maria da Feira	72389
10	110	Ilhavo	25394
11	111	Mealhada	11546
12	112	Murtosa	8570
13	113	Oliveira de Azeméis	33331
14	114	Oliveira do Bairro	13740
15	115	Ovar	32390
16	116	São João da Madeira	13526
17	117	Sever do Vouga	7693
18	118	Vagos	15568
19	119	Vale de Cambra	13471
20	201	Aljustrel	6058
21	202	Almodôvar	5693
22	203	Alvito	1725
23	204	Barrancos	1239
24	205	Beja	23054

**Close**

# Group by 2



# Group by 2

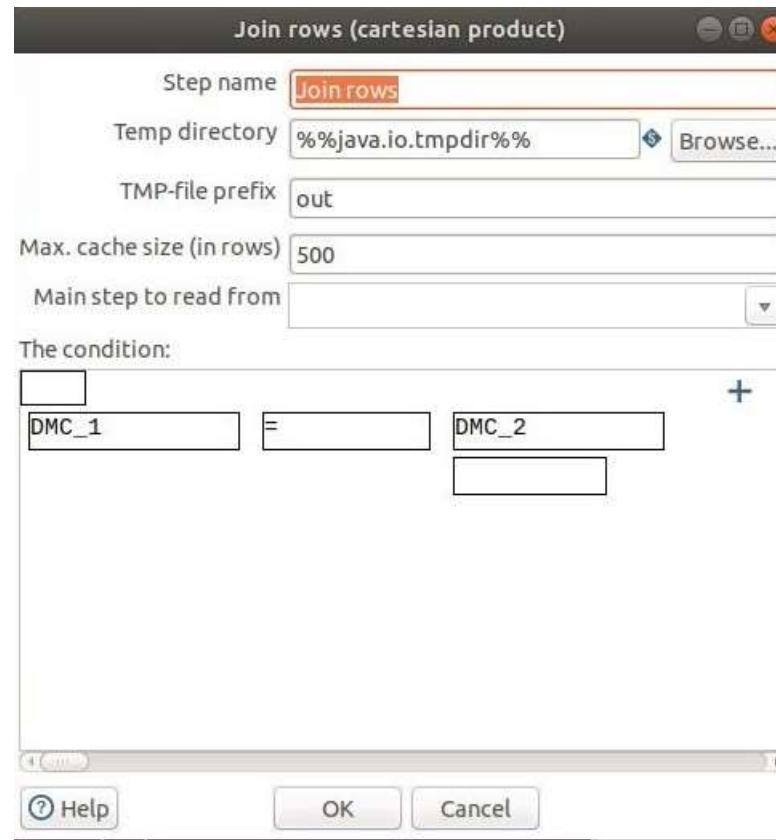
Examine preview data

Rows of step: Group by 2 (278 rows)

	DMC_2	Municipality_2	kwh
1	101	Águeda	23359302.557
2	102	Albergaria-a-Velha	14765120.434
3	103	Anadia	9902342.348
4	104	Arouca	4066420.3629999994
5	105	Aveiro	59927297.077999994
6	106	Castelo de Paiva	3449372.4540000004
7	107	Espinho	6566676.72
8	108	Estarreja	52941071.745000005
9	109	Santa Maria da Feira	50119944.23199999
10	110	Ílhavo	16115482.652
11	111	Mealhada	7650201.085999999
12	112	Murtosa	2650940.073999996
13	113	Oliveira de Azeméis	30844529.374000005
14	114	Oliveira do Bairro	10360783.478
15	115	Ovar	25900054.744000003
16	116	São João da Madeira	8928718.162
17	117	Sever do Vouga	2596915.376
18	118	Vagos	7618458.860000001
19	119	Vale de Cambra	7862887.632
20	201	Aljustrel	15121881.473000001
21	202	Almodôvar	1455934.812000002
22	203	Alvito	1220644.494
23	204	Barrancos	401405.503
24	205	Beja	45508349.42300001

[Close](#)

# Join rows



# Join rows

Examine preview data

Rows of step: Join rows (278 rows)

	DMC_1	Municipality	CPEs	DMC_2	Municipality_2	kwh
1	101	Águeda	24802	101	Águeda	23359302.557
2	102	Albergaria-a-Velha	14241	102	Albergaria-a-Velha	14765120.434
3	103	Anadia	17295	103	Anadia	9902342.348
4	104	Arouca	12670	104	Arouca	4066420.3629999994
5	105	Aveiro	51443	105	Aveiro	59927297.077999994
6	106	Castelo de Paiva	8082	106	Castelo de Paiva	3449372.454000004
7	107	Espinho	18639	107	Espinho	6566676.72
8	108	Estarreja	14469	108	Estarreja	52941071.74500005
9	109	Santa Maria da Feira	72389	109	Santa Maria da Feira	50119944.23199999
10	110	ílhavo	25394	110	ílhavo	16115482.652
11	111	Mealhada	11546	111	Mealhada	7650201.085999999
12	112	Murtosa	8570	112	Murtosa	2650940.073999996
13	113	Oliveira de Azeméis	33331	113	Oliveira de Azeméis	30844529.37400005
14	114	Oliveira do Bairro	13740	114	Oliveira do Bairro	10360783.478
15	115	Ovar	32390	115	Ovar	25900054.74400003
16	116	São João da Madeira	13526	116	São João da Madeira	8928718.162
17	117	Sever do Vouga	7693	117	Sever do Vouga	2596915.376
18	118	Vagos	15568	118	Vagos	7618458.86000001
19	119	Vale de Cambra	13471	119	Vale de Cambra	7862887.632
20	201	Aljustrel	6058	201	Aljustrel	15121881.47300001
21	202	Almodôvar	5693	202	Almodôvar	1455934.812000002
22	203	Alvito	1725	203	Alvito	1220644.494
23	204	Barrancos	1239	204	Barrancos	401405.503
24	205	Beja	23054	205	Beja	45508349.42300001

Close

# Sort rows 3

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	Municipality	Y	N	Y	0	N

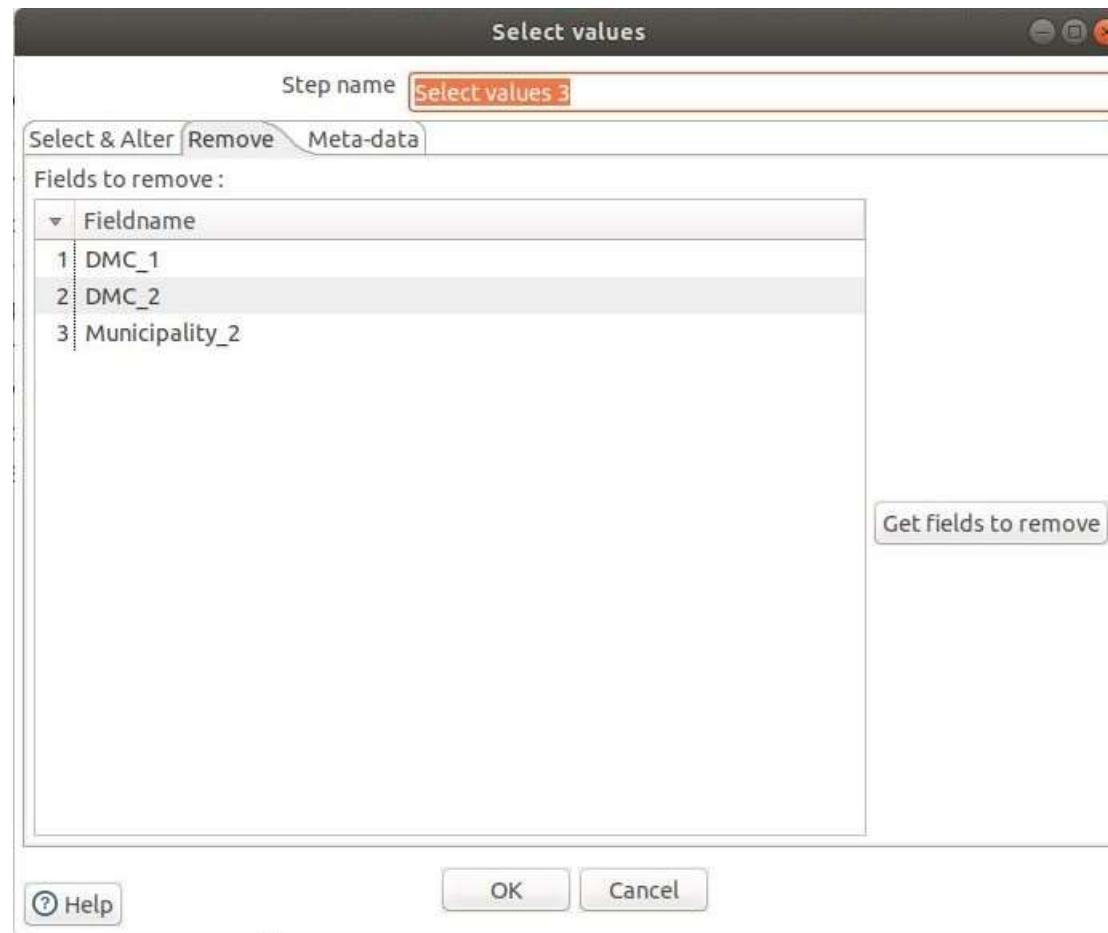
# Sort rows 3

Examine preview data

Rows of step: Sort rows 3 (278 rows)

	DMC_1	Municipality	CPEs	DMC_2	Municipality_2	kwh
1	1401	Abrantes	25180	1401	Abrantes	10567936.841000002
2	101	Águeda	24802	101	Águeda	23359302.557
3	901	Aguiar da Beira	4933	901	Aguiar da Beira	1135609.924
4	701	Alandroal	4133	701	Alandroal	1838245.245999998
5	102	Albergaria-a-Velha	14241	102	Albergaria-a-Velha	14765120.434
6	801	Albufeira	54157	801	Albufeira	33437993.117000002
7	1501	Alcácer do Sal	8480	1501	Alcácer do Sal	6333071.731
8	1402	Alcanena	8265	1402	Alcanena	5643207.701
9	1001	Alcobaça	39245	1001	Alcobaça	19699225.738
10	1502	Alcochete	11819	1502	Alcochete	6670715.332
11	802	Alcoutim	3703	802	Alcoutim	611524.557
12	1101	Alenquer	24858	1101	Alenquer	18889521.167
13	401	Alfândega da Fé	4167	401	Alfândega da Fé	758239.7180000001
14	1701	Alijó	9146	1701	Alijó	2051978.2900000003
15	803	Aljezur	6523	803	Aljezur	1880390.4440000001
16	201	Aljustrel	6058	201	Aljustrel	15121881.473000001
17	1503	Almada	116447	1503	Almada	39254590.524
18	902	Almeida	7014	902	Almeida	1219835.386
19	1403	Almeirim	14117	1403	Almeirim	6655643.254000001
20	202	Almodôvar	5693	202	Almodôvar	1455934.8120000002
21	1404	Alpiarça	4630	1404	Alpiarça	3980930.421
22	1201	Alter do Chão	2892	1201	Alter do Chão	858120.347
23	1002	Alvaiázere	5897	1002	Alvaiázere	1248745.845
24	203	Alvito	1725	203	Alvito	1220644.494

# Select values 3



# Select values 3

Examine preview data

Rows of step: Select values 3 (278 rows)

	Municipality	CPEs	kwh
1	Abrantes	25180	10567936.841000002
2	Águeda	24802	23359302.557
3	Aguiar da Beira	4933	1135609.924
4	Alandroal	4133	1838245.245999998
5	Albergaria-a-Velha	14241	14765120.434
6	Albufeira	54157	33437993.117000002
7	Alcácer do Sal	8480	6333071.731
8	Alcanena	8265	5643207.701
9	Alcobaça	39245	19699225.738
10	Alcochete	11819	6670715.332
11	Alcoutim	3703	611524.557
12	Alenquer	24858	18889521.167
13	Alfândega da Fé	4167	758239.718000001
14	Alijó	9146	2051978.290000003
15	Aljezur	6523	1880390.444000001
16	Aljustrel	6058	15121881.473000001
17	Almada	116447	39254590.524
18	Almeida	7014	1219835.386
19	Almeirim	14117	6655643.254000001
20	Almodôvar	5693	1455934.812000002
21	Alpiarça	4630	3980930.421
22	Alter do Chão	2892	858120.347
23	Alvaiázere	5897	1248745.845
24	Alvito	1725	1220644.494

# CSV file

Text file output

Step name **CSV file**

**File Content Fields**

Filename

Pass output to servlet

Create Parent folder

Do not create file at start

Accept file name from field?

File name field

Extension

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

# CSV file

Examine preview data

Rows of step: CSV file (278 rows)

	Municipality	CPEs	kwh
1	Abrantes	25180	10567936.841000002
2	Águeda	24802	23359302.557
3	Aguiar da Beira	4933	1135609.924
4	Alandroal	4133	1838245.245999998
5	Albergaria-a-Velha	14241	14765120.434
6	Albufeira	54157	33437993.117000002
7	Alcácer do Sal	8480	6333071.731
8	Alcanena	8265	5643207.701
9	Alcobaça	39245	19699225.738
10	Alcochete	11819	6670715.332
11	Alcoutim	3703	611524.557
12	Alenquer	24858	18889521.167
13	Alfândega da Fé	4167	758239.718000001
14	Alijó	9146	2051978.290000003
15	Aljezur	6523	1880390.444000001
16	Aljustrel	6058	15121881.47300001
17	Almada	116447	39254590.524
18	Almeida	7014	1219835.386
19	Almeirim	14117	6655643.25400001
20	Almodôvar	5693	1455934.812000002
21	Alpiarça	4630	3980930.421
22	Alter do Chão	2892	858120.347
23	Alvaiázere	5897	1248745.845
24	Alvito	1725	1220644.494

[Close](#)

# Output file



Municipality	CPGs	KWh
Abrantes	25181	10567936.841
Agueda	24802	23350302.557
Aguas da Beira	4933	1135609.934
Alcoutim	4133	1838245.246
Albergaria-a-Velha	14241	14765120.434
Albufeira	54157	33437993.117
Alcacer do Sal	8460	6333071.731
Alcanena	8285	5643207.701
Alcobaça	39245	19698225.738
Alcochete	11819	6670715.332
Alcoutim	3703	611524.557
Alegrete	24668	188886521.167
Altandega da Fe	4167	756239.718
Alje	9146	2051978.29
Almeida	6023	1888390.444
Almourol	6058	15121881.473
Almodôvar	116447	39254590.524
Almada	7014	1218835.366
Almeirim	14117	6655643.254
Almourol	5093	1455934.812
Alpiarça	4630	3960930.471
Alter do Chão	2992	858120.347
Alvalade	5897	1248745.845
Alvito	1725	1220644.494
Amadora	98733	38592131.582
Amareleja	31797	9305889.397
Amieira	11329	3512297.001
Amieira	17294	9902342.348
Angro	9048	2611003.207
Arcos de Valdevez	19112	7737796.646
Arganil	11140	3583461.19
Armamar	5147	1484130.589
Arouca	12870	4066420.363
Arruda dos Vinhos	4874	2092141.816
Armaches	2319	796947.881
Amadora dos Vinhos	7678	2988192.087
Aveiro	51443	59827297.078
Avintes	3182	4463050.29
Azenhas	12414	12774466.277
Ribeira	11943	79553199.179

02

# Analysis Job

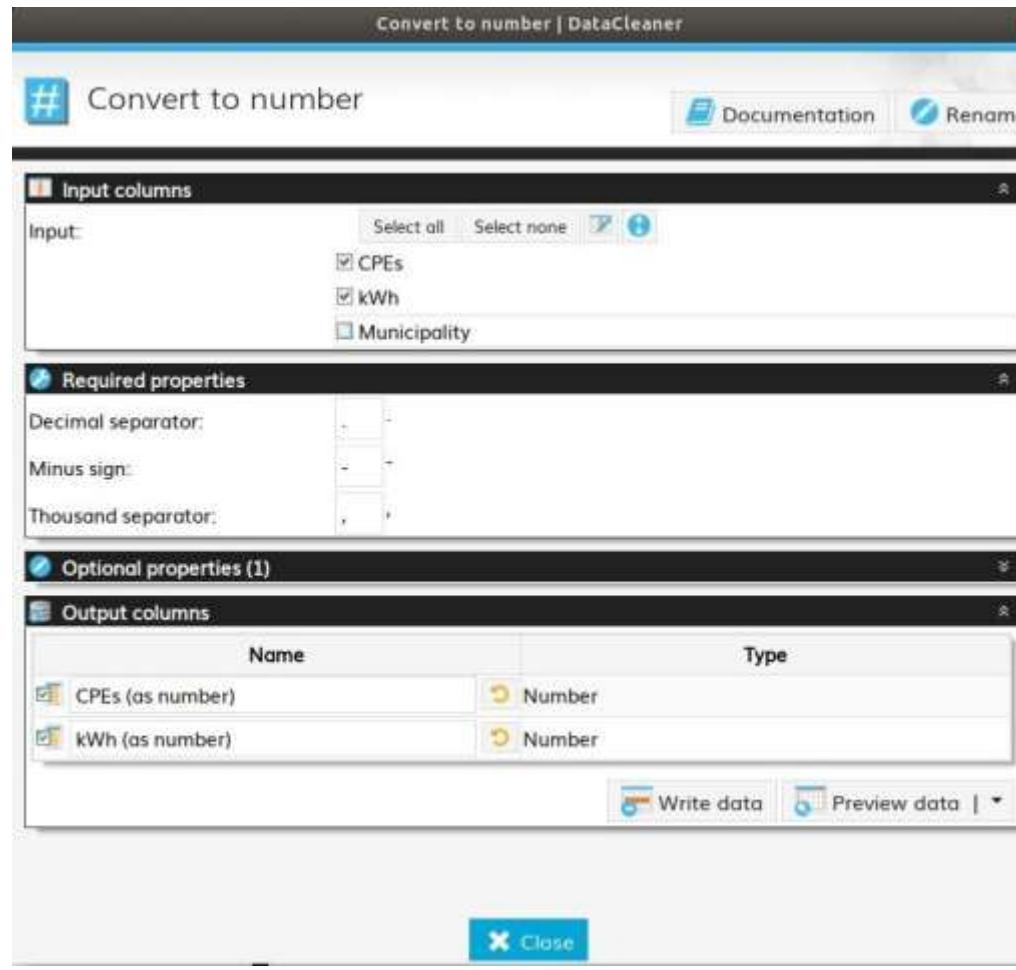
The screenshot shows a Data Cleaning tool interface with the following elements:

- Toolbar:** Includes "New", "Open", "Save", "Save As...", "More", "Execute", and a status indicator.
- Left Sidebar:** Shows a file tree with "resultados-quest-1-final", "information\_schema", "projeto", "Library", "Transform", "Improve", "Analyze", and "Write".
- Main Area:** A flow diagram with nodes:
  - A blue square node labeled "resultados-1.2.csv".
  - A green square node labeled "Scatter plot".
  - A blue square node labeled "Convert to number".Arrows indicate data flow from "resultados-1.2.csv" to "Scatter plot", and from "Scatter plot" to "Convert to number".
- Callout:** An arrow points to the "Execute" button with the text "Click here to run job".
- Bottom Left:** A play button icon with the text "Ready to execute".
- Bottom Right:** Text "Click the 'Execute' button in the upper-right corner when you're ready to run the job."
- Status Bar:** Includes "Search component library..." and "Job is correctly configured".
- Footer:** "Community edition - Up to date".

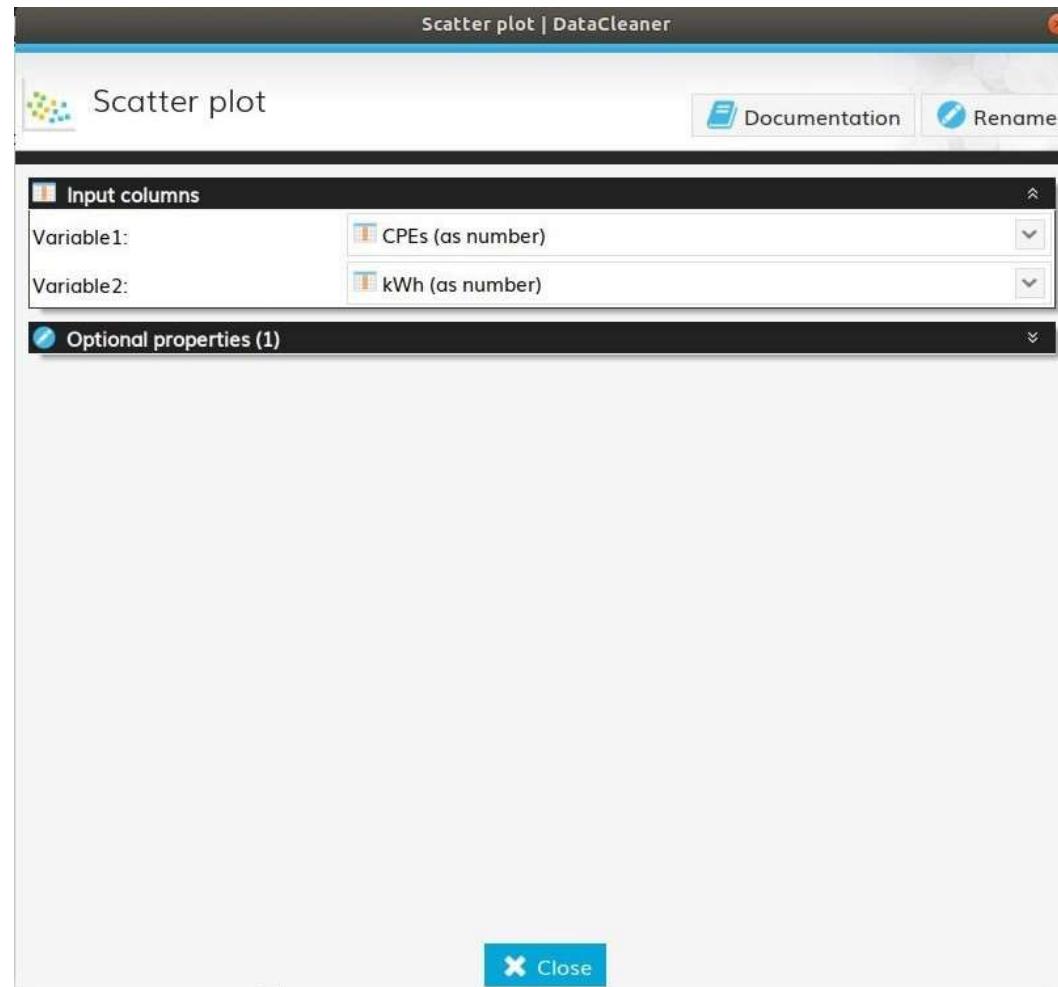
# CSV configuration



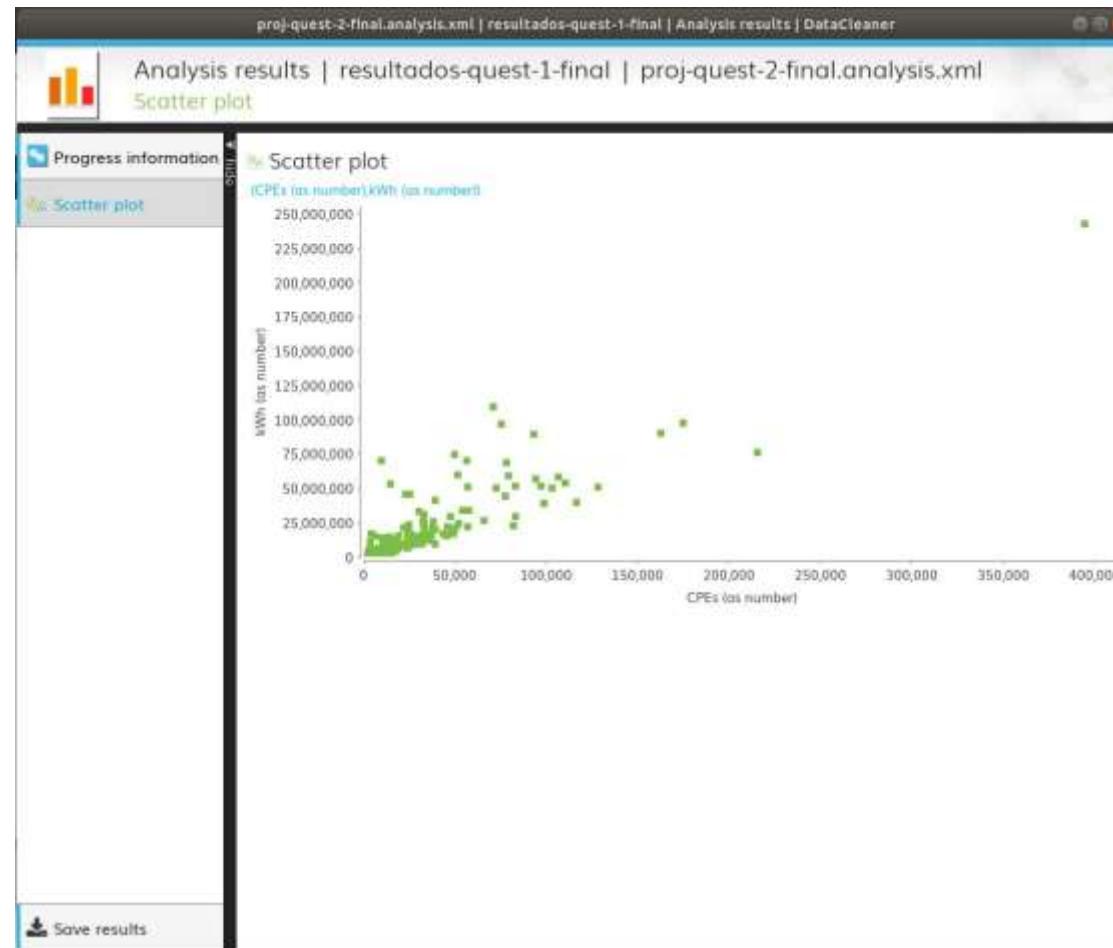
# Convert to number configuration



# Scatter plot configuration



# Analysis results



# Lisbon

Detailed results for scatter plot coordinate | DataCleaner

Detailed results

Records (1)

Municipality	CPEs	kWh	CPEs (as number)	kWh (as number)
Lisboa	394572	2426645...	394572	2.42664567353E8

# Sintra

Detailed results for scatter plot coordinate | DataCleaner

Detailed results

Records (1) Save dataset

Municipality	CPEs	kWh	CPEs (as number)	kWh (as number)
Sintra	215583	7568417...	215583	7.5684179255E7

# Porto

Detailed results for scatter plot coordinate | DataCleaner

**Detailed results**

Records (1)

Save dataset

Municipality	CPEs	kWh	CPEs (as number)	kWh (as number)
Porto	174973	9752099...	174973	9.7520991223E7

# Vila Nova de Gaia

Detailed results for scatter plot coordinate | DataCleaner

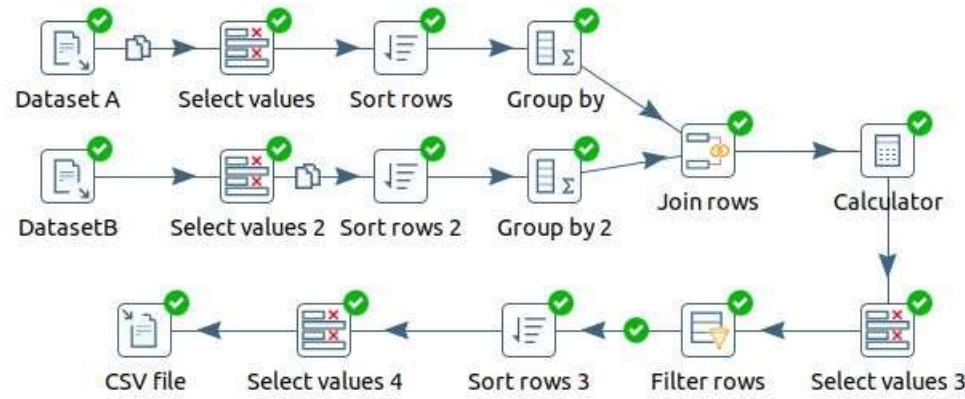
Records (1)

Save dataset

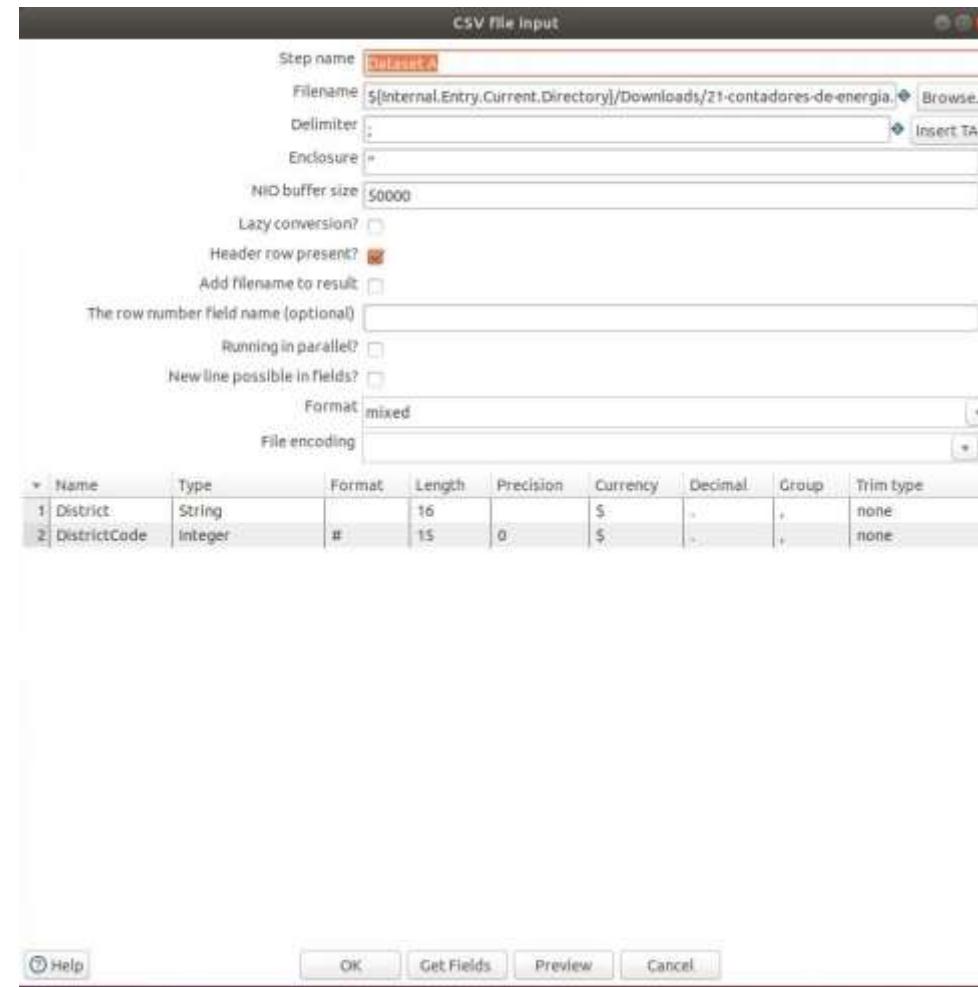
Municipality	CPEs	kWh	CPEs (as number)	kWh (as number)
Vila Nova de Gaia	162781	89994080.204	162781	8.9994080204E7

03

# Transformation



# CSV file input



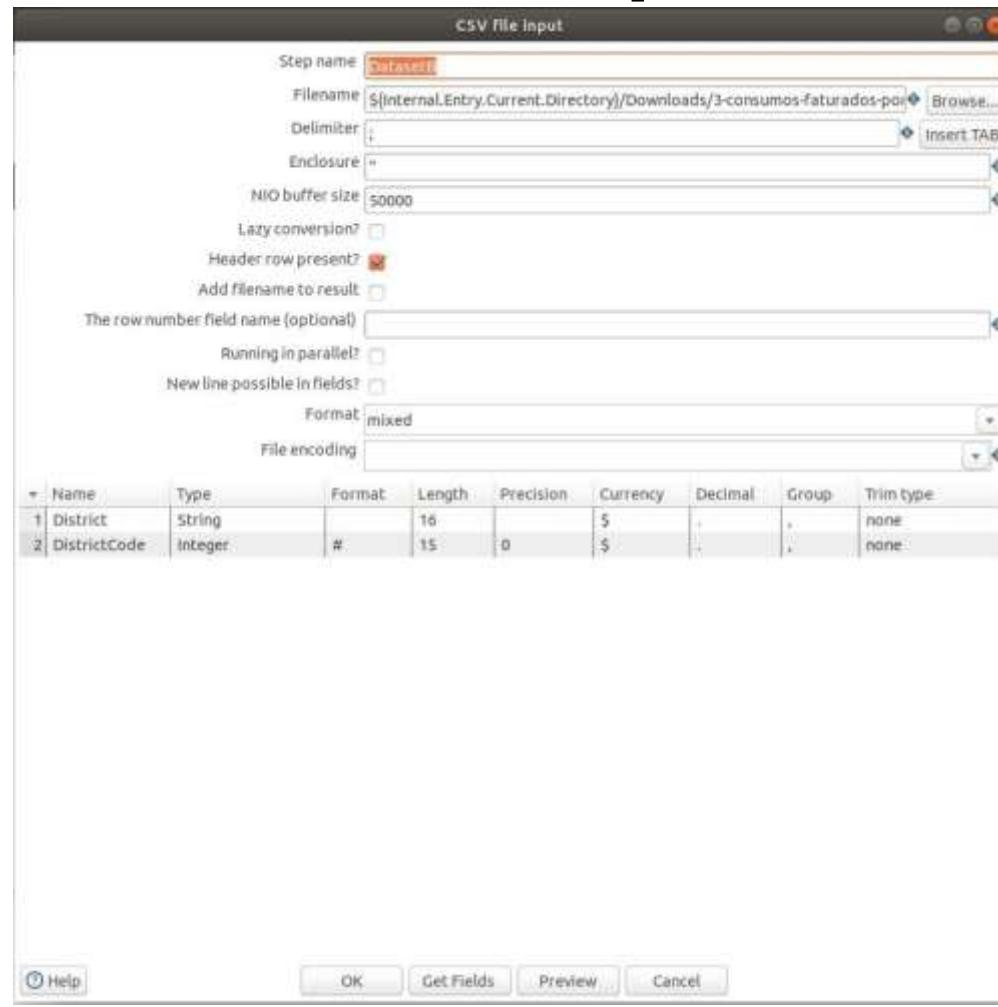
# CSV file input

Examine preview data

Rows of step: Dataset A (1000 rows)

	District	DistrictCode
1	Aveiro	1
2	Braga	3
3	Évora	7
4	Bragança	4
5	Braga	3
6	Bragança	4
7	Beja	2
8	Guarda	9
9	Braga	3
10	Porto	13
11	Braga	3
12	Leiria	10
13	Guarda	9
14	Vila Real	17
15	Viana do Castelo	16
16	Porto	13
17	Castelo Branco	5
18	Guarda	9
19	Évora	7
20	Viseu	18
21	Évora	7
22	Viseu	18
23	Guarda	9
24	Porto	13

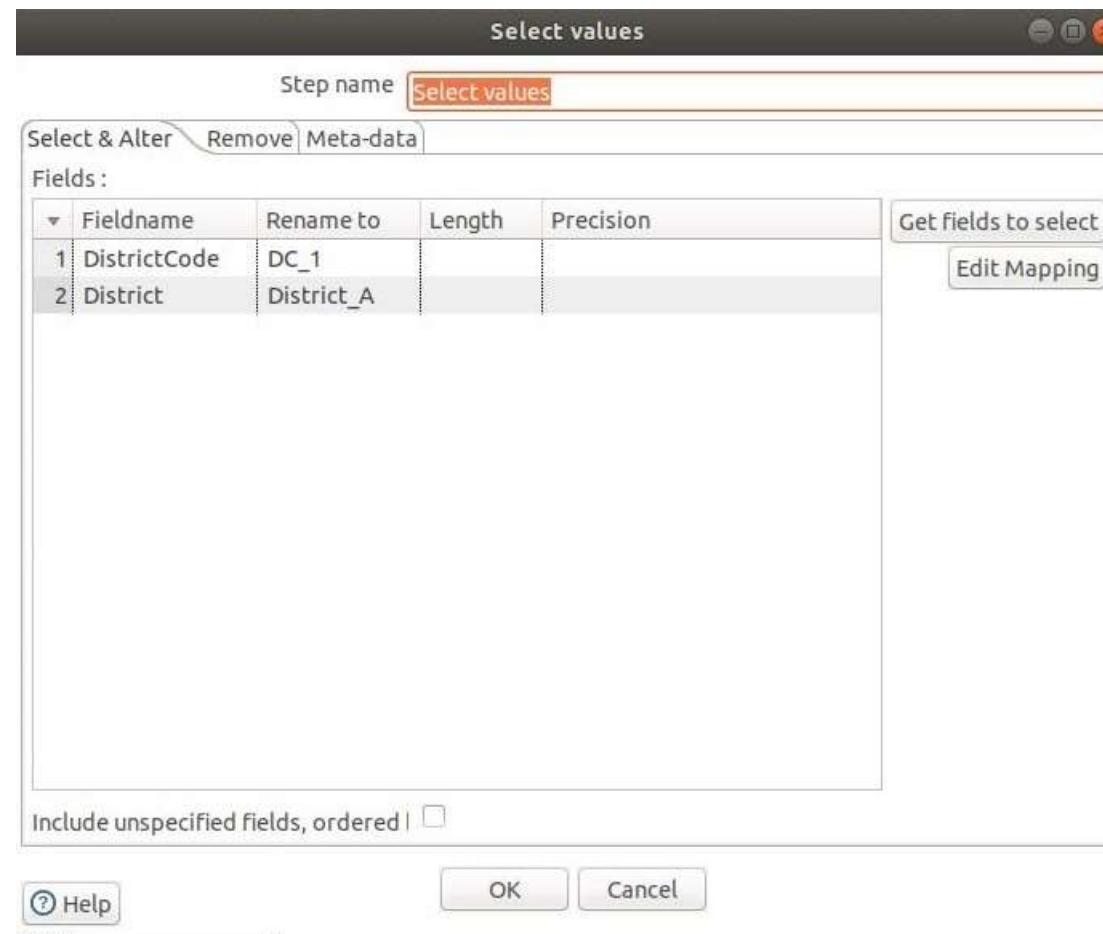
# CSV file input 2



# CSV file input 2

Rows of step: DatasetB (1000 rows)		Examine preview data
	District	DistrictCode
1	LISBOA	11
2	PORTO	13
3	AVEIRO	1
4	CASTELO BRANCO	5
5	GUARDA	9
6	GUARDA	9
7	LEIRIA	10
8	LEIRIA	10
9	LISBOA	11
10	VIANA DO CASTELO	16
11	VIANA DO CASTELO	16
12	VISEU	18
13	AVEIRO	1
14	AVEIRO	1
15	AVEIRO	1
16	BRAGA	3
17	FARO	8
18	GUARDA	9
19	LISBOA	11
20	PORTO	13
21	SANTAREM	14
22	VISEU	18
23	COIMBRA	6
24	COIMBRA	6

# Select Values



# Select Values

Examine preview data

Rows of step: Select values (1000 rows)

	DC_1	District_A
1	1	Aveiro
2	3	Braga
3	7	Évora
4	4	Bragança
5	3	Braga
6	4	Bragança
7	2	Beja
8	9	Guarda
9	3	Braga
10	13	Porto
11	3	Braga
12	10	Leiria
13	9	Guarda
14	17	Vila Real
15	16	Viana do Castelo
16	13	Porto
17	5	Castelo Branco
18	9	Guarda
19	7	Évora
20	18	Viseu
21	7	Évora
22	18	Viseu
23	9	Guarda
24	13	Porto

# Select Values 2



# Select Values 2

Examine preview data

Rows of step: Select values 2 (1000 rows)

	DC_2	District_B
1	11	LISBOA
2	13	PORTO
3	1	AVEIRO
4	5	CASTELO BRANCO
5	9	GUARDA
6	9	GUARDA
7	10	LEIRIA
8	10	LEIRIA
9	11	LISBOA
10	16	VIANA DO CASTELO
11	16	VIANA DO CASTELO
12	18	VISEU
13	1	AVEIRO
14	1	AVEIRO
15	1	AVEIRO
16	3	BRAGA
17	8	FARO
18	9	GUARDA
19	11	LISBOA
20	13	PORTO
21	14	SANTAREM
22	18	VISEU
23	6	COIMBRA
24	6	COIMBRA

# Sort Rows

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 %) `10`

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_A	Y	N	N	0	N

# Sort Rows

Examine preview data

Rows of step: Sort rows (1000 rows)

	DC_1	District_A
1	1	Aveiro
2	1	Aveiro
3	1	Aveiro
4	1	Aveiro
5	1	Aveiro
6	1	Aveiro
7	1	Aveiro
8	1	Aveiro
9	1	Aveiro
10	1	Aveiro
11	1	Aveiro
12	1	Aveiro
13	1	Aveiro
14	1	Aveiro
15	1	Aveiro
16	1	Aveiro
17	1	Aveiro
18	1	Aveiro
19	1	Aveiro
20	1	Aveiro
21	1	Aveiro
22	1	Aveiro
23	1	Aveiro
24	1	Aveiro

# Sort Rows 2

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	Presorted?
1	District_B	Y	N	N	0	N

# Sort Rows 2

Examine preview data

Rows of step: Sort rows 2 (1000 rows)

	DC_2	District_B
1	1	AVEIRO
2	1	AVEIRO
3	1	AVEIRO
4	1	AVEIRO
5	1	AVEIRO
6	1	AVEIRO
7	1	AVEIRO
8	1	AVEIRO
9	1	AVEIRO
10	1	AVEIRO
11	1	AVEIRO
12	1	AVEIRO
13	1	AVEIRO
14	1	AVEIRO
15	1	AVEIRO
16	1	AVEIRO
17	1	AVEIRO
18	1	AVEIRO
19	1	AVEIRO
20	1	AVEIRO
21	1	AVEIRO
22	1	AVEIRO
23	1	AVEIRO
24	1	AVEIRO

# Group By

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	DC_1
2	District_A

Aggregates :

▼

Name	Subject	Type	Value
1	DC_1	DC_1	Number of Distinct Values (N)

② Help

# Group By

Examine preview data

Rows of step: Group by (18 rows)

	DC_1	District_A	DC_1_1
1	1	Aveiro	1
2	2	Beja	1
3	3	Braga	1
4	4	Bragança	1
5	5	Castelo Branco	1
6	6	Coimbra	1
7	8	Faro	1
8	9	Guarda	1
9	10	Leiria	1
10	11	Lisboa	1
11	12	Portalegre	1
12	13	Porto	1
13	14	Santarém	1
14	15	Setúbal	1
15	16	Viana do Castelo	1
16	17	Vila Real	1
17	18	Viseu	1
18	7	Évora	1

[Close](#)

# Group By 2

Group by

Step name **Group by 2**

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 DC_2
2 District_B

Aggregates:

▼

Name	Subject	Type	Value	<input type="button" value="Get lookup fields"/>
1 DC_2	DC_2	Number of Distinct Values (N)		

# Group By 2

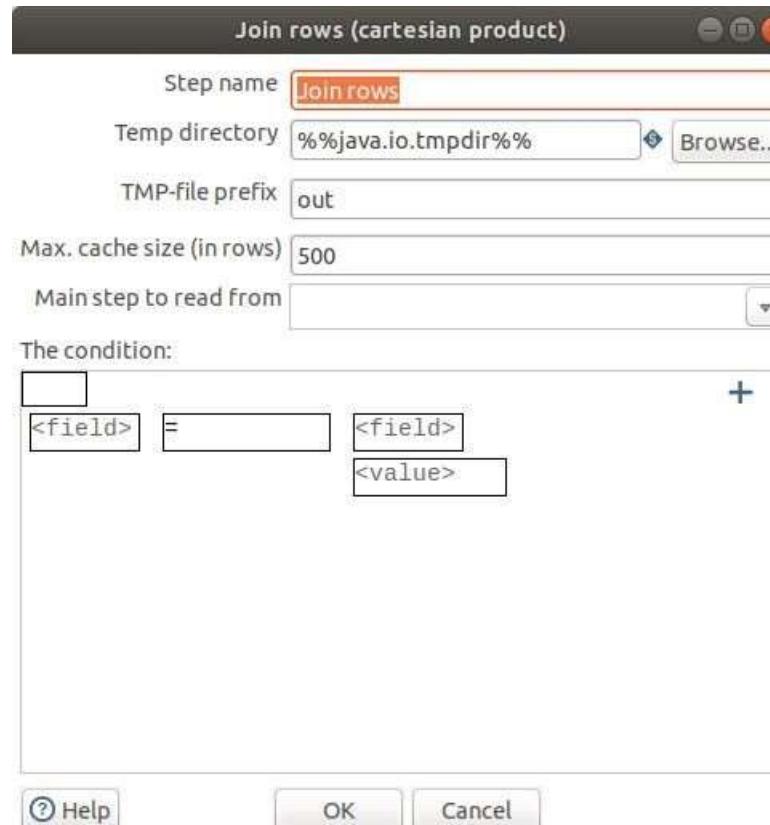
Examine preview data

Rows of step: Group by 2 (18 rows)

	DC_2	District_B	DC_2_1
1	1	AVEIRO	1
2	2	BEJA	1
3	3	BRAGA	1
4	4	BRAGANCA	1
5	5	CASTELO BRANCO	1
6	6	COIMBRA	1
7	7	EVORA	1
8	8	FARO	1
9	9	GUARDA	1
10	10	LEIRIA	1
11	11	LISBOA	1
12	12	PORTALEGRE	1
13	13	PORTO	1
14	14	SANTAREM	1
15	15	SETUBAL	1
16	16	VIANA DO CASTELO	1
17	17	VILA REAL	1
18	18	VISEU	1

[Close](#)

# Join Rows



# Join Rows

Examine preview data

Rows of step: Join rows (324 rows)

	DC_1	District_A	DC_1_1	DC_2	District_B	DC_2_1
1	1	Aveiro	1	1	AVEIRO	1
2	1	Aveiro		2	BEJA	1
3	1	Aveiro		3	BRAGA	1
4	1	Aveiro		4	BRAGANCA	1
5	1	Aveiro		5	CASTELO BRANCO	1
6	1	Aveiro		6	COIMBRA	1
7	1	Aveiro		7	EVORA	1
8	1	Aveiro		8	FARO	1
9	1	Aveiro		9	GUARDA	1
10	1	Aveiro		10	LEIRIA	1
11	1	Aveiro		11	LISBOA	1
12	1	Aveiro		12	PORTALEGRE	1
13	1	Aveiro		13	PORTO	1
14	1	Aveiro		14	SANTAREM	1
15	1	Aveiro		15	SETUBAL	1
16	1	Aveiro		16	VIANA DO CASTELO	1
17	1	Aveiro		17	VILA REAL	1
18	1	Aveiro		18	VISEU	1
19	2	Beja	1	1	AVEIRO	1
20	2	Beja	1	2	BEJA	1
21	2	Beja	1	3	BRAGA	1
22	2	Beja	1	4	BRAGANCA	1
23	2	Beja	1	5	CASTELO BRANCO	1
24	2	Beja	1	6	COIMBRA	1

[Close](#)

# Calculator

Calculator

Step name  
 

Throw an error on non existing files

Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimals
1	needleman	NeedlemanWunsch distance between String A and String B	District_A	District_B		Integer			N		

 Help  OK  Cancel

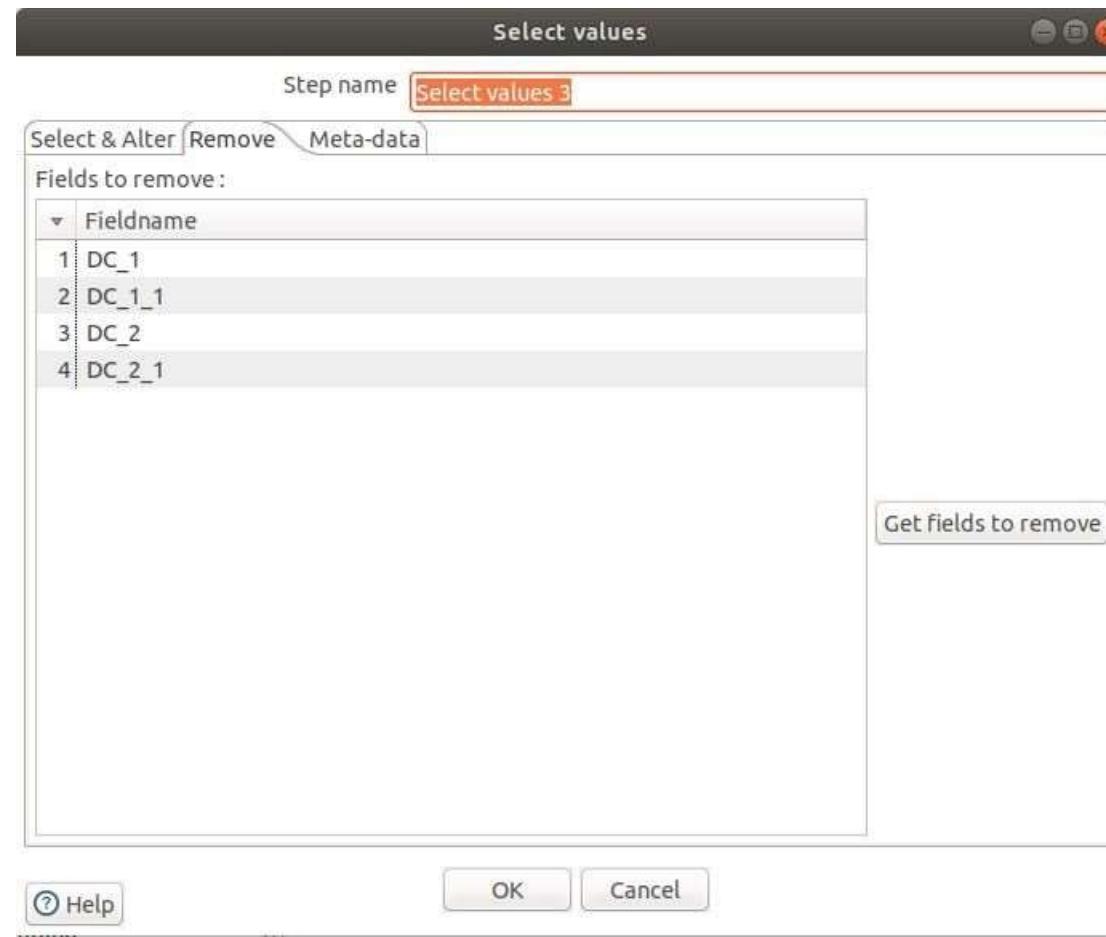
# Calculator

Examine preview data

Rows of step: Calculator (324 rows)

	DC_1	District_A	DC_1_1	DC_2	District_B	DC_2_1		needleman
1	1	Aveiro	1	1	AVEIRO	1		0
2	1	Aveiro	1	2	BEJA	1		-5
3	1	Aveiro	1	3	BRAGA	1		-6
4	1	Aveiro	1	4	BRAGANCA	1		-7
5	1	Aveiro	1	5	CASTELO BRANCO	1		-10
6	1	Aveiro	1	6	COIMBRA	1		-6
7	1	Aveiro	1	7	EVORA	1		-4
8	1	Aveiro	1	8	FARO	1		-4
9	1	Aveiro	1	9	GUARDA	1		-6
10	1	Aveiro	1	10	LEIRIA	1		-4
11	1	Aveiro	1	11	LISBOA	1		-6
12	1	Aveiro	1	12	PORTALEGRE	1		-7
13	1	Aveiro	1	13	PORTO	1		-5
14	1	Aveiro	1	14	SANTAREM	1		-6
15	1	Aveiro	1	15	SETUBAL	1		-7
16	1	Aveiro	1	16	VIANA DO CASTELO	1		-14
17	1	Aveiro	1	17	VILA REAL	1		-8
18	1	Aveiro	1	18	VISEU	1		-5
19	2	Beja	1	1	AVEIRO	1		-5
20	2	Beja	1	2	BEJA	1		0
21	2	Beja	1	3	BRAGA	1		-3
22	2	Beja	1	4	BRAGANCA	1		-6
23	2	Beja	1	5	CASTELO BRANCO	1		-12
24	2	Beja	1	6	COIMBRA	1		-6

# Select Values 3



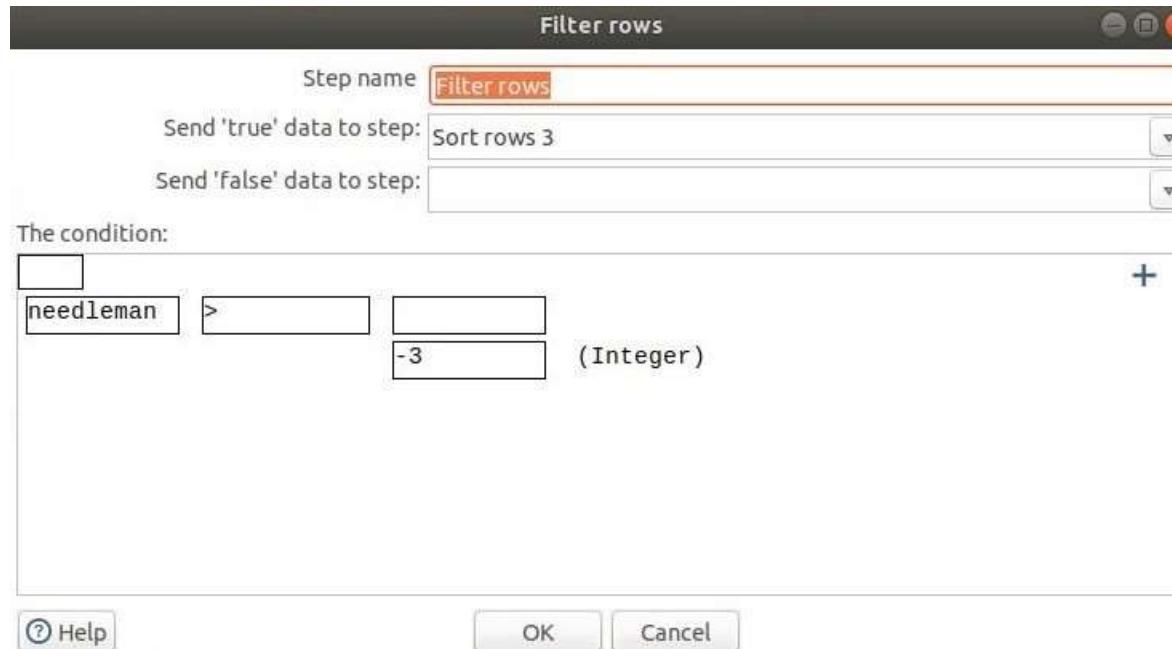
# Select Values 3

Examine preview data

Rows of step: Select values 3 (324 rows)

	District_A	District_B	needleman
1	Aveiro	AVEIRO	0
2	Aveiro	BEJA	-5
3	Aveiro	BRAGA	-6
4	Aveiro	BRAGANCA	-7
5	Aveiro	CASTELO BRANCO	-10
6	Aveiro	COIMBRA	-6
7	Aveiro	EVORA	-4
8	Aveiro	FARO	-4
9	Aveiro	GUARDA	-6
10	Aveiro	LEIRIA	-4
11	Aveiro	LISBOA	-6
12	Aveiro	PORTALEGRE	-7
13	Aveiro	PORTO	-5
14	Aveiro	SANTAREM	-6
15	Aveiro	SETUBAL	-7
16	Aveiro	VIANA DO CASTELO	-14
17	Aveiro	VILA REAL	-8
18	Aveiro	VISEU	-5
19	Beja	AVEIRO	-5
20	Beja	BEJA	0
21	Beja	BRAGA	-3
22	Beja	BRAGANCA	-6
23	Beja	CASTELO BRANCO	-12
24	Beja	COIMBRA	-6

# Filter Rows



# Filter Rows

Examine preview data

Rows of step: Filter rows (18 rows)

	District_A	District_B	needleman
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	PORCALEGRE	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

[Close](#)

# Sort Rows 3

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

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_A	Y	N	Y	0	N

# Sort Rows 3

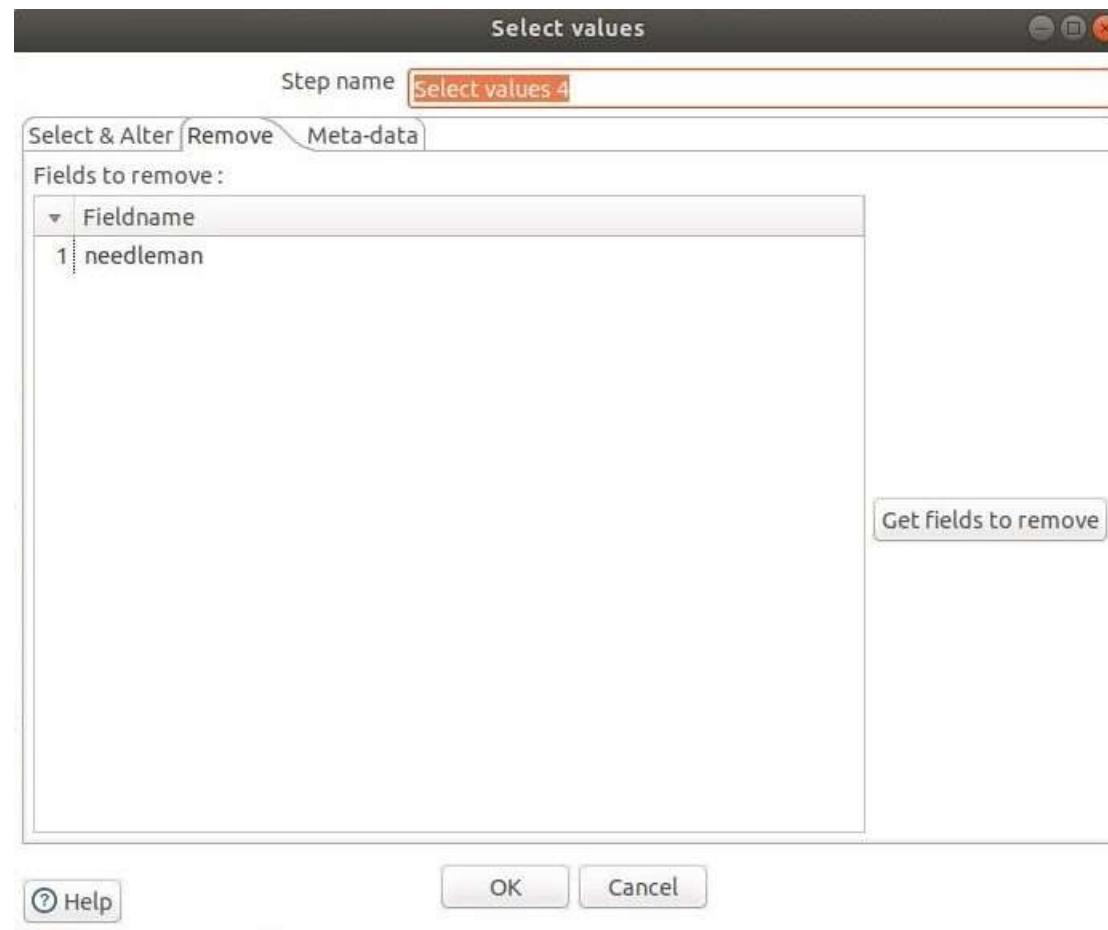
Examine preview data

Rows of step: Sort rows 3 (18 rows)

#	District_A	District_B	needleman
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	-1
8	Faro	FARO	0
9	Guarda	GUARDA	0
10	Leiria	LEIRIA	0
11	Lisboa	LISBOA	0
12	Portalegre	PORCALEGRE	0
13	Porto	PORTO	0
14	Santarém	SANTAREM	-1
15	Setúbal	SETUBAL	-1
16	Viana do Castelo	VIANA DO CASTELO	0
17	Vila Real	VILA REAL	0
18	Viseu	VISEU	0

[Close](#)

# Select Values 4



# Select Values 4

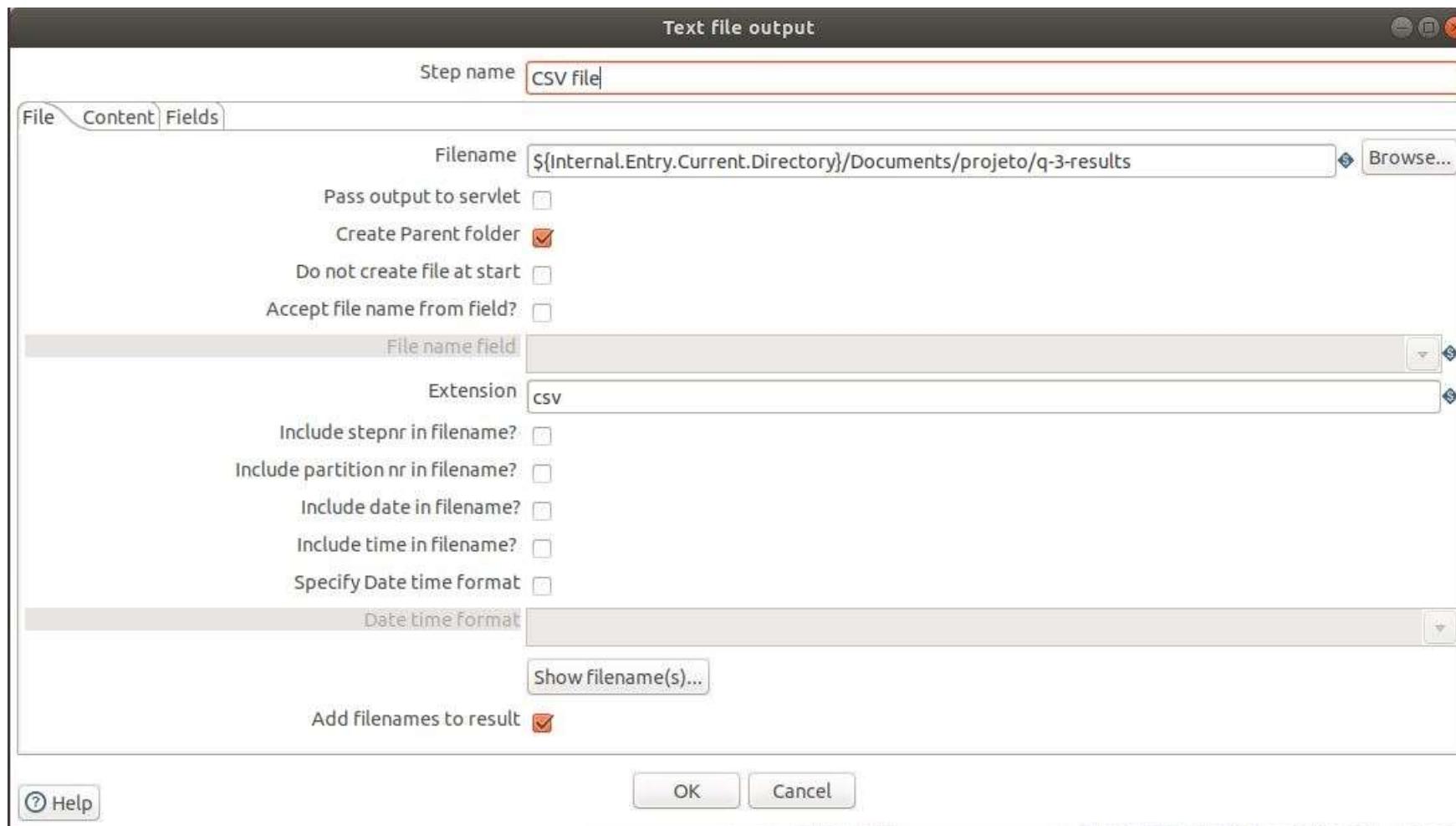
Examine preview data

Rows of step: Select values 4 (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	PORCALEGRE
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](#)

# CSV file



# CSV file

Examine preview data

Rows of step: CSV File (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	PORCALEGRE
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

# Output file

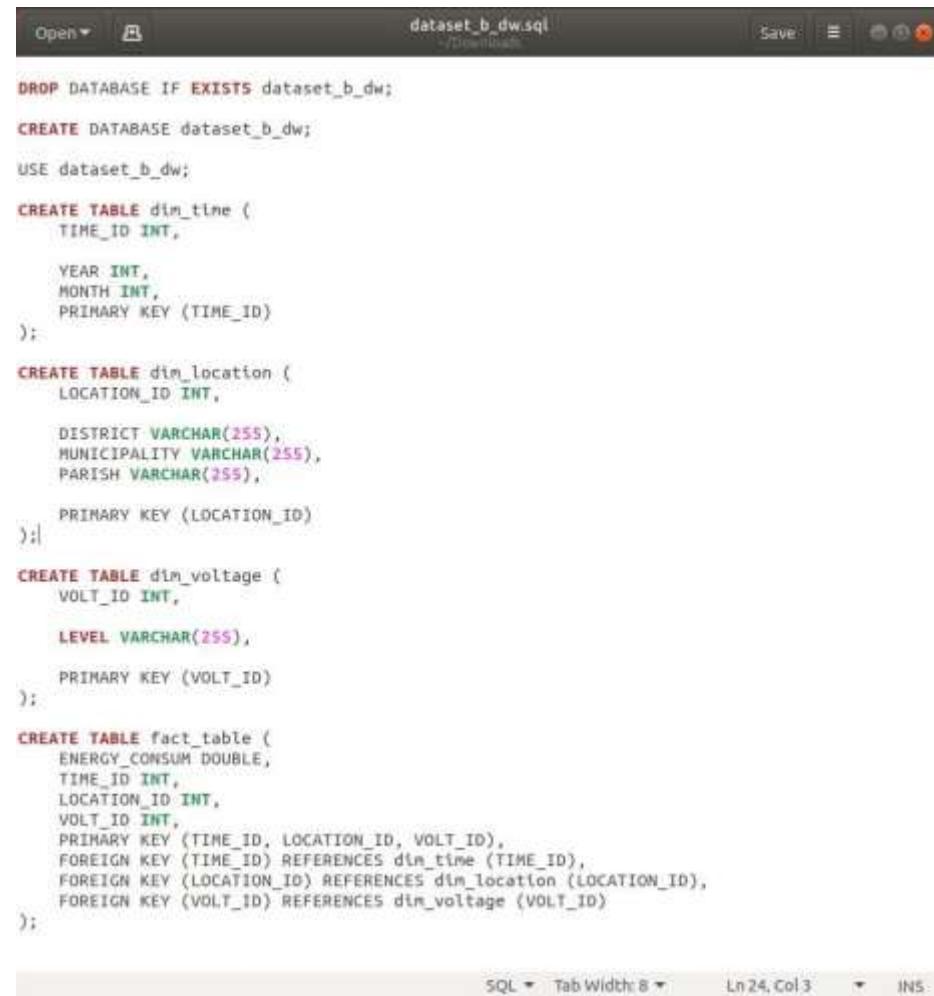
File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation Sans 10 a a a a a a District\_A

	A	B	C	D	E	F
1	District A	District_B				
2	Aveiro	AVEIRO				
3	Beja	BEJA				
4	Braga	BRAGA				
5	Braganca	BRAGANCA				
6	Castelo Branco	CASTELO BRANCO				
7	Coimbra	COIMBRA				
8	Evora	EVORA				
9	Faro	FARO				
10	Guarda	GUARDA				
11	Leiria	LEIRIA				
12	Lisboa	LISBOA				
13	Portalegre	PORTALEGRE				
14	Porto	PORTO				
15	Santarém	SANTAREM				
16	Setubal	SETUBAL				
17	Viana do Castelo	VIANA DO CASTELO				
18	Vila Real	VILA REAL				
19	Viseu	VISEU				
20						
21						
22						
23						
24						

04

# Script



The screenshot shows a SQL editor window with the following details:

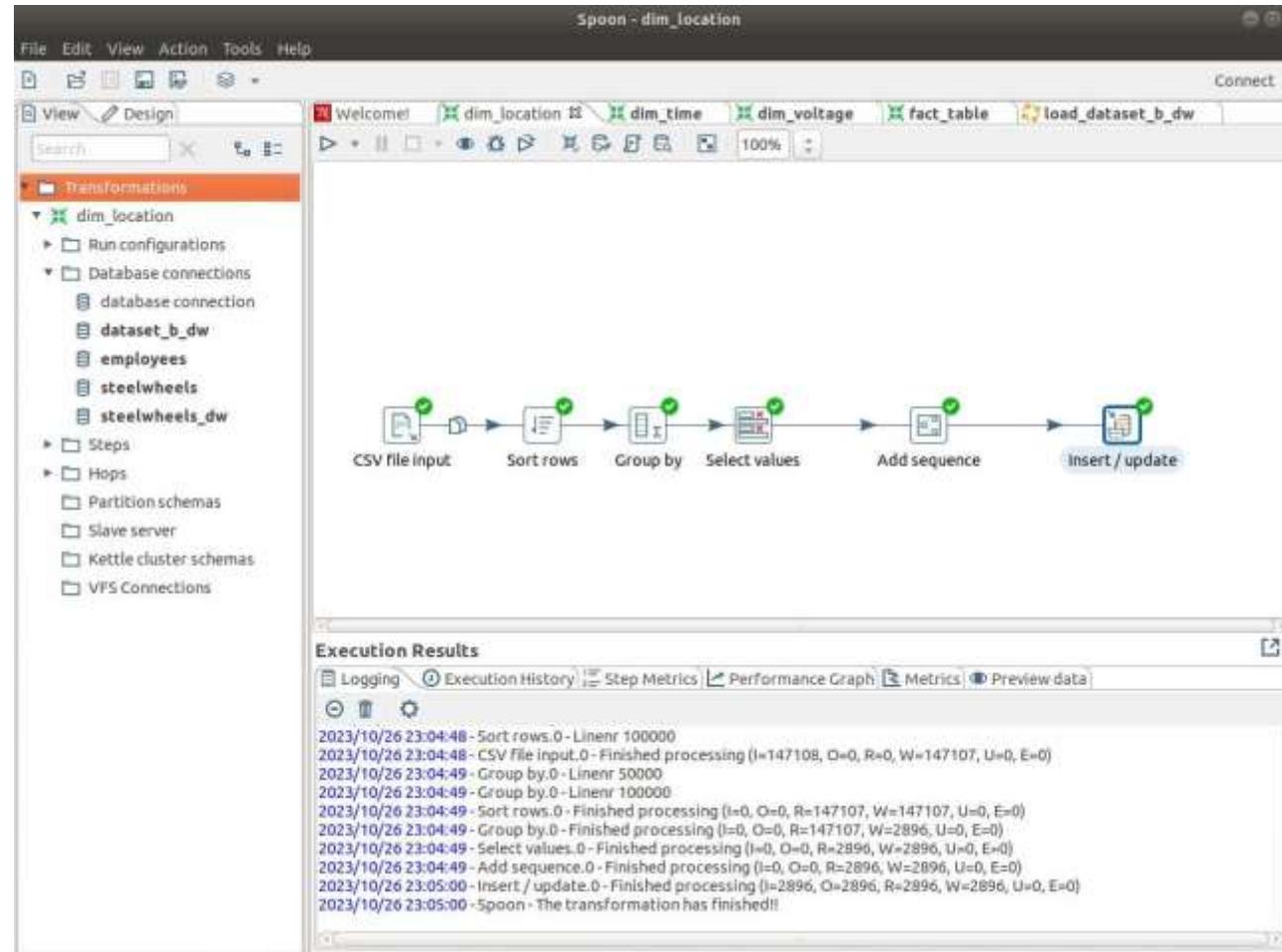
- Title Bar:** The title bar displays "dataset\_b\_dw.sql".
- Toolbar:** Standard save, undo, redo, and other file operations.
- Code Area:** The main area contains the following SQL script:

```
Open ▾ dataset_b_dw.sql Save: ▾
DROP DATABASE IF EXISTS dataset_b_dw;
CREATE DATABASE dataset_b_dw;
USE dataset_b_dw;
CREATE TABLE dim_time (
    TIME_ID INT,
    YEAR INT,
    MONTH INT,
    PRIMARY KEY (TIME_ID)
);
CREATE TABLE dim_location (
    LOCATION_ID INT,
    DISTRICT VARCHAR(255),
    MUNICIPALITY VARCHAR(255),
    PARISH VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);
CREATE TABLE dim_voltage (
    VOLT_ID INT,
    LEVEL VARCHAR(255),
    PRIMARY KEY (VOLT_ID)
);
CREATE TABLE fact_table (
    ENERGY_CONSUM DOUBLE,
    TIME_ID INT,
    LOCATION_ID INT,
    VOLT_ID INT,
    PRIMARY KEY (TIME_ID, LOCATION_ID, VOLT_ID),
    FOREIGN KEY (TIME_ID) REFERENCES dim_time (TIME_ID),
    FOREIGN KEY (LOCATION_ID) REFERENCES dim_location (LOCATION_ID),
    FOREIGN KEY (VOLT_ID) REFERENCES dim_voltage (VOLT_ID)
);
```

At the bottom of the editor, there are status indicators: "SQL", "Tab Width: 8", "Ln 24, Col 3", and "INS".

05

# Location dimension



# CSV file input

CSV file Input

Step name: CSV file input

Filename: \$internal.Entry.Current.Directory)/dataset\_b.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:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 District	String		16		\$	,		none
2 Municipality	String		27		\$	,		none
3 parish	String		30		\$	,		none

# CSV file input

Examine preview data

Rows of step: CSV file input (1000 rows)

	District	Municipality	parish
1	LISBOA	Amadora	ENCOSTA DO SOL
2	PORTO	Vila do Conde	MINDELO
3	AVEIRO	Espinho	ESPINHO
4	CASTELO BRANCO	Penamacor	SALVADOR
5	GUARDA	Guarda	FERNAO JOANES
6	GUARDA	Sabugal	VALE DE ESPINHO
7	LEIRIA	Óbidos	GAEIRAS
8	LEIRIA	Porto de Mós	MIRA DE AIRE
9	LISBOA	Amadora	ENCOSTA DO SOL
10	VIANA DO CASTELO	Caminha	UF GONDAR E ORBACEM
11	VIANA DO CASTELO	Caminha	UF MOLEDO E CRISTELO
12	VIDEU	Tondela	UF MOURAZ VILA NOVA RAINHA
13	AVEIRO	Aveiro	OLIVEIRINHA
14	AVEIRO	Ilhavo	ILHAVO (SAO SALVADOR)
15	AVEIRO	Oliveira do Bairro	OIA
16	BRAGA	Fafe	UF CEPAES E FAREJA
17	FARO	Portimão	ALVOR
18	GUARDA	Guarda	FERNAO JOANES
19	LISBOA	Lourinhã	UF MIRAGAIA E MARTELEIRA
20	PORTO	Santo Tirso	AVES
21	SANTAREM	Mação	CARVOEIRO
22	VIDEU	Vouzela	UF CAMBRA CARVALHAL VERMILHAS
23	COIMBRA	Arganil	SARZEDO
24	COIMBRA	Montemor-o-Velho	UF ABRUNHEIRA VERRIDE VN BARCA

# Sort rows

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

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	Y	N	N	0	N
2	Municipality	Y	N	N	0	N
3	parish	Y	N	N	0	N

# Sort rows

Examine preview data

Rows of step: Sort rows (1000 rows)

	District	Municipality	parish
1	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
2	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
3	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
4	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
5	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
6	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
7	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
8	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
9	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
10	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
11	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
12	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
13	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
14	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
15	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
16	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
17	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
18	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
19	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
20	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
21	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
22	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
23	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
24	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR

# Group by

Group by

Step name **Group by**

Include all rows?

Temporary files directory `%TEMP%`

THP-file profile `tp1`

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		<input type="button" value="Get Fields"/>
1	District	
2	Municipality	
3	parish	

Aggregates :

Name	Subject	Type	Value	<input type="button" value="Get lookup fields"/>
1 District	District	Number of Values (N)		
2 Municipality	Municipality	Number of Values (N)		
3 parish	parish	First value		

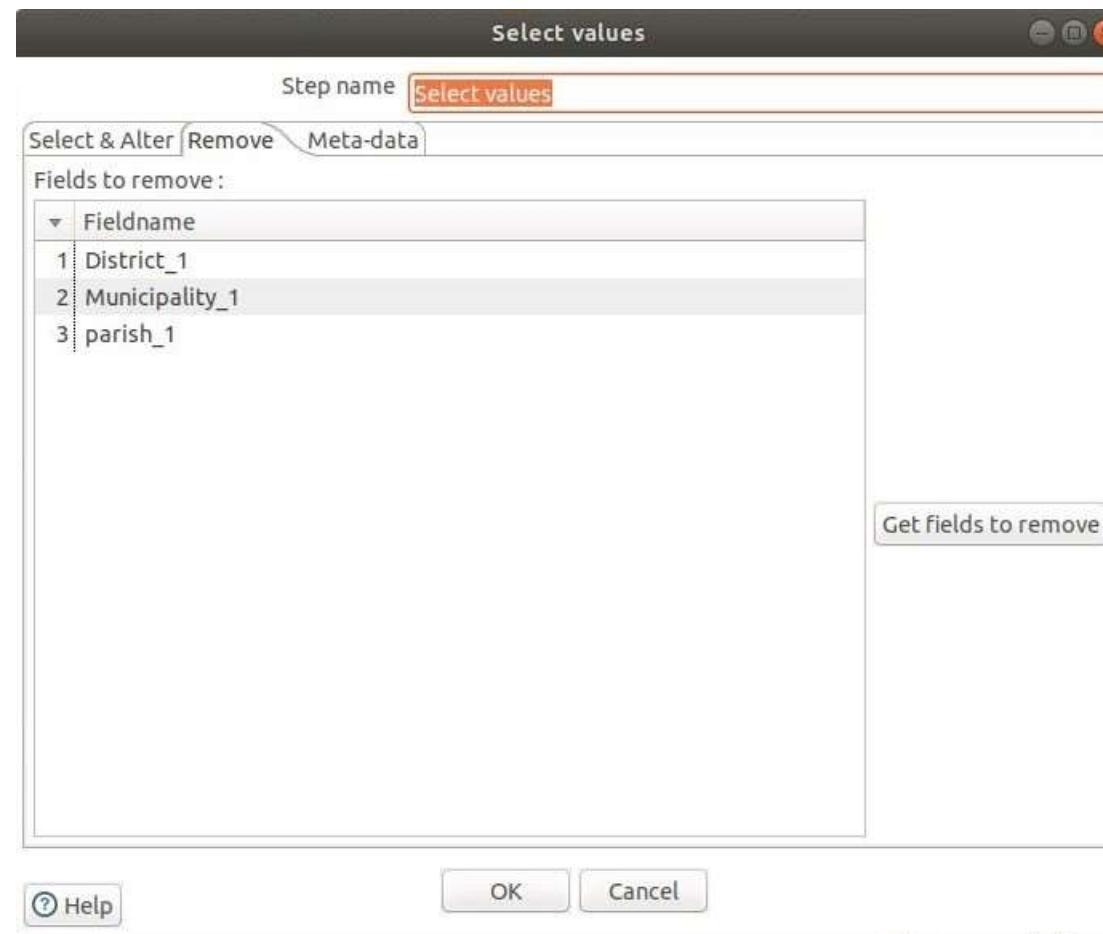
# Group by

Examine preview data

Rows of step: Group by (1000 rows)

#	District	Municipality	parish	District_1	Municipality_1	parish_1
1	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	66	66	ALBERGARIA-A-VELHA E VALMAIOR
2	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	33	33	ALQUERUBIM
3	AVEIRO	Albergaria-a-Velha	ANGEJA	33	33	ANGEJA
4	AVEIRO	Albergaria-a-Velha	BRANCA	66	66	BRANCA
5	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	33	33	RIBEIRA DE FRAGUAS
6	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	66	66	SAO JOAO DE LOURE E FROSSOS
7	AVEIRO	Anadia	AVELAS DE CAMINHO	66	66	AVELAS DE CAMINHO
8	AVEIRO	Anadia	AVELAS DE CIMA	66	66	AVELAS DE CIMA
9	AVEIRO	Anadia	MOITA	33	33	MOITA
10	AVEIRO	Anadia	SANGALHOS	66	66	SANGALHOS
11	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	66	66	SAO LOURENCO DO BAIRRO
12	AVEIRO	Anadia	UF A DA GANDARA BAIRRO E ANCAS	66	66	UF A DA GANDARA BAIRRO E ANCAS
13	AVEIRO	Anadia	UF ARCOS E MOGOFORES	66	66	UF ARCOS E MOGOFORES
14	AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	66	66	UF TAMENGOS AGUIM OIS BAIRRO
15	AVEIRO	Anadia	VILA NOVA DE MONSARROS	33	33	VILA NOVA DE MONSARROS
16	AVEIRO	Anadia	VILARINHO DO BAIRRO	66	66	VILARINHO DO BAIRRO
17	AVEIRO	Arouca	ALVARENGA	33	33	ALVARENGA
18	AVEIRO	Arouca	CHAVE	66	66	CHAVE
19	AVEIRO	Arouca	ESCARIZ	66	66	ESCARIZ
20	AVEIRO	Arouca	FERMEDO	33	33	FERMEDO
21	AVEIRO	Arouca	MANSORES	66	66	MANSORES
22	AVEIRO	Arouca	MOLDES	33	33	MOLDES
23	AVEIRO	Arouca	ROSSAS	33	33	ROSSAS
24	AVEIRO	Arouca	SANTA EULALIA	66	66	SANTA EULALIA

# Select Values



# Select Values

Examine preview data

Rows of step: Select values (1000 rows)

	District	Municipality	parish
1	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR
2	AVEIRO	Albergaria-a-Velha	ALQUERUBIM
3	AVEIRO	Albergaria-a-Velha	ANGEJA
4	AVEIRO	Albergaria-a-Velha	BRANCA
5	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS
6	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS
7	AVEIRO	Anadia	AVELAS DE CAMINHO
8	AVEIRO	Anadia	AVELAS DE CIMA
9	AVEIRO	Anadia	MOITA
10	AVEIRO	Anadia	SANGALHOS
11	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO
12	AVEIRO	Anadia	UF A DA GANDARA BAIRRO E ANCAS
13	AVEIRO	Anadia	UF ARCOS E MOGOFORES
14	AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIRRO
15	AVEIRO	Anadia	VILA NOVA DE MONSARROS
16	AVEIRO	Anadia	VILARINHO DO BAIRRO
17	AVEIRO	Arouca	ALVARENGA
18	AVEIRO	Arouca	CHAVE
19	AVEIRO	Arouca	ESCARIZ
20	AVEIRO	Arouca	FERMEDO
21	AVEIRO	Arouca	MANSORES
22	AVEIRO	Arouca	MOLDES
23	AVEIRO	Arouca	ROSSAS
24	AVEIRO	Arouca	SANTA EULALIA

Close Stop Get more rows

# Add Sequence

Add sequence

Step name **Add sequence**

Name of value **location\_id**

Use a database to generate the sequence

Use DB to get sequence?

Connection **database connection**

Schema name

Sequence name **SEQ**

Use a transformation counter to generate the sequence

Use counter to calculate sequence?

Counter name (optional)

Start at value **0**

Increment by **1**

Maximum value **999999999**

# Add Sequence

Examine preview data

Rows of step: Add sequence (1000 rows)

	District	Municipality	parish	location_id
1	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	0
2	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1
3	AVEIRO	Albergaria-a-Velha	ANGEJA	2
4	AVEIRO	Albergaria-a-Velha	BRANCA	3
5	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	4
6	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	5
7	AVEIRO	Anadia	AVELAS DE CAMINHO	6
8	AVEIRO	Anadia	AVELAS DE CIMA	7
9	AVEIRO	Anadia	MOITA	8
10	AVEIRO	Anadia	SANGALHOS	9
11	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	10
12	AVEIRO	Anadia	UF A DA GANDARA BAIRRO E ANCAS	11
13	AVEIRO	Anadia	UF ARCOS E MOGOFORES	12
14	AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	13
15	AVEIRO	Anadia	VILA NOVA DE MONSARROS	14
16	AVEIRO	Anadia	VILARINHO DO BAIRRO	15
17	AVEIRO	Arouca	ALVARENGA	16
18	AVEIRO	Arouca	CHAVE	17
19	AVEIRO	Arouca	ESCARIZ	18
20	AVEIRO	Arouca	FERMEDO	19
21	AVEIRO	Arouca	MANSORES	20
22	AVEIRO	Arouca	MOLDES	21
23	AVEIRO	Arouca	ROSSAS	22
24	AVEIRO	Arouca	SANTA EULALIA	23

# Insert/Update

Insert / update

Step name **Insert / update**

Connection **dataset\_b\_dw**

Target schema **dataset\_b\_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	<input type="button" value="Get fields"/>
1	LOCATION_ID	=	location_id		

Update fields:

	Table field	Stream field	Update	<input type="button" value="Get update fields"/>
1	DISTRICT	District	Y	
2	MUNICIPALITY	Municipality	Y	
3	PARISH	parish	Y	
4	LOCATION_ID	location_id	Y	

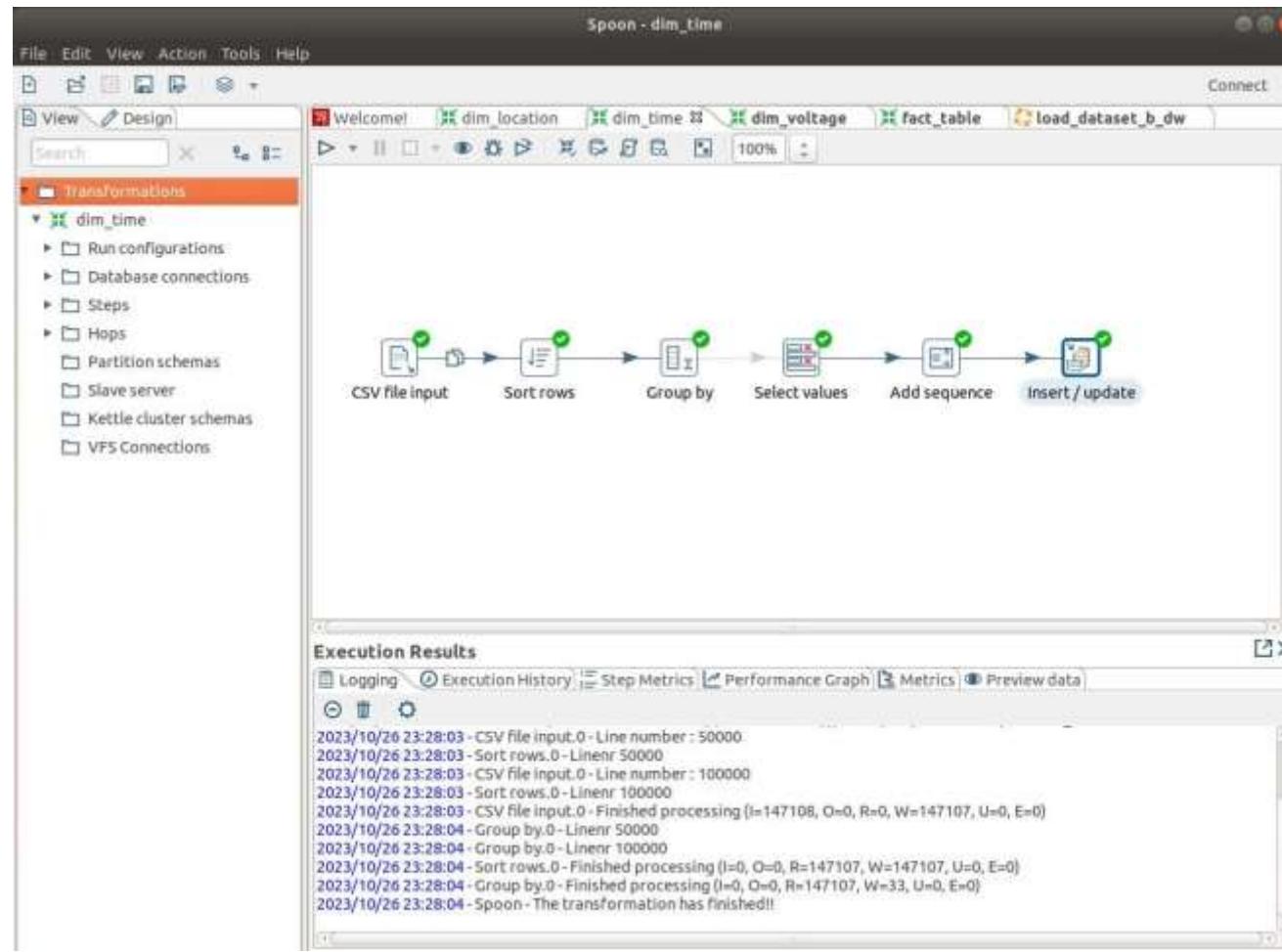
# Insert/Update

Examine preview data

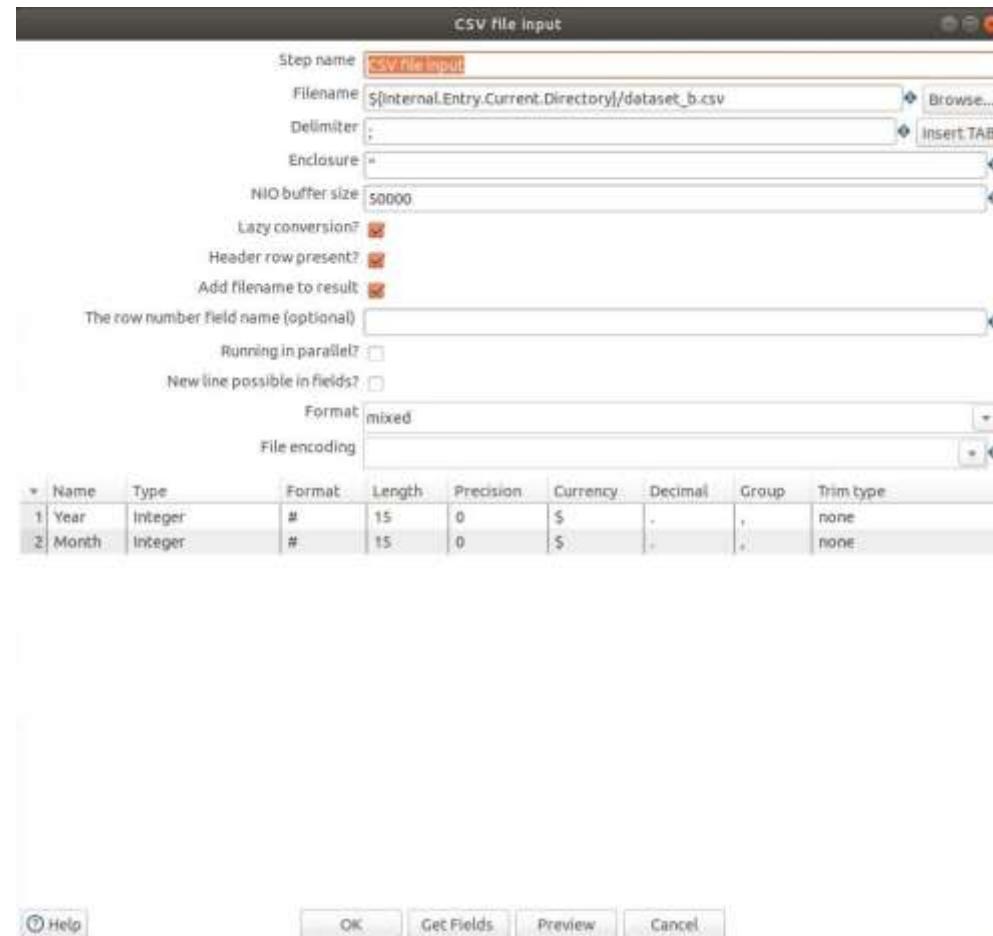
Rows of step: Insert / update (1000 rows)

	District	Municipality	parish	location_id
1	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	0
2	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1
3	AVEIRO	Albergaria-a-Velha	ANGEJA	2
4	AVEIRO	Albergaria-a-Velha	BRANCA	3
5	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	4
6	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	5
7	AVEIRO	Anadia	AVELAS DE CAMINHO	6
8	AVEIRO	Anadia	AVELAS DE CIMA	7
9	AVEIRO	Anadia	MOITA	8
10	AVEIRO	Anadia	SANGALHOS	9
11	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	10
12	AVEIRO	Anadia	UF A DA GANDARA BAIRRO E ANCAS	11
13	AVEIRO	Anadia	UF ARCOS E MOGOFORES	12
14	AVEIRO	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	13
15	AVEIRO	Anadia	VILA NOVA DE MONSARROS	14
16	AVEIRO	Anadia	VILARINHO DO BAIRRO	15
17	AVEIRO	Arouca	ALVARENGA	16
18	AVEIRO	Arouca	CHAVE	17
19	AVEIRO	Arouca	ESCARIZ	18
20	AVEIRO	Arouca	FERMEDO	19
21	AVEIRO	Arouca	MANSORES	20
22	AVEIRO	Arouca	MOLDES	21
23	AVEIRO	Arouca	ROSSAS	22
24	AVEIRO	Arouca	SANTA EULALIA	23

# Time Dimension



# CSV File Input



# CSV File Input

Examine preview data

Rows of step: CSV file input (1000 rows)

	Year	Month
1	2022	7
2	2022	7
3	2022	8
4	2022	8
5	2022	8
6	2022	8
7	2022	8
8	2022	8
9	2022	8
10	2022	8
11	2022	8
12	2022	8
13	2022	9
14	2022	9
15	2022	9
16	2022	9
17	2022	9
18	2022	9
19	2022	9
20	2022	9
21	2022	9
22	2022	9
23	2022	10
24	2022	10

Close Stop Get more rows

# Sort Rows

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

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	Year	Y	N	N	0	N
2	Month	Y	N	N	0	N

# Sort Rows

Examine preview data

Rows of step: Sort rows (1000 rows)

	Year	Month
1	2020	11
2	2020	11
3	2020	11
4	2020	11
5	2020	11
6	2020	11
7	2020	11
8	2020	11
9	2020	11
10	2020	11
11	2020	11
12	2020	11
13	2020	11
14	2020	11
15	2020	11
16	2020	11
17	2020	11
18	2020	11
19	2020	11
20	2020	11
21	2020	11
22	2020	11
23	2020	11
24	2020	11

# Group by

Group by

Step name **Group by**

Include all rows?

Temporary files directory **/tmp/sooopdir/%**

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	Year
2	Month

Aggregates:

Name	Subject	Type	Value	<input type="button" value="Get lookup fields"/>
1	Year	Year	Number of Values (N)	
2	Month	Month	First value	

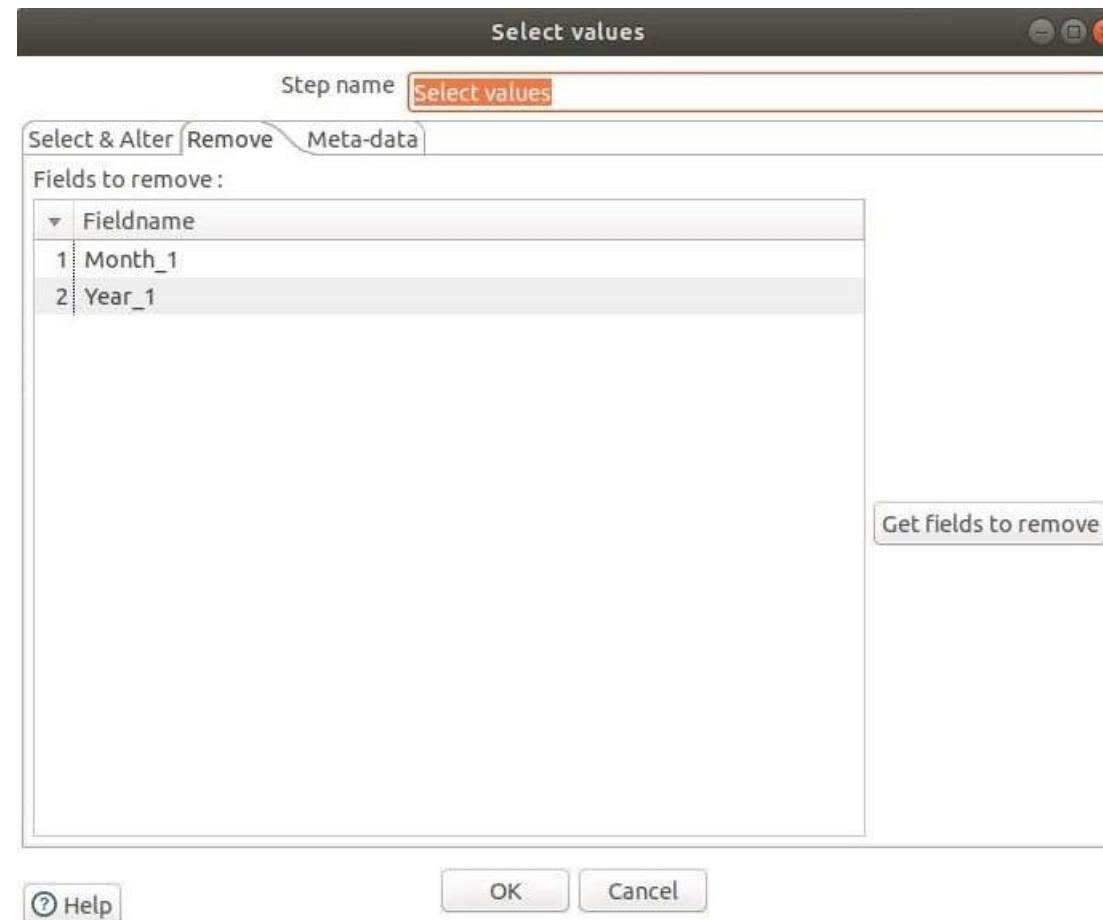
# Group by

Examine preview data

Rows of step: Group by (33 rows)

	Year	Month	Year_1	Month_1
1	2020	11	4457	11
2	2020	12	4457	12
3	2021	1	4457	1
4	2021	2	4455	2
5	2021	3	4456	3
6	2021	4	4459	4
7	2021	5	4457	5
8	2021	6	4456	6
9	2021	7	4455	7
10	2021	8	4457	8
11	2021	9	4459	9
12	2021	10	4458	10
13	2021	11	4460	11
14	2021	12	4460	12
15	2022	1	4460	1
16	2022	2	4461	2
17	2022	3	4464	3
18	2022	4	4464	4
19	2022	5	4461	5
20	2022	6	4461	6
21	2022	7	4464	7
22	2022	8	4466	8
23	2022	9	4464	9
24	2022	10	4467	10

# Select Values



# Select Values

Examine preview data

Rows of step: Select values (33 rows)

	Year	Month
1	2020	11
2	2020	12
3	2021	1
4	2021	2
5	2021	3
6	2021	4
7	2021	5
8	2021	6
9	2021	7
10	2021	8
11	2021	9
12	2021	10
13	2021	11
14	2021	12
15	2022	1
16	2022	2
17	2022	3
18	2022	4
19	2022	5
20	2022	6
21	2022	7
22	2022	8
23	2022	9
24	2022	10

# Add Sequence

Add sequence

Step name **Add sequence**

Name of value **time\_id**

Use a database to generate the sequence

Use DB to get sequence?

Connection **database connection**

Schema name

Sequence name **SEQ**

Use a transformation counter to generate the sequence

Use counter to calculate sequence?

Counter name (optional)

Start at value **0**

Increment by **1**

Maximum value **999999999**

# Add Sequence

Examine preview data

Rows of step: Add sequence (33 rows)

	Year	Month	time_id
1	2020	11	0
2	2020	12	1
3	2021	1	2
4	2021	2	3
5	2021	3	4
6	2021	4	5
7	2021	5	6
8	2021	6	7
9	2021	7	8
10	2021	8	9
11	2021	9	10
12	2021	10	11
13	2021	11	12
14	2021	12	13
15	2022	1	14
16	2022	2	15
17	2022	3	16
18	2022	4	17
19	2022	5	18
20	2022	6	19
21	2022	7	20
22	2022	8	21
23	2022	9	22
24	2022	10	23

# Insert/Update

Insert / update

Step name **Insert / update**

Connection **dataset\_b\_dw**

Target schema **dataset\_b\_dw**

Target table **dim\_time**

Commit size **100**

Don't perform any updates:

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

	Table field	Comparator	Stream field1	Stream field2	<input type="button" value="Get Fields"/>
1	TIME_ID	=	time_id		

Update fields:

	Table field	Stream field	Update	<input type="button" value="Get update fields"/>
1	YEAR	Year	Y	<input type="button" value="Edit mapping"/>
2	MONTH	Month	Y	
3	TIME_ID	time_id	Y	

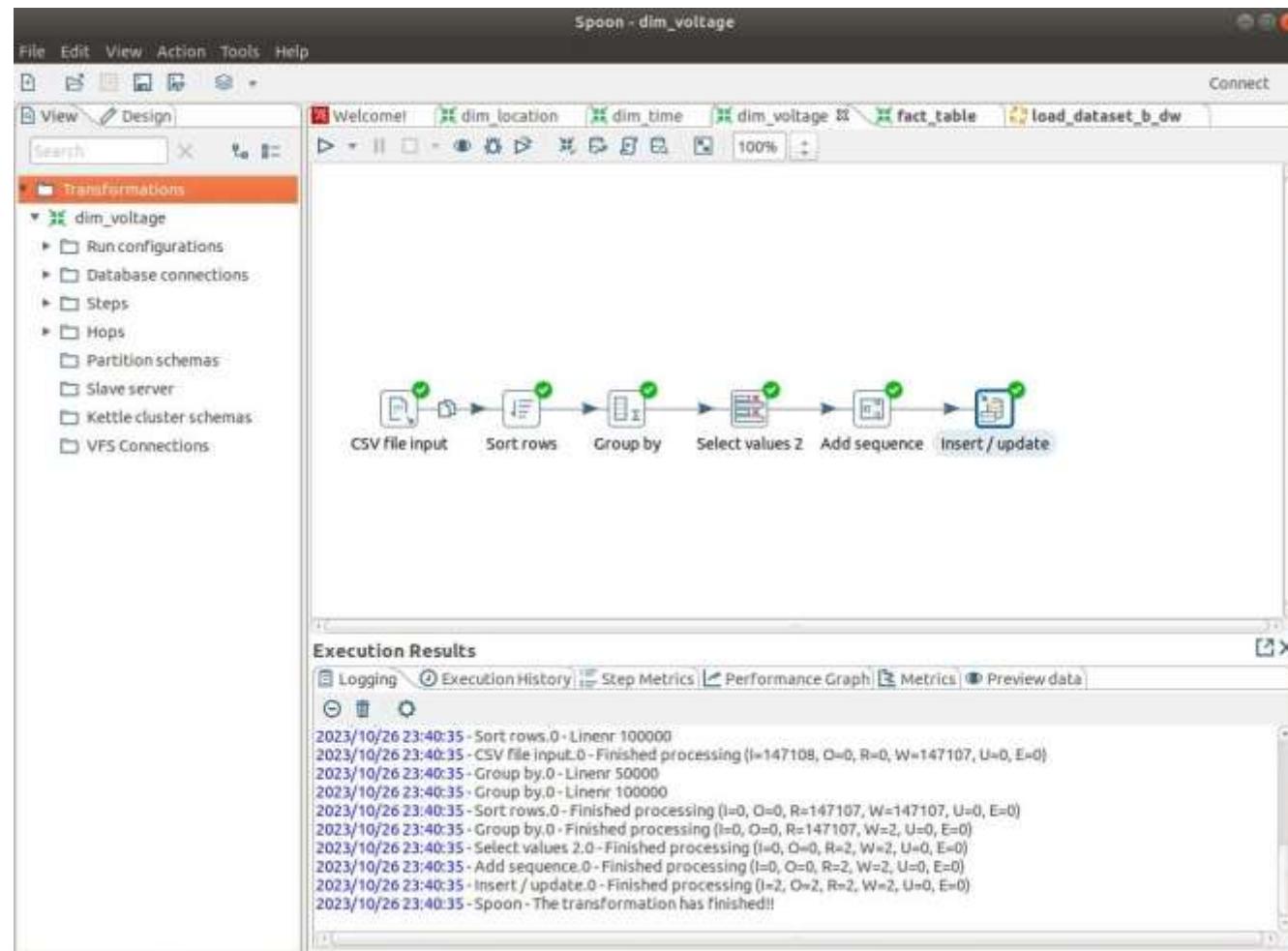
# Insert/Update

Examine preview data

Rows of step: Insert / update (33 rows)

	Year	Month	time_id
1	2020	11	0
2	2020	12	1
3	2021	1	2
4	2021	2	3
5	2021	3	4
6	2021	4	5
7	2021	5	6
8	2021	6	7
9	2021	7	8
10	2021	8	9
11	2021	9	10
12	2021	10	11
13	2021	11	12
14	2021	12	13
15	2022	1	14
16	2022	2	15
17	2022	3	16
18	2022	4	17
19	2022	5	18
20	2022	6	19
21	2022	7	20
22	2022	8	21
23	2022	9	22
24	2022	10	23

# Voltage Dimension



# CSV File Input

CSV file input

Step name: CSV file input

Filename: \${Internal.Entry.Current.Directory}/dataset\_b.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:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Voltage level	String		32		\$	,	,	none

# CSV File Input

Examine preview data	
Rows of step: CSV file input (1000 rows)	
▼	Voltage level
1	Baixa Tensão
2	Baixa Tensão
3	Baixa Tensão
4	Baixa Tensão
5	Baixa Tensão
6	Baixa Tensão
7	Baixa Tensão
8	Muito Alta, Alta e Média Tensões
9	Baixa Tensão
10	Baixa Tensão
11	Baixa Tensão
12	Muito Alta, Alta e Média Tensões
13	Baixa Tensão
14	Baixa Tensão
15	Baixa Tensão
16	Baixa Tensão
17	Baixa Tensão
18	Baixa Tensão
19	Baixa Tensão
20	Baixa Tensão
21	Baixa Tensão
22	Baixa Tensão
23	Muito Alta, Alta e Média Tensões
24	Baixa Tensão

# Sort rows

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

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	Voltage level	Y	N	N	0	N

# Sort rows

Examine preview data	
Rows of step: Sort rows (1000 rows)	
▼	Voltage level
1	Baixa Tensão
2	Baixa Tensão
3	Baixa Tensão
4	Baixa Tensão
5	Baixa Tensão
6	Baixa Tensão
7	Baixa Tensão
8	Baixa Tensão
9	Baixa Tensão
10	Baixa Tensão
11	Baixa Tensão
12	Baixa Tensão
13	Baixa Tensão
14	Baixa Tensão
15	Baixa Tensão
16	Baixa Tensão
17	Baixa Tensão
18	Baixa Tensão
19	Baixa Tensão
20	Baixa Tensão
21	Baixa Tensão
22	Baixa Tensão
23	Baixa Tensão
24	Baixa Tensão

# Group By

Group by

Step name **Group by**

include all rows?

Temporary files directory `%TEMP%`

Table prefix `grp`

Add line number, record in each group

Line number field name

Always give back a result row

The fields that make up the group:

<input type="checkbox"/> Group field	<input type="button" value="Get Fields"/>
1   Voltage level	

Aggregates:

Name	Subject	Type	Value	<input type="button" value="Get lookup Fields"/>
1   Voltage level	Voltage level	First value		

# Group By

Examine preview data (Close)

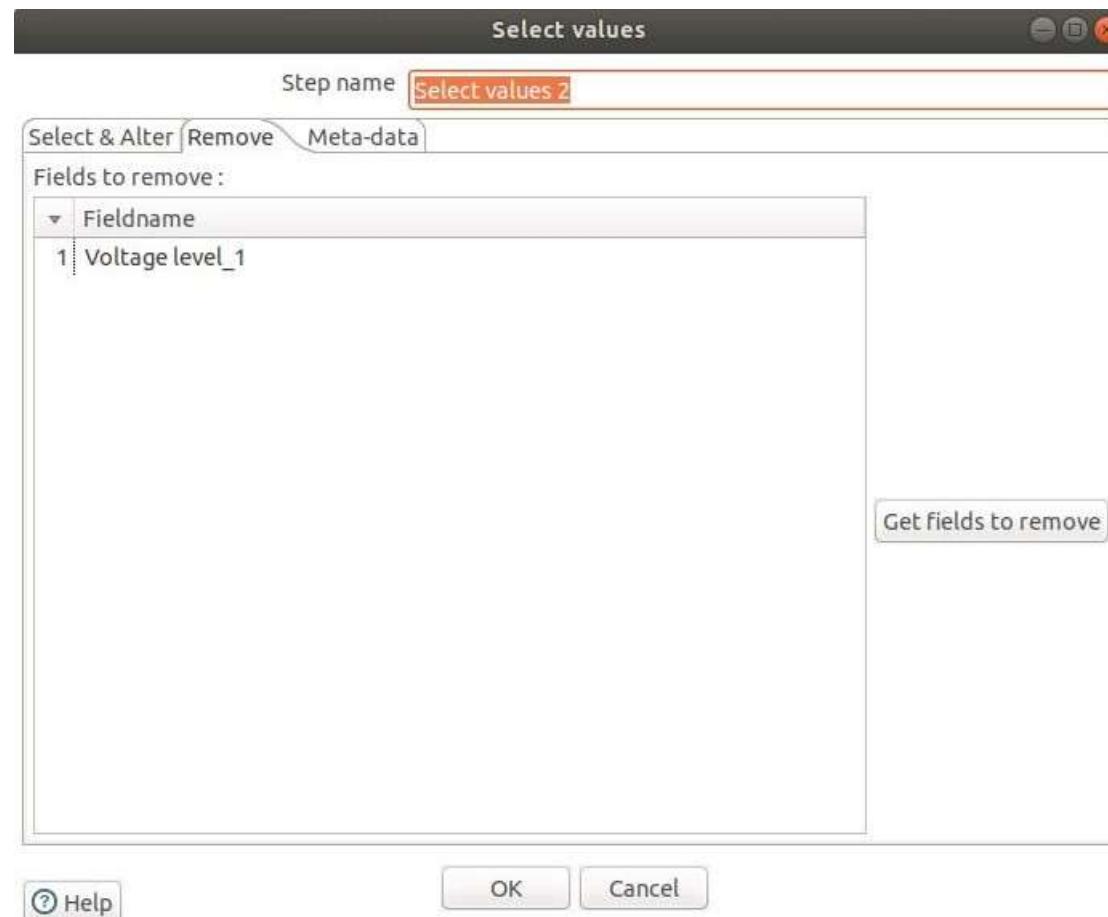
Rows of step: Group by (2 rows)

	Voltage level	Voltage level_1
1	Baixa Tensão	Baixa Tensão
2	Muito Alta, Alta e Média Tensões	Muito Alta, Alta e Média Tensões

© 2012 Pentaho Corporation. All rights reserved. Pentaho is a registered trademark of Pentaho Corporation in the United States and other countries.

Close

# Select Values



# Select Values

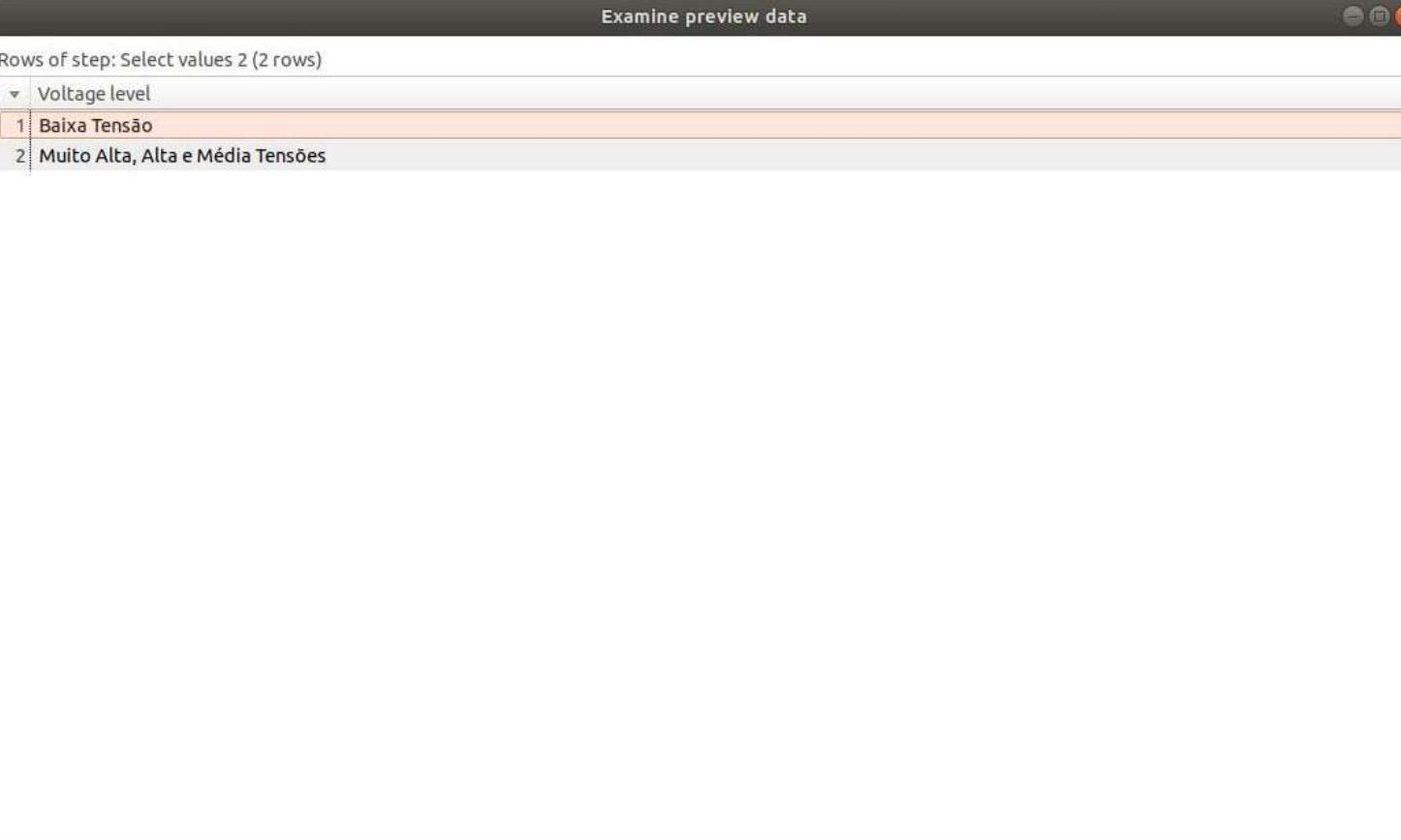
Examine preview data

Rows of step: Select values 2 (2 rows)

Voltage level

1	Baixa Tensão
2	Muito Alta, Alta e Média Tensões

**Close**



# Add Sequence

Add sequence

Step name **Add sequence**

Name of value **volt\_id**

Use a database to generate the sequence

Use DB to get sequence?

Connection

Schema name

Sequence name **SEQ**

Use a transformation counter to generate the sequence

Use counter to calculate sequence?

Counter name (optional)

Start at value **0**

Increment by **1**

Maximum value **999999999**

# Add Sequence

Examine preview data

Rows of step: Add sequence (2 rows)

Voltage level	volt_id
1 Baixa Tensão	0
2 Muito Alta, Alta e Média Tensões	1

**Close**

# Insert/Update

Insert / update

Step name **Insert / update**

Connection **dataset\_b\_dw**

Target schema **dataset\_b\_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	<input type="button" value="Get fields"/>
1	VOLT_ID	=	volt_id		

Update fields:

	Table field	Stream field	Update	<input type="button" value="Get update fields"/>
1	LEVEL	Voltage level	Y	<input type="button" value="Edit mapping"/>
2	VOLT_ID	volt_id	Y	

# Insert/Update

Examine preview data

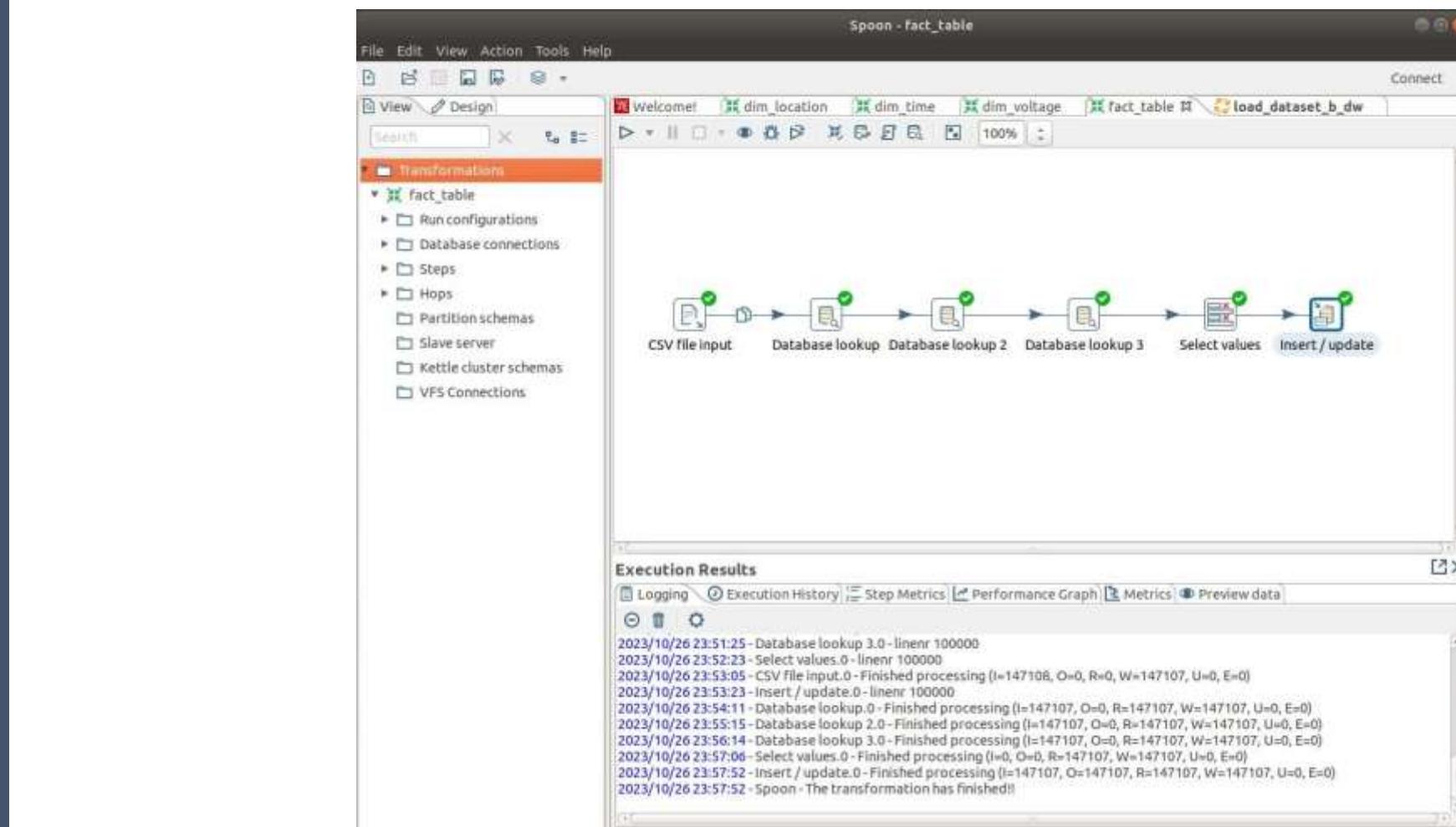
Rows of step: Insert / update (2 rows)

Voltage level	volt_id
1 Baixa Tensão	0
2 Muito Alta, Alta e Média Tensões	1

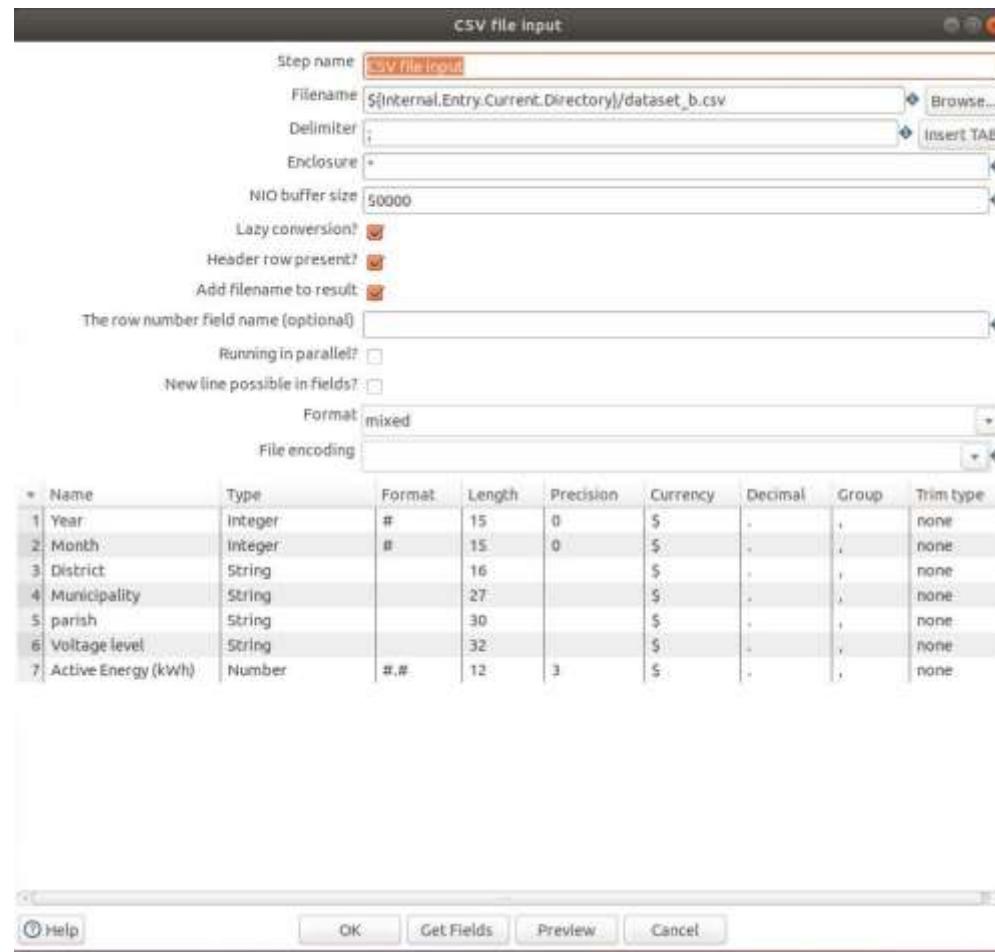
[Close](#)

06

# Fact Table



# CSV File Input



# CSV File Input

Examine preview data

Rows of step: CSV file input (1000 rows)

	Year	Month	District	Municipality	parish	Voltage level
1	2022	7	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
2	2022	7	PORTO	Vila do Conde	MINDELO	Baixa Tensão
3	2022	8	AVEIRO	Espinho	ESPINHO	Baixa Tensão
4	2022	8	CASTELO BRANCO	Penamacor	SALVADOR	Baixa Tensão
5	2022	8	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
6	2022	8	GUARDA	Sabugal	VALE DE ESPINHO	Baixa Tensão
7	2022	8	LEIRIA	Óbidos	GAEIRAS	Baixa Tensão
8	2022	8	LEIRIA	Porto de Mós	MIRA DE AIRE	Muito Alta, Alta e Média Tensões
9	2022	8	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
10	2022	8	VIANA DO CASTELO	Caminha	UF GONDAR E ORBACEM	Baixa Tensão
11	2022	8	VIANA DO CASTELO	Caminha	UF MOLEDO E CRISTELO	Baixa Tensão
12	2022	8	VISEU	Tondela	UF MOURAZ VILA NOVA RAINHA	Muito Alta, Alta e Média Tensões
13	2022	9	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão
14	2022	9	AVEIRO	Ílhavo	ILHAZO (SAO SALVADOR)	Baixa Tensão
15	2022	9	AVEIRO	Oliveira do Bairro	OIA	Baixa Tensão
16	2022	9	BRAGA	Fafe	UF CEPAES E FAREJA	Baixa Tensão
17	2022	9	FARO	Portimão	ALVOR	Baixa Tensão
18	2022	9	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
19	2022	9	LISBOA	Lourinhã	UF MIRAGAIA E MARTELEIRA	Baixa Tensão
20	2022	9	PORTO	Santo Tirso	AVES	Baixa Tensão
21	2022	9	SANTAREM	Maçao	CARVOEIRO	Baixa Tensão
22	2022	9	VISEU	Vouzela	UF CAMBRA CARVALHAL VERMILHAS	Baixa Tensão
23	2022	10	COIMBRA	Arganil	SARZEDO	Muito Alta, Alta e Média Tensões
24	2022	10	COIMBRA	Montemor-o-Velho	UF ABRUNHEIRA VERRIDE VN BARCA	Baixa Tensão

# Database Lookup

Database lookup

Step name **Database lookup**

Connection **dataset\_b\_dw**

Lookup schema **dataset\_b\_dw**

Lookup table **dim\_voltage**

Enable cache?

Cache size in rows (0=cache everything) **0**

Load all data from table

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

Table field	Comparator	Field1	Field2
1 LEVEL	=	Voltage level	

Values to return from the lookup table :

Field	New name	Default	Type
1 VOLT_ID			Integer

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

# Database Lookup

Examine preview data

Rows of step: Database lookup (1000 rows)

	Year	Month	District	Municipality	parish	Voltage level
1	2022	7	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
2	2022	7	PORTO	Vila do Conde	MINDELO	Baixa Tensão
3	2022	8	AVEIRO	Espinho	ESPINHO	Baixa Tensão
4	2022	8	CASTELO BRANCO	Penamacor	SALVADOR	Baixa Tensão
5	2022	8	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
6	2022	8	GUARDA	Sabugal	VALE DE ESPINHO	Baixa Tensão
7	2022	8	LEIRIA	Óbidos	GAEIRAS	Baixa Tensão
8	2022	8	LEIRIA	Porto de Mós	MIRA DE AIRE	Muito Alta, Alta e Média Tensões
9	2022	8	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
10	2022	8	VIANA DO CASTELO	Caminha	UF GONDAR E ORBACEM	Baixa Tensão
11	2022	8	VIANA DO CASTELO	Caminha	UF MOLEDO E CRISTELO	Baixa Tensão
12	2022	8	VISEU	Tondela	UF MOURAZ VILA NOVA RAINHA	Muito Alta, Alta e Média Tensões
13	2022	9	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão
14	2022	9	AVEIRO	Ílhavo	ILHAZO (SAO SALVADOR)	Baixa Tensão
15	2022	9	AVEIRO	Oliveira do Bairro	OIA	Baixa Tensão
16	2022	9	BRAGA	Fafe	UF CEPAES E FAREJA	Baixa Tensão
17	2022	9	FARO	Portimão	ALVOR	Baixa Tensão
18	2022	9	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
19	2022	9	LISBOA	Lourinhã	UF MIRAGAIA E MARTELEIRA	Baixa Tensão
20	2022	9	PORTO	Santo Tirso	AVES	Baixa Tensão
21	2022	9	SANTAREM	Maçao	CARVOEIRO	Baixa Tensão
22	2022	9	VISEU	Vouzela	UF CAMBRA CARVALHAL VERMILHAS	Baixa Tensão
23	2022	10	COIMBRA	Arganil	SARZEDO	Muito Alta, Alta e Média Tensões
24	2022	10	COIMBRA	Montemor-o-Velho	UF ABRUNHEIRA VERRIDE VN BARCA	Baixa Tensão

# Database Lookup 2

Database lookup

Step name **Database lookup 2**

Connection **dataset\_b\_dw**

Lookup schema **dataset\_b\_dw**

Lookup table **dim\_time**

Enable cache?

Cache size in rows (0=cache everything)

Load all data from table

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

	Table Field	Comparator	Field1	Field2
1	YEAR	=	Year	
2	MONTH	=	Month	

Values to return from the lookup table:

	Field	New name	Default	Type
1	TIME_ID			Integer

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

# Database Lookup 2

Examine preview data

Rows of step: Database lookup 2 (1000 rows)

	Year	Month	District	Municipality	parish	Voltage level
1	2022	7	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
2	2022	7	PORTO	Vila do Conde	MINDELO	Baixa Tensão
3	2022	8	AVEIRO	Espinho	ESPINHO	Baixa Tensão
4	2022	8	CASTELO BRANCO	Penamacor	SALVADOR	Baixa Tensão
5	2022	8	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
6	2022	8	GUARDA	Sabugal	VALE DE ESPINHO	Baixa Tensão
7	2022	8	LEIRIA	Óbidos	GAEIRAS	Baixa Tensão
8	2022	8	LEIRIA	Porto de Mós	MIRA DE AIRE	Muito Alta, Alta e Média Tensões
9	2022	8	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
10	2022	8	VIANA DO CASTELO	Caminha	UF GONDAR E ORBACEM	Baixa Tensão
11	2022	8	VIANA DO CASTELO	Caminha	UF MOLEDO E CRISTELO	Baixa Tensão
12	2022	8	VIDEUA	Tondela	UF MOURAZ VILA NOVA RAINHA	Muito Alta, Alta e Média Tensões
13	2022	9	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão
14	2022	9	AVEIRO	Ilhavo	ILHAZO (SAO SALVADOR)	Baixa Tensão
15	2022	9	AVEIRO	Oliveira do Bairro	OIA	Baixa Tensão
16	2022	9	BRAGA	Fafe	UF CEPAES E FAREJA	Baixa Tensão
17	2022	9	FARO	Portimão	ALVOR	Baixa Tensão
18	2022	9	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
19	2022	9	LISBOA	Lourinhã	UF MIRAGAIA E MARTELEIRA	Baixa Tensão
20	2022	9	PORTO	Santo Tirso	AVES	Baixa Tensão
21	2022	9	SANTAREM	Maçao	CARVOEIRO	Baixa Tensão
22	2022	9	VIDEUA	Vouzela	UF CAMBRA CARVALHAL VERMILHAS	Baixa Tensão
23	2022	10	COIMBRA	Arganil	SARZEDO	Muito Alta, Alta e Média Tensões
24	2022	10	COIMBRA	Montemor-o-Velho	UF ABRUNHEIRA VERRIDE VN BARCA	Baixa Tensão

# Database Lookup 3

Database lookup

Step name: Database lookup 3

Connection: dataset\_b\_dw   

Lookup schema: dataset\_b\_dw   

Lookup table: dim\_location   

Enable cache?

Cache size in rows (0=cache everything)

Load all data from table

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

	Table field	Comparator	Field1	Field2
1	DISTRICT	=	District	
2	MUNICIPALITY	=	Municipality	
3	PARISH	=	parish	

Values to return from the lookup table:

	Field	New name	Default	Type
1	LOCATION_ID			Integer

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

# Database Lookup 3

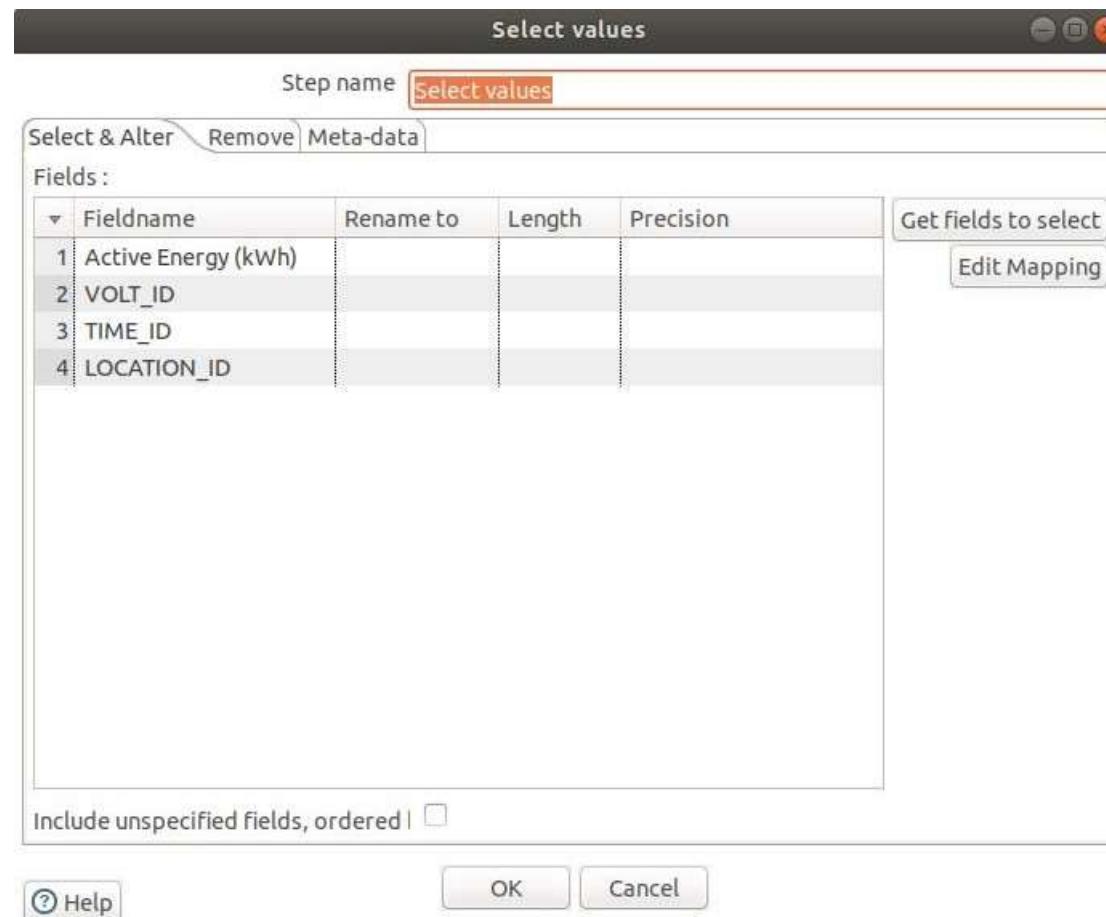
Examine preview data

Rows of step: Database lookup 3 (1000 rows)

	Year	Month	District	Municipality	parish	Voltage level
1	2022	7	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
2	2022	7	PORTO	Vila do Conde	MINDELO	Baixa Tensão
3	2022	8	AVEIRO	Espinho	ESPINHO	Baixa Tensão
4	2022	8	CASTELO BRANCO	Penamacor	SALVADOR	Baixa Tensão
5	2022	8	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
6	2022	8	GUARDA	Sabugal	VALE DE ESPINHO	Baixa Tensão
7	2022	8	LEIRIA	Óbidos	GAEIRAS	Baixa Tensão
8	2022	8	LEIRIA	Porto de Mós	MIRA DE AIRE	Muito Alta, Alta e Média Tensões
9	2022	8	LISBOA	Amadora	ENCOSTA DO SOL	Baixa Tensão
10	2022	8	VIANA DO CASTELO	Caminha	UF GONDAR E ORBACEM	Baixa Tensão
11	2022	8	VIANA DO CASTELO	Caminha	UF MOLEDO E CRISTELO	Baixa Tensão
12	2022	8	VISEU	Tondela	UF MOURAZ VILA NOVA RAINHA	Muito Alta, Alta e Média Tensões
13	2022	9	AVEIRO	Aveiro	OLIVEIRINHA	Baixa Tensão
14	2022	9	AVEIRO	Ílhavo	ILHAVO (SAO SALVADOR)	Baixa Tensão
15	2022	9	AVEIRO	Oliveira do Bairro	OIA	Baixa Tensão
16	2022	9	BRAGA	Fafe	UF CEPAES E FAREJA	Baixa Tensão
17	2022	9	FARO	Portimão	ALVOR	Baixa Tensão
18	2022	9	GUARDA	Guarda	FERNAO JOANES	Baixa Tensão
19	2022	9	LISBOA	Lourinhã	UF MIRAGAIA E MARTELEIRA	Baixa Tensão
20	2022	9	PORTO	Santo Tirso	AVES	Baixa Tensão
21	2022	9	SANTAREM	Maçao	CARVOEIRO	Baixa Tensão
22	2022	9	VISEU	Vouzela	UF CAMBRA CARVALHAL VERMILHAS	Baixa Tensão
23	2022	10	COIMBRA	Arganil	SARZEDO	Muito Alta, Alta e Média Tensões
24	2022	10	COIMBRA	Montemor-o-Velho	UF ABRUNHEIRA VERRIDE VN BARCA	Baixa Tensão

[Close](#) [Stop](#) [Get more rows](#)

# Select Values



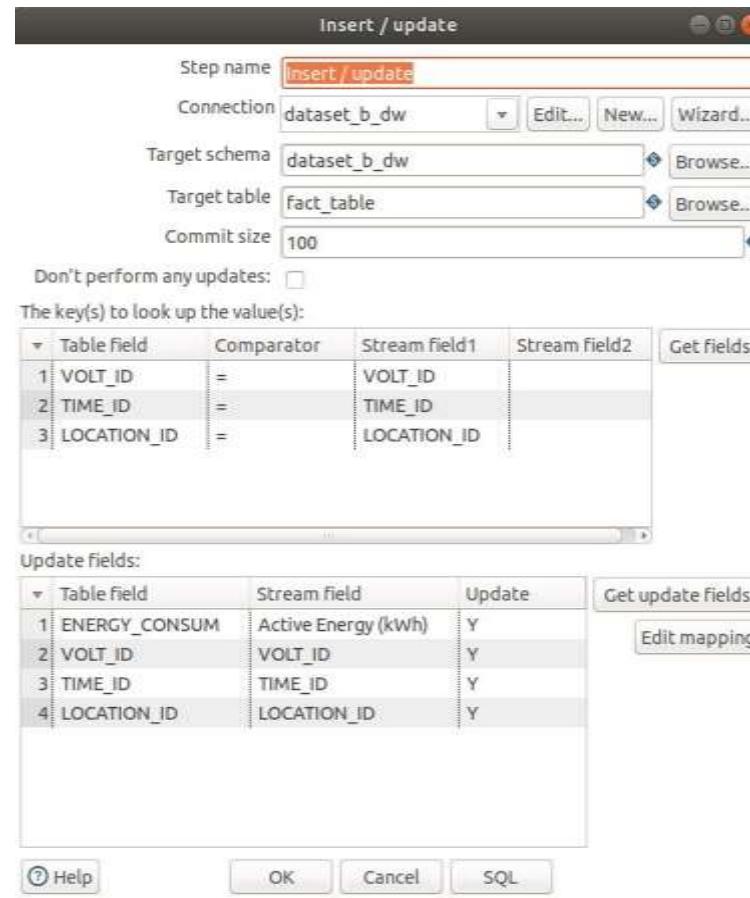
# Select Values

Examine preview data

Rows of step: Select values (1000 rows)

	Active Energy (kWh)	VOLT_ID	TIME_ID	LOCATION_ID
1	3821309.7	0	20	1579
2	757616.4	0	20	1988
3	2244683.1	0	21	49
4	46110.4	0	21	894
5	29907.3	0	21	1306
6	65586.5	0	21	1396
7	500908.2	0	21	1561
8	142244.8	1	21	1552
9	3685384.5	0	21	1579
10	65482.1	0	21	2257
11	599125.9	0	21	2258
12	256920.4	1	21	2852
13	615474.5	0	22	37
14	1922566	0	22	147
15	1191897.7	0	22	80
16	266406.9	0	22	395
17	2662235.5	0	22	1190
18	29762.4	0	22	1306
19	582614.3	0	22	1645
20	1039783.1	0	22	1958
21	71593	0	22	2078
22	178751.3	0	22	2892
23	2319197	1	23	927
24	286168.2	0	23	1012

# Insert/Update



# Insert/Update

Examine preview data

Rows of step: Insert / update (1000 rows)

	Active Energy (kWh)	VOLT_ID	TIME_ID	LOCATION_ID
1	3821309.7	0	20	1579
2	757616.4	0	20	1988
3	2244683.1	0	21	49
4	46110.4	0	21	894
5	29907.3	0	21	1306
6	65586.5	0	21	1396
7	500908.2	0	21	1561
8	142244.8	1	21	1552
9	3685384.5	0	21	1579
10	65482.1	0	21	2257
11	599125.9	0	21	2258
12	256920.4	1	21	2852
13	615474.5	0	22	37
14	1922566	0	22	147
15	1191897.7	0	22	80
16	266406.9	0	22	395
17	2662235.5	0	22	1190
18	29762.4	0	22	1306
19	582614.3	0	22	1645
20	1039783.1	0	22	1958
21	71593	0	22	2078
22	178751.3	0	22	2892
23	2319197	1	23	927
24	286168.2	0	23	1012