

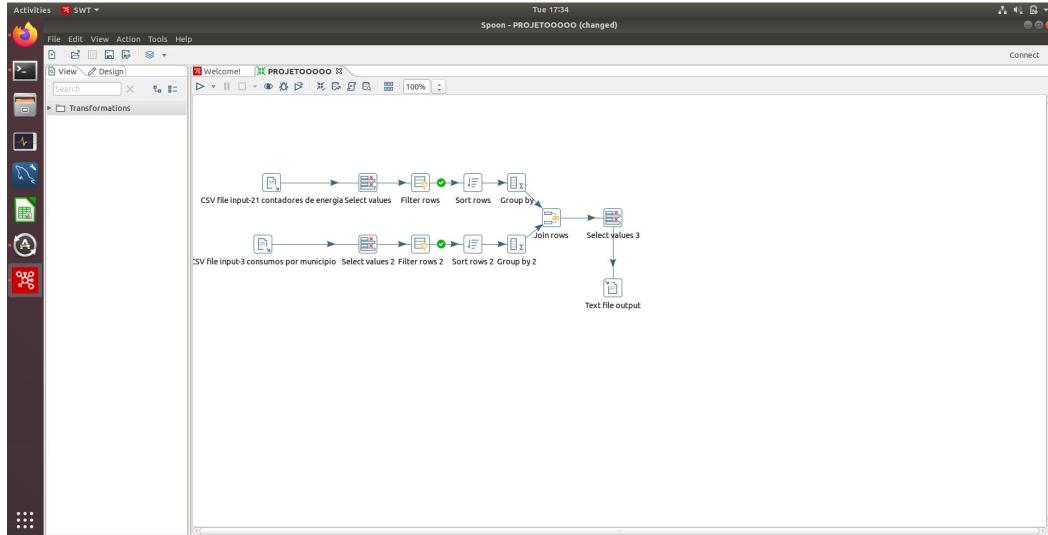
# Data Analysis and Integration Project 2023/24

Group 54

André Caseiro 99180

Miguel Mano 99286

# 1st Task



Tue 17:35 CSV File Input

Step name: CSV file input:21 contadores de energia

Filename: /home/aid/Downloads/21-contadores-de-energia.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 10000

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	16	0	\$	.	,	none
4	District	String	21	0	\$	,	,	none
5	Municipality	String	76	0	\$	,	,	none
6	Parish	String	76	0	\$	,	,	none
7	Incl. Smart Meter	String	3	0	\$	,	,	none
8	Number of CP's	Integer	#	15	0	\$	,	none
9	DistrictCode	Integer	#	15	0	\$	,	none
10	DistrictMunicipalityCode	Integer	#	15	0	\$	,	none
11	DistrictMunicipalityParishCode	String	6	0	\$	,	,	none

OK Get Fields Preview Cancel

Tue 17:36 Examine preview data

Rows of step: CSV file input:21 contadores de energia (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Incl. Smart Meter	Number of CP's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	
1	2023	4	2023-04-01	Coimbra	União das Freguesias de Assafarje e Antanhel	SIM	1688	6	603	06		
2	2023	5	2023-05-01	Viana do Castelo	União das Freguesias de Arcos de Valdevez	SIM	32	16	1601	16		
3	2023	5	2023-05-01	Vila Real	Mesão Frio	NÃO	157	17	1704	17		
4	2023	5	2023-05-01	Vila Real	Boticas	SIM	89	17	1703	17		
5	2023	8	2023-08-01	Braga	Celorico de Basto	NÃO	144	3	305	02		
6	2023	7	2023-07-01	Coimbra	Oliveira do Hospital	NÃO	178	6	611	06		
7	2023	7	2023-07-01	Braga	Gumaraes	SIM	1351	3	308	03		
8	2023	7	2023-07-01	Braga	Serzedelo	NÃO	361	3	309	03		
9	2023	6	2023-06-01	Porto	Braga	SIM	190	16	1607	16		
10	2022	12	2022-12-01	Setúbal	Ponte de Lima	NÃO	1346	14	406	14		
11	2023	2	2023-02-01	Coimbra	Carvalho	NÃO	426	4	407	04		
12	2022	2	2022-02-01	Porto	Porto	Centro da Boa Chama	NÃO	431	16	1603	16	
13	2023	2	2023-02-01	Castro Daire	Coroupa	NÃO	405	10	1001	10		
14	2022	7	2022-07-01	Leiria	Alcobaça	Malorga	SIM	974	1	169	01	
15	2022	7	2022-07-01	Aveiro	Santa Maria da Feira	Escapelas	NÃO	682	7	707	01	
16	2022	7	2022-07-01	Evora	Most	Cabeção	NÃO	169	3	914	06	
17	2022	7	2022-07-01	Guarda	Vila Nova de Foz Côa	Portela	SIM	778	6	600	06	
18	2023	4	2023-04-01	Coimbra	Côde	Alentes	NÃO	258	9	907	09	
19	2023	7	2023-07-01	Guarda	Guarda	Vila Fernando	NÃO	668	14	1411	14	
20	2022	6	2022-06-01	Sanxém	Ferreira do Zêzere	União das Freguesias de Areias e Pias	SIM	2137	18	1802	18	
21	2022	10	2022-10-01	Porto	Carregal do Sal	Carregal do Sal	SIM	996	9	996	09	
22	2022	10	2022-10-01	Guarda	Coimbra	União das Freguesias de Moimenta da Serra e Vinhã	SIM	270	9	996	09	
23	2022	10	2022-10-01	Porto	Beira	União das Freguesias de Campbell e Ovil	SIM	717	13	1302	13	
24	2022	5	2022-05-01	Viana do Castelo	Arcos de Valdevez	União das Freguesias de Vilela, São Cosme e São Damão e Sá	NÃO	383	16	1601	16	
25	2022	11	2022-11-01	Vila Real	Valpaços	Possacos	SIM	86	17	1712	17	
26	2022	7	2022-07-01	Porto	Amorim	Amorim	SIM	127	13	1303	13	
27	2022	7	2022-07-01	Braga	Gumaraes	Mesão Frio	SIM	1932	3	306	03	
28	2022	3	2022-03-01	Sanxém	Chamusca	União das Freguesias de Parreira e Chouto	NÃO	658	14	1407	14	
29	2023	3	2023-03-01	Vila Real	Chaves	Outeiro Seco	NÃO	137	17	1703	17	
30	2022	8	2022-08-01	Porto	Tebujo	Longa	NÃO	165	18	1819	18	
31	2022	10	2022-10-01	Braga	Bonfim	União das Freguesias de Aradas, Naselha e Pereira	SIM	97	4	996	04	
32	2022	8	2022-08-01	Viseu	Tebujo	União das Freguesias de Pinheiros e Vale de Figueira	NÃO	123	18	1819	18	
33	2022	8	2022-08-01	Braga	Barcelos	Barqueiros	SIM	360	3	302	03	
34	2022	8	2022-08-01	Porto	Gondomar	União das Freguesias de Foz do Sousa e Covelo	NÃO	2278	13	1304	13	
35	2022	8	2022-08-01	Ilhoca	Ilhoca	Camões da Foz da Ribeira	NÃO	1464	11	1106	11	

Close Show Log

Tue 17:37 CSV File Input

Step name: CSV file input:3 consumos por município

Filename: /home/aid/Downloads/3-consumos-faturados-por-municipio-ultimos-10-anos.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 10000

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	16	0	\$	,	,	none
4	District	String	16	0	\$	,	,	none
5	Municipality	String	22	0	\$	,	,	none
6	Parish	String	90	0	\$	,	,	none
7	Incl. Smart Meter	String	32	0	\$	,	,	none
8	Active Energy (kWh)	Number	#.	12	3	\$	,	none
9	DistrictCode	Integer	#	15	0	\$	,	none
10	DistrictMunicipalityCode	Integer	#	15	0	\$	,	none

OK Get Fields Preview Cancel

Tue 17:38 Examine preview data

Rows of step: CSV file input:3 consumos por município (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode
1	2023	2	2023-02-01	Porto	Vila do Conde	MINDELO	Baixa Tensão	2021002	11	1116
2	2022	7	2022-07-01	PORTO	AVIADO	ESPINHO	Baixa Tensão	757616.4	13	1316
3	2022	8	2022-08-01	AVIADO	CASTELO BRANCO	PAZINHOS	Baixa Tensão	274408.1	1	107
4	2022	8	2022-08-01	AVIADO	GUARDA	VALDESPINHO	Baixa Tensão	46110.4	5	507
5	2022	8	2022-08-01	GUARDA	Sabugal	VALDESPINHO	Baixa Tensão	29007.0	9	907
6	2022	8	2022-08-01	LEIRIA	LEIRIA	GAIBRAS	Baixa Tensão	509008.2	10	1012
7	2022	8	2022-08-01	LEIRIA	Monte da Beira	MBRA	Muito Alta, Alta e Média Tensões	142244.8	10	1016
8	2022	8	2022-08-01	LEIRIA	Porto de Mós	MIRA	Muito Alta, Alta e Média Tensões	3685384.5	11	1115
9	2022	8	2022-08-01	LEIRIA	Amadora	ENCOSTA DO SOL	Muito Alta, Alta e Média Tensões	16583.1	14	1400
10	2022	8	2022-08-01	LEIRIA	VILA NOVA DE CASTELO	UF MOLEDO E CRISTELO	Baixa Tensão	59125.9	16	1602
11	2022	8	2022-08-01	VISEU	TODELLA	UF VILA NOVA RAINHA	Baixa Tensão	256920.4	18	1821
12	2022	8	2022-08-01	VISEU	OLIVEIRINHA	OLIVEIRINHA	Baixa Tensão	615474.5	19	1905
13	2022	9	2022-09-01	AVIADO	BRAGA	BRAGA	Baixa Tensão	192500.0	1	1
14	2022	9	2022-09-01	AVIADO	Olivença	BAIRRO DA LARANJEIRA	Baixa Tensão	1191887.7	1	114
15	2022	9	2022-09-01	AVIADO	FAIFE	UF CEPIAS E FAIREJA	Baixa Tensão	266406.9	3	307
16	2022	9	2022-09-01	BRAGA	FAIFE	ALVOR	Baixa Tensão	266235.5	8	811
17	2022	9	2022-09-01	BRAGA	FERMOS	FERMOS	Baixa Tensão	29762.4	9	907
18	2022	9	2022-09-01	GUARDA	GUARDA	UF HAMACA E MARTELEIRA	Baixa Tensão	53831.4	11	1108
19	2022	9	2022-09-01	GUARDA	Santo Tirso	AVES	Baixa Tensão	103978.1	13	1314
20	2022	9	2022-09-01	SANTAREM	Mação	CARVOEIRO	Baixa Tensão	7193.9	14	1413
21	2022	9	2022-09-01	SANTAREM	Vila Franca de Xira	UF CARVALHEIRAS	Baixa Tensão	178751.3	18	1824
22	2022	10	2022-10-01	COMBRA	Arganil	UF SANTAREM CARVALHAL VERNALHAS	Baixa Tensão	231917.9	20	2001
23	2022	10	2022-10-01	COMBRA	Montemor-o-Velho	UF ABREUVERDE VERNDE VN MARCA	Baixa Tensão	286168.2	6	610
24	2022	10	2022-10-01	COMBRA	Fóios	FOIOS	Baixa Tensão	41248.6	9	911
25	2022	10	2022-10-01	GUARDA	Castañheira de Pêra	UF CASTAÑHEIRA PEIRA COENTRAL	Baixa Tensão	498313.0	10	1007
26	2022	10	2022-10-01	GUARDA	Porto de Márquez	UF CASTAÑHEIRA PEIRA COENTRAL	Baixa Tensão	1618893.0	13	1307
27	2022	10	2022-10-01	GUARDA	Teixeira	UF TEIXEIRA	Baixa Tensão	1345330.2	14	1421
28	2022	10	2022-10-01	VILA NOVA DO CASTELO	Arcos de Valdevez	UF ARCO VALDEVEZ 2 GONÇALO PARADA	Baixa Tensão	681243.4	16	1601
29	2022	10	2022-10-01	VILA NOVA DO CASTELO	VIAREDO	UF CABRIOS ALBERCABA SERRA	Baixa Tensão	33949.5	1	104
30	2022	11	2022-11-01	BRAGANA	Vinhais	UF SOBRIO BAIXO ALVARDEDO	Baixa Tensão	49353.8	4	412
31	2022	11	2022-11-01	BRAGANA	Lagarica	UF VILA NOVA DE CERVEIRA	Baixa Tensão	2020917.1	5	507
32	2022	11	2022-11-01	FARO	Tevira	UF LUZ TAVIRA SANTO ESTEVAO	Muito Alta, Alta e Média Tensões	643651.8	8	814
33	2022	11	2022-11-01	FARO	FORNO TELHEIRO	UF SANGUEDO	Muito Alta, Alta e Média Tensões	88103.4	9	903
34	2022	11	2022-11-01	GUARDA	Celorico de Beira	SANGUEDO	Muito Alta, Alta e Média Tensões	195979.1	1	109
35	2022	12	2022-12-01	AVIADO	Santa Maria da Feira	SANGUEDO	Muito Alta, Alta e Média Tensões			

Close Show Log

Step name: Select values

Fields:

Fieldname	Rename to	Length	Precision
1 Year	year_1		
2 Month	month_1		
3 Municipality	municipality_1		
4 Number of CPE's			
5 DistrictMunicipalityCode	DistrictMunicipalityCode1		

Get fields to select

Step name: Select values 2

Fields:

Fieldname	Rename to	Length	Precision
1 Year	year2		
2 Month	month2		
3 Municipality	municipality2		
4 Active Energy (kWh)			
5 DistrictMunicipalityCode	DistrictMunicipalityCode2		

Get fields to select

Include unspecified Fields, ordered by name

OK Cancel Help

Tue 17:44

Activities SWT

Examine preview data

Rows of step: Select values (1000 rows)

year_1	month_1	municipality_1	Number of CPE's
1	2023	4 Coimbra	1688
2	2023	5 Arcos de Valdevez	32
3	2023	14 Vila do Conde	157
4	2023	1 Braga	89
5	2023	8 Celorico de Basto	144
6	2023	7 Oliveira do Hospital	178
7	2023	7 Guimarães	1351
8	2023	7 Barcelos	361
9	2023	6 Ponte de Lima	190
10	2023	10 Viseu	13397
11	2023	2 Cartaxo	1346
12	2023	8 Mirandela	426
13	2023	6 Castro daire	431
14	2022	7 Alcochete	405
15	2022	7 São João da Feira	974
16	2022	9 Mora	600
17	2022	3 Vila Nova de Foz Côa	169
18	2023	4 Góis	778
19	2023	7 Guarda	258
20	2022	6 Ferreira do Zêzere	668
21	2022	10 Carregal do Sal	213
22	2022	10 Vouzela	270
23	2022	10 Baião	717
24	2022	5 Arcos de Valdevez	383
25	2022	11 Valpaços	86
26	2022	7 Amarante	127
27	2022	7 Guimarães	1932
28	2022	10 Vila Verde	656
29	2023	3 Chaves	137
30	2023	8 Tabuaçô	165
31	2022	12 Mirandela	97
32	2022	8 Tabuaçô	123
33	2022	8 Barcelos	360
34	2022	8 Gondomar	2278
35	2022	8 Lisboa	1984

Tue 17:46

Activities SWT

Examine preview data

Rows of step: Select values 2 (1000 rows)

year2	month2	municipality2	Active Energy (kWh)
1	2022	7 Vila do Conde	3871907.7
2	2022	8 Espinho	737916.4
3	2022	8 Penamacor	224463.1
4	2022	8 Guarda	46110.4
5	2022	8 Sabugal	29907.3
6	2022	8 Góis	65586.5
7	2022	8 Amarante	50863
8	2022	8 Vila Verde	142348.8
9	2022	8 Amadora	368534.5
10	2022	8 Caminha	65482.1
11	2022	8 Vila do Conde	59915.9
12	2022	8 Tondela	23004.6
13	2022	8 Viseu	81541.4
14	2022	9 Ilheus	1932566
15	2022	9 Oliveira do Bairro	1191897.7
16	2022	9 Fafe	266469.9
17	2022	9 Portimão	2662235.5
18	2022	9 Vila Real	297622.5
19	2022	9 Lourinhã	56454.3
20	2022	9 Santo Tirso	1039783.1
21	2022	9 Mação	71593
22	2022	9 Vouzela	178751.3
23	2022	10 Arganil	2319197
24	2022	10 Vila Franca de Xira	26943.2
25	2022	10 Sabugal	4126.8
26	2022	10 Castanheira de Pêra	498313
27	2022	10 Marco de Canaveses	161889.8
28	2022	10 Ourense	1345330.2
29	2022	10 Arcos de Valdevez	661243.4
30	2022	10 Vila Real	139593.5
31	2022	11 Viseu	43853.8
32	2022	11 Lagos	200971.1
33	2022	11 Tavira	643651.8
34	2022	11 Celorico da Beira	88103.4
35	2022	12 Santa Maria da Feira	195979

Close this dialog and leave the preview transformation in a paused state

Tue 17:47

Step name: **Filter rows**

Send 'true' data to step: **Sort rows**  
Send 'false' data to step:

The condition:

```
year_1 = [2023]
AND
month_1 = [6]
```

OK Cancel

Tue 17:49

Step name: **Filter rows**

Send 'true' data to step: **Sort rows 2**  
Send 'false' data to step:

The condition:

```
year2 = [2023]
AND
month2 = [6]
```

OK Cancel

Tue 17:48

Examine preview data

Rows of step: Filter rows (1000 rows)

	year_1	month_1	municipality_1	Number of CPEs	DistrictMunicipalityCode1
1	2023	6	Caminha	143	1803
2	2023	6	Almeida	192	1804
3	2023	6	Aveiro da Beira	192	1805
4	2023	6	São João da Pesqueira	111	1806
5	2023	6	Anadia	954	1807
6	2023	6	Porto	3244	1812
7	2023	6	Gondomar	535	1813
8	2023	6	Tremosso	89	1814
9	2023	6	Pacos de Ferreira	1009	1815
10	2023	6	Braga	188	1816
11	2023	6	Lousada	451	1817
12	2023	6	Monte da Lagoa	692	1818
13	2023	6	Torres Novas	1021	1819
14	2023	6	Vila Real de Santo António	5654	1820
15	2023	6	Penafiel	564	1821
16	2023	6	Arruda dos Vinhos	106	1822
17	2023	6	Mirandela da Beira	158	1823
18	2023	6	Oeiras	7261	1824
19	2023	6	Mangualde	415	1825
20	2023	6	Évora	4972	1826
21	2023	6	Leiria	152	1827
22	2023	6	Ourém	337	1828
23	2023	6	Aveida	147	1829
24	2023	6	Barcelos	261	1830
25	2023	6	Cinfães	78	1831
26	2023	6	Coimbra	245	1832
27	2023	6	Miranda do Corvo	751	1833
28	2023	6	Viana do Castelo	354	1834
29	2023	6	Tábua	955	1835
30	2023	6	Tomar	593	1836
31	2023	6	Arganil	179	1837
32	2023	6	Alenquer	11834	1838
33	2023	6	Sabrosa	161	1839
34	2023	6	Arouca	426	1840
35	2023	6	Póvoa de Lanhoso	909	1841
				309	1842
				1420	1843

**Close Stop Get more rows**

Tue 17:50

Examine preview data

Rows of step: Filter rows 2 (1000 rows)

	year2	month2	municipality2	Active Energy (kWh)	DistrictMunicipalityCode2
1	2023	6	Aldeia do Corvo	4253174	1803
2	2023	6	Portalegre	10039576.7	811
3	2023	6	Mirandela	1378504.1	1016
4	2023	6	Felgueiras	1361618.8	1303
5	2023	6	Arcos de Valdevez	31831.3	1601
6	2023	6	Vila Real	676374.4	1714
7	2023	6	Almeida	92030.0	202
8	2023	6	Odemira	671764.2	211
9	2023	6	Torres Vedras	79163.1	1113
10	2023	6	Fronseira	105237.8	1208
11	2023	6	Vila do Conde	39314.6	1316
12	2023	6	Guimarães	148382.8	1409
13	2023	6	Viseu	126909.2	1823
14	2023	6	Ferreira do Alentejo	2152018.5	208
15	2023	6	Braga	767437.8	303
16	2023	6	Póvoa de Lanhoso	935393.9	309
17	2023	6	Carrazeda	2506517.7	503
18	2023	6	Arganil	201297	601
19	2023	6	São Pedro do Sul	84389.1	1816
20	2023	6	Vouzela	48901.4	1824
21	2023	6	Vagos	320915.5	118
22	2023	6	Mondim de Basto	388605.5	407
23	2023	6	Alandroal	623105.5	701
24	2023	6	Aleijouer	128920.6	1101
25	2023	6	Maia	991122.9	1109
26	2023	6	Almeida	438524.1	1402
27	2023	6	Tomar	556205.5	1418
28	2023	6	Castro daire	127939.3	1803
29	2023	6	Barcelos	208995.5	302
30	2023	6	Guimarães	173463	308
31	2023	6	Penafiel	93943.2	504
32	2023	6	Póvoa	554517.7	910
33	2023	6	Torres Vedras	4761280.6	1113
34	2023	6	Lousada	560714.2	1305
35	2023	6	Vila Real	3656671.1	1714

**Close Stop Get more rows**

Step name **Sort rows**

Sort directory **%{java.io.tmpdir%}**

TMP-file prefix **out**

Sort size (rows in memory) **1000000**

Free memory threshold (in %) **1000000**

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	DistrictMunicipalityCode1	Y	N	N	0	N
2	municipality_1	Y	N	N	0	N

Activities → SWT

Step name **Sort rows**

Sort directory **%{java.io.tmpdir%}**

TMP-file prefix **out**

Sort size (rows in memory) **1000000**

Free memory threshold (in %) **1000000**

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	DistrictMunicipalityCode2	Y	N	N	0	N
2	municipality2	Y	N	N	0	N

activities → SWT

Sat 17/51

Examine preview data

Rows of step: Sort rows (1000 rows)

year_1	month_1	municipality_1	Number of CPE's
1	2023	6 Águeda	147
2	2023	6 Águeda	156
3	2023	6 Águeda	300
4	2023	6 Águeda	1608
5	2023	6 Águeda	7673
6	2023	6 Águeda	2016
7	2023	6 Águeda	323
8	2023	6 Águeda	705
9	2023	6 Águeda	1698
10	2023	6 Águeda	2150
11	2023	6 Águeda	169
12	2023	6 Águeda	168
13	2023	6 Águeda	354
14	2023	6 Águeda	56
15	2023	6 Águeda	1395
16	2023	6 Águeda	1073
17	2023	6 Águeda	1946
18	2023	6 Águeda	297
19	2023	6 Águeda	764
20	2023	6 Águeda	325
21	2023	6 Águeda	143
22	2023	6 Águeda	476
23	2023	6 Albergaria-a-Velha	1041
24	2023	6 Albergaria-a-Velha	1081
25	2023	6 Albergaria-a-Velha	650
26	2023	6 Albergaria-a-Velha	469
27	2023	6 Albergaria-a-Velha	371
28	2023	6 Albergaria-a-Velha	475
29	2023	6 Albergaria-a-Velha	679
30	2023	6 Albergaria-a-Velha	579
31	2023	6 Albergaria-a-Velha	229
32	2023	6 Albergaria-a-Velha	1015
33	2023	6 Albergaria-a-Velha	5829
34	2023	6 Albergaria-a-Velha	1824
35	2023	6 Anadia	954

Close this dialog and leave the preview transformation in a paused state.  
You can resume this transformation at any time.

activities → SWT

Sat 17/52

Examine preview data

Rows of step: Sort rows 2 (1000 rows)

year1	month1	municipality2	Active Energy (kWh)
1	2023	6 Águeda	49163.6
101	2023	6 Águeda	7042.1
101	2023	6 Águeda	58764.2
101	2023	6 Águeda	20555.5
5	2023	6 Águeda	120398.5
6	2023	6 Águeda	401248.8
7	2023	6 Águeda	200985.3
101	2023	6 Águeda	748027.7
101	2023	6 Águeda	623378.8
9	2023	6 Águeda	404615.9
101	2023	6 Águeda	1206462.7
101	2023	6 Águeda	2000000.0
13	2023	6 Águeda	323120.7
13	2023	6 Águeda	753311.3
15	2023	6 Águeda	111918.9
101	2023	6 Águeda	1000000.0
101	2023	6 Águeda	198258.9
101	2023	6 Águeda	1131063.2
19	2023	6 Águeda	649936.4
101	2023	6 Águeda	1822948.9
21	2023	6 Águeda	1000000.0
22	2023	6 Águeda	767498.1
23	2023	6 Albergaria-a-Velha	350953.7
34	2023	6 Albergaria-a-Velha	799771.0
101	2023	6 Albergaria-a-Velha	1000000.0
26	2023	6 Albergaria-a-Velha	25993.3
27	2023	6 Albergaria-a-Velha	262632
28	2023	6 Albergaria-a-Velha	862716.4
29	2023	6 Albergaria-a-Velha	2000000.0
30	2023	6 Albergaria-a-Velha	1857419.5
31	2023	6 Albergaria-a-Velha	10170977.3
32	2023	6 Anadia	552792.6
33	2023	6 Anadia	1000000.0
34	2023	6 Anadia	558566.2
35	2023	6 Anadia	362009.1

**Activities** **Tue 17/53** **Group by**

Step name: Group by  
 Include all rows?  
Temporary files directory: %java.io.tmpdir%/  
Temporary 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\_1  
2 DistrictMunicipalityCode1

Aggregates:  

Name	Subject	Type	Value
CPE's by locality	Number of CPE's	Sum	

**OK** **Cancel**

**Activities** **Tue 17/55** **Group by**

Step name: Group by  
 Include all rows?  
Temporary files directory: %java.io.tmpdir%/  
Temporary 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 municipality2  
2 DistrictMunicipalityCode2

Aggregates:  

Name	Subject	Type	Value
Active Energy by municipality(kWh)	Active Energy (kWh)	Sum	

**OK** **Cancel**

**Examine preview data**

Rows of step: Group by (278 rows)

municipality_1	DistrictMunicipalityCode1	CPE's by locality
1	Avega	101
2	Albergaria-a-Velha	102
3	Anadia	103
4	Arouca	104
5	Barcelos	105
6	Castelo de Paiva	106
7	Espinho	107
8	Estarreja	108
9	Santa Maria da Feira	109
10	Ilhavo	110
11	Marcosende	111
12	Mirrosa	112
13	Olivença Azeémis	113
14	Olivença do Bairro	114
15	Ovar	115
16	São João da Madeira	116
17	Sever do Vouga	117
18	Vagos	118
19	Vale de Cambra	119
20	Aljustrel	201
21	Almodôvar	202
22	Alva	203
23	Barrancos	204
24	Beja	205
25	Castro Verde	206
26	Cuba	207
27	Fátima do Alentejo	208
28	Mértola	209
29	Moura	210
30	Odemira	211
31	Ourique	212
32	Peniche	213
33	Vidigueira	214
34	Amarante	301
35	Barcelos	302

**Close**

**Examine preview data**

Rows of step: Group by 2 (278 rows)

municipality2	DistrictMunicipalityCode2	Active Energy by municipality(kWh)
1	Avega	101
2	Albergaria-a-Velha	102
3	Anadia	103
4	Arouca	104
5	Ilhavo	105
6	Castelo de Paiva	106
7	Espinho	107
8	Estarreja	108
9	Santa Maria da Feira	109
10	Ilhavo	110
11	Marcosende	111
12	Mirrosa	112
13	Olivença de Azeméis	113
14	Olivença do Bairro	114
15	Ovar	115
16	São João da Madeira	116
17	Sever do Vouga	117
18	Vagos	118
19	Vale de Cambra	201
20	Aljustrel	202
21	Almodôvar	203
22	Alva	204
23	Barrancos	205
24	Beja	206
25	Castro Verde	207
26	Cuba	208
27	Fátima do Alentejo	209
28	Mértola	210
29	Moura	211
30	Odemira	212
31	Ourique	213
32	Serpa	214
33	Vidigueira	301
34	Amarante	302

**Close**

Tue 17:57

Activities > SWT >

**Join rows (cartesian product)**

Step name: **municipality\_1**

Temp directory: **%{java.io.tmpdir%}**

Temp file prefix: **out**

Max. cache size (in rows): **500**

Main step to read from:

The condition:

```
municipality_1 = municipality2
```

AND

```
DistrictMunicipalityCode1 = DistrictMunicipalityCode2
```

OK Cancel Help

Tue 17:58

Activities > SWT >

**Select values**

Step name: **municipality\_1**

Select & Alter > Remove Meta-data

Fieldname	Rename to	Length	Precision
1 municipality_1	Municipality		
2 CPE's by locality			
3 Active Energy by municipality(0/w)			

Get fields to select:

Include unspecified fields, ordered by name

OK Cancel Help

Tue 17:58

Activities > SWT >

**Rows of step: Join rows (278 rows)**

Examine preview data

municipality_1	DistrictMunicipalityCode1	CPE's by locality	municipality2	DistrictMunicipalityCode2	Active Energy by municipality(0/w)
1 Aguda	101	24802	Aguarda	101	23359302.557
2 Albergaria-a-Velha	102	14241	Albergaria-a-Velha	102	14765120.434
3 Anadia	103	17295	Anadia	103	9902342.348
4 Aveiro	104	12679	Aveiro	104	40646423.36299994
5 Averio	105	54462	Averio	105	5997228.36999994
6 Castelo de Paiva	106	8082	Castelo de Paiva	106	3493372.45400000
7 Espinho	107	18639	Espinho	107	6566676.72
8 Estarreja	108	14469	Estarreja	108	52194107.14500005
9 Santa Maria da Feira	109	72390	Santa Maria da Feira	109	56119912.21399999
10 Vagos	110	25100	Vagos	110	9111822.352
11 Meiahdas	111	1546	Meiahdas	111	7650201.06599999
12 Murtosa	112	8570	Murtosa	112	2650942.073999996
13 Oliveira de Azeméis	113	3331	Oliveira de Azeméis	113	30844529.74000005
14 Oliveira do Bairro	114	13740	Oliveira do Bairro	114	1000000.00000000
15 Our	115	3999	Our	115	25990054.74400000
16 São João da Madeira	116	13526	São João da Madeira	116	892818.162
17 Sever do Vouga	117	7693	Sever do Vouga	117	259915.376
18 Vagos	118	15568	Vagos	118	7018458.86000001
19 Vale de Cambra	119	13471	Vale de Cambra	119	7862887.632
20 Vila do Conde	201	15041	Vila do Conde	201	15212818.00000001
21 Almodôvar	202	5601	Almodôvar	202	1455934.812000002
22 Alvião	203	1721	Alvião	203	120644.494
23 Barrancos	204	12939	Barrancos	204	401405.503
24 Beja	205	269	Beja	205	4550849.42300001
25 Castro Verde	206	4993	Castro Verde	206	144474.0000000000
26 Cuba	207	3160	Cuba	207	1776994.995
27 Ferreira do Alentejo	208	5357	Ferreira do Alentejo	208	585930.763
28 Mertola	209	8018	Mertola	209	1319320.814
29 Moura	210	10421	Moura	210	4675012.562
30 Odemira	211	20930	Odemira	211	742954.789000005
31 Ourique	212	4170	Ourique	212	1477600.494
32 Serpa	213	10734	Serpa	213	9775084.33399998
33 Vidigueira	214	4049	Vidigueira	214	5859727.24000001
34 Amares	301	11329	Amares	301	3512297.001
35 Barcelos	302	57289	Barcelos	302	33547153.26599998

Close Close

Tue 17:59

Activities > SWT >

**Rows of step: Select values 3 (278 rows)**

Examine preview data

Municipality	CPE's by locality
1 Aguda	24802
2 Albergaria-a-Velha	14241
3 Anadia	17295
4 Aveiro	12679
5 Averio	51443
6 Castelo de Paiva	8082
7 Espinho	18639
8 Estarreja	14469
9 Santa Maria da Feira	72390
10 Vagos	25100
11 Meiahdas	1546
12 Murtosa	8570
13 Oliveira de Azeméis	33311
14 Oliveira do Bairro	13740
15 Our	3999
16 São João da Madeira	13526
17 Sever do Vouga	7693
18 Vagos	15568
19 Vale de Cambra	13471
20 Vila do Conde	15041
21 Almodôvar	5601
22 Alvião	1721
23 Barrancos	12939
24 Beja	269
25 Castro Verde	4993
26 Cuba	3160
27 Ferreira do Alentejo	5357
28 Mertola	8018
29 Moura	10421
30 Odemira	20930
31 Ourique	4170
32 Serpa	10734
33 Vidigueira	4049
34 Amares	11329
35 Barcelos	57299

Active Energy by municipality(0/w)

23359302.557
14765120.434
9902342.348
40646423.36299994
5997228.36999994
3493372.45400000
6566676.72
52194107.14500005
56119912.21399999
9111822.352
7650201.06599999
2650942.073999996
30844529.74000005
1000000.00000000
25990054.74400000
892818.162
259915.376
7018458.86000001
7862887.632
15212818.00000001
1455934.812000002
120644.494
401405.503
4550849.42300001
144474.0000000000
1776994.995
585930.763
1319320.814
4675012.562
742954.789000005
1477600.494
9775084.33399998
5859727.24000001
3512297.001

Close Close

The screenshot shows the Spoon IDE interface with the 'Text file output' step configuration dialog open. The 'Step name' is set to 'TESTE10002'. The 'Filename' field contains the expression 'External.Entry.Current.Directory() / ponto1PROJETO'. The 'Extension' field is set to 'csv'. There are several optional checkboxes: 'Pass output to parent' (unchecked), 'Create Parent folder' (unchecked), 'Do not create file at start' (unchecked), 'Accept file name from field' (unchecked), 'Include step' (unchecked), 'Include step in filename?' (unchecked), 'Include partition in filename?' (unchecked), 'Include date in filename?' (unchecked), 'Include time in filename?' (unchecked), and 'Specify Date time format' (unchecked). Below these options is a 'Select time format' dropdown menu. At the bottom of the dialog are 'Show filename(s)...' and 'Add filenames to result' buttons, along with 'OK' and 'Cancel' buttons.

	Municipality	CPE's by locality	Active Energy by municipality (kWh)
1	Agueda	24862	3339300.157
2	Aveiro e Vila do Conde	14431	14185150.484
3	Aveiro	17295	990236.348
4	Arouca	12670	4056020.329999994
5	Avanca	51443	5997297.329999994
6	Bananeira de Paiva	9882	3449320.640000001
7	Espinho	18639	656676.72
8	Esteirinha	14469	52941071.745000005
9	Santa Maria da Feira	72389	10119493.542
10	Albergaria-a-Velha	13593	1611548.562
11	Meiahelda	11546	7650201.08599999
12	Murtosa	8570	2650940.073999996
13	Oliveira de Azeméis	33311	3084425.174000001
14	Penacova e Bairro	13269	1629883.478
15	Ovar	32390	2390054.744000003
16	São João da Madeira	13526	892871.162
17	Sever do Vouga	7693	239611.576
18	Tondela	15546	761886.0000000001
19	Vale de Cambra	13471	782897.632
20	Ajuda	6058	15121881.473000001
21	Almodôvar	5693	1453993.000000001
22	Almeida	12371	122044.464
23	Barrancos	1239	401405.503
24	Beja	23054	45508349.42300001
25	Castelo Verde	4998	144471.039999998
26	Cuba	23409	17760.00000000001
27	Ferreira do Alentejo	5357	5859593.763
28	Mertola	8018	1319320.814
29	Moura	10421	4675015.562
30	Óbidos	20291	7429933.000000001
31	Ourique	4170	147760.935
32	Serpa	10734	9773584.93999999
33	Vidigueira	4049	585972.324000001
34	Anões	11349	9512791.061
35	Barreiros	57289	3354713.529999994

## 2nd Task

The screenshot shows the DataCleaner software interface. The top menu bar includes 'Activities' (with a flame icon), 'File' (New, Open, Save, Save As..., More), 'Edit' (Copy, Paste, Cut, Delete, Undo, Redo), 'Tools' (Import, Export, Preferences, Help), and 'Execute' (Run, Stop, Refresh, Execute). The title bar displays 'ponto1PROJETO.analysis.xml | ponto1PROJETO.csv | Analysis Job | DataCleaner'. The status bar at the bottom shows 'Tue 18:09' and 'Job is correctly configured'.

The main workspace shows a data flow diagram:

- A file node labeled "ponto1PROJETO.csv" is connected to a "Convert to number" node.
- The output of the "Convert to number" node is connected to a "Scatter plot" node.

A callout bubble points to the "Execute" button in the top right corner with the text "Click here to run job".

**Ready to execute**

Click the 'Execute' button in the upper-right corner when you're ready to run the job.

Search component library... X

Community edition - Up to date

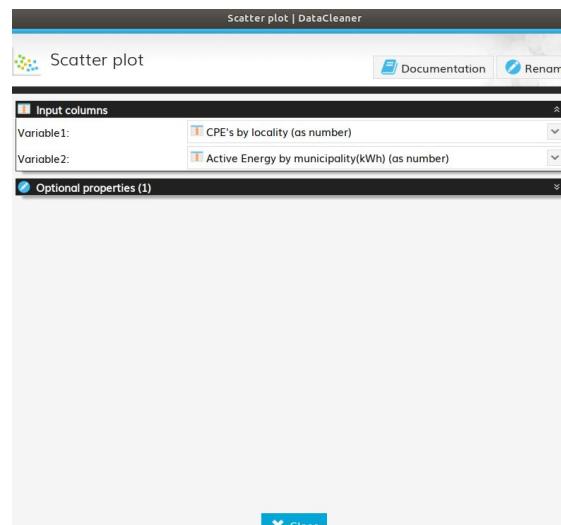
ponto1PROJETO.csv | DataCleaner

ponto1PROJETO.csv

Downloads.ponto1PROJETO.csv

Name	Type
Municipality	String
CPE's by locality	String
Active Energy by municipality(kWh)	String

X Close



Convert to number | DataCleaner

# Convert to number

Documentation Rename

Input columns

Input: Select all Select none  CPE's by locality  Active Energy by municipality(kWh)  Municipality

Required properties

Decimal separator: . -

Minus sign: -

Thousand separator: , ,

Optional properties (1)

Output columns

Name	Type
CPE's by locality (as number)	Number
Active Energy by municipality(kWh) (as number)	Number

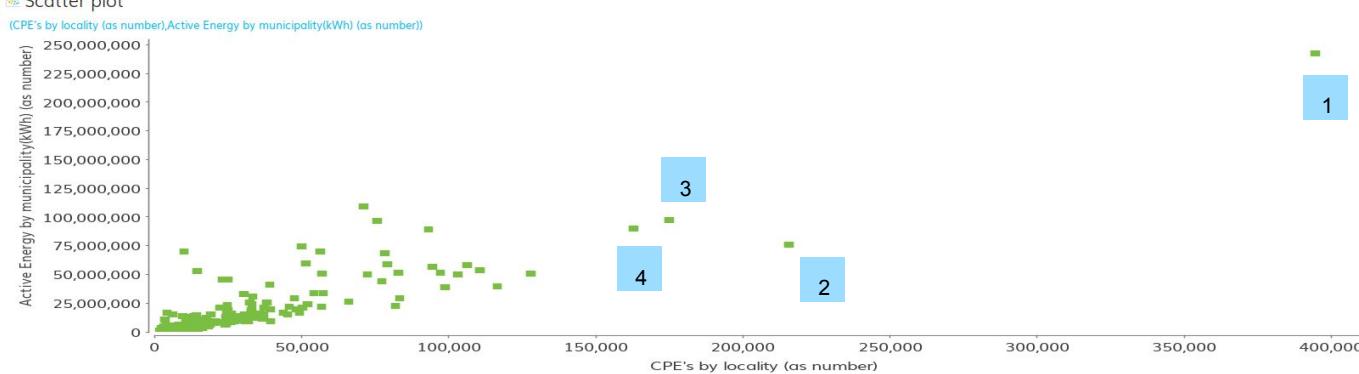
Write data Preview data | ▾

X Close

Analysis results | ponto1PROJETO.csv | ponto2Projeto.analysis.xml  
Scatter plot

Progress information

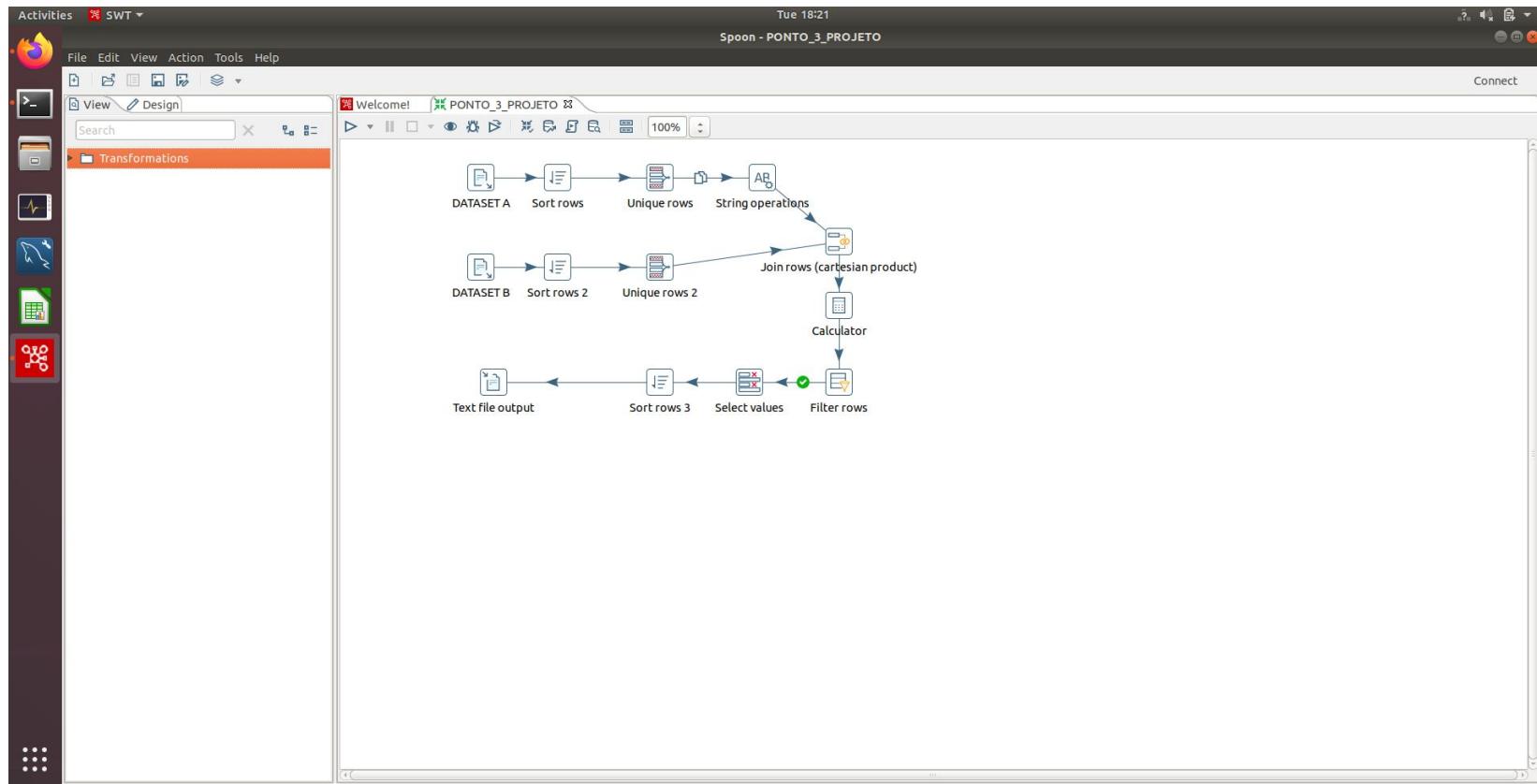
Scatter plot



Save results

1. Lisboa
2. Sintra
3. Porto
4. Vila Nova de Gaia

# 3nd Task



Tue 18:22 CSV file Input

Step name: DATASET A

Filename: /home/aid/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 type
District	String		16		\$	.	,	none

OK Get Fields Preview Cancel Help

Tue 18:23 CSV file Input

Step name: DATASET B

Filename: /home/aid/Downloads/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
District	String		16		\$	.	,	none

OK Get Fields Preview Cancel Help

Tue 18:22 Examine preview data

Rows of step: DATASET A (1000 rows)

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

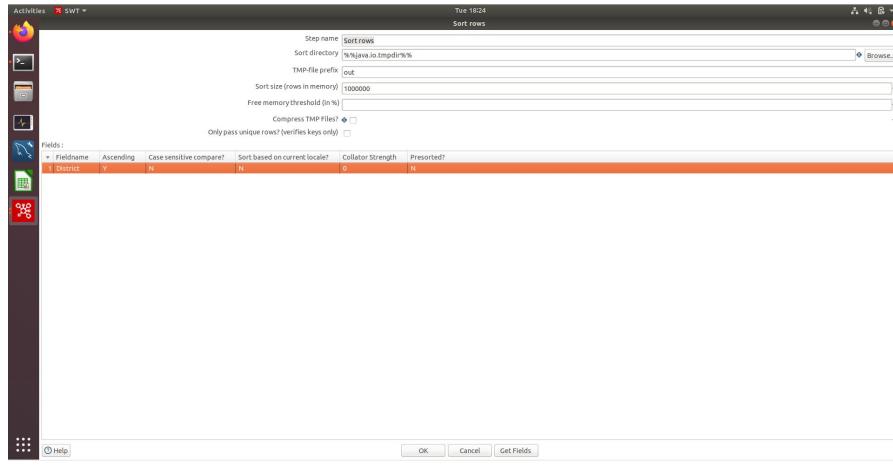
Close Show Log

Tue 18:23 Examine preview data

Rows of step: DATASET B (1000 rows)

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

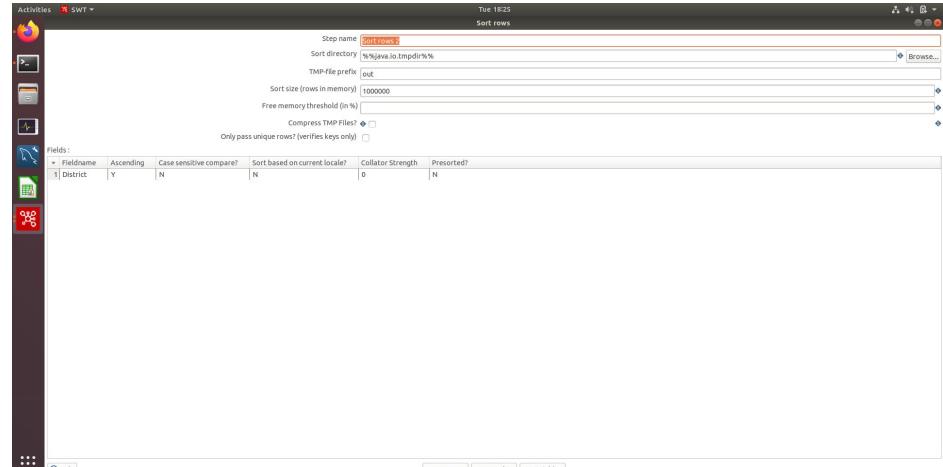
Close Show Log



Rows of step: Sort rows (1000 rows)

Row	District
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
35	Aveiro

Close Stop Get more rows



Rows of step: Sort rows 2 (1000 rows)

Examine preview data

Row	District
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
35	AVEIRO

Close Stop Get more rows

Activities  SWT

Step name: Unique rows

Settings

Add counter to output?  Counter field:

Redirect duplicate row  Error description:

Fields to compare on (no entries means: compare complete row)

+ Fieldname: Ignore case

1 District N

OK Cancel Get

Activities  SWT

Step name: Unique rows 2

Settings

Add counter to output?  Counter field:

Redirect duplicate row  Error description:

Fields to compare on (no entries means: compare complete row)

+ Fieldname: Ignore case

1 District N

2

OK Cancel Get

Activities  SWT

Tue 18:27

Examine preview data

Rows of step: Unique rows (18 rows)

District

1	Aveiro
2	Beira
3	Braga
4	Bragança
5	Castelo Branco
6	Coimbra
7	Faro
8	Guarda
9	Leiria
10	Lisboa
11	Portalegre
12	Porto
13	Santarem
14	Setubal
15	Viana do Castelo
16	Vila Real
17	Viseu
18	Evora

Close

Activities  SWT

Tue 18:28

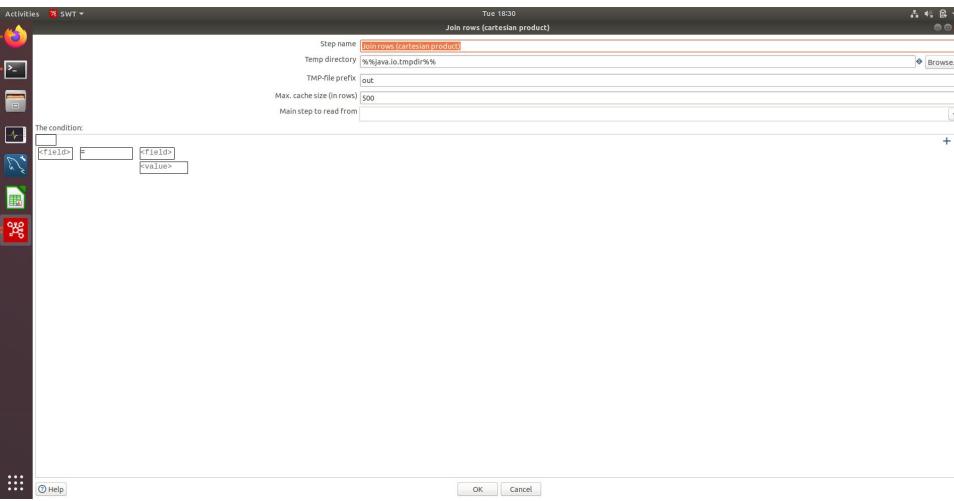
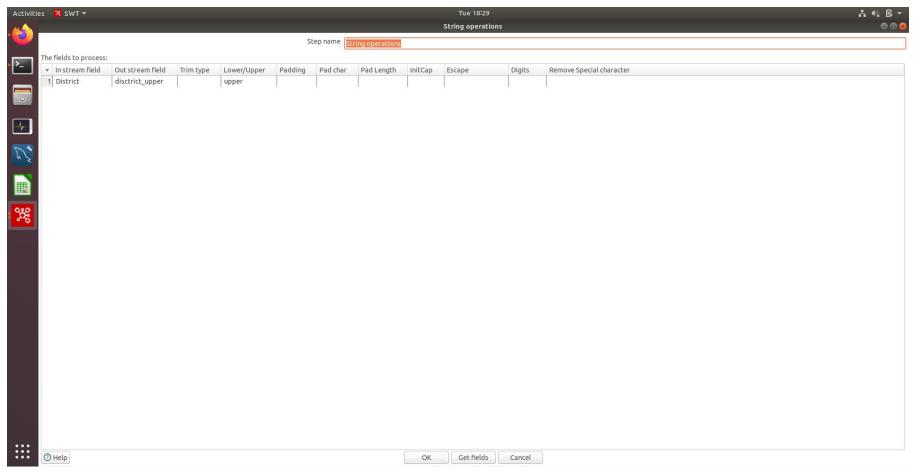
Examine preview data

Rows of step: Unique rows 2 (18 rows)

District

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

Close



Activities SWT

Tue 18:30

Examine preview data

Rows of step: String operations (18 rows)

	District	district_upper
1	Aveiro	AVEIRO
2	Beja	BEJA
3	Braga	BRAGA
4	Braganca	BRAGANCA
5	Castelo Branco	CASTELO BRANCO
6	Coimbra	COIMBRA
7	Faro	FARO
8	Guarda	GUARDA
9	Leiria	LEIRIA
10	Lisboa	LISBOA
11	Portalegre	PORCALEGRE
12	Porto	PORTO
13	Portugal	PORTUGAL
14	Setubal	SETUBAL
15	Viana do Castelo	VIANA DO CASTELO
16	Vila Real	VILA REAL
17	Viseu	VISEU
18	Evora	EVORA

Close

Activities SWT

Tue 18:31

Examine preview data

Rows of step: Join rows (cartesian product) (324 rows)

	District	District_1	district_upper
1	AVEIRO	Aveiro	AVEIRO
2	AVEIRO	Beja	BEJA
3	AVEIRO	Braga	BRAGA
4	AVEIRO	Braganca	BRAGANCA
5	AVEIRO	Castelo Branco	CASTELO BRANCO
6	AVEIRO	Coimbra	COIMBRA
7	AVEIRO	Faro	FARO
8	AVEIRO	Guarda	GUARDA
9	AVEIRO	Leiria	LEIRIA
10	AVEIRO	Lisboa	LISBOA
11	AVEIRO	Portalegre	PORCALEGRE
12	AVEIRO	Porto	PORTO
13	AVEIRO	Santarém	SANTARÉM
14	AVEIRO	Setubal	SETUBAL
15	AVEIRO	Viana do Castelo	VIANA DO CASTELO
16	AVEIRO	Vila Real	VILA REAL
17	AVEIRO	Viseu	VISEU
18	AEVRO	Evora	EVORA
19	BEJA	Aveiro	AVEIRO
20	BEJA	Beja	BEJA
21	BEJA	Braga	BRAGA
22	BEJA	Braganca	BRAGANCA
23	BEJA	Castelo Branco	CASTELO BRANCO
24	BEJA	Coimbra	COIMBRA
25	BEJA	Faro	FARO
26	BEJA	Guarda	GUARDA
27	BEJA	Leiria	LEIRIA
28	BEJA	Lisboa	LISBOA
29	BEJA	Portalegre	PORCALEGRE
30	BEJA	Porto	PORTO
31	BEJA	Santarém	SANTARÉM
32	BEJA	Setubal	SETUBAL
33	BEJA	Viana do Castelo	VIANA DO CASTELO
34	BEJA	Vila Real	VILA REAL
35	BEJA	Viseu	VISEU

Close

Activities SWT

Tue 18:32  
Calculator

Step name:

Fields:

New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol	Currency symbol
1 LEVDIS	Levenshtein Distance (source A and target B)	District	district_upper		Number		N					

Throw an error on non-existing files

OK Cancel Help

Activities SWT

Tue 18:33  
Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

LEVDIS	=	1.0	(Number)
--------	---	-----	----------

OK Cancel Help

Activities SWT

Tue 18:32  
Examine preview data

Rows of step: Calculator (324 rows)

▼ District District\_1 district\_upper

	District	District_1	district_upper	LEVDIS
1	AVERO	Aveiro	AVERO	0.0
2	AVERO	Beja	BEJA	5.0
3	AVERO	Braga	BRAGA	6.0
4	AVERO	Bragança	BRAGANÇA	7.0
5	AVERO	Castelo Branco	CASTELO BRANCO	10.0
6	AVERO	Coimbra	COIMBRA	5.0
7	AVERO	Faro	FARO	4.0
8	AVERO	Guarda	GUARDA	6.0
9	AVERO	Leiria	LEIRA	4.0
10	AVERO	Lisboa	LISBOA	6.0
11	AVERO	Portalegre	PORTALEGRE	7.0
12	AVERO	Porto	PORTO	5.0
13	AVERO	Santarém	SANTAREM	6.0
14	AVERO	Setúbal	SETUBAL	6.0
15	AVERO	Viana do Castelo	VIANA DO CASTELO	10.0
16	AVERO	Vila Real	VILA REAL	4.0
17	AVERO	Viseu	VISEU	8.0
18	AVERO	Évora	ÉVORA	5.0
19	BEJA	Aveiro	AVERO	4.0
20	BEJA	Beja	BEJA	0.0
21	BEJA	Braga	BRAGA	3.0
22	BEJA	Bragança	BRAGANÇA	6.0
23	BEJA	Castelo Branco	CASTELO BRANCO	12.0
24	BEJA	Coimbra	COIMBRA	6.0
25	BEJA	Faro	FARO	4.0
26	BEJA	Guarda	GUARDA	5.0
27	BEJA	Leiria	LEIRA	4.0
28	BEJA	Lisboa	LISBOA	5.0
29	BEJA	Portalegre	PORTALEGRE	9.0
30	BEJA	Porto	PORTO	5.0
31	BEJA	Santarém	SANTAREM	7.0
32	BEJA	Setúbal	SETUBAL	5.0
33	BEJA	Viana do Castelo	VIANA DO CASTELO	15.0
34	BEJA	Vila Real	VILA REAL	8.0
35	BEJA	Viseu	VISEU	5.0

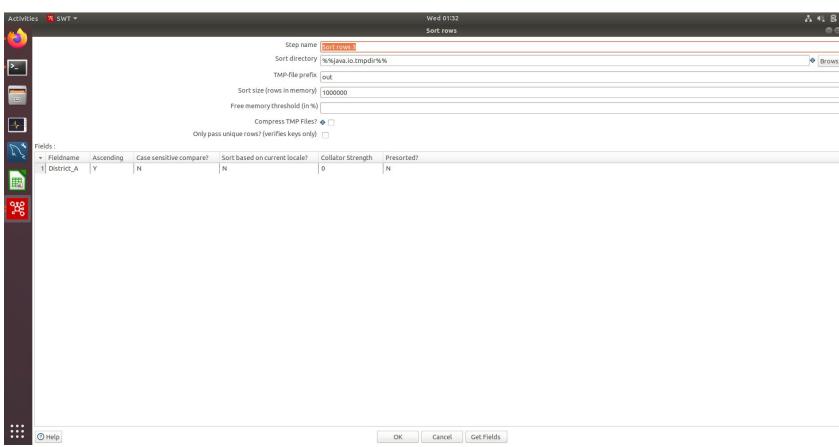
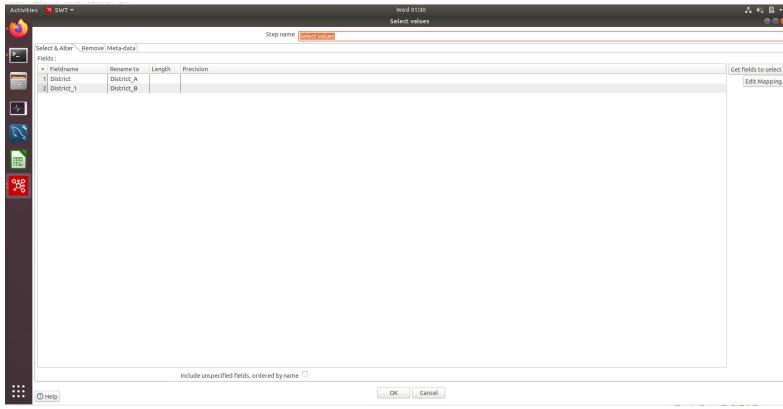
Activities SWT

Tue 18:33  
Examine preview data

Rows of step: Filter rows (18 rows)

▼ District District\_1 district\_upper

	District	District_1	district_upper	LEVDIS
1	AVEIRO	Aveiro	AVERO	0.0
2	BEJA	Beja	BEJA	0.0
3	BRAGA	Braga	BRAGA	0.0
4	BRAGANCA	Braganca	BRAGANÇA	0.0
5	CASTELO BRANCO	Castelo Branco	CASTELO BRANCO	1.0
6	COIMBRA	Coimbra	COIMBRA	0.0
7	EVORA	Évora	ÉVORA	0.0
8	FARO	Faro	FARO	0.0
9	GUARDA	Guarda	GUARDA	0.0
10	LEIRA	Leiria	LEIRA	0.0
11	LISBOA	Lisboa	LISBOA	0.0
12	PORTALEGRE	Portalegre	PORTALEGRE	0.0
13	PORTO	porto	PORTO	0.0
14	SANTAREM	Santarém	SANTAREM	1.0
15	SETUBAL	Setúbal	SETUBAL	0.0
16	VIANA DO CASTELO	Viana do Castelo	VIANA DO CASTELO	0.0
17	VILA REAL	Vila Real	VILA REAL	0.0
18	VISEU	Viseu	VISEU	0.0



Activities SWT

Rows of step: Select values (18 rows)

Examine preview data

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

Close

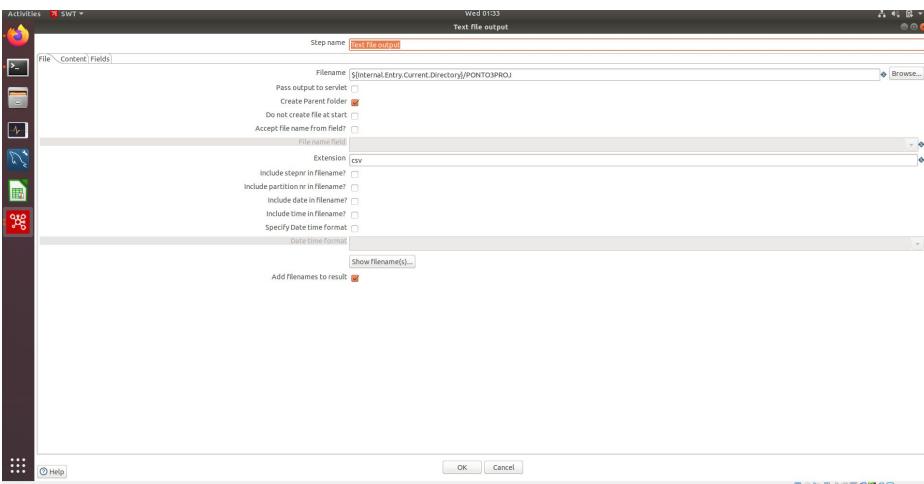
Activities SWT

Rows of step: Sort\_rows (18 rows)

Examine preview data

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

Close



Activities SWT

Rows of step: Text file output (18 rows)

Examine preview data

	District_A	District_B
1	AVIZIO	Aveiro
2	BEJA	Beja
3	BRACA	Braga
4	CASCAIS	Braganca
5	CASTELO BRANCO	Castelo Branco
6	COMBRA	Coimbra
7	EVORA	Evora
8	FARO	Faro
9	LEIRIA	Guarda
10	LENHA	Lamego
11	LISBOA	Lisboa
12	PORTALEGRE	Portalegre
13	PORTO	Porto
14	PORTUGAL	Santarem
15	SEBILHES	Sebeilh
16	VIANA DO CASTELO	Viana do Castelo
17	VILA REAL	Vila Real
18	VISEU	Viseu

Close

wed 0134

Activities LibreOffice Calc

PONTOSPROJ.csv - LibreOffice Calc

	District_A	District_B
1	Distrito_A	Distrito_B
2	AVEIRO	Aveiro
3	BEJA	Beja
4	BRAGA	Braga
5	BRAGANCA	Braganca
6	CASCAIS	Braganca
7	CASTELO BRANCO	Castelo Branco
8	COMBRA	Centro
9	EVORA	Centro
10	FARO	Faro
11	GUARDA	Guarda
12	LEIRIA	Lamego
13	LISBOA	Lisboa
14	PORTALEGRE	Portalegre
15	PORTO	Porto
16	PORTUGAL	Santarem
17	SEBILHES	Sebeilh
18	VIANA DO CASTELO	Viana do Castelo
19	VILA REAL	Vila Real
20	VISEU	Viseu
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		

# 4. SQL script to create DW tables

```
DROP DATABASE IF EXISTS dataset_b_dw;
CREATE DATABASE dataset_b_dw;
USE dataset_b_dw;

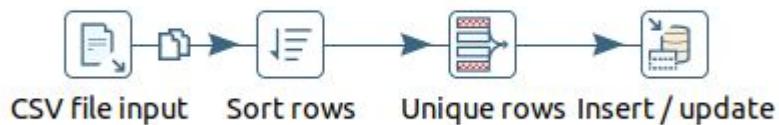
CREATE TABLE dim_time (
    TIME_ID DATETIME,
    YEAR_ID INT,
    MONTH_ID INT,
    PRIMARY KEY (TIME_ID)
);

CREATE TABLE dim_location (
    LOCATION_ID INT,
    PARISH VARCHAR(255),
    MUNICIPALITY VARCHAR(255),
    DISTRICT VARCHAR(255),
    PRIMARY KEY (LOCATION_ID)
);

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

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

## 5. Time dimension



	Year	Month	Date
1	2020	11	2020-11
2	2020	12	2020-12
3	2021	1	2021-01
4	2021	2	2021-02
5	2021	3	2021-03
6	2021	4	2021-04
7	2021	5	2021-05
8	2021	6	2021-06
9	2021	7	2021-07
10	2021	8	2021-08
11	2021	9	2021-09
12	2021	10	2021-10

# 5. Time dimension

CSV file input

Step name: CSV file input

Filename: /home/aid/Documents/proj/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:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Year	Integer	#	15	0	\$	.	,	none
2 Month	Integer	#	15	0	\$	.	,	none
3 Date	Date	yyyy-MM	7		\$	.	,	none

Examine preview data

Rows of step: CSV file input (1000 rows)

	Year	Month	Date
1	2021	3	2021-03
2	2021	3	2021-03
3	2021	3	2021-03
4	2021	3	2021-03
5	2021	3	2021-03
6	2021	4	2021-04
7	2021	4	2021-04
8	2021	4	2021-04
9	2021	4	2021-04
10	2021	4	2021-04
11	2021	4	2021-04
12	2021	4	2021-04
13	2021	4	2021-04
14	2021	4	2021-04
15	2021	5	2021-05
16	2021	5	2021-05
17	2021	5	2021-05
18	2021	5	2021-05
19	2021	5	2021-05
20	2021	5	2021-05
21	2021	5	2021-05
22	2021	5	2021-05

# 5. Time dimension

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 Date	Y	N	N	0	N

Examine preview data

Rows of step: Sort rows (1000 rows)

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

# 5. Time dimension

**Unique rows**

Step name **Unique rows**

Settings

Add counter to output?  Counter Field

Redirect duplicate row  Error description

Fields to compare on (no entries means: compare complete row)

▼	Fieldname	Ignore case
1	Date	N

Examine preview data

Rows of step: Unique rows (33 rows)

	Year	Month	Date
1	2020	11	2020-11
2	2020	12	2020-12
3	2021	1	2021-01
4	2021	2	2021-02
5	2021	3	2021-03
6	2021	4	2021-04
7	2021	5	2021-05
8	2021	6	2021-06
9	2021	7	2021-07
10	2021	8	2021-08
11	2021	9	2021-09
12	2021	10	2021-10
13	2021	11	2021-11
14	2021	12	2021-12
15	2022	1	2022-01
16	2022	2	2022-02
17	2022	3	2022-03
18	2022	4	2022-04
19	2022	5	2022-05
20	2022	6	2022-06
21	2022	7	2022-07
22	2022	8	2022-08

# 5. Time dimension

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	Get fields
1 TIME_ID	=	Date		

Update fields:

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

Examine preview data

Rows of step: Insert / update (33 rows)

	Year	Month	Date
1	2020	11	2020-11
2	2020	12	2020-12
3	2021	1	2021-01
4	2021	2	2021-02
5	2021	3	2021-03
6	2021	4	2021-04
7	2021	5	2021-05
8	2021	6	2021-06
9	2021	7	2021-07
10	2021	8	2021-08
11	2021	9	2021-09
12	2021	10	2021-10
13	2021	11	2021-11
14	2021	12	2021-12
15	2022	1	2022-01
16	2022	2	2022-02
17	2022	3	2022-03
18	2022	4	2022-04
19	2022	5	2022-05
20	2022	6	2022-06
21	2022	7	2022-07
22	2022	8	2022-08

# 5. Location dimension



CSV file input   Sort rows   Unique rows   Add sequence   Insert / update

**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows  Last rows  Off

	District	Municipality	parish
1	PORTE	Felgueiras	REGILDE
2	PORTE	Paredes	SOBROSA
3	SANTAREM	Golegā	POMBALINHO
4	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)
5	VISEU	Viseu	FRAGOSELA
6	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA
7	EVORA	Vila Viçosa	BENCATEL
8	GUARDA	Gouveia	VILA NOVA DE TAZEM
9	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)
10	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA
11	LISBOA	Sintra	COLARES

# 5. Location dimension

**CSV file input**

Step name **CSV file input**

Filename

Delimiter

Enclosure

NIO buffer size

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional)

Running in parallel?

New line possible in fields?

Format

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

**Examine preview data**

Rows of step: CSV file input (1000 rows)

	District	Municipality	parish
1	PORTO	Felgueiras	REGILDE
2	PORTO	Paredes	SOBROSA
3	SANTAREM	Golegã	POMBALINHO
4	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)
5	VISEU	Viseu	FRAGOSELA
6	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA
7	EVORA	Vila Viçosa	BENCATEL
8	GUARDA	Gouveia	VILA NOVA DE TAZEM
9	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)
10	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA
11	LISBOA	Sintra	COLARES
12	PORTO	Penafiel	CABECA SANTA
13	VIANA DO CASTELO	Arcos de Valdevez	SISTELO
14	VISEU	Mortágua	PALA
15	BRAGA	Barcelos	UF BARCELOS V BOA V FRESCAINHA
16	CASTELO BRANCO	Sertã	CABECUDO
17	COIMBRA	Arganil	CELAVISA
18	LISBOA	ALENQUER	UF RIBAFRIA PEREIRO PALHACANA
19	LISBOA	Azambuja	ALCOENTRE
20	SANTAREM	Tomar	SAO PEDRO DE TOMAR
21	VIANA DO CASTELO	Caminha	VILE

# 5. Location dimension

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

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	AI RFRGARIA-A-VFI HA F VAI MAIOR

# 5. Location dimension

Unique rows

Step name Unique rows

Settings

Add counter to output?  Counter field

Redirect duplicate row  Error description

Fields to compare on (no entries means: compare complete row)

Fieldname	Ignore case
1 District	N
2 Municipality	N
3 parish	N

Examine preview data

Rows of step: Unique rows (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

# 5. Location dimension

**Add sequence**

Step name **Add sequence**

Name of value **id**

Use a database to generate the sequence

Use DB to get sequence?

Connection **Connection**

Schema name **Schema name**

Sequence name **SEQ**

Use a transformation counter to generate the sequence

Use counter to calculate sequence?

Counter name (optional) **Counter name (optional)**

Start at value **1**

Increment by **1**

Maximum value **999999999**

Examine preview data

Rows of step: Add sequence (1000 rows)

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

# 5. Location dimension

**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
1 LOCATION_ID	=	id	

Update fields:

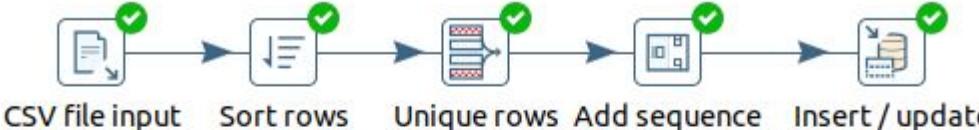
Table field	Stream field	Update
1 LOCATION_ID	id	N
2 PARISH	parish	N
3 MUNICIPALITY	Municipality	N
4 DISTRICT	District	N

**Examine preview data**

Rows of step: Insert / update (1000 rows)

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

## 5. Voltage dimension



The diagram illustrates a data processing workflow consisting of five sequential steps:

- CSV file input
- Sort rows
- Unique rows
- Add sequence
- Insert / update

Each step is represented by a rectangular icon with a green checkmark in the top right corner, indicating successful execution.

**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows  Last rows  Off

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

# 5. Voltage dimension

CSV file input

Step name: CSV file input

Filename: /home/aid/Documents/proj/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:

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

Examine preview data

Rows of step: CSV File input (1000 rows)

Voltage level

Row	Voltage level
1	Baixa Tensão
2	Muito Alta, Alta e Média Tensões
3	Baixa Tensão
4	Muito Alta, Alta e Média Tensões
5	Baixa Tensão
6	Baixa Tensão
7	Baixa Tensão
8	Muito Alta, Alta e Média Tensões
9	Baixa Tensão
10	Muito Alta, Alta e Média Tensões
11	Baixa Tensão
12	Muito Alta, Alta e Média Tensões
13	Baixa Tensão
14	Baixa Tensão
15	Muito Alta, Alta e Média Tensões
16	Baixa Tensão
17	Baixa Tensão
18	Baixa Tensão
19	Baixa Tensão
20	Baixa Tensão
21	Baixa Tensão
22	Muito Alta, Alta e Média Tensões

# 5. Voltage dimension

**Sort rows**

Step name **Sort rows**

Sort directory `%java.io.tmpdir%` [Browse...](#)

TMP-file prefix `out`

Sort size (rows in memory) `1000000`

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

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

**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

# 5. Voltage dimension

### Unique rows

Step name **Unique rows**

Settings

Add counter to output?  Counter field

Redirect duplicate row  Error description

Fields to compare on (no entries means: compare complete row)

Fieldname	Ignore case
1 Voltage level	N

### Examine preview data

Rows of step: Unique rows (2 rows)

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

# 5. Voltage dimension

**Add sequence**

Step name **Add sequence**

Name of value **id**

Use a database to generate the sequence

Use DB to get sequence?

Connection **Connection**

Schema name **Schema name**

Sequence name **SEQ**

Use a transformation counter to generate the sequence

Use counter to calculate sequence?

Counter name (optional)

Start at value **1**

Increment by **1**

Maximum value **999999999**

**Examine preview data**

Rows of step: Add sequence (2 rows)

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

# 5. Voltage dimension

**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	Get fields
1 VOLTAGE_ID	=	id		

Update fields:

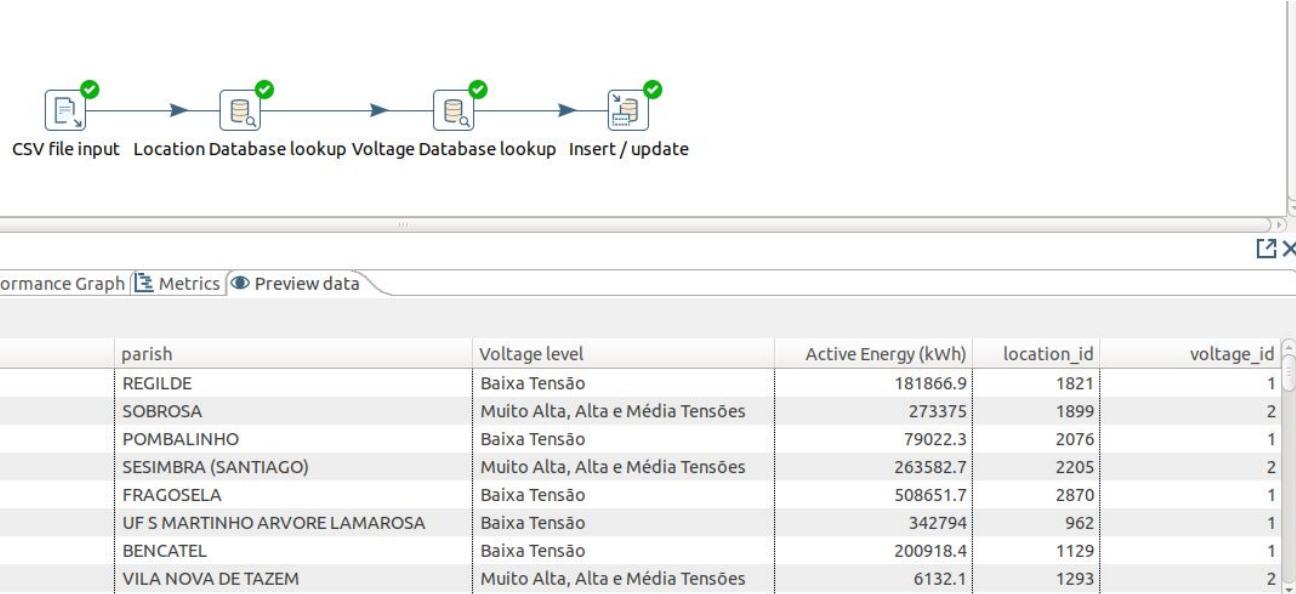
Table field	Stream field	Update	Get update fields	Edit mapping
1 VOLTAGE_ID	id	N		
2 VOLTAGE_LEVEL	Voltage level	N		

**Examine preview data**

Rows of step: Insert / update (2 rows)

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

# 6. Fact table



# 6. Fact table

**CSV file input**

Step name **CSV file input**

Filename

Delimiter

Enclosure

NIO buffer size

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional)

Running in parallel?

New line possible in fields?

Format

File encoding

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1 Date	Date	yyyy-MM	7		\$	.	,	none
2 District	String		16		\$	.	,	none
3 Municipality	String		22		\$	.	,	none
4 parish	String		30		\$	.	,	none
5 Voltage level	String		32		\$	.	,	none
6 Active Energy (kWh)	Number	#.##	12	3	\$	.	,	none

	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)
1	2021-03	PORTO	Felgueiras	REGILDE	Baixa Tensão	181866.9
2	2021-03	PORTO	Paredes	SOBROSA	Muito Alta, Alta e Média Tensões	273375
3	2021-03	SANTAREM	Golegã	POMBALINHO	Baixa Tensão	79022.3
4	2021-03	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)	Muito Alta, Alta e Média Tensões	263582.7
5	2021-03	VISEU	Viseu	FRAGOSELA	Baixa Tensão	508651.7
6	2021-04	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA	Baixa Tensão	342794
7	2021-04	EVORA	Vila Viçosa	BENCATEL	Baixa Tensão	200918.4
8	2021-04	GUARDA	Gouveia	VILA NOVA DE TAZEM	Muito Alta, Alta e Média Tensões	6132.1
9	2021-04	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)	Baixa Tensão	303825
10	2021-04	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA	Muito Alta, Alta e Média Tensões	237087.9
11	2021-04	LISBOA	Sintra	COLARES	Baixa Tensão	2110346.3
12	2021-04	PORTO	Penafiel	CABECA SANTA	Muito Alta, Alta e Média Tensões	370909.4
13	2021-04	VIANA DO CASTELO	Arcos de Valdevez	SISTELO	Baixa Tensão	26991.6
14	2021-04	VISEU	Mortágua	PALA	Baixa Tensão	84457.9
15	2021-05	BRAGA	Barcelos	UF BARCELOS V BOA V FRESCAINHA	Muito Alta, Alta e Média Tensões	2794312.3
16	2021-05	CASTELO BRANCO	Sertã	CABECUDO	Baixa Tensão	120004.3
17	2021-05	COIMBRA	Arganil	CELAVISA	Baixa Tensão	18459.6
18	2021-05	LISBOA	Alenquer	UF RIBAFRIA PEREIRO PALHACANA	Baixa Tensão	199535.6
19	2021-05	LISBOA	Azambuja	ALCOENTRE	Baixa Tensão	409862.9
20	2021-05	SANTAREM	Tomar	SAO PEDRO DE TOMAR	Baixa Tensão	432834.4
21	2021-05	VIANA DO CASTELO	Caminha	VILE	Baixa Tensão	37537.5
22	2021-05	VILA REAL	Montalegre	UF PARADEI A CONTIM E FIAES	Muito Alta, Alta e Média Tensões	15211.1

# 6. Fact table

**Database lookup**

Step name: Location Database lookup

Connection: dataset\_b\_dw    Edit...    New...    Wizard...

Lookup schema: dataset\_b\_dw    Browse...

Lookup table: dim\_location    Browse...

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 PARISH	=	parish	
2 MUNICIPALITY	=	Municipality	
3 DISTRICT	=	District	

Values to return from the lookup table:

Field	New name	Default	Type
1 LOCATION_ID	location_id		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by:

#	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	location_id
1	2021-03	PORTO	Felgueiras	REGILDE	Baixa Tensão	181866.9	1821
2	2021-03	PORTO	Paredes	SOBROSA	Muito Alta, Alta e Média Tensões	273375	1899
3	2021-03	SANTAREM	Golegã	POMBALINHO	Baixa Tensão	79022.3	2076
4	2021-03	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)	Muito Alta, Alta e Média Tensões	263582.7	2205
5	2021-03	VISEU	Viseu	FRAGOSELA	Baixa Tensão	508651.7	2870
6	2021-04	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA	Baixa Tensão	342794	962
7	2021-04	EVORA	Vila Viçosa	BENCATEL	Baixa Tensão	200918.4	1129
8	2021-04	GUARDA	Gouveia	VILA NOVA DE TAZEM	Muito Alta, Alta e Média Tensões	6132.1	1293
9	2021-04	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)	Baixa Tensão	303825	1337
10	2021-04	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA	Muito Alta, Alta e Média Tensões	237087.9	1455
11	2021-04	LISBOA	Sintra	COLARES	Baixa Tensão	2110346.3	1671
12	2021-04	PORTO	Penafiel	CABECA SANTA	Muito Alta, Alta e Média Tensões	370909.4	1917
13	2021-04	VIANA DO CASTELO	Arcos de Valdevez	SISTELO	Baixa Tensão	26991.6	2234
14	2021-04	VISEU	Mortágua	PALA	Baixa Tensão	84457.9	2718
15	2021-05	BRAGA	Barcelos	UF BARCELOS V BOA V FRESCAINHA	Muito Alta, Alta e Média Tensões	2794312.3	285
16	2021-05	CASTELO BRANCO	Sertã	CABECUDO	Baixa Tensão	120004.3	903
17	2021-05	COIMBRA	Arganil	CELAVISA	Baixa Tensão	18459.6	922
18	2021-05	LISBOA	Alenquer	UF RIBAFRIA PEREIRO PALHACANA	Baixa Tensão	199535.6	1575
19	2021-05	LISBOA	Azambuja	ALCOENTRE	Baixa Tensão	409862.9	1588
20	2021-05	SANTAREM	Toramar	SAO PEDRO DE TOMAR	Baixa Tensão	432834.4	2138
21	2021-05	VIANA DO CASTELO	Caminha	VILE	Baixa Tensão	37537.5	2263
22	2021-05	VILA REAL	Montalegre	UF PARADELA CONTIM E FIAES	Muito Alta, Alta e Média Tensões	15211.1	2516

# 6. Fact table

**Database lookup**

Step name **Voltage Database lookup**

Connection **dataset\_b\_dw**

Lookup schema **dataset\_b\_dw**

Lookup table **dim\_voltage**

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 VOLTAGE_LEVEL	=	Voltage level	

Values to return from the lookup table :

Field	New name	Default	Type
1 VOLTAGE_ID	voltage_id		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Date	District	Municipality	parish	Voltage level	Active Energy (kwh)	location_id	voltage_id
1 2021-03	PORTO	Felgueiras	REGILDE	Baixa Tensão	181866.9	1821	1
2 2021-03	PORTO	Paredes	SOBROSA	Muito Alta, Alta e Média Tensões	273375	1899	2
3 2021-03	SANTAREM	Golegã	POMBALINHO	Baixa Tensão	79022.3	2076	1
4 2021-03	SETUBAL	Sesimbra	SESIMBRA (SANTIAGO)	Muito Alta, Alta e Média Tensões	263582.7	2205	2
5 2021-03	VIDEU	Viseu	FRAGOSOLA	Baixa Tensão	508651.7	2870	1
6 2021-04	COIMBRA	Coimbra	UF S MARTINHO ARVORE LAMAROSA	Baixa Tensão	342794	962	1
7 2021-04	EVORA	Vila Viçosa	BENCATEL	Baixa Tensão	200918.4	1129	1
8 2021-04	GUARDA	Gouveia	VILA NOVA DE TAZEM	Muito Alta, Alta e Média Tensões	6132.1	1293	2
9 2021-04	GUARDA	Manteigas	MANTEIGAS (SANTA MARIA)	Baixa Tensão	303825	1337	1
10 2021-04	GUARDA	Vila Nova de Foz Côa	VILA NOVA DE FOZ COA	Muito Alta, Alta e Média Tensões	237087.9	1455	2
11 2021-04	LISBOA	Sintra	COLARES	Baixa Tensão	2110346.3	1671	1
12 2021-04	PORTO	Penafiel	CABECA SANTA	Muito Alta, Alta e Média Tensões	370909.4	1917	2
13 2021-04	VIANA DO CASTELO	Arcos de Valdevez	SISTELO	Baixa Tensão	26991.6	2234	1
14 2021-04	VIDEU	Mortágua	PALA	Baixa Tensão	84457.9	2718	1
15 2021-05	BRAGA	Barcelos	UF BARCELOS V BOA V FRESCAINHA	Muito Alta, Alta e Média Tensões	2794312.3	285	2
16 2021-05	CASTELO BRANCO	Sertã	CABECUDO	Baixa Tensão	120004.3	903	1
17 2021-05	COIMBRA	Arganil	CELAVISA	Baixa Tensão	18459.6	922	1
18 2021-05	LISBOA	Alenquer	UF RIBAFRIA PEREIRO PALHACANA	Baixa Tensão	199535.6	1575	1
19 2021-05	LISBOA	Azambuja	ALCENTRE	Baixa Tensão	409862.9	1588	1
20 2021-05	SANTAREM	Tomar	SAO PEDRO DE TOMAR	Baixa Tensão	432834.4	2138	1
21 2021-05	VIANA DO CASTELO	Caminha	VILE	Baixa Tensão	37537.5	2263	1
22 2021-05	VILA REAL	Montalegre	UF PARADELA CONTIM E FIAES	Muito Alta, Alta e Média Tensões	15211.1	2516	2

# 6. Fact table

Insert / update

Step name **Insert / update**

Connection **dataset\_b\_dw**

Target schema **dataset\_b\_dw**

Target table **fact\_consumption**

Commit size **100**

Don't perform any updates:

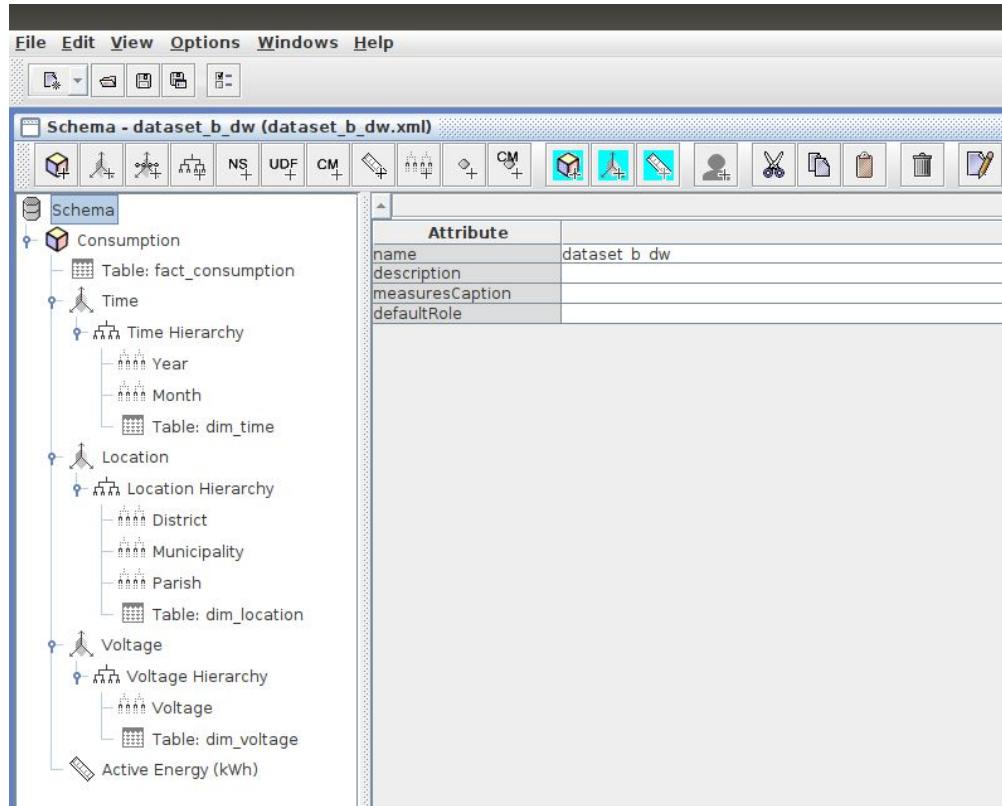
The key(s) to look up the value(s):

▼	Table field	Comparator	Stream field1	Stream field2	<input type="button" value="Get fields"/>
1	TIME_ID	=	Date		
2	LOCATION_ID	=	location_id		
3	VOLTAGE_ID	=	voltage_id		

Update fields:

▼	Table field	Stream field	Update	<input type="button" value="Get update fields"/>
1	TIME_ID	Date	Y	
2	LOCATION_ID	location_id	Y	
3	VOLTAGE_ID	voltage_id	Y	
4	ENERGY_CONSUMPTION	Active Energy (kWh)	Y	<input type="button" value="Edit mapping"/>

# 7. Pentaho Schema Workbench



# 7. XML file

```
<Schema name="dataset_b_dw">
  <Cube name="Consumption" visible="true" cache="true" enabled="true">
    <Table name="fact_consumption">
      </Table>
    <Dimension type="TimeDimension" visible="true" foreignKey="TIME_ID" highCardinality="false" 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" hideMemberIf="Never">
          </Level>
        <Level name="Month" visible="true" column="MONTH_ID" type="Integer" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
          </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="LOCATION_ID" highCardinality="false" 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" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
        <Level name="Municipality" visible="true" column="MUNICIPALITY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
        <Level name="Parish" visible="true" column="PARISH" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="VOLTAGE_ID" highCardinality="false" name="Voltage">
      <Hierarchy name="Voltage Hierarchy" visible="true" hasAll="true" allMemberName="All Voltage Levels" primaryKey="VOLTAGE_ID">
        <Table name="dim_voltage">
          </Table>
        <Level name="Voltage" visible="true" column="VOLTAGE_LEVEL" ordinalColumn="VOLTAGE_ID" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="Active Energy (kWh)" column="ENERGY_CONSUMPTION" datatype="Numeric" formatString="#,###.000" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

# 8. Consumption by district and year and evolution from 2021 to 2022 by voltage type

```

1 SELECT {Time.Year.[2021], Time.Year.[2022]} ON COLUMNS,
2 Location.District.Members ON ROWS
3 FROM Consumption
4 WHERE Measures.[Active Energy (kWh)]

```

1, 43

District	2021	2022
AVEIRO	4,381,127,458.000	4,351,723,261.454
BEJA	1,087,511,836.000	1,136,176,040.000
BRAGA	3,407,041,095.000	3,454,781,865.220
BRAGANCA	379,473,577.000	388,817,972.000
CASTELO BRANCO	771,822,945.850	816,148,419.150
COIMBRA	2,689,109,688.000	2,612,783,627.000
EVORA	712,514,243.000	780,338,813.000
FARO	2,166,312,114.000	2,367,894,146.310
GUARDA	467,885,614.000	482,750,861.000
LEIRIA	2,517,152,013.000	2,553,090,826.970
LISBOA	8,270,141,703.000	8,670,500,520.346
PORTALEGRE	468,289,813.000	482,877,968.406
PORTO	7,314,575,940.720	7,386,737,067.874
SANTAREM	2,161,810,637.000	2,158,709,002.642
SETUBAL	4,776,806,313.000	4,853,113,930.288
VIANA DO CASTELO	1,083,665,399.000	1,098,441,193.290
VILA REAL	547,451,452.000	572,536,374.612
VIDEU	1,384,383,867.000	1,389,813,573.596

```

1 WITH MEMBER Measures.[Previous Year] AS Time.Year.CurrentMember.PrevMember
2 MEMBER Measures.[Consumption Growth] AS Measures.[Active Energy (kWh)] - Measures.[Previous Year]
3 SELECT {Voltage.Members} ON COLUMNS,
4 Location.District.Members ON ROWS
5 FROM Consumption
6 WHERE {Time.Year.[2022], Measures.[Consumption Growth])}

```

6, 55

District	All Voltage Levels		
	Baixa Tensão	Muito Alta, Alta e Média Tensões	
AVEIRO	-29,404,196.546	7,047,811.334	-36,452,007.880
BEJA	48,664,204.000	2,375,766.000	46,288,438.000
BRAGA	47,740,770.220	40,119,973.220	7,620,797.000
BRAGANCA	9,344,395.000	7,350,108.000	1,994,287.000
CASTELO BRANCO	44,325,473.300	15,337,120.300	28,988,353.000
COIMBRA	-76,326,061.000	10,955,923.000	-87,281,984.000
EVORA	67,824,570.000	6,444,965.000	61,379,605.000
FARO	201,582,032.310	121,937,385.310	79,644,647.000
GUARDA	14,865,247.000	5,449,557.000	9,415,690.000
LEIRIA	35,938,813.970	20,514,440.970	15,424,373.000
LISBOA	400,358,817.346	151,739,665.346	248,619,152.000
PORTALEGRE	14,588,155.406	1,631,679.406	12,956,476.000
PORTO	72,161,127.154	41,054,966.154	31,106,161.000
SANTAREM	-3,101,634.358	20,601,495.642	-23,703,130.000
SETUBAL	76,307,617.288	13,740,947.288	62,566,670.000
VIANA DO CASTELO	14,775,794.290	16,517,985.290	-1,742,191.000
VILA REAL	25,084,922.612	11,448,945.612	13,635,977.000
VIDEU	5,429,706.596	19,991,602.596	-14,561,896.000