TÉCNICO LISBOA

Data Analysis and Integration

Project G12

Ana Margarida Cardoso Almeida - 102618
Vasco Manuel Dias Paisana - 102533

In this project, we will be using two datasets *(version from 3rd October)* by E-REDES available in the Open Data Portal:

Dataset A:

Number of active energy contracts by meter type

Dataset B:

Monthly consumption by municipality

Updated datasets are available for download as a CSV file on the following website:

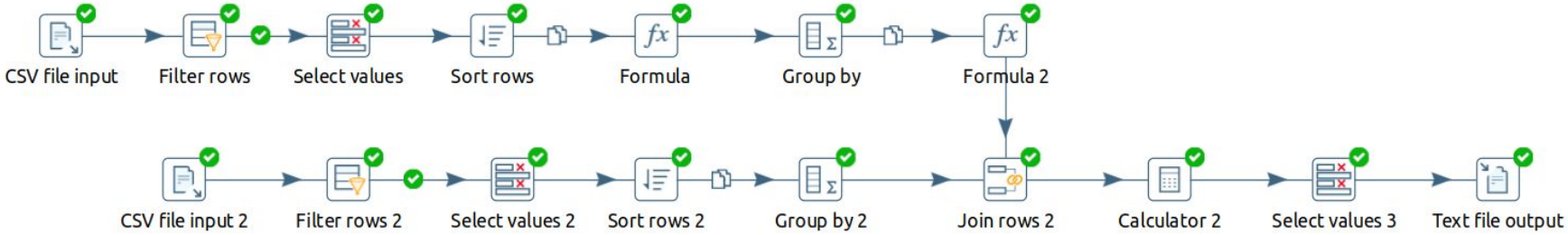
https://e-redes.opendatasoft.com/explore/

These datasets contain real-world data about the energy distribution network in Portugal.

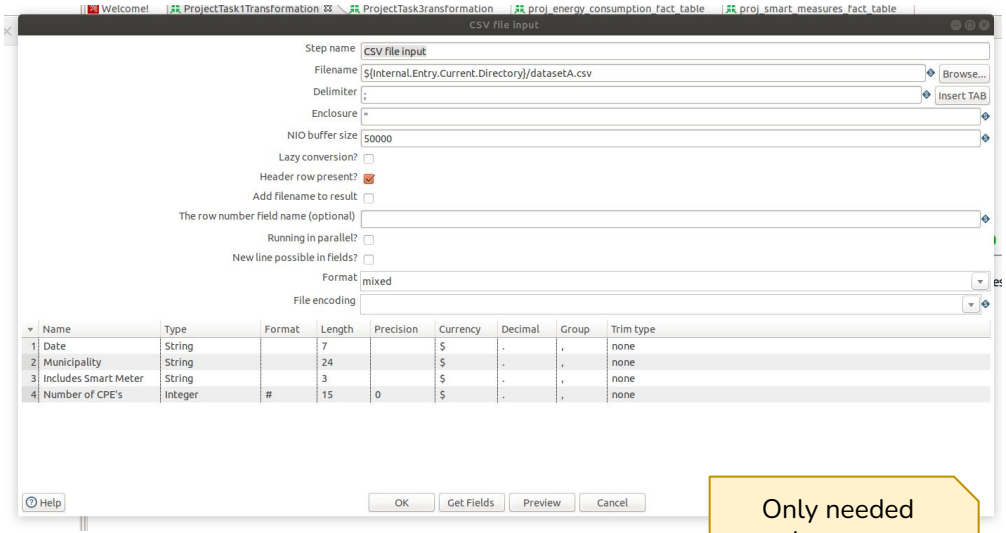
Task 1

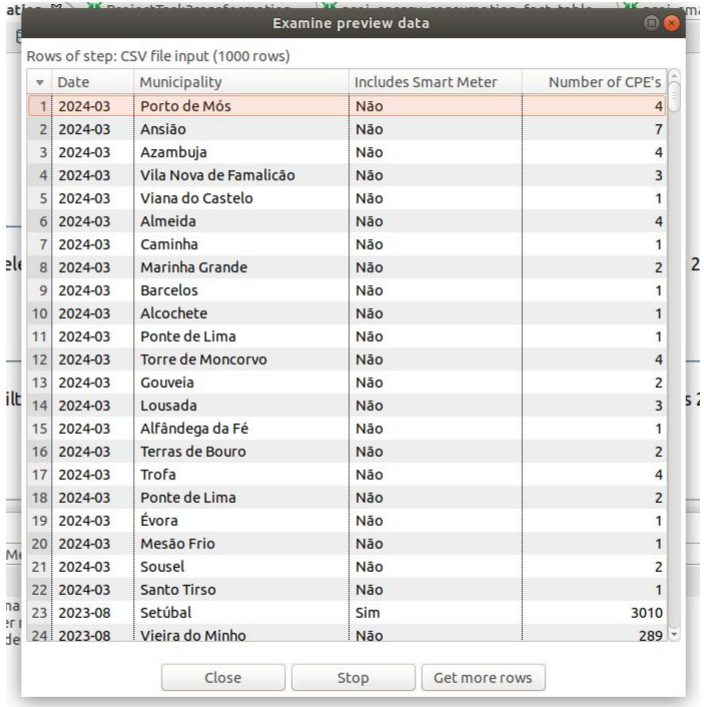
Build a transformation that provides the percentage of smart meters and the average energy consumption per contract.

Transformation



CSV file input

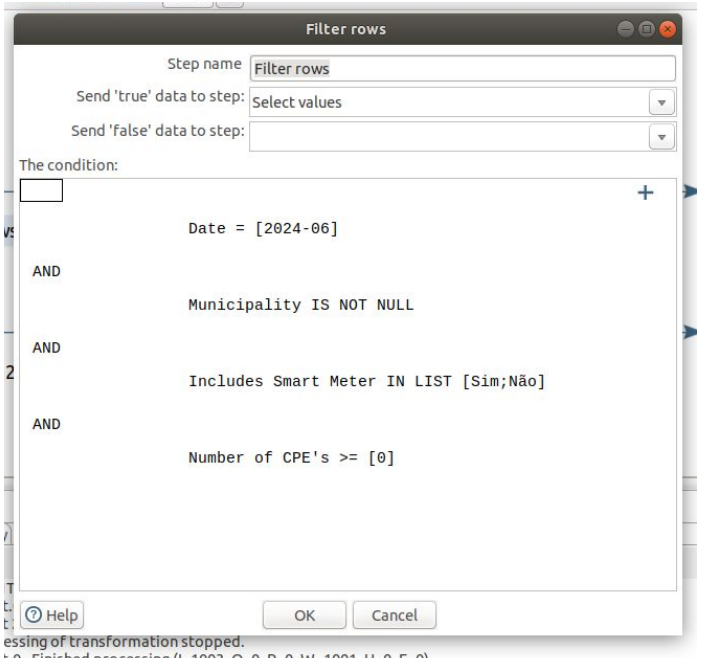


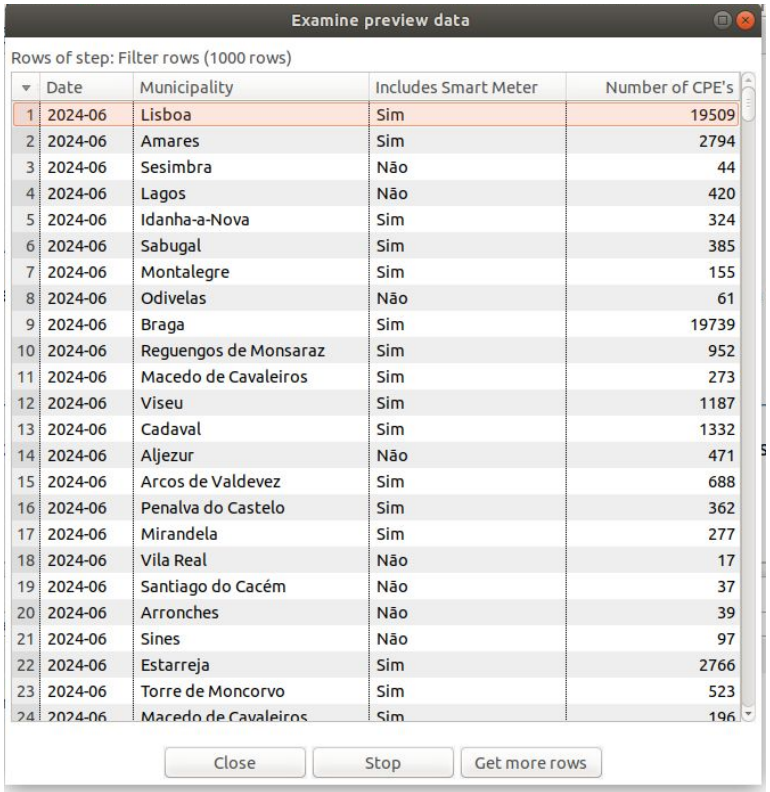


Only needed columns are retrieved from the dataset.

Filter rows

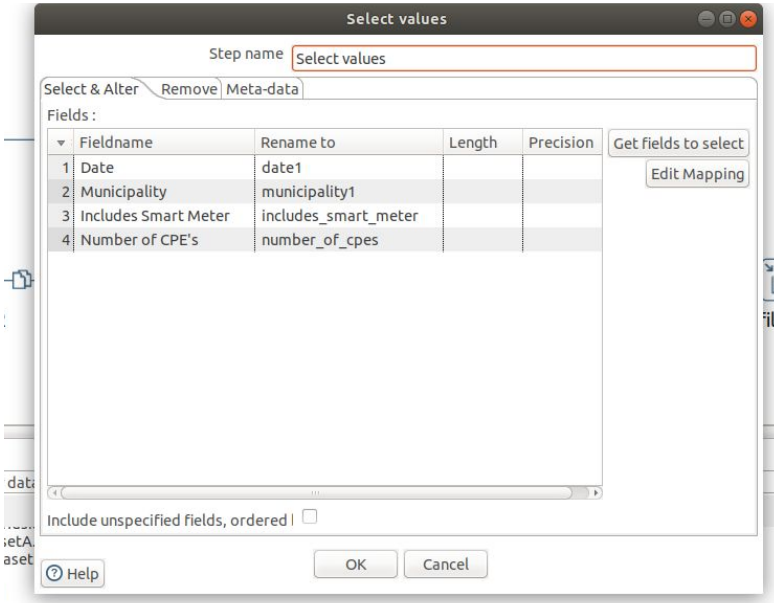
This step ensures validity of data and limits its date to June 2024

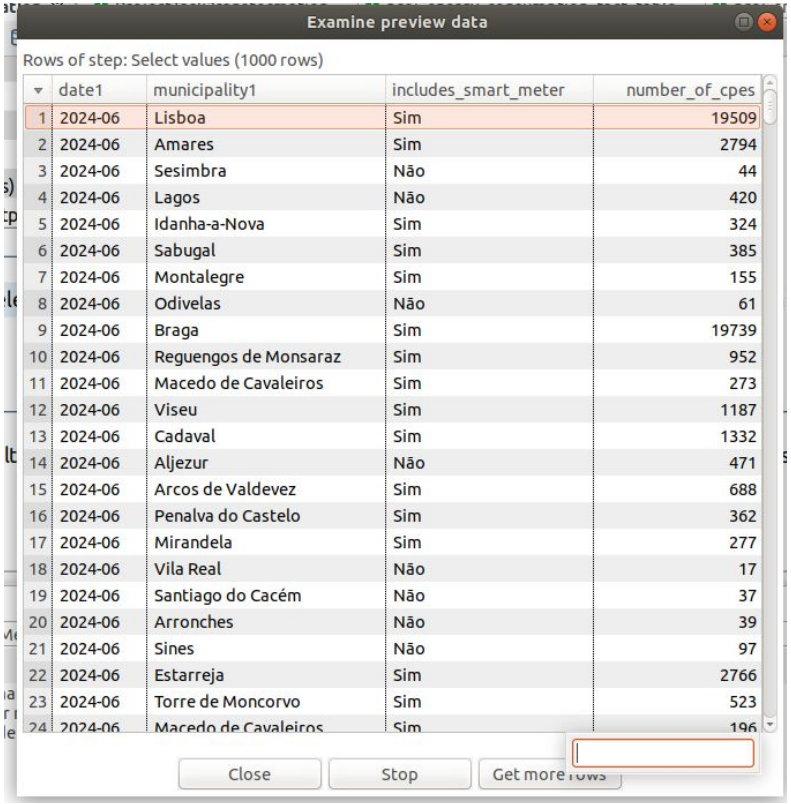




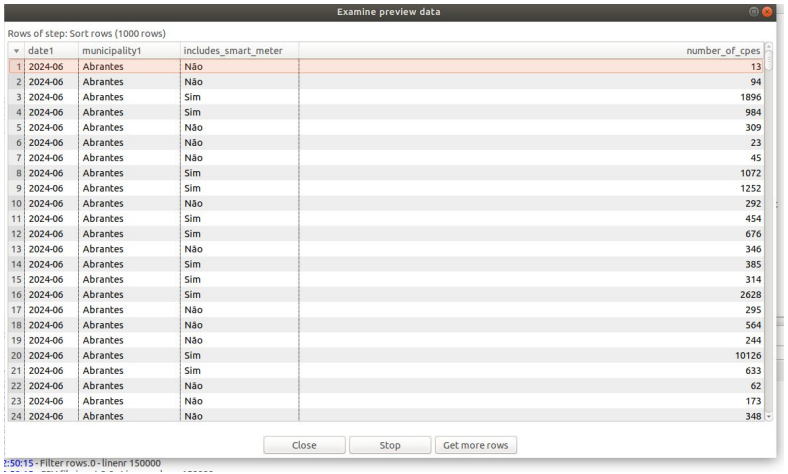
	Date	Municipality	Includes Smart Meter	Number of CPE's
1	2024-06	Lisboa	Sim	19509
2	2024-06	Amares	Sim	2794
3	2024-06	Sesimbra	Não	44
4	2024-06	Lagos	Não	420
5	2024-06	Idanha-a-Nova	Sim	324
6	2024-06	Sabugal	Sim	385
7	2024-06	Montalegre	Sim	155
8	2024-06	Odivelas	Não	61
9	2024-06	Braga	Sim	19739
10	2024-06	Reguengos de Monsaraz	Sim	952
11	2024-06	Macedo de Cavaleiros	Sim	273
12	2024-06	Viseu	Sim	1187
13	2024-06	Cadaval	Sim	1332
14	2024-06	Aljezur	Não	471
15	2024-06	Arcos de Valdevez	Sim	688
16	2024-06	Penalva do Castelo	Sim	362
17	2024-06	Mirandela	Sim	277
18	2024-06	Vila Real	Não	17
19	2024-06	Santiago do Cacém	Não	37
20	2024-06	Arronches	Não	39
21	2024-06	Sines	Não	97
22	2024-06	Estarreja	Sim	2766
23	2024-06	Torre de Moncorvo	Sim	523
24	2024-06	Macedo de Cavaleiros	Sim	196

Select Values

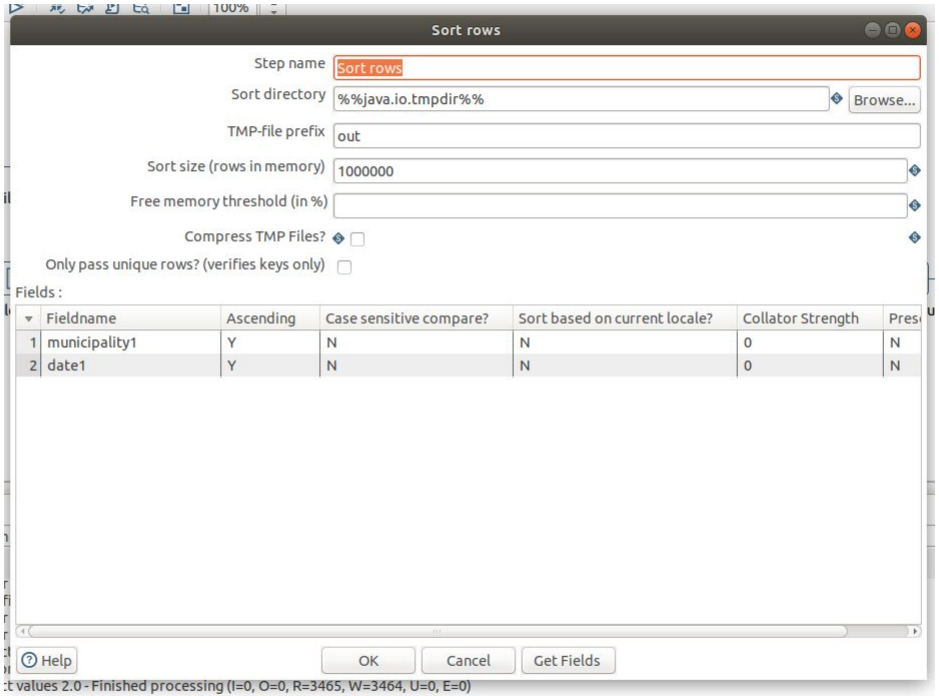




Sort rows

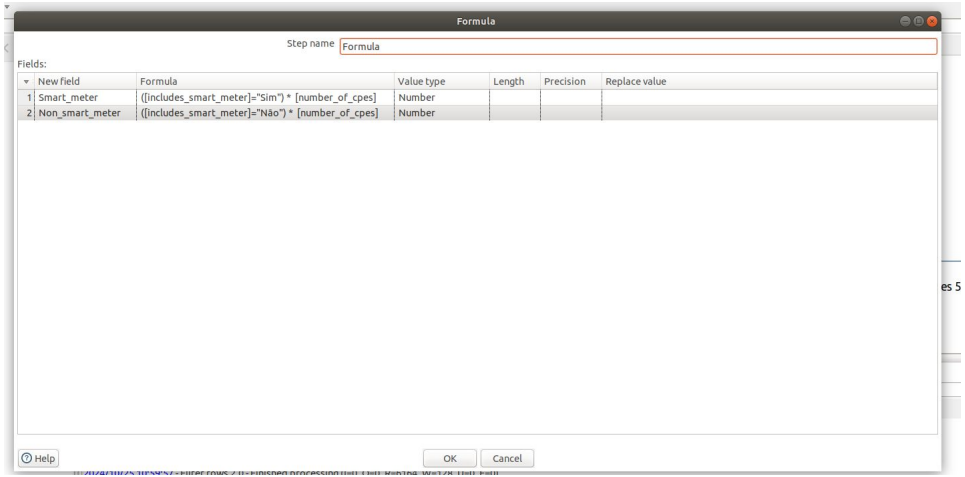


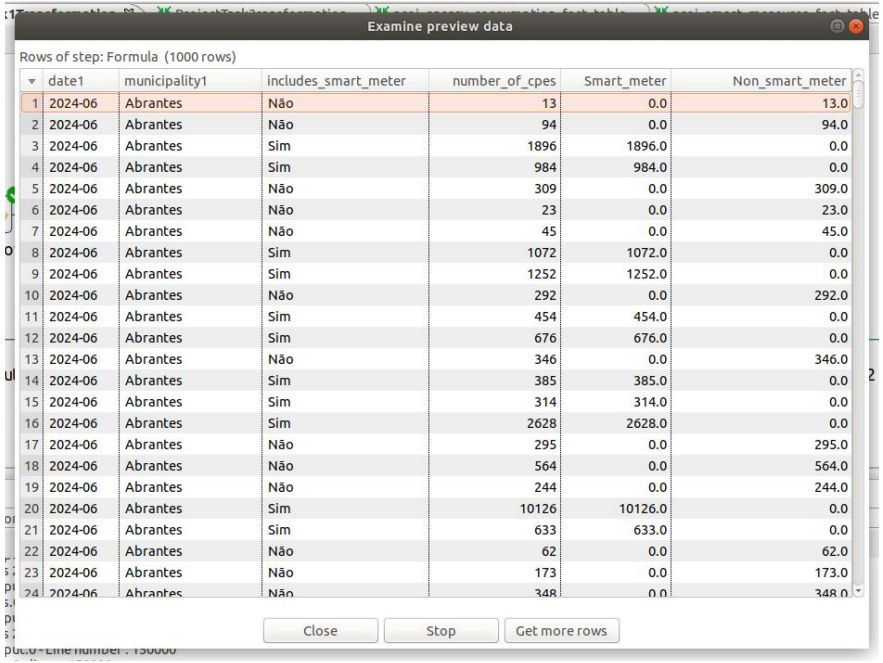
	date1	municipality1	includes_smart_meter	number_of_cpes
1	2024-06	Abrantes	Não	13
2	2024-06	Abrantes	Não	94
3	2024-06	Abrantes	Sim	1896
4	2024-06	Abrantes	Sim	984
5	2024-06	Abrantes	Não	309
6	2024-06	Abrantes	Não	23
7	2024-06	Abrantes	Não	45
8	2024-06	Abrantes	Sim	1072
9	2024-06	Abrantes	Sim	1252
10	2024-06	Abrantes	Não	292
11	2024-06	Abrantes	Sim	454
12	2024-06	Abrantes	Sim	676
13	2024-06	Abrantes	Não	346
14	2024-06	Abrantes	Sim	385
15	2024-06	Abrantes	Sim	314
16	2024-06	Abrantes	Sim	2628
17	2024-06	Abrantes	Não	295
18	2024-06	Abrantes	Não	564
19	2024-06	Abrantes	Não	244
20	2024-06	Abrantes	Sim	10126
21	2024-06	Abrantes	Sim	633
22	2024-06	Abrantes	Não	62
23	2024-06	Abrantes	Não	173
24	2024-06	Abrantes	Não	348



	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Pres
1	municipality1	Y	N	N	0	N
2	date1	Y	N	N	0	N

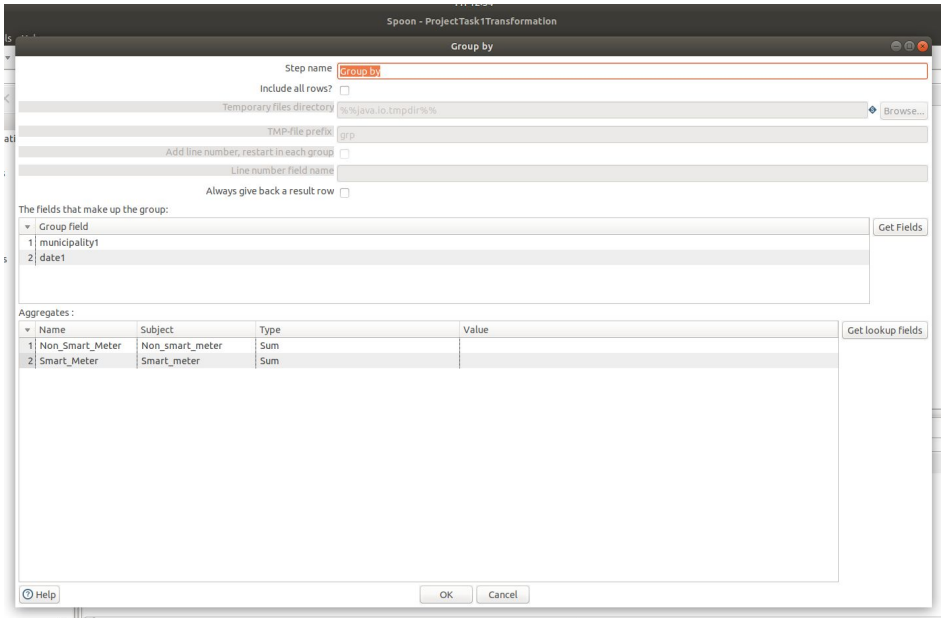
Formula

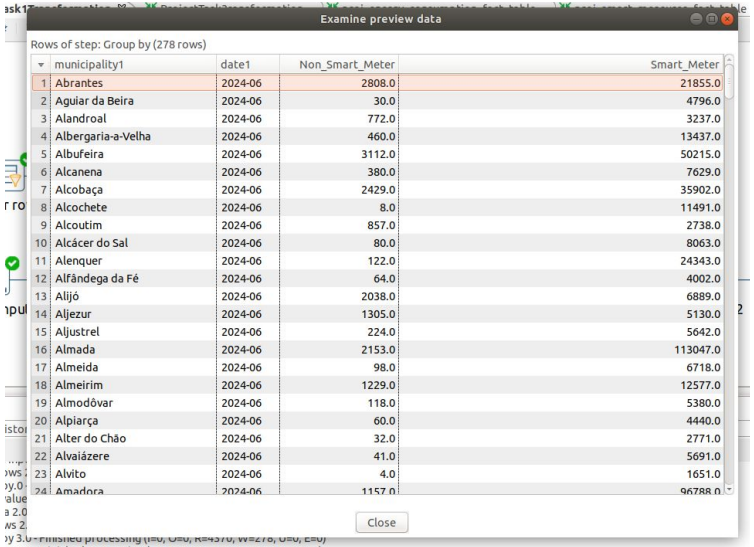




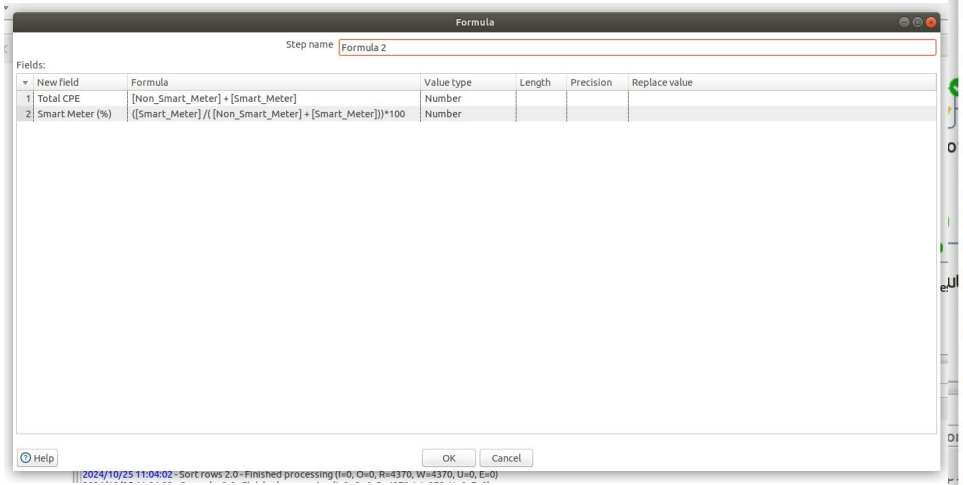
	date1	municipality1	includes_smart_meter	number_of_cpes	Smart_meter	Non_smart_meter
1	2024-06	Abrantes	Não	13	0.0	13.0
2	2024-06	Abrantes	Não	94	0.0	94.0
3	2024-06	Abrantes	Sim	1896	1896.0	0.0
4	2024-06	Abrantes	Sim	984	984.0	0.0
5	2024-06	Abrantes	Não	309	0.0	309.0
6	2024-06	Abrantes	Não	23	0.0	23.0
7	2024-06	Abrantes	Não	45	0.0	45.0
8	2024-06	Abrantes	Sim	1072	1072.0	0.0
9	2024-06	Abrantes	Sim	1252	1252.0	0.0
10	2024-06	Abrantes	Não	292	0.0	292.0
11	2024-06	Abrantes	Sim	454	454.0	0.0
12	2024-06	Abrantes	Sim	676	676.0	0.0
13	2024-06	Abrantes	Não	346	0.0	346.0
14	2024-06	Abrantes	Sim	385	385.0	0.0
15	2024-06	Abrantes	Sim	314	314.0	0.0
16	2024-06	Abrantes	Sim	2628	2628.0	0.0
17	2024-06	Abrantes	Não	295	0.0	295.0
18	2024-06	Abrantes	Não	564	0.0	564.0
19	2024-06	Abrantes	Não	244	0.0	244.0
20	2024-06	Abrantes	Sim	10126	10126.0	0.0
21	2024-06	Abrantes	Sim	633	633.0	0.0
22	2024-06	Abrantes	Não	62	0.0	62.0
23	2024-06	Abrantes	Não	173	0.0	173.0
24	2024-06	Abrantes	Não	348	0.0	348.0

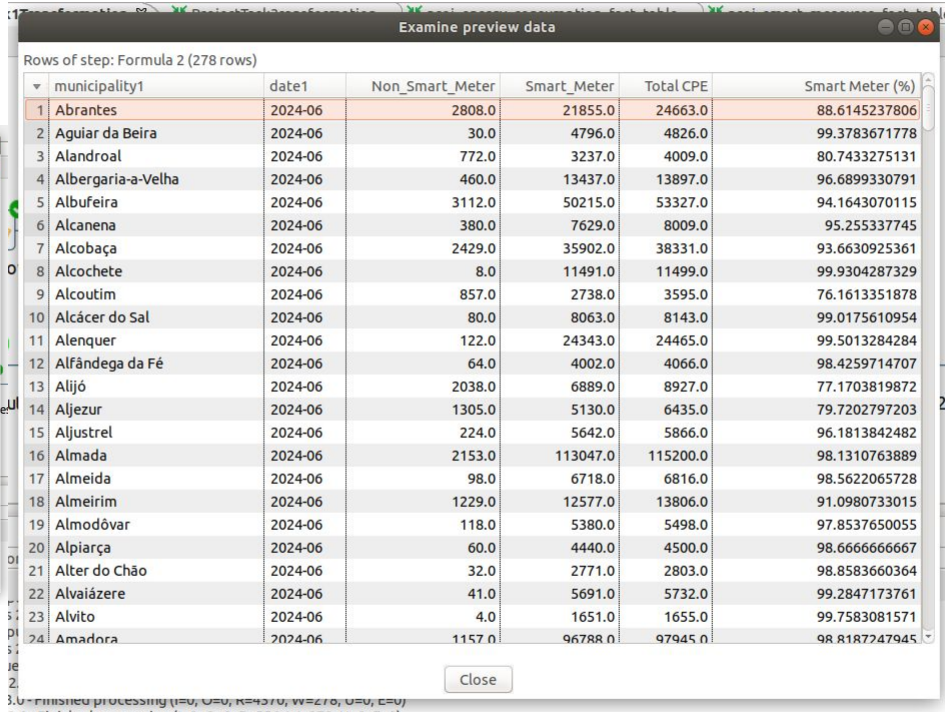
Group by


Spoon - ProjectTask1Transformation
Group by
Step name
Group by
Include all rows?
Temporary files directory
%%java.io.tmpdir%%
Browse...
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
Get Fields
1 municipality1
2 date1
Aggregates :
Name
Subject
Type
Value
Get lookup fields
1 Non_Smart_Meter Non_smart_meter Sum
2 Smart_Meter Smart_meter Sum
Help
OK
Cancel

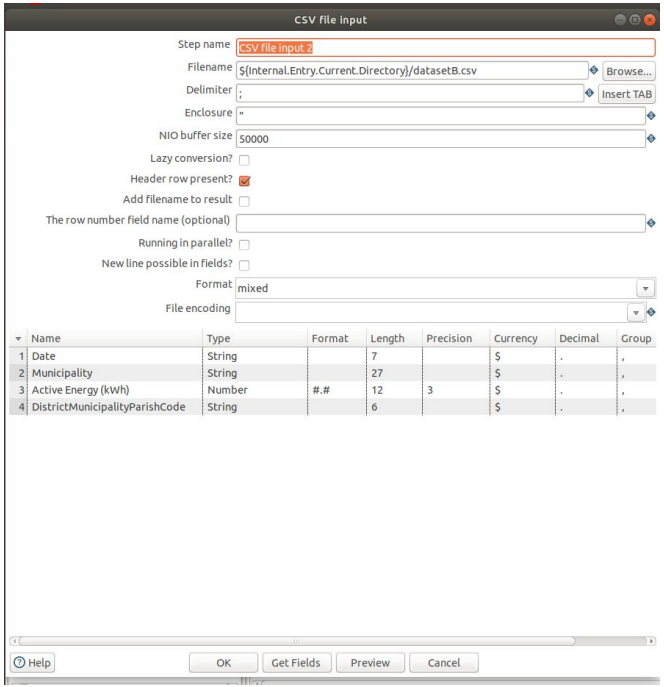

Examine preview data
Rows of step: Group by (278 rows)
municipality1 date1 Non_Smart_Meter Smart_Meter
1 Abrantes 2024-06 2808.0 21855.0
2 Aguiar da Beira 2024-06 30.0 4796.0
3 Alandroal 2024-06 772.0 3237.0
4 Albergaria-a-Velha 2024-06 460.0 13437.0
5 Albufeira 2024-06 3112.0 50215.0
6 Alcanena 2024-06 380.0 7629.0
7 Alcobaça 2024-06 2429.0 35902.0
8 Alcochete 2024-06 8.0 11491.0
9 Alcoutim 2024-06 857.0 2738.0
10 Alcácer do Sal 2024-06 80.0 8063.0
11 Alenquer 2024-06 122.0 24343.0
12 Alfândega da Fé 2024-06 64.0 4002.0
13 Alijó 2024-06 2038.0 6889.0
14 Aljezur 2024-06 1305.0 5130.0
15 Aljustrel 2024-06 224.0 5642.0
16 Almada 2024-06 2153.0 113047.0
17 Almeida 2024-06 98.0 6718.0
18 Almeirim 2024-06 1229.0 12577.0
19 Almodôvar 2024-06 118.0 5380.0
20 Alpiarça 2024-06 60.0 4440.0
21 Alter do Chão 2024-06 32.0 2771.0
22 Alvaiázere 2024-06 41.0 5691.0
23 Alvito 2024-06 4.0 1651.0
24 Amadora 2024-06 1157.0 96788.0
Close

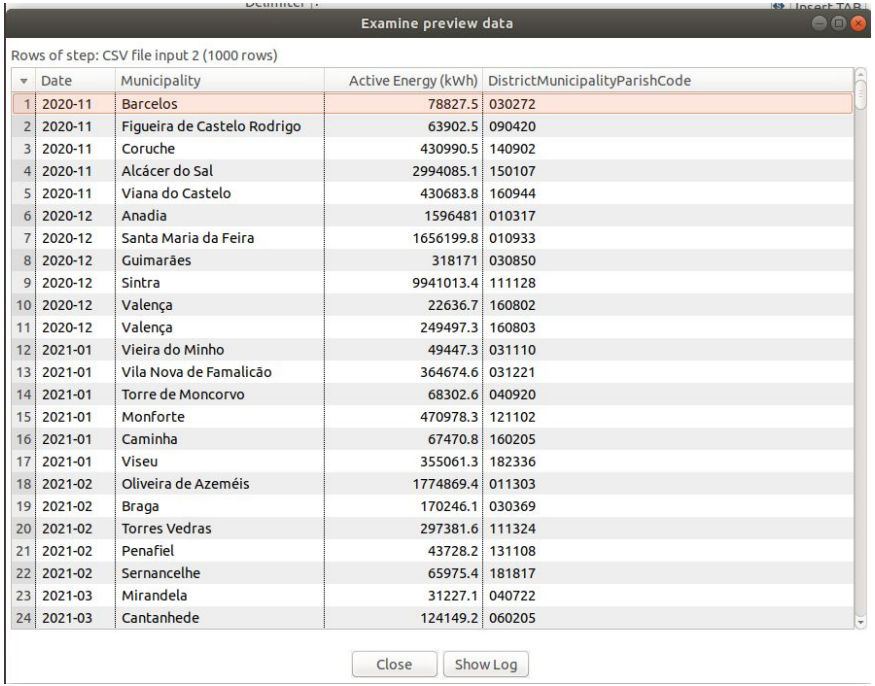
Formula 2


Formula
Step name
Formula 2
Fields:
New field
Formula
Value type
Length
Precision
Replace value
1
Total CPE
[Non_Smart_Meter] + [Smart_Meter]
Number
2
Smart Meter (%)
([Smart_Meter] /([Non_Smart_Meter] + [Smart_Meter]))*100
Number
Help
OK
Cancel
2024/10/25 11:04:02 - Sort rows 2.0 - Finished processing (I=0, O=0, R=4370, W=4370, U=0, E=0)


Examine preview data
Rows of step: Formula 2 (278 rows)
municipality1
date1
Non_Smart_Meter
Smart_Meter
Total CPE
Smart Meter (%)
1 Abrantes 2024-06 2808.0 21855.0 24663.0 88.6145237806
2 Aguiar da Beira 2024-06 30.0 4796.0 4826.0 99.3783671778
3 Alandroal 2024-06 772.0 3237.0 4009.0 80.7433275131
4 Albergaria-a-Velha 2024-06 460.0 13437.0 13897.0 96.6899330791
5 Albufeira 2024-06 3112.0 50215.0 53327.0 94.1643070115
6 Alcanena 2024-06 380.0 7629.0 8009.0 95.255337745
7 Alcobaça 2024-06 2429.0 35902.0 38331.0 93.6630925361
8 Alcochete 2024-06 8.0 11491.0 11499.0 99.9304287329
9 Alcoutim 2024-06 857.0 2738.0 3595.0 76.1613351878
10 Alcácer do Sal 2024-06 80.0 8063.0 8143.0 99.0175610954
11 Alenquer 2024-06 122.0 24343.0 24465.0 99.5013284284
12 Alfândega da Fé 2024-06 64.0 4002.0 4066.0 98.4259714707
13 Alijó 2024-06 2038.0 6889.0 8927.0 77.1703819872
14 Aljezur 2024-06 1305.0 5130.0 6435.0 79.7202797203
15 Aljustrel 2024-06 224.0 5642.0 5866.0 96.1813842482
16 Almada 2024-06 2153.0 113047.0 115200.0 98.1310763889
17 Almeida 2024-06 98.0 6718.0 6816.0 98.5622065728
18 Almeirim 2024-06 1229.0 12577.0 13806.0 91.0980733015
19 Almodôvar 2024-06 118.0 5380.0 5498.0 97.8537650055
20 Alpiarça 2024-06 60.0 4440.0 4500.0 98.6666666667
21 Alter do Chão 2024-06 32.0 2771.0 2803.0 98.8583660364
22 Alvaiázere 2024-06 41.0 5691.0 5732.0 99.2847173761
23 Alvito 2024-06 4.0 1651.0 1655.0 99.7583081571
24 Amadora 2024-06 1157.0 96788.0 97945.0 98.8187247945
Close

CSV file input 2





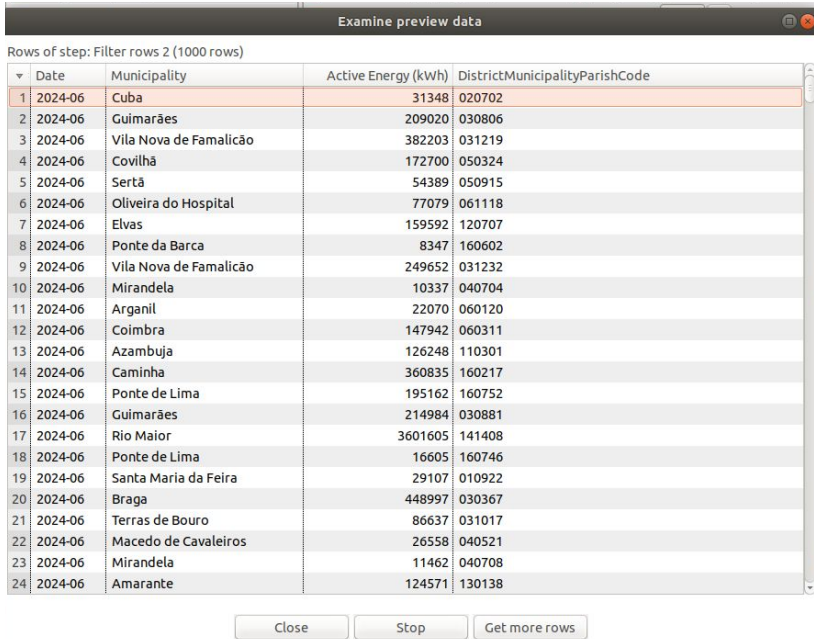
	Date	Municipality	Active Energy (kWh)	DistrictMunicipalityParishCode
1	2020-11	Barcelos	78827.5	030272
2	2020-11	Figueira de Castelo Rodrigo	63902.5	090420
3	2020-11	Coruche	430990.5	140902
4	2020-11	Alcácer do Sal	2994085.1	150107
5	2020-11	Viana do Castelo	430683.8	160944
6	2020-12	Anadia	1596481	010317
7	2020-12	Santa Maria da Feira	1656199.8	010933
8	2020-12	Guimarães	318171	030850
9	2020-12	Sintra	9941013.4	111128
10	2020-12	Valença	22636.7	160802
11	2020-12	Valença	249497.3	160803
12	2021-01	Vieira do Minho	49447.3	031110
13	2021-01	Vila Nova de Famalicão	364674.6	031221
14	2021-01	Torre de Moncorvo	68302.6	040920
15	2021-01	Monforte	470978.3	121102
16	2021-01	Caminha	67470.8	160205
17	2021-01	Viseu	355061.3	182336
18	2021-02	Oliveira de Azeméis	1774869.4	011303
19	2021-02	Braga	170246.1	030369
20	2021-02	Torres Vedras	297381.6	111324
21	2021-02	Penafiel	43728.2	131108
22	2021-02	Sernancelhe	65975.4	181817
23	2021-03	Mirandela	31227.1	040722
24	2021-03	Cantanhede	124149.2	060205

Filter rows 2

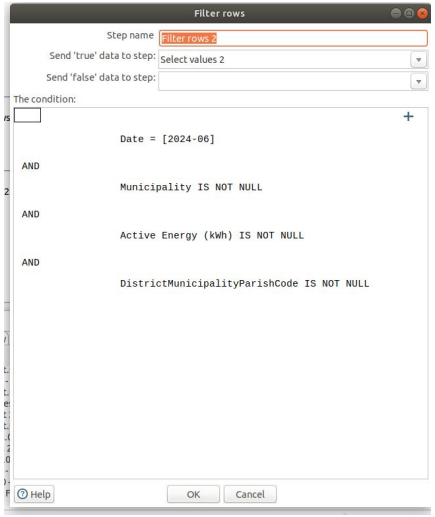
For General Data Protection Regulation reasons, some consumptions appear aggregated in a category which includes all consumption in the District that, for the reasons mentioned, cannot be made available in the respective Parish/Municipality

The corresponding fields have null parish code, and municipality name equal to the belonging district.

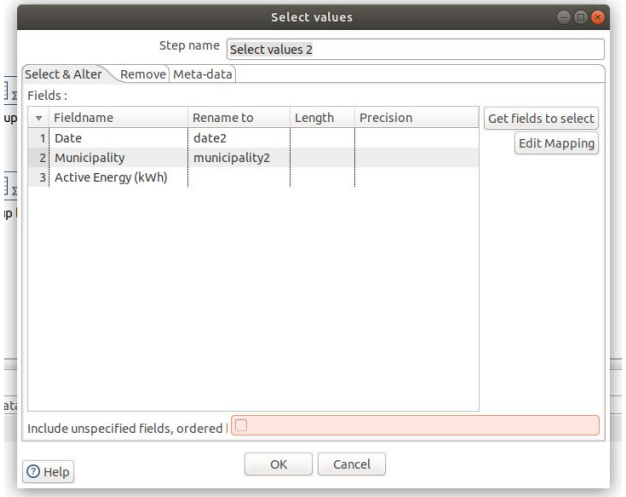
Therefore this data was filtered to ensure reliability of data when grouping by municipality.

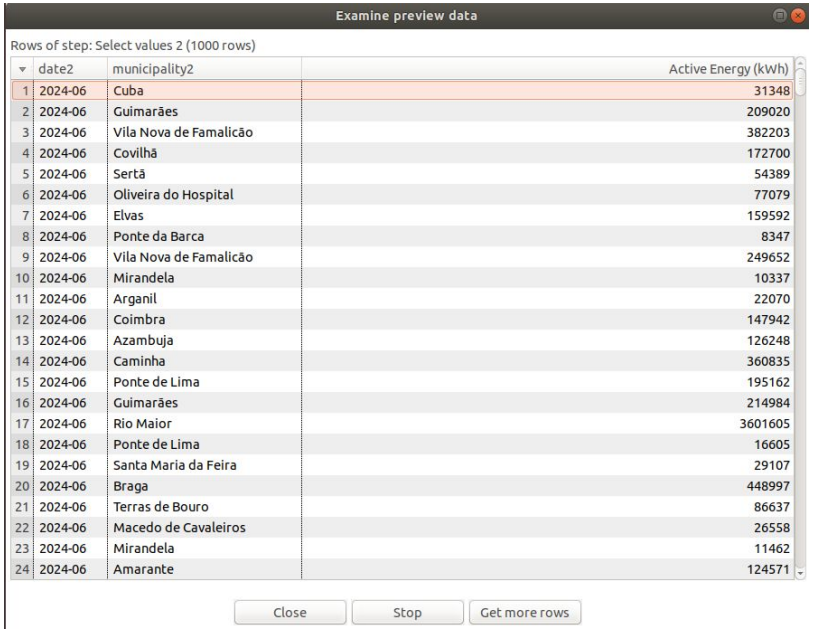


	Date	Municipality	Active Energy (kWh)	DistrictMunicipalityParishCode
1	2024-06	Cuba	31348	020702
2	2024-06	Guimarães	209020	030806
3	2024-06	Vila Nova de Famalicão	382203	031219
4	2024-06	Covilhã	172700	050324
5	2024-06	Sertã	54389	050915
6	2024-06	Oliveira do Hospital	77079	061118
7	2024-06	Elvas	159592	120707
8	2024-06	Ponte da Barca	8347	160602
9	2024-06	Vila Nova de Famalicão	249652	031232
10	2024-06	Mirandela	10337	040704
11	2024-06	Arganil	22070	060120
12	2024-06	Coimbra	147942	060311
13	2024-06	Azambuja	126248	110301
14	2024-06	Caminha	360835	160217
15	2024-06	Ponte de Lima	195162	160752
16	2024-06	Guimarães	214984	030881
17	2024-06	Rio Maior	3601605	141408
18	2024-06	Ponte de Lima	16605	160746
19	2024-06	Santa Maria da Feira	29107	010922
20	2024-06	Braga	448997	030367
21	2024-06	Terras de Bouro	86637	031017
22	2024-06	Macedo de Cavaleiros	26558	040521
23	2024-06	Mirandela	11462	040708
24	2024-06	Amarante	124571	130138

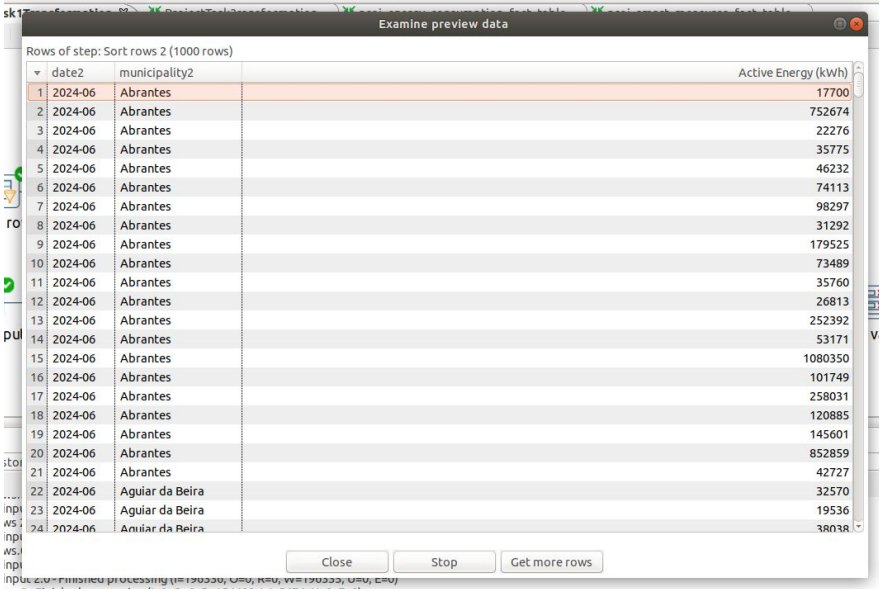


Select values 2

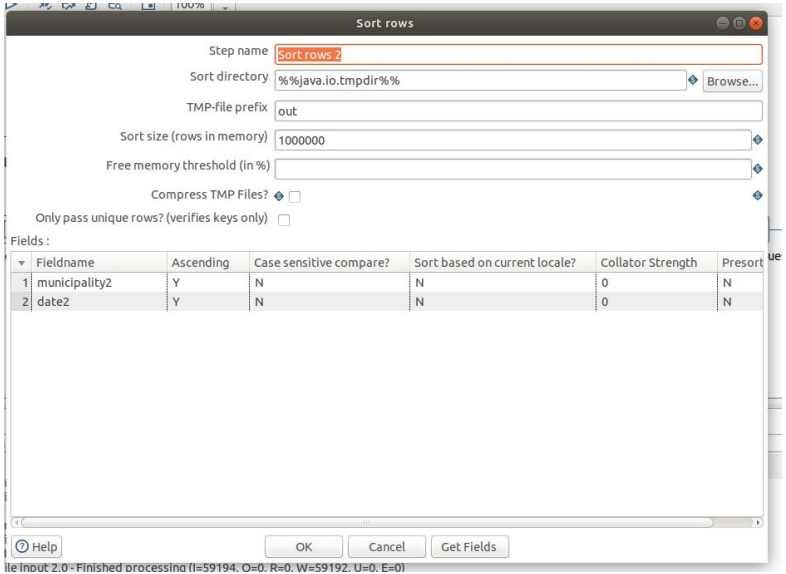




Sort rows 2

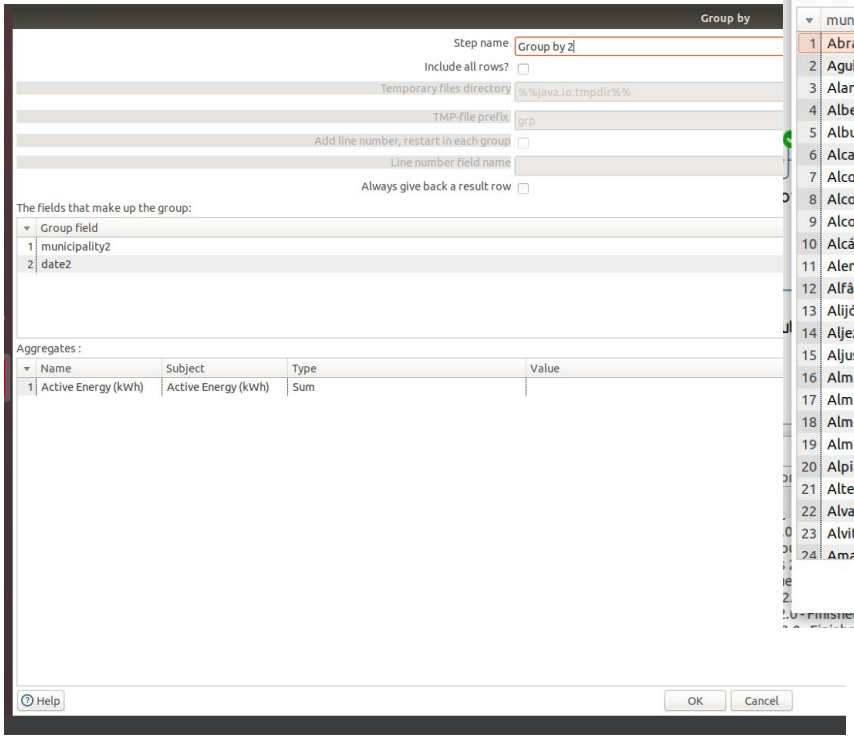


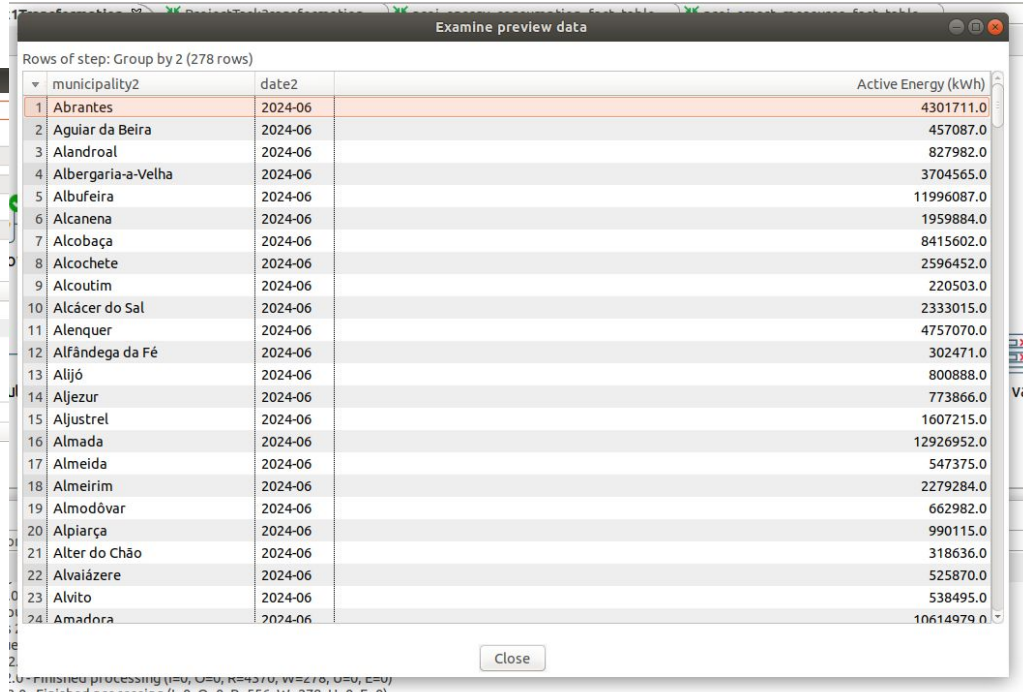
	date2	municipality2	Active Energy (kWh)
1	2024-06	Abrantes	17700
2	2024-06	Abrantes	752674
3	2024-06	Abrantes	22276
4	2024-06	Abrantes	35775
5	2024-06	Abrantes	46232
6	2024-06	Abrantes	74113
7	2024-06	Abrantes	98297
8	2024-06	Abrantes	31292
9	2024-06	Abrantes	179525
10	2024-06	Abrantes	73489
11	2024-06	Abrantes	35760
12	2024-06	Abrantes	26813
13	2024-06	Abrantes	252392
14	2024-06	Abrantes	53171
15	2024-06	Abrantes	1080350
16	2024-06	Abrantes	101749
17	2024-06	Abrantes	258031
18	2024-06	Abrantes	120885
19	2024-06	Abrantes	145601
20	2024-06	Abrantes	852859
21	2024-06	Abrantes	42727
22	2024-06	Aguiar da Beira	32570
23	2024-06	Aguiar da Beira	19536
24	2024-06	Aguiar da Beira	38038



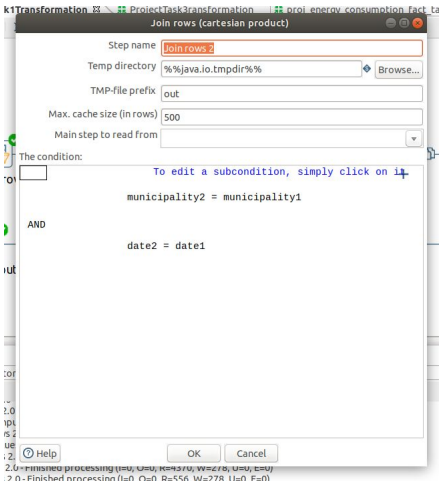
	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presort
1	municipality2	Y	N	N	0	N
2	date2	Y	N	N	0	N

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 municipality2
2 date2
Aggregates :
Name
Subject
Type
Value
1 Active Energy (kWh)
Active Energy (kWh)
Sum
Help
OK
Cancel


Examine preview data
Rows of step: Group by 2 (278 rows)
municipality2
date2
Active Energy (kWh)
1 Abrantes 2024-06 4301711.0
2 Aguiar da Beira 2024-06 457087.0
3 Alandroal 2024-06 827982.0
4 Albergaria-a-Velha 2024-06 3704565.0
5 Albufeira 2024-06 11996087.0
6 Alcanena 2024-06 1959884.0
7 Alcobaça 2024-06 8415602.0
8 Alcochete 2024-06 2596452.0
9 Alcoutim 2024-06 220503.0
10 Alcácer do Sal 2024-06 2333015.0
11 Alenquer 2024-06 4757070.0
12 Alfândega da Fé 2024-06 302471.0
13 Alijó 2024-06 800888.0
14 Aljezur 2024-06 773866.0
15 Aljustrel 2024-06 1607215.0
16 Almada 2024-06 12926952.0
17 Almeida 2024-06 547375.0
18 Almeirim 2024-06 2279284.0
19 Almodôvar 2024-06 662982.0
20 Alpiarça 2024-06 990115.0
21 Alter do Chão 2024-06 318636.0
22 Alvaiázere 2024-06 525870.0
23 Alvito 2024-06 538495.0
24 Amadora 2024-06 10614979.0
Close

Join rows 2



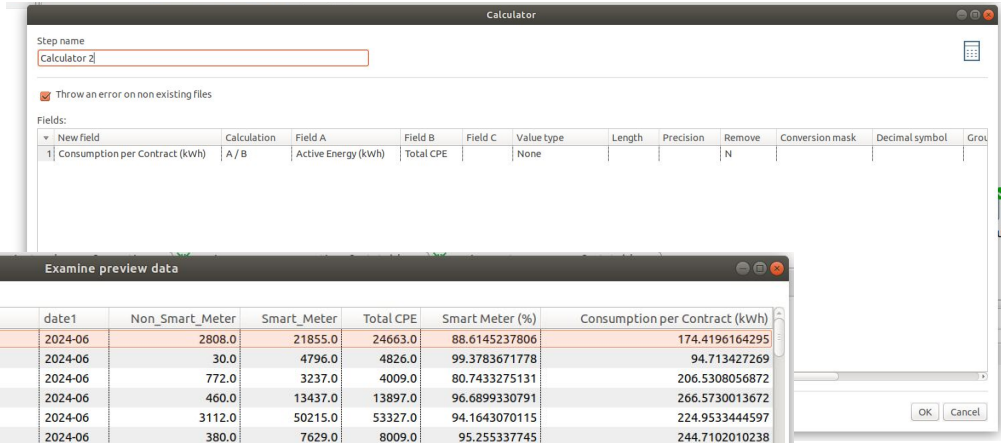
Examine preview data

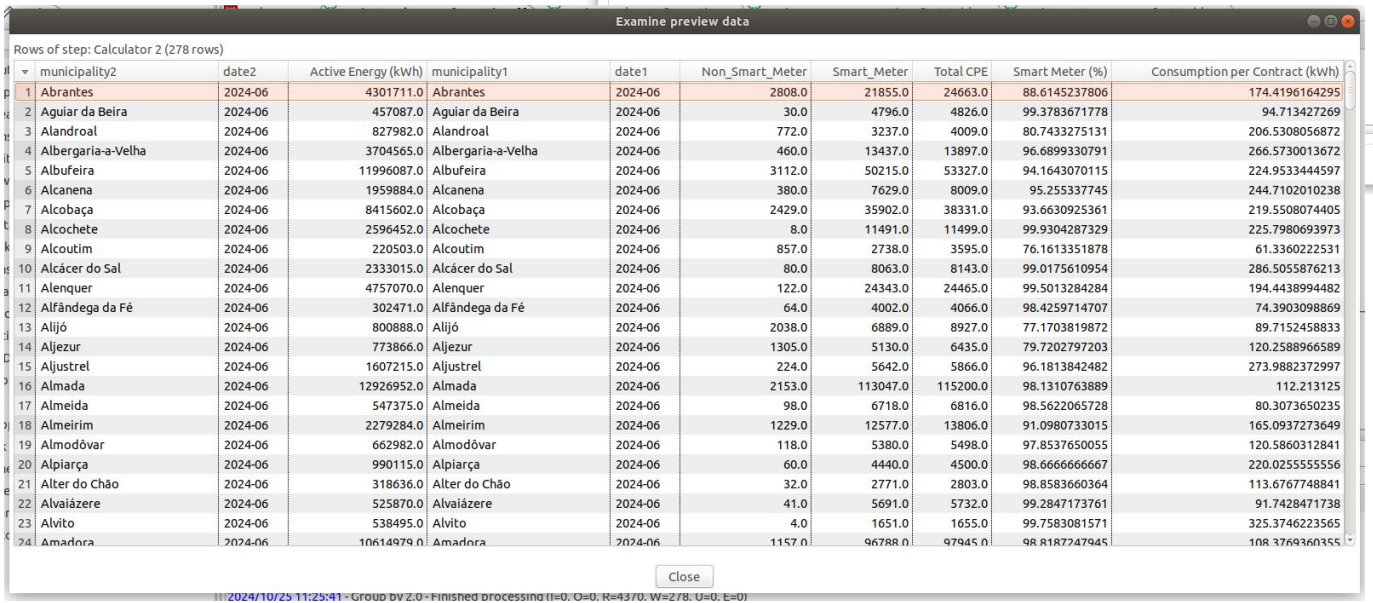
Rows of step: Join rows 2 (278 rows)

	municipality2	date2	Active Energy (kWh)	municipality1	date1	Non_Smart_Meter	Smart_Meter	Total CPE	Smart Meter (%)
1	Abrantes	2024-06	4301711.0	Abrantes	2024-06	2808.0	21855.0	24663.0	88.6145237806
2	Aguiar da Beira	2024-06	457087.0	Aguiar da Beira	2024-06	30.0	4796.0	4826.0	99.3783671778
3	Alandroal	2024-06	827982.0	Alandroal	2024-06	772.0	3237.0	4009.0	80.7433275131
4	Albergaria-a-Velha	2024-06	3704565.0	Albergaria-a-Velha	2024-06	460.0	13437.0	13897.0	96.6899330791
5	Albufeira	2024-06	11996087.0	Albufeira	2024-06	3112.0	50215.0	53327.0	94.1643070115
6	Alcanena	2024-06	1959884.0	Alcanena	2024-06	380.0	7629.0	8009.0	95.255337745
7	Alcobaça	2024-06	8415602.0	Alcobaça	2024-06	2429.0	35902.0	38331.0	93.6630925361
8	Alcochete	2024-06	2596452.0	Alcochete	2024-06	8.0	11491.0	11499.0	99.9304287329
9	Alcoutim	2024-06	220503.0	Alcoutim	2024-06	857.0	2738.0	3595.0	76.1613351878
10	Alcácer do Sal	2024-06	2333015.0	Alcácer do Sal	2024-06	80.0	8063.0	8143.0	99.0175610954
11	Alenquer	2024-06	4757070.0	Alenquer	2024-06	122.0	24343.0	24465.0	99.5013284284
12	Alfândega da Fé	2024-06	302471.0	Alfândega da Fé	2024-06	64.0	4002.0	4066.0	98.4259714707
13	Alijó	2024-06	800888.0	Alijó	2024-06	2038.0	6889.0	8927.0	77.1703819872
14	Aljezur	2024-06	773866.0	Aljezur	2024-06	1305.0	5130.0	6435.0	79.7202797203
15	Aljustrel	2024-06	1607215.0	Aljustrel	2024-06	224.0	5642.0	5866.0	96.1813842482
16	Almada	2024-06	12926952.0	Almada	2024-06	2153.0	113047.0	115200.0	98.1310763889
17	Almeida	2024-06	547375.0	Almeida	2024-06	98.0	6718.0	6816.0	98.5622065728
18	Almeirim	2024-06	2279284.0	Almeirim	2024-06	1229.0	12577.0	13806.0	91.0980733015
19	Almodôvar	2024-06	662982.0	Almodôvar	2024-06	118.0	5380.0	5498.0	97.8537650055
20	Alpiarça	2024-06	990115.0	Alpiarça	2024-06	60.0	4440.0	4500.0	98.6666666667
21	Alter do Chão	2024-06	318636.0	Alter do Chão	2024-06	32.0	2771.0	2803.0	98.8583660364
22	Alvaiázere	2024-06	525870.0	Alvaiázere	2024-06	41.0	5691.0	5732.0	99.2847173761
23	Alvito	2024-06	538495.0	Alvito	2024-06	4.0	1651.0	1655.0	99.7583081571
24	Amadora	2024-06	10614979.0	Amadora	2024-06	1157.0	96788.0	97945.0	98.8187247945

Close

Calculator 2

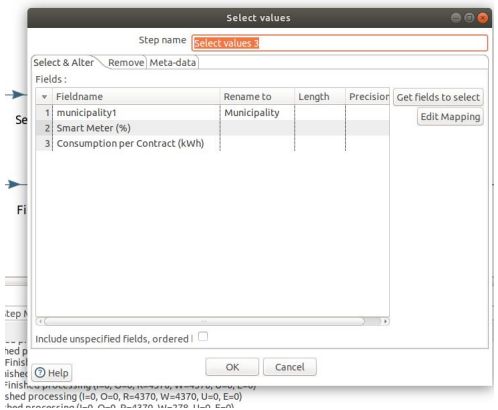


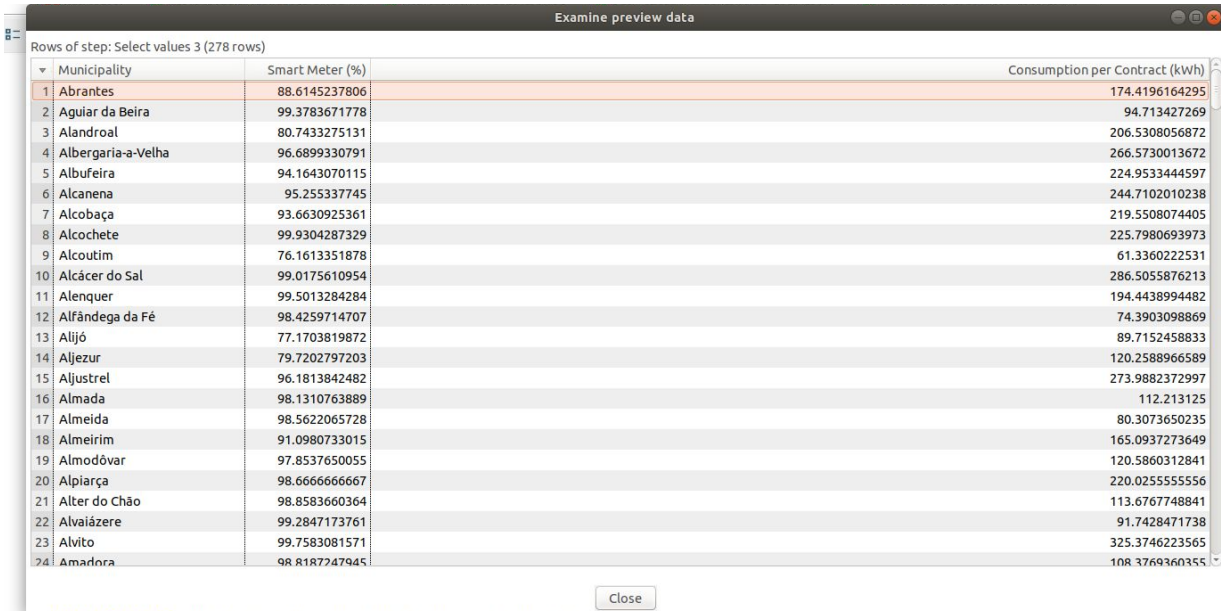


	municipality2	date2	Active Energy (kWh)	municipality1	date1	Non_Smart_Meter	Smart_Meter	Total CPE	Smart Meter (%)	Consumption per Contract (kWh)
1	Abrantes	2024-06	4301711.0	Abrantes	2024-06	2808.0	21855.0	24663.0	88.6145237806	174.4196164295
2	Aguiar da Beira	2024-06	457087.0	Aguiar da Beira	2024-06	30.0	4796.0	4826.0	99.3783671778	94.713427269
3	Alandroal	2024-06	827982.0	Alandroal	2024-06	772.0	3237.0	4009.0	80.7433275131	206.5308056872
4	Albergaria-a-Velha	2024-06	3704565.0	Albergaria-a-Velha	2024-06	460.0	13437.0	13897.0	96.6899330791	266.5730013672
5	Albufeira	2024-06	11996087.0	Albufeira	2024-06	3112.0	50215.0	53327.0	94.1643070115	224.9533444597
6	Alcanena	2024-06	1959884.0	Alcanena	2024-06	380.0	7629.0	8009.0	95.255337745	244.7102010238
7	Alcobaça	2024-06	8415602.0	Alcobaça	2024-06	2429.0	35902.0	38331.0	93.6630925361	219.5508074405
8	Alcochete	2024-06	2596452.0	Alcochete	2024-06	8.0	11491.0	11499.0	99.9304287329	225.7980693973
9	Alcoutim	2024-06	220503.0	Alcoutim	2024-06	857.0	2738.0	3595.0	76.1613351878	61.3360222531
10	Alcácer do Sal	2024-06	2333015.0	Alcácer do Sal	2024-06	80.0	8063.0	8143.0	99.0175610954	286.5055876213
11	Alenquer	2024-06	4757070.0	Alenquer	2024-06	122.0	24343.0	24465.0	99.5013284284	194.4438994482
12	Alfândega da Fé	2024-06	302471.0	Alfândega da Fé	2024-06	64.0	4002.0	4066.0	98.4259714707	74.3903098869
13	Alijó	2024-06	800888.0	Alijó	2024-06	2038.0	6889.0	8927.0	77.1703819872	89.7152458833
14	Aljezur	2024-06	773866.0	Aljezur	2024-06	1305.0	5130.0	6435.0	79.7202797203	120.2588966589
15	Aljustrel	2024-06	1607215.0	Aljustrel	2024-06	224.0	5642.0	5866.0	96.1813842482	273.9882372997
16	Almada	2024-06	12926952.0	Almada	2024-06	2153.0	113047.0	115200.0	98.1310763889	112.213125
17	Almeida	2024-06	547375.0	Almeida	2024-06	98.0	6718.0	6816.0	98.5622065728	80.3073650235
18	Almeirim	2024-06	2279284.0	Almeirim	2024-06	1229.0	12577.0	13806.0	91.0980733015	165.0937273649
19	Almodôvar	2024-06	662982.0	Almodôvar	2024-06	118.0	5380.0	5498.0	97.8537650055	120.5860312841
20	Alpiarça	2024-06	990115.0	Alpiarça	2024-06	60.0	4440.0	4500.0	98.6666666667	220.0255555556
21	Alter do Chão	2024-06	318636.0	Alter do Chão	2024-06	32.0	2771.0	2803.0	98.8583660364	113.6767748841
22	Alvaiázere	2024-06	525870.0	Alvaiázere	2024-06	41.0	5691.0	5732.0	99.2847173761	91.7428471738
23	Alvito	2024-06	538495.0	Alvito	2024-06	4.0	1651.0	1655.0	99.7583081571	325.3746223565
24	Amadora	2024-06	10614979.0	Amadora	2024-06	1157.0	96788.0	97945.0	98.8187247945	108.3769360355

2024/10/25 11:25:41 - Group by 2.0 - Finished processing (I=0, O=0, R=4370, W=278, U=0, E=0)

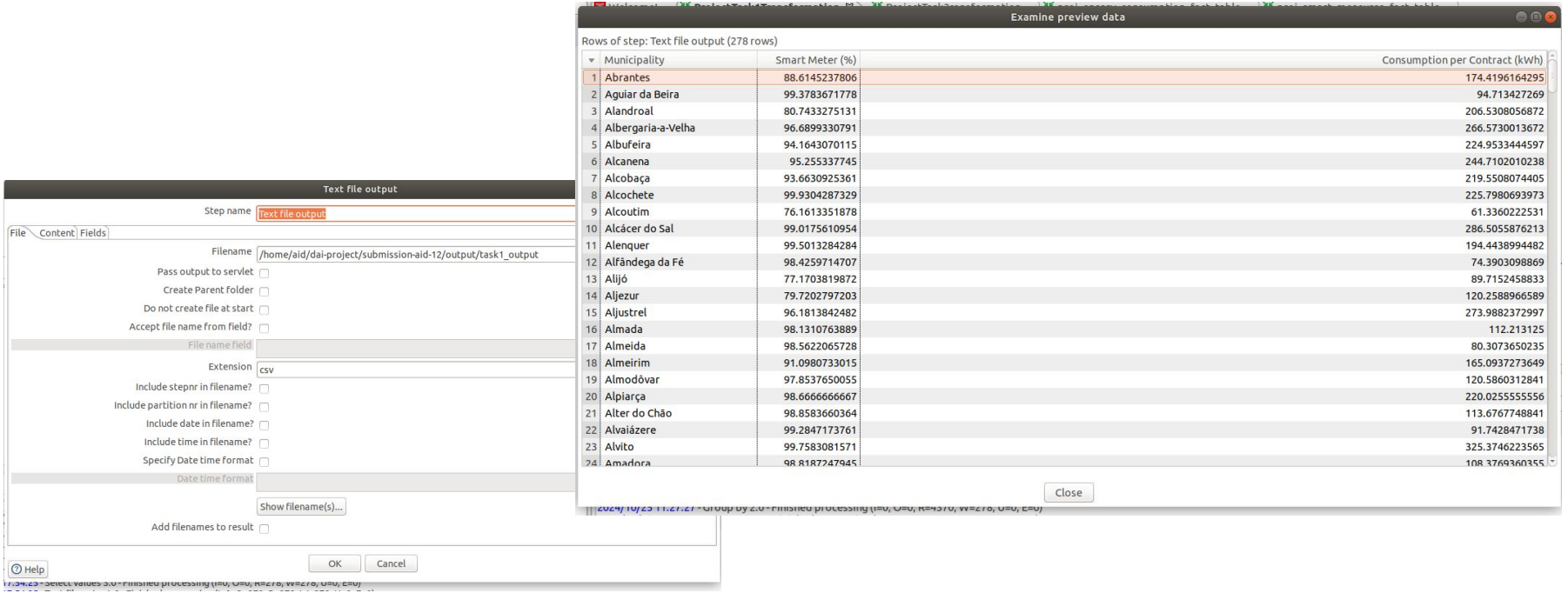
Select values 3



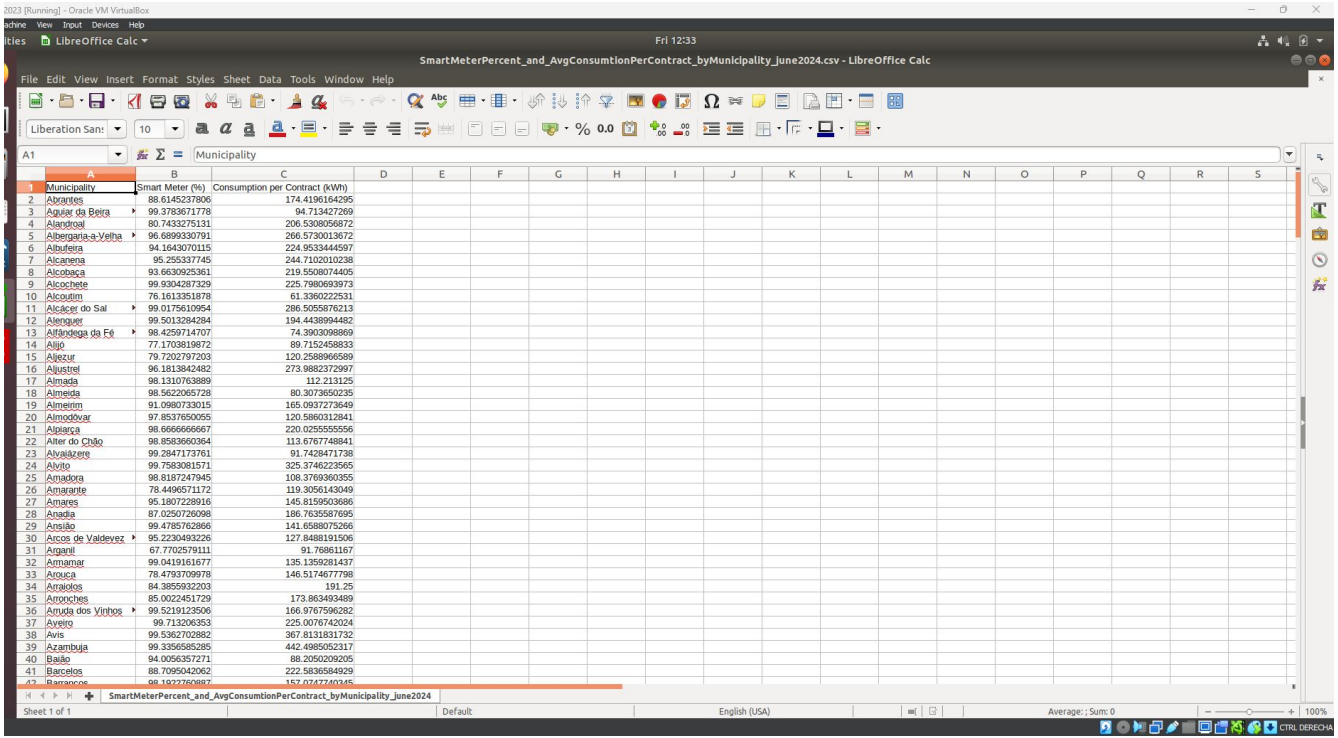


	Municipality	Smart Meter (%)	Consumption per Contract (kWh)
1	Abrantes	88.6145237806	174.4196164295
2	Aguiar da Beira	99.3783671778	94.713427269
3	Alandroal	80.7433275131	206.5308056872
4	Albergaria-a-Velha	96.6899330791	266.5730013672
5	Albufeira	94.1643070115	224.9533444597
6	Alcanena	95.255337745	244.7102010238
7	Alcobaça	93.6630925361	219.5508074405
8	Alcochete	99.9304287329	225.7980693973
9	Alcoutim	76.1613351878	61.3360222531
10	Alcácer do Sal	99.0175610954	286.5055876213
11	Alenquer	99.5013284284	194.4438994482
12	Alfândega da Fé	98.4259714707	74.3903098869
13	Alijó	77.1703819872	89.7152458833
14	Aljezur	79.7202797203	120.2588966589
15	Aljustrel	96.1813842482	273.9882372997
16	Almada	98.1310763889	112.213125
17	Almeida	98.5622065728	80.3073650235
18	Almeirim	91.0980733015	165.0937273649
19	Almodôvar	97.8537650055	120.5860312841
20	Alpiarça	98.6666666667	220.0255555556
21	Alter do Chão	98.8583660364	113.6767748841
22	Alvaiázere	99.2847173761	91.7428417138
23	Alvito	99.7583081571	325.3746223565
24	Amadora	98.8187247945	108.3769360355

Text file output



Output File

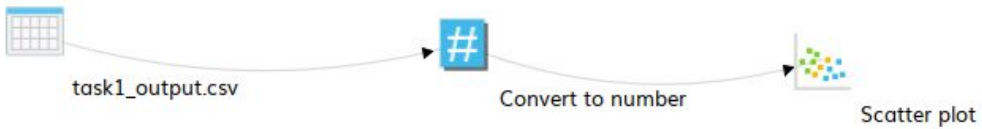


	A	B	C
1	Municipality	Smart Meter (%)	Consumption per Contract (kWh)
2	Abrantes	88.6145237806	174.4196164295
3	Aguiar da Beira	99.3783671778	94.713427269
4	Alandroal	80.7433275131	206.5308056872
5	Albergaria-a-Velha	96.6899330791	266.5730013672
6	Albufeira	94.1643070115	224.9533444597
7	Alcanena	95.255337745	244.7102010238
8	Alcobaça	93.6630925361	219.5508074405
9	Alcochete	99.9304287329	225.7980693973
10	Alcoutim	76.1613351878	61.3360222531
11	Alcácer do Sal	99.0175610954	286.5055876213
12	Alenquer	99.5013284284	194.4438994482
13	Alfândega da Fé	98.4259714707	74.3903098869
14	Alijó	77.1703819872	89.7152458833
15	Aljezur	79.7202797203	120.2588966589
16	Aljustrel	96.1813842482	273.9882372997
17	Almada	98.1310763889	112.213125
18	Almeida	98.5622065728	80.3073650235
19	Almeirim	91.0980733015	165.0937273649
20	Almodôvar	97.8537650055	120.5860312841
21	Alpiarça	98.6666666667	220.0255555556
22	Alter do Chão	98.8583660364	113.6767748841
23	Alvaiázere	99.2847173761	91.7428471738
24	Alvito	99.7583081571	325.3746223565
25	Amadora	98.8187247945	108.3769360355
26	Amarante	78.4496571172	119.3056143049
27	Amares	95.1807228916	145.8159503686
28	Anadia	87.0250726098	186.7635587695
29	Ansião	99.4785762866	141.6588075266
30	Arcos de Valdevez	95.2230493226	127.8488191506
31	Arganil	67.7702579111	91.76861167
32	Armamar	99.0419161677	135.1359281437
33	Arouca	78.4793709978	146.5174677798
34	Arraiolos	84.3855932203	191.25
35	Arronches	85.0022451729	173.863493489
36	Arruda dos Vinhos	99.5219123506	166.9767596282
37	Aveiro	99.713206353	225.0076742024
38	Avis	99.5362702882	367.8131831732
39	Azambuja	99.3356585285	442.4985052317
40	Baião	94.0056357271	88.2050209205
41	Barcelos	88.7095042062	222.5836584929
42	Barreiros	98.1922260987	157.0747240345

Task 2

Generate a scatter plot to study the correlation between the percentage of smart meters (%, x-axis) and energy consumption per contract (kWh/CPE, y-axis).

Analysis Job



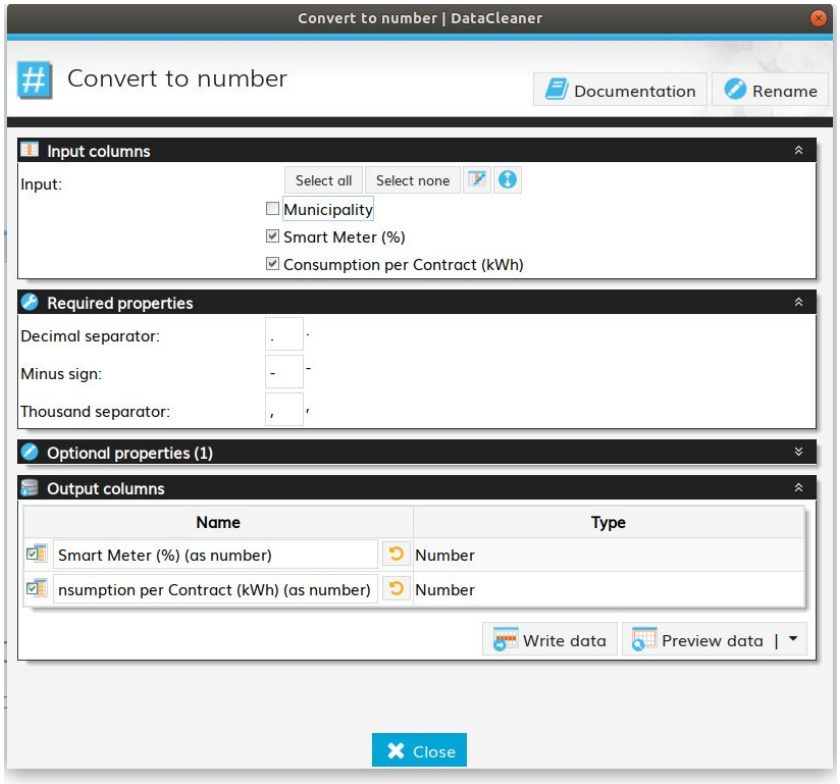
Step Preview Table

DataSet: SELECT task1_output.csv.Municipality, task1_output.csv.Smart Meter (%), task1_output.csv.Consumption per Contract (kWh) FROM out...

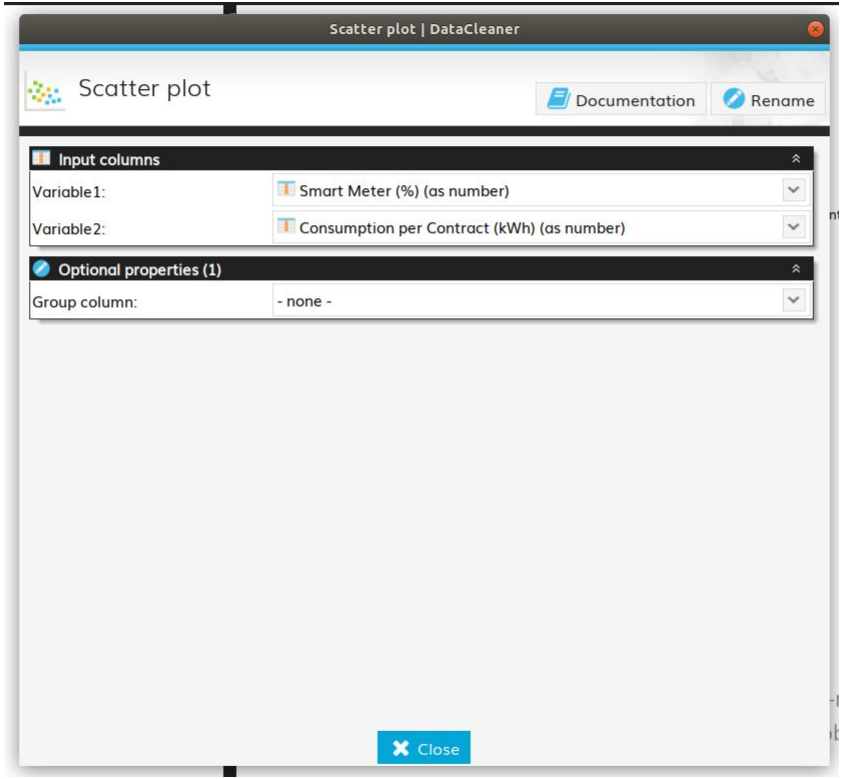
Municipality	Smart Meter (%)	Consumption per Contract (kWh)
Abrantes	88.6145237806	174.4196164295
Aguiar da Beira	99.3783671778	94.713427269
Alandroal	80.7433275131	206.5308056872
Albergaria-a-Velha	96.6899330791	266.5730013672
Albufeira	94.1643070115	224.9533444597
Alcanena	95.255337745	244.7102010238
Alcobaça	93.6630925361	219.5508074405
Alcochete	99.9304287329	225.7980693973
Alcoutim	76.1613351878	61.3360222531
Alcácer do Sal	99.0175610954	286.5055876213
Alenquer	99.5013284284	194.4438994482
Alfândega da Fé	98.4259714707	74.3903098869
Alijó	77.1703819872	89.7152458833
Aljezur	79.7202797203	120.2588966589
Aljustrel	96.1813842482	273.9882372997
Almada	98.1310763889	112.213125
Almeida	98.5622065728	80.3073650235
Almeirim	91.0980733015	165.0937273649
Almodôvar	97.8537650055	120.5860312841
Alpiarça	98.6666666667	220.0255555556

Previous page Next page

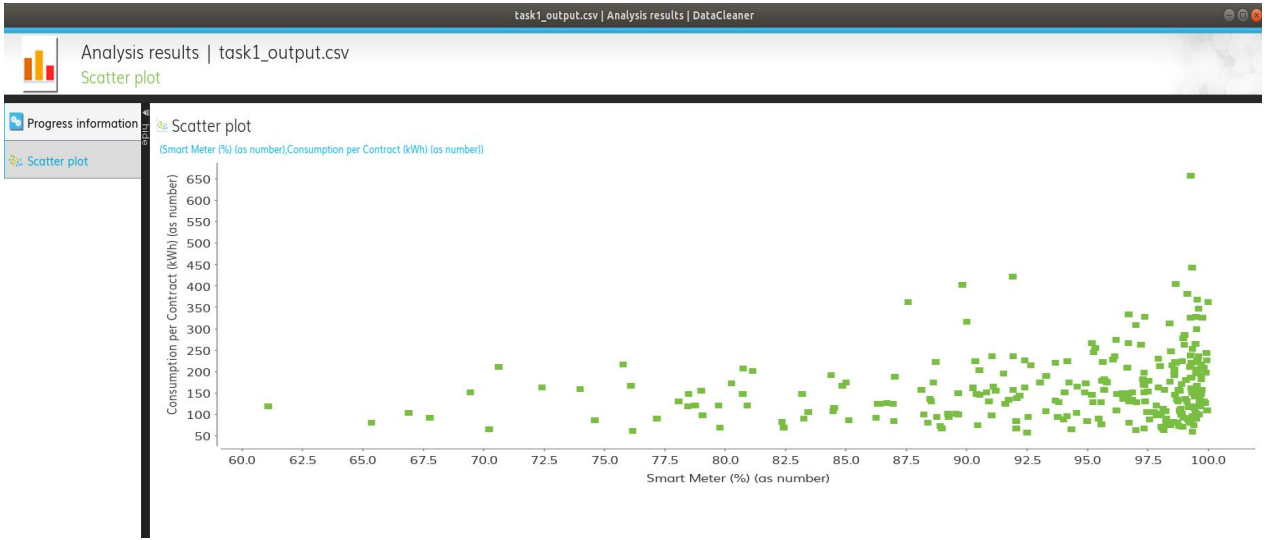
Step Convert to Number



Step Scatter plot



Scatter Plot



Is there a correlation between these variables? If yes, how strong is it?

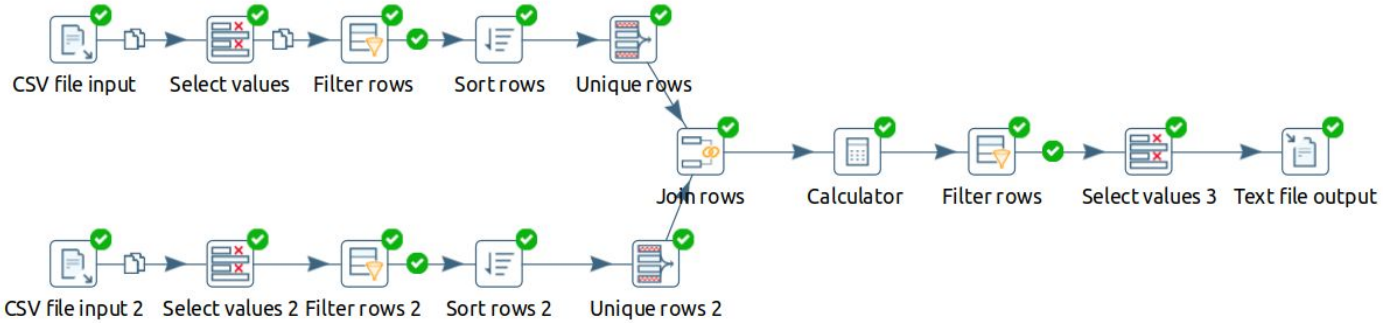
There is no correlation, if there was a correlation we would see a pattern with a slope in the Scatter Plot.

Task 3

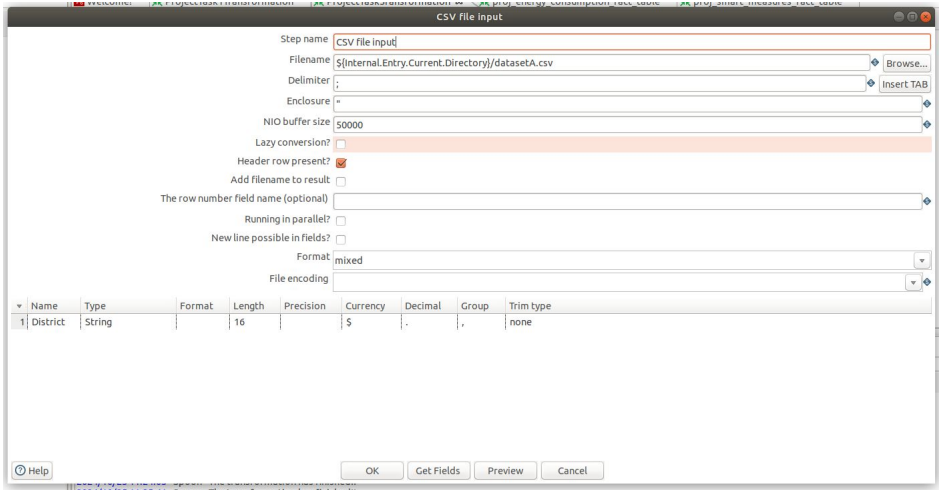
Create a transformation to identify the correct mapping of regions (‘districts’ in Portuguese) in dataset A to regions in dataset B.

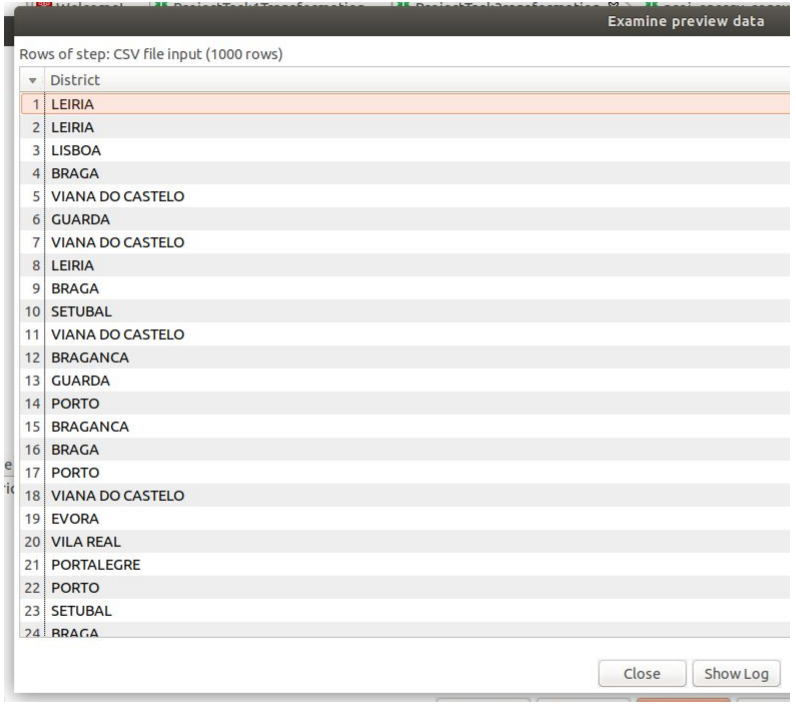
The files have been updated by E-Redes and the names of the districts are now exactly the same in both datasets. We still assumed that the names could be slightly different and that the problem may recur.

Transformation

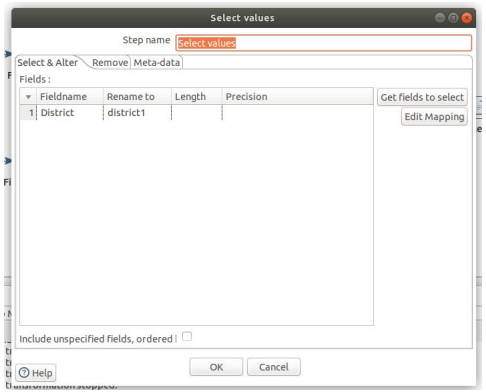


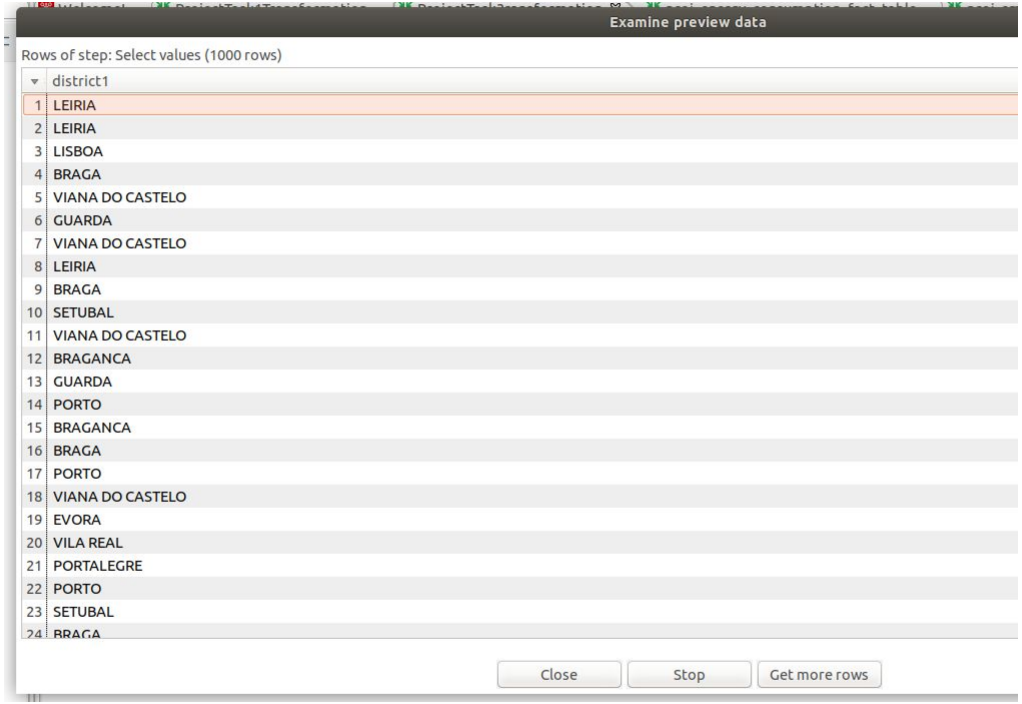
CSV file input





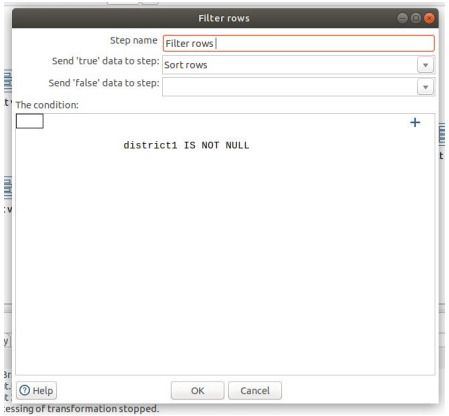
Select values

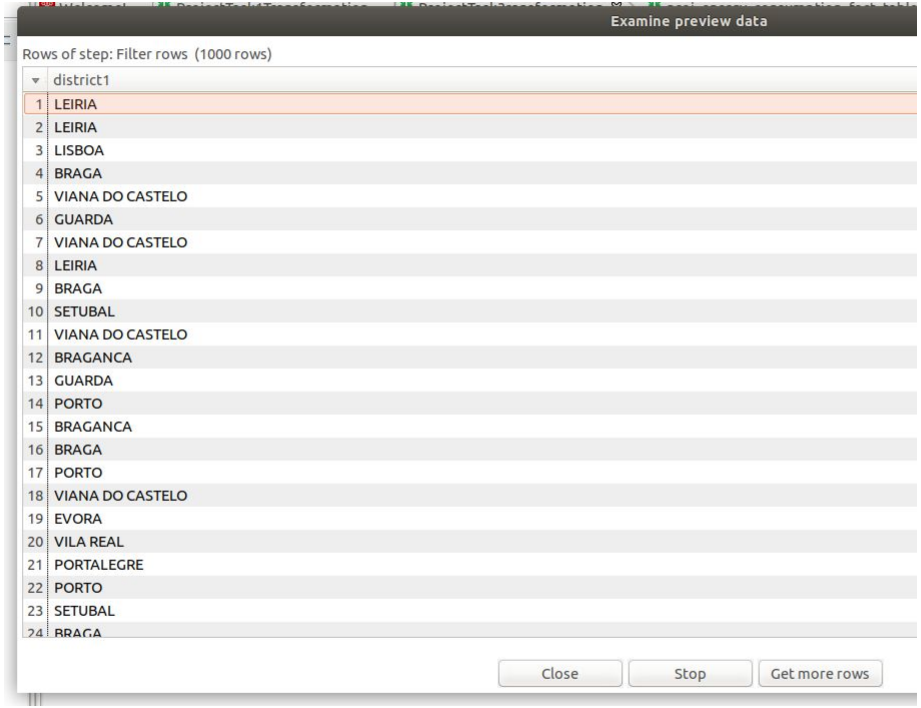




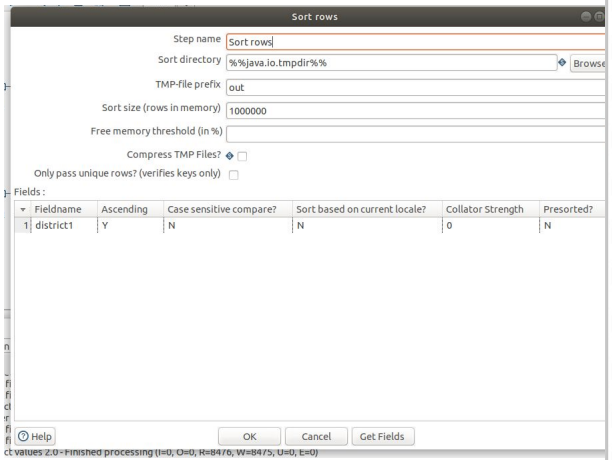
Filter rows

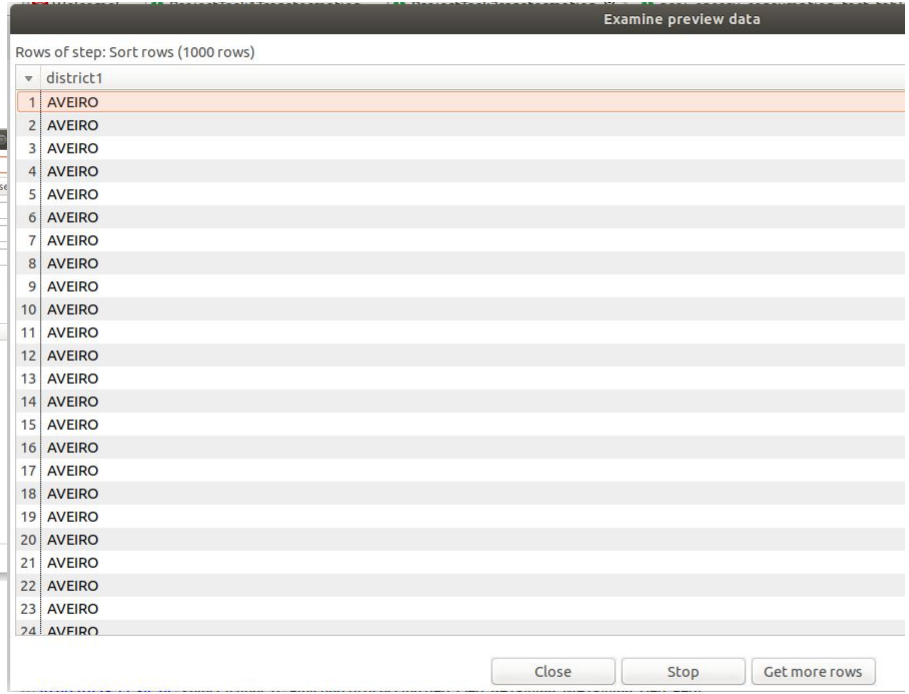
This step ensures validity of data.



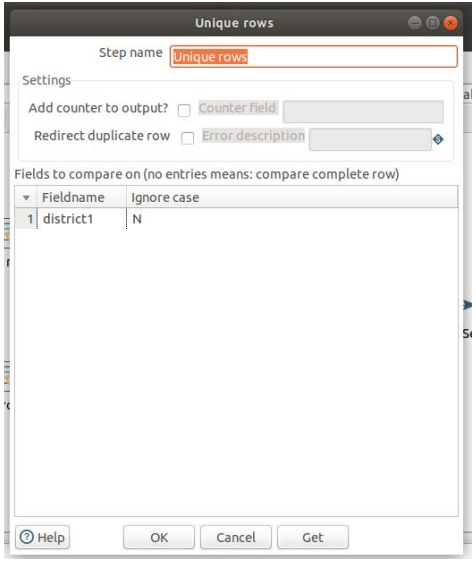


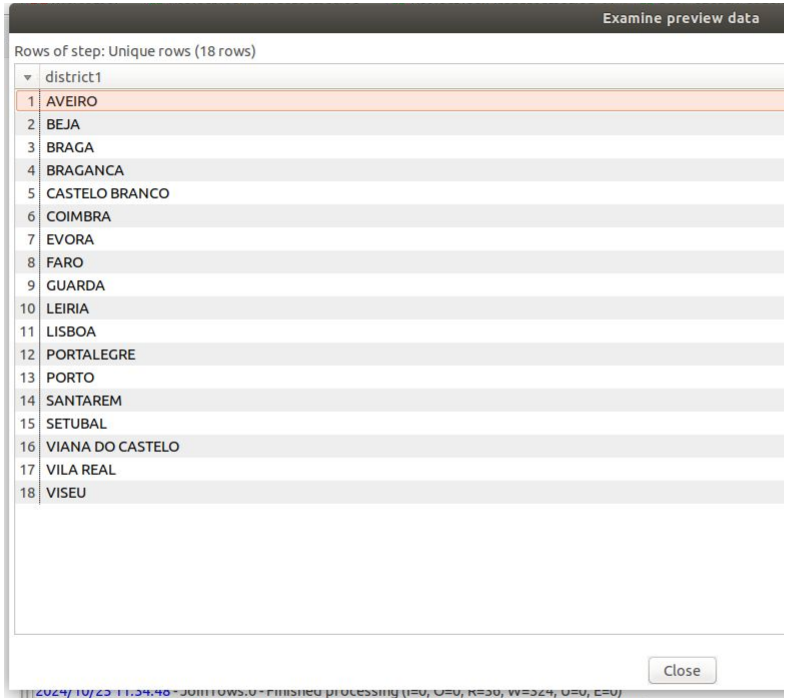
Sort rows


Sort rows
Step name
Sort rows
Sort directory
%%java.io.tmpdir%%
Browse
TMP-file prefix
out
Sort size (rows in memory)
1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys only)
Fields :
Fieldname
Ascending
Case sensitive compare?
Sort based on current locale?
Collator Strength
Presorted?
1
district1
Y
N
N
0
N
Help
OK
Cancel
Get Fields

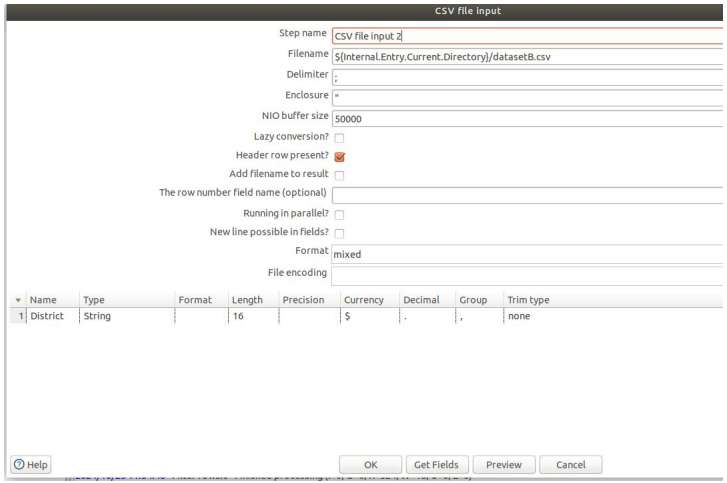

Examine preview data
Rows of step: Sort rows (1000 rows)
district1
1 AVEIRO
2 AVEIRO
3 AVEIRO
4 AVEIRO
5 AVEIRO
6 AVEIRO
7 AVEIRO
8 AVEIRO
9 AVEIRO
10 AVEIRO
11 AVEIRO
12 AVEIRO
13 AVEIRO
14 AVEIRO
15 AVEIRO
16 AVEIRO
17 AVEIRO
18 AVEIRO
19 AVEIRO
20 AVEIRO
21 AVEIRO
22 AVEIRO
23 AVEIRO
24 AVEIRO
Close
Stop
Get more rows

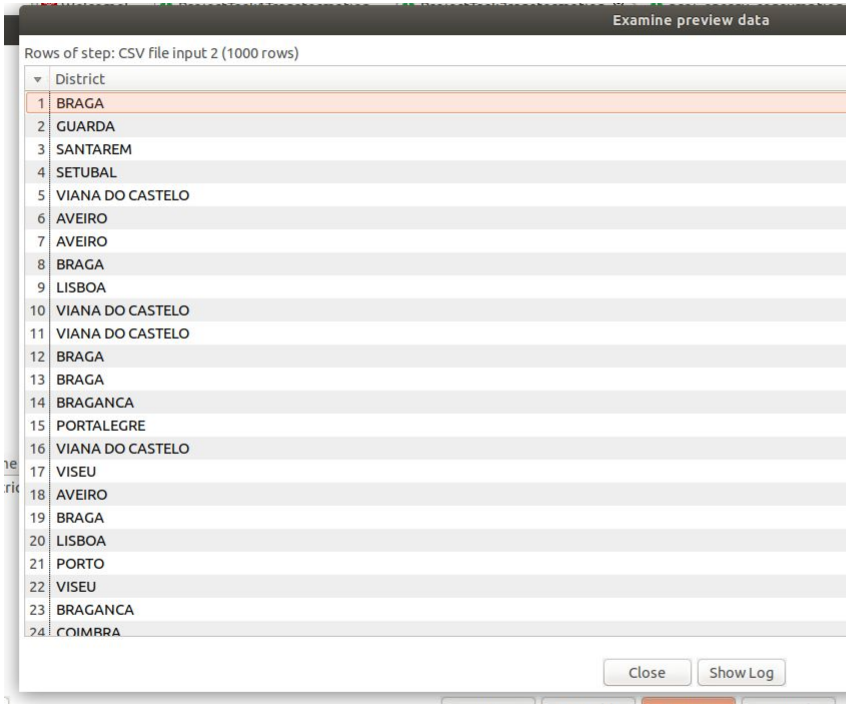
Unique rows


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
district1
N
Help
OK
Cancel
Get

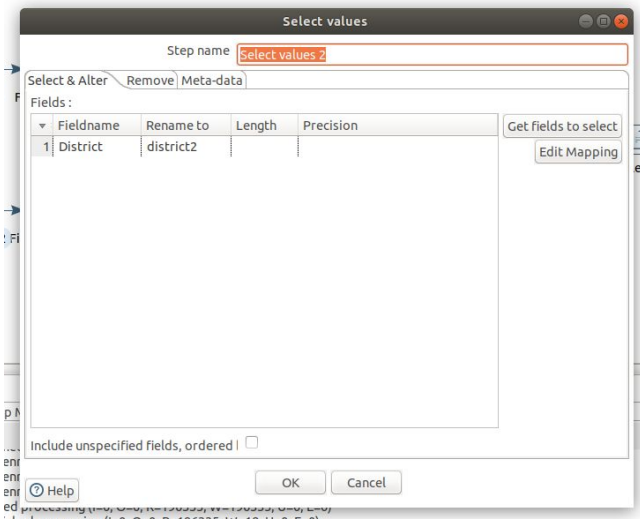

Examine preview data
Rows of step: Unique rows (18 rows)
district1
1 AVEIRO
2 BEJA
3 BRAGA
4 BRAGANCA
5 CASTELO BRANCO
6 COIMBRA
7 EVORA
8 FARO
9 GUARDA
10 LEIRIA
11 LISBOA
12 PORTALEGRE
13 PORTO
14 SANTAREM
15 SETUBAL
16 VIANA DO CASTELO
17 VILA REAL
18 VISEU
Close

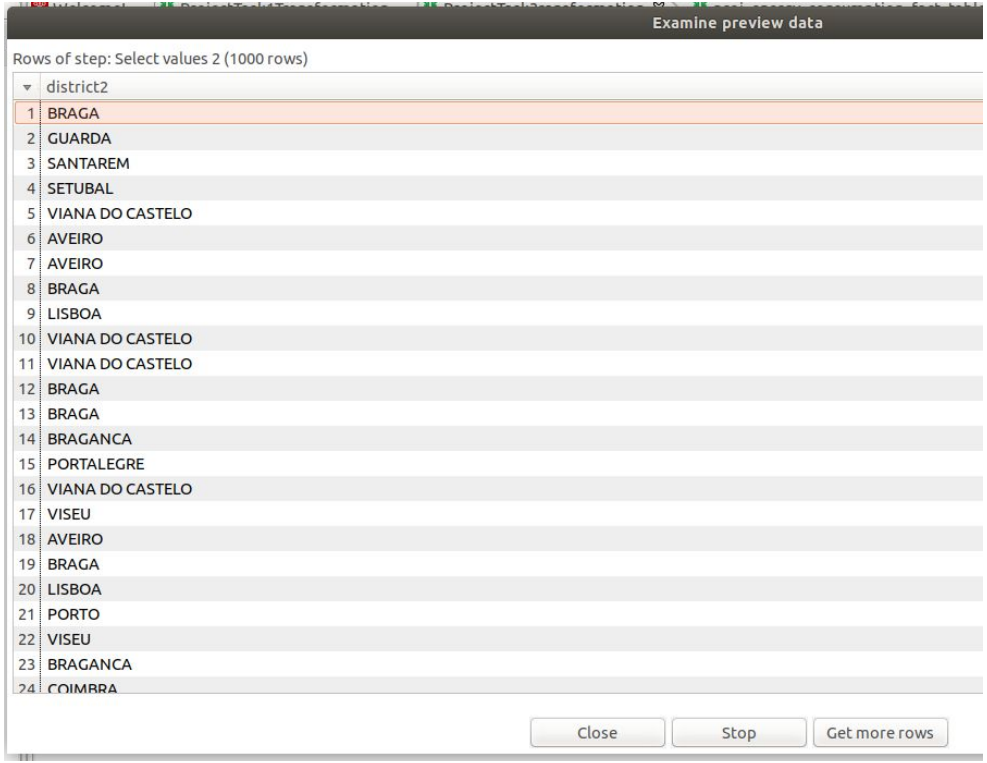
CSV file input 2



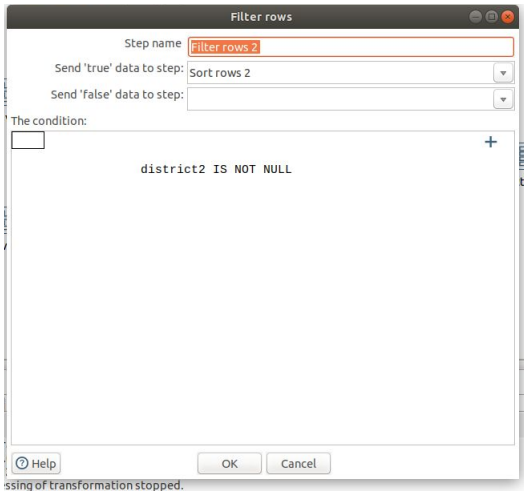


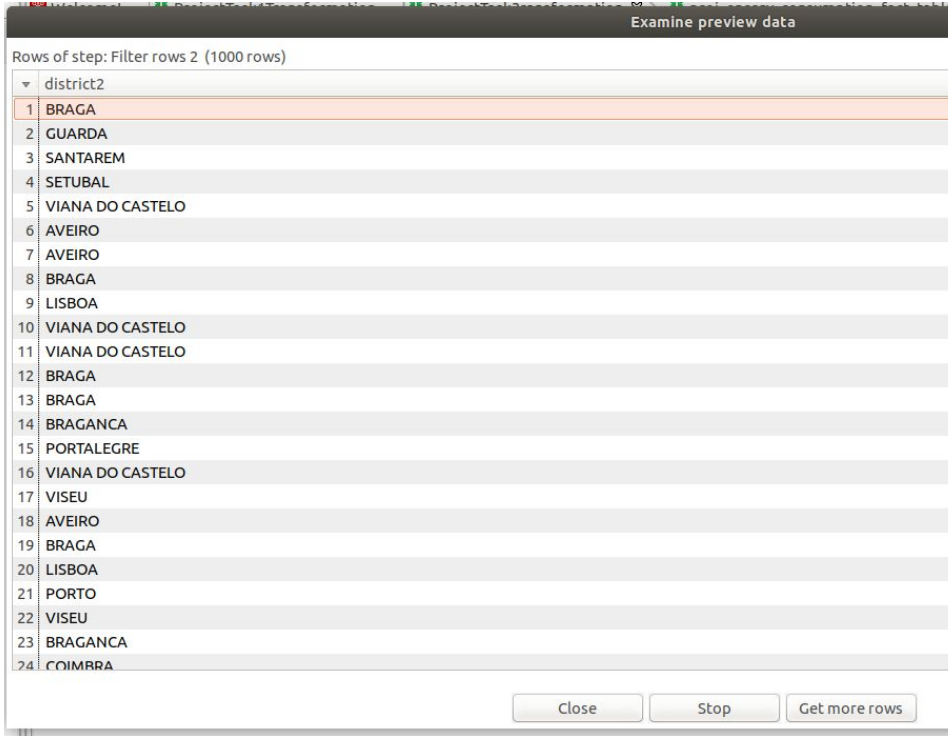
Select values 2



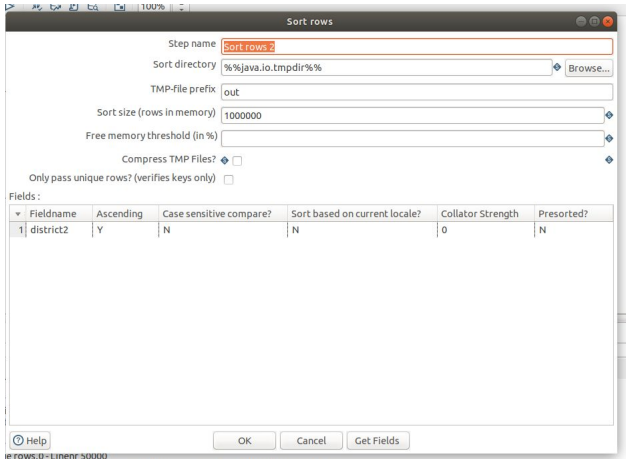


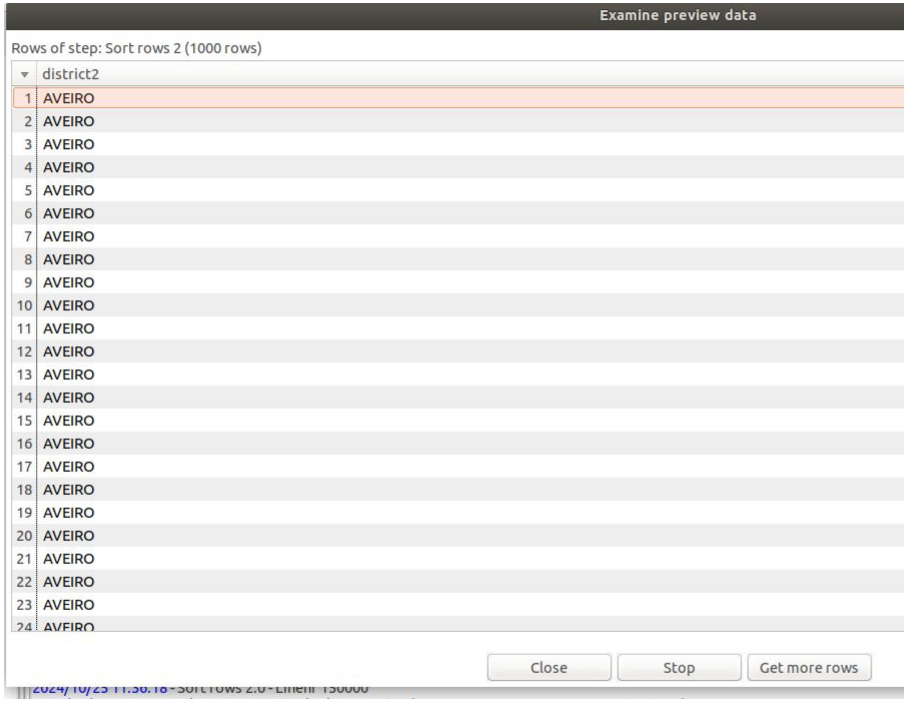
Filter rows 2



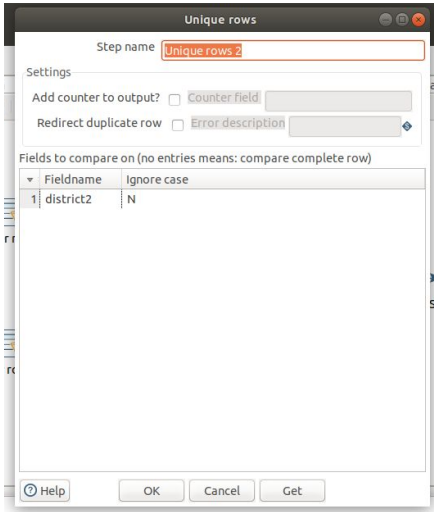


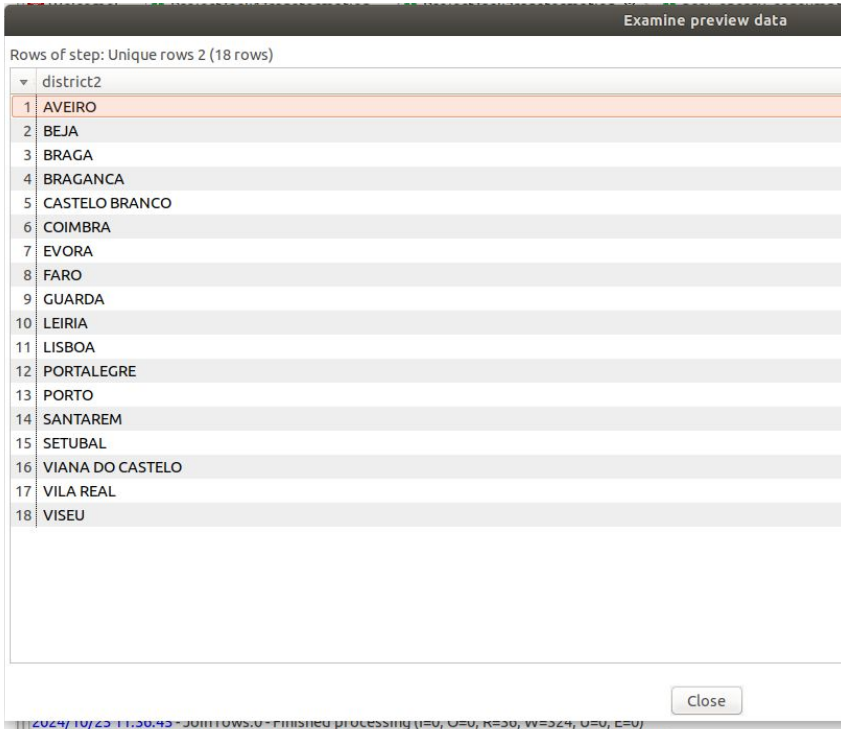
Sort rows 2


Sort rows
Step name
Sort rows 2
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
district2
Y
N
N
0
N
Help
OK
Cancel
Get Fields


Examine preview data
Rows of step: Sort rows 2 (1000 rows)
district2
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
Close
Stop
Get more rows

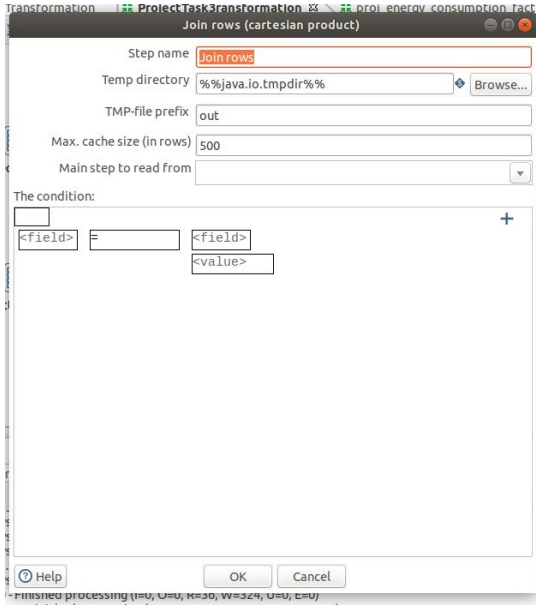
Unique rows 2

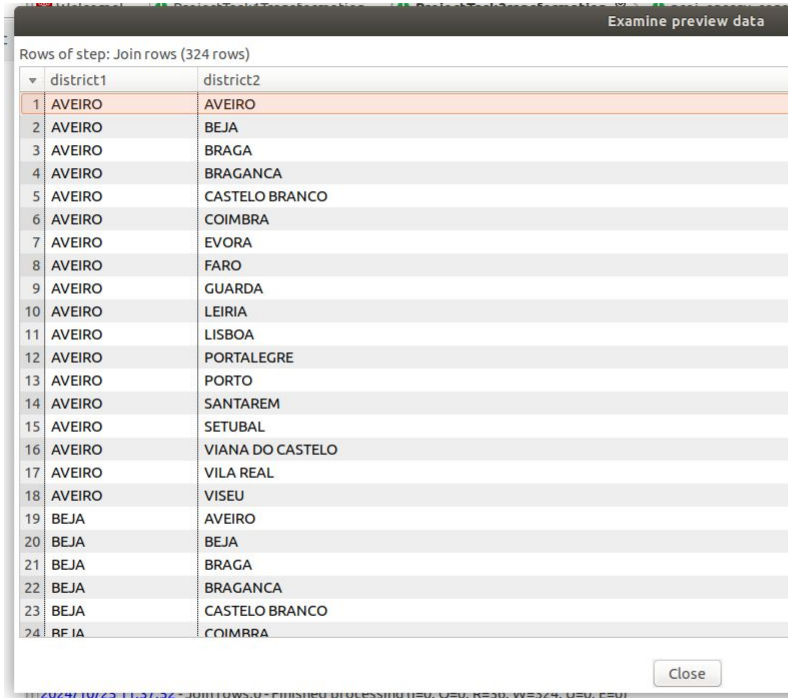




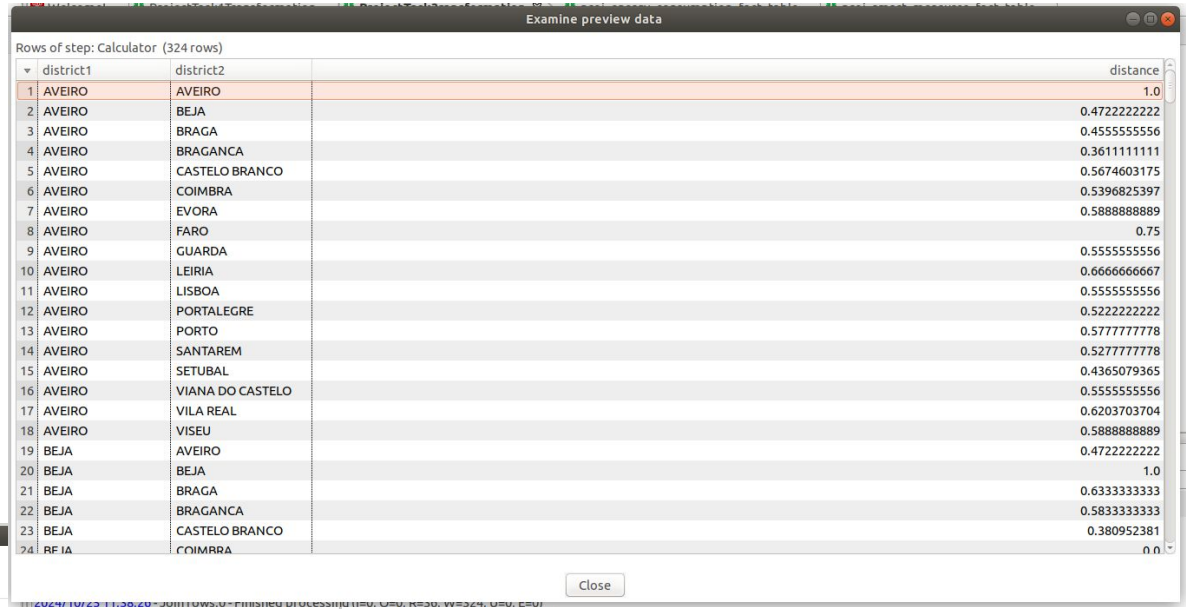
Join rows

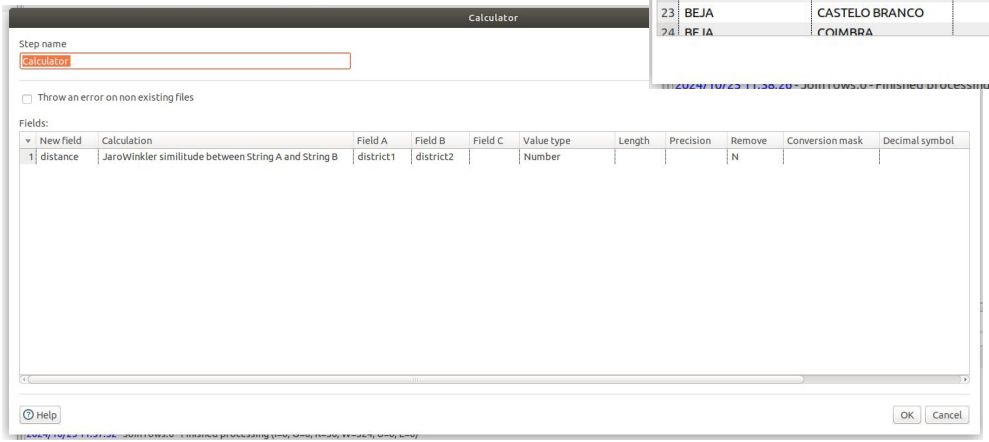
To get all combinations between district in both datasets no conditions were added.





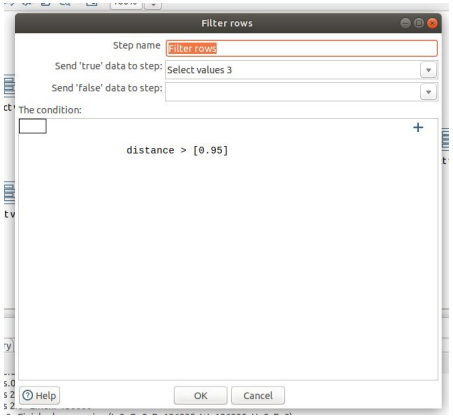
Calculator





Jaro Winkler similitude was used, nonetheless other string matching measures would had similar effect (with the right threshold).

Filter rows



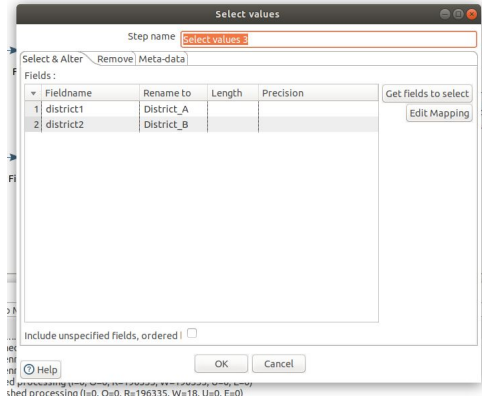
Examine preview data

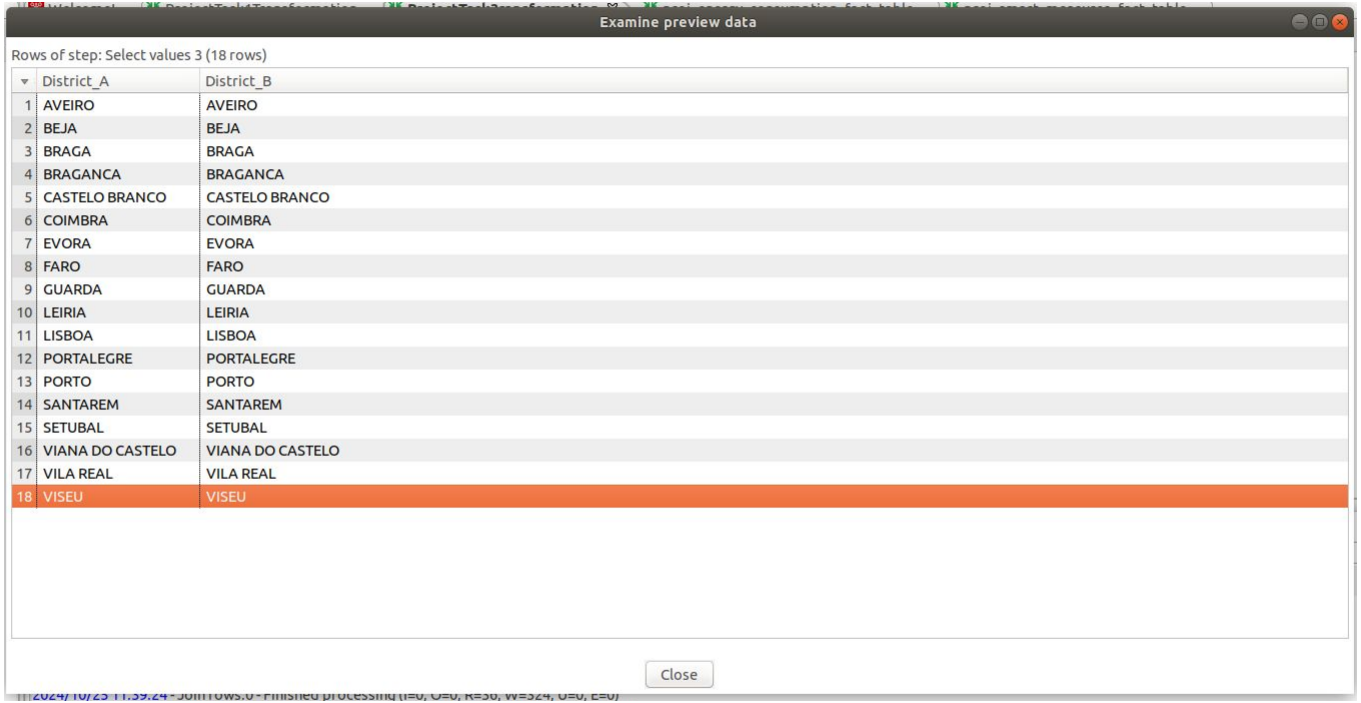
Rows of step: Filter rows (18 rows)

	district1	district2	distance
1	AVEIRO	AVEIRO	1.0
2	BEJA	BEJA	1.0
3	BRAGA	BRAGA	1.0
4	BRAGANCA	BRAGANCA	1.0
5	CASTELO BRANCO	CASTELO BRANCO	1.0
6	COIMBRA	COIMBRA	1.0
7	EVORA	EVORA	1.0
8	FARO	FARO	1.0
9	GUARDA	GUARDA	1.0
10	LEIRIA	LEIRIA	1.0
11	LISBOA	LISBOA	1.0
12	PORTALEGRE	PORTALEGRE	1.0
13	PORTO	PORTO	1.0
14	SANTAREM	SANTAREM	1.0
15	SETUBAL	SETUBAL	1.0
16	VIANA DO CASTELO	VIANA DO CASTELO	1.0
17	VILA REAL	VILA REAL	1.0
18	VISEU	VISEU	1.0

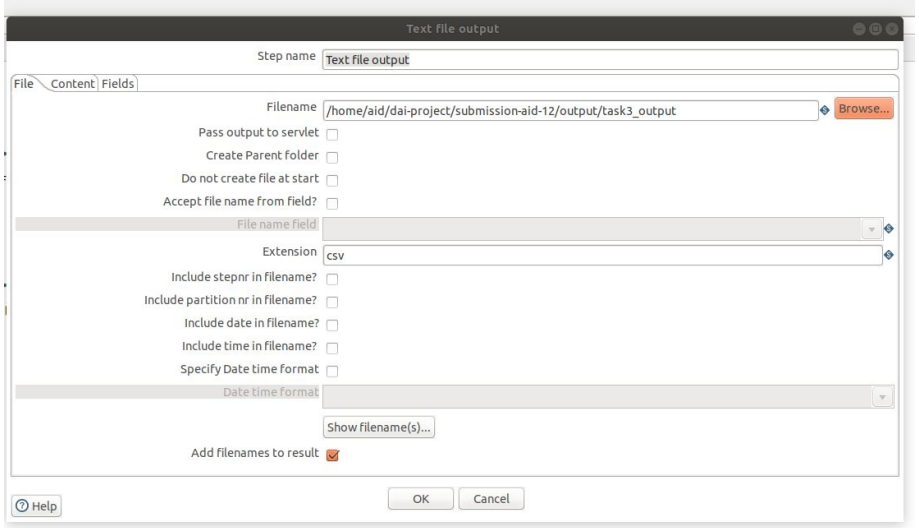
Close

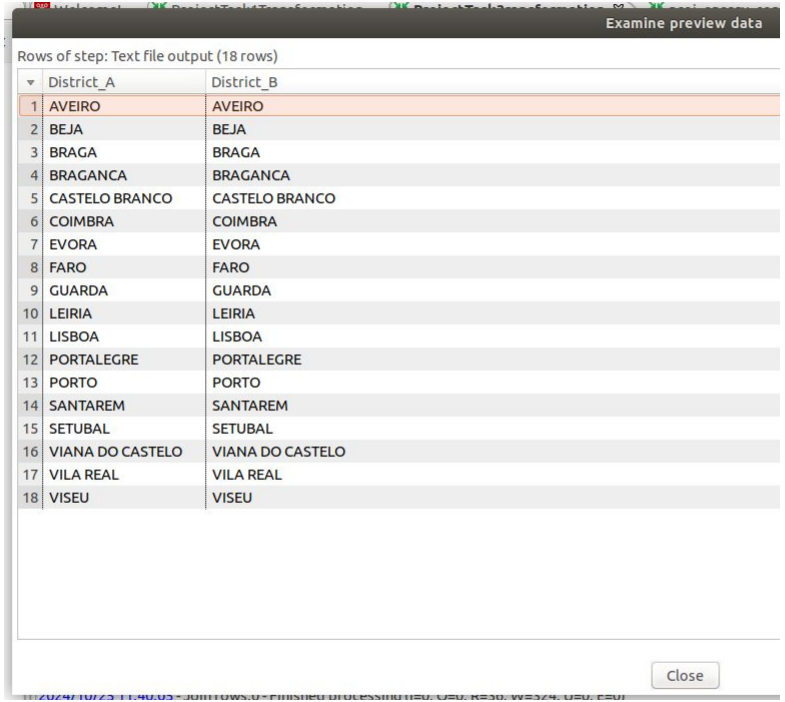
Select values 3





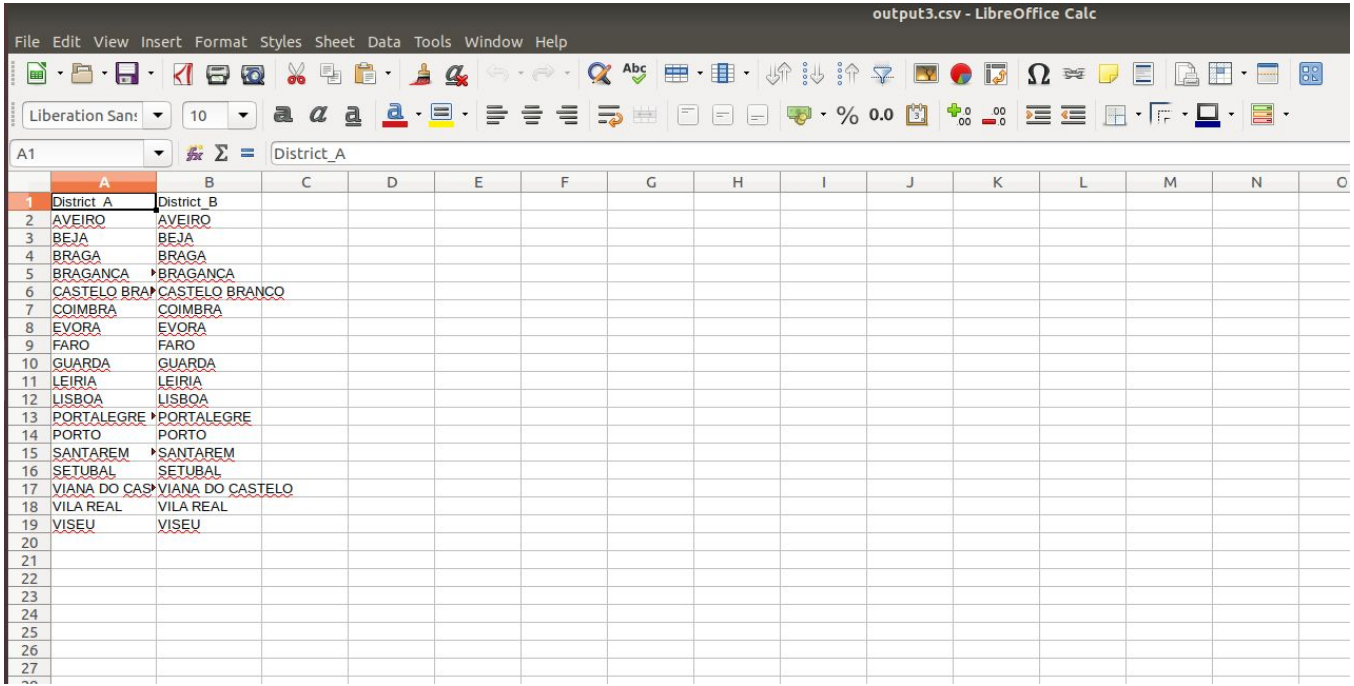
Text file output:





	District_A	District_B
1	AVEIRO	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

Output File



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	District_A	District_B													
2	AVEIRO	AVEIRO													
3	BEJA	BEJA													
4	BRAGA	BRAGA													
5	BRAGANCA	BRAGANCA													
6	CASTELO BRA	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	SANTAREM	SANTAREM													
16	SETUBAL	SETUBAL													
17	VIANA DO CAS	VIANA DO CASTELO													
18	VILA REAL	VILA REAL													
19	VISEU	VISEU													
20															
21															
22															
23															
24															
25															
26															
27															

Task 4

Create an SQL schema creation script to create the tables for a data warehouse.

SQL Script

The time dimension was built and the corresponding procedure called directly in the sql script. This is not only a more pragmatic way than doing it the ETL tool but also ensures chronological integrity, since creating it from datasets could violate that if (in this case) a month was missing.

```sql
DROP DATABASE IF EXISTS energy_dw;

CREATE DATABASE energy_dw;

USE energy_dw;

-- Drop tables if they exist
DROP TABLE IF EXISTS energy_consumption;
DROP TABLE IF EXISTS dim_time;
DROP TABLE IF EXISTS dim_location;

CREATE TABLE dim_time (
    time_id INT PRIMARY KEY,
    year_id INT NOT NULL,
    date_value CHAR(25) NOT NULL,
    season_name VARCHAR(25) NOT NULL,
    season_id INT NOT NULL,
    month_name VARCHAR(25) NOT NULL,
    month_id INT NOT NULL
);

CREATE TABLE dim_location (
    location_id INT PRIMARY KEY,
    district VARCHAR(100) NOT NULL,
    district_code INT NOT NULL,
    municipality VARCHAR(100) NOT NULL,
    municipality_code INT NOT NULL,
    parish VARCHAR(100) NOT NULL,
    parish_code VARCHAR(100) NOT NULL
);

CREATE TABLE fact_energy_consumption (
    time_id INT NOT NULL,
    location_id INT NOT NULL,
    energy_consumption FLOAT NOT NULL,
    PRIMARY KEY (time_id, location_id),
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id) ON DELETE CASCADE ON UPDATE CASCADE,
    FOREIGN KEY (location_id) REFERENCES dim_location(location_id) ON DELETE CASCADE ON UPDATE CASCADE
);

CREATE TABLE fact_smart_measures (
    time_id INT NOT NULL,
    location_id INT NOT NULL,
    smart_meter_qty INT NOT NULL,
    non_smart_meter_qty INT NOT NULL,
    PRIMARY KEY (time_id, location_id),
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id) ON DELETE CASCADE ON UPDATE CASCADE,
    FOREIGN KEY (location_id) REFERENCES dim_location(location_id) ON DELETE CASCADE ON UPDATE CASCADE
);
```

SQL Script

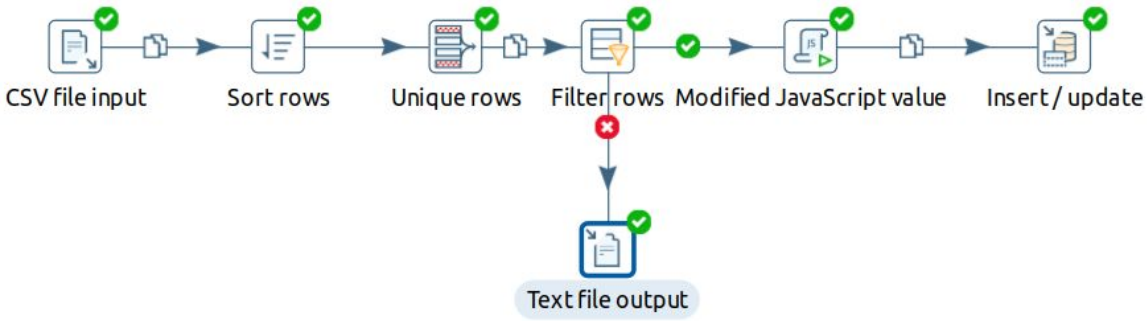
```sql
CREATE PROCEDURE populate_time_dimension()
BEGIN
    DECLARE todays_date DATE;
    DECLARE end_date DATE;

    -- Initialize the start and end dates
    SET todays_date = '2020-01-01';
    SET end_date = '2030-12-31';

    -- Loop through each date
    WHILE todays_date <= end_date DO
        INSERT INTO dim_time (
            time_id,
            date_value,
            year_id,
            season_id,
        season_name,
            month_name,
            month_id
        )
        VALUES (
            -- Date ID format: YYYYMM
            YEAR(todays_date) * 100 + MONTH(todays_date),
            -- Date as a string
            DATE_FORMAT(todays_date, '%Y-%m'),
            -- Year number
            YEAR(todays_date),
            -- Season number
            CASE
                WHEN MONTH(todays_date) IN (12, 1, 2) THEN 1
                WHEN MONTH(todays_date) IN (3, 4, 5) THEN 2
                WHEN MONTH(todays_date) IN (6, 7, 8) THEN 3
                WHEN MONTH(todays_date) IN (9, 10, 11) THEN 4
            END,
        -- Season name
        VALUES (
            -- Date ID format: YYYYMM
            YEAR(todays_date) * 100 + MONTH(todays_date),
            -- Date as a string
            DATE_FORMAT(todays_date, '%Y-%m'),
            -- Year number
            YEAR(todays_date),
            -- Season number
            CASE
                WHEN MONTH(todays_date) IN (12, 1, 2) THEN 1
                WHEN MONTH(todays_date) IN (3, 4, 5) THEN 2
                WHEN MONTH(todays_date) IN (6, 7, 8) THEN 3
                WHEN MONTH(todays_date) IN (9, 10, 11) THEN 4
            END,
        -- Season name
        CASE
                WHEN MONTH(todays_date) IN (12, 1, 2) THEN 'Winter'
                WHEN MONTH(todays_date) IN (3, 4, 5) THEN 'Spring'
                WHEN MONTH(todays_date) IN (6, 7, 8) THEN 'Summer'
                WHEN MONTH(todays_date) IN (9, 10, 11) THEN 'Autumn'
            END,
            -- Month name
            MONTHNAME(todays_date),
            -- Month number
            MONTH(todays_date)
           );

        -- Increment the date by one day
        SET todays_date = DATE_ADD(todays_date, INTERVAL 1 MONTH);
    END WHILE;
END$$
```

Task 5

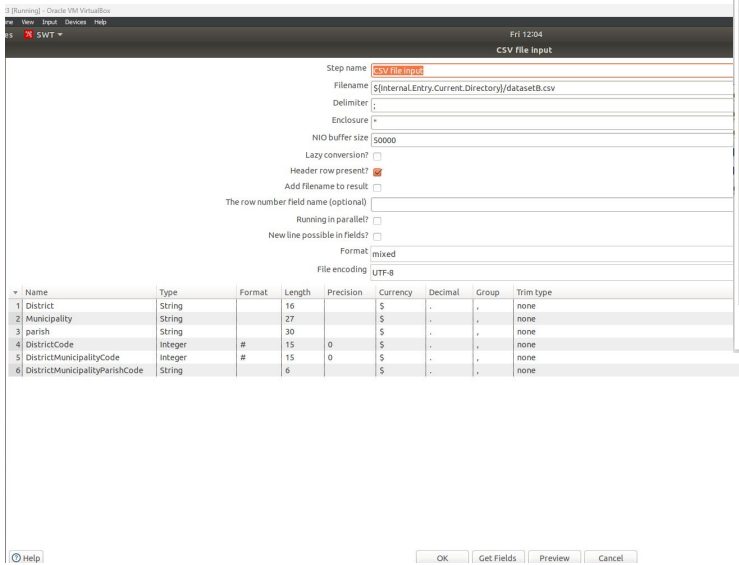
Develop the transformation or transformations to populate the dimension tables.

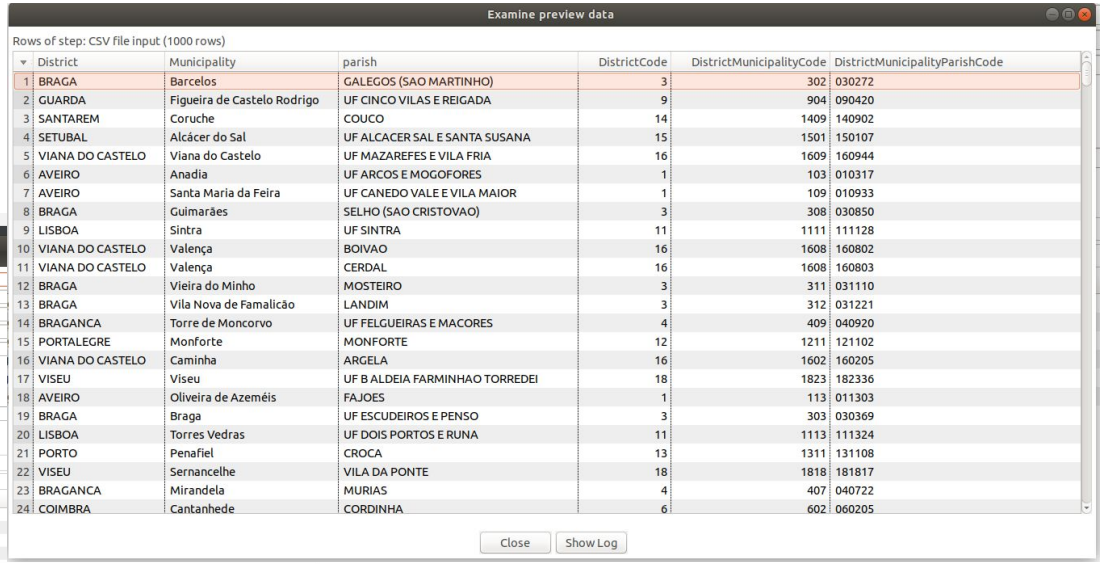
Location Dimension

Transformation



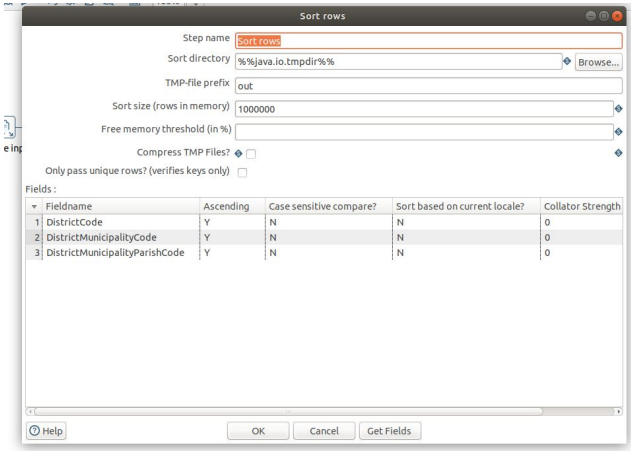
CSV file input

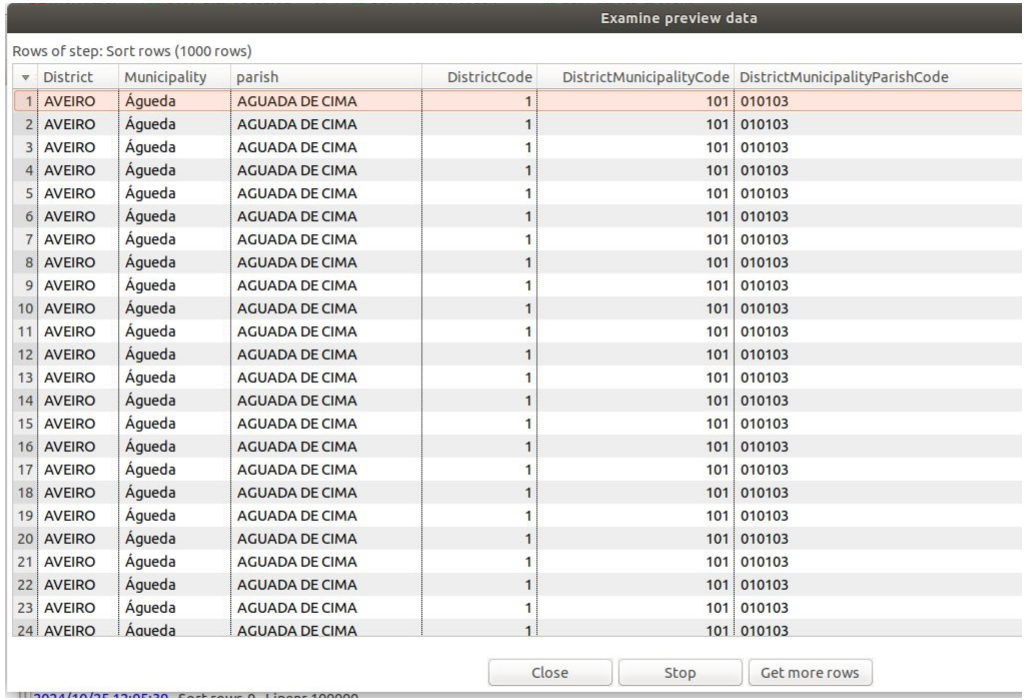




	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	BRAGA	Barcelos	GALEGOS (SAO MARTINHO)	3	302	030272
2	GUARDA	Figueira de Castelo Rodrigo	UF CINCO VILAS E REIGADA	9	904	090420
3	SANTAREM	Coruche	COUCO	14	1409	140902
4	SETUBAL	Alcácer do Sal	UF ALCACER SAL E SANTA SUSANA	15	1501	150107
5	VIANA DO CASTELO	Viana do Castelo	UF MAZAREFES E VILA FRIA	16	1609	160944
6	AVEIRO	Anadia	UF ARCOS E MOGOFORES	1	103	010317
7	AVEIRO	Santa Maria da Feira	UF CANEDO VALE E VILA MAIOR	1	109	010933
8	BRAGA	Guimarães	SELHO (SAO CRISTOVAO)	3	308	030850
9	LISBOA	Sintra	UF SINTRA	11	1111	111128
10	VIANA DO CASTELO	Valença	BOIVAO	16	1608	160802
11	VIANA DO CASTELO	Valença	CERDAL	16	1608	160803
12	BRAGA	Vieira do Minho	MOSTEIRO	3	311	031110
13	BRAGA	Vila Nova de Famalicão	LANDIM	3	312	031221
14	BRAGANCA	Torre de Moncorvo	UF FELGUEIRAS E MACORES	4	409	040920
15	PORTALEGRE	Monforte	MONFORTE	12	1211	121102
16	VIANA DO CASTELO	Caminha	ARGELA	16	1602	160205
17	VISEU	Viseu	UF B ALDEIA FARMINHAO TORREDEI	18	1823	182336
18	AVEIRO	Oliveira de Azeméis	FAJOES	1	113	011303
19	BRAGA	Braga	UF ESCUDEIROS E PENSO	3	303	030369
20	LISBOA	Torres Vedras	UF DOIS PORTOS E RUNA	11	1113	111324
21	PORTO	Penafiel	CROCA	13	1311	131108
22	VISEU	Sernancelhe	VILA DA PONTE	18	1818	181817
23	BRAGANCA	Mirandela	MURIAS	4	407	040722
24	COIMBRA	Cantanhede	CORDINHA	6	602	060205

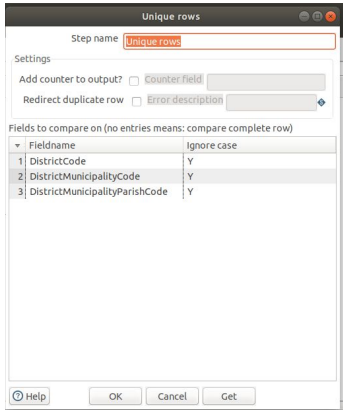
Sort rows

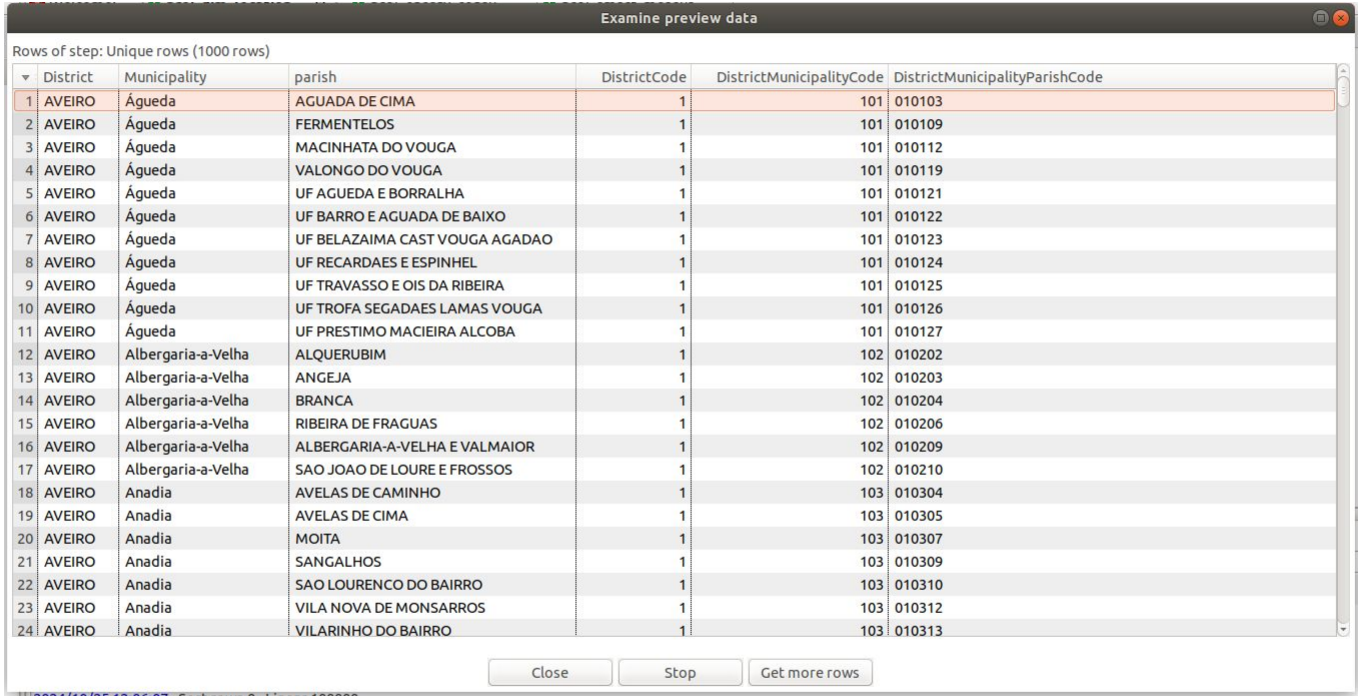




	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
2	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
3	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
4	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
5	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
6	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
7	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
8	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
9	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
10	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
11	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
12	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
13	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
14	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
15	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
16	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
17	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
18	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
19	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
20	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
21	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
22	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
23	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
24	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103

Unique rows

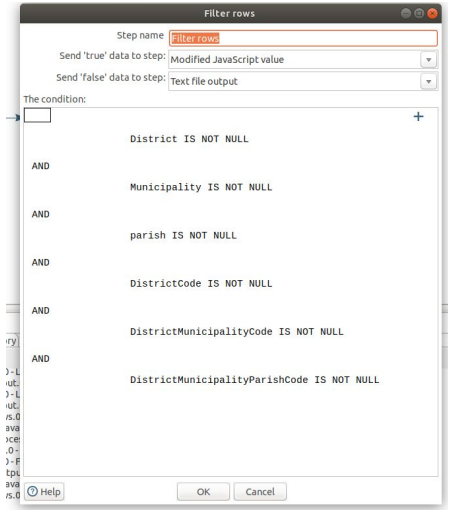




	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
2	AVEIRO	Águeda	FERMENTELOS	1	101	010109
3	AVEIRO	Águeda	MACINHATA DO VOUGA	1	101	010112
4	AVEIRO	Águeda	VALONGO DO VOUGA	1	101	010119
5	AVEIRO	Águeda	UF AGUEDA E BORRALHA	1	101	010121
6	AVEIRO	Águeda	UF BARRO E AGUADA DE BAIXO	1	101	010122
7	AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA AGADAO	1	101	010123
8	AVEIRO	Águeda	UF RECARDAES E ESPINHEL	1	101	010124
9	AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA	1	101	010125
10	AVEIRO	Águeda	UF TROFA SEGADAES LAMAS VOUGA	1	101	010126
11	AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCOBA	1	101	010127
12	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1	102	010202
13	AVEIRO	Albergaria-a-Velha	ANGEJA	1	102	010203
14	AVEIRO	Albergaria-a-Velha	BRANCA	1	102	010204
15	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	1	102	010206
16	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	1	102	010209
17	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	1	102	010210
18	AVEIRO	Anadia	AVELAS DE CAMINHO	1	103	010304
19	AVEIRO	Anadia	AVELAS DE CIMA	1	103	010305
20	AVEIRO	Anadia	MOITA	1	103	010307
21	AVEIRO	Anadia	SANGALHOS	1	103	010309
22	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	1	103	010310
23	AVEIRO	Anadia	VILA NOVA DE MONSARROS	1	103	010312
24	AVEIRO	Anadia	VILARINHO DO BAIRRO	1	103	010313

The fields with null parish code due to GDPR (or other irregularities) were filtered and sent to an output file with the purpose of further analysis and transformation.

Filter rows



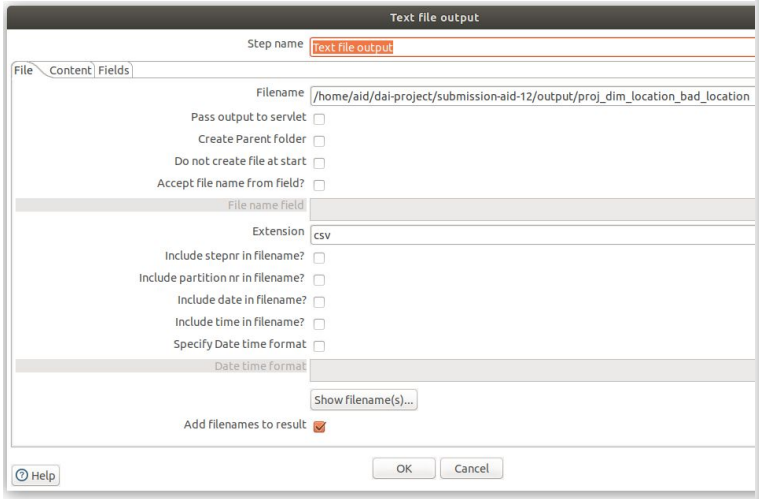
Examine preview data

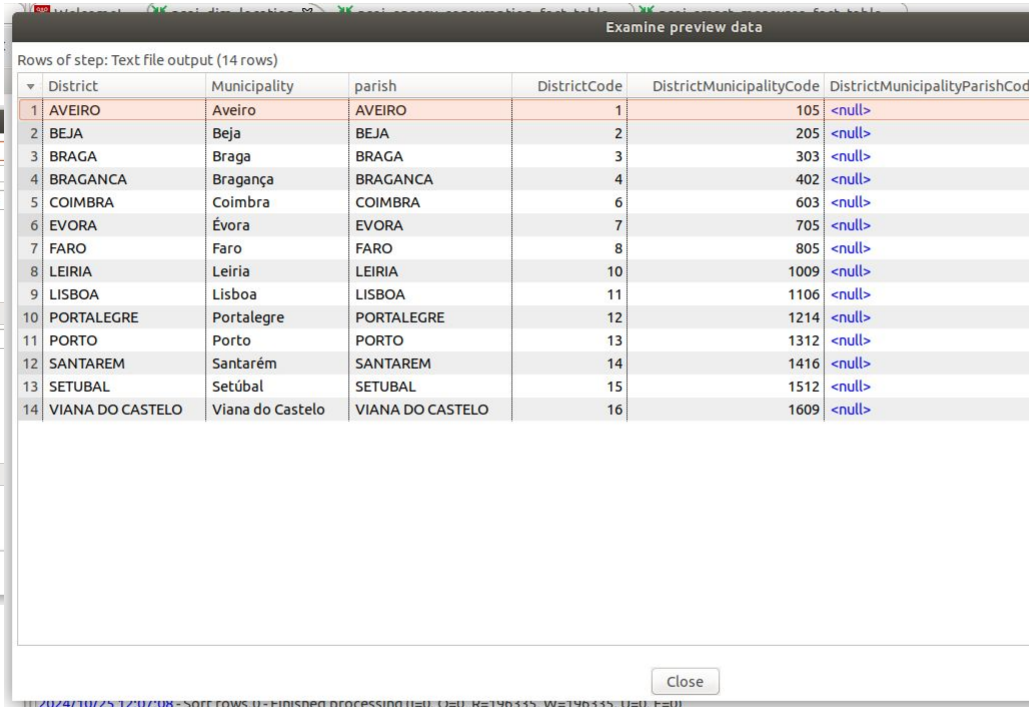
Rows of step: Filter rows (1000 rows)

	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103
2	AVEIRO	Águeda	FERMENTELOS	1	101	010109
3	AVEIRO	Águeda	MACINHATA DO VOUGA	1	101	010112
4	AVEIRO	Águeda	VALONGO DO VOUGA	1	101	010119
5	AVEIRO	Águeda	UF AGUEDA E BORRALHA	1	101	010121
6	AVEIRO	Águeda	UF BARRO E AGUADA DE BAIXO	1	101	010122
7	AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA AGADAO	1	101	010123
8	AVEIRO	Águeda	UF RECARDAES E ESPINHEL	1	101	010124
9	AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA	1	101	010125
10	AVEIRO	Águeda	UF TROFA SEGADAES LAMAS VOUGA	1	101	010126
11	AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCOBA	1	101	010127
12	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1	102	010202
13	AVEIRO	Albergaria-a-Velha	ANGEJA	1	102	010203
14	AVEIRO	Albergaria-a-Velha	BRANCA	1	102	010204
15	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	1	102	010206
16	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	1	102	010209
17	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	1	102	010210
18	AVEIRO	Anadia	AVELAS DE CAMINHO	1	103	010304
19	AVEIRO	Anadia	AVELAS DE CIMA	1	103	010305
20	AVEIRO	Anadia	MOITA	1	103	010307
21	AVEIRO	Anadia	SANGALHOS	1	103	010309
22	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	1	103	010310
23	AVEIRO	Anadia	VILA NOVA DE MONSARROS	1	103	010312
24	AVEIRO	Anadia	VILARINHO DO BAIRRO	1	103	010313

Close | Stop | Get more rows

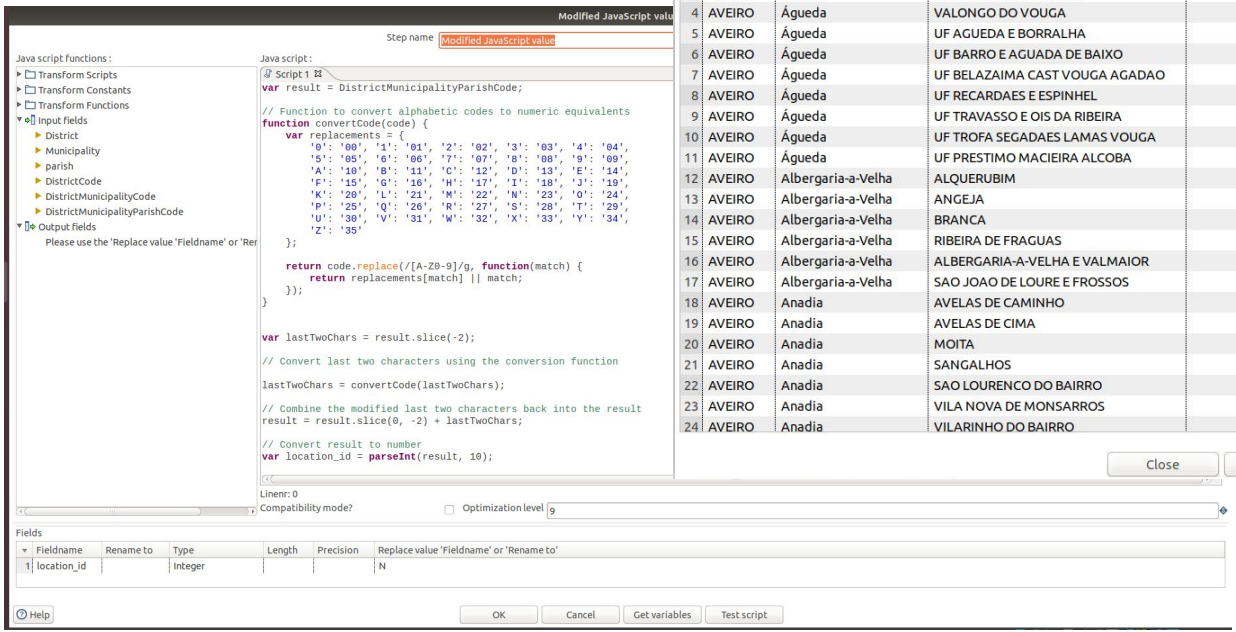
Text file output

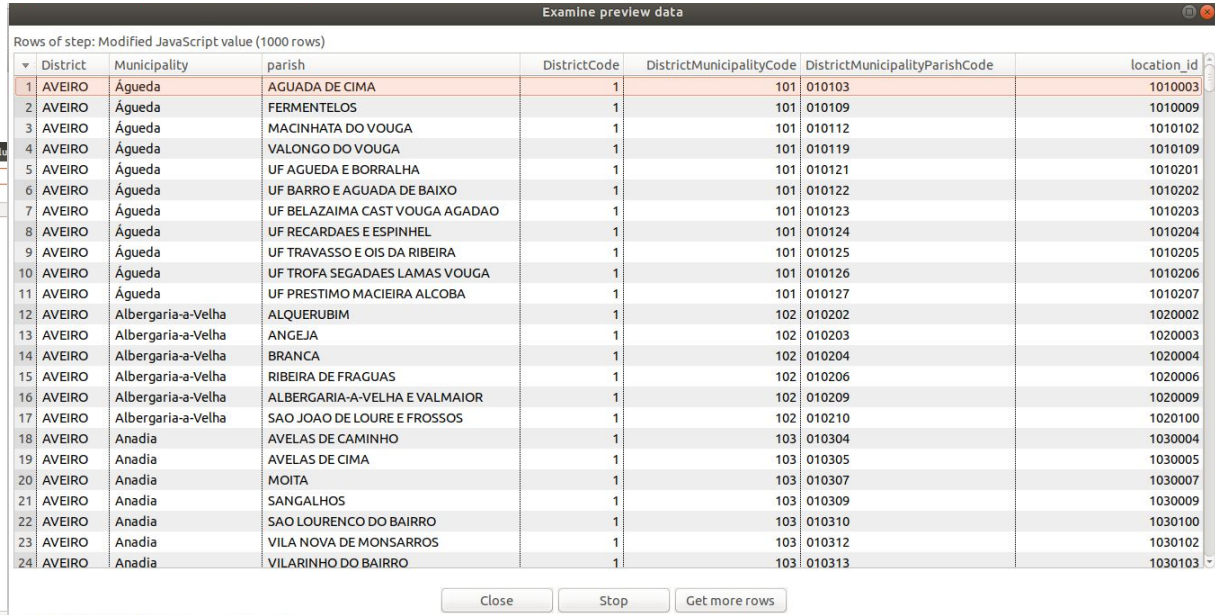




Modified JavaScript value

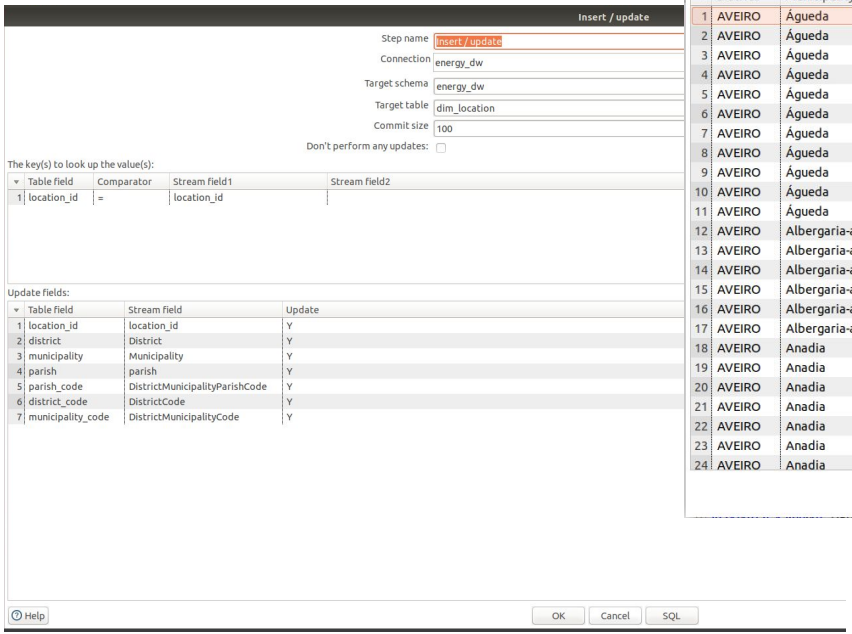
Some parish codes (namely from the municipality of Barcelos) include letters. As this field was used to build a surrogate key, a script was needed to build integer keys.

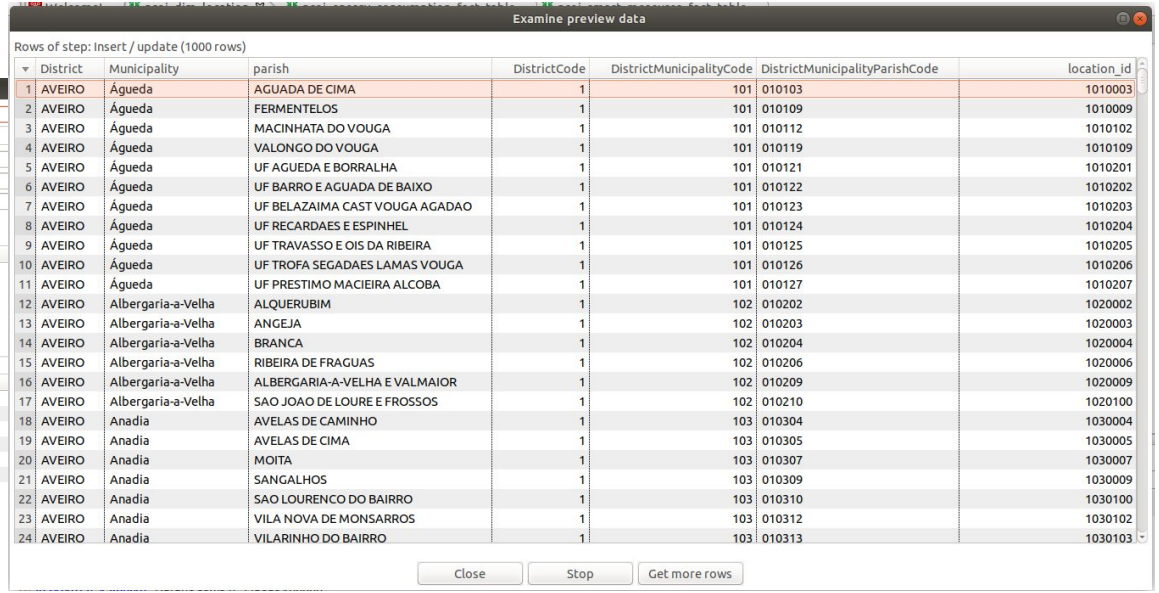




	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	location_id
1	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103	1010003
2	AVEIRO	Águeda	FERMENTELOS	1	101	010109	1010009
3	AVEIRO	Águeda	MACINHATA DO VOUGA	1	101	010112	1010102
4	AVEIRO	Águeda	VALONGO DO VOUGA	1	101	010119	1010109
5	AVEIRO	Águeda	UF AGUEDA E BORRALHA	1	101	010121	1010201
6	AVEIRO	Águeda	UF BARRO E AGUADA DE BAIXO	1	101	010122	1010202
7	AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA AGADAO	1	101	010123	1010203
8	AVEIRO	Águeda	UF RECARDAES E ESPINHEL	1	101	010124	1010204
9	AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA	1	101	010125	1010205
10	AVEIRO	Águeda	UF TROFA SEGADAES LAMAS VOUGA	1	101	010126	1010206
11	AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCOBA	1	101	010127	1010207
12	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1	102	010202	1020002
13	AVEIRO	Albergaria-a-Velha	ANGEJA	1	102	010203	1020003
14	AVEIRO	Albergaria-a-Velha	BRANCA	1	102	010204	1020004
15	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	1	102	010206	1020006
16	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	1	102	010209	1020009
17	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	1	102	010210	1020100
18	AVEIRO	Anadia	AVELAS DE CAMINHO	1	103	010304	1030004
19	AVEIRO	Anadia	AVELAS DE CIMA	1	103	010305	1030005
20	AVEIRO	Anadia	MOITA	1	103	010307	1030007
21	AVEIRO	Anadia	SANGALHOS	1	103	010309	1030009
22	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	1	103	010310	1030100
23	AVEIRO	Anadia	VILA NOVA DE MONSARROS	1	103	010312	1030102
24	AVEIRO	Anadia	VILARINHO DO BAIRRO	1	103	010313	1030103

Insert/Update





	District	Municipality	parish	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	location_id
1	AVEIRO	Águeda	AGUADA DE CIMA	1	101	010103	1010003
2	AVEIRO	Águeda	FERMENTELOS	1	101	010109	1010009
3	AVEIRO	Águeda	MACINHATA DO VOUGA	1	101	010112	1010102
4	AVEIRO	Águeda	VALONGO DO VOUGA	1	101	010119	1010109
5	AVEIRO	Águeda	UF AGUEDA E BORRALHA	1	101	010121	1010201
6	AVEIRO	Águeda	UF BARRO E AGUADA DE BAIXO	1	101	010122	1010202
7	AVEIRO	Águeda	UF BELAZAIMA CAST VOUGA AGADAO	1	101	010123	1010203
8	AVEIRO	Águeda	UF RECARDAES E ESPINHEL	1	101	010124	1010204
9	AVEIRO	Águeda	UF TRAVASSO E OIS DA RIBEIRA	1	101	010125	1010205
10	AVEIRO	Águeda	UF TROFA SEGADAES LAMAS VOUGA	1	101	010126	1010206
11	AVEIRO	Águeda	UF PRESTIMO MACIEIRA ALCOBA	1	101	010127	1010207
12	AVEIRO	Albergaria-a-Velha	ALQUERUBIM	1	102	010202	1020002
13	AVEIRO	Albergaria-a-Velha	ANGEJA	1	102	010203	1020003
14	AVEIRO	Albergaria-a-Velha	BRANCA	1	102	010204	1020004
15	AVEIRO	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	1	102	010206	1020006
16	AVEIRO	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	1	102	010209	1020009
17	AVEIRO	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	1	102	010210	1020100
18	AVEIRO	Anadia	AVELAS DE CAMINHO	1	103	010304	1030004
19	AVEIRO	Anadia	AVELAS DE CIMA	1	103	010305	1030005
20	AVEIRO	Anadia	MOITA	1	103	010307	1030007
21	AVEIRO	Anadia	SANGALHOS	1	103	010309	1030009
22	AVEIRO	Anadia	SAO LOURENCO DO BAIRRO	1	103	010310	1030100
23	AVEIRO	Anadia	VILA NOVA DE MONSARROS	1	103	010312	1030102
24	AVEIRO	Anadia	VILARINHO DO BAIRRO	1	103	010313	1030103

Output Table

```
mysql> select * from dim_location limit 25;
+-------------+----------+---------------+--------------------+-------------------+--------------------------------+-------------+
| location_id | district | district_code | municipality       | municipality_code | parish                         | parish_code |
+-------------+----------+---------------+--------------------+-------------------+--------------------------------+-------------+
|     1010003 | AVEIRO   |             1 | Águeda             |               101 | AGUADA DE CIMA                 | 010103      |
|     1010009 | AVEIRO   |             1 | Águeda             |               101 | FERMENTELOS                    | 010109      |
|     1010102 | AVEIRO   |             1 | Águeda             |               101 | MACINHATA DO VOUGA             | 010112      |
|     1010109 | AVEIRO   |             1 | Águeda             |               101 | VALONGO DO VOUGA               | 010119      |
|     1010201 | AVEIRO   |             1 | Águeda             |               101 | UF AGUEDA E BORRALHA           | 010121      |
|     1010202 | AVEIRO   |             1 | Águeda             |               101 | UF BARRO E AGUADA DE BAIXO     | 010122      |
|     1010203 | AVEIRO   |             1 | Águeda             |               101 | UF BELAZAIMA CAST VOUGA AGADAO | 010123      |
|     1010204 | AVEIRO   |             1 | Águeda             |               101 | UF RECARDAES E ESPINHEL        | 010124      |
|     1010205 | AVEIRO   |             1 | Águeda             |               101 | UF TRAVASSO E OIS DA RIBEIRA   | 010125      |
|     1010206 | AVEIRO   |             1 | Águeda             |               101 | UF TROFA SEGADAES LAMAS VOUGA  | 010126      |
|     1010207 | AVEIRO   |             1 | Águeda             |               101 | UF PRESTIMO MACIEIRA ALCOBA    | 010127      |
|     1020002 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | ALQUERUBIM                     | 010202      |
|     1020003 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | ANGEJA                         | 010203      |
|     1020004 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | BRANCA                         | 010204      |
|     1020006 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | RIBEIRA DE FRAGUAS             | 010206      |
|     1020009 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | ALBERGARIA-A-VELHA E VALMAIOR  | 010209      |
|     1020100 | AVEIRO   |             1 | Albergaria-a-Velha |               102 | SAO JOAO DE LOURE E FROSSOS    | 010210      |
|     1030004 | AVEIRO   |             1 | Anadia             |               103 | AVELAS DE CAMINHO              | 010304      |
|     1030005 | AVEIRO   |             1 | Anadia             |               103 | AVELAS DE CIMA                 | 010305      |
|     1030007 | AVEIRO   |             1 | Anadia             |               103 | MOITA                          | 010307      |
|     1030009 | AVEIRO   |             1 | Anadia             |               103 | SANGALHOS                      | 010309      |
|     1030100 | AVEIRO   |             1 | Anadia             |               103 | SAO LOURENCO DO BAIRRO         | 010310      |
|     1030102 | AVEIRO   |             1 | Anadia             |               103 | VILA NOVA DE MONSARROS         | 010312      |
|     1030103 | AVEIRO   |             1 | Anadia             |               103 | VILARINHO DO BAIRRO            | 010313      |
|     1030106 | AVEIRO   |             1 | Anadia             |               103 | UF A DA GANDARA BAIRRO E ANCAS | 010316      |
+-------------+----------+---------------+--------------------+-------------------+--------------------------------+-------------+
25 rows in set (0.01 sec)
```

Time Dimension

Output Table

```
mysql> select * from dim_time limit 25;
+---------+---------+------------+-------------+-----------+------------+----------+
| time_id | year_id | date_value | season_name | season_id | month_name | month_id |
+---------+---------+------------+-------------+-----------+------------+----------+
|  202001 |    2020 | 2020-01    | Winter      |         1 | January    |        1 |
|  202002 |    2020 | 2020-02    | Winter      |         1 | February   |        2 |
|  202003 |    2020 | 2020-03    | Spring      |         2 | March      |        3 |
|  202004 |    2020 | 2020-04    | Spring      |         2 | April      |        4 |
|  202005 |    2020 | 2020-05    | Spring      |         2 | May        |        5 |
|  202006 |    2020 | 2020-06    | Summer      |         3 | June       |        6 |
|  202007 |    2020 | 2020-07    | Summer      |         3 | July       |        7 |
|  202008 |    2020 | 2020-08    | Summer      |         3 | August     |        8 |
|  202009 |    2020 | 2020-09    | Autumn      |         4 | September  |        9 |
|  202010 |    2020 | 2020-10    | Autumn      |         4 | October    |       10 |
|  202011 |    2020 | 2020-11    | Autumn      |         4 | November   |       11 |
|  202012 |    2020 | 2020-12    | Winter      |         1 | December   |       12 |
|  202101 |    2021 | 2021-01    | Winter      |         1 | January    |        1 |
|  202102 |    2021 | 2021-02    | Winter      |         1 | February   |        2 |
|  202103 |    2021 | 2021-03    | Spring      |         2 | March      |        3 |
|  202104 |    2021 | 2021-04    | Spring      |         2 | April      |        4 |
|  202105 |    2021 | 2021-05    | Spring      |         2 | May        |        5 |
|  202106 |    2021 | 2021-06    | Summer      |         3 | June       |        6 |
|  202107 |    2021 | 2021-07    | Summer      |         3 | July       |        7 |
|  202108 |    2021 | 2021-08    | Summer      |         3 | August     |        8 |
|  202109 |    2021 | 2021-09    | Autumn      |         4 | September  |        9 |
|  202110 |    2021 | 2021-10    | Autumn      |         4 | October    |       10 |
|  202111 |    2021 | 2021-11    | Autumn      |         4 | November   |       11 |
|  202112 |    2021 | 2021-12    | Winter      |         1 | December   |       12 |
|  202201 |    2022 | 2022-01    | Winter      |         1 | January    |        1 |
+---------+---------+------------+-------------+-----------+------------+----------+
25 rows in set (0.00 sec)
```

The time dimension was built and the corresponding procedure called directly in the sql script.

Task 6

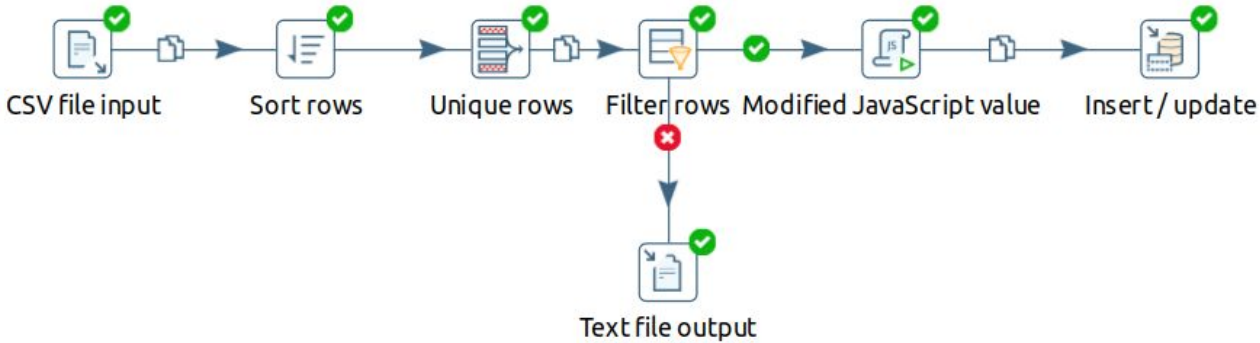
Develop the transformation to populate the fact table.

We opted to adopt a constellation schema instead of a star schema, consequently two fact tables that share the same dimension tables (location and time) were created.

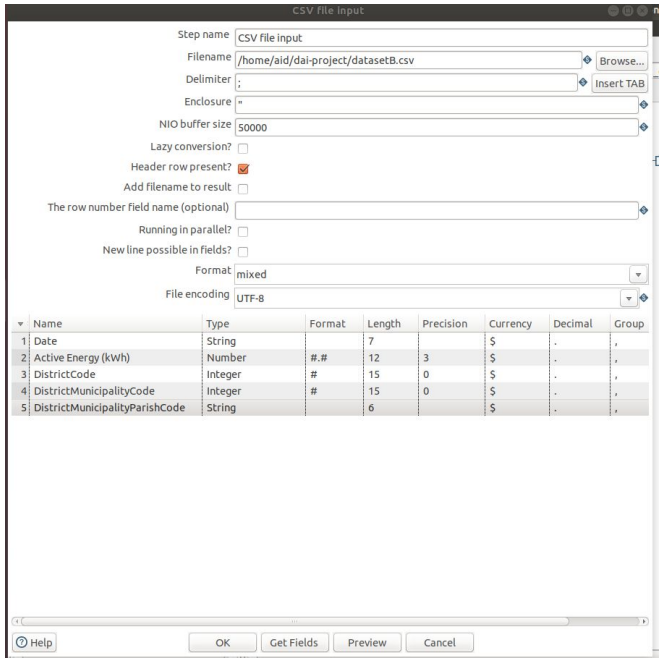
This comes from the point that both datasets have chronological differences and we aimed to gather as much data as possible and avoid filtering out rows without information about both energy consumption and smart meter at the same time and location, in the case of a single fact table.

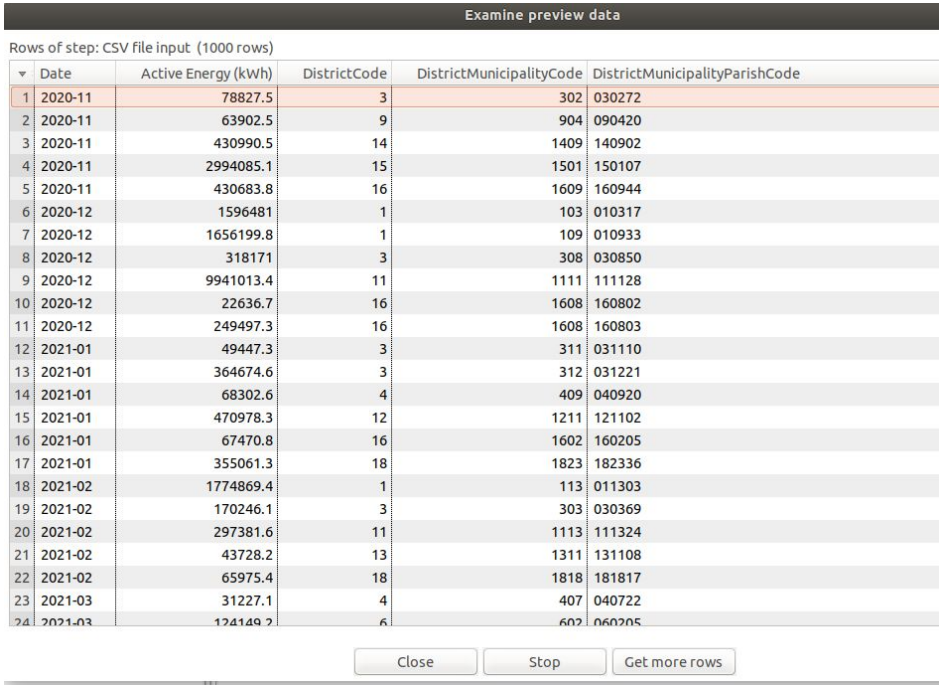
Energy Consumption Fact Table

Transformation



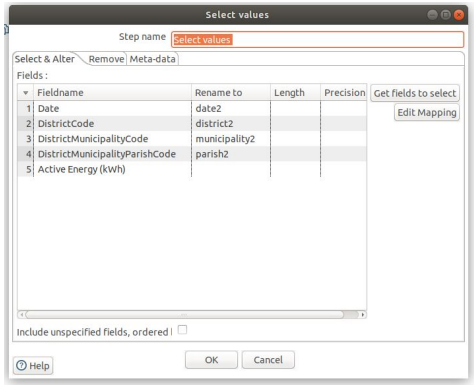
CSV file input

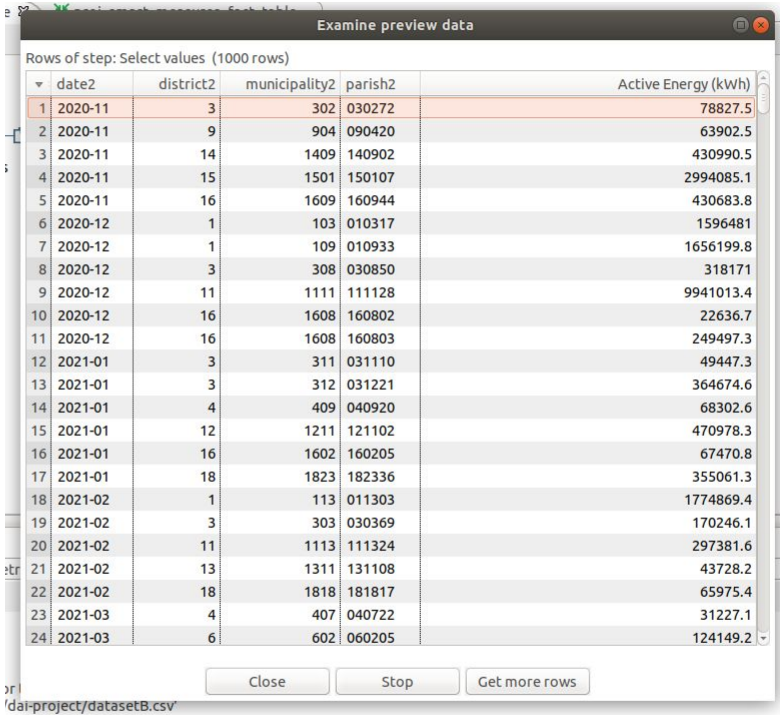




	Date	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2020-11	78827.5	3	302	030272
2	2020-11	63902.5	9	904	090420
3	2020-11	430990.5	14	1409	140902
4	2020-11	2994085.1	15	1501	150107
5	2020-11	430683.8	16	1609	160944
6	2020-12	1596481	1	103	010317
7	2020-12	1656199.8	1	109	010933
8	2020-12	318171	3	308	030850
9	2020-12	9941013.4	11	1111	111128
10	2020-12	22636.7	16	1608	160802
11	2020-12	249497.3	16	1608	160803
12	2021-01	49447.3	3	311	031110
13	2021-01	364674.6	3	312	031221
14	2021-01	68302.6	4	409	040920
15	2021-01	470978.3	12	1211	121102
16	2021-01	67470.8	16	1602	160205
17	2021-01	355061.3	18	1823	182336
18	2021-02	1774869.4	1	113	011303
19	2021-02	170246.1	3	303	030369
20	2021-02	297381.6	11	1113	111324
21	2021-02	43728.2	13	1311	131108
22	2021-02	65975.4	18	1818	181817
23	2021-03	31227.1	4	407	040722
24	2021-03	124149.2	6	602	060205

Select values

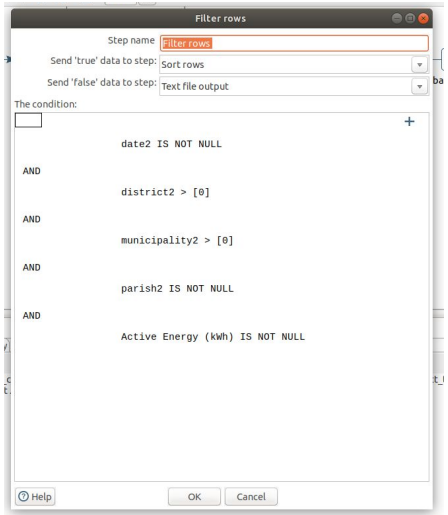


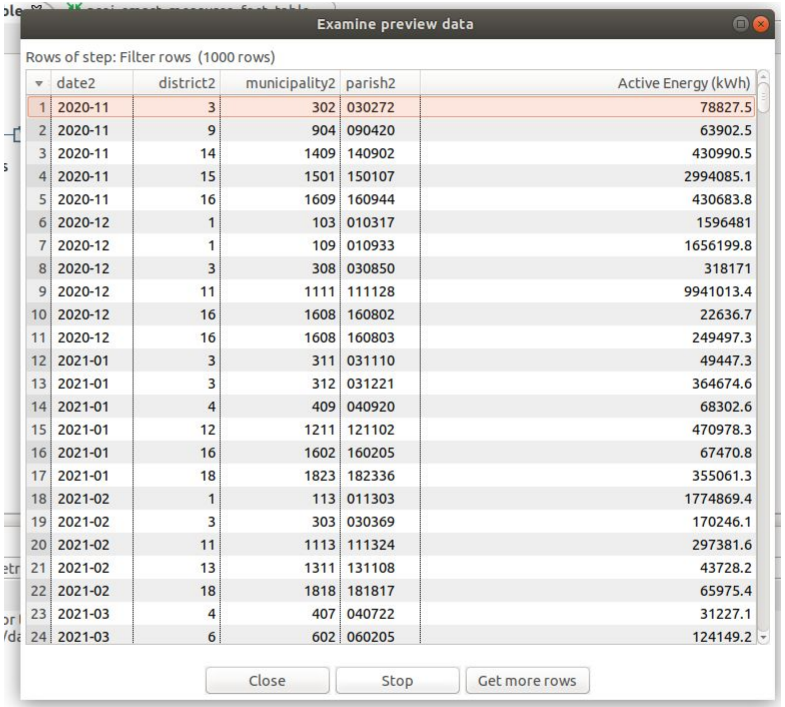


	date2	district2	municipality2	parish2	Active Energy (kWh)
1	2020-11	3	302	030272	78827.5
2	2020-11	9	904	090420	63902.5
3	2020-11	14	1409	140902	430990.5
4	2020-11	15	1501	150107	2994085.1
5	2020-11	16	1609	160944	430683.8
6	2020-12	1	103	010317	1596481
7	2020-12	1	109	010933	1656199.8
8	2020-12	3	308	030850	318171
9	2020-12	11	1111	111128	9941013.4
10	2020-12	16	1608	160802	22636.7
11	2020-12	16	1608	160803	249497.3
12	2021-01	3	311	031110	49447.3
13	2021-01	3	312	031221	364674.6
14	2021-01	4	409	040920	68302.6
15	2021-01	12	1211	121102	470978.3
16	2021-01	16	1602	160205	67470.8
17	2021-01	18	1823	182336	355061.3
18	2021-02	1	113	011303	1774869.4
19	2021-02	3	303	030369	170246.1
20	2021-02	11	1113	111324	297381.6
21	2021-02	13	1311	131108	43728.2
22	2021-02	18	1818	181817	65975.4
23	2021-03	4	407	040722	31227.1
24	2021-03	6	602	060205	124149.2

Filter rows

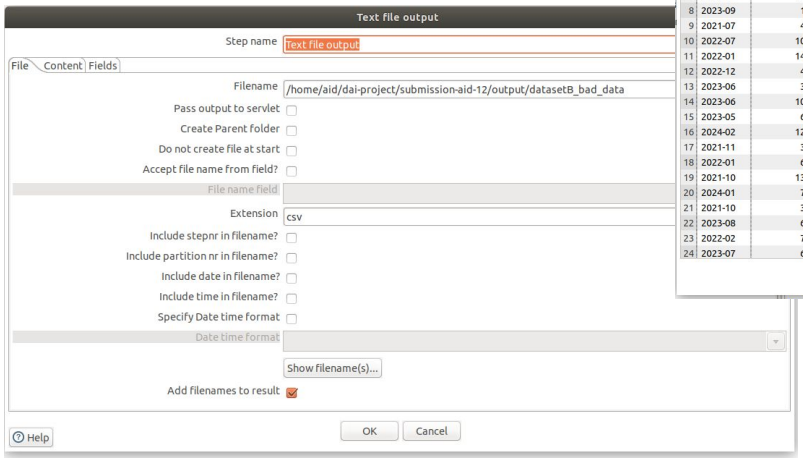
The fields with null parish code due to GDPR (or other irregularities) were filtered and sent to an output file with the purpose of further analysis and transformation.

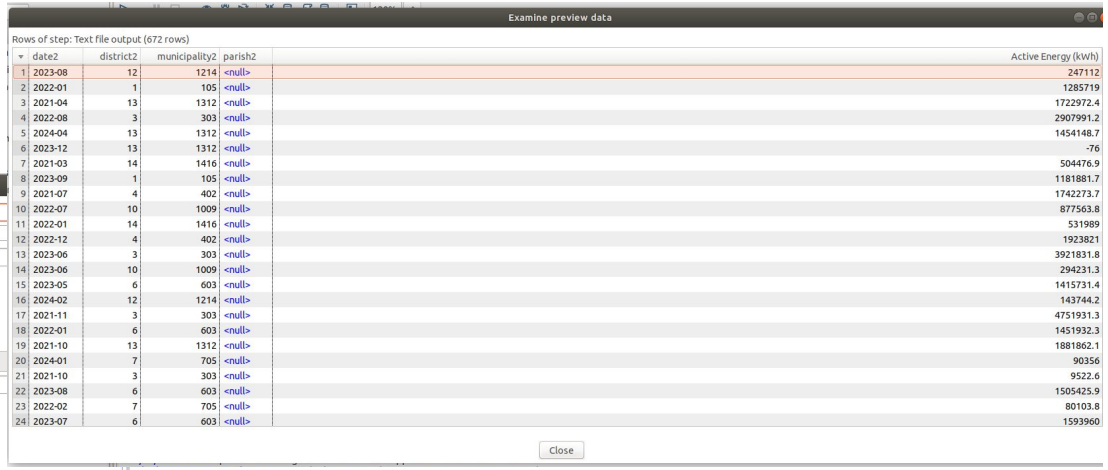




	date2	district2	municipality2	parish2	Active Energy (kWh)
1	2020-11	3	302	030272	78827.5
2	2020-11	9	904	090420	63902.5
3	2020-11	14	1409	140902	430990.5
4	2020-11	15	1501	150107	2994085.1
5	2020-11	16	1609	160944	430683.8
6	2020-12	1	103	010317	1596481
7	2020-12	1	109	010933	1656199.8
8	2020-12	3	308	030850	318171
9	2020-12	11	1111	111128	9941013.4
10	2020-12	16	1608	160802	22636.7
11	2020-12	16	1608	160803	249497.3
12	2021-01	3	311	031110	49447.3
13	2021-01	3	312	031221	364674.6
14	2021-01	4	409	040920	68302.6
15	2021-01	12	1211	121102	470978.3
16	2021-01	16	1602	160205	67470.8
17	2021-01	18	1823	182336	355061.3
18	2021-02	1	113	011303	1774869.4
19	2021-02	3	303	030369	170246.1
20	2021-02	11	1113	111324	297381.6
21	2021-02	13	1311	131108	43728.2
22	2021-02	18	1818	181817	65975.4
23	2021-03	4	407	040722	31227.1
24	2021-03	6	602	060205	124149.2

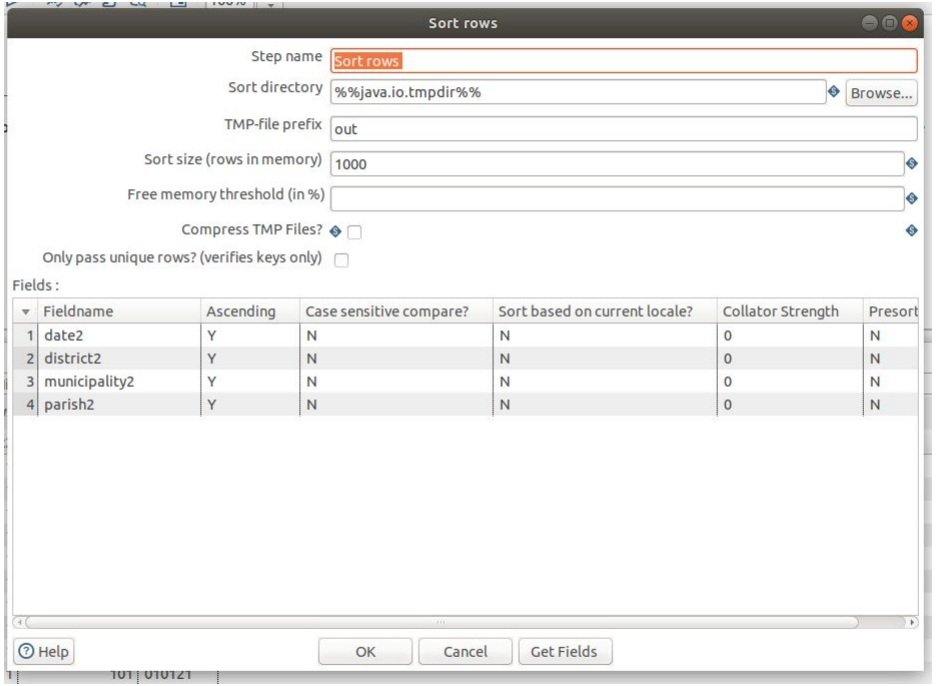
Text file output

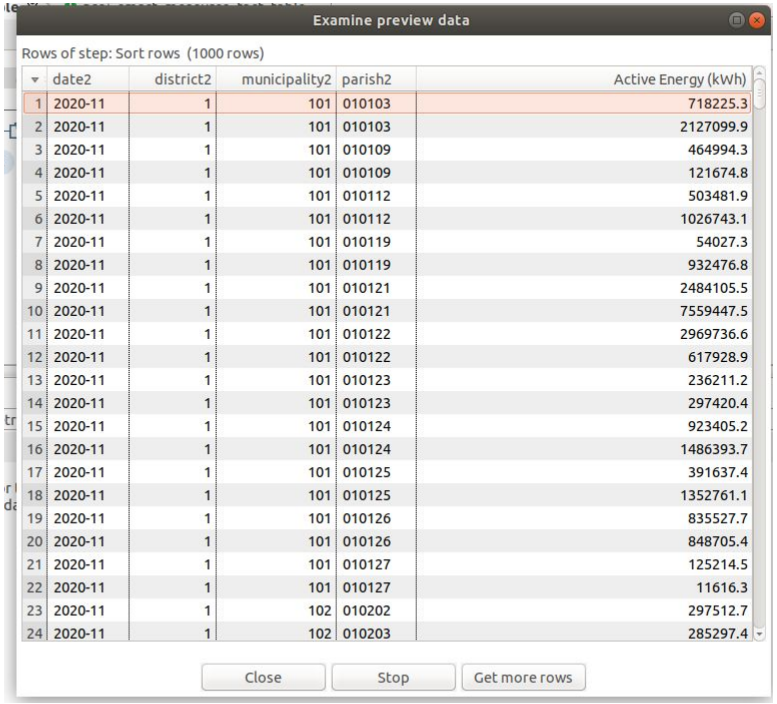




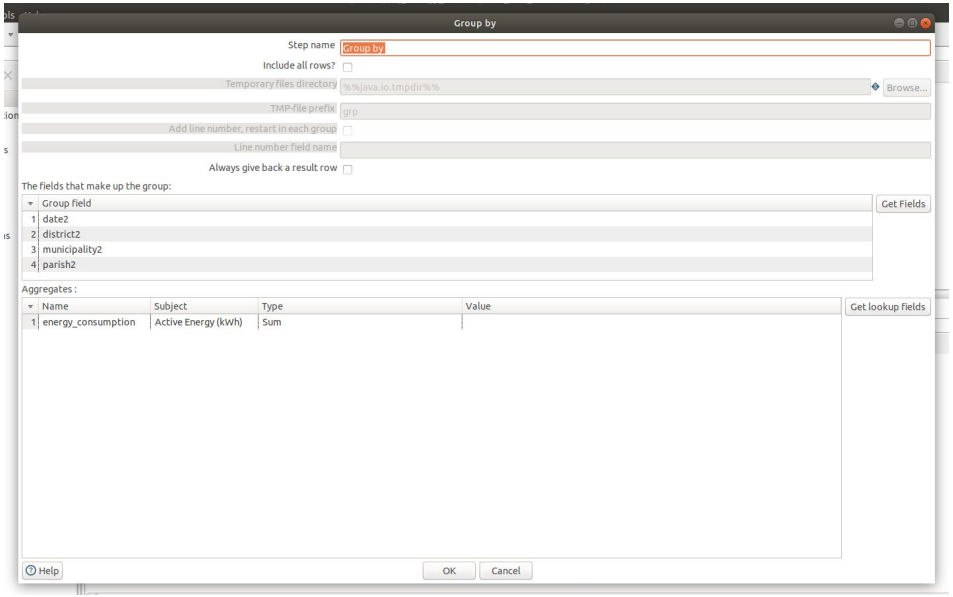
	date2	district2	municipality2	parish2	Active Energy (kWh)
1	2023-08	12	1214	<null>	247112
2	2022-01	1	105	<null>	1285719
3	2021-04	13	1312	<null>	1722972.4
4	2022-08	3	303	<null>	2907991.2
5	2024-04	13	1312	<null>	1454148.7
6	2023-12	13	1312	<null>	-76
7	2021-03	14	1416	<null>	504476.9
8	2023-09	1	105	<null>	1181881.7
9	2021-07	4	402	<null>	1742273.7
10	2022-07	10	1009	<null>	877563.8
11	2022-01	14	1416	<null>	531989
12	2022-12	4	402	<null>	1923821
13	2023-06	3	303	<null>	3921831.8
14	2023-06	10	1009	<null>	294231.3
15	2023-05	6	603	<null>	1415731.4
16	2024-02	12	1214	<null>	143744.2
17	2021-11	3	303	<null>	4751931.3
18	2022-01	6	603	<null>	1451932.3
19	2021-10	13	1312	<null>	1881862.1
20	2024-01	7	705	<null>	90356
21	2021-10	3	303	<null>	9522.6
22	2023-08	6	603	<null>	1505425.9
23	2022-02	7	705	<null>	80103.8
24	2023-07	6	603	<null>	1593960

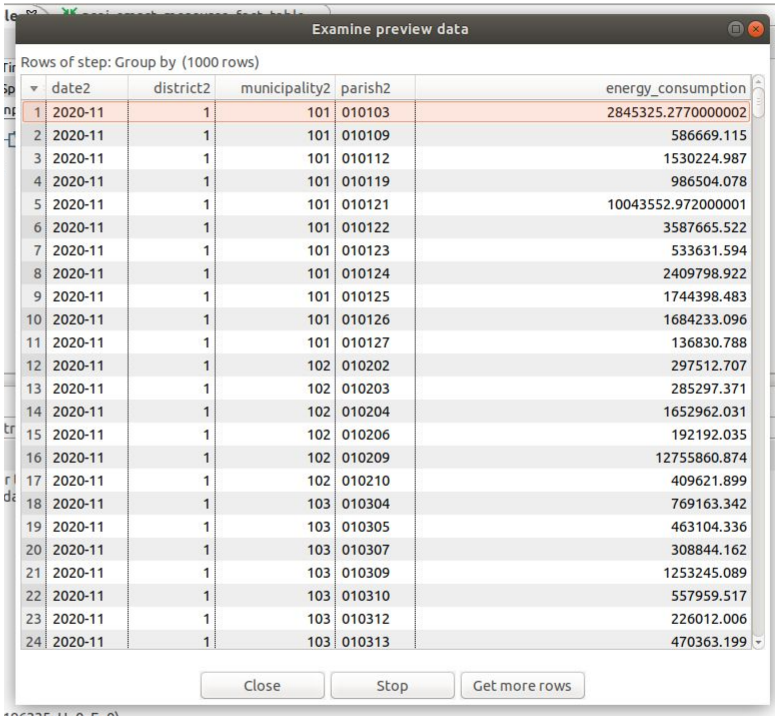
Sort rows


Sort rows
Step name
Sort rows
Sort directory
%%java.io.tmpdir%%
Browse...
TMP-file prefix
out
Sort size (rows in memory)
1000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys only)
Fields :
Fieldname
Ascending
Case sensitive compare?
Sort based on current locale?
Collator Strength
Presort
1 date2 Y N N 0 N
2 district2 Y N N 0 N
3 municipality2 Y N N 0 N
4 parish2 Y N N 0 N
Help
OK
Cancel
Get Fields

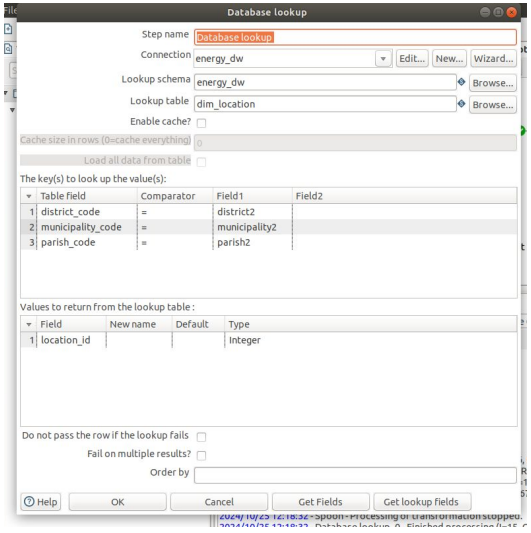

Examine preview data
Rows of step: Sort rows (1000 rows)
date2 district2 municipality2 parish2 Active Energy (kWh)
1 2020-11 1 101 010103 718225.3
2 2020-11 1 101 010103 2127099.9
3 2020-11 1 101 010109 464994.3
4 2020-11 1 101 010109 121674.8
5 2020-11 1 101 010112 503481.9
6 2020-11 1 101 010112 1026743.1
7 2020-11 1 101 010119 54027.3
8 2020-11 1 101 010119 932476.8
9 2020-11 1 101 010121 2484105.5
10 2020-11 1 101 010121 7559447.5
11 2020-11 1 101 010122 2969736.6
12 2020-11 1 101 010122 617928.9
13 2020-11 1 101 010123 236211.2
14 2020-11 1 101 010123 297420.4
15 2020-11 1 101 010124 923405.2
16 2020-11 1 101 010124 1486393.7
17 2020-11 1 101 010125 391637.4
18 2020-11 1 101 010125 1352761.1
19 2020-11 1 101 010126 835527.7
20 2020-11 1 101 010126 848705.4
21 2020-11 1 101 010127 125214.5
22 2020-11 1 101 010127 11616.3
23 2020-11 1 102 010202 297512.7
24 2020-11 1 102 010203 285297.4
Close
Stop
Get more rows

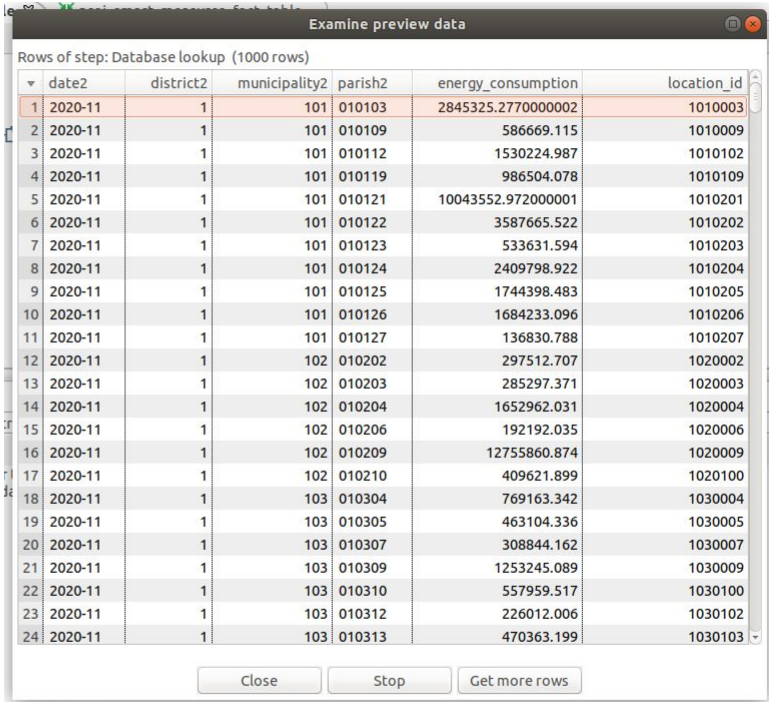
Group by


Group by
Step name
Group by
Include all rows?
Temporary files directory
%%java.io.tmpdir%%
Browse...
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
Get Fields
1 date2
2 district2
3 municipality2
4 parish2
Aggregates :
Name
Subject
Type
Value
Get lookup fields
1 energy_consumption
Active Energy (kWh)
Sum
Help
OK
Cancel


Examine preview data
Rows of step: Group by (1000 rows)
date2
district2
municipality2
parish2
energy_consumption
1 2020-11 1 101 010103 2845325.2770000002
2 2020-11 1 101 010109 586669.115
3 2020-11 1 101 010112 1530224.987
4 2020-11 1 101 010119 986504.078
5 2020-11 1 101 010121 10043552.972000001
6 2020-11 1 101 010122 3587665.522
7 2020-11 1 101 010123 533631.594
8 2020-11 1 101 010124 2409798.922
9 2020-11 1 101 010125 1744398.483
10 2020-11 1 101 010126 1684233.096
11 2020-11 1 101 010127 136830.788
12 2020-11 1 102 010202 297512.707
13 2020-11 1 102 010203 285297.371
14 2020-11 1 102 010204 1652962.031
15 2020-11 1 102 010206 192192.035
16 2020-11 1 102 010209 12755860.874
17 2020-11 1 102 010210 409621.899
18 2020-11 1 103 010304 769163.342
19 2020-11 1 103 010305 463104.336
20 2020-11 1 103 010307 308844.162
21 2020-11 1 103 010309 1253245.089
22 2020-11 1 103 010310 557959.517
23 2020-11 1 103 010312 226012.006
24 2020-11 1 103 010313 470363.199
Close
Stop
Get more rows

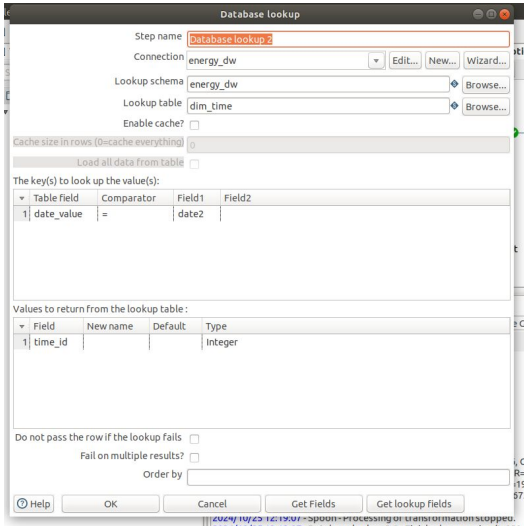
Database lookup

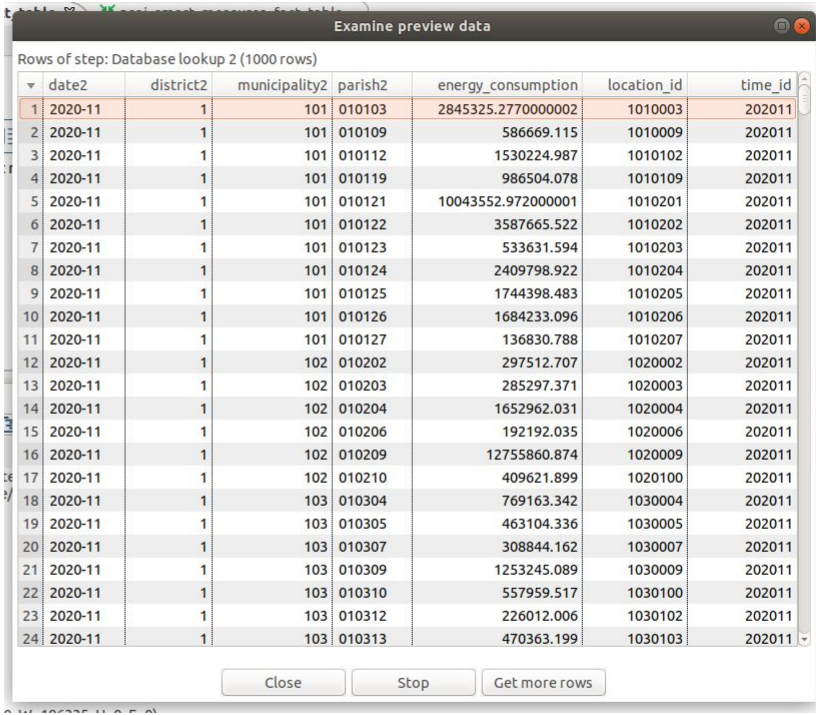




	date2	district2	municipality2	parish2	energy_consumption	location_id
1	2020-11	1	101	010103	2845325.2770000002	1010003
2	2020-11	1	101	010109	586669.115	1010009
3	2020-11	1	101	010112	1530224.987	1010102
4	2020-11	1	101	010119	986504.078	1010109
5	2020-11	1	101	010121	10043552.972000001	1010201
6	2020-11	1	101	010122	3587665.522	1010202
7	2020-11	1	101	010123	533631.594	1010203
8	2020-11	1	101	010124	2409798.922	1010204
9	2020-11	1	101	010125	1744398.483	1010205
10	2020-11	1	101	010126	1684233.096	1010206
11	2020-11	1	101	010127	136830.788	1010207
12	2020-11	1	102	010202	297512.707	1020002
13	2020-11	1	102	010203	285297.371	1020003
14	2020-11	1	102	010204	1652962.031	1020004
15	2020-11	1	102	010206	192192.035	1020006
16	2020-11	1	102	010209	12755860.874	1020009
17	2020-11	1	102	010210	409621.899	1020100
18	2020-11	1	103	010304	769163.342	1030004
19	2020-11	1	103	010305	463104.336	1030005
20	2020-11	1	103	010307	308844.162	1030007
21	2020-11	1	103	010309	1253245.089	1030009
22	2020-11	1	103	010310	557959.517	1030100
23	2020-11	1	103	010312	226012.006	1030102
24	2020-11	1	103	010313	470363.199	1030103

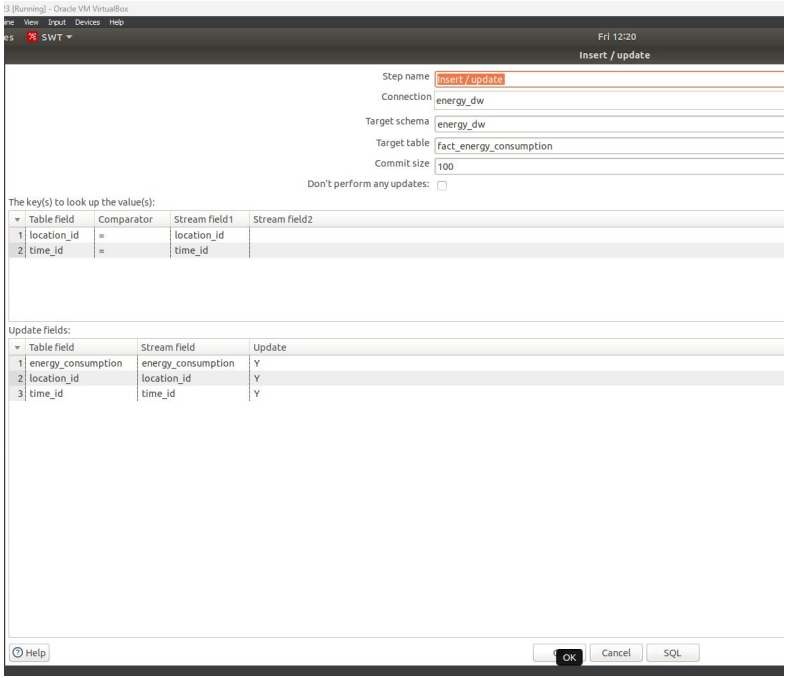
Database lookup 2

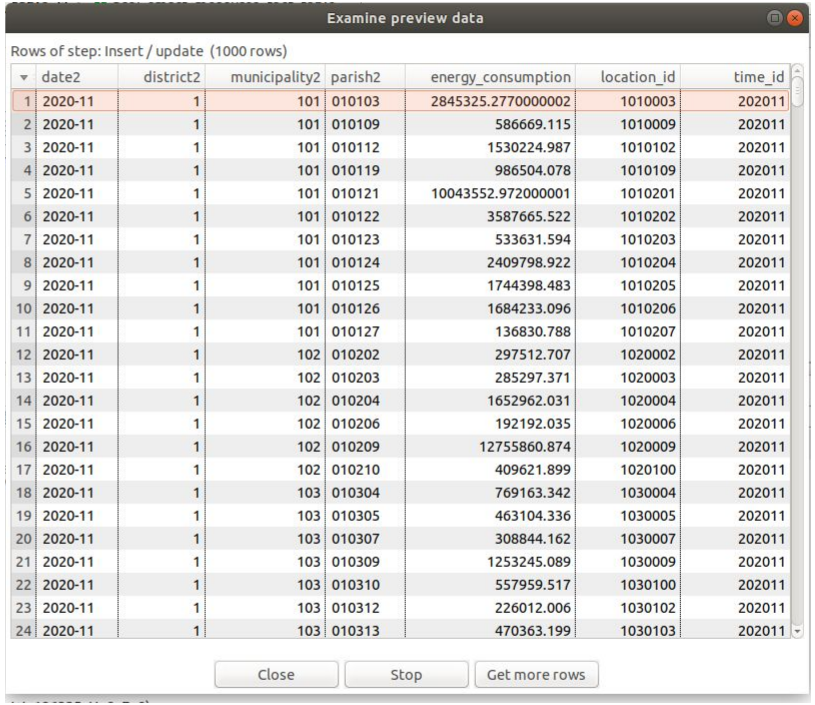




	date2	district2	municipality2	parish2	energy_consumption	location_id	time_id
1	2020-11	1	101	010103	2845325.2770000002	1010003	202011
2	2020-11	1	101	010109	586669.115	1010009	202011
3	2020-11	1	101	010112	1530224.987	1010102	202011
4	2020-11	1	101	010119	986504.078	1010109	202011
5	2020-11	1	101	010121	10043552.972000001	1010201	202011
6	2020-11	1	101	010122	3587665.522	1010202	202011
7	2020-11	1	101	010123	533631.594	1010203	202011
8	2020-11	1	101	010124	2409798.922	1010204	202011
9	2020-11	1	101	010125	1744398.483	1010205	202011
10	2020-11	1	101	010126	1684233.096	1010206	202011
11	2020-11	1	101	010127	136830.788	1010207	202011
12	2020-11	1	102	010202	297512.707	1020002	202011
13	2020-11	1	102	010203	285297.371	1020003	202011
14	2020-11	1	102	010204	1652962.031	1020004	202011
15	2020-11	1	102	010206	192192.035	1020006	202011
16	2020-11	1	102	010209	12755860.874	1020009	202011
17	2020-11	1	102	010210	409621.899	1020100	202011
18	2020-11	1	103	010304	769163.342	1030004	202011
19	2020-11	1	103	010305	463104.336	1030005	202011
20	2020-11	1	103	010307	308844.162	1030007	202011
21	2020-11	1	103	010309	1253245.089	1030009	202011
22	2020-11	1	103	010310	557959.517	1030100	202011
23	2020-11	1	103	010312	226012.006	1030102	202011
24	2020-11	1	103	010313	470363.199	1030103	202011

Insert/Update





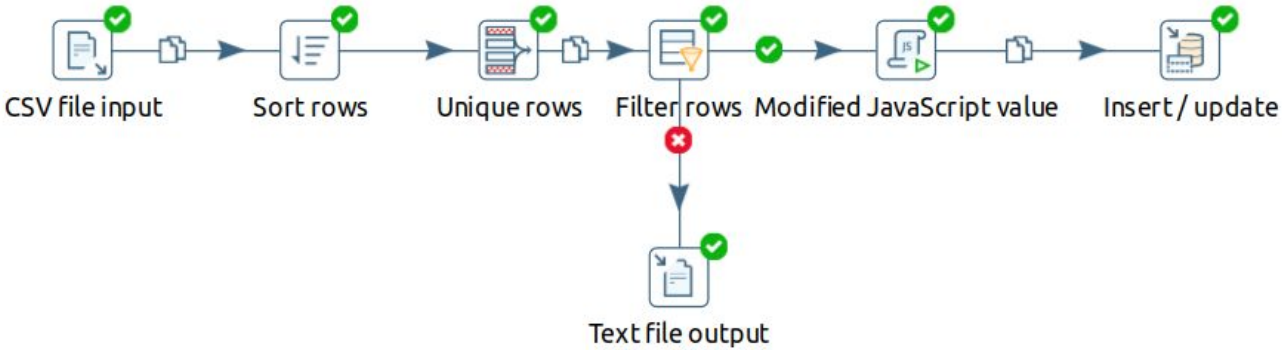
	date2	district2	municipality2	parish2	energy_consumption	location_id	time_id
1	2020-11	1	101	010103	2845325.2770000002	1010003	202011
2	2020-11	1	101	010109	586669.115	1010009	202011
3	2020-11	1	101	010112	1530224.987	1010102	202011
4	2020-11	1	101	010119	986504.078	1010109	202011
5	2020-11	1	101	010121	10043552.972000001	1010201	202011
6	2020-11	1	101	010122	3587665.522	1010202	202011
7	2020-11	1	101	010123	533631.594	1010203	202011
8	2020-11	1	101	010124	2409798.922	1010204	202011
9	2020-11	1	101	010125	1744398.483	1010205	202011
10	2020-11	1	101	010126	1684233.096	1010206	202011
11	2020-11	1	101	010127	136830.788	1010207	202011
12	2020-11	1	102	010202	297512.707	1020002	202011
13	2020-11	1	102	010203	285297.371	1020003	202011
14	2020-11	1	102	010204	1652962.031	1020004	202011
15	2020-11	1	102	010206	192192.035	1020006	202011
16	2020-11	1	102	010209	12755860.874	1020009	202011
17	2020-11	1	102	010210	409621.899	1020100	202011
18	2020-11	1	103	010304	769163.342	1030004	202011
19	2020-11	1	103	010305	463104.336	1030005	202011
20	2020-11	1	103	010307	308844.162	1030007	202011
21	2020-11	1	103	010309	1253245.089	1030009	202011
22	2020-11	1	103	010310	557959.517	1030100	202011
23	2020-11	1	103	010312	226012.006	1030102	202011
24	2020-11	1	103	010313	470363.199	1030103	202011

Output Table

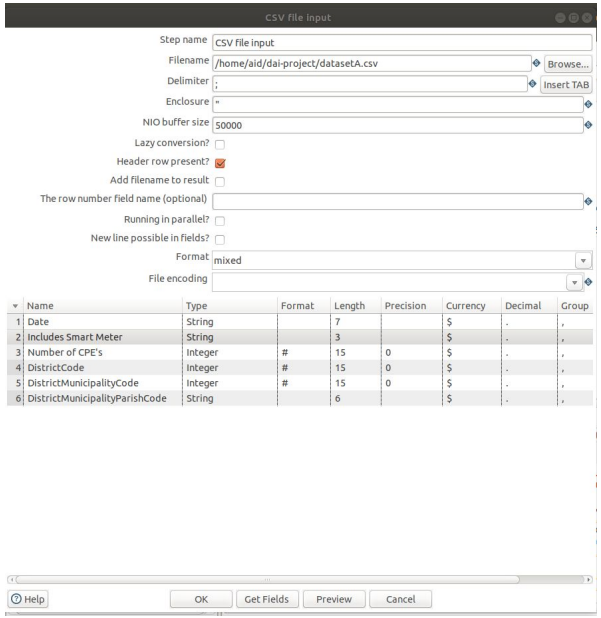
```
mysql> select * from fact_energy_consumption limit 25;
+---------+-------------+--------------------+
| time_id | location_id | energy_consumption |
+---------+-------------+--------------------+
|  202011 |     1010003 |            2845330 |
|  202011 |     1010009 |             586669 |
|  202011 |     1010102 |            1530220 |
|  202011 |     1010109 |             986504 |
|  202011 |     1010201 |           10043600 |
|  202011 |     1010202 |            3587670 |
|  202011 |     1010203 |             533632 |
|  202011 |     1010204 |            2409800 |
|  202011 |     1010205 |            1744400 |
|  202011 |     1010206 |            1684230 |
|  202011 |     1010207 |             136831 |
|  202011 |     1020002 |             297513 |
|  202011 |     1020003 |             285297 |
|  202011 |     1020004 |            1652960 |
|  202011 |     1020006 |             192192 |
|  202011 |     1020009 |           12755900 |
|  202011 |     1020100 |             409622 |
|  202011 |     1030004 |             769163 |
|  202011 |     1030005 |             463104 |
|  202011 |     1030007 |             308844 |
|  202011 |     1030009 |            1253250 |
|  202011 |     1030100 |             557960 |
|  202011 |     1030102 |             226012 |
|  202011 |     1030103 |             470363 |
|  202011 |     1030106 |             633762 |
+---------+-------------+--------------------+
25 rows in set (0.00 sec)
```

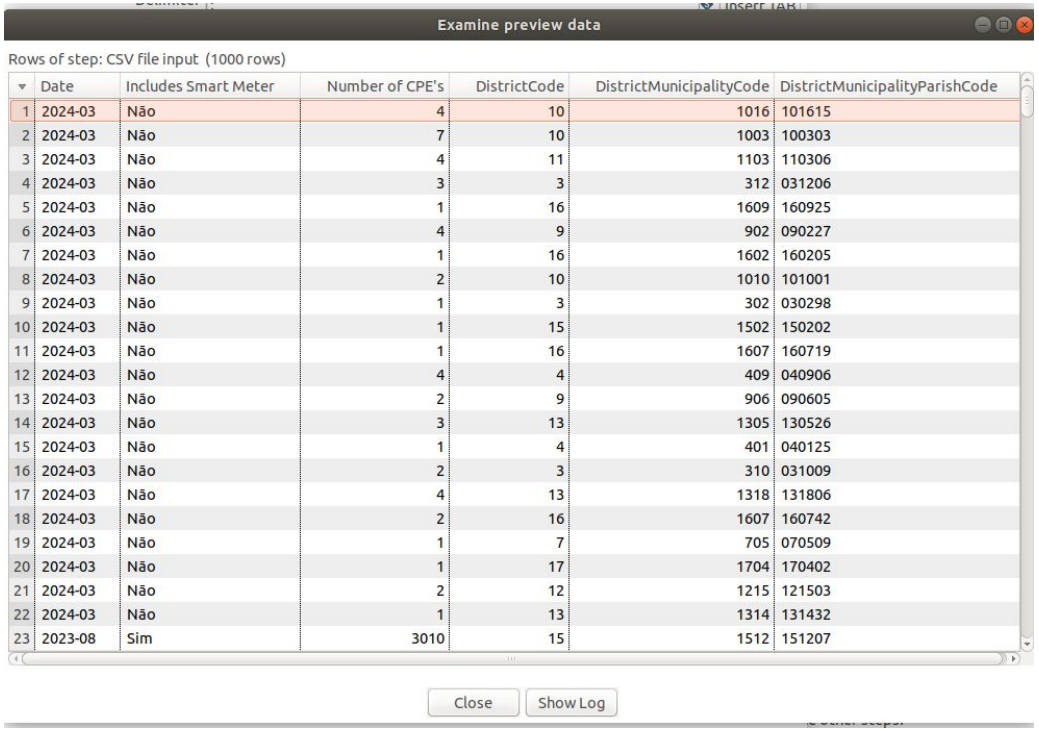
Smart Measures Fact Table

Transformation:



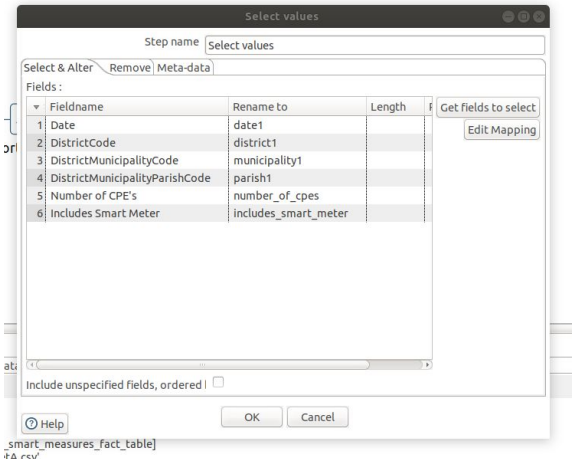
CSV file input

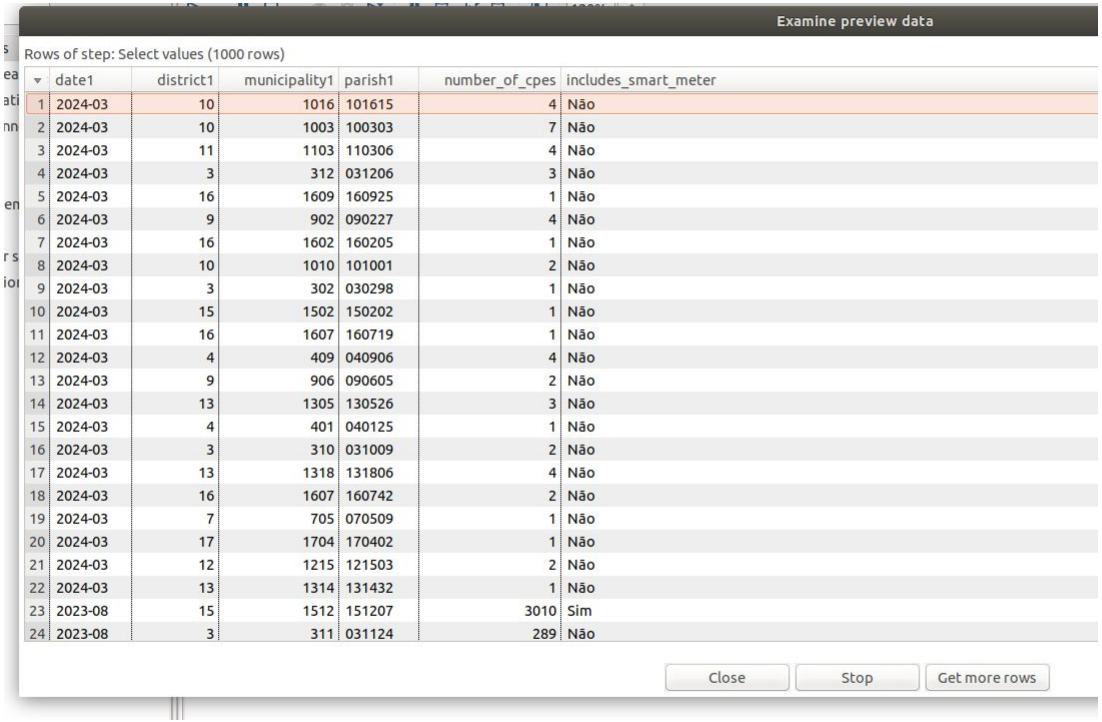




	Date	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode
1	2024-03	Não	4	10	1016	101615
2	2024-03	Não	7	10	1003	100303
3	2024-03	Não	4	11	1103	110306
4	2024-03	Não	3	3	312	031206
5	2024-03	Não	1	16	1609	160925
6	2024-03	Não	4	9	902	090227
7	2024-03	Não	1	16	1602	160205
8	2024-03	Não	2	10	1010	101001
9	2024-03	Não	1	3	302	030298
10	2024-03	Não	1	15	1502	150202
11	2024-03	Não	1	16	1607	160719
12	2024-03	Não	4	4	409	040906
13	2024-03	Não	2	9	906	090605
14	2024-03	Não	3	13	1305	130526
15	2024-03	Não	1	4	401	040125
16	2024-03	Não	2	3	310	031009
17	2024-03	Não	4	13	1318	131806
18	2024-03	Não	2	16	1607	160742
19	2024-03	Não	1	7	705	070509
20	2024-03	Não	1	17	1704	170402
21	2024-03	Não	2	12	1215	121503
22	2024-03	Não	1	13	1314	131432
23	2023-08	Sim	3010	15	1512	151207

Select values

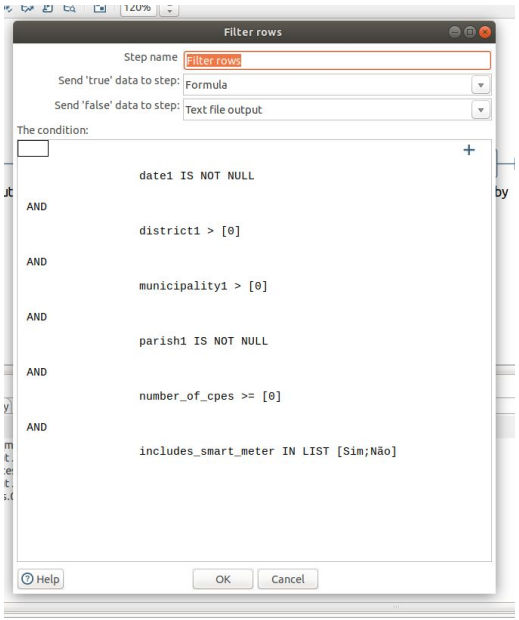


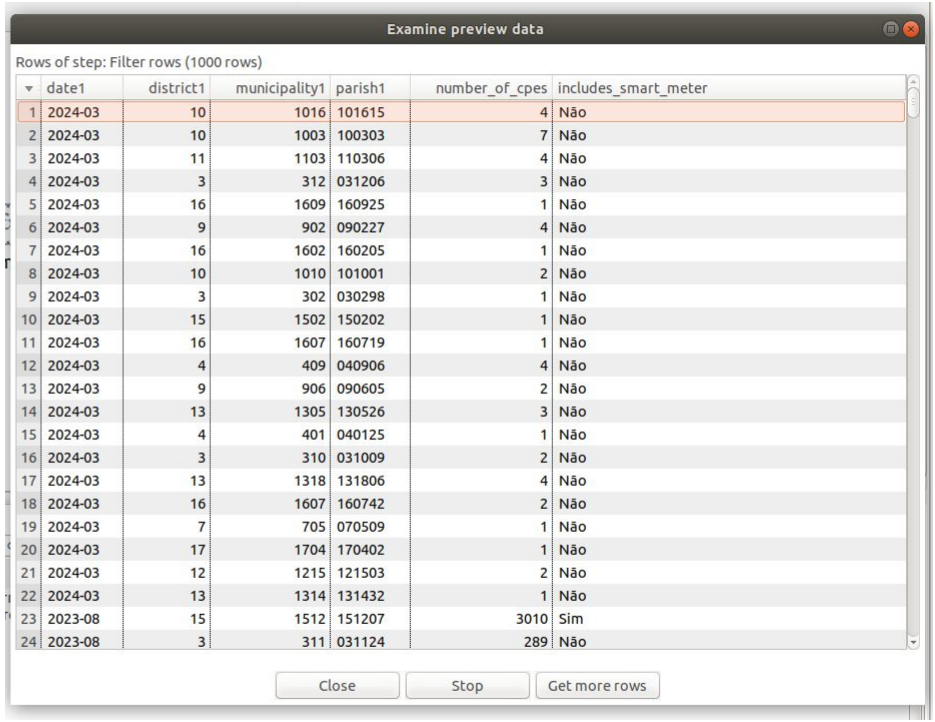


	date1	district1	municipality1	parish1	number_of_cpes	includes_smart_meter
1	2024-03	10	1016	101615	4	Não
2	2024-03	10	1003	100303	7	Não
3	2024-03	11	1103	110306	4	Não
4	2024-03	3	312	031206	3	Não
5	2024-03	16	1609	160925	1	Não
6	2024-03	9	902	090227	4	Não
7	2024-03	16	1602	160205	1	Não
8	2024-03	10	1010	101001	2	Não
9	2024-03	3	302	030298	1	Não
10	2024-03	15	1502	150202	1	Não
11	2024-03	16	1607	160719	1	Não
12	2024-03	4	409	040906	4	Não
13	2024-03	9	906	090605	2	Não
14	2024-03	13	1305	130526	3	Não
15	2024-03	4	401	040125	1	Não
16	2024-03	3	310	031009	2	Não
17	2024-03	13	1318	131806	4	Não
18	2024-03	16	1607	160742	2	Não
19	2024-03	7	705	070509	1	Não
20	2024-03	17	1704	170402	1	Não
21	2024-03	12	1215	121503	2	Não
22	2024-03	13	1314	131432	1	Não
23	2023-08	15	1512	151207	3010	Sim
24	2023-08	3	311	031124	289	Não

Filter rows

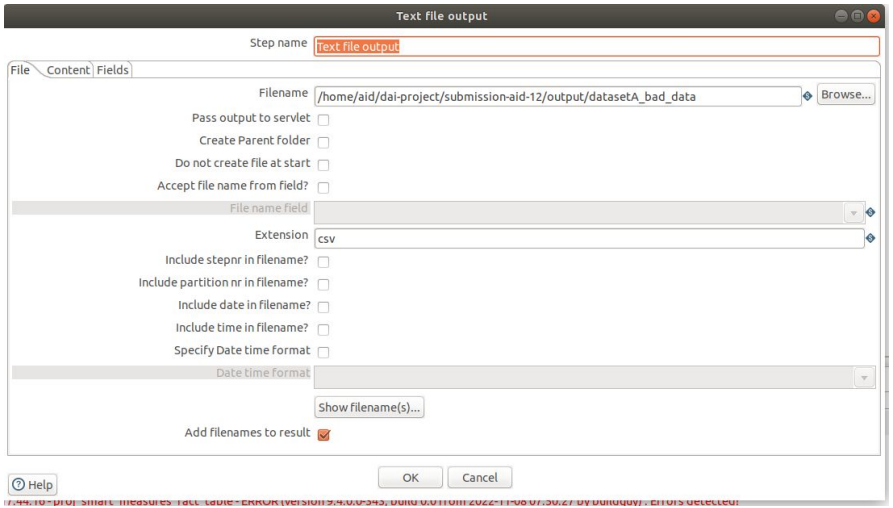
The fields with null parish code due to GDPR (or other irregularities) were filtered and sent to an output file with the purpose of further analysis and transformation.

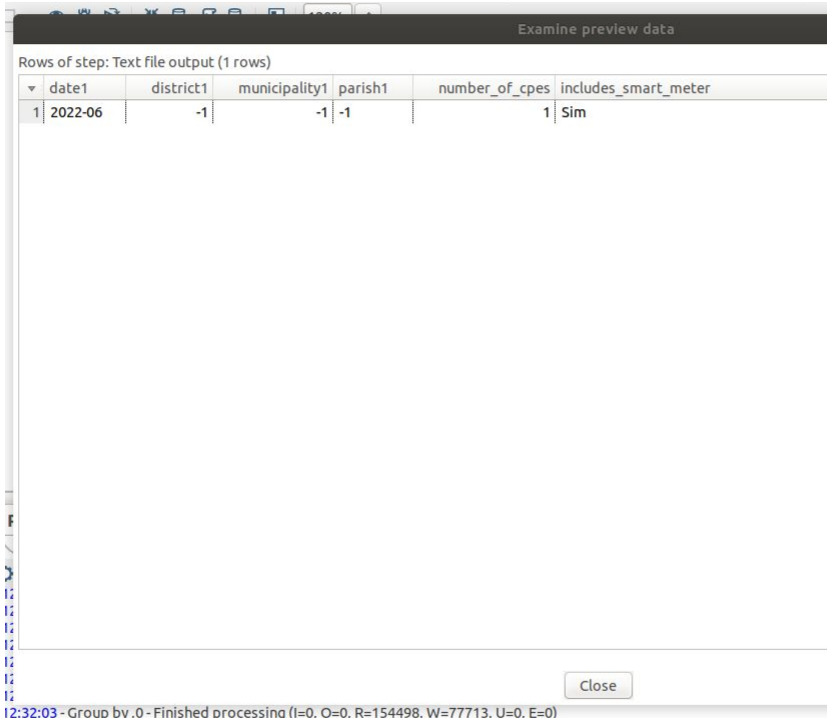




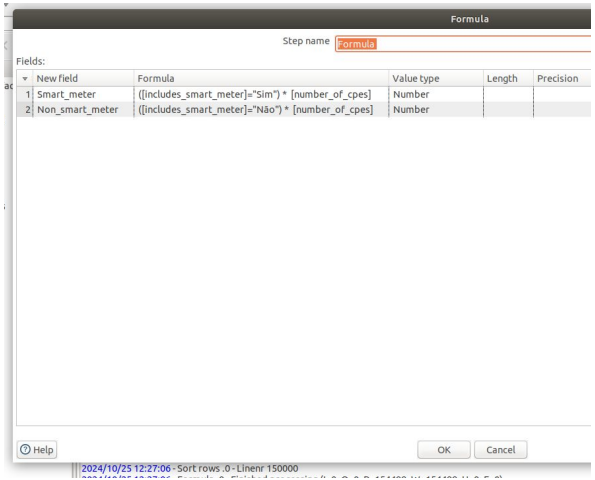
	date1	district1	municipality1	parish1	number_of_cpes	includes_smart_meter
1	2024-03	10	1016	101615	4	Não
2	2024-03	10	1003	100303	7	Não
3	2024-03	11	1103	110306	4	Não
4	2024-03	3	312	031206	3	Não
5	2024-03	16	1609	160925	1	Não
6	2024-03	9	902	090227	4	Não
7	2024-03	16	1602	160205	1	Não
8	2024-03	10	1010	101001	2	Não
9	2024-03	3	302	030298	1	Não
10	2024-03	15	1502	150202	1	Não
11	2024-03	16	1607	160719	1	Não
12	2024-03	4	409	040906	4	Não
13	2024-03	9	906	090605	2	Não
14	2024-03	13	1305	130526	3	Não
15	2024-03	4	401	040125	1	Não
16	2024-03	3	310	031009	2	Não
17	2024-03	13	1318	131806	4	Não
18	2024-03	16	1607	160742	2	Não
19	2024-03	7	705	070509	1	Não
20	2024-03	17	1704	170402	1	Não
21	2024-03	12	1215	121503	2	Não
22	2024-03	13	1314	131432	1	Não
23	2023-08	15	1512	151207	3010	Sim
24	2023-08	3	311	031124	289	Não

Text file output





Formula



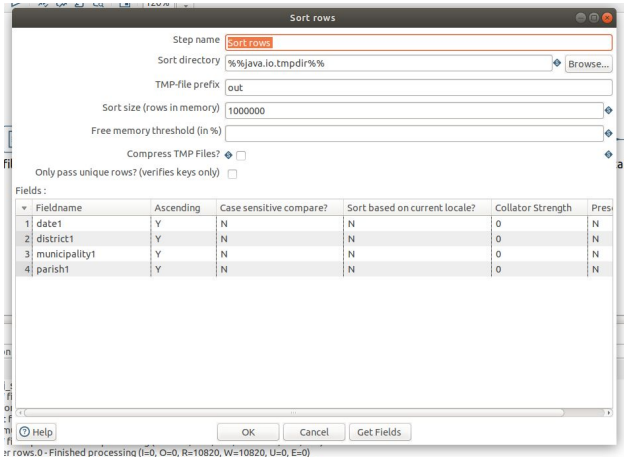
Examine preview data

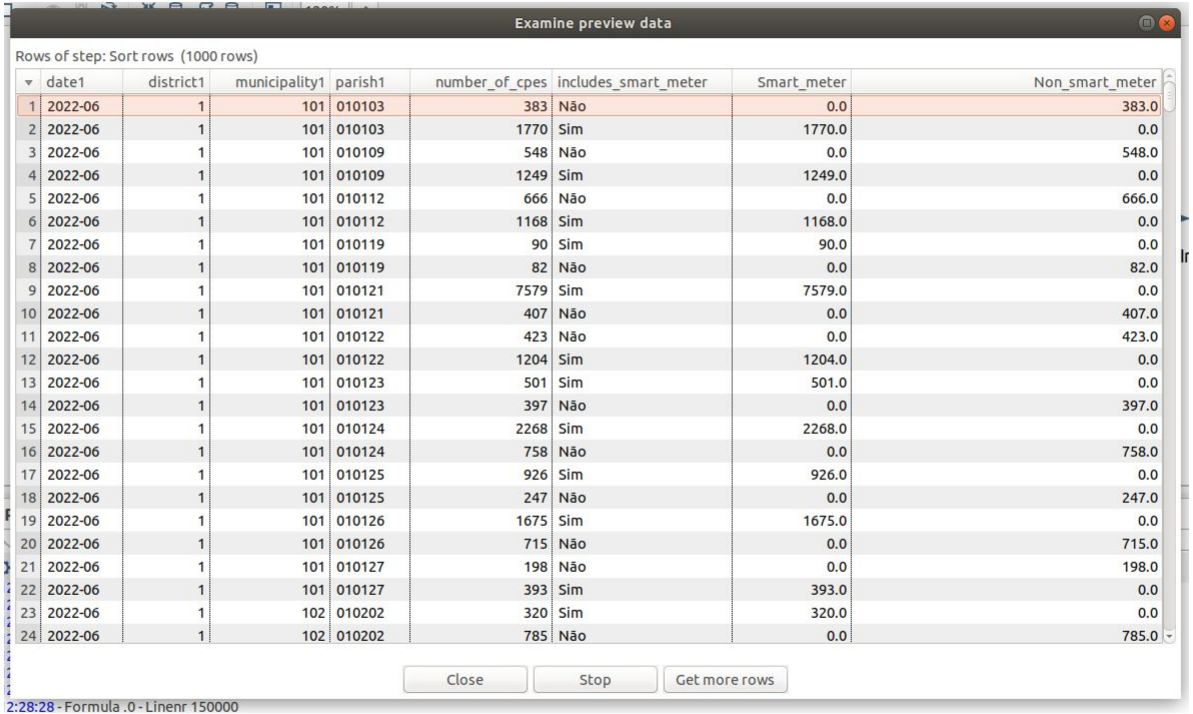
Rows of step: Formula (1000 rows)

	date1	district1	municipality1	parish1	number_of_cpes	includes_smart_meter	Smart_meter	Non_smart_meter
1	2024-03	10	1016	101615	4	Não	0.0	4.0
2	2024-03	10	1003	100303	7	Não	0.0	7.0
3	2024-03	11	1103	110306	4	Não	0.0	4.0
4	2024-03	3	312	031206	3	Não	0.0	3.0
5	2024-03	16	1609	160925	1	Não	0.0	1.0
6	2024-03	9	902	090227	4	Não	0.0	4.0
7	2024-03	16	1602	160205	1	Não	0.0	1.0
8	2024-03	10	1010	101001	2	Não	0.0	2.0
9	2024-03	3	302	030298	1	Não	0.0	1.0
10	2024-03	15	1502	150202	1	Não	0.0	1.0
11	2024-03	16	1607	160719	1	Não	0.0	1.0
12	2024-03	4	409	040906	4	Não	0.0	4.0
13	2024-03	9	906	090605	2	Não	0.0	2.0
14	2024-03	13	1305	130526	3	Não	0.0	3.0
15	2024-03	4	401	040125	1	Não	0.0	1.0
16	2024-03	3	310	031009	2	Não	0.0	2.0
17	2024-03	13	1318	131806	4	Não	0.0	4.0
18	2024-03	16	1607	160742	2	Não	0.0	2.0
19	2024-03	7	705	070509	1	Não	0.0	1.0
20	2024-03	17	1704	170402	1	Não	0.0	1.0
21	2024-03	12	1215	121503	2	Não	0.0	2.0
22	2024-03	13	1314	131432	1	Não	0.0	1.0
23	2023-08	15	1512	151207	3010	Sim	3010.0	0.0
24	2023-08	3	311	031124	289	Não	0.0	289.0

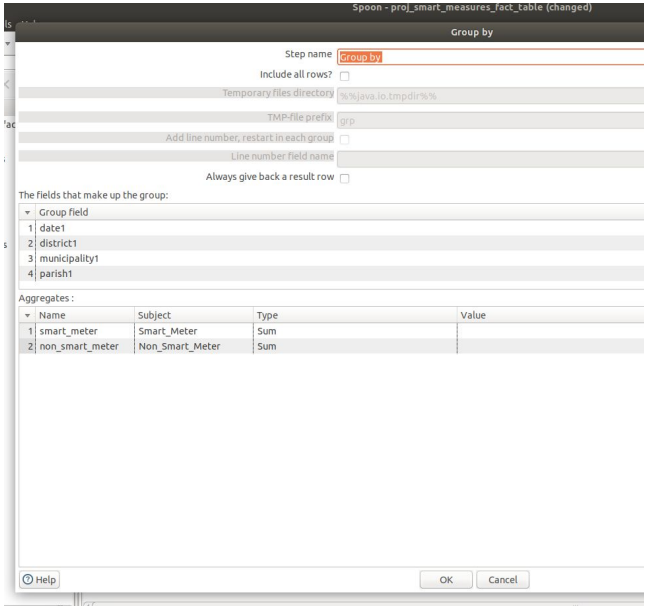
Close | Stop | Get more rows

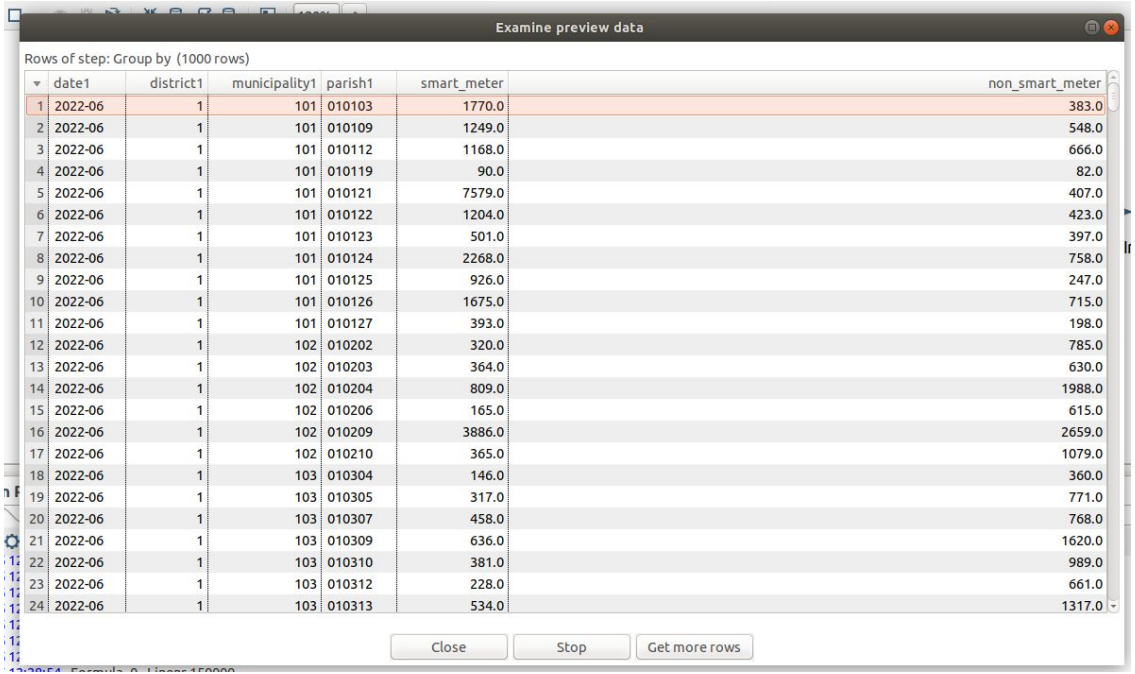
Sort rows


Sort rows
Step name
Sort rows
Sort directory
%%java.io.tmpdir%%
Browse...
TMP-file prefix
out
Sort size (rows in memory)
1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys only)
Fields :
Fieldname
Ascending
Case sensitive compare?
Sort based on current locale?
Collator Strength
1 date1 Y N N 0 N
2 district1 Y N N 0 N
3 municipality1 Y N N 0 N
4 parish1 Y N N 0 N
Help
OK
Cancel
Get Fields

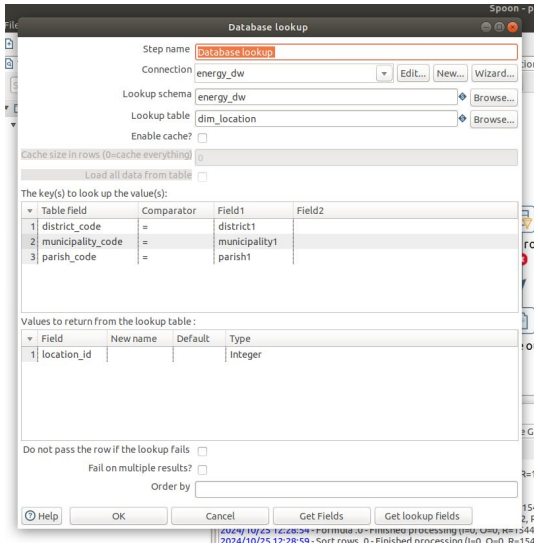

Examine preview data
Rows of step: Sort rows (1000 rows)
date1 district1 municipality1 parish1 number_of_cpes includes_smart_meter Smart_meter Non_smart_meter
1 2022-06 1 101 010103 383 Não 0.0 383.0
2 2022-06 1 101 010103 1770 Sim 1770.0 0.0
3 2022-06 1 101 010109 548 Não 0.0 548.0
4 2022-06 1 101 010109 1249 Sim 1249.0 0.0
5 2022-06 1 101 010112 666 Não 0.0 666.0
6 2022-06 1 101 010112 1168 Sim 1168.0 0.0
7 2022-06 1 101 010119 90 Sim 90.0 0.0
8 2022-06 1 101 010119 82 Não 0.0 82.0
9 2022-06 1 101 010121 7579 Sim 7579.0 0.0
10 2022-06 1 101 010121 407 Não 0.0 407.0
11 2022-06 1 101 010122 423 Não 0.0 423.0
12 2022-06 1 101 010122 1204 Sim 1204.0 0.0
13 2022-06 1 101 010123 501 Sim 501.0 0.0
14 2022-06 1 101 010123 397 Não 0.0 397.0
15 2022-06 1 101 010124 2268 Sim 2268.0 0.0
16 2022-06 1 101 010124 758 Não 0.0 758.0
17 2022-06 1 101 010125 926 Sim 926.0 0.0
18 2022-06 1 101 010125 247 Não 0.0 247.0
19 2022-06 1 101 010126 1675 Sim 1675.0 0.0
20 2022-06 1 101 010126 715 Não 0.0 715.0
21 2022-06 1 101 010127 198 Não 0.0 198.0
22 2022-06 1 101 010127 393 Sim 393.0 0.0
23 2022-06 1 102 010202 320 Sim 320.0 0.0
24 2022-06 1 102 010202 785 Não 0.0 785.0
Close
Stop
Get more rows
2:28:28 - Formula .0 - Linenr 150000

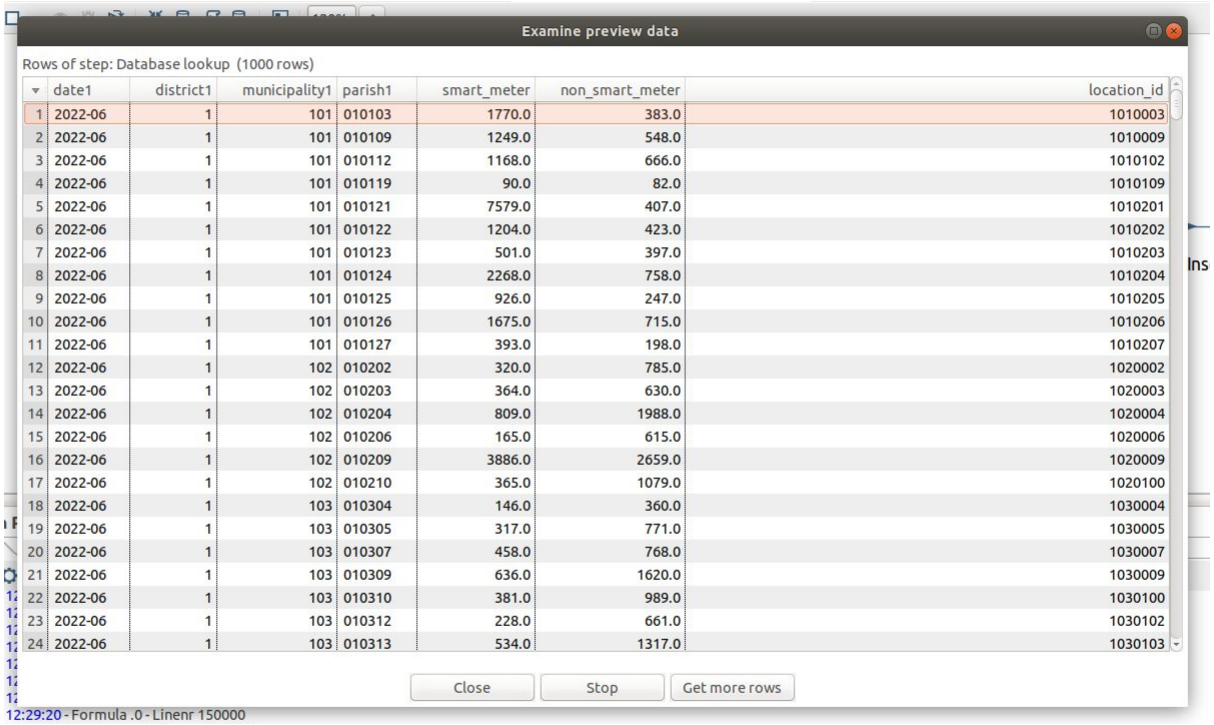
Group by


Spoon - proj_smart_measures_fact_table (changed)
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 date1
2 district1
3 municipality1
4 parish1
Aggregates :
Name Subject Type Value
1 smart_meter Smart_Meter Sum
2 non_smart_meter Non_Smart_Meter Sum
Help
OK
Cancel


Examine preview data
Rows of step: Group by (1000 rows)
date1 district1 municipality1 parish1 smart_meter non_smart_meter
1 2022-06 1 101 010103 1770.0 383.0
2 2022-06 1 101 010109 1249.0 548.0
3 2022-06 1 101 010112 1168.0 666.0
4 2022-06 1 101 010119 90.0 82.0
5 2022-06 1 101 010121 7579.0 407.0
6 2022-06 1 101 010122 1204.0 423.0
7 2022-06 1 101 010123 501.0 397.0
8 2022-06 1 101 010124 2268.0 758.0
9 2022-06 1 101 010125 926.0 247.0
10 2022-06 1 101 010126 1675.0 715.0
11 2022-06 1 101 010127 393.0 198.0
12 2022-06 1 102 010202 320.0 785.0
13 2022-06 1 102 010203 364.0 630.0
14 2022-06 1 102 010204 809.0 1988.0
15 2022-06 1 102 010206 165.0 615.0
16 2022-06 1 102 010209 3886.0 2659.0
17 2022-06 1 102 010210 365.0 1079.0
18 2022-06 1 103 010304 146.0 360.0
19 2022-06 1 103 010305 317.0 771.0
20 2022-06 1 103 010307 458.0 768.0
21 2022-06 1 103 010309 636.0 1620.0
22 2022-06 1 103 010310 381.0 989.0
23 2022-06 1 103 010312 228.0 661.0
24 2022-06 1 103 010313 534.0 1317.0
Close
Stop
Get more rows

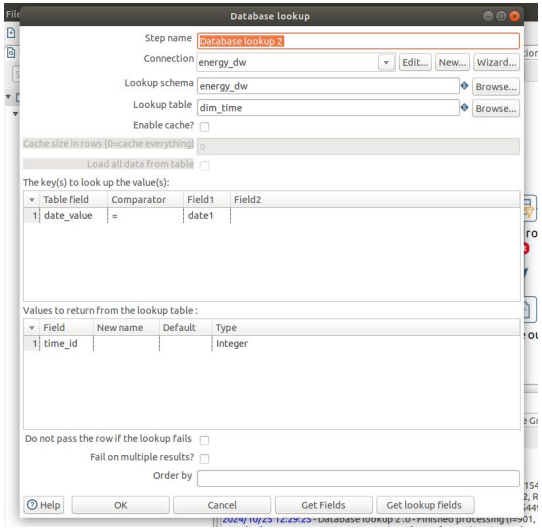
Database lookup

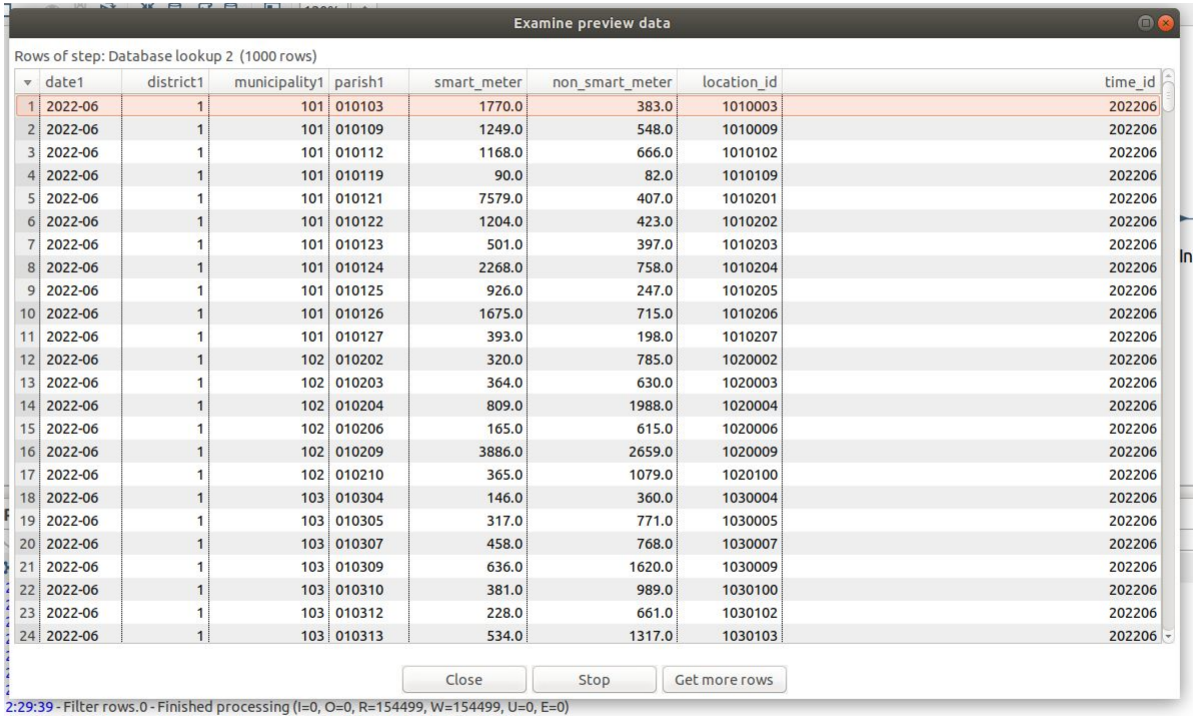




	date1	district1	municipality1	parish1	smart_meter	non_smart_meter	location_id
1	2022-06	1	101	010103	1770.0	383.0	1010003
2	2022-06	1	101	010109	1249.0	548.0	1010009
3	2022-06	1	101	010112	1168.0	666.0	1010102
4	2022-06	1	101	010119	90.0	82.0	1010109
5	2022-06	1	101	010121	7579.0	407.0	1010201
6	2022-06	1	101	010122	1204.0	423.0	1010202
7	2022-06	1	101	010123	501.0	397.0	1010203
8	2022-06	1	101	010124	2268.0	758.0	1010204
9	2022-06	1	101	010125	926.0	247.0	1010205
10	2022-06	1	101	010126	1675.0	715.0	1010206
11	2022-06	1	101	010127	393.0	198.0	1010207
12	2022-06	1	102	010202	320.0	785.0	1020002
13	2022-06	1	102	010203	364.0	630.0	1020003
14	2022-06	1	102	010204	809.0	1988.0	1020004
15	2022-06	1	102	010206	165.0	615.0	1020006
16	2022-06	1	102	010209	3886.0	2659.0	1020009
17	2022-06	1	102	010210	365.0	1079.0	1020100
18	2022-06	1	103	010304	146.0	360.0	1030004
19	2022-06	1	103	010305	317.0	771.0	1030005
20	2022-06	1	103	010307	458.0	768.0	1030007
21	2022-06	1	103	010309	636.0	1620.0	1030009
22	2022-06	1	103	010310	381.0	989.0	1030100
23	2022-06	1	103	010312	228.0	661.0	1030102
24	2022-06	1	103	010313	534.0	1317.0	1030103

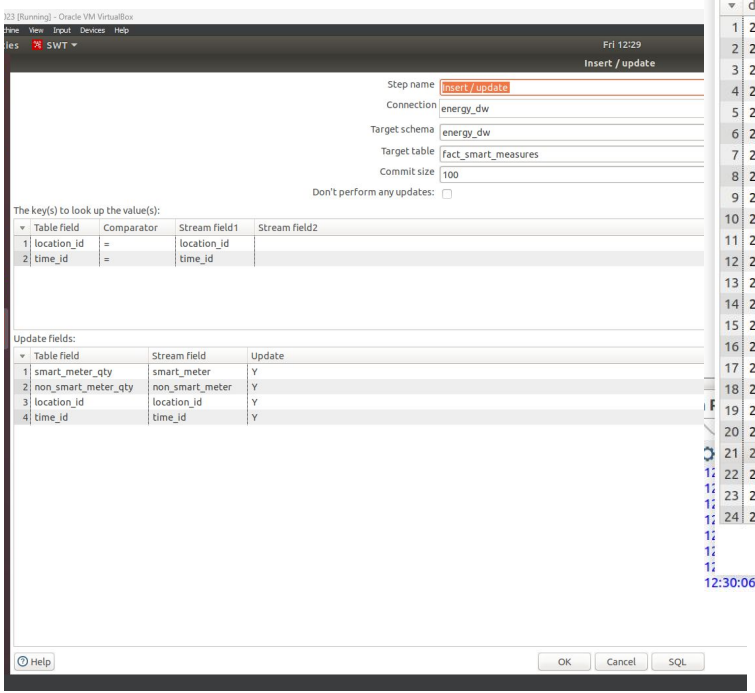
Database lookup 2

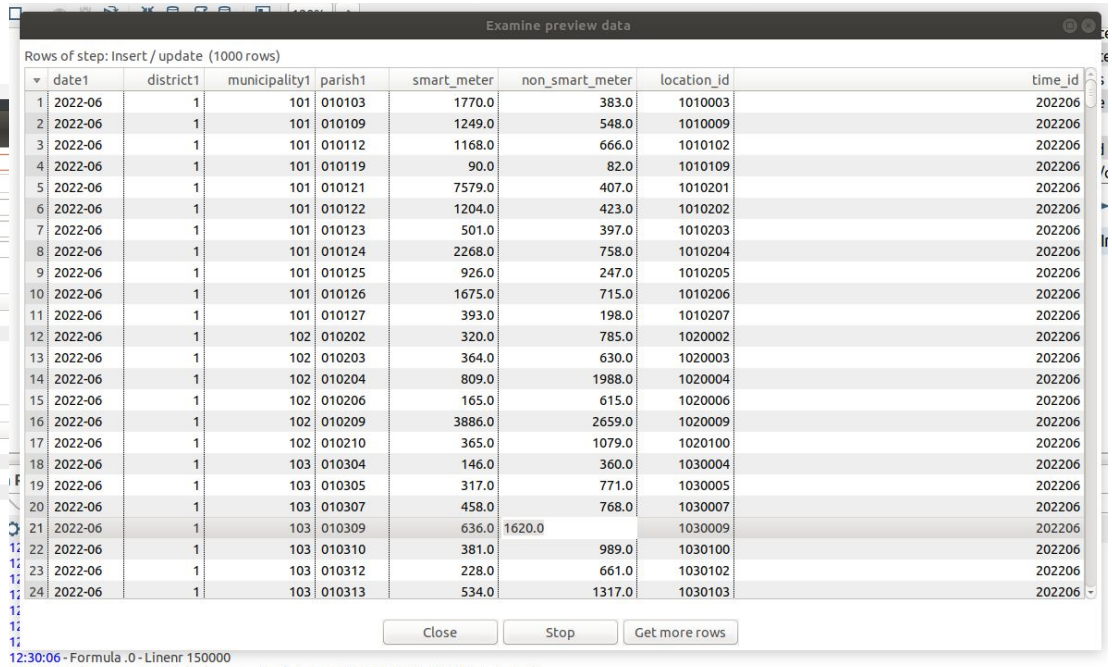




	date1	district1	municipality1	parish1	smart_meter	non_smart_meter	location_id	time_id
1	2022-06	1	101	010103	1770.0	383.0	1010003	202206
2	2022-06	1	101	010109	1249.0	548.0	1010009	202206
3	2022-06	1	101	010112	1168.0	666.0	1010102	202206
4	2022-06	1	101	010119	90.0	82.0	1010109	202206
5	2022-06	1	101	010121	7579.0	407.0	1010201	202206
6	2022-06	1	101	010122	1204.0	423.0	1010202	202206
7	2022-06	1	101	010123	501.0	397.0	1010203	202206
8	2022-06	1	101	010124	2268.0	758.0	1010204	202206
9	2022-06	1	101	010125	926.0	247.0	1010205	202206
10	2022-06	1	101	010126	1675.0	715.0	1010206	202206
11	2022-06	1	101	010127	393.0	198.0	1010207	202206
12	2022-06	1	102	010202	320.0	785.0	1020002	202206
13	2022-06	1	102	010203	364.0	630.0	1020003	202206
14	2022-06	1	102	010204	809.0	1988.0	1020004	202206
15	2022-06	1	102	010206	165.0	615.0	1020006	202206
16	2022-06	1	102	010209	3886.0	2659.0	1020009	202206
17	2022-06	1	102	010210	365.0	1079.0	1020100	202206
18	2022-06	1	103	010304	146.0	360.0	1030004	202206
19	2022-06	1	103	010305	317.0	771.0	1030005	202206
20	2022-06	1	103	010307	458.0	768.0	1030007	202206
21	2022-06	1	103	010309	636.0	1620.0	1030009	202206
22	2022-06	1	103	010310	381.0	989.0	1030100	202206
23	2022-06	1	103	010312	228.0	661.0	1030102	202206
24	2022-06	1	103	010313	534.0	1317.0	1030103	202206

Insert/update





Output Table

```
mysql> select * from fact_smart_measures limit 25;
+---------+-------------+-----------------+---------------------+
| time_id | location_id | smart_meter_qty | non_smart_meter_qty |
+---------+-------------+-----------------+---------------------+
|  202206 |     1010003 |            1770 |                 383 |
|  202206 |     1010009 |            1249 |                 548 |
|  202206 |     1010102 |            1168 |                 666 |
|  202206 |     1010109 |              90 |                  82 |
|  202206 |     1010201 |            7579 |                 407 |
|  202206 |     1010202 |            1204 |                 423 |
|  202206 |     1010203 |             501 |                 397 |
|  202206 |     1010204 |            2268 |                 758 |
|  202206 |     1010205 |             926 |                 247 |
|  202206 |     1010206 |            1675 |                 715 |
|  202206 |     1010207 |             393 |                 198 |
|  202206 |     1020002 |             320 |                 785 |
|  202206 |     1020003 |             364 |                 630 |
|  202206 |     1020004 |             809 |                1988 |
|  202206 |     1020006 |             165 |                 615 |
|  202206 |     1020009 |            3886 |                2659 |
|  202206 |     1020100 |             365 |                1079 |
|  202206 |     1030004 |             146 |                 360 |
|  202206 |     1030005 |             317 |                 771 |
|  202206 |     1030007 |             458 |                 768 |
|  202206 |     1030009 |             636 |                1620 |
|  202206 |     1030100 |             381 |                 989 |
|  202206 |     1030102 |             228 |                 661 |
|  202206 |     1030103 |             534 |                1317 |
|  202206 |     1030106 |             419 |                1174 |
+---------+-------------+-----------------+---------------------+
25 rows in set (0.00 sec)
```

Given this fact table would be used to build a cube, storing smart measures as percentage did not seem the most correct choice.

With cube aggregation functions this measure would be distorted (it is a non-additive measure) even using the mean or maximum for example, would not assess the reality and could lead to underestimation or overestimation of data from different locations.

Task 7

Define an OLAP cube based on this data warehouse.

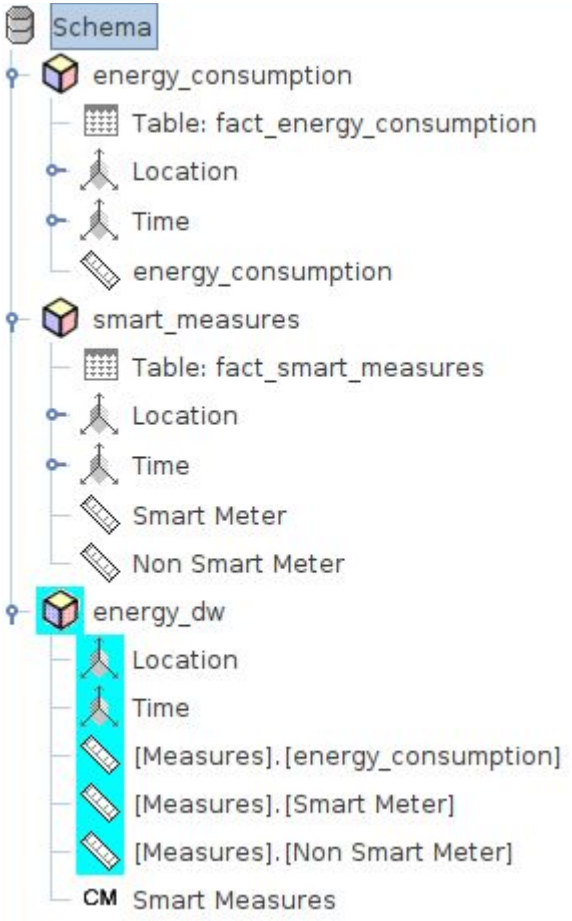
Pentaho only supports star schema based cubes. To outmaneuver this, virtual cubes appear as an option to build a cube based on a constellation schema, that aggregates the measures with common dimensions from the two cubes built based on both fact tables.

OLAP Cube

```
1   <Schema name="energy_dw">
2   >  <Cube name="energy_consumption" visible="true" cache="true" enabled="true">···
31     </Cube>
32  >  <Cube name="smart_measures" visible="true" cache="true" enabled="true">···
63     </Cube>
64  >  <VirtualCube enabled="true" name="energy_dw" visible="true">···
77     </VirtualCube>
78  </Schema>
```

Since our XML file became too large for a single slide, we will split the screenshots into three parts:

1. energy_consumption cube
2. smart_measures cube
3. energy_dw virtual cube (used for the analyses)



Energy Consumption Cube

```
<Cube name="energy_consumption" visible="true" cache="true" enabled="true">
  <Table name="fact_energy_consumption">
  </Table>
  <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" ordinalColumn="district_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Municipality" visible="true" column="municipality" ordinalColumn="municipality_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Parish" visible="true" column="parish" ordinalColumn="parish_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <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="Season" visible="true" column="season_name" ordinalColumn="season_id" type="String" uniqueMembers="false" levelType="TimeUndefined" hideMemberIf="Never">
      </Level>
      <Level name="Month" visible="true" column="month_name" ordinalColumn="month_id" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Measure name="energy_consumption" column="energy_consumption" datatype="Numeric" aggregator="sum" visible="true">
  </Measure>
</Cube>
```

Smart Measures Cube

```
<Cube name="smart_measures" visible="true" cache="true" enabled="true">
  <Table name="fact_smart_measures">
  </Table>
  <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" ordinalColumn="district_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Municipality" visible="true" column="municipality" ordinalColumn="municipality_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
      <Level name="Parish" visible="true" column="parish" ordinalColumn="parish_code" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <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="Season" visible="true" column="season_name" ordinalColumn="season_id" type="String" uniqueMembers="false" levelType="TimeUndefined" hideMemberIf="Never">
      </Level>
      <Level name="Month" visible="true" column="month_name" ordinalColumn="month_id" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Measure name="Smart Meter" column="smart_meter_qty" datatype="Numeric" aggregator="sum" visible="true">
  </Measure>
  <Measure name="Non Smart Meter" column="non_smart_meter_qty" aggregator="sum" visible="true">
  </Measure>
</Cube>
```

Energy DW Virtual Cube

```
<VirtualCube enabled="true" name="energy_dw" visible="true">
  <VirtualCubeDimension cubeName="energy_consumption" visible="true" foreignKey="location_id" highCardinality="false" name="Location">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="energy_consumption" visible="true" foreignKey="time_id" highCardinality="false" name="Time">
  </VirtualCubeDimension>
  <VirtualCubeMeasure cubeName="energy_consumption" name="[Measures].[energy_consumption]" visible="true">
  </VirtualCubeMeasure>
  <VirtualCubeMeasure cubeName="smart_measures" name="[Measures].[Smart Meter]" visible="true">
  </VirtualCubeMeasure>
  <VirtualCubeMeasure cubeName="smart_measures" name="[Measures].[Non Smart Meter]" visible="true">
  </VirtualCubeMeasure>
  <CalculatedMember name="Smart Measures" formula="[Measures].[Smart Meter] / ([Measures].[Smart Meter] + [Measures].[Non Smart Meter]) * 100" dimension="Measures" visible="true">
  </CalculatedMember>
</VirtualCube>
```

Task 8

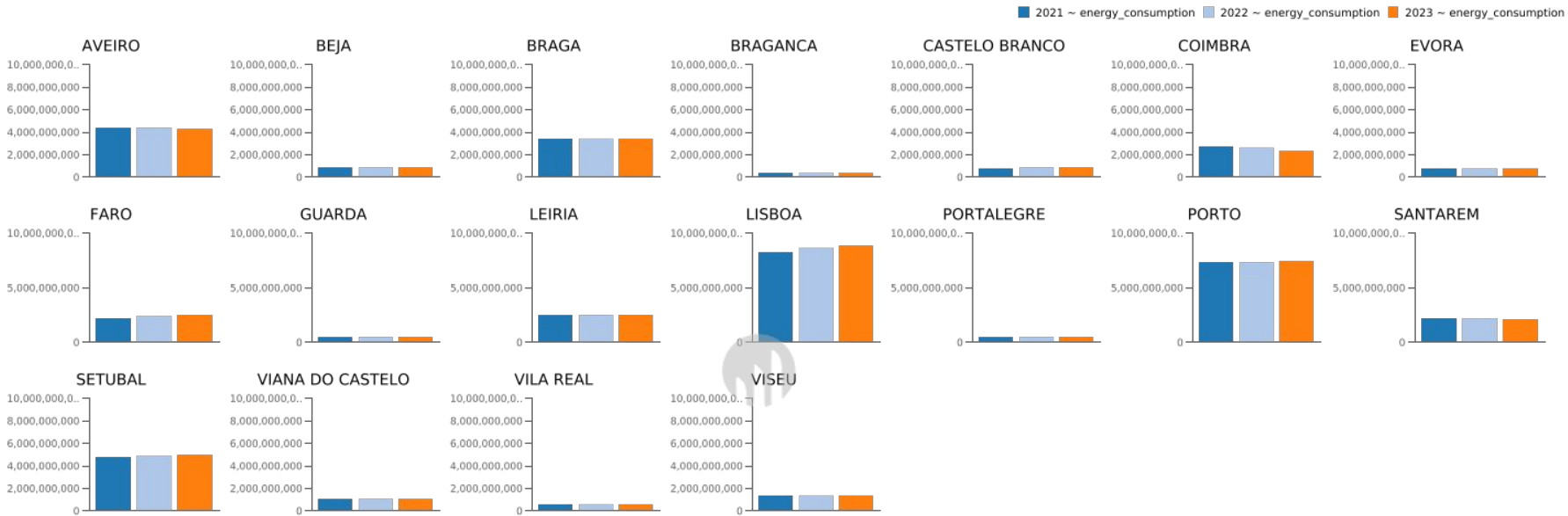
Perform an analysis on the data cube.

A: Has the consumption and the percentage of smart meters increased or decreased from one year to the next?

```
WITH
SET [~YEARS_WITH_ALL_MONTHS] AS
    Filter(
        [Time.Time Hierarchy].[Year].Members,
        Count(
            Filter(
                GENERATE(
                    [Time.Time Hierarchy].[Year].CurrentMember,
                    DESCENDANTS(
                         [Time.Time Hierarchy].CurrentMember,
                         [Time.Time Hierarchy].Month
                        )
                    ),
                NOT IsEmpty(([Measures].[energy_consumption],
                             [Time.Time Hierarchy].CurrentMember))
            )
        ) = 12
    )
SET [~YEARS_AND_SEASONS] AS
    DESCENDANTS([~YEARS_WITH_ALL_MONTHS], [Time.Time Hierarchy].[Season], SELF_AND_BEFORE)
SET [~ROWS] AS
    {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY CrossJoin({[~YEARS_WITH_ALL_MONTHS]}, {[Measures].[energy_consumption]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [energy_dw]
```

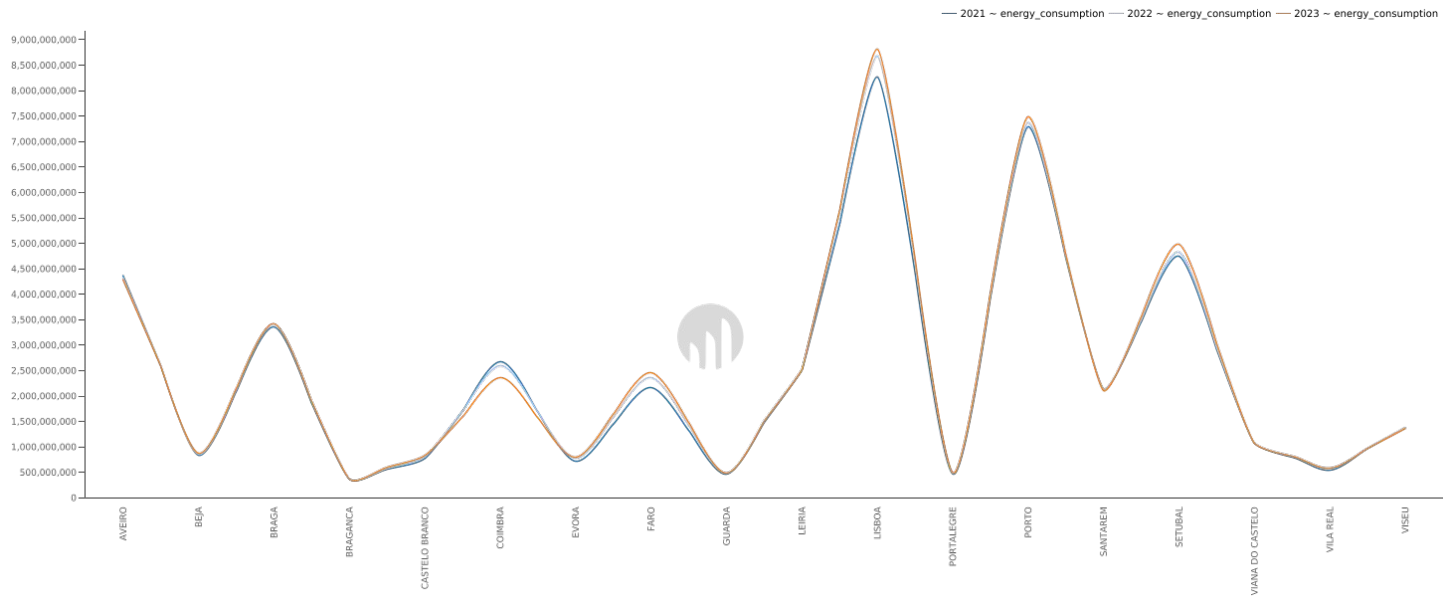
Analysis Query: **Energy Consumption** by **Years** across **Districts**

A: Has the energy consumption increased or decreased from one year to the next?



Multiple Bar Chart: **Energy Consumption** by **Years** across **Districts**

A: Has the energy consumption increased or decreased from one year to the next?



Line Chart: **Energy Consumption** by **Years** across **Districts**

A: Has the energy consumption increased or decreased from one year to the next?

	2021	2022	2023
District	energy_consumption	energy_consumption	energy_consumption
AVEIRO	4,367,415,322.888	4,337,651,601.877	4,292,519,285.717
BEJA	828,496,991.438	856,115,227.102	882,930,144.25
BRAGA	3,352,168,003.98	3,409,314,708.104	3,409,434,628.635
BRAGANCA	358,135,574.176	367,421,572.446	367,850,454.265
CASTELO BRANCO	771,720,596.076	816,629,476.959	828,403,340.268
COIMBRA	2,674,881,792.557	2,599,261,723.246	2,364,239,927.299
EVORA	711,412,031.953	778,912,645.328	789,096,239.66
FARO	2,164,425,491.223	2,365,507,704.766	2,468,932,472.211
GUARDA	467,838,307.456	482,843,620.032	482,685,491.942
LEIRIA	2,511,347,694.773	2,543,306,104.766	2,521,946,640.148
LISBOA	8,267,446,869.688	8,667,874,136.461	8,808,983,339.531
PORTALEGRE	466,189,012.254	480,855,517.918	479,473,222.707
PORTO	7,292,810,852.363	7,368,187,540.58	7,477,009,082.496
SANTAREM	2,154,799,038.299	2,151,826,323.938	2,116,235,772.021
SETUBAL	4,755,377,522.789	4,832,714,901.797	4,989,770,224.133
VIANA DO CASTELO	1,064,143,215.611	1,078,648,954.009	1,075,076,121.051
VILA REAL	547,356,482.146	573,324,500.667	574,558,214.852
VISEU	1,380,978,355.764	1,385,994,150.708	1,373,734,561.407

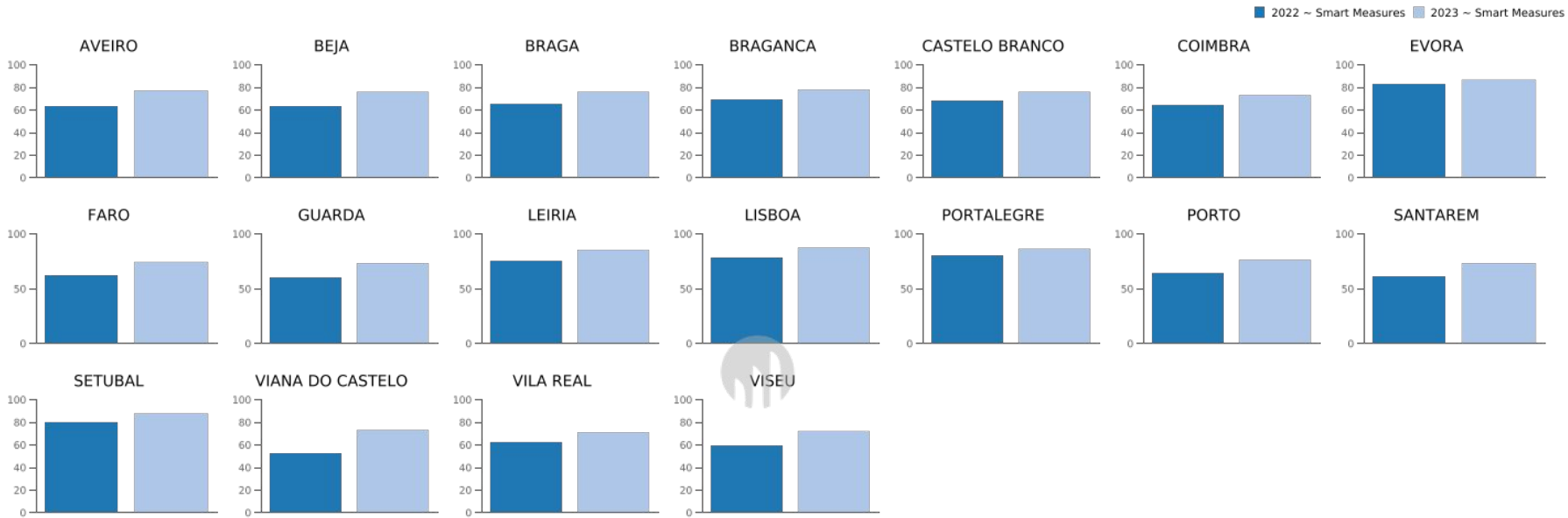
Table: **Energy Consumption** by **Years** across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?

```
WITH
SET [~YEARS_WITH_ALL_MONTHS] AS
    Filter(
        [Time.Time Hierarchy].[Year].Members,
        Count(
            Filter(
                GENERATE(
                    [Time.Time Hierarchy].[Year].CurrentMember,
                    DESCENDANTS(
                         [Time.Time Hierarchy].CurrentMember,
                         [Time.Time Hierarchy].Month
                        )
                    ),
                NOT IsEmpty(([Measures].[energy_consumption],
                             [Time.Time Hierarchy].CurrentMember))
            )
        ) = 12
    )
SET [~ROWS] AS
    {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY CrossJoin({[~YEARS_WITH_ALL_MONTHS]}, {[Measures].[Smart Measures]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [energy_dw]
```

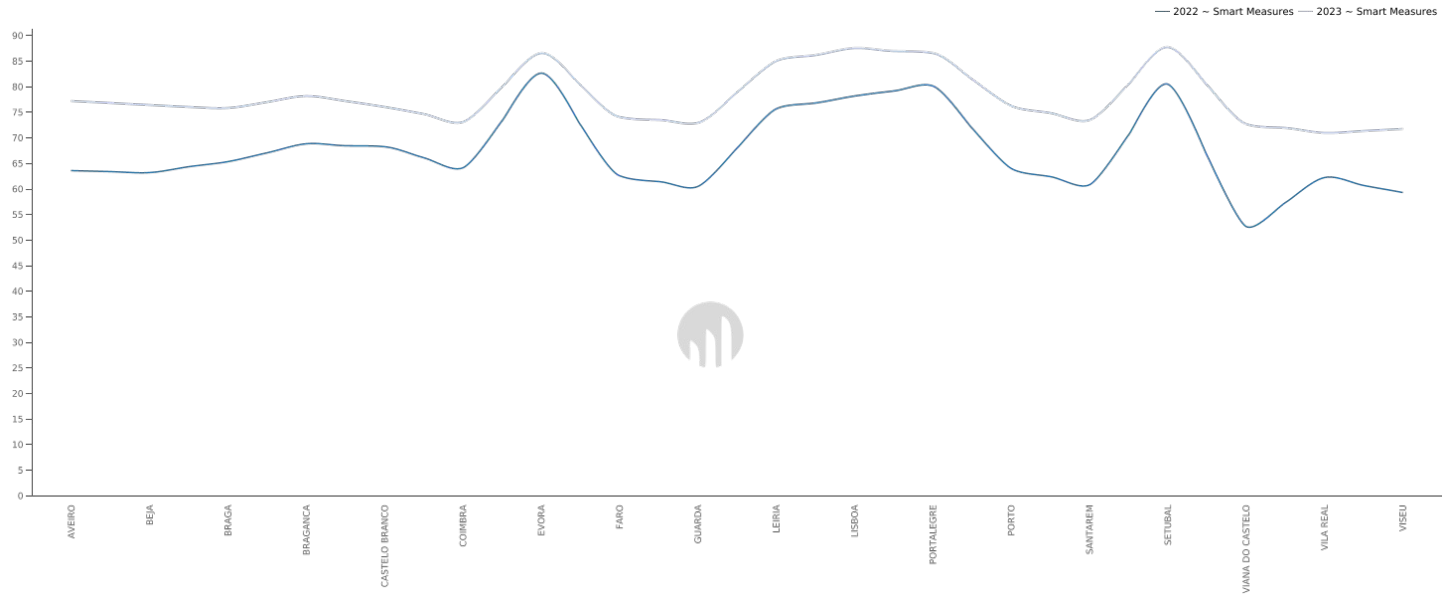
Analysis Query: **Percentage of Smart Meters** by **Years** across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?



Multiple Bar Chart: **Percentage of Smart Meters** by **Years** across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?



Line Chart: **Percentage of Smart Meters** by **Years** across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?

Since the previous query was only for years that had energy consumption with complete data from January to December, we thought it would be appropriate to do the same for the percentage of smart meters (2022 and 2024 have incomplete data for smart meters).

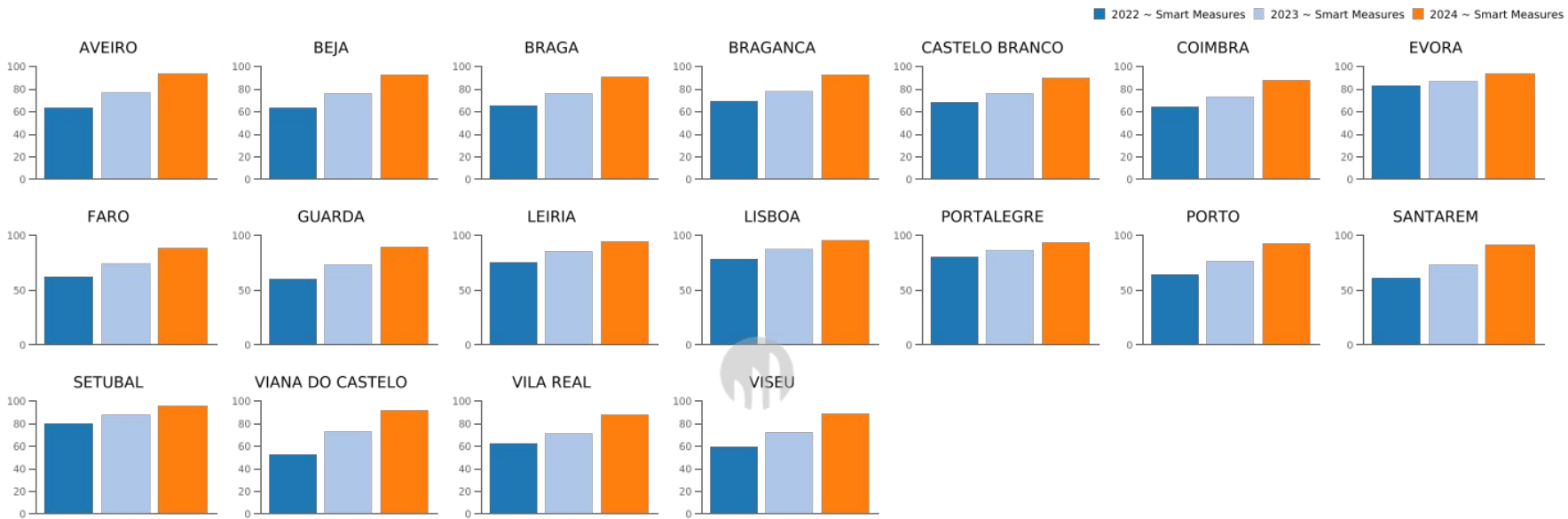
However, the only resulting year was 2023, which did not allow for analysis of the evolution from one year to another.

Thus, here is a query to analyze the values for all years where there is information on Smart Meters (although not entirely accurate, it may be relevant for making more comprehensive analyses).

```
WITH
SET [~ROWS] AS
    {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY CrossJoin({[Time.Time Hierarchy].[Year].Members}, {[Measures].[Smart Measures]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [energy_dw]
```

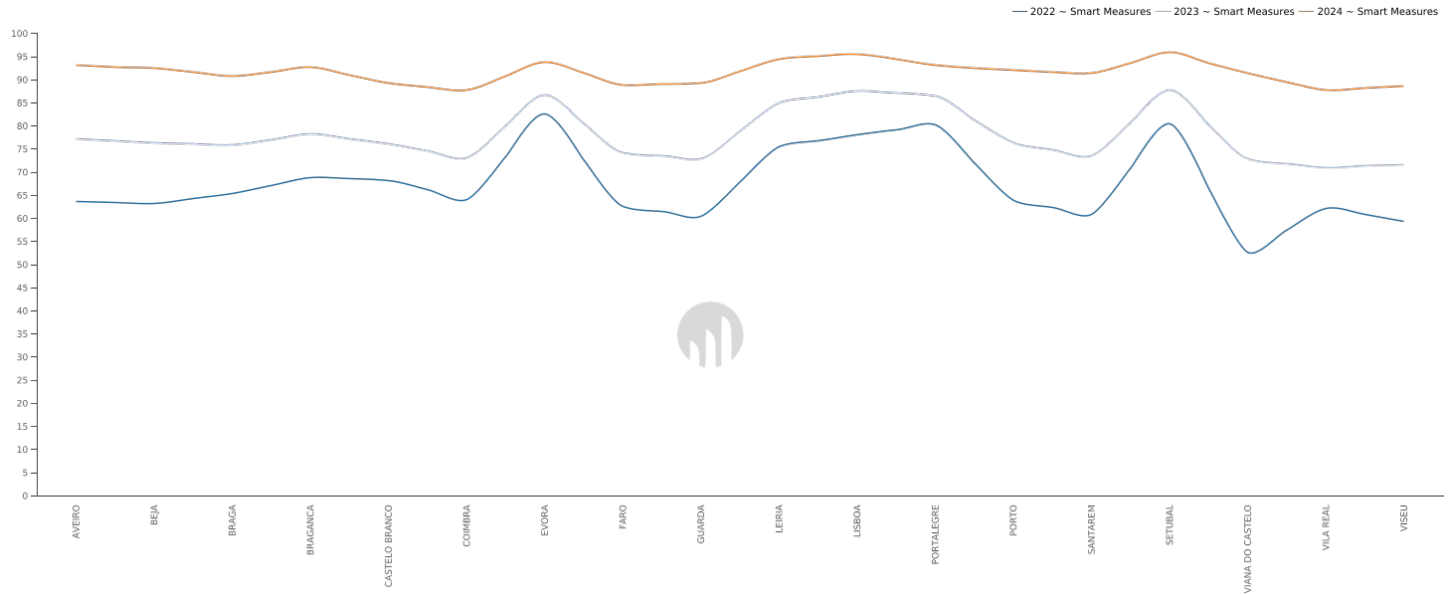
Analysis Query: **Percentage of Smart Meters** by **Years** (all years) across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?



Multiple Bar Charts: **Percentage of Smart Meters** by **Years** (all years) across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?



Line Chart: **Percentage of Smart Meters** by **Years** (all years) across **Districts**

A: Has the percentage of smart meters increased or decreased from one year to the next?

	2022	2023	2024
District	Smart Measures	Smart Measures	Smart Measures
AVEIRO	63.623	77.328	93.255
BEJA	63.239	76.411	92.436
BRAGA	65.456	75.855	90.748
BRAGANCA	68.796	78.252	92.665
CASTELO BRANCO	68.268	76.145	89.319
COIMBRA	64.177	73.241	87.71
EVORA	82.667	86.654	93.927
FARO	62.557	74.178	88.875
GUARDA	60.432	73.015	89.316
LEIRIA	75.629	85.029	94.557
LISBOA	78.201	87.506	95.648
PORTALEGRE	80.19	86.598	93.101
PORTO	63.956	76.276	92.058
SANTAREM	60.942	73.565	91.434
SETUBAL	80.477	87.693	95.873
VIANA DO CASTELO	52.612	72.837	91.389
VILA REAL	62.248	70.967	87.739
VISEU	59.308	71.723	88.743

Line Chart: **Percentage of Smart Meters** by **Years** (all years) across **Districts**

A: Has the consumption percentage of smart meters increased or decreased from one year to the next?

Conclusions:

The energy consumption has **increased** over the years in districts of Castelo Branco, Évora, Faro, Lisboa, Porto and Setúbal. The energy consumption **remained roughly the same** in the districts of Aveiro, Braga, Beja, Bragança, Guarda, Leiria, Portalegre, Santarém, Viana do Castelo, Vila Real and Viseu. Has **decreased** in Coimbra.

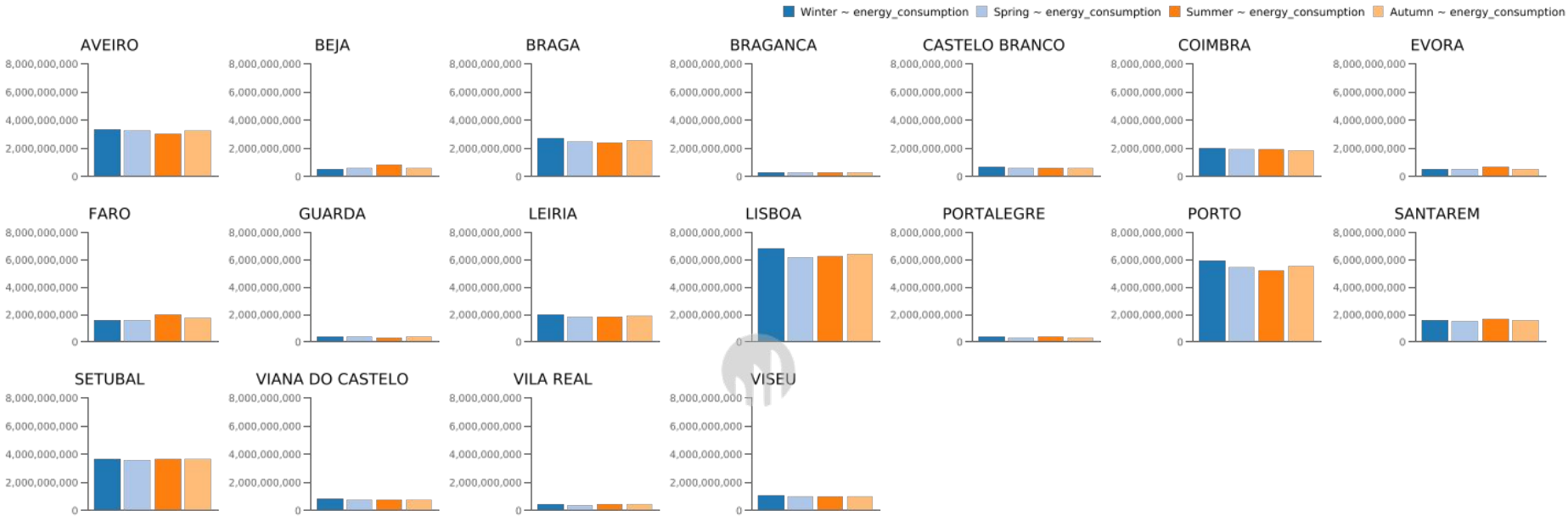
The percentage of smart meters increased for all districts over the years.

B: What is the influence of the season on consumption?

```
WITH
SET [~YEARS_WITH_ALL_MONTHS] AS
    Filter(
        [Time.Time Hierarchy].[Year].Members,
        Count(
            Filter(
                GENERATE(
                    [Time.Time Hierarchy].[Year].CurrentMember,
                    DESCENDANTS(
                         [Time.Time Hierarchy].CurrentMember,
                         [Time.Time Hierarchy].Month
                        )
                    ),
                NOT IsEmpty(([Measures].[energy_consumption],
                             [Time.Time Hierarchy].CurrentMember))
            )
        ) = 12
    )
SET [~YEARS_AND_SEASONS] AS
    DESCENDANTS([~YEARS_WITH_ALL_MONTHS], [Time.Time Hierarchy].[Season], SELF)
SET [~ROWS] AS
    {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY   CrossJoin( {[~YEARS_AND_SEASONS]}, {[Measures].[energy_consumption]} ) ON COLUMNS,
NON EMPTY[~ROWS]  ON ROWS
FROM [energy_dw]
```

Analysis Query: **Energy Consumption** by **Season** across **Districts**

B: What is the influence of the season on consumption?

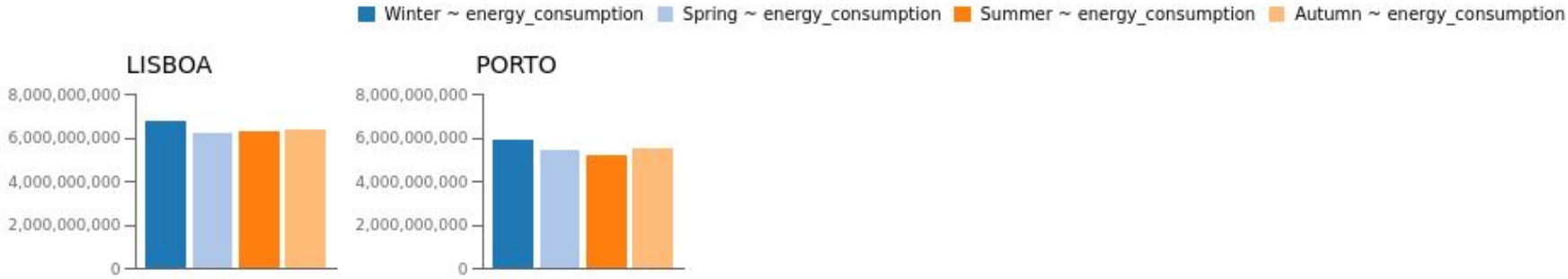


Multiple Bar Chart: **Energy Consumption** by **Season** across **Districts**

B: What is the influence of the season on consumption?

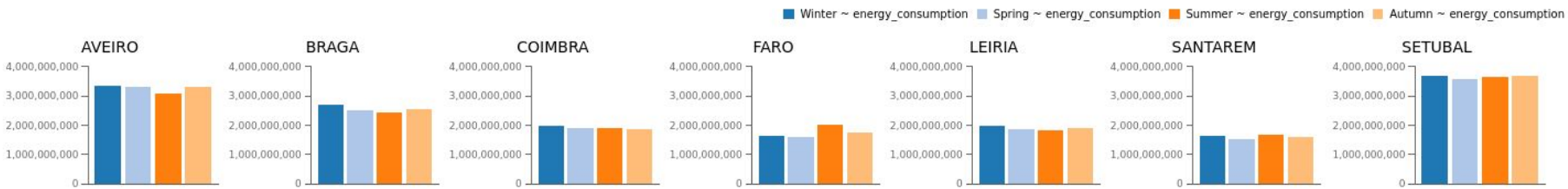
```
SET [~ROWS] AS
    {
      [Location.Location Hierarchy].[District].&[LISBOA],
      [Location.Location Hierarchy].[District].&[PORTO]
    }
```

Due to the significant differences in consumption values and in order to better analyze each district, we applied some filtering.



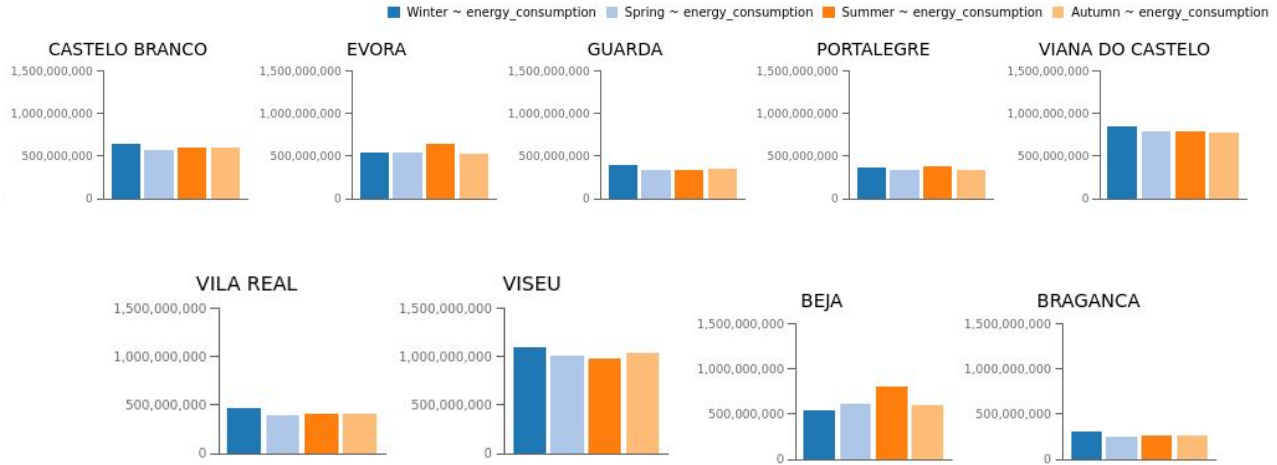
B: What is the influence of the season on consumption?

```
SET [~ROWS] AS
  Filter(
    Except(
        {[Location.Location Hierarchy].[District].Members},
        {
          [Location.Location Hierarchy].[District].&[LISBOA],
          [Location.Location Hierarchy].[District].&[PORTO]
        }
    ),
    [Measures].[energy_consumption] >= 7000000000
  )
```



B: What is the influence of the season on consumption?

```
SET [~ROWS] AS
  Filter(
    {[Location.Location Hierarchy].[District].Members},
    [Measures].[energy_consumption] <= 7000000000
  )
```



B: What is the influence of the season on energy consumption?

Conclusion:

There is a clear influence of the season on energy consumption.

We verify that in portuguese southern regions of Algarve and Alentejo (Faro, Beja, Portalegre, Évora) have higher energy consumption in Summer.

Northern and Central districts (Lisboa appears as an exception) such as Vila Real, Viseu, Castelo Branco, Aveiro, Leiria Guarda, Viana do Castelo, Bragança, Braga and Porto, winters are more demanding.

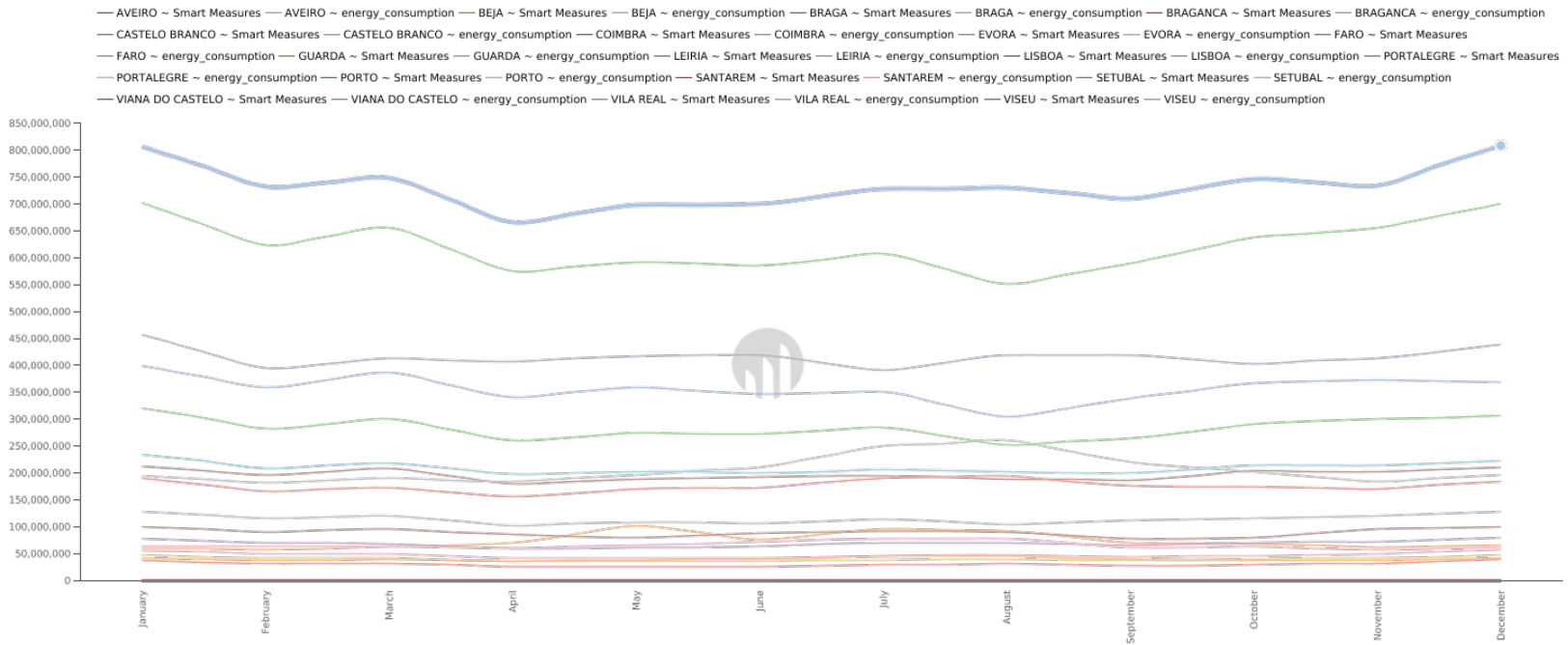
Milder districts like Coimbra, Setúbal and Santarém present an even energy consumption over all year.

We can suppose that these seasonal influences on energy consumption may stem from factors such as increased tourism or temperature changes that lead to greater use of air conditioning or heating or energy for other sources.

C: Is the percentage of smart meters having an impact on electricity consumption?

```
WITH
SET [~YEARS_WITH_ALL_MONTHS] AS
  Filter(
    [Time.Time Hierarchy].[Year].Members,
    Count(
      Filter(
        GENERATE(
          [Time.Time Hierarchy].[Year].CurrentMember,
          DESCENDANTS(
            [Time.Time Hierarchy].CurrentMember,
            [Time.Time Hierarchy].Month
          )
        ),
        NOT IsEmpty(([Measures].[energy_consumption], [Time.Time Hierarchy].CurrentMember))
        AND
        NOT IsEmpty(([Measures].[Smart Measures], [Time.Time Hierarchy].CurrentMember))
      )
    ) = 12
  )
SET [~MONTHS_WITHIN_YEARS] AS
    GENERATE(
        [~YEARS_WITH_ALL_MONTHS],
        DESCENDANTS(
            [Time.Time Hierarchy].[Year].CurrentMember,
            [Time.Time Hierarchy].Month
        )
    )
SET [~ROWS] AS
  {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY CrossJoin({[~ROWS]}, {[Measures].[Smart Measures],  [Measures].[energy_consumption]}) ON COLUMNS,
NON EMPTY  ORDER(
        [~MONTHS_WITHIN_YEARS],
        [Measures].[Smart Measures], BASC
    ) ON ROWS
FROM [energy_dw]
```

C: Is the percentage of smart meters having an impact on electricity consumption?

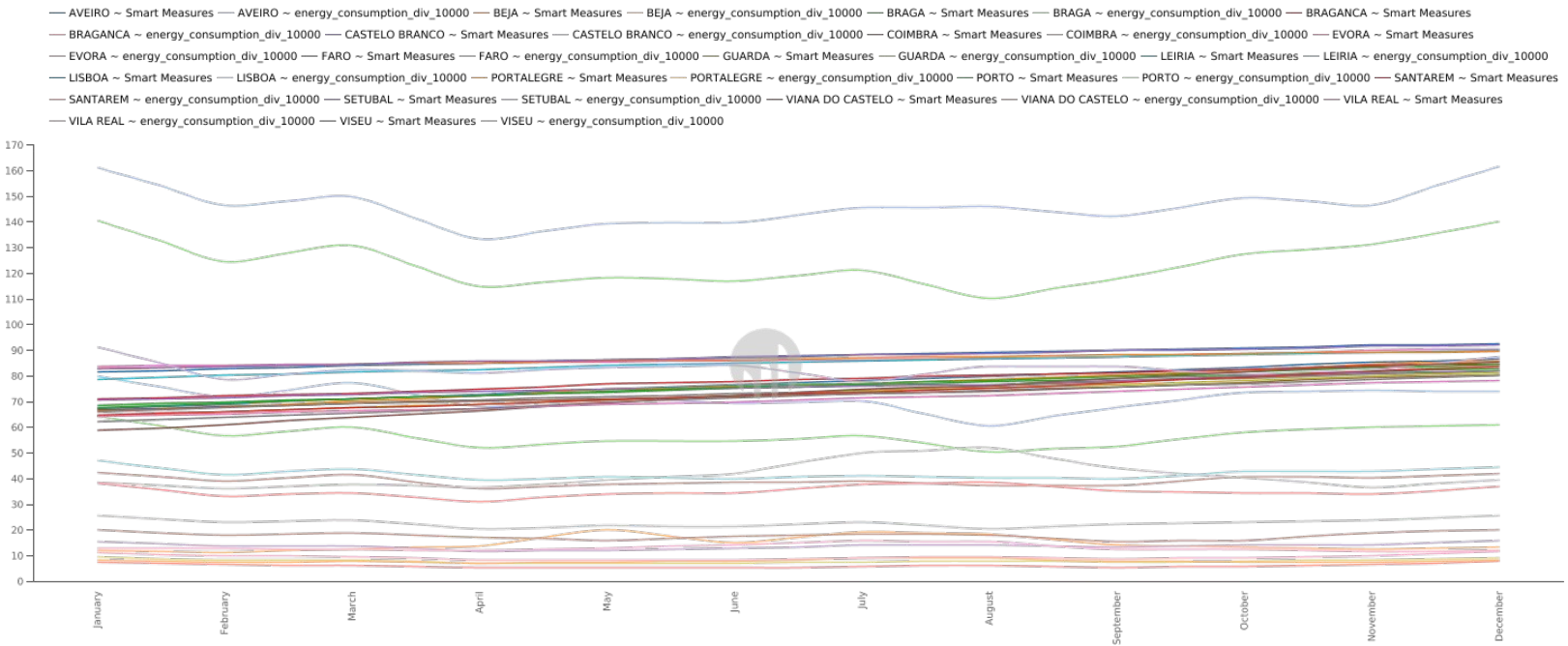


C: Is the percentage of smart meters having an impact on electricity consumption?

Since the percentage of smart meters is significantly lower than energy consumption, it is not perceptible in the chart.
Therefore, we decided to proportionally reduce energy consumption in order to have a clearer analysis.

```
WITH
SET [~YEARS_WITH_ALL_MONTHS] AS
  Filter(
    [Time.Time Hierarchy].[Year].Members,
    Count(
      Filter(
        GENERATE(
          [Time.Time Hierarchy].[Year].CurrentMember,
          DESCENDANTS(
            [Time.Time Hierarchy].CurrentMember,
            [Time.Time Hierarchy].Month
          )
        ),
        NOT IsEmpty(([Measures].[energy_consumption], [Time.Time Hierarchy].CurrentMember))
        AND
        NOT IsEmpty(([Measures].[Smart Measures], [Time.Time Hierarchy].CurrentMember))
      )
    ) = 12
  )
SET [~MONTHS_WITHIN_YEARS] AS
    GENERATE(
        [~YEARS_WITH_ALL_MONTHS],
        DESCENDANTS(
            [Time.Time Hierarchy].[Year].CurrentMember,
            [Time.Time Hierarchy].Month
        )
    )
MEMBER [Measures].[energy_consumption_reduced] AS
    [Measures].[energy_consumption] / 5000000
SET [~ROWS] AS
  {[Location.Location Hierarchy].[District].Members}
SELECT
NON EMPTY CrossJoin({[~ROWS]}, {[Measures].[Smart Measures],  [Measures].[energy_consumption_reduced]}) ON COLUMNS,
NON EMPTY  ORDER(
        [~MONTHS_WITHIN_YEARS],
        [Measures].[Smart Measures], BASC
    ) ON ROWS
FROM [energy_dw]
```

C: Is the percentage of smart meters having an impact on electricity consumption?



C: Is the percentage of smart meters having an impact on electricity consumption?

Conclusion:

We can observe energy consumption has not being impacted by the percentage of smart meters (this is expected from the conclusions drawn in task 2).

Thank You For Your Attention!

Ana Margarida Cardoso Almeida - 102618
Vasco Manuel Dias Paisana - 102533