

# Data Analysis and Integration

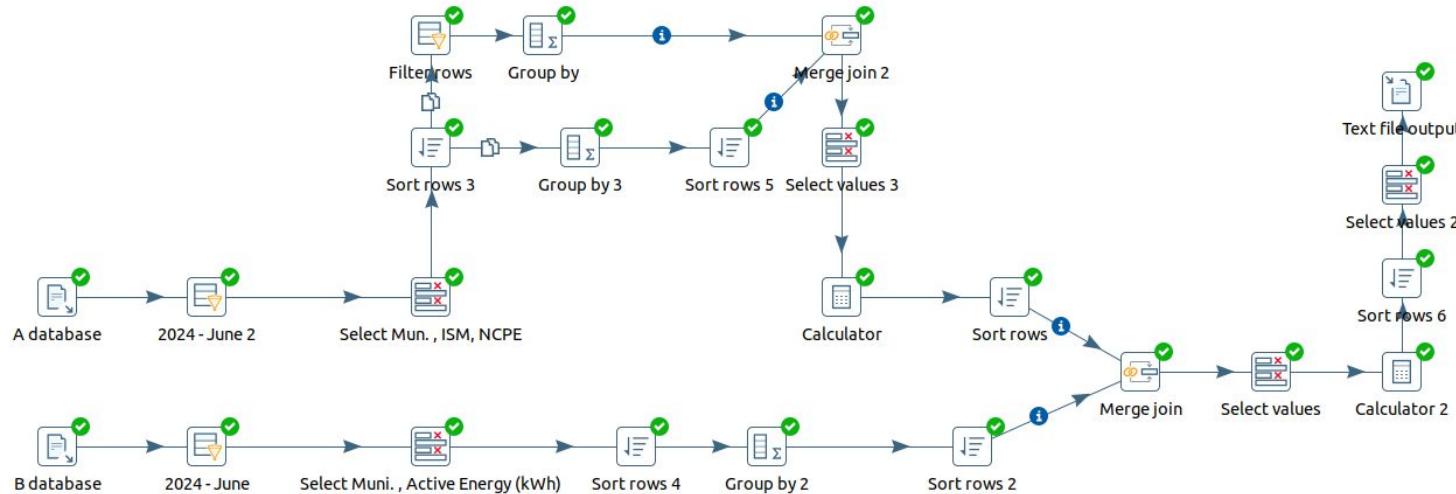
## Project

Grupo 52

Alessandro Campagna 106751  
João Rocha 103465

# Task 1

# Task 1



# Task 1 - A database (Csv File Input)

**CSV file input**

Step name: **A database**

Filename: **/home/aid/Downloads/dataset\_A.csv**

Delimiter: **,**

Enclosure: **"**

NIO buffer size: **50000**

Lazy conversion?

Header row present?

Add filename to result?

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: **mixed**

File encoding: **UTF-8**

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	.	x
2 Month	Integer	#	15	0	\$	.	x
3 Date	String		7		\$	.	x
4 District	String		16		\$	.	x
5 Municipality	String		22		\$	.	x
6 Parish	String		51		\$	.	x
7 Includes Smart Meter	String		3		\$	.	x
8 Number of CPE's	Integer	#	15	0	\$	.	x
9 DistrictCode	Integer	#	15	0	\$	.	x
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	x
11 DistrictMunicipalityParishCode	String		6		\$	.	x
12 Active Contract	String		3		\$	.	x

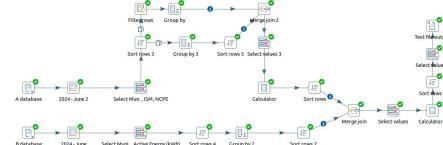
Buttons: **Help**, **OK**, **Get Fields**, **Preview**, **Cancel**

**Examine preview data**

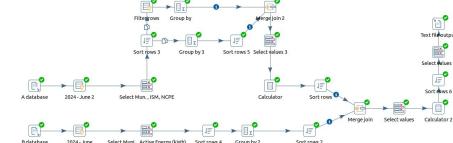
Rows of step: A database (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active Contract
1	2024	5	2024-05	Vila Real	Ribeira de Pena	Alvadia	Sim	155	17	170901	170901	Sim
2	2024	5	2024-05	Bragança	Mogadouro	Travanca	Não	15	4	408	040820	Sim
3	2024	5	2024-05	Portalegre	Ponte de Sor	Montargil	Não	475	12	1213	121302	Sim
4	2024	5	2024-05	Viseu	Santa Comba Dão	Santa Comba Dão	Sim	3141	18	1814	181411	Sim
5	2024	5	2024-05	Portalegre	Arronches	Assunção	Sim	1248	12	1202	120201	Sim
6	2024	5	2024-05	Santarém	Tomar	Asseiceira	Sim	1549	14	1418	141802	Sim
7	2024	5	2024-05	Coimbra	Mira	Carapinhos	Sim	488	6	608	060803	Sim
8	2024	5	2024-05	Guarda	Celorico da Beira	Prados	Sim	221	9	903	090312	Sim
9	2024	5	2024-05	Viana do Castelo	Paredes de Coura	Agualonga	Sim	175	16	1605	160501	Sim
10	2024	5	2024-05	Beja	Odemira	Reliúquias	Sim	334	2	211	021102	Sim
11	2024	5	2024-05	Bragança	Macedo de Cavaleiros	Talhinhos e Baga	Sim	319	4	405	040544	Sim
12	2024	5	2024-05	Coimbra	Condeixa-a-Nova	Zambujal	Não	1	6	604	060410	Sim
13	2024	5	2024-05	Braganca	Vila Flor	Santa Comba de Poiares	Não	19	4	410	041012	Sim
14	2024	5	2024-05	Viana do Castelo	Melgaço	Cousso	Não	96	16	1603	160304	Sim
15	2024	5	2024-05	Castelo Branco	Pronaça-a-Nova	Sobreira Forma	Sim	1525	5	508	050808	Sim
16	2024	5	2024-05	Guarda	Trancoso	Rio de Mel	Sim	186	9	913	091316	Sim
17	2024	5	2024-05	Braga	Barcelos	Tanel (São Veríssimo)	Sim	1353	3	302	030377	Sim
18	2024	5	2024-05	Beja	Moura	Amaraldeja	Não	63	2	210	021001	Sim
19	2024	5	2024-05	Bragança	Vinhais	Edral	Sim	151	4	412	041206	Sim
20	2024	5	2024-05	Porto	Paços de Ferreira	Raimonda	Sim	1069	13	1309	130914	Sim
21	2024	5	2024-05	Vila Real	Murça	Murça	Sim	1346	17	1707	170705	Sim
22	2024	5	2024-05	Guarda	Seia	Sazes da Beira	Não	107	9	912	091219	Sim
23	2024	5	2024-05	Braga	Banhos	Vila Cova e Feliz	Não	196	3	302	03027H	Sim
24	2024	5	2024-05	Braga	Barcelos	Vila Seca	Sim	342	3	302	030287	Sim
25	2024	5	2024-05	Viseu	Sátão	Milfonte	Não	14	18	1817	181706	Sim
26	2024	5	2024-05	Coimbra	Tábua	Póvoa de Mós	Não	14	6	616	061611	Sim
27	2024	5	2024-05	Lisboa	Torres Vedras	Ramalhal	Não	594	11	1113	111311	Sim
28	2024	5	2024-05	Lisboa	Cadaval	Alguber	Sim	552	11	1104	110401	Sim
29	2024	5	2024-05	Viana do Castelo	Monção	Celvões e Badim	Sim	521	16	1604	160435	Sim
30	2024	5	2024-05	Santarém	Santarém	Castelo e Vaqueiro	Sim	444	14	1416	141631	Sim

Buttons: **Close**, **Stop**, **Get more rows**



# Task 1 - 2024 - June 2 (Filter rows)



**Filter rows**

Step name: **2024 - June 2**

Send 'true' data to step:

Send 'false' data to step:

The condition:

+

Year = [2024]

AND

Month = [6]

OK Cancel

Examine preview data

Rows of step: 2024 - June 2 (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active Contract
1	2024	6	2024-06	Faro	Lagos	Luz	Não	420	8	807	080703	Sim
2	2024	6	2024-06	Guarda	Guarda	Codesseiro	Sim	139	9	907	090714	Sim
3	2024	6	2024-06	Castelo Br	Idanha-a-Nova	Aldeia de Santa Ma	Sim	324	5	505	050502	Sim
4	2024	6	2024-06	Viana do C	Viana do Cast	São Romão de Neiv	Sim	713	16	1609	160923	Sim
5	2024	6	2024-06	Guarda	Sabugue	Malacata	Sim	385	9	911	091118	Sim
6	2024	6	2024-06	Vila Real	Montalegre	Morgadie	Sim	155	17	1706	170616	Sim
7	2024	6	2024-06	Bragança	Mirandela	Torre de Dona Char	Não	23	4	407	040730	Sim
8	2024	6	2024-06	Braga	Barcelos	Cambeses	Não	32	3	302	030216	Sim
9	2024	6	2024-06	Castelo Br	Belmonte	Magalhães	Sim	300	5	501	050105	Sim
10	2024	6	2024-06	Castelo Br	Fundão	Castelo Novo	Não	12	5	504	050413	Sim
11	2024	6	2024-06	Ler	Porto de Mós	Serra Ventoso	Não	8	10	1016	101613	Sim
12	2024	6	2024-06	Viana do C	Aveiro	Sil	Não	2675	16	1109	110905	Sim
13	2024	6	2024-06	Évora	Reguengos de Corval	Sil	Não	952	7	711	071102	Sim
14	2024	6	2024-06	Braga	Góis	Almeida e São Pa	Sim	1016	3	308	030801	Sim
15	2024	6	2024-06	Castelo Br	Penedono	Freixo da Fi	Não	12	6	605	060314	Sim
16	2024	6	2024-06	Bragança	Almeida	Moedas de C	Sim	273	41	405	040915	Sim
17	2024	6	2024-06	Guarda	Carrazeda de C	São Pedro e Santa	Sim	1791	9	903	090324	Sim
18	2024	6	2024-06	Faro	Aljezur	Não	Não	471	8	803	080301	Sim
19	2024	6	2024-06	Vila Real	Parada de Cunhos	Não	Não	17	17	1714	171420	Sim
20	2024	6	2024-06	Vila Real	Peso da Rég	Eduílhas	Não	63	17	1708	170809	Sim
21	2024	6	2024-06	Porto	Marcos de Cai	Sobretâmega	Não	18	13	1307	130723	Sim
22	2024	6	2024-06	Braga	Braga	Torre de Mon	Sim	10841	3	303	030365	Sim
23	2024	6	2024-06	Bragança	Torre de Mon	Adegaña e Cardar	Sim	523	4	409	040918	Sim
24	2024	6	2024-06	Castelo Br	Penamacor	Aranhas	Sim	458	5	507	050704	Sim
25	2024	6	2024-06	Guarda	Fornos de Al	Sobral Pichorro e F	Não	2	9	905	090519	Sim
26	2024	6	2024-06	Santarém	Santarém	Almoster	Não	116	14	1416	141606	Sim
27	2024	6	2024-06	Faro	Silves	Algés e Tunes	Não	638	8	813	081310	Sim
28	2024	6	2024-06	Viseu	Vouzela	São Miguel do Mat	Não	22	18	1824	182410	Sim
29	2024	6	2024-06	Faro	Tavira	Cachopo	Sim	344	8	814	081401	Sim
30	2024	6	2024-06	Guarda	Almeida	Leomil, Mido, Sen	Não	6	9	902	090234	Sim

Close Stop Get more rows

# Task 1 - Select Mun, ISM, NCPE



**Select values**

Step name **Select Mun. , ISM, NCPE**

Select & Alter  Remove  Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 DistrictMunicipalityCode			
2 Municipality			
3 Includes Smart Meter			
4 Number of CPE's			

Include unspecified fields, ordered

OK Cancel

Help

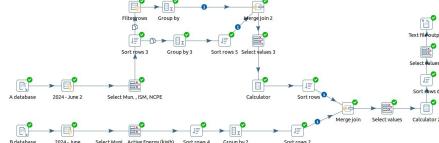
Examine preview data

Rows of step: Select Mun. , ISM, NCPE (1000 rows)

	DistrictMunicipalityCode	Municipality	Includes Smart Meter	Number of CPE's
1	807	Lagos	Não	420
2	907	Guarda	Sim	139
3	505	Idanha-a-Nova	Sim	324
4	1609	Viana do Castelo	Sim	713
5	911	Sabugal	Sim	385
6	1706	Montalegre	Sim	155
7	407	Mirandela	Não	23
8	302	Barcelos	Não	32
9	501	Belmonte	Sim	300
10	504	Fundão	Não	12
11	1016	Porto de Mós	Não	8
12	1609	Viana do Castelo	Sim	2675
13	711	Reguengos de Mons.	Sim	952
14	308	Guimarães	Sim	1016
15	605	Figueira da Foz	Não	12
16	405	Macedo de Cavaleiros	Sim	723

**Close Stop Get more rows**

# Task 1 - Sort rows 3



Sort rows

Step name: Sort rows 3

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields:

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

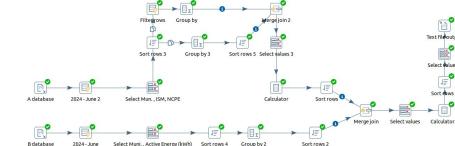
OK Cancel Get Fields

Examine preview data

Rows of step: Sort rows 3 (1000 rows)

	DistrictMunicipalityCode	Municipality	Includes Smart Meter	Number of CPE's
1	101	Águeda	Sim	867
2	101	Águeda	Não	16
3	101	Águeda	Sim	2410
4	101	Águeda	Sim	1843
5	101	Águeda	Não	36
6	101	Águeda	Sim	1172
7	101	Águeda	Sim	3065
8	101	Águeda	Não	37
9	101	Águeda	Sim	171
10	101	Águeda	Não	11
11	101	Águeda	Não	4
12	101	Águeda	Sim	584
13	101	Águeda	Não	21
14	101	Águeda	Não	12
15	101	Águeda	Não	20
16	101	Águeda	Não	20

# Task 1 - Filter rows



**Filter rows**

Step name: **Filter rows**

Send 'true' data to step:

Send 'false' data to step:

The condition:

+

Includes Smart Meter	=	
		Sim (String)

OK Cancel

**Examine preview data**

Rows of step: Filter rows (1000 rows)

	DistrictMunicipalityCode	Municipality	Includes Smart Meter	Number of CPE's
1	101	Águeda	Sim	867
2	101	Águeda	Sim	2410
3	101	Águeda	Sim	1843
4	101	Águeda	Sim	1172
5	101	Águeda	Sim	3065
6	101	Águeda	Sim	171
7	101	Águeda	Sim	584
8	101	Águeda	Sim	2165
9	101	Águeda	Sim	1617
10	101	Águeda	Sim	1801
11	101	Águeda	Sim	8079
12	102	Albergaria-a-Velha	Sim	1291
13	102	Albergaria-a-Velha	Sim	1090
14	102	Albergaria-a-Velha	Sim	990
15	102	Albergaria-a-Velha	Sim	2735
16	102	Albergaria-a-Velha	Sim	6620

Close Stop Get more rows

# Task 1 - Group by

**Group by**

Step name **Group by**

Include all rows?

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

TMP-file prefix `grp`

Add line number, restart in each group

Line number field name

Always give back a result row

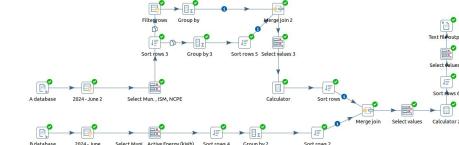
The fields that make up the group:

Group field

1	DistrictMunicipalityCode
2	Municipality

Aggregates :

Name	Subject	Type
1	Number_Smart	Number of CPE's

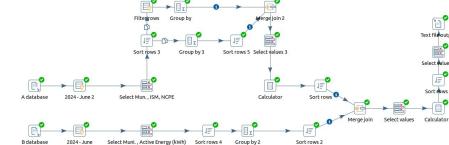


Examine preview data

Rows of step: Group by (278 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart
1	101	Águeda	23774
2	102	Albergaria-a-Velha	13437
3	103	Anadia	14682
4	104	Arouca	9682
5	105	Aveiro	50414
6	106	Castelo de Paiva	7834
7	107	Espinho	18258
8	108	Estarreja	14009
9	109	Santa Maria da Feira	65079
10	110	Ilhavo	24983
11	111	Mealhada	11200
12	112	Murtosa	8345
13	113	Oliveira de Azeméis	28372
14	114	Oliveira do Bairro	13304
15	115	Ovar	31361
16	116	São João da Madeira	12122

# Task 1 - Group by 3



**Group by**

Step name	<input type="text" value="Group by 3"/>
Include all rows?	<input type="checkbox"/>
Temporary files directory	<input type="text" value="%%java.io.tmpdir%%"/> <a href="#">Browse...</a>
TMP-file prefix	<input type="text" value="grp"/>
Add line number, restart in each group	<input type="checkbox"/>
Line number field name	<input type="text"/>
Always give back a result row	<input type="checkbox"/>

The fields that make up the group:

<input type="checkbox"/>	Group field	<a href="#">Get Fields</a>
1	DistrictMunicipalityCode	
2	Municipality	

Aggregates :

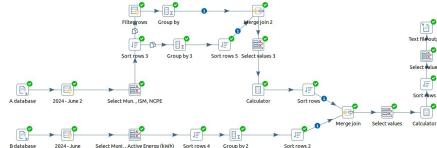
<input type="checkbox"/>	Name	Subject	Type	<a href="#">Get lookup fields</a>
1	Count_CPE	Number of CPE's	Sum	

Examine preview data

Rows of step: Group by 3 (278 rows)

	DistrictMunicipalityCode	Municipality	Count_CPE
1	101	Águeda	24011
2	102	Albergaria-a-Velha	13897
3	103	Anadia	16871
4	104	Arouca	12337
5	105	Aveiro	50559
6	106	Castelo de Paiva	7904
7	107	Espinho	18385
8	108	Estarreja	14156
9	109	Santa Maria da Feira	70801
10	110	Ilhavo	25089
11	111	Mealhada	11333
12	112	Murtosa	8450
13	113	Oliveira de Azeméis	32400
14	114	Oliveira do Bairro	13374
15	115	Ovar	31704
16	116	São João da Madeira	12156

# Task 1 - Sort rows 5



**Sort rows**

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields:

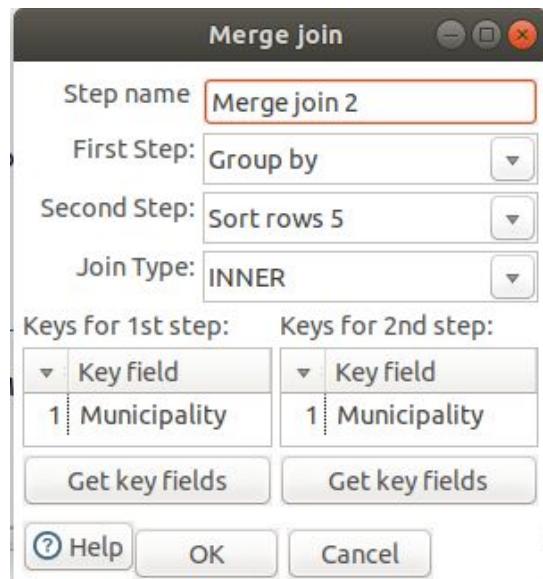
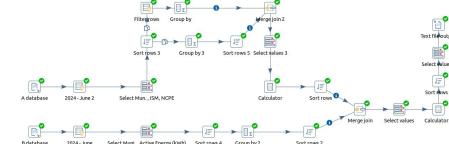
Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
DistrictMunicipalityCode	Y	N	N	0	N

**Examine preview data**

Rows of step: Sort rows 5 (278 rows)

	DistrictMunicipalityCode	Municipality	Count_CPE
1	101	Águeda	24011
2	102	Albergaria-a-Velha	13897
3	103	Anadia	16871
4	104	Arouca	12337
5	105	Aveiro	50559
6	106	Castelo de Paiva	7904
7	107	Espinho	18385
8	108	Estarreja	14156
9	109	Santa Maria da Feira	70801
10	110	Ilhavo	25089
11	111	Mealhada	11333
12	112	Murtosa	8450
13	113	Oliveira de Azeméis	32400
14	114	Oliveira do Bairro	13374
15	115	Ovar	31704
16	116	São João da Madeira	12156

# Task 1 - Merge join 2



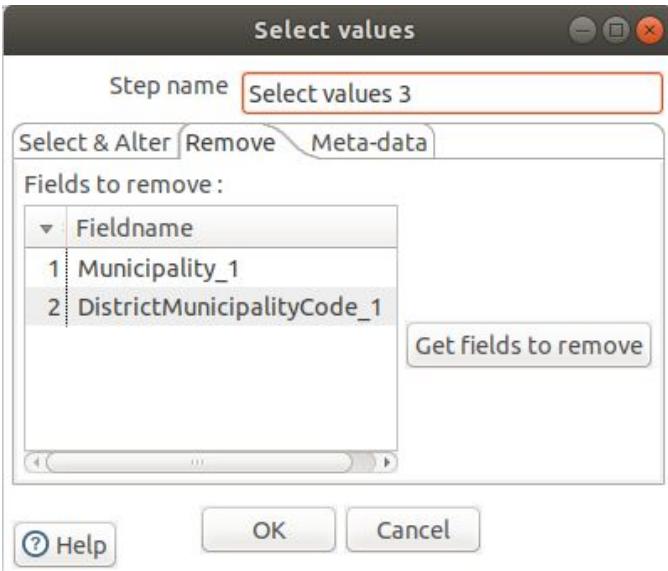
Examine preview data

Rows of step: Merge join 2 (278 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	DistrictMunicipalityCode_1	Municipality_1	Count_CPE
1	101	Águeda	23774	101	Águeda	24011
2	102	Albergaria-a-Velha	13437	102	Albergaria-a-Velha	13897
3	103	Anadia	14682	103	Anadia	16871
4	104	Arouca	9682	104	Arouca	12337
5	105	Aveiro	50414	105	Aveiro	50559
6	106	Castelo de Paiva	7834	106	Castelo de Paiva	7904
7	107	Espinho	18258	107	Espinho	18385
8	108	Estarreja	14009	108	Estarreja	14156
9	109	Santa Maria da Feira	65079	109	Santa Maria da Feira	70801
10	110	Ilhavo	24983	110	Ilhavo	25089
11	111	Mealhada	11200	111	Mealhada	11333
12	112	Murtosa	8345	112	Murtosa	8450
13	113	Oliveira de Azeméis	28372	113	Oliveira de Azeméis	32400
14	114	Oliveira do Bairro	13304	114	Oliveira do Bairro	13374
15	115	Ovar	31361	115	Ovar	31704
16	116	São João da Madeira	12122	116	São João da Madeira	12156

**Close**

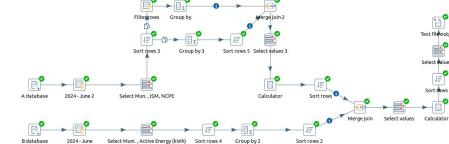
# Task 1 - Select values 3



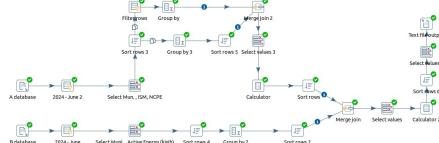
Examine preview data

Rows of step: Select values 3 (278 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE
1	101	Águeda	23774	24011
2	102	Albergaria-a-Velha	13437	13897
3	103	Anadia	14682	16871
4	104	Arouca	9682	12337
5	105	Aveiro	50414	50559
6	106	Castelo de Paiva	7834	7904
7	107	Espinho	18258	18385
8	108	Estarreja	14009	14156
9	109	Santa Maria da Feira	65079	70801
10	110	Ílhavo	24983	25089
11	111	Mealhada	11200	11333
12	112	Murtosa	8345	8450
13	113	Oliveira de Azeméis	28372	32400
14	114	Oliveira do Bairro	13304	13374
15	115	Ovar	31361	31704
16	116	São João da Madeira	12122	12156



# Task 1 - Calculator



Calculator

Step name  
Calculator

Throw an error on non existing files

Fields:

New field	Calculation	Field A	Field B	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol	Currency symb
1 SmartMeterPercentage	100 * A / B	Number Smart	Count_CPE	None		N					

OK Cancel

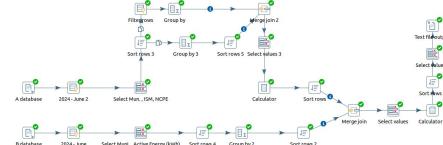
Examine preview data

Rows of step: Calculator (278 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage
1	101	Agueda	23774	24011	99.0
2	102	Albergaria-a-Velha	13437	13897	96.0
3	103	Anadia	14682	16871	87.0
4	104	Arouca	9682	12337	78.0
5	105	Aveiro	50414	50559	99.0
6	106	Castelo de Paiva	7834	7904	99.0
7	107	Espinho	18258	18385	99.0
8	108	Estarreja	14009	14156	98.0
9	109	Santa Maria da Feira	65079	70801	91.0
10	110	Ilhavo	24983	25089	99.0
11	111	Mealhada	11200	11333	98.0
12	112	Murtosa	8345	8450	96.0
13	113	Oliveira de Azeméis	28372	32400	87.0
14	114	Oliveira do Bairro	13304	13374	99.0
15	115	Ovar	31361	31704	98.0
...	...	...	...	...	...

Close

# Task 1 - Sort rows



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

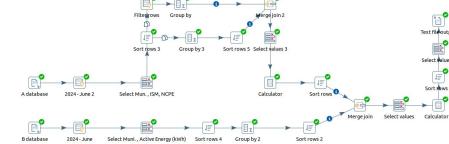
#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	DistrictMunicipalityCode	Y	N	N	0	N

Examine preview data

Rows of step: Sort rows (278 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage
1	101	Águeda	23774	24011	99,0
2	102	Albergaria-a-Velha	13437	13897	96,0
3	103	Anadia	14682	16871	87,0
4	104	Arouca	9682	12337	78,0
5	105	Aveiro	50414	50559	99,0
6	106	Castelo de Paiva	7834	7904	99,0
7	107	Espinho	18258	18385	99,0
8	108	Estarreja	14009	14156	98,0
9	109	Santa Maria da Feira	65079	70801	91,0
10	110	Ilhavo	24983	25089	99,0
11	111	Mealhada	11200	11333	98,0
12	112	Murtosa	8345	8450	98,0
13	113	Oliveira de Azeméis	28372	32400	87,0
14	114	Oliveira do Bairro	13304	13374	99,0
15	115	Ovar	31361	31704	98,0

# Task 1 - B database (Csv File Input)



**CSV file input**

Step name: B database

Filename: /home/aid/Downloads/dataset\_B.csv

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result?

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: mixed

File encoding: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	,	,
2 Month	Integer	#	15	0	\$	,	,
3 Date	String		7		\$	.	,
4 District	String		14		\$	.	,
5 Municipality	String		21		\$	.	,
6 parish	String		30		\$	.	,
7 Voltage level	String		32		\$	.	,
8 Active Energy (kWh)	Number	#.##	11	3	\$	.	,
9 DistrictCode	Integer	#	15	0	\$	,	,
10 DistrictMunicipalityCode	String		4		\$	.	,
11 DistrictMunicipalityParishCode	String		6		\$	.	,
12 mes_int	Integer	#	15	0	\$	.	,

OK Get Fields Preview Cancel

**Examine preview data**

Rows of step: B database (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1
2	2021	1	2021-01	Bragança	Mirandela	MASCARENHAS	Baixa Tensão	88605.5	4	0407	040720	1
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRO VELHO	Baixa Tensão	44866.8	4	0407	040727	1
4	2021	1	2021-01	Bragança	Mirandela	UF AVIADOGOS NAVALI	Baixa Tensão	82565.5	4	0407	040739	1
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCO E VILA BO	Baixa Tensão	49094.8	4	0407	040741	1
6	2021	1	2021-01	Bragança	Mogadouro	MEIRINHOS	Baixa Tensão	29669.1	4	0408	040809	1
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA BEMPOST	Baixa Tensão	34596.4	4	0408	040813	1
8	2021	1	2021-01	Bragança	Torre de Mon	CABECA BOA	Baixa Tensão	52718.7	4	0409	040903	1
9	2021	1	2021-01	Bragança	Torre de Mon	MOS	Baixa Tensão	22259.1	4	0409	040913	1
10	2021	1	2021-01	Bragança	Torre de Mon	UF ADEGANHA E CARE	Muito Alta, Al	355083.9	4	0409	040918	1
11	2021	1	2021-01	Bragança	Torre de Mon	UF URROS PEREDO CA	Baixa Tensão	60462.4	4	0409	040921	1
12	2021	1	2021-01	Bragança	Vila Flor	SEIXO DE MANHOSES	Baixa Tensão	61428.5	4	0410	041013	1
13	2021	1	2021-01	Bragança	Vila Flor	UF VILAS BOAS VILAR	Baixa Tensão	118577.9	4	0410	041024	1
14	2021	1	2021-01	Bragança	Vimioso	UF ALGOSO CAMPO V	Baixa Tensão	99930.1	4	0411	041115	1
15	2021	1	2021-01	Bragança	Vinhais	CELAS	Baixa Tensão	30072.6	4	0412	041204	1
16	2021	1	2021-01	Bragança	Vinhais	SANTALHA	Baixa Tensão	30792.1	4	0412	041221	1
17	2021	1	2021-01	Bragança	Vinhais	VALE DAS FONTES	Baixa Tensão	49676.7	4	0412	041227	1
18	2021	1	2021-01	Bragança	Vinhais	UF MOIMENTA E MON	Baixa Tensão	37859.3	4	0412	041237	1
19	2021	1	2021-01	Bragança	Vinhais	UF SOBREIRO BAIXA O	Baixa Tensão	51699.9	4	0412	041240	1
20	2021	1	2021-01	Castelo Br	OUTROS CASTELO BR	Muito Alta, Al	2230407.6	5	05-	05-	05-	1
21	2021	1	2021-01	Castelo Br	Belmonte	MACAINHAS	Baixa Tensão	50757.7	5	0501	050105	1
22	2021	1	2021-01	Castelo Br	Belmonte	UF BELMONTE E COLN	Baixa Tensão	870454.6	5	0501	050106	1
23	2021	1	2021-01	Castelo Br	Castelo Branc	ALCAINS	Muito Alta, Al	332188.4	5	0502	050201	1
24	2021	1	2021-01	Castelo Br	Castelo Branc	BENQUERENCAS	Baixa Tensão	292019.6	5	0502	050203	1

Close Show Log

# Task 1 - 2024 - June (Filter rows)

**Filter rows**

Step name: **2024 - June**

Send 'true' data to step:

Send 'false' data to step:

The condition:

+

Year = [2024]

AND

Month = [6]

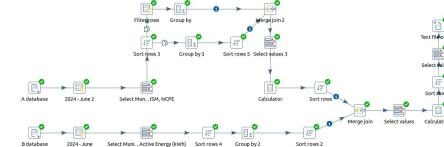
Help      OK      Cancel

**Examine preview data**

Rows of step: 2024 - June (1000 rows)

Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int	
1	2024	6	2024-06	Aveiro	Águeda	FERMENTELOS	Muito Alta, Alta	114091.3	1	0101	010109	6
2	2024	6	2024-06	Aveiro	Águeda	UF BARRO E AGU.	Baixa Tensão	572932.5	1	0101	010122	6
3	2024	6	2024-06	Aveiro	Águeda	UF BARRO E AGU.	Muito Alta, Alta	2673074.3	1	0101	010122	6
4	2024	6	2024-06	Aveiro	Águeda	UF TRAVASSO E C.	Baixa Tensão	307590.3	1	0101	010125	6
5	2024	6	2024-06	Aveiro	Águeda	UF PRESTIMO MA	Muito Alta, Alta	7718.7	1	0101	010127	6
6	2024	6	2024-06	Aveiro	ALBERGARIA-a-	ALQUERUBIM	Baixa Tensão	275300.8	1	0102	010202	6
7	2024	6	2024-06	Aveiro	Albergaria-a-	ANGEJA	Baixa Tensão	247504.8	1	0102	010203	6
8	2024	6	2024-06	Aveiro	Albergaria-a-	ALBERGARIA-A-V	Baixa Tensão	1828453.3	1	0102	010209	6
9	2024	6	2024-06	Aveiro	Albergaria-a-	ALBERGARIA-A-V	Muito Alta, Alta	11207047	1	0102	010209	6
10	2024	6	2024-06	Aveiro	Albergaria-a-	SAO JOAO DE LOI	Muito Alta, Alta	23651.6	1	0102	010210	6
11	2024	6	2024-06	Aveiro	Albergaria-a-	SAO JOAO DE LOI	Baixa Tensão	352699.1	1	0102	010210	6
12	2024	6	2024-06	Aveiro	Anadia	AVELAS DE CAMI	Baixa Tensão	133704.9	1	0103	010304	6
13	2024	6	2024-06	Aveiro	Anadia	AVELAS DE CIMA	Baixa Tensão	231381.7	1	0103	010305	6
14	2024	6	2024-06	Aveiro	Anadia	SANGALHOS	Baixa Tensão	579587.9	1	0103	010309	6
15	2024	6	2024-06	Aveiro	Anadia	SANGALHOS	Muito Alta, Alta	558026.8	1	0103	010309	6
16	2024	6	2024-06	Aveiro	Anadia	VILA NOVA DE M.	Baixa Tensão	200216.1	1	0103	010312	6
17	2024	6	2024-06	Aveiro	Arouca	CHAVE	Baixa Tensão	209192	1	0104	010407	6
18	2024	6	2024-06	Aveiro	Arouca	ESCARIZ	Baixa Tensão	345331.6	1	0104	010409	6
19	2024	6	2024-06	Aveiro	Arouca	FERMEDO	Baixa Tensão	159665.8	1	0104	010411	6
20	2024	6	2024-06	Aveiro	Arouca	MANSORES	Muito Alta, Alta	68662.5	1	0104	010413	6
21	2024	6	2024-06	Aveiro	Arouca	SANTA EULALIA	Baixa Tensão	260861.2	1	0104	010416	6
22	2024	6	2024-06	Aveiro	Arouca	SAO MIGUEL DO	Baixa Tensão	73240	1	0104	010417	6
23	2024	6	2024-06	Aveiro	Arouca	TROPECO	Muito Alta, Alta	58094.8	1	0104	010418	6
24	2024	6	2024-06	Aveiro	Arouca	URRO	Baixa Tensão	99178.4	1	0104	010419	6

Close   Stop   Get more rows



# Task 1 - Select Muni, Active Energy

**Select values**

Step name **Select Muni. , Active Energy (kWh)**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 DistrictMunicipalityCode			
2 Municipality			
3 Active Energy (kWh)			

Get fields to select Edit Mapping

Include unspecified fields, ordered

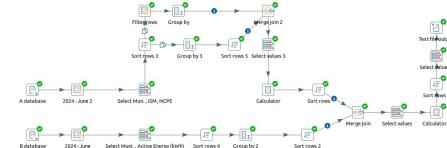
① Help OK Cancel

**Examine preview data**

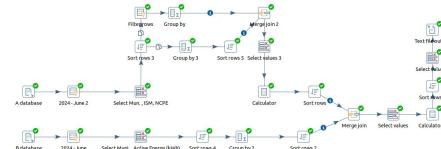
Rows of step: Select Muni. , Active Energy (kWh) (1000 rows)

	DistrictMunicipalityCode	Municipality	Active Energy (kWh)
1	0101	Águeda	114091.3
2	0101	Águeda	572932.5
3	0101	Águeda	2673074.3
4	0101	Águeda	307590.3
5	0101	Águeda	7718.7
6	0102	Albergaria-a-Velha	275300.8
7	0102	Albergaria-a-Velha	247504.8
8	0102	Albergaria-a-Velha	1828453.3
9	0102	Albergaria-a-Velha	11207047
10	0102	Albergaria-a-Velha	23651.6

Close Stop Get more rows



# Task 1 - Sort rows 4



**Sort rows**

Step name	<input type="text" value="Sort rows 4"/>				
Sort directory	<input type="text" value="%%java.io.tmpdir%%"/> <input type="button" value="Browse..."/>				
TMP-file prefix	<input type="text" value="out"/>				
Sort size (rows in memory)	<input type="text" value="1000000"/>				
Free memory threshold (in %)	<input type="text"/>				
Compress TMP Files?	<input type="checkbox"/>				
Only pass unique rows? (verifies keys only)	<input type="checkbox"/>				
<b>Fields :</b>					
Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1 DistrictMunicipalityCode	Y	N	N	0	N

Examine preview data

Rows of step: Sort rows 4 (1000 rows)

	DistrictMunicipalityCode	Municipality	Active Energy (kWh)
1	01-	OUTROS AVEIRO	1155984.1
2	0101	Águeda	114091.3
3	0101	Águeda	572932.5
4	0101	Águeda	2673074.3
5	0101	Águeda	307590.3
6	0101	Águeda	7718.7
7	0101	Águeda	1734175.6
8	0101	Águeda	399145.2
9	0101	Águeda	799683
10	0101	Águeda	746803.8

# Task 1 - 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 DistrictMunicipalityCode  
2 Municipality

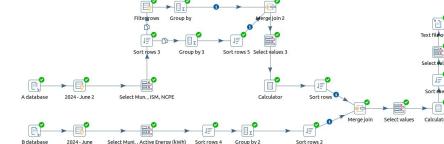
Aggregates :

Name	Subject	Type
1 Sum_ActiveEnergy	Active Energy (kWh)	Sum

**Examine preview data**

Rows of step: Group by 2 (296 rows)

	DistrictMunicipalityCode	Municipality	Sum_ActiveEnergy
1	01-	OUTROS AVEIRO	1155984.144
2	0101	Águeda	22326014.084
3	0102	Albergaria-a-Velha	15763323.427000001
4	0103	Anadia	9956370.766
5	0104	Arouca	4096619.999
6	0105	Aveiro	62176296.87100001
7	0106	Castelo de Paiva	3421087.338000005
8	0107	Espinho	6600028.11
9	0108	Estarreja	54683795.56999999
10	0109	Santa Maria da Feira	46639548.62999999



# Task 1 - Sort rows 2



Sort rows

Step name **Sort rows 2**

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

TMP-file prefix **out**

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

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

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

Examine preview data

Rows of step: Sort rows 2 (296 rows)

	DistrictMunicipalityCode	Municipality	Sum_ActiveEnergy
1	01-	OUTROS AVEIRO	1155984.144
2	0101	Águeda	22326014.084
3	0102	Albergaria-a-Velha	15763323.427000001
4	0103	Anadia	9956370.766
5	0104	Arouca	4096619.999
6	0105	Aveiro	62176296.87100001
7	0106	Castelo de Paiva	3421087.338000005
8	0107	Espinho	6600028.11
9	0108	Estarreja	54683795.56999999
10	0109	Santa Maria da Feira	46639548.62999999

# Task 1 - Merge join



Merge join

Step name:

First Step:

Second Step:

Join Type:

Keys for 1st step:

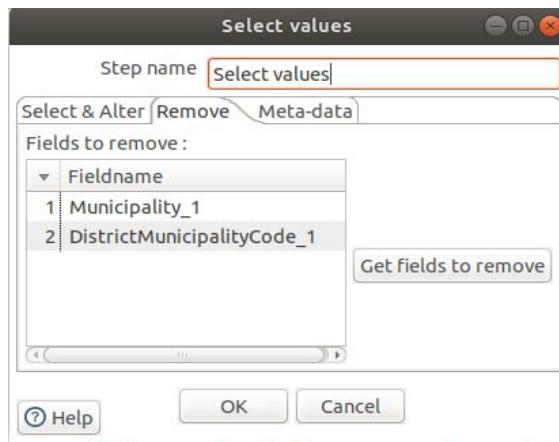
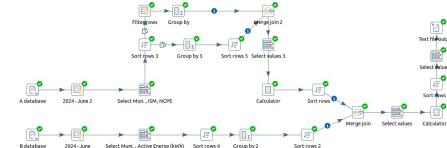
<input type="button" value="Key field"/>	<input type="button" value="Key field"/>
1 Municipality	1 Municipality

Examine preview data

Rows of step: Merge join (135 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage	DistrictMunicipalityCode_1	Municipality_1	Sum_ActiveEnergy
1	101	Águeda	23774	24011	99.0	0101	Águeda	22326014.084
2	102	Albergaria-a-Velha	13437	13897	96.0	0102	Albergaria-a-Velha	15763323.427000001
3	103	Anadia	14682	16871	87.0	0103	Anadia	9956370.766
4	104	Arouca	9682	12337	78.0	0104	Arouca	4096619.999
5	105	Aveiro	50414	50559	99.0	0105	Aveiro	62176296.87100001
6	106	Castelo de Paiva	7834	7904	99.0	0106	Castelo de Paiva	3421087.338000005
7	107	Espinho	18258	18385	99.0	0107	Espinho	6600028.11
8	108	Estarreja	14009	14156	98.0	0108	Estarreja	54683795.56999999
9	109	Santa Maria da Feira	65079	70801	91.0	0109	Santa Maria da Feira	46639548.62999999
10	110	Ilhavo	24983	25089	99.0	0110	Ilhavo	16741754.398000002

# Task 1 - Select values



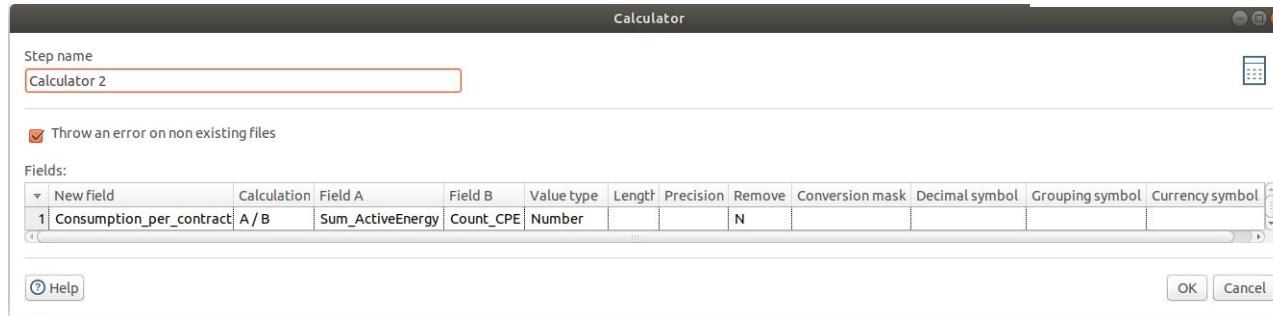
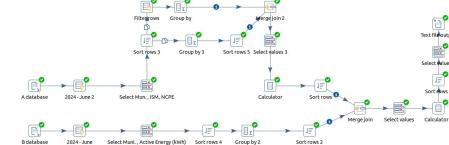
**Examine preview data**

Rows of step: Select values (135 rows)

	DistrictMunicipalityCod	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage	Sum_ActiveEnergy
1	101	Águeda	23774	24011	99.0	22326014.084
2	102	Albergaria-a-Velha	13437	13897	96.0	15763323.427000001
3	103	Anadia	14682	16871	87.0	9956370.766
4	104	Arouca	9682	12337	78.0	4096619.999
5	105	Aveiro	50414	50559	99.0	62176296.87100001
6	106	Castelo de Paiva	7834	7904	99.0	3421087.338000005
7	107	Espinho	18258	18385	99.0	6600028.11
8	108	Estarreja	14009	14156	98.0	54683795.56999999
9	109	Santa Maria da Feira	65079	70801	91.0	46639548.62999999
10	110	Ílhavo	24983	25089	99.0	16741754.398000002

Close

# Task 1 - Calculator 2

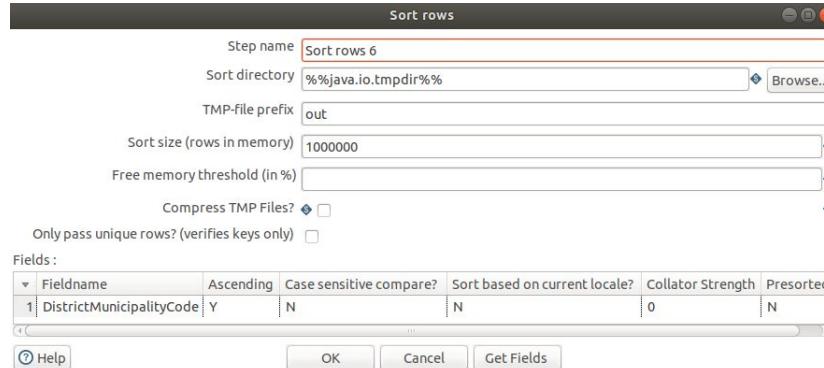
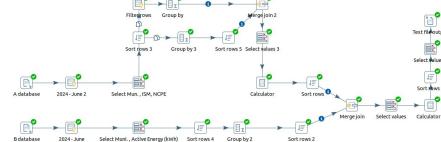


Examine preview data

Rows of step: Calculator 2 (135 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage	Sum_ActiveEnergy	Consumption_per_contract
1	101	Águeda	23774	24011	99.0	22326014.084	929.8244173087
2	102	Albergaria-a-Velha	13437	13897	96.0	15763323.427000001	1134.2968573793
3	103	Anadia	14682	16871	87.0	9956370.766	590.1470432102
4	104	Arouca	9682	12337	78.0	4096619.999	332.0596578585
5	105	Aveiro	50414	50559	99.0	62176296.87100001	1229.7770302221
6	106	Castelo de Paiva	7834	7904	99.0	3421087.338000005	432.8298757591
7	107	Espinho	18258	18385	99.0	6600028.11	358.9898346478
8	108	Estarreja	14009	14156	98.0	54683795.56999999	3862.9411959593
9	109	Santa Maria da Feira	65079	70801	91.0	46639548.62999999	658.7413826076
10	110	Ilhavo	24983	25089	99.0	16741754.39800002	667.2946071187

# Task 1 - Sort rows 6



Examine preview data

Rows of step: Sort rows 6 (135 rows)

	DistrictMunicipalityCode	Municipality	Number_Smart	Count_CPE	SmartMeterPercentage	Sum_ActiveEnergy	Consumption_per_contract
1	101	Águeda	23774	24011	99.0	22326014.084	929.8244173087
2	102	Albergaria-a-V	13437	13897	96.0	15763323.427000001	1134.2968573793
3	103	Anadia	14682	16871	87.0	9956370.766	590.1470432102
4	104	Arouca	9682	12337	78.0	4096619.999	332.0596578585
5	105	Aveiro	50414	50559	99.0	62176296.87100001	1229.7770302221
6	106	Castelo de Pa	7834	7904	99.0	3421087.338000005	432.8298757591
7	107	Espinho	18258	18385	99.0	6600028.11	358.9898346478
8	108	Estarreja	14009	14156	98.0	54683795.56999999	3862.9411959593
9	109	Santa Maria da	65079	70801	91.0	46639548.62999999	658.7413826076
10	110	Ilhavo	24983	25089	99.0	16741754.398000002	667.2946071187

Close

# Task 1 - Select values 2

Select values

Step name **Select values 2**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length
1 Municipality		
2 Consumption_per_contract	Consumption Per Contract	
3 SmartMeterPercentage	Percentage Of Smart Meters	

Include unspecified fields, ordered

Get fields to select Edit Mapping

OK Cancel Help

Examine preview data

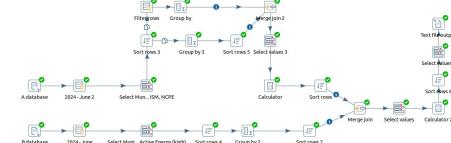
Rows of step: Select values 2 (135 rows)

Municipality	Consumption Per Contract	Percentage Of Smart Meters
1 Águeda	929.8244173087	99.0
2 Albergaria-a-Velha	1134.2968573793	96.0
3 Anadia	590.1470432102	87.0
4 Arouca	332.0596578585	78.0
5 Aveiro	1229.7770302221	99.0
6 Castelo de Paiva	432.8298757591	99.0
7 Espinho	358.9898346478	99.0
8 Estarreja	3862.9411959593	98.0
9 Santa Maria da Feira	658.7413826076	91.0
10 Ilhavo	667.2946071187	99.0

Close



# Task 1 - Text file output



**Text file output**

Step name **Text file output**

**File** **Content** **Fields**

Filename `$(Internal.Entry.Current.Directory)/..../output/task1`

Pass output to servlet

Create Parent folder

Do not create file at start

Accept file name from field?

File name field:

Extension: `csv`

Include stepnr in filename?

Include partition nr in filename?

Include date in filename?

Include time in filename?

Specify Date time Format

Date time format:

Show filename(s)...

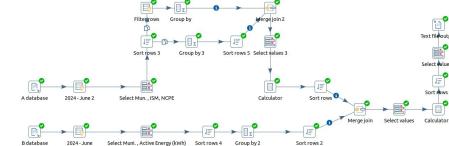
Add filenames to result

**Examine preview data**

Rows of step: Text file output (135 rows)

	Municipality	Consumption Per Contract	Percentage Of Smart Meters
1	Águeda	929.8244173087	99.0
2	Albergaria-a-Velha	1134.2968573793	96.0
3	Anadia	590.1470432102	87.0
4	Arouca	332.0596578585	78.0
5	Aveiro	1229.7770302221	99.0
6	Castelo de Paiva	432.8298757591	99.0
7	Espinho	358.9898346478	99.0
8	Estarreja	3862.9411959593	98.0
9	Santa Maria da Feira	658.7413826076	91.0
10	Ilhavo	667.2946071187	99.0

# Task 1 - output

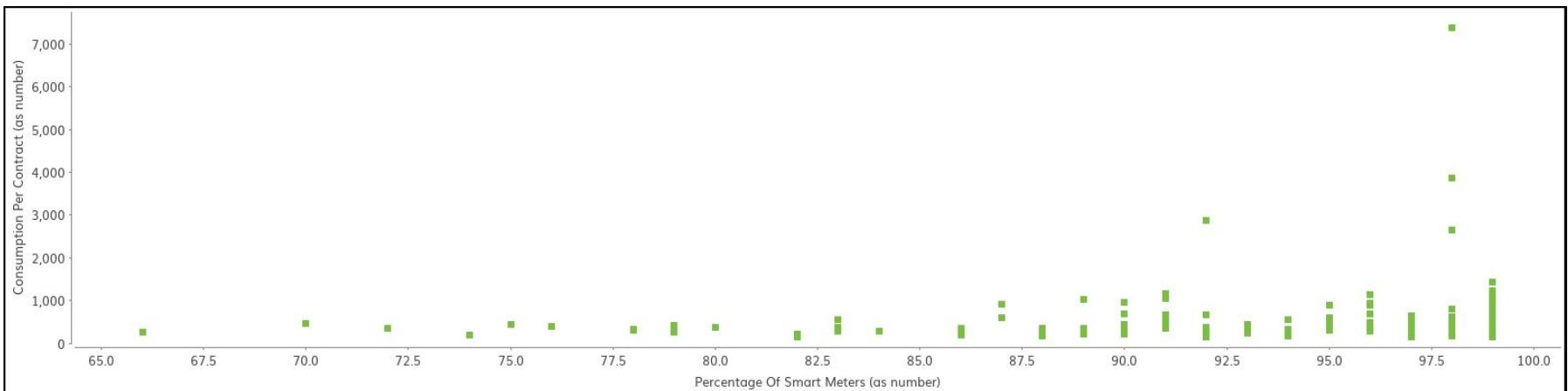


Screenshot of a spreadsheet application showing the output of Task 1. The data is presented in a table with three columns: Municipality, Consumption Per Contract, and Percentage Of Smart Meters.

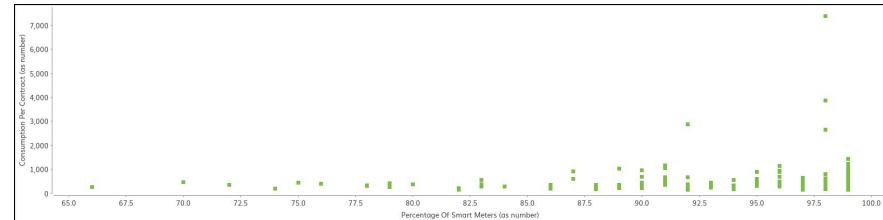
	A	B	C
1	Municipality	Consumption Per Contract	Percentage Of Smart Meters
2	Águeda	929.82442	99
3	Albergaria-a-Velha	1134.29686	96
4	Anadia	590.14704	87
5	Arouca	332.05966	78
6	Aveiro	1229.77703	99
7	Castelo de Paiva	432.82988	99
8	Espinho	358.98983	99
9	Estarreja	3862.9412	98
10	Santa Maria da Feira	658.74138	91
11	Ílhavo	667.29461	99
12	Mealhada	608.46688	98
13	Murtosa	325.25107	98
14	Oliveira de Azeméis	908.12575	87
15	Oliveira do Bairro	909.26477	99
16	Ovar	808.21901	98
17	São João da Madeira	653.5279	99
18	Sever do Vouga	342.0257	96

# Task 2

# Task 2 - Scatter plot



# Task 2 - Correlation



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

There is a correlation between the percentage of smart meters and energy consumption per contract, but it appears to be moderate to weak overall.

In the lower range (65% to 90%): There is little to no correlation, as energy consumption remains mostly low and stable, regardless of the percentage of smart meters.

In the higher range (90% to 100%): There is a clearer positive correlation, with energy consumption increasing as the percentage of smart meters approaches 100%. However, the correlation is not consistent across all contracts, as some contracts still have low consumption while others spike significantly.

Overall, the correlation is not strong across the entire range, but in the higher range it becomes more noticeable.

# Task 2 - Configuration



**task1.csv | DataCleaner**

task1.csv

output.task1.csv

Name	Type
Consumption Per Contract	String
Percentage Of Smart Meters	String

**Scatter plot | DataCleaner**

Scatter plot

**Input columns**

Variable1:	Variable2:
Percentage Of Smart Meters (as number)	Consumption Per Contract (as number)

**Optional properties (1)**

**Close**

**Convert to number | DataCleaner**

# Convert to number

**Input columns**

Input: Select all Select none  Consumption Per Contract  Percentage Of Smart Meters

**Required properties**

Decimal separator:

Minus sign:

Thousand separator: ,

**Optional properties (1)**

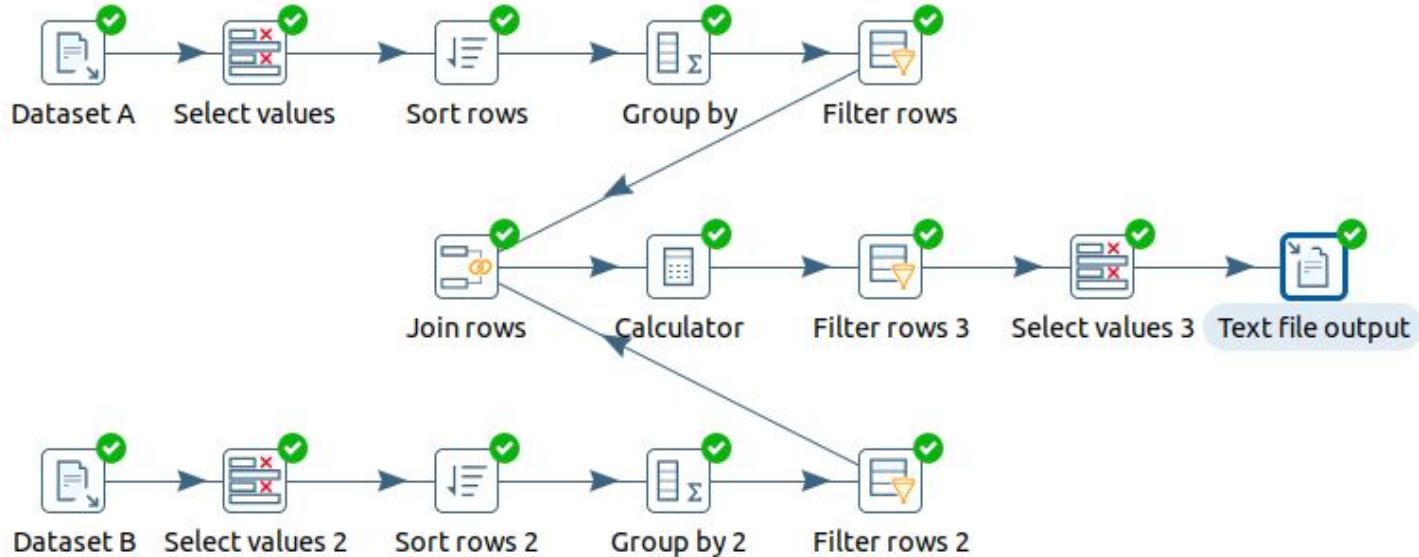
**Output columns**

Name	Type
Consumption Per Contract (as number)	Number
Percentage Of Smart Meters (as number)	Number

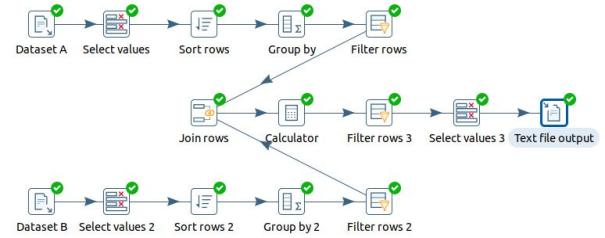
**Write data** **Preview data** **Close**

# Task 3

# Task 3



# Task 3 - Csv File Input Configuration



**CSV file input**

Step name: Dataset A

Filename: /home/aid/Downloads/dataset\_A.csv

Delimiter: ; Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result:

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: mixed

File encoding: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	.	,
2 Month	Integer	#	15	0	\$	.	,
3 Date	String		7		\$	.	,
4 District	String		16		\$	.	,
5 Municipality	String		22		\$	.	,
6 Parish	String		51		\$	.	,
7 Includes Smart Meter	String		3		\$	.	,
8 Number of CPE's	Integer	#	15	0	\$	.	,
9 DistrictCode	Integer	#	15	0	\$	.	,
10 DistrictMunicipalityCode	Integer	#	15	0	\$	.	,
11 DistrictMunicipalityParishCode	String		6		\$	.	,
12 Active Contract	String		3		\$	.	,

Help      OK      Get Fields      Preview      Cancel

**CSV file input**

Step name: Dataset B

Filename: /home/aid/Downloads/dataset\_B.csv

Delimiter: ; Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result:

The row number field name (optional):

Running in parallel?

New line possible in fields?

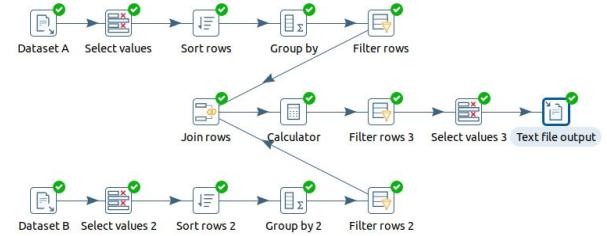
Format: mixed

File encoding: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	.	,
2 Month	Integer	#	15	0	\$	.	,
3 Date	String		7		\$	.	,
4 District	String		14		\$	.	,
5 Municipality	String		21		\$	.	,
6 parish	String		30		\$	.	,
7 Voltage level	String		32		\$	.	,
8 Active Energy (kWh)	Number	#,.#	11	3	\$	.	,
9 DistrictCode	Integer	#	15	0	\$	.	,
10 DistrictMunicipalityCode	String		4		\$	.	,
11 DistrictMunicipalityParishCode	String		6		\$	.	,
12 mes_int	Integer	#	15	0	\$	.	,

Help      OK      Get Fields      Preview      Cancel

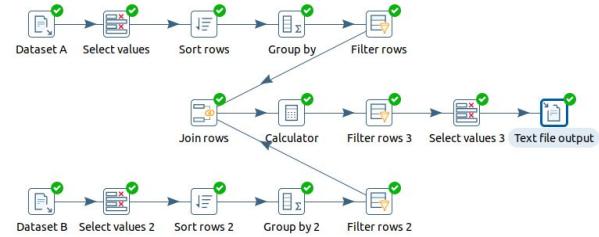
# Task 3 - Csv File Input Preview



Examine preview data													
Rows of step: Dataset A (1000 rows)													
#	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active Contract	
1	2024	5	2024-05	Vila Real	Ribeira de Peña	Alvadia	Sim	155	17	1709	170901	Sim	
2	2024	5	2024-05	Bragança	Mogadouro	Travanca	Não	15	4	408	040820	Sim	
3	2024	5	2024-05	Portalegre	Ponte de Sor	Montargil	Não	475	12	1213	121302	Sim	
4	2024	5	2024-05	Viseu	Santa Comba Dão	Santa Comba Dão e Couto do Mosteiro	Sim	3141	18	1814	181411	Sim	
5	2024	5	2024-05	Portalegre	Arronches	Assunção	Sim	1248	12	1202	120201	Sim	
6	2024	5	2024-05	Santarém	Tomar	Asseiceira	Sim	1549	14	1418	141802	Sim	
7	2024	5	2024-05	Coimbra	Mira	Carapelhos	Sim	488	6	608	060803	Sim	
8	2024	5	2024-05	Guarda	Celorico da Beira	Prados	Sim	221	9	903	090312	Sim	
9	2024	5	2024-05	Viana do Castelo	Paredes de Coura	Agualonga	Sim	175	16	1605	160501	Sim	
10	2024	5	2024-05	Beja	Odemira	Reliquias	Sim	334	2	211	021102	Sim	

Examine preview data												
Rows of step: Dataset B (1000 rows)												
#	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1
2	2021	1	2021-01	Bragança	Mirandela	MASCARENHAS	Baixa Tensão	88605.5	4	0407	040720	1
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRO VELHO	Baixa Tensão	44866.8	4	0407	040727	1
4	2021	1	2021-01	Bragança	Mirandela	UF AVIDAGOS NAVALI	Baixa Tensão	82565.5	4	0407	040739	1
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCO E VILA BC	Baixa Tensão	49094.8	4	0407	040741	1
6	2021	1	2021-01	Bragança	Mogadouro	MEIRINHOS	Baixa Tensão	29669.1	4	0408	040809	1
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA BEMPOST	Baixa Tensão	34596.4	4	0408	040813	1
8	2021	1	2021-01	Bragança	Torre de Moncorvo	CABECA BOA	Baixa Tensão	52718.7	4	0409	040903	1
9	2021	1	2021-01	Bragança	Torre de Moncorvo	MOS	Baixa Tensão	22259.1	4	0409	040913	1
10	2021	1	2021-01	Bragança	Torre de Moncorvo	UF ADEGANHA E CAR	Muito Alta, Alt	355083.9	4	0409	040918	1

# Task 3 - Select Values Configuration



Select values

Step name **Select values**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length	Precision
1 District	District_A		

Include unspecified fields, ordered |

OK Cancel

Select values

Step name **Select values 2**

Select & Alter Remove Meta-data

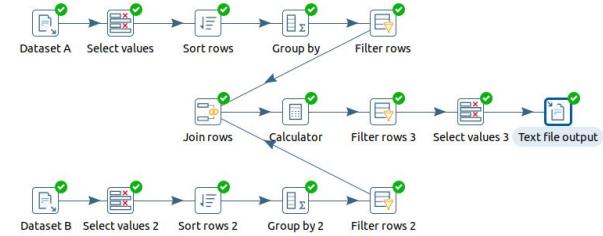
Fields :

Fieldname	Rename to	Length	Precision
1 District	District_B		

Include unspecified fields, ordered |

OK Cancel

# Task 3 - Select Values Preview



Examine preview data

Rows of step: Select values (1000 rows)

District_A
1 Vila Real
2 Bragança
3 Portalegre
4 Viseu
5 Portalegre
6 Santaém
7 Coimbra
8 Guarda
9 Viana do Castelo
10 Beja

Close Stop Get more rows

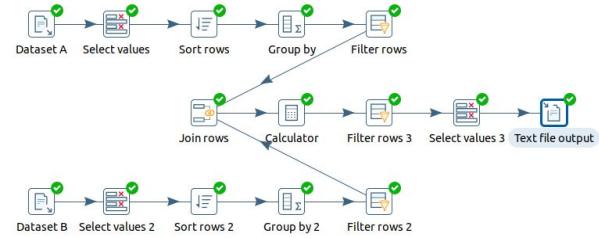
Examine preview data

Rows of step: Select values 2 (1000 rows)

District_B
1 Bragança
2 Bragança
3 Bragança
4 Bragança
5 Bragança
6 Bragança
7 Bragança
8 Bragança
9 Bragança
10 Bragança

Close Stop Get more rows

# Task 3 - Sort Rows Configuration



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys)

Fields :

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

Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

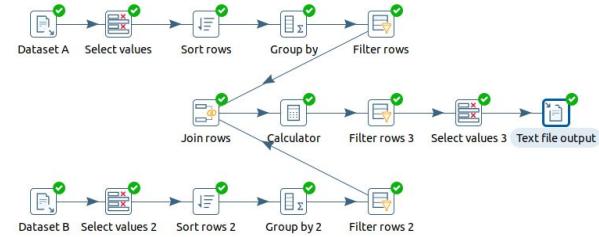
Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

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

# Task 3 - Sort Rows Preview



**Examine preview data**

Rows of step: Sort rows (1000 rows)

District_A
1 <null>
2 Aveiro
3 Aveiro
4 Aveiro
5 Aveiro
6 Aveiro
7 Aveiro
8 Aveiro
9 Aveiro
10 Aveiro

**Close** **Stop** **Get more rows**

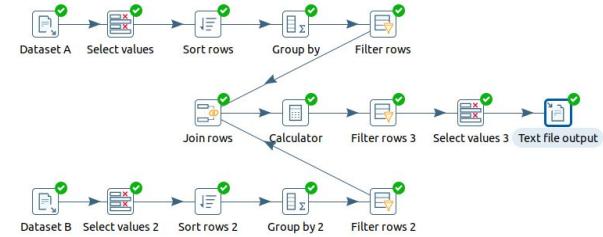
**Examine preview data**

Rows of step: Sort rows 2 (1000 rows)

District_B
1 Aveiro
2 Aveiro
3 Aveiro
4 Aveiro
5 Aveiro
6 Aveiro
7 Aveiro
8 Aveiro
9 Aveiro
10 Aveiro

**Close** **Stop** **Get more rows**

# Task 3 - Group by Configuration



**Group by**

Step name **Group by**

Include all rows?

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

TMP-file prefix `grp`

Add line number, res

Line number field nam

Always give back a res

The fields that make up the group:

Group field   
1 District\_A

Aggregates :

Name	Subject	Type	Value
1			

**Group by**

Step name **Group by 2**

Include all rows?

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

TMP-file prefix `grp`

Add line number, res

Line number field nam

Always give back a res

The fields that make up the group:

Group field   
1 District\_B

Aggregates :

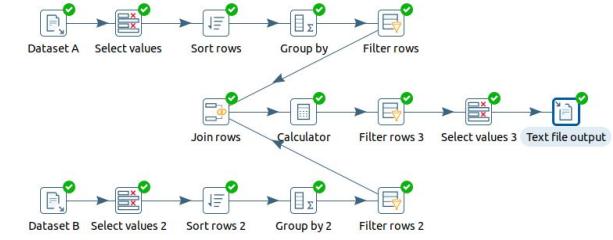
Name	Subject	Type	Value
1			

# Task 3 - Group by Preview

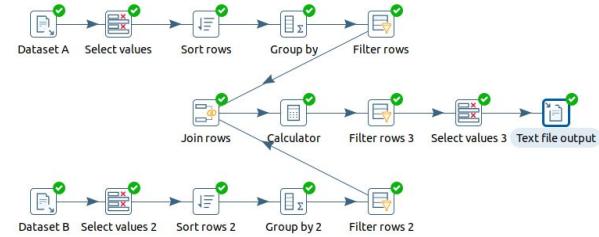
Examine preview data	
Rows of step: Group by (19 rows)	
▼	District_A
1	<null>
2	Aveiro
3	Beja
4	Braga
5	Bragança
6	Castelo Branco
7	Coimbra
8	Faro
9	Guarda
10	Leiria
11	Lisboa
12	Portalegre
13	Porto
14	Santarém
15	Setúbal
16	Viana do Castelo
17	Vila Real
18	Viseu
19	Évora

Close

Examine preview data	
Rows of step: Group by 2 (18 rows)	
▼	District_B
1	Aveiro
2	Beja
3	Braga
4	Bragança
5	Castelo Branco
6	Coimbra
7	Faro
8	Guarda
9	Leiria
10	Lisboa
11	Portalegre
12	Porto
13	Santarém
14	Setúbal
15	Viana do Castelo
16	Vila Real
17	Viseu
18	Évora



# Task 3 - Filter Rows Configuration



**Filter rows**

Step name: **Filter rows**

Send 'true' data to step:

Send 'false' data to step:

The condition:

District\_A IS NOT NULL

Help OK Cancel

**Filter rows**

Step name: **Filter rows 2**

Send 'true' data to step:

Send 'false' data to step:

The condition:

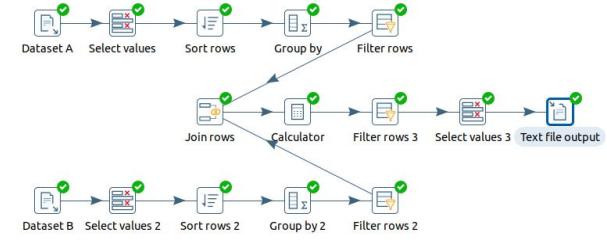
District\_B IS NOT NULL

Help OK Cancel

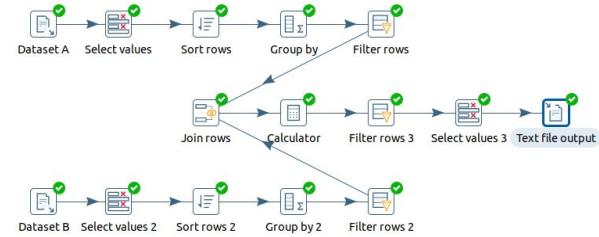
# Task 3 - Filter Rows Preview

Examine preview data	
Rows of step: Filter rows (18 rows)	
	District_A
1	Aveiro
2	Beja
3	Braga
4	Bragança
5	Castelo Branco
6	Coimbra
7	Faro
8	Guarda
9	Leiria
10	Lisboa
11	Portalegre
12	Porto
13	Santarém
14	Setúbal
15	Viana do Castelo
16	Vila Real
17	Viseu
18	Évora

Examine preview data	
Rows of step: Filter rows 2 (18 rows)	
	District_B
1	Aveiro
2	Beja
3	Braga
4	Bragança
5	Castelo Branco
6	Coimbra
7	Faro
8	Guarda
9	Leiria
10	Lisboa
11	Portalegre
12	Porto
13	Santarém
14	Setúbal
15	Viana do Castelo
16	Vila Real
17	Viseu
18	Évora



# Task 3 - Join Rows



**Join rows (cartesian product)**

Step name: **Join rows**

Temp directory: `%%java.io.tmpdir%%`

TMP-file prefix: `out`

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

Main step to read from:

The condition:

`To edit a subcondition + si`

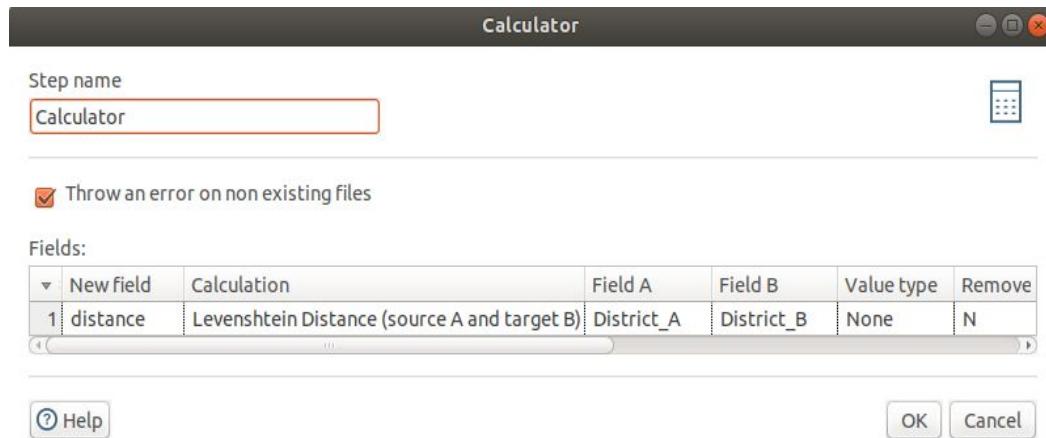
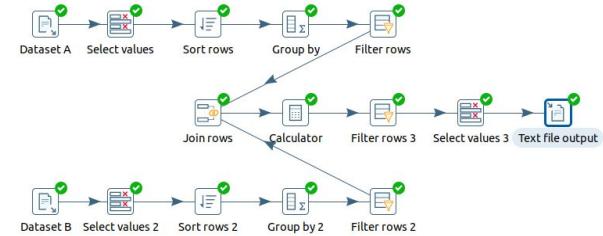
`District_A <= District_B`

**Examine preview data**

Rows of step: Join rows (171 rows)

	District_A	District_B
1	Aveiro	Aveiro
2	Aveiro	Beja
3	Aveiro	Braga
4	Aveiro	Bragança
5	Aveiro	Castelo Branco
6	Aveiro	Coimbra
7	Aveiro	Faro
8	Aveiro	Guarda
9	Aveiro	Leiria
10	Aveiro	Lisboa

# Task 3 - Calculator

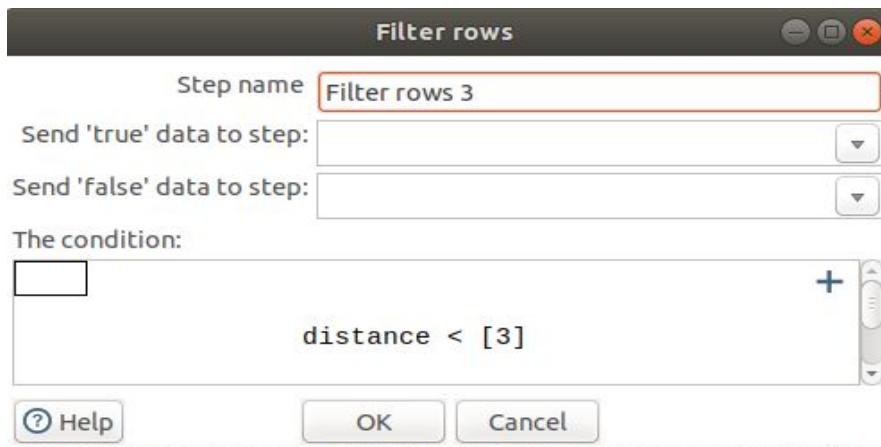
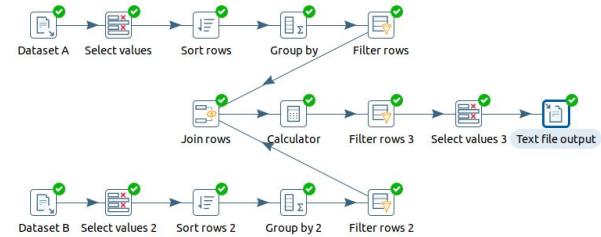


Examine preview data  
Rows of step: Calculator (171 rows)

	District_A	District_B	distance
1	Aveiro	Aveiro	0
2	Aveiro	Beja	5
3	Aveiro	Braga	6
4	Aveiro	Bragança	8
5	Aveiro	Castelo Branco	11
6	Aveiro	Coimbra	6
7	Aveiro	Faro	4
8	Aveiro	Guarda	6
9	Aveiro	Leiria	4
10	Aveiro	Lisboa	6

Close

# Task 3 - Filter Rows 3



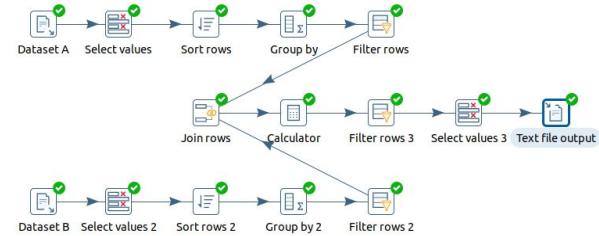
Examine preview data

Rows of step: Filter rows 3 (18 rows)

	District_A	District_B	distance
1	Aveiro	Aveiro	0
2	Beja	Beja	0
3	Braga	Braga	0
4	Bragança	Bragança	0
5	Castelo Branco	Castelo Branco	0
6	Coimbra	Coimbra	0
7	Faro	Faro	0
8	Guarda	Guarda	0
9	Leiria	Leiria	0
10	Lisboa	Lisboa	0
11	Portalegre	Portalegre	0
12	Porto	Porto	0
13	Santarém	Santarém	0
14	Setúbal	Setúbal	0
15	Viana do Castelo	Viana do Castelo	0
16	Vila Real	Vila Real	0
17	Viseu	Viseu	0
18	Évora	Évora	0

Close

# Task 3 - Select Values 3



Select values

Step name **Select values 3**

Select & Alter Remove Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	District_A			
2	District_B			

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

Include unspecified fields, or

**OK** **Cancel** **Help**

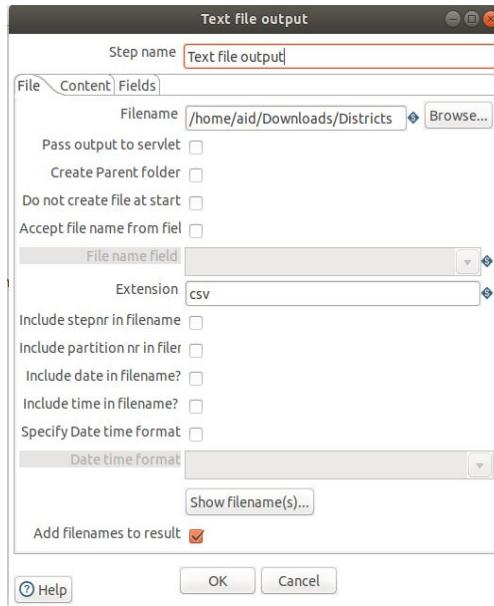
Examine preview data

Rows of step: Select values 3 (18 rows)

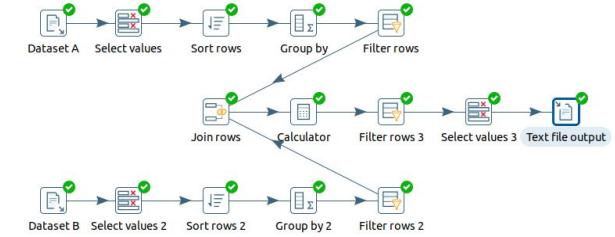
	District_A	District_B
1	Aveiro	Aveiro
2	Beja	Beja
3	Braga	Braga
4	Bragança	Bragança
5	Castelo Branco	Castelo Branco
6	Coimbra	Coimbra
7	Faro	Faro
8	Guarda	Guarda
9	Leiria	Leiria
10	Lisboa	Lisboa
11	Portalegre	Portalegre
12	Porto	Porto
13	Santarém	Santarém
14	Setúbal	Setúbal
15	Viana do Castelo	Viana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

**Close**

# Task 3 - Text File Output



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



Districts.csv - LibreOffice Calc		
	District_A	District_B
1	District_A	District_B
2	Aveiro	Aveiro
3	Beja	Beja
4	Braga	Braga
5	Bragança	Bragança
6	Castelo Branco	Castelo Branco
7	Coimbra	Coimbra
8	Faro	Faro
9	Guarda	Guarda
10	Leiria	Leiria
11	Lisboa	Lisboa
12	Portalegre	Portalegre
13	Porto	Porto
14	Santarém	Santarém
15	Setúbal	Setúbal
16	Viana do Castelo	Viana do Castelo
17	Vila Real	Vila Real
18	Viseu	Viseu
19	Évora	Évora
20		

# Task 4

# Task 4 - SQL

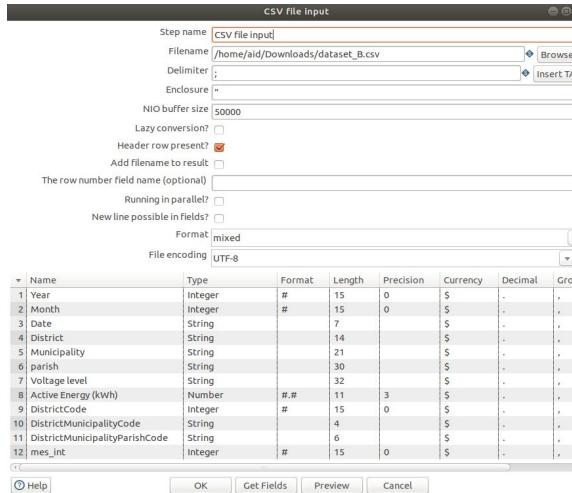
```
1  -- Create schema for the data warehouse
2
3  DROP DATABASE IF EXISTS EnergyDataWarehouse;
4
5  CREATE DATABASE EnergyDataWarehouse;
6
7  USE EnergyDataWarehouse;
8
9  -- Dimension tables
10
11 CREATE TABLE dim_time (
12     time_id INT PRIMARY KEY AUTO_INCREMENT,
13     year INT NOT NULL,
14     season ENUM('Spring', 'Summer', 'Autumn', 'Winter') NOT NULL,
15     month INT NOT NULL,
16     UNIQUE KEY (year, month)
17 );
18
19 CREATE TABLE dim_location (
20     location_id INT PRIMARY KEY AUTO_INCREMENT,
21     district VARCHAR(255) NOT NULL,
22     municipality VARCHAR(255) NOT NULL,
23     parish VARCHAR(255),
24     municipalityparishcode INT,
25     UNIQUE KEY (district, municipality, parish, municipalityparishcode)
26 );
27
28 -- Fact tables
29
30 CREATE TABLE fact_energy_consumption (
31     time_id INT ,
32     location_id INT ,
33     energy_consumption DECIMAL(15, 3) NOT NULL,
34     count_smart INT NOT NULL,
35     count_non_smart INT NOT NULL,
36     FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
37     FOREIGN KEY (location_id) REFERENCES dim_location(location_id),
38     PRIMARY KEY (time_id, location_id)
39 );
```

# Task 5

# Task 5 - Time Dimension (dim\_time)



# Task 5 - Time Dimension - Csv file input



**Examine preview data**

Rows of step: CSV file input (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1
2	2021	1	2021-01	Bragança	Mirandela	MASCARENHAS	Baixa Tensão	88605.5	4	0407	040720	1
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRO VELHO	Baixa Tensão	44806.8	4	0407	040727	1
4	2021	1	2021-01	Bragança	Mirandela	UF AVIADIGOS NAVALHO	Baixa Tensão	82565.5	4	0407	040739	1
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCO E VILA BOA	Baixa Tensão	49094.8	4	0407	040741	1
6	2021	1	2021-01	Bragança	Mogadouro	MEIRINHOS	Baixa Tensão	29669.1	4	0408	040809	1
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA BEMPOSTA	Baixa Tensão	34596.4	4	0408	040813	1
8	2021	1	2021-01	Bragança	Torre de Moncorvo	CABECA BOA	Baixa Tensão	52718.7	4	0409	040903	1
9	2021	1	2021-01	Bragança	Torre de Moncorvo	MOS	Baixa Tensão	22259.1	4	0409	040913	1
10	2021	1	2021-01	Bragança	Torre de Moncorvo	UF ADEGANHA E CARDAL	Muito Alta, A	355083.9	4	0409	040918	1
11	2021	1	2021-01	Bragança	Torre de Moncorvo	UF URROS PEREDO CAST	Baixa Tensão	60646.2	4	0409	040921	1
12	2021	1	2021-01	Bragança	Vila Flor	SEIXO DE MANHOSES	Baixa Tensão	61428.5	4	0410	041013	1
13	2021	1	2021-01	Bragança	Vila Flor	UF VILAS BOAS VILARIN	Baixa Tensão	118577.9	4	0410	041024	1
14	2021	1	2021-01	Bragança	Vimioso	UF ALCOSO CAMPO VIB	Baixa Tensão	99930.1	4	0411	041115	1
15	2021	1	2021-01	Bragança	Vinhas	CELAS	Baixa Tensão	30072.6	4	0412	041204	1
16	2021	1	2021-01	Bragança	Vinhas	SANTALHA	Baixa Tensão	30792.1	4	0412	041221	1
17	2021	1	2021-01	Bragança	Vinhas	VALE DAS FONTES	Baixa Tensão	49767.7	4	0412	041227	1
18	2021	1	2021-01	Bragança	Vinhas	UF MOIMENTA E MONT	Baixa Tensão	37859.3	4	0412	041237	1
19	2021	1	2021-01	Bragança	Vinhas	UF SOBREIRO BAIXO ALV	Baixa Tensão	51699.9	4	0412	041240	1
20	2021	1	2021-01	Castelo Br	OUTROS CASTELO E	OUTROS CASTELO BRAN	Muito Alta, A	2230407.6	5	05-	05---	1
21	2021	1	2021-01	Castelo Br	Belmonte	MACAINHAS	Baixa Tensão	50577.5	5	0501	050105	1
22	2021	1	2021-01	Castelo Br	Belmonte	UF BELMONTE E COLME	Baixa Tensão	870454.6	5	0501	050106	1
23	2021	1	2021-01	Castelo Br	Castelo Branco	ALCAINES	Muito Alta, A	332188.4	5	0502	050201	1
24	2021	1	2021-01	Castelo Br	Castelo Branco	BENQUERENCIAS	Baixa Tensão	292019.6	5	0502	050203	1
25	2021	1	2021-01	Castelo Br	Castelo Branco	CASTELO BRANCO	Muito Alta, A	3976744.7	5	0502	050205	1

Buttons: Close, Show Log.

# Task 5 - Time Dimension - Select values



**Select values**

Step name **Select values**

Select & Alter Remove Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	Year	year		
2	Month	month		

**Get fields to select**

**Edit Mapping**

Include unspecified fields, ordered

OK Cancel

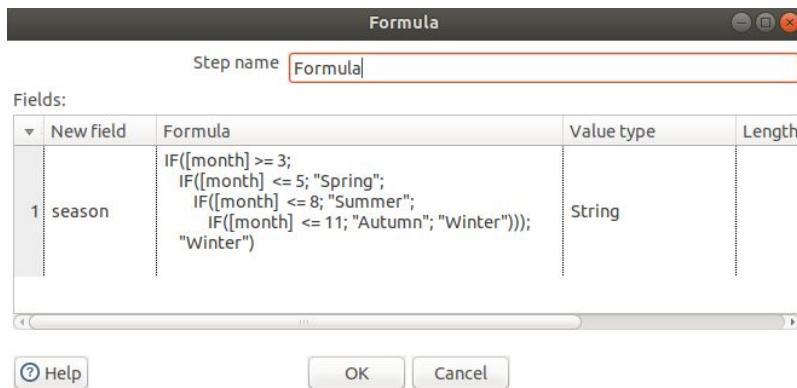
**Examine preview data**

Rows of step: Select values (1000 rows)

	year	month
1	2021	1
2	2021	1
3	2021	1
4	2021	1
5	2021	1
6	2021	1
7	2021	1
8	2021	1
9	2021	1
10	2021	1
...		

Close Stop Get more rows

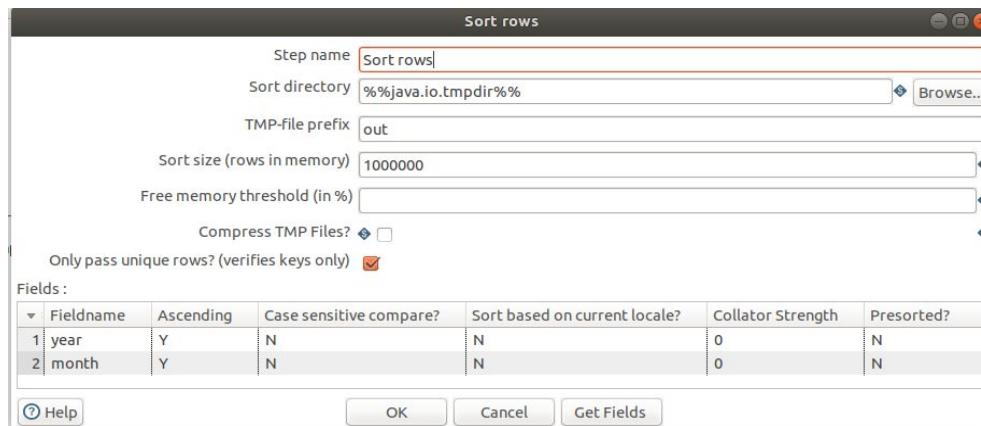
# Task 5 - Time Dimension - Formula



Examine preview data			
	year	month	season
1	2021	1	Winter
2	2021	1	Winter
3	2021	1	Winter
4	2021	1	Winter
5	2021	1	Winter
6	2021	1	Winter
7	2021	1	Winter
8	2021	1	Winter
9	2021	1	Winter
10	2021	1	Winter

Buttons at the bottom: Close, Stop, Get more rows.

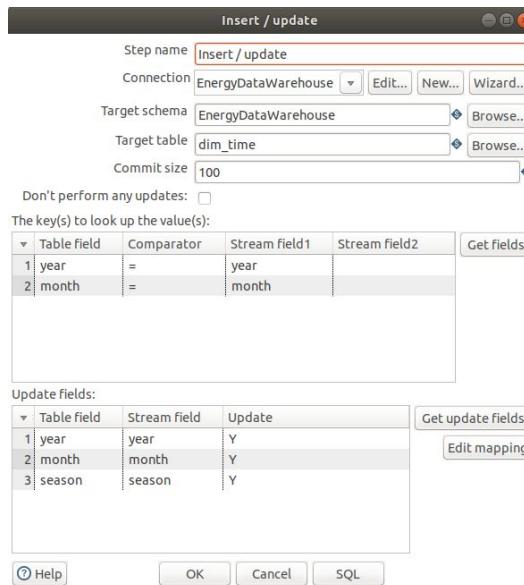
# Task 5 - Time Dimension - Sort Rows



Examine preview data			
	year	month	season
1	2020	11	Autumn
2	2020	12	Winter
3	2021	1	Winter
4	2021	2	Winter
5	2021	3	Spring
6	2021	4	Spring
7	2021	5	Spring
8	2021	6	Summer
9	2021	7	Summer
10	2021	8	Summer

Close

# Task 5 - Time Dimension - Insert / Update



Examine preview data

Rows of step: Insert / update (45 rows)

	year	month	season
1	2020	11	Autumn
2	2020	12	Winter
3	2021	1	Winter
4	2021	2	Winter
5	2021	3	Spring
6	2021	4	Spring
7	2021	5	Spring
8	2021	6	Summer
9	2021	7	Summer
10	2021	8	Summer
11	2021	9	Autumn
12	2021	10	Autumn
13	2021	11	Autumn
14	2021	12	Winter
15	2022	1	Winter
16	2022	2	Winter
17	2022	3	Spring
18	2022	4	Spring
19	2022	5	Spring
20	2022	6	Summer
21	2022	7	Summer
22	2022	8	Summer
23	2022	9	Autumn
24	2022	10	Autumn
25	2022	11	Autumn
26	2022	12	Winter
27	2023	1	Winter
28	2023	2	Winter
29	2023	3	Spring

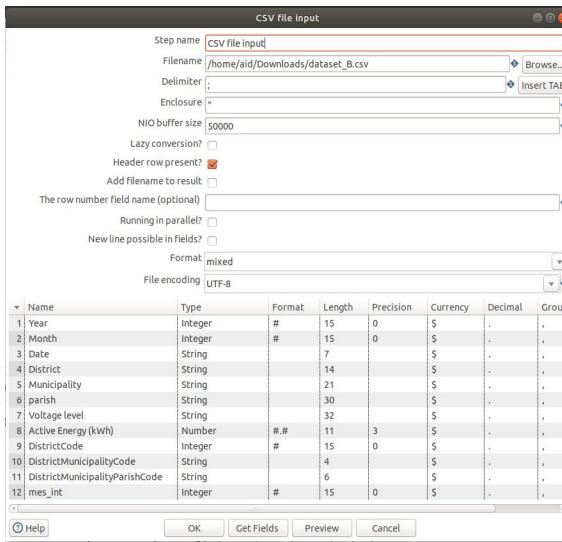
Close

```
mysql> select * from dim_time;
+-----+-----+-----+-----+
| time_id | year | season | month |
+-----+-----+-----+-----+
| 1 | 2020 | Autumn | 11 |
| 2 | 2020 | Winter | 12 |
| 3 | 2021 | Winter | 1 |
| 4 | 2021 | Winter | 2 |
| 5 | 2021 | Spring | 3 |
| 6 | 2021 | Spring | 4 |
| 7 | 2021 | Spring | 5 |
| 8 | 2021 | Summer | 6 |
| 9 | 2021 | Summer | 7 |
| 10 | 2021 | Summer | 8 |
| 11 | 2021 | Autumn | 9 |
| 12 | 2021 | Autumn | 10 |
| 13 | 2021 | Autumn | 11 |
| 14 | 2021 | Winter | 12 |
| 15 | 2022 | Winter | 1 |
| 16 | 2022 | Winter | 2 |
| 17 | 2022 | Spring | 3 |
| 18 | 2022 | Spring | 4 |
| 19 | 2022 | Spring | 5 |
| 20 | 2022 | Summer | 6 |
| 21 | 2022 | Summer | 7 |
| 22 | 2022 | Summer | 8 |
| 23 | 2022 | Autumn | 9 |
| 24 | 2022 | Autumn | 10 |
| 25 | 2022 | Autumn | 11 |
| 26 | 2022 | Winter | 12 |
| 27 | 2023 | Winter | 1 |
| 28 | 2023 | Winter | 2 |
| 29 | 2023 | Spring | 3 |
```

# Task 5 - Location Dimension (dim\_location)



# Task 5 - Location Dimension - Csv file input



**Examine preview data**

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1
2	2021	1	2021-01	Bragança	Mirandela	MASCAREN	Baixa Tensão	88605.5	4	0407	040720	1
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRC	Baixa Tensão	44866.8	4	0407	040727	1
4	2021	1	2021-01	Bragança	Mirandela	UF AVIDAGI	Baixa Tensão	82565.5	4	0407	040739	1
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCC	Baixa Tensão	49094.8	4	0407	040741	1
6	2021	1	2021-01	Bragança	Mogadouro	MEIRINHOS	Baixa Tensão	29669.1	4	0408	040809	1
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA	Baixa Tensão	34596.4	4	0408	040813	1
8	2021	1	2021-01	Bragança	Torre de Mor	CABECA BC	Baixa Tensão	52718.7	4	0409	040903	1
9	2021	1	2021-01	Bragança	Torre de Mor	MOS	Baixa Tensão	22259.1	4	0409	040913	1
10	2021	1	2021-01	Bragança	Torre de Mor	UF ADEGAN	Muito Alta, Alta	355083.9	4	0409	040918	1
11	2021	1	2021-01	Bragança	Torre de Mor	UF URROS F	Baixa Tensão	60646.2	4	0409	040921	1
12	2021	1	2021-01	Bragança	Vila Flor	SEIXO DE M	Baixa Tensão	61428.5	4	0410	041013	1
13	2021	1	2021-01	Bragança	Vila Flor	UF VILAS BI	Baixa Tensão	118577.9	4	0410	041024	1
14	2021	1	2021-01	Bragança	Vimioso	UF ALGOSC	Baixa Tensão	99930.1	4	0411	041115	1
15	2021	1	2021-01	Bragança	Vinhais	CELAS	Baixa Tensão	30072.6	4	0412	041204	1
16	2021	1	2021-01	Bragança	Vinhais	SANTALHA	Baixa Tensão	30792.1	4	0412	041221	1
17	2021	1	2021-01	Bragança	Vinhais	VALE DAS F	Baixa Tensão	49676.7	4	0412	041227	1
18	2021	1	2021-01	Bragança	Vinhais	UF MOIMEN	Baixa Tensão	37859.3	4	0412	041237	1
19	2021	1	2021-01	Bragança	Vinhais	UF SOBREIF	Baixa Tensão	51699.9	4	0412	041240	1
20	2021	1	2021-01	Castelo B	OUTROS CAS	OUTROS CA	Muito Alta, Alta	2230407.6	5	05-	05--	1

Close Show Log

# Task 5 - Location Dimension - Select values



**Select values**

Step name **Select values**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length
1 Municipality	municipality	
2 District	district	
3 parish		
4 DistrictMunicipalityParishCode	municipalityparishcode	

Include unspecified fields, ordered

Get fields to select Edit Mapping

OK Cancel

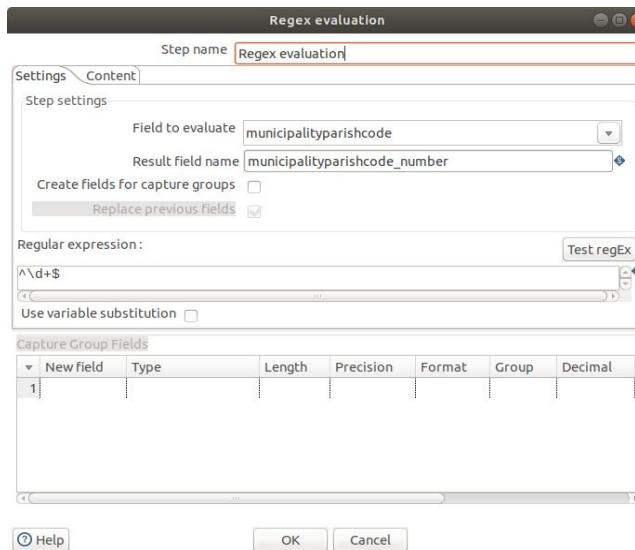
**Examine preview data**

Rows of step: Select values (1000 rows)

	municipality	district	parish	municipalityparishcode
1	Mirandela	Bragança	FRECHAS	040716
2	Mirandela	Bragança	MASCARENHAS	040720
3	Mirandela	Bragança	SAO PEDRO VELHO	040727
4	Mirandela	Bragança	UF AVIDAGOS NA	040739
5	Mirandela	Bragança	UF FRANCO E VIL	040741
6	Mogadouro	Bragança	MEIRINHOS	040809
7	Mogadouro	Bragança	PEREDO DA BEMposta	040813
8	Torre de Moel	Bragança	CABECA BOA	040903
9	Torre de Moel	Bragança	MOS	040913
10	Torre de Moel	Bragança	UF ADEGANHA E	040918

Close Stop Get more rows

# Task 5 - Location Dimension - Regex evaluation



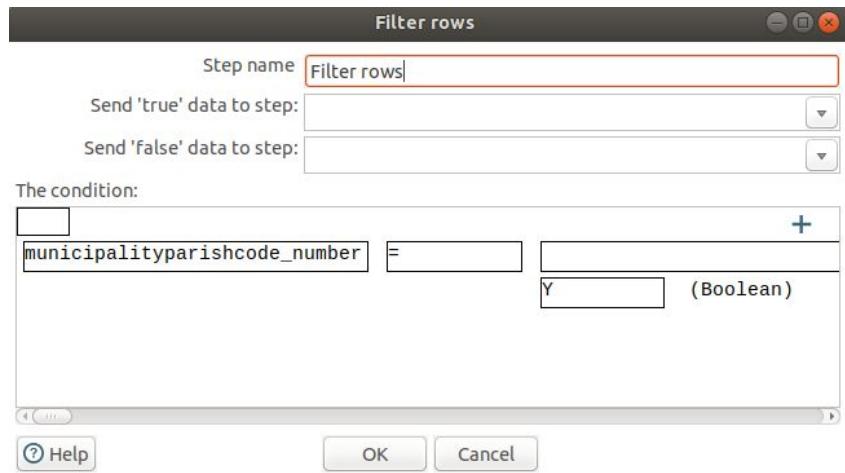
**Examine preview data**

Rows of step: **Regex evaluation (1000 rows)**

	municipality	district	parish	municipalityparishcode	municipalityparishcode_number
1	Mirandela	Bragança	FRECHAS	040716	Y
2	Mirandela	Bragança	MASCARENHAS	040720	Y
3	Mirandela	Bragança	SAO PEDRO VELHO	040727	Y
4	Mirandela	Bragança	UF AVIDAGOS NAVALH	040739	Y
5	Mirandela	Bragança	UF FRANCO E VILA BOA	040741	Y
6	Mogadouro	Bragança	MEIRINHOS	040809	Y
7	Mogadouro	Bragança	PEREDO DA BEMPOSTA	040813	Y
8	Torre de Mon	Bragança	CABECA BOA	040903	Y
9	Torre de Mon	Bragança	MOS	040913	Y
10	Torre de Mon	Bragança	UF ADEGANHA E CARD	040918	Y
11	Torre de Mon	Bragança	UF URROS PEREDO CAS	040921	Y
12	Vila Flor	Bragança	SEIXO DE MANHOSES	041013	Y
13	Vila Flor	Bragança	UF VILAS BOAS VILARII	041024	Y
14	Vimioso	Bragança	UF ALGOSO CAMPO VII	041115	Y

Close Stop Get more rows

# Task 5 - Location Dimension - Filter rows



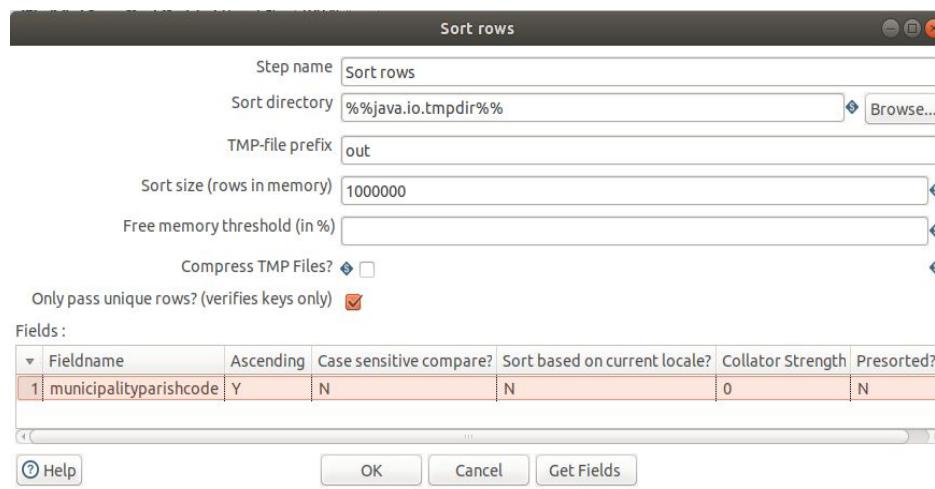
Examine preview data

Rows of step: Filter rows (1000 rows)

	municipality	district	parish	municipalityparishcode	municipalityparishcode_number
1	Mirandela	Bragança	FRECHAS	040716	Y
2	Mirandela	Bragança	MASCARENHAS	040720	Y
3	Mirandela	Bragança	SAO PEDRO VELHO	040727	Y
4	Mirandela	Bragança	UF AVIDAGOS NAVAI	040739	Y
5	Mirandela	Bragança	UF FRANCO EVILA	040741	Y
6	Mogadouro	Bragança	MEIRINHOS	040809	Y
7	Mogadouro	Bragança	PEREDO DA BEMPO	040813	Y
8	Torre de Monc	Bragança	CABECA BOA	040903	Y
9	Torre de Monc	Bragança	MOS	040913	Y
10	Torre de Monc	Bragança	UF ADEGANHA E CA	040918	Y
11	Torre de Monc	Bragança	UF URROS PEREDO	040921	Y
12	Vila Flor	Bragança	SEIXO DE MANHOSI	041013	Y
13	Vila Flor	Bragança	UF VILAS BOAS VIL	041024	Y
14	Vimioso	Bragança	UF ALGOSO CAMPC	041115	Y

Close Stop Get more rows

# Task 5 - Location Dimension - Sort rows

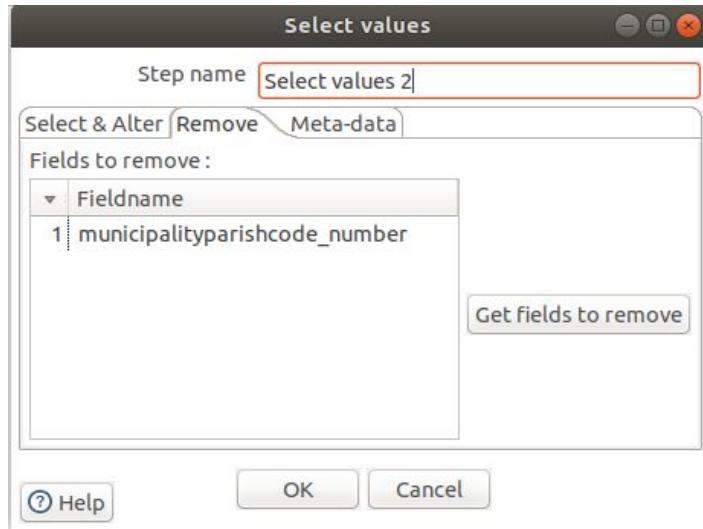


**Examine preview data**

Rows of step: Sort rows (1000 rows)

	municipality	district	parish	municipalityparishcode	municipalityparishcode_number
1	Águeda	Aveiro	AGUADA DE CIMA	010103	Y
2	Águeda	Aveiro	FERMENTELOS	010109	Y
3	Águeda	Aveiro	MACINHATA DO VOUGA	010112	Y
4	Águeda	Aveiro	VALONGO DO VOUGA	010119	Y
5	Águeda	Aveiro	UF AGUEDA E BORRALHA	010121	Y
6	Águeda	Aveiro	UF BARRO E AGUDA DE	010122	Y
7	Águeda	Aveiro	UF BELAZAIMA CAST VOL	010123	Y
8	Águeda	Aveiro	UF RECARDAES E ESPINHE	010124	Y
9	Águeda	Aveiro	UF TRAVASSO E OIS DA RI	010125	Y
10	Águeda	Aveiro	UF TROFA SEGADAES LAN	010126	Y
11	Águeda	Aveiro	UF PRESTIMO MACIEIRA A	010127	Y
12	Albergaria-a	Aveiro	ALQUERUBIM	010202	Y
13	Albergaria-a	Aveiro	ANGEJA	010203	Y
14	Albergaria-a	Aveiro	BRANCA	010204	Y

# Task 5 - Location Dimension - Select values



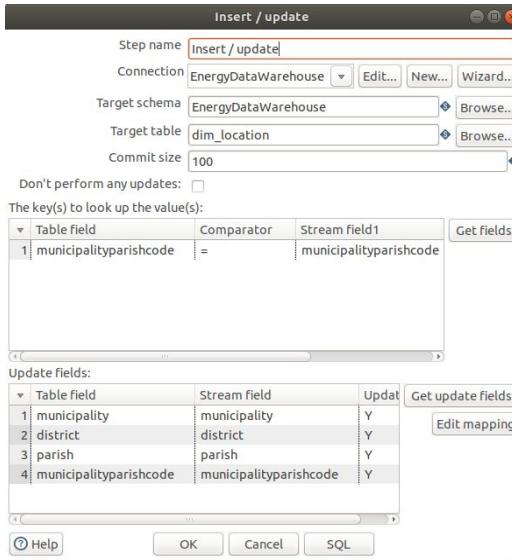
**Examine preview data**

Rows of step: Select values 2 (1000 rows)

	municipality	district	parish	municipalityparishcode
1	Águeda	Aveiro	AGUADA DE CIMA	010103
2	Águeda	Aveiro	FERMENTELOS	010109
3	Águeda	Aveiro	MACINHATA DO VOU	010112
4	Águeda	Aveiro	VALONGO DO VOUC	010119
5	Águeda	Aveiro	UF AGUEDA E BORR	010121
6	Águeda	Aveiro	UF BARRO E AGUAD	010122
7	Agueda	Aveiro	UF BELAZAIMA CAS	010123
8	Águeda	Aveiro	UF RECARDAES E ES	010124
9	Águeda	Aveiro	UF TRAVASSO E OIS	010125
10	Águeda	Aveiro	UF TROFA SEGADAE	010126
11	Águeda	Aveiro	UF PRESTIMO MACIE	010127
12	Albergaria-a-Velha	Aveiro	ALQUERUBIM	010202
13	Albergaria-a-Velha	Aveiro	ANGEJA	010203
14	Albergaria-a-Velha	Aveiro	BRANCA	010204

Close Stop Get more rows

# Task 5 - Location Dimension - Insert / Update



Examine preview data

Rows of step: Insert / update (1000 rows)

	municipality	district	parish	municipalityparishcode
1	Águeda	Aveiro	AGUADA DE CIMA	010103
2	Águeda	Aveiro	FERMENTELOS	010109
3	Águeda	Aveiro	MACINHATA DO VOUGA	010112
4	Águeda	Aveiro	VALONGO DO VOUGA	010119
5	Águeda	Aveiro	UF AGUEDA E BORRALHA	010121
6	Águeda	Aveiro	UF BARRO E AGUADA DE BAIXO	010122
7	Águeda	Aveiro	UF BELAZAIMA CAST VOUGA AGADAO	010123
8	Águeda	Aveiro	UF RECARDAES E ESPINHEL	010124
9	Águeda	Aveiro	UF TRAVASSO E OIS DA RIBEIRA	010125
10	Águeda	Aveiro	UF TROFA SEGADAES LAMAS VOUGA	010126
11	Águeda	Aveiro	UF PRESTIMO MACIEIRA ALCoba	010127
12	Albergaria-a-Velha	Aveiro	ALQUERUBIM	010202
13	Albergaria-a-Velha	Aveiro	ANGEJA	010203
14	Albergaria-a-Velha	Aveiro	BRANCA	010204
15	Albergaria-a-Velha	Aveiro	RIBEIRA DE FRAGUAS	010206

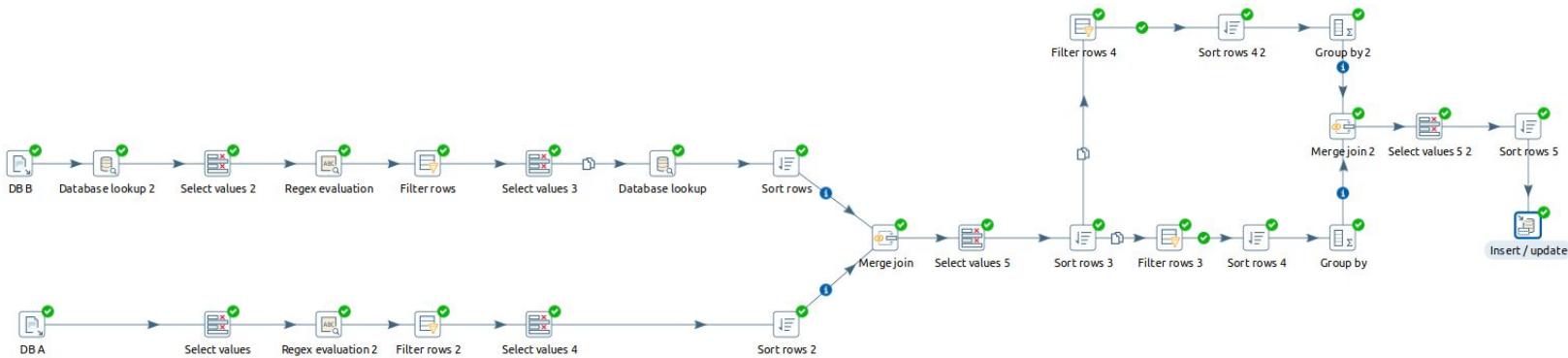
Close Stop Get more rows

# Task 5 - Location Dimension - Table output

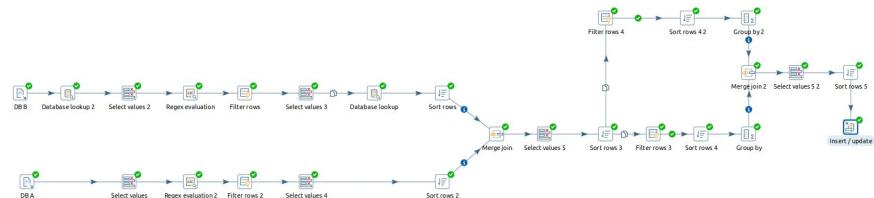
mysql> select * from dim_location;				
location_id	district	municipality	parish	municipalityparishcode
1	Aveiro	Aqueda	AGUADA DE CIMA	10103
2	Aveiro	Aqueda	FERMENTELOS	10109
3	Aveiro	Aqueda	MACINHATA DO VOUGA	10112
5	Aveiro	Aqueda	UF AGUEDA E BORRALHA	10121
6	Aveiro	Aqueda	UF BARRO E AGUADA DE BAIXO	10122
7	Aveiro	Aqueda	UF BELAZAIMA CAST VOUGA AGADAO	10123
11	Aveiro	Aqueda	UF PRESTIMO MACIEIRIA ALCOBA	10127
8	Aveiro	Aqueda	UF RECARDAES E ESPINHEI	10124
9	Aveiro	Aqueda	UF TRAVASSO E OIS DA RIBEIRA	10125
10	Aveiro	Aqueda	UF TROFA SEGADAS LAMAS VOUGA	10126
4	Aveiro	Aqueda	VALONGO DO VOUGA	10119
16	Aveiro	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	10209
12	Aveiro	Albergaria-a-Velha	ALQUERUBIM	10202
13	Aveiro	Albergaria-a-Velha	ANGEJA	10203
14	Aveiro	Albergaria-a-Velha	BRANCA	10204
15	Aveiro	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	10206
17	Aveiro	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	10210
18	Aveiro	Anadia	AVELAS DE CAMINHO	10304
19	Aveiro	Anadia	AVELAS DE CIMA	10305
20	Aveiro	Anadia	MOTIA	10307
21	Aveiro	Anadia	SANGALHOS	10309
22	Aveiro	Anadia	SAO LOURENCO DO BAIRRO	10310
25	Aveiro	Anadia	UF A DA GANDARA BATRRO E ANCAS	10316
26	Aveiro	Anadia	UF ARCOS E MOCOFORES	10317
27	Aveiro	Anadia	UF TAMENGOS AGUM OIS BAIRRO	10318
23	Aveiro	Anadia	VILA NOVA DE MONSARROS	10312
24	Aveiro	Anadia	VILARINHO DO BAIRRO	10313
28	Aveiro	Arouca	ALVARENGA	10402
29	Aveiro	Arouca	CHAVE	10407
30	Aveiro	Arouca	ESCARIZ	10409
31	Aveiro	Arouca	FERMEDO	10411
32	Aveiro	Arouca	MANSORES	10413
33	Aveiro	Arouca	MOLDES	10414
34	Aveiro	Arouca	ROSSAS	10415
35	Aveiro	Arouca	SANTA EULALIA	10416
36	Aveiro	Arouca	SAO MIGUEL DO MATO	10417
37	Aveiro	Arouca	TROPECO	10418
40	Aveiro	Arouca	UF AROUCA E BURGO	10421
41	Aveiro	Arouca	UF CABRETTOS ALBERGARIA SERRA	10422
42	Aveiro	Arouca	UF CANELAS E ESPUNCA	10423
43	Aveiro	Arouca	UF COVELO DE PAIVO E JANARDE	10424
38	Aveiro	Arouca	URRO	10419
39	Aveiro	Arouca	VARZEA	10420
44	Aveiro	Aveiro	ABADAS	10501
45	Aveiro	Aveiro	CACIA	10502
51	Aveiro	Aveiro	EIXO E EIROL	10515
46	Aveiro	Aveiro	ESQUEIRA	10505

# Task 6

# Task 6



# Task 6 - DB B



**CSV file input**

Step name: DB B

Filename: /home/aid/Downloads/dataset\_B.csv

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: mixed

File encoding: UTF-8

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	.	,
2 Month	Integer	#	15	0	\$	.	,
3 Date	String		7		\$	.	,
4 District	String		14		\$	.	,
5 Municipality	String		21		\$	.	,
6 parish	String		30		\$	.	,
7 Voltage level	String		32		\$	.	,
8 Active Energy (kWh)	Number	#.##	11	3	\$	.	,
9 DistrictCode	Integer	#	15	0	\$	.	,
10 DistrictMunicipalityCode	String		4		\$	.	,
11 DistrictMunicipalityParishCode	String		6		\$	.	,
12 mes_int	Integer	#	15	0	\$	.	,

OK   Get Fields   Preview   Cancel

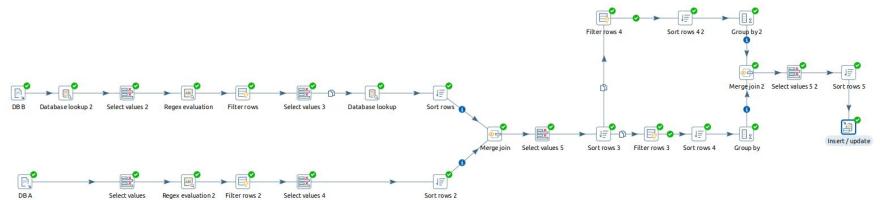
**Examine preview data**

Rows of step: DB B (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kW)	DistrictCo	DistrictMunicipalityCode	DistrictMunicipality	mes_int
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1
2	2021	1	2021-01	Bragança	Mirandela	MASCARENHAS	Baixa Tensão	88605.5	4	0407	040720	1
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRO VELHO	Baixa Tensão	44866.8	4	0407	040727	1
4	2021	1	2021-01	Bragança	Mirandela	UF AVIDAGOS NAVA	Baixa Tensão	82565.5	4	0407	040739	1
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCO E VILA E	Baixa Tensão	49094.8	4	0407	040741	1
6	2021	1	2021-01	Bragança	Mogadouro	MERINHOS	Baixa Tensão	29669.1	4	0408	040809	1
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA BEMPOS	Baixa Tensão	34596.4	4	0408	040813	1
8	2021	1	2021-01	Bragança	Torre de Moncorvo	CABECA BOA	Baixa Tensão	52718.7	4	0409	040903	1
9	2021	1	2021-01	Bragança	Torre de Moncorvo	MOS	Baixa Tensão	22259.1	4	0409	040913	1
10	2021	1	2021-01	Bragança	Torre de Moncorvo	UF ADEGANHA E CA	Multa Alta, A	355083.9	4	0409	040918	1
11	2021	1	2021-01	Bragança	Torre de Moncorvo	UF URROS PEREDO C	Baixa Tensão	60464.2	4	0409	040921	1
12	2021	1	2021-01	Bragança	Vila Flor	SEIXO DE MANHOSE	Baixa Tensão	61428.5	4	0410	041013	1
13	2021	1	2021-01	Bragança	Vila Flor	UF VILAS BOAS VILA	Baixa Tensão	118577.9	4	0410	041024	1
14	2021	1	2021-01	Bragança	Vimioso	UF ALGOSO CAMPO	Baixa Tensão	99930.1	4	0411	041115	1

**Close**   **Show Log**

# Task 6 - Database lookup 2



**Database lookup**

Step name: Database lookup 2

Connection: EnergyDataWarehouse

Lookup schema: EnergyDataWarehouse

Lookup table: dim\_time

Enable cache?

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

Load all data from table

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

Table field	Comparator	Field1	Field2
1 year	=	Year	
2 month	=	Month	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	time_id			None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by:

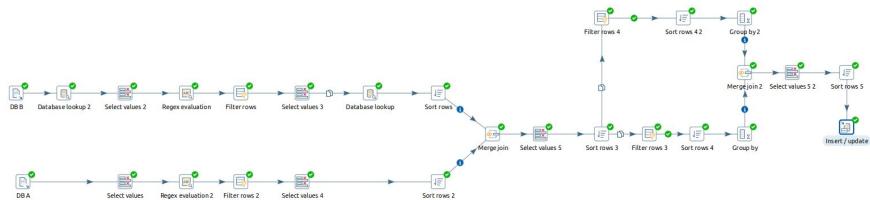
Buttons: Help, OK, Cancel, Get Fields, Get lookup fields

Rows of step: Database lookup 2 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	mes_int	time_id
1	2021	1	2021-01	Bragança	Mirandela	FRECHAS	Baixa Tensão	148232.2	4	0407	040716	1	3
2	2021	1	2021-01	Bragança	Mirandela	MASCARENHAS	Baixa Tensão	88605.5	4	0407	040720	1	3
3	2021	1	2021-01	Bragança	Mirandela	SAO PEDRO VELHO	Baixa Tensão	44866.8	4	0407	040727	1	3
4	2021	1	2021-01	Bragança	Mirandela	UF AVIDAGOS NAVALHO	Baixa Tensão	82565.5	4	0407	040739	1	3
5	2021	1	2021-01	Bragança	Mirandela	UF FRANCO E VILA BOA	Baixa Tensão	49094.8	4	0407	040741	1	3
6	2021	1	2021-01	Bragança	Mogadouro	MEIRINHOS	Baixa Tensão	29669.1	4	0408	040809	1	3
7	2021	1	2021-01	Bragança	Mogadouro	PEREDO DA BEMPOSTA	Baixa Tensão	34596.4	4	0408	040813	1	3
8	2021	1	2021-01	Bragança	Torre de Monc	CABECA BOA	Baixa Tensão	52718.7	4	0409	040903	1	3
9	2021	1	2021-01	Bragança	Torre de Monc	MOS	Baixa Tensão	22259.1	4	0409	040913	1	3
10	2021	1	2021-01	Bragança	Torre de Monc	UF ADEGANHA E CARDAN	Muito Alta, Alt	355083.9	4	0409	040918	1	3
11	2021	1	2021-01	Bragança	Torre de Monc	UF URROS PEREDO CASTE	Baixa Tensão	60646.2	4	0409	040921	1	3
12	2021	1	2021-01	Bragança	Vila Flor	SEIXO DE MANHOSES	Baixa Tensão	61428.5	4	0410	041013	1	3
13	2021	1	2021-01	Bragança	Vila Flor	UF VILAS BOAS VILARINH	Baixa Tensão	118577.9	4	0410	041024	1	3
14	2021	1	2021-01	Bragança	Vimioso	UF ALGOSO CAMPO VIBO	Baixa Tensão	99930.1	4	0411	041115	1	3

Buttons: Close, Stop, Get more rows.

# Task 6 - Select Values 2



**Select values**

Step name **Select values 2**

Select & Alter **Remove** Meta-data

Fields :

Fieldname	Rename to	Length
1 DistrictMunicipalityParishCode	municipalityparishcode	
2 Active Energy (kWh)		
3 time_id		

Get fields to select **Edit Mapping**

Include unspecified fields, ordered

**OK** **Cancel** **Help**

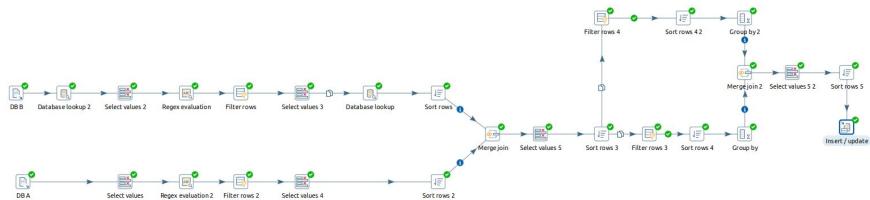
**Examine preview data**

Rows of step: Select values 2 (1000 rows)

	municipalityparishcode	Active Energy (kWh)	time_id
1	040716	148232.2	3
2	040720	88605.5	3
3	040727	44866.8	3
4	040739	82565.5	3
5	040741	49094.8	3
6	040809	29669.1	3
7	040813	34596.4	3
8	040903	52718.7	3
9	040913	22259.1	3
10	040918	355083.9	3
11	040921	60646.2	3
12	041013	61428.5	3
13	041024	118577.9	3
14	041115	99930.1	3
		20072.6	

**Close** **Stop** **Get more rows**

# Task 6 - Regex evaluation



**Regex evaluation**

Step name: **Regex evaluation**

**Settings** **Content**

Step settings

Field to evaluate: **municipalityparishcode**

Result field name: **municipalityparishcode\_number**

Create fields for capture groups:

Replace previous fields:

Regular expression: **^\d+\$**

Test regEx

Use variable substitution:

Capture Group Fields

New field	Type	Length	Precision	Format	Group	Decimal	Currency	Null If	Default
1									

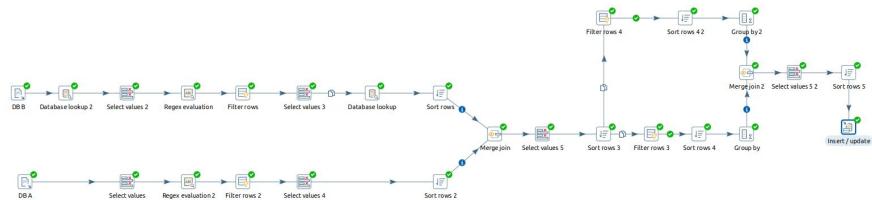
Help

Examine preview data

Rows of step: Regex evaluation (1000 rows)

	municipalityparishcode	Active Energy (kWh)	time_id	municipalityparishcode_number
1	040716	148232.2	3	Y
2	040720	88605.5	3	Y
3	040727	44866.8	3	Y
4	040739	82565.5	3	Y
5	040741	49094.8	3	Y
6	040809	29669.1	3	Y
7	040813	34596.4	3	Y
8	040903	52718.7	3	Y
9	040913	22259.1	3	Y
10	040918	355083.9	3	Y
11	040921	60646.2	3	Y
12	041013	61428.5	3	Y
13	041024	118577.9	3	Y
14	041115	99930.1	3	Y

# Task 6 - Filter Rows



### Filter rows

Step name

Send 'true' data to step:

Send 'false' data to step:

The condition:

+   
 **municipalityparishcode\_number** =   
Y  (Boolean)

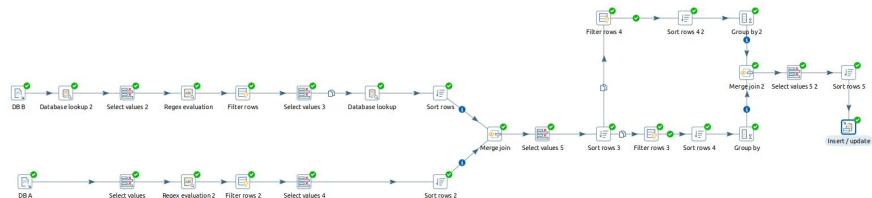
4 (11)

### Examine preview data

Rows of step: Filter rows (1000 rows)

#	municipalityparishcode	Active Energy (kWh)	time_id	municipalityparishcode_number
1	040716	148232.2	3	Y
2	040720	88605.5	3	Y
3	040727	44866.8	3	Y
4	040739	82565.5	3	Y
5	040741	49094.8	3	Y
6	040809	29669.1	3	Y
7	040813	34596.4	3	Y
8	040903	52718.7	3	Y
9	040913	22259.1	3	Y
10	040918	355083.9	3	Y
11	040921	60646.2	3	Y
12	041013	61428.5	3	Y
13	041024	118577.9	3	Y
14	041115	99930.1	3	Y

# Task 6 - Select Values 3



**Select values**

Step name **Select values 3**

Select & Alter Remove Meta-data

Fields to remove :

Fieldname
1 municipalityparishcode_number

**Get fields to remove**

OK Cancel Help

**Examine preview data**

Rows of step: Select values 3 (1000 rows)

	municipalityparishcode	Active Energy (kWh)	time_id
1	040716	148232.2	3
2	040720	88605.5	3
3	040727	44866.8	3
4	040739	82565.5	3
5	040741	49094.8	3
6	040809	29669.1	3
7	040813	34596.4	3
8	040903	52718.7	3
9	040913	22259.1	3
10	040918	355083.9	3
11	040921	60646.2	3
12	041013	61428.5	3
13	041024	118577.9	3
14	041115	99930.1	3
15	041124	20072.6	3

Close Stop Get more rows

# Task 6 - Database lookup

**Database lookup**

Step name **Database lookup**

Connection **EnergyDataWarehouse**

Lookup schema **EnergyDataWarehouse**

Lookup table **dim\_location**

Enable cache?

Cache size in rows (0=cache everything)

Load all data from table

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

Table field	Comparator	Field1	Field2
1 municipalityparishcode	=	municipalityparishcode	

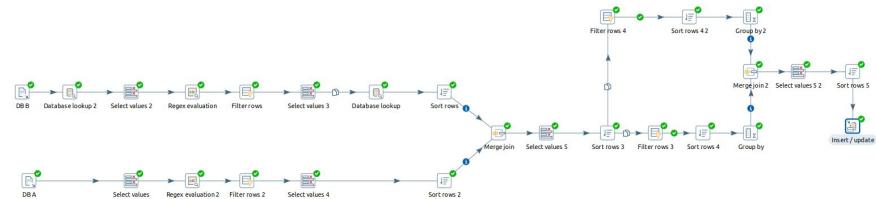
Values to return from the lookup table :

Field	New name	Default	Type
1 location_id			None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

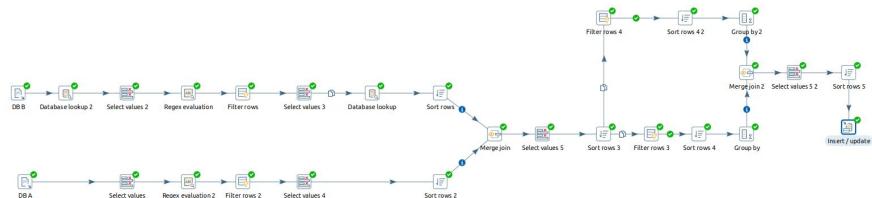


**Examine preview data**

Rows of step: Database lookup (1000 rows)

	municipalityparishcode	Active Energy (kWh)	time_id	location_id
1	040716	148232.2	3	685
2	040720	88605.5	3	687
3	040727	44866.8	3	691
4	040739	82565.5	3	700
5	040741	49094.8	3	702
6	040809	29669.1	3	710
7	040813	34596.4	3	713
8	040903	52718.7	3	726
9	040913	22259.1	3	732
10	040918	355083.9	3	734
11	040921	60646.2	3	737
12	041013	61428.5	3	744
13	041024	118577.9	3	751
14	041115	99930.1	3	759
15	041201	20072.6	3	761

# Task 6 - Sort rows



### Sort rows

Step name: Sort rows

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?

Only pass unique rows? (verifies keys only)

**Fields:**

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

OK Cancel Get Fields

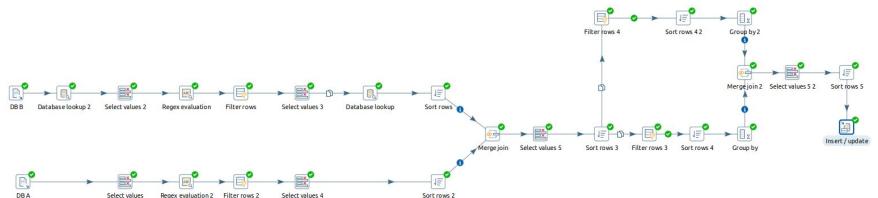
Examine preview data

Rows of step: Sort rows (1000 rows)

municipalityparishcode	Active Energy (kWh)	time_id	location_id
1 010103	1830933.5	37	1
2 010103	2166786.9	6	1
3 010103	1777962	26	1
4 010103	709818.4	25	1
5 010103	729692.4	28	1
6 010103	1928826.5	28	1
7 010103	1112395.9	34	1
8 010103	620929	23	1
9 010103	2083921.7	4	1
10 010103	644290.5	31	1
11 010103	630531	30	1
12 010103	1823366.8	32	1
13 010103	1899922.2	14	1
14 010103	2014952.9	20	1
15 010103	2070205.6	20	1

Close Stop Get more rows

# Task 6 - DB A



**CSV File Input**

Step name: DB A

Filename: /home/aid/Downloads/dataset\_A.csv

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result?

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: mixed

File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group
1 Year	Integer	#	15	0	\$	.	,
2 Month	Integer	#	15	0	\$	.	,
3 Date	String		7		\$	-	,
4 District	String		16		\$	-	,
5 Municipality	String		22		\$	-	,
6 Parish	String		51		\$	-	,
7 Includes Smart Meter	String		3		\$	-	,
8 Number of CPE's	Integer	#	15	0	\$	-	,
9 DistrictCode	Integer	#	15	0	\$	-	,
10 DistrictMunicipalityCode	Integer	#	15	0	\$	-	,
11 DistrictMunicipalityParishCode	String		6		\$	-	,
12 Active Contract	String		3		\$	-	,

OK   Get Fields   Preview   Cancel

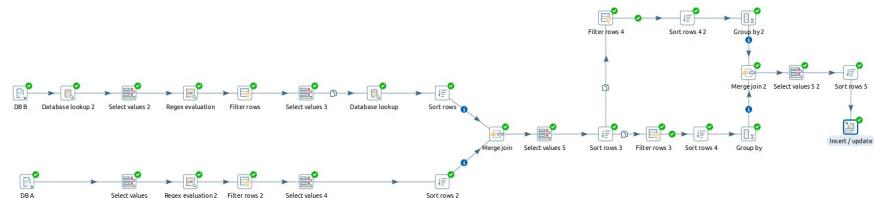
**Examine preview data**

Rows of step: DB A (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active Contract
1	2024	5	2024-05	Vila Real	Ribeira de Pena	Alvadia	Sim	155	17	1709	170901	Sim
2	2024	5	2024-05	Bragança	Mogadouro	Travanca	Não	15	4	408	040820	Sim
3	2024	5	2024-05	Portalegre	Ponte de Sor	Montarigil	Não	475	12	1213	121302	Sim
4	2024	5	2024-05	Viseu	Santa Comba Dão	Santa Comba Dão	Sim	3141	18	1814	181411	Sim
5	2024	5	2024-05	Portalegre	Arronches	Assunção	Sim	1248	12	1202	120201	Sim
6	2024	5	2024-05	Santarém	Tomar	Asseiceira	Sim	1549	14	1418	141802	Sim
7	2024	5	2024-05	Coimbra	Mira	Carapinhos	Sim	488	6	608	060803	Sim
8	2024	5	2024-05	Guarda	Celorico da Beira	Prados	Sim	221	9	903	090312	Sim
9	2024	5	2024-05	Viana do C	Paredes de Coura	Agualonga	Sim	175	16	1605	160501	Sim
10	2024	5	2024-05	Beja	Odemira	Reliquias	Sim	334	2	211	021102	Sim
11	2024	5	2024-05	Bragança	Macedo de Cavale	Talhinhas e Bagu	Sim	319	4	405	040544	Sim
12	2024	5	2024-05	Coimbra	Condeixa-a-Nova	Zambujal	Não	1	6	604	060410	Sim
13	2024	5	2024-05	Bragança	Vila Flor	Santa Comba de	Não	19	4	410	041012	Sim
14	2024	5	2024-05	Melgaço	Cousso	Melgaço	Não	96	16	1603	160304	Sim
15	2024	5	2024-05	Castelo Branco	Almeida	Chaves	Ferreira	455	5	500	050000	Sim

**Close**   **Stop**   **Get more rows**

# Task 6 - Select values



**Select values**

Step name **Select values**

Select & Alter Remove Meta-data

Fields :

Fieldname	Rename to	Length
1 DistrictMunicipalityParishCode	municipalityparishcode	
2 Includes Smart Meter		
3 Number of CPE's		

Get fields to select  
Edit Mapping

Include unspecified fields, ordered

OK Cancel

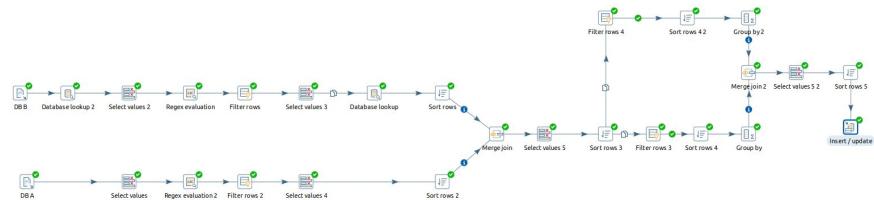
Examine preview data

Rows of step: Select values (1000 rows)

	municipalityparishcode	Includes Smart Meter	Number of CPE's
1	170901	Sim	155
2	040820	Não	15
3	121302	Não	475
4	181411	Sim	3141
5	120201	Sim	1248
6	141802	Sim	1549
7	060803	Sim	488
8	090312	Sim	221
9	160501	Sim	175
10	021102	Sim	334
11	040544	Sim	319
12	060410	Não	1
13	041012	Não	19
14	160304	Não	96
15	050000	Sim	1525

Close Stop Get more rows

# Task 6 - Regex evaluation 2



**Regex evaluation**

Step name **Regex evaluation 2**

**Settings Content**

Step settings

Field to evaluate **municipalityparishcode**

Result field name **municipalityparishcode\_number**

Create fields for capture groups

Replace previous fields

Regular expression: **^\d+\$**

**Test regEx**

Use variable substitution

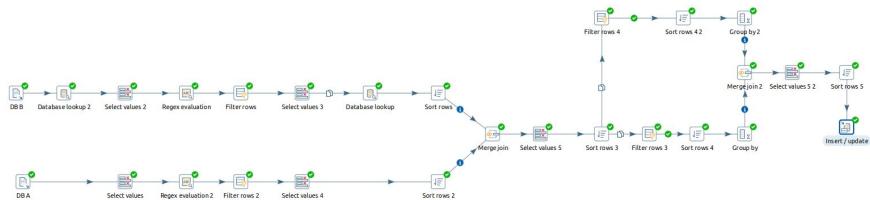
**Capture Group Fields**

New Field	Type	Length	Precision	Format	Group	Decimal	Currency	Null If	Default
1									

**OK** **Cancel**

Examine preview data				
Rows of step: Regex evaluation 2 (1000 rows)				
	municipalityparishcode	Includes Smart Meter	Number of CPE's	municipalityparishcode_number
1	170901	Sim	155	Y
2	040820	Não	15	Y
3	121302	Não	475	Y
4	181411	Sim	3141	Y
5	120201	Sim	1248	Y
6	141802	Sim	1549	Y
7	060803	Sim	488	Y
8	090312	Sim	221	Y
9	160501	Sim	175	Y
10	021102	Sim	334	Y
11	040544	Sim	319	Y
12	060410	Não	1	Y
13	041012	Não	19	Y
14	160304	Não	96	Y
15	050909	Sim	1725	Y

# Task 6 - Filter rows 2



**Filter rows**

Step name: **Filter rows 2**

Send 'true' data to step:

Send 'false' data to step:

The condition:

municipalityparishcode\_number =  +  Y (Boolean)

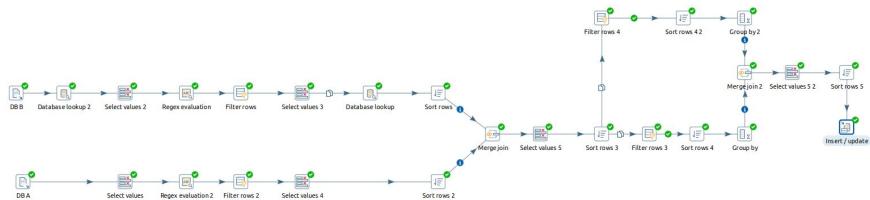
OK Cancel

Examine preview data

	municipalityparishcode	Includes Smart Meter	Number of CPE's	municipalityparishcode_number
1	170901	Sim	155	Y
2	040820	Não	15	Y
3	121302	Não	475	Y
4	181411	Sim	3141	Y
5	120201	Sim	1248	Y
6	141802	Sim	1549	Y
7	060803	Sim	488	Y
8	090312	Sim	221	Y
9	160501	Sim	175	Y
10	021102	Sim	334	Y
11	040544	Sim	319	Y
12	060410	Não	1	Y
13	041012	Não	19	Y
14	160304	Não	96	Y

Close Stop Get more rows

# Task 6 - Select values 4



Select values

Step name **Select values 4**

Select & Alter Remove Meta-data

Fields to remove :

▼ Fieldname  
1 municipalityparishcode\_number

Get fields to remove

OK Cancel

Help

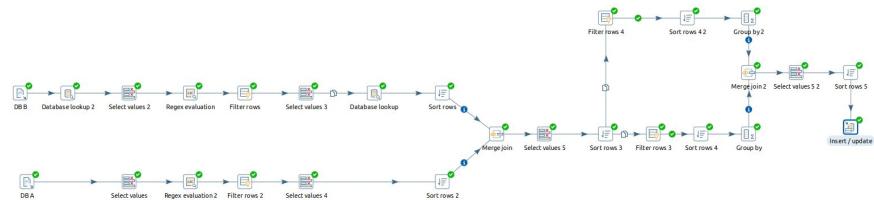
Examine preview data

Rows of step: Select values 4 (1000 rows)

municipalityparishcode	Includes Smart Meter	Number of CPE's
1 170901	Sim	155
2 040820	Não	15
3 121302	Não	475
4 181411	Sim	3141
5 120201	Sim	1248
6 141802	Sim	1549
7 060803	Sim	488
8 090312	Sim	221
9 160501	Sim	175
10 021102	Sim	334
11 040544	Sim	319
12 060410	Não	1
13 041012	Não	19
14 160304	Não	96
...	...	...

Close Stop Get more rows

# Task 6 - 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
1 municipalityparishcode	Y	N	N	0

OK Cancel Get Fields Help

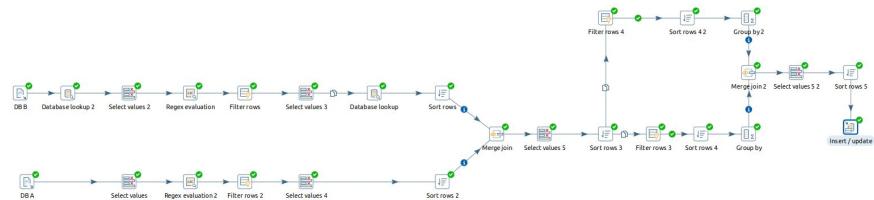
Examine preview data

Rows of step: Sort rows 2 (1000 rows)

municipalityparishcode	Includes Smart Meter	Number of CPE's
1 010103	Sim	2159
2 010103	Não	221
3 010103	Sim	1953
4 010103	Não	211
5 010103	Sim	1928
6 010103	Sim	1880
7 010103	Não	1972
8 010103	Sim	69
9 010103	Não	228
10 010103	Sim	1934
11 010103	Não	361
12 010103	Não	102
13 010103	Sim	2061
14 010103	Sim	2171

**Close Stop Get more rows**

# Task 6 - Merge join



**Merge join**

Step name: **Merge join**

First Step: **Sort rows**

Second Step: **Sort rows 2**

Join Type: **INNER**

Keys for 1st step: **municipalitypar**

Keys for 2nd step: **municipalitypar**

**Get key fields**      **Get key fields**

**Help**      **OK**      **Cancel**

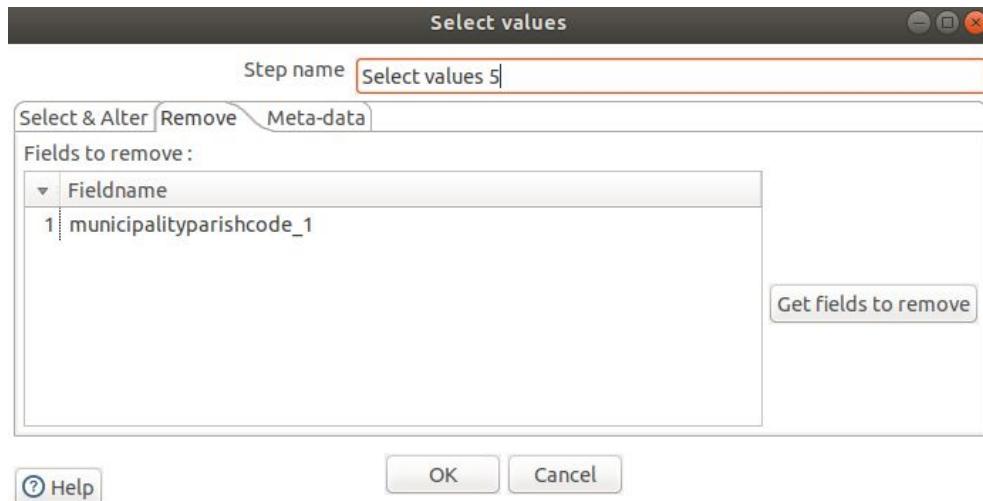
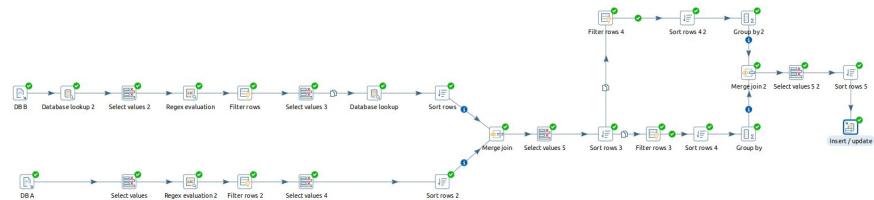
**Examine preview data**

Rows of step: Merge join (1000 rows)

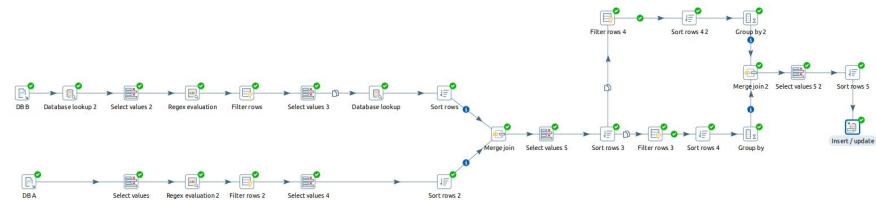
	municipalityparishcode	Active Energy (kWh)	time_id	location_id	municipalityparishcode_1	Includes Smart Meter	Number of CPE's
1	010103	1830933.5	37	1	010103	Sim	2159
2	010103	1830933.5	37	1	010103	Não	221
3	010103	1830933.5	37	1	010103	Sim	1953
4	010103	1830933.5	37	1	010103	Não	211
5	010103	1830933.5	37	1	010103	Sim	1928
6	010103	1830933.5	37	1	010103	Sim	1880
7	010103	1830933.5	37	1	010103	Sim	1972
8	010103	1830933.5	37	1	010103	Não	69
9	010103	1830933.5	37	1	010103	Não	228
10	010103	1830933.5	37	1	010103	Sim	1934
11	010103	1830933.5	37	1	010103	Não	361
12	010103	1830933.5	37	1	010103	Não	102
13	010103	1830933.5	37	1	010103	Sim	2061
14	010103	1830933.5	37	1	010103	Sim	2171

**Close**      **Stop**      **Get more rows**

# Task 6 - Select Values 5



# Task 6 - Sort rows 3



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

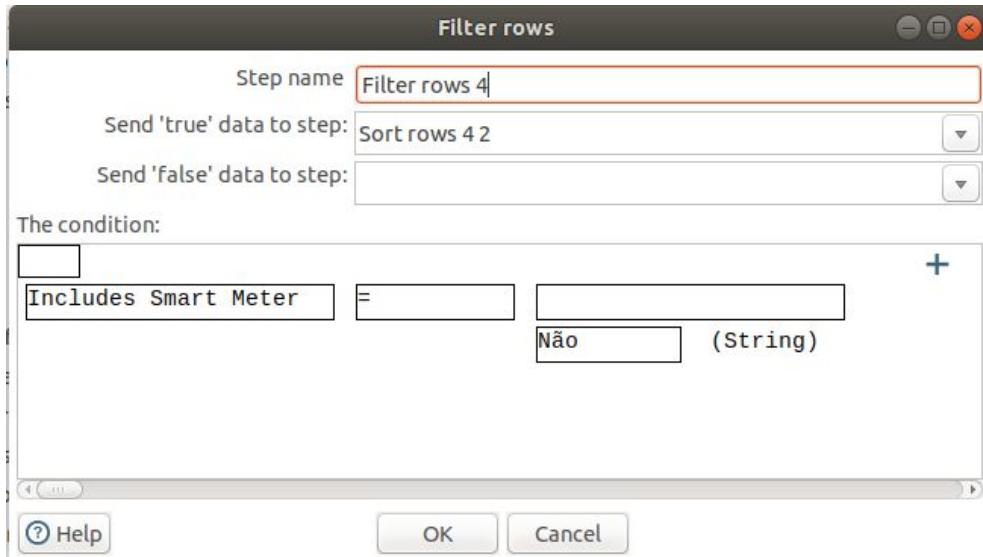
Compress TMP Files?

Only pass unique rows? (verifies keys only)

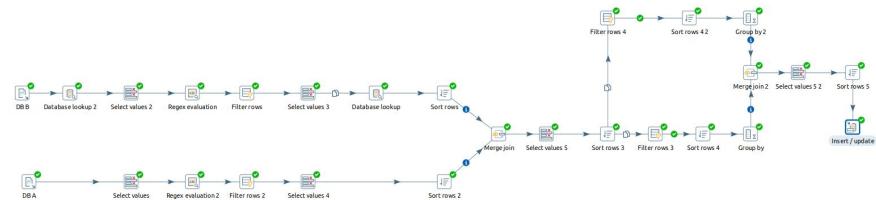
Fields :

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

# Task 6 - Filter rows 4



# Task 6 - Sort rows 4 2



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

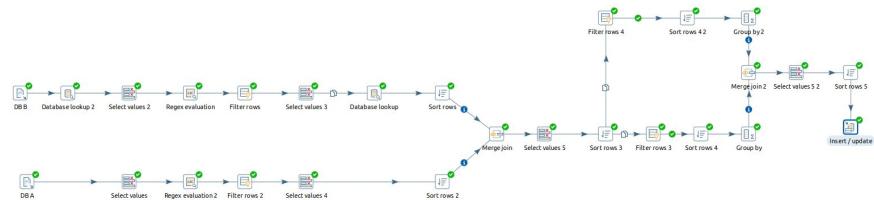
Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

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

# Task 6 - Group by 2



**Group by**

Step name

Include all rows?

Temporary files directory

TMP-file prefix

Add line number, restart in each group

Always give back a result row

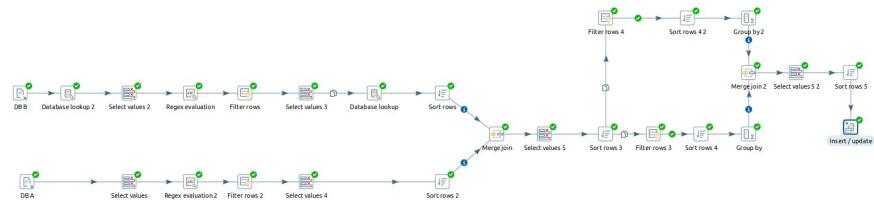
The fields that make up the group:

Group field  
1 location\_id  
2 time\_id

Aggregates:

Name	Subject	Type
1 count_non_smart	Number of CPE's	Sum

# Task 6 - Filter Rows 3



**Filter rows**

Step name:

Send 'true' data to step:

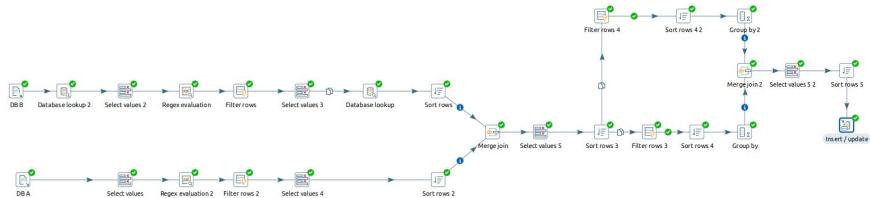
Send 'false' data to step:

The condition:

+   
Includes Smart Meter =   
 Sim (String)

OK Cancel

# Task 6 - Sort Rows 4



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields :

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

# Task 6 - Group by

Group by

Step name

Include all rows?

Temporary files directory

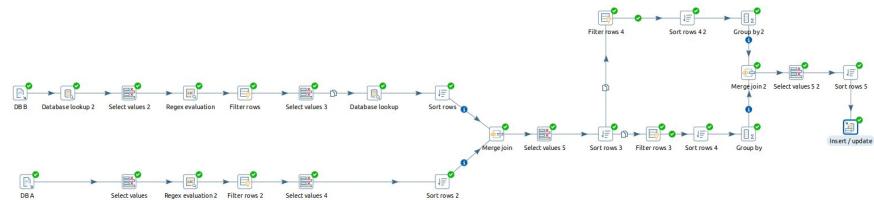
TMP-file prefix

Add line number, restart in each group

Always give back a result row

The fields that make up the group:

Name	Subject	Type
1 count_smart	Number of CPE's	Sum
2 energy_consumption	Active Energy (kWh)	Sum



# Task 6 - Merge join 2

**Merge join**

Step name: **Merge join 2**

First Step: **Group by 2**

Second Step: **Group by**

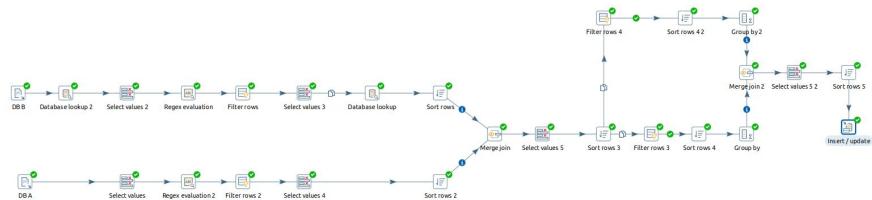
Join Type: **INNER**

Keys for 1st step:      Keys for 2nd step:

▼ Key field	▼ Key field
1 location_id	1 location_id

**Get key fields**      **Get key fields**

**Help**      **OK**      **Cancel**



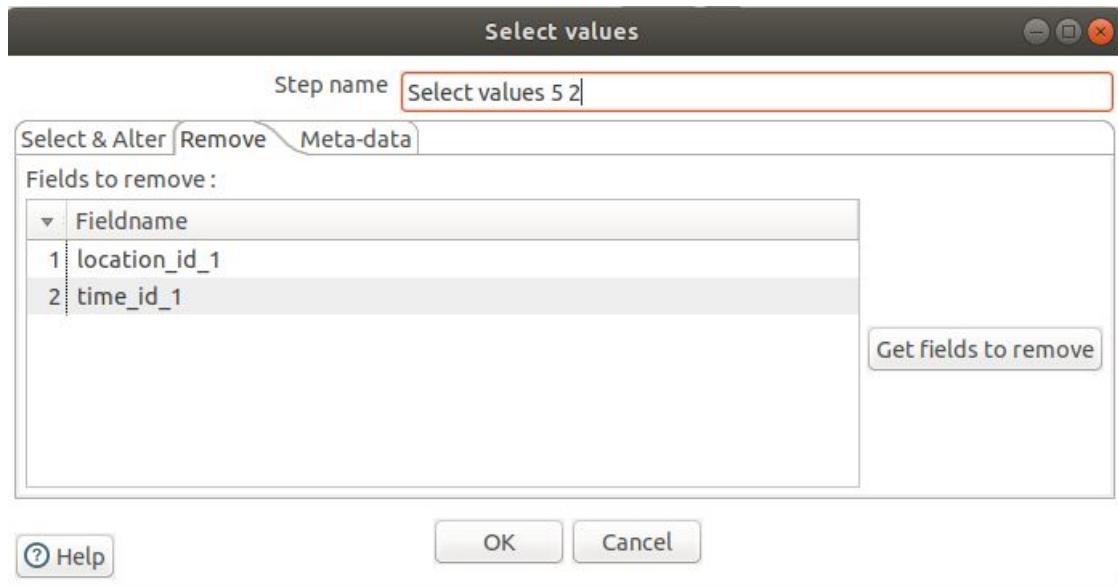
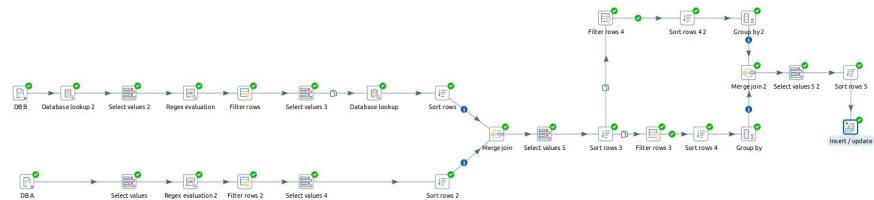
**Examine preview data**

Rows of step: Merge.join 2 (1000 rows)

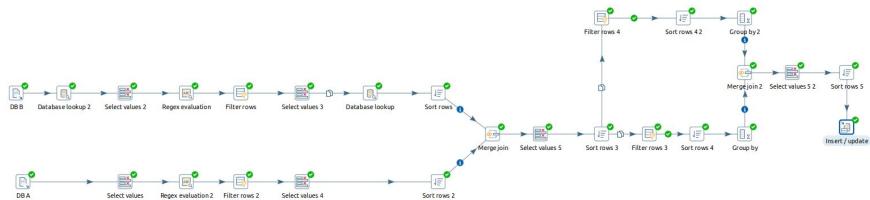
	location_id	time_id	count_non_smari	location_id_1	time_id_1	count_smart	energy_consumption
1	1	37	4619	1	37	56010	51266136.82399998
2	1	37	4619	1	6	56010	60670033.116
3	1	37	4619	1	26	56010	49782936.84000002
4	1	37	4619	1	25	56010	19874914.892
5	1	37	4619	1	28	112020	74438527.32399996
6	1	37	4619	1	34	56010	31147085.284000017
7	1	37	4619	1	23	56010	17386012.112000007
8	1	37	4619	1	4	56010	58349807.01199998
9	1	37	4619	1	31	56010	18040134.112000007
10	1	37	4619	1	30	56010	17654868.89599998
11	1	37	4619	1	32	56010	51054269.53200002
12	1	37	4619	1	14	56010	53197821.79600001
13	1	37	4619	1	20	56010	56418680.92000001
14	1	37	4619	1	29	56010	58220024.54799996
4,5	1	37	4619	1	27	112020	62420242.220

**Close**      **Stop**      **Get more rows**

# Task 6 - Select values 5 2



# Task 6 - Select values 5



Sort rows

Step name

Sort directory

TMP-file prefix

Sort size (rows in memory)

Free memory threshold (in %)

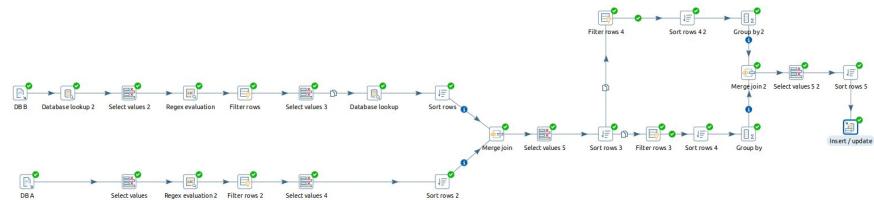
Compress TMP Files?

Only pass unique rows? (verifies keys only)

Fields:

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	location_id	Y	N	N	0	N
2	time_id	Y	N	N	0	N

# Task 6 - Insert Update



Insert / update

Step name **Insert / update**

Connection **EnergyDataWarehouse**

Target schema **EnergyDataWarehouse**

Target table **fact\_energy\_consumption**

Commit size **100**

Don't perform any updates:

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

Table field	Comparator	Stream field1	Stream field2
1 location_id	=	location_id	
2 time_id	=	time_id	

Get fields

Update fields:

Table field	Stream field	Update
1 location_id	location_id	Y
2 time_id	time_id	Y
3 count_non_smart	count_non_smart	Y
4 count_smart	count_smart	Y
5 energy_consumption	energy_consumption	Y

Get update fields

# Task 6 - Output

```
mysql> select * from fact_energy_consumption limit 10;
+-----+-----+-----+-----+-----+
| time_id | location_id | energy_consumption | count_smart | count_non_smart |
+-----+-----+-----+-----+-----+
|       1 |         1 |    51266136.824 |      56010 |        4619 |
|       1 |         2 |    3446160.172 |      46058 |        4459 |
|       1 |         3 |   12232451.560 |      44808 |       6834 |
|       1 |         4 |    1634083.836 |      3925 |        923 |
|       1 |         5 |   195414235.268 |     220154 |       5665 |
|       1 |         6 |  20290229.204 |      40533 |      10622 |
|       1 |         7 |    9787809.864 |      20450 |        4743 |
|       1 |         8 |  24777097.156 |      78615 |      13918 |
|       1 |         9 |  10149434.148 |      30269 |        5478 |
|       1 |        10 |   24247578.096 |      60219 |       7381 |
+-----+-----+-----+-----+-----+
10 rows in set (0.00 sec)
```

# Task 7

# Task 7 - OLAP cube

Schema Workbench

File Edit View Options Windows Help

Schema - EnergyDataWarehouse (Task7.xml)

Schema

EnergyConsumption

- Table: fact\_energy\_consumption

Time

- Time Hierarchy
  - Year
  - Season
  - Month
- Table: dim\_time

Location

- Location Hierarchy
  - District
  - Municipality
  - Municipality Parish Code
  - Parish
- Table: dim\_location

Energy consumption

- Smart Meters
- Non Smart Meters

Database - EnergyDataWarehouse (MySQL)

Schema

Attribute	Value
name	EnergyDataWarehouse
description	
measuresCaption	
defaultRole	

```
<Schema name="EnergyDataWarehouse">
  <Cube name="EnergyConsumption" visible="true" cache="true" enabled="true">
    <Table name="fact_energy_consumption">
    </Table>
    <Dimension type="TimeDimension" visible="true" foreignKey="time_id" highCardinality="false" name="Time">
      <Hierarchy name="Time Hierarchy" visible="true" hasAll="true" allMemberName="All Years" primaryKey="time_id">
        <Table name="dim_time">
        </Table>
        <level name="Year" visible="true" column="year" type="Integer" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">
        </Level>
        <level name="Season" visible="true" column="season" type="String" uniqueMembers="false" levelType="TimeQuarters" hideMemberIf="Never">
        </Level>
        <level name="Month" visible="true" column="month" type="Integer" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="location_id" highCardinality="false" name="Location">
      <Hierarchy name="Location Hierarchy" visible="true" hasAll="true" allMemberName="All Locations" primaryKey="location_id">
        <Table name="dim_location">
        </Table>
        <level name="District" visible="true" column="district" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <level name="Municipality" visible="true" column="municipality" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <level name="Municipality Parish Code" visible="true" column="municipalityparishcode" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <level name="Parish" visible="true" column="parish" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="Energy consumption" column="energy_consumption" datatype="Numeric" formatString="#,###.00" aggregator="sum" visible="true">
    </Measure>
    <Measure name="Smart Meters" column="count_smart" datatype="Integer" formatString="#,###" aggregator="sum" visible="true">
    </Measure>
    <Measure name="Non Smart Meters" column="count_non_smart" datatype="Integer" formatString="#,###" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

# Task 8

# Task 8 - Queries

a)

Measures	▼
Energy consumption	
Smart Meter Percentage	
✖	
Columns	▼
✖	
Rows	▼
Location Hierarchy	
District	
Time Hierarchy	
Year	
✖	
Filter	▼
✖	

b)

Measures	▼
Energy consumption	
✖	
Columns	▼
✖	
Rows	▼
Time Hierarchy	
Year	
Season	
Location Hierarchy	
District	
✖	
Filter	▼
✖	

c)

Measures	▼
Smart Meter Percentage	
Energy consumption	
✖	
Columns	▼
✖	
Rows	▼
Time Hierarchy	
Year	
Location Hierarchy	
District	
✖	
Filter	▼
✖	

# Task 8 - Analysis a)

Year	District	Smart Meter Percentage	Energy consumption
2021	Aveiro	77.15%	73,505,890,544.69
	Beja	77.14%	15,601,908,439.15
	Braga	78.52%	59,916,833,250.31
	Bragança	80.68%	7,375,929,716.74
	Castelo Branco	81.3%	18,500,153,653.63
	Coimbra	76.39%	33,861,484,890.80
	Évora	87.9%	20,878,312,089.26
	Faro	75.13%	44,176,376,180.50
	Guarda	78.23%	11,459,035,069.78
	Leiria	84.4%	46,865,141,973.89
	Lisboa	87.66%	162,931,680,095.86
	Portalegre	86.99%	7,936,245,942.91
	Porto	76.72%	133,098,158,119.36
	Santarém	77.17%	39,224,575,077.41
	Setúbal	87.15%	86,633,842,863.79
	Viana do Castelo	73.05%	22,331,138,823.50
	Vila Real	75.59%	10,369,689,466.75
	Viseu	72.63%	19,714,872,800.78

2022	Aveiro	77.06%	73,505,890,544.69
	Beja	77.3%	15,601,908,439.15
	Braga	78.34%	59,916,833,250.31
	Bragança	80.56%	7,375,929,716.74
	Castelo Branco	81.12%	18,500,153,653.63
	Coimbra	76.5%	33,861,734,062.80
	Évora	88.02%	20,878,312,089.26
	Faro	75.21%	44,176,376,180.50
	Guarda	77.97%	11,459,035,069.78
	Leiria	84.56%	46,865,141,973.89
	Lisboa	87.6%	162,931,680,095.86
	Portalegre	86.7%	7,936,245,942.91
	Porto	76.51%	133,098,158,119.36
	Santarém	76.78%	39,224,575,077.41
	Setúbal	87.72%	86,633,842,863.79
	Viana do Castelo	73.22%	22,331,138,823.50
	Vila Real	75.77%	10,369,689,466.75
	Viseu	72.64%	19,714,872,800.78

2023	Aveiro	77.01%	73,505,890,544.69
	Beja	76.44%	15,601,908,439.15
	Braga	78.17%	59,916,833,250.31
	Bragança	80.6%	7,375,929,716.74
	Castelo Branco	80.75%	18,500,153,653.63
	Coimbra	76.12%	33,861,734,062.80
	Évora	87.95%	20,878,312,089.26
	Faro	75.48%	44,176,376,180.50
	Guarda	78.05%	11,459,035,069.78
	Leiria	84.4%	46,865,141,973.89
	Lisboa	87.54%	162,931,680,095.86
	Portalegre	86.81%	7,936,245,942.91
	Porto	76.64%	133,108,739,065.96
	Santarém	76.31%	39,224,575,077.41
	Setúbal	87.12%	86,633,842,863.79
	Viana do Castelo	72.87%	22,340,288,920.04
	Vila Real	75.3%	10,369,689,466.75
	Viseu	72.63%	19,714,872,800.78

# Task 8 - Analysis b)

District	Year	Season	Energy consumption												
Aveiro	2021	Autumn	18,376,472,636.17	Braga	2021	Autumn	14,979,208,312.58	Castelo Branco	2021	Autumn	4,625,038,413.41	Faro	2021	Autumn	11,044,094,045.12
		Spring	18,376,472,636.17			Spring	14,979,208,312.58			Spring	4,625,038,413.41			Spring	11,044,094,045.12
		Summer	18,376,472,636.17			Summer	14,979,208,312.58			Summer	4,625,038,413.41			Summer	11,044,094,045.12
		Winter	18,376,472,636.17			Winter	14,979,208,312.58			Winter	4,625,038,413.41			Winter	11,044,094,045.12
	2022	Autumn	18,376,472,636.17		2022	Autumn	14,979,208,312.58		2022	Autumn	4,625,038,413.41		2022	Autumn	11,044,094,045.12
		Spring	18,376,472,636.17			Spring	14,979,208,312.58			Spring	4,625,038,413.41			Spring	11,044,094,045.12
		Summer	18,376,472,636.17			Summer	14,979,208,312.58			Summer	4,625,038,413.41			Summer	11,044,094,045.12
		Winter	18,376,472,636.17			Winter	14,979,208,312.58			Winter	4,625,038,413.41			Winter	11,044,094,045.12
	2023	Autumn	18,376,472,636.17		2023	Autumn	14,979,208,312.58		2023	Autumn	4,625,038,413.41		2023	Autumn	11,044,094,045.12
		Spring	18,376,472,636.17			Spring	14,979,208,312.58			Spring	4,625,038,413.41			Spring	11,044,094,045.12
		Summer	18,376,472,636.17			Summer	14,979,208,312.58			Summer	4,625,038,413.41			Summer	11,044,094,045.12
		Winter	18,376,472,636.17			Winter	14,979,208,312.58			Winter	4,625,038,413.41			Winter	11,044,094,045.12
Beja	2021	Autumn	3,900,477,109.79	Bragança	2021	Autumn	1,843,982,429.18	Coimbra	2021	Autumn	8,465,433,515.70	Guarda	2021	Autumn	2,864,758,767.44
		Spring	3,900,477,109.79			Spring	1,843,982,429.18			Spring	8,465,433,515.70			Spring	2,864,758,767.44
		Summer	3,900,477,109.79			Summer	1,843,982,429.18			Summer	8,465,433,515.70			Summer	2,864,758,767.44
		Winter	3,900,477,109.79			Winter	1,843,982,429.18			Winter	8,465,184,343.70			Winter	2,864,758,767.44
	2022	Autumn	3,900,477,109.79		2022	Autumn	1,843,982,429.18		2022	Autumn	8,465,433,515.70		2022	Autumn	2,864,758,767.44
		Spring	3,900,477,109.79			Spring	1,843,982,429.18			Spring	8,465,433,515.70			Spring	2,864,758,767.44
		Summer	3,900,477,109.79			Summer	1,843,982,429.18			Summer	8,465,433,515.70			Summer	2,864,758,767.44
		Winter	3,900,477,109.79			Winter	1,843,982,429.18			Winter	8,465,433,515.70			Winter	2,864,758,767.44
	2023	Autumn	3,900,477,109.79		2023	Autumn	1,843,982,429.18		2023	Autumn	8,465,433,515.70		2023	Autumn	2,864,758,767.44
		Spring	3,900,477,109.79			Spring	1,843,982,429.18			Spring	8,465,433,515.70			Spring	2,864,758,767.44
		Summer	3,900,477,109.79			Summer	1,843,982,429.18			Summer	8,465,433,515.70			Summer	2,864,758,767.44
		Winter	3,900,477,109.79			Winter	1,843,982,429.18			Winter	8,465,433,515.70			Winter	2,864,758,767.44

# Task 8 - Analysis b)

Leiria	2021	Autumn	11,716,285,493.47	Portalegre	2021	Autumn	1,984,061,485.73	Santarém	2021	Autumn	9,806,143,769.35	Viana do Castelo	2021	Autumn	5,582,784,705.88
		Spring	11,716,285,493.47			Spring	1,984,061,485.73			Spring	9,806,143,769.35			Spring	5,582,784,705.88
		Summer	11,716,285,493.47			Summer	1,984,061,485.73			Summer	9,806,143,769.35			Summer	5,582,784,705.88
		Winter	11,716,285,493.47			Winter	1,984,061,485.73			Winter	9,806,143,769.35			Winter	5,582,784,705.88
	2022	Autumn	11,716,285,493.47	2022	Autumn	1,984,061,485.73	2022	Autumn	9,806,143,769.35	2022	Autumn	9,806,143,769.35	2022	Autumn	5,582,784,705.88
Lisboa		Spring	11,716,285,493.47		Spring	1,984,061,485.73	Spring	9,806,143,769.35	Spring	5,582,784,705.88					
		Summer	11,716,285,493.47		Summer	1,984,061,485.73	Summer	9,806,143,769.35	Summer	5,582,784,705.88					
		Winter	11,716,285,493.47		Winter	1,984,061,485.73	Winter	9,806,143,769.35	Winter	5,582,784,705.88					
2023	Autumn	11,716,285,493.47	2023	Autumn	1,984,061,485.73	2023	Autumn	9,806,143,769.35	Autumn	9,806,143,769.35	2023	Autumn	5,582,784,705.88		
	Spring	11,716,285,493.47		Spring	1,984,061,485.73	Spring	9,806,143,769.35	Spring	5,591,934,802.41						
	Lisboa			Summer	11,716,285,493.47	Summer	1,984,061,485.73	Summer	9,806,143,769.35	Summer		5,582,784,705.88			
				Winter	11,716,285,493.47	Winter	1,984,061,485.73	Winter	9,806,143,769.35	Winter		5,582,784,705.88			
2021	Autumn	40,732,920,023.96	Porto	2021	Autumn	33,274,539,529.84	Setúbal	2021	Autumn	21,658,460,715.95	Vila Real	2021	Autumn	2,592,422,366.69	
	Spring	40,732,920,023.96			Spring	33,274,539,529.84			Spring	21,658,460,715.95			Spring	2,592,422,366.69	
	Summer	40,732,920,023.96			Summer	33,274,539,529.84			Summer	21,658,460,715.95			Summer	2,592,422,366.69	
	Winter	40,732,920,023.96			Winter	33,274,539,529.84			Winter	21,658,460,715.95			Winter	2,592,422,366.69	
Lisboa	2022	Autumn	40,732,920,023.96	2022	Autumn	33,274,539,529.84	2022	Autumn	21,658,460,715.95	2022	Autumn	21,658,460,715.95	2022	Autumn	2,592,422,366.69
		Spring	40,732,920,023.96		Spring	33,274,539,529.84	Spring	21,658,460,715.95	Spring	2,592,422,366.69					
		Summer	40,732,920,023.96		Summer	33,274,539,529.84	Summer	21,658,460,715.95	Summer	2,592,422,366.69					
		Winter	40,732,920,023.96		Winter	33,274,539,529.84	Winter	21,658,460,715.95	Winter	2,592,422,366.69					
	2023	Autumn	40,732,920,023.96	2023	Autumn	33,285,120,476.44	2023	Autumn	21,658,460,715.95	2023	Autumn	21,658,460,715.95	2023	Autumn	2,592,422,366.69
		Spring	40,732,920,023.96		Spring	33,274,539,529.84	Spring	21,658,460,715.95	Spring	2,592,422,366.69					
		Summer	40,732,920,023.96		Summer	33,274,539,529.84	Summer	21,658,460,715.95	Summer	2,592,422,366.69					
		Winter	40,732,920,023.96		Winter	33,274,539,529.84	Winter	21,658,460,715.95	Winter	2,592,422,366.69					

# Task 8 - Analysis b)

Viseu	2021	Autumn	4,928,718,200.20
		Spring	4,928,718,200.20
		Summer	4,928,718,200.20
		Winter	4,928,718,200.20
	2022	Autumn	4,928,718,200.20
		Spring	4,928,718,200.20
		Summer	4,928,718,200.20
		Winter	4,928,718,200.20
	2023	Autumn	4,928,718,200.20
		Spring	4,928,718,200.20
		Summer	4,928,718,200.20
		Winter	4,928,718,200.20
Évora	2021	Autumn	5,219,578,022.32
		Spring	5,219,578,022.32
		Summer	5,219,578,022.32
		Winter	5,219,578,022.32
	2022	Autumn	5,219,578,022.32
		Spring	5,219,578,022.32
		Summer	5,219,578,022.32
		Winter	5,219,578,022.32
	2023	Autumn	5,219,578,022.32
		Spring	5,219,578,022.32
		Summer	5,219,578,022.32
		Winter	5,219,578,022.32

# Task 8 - Analysis c)

Year	District	Smart Meter Percentage	Energy consumption			
2021	Aveiro	77.15%	73,505,890,544.69			
	Beja	77.14%	15,601,908,439.15			
	Braga	78.52%	59,916,833,250.31			
	Bragança	80.68%	7,375,929,716.74			
	Castelo Branco	81.3%	18,500,153,653.63			
	Coimbra	76.39%	33,861,484,890.80			
	Évora	87.9%	20,878,312,089.26			
	Faro	75.13%	44,176,376,180.50			
	Guarda	78.23%	11,459,035,069.78			
	Leiria	84.4%	46,865,141,973.89			
	Lisboa	87.66%	162,931,680,095.86			
	Portalegre	86.99%	7,936,245,942.91			
	Porto	76.72%	133,098,158,119.36			
	Santarém	77.17%	39,224,575,077.41			
	Setúbal	87.15%	86,633,842,863.79			
	Viana do Castelo	73.05%	22,331,138,823.50			
	Vila Real	75.59%	10,369,689,466.75			
	Viseu	72.63%	19,714,872,800.78			
2022	Aveiro	77.06%	73,505,890,544.69			
	Beja	77.3%	15,601,908,439.15			
	Braga	78.34%	59,916,833,250.31			
	Bragança	80.56%	7,375,929,716.74			
	Castelo Branco	81.12%	18,500,153,653.63			
	Coimbra	76.5%	33,861,734,062.80			
	Évora	88.02%	20,878,312,089.26			
	Faro	75.21%	44,176,376,180.50			
	Guarda	77.97%	11,459,035,069.78			
	Leiria	84.56%	46,865,141,973.89			
	Lisboa	87.6%	162,931,680,095.86			
	Portalegre	86.7%	7,936,245,942.91			
	Porto	76.51%	133,098,158,119.36			
	Santarém	76.78%	39,224,575,077.41			
	Setúbal	87.72%	86,633,842,863.79			
	Viana do Castelo	73.22%	22,331,138,823.50			
	Vila Real	75.77%	10,369,689,466.75			
	Viseu	72.64%	19,714,872,800.78			
2023	Aveiro	77.01%	73,505,890,544.69			
	Beja	76.44%	15,601,908,439.15			
	Braga	78.17%	59,916,833,250.31			
	Bragança	80.6%	7,375,929,716.74			
	Castelo Branco	80.75%	18,500,153,653.63			
	Coimbra	76.12%	33,861,734,062.80			
	Évora	87.95%	20,878,312,089.26			
	Faro	75.48%	44,176,376,180.50			
	Guarda	78.05%	11,459,035,069.78			
	Leiria	84.4%	46,865,141,973.89			
	Lisboa	87.54%	162,931,680,095.86			
	Portalegre	86.81%	7,936,245,942.91			
	Porto	76.64%	133,108,739,065.96			
	Santarém	76.31%	39,224,575,077.41			
	Setúbal	87.12%	86,633,842,863.79			
	Viana do Castelo	72.87%	22,340,288,920.04			
	Vila Real	75.3%	10,369,689,466.75			
	Viseu	72.63%	19,714,872,800.78			