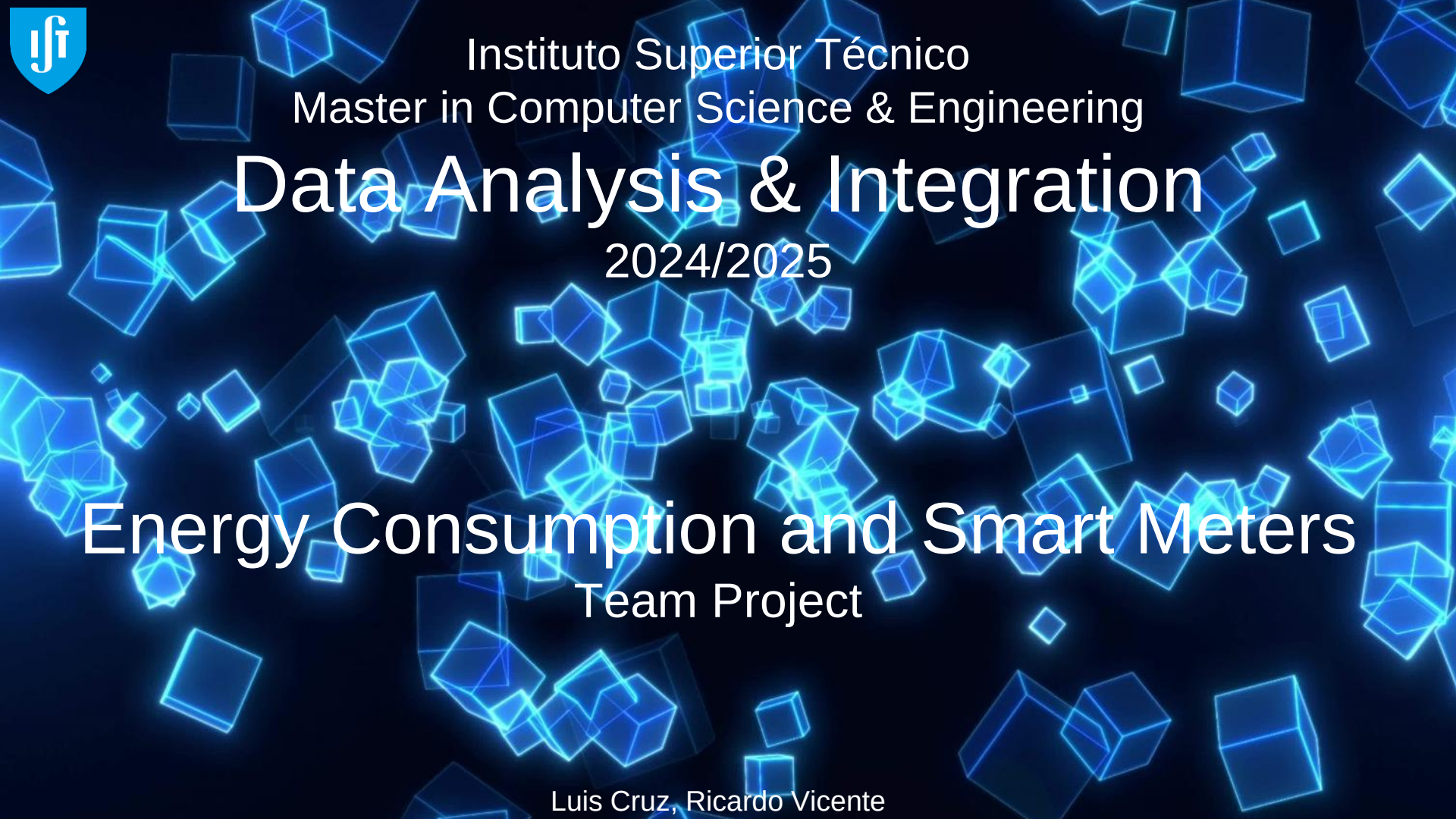


Instituto Superior Técnico

Master in Computer Science & Engineering

Data Analysis & Integration

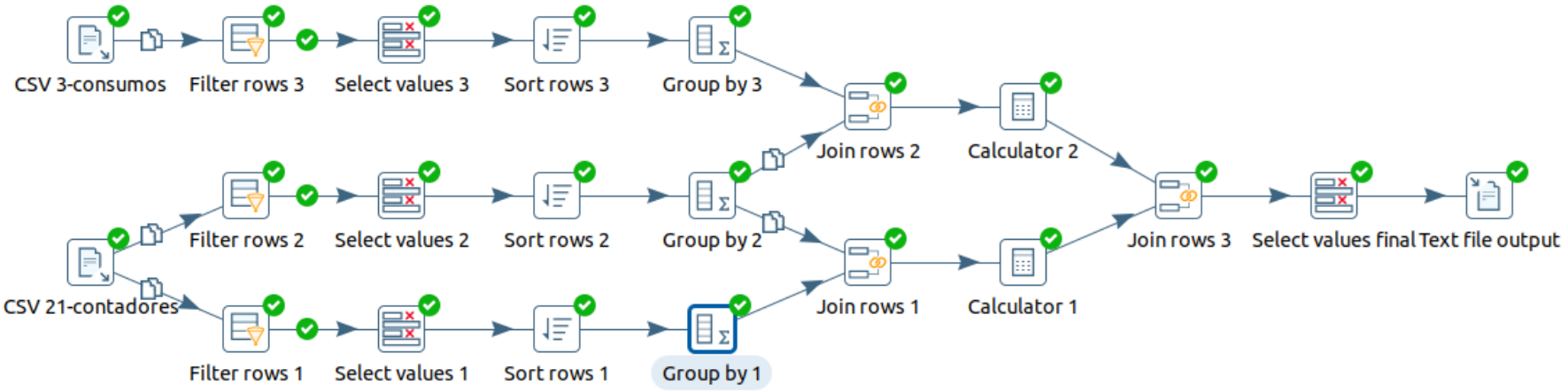
2024/2025

Energy Consumption and Smart Meters

Team Project

Luis Cruz, Ricardo Vicente

Question 1 - Transformation



Question 1 - Transformation: Execution Results

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/22 18:46:20 - project_aid - Dispatching started for transformation [project_aid]
2024/10/22 18:46:21 - CSV 21-contadores.0 - Header row skipped in file '/home/aid/Downloads/Project_AID/21-contadores-de-energia.csv'
2024/10/22 18:46:21 - CSV 3-consumos.0 - Header row skipped in file '/home/aid/Downloads/Project_AID/3-consumos-faturados-por-municipio-ultimos-10-anos.csv'
2024/10/22 18:46:22 - CSV 21-contadores.0 - Line number : 50000
2024/10/22 18:46:22 - Filter rows 1.0 - linenr 50000
2024/10/22 18:46:22 - Filter rows 2.0 - linenr 50000
2024/10/22 18:46:23 - CSV 3-consumos.0 - Line number : 50000
2024/10/22 18:46:23 - Filter rows 3.0 - linenr 50000
2024/10/22 18:46:24 - CSV 21-contadores.0 - Line number : 100000
2024/10/22 18:46:24 - Filter rows 1.0 - linenr 100000
2024/10/22 18:46:24 - Filter rows 2.0 - linenr 100000
2024/10/22 18:46:25 - CSV 3-consumos.0 - Line number : 100000
2024/10/22 18:46:25 - Filter rows 3.0 - linenr 100000
2024/10/22 18:46:26 - CSV 21-contadores.0 - Finished processing (I=148852, O=0, R=0, W=297702, U=0, E=0)
2024/10/22 18:46:26 - Filter rows 2.0 - Finished processing (I=0, O=0, R=148851, W=5674, U=0, E=0)
2024/10/22 18:46:26 - Filter rows 1.0 - Finished processing (I=0, O=0, R=148851, W=2878, U=0, E=0)
2024/10/22 18:46:26 - Select values 2.0 - Finished processing (I=0, O=0, R=5674, W=5674, U=0, E=0)
2024/10/22 18:46:26 - Select values 1.0 - Finished processing (I=0, O=0, R=2878, W=2878, U=0, E=0)
2024/10/22 18:46:26 - Sort rows 1.0 - Finished processing (I=0, O=0, R=2878, W=2878, U=0, E=0)
2024/10/22 18:46:26 - Group by 1.0 - Finished processing (I=0, O=0, R=2878, W=278, U=0, E=0)
2024/10/22 18:46:26 - Sort rows 2.0 - Finished processing (I=0, O=0, R=5674, W=5674, U=0, E=0)
2024/10/22 18:46:26 - Group by 2.0 - Finished processing (I=0, O=0, R=5674, W=556, U=0, E=0)
2024/10/22 18:46:26 - CSV 3-consumos.0 - Line number : 150000
2024/10/22 18:46:26 - Filter rows 3.0 - linenr 150000
2024/10/22 18:46:27 - Join rows 1.0 - Finished processing (I=0, O=0, R=556, W=278, U=0, E=0)
2024/10/22 18:46:27 - Calculator 1.0 - Finished processing (I=0, O=0, R=278, W=278, U=0, E=0)
2024/10/22 18:46:27 - CSV 3-consumos.0 - Finished processing (I=196336, O=0, R=0, W=196335, U=0, E=0)
2024/10/22 18:46:27 - Filter rows 3.0 - Finished processing (I=0, O=0, R=196335, W=4370, U=0, E=0)
2024/10/22 18:46:27 - Select values 3.0 - Finished processing (I=0, O=0, R=4370, W=4370, U=0, E=0)
2024/10/22 18:46:27 - Sort rows 3.0 - Finished processing (I=0, O=0, R=4370, W=4370, U=0, E=0)
2024/10/22 18:46:27 - Group by 3.0 - Finished processing (I=0, O=0, R=4370, W=278, U=0, E=0)
2024/10/22 18:46:27 - Join rows 2.0 - Finished processing (I=0, O=0, R=556, W=278, U=0, E=0)
2024/10/22 18:46:27 - Calculator 2.0 - Finished processing (I=0, O=0, R=278, W=278, U=0, E=0)
2024/10/22 18:46:27 - Join rows 3.0 - Finished processing (I=0, O=0, R=556, W=278, U=0, E=0)
2024/10/22 18:46:27 - Select values final.0 - Finished processing (I=0, O=0, R=278, W=278, U=0, E=0)
2024/10/22 18:46:27 - Text file output.0 - Finished processing (I=0, O=279, R=278, W=278, U=0, E=0)
2024/10/22 18:46:27 - Spoon - The transformation has finished!!
```

Question 1 - Initial Operations: CSV 3 - Consumos

CSV File Input

Step name	CSV 3-consumos
Filename	${Internal.Entry.Current.Directory}/CSV/3-consumos-faturados-por
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
1	Year	Integer	#	15	0	$
2	Month	Integer	#	15	0	$
3	Date	String		7		$

Rows of step: CSV 3-consumos (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)
1	2021	5	2021-05	BEJA	Almodôvar	ALDEIA DOS FERNANDES	Baixa Tensão	70014.8
2	2021	5	2021-05	EVORA	Borba	RIO DE MOINHOS	Muito Alta, Alta e Média Tensões	92461.7
3	2021	5	2021-05	EVORA	Viana do Alentejo	AGUIAR	Baixa Tensão	139795.3
4	2021	5	2021-05	LEIRIA	Leiria	UF PARCEIROS E AZOIA	Baixa Tensão	1728071.9
5	2021	5	2021-05	LISBOA	Sintra	UF CACEM E SAO MARCOS	Baixa Tensão	3863416.5
6	2021	5	2021-05	PORTO	Vila Nova de Gaia	MADALENA	Baixa Tensão	1505987.9
7	2021	5	2021-05	SANTAREM	Ourém	FATIMA	Muito Alta, Alta e Média Tensões	2775022
8	2021	5	2021-05	VISEU	Resende	UF OVADAS E PANCHORRA	Baixa Tensão	26305.7
9	2021	6	2021-06	BRAGA	Guimarães	BRITO	Muito Alta, Alta e Média Tensões	1677884.5
10	2021	6	2021-06	COIMBRA	Cantanhede	OURENTA	Baixa Tensão	126207.7
11	2021	6	2021-06	PORTO	Gondomar	UF FANZERES SAO PEDRO COVA	Baixa Tensão	4472527.5

Question 1 - Initial Operations: CSV 3 - Consumos

Filter rows

Step name: Filter rows 3

Send 'true' data to step: Select values 3

Send 'false' data to step:

The condition:

```
Year = [2024]
AND
Month = [6]
```

Rows of step: Filter rows 3 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	Di
1	2024	6	2024-06	COIMBRA	Condeixa-a-Nova	EGA	Muito Alta, Alta e Média Tensões	20559	
2	2024	6	2024-06	VIANA DO CASTELO	Caminha	UF VENADE E AZEVEDO	Baixa Tensão	67713	
3	2024	6	2024-06	VIANA DO CASTELO	Monção	MOREIRA	Baixa Tensão	35454	
4	2024	6	2024-06	VIANA DO CASTELO	Ponte de Lima	GEMIEIRA	Baixa Tensão	38425	
5	2024	6	2024-06	VISEU	Moimenta da Beira	UF PERA VELHA ALD NACOMBA ARIZ	Baixa Tensão	26448	
6	2024	6	2024-06	BRAGA	Barcelos	UF SILVEIROS E RIO COVO	Muito Alta, Alta e Média Tensões	36717	
7	2024	6	2024-06	BRAGA	Barcelos	VARZEA	Baixa Tensão	180725	
8	2024	6	2024-06	BRAGA	Vila Nova de Famalicão	GAVIAO	Baixa Tensão	218184	
9	2024	6	2024-06	BRAGANCA	Bragança	GONDESENDE	Baixa Tensão	7800	
10	2024	6	2024-06	CASTELO BRANCO	Covilhã	UF PESO E VALES DO RIO	Baixa Tensão	79682	
11	2024	6	2024-06	COIMBRA	Coimbra	CERNACHE	Muito Alta, Alta e Média Tensões	471267	
12	2024	6	2024-06	EVORA	Estremoz	UF S BENTO CORTICO E S ESTEVAO	Baixa Tensão	58525	
13	2024	6	2024-06	PORTO	Marco de Canaveses	BEM VIVER	Baixa Tensão	227450	

Question 1 - Initial Operations: CSV 3 - Consumos

Step name: Select values 3

Select & Alter | Remove | Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_3		
2	District	District_3		
3	DistrictMunicipalityCode	MunicipalityCode_3		
4	Municipality	Municipality_3		
5	Active Energy (kWh)			

Rows of step: Select values 3 (1000 rows)

	Dis	District_3	Municipa	Municipality_3	Active Energy (kWh)
1	6	COIMBRA	604	Condeixa-a-Nova	20559
2	16	VIANA DO CASTELO	1602	Caminha	67713
3	16	VIANA DO CASTELO	1604	Monção	35454
4	16	VIANA DO CASTELO	1607	Ponte de Lima	38425
5	18	VISEU	1807	Moimenta da Beira	26448
6	3	BRAGA	302	Barcelos	36717
7	3	BRAGA	302	Barcelos	180725
8	3	BRAGA	312	Vila Nova de Famalicão	218184
9	4	BRAGANCA	402	Bragança	7800

Question 1 - Initial Operations: CSV 3 - Consumos

Step name: Sort rows 3
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)

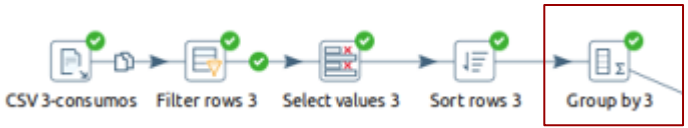
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presor
1	DistrictCode_3	Y	N	N	0	N
2	MunicipalityCode_3	Y	N	N	0	N

Rows of step: Sort rows 3 (1000 rows)

	DistrictCod	District_3	Municipality	Municipality_3	Active Energy (kWh)
1	1	AVEIRO	101	Águeda	699874
2	1	AVEIRO	101	Águeda	81119
3	1	AVEIRO	101	Águeda	3555
4	1	AVEIRO	101	Águeda	305487
5	1	AVEIRO	101	Águeda	58428
6	1	AVEIRO	101	Águeda	504891
7	1	AVEIRO	101	Águeda	687897
8	1	AVEIRO	101	Águeda	295088
9	1	AVEIRO	101	Águeda	290943
10	1	AVEIRO	101	Águeda	117348
11	1	AVEIRO	101	Águeda	89338

Question 1 - Initial Operations: CSV 3 - Consumos



Step name: Group by 3

Include all rows? ☐

Temporary files directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: grp

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

	Group field
1	DistrictCode_3
2	MunicipalityCode_3

Get Fields

Aggregates :

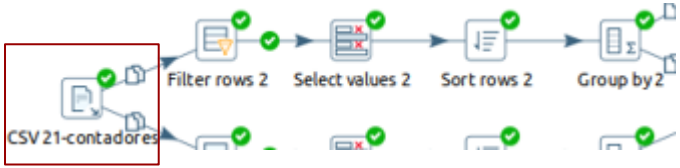
	Name	Subject	Type
1	total_energy_munic	Active Energy (kWh)	Sum

Get lookup fields

Rows of step: Group by 3 (278 rows)

	DistrictCode_3	MunicipalityCode_3	total_energy_munic
1	1	101	6287917.0
2	1	102	3704565.0
3	1	103	3150888.0
4	1	104	1807586.0
5	1	105	11376163.0
6	1	106	1020907.0
7	1	107	2772488.0
8	1	108	3928678.0
9	1	109	16672271.0
10	1	110	5683603.0
11	1	111	1760919.0

Question 1 - Initial Operations: CSV 21 - Contadores



CSV file input

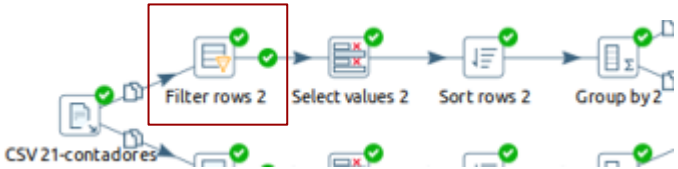
Step name	CSV 21-contadores
Filename	${Internal.Entry.Current.Directory}/CSV/21-contadores-de-energia. (Browse...)
Delimiter	; (Insert TAB)
Enclosure	"
NIO buffer size	50000
Lazy conversion?	☐
Header row present?	☑
Add filename to result	☐
The row number field name (optional)	
Running in parallel?	☐
New line possible in fields?	☐
Format	mixed
File encoding	UTF-8

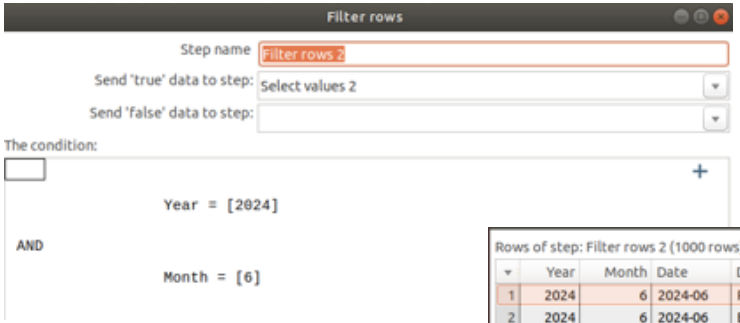
	Name	Type	Format	Length	Precision	Currency	Decima
1	Year	Integer	#	15	0	$	.
2	Month	Integer	#	15	0	$	.
3	Date	String		7		$	.

Rows of step: CSV 21-contadores (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode
1	2024	5	2024-05	SETUBAL	Sines	SINES	Não	128	15
2	2024	5	2024-05	PORTALEGRE	Ponte de Sor	MONTARGIL	Não	475	12
3	2024	5	2024-05	VISEU	Santa Comba Dão	UF S COMBA DAO COUTO MOSTEIRO	Sim	3141	18
4	2024	5	2024-05	CASTELO BRANCO	Belmonte	CARIA	Sim	1378	5
5	2024	5	2024-05	CASTELO BRANCO	Castelo Branco	SAO VICENTE DA BEIRA	Sim	620	5
6	2024	5	2024-05	FARO	Albufeira	GUIA	Não	1335	8
7	2024	5	2024-05	SANTAREM	Ourém	UF MATAS E CERCAL	Sim	1056	14
8	2024	5	2024-05	VIANA DO CASTELO	Paredes de Coura	AGUALONGA	Sim	175	16
9	2024	5	2024-05	BEJA	Odemira	RELIQUIAS	Sim	334	2

Question 1 - Initial Operations: CSV 21 - Contadores

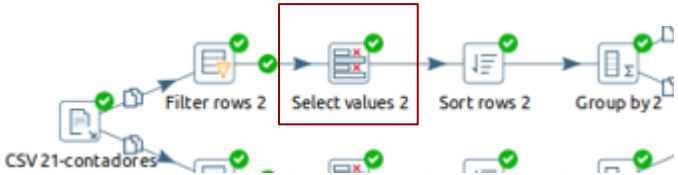




Rows of step: Filter rows 2 (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode
1	2024	6	2024-06	PORTO	Marco de Canaveses	SOALHAES	Não	406	13
2	2024	6	2024-06	EVORA	Mora	MORA	Não	228	7
3	2024	6	2024-06	PORTO	Paredes	SOBREIRA	Sim	1678	13
4	2024	6	2024-06	PORTO	Vila Nova de Gaia	SAO FELIX DA MARINHA	Não	1588	13
5	2024	6	2024-06	BRAGA	Amares	GOAES	Sim	315	3
6	2024	6	2024-06	GUARDA	Guarda	CODESSEIRO	Sim	139	9
7	2024	6	2024-06	SANTAREM	Mação	ENVENDOS	Sim	916	14
8	2024	6	2024-06	VILA REAL	Ribeira de Pena	SANTA MARINHA	Sim	419	17
9	2024	6	2024-06	BRAGA	Barcelos	CAMBESES	Não	32	3

Question 1 - Initial Operations: CSV 21 - Contadores



Step name: Select values 2

Select & Alter | Remove | Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_2		
2	District	District_2		
3	DistrictMunicipalityCode	MunicipalityCode_2		
4	Municipality	Municipality_2		
5	Number of CPE's			

Rows of step: Select values 2 (1000 rows)

	DistrictCode_2	District_2	MunicipalityCode_2	Municipality_2	Number of CPE's
1	13	PORTO	1307	Marco de Canaveses	406
2	7	EVORA	707	Mora	228
3	13	PORTO	1310	Paredes	1678
4	13	PORTO	1317	Vila Nova de Gaia	1588
5	3	BRAGA	301	Amares	315
6	9	GUARDA	907	Guarda	139
7	14	SANTAREM	1413	Mação	916
8	17	VILA REAL	1709	Ribeira de Pena	419
9	3	BRAGA	302	Barcelos	32
10	17	VILA REAL	1711	Santa Marta de Penaguião	13
11	10	LEIRIA	1016	Porto de Mós	8
12	16	VIANA DO CASTELO	1609	Viana do Castelo	2675

Question 1 - Initial Operations: CSV 21 - Contadores

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?	Collat
1	DistrictCode_2	Y	N	N	0
2	MunicipalityCode_2	Y	N	N	0

Rows of step: Sort rows 2 (1000 rows)

	DistrictCode_2	District_2	MunicipalityCode_2	Municipality_2	Number of CPE's
1	1	AVEIRO	101	Águeda	867
2	1	AVEIRO	101	Águeda	16
3	1	AVEIRO	101	Águeda	11
4	1	AVEIRO	101	Águeda	27
5	1	AVEIRO	101	Águeda	36
6	1	AVEIRO	101	Águeda	21
7	1	AVEIRO	101	Águeda	20
8	1	AVEIRO	101	Águeda	2165
9	1	AVEIRO	101	Águeda	2410
10	1	AVEIRO	101	Águeda	1843
11	1	AVEIRO	101	Águeda	4
12	1	AVEIRO	101	Águeda	12
13	1	AVEIRO	101	Águeda	3065

Question 1 - Initial Operations: CSV 21 - Contadores

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	DistrictCode_2
2	MunicipalityCode_2

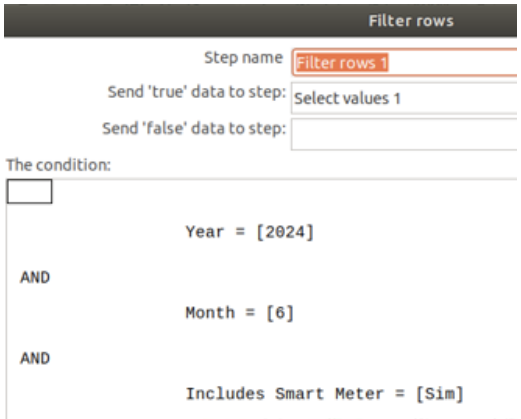
Aggregates :

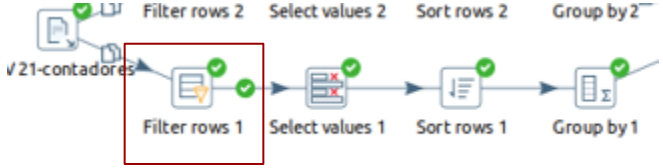
	Name	Subject	Type	Value
1	total_cpes	Number of CPE's	Sum	

Rows of step: Group by 2 (278 rows)

	DistrictCode_2	MunicipalityCode_2	total_cpes
1	1	101	24011
2	1	102	13897
3	1	103	16871
4	1	104	12337
5	1	105	50559
6	1	106	7904
7	1	107	18385
8	1	108	14156
9	1	109	70801
10	1	110	25089
11	1	111	11333

Question 1 - Initial Operations: CSV 21 - Contadores





Rows of step: Filter rows 1 (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode
1	2024	6	2024-06	PORTO	Paredes	SOBREIRA	Sim	1678	13
2	2024	6	2024-06	BRAGA	Amares	GOAES	Sim	315	3
3	2024	6	2024-06	GUARDA	Guarda	CODESSEIRO	Sim	139	9
4	2024	6	2024-06	SANTAREM	Mação	ENVENDOS	Sim	916	14
5	2024	6	2024-06	VILA REAL	Ribeira de Pena	SANTA MARINHA	Sim	419	17
6	2024	6	2024-06	VIANA DO CASTELO	Viana do Castelo	AREOSA	Sim	2675	16
7	2024	6	2024-06	SANTAREM	Abrantes	CARVALHAL	Sim	385	14
8	2024	6	2024-06	BRAGA	Braga	UF BRAGA JOSE S LAZARO E SOUTO	Sim	10841	3

Question 1 - Initial Operations: CSV 21 - Contadores

Step name: Select values 1

Select & Alter | Remove | Meta-data

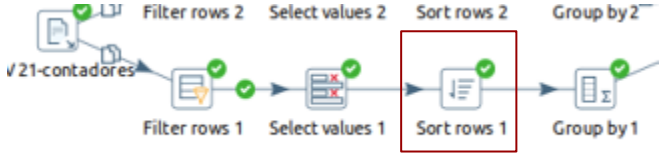
Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_1		
2	District	District_1		
3	DistrictMunicipalityCode	MunicipalityCode_1		
4	Municipality	Municipality_1		
5	Number of CPE's			

Rows of step: Select values 1 (1000 rows)

	Distri	District_1	MunicipalityCode_1	Municipality_1	Number of CPE's
1	13	PORTO	1310	Paredes	1678
2	3	BRAGA	301	Amares	315
3	9	GUARDA	907	Guarda	139
4	14	SANTAREM	1413	Mação	916
5	17	VILA REAL	1709	Ribeira de Pena	419
6	16	VIANA DO CASTELO	1609	Viana do Castelo	2675
7	14	SANTAREM	1401	Abrantes	385
8	3	BRAGA	303	Braga	10841
9	18	VISEU	1823	Viseu	3550
10	5	CASTELO BRANCO	507	Penamacor	458
11	16	VIANA DO CASTELO	1608	Valença	486

Question 1 - Initial Operations: CSV 21 - Contadores



Step name: Sort rows 1

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) ☐

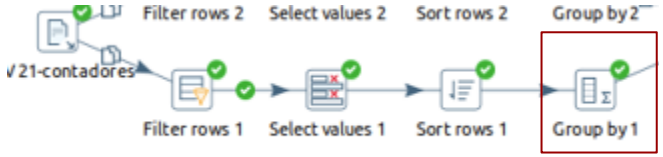
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?
1	DistrictCode_1	Y	N	N
2	MunicipalityCode_1	Y	N	N

Rows of step: Sort rows 1 (1000 rows)

	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	Number of CPE's
1	1	AVEIRO	101	Águeda	867
2	1	AVEIRO	101	Águeda	2165
3	1	AVEIRO	101	Águeda	2410
4	1	AVEIRO	101	Águeda	1843
5	1	AVEIRO	101	Águeda	3065
6	1	AVEIRO	101	Águeda	171
7	1	AVEIRO	101	Águeda	1617
8	1	AVEIRO	101	Águeda	1801
9	1	AVEIRO	101	Águeda	584
10	1	AVEIRO	101	Águeda	8079
11	1	AVEIRO	101	Águeda	1172
12	1	AVEIRO	102	Albergaria-a-Velha	711
13	1	AVEIRO	102	Albergaria-a-Velha	1291
14	1	AVEIRO	102	Albergaria-a-Velha	990
15	1	AVEIRO	102	Albergaria-a-Velha	2735

Question 1 - Initial Operations: CSV 21 - Contadores



Step name: Group by 1

Include all rows? ☐

Temporary files directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: grp

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

	Group field
1	DistrictCode_1
2	MunicipalityCode_1
3	Municipality_1

Get Fields

Aggregates :

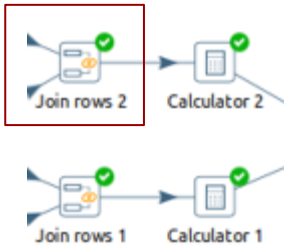
	Name	Subject	Type	V
1	sum_CPE_sm_SIM	Number of CPE's	Sum	

Get lookup fields

Rows of step: Group by 1 (278 rows)

	DistrictCode_1	MunicipalityCode_1	Municipality_1	sum_CPE_sm_SIM
1	1	101	Águeda	23774
2	1	102	Albergaria-a-Velha	13437
3	1	103	Anadia	14682
4	1	104	Arouca	9682
5	1	105	Aveiro	50414
6	1	106	Castelo de Paiva	7834
7	1	107	Espinho	18258
8	1	108	Estarreja	14009
9	1	109	Santa Maria da Feira	65079
10	1	110	Ílhavo	24983
11	1	111	Mealhada	11200
12	1	112	Murtosa	8345

Question 1 - Joins and Calculators



Step name: Join rows 2
Temp directory: %%java.io.tmpdir%% Browse...
TMP-file prefix: out
Max. cache size (in rows): 500
Main step to read from:

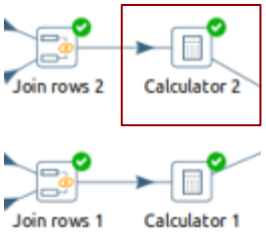
The condition:

```
DistrictCode_2 = DistrictCode_3

AND

MunicipalityCode_2 = MunicipalityCode_3
```

Rows of step: Join rows 2 (278 rows)

	DistrictCode_3	MunicipalityCode_3	total_energy_munic	DistrictCode_2	MunicipalityCode_2	total_cpes
1	1	101	6287917.0	1	101	24011
2	1	102	3704565.0	1	102	13897
3	1	103	3150888.0	1	103	16871
4	1	104	1807586.0	1	104	12337
5	1	105	11376163.0	1	105	50559
6	1	106	1020907.0	1	106	7904
7	1	107	2772488.0	1	107	18385
8	1	108	3928678.0	1	108	14156
9	1	109	16672271.0	1	109	70801
10	1	110	5683603.0	1	110	25089

Question 1 - Joins and Calculators



Step name

Calculator 2

Throw an error on non existing files

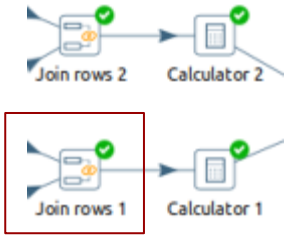
Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Convers
1	cons_per_contr	A / B	total_energy_munic	total_cpes		Number			N	

Rows of step: Calculator 2 (278 rows)

	DistrictCode_3	MunicipalityCode_	total_energy_munic	DistrictCode_2	MunicipalityCode_2	total_cpes	cons_per_contr
1	1	101	6287917.0	1	101	24011	261.8765149307
2	1	102	3704565.0	1	102	13897	266.5730013672
3	1	103	3150888.0	1	103	16871	186.7635587695
4	1	104	1807586.0	1	104	12337	146.5174677798
5	1	105	11376163.0	1	105	50559	225.0076742024
6	1	106	1020907.0	1	106	7904	129.1633350202
7	1	107	2772488.0	1	107	18385	150.801631765
8	1	108	3928678.0	1	108	14156	277.5274088726
9	1	109	16672271.0	1	109	70801	235.4807276733
10	1	110	5683603.0	1	110	25089	226.5376459803
11	1	111	1760919.0	1	111	11333	155.3797758758
12	1	112	1096356.0	1	112	8450	129.7462721893

Question 1 - Joins and Calculators



Step name: Join rows 1
Temp directory: %%java.io.tmpdir%% Browse...
TMP-file prefix: out
Max. cache size (in rows): 500
Main step to read from:

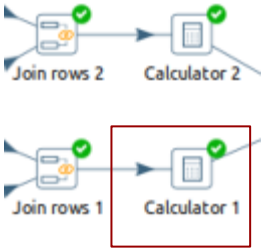
The condition:

```
DistrictCode_1 = DistrictCode_2

AND

MmunicipalityCode_1 = MunicipalityCode_2
```

Rows of step: Join rows 1 (278 rows)

	DistrictCode_2	MunicipalityCode_2	total_cpes	DistrictCode_1	MunicipalityCode_1	Municipality_1	sum_CPE_sm_SIM
1	1	101	24011	1	101	Águeda	23774
2	1	102	13897	1	102	Albergaria-a-Velha	13437
3	1	103	16871	1	103	Anadia	14682
4	1	104	12337	1	104	Arouca	9682
5	1	105	50559	1	105	Aveiro	50414
6	1	106	7904	1	106	Castelo de Paiva	7834
7	1	107	18385	1	107	Espinho	18258
8	1	108	14156	1	108	Estarreja	14009
9	1	109	70801	1	109	Santa Maria da Feira	65079
10	1	110	25089	1	110	Ílhavo	24983
11	1	111	11333	1	111	Mealhada	11200
12	1	112	8450	1	112	Murtosa	8345
13	1	113	32400	1	113	Oliveira de Azeméis	28372

Question 1 - Joins and Calculators



Step name

Calculator 1

Throw an error on non existing files

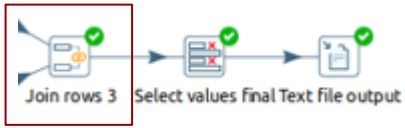
Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decim
1	percent_smart_cpe	100 * A / B	sum_CPE_sm_SIM	total_cpes		Number			N		

Rows of step: Calculator 1 (278 rows)

	DistrictCode_2	MunicipalityCode_2	total_cpes	DistrictCode_1	MunicipalityCode_1	Municipality_1	sum_CPE_sm_SIM	percent_smart_cpe
1	1	101	24011	1	101	Águeda	23774	99.0
2	1	102	13897	1	102	Albergaria-a-Velha	13437	96.0
3	1	103	16871	1	103	Anadia	14682	87.0
4	1	104	12337	1	104	Arouca	9682	78.0
5	1	105	50559	1	105	Aveiro	50414	99.0
6	1	106	7904	1	106	Castelo de Paiva	7834	99.0
7	1	107	18385	1	107	Espinho	18258	99.0
8	1	108	14156	1	108	Estarreja	14009	98.0
9	1	109	70801	1	109	Santa Maria da Feira	65079	91.0
10	1	110	25089	1	110	Ílhavo	24983	99.0
11	1	111	11333	1	111	Mealhada	11200	98.0
12	1	112	8450	1	112	Murtosa	8345	98.0
13	1	113	32400	1	113	Oliveira de Azeméis	28372	87.0
14	1	114	13374	1	114	Oliveira do Bairro	13304	99.0

Question 1 - Final Join and Output to File



Step name: Join rows 3
Temp directory: %%java.io.tmpdir%% Browse...
TMP-file prefix: out
Max. cache size (in rows): 500
Main step to read from:

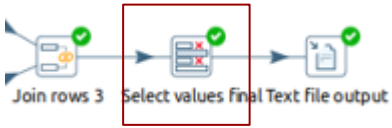
The condition:

```
DistrictCode_1 = DistrictCode_3

AND

MmunicipalityCode_1 = MunicipalityCode_3
```

Join rows 3 (278 rows)

total_cpes	DistrictCode_1	MunicipalityCode_1	Municipality_1	sum_CPE_sm_SIM	percent_smart_cpe	DistrictCode_3	MunicipalityCode_3	total_energy_munic	DistrictCode_2_1	MunicipalityCode_2_1	total_cpes_1	cons_per_contr
24011	1	101	Águeda	23774	99.0	1	101	6287917.0	1	101	24011	261.8765149307
13897	1	102	Albergaria-a-Velha	13437	96.0	1	102	3704565.0	1	102	13897	266.5730013672
16871	1	103	Anadia	14682	87.0	1	103	3150888.0	1	103	16871	186.7635587695
12337	1	104	Arouca	9682	78.0	1	104	1807586.0	1	104	12337	146.5174677798
50559	1	105	Aveiro	50414	99.0	1	105	11376163.0	1	105	50559	225.0076742024
7904	1	106	Castelo de Paiva	7834	99.0	1	106	1020907.0	1	106	7904	129.1633350202
18385	1	107	Espinho	18258	99.0	1	107	2772488.0	1	107	18385	150.801631765
14156	1	108	Estarreja	14009	98.0	1	108	3928678.0	1	108	14156	277.5274088726
70801	1	109	Santa Maria da Feira	65079	91.0	1	109	16672271.0	1	109	70801	235.4807276733

Question 1 - Final Join and Output to File



Select values

Step name: Select values final

Select & Alter | Remove | Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	Municipality_1	Municipality		
2	percent_smart_cpe	PercentageSmartMeter		
3	cons_per_contr	ConsumptionPerContract		

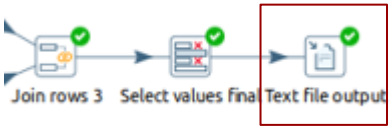
Get fields to select

Edit Mapping

Rows of step: Select values final (278 rows)

	Municipality	PercentageSmartMeter	ConsumptionPerContract
1	Albergaria-a-Velha	96.0	266.5730013672
2	Anadia	87.0	186.7635587695
3	Arouca	78.0	146.5174677798
4	Aveiro	99.0	225.0076742024
5	Castelo de Paiva	99.0	129.1633350202
6	Espinho	99.0	150.801631765
7	Estarreja	98.0	277.5274088726
8	Mealhada	98.0	155.3797758758

Question 1 - Final Join and Output to File



Step name: Text file output

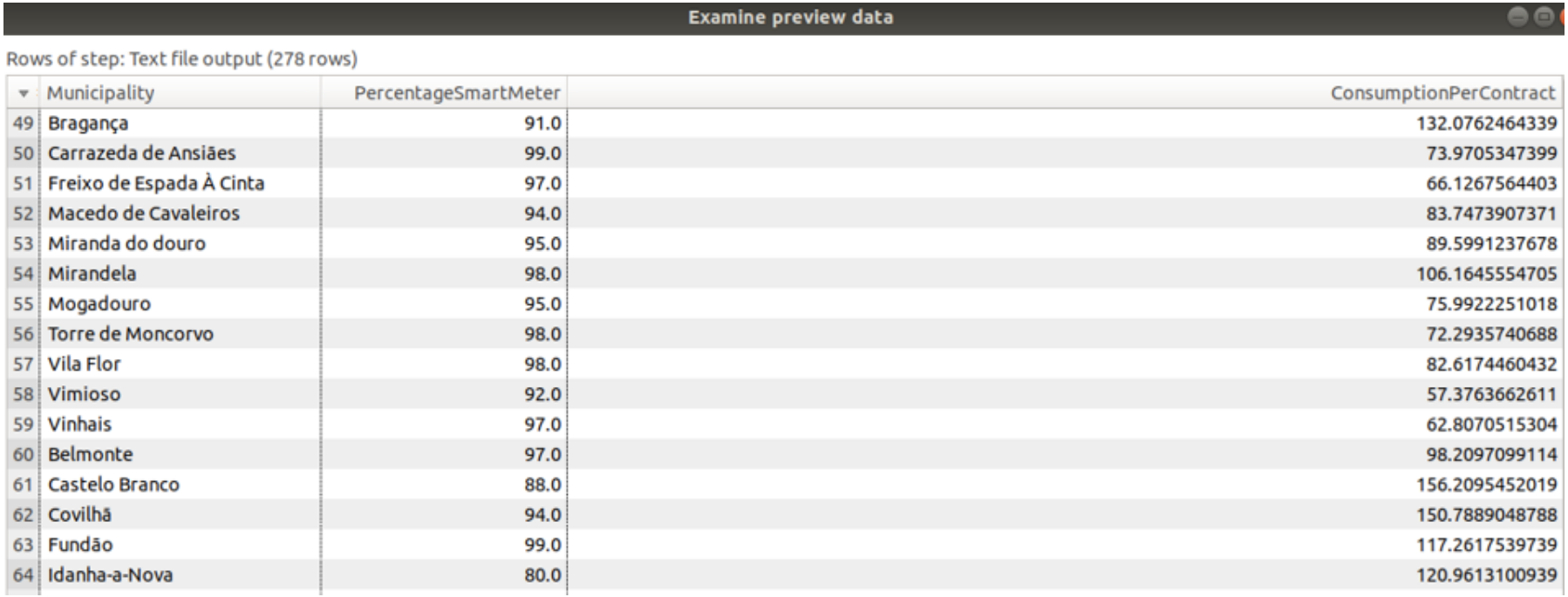
Filename: ${Internal.Entry.Current.Directory}/1_output

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 ☑

Rows of step: Text file output (278 rows)

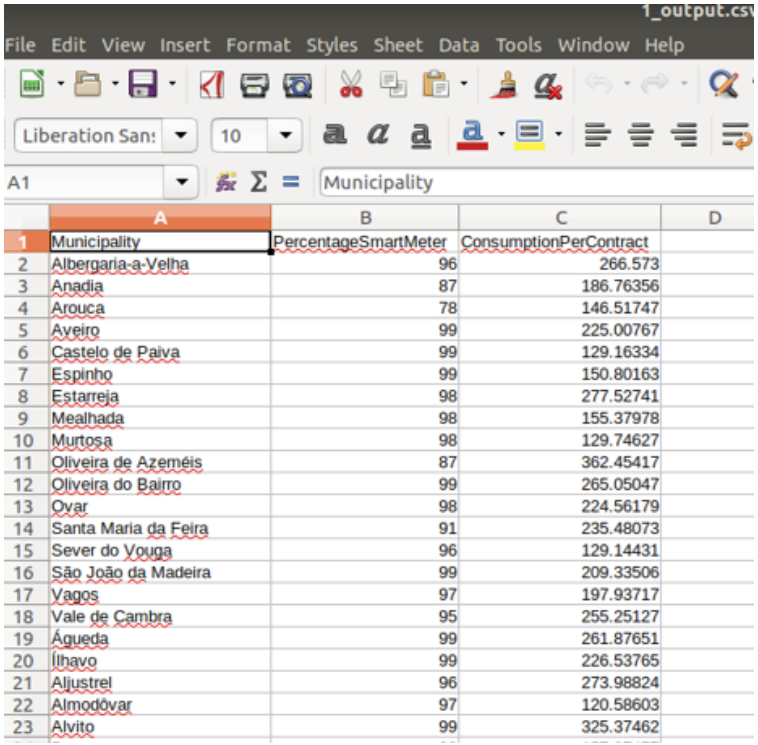
	Municipality	PercentageSmartMeter	ConsumptionPerContract
1	Albergaria-a-Velha	96.0	266.5730013672
2	Anadia	87.0	186.7635587695
3	Arouca	78.0	146.5174677798
4	Aveiro	99.0	225.0076742024
5	Castelo de Paiva	99.0	129.1633350202
6	Espinho	99.0	150.801631765
7	Estarreja	98.0	277.5274088726
8	Mealhada	98.0	155.3797758758
9	Murtosa	98.0	129.7462721893
10	Oliveira de Azeméis	87.0	362.4541666667

Question 1 - Final Result in Pentaho



	Municipality	PercentageSmartMeter	ConsumptionPerContract
49	Bragança	91.0	132.0762464339
50	Carrazeda de Ansiães	99.0	73.9705347399
51	Freixo de Espada À Cinta	97.0	66.1267564403
52	Macedo de Cavaleiros	94.0	83.7473907371
53	Miranda do douro	95.0	89.5991237678
54	Mirandela	98.0	106.1645554705
55	Mogadouro	95.0	75.9922251018
56	Torre de Moncorvo	98.0	72.2935740688
57	Vila Flor	98.0	82.6174460432
58	Vimioso	92.0	57.3763662611
59	Vinhais	97.0	62.8070515304
60	Belmonte	97.0	98.2097099114
61	Castelo Branco	88.0	156.2095452019
62	Covilhã	94.0	150.7889048788
63	Fundão	99.0	117.2617539739
64	Idanha-a-Nova	80.0	120.9613100939

Question 1 - Final Result in File



1_output.csv

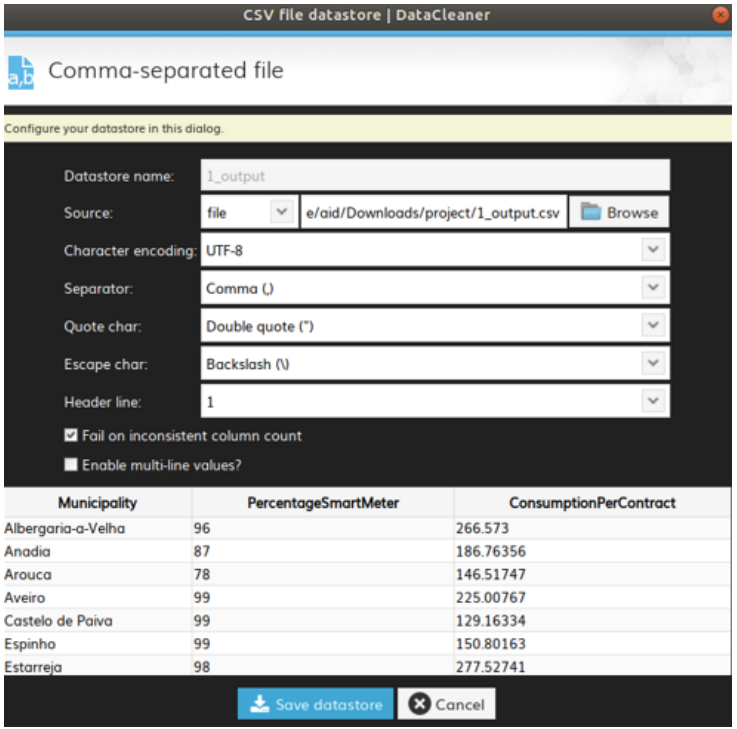
File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation San: 10

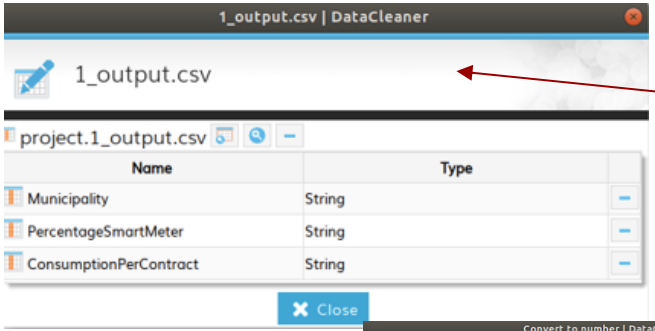
A1 Municipality

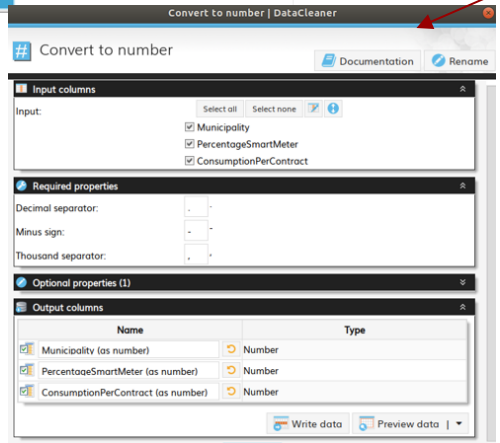
	A	B	C	D
1	Municipality	PercentageSmartMeter	ConsumptionPerContract	
2	Albergaria-a-Velha	96	266.573	
3	Anadia	87	186.76356	
4	Arouca	78	146.51747	
5	Aveiro	99	225.00767	
6	Castelo de Paiva	99	129.16334	
7	Espinho	99	150.80163	
8	Estarreja	98	277.52741	
9	Mealhada	98	155.37978	
10	Murtosa	98	129.74627	
11	Oliveira de Azeméis	87	362.45417	
12	Oliveira do Bairro	99	265.05047	
13	Ovar	98	224.56179	
14	Santa Maria da Feira	91	235.48073	
15	Sever do Vouga	96	129.14431	
16	São João da Madeira	99	209.33506	
17	Vagos	97	197.93717	
18	Vale de Cambra	95	255.25127	
19	Águeda	99	261.87651	
20	Ílhavo	99	226.53765	
21	Aljustrel	96	273.98824	
22	Almodôvar	97	120.58603	
23	Alvito	99	325.37462	

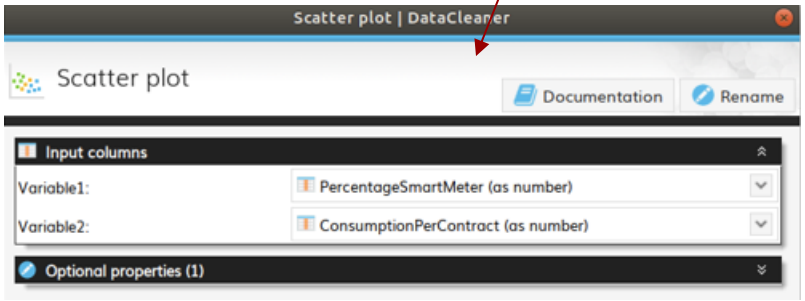
Question 2 - DataCleaner



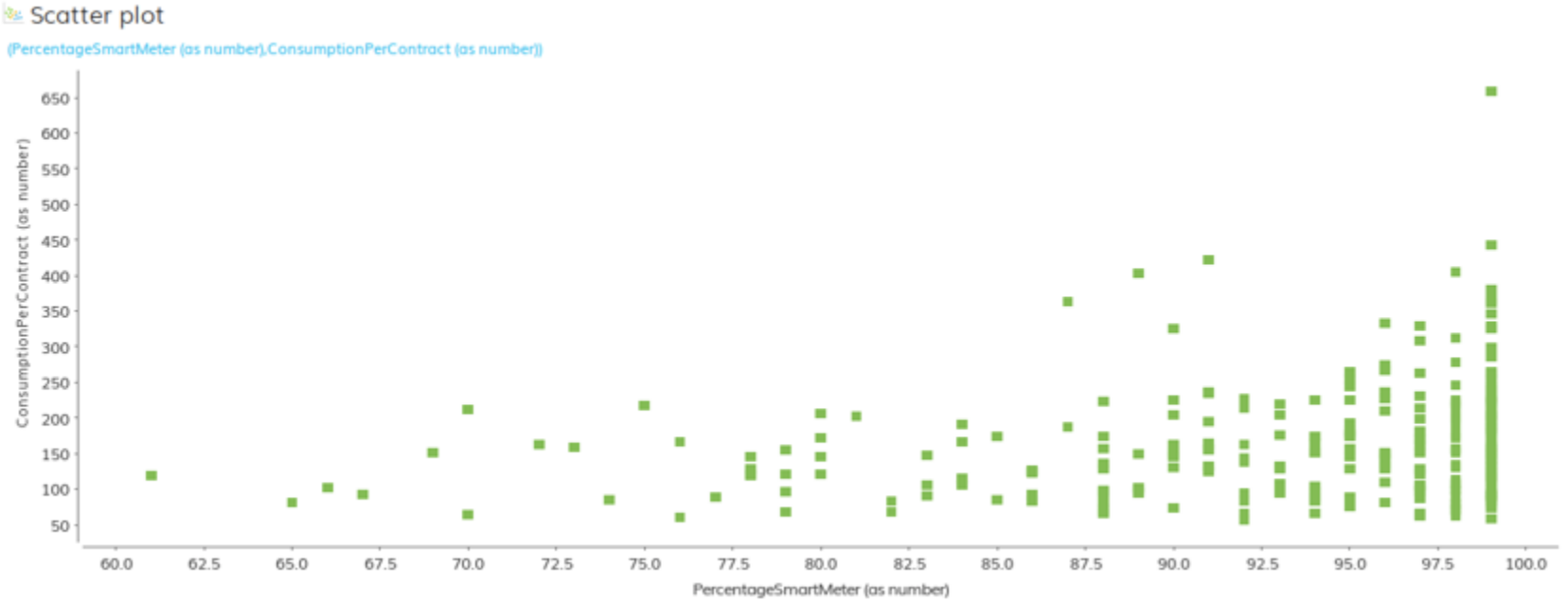
Question 2 - DataCleaner: Job







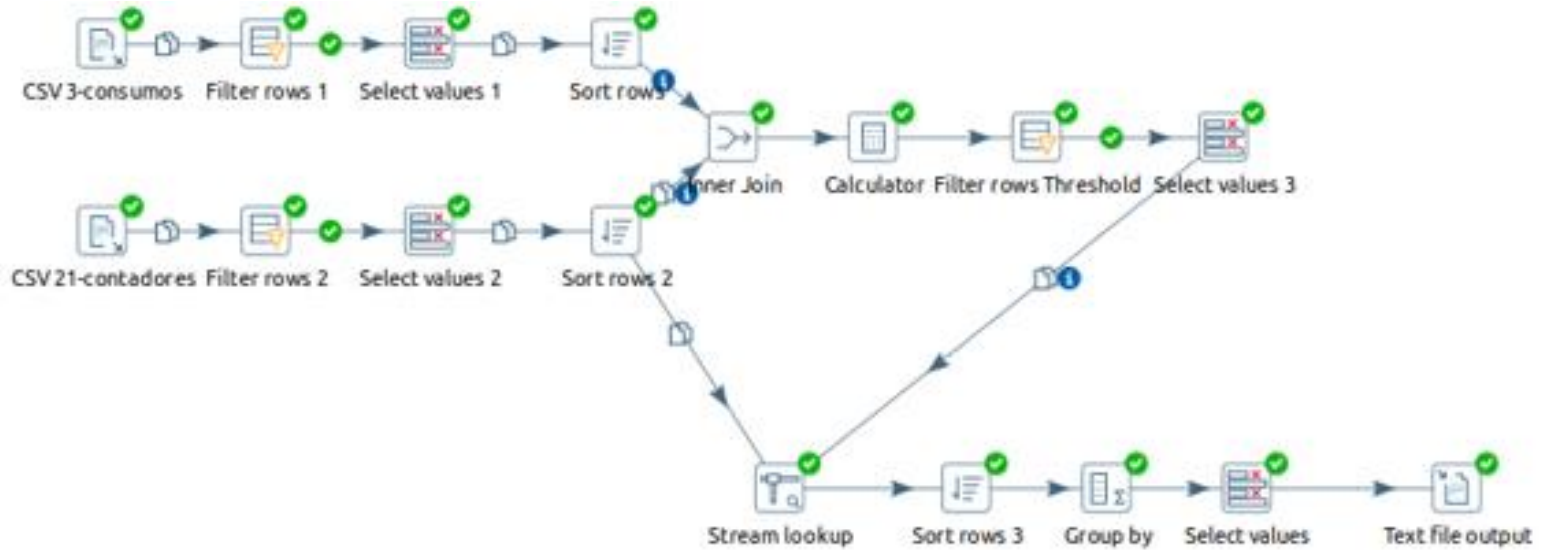
Question 2 - DataCleaner: Scatter Plot



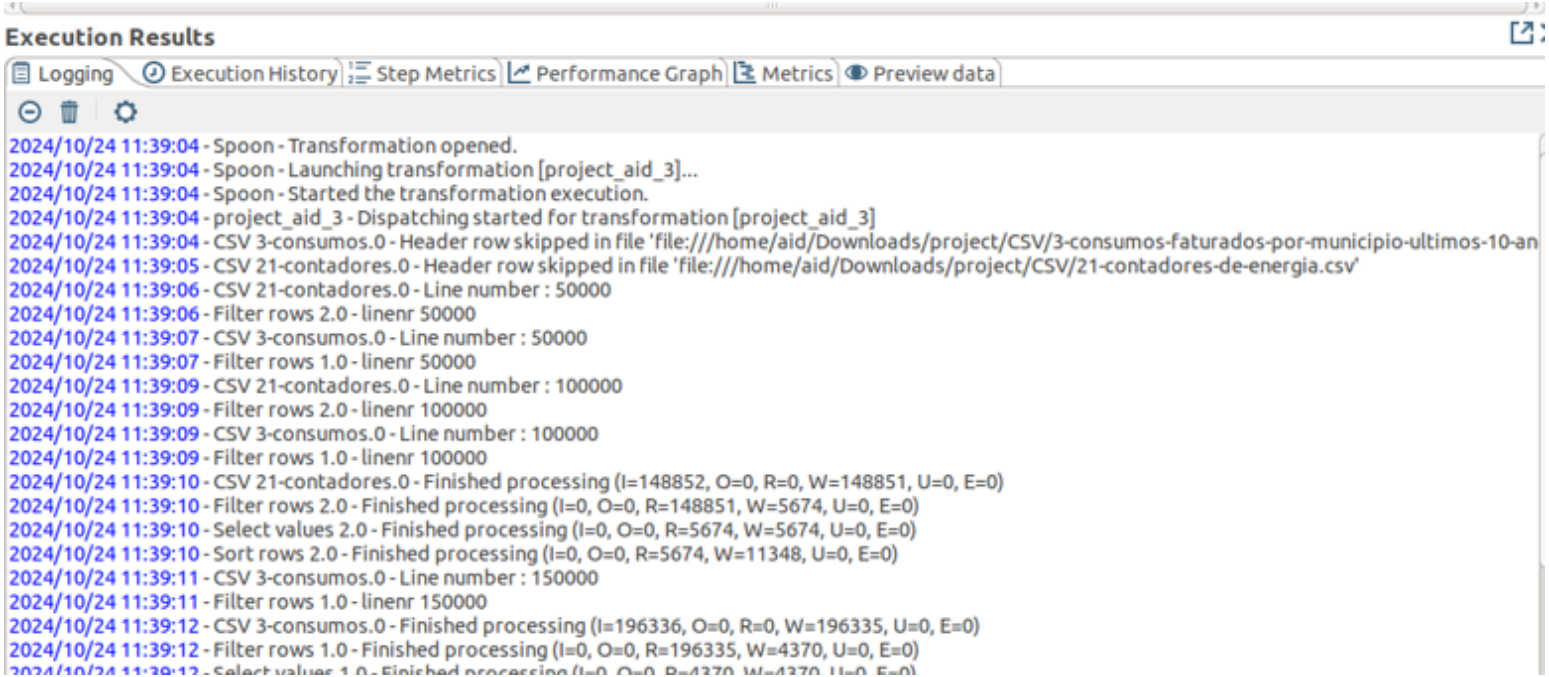
Pergunta 2 - DataCleaner, Scatter Plot

From observing the graph on the previous slide, it is not possible to deduce any correlation between the percentage of 'smart meters' and 'energy consumption'. In fact, when the percentage of 'smart meters' approaches 100%, the energy consumption falls within a very wide range of values, specifically between 50 and 350 kWh.

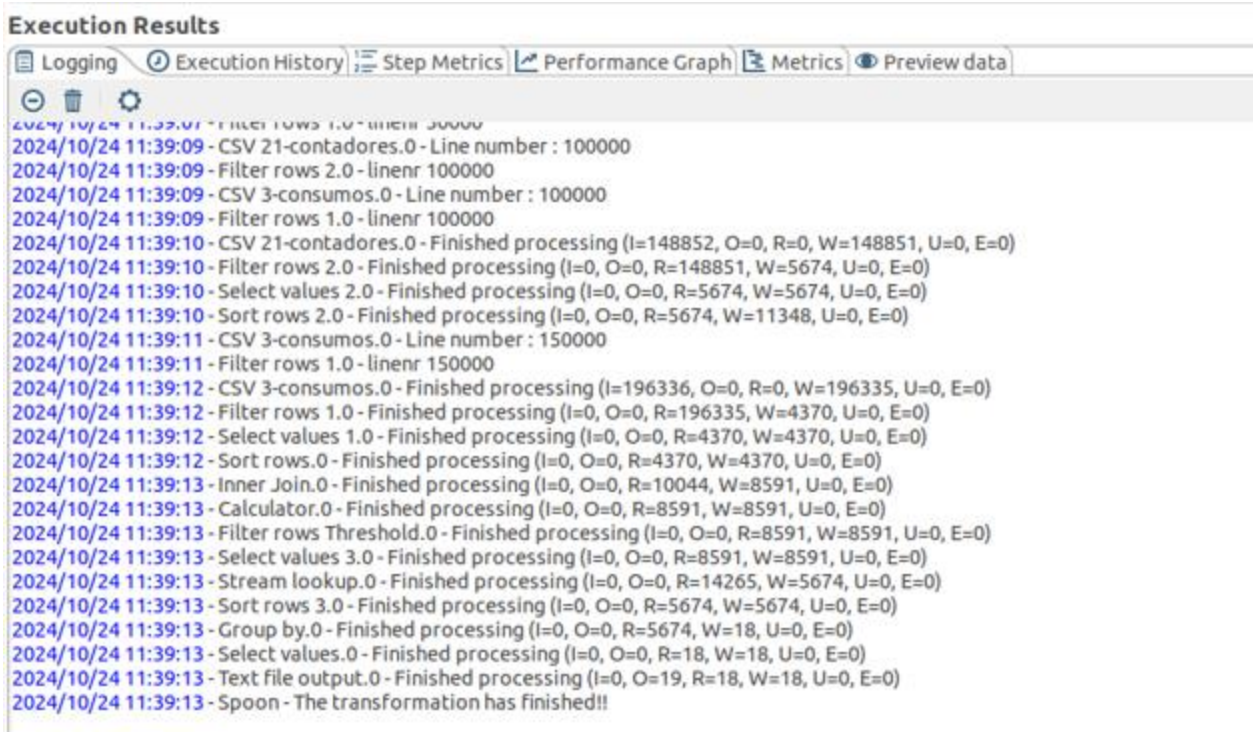
Question 3 - String Matching: Transformation



Question 3 - String Matching: Execution Results



Question 3 - String Matching: Execution Results



Question 3 - String Matching: CSV Consumos

Step name: CSV 3-consumos
Filename: ${Internal.Entry.Current.Directory}/3-consumos-faturados-por-munici
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	Grou
1	Year	Integer	#	15	0	$	.	,
2	Month	Integer	#	15	0	$	.	,
3	Date	String		7		$	.	,
4	District	String		16		$	.	,
5	Municipality	String		27		$	.	,
6	parish	String		30		$	.	,
7	Voltage level	String		32		$	.	,
8	Active Energy (kWh)	Number	#.#	12	3	$	.	,
9	DistrictCode	Integer	#	15	0	$	.	,
10	DistrictMunicipalityCode	Integer	#	15	0	$	.	,
11	DistrictMunicipalityParishCode	String		6		$	.	,
12	mes_int	Integer	#	15	0	$	.	,

Rows of step: CSV 3-consumos (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCc	mes_int
1	2021	5	2021-05	BEJA	Almodôvar	ALDEIA DOS FERNANDES	Baixa Tensão	70014.8	2	202	020208	5
2	2021	5	2021-05	EVORA	Borba	RIO DE MOINHOS	Muito Alta, Alta e Média Tensões	92461.7	7	703	070303	5
3	2021	5	2021-05	EVORA	Viana do Alentejo	AGUIAR	Baixa Tensão	139795.3	7	713	071303	5
4	2021	5	2021-05	LEIRIA	Leiria	UF PARCEIROS E AZOIA	Baixa Tensão	1728071.9	10	1009	100937	5
5	2021	5	2021-05	LISBOA	Sintra	UF CACEM E SAO MARCOS	Baixa Tensão	3863416.5	11	1111	111124	5
6	2021	5	2021-05	PORTO	Vila Nova de Gaia	MADALENA	Baixa Tensão	1505987.9	13	1317	131709	5
7	2021	5	2021-05	SANTAREM	Ourém	FATIMA	Muito Alta, Alta e Média Tensões	2775022	14	1421	142106	5
8	2021	5	2021-05	VISEU	Resende	UF OVADAS E PANCHORRA	Baixa Tensão	26305.7	18	1813	181319	5
9	2021	6	2021-06	BRAGA	Guimarães	BRITO	Muito Alta, Alta e Média Tensões	1677884.5	3	308	030807	6
10	2021	6	2021-06	COIMBRA	Cantanhede	OURENTA	Baixa Tensão	126207.7	6	602	060209	6

Question 3 - String Matching: CSV Consumos

Filter rows

Step name: Filter rows 1

Send 'true' data to step: Select values 1

Send 'false' data to step:

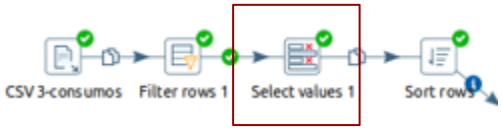
The condition:

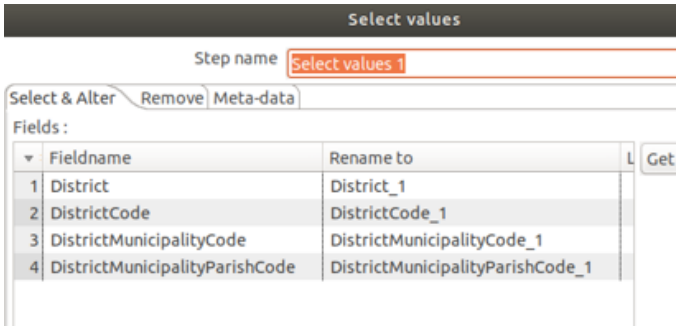
```
Year = [2024]

AND

Month = [6]
```

Rows of step: Filter rows 1 (1000 rows)

	Year	Montl	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCc	mes_int
1	2024	6	2024-06	COIMBRA	Condeixa-a-Nova	EGA	Muito Alta, Alta e Média Tensões	20559	6	604	060406	6
2	2024	6	2024-06	VIANA DO CASTELO	Caminha	UF VENADE E AZEVEDO	Baixa Tensão	67713	16	1602	160225	6
3	2024	6	2024-06	VIANA DO CASTELO	Monção	MOREIRA	Baixa Tensão	35454	16	1604	160418	6
4	2024	6	2024-06	VIANA DO CASTELO	Ponte de Lima	GEMIEIRA	Baixa Tensão	38425	16	1607	160727	6
5	2024	6	2024-06	VISEU	Moimenta da Beira	UF PERA VELHA ALD NACOMBA ARIZ	Baixa Tensão	26448	18	1807	180722	6
6	2024	6	2024-06	BRAGA	Barcelos	UF SILVEIROS E RIO COVO	Muito Alta, Alta e Média Tensões	36717	3	302	0302FE	6
7	2024	6	2024-06	BRAGA	Barcelos	VARZEA	Baixa Tensão	180725	3	302	030283	6
8	2024	6	2024-06	BRAGA	Vila Nova de Famalicão	GAVIAO	Baixa Tensão	218184	3	312	031216	6
9	2024	6	2024-06	BRAGANCA	Bragança	GONDESENDE	Baixa Tensão	7800	4	402	040217	6
10	2024	6	2024-06	CASTELO BRANCO	Covilhã	UF PESO E VALES DO RIO	Baixa Tensão	79682	5	503	050336	6

Question 3 - String Matching: CSV Consumos





Rows of step: Select values 1 (1000 rows)

	District_1	DistrictCode_1	DistrictMunicipalityCode_1	DistrictMunicipalityParishCode_1
1	COIMBRA	6	604	060406
2	VIANA DO CASTELO	16	1602	160225
3	VIANA DO CASTELO	16	1604	160418
4	VIANA DO CASTELO	16	1607	160727
5	VISEU	18	1807	180722
6	BRAGA	3	302	0302FE
7	BRAGA	3	302	030283
8	BRAGA	3	312	031216
9	BRAGANCA	4	402	040217
10	CASTELO BRANCO	5	503	050336

Question 3 - String Matching: CSV Consumos

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)	☐

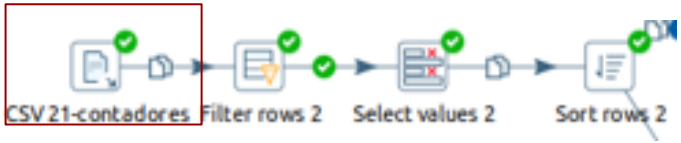
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Coll
1	DistrictCode_1	Y	N	N	0
2	DistrictMunicipalityCode_1	Y	N	N	0
3	DistrictMunicipalityParishCode_1	Y	N	N	0

Rows of step: Sort rows (1000 rows)

	District_1	DistrictCode_1	DistrictMunicipalityCode_1	DistrictMunicipalityParishCode_1
1	AVEIRO	1	101	010103
2	AVEIRO	1	101	010103
3	AVEIRO	1	101	010109
4	AVEIRO	1	101	010109
5	AVEIRO	1	101	010112
6	AVEIRO	1	101	010112
7	AVEIRO	1	101	010119
8	AVEIRO	1	101	010119
9	AVEIRO	1	101	010121
10	AVEIRO	1	101	010121

Question 3 - String Matching: CSV Consumos



Step name: CSV 21-contadores
Filename: ${Internal.Entry.Current.Directory}/21-contadores-d
Browse...
Delimiter: ;
Insert TAB
Enclosure: "
NIO buffer size: 50000
Lazy conversion?
Header row present?
Add filename to result
The row number field name (option
Running in parallel?
New line possible in fields?
Format: mixed
File encoding: UTF-8

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

Rows of step: CSV 21-contadores (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	SETUBAL	Sines	SINES	Não	128	15	1513	151301	Sim
2	2024	5	2024-05	PORTALEGRE	Ponte de Sor	MONTARGIL	Não	475	12	1213	121302	Sim
3	2024	5	2024-05	VISEU	Santa Comba Dão	UF S COMBA DAO COUTO MOSTEIRO	Sim	3141	18	1814	181411	Sim
4	2024	5	2024-05	CASTELO BRANCO	Belmonte	CARIA	Sim	1378	5	501	050102	Sim
5	2024	5	2024-05	CASTELO BRANCO	Castelo Branco	SAO VICENTE DA BEIRA	Sim	620	5	502	050222	Sim
6	2024	5	2024-05	FARO	Albufeira	GUIA	Não	1335	8	801	080102	Sim
7	2024	5	2024-05	SANTAREM	Ourém	UF MATAS E CERCAL	Sim	1056	14	1421	142121	Sim
8	2024	5	2024-05	VIANA DO CASTELO	Paredes de Coura	AGUALONGA	Sim	175	16	1605	160501	Sim
9	2024	5	2024-05	BEJA	Odemira	RELIQUIAS	Sim	334	2	211	021102	Sim
10	2024	5	2024-05	COIMBRA	Condeixa-a-Nova	ZAMBUJAL	Não	1	6	604	060410	Sim

Question 3 - String Matching: CSV Consumos

Filter rows

Step name: Filter rows 2

Send 'true' data to step: Select values 2

Send 'false' data to step:

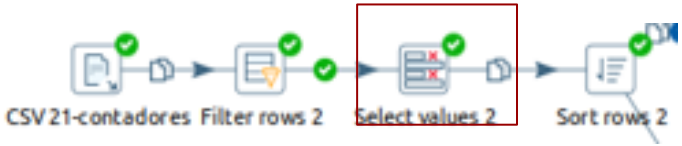
The condition:

```
Year = [2024]
AND
Month = [6]
```

Rows of step: Filter rows 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	PORTO	Marco de Canaveses	SOALHAES	Não	406	13	1307	130722	Sim
2	2024	6	2024-06	EVORA	Mora	MORA	Não	228	7	707	070703	Sim
3	2024	6	2024-06	PORTO	Paredes	SOBREIRA	Sim	1678	13	1310	131020	Sim
4	2024	6	2024-06	PORTO	Vila Nova de Gaia	SAO FELIX DA MARINHA	Não	1588	13	1317	131717	Sim
5	2024	6	2024-06	BRAGA	Amares	GOAES	Sim	315	3	301	030112	Sim
6	2024	6	2024-06	GUARDA	Guarda	CODESSEIRO	Sim	139	9	907	090714	Sim
7	2024	6	2024-06	SANTAREM	Mação	ENVENDOS	Sim	916	14	1413	141305	Sim
8	2024	6	2024-06	VILA REAL	Ribeira de Pena	SANTA MARINHA	Sim	419	17	1709	170906	Sim
9	2024	6	2024-06	BRAGA	Barcelos	CAMBESES	Não	32	3	302	030216	Sim
10	2024	6	2024-06	VILA REAL	Santa Marta de Penaguião	ALVACOES DO CORGO	Não	13	17	1711	171101	Sim

Question 3 - String Matching: CSV Consumos



Select values

Step name: Select values 2

Select & Alter | Remove | Meta-data

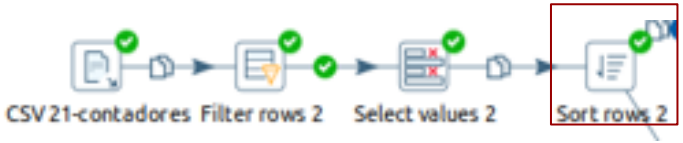
Fields :

	Fieldname	Rename to
1	District	District_2
2	DistrictCode	DistrictCode_2
3	DistrictMunicipalityCode	DistrictMunicipalityCode_2
4	DistrictMunicipalityParishCode	DistrictMunicipalityParishCode_2

Rows of step: Select values 2 (1000 rows)

	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2
1	PORTO	13	1307	130722
2	EVORA	7	707	070703
3	PORTO	13	1310	131020
4	PORTO	13	1317	131717
5	BRAGA	3	301	030112
6	GUARDA	9	907	090714
7	SANTAREM	14	1413	141305
8	VILA REAL	17	1709	170906
9	BRAGA	3	302	030216
10	VILA REAL	17	1711	171101

Question 3 - String Matching: CSV Consumos



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)

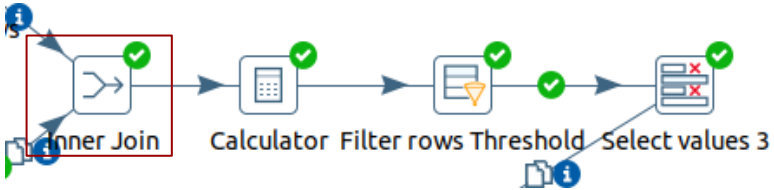
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator
1	DistrictCode_2	Y	N	N	0
2	DistrictMunicipalityCode_2	Y	N	N	0
3	DistrictMunicipalityParishCode_2	Y	N	N	0

Rows of step: Sort rows 2 (1000 rows)

	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2
1	AVEIRO	1	101	010103
2	AVEIRO	1	101	010103
3	AVEIRO	1	101	010109
4	AVEIRO	1	101	010109
5	AVEIRO	1	101	010112
6	AVEIRO	1	101	010112
7	AVEIRO	1	101	010119
8	AVEIRO	1	101	010119
9	AVEIRO	1	101	010121
10	AVEIRO	1	101	010121

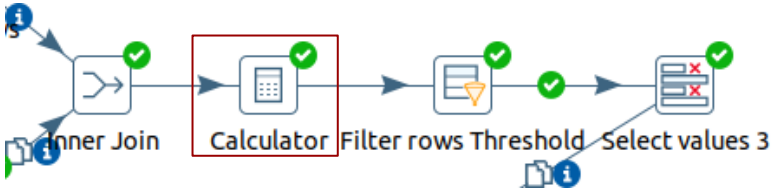
Question 3 - String Matching: Calculator and Threshold



Rows of step: Inner Join (1000 rows)

	District_1	DistrictCode_1	DistrictMunicipalityCode_1	DistrictMunicipalityParishCode_1	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2
1	AVEIRO	1	101	010103	AVEIRO	1	101	010103
2	AVEIRO	1	101	010103	AVEIRO	1	101	010103
3	AVEIRO	1	101	010103	AVEIRO	1	101	010103
4	AVEIRO	1	101	010103	AVEIRO	1	101	010103
5	AVEIRO	1	101	010109	AVEIRO	1	101	010109
6	AVEIRO	1	101	010109	AVEIRO	1	101	010109
7	AVEIRO	1	101	010109	AVEIRO	1	101	010109
8	AVEIRO	1	101	010109	AVEIRO	1	101	010109
9	AVEIRO	1	101	010112	AVEIRO	1	101	010112
10	AVEIRO	1	101	010112	AVEIRO	1	101	010112

Question 3 - String Matching: Calculator and Threshold



Step name

Calculator

Throw an error on non existing files

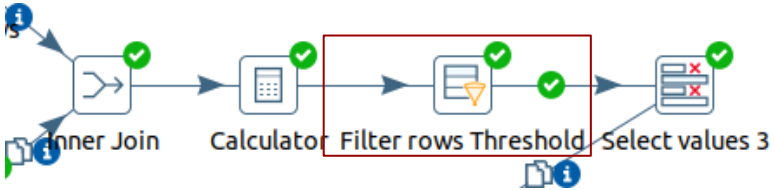
Fields:

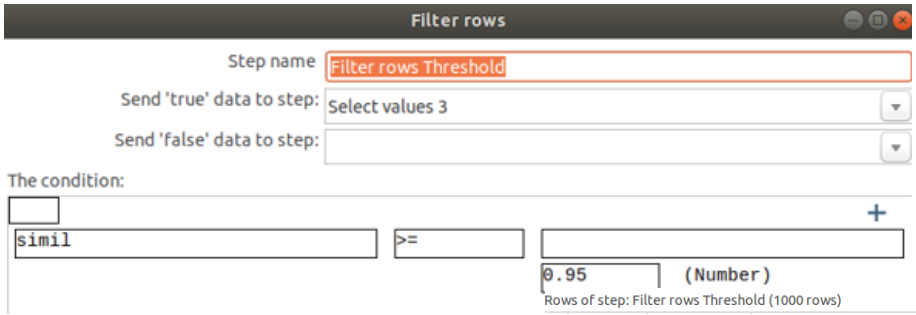
	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	simil	Jaro similitude between String A and String B	District_1	District_2		Number			N

Rows of step: Calculator (1000 rows)

#	District_1	DistrictCode_1	DistrictMunicipalityCode_1	DistrictMunicipalityParishCode_1	District_2	DistrictCode	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2	simil
1	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
2	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
3	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
4	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
5	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
6	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
7	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
8	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
9	AVEIRO	1	101	010112	AVEIRO	1	101	010112	1.0
10	AVEIRO	1	101	010112	AVEIRO	1	101	010112	1.0

Question 3 - String Matching: Calculator and Threshold

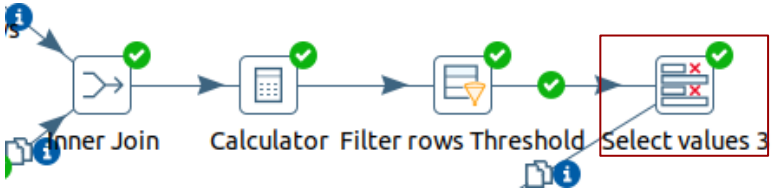




Rows of step: Filter rows Threshold (1000 rows)

	District_1	DistrictCode_1	DistrictMunicipalityCode_1	DistrictMunicipalityParishCode_1	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2	simil
1	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
2	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
3	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
4	AVEIRO	1	101	010103	AVEIRO	1	101	010103	1.0
5	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
6	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
7	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
8	AVEIRO	1	101	010109	AVEIRO	1	101	010109	1.0
9	AVEIRO	1	101	010112	AVEIRO	1	101	010112	1.0
10	AVEIRO	1	101	010112	AVEIRO	1	101	010112	1.0

Question 3 - String Matching: Calculator and Threshold



Select values

Step name: Select values 3

Select & Alter | Remove | Meta-data

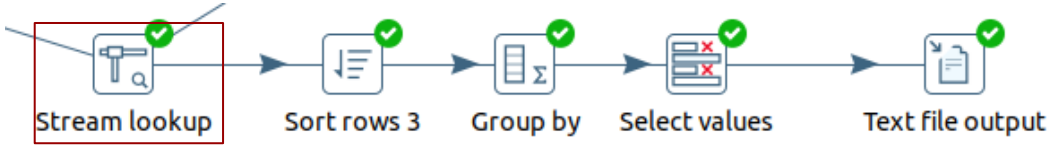
Fields :

	Fieldname	Rename to	Length	Precision	Ge
1	District_1	District_11			
2	District_2	District_22			

Rows of step: Select values 3 (1000 r

	District_11	District_22
1	AVEIRO	AVEIRO
2	AVEIRO	AVEIRO
3	AVEIRO	AVEIRO
4	AVEIRO	AVEIRO
5	AVEIRO	AVEIRO
6	AVEIRO	AVEIRO
7	AVEIRO	AVEIRO
8	AVEIRO	AVEIRO
9	AVEIRO	AVEIRO
10	AVEIRO	AVEIRO

Question 3 - String Matching: Lookup and Output



Stream lookup

Step name: Stream lookup

Lookup step: Select values 3

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

	Field	LookupField
1	District_2	District_22
2		

Specify the fields to retrieve :

	Field	New name	Default	Type
1	District_11	duplicate		None

Preserve memory (costs CPU)

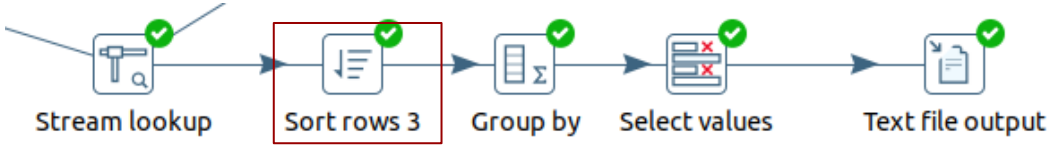
Key and value are exactly one integer field

Use sorted list (i.s.o. hashtable)

Rows of step: Stream lookup (1000 rows)

	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2	duplicate
1	AVEIRO	1	101	010103	AVEIRO
2	AVEIRO	1	101	010103	AVEIRO
3	AVEIRO	1	101	010109	AVEIRO
4	AVEIRO	1	101	010109	AVEIRO
5	AVEIRO	1	101	010112	AVEIRO
6	AVEIRO	1	101	010112	AVEIRO
7	AVEIRO	1	101	010119	AVEIRO
8	AVEIRO	1	101	010119	AVEIRO
9	AVEIRO	1	101	010121	AVEIRO
10	AVEIRO	1	101	010121	AVEIRO

Question 3 - String Matching: Lookup and Output



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 b
1	DistrictCode_2	Y	N	N
2	duplicate	Y	N	N

Rows of step: Sort rows 3 (1000 rows)

	District_2	DistrictCode_2	DistrictMunicipalityCode_2	DistrictMunicipalityParishCode_2	duplicate
1	AVEIRO	1	101	010103	AVEIRO
2	AVEIRO	1	101	010103	AVEIRO
3	AVEIRO	1	101	010109	AVEIRO
4	AVEIRO	1	101	010109	AVEIRO
5	AVEIRO	1	101	010112	AVEIRO
6	AVEIRO	1	101	010112	AVEIRO
7	AVEIRO	1	101	010119	AVEIRO
8	AVEIRO	1	101	010119	AVEIRO
9	AVEIRO	1	101	010121	AVEIRO
10	AVEIRO	1	101	010121	AVEIRO

Question 3 - String Matching: Lookup and Output

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	District_2
2	duplicate

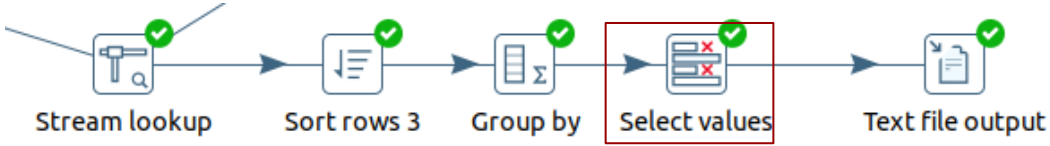
Aggregates :

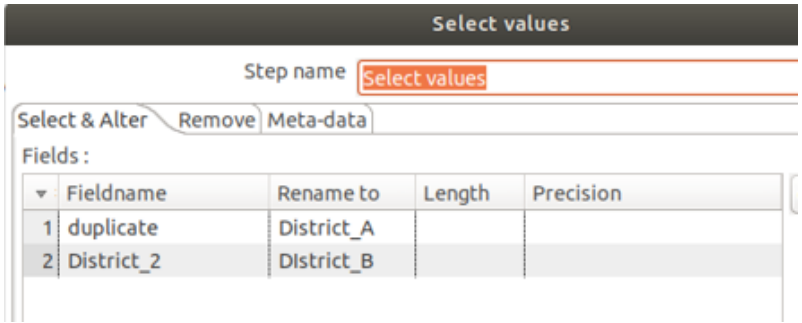
	Name	Subject	Type	Value
1				

Rows of step: Group by (18 rows)

	District_2	duplicate
1	AVEIRO	AVEIRO
2	BEJA	BEJA
3	BRAGA	BRAGA
4	BRAGANCA	BRAGANCA
5	CASTELO BRANCO	CASTELO BRANCO
6	COIMBRA	COIMBRA
7	EVORA	EVORA
8	FARO	FARO
9	GUARDA	GUARDA
10	LEIRIA	LEIRIA
11	LISBOA	LISBOA
12	PORTALEGRE	PORTALEGRE
13	PORTO	PORTO
14	SANTAREM	SANTAREM
15	SETUBAL	SETUBAL
16	VIANA DO CASTELO	VIANA DO CASTELO
17	VILA REAL	VILA REAL
18	VISEU	VISEU

Question 3 - String Matching: Lookup and Output

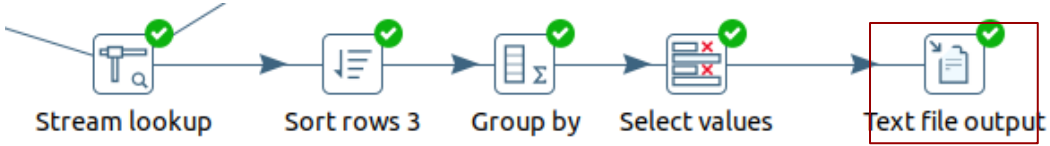




Rows of step: Select values (18 rows)

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

Question 3 - String Matching: Lookup and Output



Step name: Text file output

Filename: /home/aid/Downloads/project/3_output

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)...

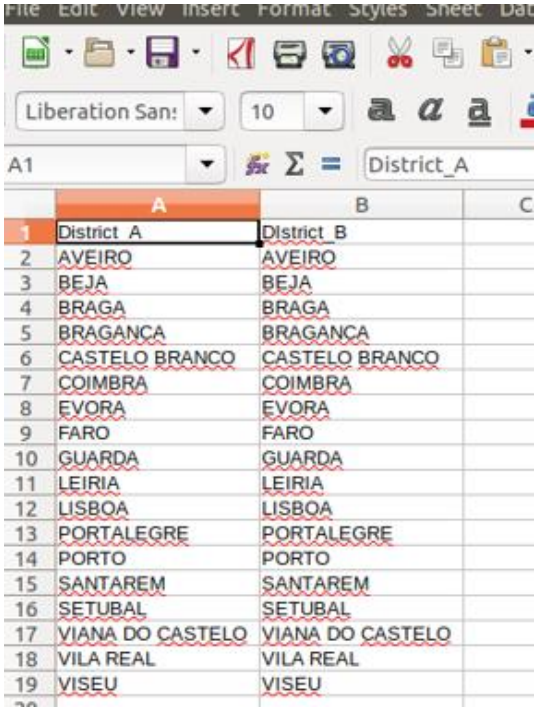
Add filenames to result

Step name: Text file output

File | Content | Fields

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

Question 3 - String Matching: File



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

Question 3 - String Matching: Final Observations

We based our approach on the Jaro proximity algorithm, achieving the maximum value of 1. We also applied the Levenshtein distance algorithm, obtaining a value of 0.

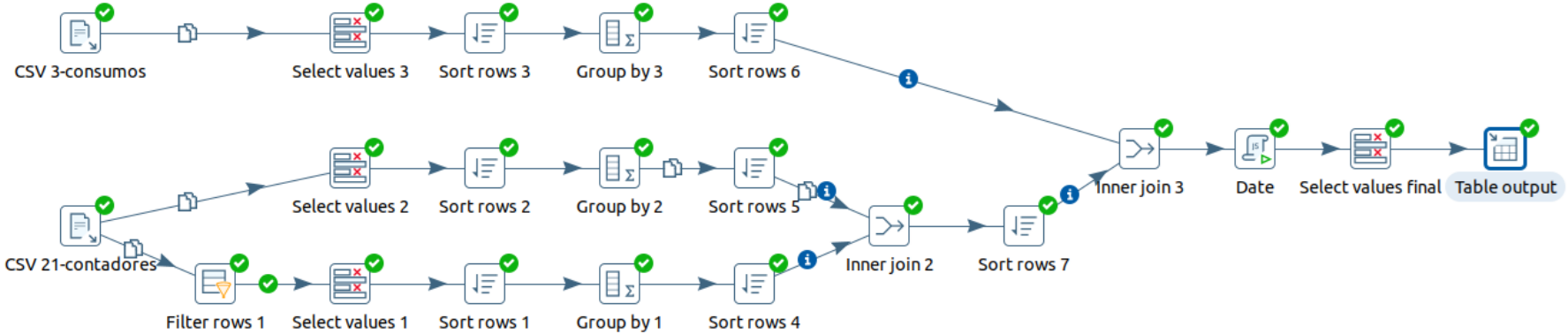
The threshold set for the Jaro algorithm was 0.95, with all values passing.

Question 4 - DataWareHouse: SQL Script

```sql
DROP DATABASE IF EXISTS datawarehouse_project;

CREATE DATABASE datawarehouse_project;

USE datawarehouse_project;

DROP TABLE IF EXISTS dim_time;
DROP TABLE IF EXISTS dim_location;
DROP TABLE IF EXISTS energy_consumption;

CREATE TABLE dim_time (
    time_id DATETIME,
    year_id INT,
    season_id INT,
    season VARCHAR(10),
    month_id INT,
    month VARCHAR(15),
    PRIMARY KEY (time_id)
);

```

```sql
CREATE TABLE dim_location (
    location_id INT,
    region_id INT,
    region VARCHAR(100),
    municipality_id INT,
    municipality VARCHAR(100),
    parish_id VARCHAR(20),
    parish VARCHAR(100),
    version INT,
    date_from DATETIME,
    date_to DATETIME,
    PRIMARY KEY (location_id)
);

CREATE TABLE energy_smart (
    time_id DATETIME,
    location_id INT,
    yes_smartmeters INT,
    total_smartmeters INT,
    active_energy_kwh DECIMAL(20,3),
    PRIMARY KEY (time_id, location_id),
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (location_id) REFERENCES dim_location(location_id)
);
```

Question 4 - Extract, Transform and Load into DB



Question 4 - ETL: Execution Results

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/24 09:55:58 - Spoon - Transformation opened.
2024/10/24 09:55:58 - Spoon - Launching transformation [4_project_aid_create_table]...
2024/10/24 09:55:58 - Spoon - Started the transformation execution.
2024/10/24 09:55:58 - 4_project_aid_create_table - Dispatching started for transformation [4_project_aid_create_table]
2024/10/24 09:55:58 - Table output.0 - Connected to database [input5] (commit=1000)
2024/10/24 09:55:58 - CSV 3-consumos.0 - Header row skipped in file '/home/aid/Downloads/Project_AID/3-consumos-faturados-por-municipio-ultimos-10-anos.csv'
2024/10/24 09:55:58 - CSV 21-contadores.0 - Header row skipped in file '/home/aid/Downloads/Project_AID/21-contadores-de-energia.csv'
2024/10/24 09:56:01 - CSV 3-consumos.0 - Line number : 50000
2024/10/24 09:56:01 - Select values 3.0 - linenr 50000
2024/10/24 09:56:01 - CSV 21-contadores.0 - Line number : 50000
2024/10/24 09:56:01 - Sort rows 3.0 - Linenr 50000
2024/10/24 09:56:01 - Filter rows 1.0 - linenr 50000
2024/10/24 09:56:01 - Select values 2.0 - linenr 50000
2024/10/24 09:56:01 - Sort rows 2.0 - Linenr 50000
2024/10/24 09:56:03 - CSV 21-contadores.0 - Line number : 100000
2024/10/24 09:56:03 - Filter rows 1.0 - linenr 100000
2024/10/24 09:56:03 - Select values 1.0 - linenr 50000
2024/10/24 09:56:03 - Select values 2.0 - linenr 100000
2024/10/24 09:56:03 - Sort rows 2.0 - Linenr 100000
2024/10/24 09:56:03 - Sort rows 1.0 - Linenr 50000
2024/10/24 09:56:04 - CSV 3-consumos.0 - Line number : 100000
2024/10/24 09:56:04 - Select values 3.0 - linenr 100000
2024/10/24 09:56:04 - Sort rows 3.0 - Linenr 100000
2024/10/24 09:56:05 - CSV 21-contadores.0 - Finished processing (I=148852, O=0, R=0, W=297702, U=0, E=0)
2024/10/24 09:56:05 - Filter rows 1.0 - Finished processing (I=0, O=0, R=148851, W=74811, U=0, E=0)
2024/10/24 09:56:05 - Select values 1.0 - Finished processing (I=0, O=0, R=74811, W=74811, U=0, E=0)
2024/10/24 09:56:05 - Select values 2.0 - Finished processing (I=0, O=0, R=148851, W=148851, U=0, E=0)
2024/10/24 09:56:06 - CSV 3-consumos.0 - Line number : 150000
2024/10/24 09:56:06 - Select values 3.0 - linenr 150000
2024/10/24 09:56:06 - Sort rows 3.0 - Linenr 150000
2024/10/24 09:56:07 - CSV 3-consumos.0 - Finished processing (I=196336, O=0, R=0, W=196335, U=0, E=0)
2024/10/24 09:56:07 - Select values 3.0 - Finished processing (I=0, O=0, R=196335, W=196335, U=0, E=0)
```

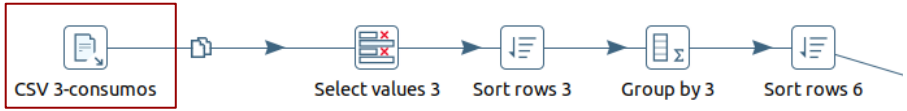
Question 4 - ETL: Execution Results

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/24 09:56:09 - Group by 1.0 - Linenr 50000
2024/10/24 09:56:09 - Sort rows 4.0 - Linenr 50000
2024/10/24 09:56:09 - Group by 3.0 - Linenr 50000
2024/10/24 09:56:09 - Sort rows 1.0 - Finished processing (I=0, O=0, R=74811, W=74811, U=0, E=0)
2024/10/24 09:56:09 - Sort rows 6.0 - Linenr 50000
2024/10/24 09:56:09 - Group by 1.0 - Finished processing (I=0, O=0, R=74811, W=74811, U=0, E=0)
2024/10/24 09:56:09 - Group by 2.0 - Linenr 50000
2024/10/24 09:56:09 - Group by 3.0 - Linenr 100000
2024/10/24 09:56:10 - Group by 2.0 - Linenr 100000
2024/10/24 09:56:10 - Sort rows 5.0 - Linenr 50000
2024/10/24 09:56:10 - Group by 3.0 - Linenr 150000
2024/10/24 09:56:10 - Sort rows 6.0 - Linenr 100000
2024/10/24 09:56:10 - Sort rows 2.0 - Finished processing (I=0, O=0, R=148851, W=148851, U=0, E=0)
2024/10/24 09:56:10 - Group by 2.0 - Finished processing (I=0, O=0, R=148851, W=74835, U=0, E=0)
2024/10/24 09:56:10 - Sort rows 3.0 - Finished processing (I=0, O=0, R=196335, W=196335, U=0, E=0)
2024/10/24 09:56:10 - Group by 3.0 - Finished processing (I=0, O=0, R=196335, W=127424, U=0, E=0)
2024/10/24 09:56:10 - Inner join 2.0 - linenr50000
2024/10/24 09:56:11 - Inner join 2.0 - linenr100000
2024/10/24 09:56:11 - Sort rows 7.0 - Linenr 50000
2024/10/24 09:56:11 - Sort rows 5.0 - Finished processing (I=0, O=0, R=74835, W=74835, U=0, E=0)
2024/10/24 09:56:11 - Sort rows 4.0 - Finished processing (I=0, O=0, R=74811, W=74811, U=0, E=0)
2024/10/24 09:56:11 - Inner join 2.0 - Finished processing (I=0, O=0, R=149646, W=74811, U=0, E=0)
2024/10/24 09:56:11 - Date.0 - Optimization level set to 9.
2024/10/24 09:56:12 - Inner join 3.0 - linenr50000
2024/10/24 09:56:18 - Inner join 3.0 - linenr100000
2024/10/24 09:56:39 - Inner join 3.0 - linenr150000
2024/10/24 09:56:46 - Date.0 - linenr 50000
2024/10/24 09:56:49 - Sort rows 7.0 - Finished processing (I=0, O=0, R=74811, W=74811, U=0, E=0)
2024/10/24 09:56:53 - Sort rows 6.0 - Finished processing (I=0, O=0, R=127424, W=127424, U=0, E=0)
2024/10/24 09:56:57 - Select values final.0 - linenr 50000
2024/10/24 09:56:59 - Inner join 3.0 - linenr200000
2024/10/24 09:57:00 - Inner join 3.0 - Finished processing (I=0, O=0, R=202235, W=71932, U=0, E=0)
2024/10/24 09:57:09 - Table output.0 - linenr 50000
2024/10/24 09:57:10 - Date.0 - Finished processing (I=0, O=0, R=71932, W=71932, U=0, E=0)
2024/10/24 09:57:20 - Select values final.0 - Finished processing (I=0, O=0, R=71932, W=71932, U=0, E=0)
2024/10/24 09:57:32 - Table output.0 - Finished processing (I=0, O=71932, R=71932, W=71932, U=0, E=0)
2024/10/24 09:57:32 - Spoon - The transformation has finished!!
```

Question 4 - ETL: Initial Operations, CSV Consumos



CSV file input

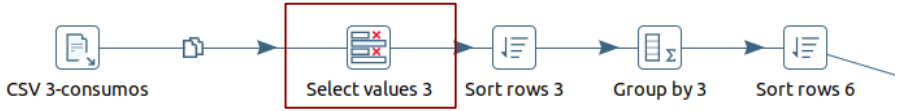
Step name: CSV 3-consumos
Filename: ${Internal.Entry.Current.Directory}/CSV/3-consumos-faturados-por
Browse...
Delimiter: ;
Insert TAB
Enclosure: "
NIO buffer size: 50000
Lazy conversion?
Header row present?
Add filename to result
The row number field name (optional)
Running in parallel?
New line possible in fields?
Format: mixed
File encoding: UTF-8

	Name	Type	Format	Length	Precision	Currency	Decima
1	Year	Integer	#	15	0	$	.
2	Month	Integer	#	15	0	$	.
3	Date	String		7		$	.

Rows of step: CSV 3-consumos (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level
1	2021	5	2021-05	BEJA	Almodôvar	ALDEIA DOS FERNANDES	Baixa Tensão
2	2021	5	2021-05	EVORA	Borba	RIO DE MOINHOS	Muito Alta, Alta e Média Tensões
3	2021	5	2021-05	EVORA	Viana do Alentejo	AGUIAR	Baixa Tensão
4	2021	5	2021-05	LEIRIA	Leiria	UF PARCEIROS E AZOIA	Baixa Tensão
5	2021	5	2021-05	LISBOA	Sintra	UF CACEM E SAO MARCOS	Baixa Tensão
6	2021	5	2021-05	PORTO	Vila Nova de Gaia	MADALENA	Baixa Tensão
7	2021	5	2021-05	SANTAREM	Ourém	FATIMA	Muito Alta, Alta e Média Tensões
8	2021	5	2021-05	VISEU	Resende	UF OVADAS E PANCHORRA	Baixa Tensão
9	2021	6	2021-06	BRAGA	Guimarães	BRITO	Muito Alta, Alta e Média Tensões
10	2021	6	2021-06	COIMBRA	Cantanhede	OURENTA	Baixa Tensão

Question 4 - ETL: Initial Operations, CSV Consumos



Step name: Select values 3

Select & Alter | Remove | Meta-data

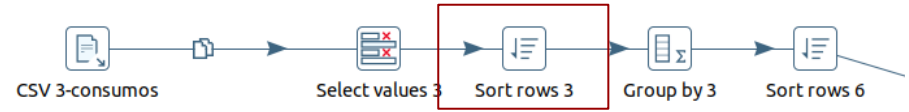
Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_3		
2	District	District_3		
3	DistrictMunicipalityCode	MunicipalityCode_3		
4	Municipality	Municipality_3		
5	DistrictMunicipalityParishCode	ParishCode_3		
6	parish	Parish_3		
7	Date	Date_3		
8	Active Energy (kWh)			

Rows of step: Select values 3 (1000 rows)

	DistrictCode_3	District_3	MunicipalityCode_3	Municipality_3	ParishCode_3	Parish_3	Date_3	Active Energy (kWh)
1	2	BEJA	202	Almodôvar	020208	ALDEIA DOS FERNANDES	2021-05	70014.8
2	7	EVORA	703	Borba	070303	RIO DE MOINHOS	2021-05	92461.7
3	7	EVORA	713	Viana do Alentejo	071303	AGUIAR	2021-05	139795.3
4	10	LEIRIA	1009	Leiria	100937	UF PARCEIROS E AZOIA	2021-05	1728071.9
5	11	LISBOA	1111	Sintra	111124	UF CACEM E SAO MARCOS	2021-05	3863416.5
6	13	PORTO	1317	Vila Nova de Gaia	131709	MADALENA	2021-05	1505987.9
7	14	SANTAREM	1421	Ourém	142106	FATIMA	2021-05	2775022
8	18	VISEU	1813	Resende	181319	UF OVADAS E PANCHORRA	2021-05	26305.7
9	3	BRAGA	308	Guimarães	030807	BRITO	2021-06	1677884.5
10	6	COIMBRA	602	Cantanhede	060209	OURENTA	2021-06	126207.7
11	13	PORTO	1304	Gondomar	130413	UF FANZERES SAO PEDRO COVA	2021-06	4472527.5

Question 4 - ETL: Initial Operations, CSV Consumos



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)

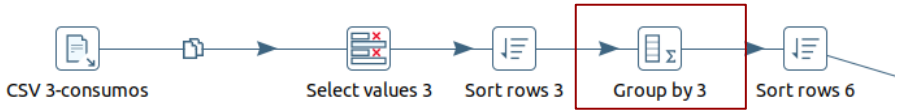
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort ba
1	DistrictCode_3	Y	N	N
2	MmunicipalityCode_3	Y	N	N
3	ParishCode_3	Y	N	N
4	Date_3	Y	N	N

Rows of step: Sort rows 3 (1000 rows)

	DistrictCode_3	District_3	MunicipalityCode_3	Municipality_3	ParishCode_3	Parish_3	Date_3	Active Energy (kWh)
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-11	718225.3
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-11	2127099.9
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-12	832788.7
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-12	1785499.8
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-01	2157978.4
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-01	858867.6
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-02	712461
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-02	2083921.7
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-03	690481
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-03	2353177
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-04	2166786.9

Question 4 - ETL: Initial Operations, CSV Consumos



Step name: Group by 3
Include all rows?
Temporary files directory: %%java.io.tmpdir%%
TMP-file prefix: grp
Add line number, restart in each group
Line number field name:
Always give back a result row

The fields that make up the group:

	Group field
1	DistrictCode_3
2	District_3
3	MunicipalityCode_3
4	Municipality_3
5	ParishCode_3
6	Parish_3

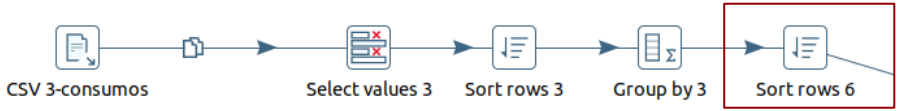
Aggregates :

	Name	Subject	Type
1	total_energy_parish	Active Energy (kWh)	Sum

Rows of step: Group by 3 (1000 rows)

	DistrictCode_3	District_3	MunicipalityCode_3	Municipality_3	ParishCode_3	Parish_3	Date_3	total_energy_parish
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-11	2845325.2770000002
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-12	2618288.47
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-01	3016845.9579999996
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-02	2796382.6890000002
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-03	3043657.93
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-04	2817581.5599999996
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-05	2890221.666
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-06	2823296.517
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-07	2824977.717
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-08	2217161.425
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-09	2633307.253

Question 4 - ETL: Initial Operations, CSV Consumos



Step name: Sort rows 6
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)

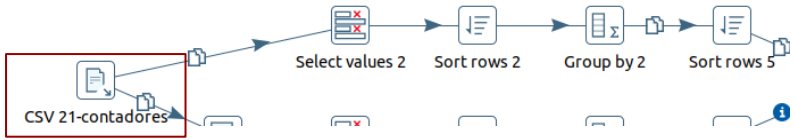
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort
1	DistrictCode_3	Y	N	N
2	MunicipalityCode_3	Y	N	N
3	ParishCode_3	Y	N	N
4	Date_3	Y	N	N

Rows of step: Sort rows 6 (1000 rows)

	DistrictCode_3	District_3	MunicipalityCode_3	Municipality_3	ParishCode_3	Parish_3	Date_3	total_energy_parish
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-11	2845325.2770000002
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2020-12	2618288.47
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-01	3016845.9579999996
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-02	2796382.6890000002
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-03	3043657.93
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-04	2817581.5599999996
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-05	2890221.666
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-06	2823296.517
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-07	2824977.717
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-08	2217161.425
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2021-09	2633307.253

Question 4 - ETL: Initial Operations, CSV Contadores



CSV file input

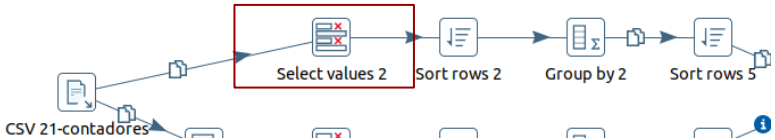
Step name: CSV 21-contadores
Filename: ${Internal.Entry.Current.Directory}/CSV/21-contadores-de-energia. (Browse...)
Delimiter: ; (Insert TAB)
Enclosure: "
NIO buffer size: 50000
Lazy conversion?
Header row present?
Add filename to result
The row number field name (optional)
Running in parallel?
New line possible in fields?
Format: mixed
File encoding: UTF-8

	Name	Type	Format	Length	Precision	Currency	Decim
1	Year	Integer	#	15	0	$	.
2	Month	Integer	#	15	0	$	.
3	Date	String		7		$	.

Rows of step: CSV 21-contadores (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Nu
1	2024	5	2024-05	SETUBAL	Sines	SINES	Não	
2	2024	5	2024-05	PORTALEGRE	Ponte de Sor	MONTARGIL	Não	
3	2024	5	2024-05	VISEU	Santa Comba Dão	UF S COMBA DAO COUTO MOSTEIRO	Sim	
4	2024	5	2024-05	CASTELO BRANCO	Belmonte	CARIA	Sim	
5	2024	5	2024-05	CASTELO BRANCO	Castelo Branco	SAO VICENTE DA BEIRA	Sim	
6	2024	5	2024-05	FARO	Albufeira	GUIA	Não	

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Select values 2

Select & Alter | Remove | Meta-data

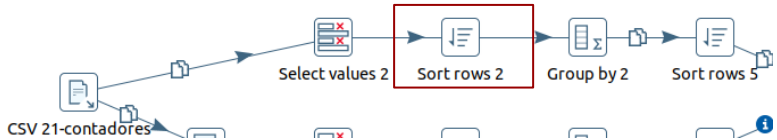
Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_2		
2	District	District_2		
3	DistrictMunicipalityCode	MunicipalityCode_2		
4	Municipality	Municipality_2		
5	DistrictMunicipalityParishCode	ParishCode_2		
6	Parish	Parish_2		
7	Date	Date_2		
8	Number of CPE's			

Rows of step: Select values 2 (1000 rows)

	DistrictCode_2	District_2	MunicipalityCode_2	Municipality_2	ParishCode_2	Parish_2	Date_2	Number of CPE's
1	15	SETUBAL	1513	Sines	151301	SINES	2024-05	128
2	12	PORTALEGRE	1213	Ponte de Sor	121302	MONTARGIL	2024-05	475
3	18	VISEU	1814	Santa Comba Dão	181411	UF S COMBA DAO COUTO MOSTEIRO	2024-05	3141
4	5	CASTELO BRANCO	501	Belmonte	050102	CARIA	2024-05	1378
5	5	CASTELO BRANCO	502	Castelo Branco	050222	SAO VICENTE DA BEIRA	2024-05	620
6	8	FARO	801	Albufeira	080102	GUIA	2024-05	1335
7	14	SANTAREM	1421	Ourém	142121	UF MATAS E CERCAL	2024-05	1056
8	16	VIANA DO CASTELO	1605	Paredes de Coura	160501	AGUALONGA	2024-05	175
9	2	BEJA	211	Odemira	021102	RELIQUIAS	2024-05	334
10	6	COIMBRA	604	Condeixa-a-Nova	060410	ZAMBUJAL	2024-05	1

Question 4 - ETL: Initial Operations, CSV Contadores



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)

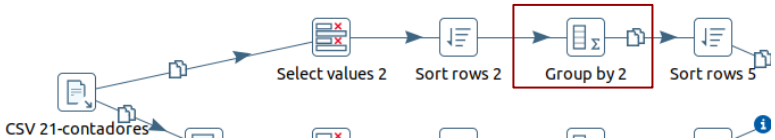
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based or
1	DistrictCode_2	Y	N	N
2	MunicipalityCode_2	Y	N	N
3	ParishCode_2	Y	N	N
4	Date_2	Y	N	N

Rows of step: Sort rows 2 (1000 rows)

#	DistrictCode_2	District_	Munic	Municipality_2	ParishCode_2	Parish_2	Date_2	Number of CPE's
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	383
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	361
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	354
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	335
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	282
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873

Question 4 - ETL: Initial Operations, CSV Contadores



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	DistrictCode_2
2	District_2
3	MunicipalityCode_2
4	Municipality_2
5	ParishCode_2
6	Parish_2

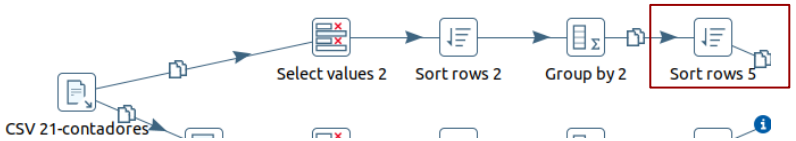
Aggregates :

	Name	Subject	Type	Value
1	total_cpes	Number of CPE's	Sum	

Rows of step: Group by 2 (1000 rows)

#	DistrictCode_2	District_2	MunicipalityCode_2	Municipality_2	ParishCode_2	Parish_2	Date_2	total_cpes
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	2153
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	2156
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	2153
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	2155
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	2155
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	2156
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	2157
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	2156
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	2155
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	2157
12	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	2158

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Sort rows 5

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) ☐

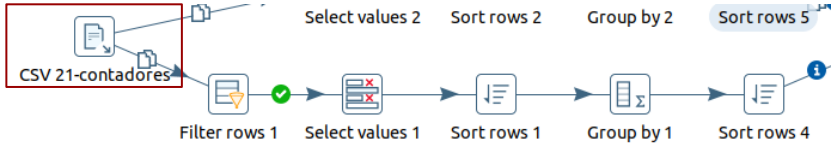
Fields :

	Fieldname	Ascending	Case sensitive compare?
1	DistrictCode_2	Y	N
2	MunicipalityCode_2	Y	N
3	ParishCode_2	Y	N
4	Date_2	Y	N

Rows of step: Sort rows 5 (1000 rows)

#	DistrictCode_2	District_2	MunicipalityCode_2	Municipality_2	ParishCode_2	Parish_2	Date_2	total_cpes
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	2153
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	2156
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	2153
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	2155
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	2155
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	2156
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	2157
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	2156
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	2155
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	2157
12	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	2158

Question 4 - ETL: Initial Operations, CSV Contadores



Filter rows

Step name: Filter rows 1

Send 'true' data to step: Select values 1

Send 'false' data to step:

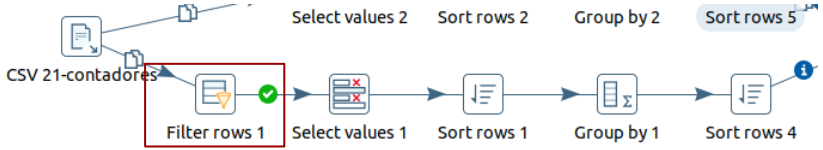
The condition:

```
Includes Smart Meter = [Sim]
```

Rows of step: Filter rows 1 (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Numb
1	2024	5	2024-05	VISEU	Santa Comba Dão	UF S COMBA DAO COUTO MOSTEIRO	Sim	
2	2024	5	2024-05	CASTELO BRANCO	Belmonte	CARIA	Sim	
3	2024	5	2024-05	CASTELO BRANCO	Castelo Branco	SAO VICENTE DA BEIRA	Sim	
4	2024	5	2024-05	SANTAREM	Ourém	UF MATAS E CERCAL	Sim	
5	2024	5	2024-05	VIANA DO CASTELO	Paredes de Coura	AGUALONGA	Sim	
6	2024	5	2024-05	BEJA	Odemira	RELIQUIAS	Sim	
7	2024	5	2024-05	CASTELO BRANCO	Proença-a-Nova	UF SOBREIRA FORMOSA ALVT BEIRA	Sim	

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Filter rows 1

Send 'true' data to step: Select values 1

Send 'false' data to step:

The condition:

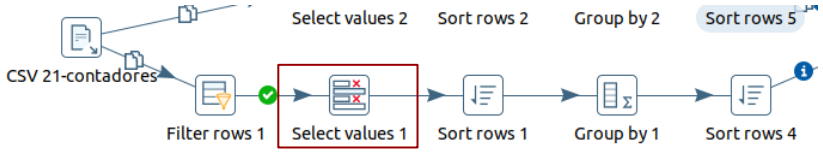
```
Includes Smart Meter = [Sim]
```

Help

Rows of step: Filter rows 1 (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE'	DistrictCode	DistrictMunicipalityCode
1	2024	5	2024-05	VISEU	Santa Comba Dão	UF S COMBA DAO COUTO MOSTEIRO	Sim	3141	18	1814
2	2024	5	2024-05	CASTELO BRANCO	Belmonte	CARIA	Sim	1378	5	501
3	2024	5	2024-05	CASTELO BRANCO	Castelo Branco	SAO VICENTE DA BEIRA	Sim	620	5	502
4	2024	5	2024-05	SANTAREM	Ourém	UF MATAS E CERCAL	Sim	1056	14	1421
5	2024	5	2024-05	VIANA DO CASTELO	Paredes de Coura	AGUALONGA	Sim	175	16	1605
6	2024	5	2024-05	BEJA	Odemira	RELIQUIAS	Sim	334	2	211
7	2024	5	2024-05	CASTELO BRANCO	Proença-a-Nova	UF SOBREIRA FORMOSA ALVT BEIRA	Sim	1525	5	508

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Select values 1

Select & Alter | Remove | Meta-data

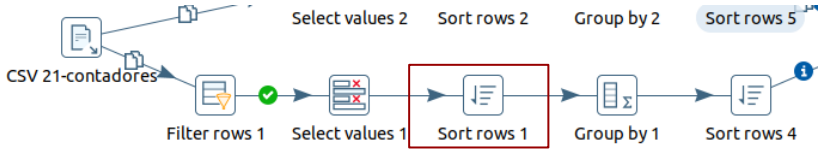
Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode	DistrictCode_1		
2	District	District_1		
3	DistrictMunicipalityCode	MunicipalityCode_1		
4	Municipality	Municipality_1		
5	DistrictMunicipalityParishCode	ParishCode_1		
6	Parish	Parish_1		
7	Date	Date_1		
8	Number of CPE's			

Rows of step: Select values 1 (1000 rows)

	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	Number of CPE's
1	18	VISEU	1814	Santa Comba Dão	181411	UF S COMBA DAO COUTO MOSTEIRO	2024-05	3141
2	5	CASTELO BRANCO	501	Belmonte	050102	CARIA	2024-05	1378
3	5	CASTELO BRANCO	502	Castelo Branco	050222	SAO VICENTE DA BEIRA	2024-05	620
4	14	SANTAREM	1421	Ourém	142121	UF MATAS E CERCAL	2024-05	1056
5	16	VIANA DO CASTELO	1605	Paredes de Coura	160501	AGUALONGA	2024-05	175
6	2	BEJA	211	Odemira	021102	RELIQUIAS	2024-05	334
7	5	CASTELO BRANCO	508	Proença-a-Nova	050808	UF SOBREIRA FORMOSA ALVT BEIRA	2024-05	1525
8	3	BRAGA	311	Vieira do Minho	031105	CANTELAES	2024-05	348
9	6	COIMBRA	601	Arganil	060121	UF COJA E BARRIL DE ALVA	2024-05	949
10	18	VISEU	1817	Sátão	181706	MIOMA	2024-05	830

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Sort rows 1
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)

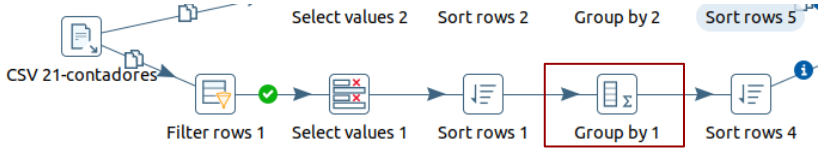
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on curre
1	DistrictCode_1	Y	N	N
2	MunicipalityCode_1	Y	N	N
3	ParishCode_1	Y	N	N
4	Date_1	Y	N	N

Rows of step: Sort rows 1 (1000 rows)

#	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	Number of CPE's
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946
12	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	1953
13	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-05	1972
14	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-06	1983

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Group by 1

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	DistrictCode_1
2	District_1
3	MunicipalityCode_1
4	Municipality_1
5	ParishCode_1
6	Parish_1

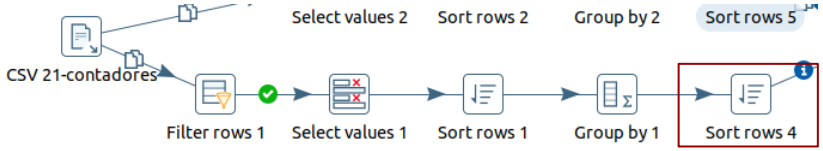
Aggregates :

	Name	Subject	Type
1	sum_CPE_sm_SIM	Number of CPE's	Sum

Rows of step: Group by 1 (1000 rows)

#	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946

Question 4 - ETL: Initial Operations, CSV Contadores



Step name: Sort rows 4
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)

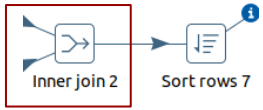
Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort base
1	DistrictCode_1	Y	N	N
2	MunicipalityCode_1	Y	N	N
3	ParishCode_1	Y	N	N
4	Date_1	Y	N	N

Rows of step: Sort rows 4 (1000 rows)

#	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946
12	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	1953

Question 4 - ETL: Joins



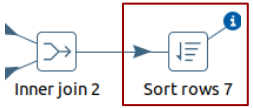
Multiway merge join

Step name	Inner join 2			
Input Step1	Sort rows 5	Join Keys	DistrictCode_2,Municip	Select Keys
Input Step2	Sort rows 4	Join Keys	DistrictCode_1,Municip	Select Keys
Join Type:	INNER			

Help | OK | Cancel

Date_2	total_cpes	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM
2022-06	2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
2022-07	2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
2022-08	2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
2022-09	2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
2022-10	2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873
2022-11	2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880
2022-12	2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912
2023-01	2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928
2023-02	2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934
2023-03	2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946
2023-04	2158	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	1953
2023-05	2159	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-05	1972
2023-06	2159	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-06	1983
2023-07	2158	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-07	1994

Question 4 - ETL: Joins



Step name: Sort rows 7

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?

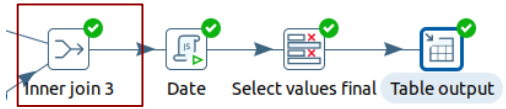
Only pass unique rows? (verifies keys only)

Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort b
1	DistrictCode_2	Y	N	N
2	MunicipalityCode_2	Y	N	N
3	ParishCode_2	Y	N	N
4	Date_2	Y	N	N

total_cpes	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM
325	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-06	81
325	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-11	81
325	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-12	81
326	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-07	82
326	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-08	82
326	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-09	82
326	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2022-10	82
327	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2023-01	84
327	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2023-02	86
327	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2023-03	88
327	1	AVEIRO	104	Arouca	010417	SAO MIGUEL DO M	2023-04	88

Question 4 - ETL: Joins



Multiway merge join

Step name: Inner join 3

Input Step1: Sort rows 6 — Join Keys: DistrictCode_3,Municipalit — Select Keys

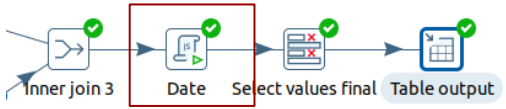
Input Step2: Sort rows 7 — Join Keys: DistrictCode_2,Municipalit — Select Keys

Join Type: INNER

Help OK

total_cpes	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM
2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795
2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880
2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934
2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946
2158	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-04	1953

Question 4 - ETL: Final Operations



Step name: Date

Java script functions:

- Transform Scripts
- Transform Constants
- Transform Functions
- Input fields
  - DistrictCode_3
  - District_3
  - MunicipalityCode_3
  - Municipality_3
  - ParishCode_3

Java script:

Script 1

```
//Script here

var data_string = Date_1;
var year_month = data_string.split("-");
var YearMonthDay = new Date(year_month[0], year_month[1] - 1, 1);
```

Linenr: 0

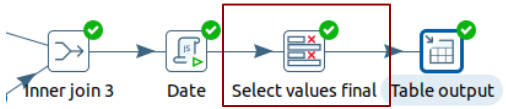
Compatibility mode? Opti

Fields

	Fieldname	Rename to	Type
1	YearMonthDay		Date

total_cpes	DistrictCode_1	District_1	MunicipalityCode_1	Municipality_1	ParishCode_1	Parish_1	Date_1	sum_CPE_sm_SIM	YearMonthDay
2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06	1770	2022/06/01 00:00:00.000
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07	1795	2022/07/01 00:00:00.000
2153	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08	1799	2022/08/01 00:00:00.000
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09	1820	2022/09/01 00:00:00.000
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10	1873	2022/10/01 00:00:00.000
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-11	1880	2022/11/01 00:00:00.000
2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-12	1912	2022/12/01 00:00:00.000
2156	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-01	1928	2023/01/01 00:00:00.000
2155	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-02	1934	2023/02/01 00:00:00.000
2157	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023-03	1946	2023/03/01 00:00:00.000

Question 4 - ETL: Final Operations



Step name: Select values final

Select & Alter | Remove | Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	DistrictCode_1	DistrictCode		
2	District_1	District		
3	MunicipalityCode_1	MunicipalityCode		
4	Municipality_1	Municipality		
5	ParishCode_1	ParishCode		
6	Parish_1	Parish		
7	YearMonthDay			
8	sum_CPE_sm_SIM	Yes_Smartmeters		
9	total_cpes	Total_Smartmeters		
10	total_energy_parish	EnergyConsumption		

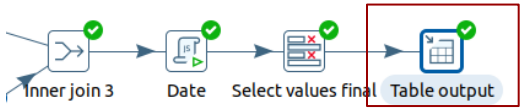
Get fields to select

Edit Mapping

Rows of step: Select values final (1000 rows)

	DistrictCode	District	MunicipalityCode	Municipality	ParishCode	Parish	YearMonthDay	Yes_Smartmeters	Total_Smartmeters	EnergyConsumption
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/06/01 00:00:00.000	1770	2153	2625319.534
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/07/01 00:00:00.000	1795	2156	2655651.065
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/08/01 00:00:00.000	1799	2153	2124549.08
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/09/01 00:00:00.000	1820	2155	2424573.401
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/10/01 00:00:00.000	1873	2155	2846325.2309999997
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/11/01 00:00:00.000	1880	2156	2804899.185
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022/12/01 00:00:00.000	1912	2157	2559577.038
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023/01/01 00:00:00.000	1928	2156	2979516.235
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023/02/01 00:00:00.000	1934	2155	2658704.36
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2023/03/01 00:00:00.000	1946	2157	2801696.182

Question 4 - ETL: Output



Step name: Table output

Connection: input5 [Edit...] [New...] [Wizard...]

Target schema: input5 [Browse...]

Target table: input5 [Browse...]

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☑

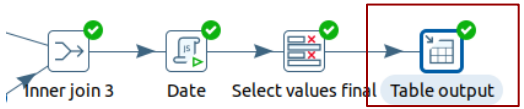
Main options | Database fields

Fields to insert:

[Get fields] [Enter field mapping]

	Table field	Stream field
1	DistrictCode	DistrictCode
2	District	District
3	MunicipalityC	MunicipalityCode
4	Municipality	Municipality
5	ParishCode	ParishCode
6	Parish	Parish
7	YearMonthDa	YearMonthDay
8	Yes_Smartme	Yes_Smartmeters
9	Total_Smartm	Total_Smartmeters
10	EnergyConsur	EnergyConsumption

Question 4 - ETL: Output



Field	Type	Null	Key	Default	Extra
DistrictCode	bigint(20)	YES		NULL	
District	varchar(16)	YES		NULL	
MunicipalityCode	bigint(20)	YES		NULL	
Municipality	varchar(27)	YES		NULL	
ParishCode	varchar(6)	YES		NULL	
Parish	varchar(30)	YES		NULL	
YearMonthDay	datetime	YES		NULL	
Yes_Smartmeters	int(11)	YES		NULL	
Total_Smartmeters	int(11)	YES		NULL	
EnergyConsumption	double	YES		NULL	

DistrictCode	District	MunicipalityCode	Municipality	ParishCode	Parish	YearMonthDay	Yes_Smartmeters	Total_Smartmeters	EnergyConsumption
1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-06-01 00:00:00	1770	2153	2625319.534
1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-07-01 00:00:00	1795	2156	2655651.065
1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-08-01 00:00:00	1799	2153	2124549.08
1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-09-01 00:00:00	1820	2155	2424573.401
1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	2022-10-01 00:00:00	1873	2155	2846325.231

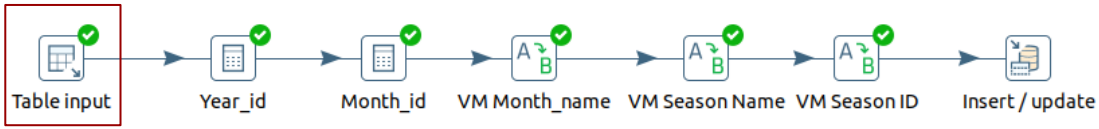
Question 5 – Time Dimension

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/24 09:55:58 - Spoon - Transformation opened.
2024/10/24 09:55:58 - Spoon - Launching transformation [4_project_aid_create_table]...
2024/10/24 09:55:58 - Spoon - Started the transformation execution.
2024/10/24 09:57:32 - Spoon - The transformation has finished!!
2024/10/24 10:05:10 - Spoon - Processing of transformation stopped.
2024/10/24 10:07:57 - Spoon - Processing of transformation stopped.
2024/10/24 10:12:33 - Spoon - Transformation opened.
2024/10/24 10:12:33 - Spoon - Launching transformation [dim_time_5]...
2024/10/24 10:12:33 - Spoon - Started the transformation execution.
2024/10/24 10:12:33 - dim_time_5 - Dispatching started for transformation [dim_time_5]
2024/10/24 10:12:46 - Table input.0 - linenr 50000
2024/10/24 10:12:46 - Year_id.0 - Linenr 50000
2024/10/24 10:13:04 - Month_id.0 - Linenr 50000
2024/10/24 10:13:10 - Table input.0 - Finished reading query, closing connection
2024/10/24 10:13:10 - Table input.0 - Finished processing (I=72107, O=0, R=0, W=72107, U=0, E=0)
2024/10/24 10:13:34 - Year_id.0 - Finished processing (I=0, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:14:00 - Month_id.0 - Finished processing (I=0, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:14:25 - VM Month_name.0 - Finished processing (I=0, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:14:31 - Spoon - Transformation opened.
2024/10/24 10:14:31 - Spoon - Launching transformation [dim_location_5]...
2024/10/24 10:14:31 - Spoon - Started the transformation execution.
2024/10/24 10:14:49 - Insert / update.0 - linenr 50000
2024/10/24 10:14:56 - VM Season Name.0 - Finished processing (I=0, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:15:06 - Spoon - The transformation has finished!!
2024/10/24 10:15:22 - VM Season ID.0 - Finished processing (I=0, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:15:47 - Insert / update.0 - Finished processing (I=72107, O=25, R=72107, W=72107, U=0, E=0)
2024/10/24 10:15:47 - Spoon - The transformation has finished!!
```

Question 5 – Time Dimension



Step name: Table input

Connection: input5 | Edit... | New... | Wizard...

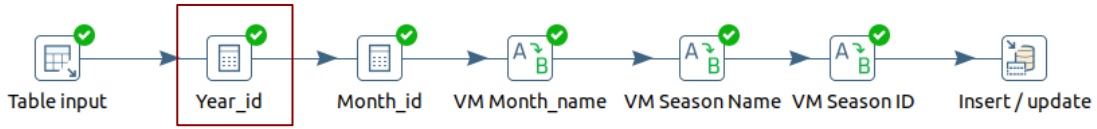
SQL | Get SQL select statement...

```
SELECT
  YearMonthDay
FROM input5.input5
```

Rows of step: Table input (1000 rows)

	YearMonthDay
1	2022/06/01 00:00:00.000000000
2	2022/07/01 00:00:00.000000000
3	2022/08/01 00:00:00.000000000
4	2022/09/01 00:00:00.000000000
5	2022/10/01 00:00:00.000000000
6	2022/11/01 00:00:00.000000000
7	2022/12/01 00:00:00.000000000
8	2023/01/01 00:00:00.000000000
9	2023/02/01 00:00:00.000000000
10	2023/03/01 00:00:00.000000000

Question 5 – Time Dimension



Step name

Year_id

Throw an error on non existing files

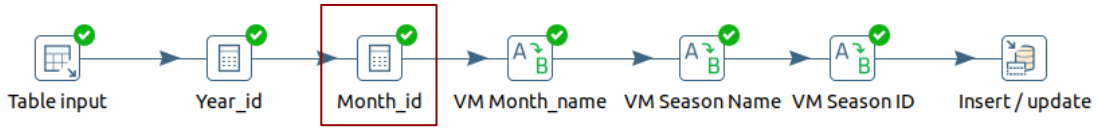
Fields:

	New field	Calculation	Field A	Field B	Field C	Value type
1	Year_id	Year of date A	YearMonthDay			Integer

Rows of step: Year_id (1000 rows)

	YearMonthDay	Year_id
1	2022/06/01 00:00:00.000000000	2022
2	2022/07/01 00:00:00.000000000	2022
3	2022/08/01 00:00:00.000000000	2022
4	2022/09/01 00:00:00.000000000	2022
5	2022/10/01 00:00:00.000000000	2022
6	2022/11/01 00:00:00.000000000	2022
7	2022/12/01 00:00:00.000000000	2022
8	2023/01/01 00:00:00.000000000	2023
9	2023/02/01 00:00:00.000000000	2023
10	2023/03/01 00:00:00.000000000	2023

Question 5 – Time Dimension



Step name

Month_id

Throw an error on non existing files

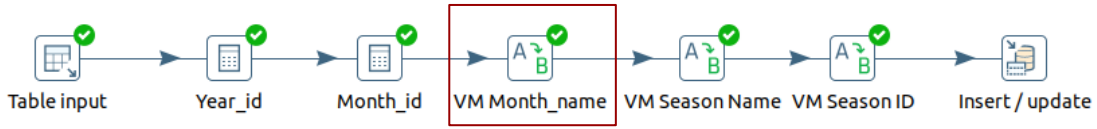
Fields:

	New field	Calculation	Field A	Field B	Field C	Value type
1	Month_id	Month of date A	YearMonthDay			Integer

Rows of step: Month_id (1000 rows)

	YearMonthDay	Year_id	Month_id
1	2022/06/01 00:00:00.000000000	2022	6
2	2022/07/01 00:00:00.000000000	2022	7
3	2022/08/01 00:00:00.000000000	2022	8
4	2022/09/01 00:00:00.000000000	2022	9
5	2022/10/01 00:00:00.000000000	2022	10
6	2022/11/01 00:00:00.000000000	2022	11
7	2022/12/01 00:00:00.000000000	2022	12
8	2023/01/01 00:00:00.000000000	2023	1
9	2023/02/01 00:00:00.000000000	2023	2
10	2023/03/01 00:00:00.000000000	2023	3

Question 5 – Time Dimension



Step name : VM Month_name

Fieldname to use : Month_id

Target field name (empty=overwrite) : Month_name

Default upon non-matching :

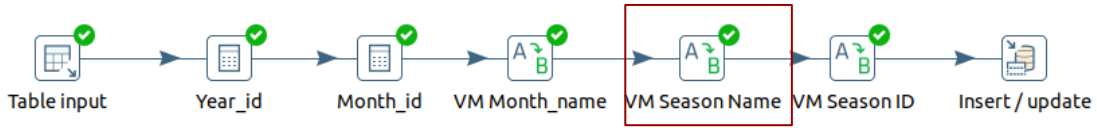
Field values:

	Source value	Target value
1	1	Jan
2	2	Feb
3	3	Mar
4	4	Apr
5	5	May
6	6	Jun
7	7	Jul
8	8	Aug
9	9	Sep
10	10	Oct
11	11	Nov
12	12	Dec

Rows of step: VM Month_name (1000 rows)

	YearMonthDay	Year_id	Month_id	Month_name
1	2022/06/01 00:00:00.000000000	2022	6	Jun
2	2022/07/01 00:00:00.000000000	2022	7	Jul
3	2022/08/01 00:00:00.000000000	2022	8	Aug
4	2022/09/01 00:00:00.000000000	2022	9	Sep
5	2022/10/01 00:00:00.000000000	2022	10	Oct
6	2022/11/01 00:00:00.000000000	2022	11	Nov
7	2022/12/01 00:00:00.000000000	2022	12	Dec
8	2023/01/01 00:00:00.000000000	2023	1	Jan
9	2023/02/01 00:00:00.000000000	2023	2	Feb
10	2023/03/01 00:00:00.000000000	2023	3	Mar

Question 5 – Time Dimension



Step name : VM Season Name

Fieldname to use: Month_id

Target field name: Season

Default upon nor:

Field values:

	Source value	Target value
1	1	winter
2	2	winter
3	3	winter
4	4	spring
5	5	spring
6	6	spring
7	7	summer
8	8	summer
9	9	summer
10	10	autumn
11	11	autumn
12	12	autumn

Rows of step: VM Season Name (1000 rows)

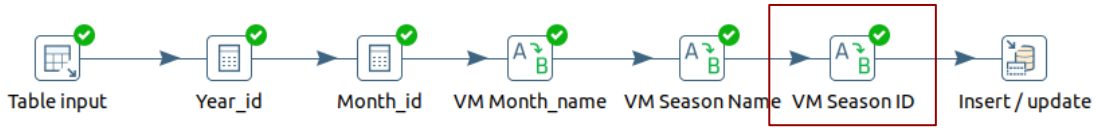
	YearMonthDay	Year_id	Month_id	Month_name	Season
1	2022/06/01 00:00:00.000000000	2022	6	Jun	spring
2	2022/07/01 00:00:00.000000000	2022	7	Jul	summer
3	2022/08/01 00:00:00.000000000	2022	8	Aug	summer
4	2022/09/01 00:00:00.000000000	2022	9	Sep	summer
5	2022/10/01 00:00:00.000000000	2022	10	Oct	autumn
6	2022/11/01 00:00:00.000000000	2022	11	Nov	autumn
7	2022/12/01 00:00:00.000000000	2022	12	Dec	autumn
8	2023/01/01 00:00:00.000000000	2023	1	Jan	winter
9	2023/02/01 00:00:00.000000000	2023	2	Feb	winter
10	2023/03/01 00:00:00.000000000	2023	3	Mar	winter

Question 5 – Time Dimension

In order to simplify the transformation, we chose to assign full months to the seasons, specifically:

1. winter: Jan, Feb, Mar;
2. spring: Apr, May, Jun;
3. summer: Jul, Aug, Sep;
4. autumn: Oct, Nov, Dec.

Question 5 – Time Dimension



Step name : VM Season ID

Fieldname tc Season

Target field r Season_id

Default upor

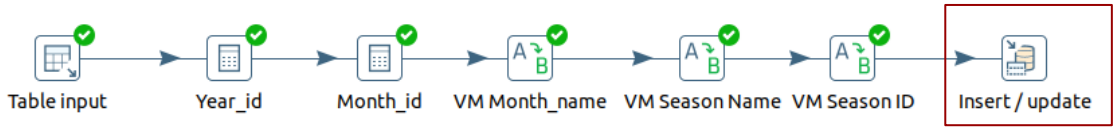
Field values:

	Source value	Target value
1	winter	1
2	spring	2
3	summer	3
4	autumn	4

Rows of step: VM Season ID (1000 rows)

	YearMonthDay	Year_id	Month_id	Month_name	Season	Season_id
1	2022/06/01 00:00:00.000000000	2022	6	Jun	spring	2
2	2022/07/01 00:00:00.000000000	2022	7	Jul	summer	3
3	2022/08/01 00:00:00.000000000	2022	8	Aug	summer	3
4	2022/09/01 00:00:00.000000000	2022	9	Sep	summer	3
5	2022/10/01 00:00:00.000000000	2022	10	Oct	autumn	4
6	2022/11/01 00:00:00.000000000	2022	11	Nov	autumn	4
7	2022/12/01 00:00:00.000000000	2022	12	Dec	autumn	4
8	2023/01/01 00:00:00.000000000	2023	1	Jan	winter	1
9	2023/02/01 00:00:00.000000000	2023	2	Feb	winter	1
10	2023/03/01 00:00:00.000000000	2023	3	Mar	winter	1

Question 5 – Time Dimension



time_id	year_id	season_id	season	month_id	month
2022-06-01 00:00:00	2022	2	spring	6	Jun
2022-07-01 00:00:00	2022	3	summer	7	Jul
2022-08-01 00:00:00	2022	3	summer	8	Aug
2022-09-01 00:00:00	2022	3	summer	9	Sep
2022-10-01 00:00:00	2022	4	autumn	10	Oct

Question 5 – Location Dimension

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/20 21:28:42 - Spoon - Transformation opened.
2024/10/20 21:28:42 - Spoon - Launching transformation [dim_location_5]...
2024/10/20 21:28:42 - Spoon - Started the transformation execution.
2024/10/20 21:28:42 - dim_location_5 - Dispatching started for transformation [dim_location_5]
2024/10/20 21:29:04 - Table input.0 - linenr 50000
2024/10/20 21:29:09 - Dimension lookup/update.0 - linenr 50000
2024/10/20 21:29:17 - Table input.0 - Finished reading query, closing connection
2024/10/20 21:29:17 - Table input.0 - Finished processing (I=71932, O=0, R=0, W=71932, U=0, E=0)
2024/10/20 21:29:23 - Dimension lookup/update.0 - Finished processing (I=5363, O=2837, R=71932, W=71932, U=0, E=0)
2024/10/20 21:29:23 - Spoon - The transformation has finished!!
```

Question 5 – Location Dimension

Step name: Table input

Connection: input5 | Edit... | New... | Wizard...

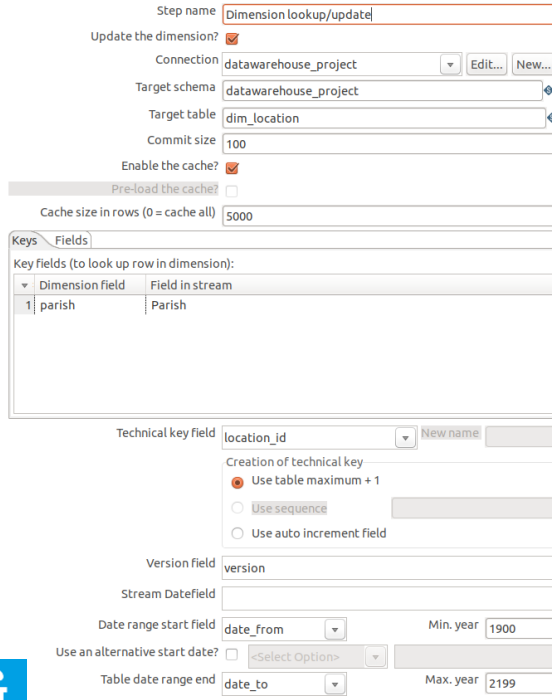
SQL | Get SQL select statement...

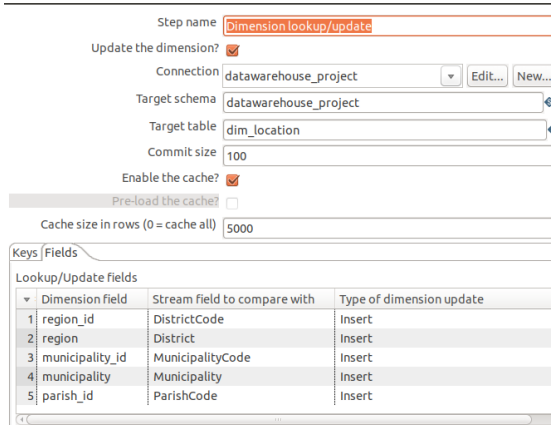
```
SELECT
  DistrictCode
, District
, MunicipalityCode
, Municipality
, ParishCode
, Parish
FROM input5.input5
```

Rows of step: Table input (1000 rows)

	DistrictCode	District	MunicipalityCode	Municipality	ParishCode	Parish
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
11	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
12	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA

Question 5 – Location Dimension

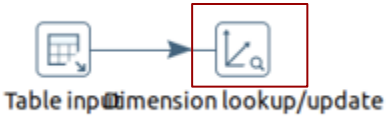




Rows of step: Dimension lookup/update (1000 rows)

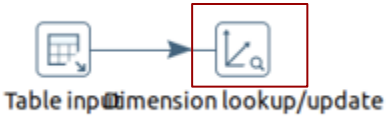
	DistrictCode	District	MunicipalityCode	Municipality	ParishCode	Parish
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
2	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
3	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
4	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
5	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
6	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
7	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
8	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
9	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA
10	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA

Question 5 – Location Dimension



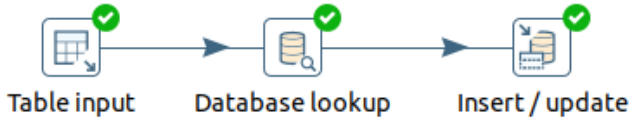
Considering that the “parish” name may vary over time, we chose to implement a Type 2 Slowly Changing Dimension.

Question 5 – Location Dimension



location_id	region_id	region	municipality_id	municipality	parish_id	parish	version	date_from	date_to
0	NULL	NULL	NULL	NULL	NULL	NULL	1	NULL	NULL
1	1	AVEIRO	101	Águeda	010103	AGUADA DE CIMA	1	1900-01-01 00:00:00	2200-01-01 00:00:00
2	1	AVEIRO	101	Águeda	010109	FERMENTELOS	1	1900-01-01 00:00:00	2200-01-01 00:00:00
3	1	AVEIRO	101	Águeda	010112	MACINHATA DO VOUGA	1	1900-01-01 00:00:00	2200-01-01 00:00:00
4	1	AVEIRO	101	Águeda	010119	VALONGO DO VOUGA	1	1900-01-01 00:00:00	2200-01-01 00:00:00

Question 6 – Facts Table

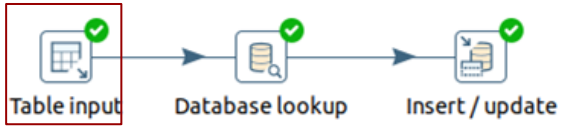


Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

```
2024/10/24 10:28:36 - Spoon - Transformation opened.
2024/10/24 10:28:36 - Spoon - Launching transformation [6_fact_energy_smart]...
2024/10/24 10:28:36 - Spoon - Started the transformation execution.
2024/10/24 10:28:37 - 6_fact_energy_smart - Dispatching started for transformation [6_fact_energy_smart]
2024/10/24 10:30:05 - Table input.0 - linenr 50000
2024/10/24 10:30:35 - Database lookup.0 - linenr 50000
2024/10/24 10:31:04 - Insert / update.0 - linenr 50000
2024/10/24 10:31:10 - Table input.0 - Finished reading query, closing connection
2024/10/24 10:31:10 - Table input.0 - Finished processing (I=72107, O=0, R=0, W=72107, U=0, E=0)
2024/10/24 10:31:37 - Database lookup.0 - Finished processing (I=72107, O=0, R=72107, W=72107, U=0, E=0)
2024/10/24 10:32:11 - Insert / update.0 - Finished processing (I=72107, O=67532, R=72107, W=72107, U=4400, E=0)
2024/10/24 10:32:11 - Spoon - The transformation has finished!!
```

Question 6 – Facts Table



Step name: Table input

Connection: input5 | Edit... | New... | Wizard...

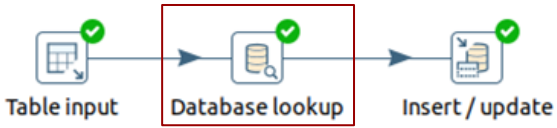
SQL | Get SQL select statement...

```
SELECT
  Parish
, YearMonthDay
, Yes_Smartmeters
, Total_Smartmeters
, EnergyConsumption
FROM input5.input5
```

Rows of step: Table input (1000 rows)

	Parish	YearMonthDay	Yes_Smartmeters	Total_Smartmeters	EnergyConsumption
1	AGUADA DE CIMA	2022/06/01 00:00:00.000000000	1770	2153	2625319.534
2	AGUADA DE CIMA	2022/07/01 00:00:00.000000000	1795	2156	2655651.065
3	AGUADA DE CIMA	2022/08/01 00:00:00.000000000	1799	2153	2124549.08
4	AGUADA DE CIMA	2022/09/01 00:00:00.000000000	1820	2155	2424573.401
5	AGUADA DE CIMA	2022/10/01 00:00:00.000000000	1873	2155	2846325.231
6	AGUADA DE CIMA	2022/11/01 00:00:00.000000000	1880	2156	2804899.185
7	AGUADA DE CIMA	2022/12/01 00:00:00.000000000	1912	2157	2559577.038
8	AGUADA DE CIMA	2023/01/01 00:00:00.000000000	1928	2156	2979516.235
9	AGUADA DE CIMA	2023/02/01 00:00:00.000000000	1934	2155	2658704.36
10	AGUADA DE CIMA	2023/03/01 00:00:00.000000000	1946	2157	2801696.182

Question 6 – Facts Table



Step name: Database lookup
Connection: datawarehouse_project
Lookup schema: datawarehouse_project
Lookup table: dim_location
Enable cache?
Cache size in rows (0=cache everything): 0
Load all data from table

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

	Table field	Comparator	Field1	Field2
1	parish	=	Parish	
2	date_from	<=	YearMonthDay	
3	date_to	>	YearMonthDay	

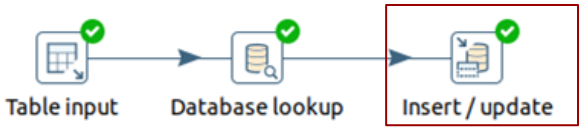
Values to return from the lookup table :

	Field	New name	Default	Type
1	location_id			Integer

Rows of step: Database lookup (1000 rows)

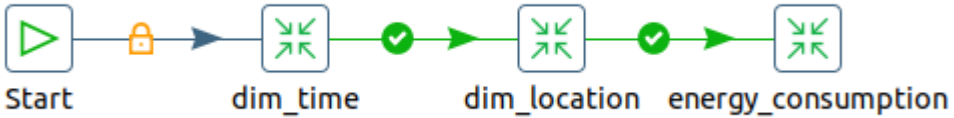
	Parish	YearMonthDay	Yes_Smartmeters	Total_Smartmeters	EnergyConsumption	location_id
1	AGUADA DE CIMA	2022/06/01 00:00:00.000000000	1770	2153	2625319.534	1
2	AGUADA DE CIMA	2022/07/01 00:00:00.000000000	1795	2156	2655651.065	1
3	AGUADA DE CIMA	2022/08/01 00:00:00.000000000	1799	2153	2124549.08	1
4	AGUADA DE CIMA	2022/09/01 00:00:00.000000000	1820	2155	2424573.401	1
5	AGUADA DE CIMA	2022/10/01 00:00:00.000000000	1873	2155	2846325.231	1
6	AGUADA DE CIMA	2022/11/01 00:00:00.000000000	1880	2156	2804899.185	1
7	AGUADA DE CIMA	2022/12/01 00:00:00.000000000	1912	2157	2559577.038	1
8	AGUADA DE CIMA	2023/01/01 00:00:00.000000000	1928	2156	2979516.235	1
9	AGUADA DE CIMA	2023/02/01 00:00:00.000000000	1934	2155	2658704.36	1
10	AGUADA DE CIMA	2023/03/01 00:00:00.000000000	1946	2157	2801696.182	1

Question 6 – Facts Table

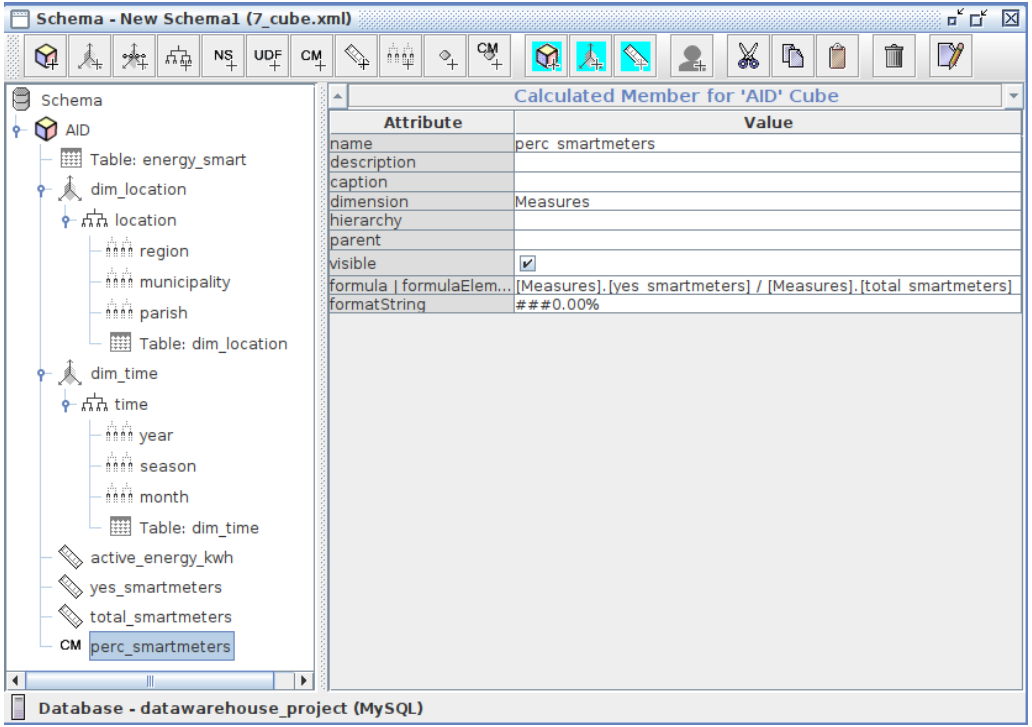


time_id	location_id	yes_smartmeters	total_smartmeters	active_energy_kwh
2022-06-01 00:00:00	1	1770	2153	2625319.534
2022-06-01 00:00:00	2	1249	1797	518598.864
2022-06-01 00:00:00	3	1168	1834	1566676.716
2022-06-01 00:00:00	4	90	172	811643.919
2022-06-01 00:00:00	5	7579	7986	8620044.239

Question 6 – Job [pipeline]



Question 7 – OLAP Cube: Pentaho Schema Workbench



Question 7 – OLAP Cube: XML File

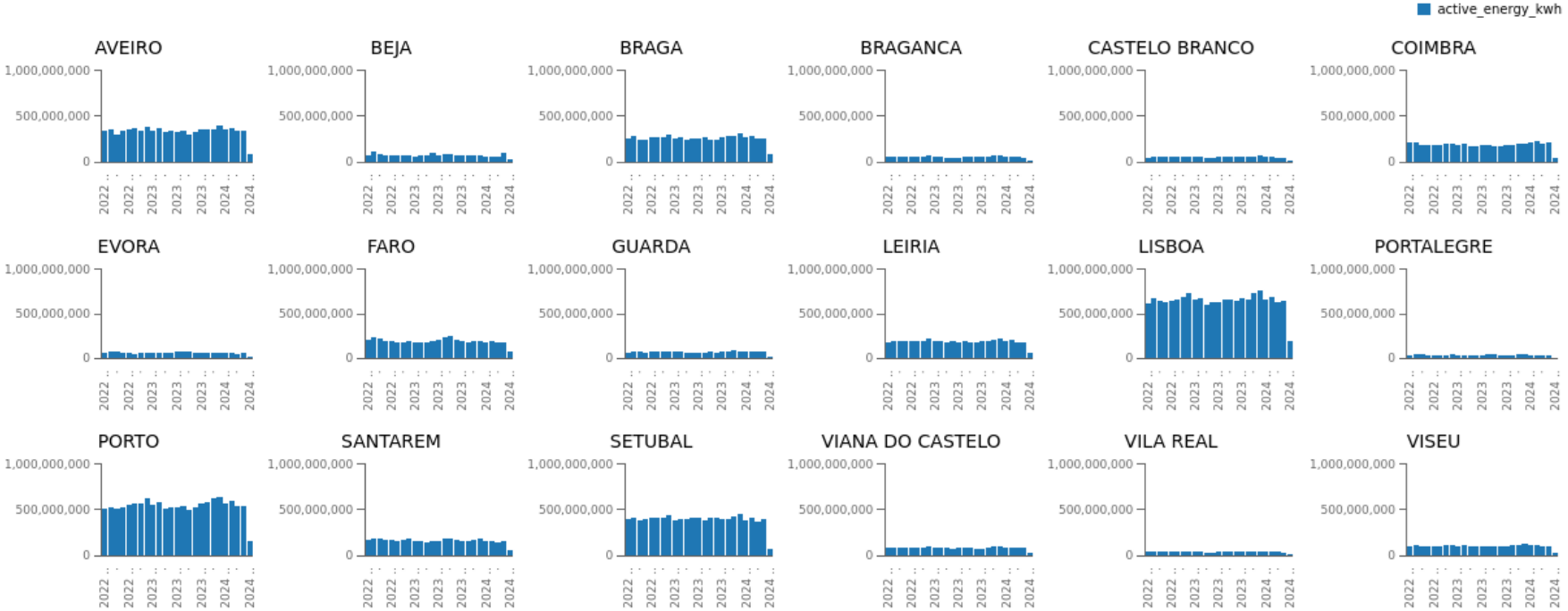
```xml
<Schema name="New Schema1">
  <Cube name="AID" visible="true" cache="true" enabled="true">
    <Table name="energy_smart">
    </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="location_id" highCardinality="false" name="dim_location">
      <Hierarchy name="location" visible="true" hasAll="true" allMemberName="allLocations" primaryKey="location_id">
        <Table name="dim_location">
        </Table>
        <Level name="region" visible="true" column="region" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <Level name="municipality" visible="true" column="municipality" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <Level name="parish" visible="true" column="parish" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="TimeDimension" visible="true" foreignKey="time_id" highCardinality="false" name="dim_time">
      <Hierarchy name="time" visible="true" hasAll="true" allMemberName="allTimes" primaryKey="time_id">
        <Table name="dim_time">
        </Table>
        <Level name="year" visible="true" column="year_id" type="Integer" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">
        </Level>
        <Level name="season" visible="true" column="season" ordinalColumn="season_id" type="String" uniqueMembers="false" levelType="TimeQuarters" hideMemberIf="Never">
        </Level>
        <Level name="month" visible="true" column="month" ordinalColumn="month_id" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="active_energy_kwh" column="active_energy_kwh" datatype="Numeric" formatString="#######,##" aggregator="sum" visible="true">
    </Measure>
    <Measure name="yes_smartmeters" column="yes_smartmeters" datatype="Numeric" aggregator="sum" visible="false">
    </Measure>
    <Measure name="total_smartmeters" column="total_smartmeters" datatype="Numeric" aggregator="sum" visible="false">
    </Measure>
    <CalculatedMember name="perc_smartmeters" formatString="###0.00%" formula="[Measures].[yes_smartmeters] / [Measures].[total_smartmeters]" dimension="Measures" visible="true">
    </CalculatedMember>
  </Cube>
</Schema>
```

Question 8 - a) Consumption

region	year	month	active_energy_kwh
AVEIRO	2022	Jun	344,487,270
		Jul	359,731,326
		Aug	306,869,331
		Sep	345,391,042
		Oct	359,835,060
		Nov	365,523,247
		Dec	345,125,505
	2023	Jan	385,166,029
		Feb	346,978,244
		Mar	373,399,531
		Apr	329,530,429
		May	347,381,269
		Jun	334,956,948
		Jul	339,453,088
		Aug	294,948,615
		Sep	327,086,895
		Oct	355,145,313
		Nov	359,406,189
		Dec	357,119,453
	2024	Jan	394,910,301
		Feb	362,162,249
		Mar	371,117,006
		Apr	341,429,700
		May	344,261,174
		Jun	87,577,000
	2022	Jun	79,019,426

```
WITH
SET [~ROWS_dim_location_dim_location.location] AS
    {[dim_location.location].[region].Members}
SET [~ROWS_dim_time_dim_time.time] AS
    Hierarchize({{{[dim_time.time].[year].Members}, {[dim_time.time].[month].Members}})
SELECT
NON EMPTY {[Measures].[active_energy_kwh]} ON COLUMNS,
NON EMPTY NonEmptyCrossJoin([~ROWS_dim_location_dim_location.location], [~ROWS_dim_time_dim_time.time]) ON ROWS
FROM [AID]
```

Question 8 - a) Consumption



Question 8 - a) Consumption

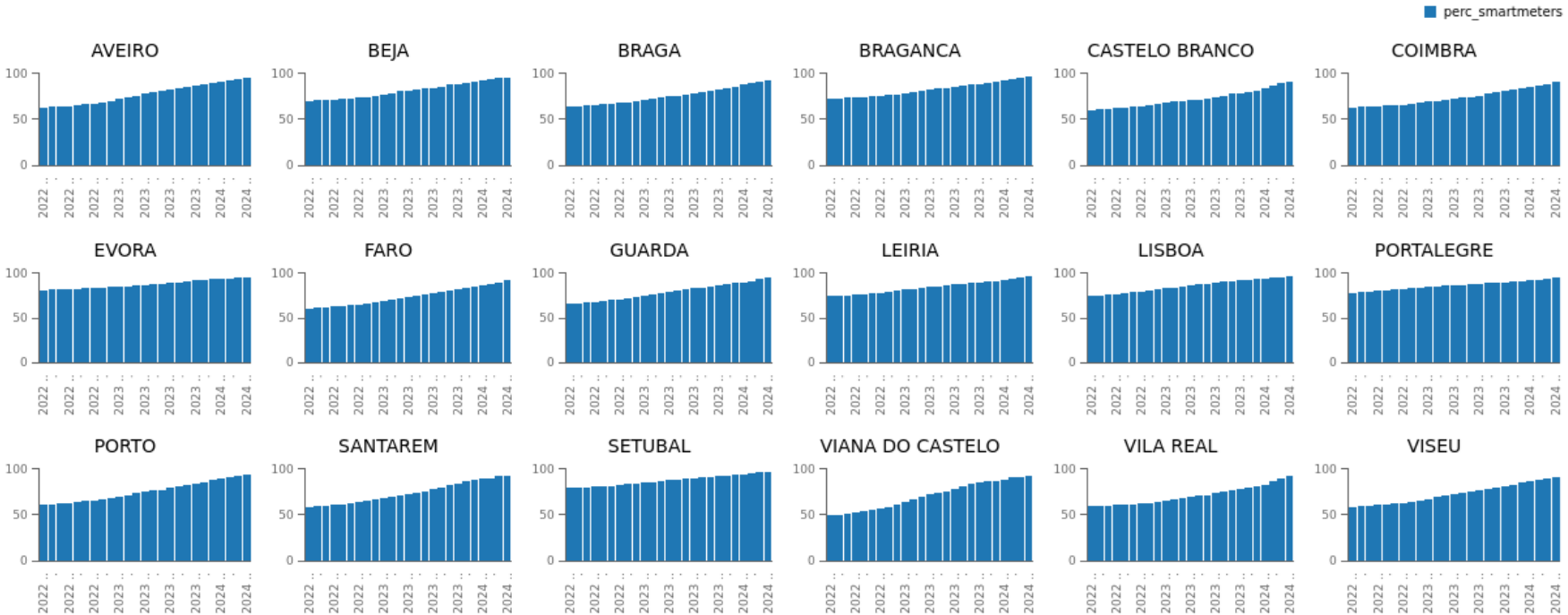
From observing the district graphs, it cannot be concluded that there was an increase or decrease in energy consumption over the years, which, with small seasonal fluctuations, remained stable.

Question 8 - a) “Smart meters” Percentage

region	year	month	perc_smartmeters
AVEIRO	2022	Jun	63.14%
		Jul	63.71%
		Aug	64.21%
		Sep	64.85%
		Oct	65.75%
		Nov	66.63%
		Dec	67.45%
	2023	Jan	68.69%
		Feb	70.33%
		Mar	72.29%
		Apr	74.03%
		May	76.04%
		Jun	77.75%
		Jul	79.27%
		Aug	80.62%
		Sep	82.35%
		Oct	84.03%
		Nov	86.02%
		Dec	87.15%
	2024	Jan	88.67%
		Feb	90.21%
		Mar	91.58%
		Apr	93.02%
		May	94.34%
		Jun	95.43%
	2022	Jun	70.17%

```
1  WITH
2  SET [~ROWS_dim_location_dim_location.location] AS
3      {[dim_location.location].[region].Members}
4  SET [~ROWS_dim_time_dim_time.time] AS
5      Hierarchize({{[dim_time.time].[year].Members}, {[dim_time.time].[month].Members}})
6  SELECT
7  NON EMPTY {[Measures].[perc_smartmeters]} ON COLUMNS,
8  NON EMPTY NonEmptyCrossJoin([~ROWS_dim_location_dim_location.location], [~ROWS_dim_time_dim_time.time]) ON ROWS
9  FROM [AID]
```

Question 8 - a) “Smart meters” Percentage



Question 8 - a) “Smart meters” Percentage

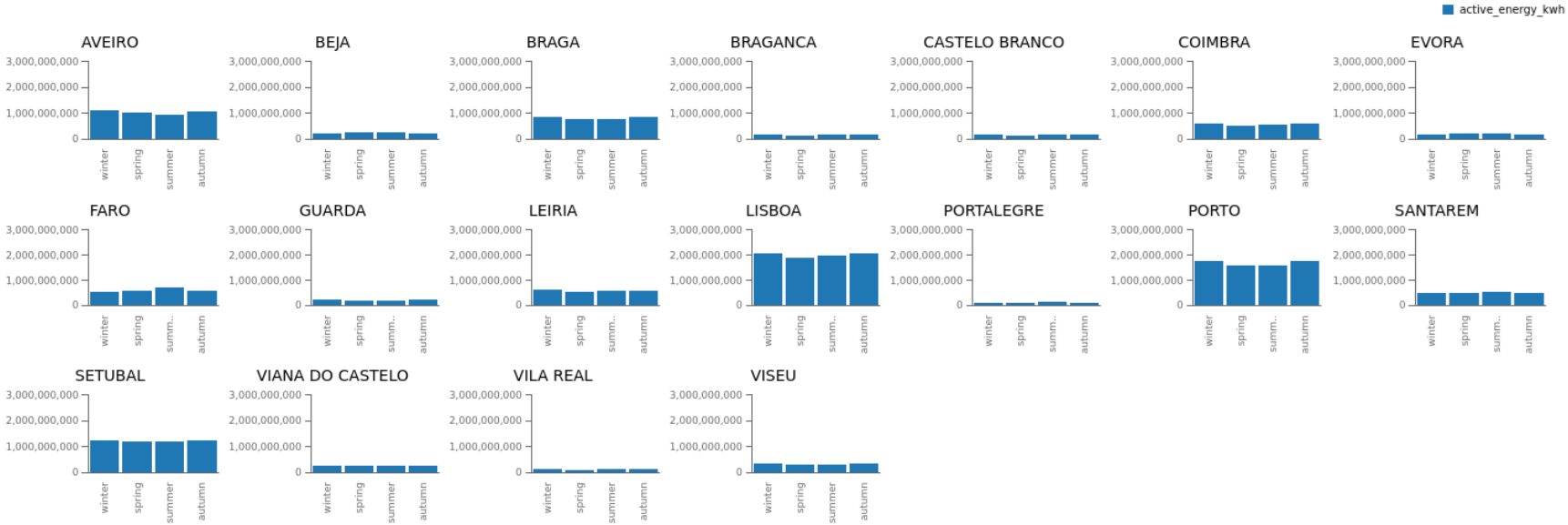
In all districts, there was a gradual increase in the percentage of 'smart meters' over the years.

Question 8 - b) Influence of the season on consumption

region	season	active_energy_kwh
AVEIRO	winter	1,105,543,803
	spring	1,011,868,646
	summer	961,488,599
	autumn	1,071,670,954
BEJA	winter	205,494,975
	spring	250,462,671
	summer	257,741,672
	autumn	220,926,555
BRAGA	winter	842,467,736
	spring	752,864,706
	summer	753,666,958
	autumn	837,623,219
BRAGANCA	winter	185,861,202
	spring	151,836,120
	summer	164,453,851
	autumn	183,563,396
CASTELO BRANCO	winter	172,232,725
	spring	149,008,741
	summer	167,688,635
	autumn	179,704,029
COIMBRA	winter	583,053,099

```
1  WITH
2  SET [~ROWS_dim_location_dim_location.location] AS
3      {[dim_location.location].[region].Members}
4
5  SET [~ROWS_dim_time_dim_time.time_season] AS
6      Descendants([dim_time.time].[year].[2023], [dim_time.time].[season])
7
8  SELECT
9  NON EMPTY {[Measures].[active_energy_kwh]} ON COLUMNS,
10 NON EMPTY NonEmptyCrossJoin([~ROWS_dim_location_dim_location.location], [~ROWS_dim_time_dim_time.time_season]) ON ROWS
11 FROM [AID]
12
```

Question 8 - b) Influence of the season on consumption



Question 8 - b) Influence of the season on consumption

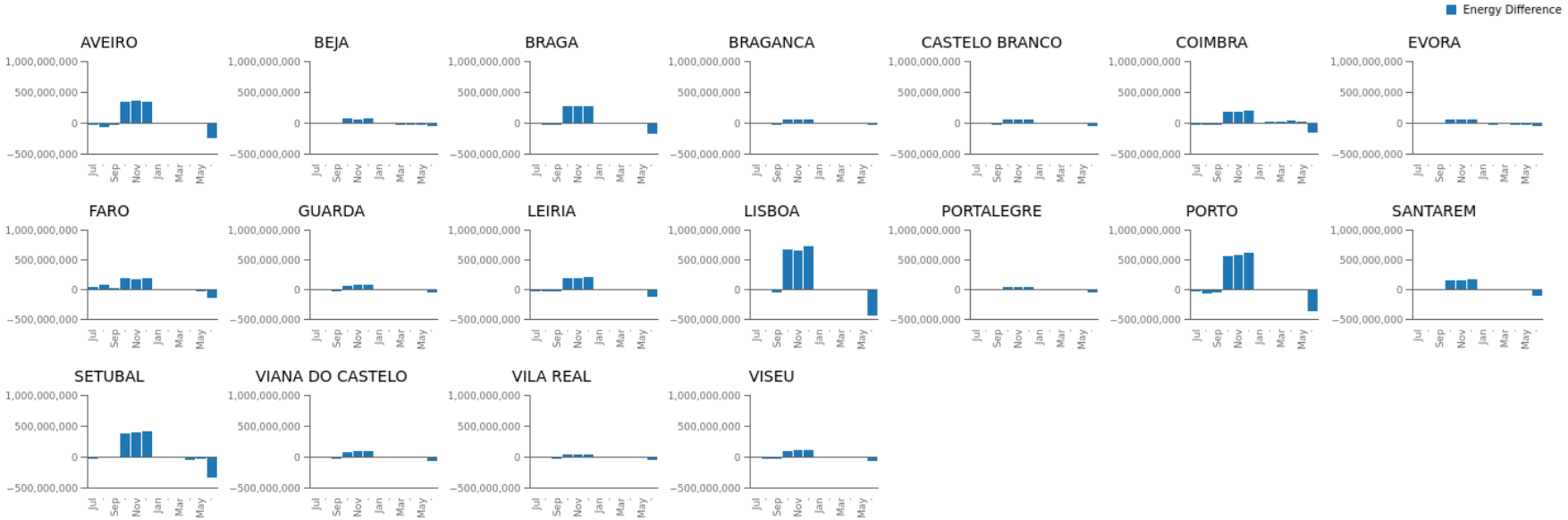
In the particular case of 2023, for which we have records for all months, it is observed that, with the exception of some southern districts, such as Faro and Beja, where it is warmer in the summer and consumption is higher in spring and summer due to air conditioning use, in almost all districts consumption is higher in autumn and winter due to heating needs.

Question 8 - c) Impact of 'smart meters' on consumption

region	month	Energy Difference
AVEIRO	Jul	-20,381,972
	Aug	-70,574,632
	Sep	-18,038,610
	Oct	355,145,313
	Nov	359,406,189
	Dec	357,119,453
	Jan	9,744,272
	Feb	15,184,006
	Mar	-2,282,525
	Apr	11,899,272
	May	-3,120,095
	Jun	-247,379,948
BEJA	Jul	22,019,185

```
WITH
SET [~ROWS_dim_location_dim_location.location] AS
    {[dim_location.location].[region].Members}

SET [~ROWS_dim_time_dim_time.time_season] AS
    {[dim_time.time].[year].[2023].[summer].[Jul] : [dim_time.time].[year].[2024].[spring].[Jun]}

MEMBER [Measures].[Energy Difference] AS
    [Measures].[active_energy_kwh] -
    (ParallelPeriod([dim_time.time].[year], 1, [dim_time.time].CurrentMember), [Measures].[active_energy_kwh])

SELECT
NON EMPTY {[Measures].[Energy Difference]} ON COLUMNS,
NON EMPTY NonEmptyCrossJoin([~ROWS_dim_location_dim_location.location], [~ROWS_dim_time_dim_time.time_season]) ON ROWS
FROM [AID]
```

Question 8 - c) Impact of 'smart meters' on consumption

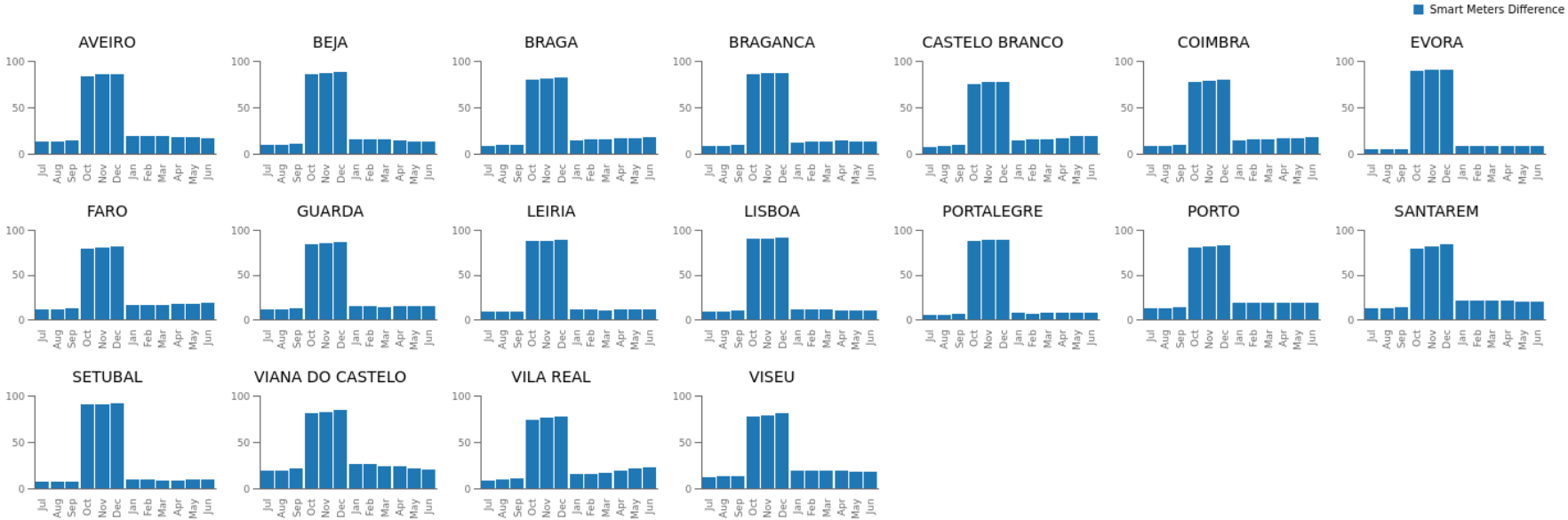


Question 8 - c) Impact of 'smart meters' on consumption

region	month	Smart Meters Difference
AVEIRO	Jul	13.52%
	Aug	14.00%
	Sep	14.91%
	Oct	84.03%
	Nov	86.02%
	Dec	87.15%
	Jan	19.98%
	Feb	19.88%
	Mar	19.29%
	Apr	18.99%
	May	18.31%
	Jun	17.69%
BEJA	Jul	9.81%

```
WITH
SET [~ROWS_dim_location_dim_location.location] AS
    {[dim_location.location].[region].Members}

SET [~ROWS_dim_time_dim_time.time_season] AS
    {[dim_time.time].[year].[2023].[summer].[Jul] : [dim_time.time].[year].[2024].[spring].[Jun]}


MEMBER [Measures].[Smart Meters Difference] AS
    [Measures].[perc_smartmeters] -
    (ParallelPeriod([dim_time.time].[year], 1, [dim_time.time].CurrentMember), [Measures].[perc_smartmeters])

SELECT
NON EMPTY {[Measures].[Smart Meters Difference]} ON COLUMNS,
NON EMPTY NonEmptyCrossJoin([~ROWS_dim_location_dim_location.location], [~ROWS_dim_time_dim_time.time_season]) ON ROWS
FROM [AID]

```

Question 8 - c) Impact of 'smart meters' on consumption



Question 8 - c) Impact of 'smart meters' on consumption

Considering the variations observed relative to the same months in the period between July 2023 and June 2024;

Given

the gradual increase in "smart meters," especially from October to December,
the rise in energy consumption during the winter,
and the decrease in consumption during the summer,

it is not possible to infer a linear impact of "smart meters" on consumption, as there are certainly other factors influencing the consumption variations.