Data Analysis and Integration Project

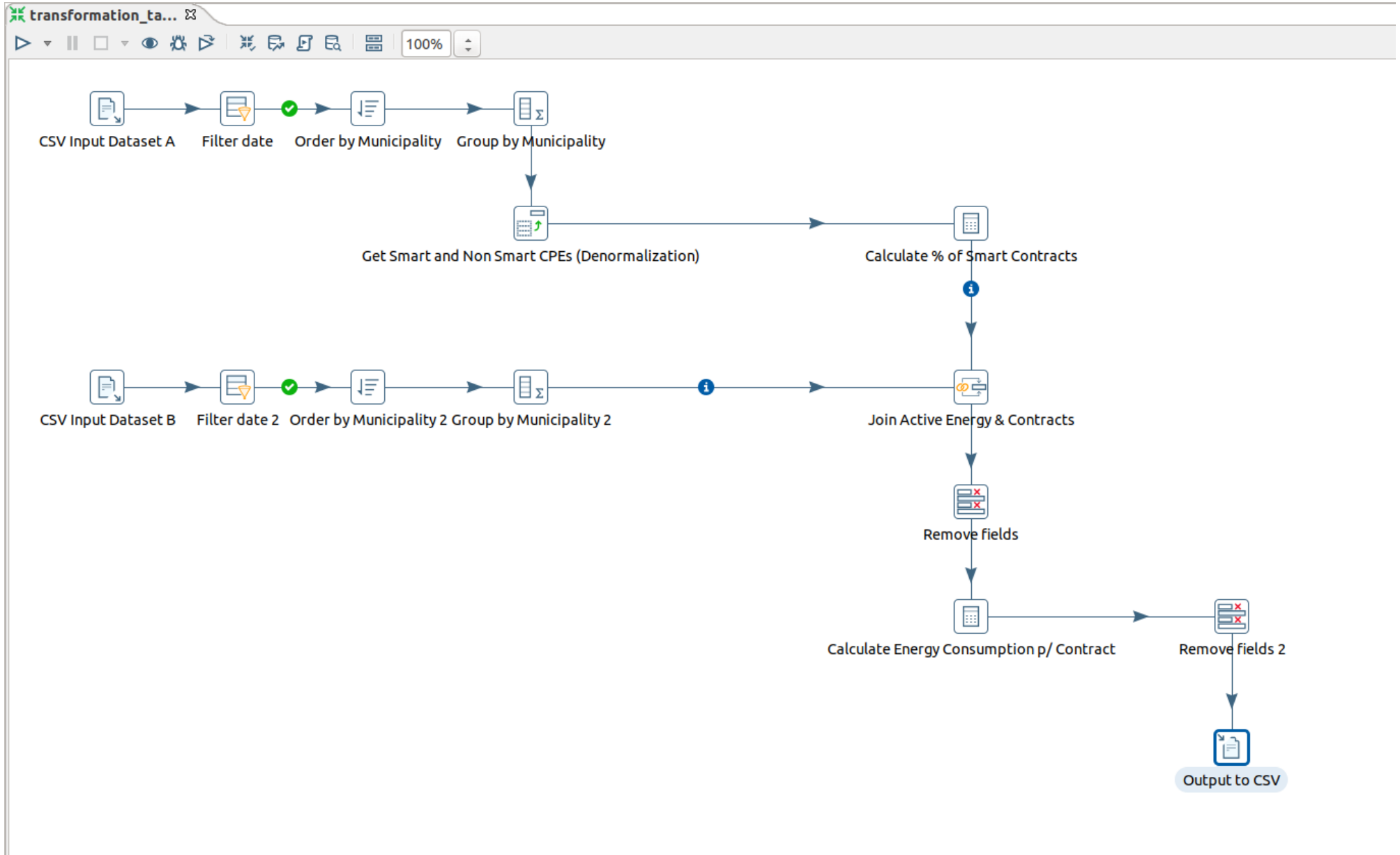
Group 16

Francesco Pelizzari – Nr. 99217

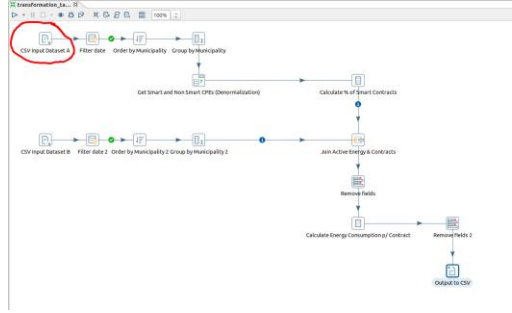
Simão Silva – Nr. 99329

Task 1

Transformation



CSV Input Dataset A | Configuration



CSV file input

Step name: CSV Input Dataset A

Filename: /home/aid/adi-project/data/21-contadores-de-energia.csv — 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? ☐

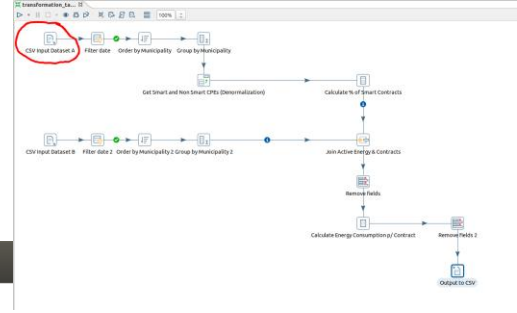
Format: mixed

File encoding: UTF-8

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Year	Integer	#	15	0	$	.	,	none
2	Month	Integer	#	15	0	$	.	,	none
3	Date	String		7		$	.	,	none
4	District	String		16		$	.	,	none
5	Municipality	String		27		$	.	,	none
6	Parish	String		30		$	.	,	none
7	Includes Smart Meter	String		3		$	.	,	none
8	Number of CPE's	Integer	#	15	0	$	.	,	none
9	DistrictCode	Integer	#	15	0	$	.	,	none
10	DistrictMunicipalityCode	Integer	#	15	0	$	.	,	none
11	DistrictMunicipalityParishCode	String	#	15	0	$	.	,	none
12	Active Contract	String		3		$	.	,	none

Help | OK | Get Fields | Preview | Cancel

CSV Input Dataset A | Preview

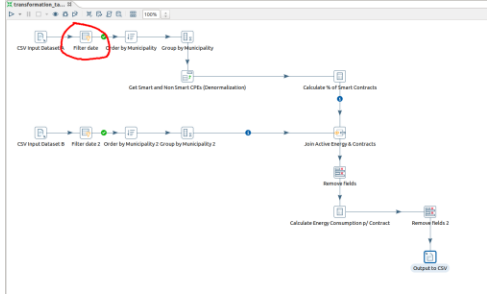


Examine preview data

Rows of step: CSV Input Dataset A (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active
1	2023	10	2023-10	GUARDA	Mêda	AVELOSO	Não	8	9	909	090901	Sim
2	2023	10	2023-10	SANTAREM	Santarém	ALCANHOES	Não	23	14	1416	141605	Sim
3	2023	10	2023-10	SETUBAL	Alcochete	SAMOUCO	Não	2	15	1502	150202	Sim
4	2023	10	2023-10	VISEU	Tabuaço	ADORIGO	Não	35	18	1819	181901	Sim
5	2023	10	2023-10	BRAGA	Terras de Bouro	RIO CALDO	Não	8	3	310	031013	Sim
6	2023	10	2023-10	BRAGANCA	Bragança	CASTRO DE AVELAS	Não	14	4	402	040209	Sim
7	2023	10	2023-10	VILA REAL	Murça	VALONGO DE MILHAIS	Não	8	17	1707	170708	Sim
8	2023	10	2023-10	LISBOA	Alenquer	MECA	Não	14	11	1101	110107	Sim
9	2023	10	2023-10	VIANA DO CASTELO	Caminha	ARGELA	Não	6	16	1602	160205	Sim
10	2023	10	2023-10	GUARDA	Mêda	BARREIRA	Não	2	9	909	090902	Sim
11	2023	10	2023-10	VIANA DO CASTELO	Paredes de Coura	CUNHA	Não	1	16	1605	160507	Sim
12	2023	10	2023-10	VISEU	Tabuaço	VALENCA DO DOURO	Não	2	18	1819	181917	Sim
13	2023	10	2023-10	LEIRIA	Óbidos	USSEIRA	Não	7	10	1012	101209	Sim
14	2023	10	2023-10	VIANA DO CASTELO	Ponte de Lima	BERTIANDOS	Não	2	16	1607	160708	Sim
15	2023	10	2023-10	BRAGANCA	Carrazeda de Ansiães	PARAMBOS	Não	1	4	403	040311	Sim
16	2023	10	2023-10	PORTO	Póvoa de Varzim	UF AGUCADOURA E NAVAIS	Não	1031	13	1313	131314	Sim
17	2023	10	2023-10	COIMBRA	Lousã	UF LOUSA E VILARINHO	Não	34	6	607	060708	Sim
18	2023	10	2023-10	LEIRIA	Caldas da Rainha	UF TORNADA E SALIR DO PORTO	Não	330	10	1006	100619	Sim
19	2023	10	2023-10	CASTELO BRANCO	Castelo Branco	UF FREIXIAL E JUNCAL DO CAMPO	Não	517	5	502	050229	Sim
20	2023	10	2023-10	LEIRIA	Porto de Mós	SAO BENTO	Sim	449	10	1016	101610	Sim
21	2023	10	2023-10	LISBOA	Sintra	COLARES	Sim	4376	11	1111	111105	Sim
22	2023	10	2023-10	BRAGA	Guimarães	FERMENTOES	Não	53	3	308	030815	Sim
23	2023	10	2023-10	VISEU	Mortágua	MARMELEIRA	Não	22	18	1808	180805	Sim
24	2023	10	2023-10	GUARDA	Seia	SANDOMIL	Não	225	9	912	091212	Sim
25	2023	10	2023-10	VISEU	Penalva do Castelo	ESMOLFE	Sim	340	18	1811	181103	Sim
26	2023	10	2023-10	GUARDA	Guarda	UF AVELAS DE AMBOM E ROCAMONDO	Sim	100	9	907	090761	Sim
27	2023	10	2023-10	PORTO	Maia	MOREIRA	Não	2157	13	1306	130609	Sim
28	2023	10	2023-10	AVEIRO	Castelo de Paiva	REAL	Não	297	1	106	010606	Sim
29	2023	10	2023-10	VISEU	Santa Comba Dão	PINHEIRO DE AZERE	Sim	364	18	1814	181403	Sim
30	2023	10	2023-10	LEIRIA	Leiria	AMOR	Sim	2267	10	1009	100901	Sim
31	2023	10	2023-10	VISEU	Lamego	VARZEA DE ABRUNHAIS	Sim	275	18	1805	180523	Sim

Filter Date | Configuration



Filter rows

Step name: Filter date

Send 'true' data to step: Order by Municipality

Send 'false' data to step:

The condition:

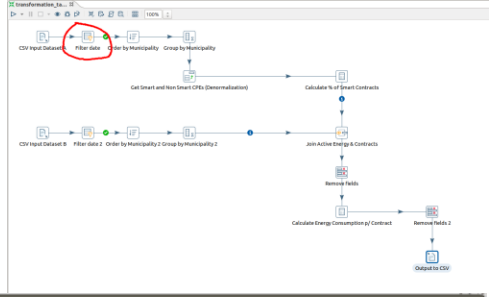
```
Year = [2024]

AND

Month = [6]
```

Help | OK | Cancel

Filter Date | Preview

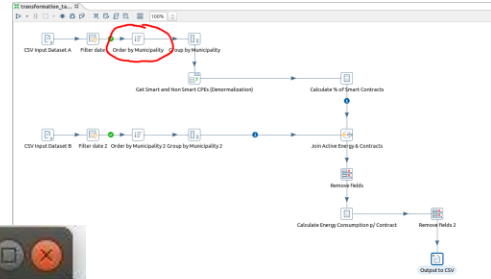


Examine preview data

Rows of step: Filter date (1000 rows)

	Year	Month	Date	District	Municipality	Parish	Includes Smart Meter	Number of CPE's	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCode	Active C
1	2024	6	2024-06	COIMBRA	Figueira da Foz	TAVAREDE	Sim	6341	6	605	060512	Sim
2	2024	6	2024-06	BRAGA	Fafe	REGADAS	Sim	546	3	307	030723	Sim
3	2024	6	2024-06	BRAGA	Guimarães	BRITO	Não	19	3	308	030807	Sim
4	2024	6	2024-06	VILA REAL	Boticas	PINHO	Sim	344	17	1702	170213	Sim
5	2024	6	2024-06	VIANA DO CASTELO	Viana do Castelo	SAO ROMAO DE NEIVA	Sim	713	16	1609	160923	Sim
6	2024	6	2024-06	BRAGANCA	Mirandela	TORRE DE DONA CHAMA	Não	23	4	407	040730	Sim
7	2024	6	2024-06	CASTELO BRANCO	Belmonte	MACAINHAS	Sim	300	5	501	050105	Sim
8	2024	6	2024-06	CASTELO BRANCO	Fundão	CASTELO NOVO	Não	12	5	504	050413	Sim
9	2024	6	2024-06	SANTAREM	Ourém	CAXARIAS	Sim	1554	14	1421	142104	Sim
10	2024	6	2024-06	BRAGA	Guimarães	UF TABUADELO E SAO FAUSTINO	Sim	1016	3	308	030891	Sim
11	2024	6	2024-06	COIMBRA	Figueira da Foz	SAO PEDRO	Não	12	6	605	060514	Sim
12	2024	6	2024-06	GUARDA	Celorico da Beira	UF CELORICO E VILA BOA MONDEGO	Sim	1791	9	903	090324	Sim
13	2024	6	2024-06	VISEU	Tondela	UF MOURAZ VILA NOVA RAINHA	Sim	484	18	1821	182129	Sim
14	2024	6	2024-06	VISEU	São Pedro do Sul	VILA MAIOR	Não	4	18	1816	181619	Sim
15	2024	6	2024-06	PORTALEGRE	Avis	FIGUEIRA E BARROS	Sim	225	12	1203	120306	Sim
16	2024	6	2024-06	CASTELO BRANCO	Penamacor	UF A BISPO AGUAS A JOAO PIRES	Não	25	5	507	050713	Sim
17	2024	6	2024-06	BRAGA	Fafe	FAFE	Sim	10095	3	307	030709	Sim
18	2024	6	2024-06	VIANA DO CASTELO	Monção	UF MESSEGAES VALADARES E SA	Sim	449	16	1604	160437	Sim
19	2024	6	2024-06	VILA REAL	Vila Pouca de Aguiar	BRAGADO	Sim	237	17	1713	171304	Sim
20	2024	6	2024-06	FARO	Tavira	CACHOPO	Sim	344	8	814	081401	Sim
21	2024	6	2024-06	PORTALEGRE	Monforte	SANTO ALEIXO	Não	19	12	1211	121103	Sim
22	2024	6	2024-06	BRAGANCA	Torre de Moncorvo	HORTA DA VILARICA	Não	8	4	409	040909	Sim
23	2024	6	2024-06	LISBOA	Lisboa	LUMIAR	Sim	27507	11	1106	110618	Sim
24	2024	6	2024-06	COIMBRA	Mira	PRAIA DE MIRA	Sim	3655	6	608	060804	Sim
25	2024	6	2024-06	LISBOA	Vila Franca de Xira	UF POVOA SANTA IRIA FORTE CASA	Não	2458	11	1114	111415	Sim
26	2024	6	2024-06	SANTAREM	Torres Novas	UF OLAIA E PACO	Sim	966	14	1419	141919	Sim
27	2024	6	2024-06	BRAGA	Braga	UF CRESPOS E POUSADA	Não	82	3	303	030368	Sim
28	2024	6	2024-06	BRAGA	Vila Verde	VALDREU	Não	154	3	313	031355	Sim
29	2024	6	2024-06	BRAGA	Barcelos	UF SEQUEADE BASTUCO	Não	380	3	302	0302FD	Sim
30	2024	6	2024-06	BRAGA	Póvoa de Lanhoso	FERREIROS	Não	21	3	309	030908	Sim
31	2024	6	2024-06	VISEU	Nelas	CANAS DE SENHORIM	Sim	2139	18	1809	180901	Sim

Order by Municipality | Configuration



Sort rows

Step name: Order by Municipality

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?

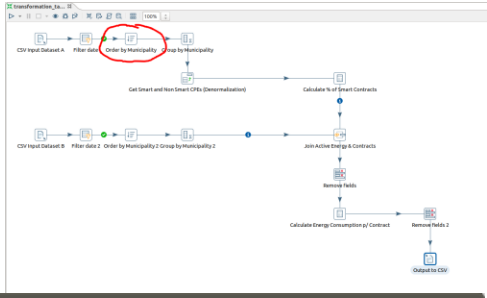
Only pass unique rows? (verifies keys only)

Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength
1	Municipality	Y	N	N	0
2	Includes Smart Meter	Y	N	N	0

Help OK Cancel Get Fields

Order by Municipality | Preview

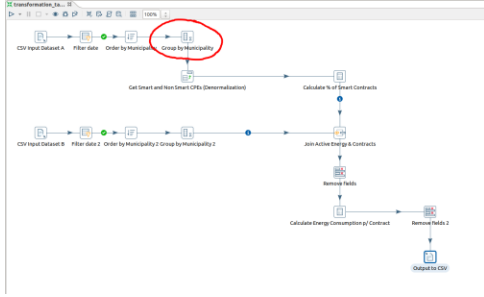


Examine preview data

Rows of step: Order by Municipality (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	SANTAREM	Abrantes	MOURISCAS	Não	309	14	1401	140106	Sim
2	2024	6	2024-06	SANTAREM	Abrantes	UF ABRANTES E ALFERRAREDE	Não	23	14	1401	140120	Sim
3	2024	6	2024-06	SANTAREM	Abrantes	RIO DE MOINHOS	Não	45	14	1401	140108	Sim
4	2024	6	2024-06	SANTAREM	Abrantes	UF S M RIO TORTO ROSSIO S TEJO	Não	292	14	1401	140124	Sim
5	2024	6	2024-06	SANTAREM	Abrantes	PEGO	Não	346	14	1401	140107	Sim
6	2024	6	2024-06	SANTAREM	Abrantes	FONTES	Não	295	14	1401	140118	Sim
7	2024	6	2024-06	SANTAREM	Abrantes	UF SAO FACUNDO E VALE DAS MOS	Não	564	14	1401	140123	Sim
8	2024	6	2024-06	SANTAREM	Abrantes	CARVALHAL	Não	244	14	1401	140119	Sim
9	2024	6	2024-06	SANTAREM	Abrantes	BEMPOSTA	Não	62	14	1401	140104	Sim
10	2024	6	2024-06	SANTAREM	Abrantes	MARTINCHEL	Não	173	14	1401	140105	Sim
11	2024	6	2024-06	SANTAREM	Abrantes	UF ALDEIA DO MATO E SOUTO	Não	348	14	1401	140121	Sim
12	2024	6	2024-06	SANTAREM	Abrantes	TRAMAGAL	Não	13	14	1401	140115	Sim
13	2024	6	2024-06	SANTAREM	Abrantes	UF ALVEGA E CONCAVADA	Não	94	14	1401	140122	Sim
14	2024	6	2024-06	SANTAREM	Abrantes	BEMPOSTA	Sim	984	14	1401	140104	Sim
15	2024	6	2024-06	SANTAREM	Abrantes	PEGO	Sim	1072	14	1401	140107	Sim
16	2024	6	2024-06	SANTAREM	Abrantes	UF ALVEGA E CONCAVADA	Sim	1252	14	1401	140122	Sim
17	2024	6	2024-06	SANTAREM	Abrantes	UF SAO FACUNDO E VALE DAS MOS	Sim	454	14	1401	140123	Sim
18	2024	6	2024-06	SANTAREM	Abrantes	UF ALDEIA DO MATO E SOUTO	Sim	676	14	1401	140121	Sim
19	2024	6	2024-06	SANTAREM	Abrantes	CARVALHAL	Sim	385	14	1401	140119	Sim
20	2024	6	2024-06	SANTAREM	Abrantes	FONTES	Sim	314	14	1401	140118	Sim
21	2024	6	2024-06	SANTAREM	Abrantes	UF S M RIO TORTO ROSSIO S TEJO	Sim	2628	14	1401	140124	Sim
22	2024	6	2024-06	SANTAREM	Abrantes	UF ABRANTES E ALFERRAREDE	Sim	10126	14	1401	140120	Sim
23	2024	6	2024-06	SANTAREM	Abrantes	RIO DE MOINHOS	Sim	633	14	1401	140108	Sim
24	2024	6	2024-06	SANTAREM	Abrantes	MARTINCHEL	Sim	246	14	1401	140105	Sim
25	2024	6	2024-06	SANTAREM	Abrantes	MOURISCAS	Sim	1189	14	1401	140106	Sim
26	2024	6	2024-06	SANTAREM	Abrantes	TRAMAGAL	Sim	1896	14	1401	140115	Sim
27	2024	6	2024-06	GUARDA	Aguiar da Beira	PENA VERDE	Não	7	9	901	090109	Sim
28	2024	6	2024-06	GUARDA	Aguiar da Beira	UF AGUIAR DA BEIRA E CORUCHE	Não	5	9	901	090114	Sim
29	2024	6	2024-06	GUARDA	Aguiar da Beira	PINHEIRO	Não	4	9	901	090110	Sim
30	2024	6	2024-06	GUARDA	Aguiar da Beira	EIRADO	Não	3	9	901	090106	Sim
31	2024	6	2024-06	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Não	6	9	901	090116	Sim
32	2024	6	2024-06	GUARDA	Aguiar da Beira	CORTICADA	Não	3	9	901	090103	Sim

Group by Municipality | Configuration



Group by

Step name: Group by Municipality

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	Municipality
2	Includes Smart Meter

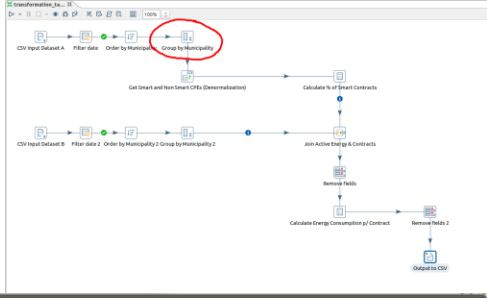
Get Fields

Aggregates :

	Name	Subject	Type
1	Number of Contracts	Number of CPE's	Sum

Get lookup fields

Group by Municipality | Preview

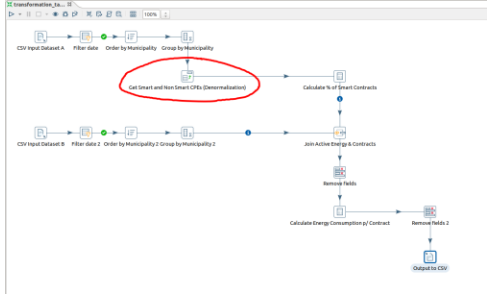


Examine preview data

Rows of step: Group by Municipality (556 rows)

	Municipality	Includes Smart Meter	Number of Contracts
1	Abrantes	Não	2808
2	Abrantes	Sim	21855
3	Aguiar da Beira	Não	30
4	Aguiar da Beira	Sim	4796
5	Alandroal	Não	772
6	Alandroal	Sim	3237
7	Albergaria-a-Velha	Não	460
8	Albergaria-a-Velha	Sim	13437
9	Albufeira	Não	3112
10	Albufeira	Sim	50215
11	Alcanena	Não	380
12	Alcanena	Sim	7629
13	Alcobaça	Não	2429
14	Alcobaça	Sim	35902
15	Alcochete	Não	8
16	Alcochete	Sim	11491
17	Alcoutim	Não	857
18	Alcoutim	Sim	2738
19	Alcácer do Sal	Não	80
20	Alcácer do Sal	Sim	8063
21	Alenquer	Não	122
22	Alenquer	Sim	24343
23	Alfândega da Fé	Não	64
24	Alfândega da Fé	Sim	4002
25	Alijó	Não	2038
26	Alijó	Sim	6889
27	Aljezur	Não	1305
28	Aljezur	Sim	5130
29	Aljustrel	Não	224
30	Aljustrel	Sim	5642
31	Almada	Não	2153
32	Almada	Sim	113047

Get Smart and Non Smart CPEs (Denormalization) - Configuration



Row denormaliser

Step name: Get Smart and Non Smart CPEs (Denormalization)

The key field: Includes Smart Meter

The fields that make up the grouping:

	Group field
1	Municipality

Get Fields

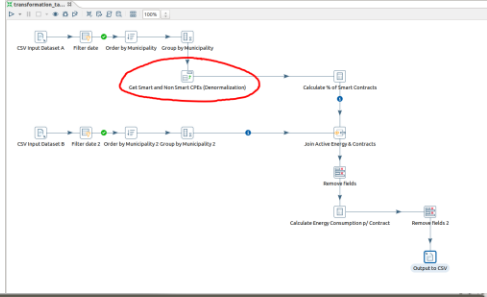
Target fields:

	Target fieldname	Value fieldname	Key value	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Aggregation
1	Number of Smart Contracts	Number of Contracts	Sim	Number			0					Sum
2	Number of Non Smart Contracts	Number of Contracts	Não	Number								Sum

Get lookup fields

Help OK Cancel

Get Smart and Non Smart CPEs (Denormalization) - Preview

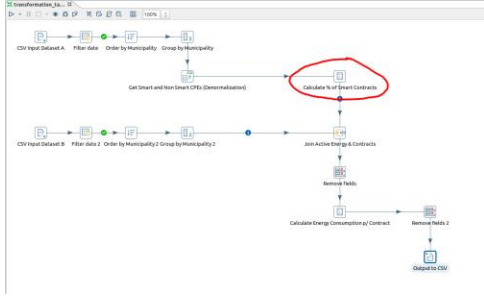


Examine preview data

Rows of step: Get Smart and Non Smart CPEs (Denormalization) (278 rows)

	Municipality	Number of Smart Contracts	Number of Non Smart Contracts
1	Abrantes	21855.0	2808.0
2	Aguiar da Beira	4796.0	30.0
3	Alandroal	3237.0	772.0
4	Albergaria-a-Velha	13437.0	460.0
5	Albufeira	50215.0	3112.0
6	Alcanena	7629.0	380.0
7	Alcobaça	35902.0	2429.0
8	Alcochete	11491.0	8.0
9	Alcoutim	2738.0	857.0
10	Alcácer do Sal	8063.0	80.0
11	Alenquer	24343.0	122.0
12	Alfândega da Fé	4002.0	64.0
13	Alijó	6889.0	2038.0
14	Aljezur	5130.0	1305.0
15	Aljustrel	5642.0	224.0
16	Almada	113047.0	2153.0
17	Almeida	6718.0	98.0
18	Almeirim	12577.0	1229.0
19	Almodôvar	5380.0	118.0
20	Alpiarça	4440.0	60.0
21	Alter do Chão	2771.0	32.0
22	Alvaiázere	5691.0	41.0
23	Alvito	1651.0	4.0
24	Amadora	96788.0	1157.0
25	Amarante	24481.0	6725.0
26	Amares	10586.0	536.0
27	Anadia	14682.0	2189.0
28	Ansião	8776.0	46.0
29	Arcos de Valdevez	17781.0	892.0
30	Arganil	7410.0	3524.0
31	Armamar	4962.0	48.0
32	Arouca	9682.0	2655.0

Calculate % of Smart Contracts | Configuration



Calculator

Step name

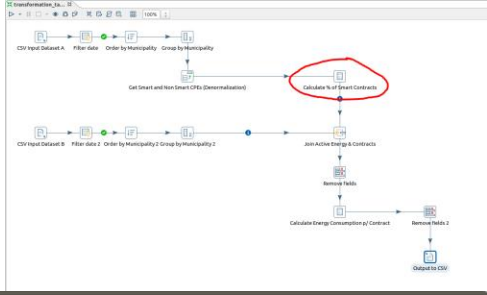
Calculate % of Smart Contracts

Throw an error on non existing files

Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conve
1	Total Contracts	A + B	Number of Smart Contracts	Number of Non Smart Contracts		Number			Y	
2	Percentage of Smart Contracts	A / B	Number of Smart Contracts	Total Contracts		Number			N	####
3	Total Contracts	Create a copy of field A	Total Contracts			Integer			N	

Calculate % of Smart Contracts | Preview

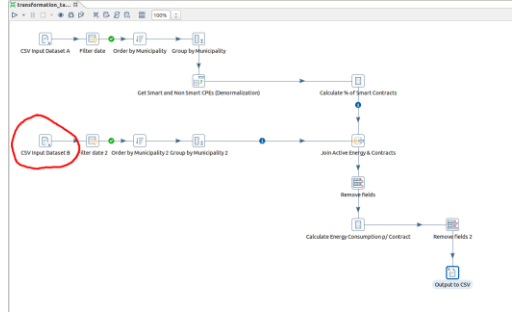


Examine preview data

Rows of step: Calculate % of Smart Contracts (278 rows)

	Municipality	Number of Smart Contracts	Number of Non Smart Contracts	Percentage of Smart Contracts	Total Contracts
1	Abrantes	21855.0	2808.0	88.61%	24663
2	Aguiar da Beira	4796.0	30.0	99.38%	4826
3	Alandroal	3237.0	772.0	80.74%	4009
4	Albergaria-a-Velha	13437.0	460.0	96.69%	13897
5	Albufeira	50215.0	3112.0	94.16%	53327
6	Alcanena	7629.0	380.0	95.26%	8009
7	Alcobaça	35902.0	2429.0	93.66%	38331
8	Alcochete	11491.0	8.0	99.93%	11499
9	Alcoutim	2738.0	857.0	76.16%	3595
10	Alcácer do Sal	8063.0	80.0	99.02%	8143
11	Alenquer	24343.0	122.0	99.5%	24465
12	Alfândega da Fé	4002.0	64.0	98.43%	4066
13	Alijó	6889.0	2038.0	77.17%	8927
14	Aljezur	5130.0	1305.0	79.72%	6435
15	Aljustrel	5642.0	224.0	96.18%	5866
16	Almada	113047.0	2153.0	98.13%	115200
17	Almeida	6718.0	98.0	98.56%	6816
18	Almeirim	12577.0	1229.0	91.1%	13806
19	Almodôvar	5380.0	118.0	97.85%	5498
20	Alpiarça	4440.0	60.0	98.67%	4500
21	Alter do Chão	2771.0	32.0	98.86%	2803
22	Alvaiázere	5691.0	41.0	99.28%	5732
23	Alvito	1651.0	4.0	99.76%	1655
24	Amadora	96788.0	1157.0	98.82%	97945
25	Amarante	24481.0	6725.0	78.45%	31206
26	Amares	10586.0	536.0	95.18%	11122
27	Anadia	14682.0	2189.0	87.03%	16871
28	Ansião	8776.0	46.0	99.48%	8822
29	Arcos de Valdevez	17781.0	892.0	95.22%	18673
30	Arganil	7410.0	3524.0	67.77%	10934
31	Armamar	4962.0	48.0	99.04%	5010
32	Arouca	9682.0	2655.0	78.48%	12337

CSV Input Dataset B | Configuration



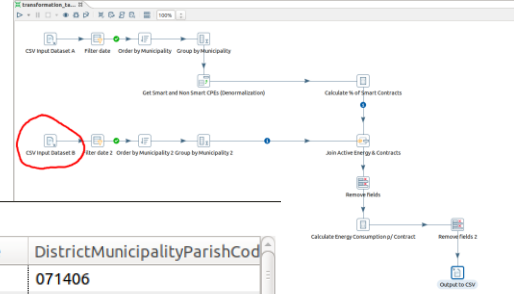
CSV file input

Step name	CSV Input Dataset B
Filename	/home/aid/adi-project/data/3-consumos-faturados-por-municipio-ultimos-10-anos.c
Delimiter	;
Enclosure	"
NIO buffer size	50000
Lazy conversion?	☐
Header row present?	☑
Add filename to result	☐
The row number field name (optional)	
Running in parallel?	☐
New line possible in fields?	☐
Format	mixed
File encoding	UTF-8

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Year	Integer	#	15	0	$	.	,	none
2	Month	Integer	#	15	0	$	.	,	none
3	Date	String		7		$	.	,	none
4	District	String		16		$	.	,	none
5	Municipality	String		22		$	.	,	none
6	parish	String		30		$	.	,	none
7	Voltage level	String		32		$	.	,	none
8	Active Energy (kWh)	Number	#.#	11	3	$	.	,	none
9	DistrictCode	Integer	#	15	0	$	.	,	none
10	DistrictMunicipalityCode	Integer	#	15	0	$	.	,	none
11	DistrictMunicipalityParishCode	String		6		$	.	,	none
12	mes_int	Integer	#	15	0	$	.	,	none

Help | OK | Get Fields | Preview | Cancel

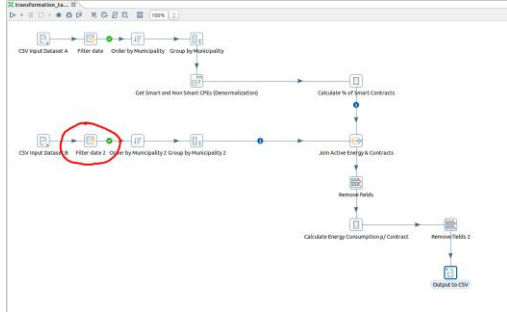
CSV Input Dataset B | Preview



Rows of step: CSV Input Dataset B (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCod
1	2020	11	2020-11	Évora	Vila Viçosa	N S CONCEICAO E SAO BARTOLOMEU	Baixa Tensão	932900.4	7	0714	071406
2	2020	11	2020-11	Faro	Albufeira	PADERNE	Muito Alta, Alta e Média Tensões	300614.9	8	0801	080103
3	2020	11	2020-11	Faro	Alcoutim	GIOES	Baixa Tensão	33531.7	8	0802	080202
4	2020	11	2020-11	Faro	Alcoutim	UF ALCOUTIM E PEREIRO	Baixa Tensão	315064.6	8	0802	080206
5	2020	11	2020-11	Faro	Alcoutim	UF ALCOUTIM E PEREIRO	Muito Alta, Alta e Média Tensões	14959.2	8	0802	080206
6	2020	11	2020-11	Faro	Aljezur	ODECEIXE	Baixa Tensão	251829.5	8	0803	080303
7	2020	11	2020-11	Faro	Castro Marim	AZINHAL	Baixa Tensão	108590.8	8	0804	080401
8	2020	11	2020-11	Faro	Faro	SANTA BARBARA DE NEXE	Muito Alta, Alta e Média Tensões	90187.8	8	0805	080503
9	2020	11	2020-11	Faro	Faro	UF CONCEICAO E ESTOI	Baixa Tensão	2021643.6	8	0805	080507
10	2020	11	2020-11	Faro	Lagoa	UF ESTOMBAR E PARCHAL	Muito Alta, Alta e Média Tensões	704903.9	8	0806	080607
11	2020	11	2020-11	Faro	Lagos	ODIAXERE	Baixa Tensão	733281.9	8	0807	080704
12	2020	11	2020-11	Faro	Loulé	LOULE (SAO CLEMENTE)	Muito Alta, Alta e Média Tensões	605332.8	8	0808	080808
13	2020	11	2020-11	Faro	Loulé	UF QUERENCA TOR E BENAFIM	Muito Alta, Alta e Média Tensões	43821.8	8	0808	080812
14	2020	11	2020-11	Faro	Olhão	OLHAO	Baixa Tensão	2341014.4	8	0810	081003
15	2020	11	2020-11	Faro	São Brás de Alportel	SAO BRAS DE ALPORTEL	Baixa Tensão	2365711.4	8	0812	081201
16	2020	11	2020-11	Faro	Vila do Bispo	SAGRES	Muito Alta, Alta e Média Tensões	250120.3	8	0815	081504
17	2020	11	2020-11	Faro	Vila Real de Santo António	VILA NOVA DE CACELA	Baixa Tensão	1309636.7	8	0816	081601
18	2020	11	2020-11	Guarda	Aguiar da Beira	PINHEIRO	Baixa Tensão	58706.4	9	0901	090110
19	2020	11	2020-11	Guarda	Almeida	MALHADA SORDA	Baixa Tensão	47519.8	9	0902	090213
20	2020	11	2020-11	Guarda	Almeida	UF JUNCA E NAVES	Baixa Tensão	32638.8	9	0902	090233
21	2020	11	2020-11	Guarda	Celorico da Beira	MESQUITELA	Baixa Tensão	30188.4	9	0903	090310
22	2020	11	2020-11	Guarda	Figueira de Castelo Rodrigo	UF COLMEAL E VILAR TORPIM	Muito Alta, Alta e Média Tensões	4650.2	9	0904	090422
23	2020	11	2020-11	Guarda	Fornos de Algodres	MUXAGATA	Baixa Tensão	28518.2	9	0905	090511
24	2020	11	2020-11	Guarda	Gouveia	SAO PAIO	Muito Alta, Alta e Média Tensões	20579.8	9	0906	090617
25	2020	11	2020-11	Guarda	Gouveia	UF ALDEIAS MANGUALDE SERRA	Muito Alta, Alta e Média Tensões	96292.7	9	0906	090623
26	2020	11	2020-11	Guarda	Gouveia	UF MELO E NABAIS	Baixa Tensão	130899.2	9	0906	090626
27	2020	11	2020-11	Guarda	Guarda	GONCALO BOCAS	Baixa Tensão	41909.1	9	0907	090721
28	2020	11	2020-11	Guarda	Guarda	PANOIAS DE CIMA	Baixa Tensão	98965.9	9	0907	090728
29	2020	11	2020-11	Guarda	Guarda	GUARDA	Muito Alta, Alta e Média Tensões	3676765	9	0907	090758
30	2020	11	2020-11	Guarda	Guarda	UF MIZARELA P SOARES V SOEIRO	Muito Alta, Alta e Média Tensões	248352	9	0907	090763
31	2020	11	2020-11	Guarda	Manteigas	MANTEIGAS (SANTA MARIA)	Baixa Tensão	351592.8	9	0908	090802
32	2020	11	2020-11	Guarda	Pinhel	ALVERCA DA BEIRA/BOUCA COVA	Baixa Tensão	77498.9	9	0910	091029
33	2020	11	2020-11	Guarda	Sabugal	BENDADA	Baixa Tensão	77998	9	0911	091110
34	2020	11	2020-11	Guarda	Sabugal	SOUTO	Baixa Tensão	228642.3	9	0911	091134
35	2020	11	2020-11	Guarda	Seia	SANTIAGO	Baixa Tensão	168645.2	9	0912	091216
36	2020	11	2020-11	Guarda	Seia	VALEZIM	Baixa Tensão	39821.3	9	0912	091225
37	2020	11	2020-11	Guarda	Trancoso	CASTANHEIRA	Baixa Tensão	19354.3	9	0913	091303
38	2020	11	2020-11	Guarda	Trancoso	POVOA DO CONCELHO	Baixa Tensão	34637.1	9	0913	091314
39	2020	11	2020-11	Guarda	Trancoso	VALDUJO	Baixa Tensão	32161.9	9	0913	091325
40	2020	11	2020-11	Guarda	Trancoso	UF TRANCOSO E SOUTO MAIOR	Baixa Tensão	790462.7	9	0913	091332
41	2020	11	2020-11	Leiria	Alcobaça	BARRIO	Baixa Tensão	271538.3	10	1001	100104
42	2020	11	2020-11	Leiria	Alcobaça	CELA	Baixa Tensão	671277.3	10	1001	100106
43	2020	11	2020-11	Leiria	Alcobaça	VIMEIRO	Muito Alta, Alta e Média Tensões	104399.9	10	1001	100116
44	2020	11	2020-11	Leiria	Alcobaça	ALJUBARROTA	Baixa Tensão	1002307.8	10	1001	100120

Filter Date 2 | Configuration



Filter rows

Step name: Filter date 2

Send 'true' data to step: Order by Municipality 2

Send 'false' data to step:

The condition:

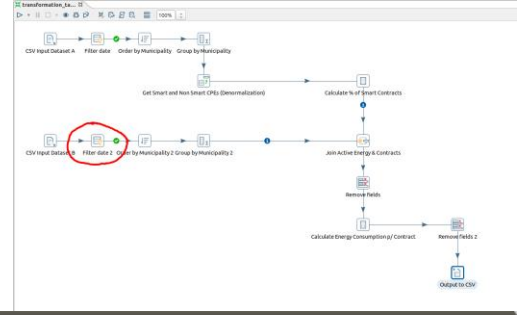
```
Year = [2024]

AND

Month = [6]
```

Help | OK | Cancel

Filter Date 2 | Preview

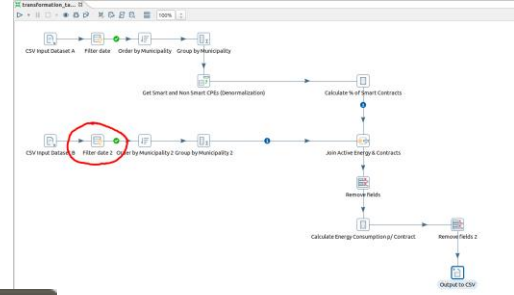


Examine preview data

Rows of step: Filter date 2 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipal	DistrictMunicipalityParishCode	mes_int
1	2024	6	2024-06	BEJA	Mértola	CORTE DO PINTO	Baixa Tensão	59617	2	209	020902	6
2	2024	6	2024-06	BRAGA	Fafe	AROES (SAO ROMAO)	Muito Alta, Alta e Média Tensões	21795	3	307	030730	6
3	2024	6	2024-06	BRAGA	Vila Nova de Famalicão	RIBA DE AVE	Baixa Tensão	179626	3	312	031234	6
4	2024	6	2024-06	CASTELO BRANCO	Castelo Branco	UF FREIXIAL E JUNCAL DO CAMPO	Baixa Tensão	39445	5	502	050229	6
5	2024	6	2024-06	GUARDA	Celorico da Beira	MINHOCAL	Baixa Tensão	8908	9	903	090311	6
6	2024	6	2024-06	VILA REAL	Montalegre	UF MONTALEGRE E PADROSO	Muito Alta, Alta e Média Tensões	35658	17	1706	170638	6
7	2024	6	2024-06	BRAGA	Vila Nova de Famalicão	UF ESMERIZ E CABECUDOS	Muito Alta, Alta e Média Tensões	174709	3	312	031254	6
8	2024	6	2024-06	EVORA	Mora	PAVIA	Baixa Tensão	78059	7	707	070704	6
9	2024	6	2024-06	PORTO	Vila do Conde	UF VILAR E MOSTEIRO	Baixa Tensão	172450	13	1316	131637	6
10	2024	6	2024-06	SANTAREM	Alcanena	MOITAS VENDA	Baixa Tensão	40326	14	1402	140207	6
11	2024	6	2024-06	VIANA DO CASTELO	Ponte de Lima	CALVELO	Baixa Tensão	43187	16	1607	160714	6
12	2024	6	2024-06	VILA REAL	Ribeira de Pena	UF RIBEIRA PENA SA ALEM TAMEGA	Baixa Tensão	203438	17	1709	170909	6
13	2024	6	2024-06	AVEIRO	Oliveira de Azeméis	CARREGOSA	Muito Alta, Alta e Média Tensões	250906	1	113	011301	6
14	2024	6	2024-06	BEJA	Odemira	SABOIA	Baixa Tensão	49546	2	211	021103	6
15	2024	6	2024-06	EVORA	Mourão	LUZ	Baixa Tensão	24176	7	708	070802	6
16	2024	6	2024-06	GUARDA	Trancoso	ALDEIA NOVA	Baixa Tensão	16933	9	913	091301	6
17	2024	6	2024-06	BEJA	Odemira	SAO SALVADOR E SANTA MARIA	Baixa Tensão	203390	2	211	021120	6
18	2024	6	2024-06	PORTALEGRE	Marvão	SANTO ANTONIO DAS AREIAS	Baixa Tensão	69556	12	1210	121003	6
19	2024	6	2024-06	PORTALEGRE	Ponte de Sor	LONGOMEL	Muito Alta, Alta e Média Tensões	2338	12	1213	121305	6
20	2024	6	2024-06	BRAGA	Guimarães	UF AIRAO S M AIRAO S J VERMIL	Muito Alta, Alta e Média Tensões	85846	3	308	030876	6
21	2024	6	2024-06	CASTELO BRANCO	Covilhã	SAO JORGE DA BEIRA	Baixa Tensão	42656	5	503	050318	6
22	2024	6	2024-06	COIMBRA	Tábua	UF COVAS VILA NOVA OLIVEIRINHA	Baixa Tensão	66740	6	616	061617	6
23	2024	6	2024-06	GUARDA	Trancoso	CASTANHEIRA	Baixa Tensão	5899	9	913	091303	6
24	2024	6	2024-06	LISBOA	Torres Vedras	TURCIFAL	Muito Alta, Alta e Média Tensões	597377	11	1113	111317	6
25	2024	6	2024-06	VISEU	Mangualde	UF TAVARES	Baixa Tensão	80469	18	1806	180622	6
26	2024	6	2024-06	VISEU	Sernancelhe	VILA DA PONTE	Muito Alta, Alta e Média Tensões	5178	18	1818	181817	6
27	2024	6	2024-06	VISEU	Vila Nova de Paiva	UF V NOVA PAIVA ALHAIS FRAGUAS	Baixa Tensão	145400	18	1822	182208	6
28	2024	6	2024-06	VISEU	Viseu	CALDE	Muito Alta, Alta e Média Tensões	5615	18	1823	182305	6
29	2024	6	2024-06	VISEU	Viseu	RIO DE LOBA	Baixa Tensão	507699	18	1823	182323	6
30	2024	6	2024-06	BEJA	Beja	CABECA GORDA	Baixa Tensão	82636	2	205	020504	6
31	2024	6	2024-06	BRAGA	Fafe	FORNELOS	Baixa Tensão	86722	3	307	030712	6
32	2024	6	2024-06	BRAGA	Guimarães	UF B S SALVADOR E B S LEOCADIA	Muito Alta, Alta e Média Tensões	72856	3	308	030880	6

Order by Municipality 2 | Configuration



Sort rows

Step name: Order by Municipality 2

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

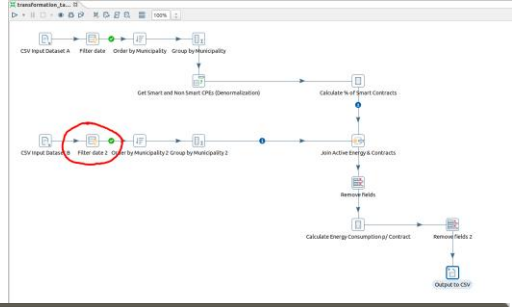
Only pass unique rows? (verifies keys only) ☐

Fields :

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

Help OK Cancel Get Fields

Order by Municipality 2 | Preview

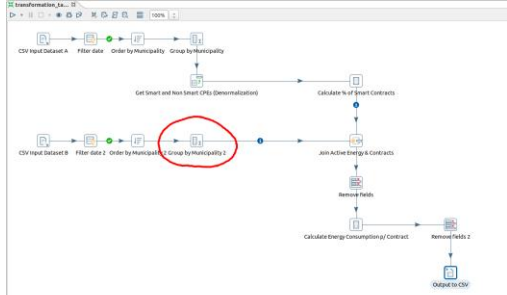


Examine preview data

Rows of step: Order by Municipality 2 (1000 rows)

	Year	Month	Date	District	Municipality	parish	Voltage level	Active Energy (kWh)	DistrictCode	DistrictMunicipalityCode	DistrictMunicipalityParishCod	mes_int
1	2024	6	2024-06	SANTAREM	Abrantes	UF ALDEIA DO MATO E SOUTO	Baixa Tensão	74113	14	1401	140121	6
2	2024	6	2024-06	SANTAREM	Abrantes	BEMPOSTA	Baixa Tensão	98297	14	1401	140104	6
3	2024	6	2024-06	SANTAREM	Abrantes	MARTINCHEL	Baixa Tensão	31292	14	1401	140105	6
4	2024	6	2024-06	SANTAREM	Abrantes	TRAMAGAL	Baixa Tensão	179525	14	1401	140115	6
5	2024	6	2024-06	SANTAREM	Abrantes	UF SAO FACUNDO E VALE DAS MOS	Baixa Tensão	73489	14	1401	140123	6
6	2024	6	2024-06	SANTAREM	Abrantes	CARVALHAL	Baixa Tensão	35760	14	1401	140119	6
7	2024	6	2024-06	SANTAREM	Abrantes	UF S M RIO TORTO ROSSIO S TEJO	Muito Alta, Alta e Média Tensões	26813	14	1401	140124	6
8	2024	6	2024-06	SANTAREM	Abrantes	TRAMAGAL	Muito Alta, Alta e Média Tensões	252392	14	1401	140115	6
9	2024	6	2024-06	SANTAREM	Abrantes	RIO DE MOINHOS	Baixa Tensão	53171	14	1401	140108	6
10	2024	6	2024-06	SANTAREM	Abrantes	UF ABRANTES E ALFERRAREDE	Baixa Tensão	1080350	14	1401	140120	6
11	2024	6	2024-06	SANTAREM	Abrantes	MOURISCAS	Muito Alta, Alta e Média Tensões	17700	14	1401	140106	6
12	2024	6	2024-06	SANTAREM	Abrantes	PEGO	Muito Alta, Alta e Média Tensões	752674	14	1401	140107	6
13	2024	6	2024-06	SANTAREM	Abrantes	UF ALVEGA E CONCAVADA	Muito Alta, Alta e Média Tensões	22276	14	1401	140122	6
14	2024	6	2024-06	SANTAREM	Abrantes	FONTES	Baixa Tensão	35775	14	1401	140118	6
15	2024	6	2024-06	SANTAREM	Abrantes	UF ALDEIA DO MATO E SOUTO	Muito Alta, Alta e Média Tensões	46232	14	1401	140121	6
16	2024	6	2024-06	SANTAREM	Abrantes	UF ALVEGA E CONCAVADA	Baixa Tensão	101749	14	1401	140122	6
17	2024	6	2024-06	SANTAREM	Abrantes	UF S M RIO TORTO ROSSIO S TEJO	Baixa Tensão	258031	14	1401	140124	6
18	2024	6	2024-06	SANTAREM	Abrantes	MOURISCAS	Baixa Tensão	120885	14	1401	140106	6
19	2024	6	2024-06	SANTAREM	Abrantes	PEGO	Baixa Tensão	145601	14	1401	140107	6
20	2024	6	2024-06	SANTAREM	Abrantes	UF ABRANTES E ALFERRAREDE	Muito Alta, Alta e Média Tensões	852859	14	1401	140120	6
21	2024	6	2024-06	SANTAREM	Abrantes	BEMPOSTA	Muito Alta, Alta e Média Tensões	42727	14	1401	140104	6
22	2024	6	2024-06	GUARDA	Aguiar da Beira	FORNINHOS	Baixa Tensão	13033	9	901	090107	6
23	2024	6	2024-06	GUARDA	Aguiar da Beira	UF SEQUEIROS E GRADIZ	Baixa Tensão	33783	9	901	090115	6
24	2024	6	2024-06	GUARDA	Aguiar da Beira	PENA VERDE	Baixa Tensão	43286	9	901	090109	6
25	2024	6	2024-06	GUARDA	Aguiar da Beira	UF AGUIAR DA BEIRA E CORUCHE	Muito Alta, Alta e Média Tensões	61114	9	901	090114	6
26	2024	6	2024-06	GUARDA	Aguiar da Beira	DORNELAS	Baixa Tensão	32570	9	901	090105	6
27	2024	6	2024-06	GUARDA	Aguiar da Beira	PINHEIRO	Baixa Tensão	19536	9	901	090110	6
28	2024	6	2024-06	GUARDA	Aguiar da Beira	UF SOUTO AGUIAR BEIRA VALVERDE	Baixa Tensão	38038	9	901	090116	6
29	2024	6	2024-06	GUARDA	Aguiar da Beira	UF AGUIAR DA BEIRA E CORUCHE	Baixa Tensão	151170	9	901	090114	6
30	2024	6	2024-06	GUARDA	Aguiar da Beira	CORTICADA	Baixa Tensão	14611	9	901	090103	6
31	2024	6	2024-06	GUARDA	Aguiar da Beira	EIRADO	Baixa Tensão	15369	9	901	090106	6
32	2024	6	2024-06	GUARDA	Aguiar da Beira	CARAPITO	Baixa Tensão	34577	9	901	090102	6

Group by Municipality 2 | Configuration



Group by

Step name: Group by Municipality 2

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	Municipality

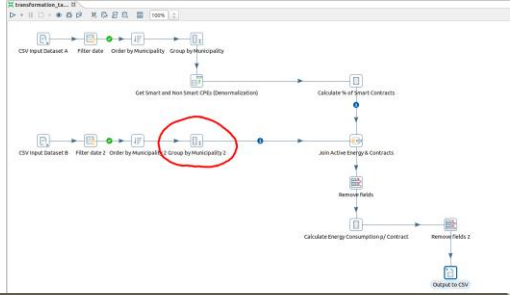
Get Fields

Aggregates :

	Name	Subject	Type
1	Active Energy (kWh)	Active Energy (kWh)	Sum

Get lookup fields

Group by Municipality 2 | Preview

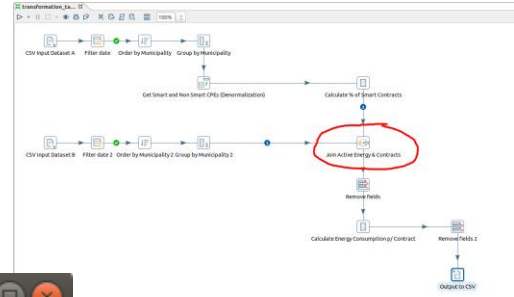


Examine preview data

Rows of step: Group by Municipality 2 (278 rows)

	Municipality	Active Energy (kWh)
1	Abrantes	4301711.0
2	Aguiar da Beira	457087.0
3	Alandroal	827982.0
4	Albergaria-a-Velha	3704565.0
5	Albufeira	11996087.0
6	Alcanena	1959884.0
7	Alcobaça	8415602.0
8	Alcochete	2596452.0
9	Alcoutim	220503.0
10	Alcácer do Sal	2333015.0
11	Alenquer	4757070.0
12	Alfândega da Fé	302471.0
13	Alijó	800888.0
14	Aljezur	773866.0
15	Aljustrel	1607215.0
16	Almada	12926952.0
17	Almeida	547375.0
18	Almeirim	2279284.0
19	Almodôvar	662982.0
20	Alpiarça	990115.0
21	Alter do Chão	318636.0
22	Alvaiázere	525870.0
23	Alvito	538495.0
24	Amadora	10614979.0
25	Amarante	3723051.0
26	Amares	1621765.0
27	Anadia	3150888.0
28	Ansião	1249714.0
29	Arcos de Valdevez	2387321.0
30	Arganil	1003398.0
31	Armamar	677031.0
32	Arouca	1807586.0

Join Active Energy & Contracts | Configuration



Merge join

Step name: Join Active Energy & Contracts

First Step: Group by Municipality 2

Second Step: Calculate % of Smart Contracts

Join Type: INNER

Keys for 1st step:

	Key field
1	Municipality

Keys for 2nd step:

	Key field
1	Municipality

Get key fields

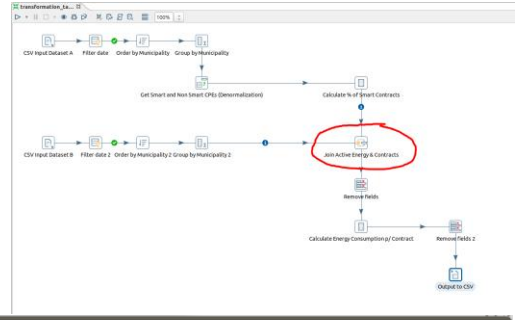
Get key fields

Help

OK

Cancel

Join Active Energy & Contracts | Preview

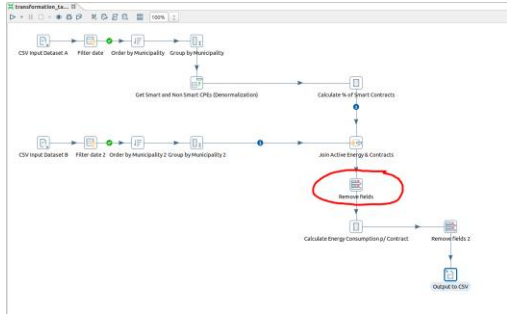


Examine preview data

Rows of step: Join Active Energy Contracts (278 rows)

	Municipality	Active Energy (kWh)	Municipality_1	Number of Smart Contracts	Number of Non Smart Contracts	Percentage of Smart Contracts	Total Contracts
1	Abrantes	4301711.0	Abrantes	21855.0	2808.0	88.61%	24663
2	Aguiar da Beira	457087.0	Aguiar da Beira	4796.0	30.0	99.38%	4826
3	Alandroal	827982.0	Alandroal	3237.0	772.0	80.74%	4009
4	Albergaria-a-Velha	3704565.0	Albergaria-a-Velha	13437.0	460.0	96.69%	13897
5	Albufeira	11996087.0	Albufeira	50215.0	3112.0	94.16%	53327
6	Alcanena	1959884.0	Alcanena	7629.0	380.0	95.26%	8009
7	Alcobaça	8415602.0	Alcobaça	35902.0	2429.0	93.66%	38331
8	Alcochete	2596452.0	Alcochete	11491.0	8.0	99.93%	11499
9	Alcoutim	220503.0	Alcoutim	2738.0	857.0	76.16%	3595
10	Alcácer do Sal	2333015.0	Alcácer do Sal	8063.0	80.0	99.02%	8143
11	Alenquer	4757070.0	Alenquer	24343.0	122.0	99.5%	24465
12	Alfândega da Fé	302471.0	Alfândega da Fé	4002.0	64.0	98.43%	4066
13	Alijó	800888.0	Alijó	6889.0	2038.0	77.17%	8927
14	Aljezur	773866.0	Aljezur	5130.0	1305.0	79.72%	6435
15	Aljustrel	1607215.0	Aljustrel	5642.0	224.0	96.18%	5866
16	Almada	12926952.0	Almada	113047.0	2153.0	98.13%	115200
17	Almeida	547375.0	Almeida	6718.0	98.0	98.56%	6816
18	Almeirim	2279284.0	Almeirim	12577.0	1229.0	91.1%	13806
19	Almodôvar	662982.0	Almodôvar	5380.0	118.0	97.85%	5498
20	Alpiarça	990115.0	Alpiarça	4440.0	60.0	98.67%	4500
21	Alter do Chão	318636.0	Alter do Chão	2771.0	32.0	98.86%	2803
22	Alvaiázere	525870.0	Alvaiázere	5691.0	41.0	99.28%	5732
23	Alvito	538495.0	Alvito	1651.0	4.0	99.76%	1655
24	Amadora	10614979.0	Amadora	96788.0	1157.0	98.82%	97945
25	Amarante	3723051.0	Amarante	24481.0	6725.0	78.45%	31206
26	Amares	1621765.0	Amares	10586.0	536.0	95.18%	11122
27	Anadia	3150888.0	Anadia	14682.0	2189.0	87.03%	16871
28	Ansião	1249714.0	Ansião	8776.0	46.0	99.48%	8822
29	Arcos de Valdevez	2387321.0	Arcos de Valdevez	17781.0	892.0	95.22%	18673
30	Arganil	1003398.0	Arganil	7410.0	3524.0	67.77%	10934
31	Armamar	677031.0	Armamar	4962.0	48.0	99.04%	5010
32	Arouca	1807586.0	Arouca	9682.0	2655.0	78.48%	12337

Remove Fields | Configuration



Select values

Step name: Remove fields

Select & Alter | Remove | Meta-data

Fields to remove :

	Fieldname
1	Municipality_1
2	Number of Non Smart Contracts
3	Number of Smart Contracts

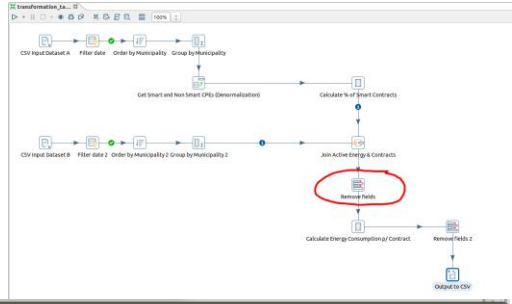
Get fields to remove

Help

OK

Cancel

Remove Fields | Preview

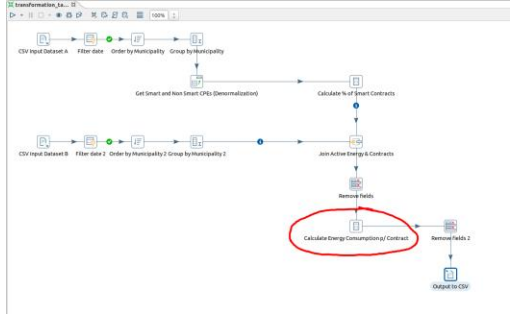


Examine preview data

Rows of step: Remove fields (278 rows)

	Municipality	Active Energy (kWh)	Percentage of Smart Contracts	Total Contracts
1	Abrantes	4301711.0	88.61%	24663
2	Aguiar da Beira	457087.0	99.38%	4826
3	Alandroal	827982.0	80.74%	4009
4	Albergaria-a-Velha	3704565.0	96.69%	13897
5	Albufeira	11996087.0	94.16%	53327
6	Alcanena	1959884.0	95.26%	8009
7	Alcobaça	8415602.0	93.66%	38331
8	Alcochete	2596452.0	99.93%	11499
9	Alcoutim	220503.0	76.16%	3595
10	Alcácer do Sal	2333015.0	99.02%	8143
11	Alenquer	4757070.0	99.5%	24465
12	Alfândega da Fé	302471.0	98.43%	4066
13	Alijó	800888.0	77.17%	8927
14	Aljezur	773866.0	79.72%	6435
15	Aljustrel	1607215.0	96.18%	5866
16	Almada	12926952.0	98.13%	115200
17	Almeida	547375.0	98.56%	6816
18	Almeirim	2279284.0	91.1%	13806
19	Almodôvar	662982.0	97.85%	5498
20	Alpiarça	990115.0	98.67%	4500
21	Alter do Chão	318636.0	98.86%	2803
22	Alvaiázere	525870.0	99.28%	5732
23	Alvito	538495.0	99.76%	1655
24	Amadora	10614979.0	98.82%	97945
25	Amarante	3723051.0	78.45%	31206
26	Amares	1621765.0	95.18%	11122
27	Anadia	3150888.0	87.03%	16871
28	Ansião	1249714.0	99.48%	8822
29	Arcos de Valdevez	2387321.0	95.22%	18673
30	Arganil	1003398.0	67.77%	10934
31	Armamar	677031.0	99.04%	5010
32	Arouca	1807586.0	78.48%	12337

Calculate Energy Consumption p/ Contract
-
Configuration



Calculator

Step name

Calculate Energy Consumption p/ Contract

Throw an error on non existing files

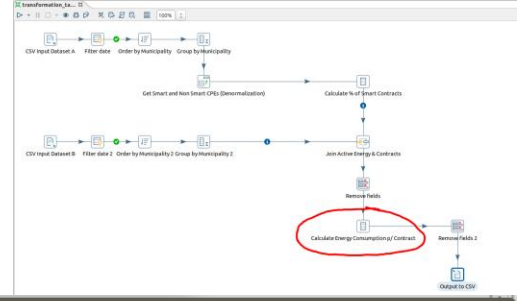
Fields:

	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Deci
1	Average Energy Consumption p/ Contract (kWh)	A / B	Active Energy (kWh)	Total Contracts		Number			N	########.##	

Calculate Energy Consumption p/ Contract

-

Preview

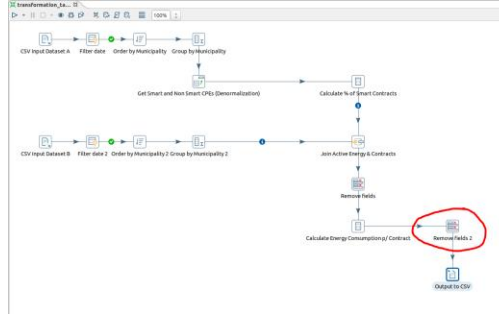


Examine preview data

Rows of step: Calculate Energy Consumption p/ Contract (278 rows)

	Municipality	Active Energy (kWh)	Percentage of Smart Contracts	Total Contracts	Average Energy Consumption p/ Contract (kWh)
1	Abrantes	4301711.0	88.61%	24663	174.42
2	Aguiar da Beira	457087.0	99.38%	4826	94.71
3	Alandroal	827982.0	80.74%	4009	206.53
4	Albergaria-a-Velha	3704565.0	96.69%	13897	266.57
5	Albufeira	11996087.0	94.16%	53327	224.95
6	Alcanena	1959884.0	95.26%	8009	244.71
7	Alcobaça	8415602.0	93.66%	38331	219.55
8	Alcochete	2596452.0	99.93%	11499	225.8
9	Alcoutim	220503.0	76.16%	3595	61.34
10	Alcácer do Sal	2333015.0	99.02%	8143	286.51
11	Alenquer	4757070.0	99.5%	24465	194.44
12	Alfândega da Fé	302471.0	98.43%	4066	74.39
13	Alijó	800888.0	77.17%	8927	89.72
14	Aljezur	773866.0	79.72%	6435	120.26
15	Aljustrel	1607215.0	96.18%	5866	273.99
16	Almada	12926952.0	98.13%	115200	112.21
17	Almeida	547375.0	98.56%	6816	80.31
18	Almeirim	2279284.0	91.1%	13806	165.09
19	Almodôvar	662982.0	97.85%	5498	120.59
20	Alpiarça	990115.0	98.67%	4500	220.03
21	Alter do Chão	318636.0	98.86%	2803	113.68
22	Alvaiázere	525870.0	99.28%	5732	91.74
23	Alvito	538495.0	99.76%	1655	325.37
24	Amadora	10614979.0	98.82%	97945	108.38
25	Amarante	3723051.0	78.45%	31206	119.31
26	Amares	1621765.0	95.18%	11122	145.82
27	Anadia	3150888.0	87.03%	16871	186.76
28	Ansião	1249714.0	99.48%	8822	141.66
29	Arcos de Valdevez	2387321.0	95.22%	18673	127.85
30	Arganil	1003398.0	67.77%	10934	91.77
31	Armamar	677031.0	99.04%	5010	135.14
32	Arouca	1807586.0	78.48%	12337	146.52

Remove Fields 2 | Configuration



Select values

Step name: Remove fields 2

Select & Alter | Remove | Meta-data

Fields to remove :

	Fieldname
1	Active Energy (kWh)
2	Total Contracts

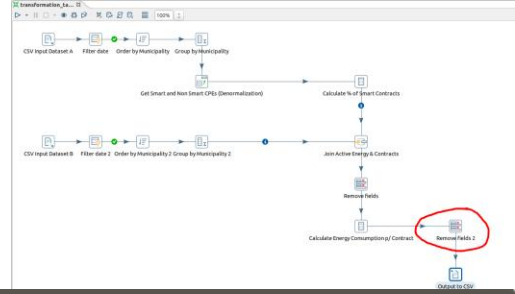
Get fields to remove

Help

OK

Cancel

Remove Fields 2 | Preview

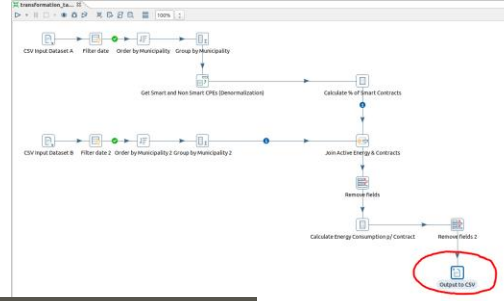


Examine preview data

Rows of step: Remove fields 2 (278 rows)

	Municipality	Percentage of Smart Contracts	Average Energy Consumption p/ Contract (kWh)
1	Abrantes	88.61%	174.42
2	Aguiar da Beira	99.38%	94.71
3	Alandroal	80.74%	206.53
4	Albergaria-a-Velha	96.69%	266.57
5	Albufeira	94.16%	224.95
6	Alcanena	95.26%	244.71
7	Alcobaça	93.66%	219.55
8	Alcochete	99.93%	225.8
9	Alcoutim	76.16%	61.34
10	Alcácer do Sal	99.02%	286.51
11	Alenquer	99.5%	194.44
12	Alfândega da Fé	98.43%	74.39
13	Alijó	77.17%	89.72
14	Aljezur	79.72%	120.26
15	Aljustrel	96.18%	273.99
16	Almada	98.13%	112.21
17	Almeida	98.56%	80.31
18	Almeirim	91.1%	165.09
19	Almodôvar	97.85%	120.59
20	Alpiarça	98.67%	220.03
21	Alter do Chão	98.86%	113.68
22	Alvaiázere	99.28%	91.74
23	Alvito	99.76%	325.37
24	Amadora	98.82%	108.38
25	Amarante	78.45%	119.31
26	Amares	95.18%	145.82
27	Anadia	87.03%	186.76
28	Ansião	99.48%	141.66
29	Arcos de Valdevez	95.22%	127.85
30	Arganil	67.77%	91.77
31	Armamar	99.04%	135.14
32	Arouca	78.48%	146.52

Output to CSV | Configuration (1/3)



Text file output

Step name: Output to CSV

File | Content | Fields

Filename: /home/aid/adi-project/delivery/output/task_1_output [Browse...]

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 ☐

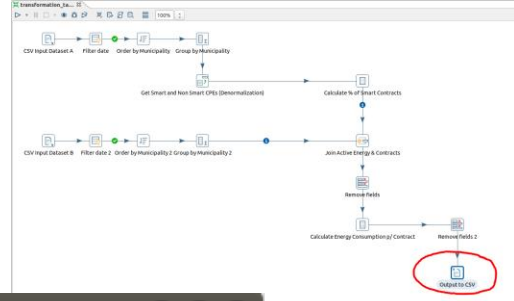
Date time format:

Show filename(s)...

Add filenames to result ☑

Help | OK | Cancel

Output to CSV | Configuration (2/3)



Text file output

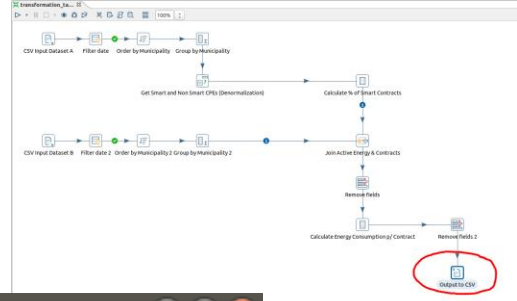
Step name: Output to CSV

File | Content | Fields

- Append: ☐
- Separator: ; — Insert TAB
- Enclosure: "
- Force the enclosure around fields?: ☐
- Disable the enclosure fix?: ☐
- Header: ☑
- Footer: ☐
- Format: CR+LF terminated (Windows, DOS)
- Compression: None
- Encoding: UTF-8
- Right pad fields: ☐
- Fast data dump (no formatting): ☐
- Split every ... rows:
- Add Ending line of file:

Help | OK | Cancel

Output to CSV | Configuration (3/3)



Text file output

Step name: Output to CSV

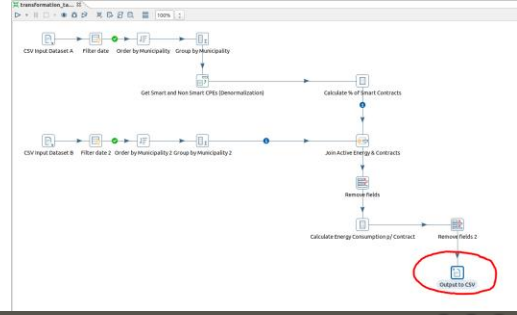
File | Content | Fields

	Name	Type	Format	Length	Precision	Currency	Decim
1	Municipality	String		22			
2	Percentage of Smart Contracts	Number	#.####				
3	Average Energy Consumption p/ Contract (kWh)	Number	########.##				

Get Fields | Minimal width

Help | OK | Cancel

Output to CSV | Preview



Examine preview data

Rows of step: Output to CSV (278 rows)

	Municipality	Percentage of Smart Contracts	Average Energy Consumption p/ Contract (kWh)
1	Abrantes	88.61%	174.42
2	Aguiar da Beira	99.38%	94.71
3	Alandroal	80.74%	206.53
4	Albergaria-a-Velha	96.69%	266.57
5	Albufeira	94.16%	224.95
6	Alcanena	95.26%	244.71
7	Alcobaça	93.66%	219.55
8	Alcochete	99.93%	225.8
9	Alcoutim	76.16%	61.34
10	Alcácer do Sal	99.02%	286.51
11	Alenquer	99.5%	194.44
12	Alfândega da Fé	98.43%	74.39
13	Alijó	77.17%	89.72
14	Aljezur	79.72%	120.26
15	Aljustrel	96.18%	273.99
16	Almada	98.13%	112.21
17	Almeida	98.56%	80.31
18	Almeirim	91.1%	165.09
19	Almodôvar	97.85%	120.59
20	Alpiarça	98.67%	220.03
21	Alter do Chão	98.86%	113.68
22	Alvaiázere	99.28%	91.74
23	Alvito	99.76%	325.37
24	Amadora	98.82%	108.38
25	Amarante	78.45%	119.31
26	Amares	95.18%	145.82
27	Anadia	87.03%	186.76
28	Ansião	99.48%	141.66
29	Arcos de Valdevez	95.22%	127.85
30	Arganil	67.77%	91.77
31	Armamar	99.04%	135.14
32	Arouca	78.48%	146.52

Output File

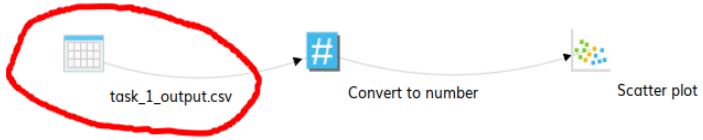
Open | task_1_output.csv | ~/adi-project/delivery/output | Save

```
Municipality;Percentage of Smart Contracts;Average Energy Consumption p/ Contract (kWh)
Abrantes              ;0.8861;174.42
Aguiar da Beira       ;0.9938;94.71
Alandroal             ;0.8074;206.53
Albergaria-a-Velha    ;0.9669;266.57
Albufeira             ;0.9416;224.95
Alcanena              ;0.9526;244.71
Alcobaça              ;0.9366;219.55
Alcochete             ;0.9993;225.8
Alcoutim              ;0.7616;61.34
Alcácer do Sal        ;0.9902;286.51
Alenquer              ;0.995;194.44
Alfândega da Fé       ;0.9843;74.39
Alijó                 ;0.7717;89.72
Aljezur               ;0.7972;120.26
Aljustrel             ;0.9618;273.99
Almada                ;0.9813;112.21
Almeida               ;0.9856;80.31
Almeirim              ;0.911;165.09
Almodôvar             ;0.9785;120.59
Alpiarça              ;0.9867;220.03
Alter do Chão         ;0.9886;113.68
Alvaiázere            ;0.9928;91.74
Alvito                ;0.9976;325.37
Amadora               ;0.9882;108.38
Amarante              ;0.7845;119.31
Amares                ;0.9518;145.82
Anadia                ;0.8703;186.76
Ansião                ;0.9948;141.66
Arcos de Valdevez     ;0.9522;127.85
Arganil               ;0.6777;91.77
Armamar               ;0.9904;135.14
Arouca                ;0.7848;146.52
Arraiolos             ;0.8439;191.25
Arronches             ;0.85;173.86
Arruda dos Vinhos     ;0.9952;166.98
Aveiro                ;0.9971;225.01
Avis                  ;0.9954;367.81
Azambuja              ;0.9934;442.5
Baião                 ;0.9401;88.21
Barcelos              ;0.8871;222.58
Barrancos             ;0.9819;157.07
Barreiro              ;0.9963;100.16
Batalha               ;0.9993;243.85
Beja                  ;0.996;240
Belmonte              ;0.9786;98.21
Benavente             ;0.9701;308.18
Bombarral             ;0.9944;141.08
Borba                 ;0.7399;158.78
Boticas               ;0.8823;98.68
Braga                 ;0.9328;205.27
Bragança              ;0.9104;132.08
Cabeceiras de Basto   ;0.8926;93.65
Cadaval               ;0.8847;136.37
Caldas da Rainha      ;0.9955;153.66
Caminha               ;0.9826;89.03
Campo Maior           ;0.984;311.84
Cantanhede            ;0.9041;148.08
Carrazeda de Ansiães  ;0.994;73.97
Carregal do Sal       ;0.9749;152.77
Cartaxo               ;0.9664;209.69
```

Task 2

DataCleaner Analysis Job

Import CSV File | Configuration



task_1_output.csv | DataCleaner

task_1_output.csv

output.task_1_output.csv

Name	Type
Percentage of Smart Contracts	String
Average Energy Consumption p/ Contra...	String

Close

Convert to Numbers | Configuration

Convert to number | DataCleaner

Convert to number

Documentation | Rename

Input columns

Input: Select all | Select none

- [x] Percentage of Smart Contracts
- [x] Average Energy Consumption p/ Contract (kWh)

Required properties

Decimal separator: .

Minus sign: -

Thousand separator: ,

Optional properties (1)

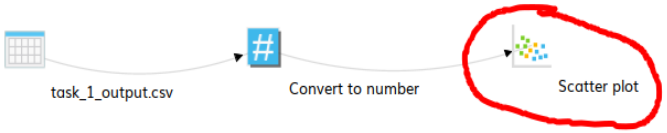
Output columns

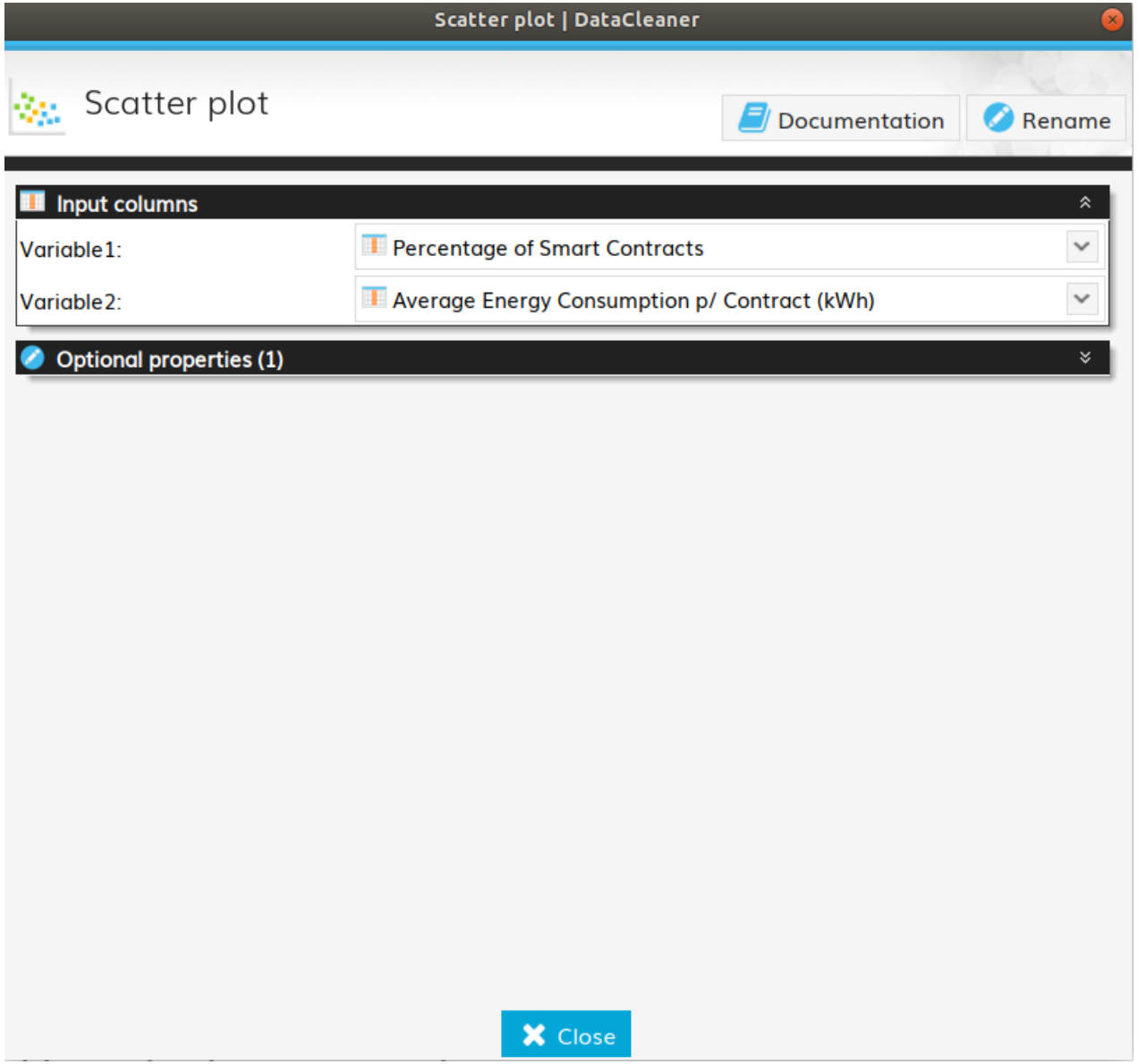
Name	Type
Percentage of Smart Contracts	Number
Average Energy Consumption p/ Contract (	Number

Write data | Preview data

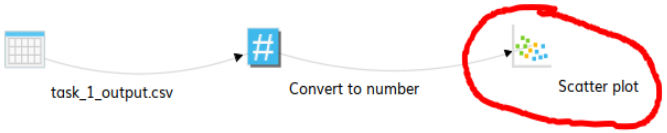
Close

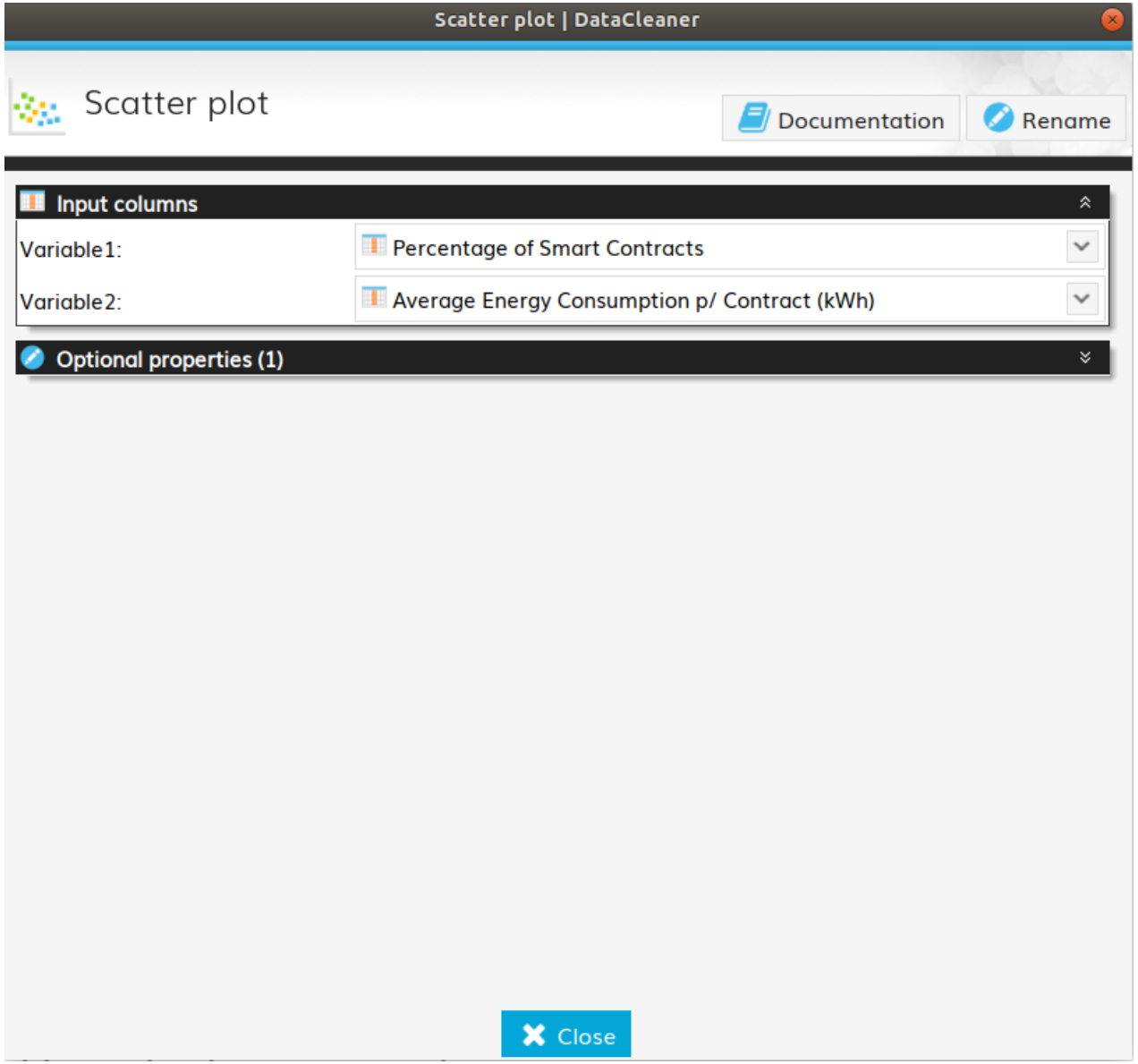
Scatter Plot | Configuration



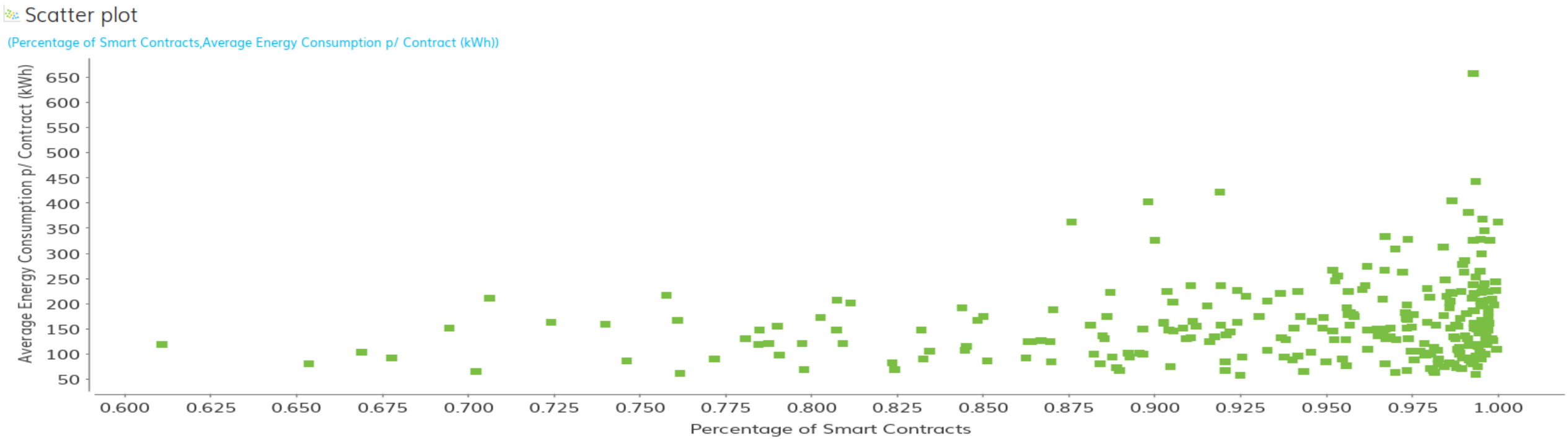


Scatter Plot | Configuration

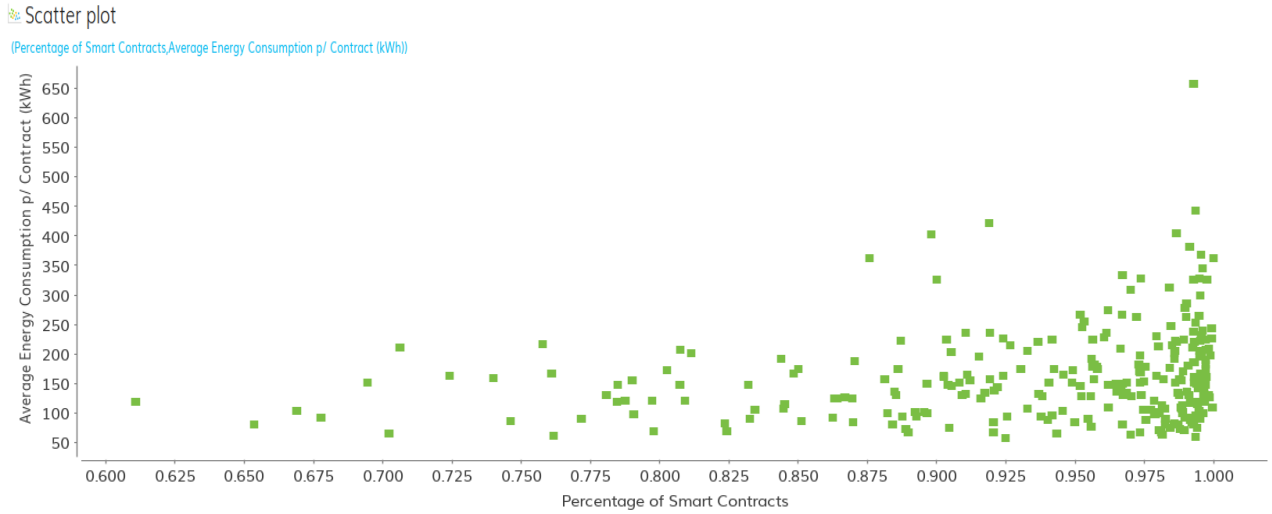




Final Plot


Scatter plot
(Percentage of Smart Contracts,Average Energy Consumption p/ Contract (kWh))
Average Energy Consumption p/ Contract (kWh)
650
600
550
500
450
400
350
300
250
200
150
100
50
0.600
0.625
0.650
0.675
0.700
0.725
0.750
0.775
0.800
0.825
0.850
0.875
0.900
0.925
0.950
0.975
1.000
Percentage of Smart Contracts

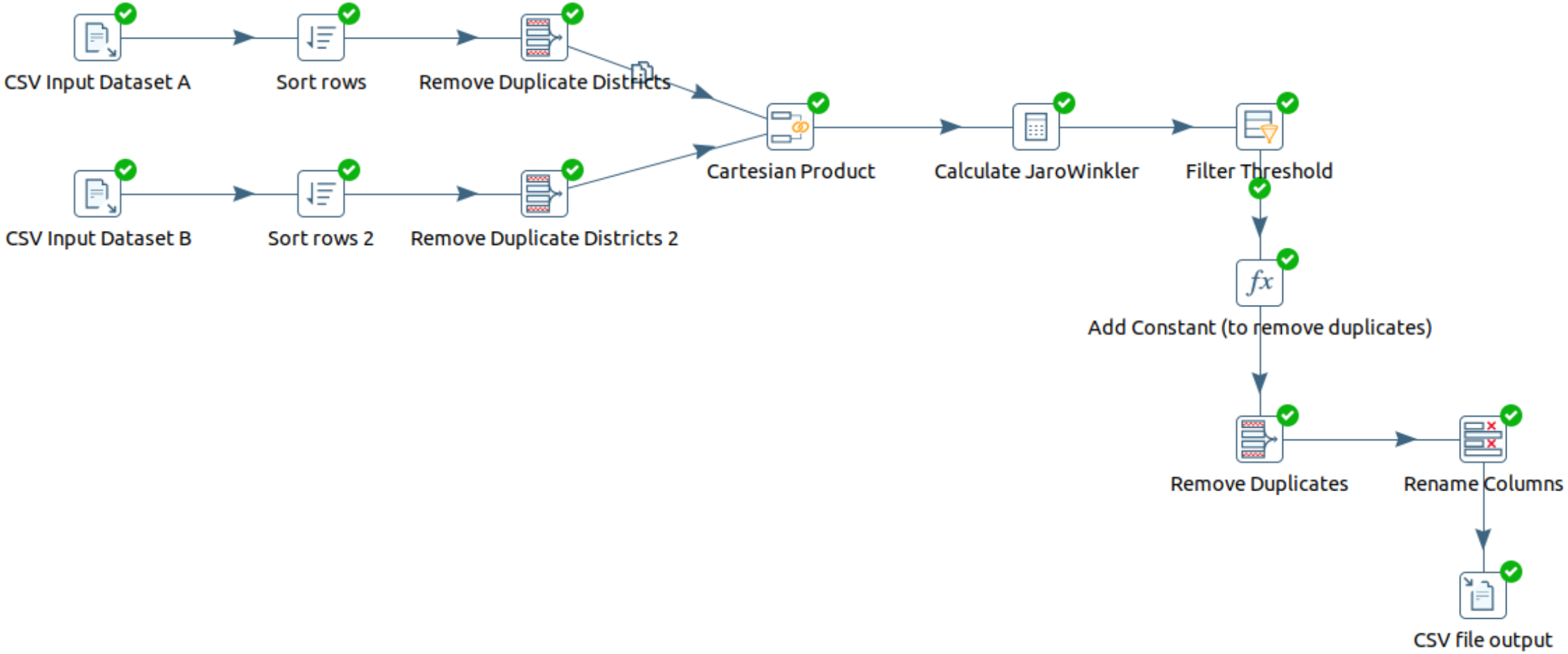
Final Plot | Considerations



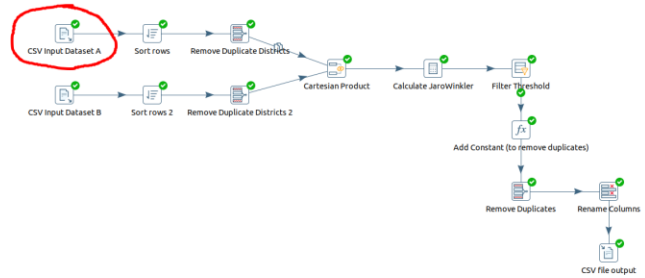
- There seems to exist a **weak positive correlation** between the percentage of smart meters and the average energy consumption per contract.
- As the percentage of smart meters **increases**, the average energy consumption per contract appears to **increase as well**
- **However**, the **relationship isn't linear**, as there are many points scattered with varying consumption values for similar percentages.
- Therefore, we can state that **there is a positive correlation** between the number of smart contracts and the energy consumption per contract, **but it is a weak correlation**

Task 3

Transformation



CSV Input Dataset A | Configuration



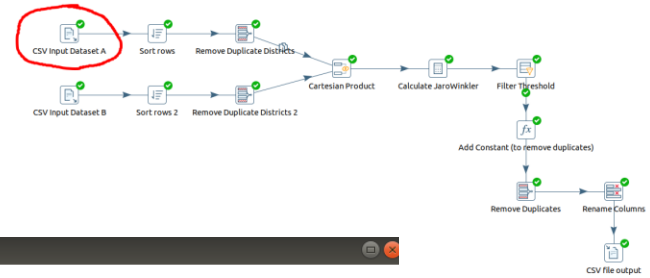
CSV file input

Step name	CSV Input Dataset A
Filename	/home/aid/adi-project/data/21-contadores-de-energia.csv [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	Decimal	Group	Trim type
1	District	String		16			.	,	none

Help | OK | Get Fields | Preview | Cancel

CSV Input Dataset A | Preview

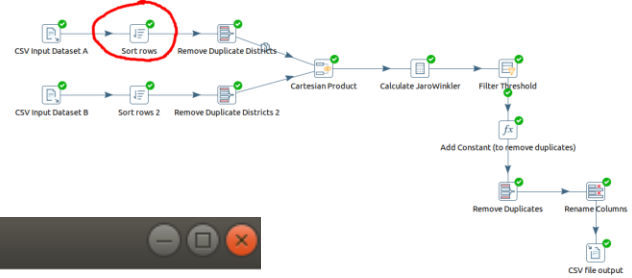


Examine preview data

Rows of step: CSV Input Dataset A (1000 rows)

	District
1	GUARDA
2	SANTAREM
3	SETUBAL
4	VISEU
5	BRAGA
6	BRAGANCA
7	VILA REAL
8	LISBOA
9	VIANA DO CASTELO
10	GUARDA
11	VIANA DO CASTELO
12	VISEU
13	LEIRIA
14	VIANA DO CASTELO
15	BRAGANCA
16	PORTO
17	COIMBRA
18	LEIRIA
19	CASTELO BRANCO
20	LEIRIA
21	LISBOA
22	BRAGA
23	VISEU
24	GUARDA
25	VISEU
26	GUARDA
27	PORTO
28	AVEIRO
29	VISEU
30	LEIRIA
31	VISEU
32	BRAGA
33	CASTELO BRANCO
34	BRAGA
35	VIANA DO CASTELO
36	VIANA DO CASTELO
37	BRAGANCA
38	BRAGANCA
39	GUARDA
40	GUARDA
41	GUARDA
42	PORTO
43	SANTAREM
44	SANTAREM
45	COIMBRA

Sort Rows | Configuration



Sort rows

Step name: Sort rows

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

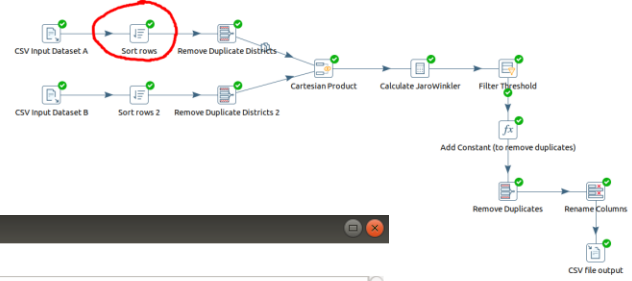
Only pass unique rows? (verifies keys only) ☐

Fields :

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

Help | OK | Cancel | Get Fields

Sort Rows | Preview

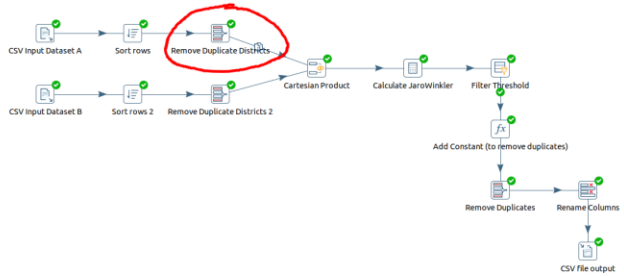


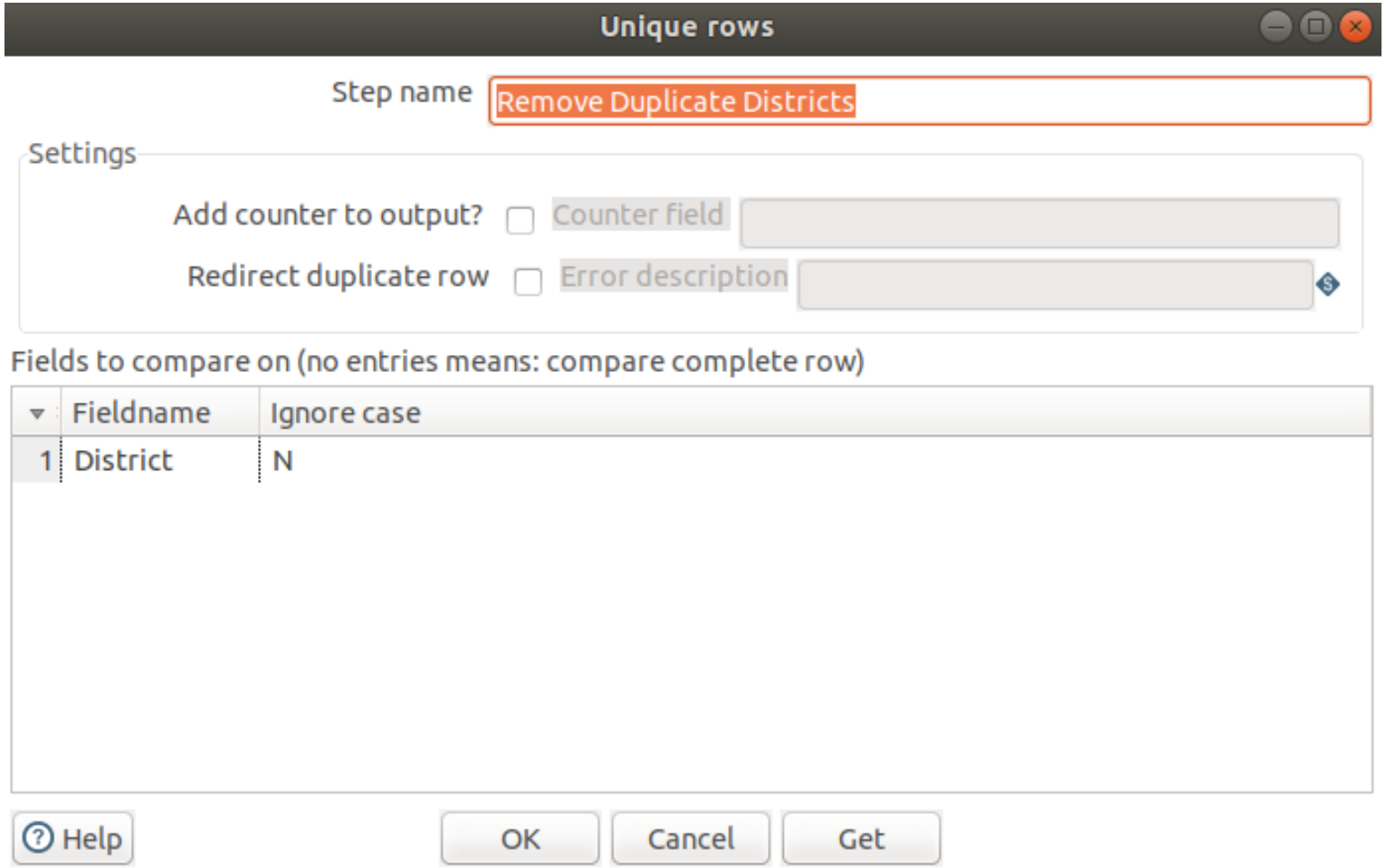
Examine preview data

Rows of step: Sort rows (1000 rows)

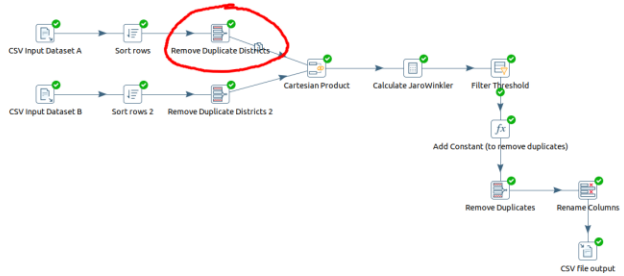
	District
1	<null>
2	AVEIRO
3	AVEIRO
4	AVEIRO
5	AVEIRO
6	AVEIRO
7	AVEIRO
8	AVEIRO
9	AVEIRO
10	AVEIRO
11	AVEIRO
12	AVEIRO
13	AVEIRO
14	AVEIRO
15	AVEIRO
16	AVEIRO
17	AVEIRO
18	AVEIRO
19	AVEIRO
20	AVEIRO
21	AVEIRO
22	AVEIRO
23	AVEIRO
24	AVEIRO
25	AVEIRO
26	AVEIRO
27	AVEIRO
28	AVEIRO
29	AVEIRO
30	AVEIRO
31	AVEIRO
32	AVEIRO
33	AVEIRO
34	AVEIRO
35	AVEIRO
36	AVEIRO
37	AVEIRO
38	AVEIRO
39	AVEIRO
40	AVEIRO
41	AVEIRO
42	AVEIRO
43	AVEIRO
44	AVEIRO
45	AVEIRO

Remove Duplicate Districts | Configuration





Remove Duplicate Districts | Preview

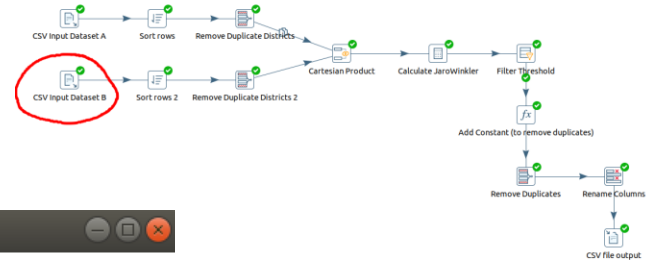


Examine preview data

Rows of step: Remove Duplicate Districts (19 rows)

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

CSV Input Dataset B | Configuration



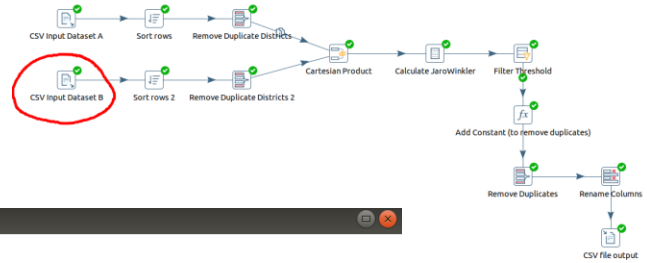
CSV file input

Step name	CSV Input Dataset B
Filename	/home/aid/adi-project/data/3-consumos-faturados-por-municipio-ultimos-10-anos.c [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	Decimal	Group	Trim type
1	District	String		16			.	,	none

Help | OK | Get Fields | Preview | Cancel

CSV Input Dataset B | Preview

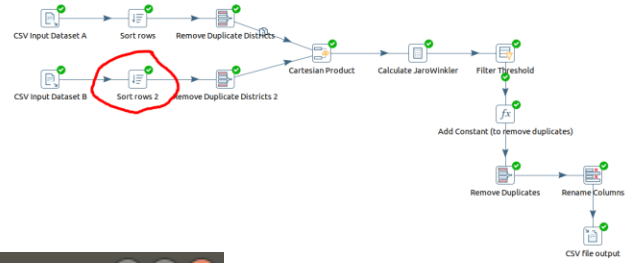


Examine preview data

Rows of step: CSV Input Dataset B (1000 rows)

	District
1	SANTAREM
2	VISEU
3	VISEU
4	AVEIRO
5	BRAGA
6	BRAGA
7	COIMBRA
8	GUARDA
9	PORTO
10	SANTAREM
11	SANTAREM
12	VIANA DO CASTELO
13	VISEU
14	AVEIRO
15	CASTELO BRANCO
16	COIMBRA
17	GUARDA
18	SANTAREM
19	SETUBAL
20	BRAGA
21	EVORA
22	GUARDA
23	GUARDA
24	LISBOA
25	LISBOA
26	PORTO
27	PORTO
28	SETUBAL
29	VIANA DO CASTELO
30	VILA REAL
31	VISEU
32	BRAGA
33	GUARDA
34	GUARDA
35	LISBOA
36	PORTO
37	PORTO
38	VIANA DO CASTELO
39	VIANA DO CASTELO
40	VIANA DO CASTELO
41	BEJA
42	BRAGA
43	BRAGA
44	BRAGA
45	BRAGANCA

Sort Rows 2 | Configuration



Sort rows

Step name: Sort rows 2

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?

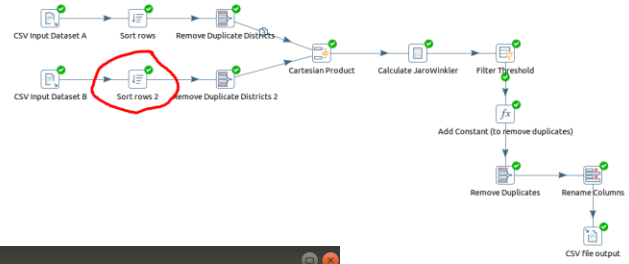
Only pass unique rows? (verifies keys only)

Fields :

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

Help OK Cancel Get Fields

Sort Rows 2 | Preview

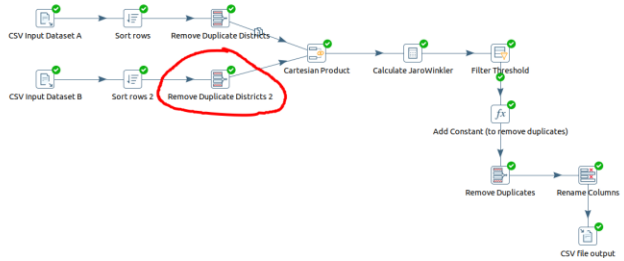


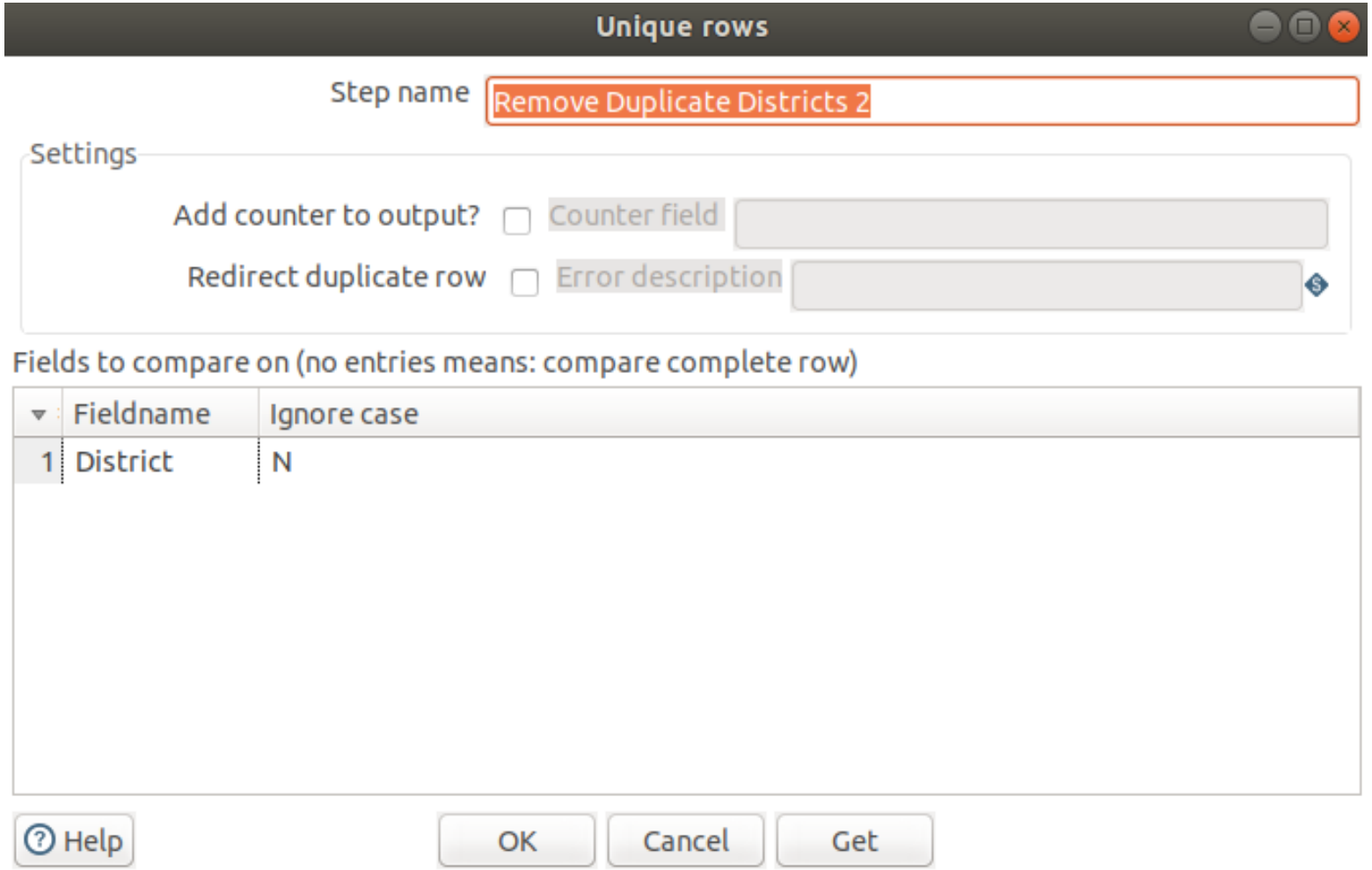
Examine preview data

Rows of step: Sort rows 2 (1000 rows)

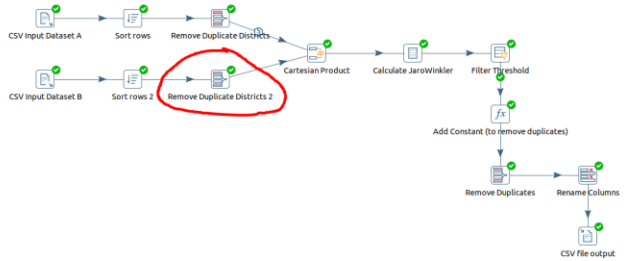
	District
1	AVEIRO
2	AVEIRO
3	AVEIRO
4	AVEIRO
5	AVEIRO
6	AVEIRO
7	AVEIRO
8	AVEIRO
9	AVEIRO
10	AVEIRO
11	AVEIRO
12	AVEIRO
13	AVEIRO
14	AVEIRO
15	AVEIRO
16	AVEIRO
17	AVEIRO
18	AVEIRO
19	AVEIRO
20	AVEIRO
21	AVEIRO
22	AVEIRO
23	AVEIRO
24	AVEIRO
25	AVEIRO
26	AVEIRO
27	AVEIRO
28	AVEIRO
29	AVEIRO
30	AVEIRO
31	AVEIRO
32	AVEIRO
33	AVEIRO
34	AVEIRO
35	AVEIRO
36	AVEIRO
37	AVEIRO
38	AVEIRO
39	AVEIRO
40	AVEIRO
41	AVEIRO
42	AVEIRO
43	AVEIRO
44	AVEIRO
45	AVEIRO

Remove Duplicate Districts 2 | Configuration





Remove Duplicate Districts 2 | Preview

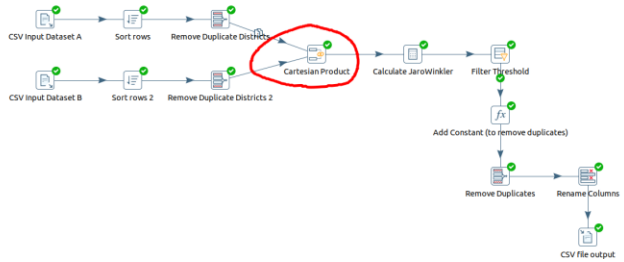


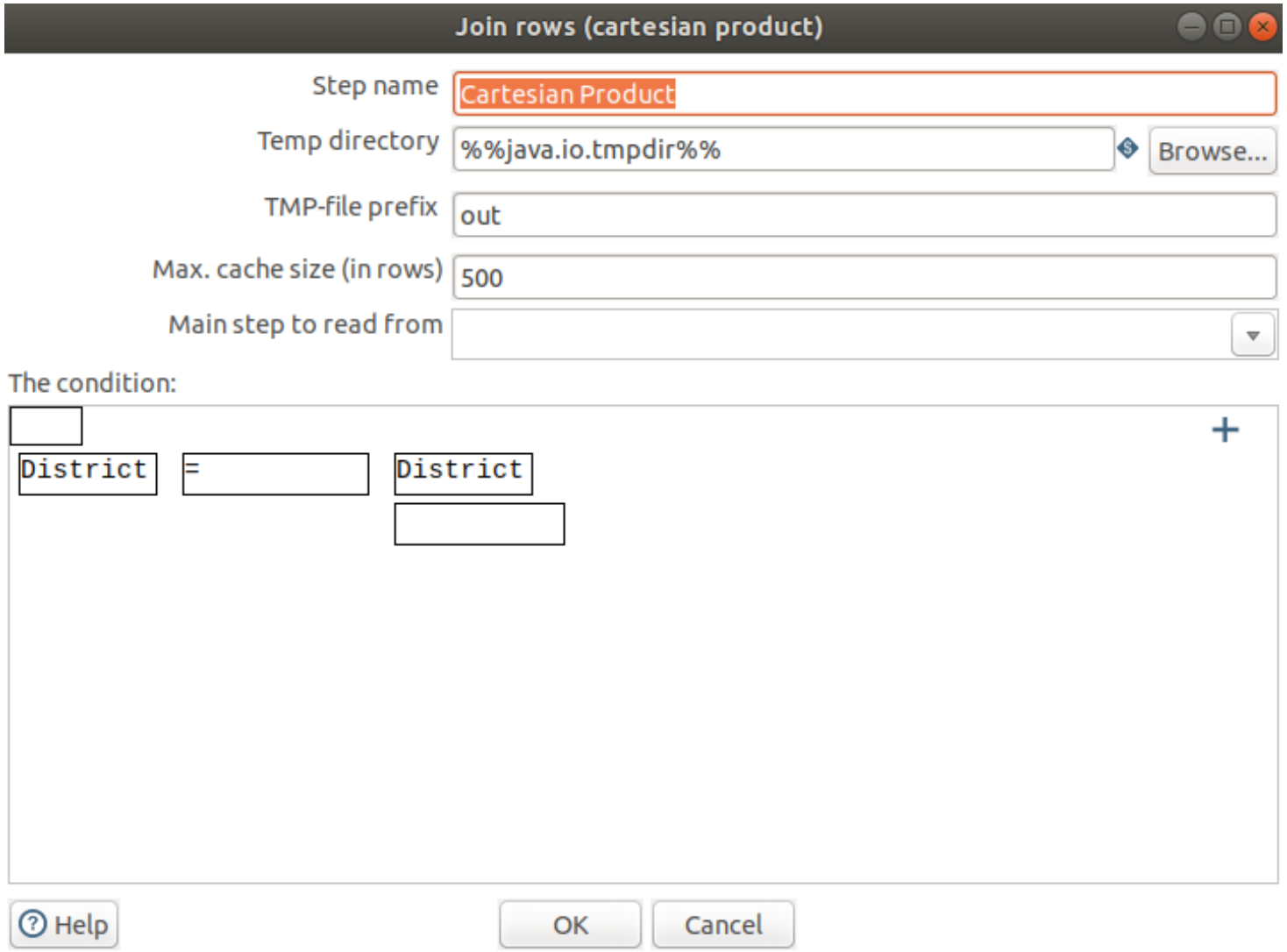
Examine preview data

Rows of step: Remove Duplicate Districts 2 (18 rows)

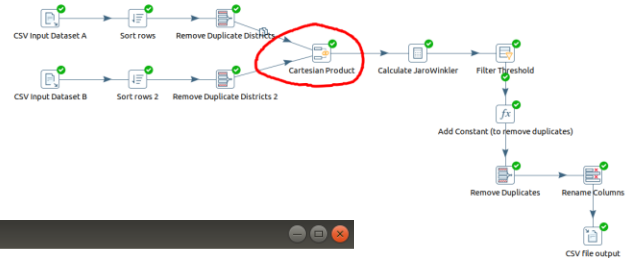
	District
1	AVEIRO
2	BEJA
3	BRAGA
4	BRAGANCA
5	CASTELO BRANCO
6	COIMBRA
7	EVORA
8	FARO
9	GUARDA
10	LEIRIA
11	LISBOA
12	PORTALEGRE
13	PORTO
14	SANTAREM
15	SETUBAL
16	VIANA DO CASTELO
17	VILA REAL
18	VISEU

Cartesian Product | Configuration





Cartesian Product | Preview

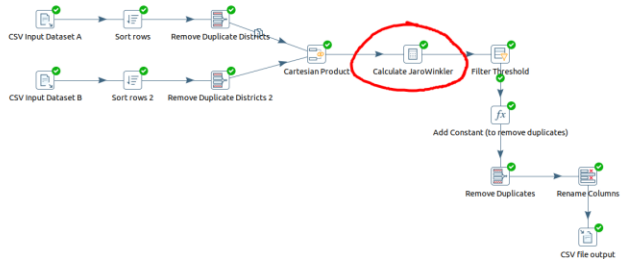


Examine preview data

Rows of step: Cartesian Product (342 rows)

	District	District_1
1	AVEIRO	<null>
2	AVEIRO	AVEIRO
3	AVEIRO	BEJA
4	AVEIRO	BRAGA
5	AVEIRO	BRAGANCA
6	AVEIRO	CASTELO BRANCO
7	AVEIRO	COIMBRA
8	AVEIRO	EVORA
9	AVEIRO	FARO
10	AVEIRO	GUARDA
11	AVEIRO	LEIRIA
12	AVEIRO	LISBOA
13	AVEIRO	PORTALEGRE
14	AVEIRO	PORTO
15	AVEIRO	SANTAREM
16	AVEIRO	SETUBAL
17	AVEIRO	VIANA DO CASTELO
18	AVEIRO	VILA REAL
19	AVEIRO	VISEU
20	BEJA	<null>
21	BEJA	AVEIRO
22	BEJA	BEJA
23	BEJA	BRAGA
24	BEJA	BRAGANCA
25	BEJA	CASTELO BRANCO
26	BEJA	COIMBRA
27	BEJA	EVORA
28	BEJA	FARO
29	BEJA	GUARDA
30	BEJA	LEIRIA
31	BEJA	LISBOA
32	BEJA	PORTALEGRE
33	BEJA	PORTO
34	BEJA	SANTAREM
35	BEJA	SETUBAL
36	BEJA	VIANA DO CASTELO
37	BEJA	VILA REAL
38	BEJA	VISEU
39	BRAGA	<null>
40	BRAGA	AVEIRO
41	BRAGA	BEJA
42	BRAGA	BRAGA
43	BRAGA	BRAGANCA
44	BRAGA	CASTELO BRANCO
45	BRAGA	COIMBRA

Calculate JaroWinkler | Configuration



Calculator

Step name

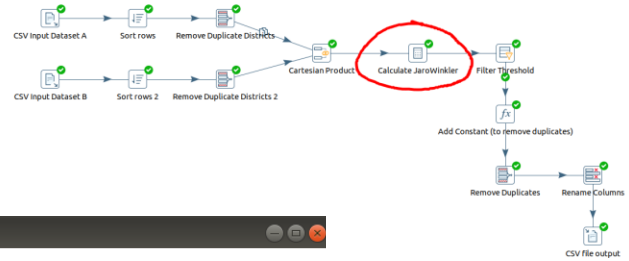
Calculate JaroWinkler

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	JaroWinkler_Measure	JaroWinkler similitude between String A and String B	District	District_1		None			N		

Calculate JaroWinkler | Preview

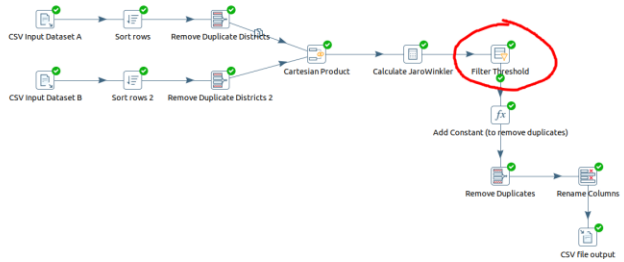


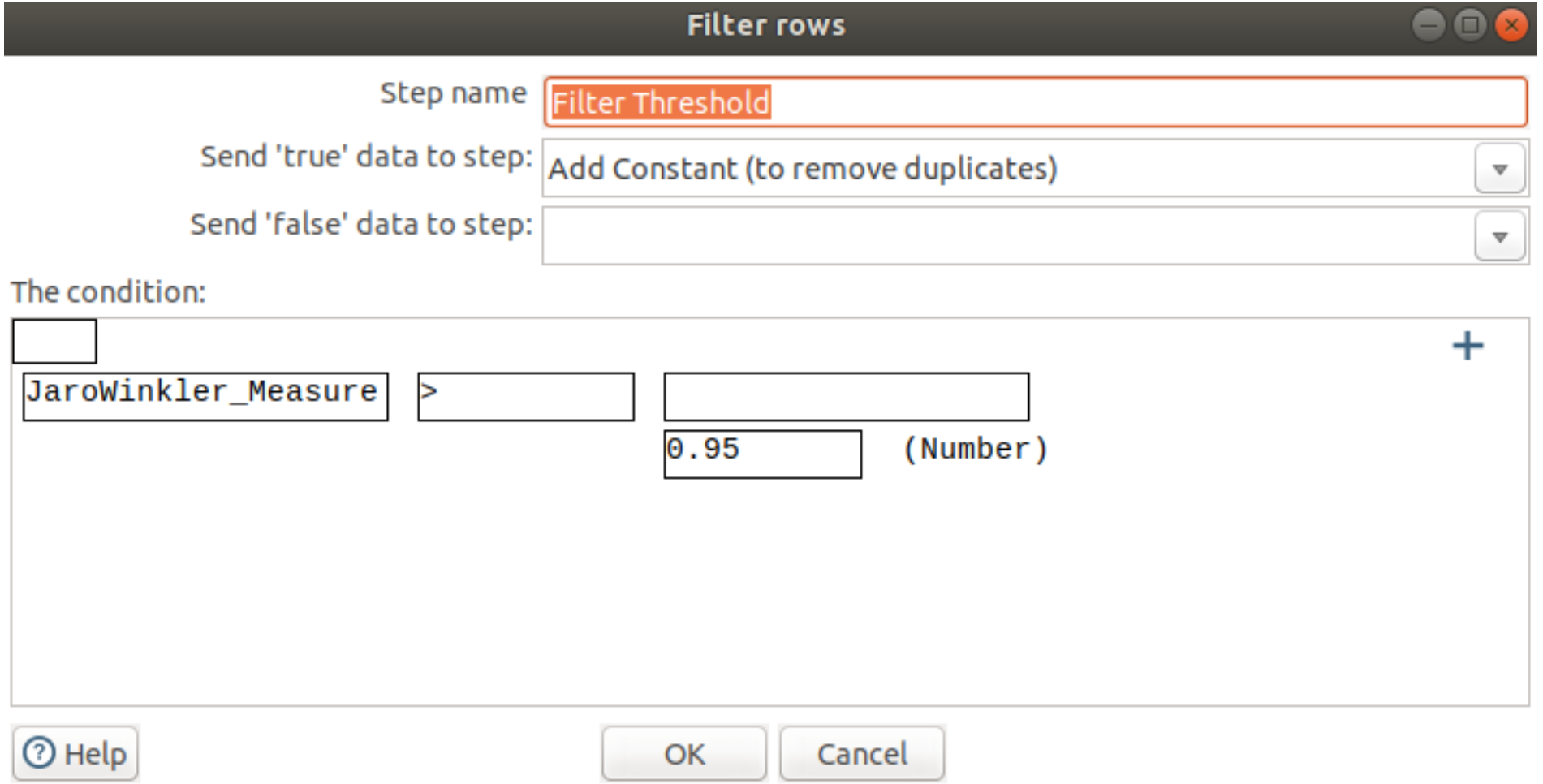
Examine preview data

Rows of step: Calculate JaroWinkler (342 rows)

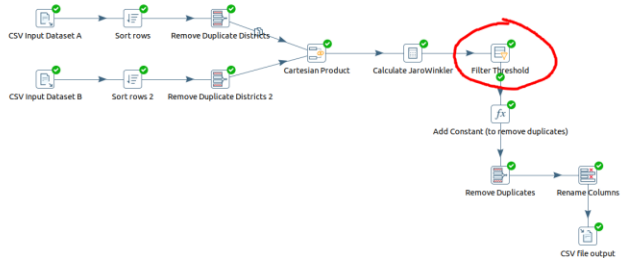
	District	District_1	JaroWinkler_Measure
1	AVEIRO	<null>	<null>
2	AVEIRO	AVEIRO	1.0
3	AVEIRO	BEJA	0.4722222222
4	AVEIRO	BRAGA	0.4555555556
5	AVEIRO	BRAGANCA	0.3611111111
6	AVEIRO	CASTELO BRANCO	0.5674603175
7	AVEIRO	COIMBRA	0.5396825397
8	AVEIRO	EVORA	0.5888888889
9	AVEIRO	FARO	0.75
10	AVEIRO	GUARDA	0.5555555556
11	AVEIRO	LEIRIA	0.6666666667
12	AVEIRO	LISBOA	0.5555555556
13	AVEIRO	PORTALEGRE	0.5222222222
14	AVEIRO	PORTO	0.5777777778
15	AVEIRO	SANTAREM	0.5277777778
16	AVEIRO	SETUBAL	0.4365079365
17	AVEIRO	VIANA DO CASTELO	0.5555555556
18	AVEIRO	VILA REAL	0.6203703704
19	AVEIRO	VISEU	0.5888888889
20	BEJA	<null>	<null>
21	BEJA	AVEIRO	0.4722222222
22	BEJA	BEJA	1.0
23	BEJA	BRAGA	0.6333333333
24	BEJA	BRAGANCA	0.5833333333
25	BEJA	CASTELO BRANCO	0.380952381
26	BEJA	COIMBRA	0.0
27	BEJA	EVORA	0.6333333333
28	BEJA	FARO	0.0
29	BEJA	GUARDA	0.4722222222
30	BEJA	LEIRIA	0.6111111111
31	BEJA	LISBOA	0.4722222222
32	BEJA	PORTALEGRE	0.45
33	BEJA	PORTO	0.0
34	BEJA	SANTAREM	0.4583333333
35	BEJA	SETUBAL	0.5952380952
36	BEJA	VIANA DO CASTELO	0.4375
37	BEJA	VILA REAL	0.4537037037
38	BEJA	VISEU	0.0
39	BRAGA	<null>	<null>
40	BRAGA	AVEIRO	0.4555555556
41	BRAGA	BEJA	0.6333333333
42	BRAGA	BRAGA	1.0
43	BRAGA	BRAGANCA	0.9375
44	BRAGA	CASTELO BRANCO	0.5142857143
45	BRAGA	COIMBRA	0.4476190476

Filter Threshold | Configuration





Filter Threshold | Preview



Examine preview data

Rows of step: Filter Threshold (18 rows)

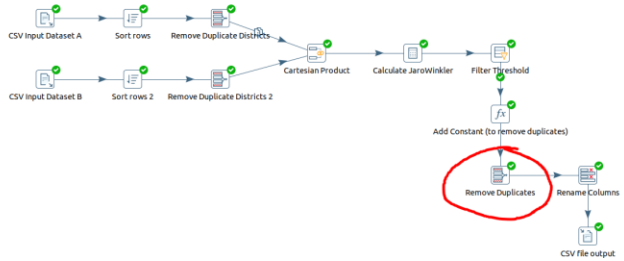
	District	District_1	JaroWinkler_Measure
1	AVEIRO	AVEIRO	1.0
2	BEJA	BEJA	1.0
3	BRAGA	BRAGA	1.0
4	BRAGANCA	BRAGANCA	1.0
5	CASTELO BRANCO	CASTELO BRANCO	1.0
6	COIMBRA	COIMBRA	1.0
7	EVORA	EVORA	1.0
8	FARO	FARO	1.0
9	GUARDA	GUARDA	1.0
10	LEIRIA	LEIRIA	1.0
11	LISBOA	LISBOA	1.0
12	PORTALEGRE	PORTALEGRE	1.0
13	PORTO	PORTO	1.0
14	SANTAREM	SANTAREM	1.0
15	SETUBAL	SETUBAL	1.0
16	VIANA DO CASTELO	VIANA DO CASTELO	1.0
17	VILA REAL	VILA REAL	1.0
18	VISEU	VISEU	1.0

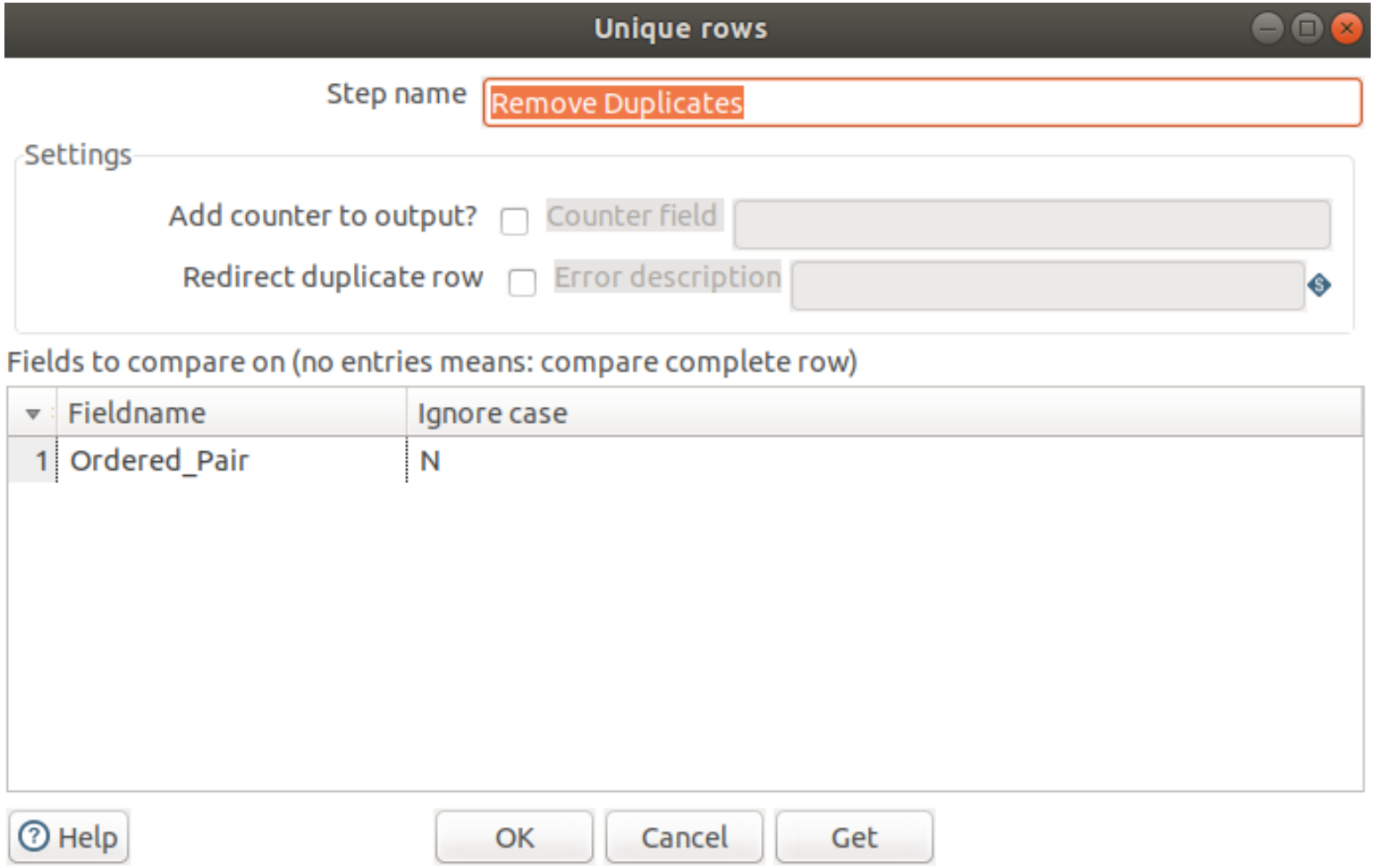
Note

- The next two steps in the transformation process (Add Constant & Remove Duplicates) **are not necessary with the current datasets** (where all the district names match).

- However, if there were similar district names, that would pass the JaroWinkler threshold (for example, Bragança-Braganca), **there would be two pairs of these values (Bragança-Braganca and Braganca-Bragança)**. Therefore, we still included the following two steps.

Add Constant (to remove duplicates) - Configuration

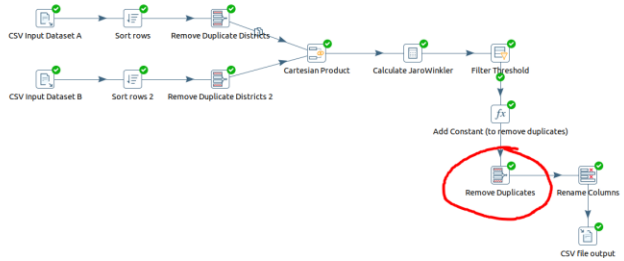




Add Constant (to remove duplicates)

-

Preview

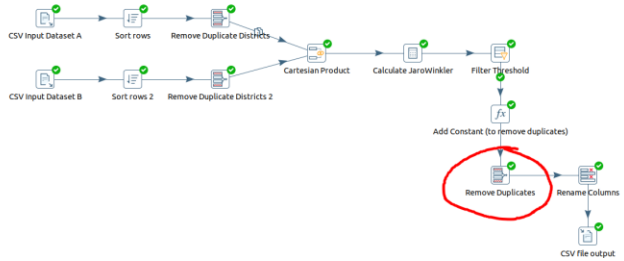


Examine preview data

Rows of step: Remove Duplicates (18 rows)

	District	District_1	JaroWinkler_Measure	Ordered_Pair
1	AVEIRO	AVEIRO	1.0	AVEIRO-AVEIRO
2	BEJA	BEJA	1.0	BEJA-BEJA
3	BRAGA	BRAGA	1.0	BRAGA-BRAGA
4	BRAGANCA	BRAGANCA	1.0	BRAGANCA-BRAGANCA
5	CASTELO BRANCO	CASTELO BRANCO	1.0	CASTELO BRANCO-CASTELO BRANCO
6	COIMBRA	COIMBRA	1.0	COIMBRA-COIMBRA
7	EVORA	EVORA	1.0	EVORA-EVORA
8	FARO	FARO	1.0	FARO-FARO
9	GUARDA	GUARDA	1.0	GUARDA-GUARDA
10	LEIRIA	LEIRIA	1.0	LEIRIA-LEIRIA
11	LISBOA	LISBOA	1.0	LISBOA-LISBOA
12	PORTALEGRE	PORTALEGRE	1.0	PORTALEGRE-PORTALEGRE
13	PORTO	PORTO	1.0	PORTO-PORTO
14	SANTAREM	SANTAREM	1.0	SANTAREM-SANTAREM
15	SETUBAL	SETUBAL	1.0	SETUBAL-SETUBAL
16	VIANA DO CASTELO	VIANA DO CASTELO	1.0	VIANA DO CASTELO-VIANA DO CASTELO
17	VILA REAL	VILA REAL	1.0	VILA REAL-VILA REAL
18	VISEU	VISEU	1.0	VISEU-VISEU

Remove Duplicates | Configuration



Unique rows

Step name: Remove Duplicates

Settings

Add counter to output? ☐ Counter field

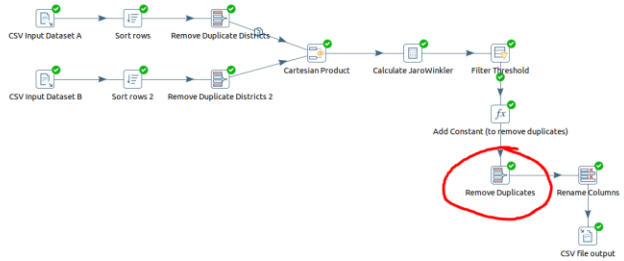
Redirect duplicate row ☐ Error description

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

	Fieldname	Ignore case
1	Ordered_Pair	N

Help | OK | Cancel | Get

Remove Duplicates | Preview

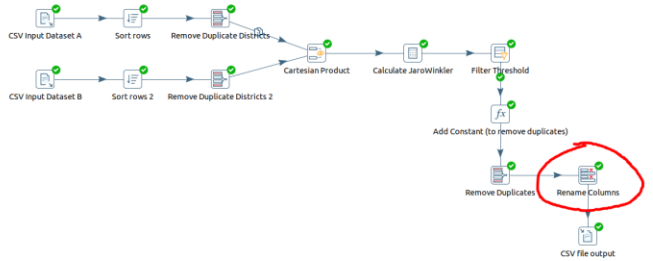


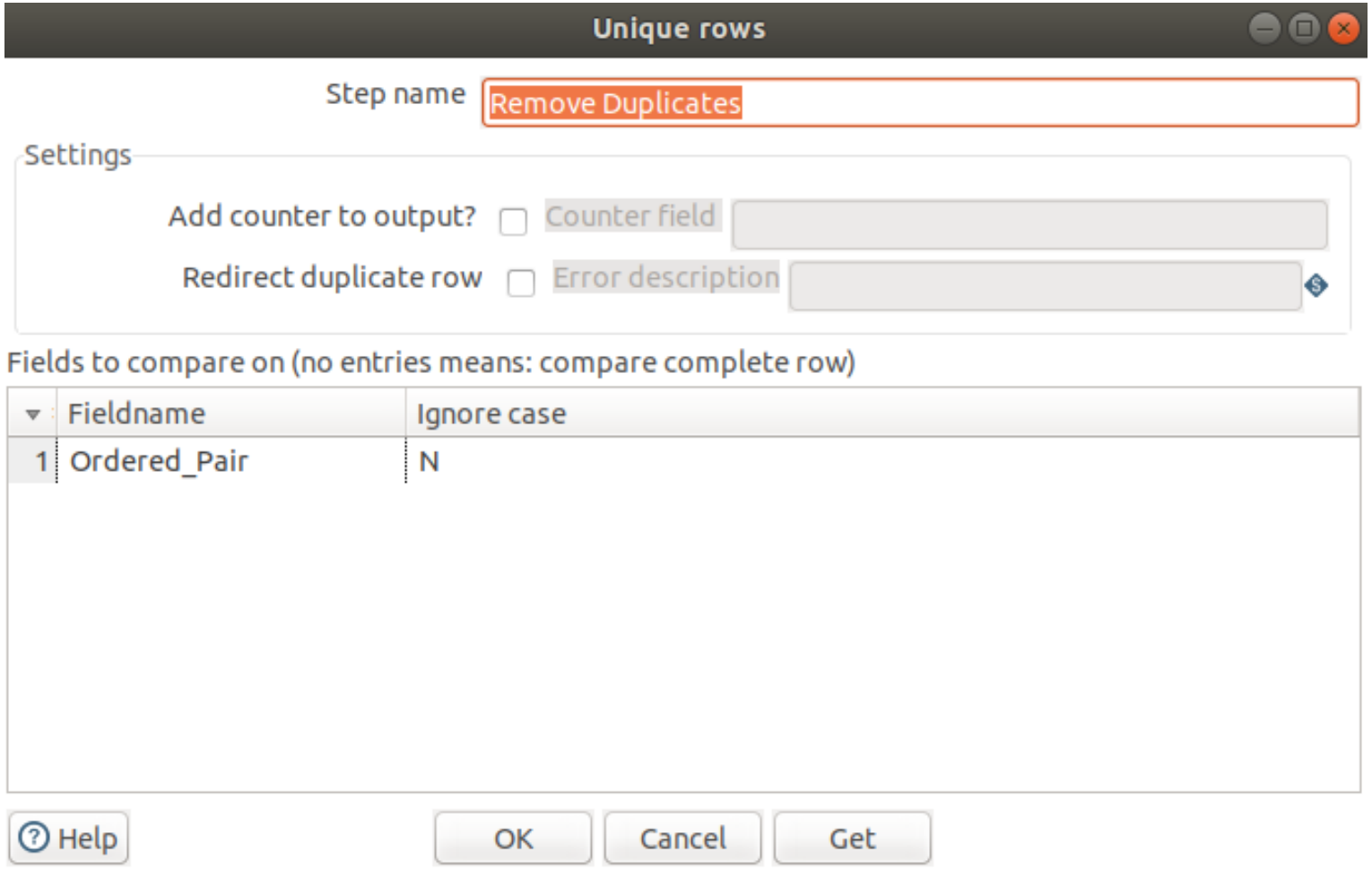
Examine preview data

Rows of step: Remove Duplicates (18 rows)

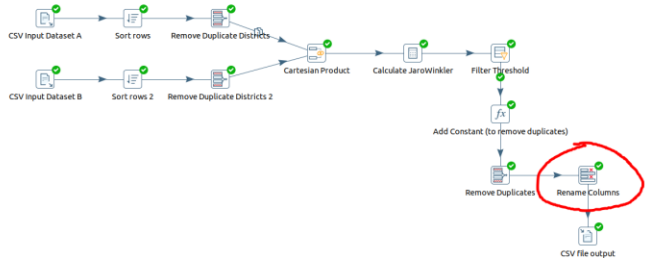
	District	District_1	JaroWinkler_Measure	Ordered_Pair
1	AVEIRO	AVEIRO	1.0	AVEIRO-AVEIRO
2	BEJA	BEJA	1.0	BEJA-BEJA
3	BRAGA	BRAGA	1.0	BRAGA-BRAGA
4	BRAGANCA	BRAGANCA	1.0	BRAGANCA-BRAGANCA
5	CASTELO BRANCO	CASTELO BRANCO	1.0	CASTELO BRANCO-CASTELO BRANCO
6	COIMBRA	COIMBRA	1.0	COIMBRA-COIMBRA
7	EVORA	EVORA	1.0	EVORA-EVORA
8	FARO	FARO	1.0	FARO-FARO
9	GUARDA	GUARDA	1.0	GUARDA-GUARDA
10	LEIRIA	LEIRIA	1.0	LEIRIA-LEIRIA
11	LISBOA	LISBOA	1.0	LISBOA-LISBOA
12	PORTALEGRE	PORTALEGRE	1.0	PORTALEGRE-PORTALEGRE
13	PORTO	PORTO	1.0	PORTO-PORTO
14	SANTAREM	SANTAREM	1.0	SANTAREM-SANTAREM
15	SETUBAL	SETUBAL	1.0	SETUBAL-SETUBAL
16	VIANA DO CASTELO	VIANA DO CASTELO	1.0	VIANA DO CASTELO-VIANA DO CASTELO
17	VILA REAL	VILA REAL	1.0	VILA REAL-VILA REAL
18	VISEU	VISEU	1.0	VISEU-VISEU

Remove Duplicates | Configuration





Rename Columns | Preview

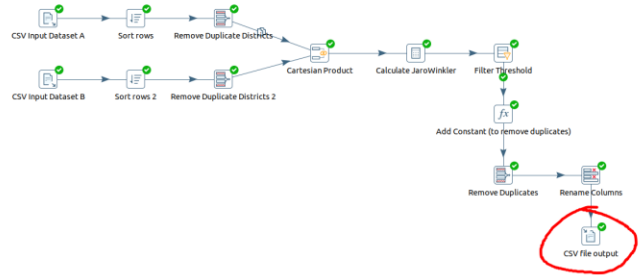


Examine preview data

Rows of step: Rename Columns (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

CSV File Output | Configuration (1/3)



Text file output

Step name: CSV file output

File | Content | Fields

Filename: /home/aid/adi-project/delivery/output/task_3_output — Browse...

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: ☐

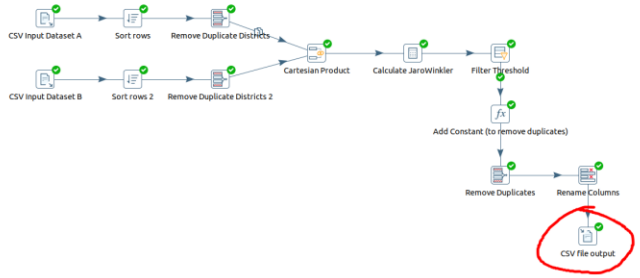
Date time format:

Show filename(s)...

Add filenames to result: ☑

Help — OK — Cancel

CSV File Output | Configuration (2/3)



Text file output

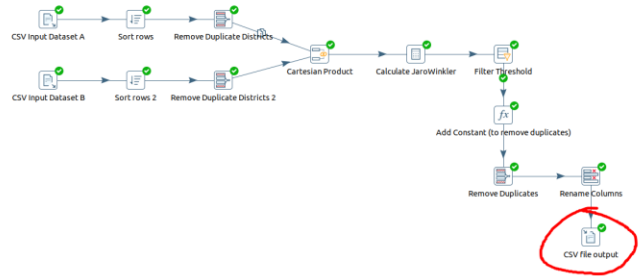
Step name: CSV file output

File | Content | Fields

- Append: ☐
- Separator: ; (Insert TAB)
- Enclosure: "
- Force the enclosure around fields?: ☐
- Disable the enclosure fix?: ☐
- Header: ☑
- Footer: ☐
- Format: CR+LF terminated (Windows, DOS)
- Compression: None
- Encoding: UTF-8
- Right pad fields: ☐
- Fast data dump (no formatting): ☐
- Split every ... rows:
- Add Ending line of file:

Help | OK | Cancel

CSV File Output | Configuration (3/3)



Text file output

Step name: CSV file output

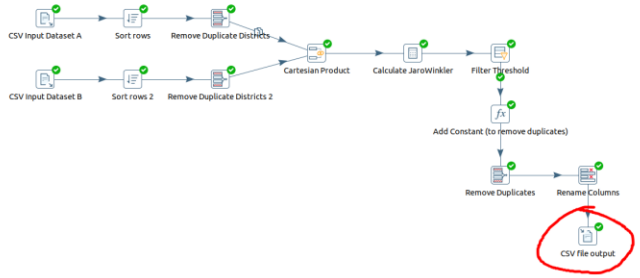
File | Content | Fields

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	District_A	String							both	
2	District_B	String							both	

Get Fields | Minimal width

Help | OK | Cancel

CSV File Output | Preview

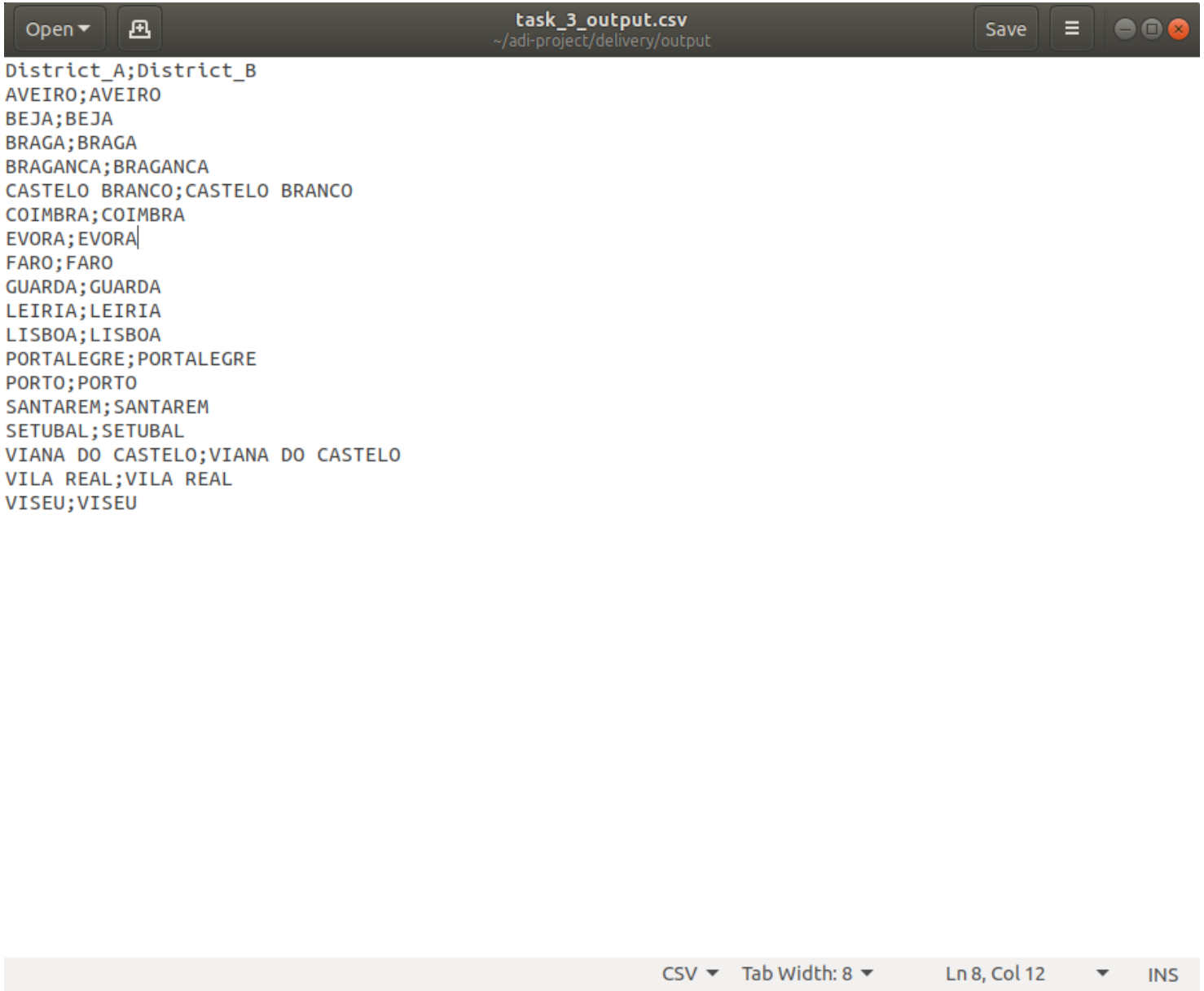


Examine preview data

Rows of step: CSV file output (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

CSV File Output | Preview



Task 4

Data Warehouse Creation SQL Script

delivery > dw > energy_dw.sql > ...

You, 1 second ago | 1 author (You) | Run | New Tab | Active Connection

```sql
DROP DATABASE IF EXISTS energy_dw;

Run | New Tab
CREATE DATABASE energy_dw;

Run | New Tab
USE energy_dw;

Run | New Tab | Copy
CREATE TABLE dim_time (
    time_id DATE PRIMARY KEY,
    year_id INT,
    season_id INT,              -- 1: Winter, 2: Spring, 3: Summer, 4: Autumn
    season_name VARCHAR(10),
    month_id INT,
    month_name VARCHAR(10)
);

Run | New Tab | Copy
CREATE TABLE dim_location (
    location_id INT AUTO_INCREMENT PRIMARY KEY,
    region VARCHAR(50),
    municipality VARCHAR(50),
    parish VARCHAR(50),
    parish_code INT -- due to discrepancies between the names of the parishes in the two datasets, we will use the code
);

-- Create the fact table for energy consumption
Run | New Tab | Copy
CREATE TABLE fact_energy_stats (
    time_id DATE,
    location_id INT,
    energy_consumption DECIMAL(10,2), -- Energy consumption in kWh
    percentage_smart_meters DECIMAL(5,4), -- Percentage with 4 decimal places
    PRIMARY KEY (time_id, location_id),
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (location_id) REFERENCES dim_location(location_id)
);      You, 1 second ago • Uncommitted changes
```

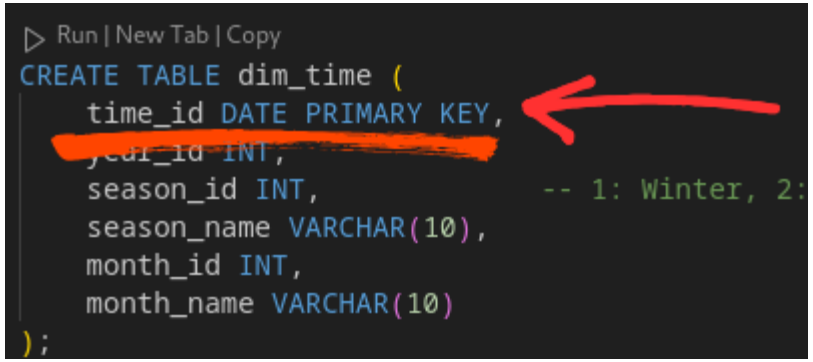
Time Dimension table

Location Dimension table

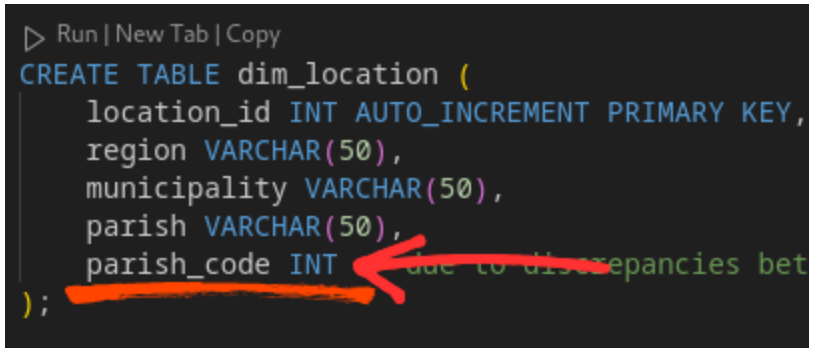
Facts table

Comments on Design Choices

- **Time Dimension table PK**
  - We decided to not use a surrogate key, as it is common practice to not use surrogate keys on time dimension tables.



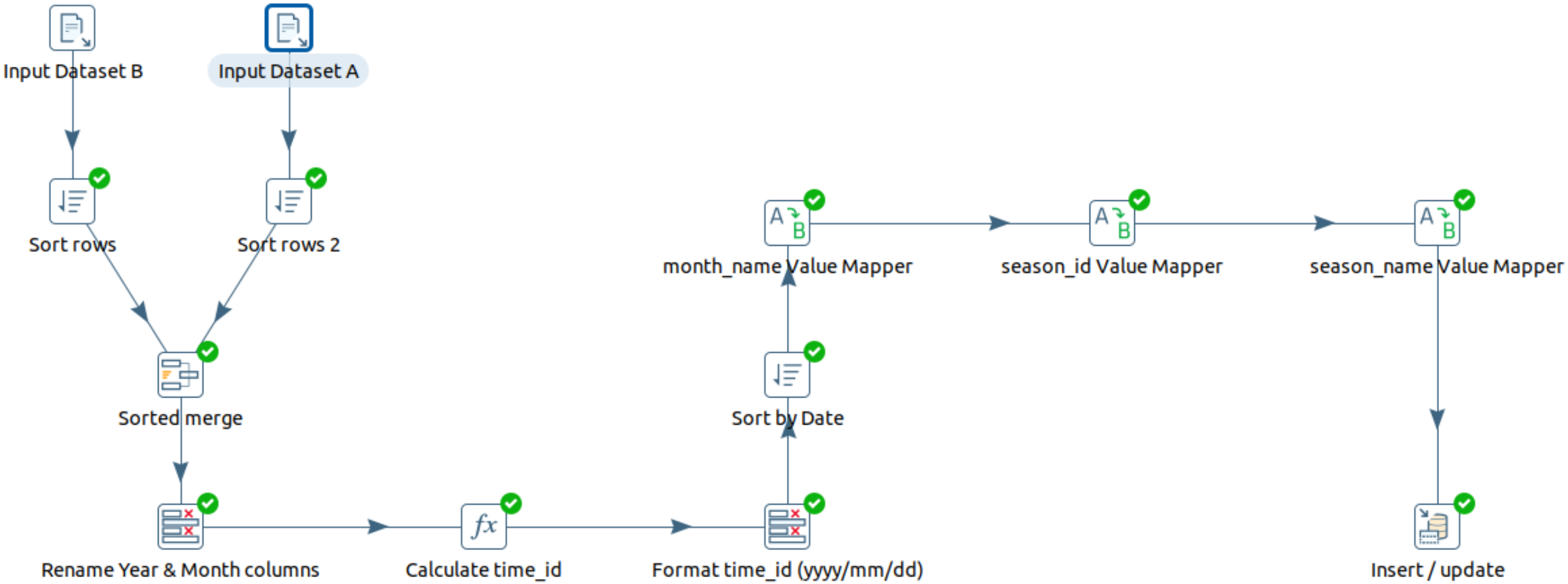
- **Parish code field in Location Dimension**
  - We detected some discrepancies in the names of parishes between Datasets A and B
  - E.g., in Dataset A existed "Azueira e Sobral da Abelheira, and in Dataset B "UF AZUEIRA SOBRAL ABELHEIRA", but it is actually the same parish
  - Due to this, we decided to add the Parish code field, that correctly and uniquely identifies all parishes across datasets



Task 5

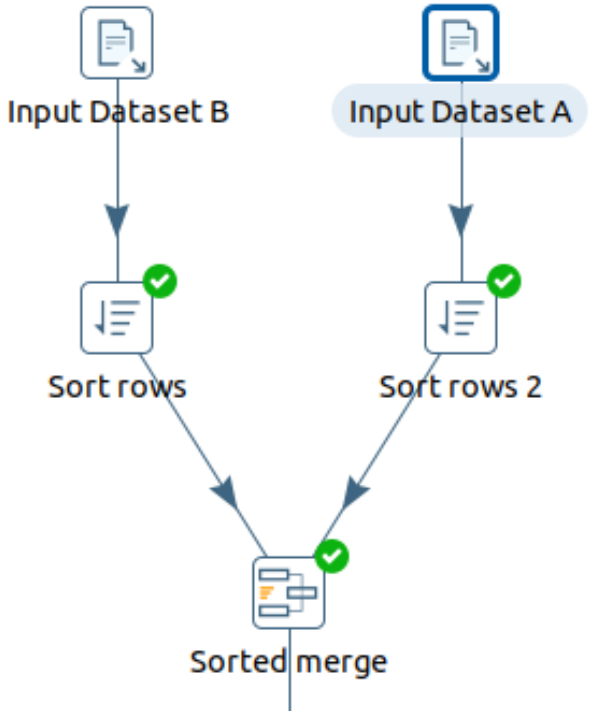
Time Dimension Population Transformation

Transformation

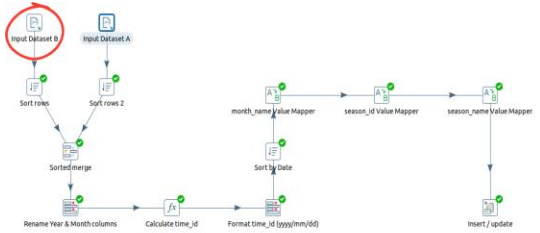


Considerations on Time Dimension Populate Transformation

- **Using both Datasets as source for Time Dimension**
  - Some months had data in Dataset A that were missing in Dataset B, and vice versa.
  - For this reason, we decided to use both Datasets A and B to populate the dimension table (even though it is stated in task 4 to base the data warehouse schema on Dataset B).



Input Dataset B | Configuration



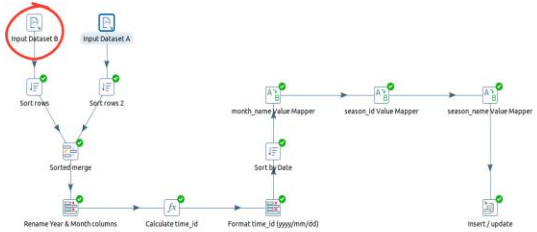
CSV file input

Step name	Input Dataset B
Filename	/home/aid/adi-project/data/3-consumos-faturados-por-municipio-ultimos-10-anos.csv [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	

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Year	Integer	#	15	0	$	.	,	none
2	Month	Integer	#	15	0	$	.	,	none

Help | OK | Get Fields | Preview | Cancel

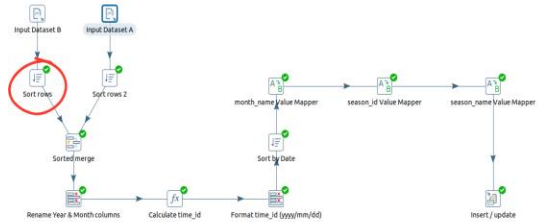
Input Dataset B | Preview



Rows of step: Input Dataset B (1000 rows)

	Year	Month
1	2020	11
2	2020	11
3	2020	11
4	2020	11
5	2020	11
6	2020	11
7	2020	11
8	2020	11
9	2020	11
10	2020	11
11	2020	11
12	2020	11
13	2020	11
14	2020	11
15	2020	11
16	2020	11
17	2020	11
18	2020	11
19	2020	11
20	2020	11
21	2020	11
22	2020	11
23	2020	11
24	2020	11
25	2020	11
26	2020	11
27	2020	11
28	2020	11
29	2020	11
30	2020	11
31	2020	11
32	2020	11
33	2020	11
34	2020	11
35	2020	11
36	2020	11
37	2020	11
38	2020	11
39	2020	11
40	2020	11
41	2020	11
42	2020	11
43	2020	11
44	2020	11
45	2020	11

Sort Rows | Configuration



Sort rows

Step name: Sort by Date

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

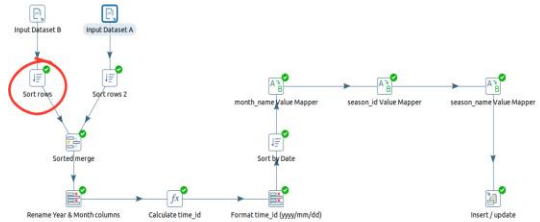
Only pass unique rows? (verifies keys only) ☑

Fields :

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

Help | OK | Cancel | Get Fields

Sort Rows | Preview

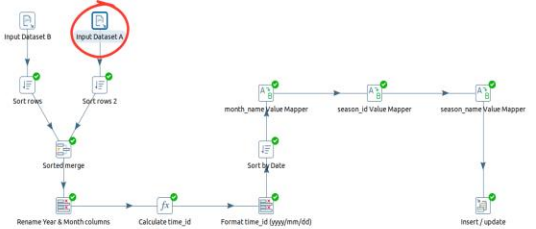


Examine preview data

Rows of step: Sort by Date (45 rows)

	Year	Month
1	2020	11
2	2020	12
3	2021	1
4	2021	2
5	2021	3
6	2021	4
7	2021	5
8	2021	6
9	2021	7
10	2021	8
11	2021	9
12	2021	10
13	2021	11
14	2021	12
15	2022	1
16	2022	2
17	2022	3
18	2022	4
19	2022	5
20	2022	6
21	2022	7
22	2022	8
23	2022	9
24	2022	10
25	2022	11
26	2022	12
27	2023	1
28	2023	2
29	2023	3
30	2023	4
31	2023	5
32	2023	6
33	2023	7
34	2023	8
35	2023	9
36	2023	10
37	2023	11
38	2023	12
39	2024	1
40	2024	2
41	2024	3
42	2024	4
43	2024	5
44	2024	6
45	2024	7

Input Dataset A | Configuration



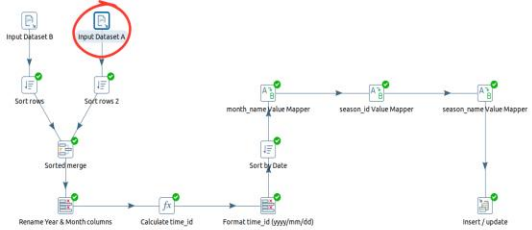
CSV file input

Step name	Input Dataset A
Filename	/home/aid/adi-project/data/21-contadores-de-energia.csv [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	

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Year	Integer	#	15	0	$	.	,	none
2	Month	Integer	#	15	0	$	.	,	none

Help | OK | Get Fields | Preview | Cancel

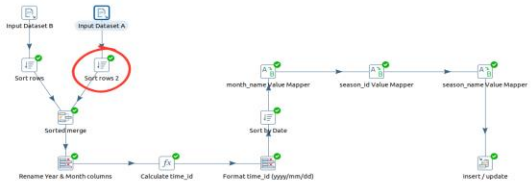
Input Dataset A | Preview



Rows of step: Input Dataset A (1000 rows)

	Year	Month
1	2024	5
2	2024	5
3	2024	5
4	2024	5
5	2024	5
6	2024	5
7	2024	5
8	2024	5
9	2024	5
10	2024	5
11	2024	5
12	2024	5
13	2024	5
14	2024	5
15	2024	5
16	2024	5
17	2024	5
18	2024	5
19	2024	5
20	2024	5
21	2024	5
22	2024	5
23	2024	5
24	2024	5
25	2024	5
26	2024	5
27	2024	5
28	2024	5
29	2024	5
30	2024	5
31	2024	5
32	2024	5
33	2024	5
34	2024	5
35	2024	5
36	2024	5
37	2024	5
38	2024	5
39	2024	5
40	2024	5
41	2024	5
42	2024	5
43	2024	5
44	2024	5
45	2024	5

Sort Rows 2 | Configuration



Sort rows

Step name: Sort by Date 2

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

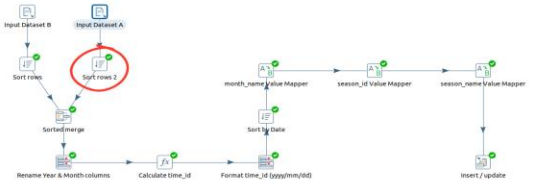
Only pass unique rows? (verifies keys only) ☑

Fields :

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

Help OK Cancel Get Fields

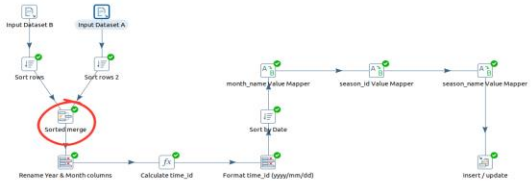
Sort Rows 2 | Preview



Rows of step: Sort rows 2 (28 rows)

	Year	Month
1	2022	6
2	2022	7
3	2022	8
4	2022	9
5	2022	10
6	2022	11
7	2022	12
8	2023	1
9	2023	2
10	2023	3
11	2023	4
12	2023	5
13	2023	6
14	2023	7
15	2023	8
16	2023	9
17	2023	10
18	2023	11
19	2023	12
20	2024	1
21	2024	2
22	2024	3
23	2024	4
24	2024	5
25	2024	6
26	2024	7
27	2024	8
28	2024	9

Sorted Merge | Configuration



Sorted merge

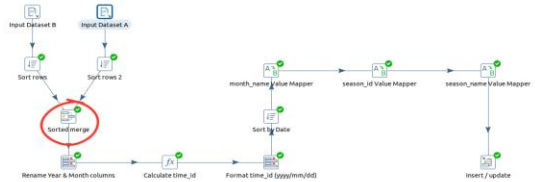
Step name: Sorted merge

Fields :

	Fieldname	Ascending
1	Year	Y
2	Month	Y

Help | OK | Cancel | Get Fields

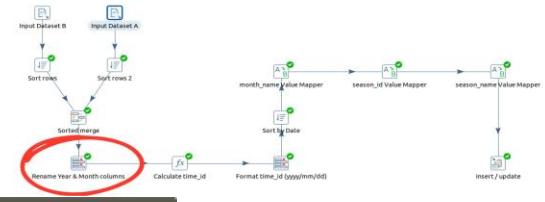
Sorted Merge | Preview



Rows of step: Sorted merge (73 rows)

	Year	Month
1	2020	11
2	2020	12
3	2021	1
4	2021	2
5	2021	3
6	2021	4
7	2021	5
8	2021	6
9	2021	7
10	2021	8
11	2021	9
12	2021	10
13	2021	11
14	2021	12
15	2022	1
16	2022	2
17	2022	3
18	2022	4
19	2022	5
20	2022	6
21	2022	6
22	2022	7
23	2022	7
24	2022	8
25	2022	8
26	2022	9
27	2022	9
28	2022	10
29	2022	10
30	2022	11
31	2022	11
32	2022	12
33	2022	12
34	2023	1
35	2023	1
36	2023	2
37	2023	2
38	2023	3
39	2023	3
40	2023	4
41	2023	4
42	2023	5
43	2023	5
44	2023	6
45	2023	6

Rename Year & Month Columns | Configuration



Select values

Step name: Rename Year & Month columns

Select & Alter | Remove | Meta-data

Fields :

	Fieldname	Rename to	Length	Precision
1	Year	year_id		
2	Month	month_id		

Get fields to select

Edit Mapping

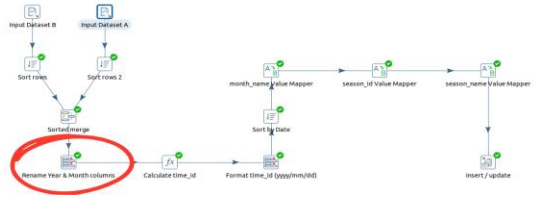
Include unspecified fields, ordered by name

Help

OK

Cancel

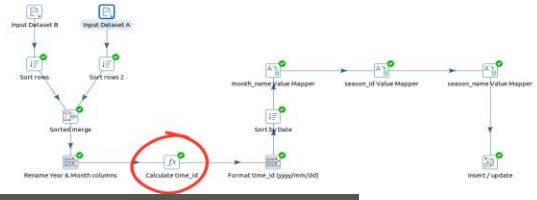
Rename Year & Month Columns | Preview



Rows of step: Rename Year Month columns (73 rows)

	year_id	month_id
1	2020	11
2	2020	12
3	2021	1
4	2021	2
5	2021	3
6	2021	4
7	2021	5
8	2021	6
9	2021	7
10	2021	8
11	2021	9
12	2021	10
13	2021	11
14	2021	12
15	2022	1
16	2022	2
17	2022	3
18	2022	4
19	2022	5
20	2022	6
21	2022	6
22	2022	7
23	2022	7
24	2022	8
25	2022	8
26	2022	9
27	2022	9
28	2022	10
29	2022	10
30	2022	11
31	2022	11
32	2022	12
33	2022	12
34	2023	1
35	2023	1
36	2023	2
37	2023	2
38	2023	3
39	2023	3
40	2023	4
41	2023	4
42	2023	5
43	2023	5
44	2023	6
45	2023	6

Calculate time_id | Configuration



Formula

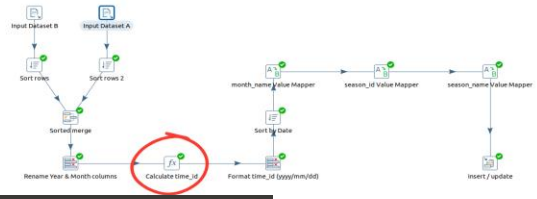
Step name: Calculate time_id

Fields:

	New field	Formula	Value type	Length	Precision	Replace value
1	time_id_unparsed	DATE([year_id];[month_id];1)	Date			

Help | OK | Cancel

Calculate time_id | Preview

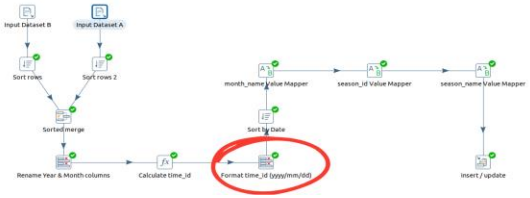


Examine preview data

Rows of step: Calculate time_id (73 rows)

	year_id	month_id	time_id_unparsed
1	2020	11	2020/11/01 00:00:00.000
2	2020	12	2020/12/01 00:00:00.000
3	2021	1	2021/01/01 00:00:00.000
4	2021	2	2021/02/01 00:00:00.000
5	2021	3	2021/03/01 00:00:00.000
6	2021	4	2021/04/01 00:00:00.000
7	2021	5	2021/05/01 00:00:00.000
8	2021	6	2021/06/01 00:00:00.000
9	2021	7	2021/07/01 00:00:00.000
10	2021	8	2021/08/01 00:00:00.000
11	2021	9	2021/09/01 00:00:00.000
12	2021	10	2021/10/01 00:00:00.000
13	2021	11	2021/11/01 00:00:00.000
14	2021	12	2021/12/01 00:00:00.000
15	2022	1	2022/01/01 00:00:00.000
16	2022	2	2022/02/01 00:00:00.000
17	2022	3	2022/03/01 00:00:00.000
18	2022	4	2022/04/01 00:00:00.000
19	2022	5	2022/05/01 00:00:00.000
20	2022	6	2022/06/01 00:00:00.000
21	2022	6	2022/06/01 00:00:00.000
22	2022	7	2022/07/01 00:00:00.000
23	2022	7	2022/07/01 00:00:00.000
24	2022	8	2022/08/01 00:00:00.000
25	2022	8	2022/08/01 00:00:00.000
26	2022	9	2022/09/01 00:00:00.000
27	2022	9	2022/09/01 00:00:00.000
28	2022	10	2022/10/01 00:00:00.000
29	2022	10	2022/10/01 00:00:00.000
30	2022	11	2022/11/01 00:00:00.000
31	2022	11	2022/11/01 00:00:00.000
32	2022	12	2022/12/01 00:00:00.000
33	2022	12	2022/12/01 00:00:00.000
34	2023	1	2023/01/01 00:00:00.000
35	2023	1	2023/01/01 00:00:00.000
36	2023	2	2023/02/01 00:00:00.000
37	2023	2	2023/02/01 00:00:00.000
38	2023	3	2023/03/01 00:00:00.000
39	2023	3	2023/03/01 00:00:00.000
40	2023	4	2023/04/01 00:00:00.000
41	2023	4	2023/04/01 00:00:00.000
42	2023	5	2023/05/01 00:00:00.000
43	2023	5	2023/05/01 00:00:00.000
44	2023	6	2023/06/01 00:00:00.000
45	2023	6	2023/06/01 00:00:00.000

Format time_id (yyyy/mm/dd) | Configuration



Select values

Step name: Format time_id (yyyy/mm/dd)

Select & Alter | Remove | Meta-data

Fields to alter the meta-data for :

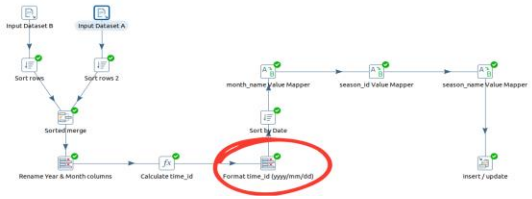
	Fieldname	Rename to	Type	Length	Precision	Binary to Normal?	Format	Date Format Lenient?	Date Locale
1	time_id_unparsed	time_id	Date			N	yyyy/MM/dd	N	

Get fields to change

Help

OK Cancel

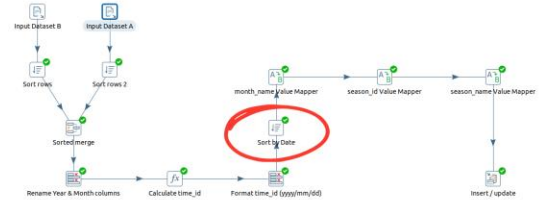
Format time_id (yyyy/mm/dd) | Preview



Rows of step: Format time_id (yyyy/mm/dd) (73 rows)

	year_id	month_id	time_id
1	2020	11	2020/11/01
2	2020	12	2020/12/01
3	2021	1	2021/01/01
4	2021	2	2021/02/01
5	2021	3	2021/03/01
6	2021	4	2021/04/01
7	2021	5	2021/05/01
8	2021	6	2021/06/01
9	2021	7	2021/07/01
10	2021	8	2021/08/01
11	2021	9	2021/09/01
12	2021	10	2021/10/01
13	2021	11	2021/11/01
14	2021	12	2021/12/01
15	2022	1	2022/01/01
16	2022	2	2022/02/01
17	2022	3	2022/03/01
18	2022	4	2022/04/01
19	2022	5	2022/05/01
20	2022	6	2022/06/01
21	2022	6	2022/06/01
22	2022	7	2022/07/01
23	2022	7	2022/07/01
24	2022	8	2022/08/01
25	2022	8	2022/08/01
26	2022	9	2022/09/01
27	2022	9	2022/09/01
28	2022	10	2022/10/01
29	2022	10	2022/10/01
30	2022	11	2022/11/01
31	2022	11	2022/11/01
32	2022	12	2022/12/01
33	2022	12	2022/12/01
34	2023	1	2023/01/01
35	2023	1	2023/01/01
36	2023	2	2023/02/01
37	2023	2	2023/02/01
38	2023	3	2023/03/01
39	2023	3	2023/03/01
40	2023	4	2023/04/01
41	2023	4	2023/04/01
42	2023	5	2023/05/01
43	2023	5	2023/05/01
44	2023	6	2023/06/01
45	2023	6	2023/06/01

Sort by Date | Configuration



Sort rows

Step name: Sort by Date

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

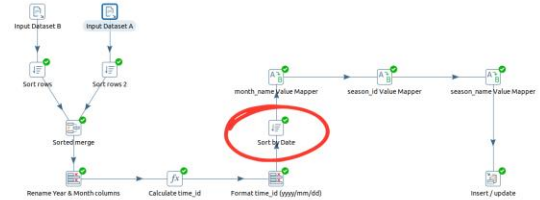
Only pass unique rows? (verifies keys only) ☑

Fields :

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

Help OK Cancel Get Fields

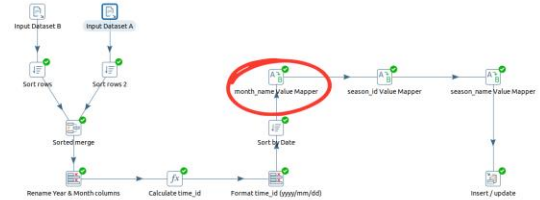
Sort by Date | Preview



Rows of step: Sort by Date (47 rows)

	year_id	month_id	time_id
1	2020	11	2020/11/01
2	2020	12	2020/12/01
3	2021	1	2021/01/01
4	2021	2	2021/02/01
5	2021	3	2021/03/01
6	2021	4	2021/04/01
7	2021	5	2021/05/01
8	2021	6	2021/06/01
9	2021	7	2021/07/01
10	2021	8	2021/08/01
11	2021	9	2021/09/01
12	2021	10	2021/10/01
13	2021	11	2021/11/01
14	2021	12	2021/12/01
15	2022	1	2022/01/01
16	2022	2	2022/02/01
17	2022	3	2022/03/01
18	2022	4	2022/04/01
19	2022	5	2022/05/01
20	2022	6	2022/06/01
21	2022	7	2022/07/01
22	2022	8	2022/08/01
23	2022	9	2022/09/01
24	2022	10	2022/10/01
25	2022	11	2022/11/01
26	2022	12	2022/12/01
27	2023	1	2023/01/01
28	2023	2	2023/02/01
29	2023	3	2023/03/01
30	2023	4	2023/04/01
31	2023	5	2023/05/01
32	2023	6	2023/06/01
33	2023	7	2023/07/01
34	2023	8	2023/08/01
35	2023	9	2023/09/01
36	2023	10	2023/10/01
37	2023	11	2023/11/01
38	2023	12	2023/12/01
39	2024	1	2024/01/01
40	2024	2	2024/02/01
41	2024	3	2024/03/01
42	2024	4	2024/04/01
43	2024	5	2024/05/01
44	2024	6	2024/06/01
45	2024	7	2024/07/01

month_name Value Mapper | Configuration



Value mapper

Step name : month_name Value Mapper

Fieldname to use : month_id

Target field name (empty=overwrite) : month_name

Default upon non-matching :

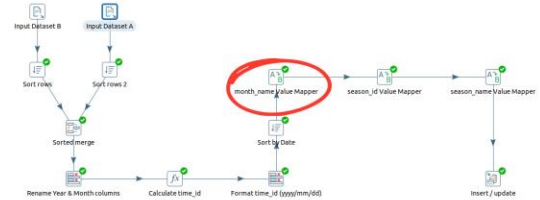
Field values:

	Source value	Target value
1	1	January
2	2	February
3	3	March
4	4	April
5	5	May
6	6	June
7	7	July
8	8	August
9	9	September
10	10	October
11	11	November
12	12	December

Help

OK Cancel

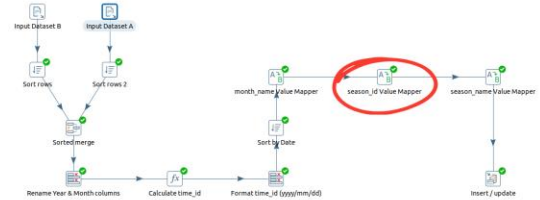
month_name Value Mapper | Preview



Rows of step: month_name Value Mapper (47 rows)

	year_id	month_id	time_id	month_name
1	2020	11	2020/11/01	November
2	2020	12	2020/12/01	December
3	2021	1	2021/01/01	January
4	2021	2	2021/02/01	February
5	2021	3	2021/03/01	March
6	2021	4	2021/04/01	April
7	2021	5	2021/05/01	May
8	2021	6	2021/06/01	June
9	2021	7	2021/07/01	July
10	2021	8	2021/08/01	August
11	2021	9	2021/09/01	September
12	2021	10	2021/10/01	October
13	2021	11	2021/11/01	November
14	2021	12	2021/12/01	December
15	2022	1	2022/01/01	January
16	2022	2	2022/02/01	February
17	2022	3	2022/03/01	March
18	2022	4	2022/04/01	April
19	2022	5	2022/05/01	May
20	2022	6	2022/06/01	June
21	2022	7	2022/07/01	July
22	2022	8	2022/08/01	August
23	2022	9	2022/09/01	September
24	2022	10	2022/10/01	October
25	2022	11	2022/11/01	November
26	2022	12	2022/12/01	December
27	2023	1	2023/01/01	January
28	2023	2	2023/02/01	February
29	2023	3	2023/03/01	March
30	2023	4	2023/04/01	April
31	2023	5	2023/05/01	May
32	2023	6	2023/06/01	June
33	2023	7	2023/07/01	July
34	2023	8	2023/08/01	August
35	2023	9	2023/09/01	September
36	2023	10	2023/10/01	October
37	2023	11	2023/11/01	November
38	2023	12	2023/12/01	December
39	2024	1	2024/01/01	January
40	2024	2	2024/02/01	February
41	2024	3	2024/03/01	March
42	2024	4	2024/04/01	April
43	2024	5	2024/05/01	May
44	2024	6	2024/06/01	June
45	2024	7	2024/07/01	July

season_id Value Mapper | Preview



Value mapper

Step name : season_id Value Mapper

Fieldname to use : month_id

Target field name (empty=overwrite) : season_id

Default upon non-matching :

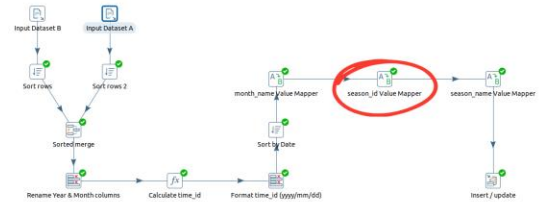
Field values:

	Source value	Target value
1	3	1
2	4	1
3	5	1
4	6	2
5	7	2
6	8	2
7	9	3
8	10	3
9	11	3
10	12	4
11	1	4
12	2	4

Help

OK Cancel

season_id Value Mapper | Configuration

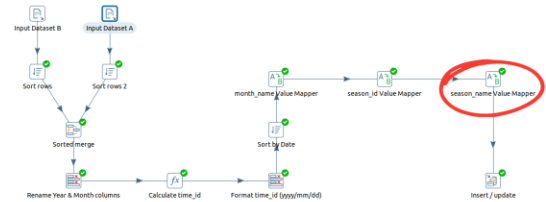


Examine preview data

Rows of step: season_id Value Mapper (47 rows)

	year_id	month_id	time_id	month_name	season_id
1	2020	11	2020/11/01	November	3
2	2020	12	2020/12/01	December	4
3	2021	1	2021/01/01	January	4
4	2021	2	2021/02/01	February	4
5	2021	3	2021/03/01	March	1
6	2021	4	2021/04/01	April	1
7	2021	5	2021/05/01	May	1
8	2021	6	2021/06/01	June	2
9	2021	7	2021/07/01	July	2
10	2021	8	2021/08/01	August	2
11	2021	9	2021/09/01	September	3
12	2021	10	2021/10/01	October	3
13	2021	11	2021/11/01	November	3
14	2021	12	2021/12/01	December	4
15	2022	1	2022/01/01	January	4
16	2022	2	2022/02/01	February	4
17	2022	3	2022/03/01	March	1
18	2022	4	2022/04/01	April	1
19	2022	5	2022/05/01	May	1
20	2022	6	2022/06/01	June	2
21	2022	7	2022/07/01	July	2
22	2022	8	2022/08/01	August	2
23	2022	9	2022/09/01	September	3
24	2022	10	2022/10/01	October	3
25	2022	11	2022/11/01	November	3
26	2022	12	2022/12/01	December	4
27	2023	1	2023/01/01	January	4
28	2023	2	2023/02/01	February	4
29	2023	3	2023/03/01	March	1
30	2023	4	2023/04/01	April	1
31	2023	5	2023/05/01	May	1
32	2023	6	2023/06/01	June	2
33	2023	7	2023/07/01	July	2
34	2023	8	2023/08/01	August	2
35	2023	9	2023/09/01	September	3
36	2023	10	2023/10/01	October	3
37	2023	11	2023/11/01	November	3
38	2023	12	2023/12/01	December	4
39	2024	1	2024/01/01	January	4
40	2024	2	2024/02/01	February	4
41	2024	3	2024/03/01	March	1
42	2024	4	2024/04/01	April	1
43	2024	5	2024/05/01	May	1
44	2024	6	2024/06/01	June	2
45	2024	7	2024/07/01	July	2

season_name Value Mapper | Configuration



Value mapper

Step name : season_name Value Mapper

Fieldname to use : season_id

Target field name (empty=overwrite) : season_name

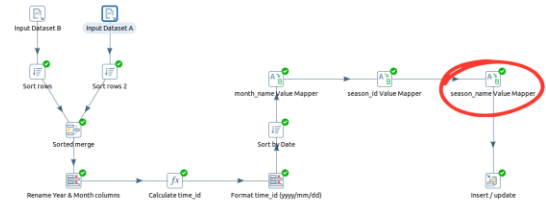
Default upon non-matching :

Field values:

	Source value	Target value
1	1	Spring
2	2	Summer
3	3	Autumn
4	4	Winter

Help OK Cancel

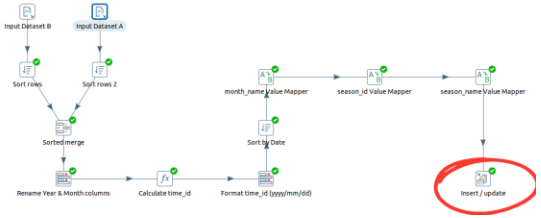
season_name Value Mapper | Preview



Rows of step: season_name Value Mapper (47 rows)

	year_id	month_id	time_id	month_name	season_id	season_name
1	2020	11	2020/11/01	November	3	Autumn
2	2020	12	2020/12/01	December	4	Winter
3	2021	1	2021/01/01	January	4	Winter
4	2021	2	2021/02/01	February	4	Winter
5	2021	3	2021/03/01	March	1	Spring
6	2021	4	2021/04/01	April	1	Spring
7	2021	5	2021/05/01	May	1	Spring
8	2021	6	2021/06/01	June	2	Summer
9	2021	7	2021/07/01	July	2	Summer
10	2021	8	2021/08/01	August	2	Summer
11	2021	9	2021/09/01	September	3	Autumn
12	2021	10	2021/10/01	October	3	Autumn
13	2021	11	2021/11/01	November	3	Autumn
14	2021	12	2021/12/01	December	4	Winter
15	2022	1	2022/01/01	January	4	Winter
16	2022	2	2022/02/01	February	4	Winter
17	2022	3	2022/03/01	March	1	Spring
18	2022	4	2022/04/01	April	1	Spring
19	2022	5	2022/05/01	May	1	Spring
20	2022	6	2022/06/01	June	2	Summer
21	2022	7	2022/07/01	July	2	Summer
22	2022	8	2022/08/01	August	2	Summer
23	2022	9	2022/09/01	September	3	Autumn
24	2022	10	2022/10/01	October	3	Autumn
25	2022	11	2022/11/01	November	3	Autumn
26	2022	12	2022/12/01	December	4	Winter
27	2023	1	2023/01/01	January	4	Winter
28	2023	2	2023/02/01	February	4	Winter
29	2023	3	2023/03/01	March	1	Spring
30	2023	4	2023/04/01	April	1	Spring
31	2023	5	2023/05/01	May	1	Spring
32	2023	6	2023/06/01	June	2	Summer
33	2023	7	2023/07/01	July	2	Summer
34	2023	8	2023/08/01	August	2	Summer
35	2023	9	2023/09/01	September	3	Autumn
36	2023	10	2023/10/01	October	3	Autumn
37	2023	11	2023/11/01	November	3	Autumn
38	2023	12	2023/12/01	December	4	Winter
39	2024	1	2024/01/01	January	4	Winter
40	2024	2	2024/02/01	February	4	Winter
41	2024	3	2024/03/01	March	1	Spring
42	2024	4	2024/04/01	April	1	Spring
43	2024	5	2024/05/01	May	1	Spring
44	2024	6	2024/06/01	June	2	Summer
45	2024	7	2024/07/01	July	2	Summer

Insert / update | Configuration



Insert / update

Step name: Insert / update

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

Target schema: energy_dw [Browse...]

Target table: dim_time [Browse...]

Commit size: 100

Don't perform any updates: ☐

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

	Table field	Comparator	Stream field1	Stream field2
1	time_id	=	time_id	

[Get fields]

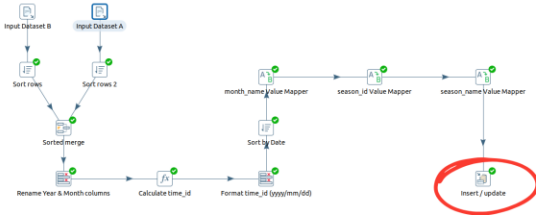
Update fields:

	Table field	Stream field	Update
1	year_id	year_id	Y
2	month_id	month_id	Y
3	month_name	month_name	Y
4	season_id	season_id	Y
5	season_name	season_name	Y
6	time_id	time_id	Y

[Get update fields] [Edit mapping]

[Help] [OK] [Cancel] [SQL]

Insert / update | Preview



Rows of step: Insert / update (47 rows)

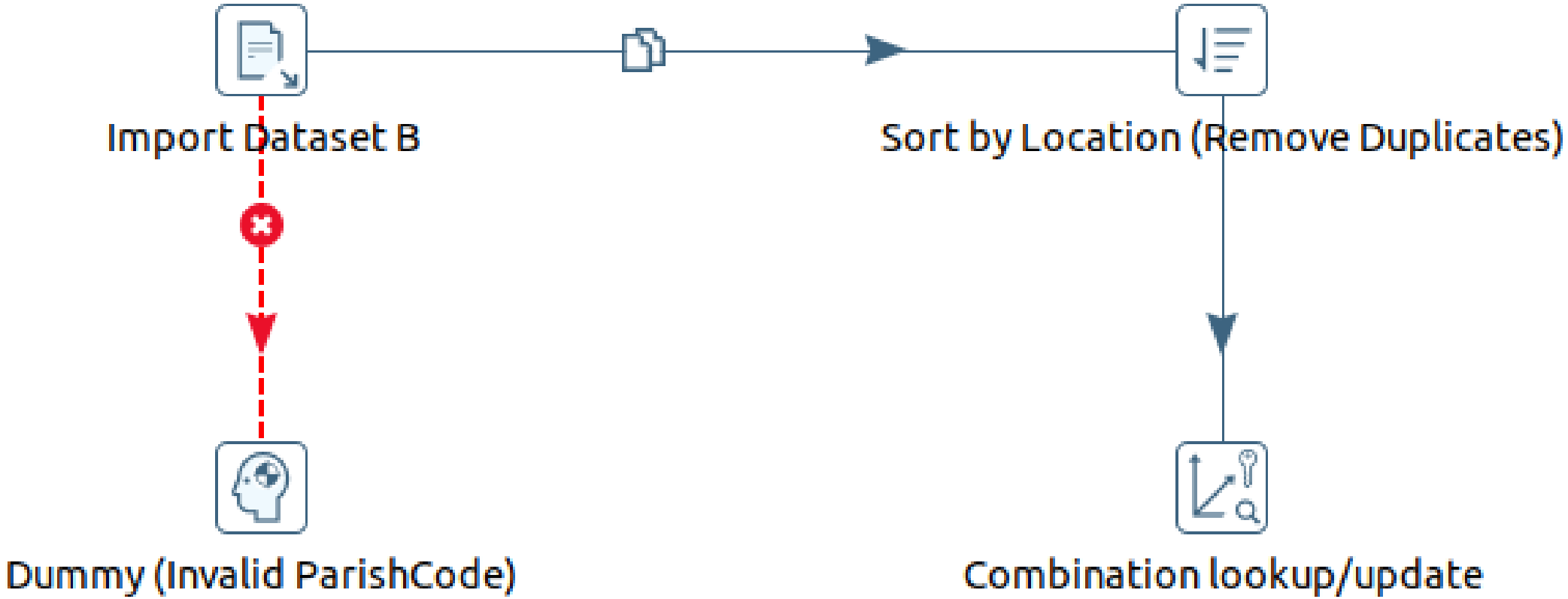
	year_id	month_id	time_id	month_name	season_id	season_name
1	2020	11	2020/11/01	November	3	Autumn
2	2020	12	2020/12/01	December	4	Winter
3	2021	1	2021/01/01	January	4	Winter
4	2021	2	2021/02/01	February	4	Winter
5	2021	3	2021/03/01	March	1	Spring
6	2021	4	2021/04/01	April	1	Spring
7	2021	5	2021/05/01	May	1	Spring
8	2021	6	2021/06/01	June	2	Summer
9	2021	7	2021/07/01	July	2	Summer
10	2021	8	2021/08/01	August	2	Summer
11	2021	9	2021/09/01	September	3	Autumn
12	2021	10	2021/10/01	October	3	Autumn
13	2021	11	2021/11/01	November	3	Autumn
14	2021	12	2021/12/01	December	4	Winter
15	2022	1	2022/01/01	January	4	Winter
16	2022	2	2022/02/01	February	4	Winter
17	2022	3	2022/03/01	March	1	Spring
18	2022	4	2022/04/01	April	1	Spring
19	2022	5	2022/05/01	May	1	Spring
20	2022	6	2022/06/01	June	2	Summer
21	2022	7	2022/07/01	July	2	Summer
22	2022	8	2022/08/01	August	2	Summer
23	2022	9	2022/09/01	September	3	Autumn
24	2022	10	2022/10/01	October	3	Autumn
25	2022	11	2022/11/01	November	3	Autumn
26	2022	12	2022/12/01	December	4	Winter
27	2023	1	2023/01/01	January	4	Winter
28	2023	2	2023/02/01	February	4	Winter
29	2023	3	2023/03/01	March	1	Spring
30	2023	4	2023/04/01	April	1	Spring
31	2023	5	2023/05/01	May	1	Spring
32	2023	6	2023/06/01	June	2	Summer
33	2023	7	2023/07/01	July	2	Summer
34	2023	8	2023/08/01	August	2	Summer
35	2023	9	2023/09/01	September	3	Autumn
36	2023	10	2023/10/01	October	3	Autumn
37	2023	11	2023/11/01	November	3	Autumn
38	2023	12	2023/12/01	December	4	Winter
39	2024	1	2024/01/01	January	4	Winter
40	2024	2	2024/02/01	February	4	Winter
41	2024	3	2024/03/01	March	1	Spring
42	2024	4	2024/04/01	April	1	Spring
43	2024	5	2024/05/01	May	1	Spring
44	2024	6	2024/06/01	June	2	Summer
45	2024	7	2024/07/01	July	2	Summer

Output Table

```
mysql> SELECT * FROM energy_dw.dim_time;
+------------+---------+-----------+-------------+----------+------------+
| time_id    | year_id | season_id | season_name | month_id | month_name |
+------------+---------+-----------+-------------+----------+------------+
| 2020-11-01 |    2020 |         3 | Autumn      |       11 | November   |
| 2020-12-01 |    2020 |         4 | Winter      |       12 | December   |
| 2021-01-01 |    2021 |         4 | Winter      |        1 | January    |
| 2021-02-01 |    2021 |         4 | Winter      |        2 | February   |
| 2021-03-01 |    2021 |         1 | Spring      |        3 | March      |
| 2021-04-01 |    2021 |         1 | Spring      |        4 | April      |
| 2021-05-01 |    2021 |         1 | Spring      |        5 | May        |
| 2021-06-01 |    2021 |         2 | Summer      |        6 | June       |
| 2021-07-01 |    2021 |         2 | Summer      |        7 | July       |
| 2021-08-01 |    2021 |         2 | Summer      |        8 | August     |
| 2021-09-01 |    2021 |         3 | Autumn      |        9 | September  |
| 2021-10-01 |    2021 |         3 | Autumn      |       10 | October    |
| 2021-11-01 |    2021 |         3 | Autumn      |       11 | November   |
| 2021-12-01 |    2021 |         4 | Winter      |       12 | December   |
| 2022-01-01 |    2022 |         4 | Winter      |        1 | January    |
| 2022-02-01 |    2022 |         4 | Winter      |        2 | February   |
| 2022-03-01 |    2022 |         1 | Spring      |        3 | March      |
| 2022-04-01 |    2022 |         1 | Spring      |        4 | April      |
| 2022-05-01 |    2022 |         1 | Spring      |        5 | May        |
| 2022-06-01 |    2022 |         2 | Summer      |        6 | June       |
| 2022-07-01 |    2022 |         2 | Summer      |        7 | July       |
| 2022-08-01 |    2022 |         2 | Summer      |        8 | August     |
| 2022-09-01 |    2022 |         3 | Autumn      |        9 | September  |
| 2022-10-01 |    2022 |         3 | Autumn      |       10 | October    |
| 2022-11-01 |    2022 |         3 | Autumn      |       11 | November   |
| 2022-12-01 |    2022 |         4 | Winter      |       12 | December   |
| 2023-01-01 |    2023 |         4 | Winter      |        1 | January    |
| 2023-02-01 |    2023 |         4 | Winter      |        2 | February   |
| 2023-03-01 |    2023 |         1 | Spring      |        3 | March      |
| 2023-04-01 |    2023 |         1 | Spring      |        4 | April      |
| 2023-05-01 |    2023 |         1 | Spring      |        5 | May        |
| 2023-06-01 |    2023 |         2 | Summer      |        6 | June       |
| 2023-07-01 |    2023 |         2 | Summer      |        7 | July       |
| 2023-08-01 |    2023 |         2 | Summer      |        8 | August     |
| 2023-09-01 |    2023 |         3 | Autumn      |        9 | September  |
| 2023-10-01 |    2023 |         3 | Autumn      |       10 | October    |
| 2023-11-01 |    2023 |         3 | Autumn      |       11 | November   |
| 2023-12-01 |    2023 |         4 | Winter      |       12 | December   |
| 2024-01-01 |    2024 |         4 | Winter      |        1 | January    |
| 2024-02-01 |    2024 |         4 | Winter      |        2 | February   |
| 2024-03-01 |    2024 |         1 | Spring      |        3 | March      |
| 2024-04-01 |    2024 |         1 | Spring      |        4 | April      |
| 2024-05-01 |    2024 |         1 | Spring      |        5 | May        |
| 2024-06-01 |    2024 |         2 | Summer      |        6 | June       |
| 2024-07-01 |    2024 |         2 | Summer      |        7 | July       |
| 2024-08-01 |    2024 |         2 | Summer      |        8 | August     |
| 2024-09-01 |    2024 |         3 | Autumn      |        9 | September  |
+------------+---------+-----------+-------------+----------+------------+
47 rows in set (0.00 sec)
```

Location Dimension Population Transformation

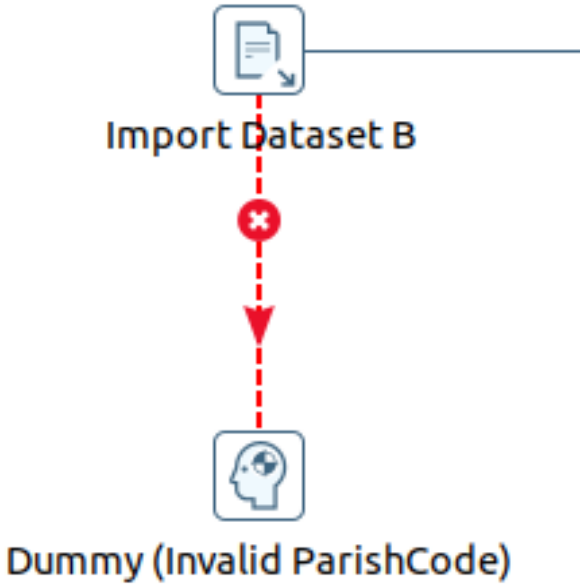
Transformation



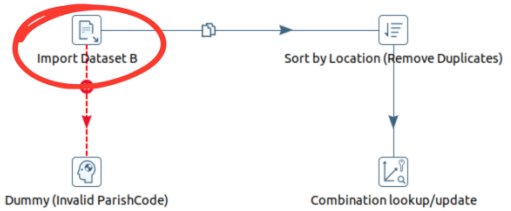
Considerations on Location Dimension Populate Transformation

- **Using only Dataset B as source for location dimension**
  - In the time dimension table, we used both Datasets A & B as our sources to populate the table
  - However, this is unnecessary in the location dimension, as Dataset B contains all the parishes that Dataset A has (plus 3 extra parishes)

- **Handling errors in importing Dataset B**
  - While importing Dataset B, there are some errors due to invalid entries (DistrictMunicipalityParishCode is imported as Integer, but some invalid entries have it as String, and the conversion fails)
  - To handle this, we redirect the invalid entries to a dummy step



Import Dataset B | Configuration



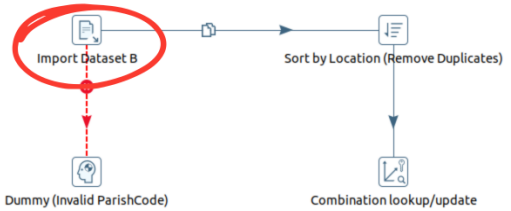
CSV file input

Step name	Import Dataset B
Filename	/home/aid/adi-project/data/3-consumos-faturados-por-municipio-ultimos-10-anos.csv [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	

	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	District	String		16		$	.	,	none
2	Municipality	String		22		$	.	,	none
3	parish	String		30		$	.	,	none
4	DistrictMunicipalityParishCode	Integer							none

Help | OK | Get Fields | Preview | Cancel

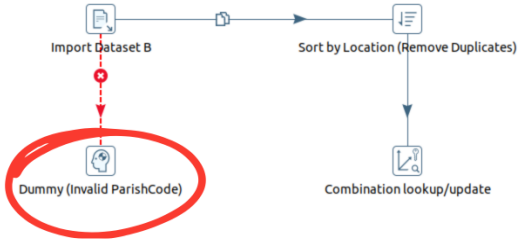
Import Dataset B | Preview



Rows of step: Import Dataset B (1000 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	Évora	Vila Viçosa	N S CONCEICAO E SAO BARTOLOMEU	71406
2	Faro	Albufeira	PADERNE	80103
3	Faro	Alcoutim	GIOES	80202
4	Faro	Alcoutim	UF ALCOUTIM E PEREIRO	80206
5	Faro	Alcoutim	UF ALCOUTIM E PEREIRO	80206
6	Faro	Aljezur	ODECEIXE	80303
7	Faro	Castro Marim	AZINHAL	80401
8	Faro	Faro	SANTA BARBARA DE NEXE	80503
9	Faro	Faro	UF CONCEICAO E ESTOI	80507
10	Faro	Lagoa	UF ESTOMBAR E PARCHAL	80607
11	Faro	Lagos	ODIAXERE	80704
12	Faro	Loulé	LOULE (SAO CLEMENTE)	80808
13	Faro	Loulé	UF QUERENCA TOR E BENAFIM	80812
14	Faro	Olhão	OLHAO	81003
15	Faro	São Brás de Alportel	SAO BRAS DE ALPORTEL	81201
16	Faro	Vila do Bispo	SAGRES	81504
17	Faro	Vila Real de Santo António	VILA NOVA DE CACELA	81601
18	Guarda	Aguiar da Beira	PINHEIRO	90110
19	Guarda	Almeida	MALHADA SORDA	90213
20	Guarda	Almeida	UF JUNCA E NAVES	90233
21	Guarda	Celorico da Beira	MESQUITELA	90310
22	Guarda	Figueira de Castelo Rodrigo	UF COLMEAL E VILAR TORPIM	90422
23	Guarda	Fornos de Algodres	MUXAGATA	90511
24	Guarda	Gouveia	SAO PAIO	90617
25	Guarda	Gouveia	UF ALDEIAS MANGUALDE SERRA	90623
26	Guarda	Gouveia	UF MELO E NABAIS	90626
27	Guarda	Guarda	GONCALO BOCAS	90721
28	Guarda	Guarda	PANOIAS DE CIMA	90728
29	Guarda	Guarda	GUARDA	90758
30	Guarda	Guarda	UF MIZARELA P SOARES V SOEIRO	90763
31	Guarda	Manteigas	MANTEIGAS (SANTA MARIA)	90802
32	Guarda	Pinhel	ALVERCA DA BEIRA/BOUCA COVA	91029
33	Guarda	Sabugal	BENDADA	91110
34	Guarda	Sabugal	SOUTO	91134
35	Guarda	Seia	SANTIAGO	91216
36	Guarda	Seia	VALEZIM	91225
37	Guarda	Trancoso	CASTANHEIRA	91303
38	Guarda	Trancoso	POVOA DO CONCELHO	91314
39	Guarda	Trancoso	VALDUJO	91325
40	Guarda	Trancoso	UF TRANCOSO E SOUTO MAIOR	91332
41	Leiria	Alcobaça	BARRIO	100104
42	Leiria	Alcobaça	CELA	100106
43	Leiria	Alcobaça	VIMEIRO	100116
44	Leiria	Alcobaça	ALJUBARROTA	100120
45	Leiria	Ansião	ANSIAO	100309

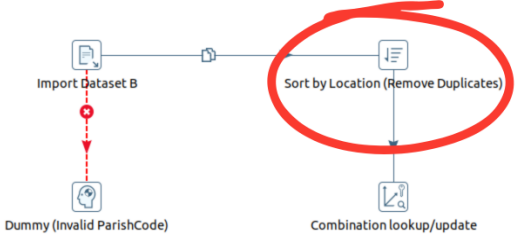
Dummy (Invalid ParishCode) | Preview



Rows of step: Dummy (Invalid ParishCode) (1000 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	Viana do Castelo	OUTROS VIANA DO CASTELO	OUTROS VIANA DO CASTELO	<null>
2	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>
3	Coimbra	OUTROS COIMBRA	OUTROS COIMBRA	<null>
4	Lisboa	OUTROS LISBOA	OUTROS LISBOA	<null>
5	Porto	OUTROS PORTO	OUTROS PORTO	<null>
6	Porto	OUTROS PORTO	OUTROS PORTO	<null>
7	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>
8	Viana do Castelo	OUTROS VIANA DO CASTELO	OUTROS VIANA DO CASTELO	<null>
9	Viseu	OUTROS VISEU	OUTROS VISEU	<null>
10	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>
11	Beja	OUTROS BEJA	OUTROS BEJA	<null>
12	Braga	OUTROS BRAGA	OUTROS BRAGA	<null>
13	Braga	Barcelos	UF MILHAZES VILAR FIGOS FARIA	<null>
14	Braga	Barcelos	UF MILHAZES VILAR FIGOS FARIA	<null>
15	Braga	Barcelos	UF QUINTIAES E AGUIAR	<null>
16	Braga	Barcelos	UF SILVEIROS E RIO COVO	<null>
17	Braga	Barcelos	UF TAMEL E VILAR DO MONTE	<null>
18	Bragança	OUTROS BRAGANCA	OUTROS BRAGANCA	<null>
19	Évora	OUTROS EVORA	OUTROS EVORA	<null>
20	Porto	OUTROS PORTO	OUTROS PORTO	<null>
21	Vila Real	OUTROS VILA REAL	OUTROS VILA REAL	<null>
22	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>
23	Beja	OUTROS BEJA	OUTROS BEJA	<null>
24	Braga	OUTROS BRAGA	OUTROS BRAGA	<null>
25	Braga	Barcelos	UF MILHAZES VILAR FIGOS FARIA	<null>
26	Braga	Barcelos	UF QUINTIAES E AGUIAR	<null>
27	Braga	Barcelos	UF SILVEIROS E RIO COVO	<null>
28	Braga	Barcelos	UF VIATOD GRIMANCEL MINHOT FRA	<null>
29	Bragança	OUTROS BRAGANCA	OUTROS BRAGANCA	<null>
30	Castelo Branco	OUTROS CASTELO BRANCO	OUTROS CASTELO BRANCO	<null>
31	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>
32	Braga	Barcelos	UF MILHAZES VILAR FIGOS FARIA	<null>
33	Braga	Barcelos	UF MILHAZES VILAR FIGOS FARIA	<null>
34	Braga	Barcelos	UF NEGREIROS E CHAVAO	<null>
35	Braga	Barcelos	UF SILVEIROS E RIO COVO	<null>
36	Braga	Barcelos	UF SILVEIROS E RIO COVO	<null>
37	Braga	Barcelos	UF TAMEL E VILAR DO MONTE	<null>
38	Braga	Barcelos	UF VIATOD GRIMANCEL MINHOT FRA	<null>
39	Coimbra	OUTROS COIMBRA	OUTROS COIMBRA	<null>
40	Faro	OUTROS FARO	OUTROS FARO	<null>
41	Guarda	OUTROS GUARDA	OUTROS GUARDA	<null>
42	Portalegre	OUTROS PORTALEGRE	OUTROS PORTALEGRE	<null>
43	Viana do Castelo	OUTROS VIANA DO CASTELO	OUTROS VIANA DO CASTELO	<null>
44	Viseu	OUTROS VISEU	OUTROS VISEU	<null>
45	Aveiro	OUTROS AVEIRO	OUTROS AVEIRO	<null>

Sort by Location (Remove Duplicates) - Configuration



Sort rows

Step name: Sort by Location (Remove Duplicates)

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

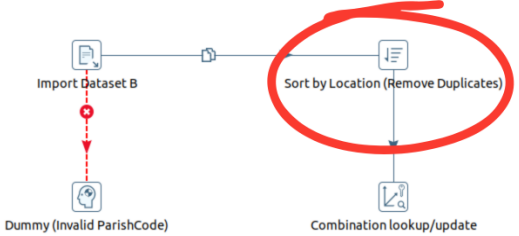
Only pass unique rows? (verifies keys only) ☑

Fields :

	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength
1	District	Y	N	N	0
2	Municipality	Y	N	N	0
3	parish	Y	N	N	0

Help | OK | Cancel | Get Fields

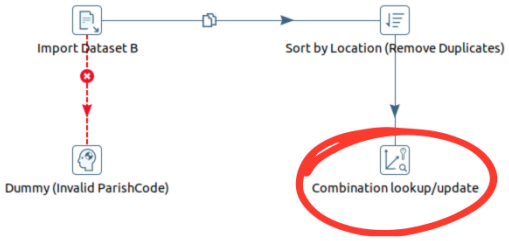
Sort by Location (Remove Duplicates) - Preview



Rows of step: Sort by Location (Remove Duplicates) (1000 rows)

	District	Municipality	parish	DistrictMunicipalityParishCode
1	Aveiro	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	10209
2	Aveiro	Albergaria-a-Velha	ALQUERUBIM	10202
3	Aveiro	Albergaria-a-Velha	ANGEJA	10203
4	Aveiro	Albergaria-a-Velha	BRANCA	10204
5	Aveiro	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	10206
6	Aveiro	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	10210
7	Aveiro	Anadia	AVELAS DE CAMINHO	10304
8	Aveiro	Anadia	AVELAS DE CIMA	10305
9	Aveiro	Anadia	MOITA	10307
10	Aveiro	Anadia	SANGALHOS	10309
11	Aveiro	Anadia	SAO LOURENCO DO BAIRRO	10310
12	Aveiro	Anadia	UF A DA GANDARA BAIRRO E ANCAS	10316
13	Aveiro	Anadia	UF ARCOS E MOGOFORES	10317
14	Aveiro	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	10318
15	Aveiro	Anadia	VILA NOVA DE MONSARROS	10312
16	Aveiro	Anadia	VILARINHO DO BAIRRO	10313
17	Aveiro	Arouca	ALVARENGA	10402
18	Aveiro	Arouca	CHAVE	10407
19	Aveiro	Arouca	ESCARIZ	10409
20	Aveiro	Arouca	FERMEDO	10411
21	Aveiro	Arouca	MANSORES	10413
22	Aveiro	Arouca	MOLDES	10414
23	Aveiro	Arouca	ROSSAS	10415
24	Aveiro	Arouca	SANTA EULALIA	10416
25	Aveiro	Arouca	SAO MIGUEL DO MATO	10417
26	Aveiro	Arouca	TROPECO	10418
27	Aveiro	Arouca	UF AROUCA E BURGO	10421
28	Aveiro	Arouca	UF CABREIROS ALBERGARIA SERRA	10422
29	Aveiro	Arouca	UF CANELAS E ESPIUNCA	10423
30	Aveiro	Arouca	UF COVELO DE PAIVO E JANARDE	10424
31	Aveiro	Arouca	URRO	10419
32	Aveiro	Arouca	VARZEA	10420
33	Aveiro	Aveiro	ARADAS	10501
34	Aveiro	Aveiro	CACIA	10502
35	Aveiro	Aveiro	EIXO E EIROL	10515
36	Aveiro	Aveiro	ESGUEIRA	10505
37	Aveiro	Aveiro	OLIVEIRINHA	10508
38	Aveiro	Aveiro	REQUEIXO N S FATIMA E NARIZ	10516
39	Aveiro	Aveiro	SANTA JOANA	10513
40	Aveiro	Aveiro	SAO BERNARDO	10510
41	Aveiro	Aveiro	SAO JACINTO	10511
42	Aveiro	Aveiro	UF GLORIA E VERA CRUZ	10517
43	Aveiro	Castelo de Paiva	FORNOS	10602
44	Aveiro	Castelo de Paiva	REAL	10606
45	Aveiro	Castelo de Paiva	SANTA MARIA DE SARDOURA	10607

Combination lookup/update | Configuration



Combination lookup/update

Step name: Combination lookup/update

Connection: energy_dw — Edit... New... Wizard...

Target schema: energy_dw — Browse...

Target table: dim_location — Browse...

Commit size: 100

Cache size: 9999

Pre-load the cache? ☐

Key fields (to look up row in table):

	Dimension field	Field in stream
1	region	District
2	municipality	Municipality
3	parish	parish
4	parish_code	DistrictMunicipalityParishCode

Technical key field: location_id

Creation of technical key

- ○ Use table maximum + 1
- ○ Use sequence
- ◉ Use auto increment field

Remove lookup fields? ☐

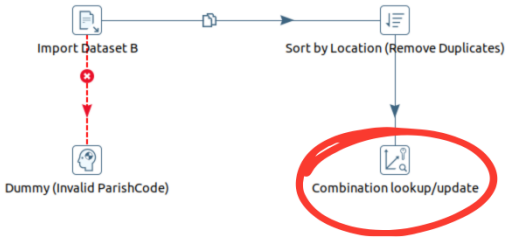
Use hashcode? ☐

Hashcode field in table

Date of last update field (optional)

Help | OK | Cancel | Get Fields | SQL

Combination lookup/update | Preview



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

	District	Municipality	parish	DistrictMunicipalityParishCode	location_id
1	Aveiro	Albergaria-a-Velha	ALBERGARIA-A-VELHA E VALMAIOR	10209	1
2	Aveiro	Albergaria-a-Velha	ALQUERUBIM	10202	2
3	Aveiro	Albergaria-a-Velha	ANGEJA	10203	3
4	Aveiro	Albergaria-a-Velha	BRANCA	10204	4
5	Aveiro	Albergaria-a-Velha	RIBEIRA DE FRAGUAS	10206	5
6	Aveiro	Albergaria-a-Velha	SAO JOAO DE LOURE E FROSSOS	10210	6
7	Aveiro	Anadia	AVELAS DE CAMINHO	10304	7
8	Aveiro	Anadia	AVELAS DE CIMA	10305	8
9	Aveiro	Anadia	MOITA	10307	9
10	Aveiro	Anadia	SANGALHOS	10309	10
11	Aveiro	Anadia	SAO LOURENCO DO BAIRRO	10310	11
12	Aveiro	Anadia	UF A DA GANDARA BAIRRO E ANCAS	10316	12
13	Aveiro	Anadia	UF ARCOS E MOGOFORES	10317	13
14	Aveiro	Anadia	UF TAMENGOS AGUIM OIS BAIRRO	10318	14
15	Aveiro	Anadia	VILA NOVA DE MONSARROS	10312	15
16	Aveiro	Anadia	VILARINHO DO BAIRRO	10313	16
17	Aveiro	Arouca	ALVARENGA	10402	17
18	Aveiro	Arouca	CHAVE	10407	18
19	Aveiro	Arouca	ESCARIZ	10409	19
20	Aveiro	Arouca	FERMEDO	10411	20
21	Aveiro	Arouca	MANSORES	10413	21
22	Aveiro	Arouca	MOLDES	10414	22
23	Aveiro	Arouca	ROSSAS	10415	23
24	Aveiro	Arouca	SANTA EULALIA	10416	24
25	Aveiro	Arouca	SAO MIGUEL DO MATO	10417	25
26	Aveiro	Arouca	TROPECO	10418	26
27	Aveiro	Arouca	UF AROUCA E BURGO	10421	27
28	Aveiro	Arouca	UF CABREIROS ALBERGARIA SERRA	10422	28
29	Aveiro	Arouca	UF CANELAS E ESPIUNCA	10423	29
30	Aveiro	Arouca	UF COVELO DE PAIVO E JANARDE	10424	30
31	Aveiro	Arouca	URRO	10419	31
32	Aveiro	Arouca	VARZEA	10420	32
33	Aveiro	Aveiro	ARADAS	10501	33
34	Aveiro	Aveiro	CACIA	10502	34
35	Aveiro	Aveiro	EIXO E EIROL	10515	35
36	Aveiro	Aveiro	ESGUEIRA	10505	36
37	Aveiro	Aveiro	OLIVEIRINHA	10508	37
38	Aveiro	Aveiro	REQUEIXO N S FATIMA E NARIZ	10516	38
39	Aveiro	Aveiro	SANTA JOANA	10513	39
40	Aveiro	Aveiro	SAO BERNARDO	10510	40
41	Aveiro	Aveiro	SAO JACINTO	10511	41
42	Aveiro	Aveiro	UF GLORIA E VERA CRUZ	10517	42
43	Aveiro	Castelo de Paiva	FORNOS	10602	43
44	Aveiro	Castelo de Paiva	REAL	10606	44
45	Aveiro	Castelo de Paiva	SANTA MARIA DE SARDOURA	10607	45

Output Table

```
mysql> SELECT * FROM dim_location;
+-------------+-----------+-----------------------------+-------------------------------+-------------+
| location_id | region    | municipality                | parish                        | parish_code |
+-------------+-----------+-----------------------------+-------------------------------+-------------+
|           1 | Aveiro    | Albergaria-a-Velha          | ALBERGARIA-A-VELHA E VALMAIOR |       10209 |
|           2 | Aveiro    | Albergaria-a-Velha          | ALQUERUBIM                    |       10202 |
|           3 | Aveiro    | Albergaria-a-Velha          | ANGEJA                        |       10203 |
|           4 | Aveiro    | Albergaria-a-Velha          | BRANCA                        |       10204 |
|           5 | Aveiro    | Albergaria-a-Velha          | RIBEIRA DE FRAGUAS            |       10206 |
|           6 | Aveiro    | Albergaria-a-Velha          | SAO JOAO DE LOURE E FROSSOS   |       10210 |
|           7 | Aveiro    | Anadia                      | AVELAS DE CAMINHO             |       10304 |
|           8 | Aveiro    | Anadia                      | AVELAS DE CIMA                |       10305 |
|           9 | Aveiro    | Anadia                      | MOITA                         |       10307 |
|          10 | Aveiro    | Anadia                      | SANGALHOS                     |       10309 |
|          11 | Aveiro    | Anadia                      | SAO LOURENCO DO BAIRRO        |       10310 |
|          12 | Aveiro    | Anadia                      | UF A DA GANDARA BAIRRO E ANCAS |       10316 |
|          13 | Aveiro    | Anadia                      | UF ARCOS E MOGOFORES          |       10317 |
|          14 | Aveiro    | Anadia                      | UF TAMENGOS AGUIM OIS BAIRRO  |       10318 |
|          15 | Aveiro    | Anadia                      | VILA NOVA DE MONSARROS        |       10312 |
|          16 | Aveiro    | Anadia                      | VILARINHO DO BAIRRO           |       10313 |
|          17 | Aveiro    | Arouca                      | ALVARENGA                     |       10402 |
|          18 | Aveiro    | Arouca                      | CHAVE                         |       10407 |
|          19 | Aveiro    | Arouca                      | ESCARIZ                       |       10409 |
|          20 | Aveiro    | Arouca                      | FERMEDO                       |       10411 |
|          21 | Aveiro    | Arouca                      | MANSORES                      |       10413 |
|          22 | Aveiro    | Arouca                      | MOLDES                        |       10414 |
|          23 | Aveiro    | Arouca                      | ROSSAS                        |       10415 |
|          24 | Aveiro    | Arouca                      | SANTA EULALIA                 |       10416 |
|          25 | Aveiro    | Arouca                      | SAO MIGUEL DO MATO            |       10417 |
|          26 | Aveiro    | Arouca                      | TROPECO                       |       10418 |
|          27 | Aveiro    | Arouca                      | UF AROUCA E BURGO             |       10421 |
|          28 | Aveiro    | Arouca                      | UF CABREIROS ALBERGARIA SERRA |       10422 |
|          29 | Aveiro    | Arouca                      | UF CANELAS E ESPIUNCA         |       10423 |
|          30 | Aveiro    | Arouca                      | UF COVELO DE PAIVO E JANARDE  |       10424 |
|          31 | Aveiro    | Arouca                      | URRO                          |       10419 |
|          32 | Aveiro    | Arouca                      | VARZEA                        |       10420 |
|          33 | Aveiro    | Aveiro                      | ARADAS                        |       10501 |
|          34 | Aveiro    | Aveiro                      | CACIA                         |       10502 |
|          35 | Aveiro    | Aveiro                      | EIXO E EIROL                  |       10515 |
|          36 | Aveiro    | Aveiro                      | ESGUEIRA                      |       10505 |
|          37 | Aveiro    | Aveiro                      | OLIVEIRINHA                   |       10508 |
|          38 | Aveiro    | Aveiro                      | REQUEIXO N S FATIMA E NARIZ   |       10516 |
|          39 | Aveiro    | Aveiro                      | SANTA JOANA                   |       10513 |
|          40 | Aveiro    | Aveiro                      | SAO BERNARDO                  |       10510 |
|          41 | Aveiro    | Aveiro                      | SAO JACINTO                   |       10511 |
|          42 | Aveiro    | Aveiro                      | UF GLORIA E VERA CRUZ         |       10517 |
|          43 | Aveiro    | Castelo de Paiva            | FORNOS                        |       10602 |
|          44 | Aveiro    | Castelo de Paiva            | REAL                          |       10606 |
|          45 | Aveiro    | Castelo de Paiva            | SANTA MARIA DE SARDOURA       |       10607 |
|          46 | Aveiro    | Castelo de Paiva            | SAO MARTINHO DE SARDOURA      |       10608 |
|          47 | Aveiro    | Castelo de Paiva            | UF RAIVA PEDORIDO E PARAISO   |       10610 |
|          48 | Aveiro    | Castelo de Paiva            | UF SOBRADO E BAIRROS          |       10611 |
|          49 | Aveiro    | Espinho                     | ESPINHO                       |       10702 |
|          50 | Aveiro    | Espinho                     | PARAMOS                       |       10704 |
```

Dimension Tables Population Job

Job


Start
Populate Time Dimension
Populate Location Dimension

Task 6