

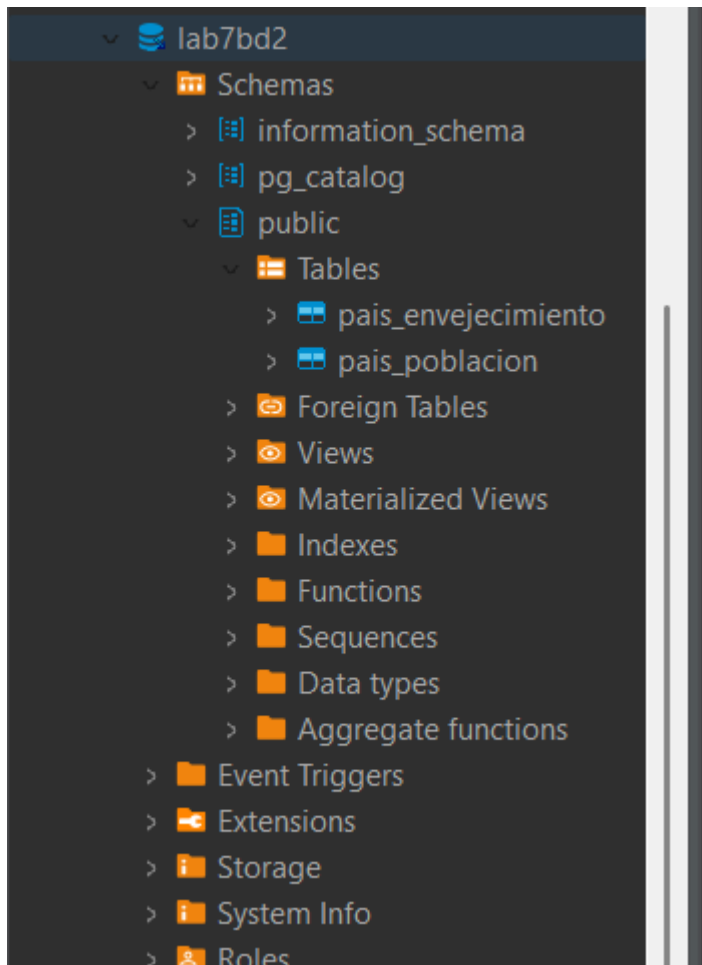
Lab 7 BDD2

Observación: podrá utilizar cualquier herramienta de procesos de ETL a su elección.

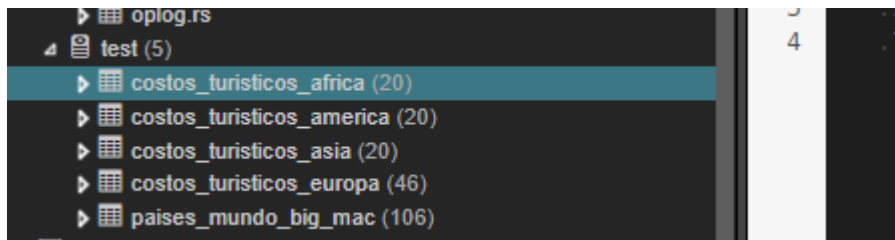
2.1 Ingeste los datos que se encuentran en la base de datos relacional. Revise si es necesario realizar algún tipo de limpieza sobre los datos.

id_pais	nombre_pais	capital	continente	region	poblacion	tasa_de_envejecimiento
1	Latvia	Riga	Europa	Norte de Europa	1,906,800	12.49
2	Romania	Bucarest	Europa	Europa del Este	19,000,000	24.01
3	Kosovo	Pristina	Europa	Balcenes	1,800,000	19.64
4	Sudáfrica	Pretoria	África	África Austral	59,308,688	16.97
5	Slovakia	Bratislava	Europa	Europa Central	5,450,000	8.12
6	Hungary				[NULL]	8.12
7	Angola				[NULL]	6.16
8	Japón				[NULL]	22.32
9	Uganda				[NULL]	17.02
10	Corea del Sur				[NULL]	19.16
11	Irán				[NULL]	5.41
12	Estados Unidos				[NULL]	24.4
13	Cuba				[NULL]	21.65
14	Denmark				[NULL]	9.25
15	México				[NULL]	8.64
16	Argentina				[NULL]	8.67
17	Etiopia				[NULL]	11.08
18	Chile				[NULL]	15.5
19	Liechtenstein				[NULL]	13.64
20	Belarus				[NULL]	10.82
21	Albania				[NULL]	17.24
22	North Macedonia				[NULL]	7.79
23	Portugal				[NULL]	10.84
24	Finland				[NULL]	12.33
25	Canadá				[NULL]	14.12
26	Colombia				[NULL]	20.7
27	Austria				[NULL]	8.99
28	Estonia				[NULL]	15.28
29	Uzbekistán				[NULL]	16.85
30	Camerún				[NULL]	5.93
31	Panamá				[NULL]	17.15
32	Malta				[NULL]	8.41
33	Ruanda				[NULL]	6.3
34	Venezuela				[NULL]	23.98
35	Montenegro				[NULL]	24.31
36	United Kingdom				[NULL]	21.17
37	Arabia Saudita				[NULL]	11.09
38	Zambia				[NULL]	6.95
39	Vatican City				[NULL]	18.68

pais	id	poblacion	anio	id	continente	costo_bajo_hospedaje	costo_promedio_comida	costo_bajo_transporte
Ruanda	1	12952218		67e2130bc29aa80afae47e7f	África	7	17	17
Namibia	2	2540905		67e2130bc29aa80afae47e80	África	8	25	11
Mozambique	3	31255435		67e2130bc29aa80afae47e7a	África	9	41	9
Camerún	4	26545863		67e2130bc29aa80afae47e7b	África	9	18	26
Angola	5	32866272		67e2130bc29aa80afae47e7c	África	9	30	15
Nepal	6	29136808		67e213cec29aa80afae47e96	Asia	10	21	37
Tanzania	7	59734218		67e2130bc29aa80afae47e75	África	11	18	29
Irán	8	83992949		67e213cec29aa80afae47e93	Asia	11	53	37
Zambia	9	18383955		67e2130bc29aa80afae47e79	África	13	14	24
Japón	10	126476461		67e213cec29aa80afae47e85	Asia	14	17	42
Guatemala	11	17915568		67e21472c29aa80afae47e9a	América	14	72	35
Etiopia	12	114963588		67e2130bc29aa80afae47e73	África	16	12	21
Uganda	13	45741007		67e2130bc29aa80afae47e78	África	16	36	26
Costa Rica	14	5094118		67e21472c29aa80afae47e9d	América	16	31	14
Paraguay	15	7132530		67e21472c29aa80afae47ea6	América	16	68	19
Perú	16	32971854		67e21472c29aa80afae47ea4	América	16	63	48
Pakistán	17	220892340		67e213cec29aa80afae47e8d	Asia	16	25	26
Túnez	18	11818619		67e2130bc29aa80afae47e77	África	18	37	11
Venezuela	19	28435940		67e21472c29aa80afae47ea2	América	18	49	27
Ghana	20	31072940		67e2130bc29aa80afae47e74	África	19	31	23
Filipinas	21	109581078		67e213cec29aa80afae47e8a	Asia	19	44	32
Marruecos	22	36910560		67e2130bc29aa80afae47e71	África	20	22	21
Colombia	23	50882891		67e21472c29aa80afae47ea1	América	20	60	34
Costa de Marfil	24	26378274		67e2130bc29aa80afae47e7d	África	20	30	7
Luxembourg	25	634814		67e211ddc29aa80afae47e56	Europa	20	46	10
Sudáfrica	26	59308690		67e2130bc29aa80afae47e6d	África	21	20	25
Arabia Saudita	27	34813871		67e213cec29aa80afae47e91	Asia	22	43	19
San Marino	28	33938		67e211ddc29aa80afae47e62	Europa	22	41	14
Moldova	29	2640438		67e211ddc29aa80afae47e58	Europa	22	33	4
Estados Unidos	30	331002651		67e21472c29aa80afae47e97	América	23	43	19
Uzbekistán	31	33469203		67e213cec29aa80afae47e90	Asia	23	54	18
Senegal	32	16743927		67e2130bc29aa80afae47e76	África	23	36	30
Tailandia	33	69799978		67e213cec29aa80afae47e88	Asia	24	22	30
Corea del Sur	34	51269185		67e213cec29aa80afae47e86	Asia	24	28	14
Denmark	35	5831404		67e211ddc29aa80afae47e48	Europa	24	34	4



2.2 Ingeste los datos que se encuentran en la base de datos no relacional. Revise si es necesario realizar algún tipo de limpieza sobre los datos.



2.3 Integre ambas fuentes de datos por medio de la herramienta de procesos de ETL.

- Proceso ETL

Se usa Prefect como software de ETL

```
18:35:40.682 | INFO      | prefect - Starting temporary server on
http://127.0.0.1:8571
See https://docs.prefect.io/3.0/manage/self-host#self-host-a-prefect-server for
more information on running a dedicated Prefect server.
18:35:51.419 | INFO      | Flow run 'sturdy-capybara' - Beginning flow run 'sturdy-
capybara' for flow 'ETL_Integracion_Total_2_Tablas'
Columnas en pais_poblacion: ['id', 'pais', 'poblacion', 'anio', '_id',
'continente', 'costo_bajo_hospedaje', 'costo_promedio_comida',
'costo_bajo_transporte', 'costo_promedio_entretenimiento']
Columnas en pais_envejecimiento: ['id_pais', 'nombre_pais', 'capital',
'continente', 'region', 'poblacion', 'tasa_de_envejecimiento']
```

Cantidad de registros después del merge: 106

18:35:52.563 | INFO | Task run 'ingest_postgres_tables-86f' - Finished in state Completed()

18:35:54.206 | INFO | Task run 'ingest_mongodb-b61' - Finished in state Completed()

18:35:54.271 | INFO | Task run 'clean_data-9e3' - Finished in state Completed()

18:35:54.351 | INFO | Task run 'clean_mongo_dfs-2b6' - Finished in state Completed()

18:35:54.475 | INFO | Task run 'integrate_all_data-257' - Finished in state Completed()

C:\Users\rodri\Documents\BDD2\Lab7\data_ETL.py:136: FutureWarning: DataFrame.applymap has been deprecated. Use DataFrame.map instead.

```
df_converted = df.applymap(convert_value)
```

```
df_converted = df.applymap(convert_value)
```

Ejemplo de registros a insertar (después de la conversión):

	id	...	costos_diarios_estimados_en_dólares	costos_turisticos_europa
0	NaN	...	{"hospedaje": {"precio_bajo_usd": 58.32, "prec...	
1	NaN	...	{"hospedaje": {"precio_bajo_usd": 24.84, "prec...	
2	NaN	...		NaN
3	NaN	...		NaN
4	NaN	...		NaN

[5 rows x 43 columns]

Número total de registros a insertar: 106

18:35:55.622 | INFO | Task run 'load_data-5a6' - Finished in state Completed()

18:35:55.635 | INFO | Task run 'load_data-176' - Finished in state Completed()

18:35:58.622 | INFO | Flow run 'sturdy-capybara' - Finished in state Completed()

```
df_converted = df.applymap(convert_value)
```

Ejemplo de registros a insertar (después de la conversión):

	id	...	costos_diarios_estimados_en_dólares	costos_turisticos_europa
0	NaN	...	{"hospedaje": {"precio_bajo_usd": 58.32, "prec...	
1	NaN	...	{"hospedaje": {"precio_bajo_usd": 24.84, "prec...	
2	NaN	...		NaN
3	NaN	...		NaN
4	NaN	...		NaN

[5 rows x 43 columns]

Número total de registros a insertar: 106

18:35:55.622 | INFO | Task run 'load_data-5a6' - Finished in state Completed()

18:35:55.635 | INFO | Task run 'load_data-176' - Finished in state Completed()

Ejemplo de registros a insertar (después de la conversión):

	id	...	costos_diarios_estimados_en_dólares	costos_turisticos_europa
0	NaN	...	{"hospedaje": {"precio_bajo_usd": 58.32, "prec...	
1	NaN	...	{"hospedaje": {"precio_bajo_usd": 24.84, "prec...	

2 NaN	...	NaN
3 NaN	...	NaN
4 NaN	...	NaN

[5 rows x 43 columns]

Número total de registros a insertar: 106

18:35:55.622 | INFO | Task run 'load_data-5a6' - Finished in state Completed()

18:35:55.635 | INFO | Task run 'load_data-176' - Finished in state Completed()

18:35:58.622 | INFO | Flow run 'sturdy-capybara' - Finished in state Completed()

18:36:00.331 | INFO | prefect - Stopping temporary server on http://127.0.0.1:8571

[5 rows x 43 columns]

Número total de registros a insertar: 106

18:35:55.622 | INFO | Task run 'load_data-5a6' - Finished in state Completed()

18:35:55.635 | INFO | Task run 'load_data-176' - Finished in state Completed()

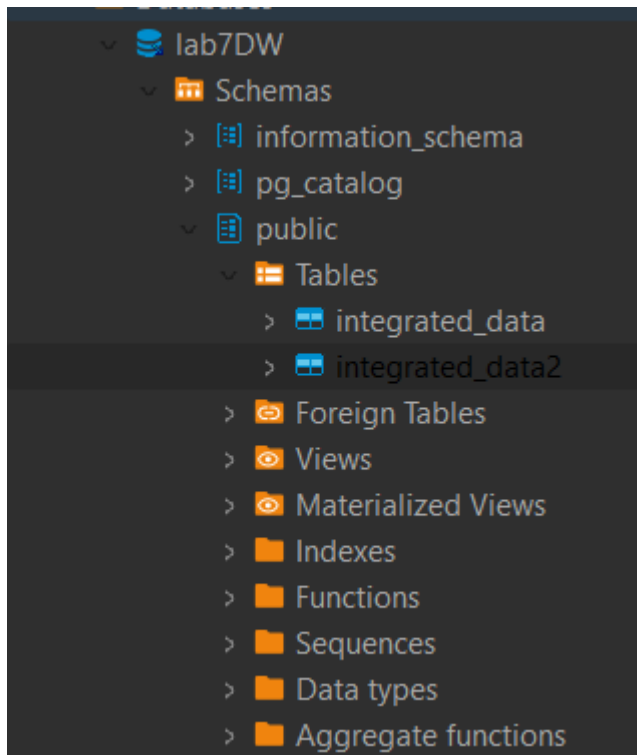
18:35:58.622 | INFO | Flow run 'sturdy-capybara' - Finished in state Completed()

18:36:00.331 | INFO | prefect - Stopping temporary server on http://127.0.0.1:8571

(venv) PS C:\Users\rodri\Documents\BDD2\Lab7> █

2.4 Configure la herramienta para que el proceso de ETL se ejecute cada cierto tiempo (la frecuencia de ejecución queda a su criterio)

2.4 Los datos integrados se deberán cargar en la base de datos que hace las veces de data warehouse, sin que se necesite su intervención



	id	pais	poblacion_x	anio	id	continente_x	costo_bajo_hospedaje	costo_promedio_comida	costo_bajo_transporte	costo_promedio_entretenimiento
1	[NULL]	albania	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
2	[NULL]	andorra	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
3	[NULL]	angola	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
4	[NULL]	arabia saudita	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
5	[NULL]	argelia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
6	[NULL]	argentina	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
7	[NULL]	austria	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
8	[NULL]	bangladés	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
9	[NULL]	belarus	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
10	[NULL]	belgium	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
11	[NULL]	bolivia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
12	[NULL]	bosnia and herzegovina	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
13	[NULL]	brasil	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
14	[NULL]	bulgaria	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
15	[NULL]	camerún	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
16	[NULL]	canadá	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
17	[NULL]	chile	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
18	[NULL]	china	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
19	[NULL]	colombia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
20	[NULL]	corea del sur	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
21	[NULL]	costa de marfil	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
22	[NULL]	costa rica	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
23	[NULL]	croatia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
24	[NULL]	cuba	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
25	[NULL]	cyprus	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
26	[NULL]	czech republic	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
27	[NULL]	denmark	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
28	[NULL]	ecuador	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
29	[NULL]	egipto	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
30	[NULL]	el salvador	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
31	[NULL]	estados unidos	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
32	[NULL]	estonia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
33	[NULL]	etiopia	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
34	[NULL]	filipinas	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
35	[NULL]	finland	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
36	[NULL]	france	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
37	[NULL]	germany	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
38	[NULL]	ghana	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]

Table Properties Enter a SQL expression to filter results (use Ctrl+Space)

	ASC	ASC	ASC	123	ASC	ASC
	id_costos_turisticos_asia	continente	región	capital_costos_turisticos_asia	población	costos_diaros_estimados_en_dólares
1	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
2	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
3	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
4	67e88454b03910499e82c8c2	Asia	Asia Occidental	Riad	34,813,871	["hospedaje": {"precio_bajo_usd": 22, "precio_promedio_usd": 32, "precio_alto_usd": 42}]
5	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
6	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
7	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
8	67e88454b03910499e82c8c8	Asia	Asia Meridional	Daca	164,689,383	["hospedaje": {"precio_bajo_usd": 39, "precio_promedio_usd": 56, "precio_alto_usd": 73}]
9	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
10	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
11	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
12	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
13	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
14	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
15	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
16	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
17	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
18	67e88454b03910499e82c8b9	Asia	Asia Oriental	Pekín	1,402,112,000	["hospedaje": {"precio_bajo_usd": 28, "precio_promedio_usd": 41, "precio_alto_usd": 54}]
19	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
20	67e88454b03910499e82c8c0	Asia	Asia Oriental	Seúl	51,269,185	["hospedaje": {"precio_bajo_usd": 24, "precio_promedio_usd": 35, "precio_alto_usd": 46}]
21	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
22	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
23	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
24	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
25	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
26	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
27	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
28	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
29	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
30	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
31	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
32	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
33	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
34	67e88454b03910499e82c8bd	Asia	Sudeste Asiático	Manila	109,581,078	["hospedaje": {"precio_bajo_usd": 19, "precio_promedio_usd": 28, "precio_alto_usd": 37}]
35	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
36	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
37	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]
38	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]	[NULL]

Refresh Save Cancel Export data 200 106 106 row(s) fetched - 0.007s (0.002s fetch), on 2025-04-01 at 18:36:29