

CIEE Análisis

Renato Vargas

Introducción

En este documento se detalla los pasos utilizados para el análisis.

Hallazgos

El primer paso consiste en cargar las librerías necesarias para el análisis.

```
library(RSQLite)
library(DBI)
```

Nos conectamos a la base de datos a través del paquete DBI y obtenemos objetos de datos usando SQL. Idealmente, la base de datos cuenta con la documentación adecuada, pero en una primera instancia se explora la base de datos para saber con qué tablas y variables cuenta. Utilizamos la función `dbConnect()` habilitada por el paquete DBI para abrir una conexión con la base de datos SQLite del Sistema de Cuentas Nacionales. Una vez concluido el análisis, se utiliza `dbDisconnect()` para cerrar la conexión como puede verse al final de esta memoria de análisis. Además de abrir la conexión, se utiliza la función `dbListTables()` para saber con qué tablas cuenta la base de datos.

```
con <- dbConnect(RSQLite::SQLite(), "C:/Users/renato/GitHub/scn_scae_gt/datos/scn.db")
dbListTables(con)
```

```
[1] "areas_columnas"      "areas_filas"         "columnas"
[4] "cuadros"             "energia"              "filas"
[7] "naeg"                "npg"                  "ntg2"
[10] "oferta_utilizacion"
```

La mayoría de tablas en esta base datos explican los valores de los campos de las tablas principales `oferta_utilizacion` y `balances_energeticos`. Por esa razón, es importante saber con qué campos cuentan estas dos tablas.

```
dbListFields(con, "oferta_utilizacion")
```

```
[1] "anio"          "id_precios" "id_cuadro"  "id_fila"     "id_columna"
[6] "valor"         "id_unidad"
```

Puesto que esta base de datos contiene los cuadros de oferta y utilización de los años 2013 a 2020 y esta, a su vez, cuenta con cientos de miles de registros, se condujo un proceso de normalización de la base datos, con el objetivo que la tabla principal cuente con la menor cantidad de columnas y dentro de estas, con la menor cantidad de valores para reducir el tamaño y evitar las redundancias de datos. Se puede inspeccionar las primeras filas de la tabla `oferta_utilizacion`. Hay varias formas de hacer esto, tanto en R, como en SQL. Para mantener la consistencia con las consultas de SQL, aquí se utilizará este lenguaje y se hace una consulta de los primeros cinco registros de la tabla `oferta_utilizacion` y todas las columnas.

```
dbGetQuery(con, "SELECT * FROM oferta_utilizacion LIMIT 5")
```

	anio	id_precios	id_cuadro	id_fila	id_columna	valor	id_unidad
1	2013	1	1	f001	oc001	5.346939	1
2	2013	1	1	f002	oc001	4734.946097	1
3	2013	1	1	f003	oc001	101.132353	1
4	2013	1	1	f004	oc001	0.000000	1
5	2013	1	1	f005	oc001	0.000000	1

Es evidente que los valores de las columnas no son descriptivos por si solos y es necesario recurrir a vínculos con las demás tablas de la base de datos para entender a qué transacciones, productos y actividades económicas se refiere la columna de `valor`. Esto se hace a través de uniones en el lenguaje SQL, identificadas por la palabra clave `JOIN`.

Por ejemplo, si estuviéramos interesados en saber cuáles fue la oferta monetaria de productos energéticos en el año 2014, debemos hacer las uniones de nuestra tabla principal `oferta_utilizacion` con las diferentes tablas necesarias para nuestra consulta, como lo son `cuadros`, `energia` y `npg`.

```
dbGetQuery(con, "
SELECT
  energia,
```

```

        filas.id_npg4 as id_npg4,
        npg4,
        sum(valor)
FROM
    oferta_utilizacion
JOIN
    filas
ON
    oferta_utilizacion.id_fila
    =
    filas.id_fila
JOIN
    energia
ON
    filas.id_npg4
    =
    energia.id_npg4
WHERE
    anio = 2014
AND
    id_cuadro = 2
AND
    id_energia > 1
GROUP BY
    energia.id_energia,
    filas.id_npg4,
    npg4
ORDER BY
    energia.id_energia,
    filas.id_npg4,
    npg4
")

```

	energia	id_npg4
1	Energía primaria	P1502
2	Energía primaria	P1503
3	Energía primaria	P1701
4	Energía primaria	P2001
5	Energía primaria	P4901
6	Energía secundaria	P4101
7	Energía secundaria	P4102
8	Energía secundaria	P4103

9	Energía secundaria	P4104
10	Energía secundaria	P4105
11	Energía secundaria	P4106
12	Energía secundaria	P4201
13	Energía secundaria	P5001

		npg4
1		Leña
2		Productos forestales no madereros
3		Petróleo crudo y gas natural
4		Carbón mineral, lignito y turba
5		Desechos de la industria alimenticia y de tabaco
6		Gasolinas
7		Diesel oil
8		Fuel oils (bunker)
9		Aceites, grasas lubricantes y otros aceites preparados
10		Gases de petróleo y otros hidrocarburos gaseosos
11	Otros productos de la refinación de petróleo n.c.p. y productos de combustible nuclear	
12		Alcoholes
13		Energía eléctrica , gas, vapor y aire acondicionado
	sum(valor)	
1	4088.2386	
2	275.3266	
3	2390.3604	
4	1297.1017	
5	619.9479	
6	14017.5870	
7	13970.1655	
8	2061.2772	
9	1196.6710	
10	5426.9628	
11	1321.6208	
12	1527.4320	
13	15412.7520	

Finalmente, es necesario cerrar la conexión que se abrió al inicio de este apartado, usando la función `dbDisconnect()`.

```
dbDisconnect(con)
```