



Caso 104 Datos del SABI

1 Casos ETL Modelado de Datos.

Jose Ignacio González Gómez

Departamento de Economía, Contabilidad y Finanzas - Universidad de La Laguna

www.jggomez.eu

V.21

Ejercicio Basado: *Desarrollado por jggomez*

Enunciado

Partimos de un fichero descargado del SABI correspondiente a un conjunto de establecimientos hoteleros con domicilio fiscal en canarias y cuya estrategia de búsqueda fue la siguiente

	Resultado etapa	último	búsqueda
1. Región/País: Las Palmas de Gran Canaria, Santa Cruz de Tenerife	64.793		64.793
2. CNAE 2009(Sólo códigos primarios): 5510 - Hoteles y alojamientos similares	14.247		834
Búsqueda booleana : 1 Y 2			
	TOTAL		834

La información de la que contamos para esta selección de empresas corresponde a las variables: número de empleados, ingresos de explotación, importe neto de la cifra de negocios y gastos de personal para el periodo 2015-2020.

El fichero bruto es

	A	B	C	D	E	F	G	H
		Nombre	Código NIF	Localidad	Código primario o CNAE 2009	Literal código CNAE 2009 primario	Número empleados Últ. año disp.	Número empleados 2020
1		NORDOTEL	A07411499	SAN	5510	Hoteles y	1.358	n.d.
2	1.	MASPALOMA	B35396639	LAS PALMAS	5510	Hoteles y	592	n.d.
3	2.	HOTEADEJE	B35619386	ADEJE	5510	Hoteles y	687	n.d.
4	3.	TOURIN	A35062116	SAN	5510	Hoteles y	645	n.d.
5	4.	EUROPEO SA	A08205915	ARONA	5510	Hoteles y	480	n.d.

Se pide

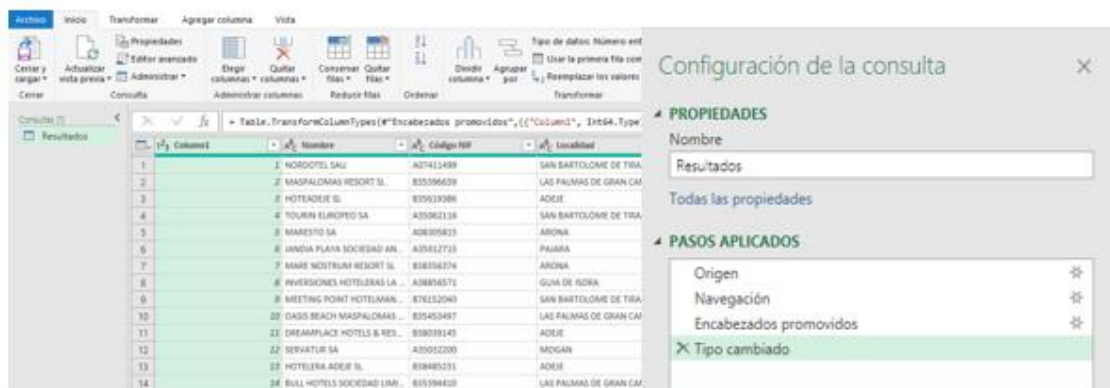
Con la información no estructurada del fichero original deseamos transformarla en un formato estructurado usando Power Query para poder ser tratada y explotada adecuadamente para nuestros informes, tablas y gráficos dinámicos.

Obteniendo una tabla similar a la siguiente para ser cargada finalmente a nuestro modelo de datos.

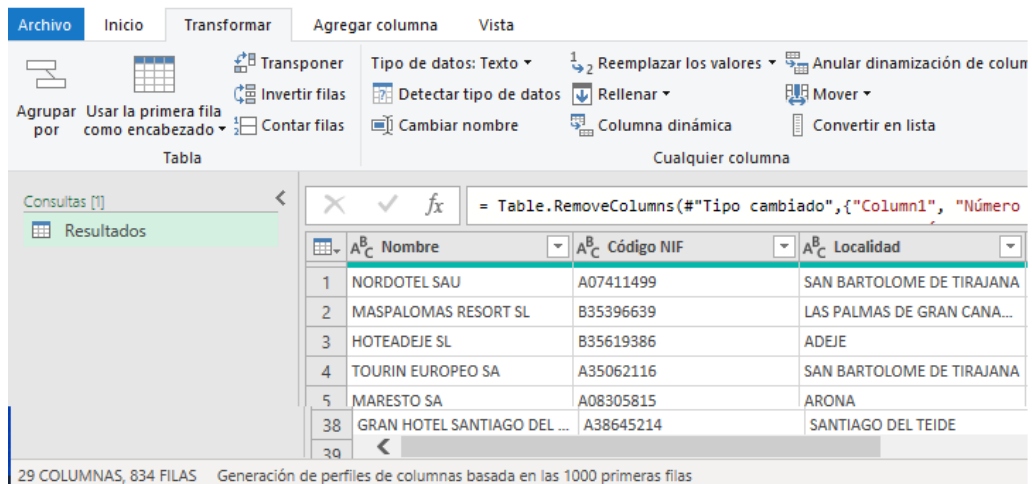
	A	B	C	D	E	F
1	Nombre	NIF	Localidad	Variable	Ejercicio	Valor
2	ACORON S.L.	B38326997	BREÑA BAJA	Número empleados	2019	103
3	ACORON S.L.	B38326997	BREÑA BAJA	Número empleados	2018	100
4	ACORON S.L.	B38326997	BREÑA BAJA	Número empleados	2017	98
5	ACORON S.L.	B38326997	BREÑA BAJA	Número empleados	2016	90
6	ACORON S.L.	B38326997	BREÑA BAJA	Número empleados	2015	70
7	ACORON S.L.	B38326997	BREÑA BAJA	Ingresos de explotaciónEUR	2019	6840282
8	ACORON S.L.	B38326997	BREÑA BAJA	Ingresos de explotaciónEUR	2018	6044493
9	ACORON S.L.	B38326997	BREÑA BAJA	Ingresos de explotaciónEUR	2017	5681588
10	ACORON S.L.	B38326997	BREÑA BAJA	Ingresos de explotaciónEUR	2016	5645460
11	ACORON S.L.	B38326997	BREÑA BAJA	Ingresos de explotaciónEUR	2015	4638573
12	ACORON S.L.	B38326997	BREÑA BAJA	Importe neto Cifra de Ventas	2019	6823209
13	ACORON S.L.	B38326997	BREÑA BAJA	Importe neto Cifra de Ventas	2018	6038066
14	ACORON S.L.	B38326997	BREÑA BAJA	Importe neto Cifra de Ventas	2017	5681588
15	ACORON S.L.	B38326997	BREÑA BAJA	Importe neto Cifra de Ventas	2016	5642435
16	ACORON S.L.	B38326997	BREÑA BAJA	Importe neto Cifra de Ventas	2015	4633946
17	ACORON S.L.	B38326997	BREÑA BAJA	Gastos de personalEUR	2019	2422296
18	ACORON S.L.	B38326997	BREÑA BAJA	Gastos de personalEUR	2018	2166839
19	ACORON S.L.	B38326997	BREÑA BAJA	Gastos de personalEUR	2017	2108014
20	ACORON S.L.	B38326997	BREÑA BAJA	Gastos de personalEUR	2016	1917966
21	ACORON S.L.	B38326997	BREÑA BAJA	Gastos de personalEUR	2015	1674706
22	ACTUAL URBAN A	B76312537	LAS PALMAS C	Número empleados	2019	6
23	ACTUAL URBAN A	B76312537	LAS PALMAS C	Número empleados	2018	2
24	ACTUAL URBAN A	B76312537	LAS PALMAS C	Ingresos de explotaciónEUR	2019	338524

Como orientación, los pasos seguidos en este proceso ETL son

1. Abrimos una nueva hoja para vincular los datos originales. CONTAMOS CON 834 FILAS
2. FASE Extracción o consulta. Iniciamos el editor Power Query (Datos – Obtener Datos)) y seleccionamos nuevo origen de datos, lo que nos permite acceder a la fuente de datos.



3. FASE Transformación. Procedimiento de limpieza y adaptación.
 - 3.1. Comenzamos quitando las columnas que no nos interesan disponer, así como cambiando el nombre a las mismas.
 - 3.2. Quitamos las columnas correspondientes a último año disponible.
 - o Numero de empleados, último año disponible
 - o Ingresos de explotación, último año disponible
 - o Importe neto de cifra de ventas, último año disponible
 - 3.3. Quitadas las principales columnas y filas, consultamos lo que nos queda.



- 3.4. Verificamos el tipo de datos de las columnas es correcto y en caso contrario lo cambiamos. En este paso podrían aparecer errores al cambiar de tipo, por ejemplo n.d por error, es decir significa que nd es no disponible y por tanto habrá que corregir (<https://docs.microsoft.com/es-es/power-query/dealing-with-errors>).
- 3.5. Cambiar Error por null. Para ello vamos a la pestaña transformar – cualquier columna – Reemplazar valores Reemplazar errores por null (en minúscula) que significa vacío. Así seleccionamos todas las columnas que deseamos cambiar el tipo y procedemos a reemplazar error por null.
- 3.6. Seleccionamos las columnas que necesita Anular Dinamización. Podemos observar como algunas columnas contiene en su descripción el nombre de la variable y el año y dentro de la misma el contenido. Por ejemplo Gastos de Personal 2019 y los valores para cada hotel y realmente lo que nos interesa es separar este contenido en tres columnas, el nombre de la variable (Gatos de Personal), Año (2019) e Importe para cada hotel.

Para ello tenemos disponible la opción TRANSFORMAR – ANULAR DINAMIZACION DE COLUMNAS.

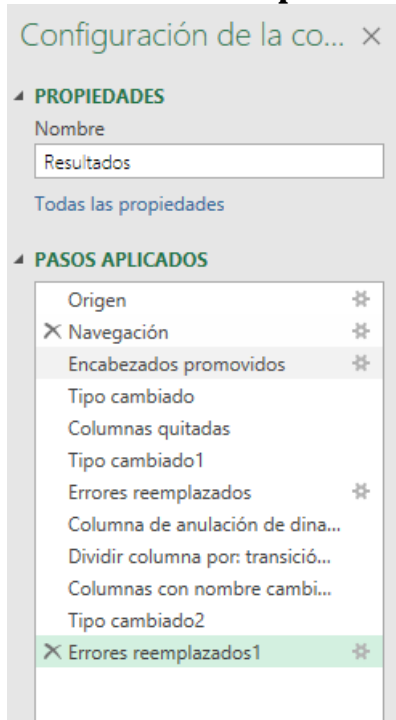
	A ^B C Nombre	A ^B C Código NIF	A ^B C Localidad	A ^B C Atributo	A ^B C Valor
1	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados 2020	n.d.
2	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados 2019	1.358
3	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados 2018	1.778
4	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados 2017	1.383
5	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados	1.281

- 3.7. Dividir columnas. En caso necesario necesitamos aplicar dividir columnas para separar por ejemplo el nombre de la variable y el año como es nuestro caso a la creada como atributo.

Para ello vamos a la opción Transformar – DIVIDIR COLUMNA: POR TRANSICION DE CARACTERES DE NO DIGITO A DIGITO pero también podríamos hacerlo de otras formas como en la opción AGREGAR COLUMNA EXTRAER como siempre al final esta el año, seleccionamos extraer los 4 últimos caracteres.

- 3.8. Renombrar las columnas (Variable-Año-Valor). Ahora nos queda renombrar el nombre de las columnas así como asignar formato a su contenido, es decir cambiamos el tipo de registro según necesidades.

Resumen de los pasos seguidos en el proceso ETL



Y el resultado final es

	A ^B _C Nombre	A ^B _C NIF	A ^B _C Localidad	A ^B _C Variable	A ^B _C Ejercicio	\$ Valor
1	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados	2019	1.358,00
2	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados	2018	1.778,00
3	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados	2017	1.383,00
4	NORDOTEL SAU	A07411499	SAN BARTOLOME DE TIRAJANA	Número empleados	2016	1.281,00

Conceptos y técnicas requeridas

Los conceptos y técnicas a aplicar son las siguientes:

Realizar una consulta con Power Query a un origen de datos Excel. Configurar la consulta empleando: Creando conexión, encabezados promovidos, quitar columnas, columna de anulación de dinamización, dividir columnas por transición de caracteres o separando por últimos caracteres, columnas por nombre cambiado, tipo cambiado de columnas (texto por numero), quitar errores, etc. Presentamos a continuación una captura de la configuración de la consulta realizada.

Cambiar el Error provocado al cambiar tipo de una columna que contiene n.d (no disponible) a numérico se genera un ERROR que queremos cambiar por null (vacío)