



# Caso 102 Ventas por año 7009 registros

1 Casos ETL Modelado de Datos.

Jose Ignacio González Gómez

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

[www.jggomez.eu](http://www.jggomez.eu)

V.21

Ejercicio Basado: Aprender a Usar Power Query en Minutos, Ejercicio Power Query Nº 4.,

Profesor: Jorge Luis Herrera [ENLACE](#)

## Enunciado

Como podemos observar en la siguiente ilustración contamos en nuestro fichero con una pestaña no normalizado de nuestra empresa “centros de abastos universal” con las ventas realizadas por vendedores durante el periodo correspondiente.

Se puede observar que la información se encuentra no estructurada y no puede ser utilizada directamente para la elaboración de análisis y estudios basados en tablas y gráficos dinámicos, Power BI y herramientas similares.

A	B	C	D	E	F	G	H	I	J	K	L
1											
2	centro de abastos universa	Fecha:	15/12/2018	nro.	12	folio	ventas por año	centro de abastos univers	regisgro	olg	ventas por año
3		Año:									
4	Datos					15-58-896					
5	Vendedores	Compañía Cliente	Ventas			Cantidad	Producto	Fecha Pedido	Código Estado	País	Categoría
6	FRANCISCO Chávez	Compañía A	1400			325	Cerveza Northwind Traders	22/11/2018	NV	EEUU	Bebidas
7	FRANCISCO Chávez	Compañía A	105			184	Ciruelas pasas Northwind Traders	17/02/2018	NV	EEUU	Frutos secos
8	PILAR, PINILLA GALLEGRO	Compañía D	300			100	Peras secas Northwind Traders	4/10/2018	NY	EEUU	Frutos secos
9	PILAR, PINILLA GALLEGRO	Compañía D	530			389	Manzanas secas Northwind Traders	26/07/2018	NY	EEUU	Frutos secos
10	PILAR, PINILLA GALLEGRO	Compañía D	35			27	Ciruelas pasas Northwind Traders	8/01/2018	NY	EEUU	Frutos secos
11	Azucena del Rio	Compañía L	270			291	Té Chai Northwind Traders	22/11/2018	NV	EEUU	Bebidas
12	Azucena Del Rio	Compañía L	920			151	Café Northwind Traders	17/02/2018	NV	EEUU	Bebidas
13	JUAN CARLOS RIVAS-	Compañía H	276			292	Galletas de chocolate surtidas Northwind Trade	4/10/2018	OR	EEUU	Productos horneados
14	FRANCISCO, Chávez	Compañía D	184			247	Galletas de chocolate surtidas Northwind Trade	26/07/2018	NY	EEUU	Productos horneados
15	PILAR PINILLA GALLEGRO	Compañía C	127,5			97	Chocolate Northwind Traders	8/01/2018	CO	EEUU	Golosinas
16	Azucena del Rio	Compañía C	1930			368	Sopa de almejas Northwind Traders	7/10/2018	CA	EEUU	Sopas
17											
18											
19											
20	LUIS, BONIFAZ -	Compañía F	680			351	Salsa curry Northwind Traders	22/11/2018	WI	EEUU	Salsas
21	FRANCISCO C-HAVES	Compañía B	13800			250	Café Northwind Traders	17/02/2018	TN	EEUU	Bebidas
22	Pilar Pinilla Gallego -	Compañía H	1275			226	Chocolate Northwind Traders	4/10/2018	OR	EEUU	Golosinas
23	Azucena Del Rio	Compañía J	598			311	Té verde Northwind Traders	26/07/2018	IL	EEUU	Bebidas

Nos vamos a encontrar con columnas y filas vacías que tendremos que corregir, así como registros con espacios en blanco y errores de caracteres que no proceden como son guiones, punto y coma y similares. Por tanto, esto nos lleva a la necesidad de una limpieza de los registros.

También contamos con una segunda pestaña Estados que es un rango de datos que tendremos que combinarla con la anterior.

## Se pide

Es por tanto necesario llevar a cabo un proceso ETL con el editor de consultas Power Query para convertir esta fuente en una base normalizada, para el caso de las dos pestañas como la siguiente:

	Vendedores	Compañía Cliente	Ventas	Cantidad	Producto
1	Francisco Chávez	Compañía A		1400	325 Cerveza Northwind Traders
2	Francisco Chávez	Compañía A		105	184 Ciruelas pasas Northwind Traders
3	Azucena Del Rio	Compañía L		270	291 Té Chai Northwind Traders
4	Azucena Del Rio	Compañía L		920	151 Café Northwind Traders
5	Pilar Pinilla Gallego	Compañía D		300	100 Peras secas Northwind Traders

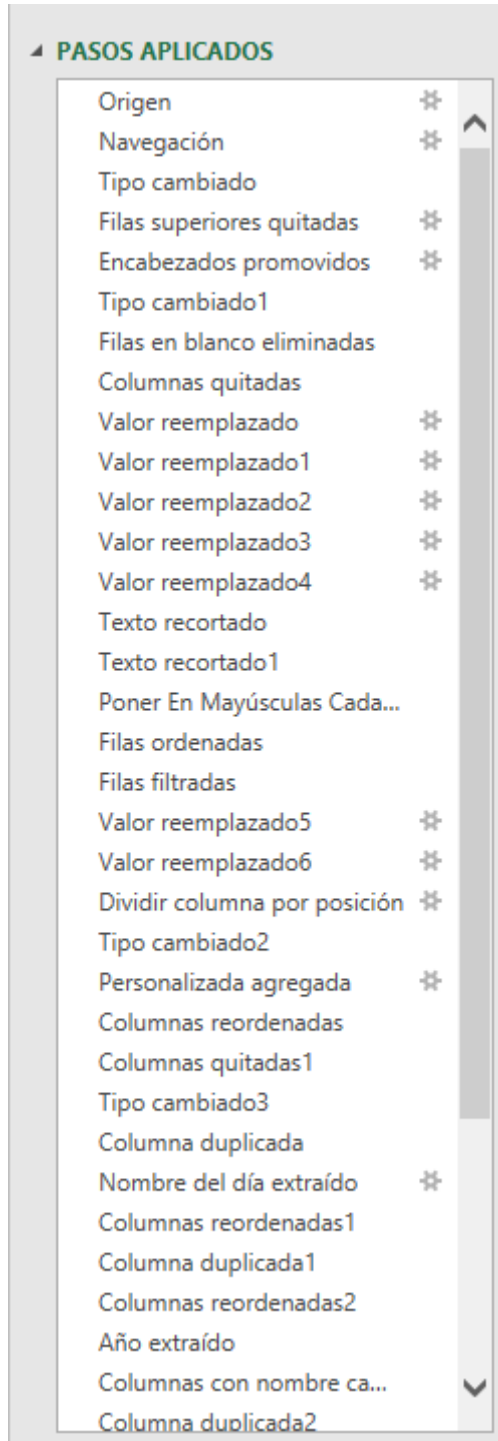
Para ello crearemos un nuevo fichero Excel y a través de PQ crearemos una conexión a la fuente original y comenzaremos el proceso de normalización.

Como orientación, los pasos seguidos en este proceso ETL presentamos los siguientes

1. FASE CREAR CONEXIÓN – CONSULTA. Así el primer paso es crear una conexión a la fuente a través de Nueva Consulta a las dos tablas
2. FASE TRANSFORMACION Y LIMPIEZA
  - 2.1. Con la pestaña Estados vamos simplemente a usar la primera fila como encabezado, a través de la opción “Encabezados Promovidos”.
  - 2.2. Pero es la pestaña ventas la que exige un mayor proceso de depuración y normalización. Vamos a quitar las primeras filas que no necesitamos y situados sobre columna 1 fila 4 y seleccionamos la opción Quitar Filas Superiores”
  - 2.3. “Encabezados Promovidos” , opción Usar la Primera fila como encabezado
  - 2.4. Podemos observar que tenemos muchas filas en blanco y para eliminarla disponemos de la opción “Quitar filas-Quitar filas en blanco”.
  - 2.5. Alcanzado este punto revisamos aquellas columnas vacías que no interesa, seleccionado todas ellas y haciendo uso de la opción “Quitar Columnas”
  - 2.6. Procedemos a continuación a depurar los registros contenidos en la columna Vendedores que contiene registros que demanda una limpiar relacionada con los espacios en blanco, caracteres extraños, etc.
    - 2.6.1. Comenzamos por confirmar que la columna es de tipo texto.
    - 2.6.2. Seleccionamos la columna y con el botón derecho seleccionamos la opción “Reemplazar valores” por ejemplo ; - , y .
    - 2.6.3. Algunos datos de vendedores se repiten y eso puede ser debido a la existencia de espacio, por tanto procederemos a eliminarlos haciendo uso de la opción también reemplazar valores. Por ejemplo reemplazar dos espacios por un espacio.
    - 2.6.4. Igualmente, con el botón derecho seleccionamos la columna vendedores y elegimos Transformar – Recortar y esto elimina todos los espacios en blanco al comienzo del texto y después del texto.
    - 2.6.5. También vamos a poner la primera letra en mayúscula con la opción señalada de transformar “Poner en Mayúscula cada Palabra”
    - 2.6.6. Vamos a ver también los vendedores que por error ortográfico aparece con acento y sin acento, por ejemplo el caso de “Francisco Chaves” sustituirlo por “Francisco Chávez”
  - 2.7. Tratamos a continuación la columna “Compañía Cliente” confirmamos que es un tipo de datos texto y donde se puede observar la necesidad de los siguientes ajustes:
    - 2.7.1. Vemos que existe espacios entre algunas palabras vamos a quitarla haciendo uso de “Reemplazar Valores” quitando el espacio por nada.
    - 2.7.2. Este ajuste anterior genera que el nombre de la compañía aparezca como CompañíaA, CompañíaB, etc.
    - 2.7.3. Necesitamos dividir este campo en dos para después unirlos. Para ello usamos la opción Dividir Columna por Numero de caracteres.
    - 2.7.4. Una vez dividida la columna procederemos a crear una columna personalizada en la opción Agregar columna de la siguiente forma [Compañía Cliente.1]&" "&[Compañía Cliente.2] y ya podemos eliminar las columnas anteriores.
    - 2.7.5. Sobre el campo fecha de pedido vamos a dividirlo por años, trimestre, semanas, etc. Así procedemos en duplicar columna y con el click derecho le decimos transformar en año, trimestre, día de la semana, etc..
  - 2.8. Ahora nos interesa relacionar las dos consultas disponibles Estados y Ventas a través de la opción Combinar-Combinar Consultas

De esta forma ya tenemos preparada todo el modelo normalizado.

## Resumen de los pasos seguidos en el proceso ETL



### Conceptos y técnicas requeridas

Los conceptos y técnicas para aplicar con Power Query relacionadas con el proceso ETL son:

Creacion de una consulta, conexión a un fichero fuente no normalizado. Relacionar dos consultas. Combinar-Combinar Consultas

Uso de las opciones “Quitar Filas Superiores” y “Encabezados Promovidos” opcion Usar la Primera fila como encabezado. Quitar filas en blanco

Quitar columnas. Cambiar tipo. Dividimos columnas por número de caracteres y creamos una columna personalizada. Creacion de una columna personalizada en la opcion Agregar columna por ejemplo: [Compañía Cliente.1]&" "&[Compañía Cliente.2]

Limpieza. Eliminar los registros que tienen espacios al comienzo y o al final, usando la opción Transformar-Recortar.

Uso de “Reemplazar Valores”, por ejemplo, cambiar null por 0. Quitar espacios en blanco y caracteres no deseados como guiones, puntos, etc.

Transformación sobre campo fecha, en año, trimestre, semanas, etc.