



Caso 100b Ventas Limpieza Basica

01 ETL y Modelo de Datos. Limpieza y transformación de datos básica. Cambiar tipo, quitar columna, auditoria de datos, remplazar valores, poner en mayúsculas cada palabra, recortar quitar espacios en blanco sobrantes...

Jose Ignacio González Gómez

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

www.jggomez.eu

V.2.4

Ejercicio Basado: **DATDATA** [Power Bi PowerQueryarrearlar datos caso 1 \(youtube.com\)](https://www.youtube.com/watch?v=...)

Archivos fuentes: 01 Datos 100b

Resumen: Contiene opciones básicas en el Proceso ETL con Power Query.

Técnicas y palabras clave: Cambiar tipo columna, quitar columna, auditoria de datos, remplazar los valores, transformar poner en Mayúsculas cada palabra, transformar recortar quitar espacios en blanco sobrantes...

Contenido

Presentación del caso	2
Se pide (preguntas base del modelo)	2
Crear conexión y a la fuente de datos y llevar el proceso de depuración	2
Realizar los siguientes analisis básicos fundamentados en Tablas y Graficos Dinámicos	3
Pasos sugeridos en el proceso ETL	3
Proceso ETL con Power Query Modelado de datos con Power Pivot	3
Resumen configuración de la consulta del editor (Power Query)	5
Referencias	5

Presentación del caso

Este proyecto nos permitirá adentrarnos en el proceso ETL a través de un fichero de ventas obtenido del ERP que contiene el siguiente conjunto de campos (columnas) y registros (filas) y que debemos revisar, limpiando y transformándolo para convertirlo en apto para el análisis.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Fecha	IDFactura	CLIENTE	Ciudad	País	Segmento	Medio	Canal	Oficina	Teléfono	Vendedor	MontoFactura
2	04/02/2024	4377	BAKER, EDWARD	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Buenos A	68335158	Sebastián Sánchez	\$ 346,06
3	04/02/2024	4377	BAKER, EDWARD	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Chile	68335158	Sebastián Sánchez	\$ 346,06
4	04/02/2024	4377	BAKER, EDWARD	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Madrid	68335158	Sebastián Sánchez	\$ 346,06
5	04/02/2024	4377	BAKER, EDWARD	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina México	68335158	Sebastián Sánchez	\$ 346,06
6	04/02/2024	4379	HOWARD, GABRIEL	Lima	Perú	Persona	Offline	Desconocido	Oficina Buenos A	54320374	Adriana Bossio	\$ 1.082,95
7	04/02/2024	4379	HOWARD, GABRIEL	Lima	Perú	Persona	Offline	Desconocido	Oficina Madrid	54320374	Adriana Bossio	\$ 1.082,95

Ilustración 1

Se pide (preguntas base del modelo)

Crear conexión y a la fuente de datos y llevar el proceso de depuración

El objetivo es disponer de una fuente depurada como la que mostramos a continuación que se caracteriza respecto a la de la Ilustración 1 en que se han realizado los siguientes ajustes:

De un primer análisis se han detectado determinadas anomalías que exigen un proceso de limpieza y transformación, en concreto debemos hacer ciertos arreglos en las características de los campos que generalmente no se encuentran bien en nuestras bases de datos o que no las necesitamos para nuestro objetivo, así se proponen los siguientes ajustes.

1. La columna Monto la convertimos a formato moneda.
2. Confirmar a través de la auditoría de datos que no tenemos errores y analizar los vacíos.

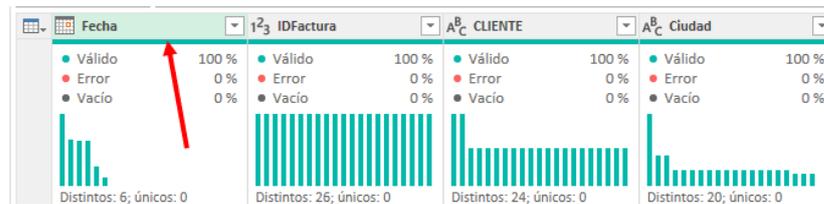


Ilustración 2

La línea verde de la ilustración anterior nos ayuda a identificar que el 100% de los datos están válidos es decir están correctamente escritos si tuviera errores saldría un pedacito de la línea de color rojo y si tenemos celdas vacías saldrá un pedacito de la línea de color negro en este caso la tabla está completamente correcta.

3. **Quitar columna Teléfono.** Disponemos de una columna teléfono y no la necesitamos para nuestros informes, por tanto, debemos quitarla con el fin de agilizar el proceso y disminuir el consumo de recursos aumentando el rendimiento y velocidad del modelo. Debemos tener en cuenta que solo la estamos quitando de la consulta, en los archivos originales seguirá existiendo aquí lo único que le estamos diciendo que no la muestre.
4. **Ajuste Cliente.** En la columna de Cliente deseamos hacer los siguientes ajustes:
 - a. En algunos registros hay una coma que separa el nombre y el apellido, pero en otros no así decidimos **quitar la coma**.
 - b. Además queremos que el nombre y apellido de los clientes no sean todos en mayúsculas sino solo la **primera letra** en mayúsculas (Ilustración 1).
5. En el **campo ciudad**, aunque no es visible, tenemos algunos casos que presentan espacios en blanco, como en el registro 6 correspondiente a Lima y nos puede dar errores al reconocer dos ciudades iguales como diferentes, la que tiene espacio y las que no lo tienen, igualmente pasaría con Caracas y si hiciéramos un análisis con Tablas o Gráficos Dinámicos tendríamos dos valores distintos, tanto para Lima como Caracas, lo cual no es cierto. Esto sucede mucho en las bases de datos entonces la idea es que vamos a arreglar automatizando el proceso.

Por tanto, el resultado del proceso ETL (Conexión-Transformación y Carga), debe ser similar al siguiente de la Ilustración 3.

Fecha	IDFactura	CLIENTE	Ciudad	Pais	Segmento	Medio	Canal	Oficina	Vendedor	MontoFactura
04/02/2024	4377	Baker Edward	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Buenos Aires	Sebastián Sánchez	346,06
04/02/2024	4377	Baker Edward	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Chile	Sebastián Sánchez	346,06
04/02/2024	4377	Baker Edward	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina Madrid	Sebastián Sánchez	346,06
04/02/2024	4377	Baker Edward	San Salvador	El Salvador	Persona	Offline	Desconocido	Oficina México	Sebastián Sánchez	346,06

Ilustración 3

Realizar los siguientes analisis básicos fundamentados en Tablas y Graficos Dinámicos

Deberá elaborarse unas tablas y graficos similares a los propuestos y con segmentadores vinculados a los objetos visuales tal y como se desprende de la Ilustración 4

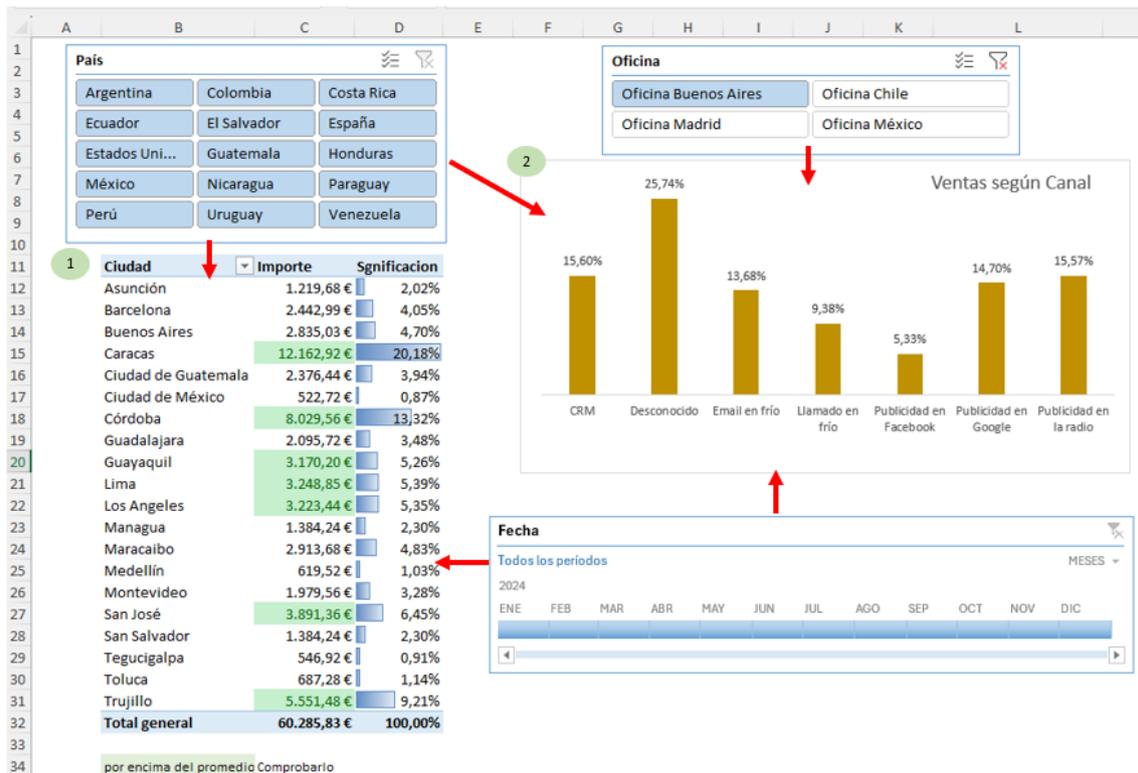


Ilustración 4

Pasos sugeridos en el proceso ETL

- Creamos nuestra hoja de solución al caso planteado

Proceso ETL con Power Query Modelado de datos con Power Pivot

De un primer analisis se han detectado determinadas anomalías que exigen un proceso de limpieza y transformación, en concreto debemos hacer ciertos arreglos en las características de los campos que generalmente no se encuentran bien en nuestras bases de datos o que no las necesitamos para nuestro objetivo, asi se proponen los siguientes ajustes.

1. **Cambiar tipo columna.** Una vez creada la conexión verificamos el tipo de dato de cada campo o columna, en especial establecemos para la columna Monto a formato moneda.
2. **Pestaña Vista: Calidad de Columna, Distribución de Columnas y Perfil de Columnas.** Confirmar a través de la auditoria de datos que no tenemos errores y analizar los vacíos.

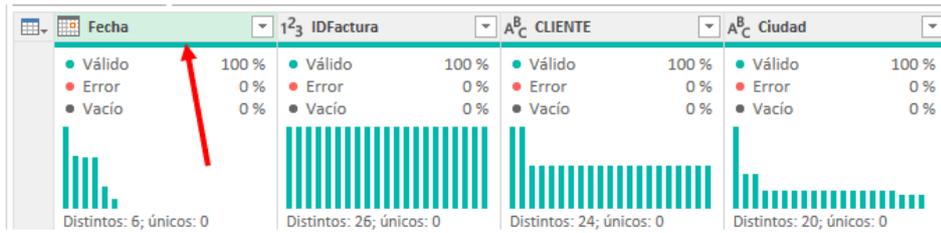


Ilustración 5

La línea verde de la ilustración anterior nos ayuda a identificar qué el 100% de los datos están válidos es decir están correctamente escritos si tuviera errores saldría un pedacito de la línea de color rojo y si tenemos celdas vacías saldrá un pedacito de la línea de color negro en este caso la tabla está completamente correcta.

3. **Quitar columna Teléfono.** Disponemos de una columna teléfono y no la necesitamos para nuestros informes, por tanto, debemos quitarla con el fin de agilizar el proceso y disminuir el consumo de recursos aumentando el rendimiento y velocidad del modelo. Debemos tener en cuenta que solo la estamos quitando de la consulta, en los archivos originales seguirá existiendo aquí lo único que le estamos diciendo que no la muestre
4. **Ajuste Cliente.** En la columna de Cliente deseamos hacer los siguientes ajustes:
 - a. **Reemplazar los valores.** En algunos registros hay una coma que separa el nombre y el apellido, pero en otros no así decidimos quitar la coma.
 - b. **Transformar poner en Mayúsculas cada palabra.** Además queremos que el nombre y apellido de los clientes no sean todos en mayúsculas sino solo la primera en mayúsculas (Ilustración 6).
5. En el **campo ciudad**, aunque no es visible, tenemos algunos casos que presentan espacios en blanco, como es el caso del registro 6 correspondiente a Lima y nos puede dar errores al reconocer dos ciudades Lima como diferentes, las que tiene espacios y las que no lo tienen, igualmente pasaría con Caracas y si hiciéramos un análisis con Tablas o Gráficos Dinámicos tendríamos dos valores distintos, tanto para Lima como para Caracas, lo cual no es cierto. Esto sucede mucho en las bases de datos entonces la idea es que vamos a arreglar automatizando el proceso y para ello debe ir a **Transformar Recortar** que nos permitirá quitar los espacios en blanco sobrantes.

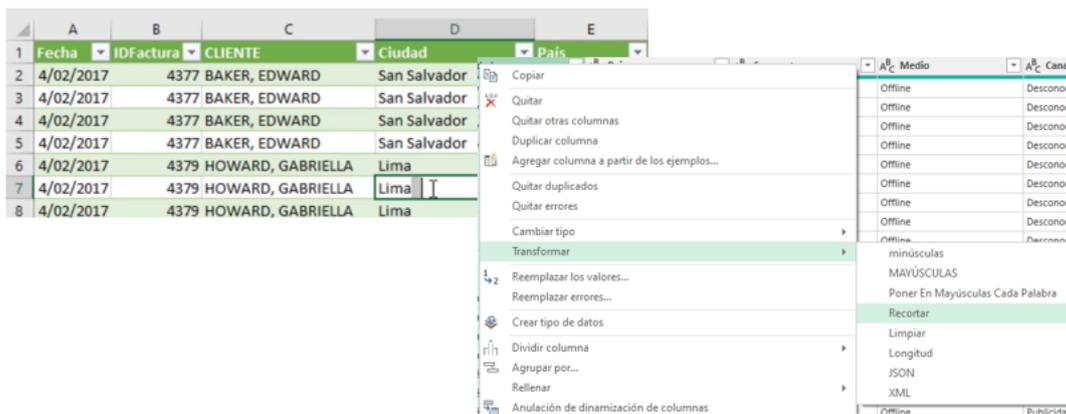


Ilustración 6

Resumen configuración de la consulta del editor (Power Query)

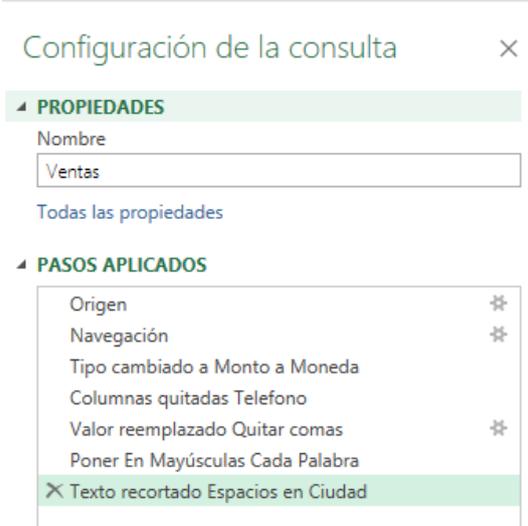


Ilustración 7

Referencias

[Power Bi PowerQuery arreglar datos caso 1 \(youtube.com\)](https://www.youtube.com/watch?v=...)