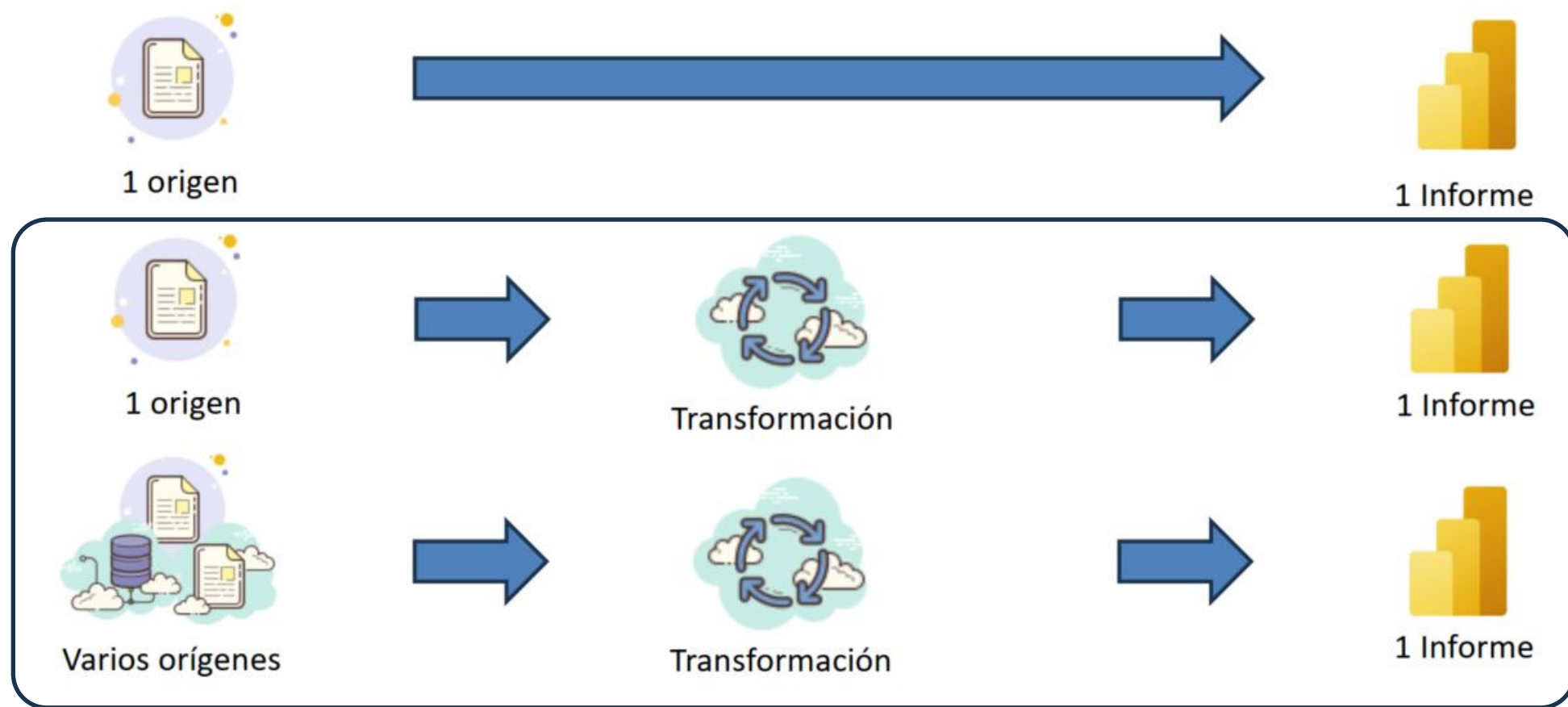


# TRANSFORMACIÓN DE DATOS EN POWER BI

Transformación digital con Power BI Básico y bases de datos relacionales

Rubén Hermoso Díez

# ¿Dónde estamos?



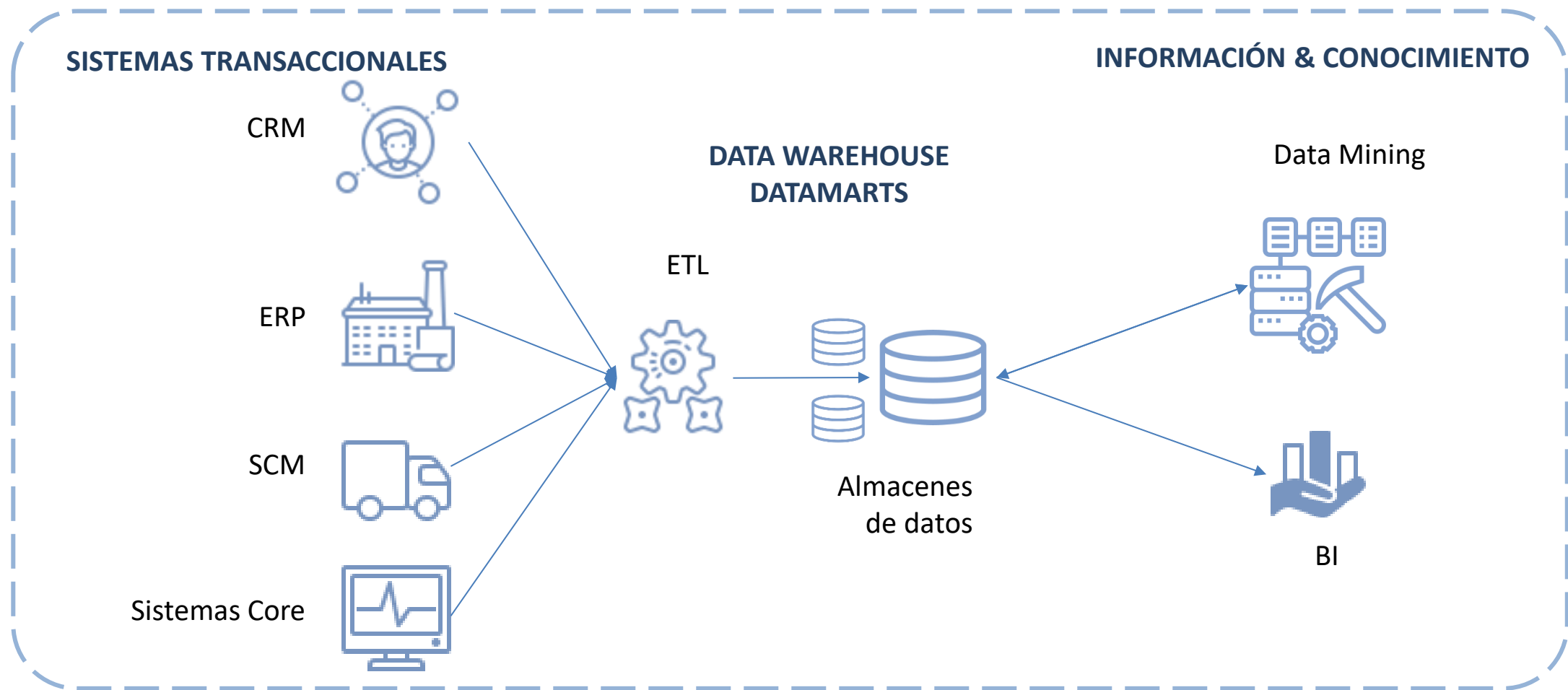
# Introducción a la transformación de datos

Un problema habitual al que se enfrentan las organizaciones es cómo recopilar datos de varios orígenes, en varios formatos. A continuación, tendrá que moverlos a uno o varios almacenes de datos. Es posible que el destino no sea el mismo tipo de almacén de datos que el origen. A menudo el formato es diferente, o bien es necesario dar forma a los datos o limpiarlos antes de cargarlos en el destino final.

Con los años se han desarrollado varias herramientas, servicios y procesos para ayudarle a afrontar estos desafíos. Independientemente del proceso que se utilice, hay una necesidad común de coordinar el trabajo y aplicar cierto nivel de transformación de datos en la canalización de datos. Las secciones siguientes resaltan los métodos más habituales utilizados para realizar estas tareas.



# ¿Qué papel juega la transformación de datos?



# ¿Por qué necesito transformar mis datos?

## Ejemplo 1:

Empleado	Salario Bruto
María	1500000
Juan	1300
Pedro	1000
Nuria	1500

## Ejemplo 2:

Empleado	Salario Bruto	% IRPF	Salario Neto
María	1500	15,00%	1275
Juan	1300		1000
Pedro	1000	8	
Nuria	1500	15,00%	1275

## Ejemplo 3:

Empleado	Salario Bruto	% IRPF	Salario Neto	Fecha contrato
María	1500	15,00%	1275	01/11/2023 0:00
Juan	1300		1000	15/10/2023 0:00:00:00
Pedro	1000	8		27/04/2020
Nuria	1500	15,00%	1275	2017-01-03

## Ejemplo 4:

Empleado	Salario Bruto	% IRPF	Salario Neto	Fecha contrato	Ciudad
María	1500	15,00%	1275	01/11/2023 0:00	Barcelona
Juan	1300		1000	15/10/2023 0:00:00:00	Murzia
Pedro	1000	8		27/04/2020	Zaragoza
Nuria	1500	15,00%	1275	2017-01-03	Zaragoza

## Beneficios

Organización mejorada



Calidad del dato



Compatibilidad



## Desafíos

Proceso Costoso



Requieren de tiempo



Múltiples fases



## Pasos en un proceso de transformación de los datos

- ❑ **Interpretación de la información** para comprender los datos que hay en la organización. Una herramienta que ayuda a interpretar la información usada dentro de una organización es el Modelado Dimensional
- ❑ **Examinar la calidad de los datos:** esta evaluación va a permitir comprobar la información para asegurarse que es veraz y no está duplicada. Esta evaluación también va a permitirnos conocer columnas con datos faltantes
- ❑ **Aplicar las transformaciones pertinentes:** estandarizar la información, rellenar valores, eliminar duplicados, ajustar el formato y el tipo de dato
- ❑ **Revisión de la nueva calidad de los datos** con el objetivo de verificar si los ajustes han cumplido su función y modificar los procesos si fuera necesario.

# INTRODUCCIÓN A POWER QUERY

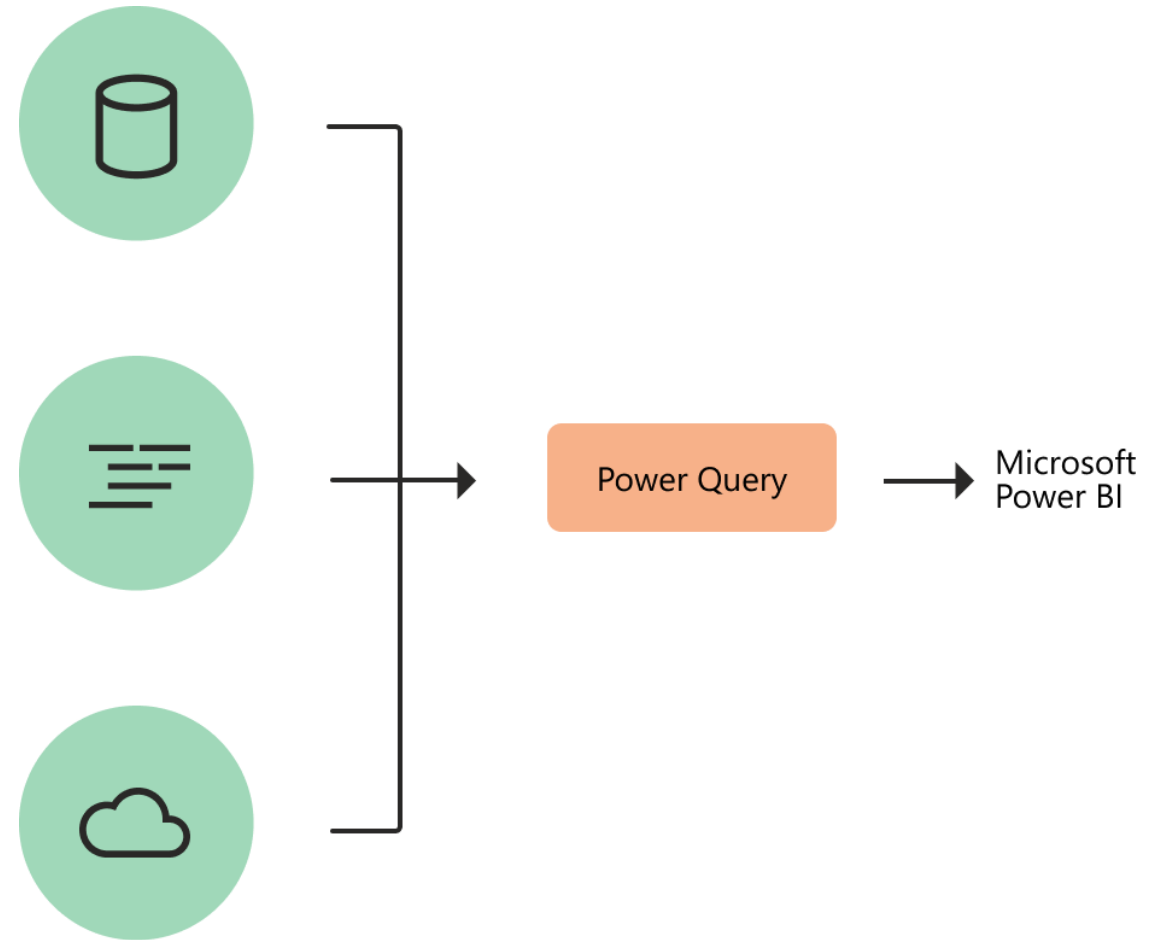


# ¿Qué es Power Query?

Power Query es un motor de preparación de datos y transformación de datos que incluye una interfaz gráfica para obtener datos de diferentes orígenes, y un Editor de Power Query para aplicar transformaciones a la información recabada

Power Query permite realizar operaciones de extracción, transformación y carga de datos (ETL)

Durante la implantación de una solución de inteligencia de negocio, se invierte hasta el 80 % del tiempo en la preparación de datos, el resto del tiempo es dedicado a la elaboración de los informes.

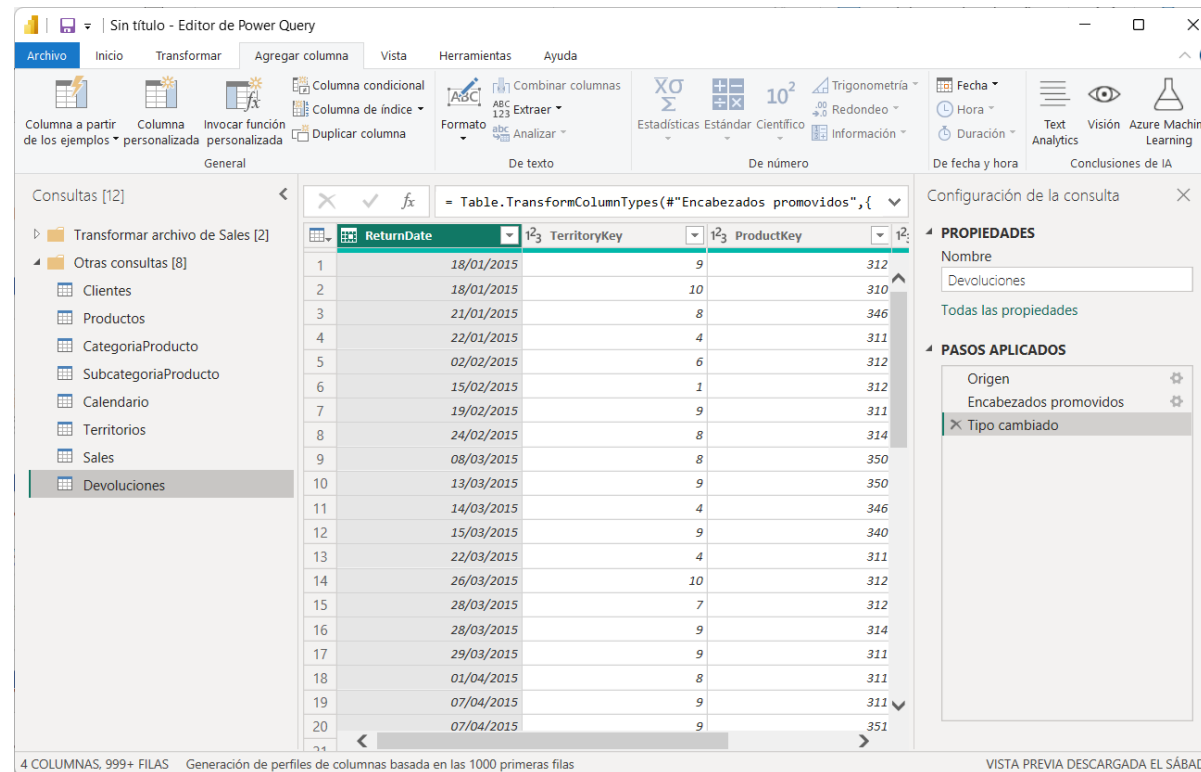
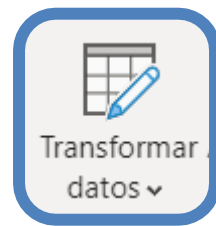


# ¿Qué es Power Query?



# Power Query

Una de las ventajas de Power Query es que puedes crear una serie de pasos para transformar tus datos y guardarlos como una consulta. Esta consulta se puede actualizar fácilmente cuando los datos de origen cambian, lo que significa que no tendrás que repetir todo el proceso de transformación manualmente.

The screenshot shows the Microsoft Power Query Editor window. The title bar reads "Sin título - Editor de Power Query". The ribbon includes "Archivo", "Inicio", "Transformar", "Agregar columna", "Vista", "Herramientas", and "Ayuda". The "Transformar" tab is active, showing options like "Columna a partir de los ejemplos", "Columna personalizada", "Invocar función personalizada", and "Duplicar columna". The main area displays a table with columns: "ReturnDate", "TerritoryKey", "ProductKey", and an unnamed column. The table contains 20 rows of data. On the left, the "Consultas" pane shows a list of queries, with "Devoluciones" selected. On the right, the "Configuración de la consulta" pane shows the "PASOS APLICADOS" section, which includes "Origen", "Encabezados promovidos", and "Tipo cambiado". The status bar at the bottom indicates "4 COLUMNAS, 999+ FILAS" and "Generación de perfiles de columnas basada en las 1000 primeras filas".

# Interfaz de Power Query

weiopjeqiojqowije

The screenshot displays the Power Query interface with the following components:

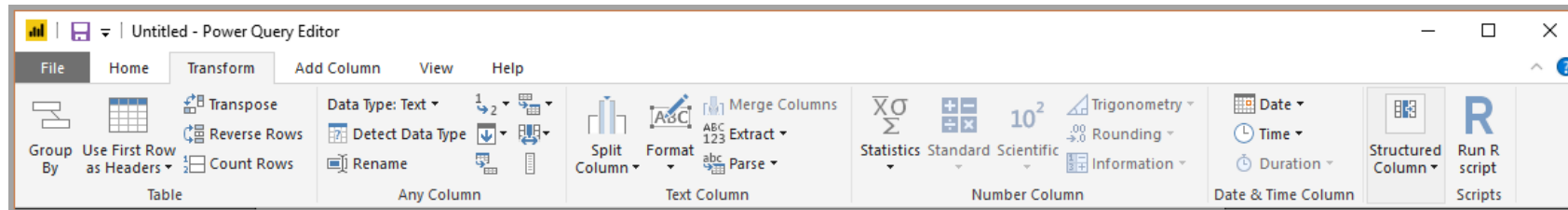
- Barra de Herramientas (Ribbon):** Includes tabs for Archivo, Inicio, Transformar, Agregar columna, Vista, Herramientas, and Ayuda. The 'Vista' tab is active, showing options for 'Diseño' (Design) and 'Vista previa de datos' (Data preview). The 'Diseño' group includes 'Configuración de la consulta' (Query settings), 'Barra de fórmulas' (Formula bar), 'Monoespaciada' (Monospaced), 'Mostrar espacio en blanco' (Show blank space), and 'Calidad de columnas' (Column quality). The 'Vista previa de datos' group includes 'Distribución de columnas' (Column distribution), 'Perfil de columna' (Column profile), 'Permitir siempre' (Allow always), 'Ir a columna' (Go to column), 'Parámetros' (Parameters), 'Editor avanzado' (Advanced editor), 'Uso avanzado' (Advanced use), 'Dependencias de la consulta' (Query dependencies), and 'Dependencias' (Dependencies).
- Consultas [1]:** A list of queries, with 'VentasMediaMarkt' selected.
- Configuración de la consulta (Query Settings):** A pane on the right showing the 'Nombre' (Name) as 'VentasMediaMarkt' and the 'Tipo' (Type) as 'Tipo cambiado' (Changed type).
- Data Preview:** A table showing the first 11 rows of data. The columns are: Fecha, Vendedor, ID Cliente, Apellidos, Nombre, Antigüedad, Edad, and Sexo. Each column has a header, a data type, and a distribution chart. The 'Fecha' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Vendedor' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'ID Cliente' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Apellidos' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Nombre' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Antigüedad' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Edad' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010. The 'Sexo' column has a distribution chart showing a range from 17/09/2015 to 21/01/2010.

Annotations on the left side of the image point to specific elements:

- Cabecera de columna:** Points to the column headers in the data preview.
- Calidad de columna:** Points to the 'Calidad de columnas' option in the ribbon.
- Distribución de columna:** Points to the distribution charts above the data preview.
- Datos:** Points to the data rows in the preview.

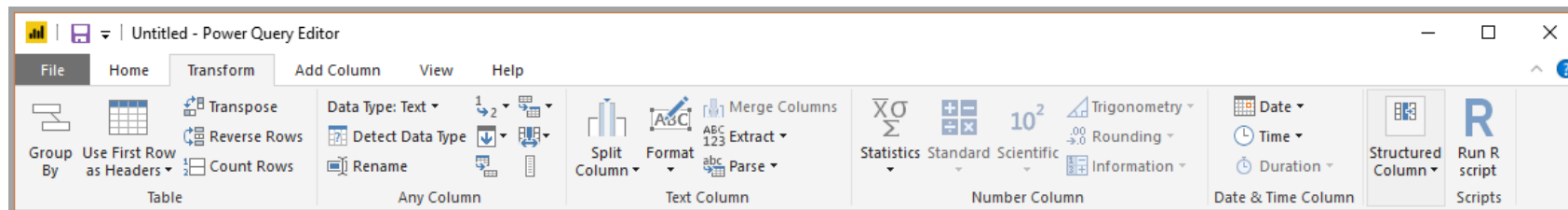
# Pestaña de Transformar Datos

Tareas comunes de transformación de datos, como agregar o quitar columnas, cambiar tipos de datos, dividir columnas y otras operaciones controladas por datos.



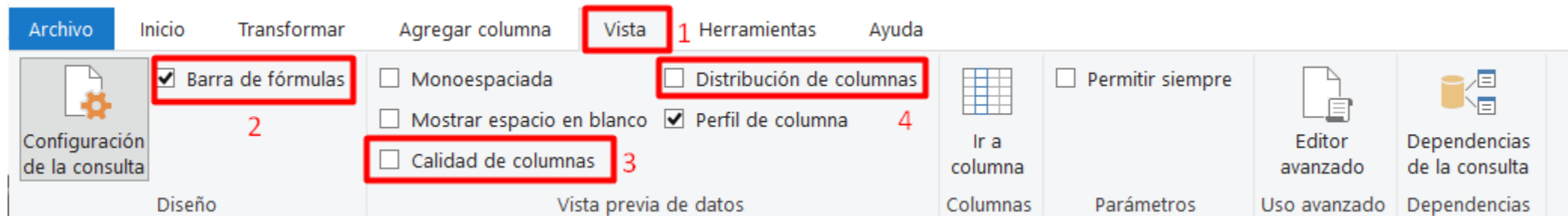
# Pestaña de Agregar Columna

Tareas adicionales asociadas a agregar una columna, aplicar formato a los datos de la columna y Agregar columnas personalizadas.



# Pestaña de vista

La pestaña de vista va a permitir activar ciertas funcionalidades interesantes de Power Query, que ayudan a los analistas de datos a ser mas conscientes de la información que están manejando y a explorar la calidad de los datos con los que se está trabajando.



**OPERACIONES COMUNES**

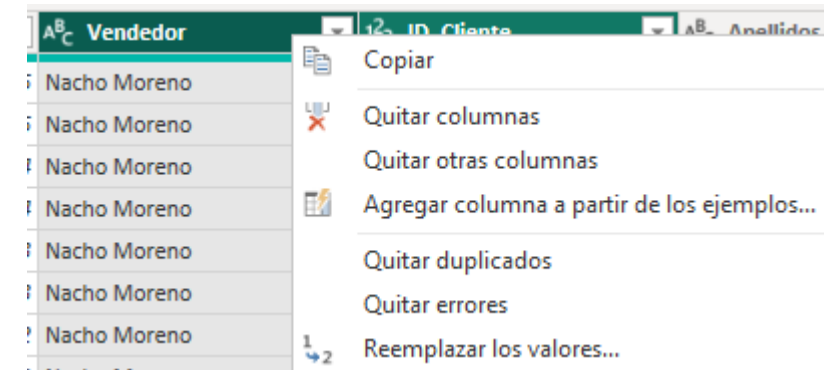
A decorative red curved shape, resembling a stylized arrow or a portion of a circle, is positioned on the right side of the slide, pointing towards the left.

# Inicio

## Eliminar columnas

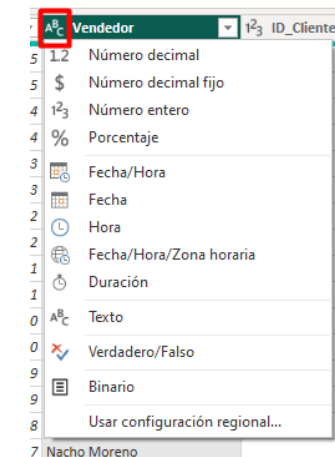
Permite quitar columnas de la tabla importada.

Se pueden eliminar varias columnas con una sola instrucción.



## Cambiar tipo de dato

- Text e.g. Bicycle
- True/False
- Whole Numbers e.g. 148
- Decimal Numbers e.g. 3.5432
- Fixed Decimal Number e.g. 5.45
- Percentage e.g. 46.8%
- Date e.g. 15/02/2020
- Time e.g. 12:35:00 PM
- Date/Time e.g. 15/02/2020 10:25:00 AM

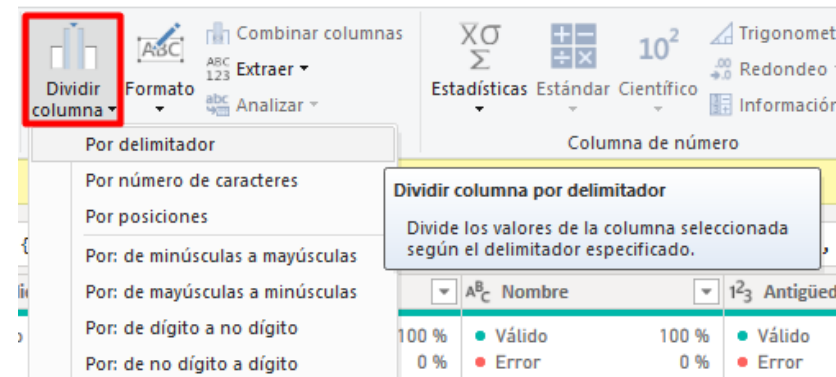




# Transformar Datos:

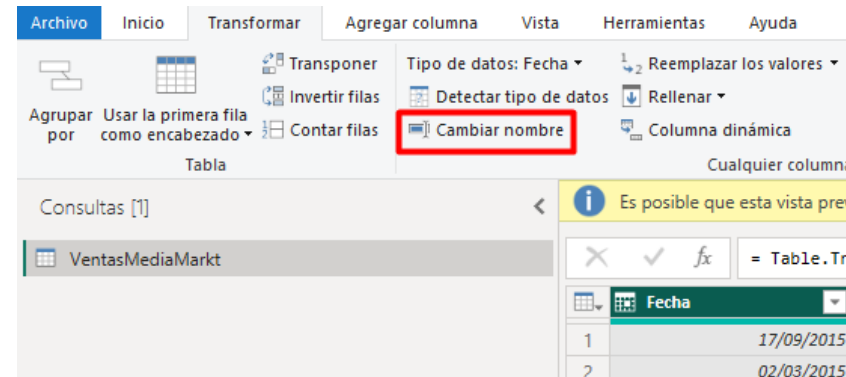
## Dividir Columna

Podemos dividir una columna separándola por diferentes criterios.



## Renombrar Campos

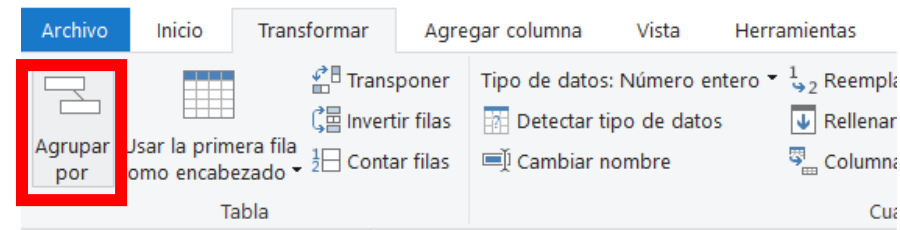
Desde la interfaz de Power Query se permite modificar los nombres internos de las columnas para que se adecuen mejor a las entidades de negocio que quieran representar



# Transformar Datos:

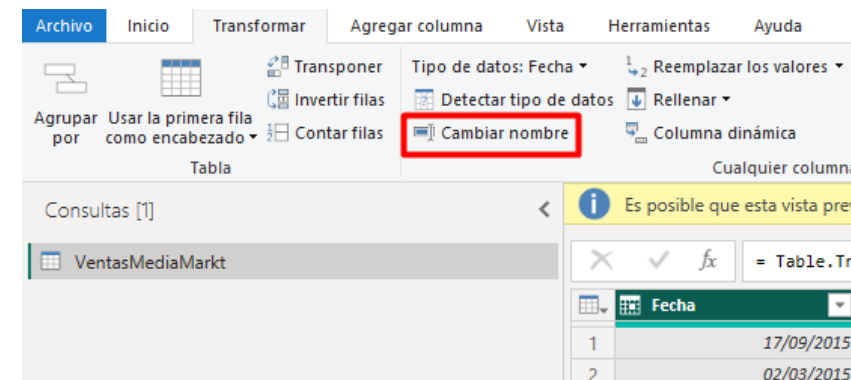
## Agrupar Por

Podemos agrupar los mismos valores de una o más columnas en una sola fila agrupada a la que se le aplica una operación (suma, mínimo, promedio...).



## Primera fila como encabezado

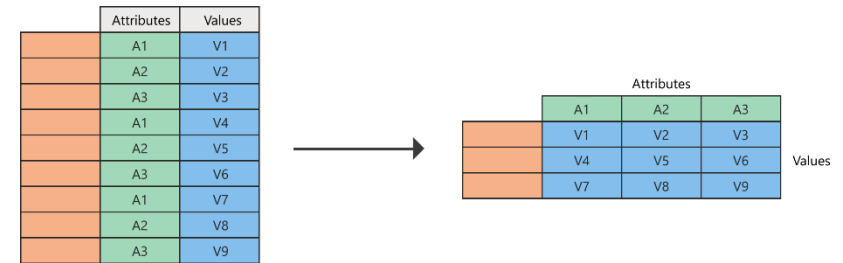
Establece como encabezado de columna la primera fila contenida en la tabla. Los nombres de las columnas de la tabla deben ser únicos. Si la fila que desea promover a una fila de encabezado contiene varias instancias de la misma cadena de texto, Power Query desambiguará los encabezados de columna agregando un sufijo



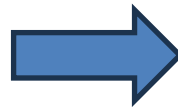
# Transformar Columnas:

## Dinamizar columnas

En Power Query, podemos crear una tabla que contenga un valor agregado para cada valor único de una columna. Power Query agrupa cada valor único, realiza un cálculo agregado para cada valor y dinamiza la columna en una nueva tabla.



	<b>A<sup>B</sup>C</b> Country	Date	1 <sup>2</sup> 3 Value
1	USA	6/1/2020	785
2	USA	7/1/2020	450
3	USA	8/1/2020	567
4	Canada	6/1/2020	357
5	Canada	7/1/2020	421
6	Canada	8/1/2020	254
7	Panama	6/1/2020	20
8	Panama	7/1/2020	40
9	Panama	8/1/2020	80

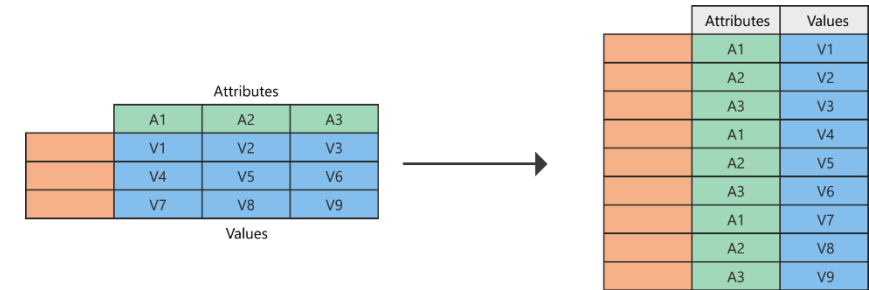


	<b>A<sup>B</sup>C</b> Country	1 <sup>2</sup> 3 6/1/2020	1 <sup>2</sup> 3 7/1/2020	1 <sup>2</sup> 3 8/1/2020
1	Canada	357	421	254
2	Panama	20	40	80
3	USA	785	450	567

# Transformar Columnas:

## Anular dinamización columnas

En Power Query, podemos crear una tabla que contenga un valor agregado para cada valor único de una columna. Power Query agrupa cada valor único, realiza un cálculo agregado para cada valor y dinamiza la columna en una nueva tabla.



	<b>A<sup>B</sup><sub>C</sub> Country</b>	<b>1<sup>2</sup><sub>3</sub> 6/1/2020</b>	<b>1<sup>2</sup><sub>3</sub> 7/1/2020</b>	<b>1<sup>2</sup><sub>3</sub> 8/1/2020</b>
1	Canada	357	421	254
2	Panama	20	40	80
3	USA	785	450	567

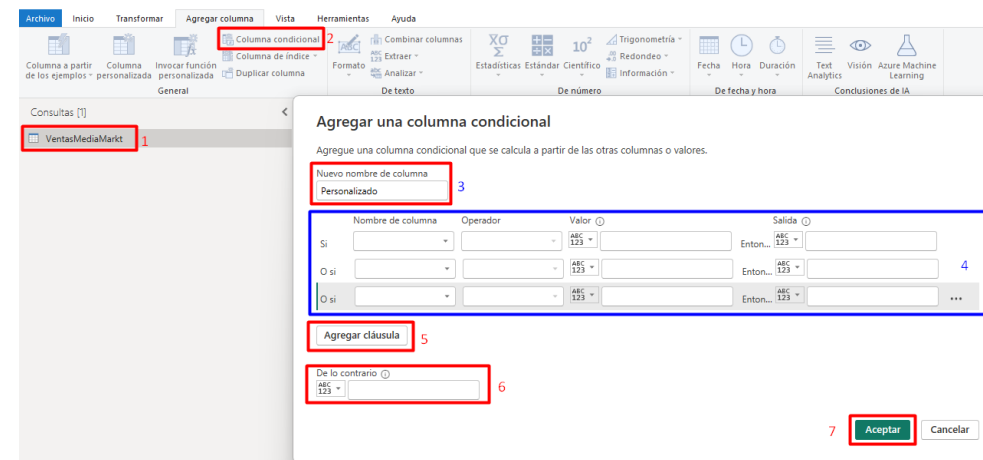


	<b>A<sup>B</sup><sub>C</sub> Country</b>	<b>Date</b>	<b>1<sup>2</sup><sub>3</sub> Value</b>
1	USA	6/1/2020	785
2	USA	7/1/2020	450
3	USA	8/1/2020	567
4	Canada	6/1/2020	357
5	Canada	7/1/2020	421
6	Canada	8/1/2020	254
7	Panama	6/1/2020	20
8	Panama	7/1/2020	40
9	Panama	8/1/2020	80

# Agregar Columnas:

## Columna condicional

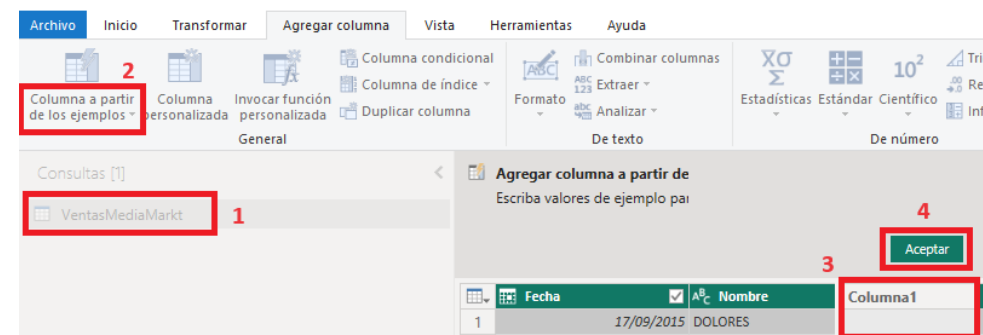
Permite crear nuevas columnas aplicando algunas reglas definidas por los analistas de negocio.



## Columna a partir de ejemplos

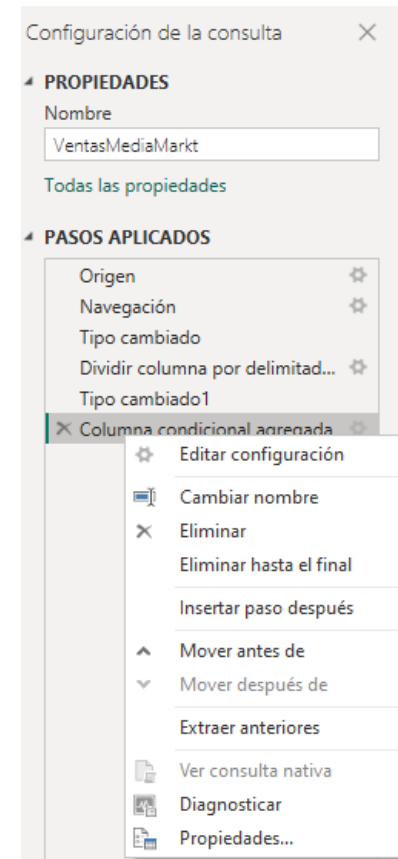
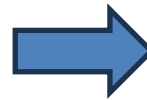
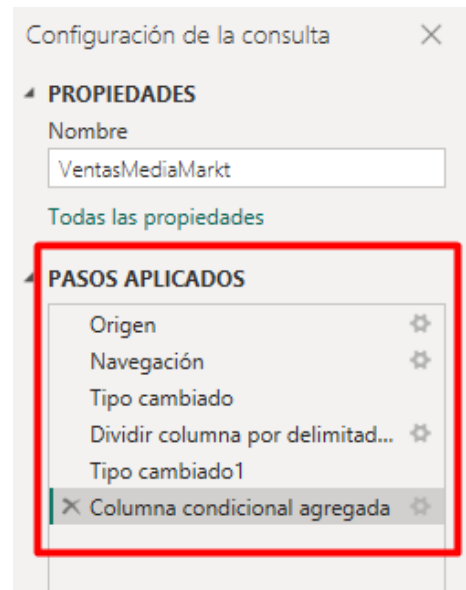
Mediante métodos de búsqueda de patrones y cierta inteligencia, detecta automáticamente la operación que se quiere aplicar sobre los datos sin necesidad de aplicar el paso específico.

Necesita que el analista le introduzca varios ejemplos del resultado deseado.



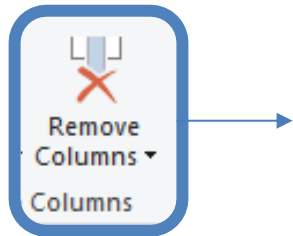
# Pasos en Power Query

Todos los pasos de transformación que vamos añadiendo se acumulan en el panel derecho, donde quedará toda la trazabilidad de las transformaciones aplicadas. Podremos además renombrar los pasos para que sean mas aclaratorios, eliminar pasos, mover los pasos de transformación...



# El lenguaje M

A través de M, Power Query traduce a sintaxis de programación los diferentes pasos seleccionados por el usuario.



# OPERACIONES COMUNES



# Análisis de datos exploratorio (EDA)



# Análisis de datos exploratorio (EDA)

## 1 ANÁLISIS DESCRIPTIVO



- **¿Qué es?** Síntesis de la información que proporciona el conjunto de datos, extrayendo sus características más representativas.
- **¿Por qué es necesario?** Permite conocer los tipos de datos, descubrir patrones y preparar los datos para futuros análisis.
- **Tratamiento:** Aplicar funciones de estadística descriptiva para explorar la estructura del conjunto de datos, examinar los datos y las variables que presenta.

## 2 RE-AJUSTE DE LOS TIPOS DE VARIABLES



- **¿Qué es?** Verificar que las variables se han almacenado con el tipo de valor correspondiente.
- **¿Por qué es necesario?** Una mala codificación de las variables puede influir negativamente en la agrupación de los datos o los resultados de los análisis.
- **Tratamiento:** Aplicar la codificación apropiada para cada una de las variables.

# Análisis de datos exploratorio (EDA)

## 3 DETECCIÓN Y TRATAMIENTO DE DATOS AUSENTES



- **¿Qué es?** Identificar la falta de algunos de los datos en la variable.
- **¿Por qué es necesario?** Los datos ausentes pueden generar problemas a la hora de aplicar técnicas de machine learning, elaborar modelos predictivos, realizar análisis estadísticos o generar representaciones gráficas.
- **Tratamiento:** Existen varias maneras de tratar los valores ausentes, como por ejemplo sustituirlos por la media o la mediana, o completar los valores faltantes con el valor anterior o posterior de la columna.

## 4 DETECCIÓN Y TRATAMIENTO DE DATOS ATÍPICOS



- **¿Qué es?** Identificar datos con valores significativamente distintos a los que presenta la variable.
- **¿Por qué es necesario su tratamiento?** Pueden modificar los resultados y restar potencia a los análisis estadísticos o técnicas de machine learning aplicadas.
- **Tratamiento:** Disminuir su influencia en análisis posteriores o, en casos muy extremos, eliminarlos del conjunto de datos.

# Análisis de datos exploratorio (EDA)

## 5 ANÁLISIS DE CORRELACIÓN DE VARIABLES



- **¿Qué es?** Analizar la relación entre dos o más variables.
- **¿Por qué es necesario?** Entre otras razones, para descartar posibles variables que aporten información redundante en el conjunto de datos, ocasionando ruido en los análisis.
- **Tratamiento:** Calcular los coeficientes de correlación para las variables para detectar coeficientes cercanos a 1 o -1.

**Fundación  
iberCaja**

