# Introducción a Power BI

## ¿Qué es Microsoft Power BI?

Según Microsoft, Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente. Es decir, es una herramienta que nos permitirá tomar datos de diferentes fuentes para analizar esa información y crear informes donde se le dé buen uso a dicha información. Estas fuentes de datos pueden ser desde ficheros Excel hasta sistemas de almacenamiento de datos en la nube, pasando muchos otros como, por ejemplo, bases de datos de Access.

## Descargando Power BI

Existen 3 formatos de Power BI: aplicación de escritorio, web y móvil. En este taller veremos la opción de escritorio, que es la más completa. Lo podemos descargar de la web oficial: <https://go.microsoft.com/fwlink/?LinkID=521662>

Para realizar algunas acciones será necesario disponer de una cuenta de PowerBI, por lo que en este taller no veremos estas características, que básicamente se reducen al uso de extensiones (de un catálogo no demasiado amplio) y a la publicación e integración con otras herramientas de Office365, como Microsoft Teams o la creación de apps. Por lo tanto, como no es algo tan limitante, podremos seguir el taller perfectamente, pues la versión gratuita nos permitirá compartir nuestros informes con otras personas que también tengan la aplicación de escritorio de Power BI instalada en su ordenador, sin falta de publicar.

En este caso utilizaremos la versión gratuita, sin embargo, para sacar el máximo partido de Power BI, como puede ser cualquier entorno productivo de una empresa, el desarrollo de los cuadros de mando terminaría con una publicación que permitiese al resto de miembros de la empresa acceder a los informes generados mediante la aplicación web, incluso pudiendo llegar a actualizar sus datos de manera automática.

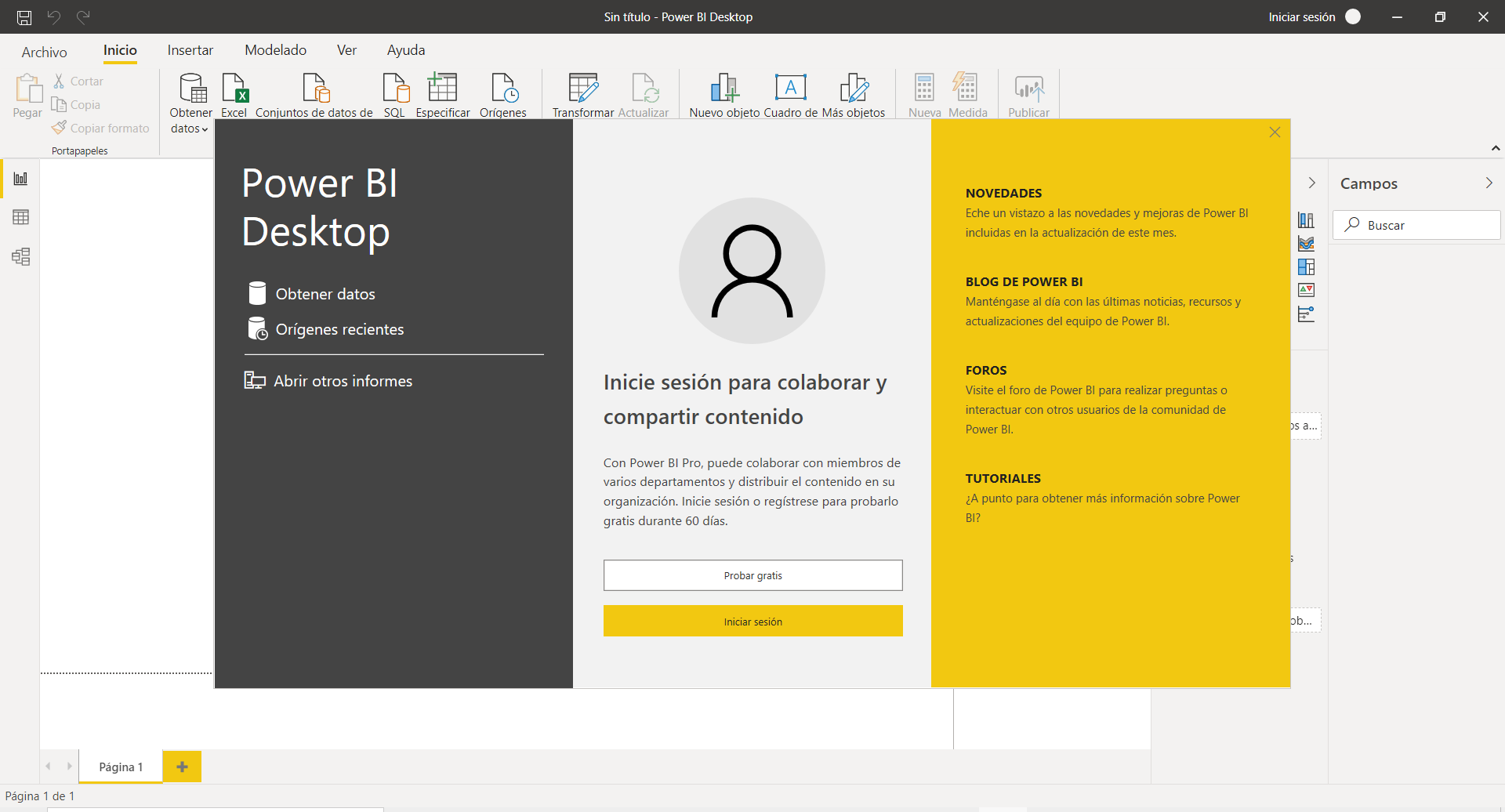
## Instalación Power BI

Una vez nos hemos hecho con el fichero de instalación, lo ejecutamos y seguimos las instrucciones que nos sugiere el asistente de instalación. Vamos, que le damos todo a “Siguiente”, aceptamos los términos de Contrato de licencia cuando nos lo pida, y seleccionamos la carpeta donde queremos instalarlo. En este caso, yo recomendaría instalarlo en la carpeta por defecto, que será algo así como “C:\Program Files\Microsoft Power BI Desktop\”, ya que en caso de fallos en la instalación será más sencillo localizarlo.

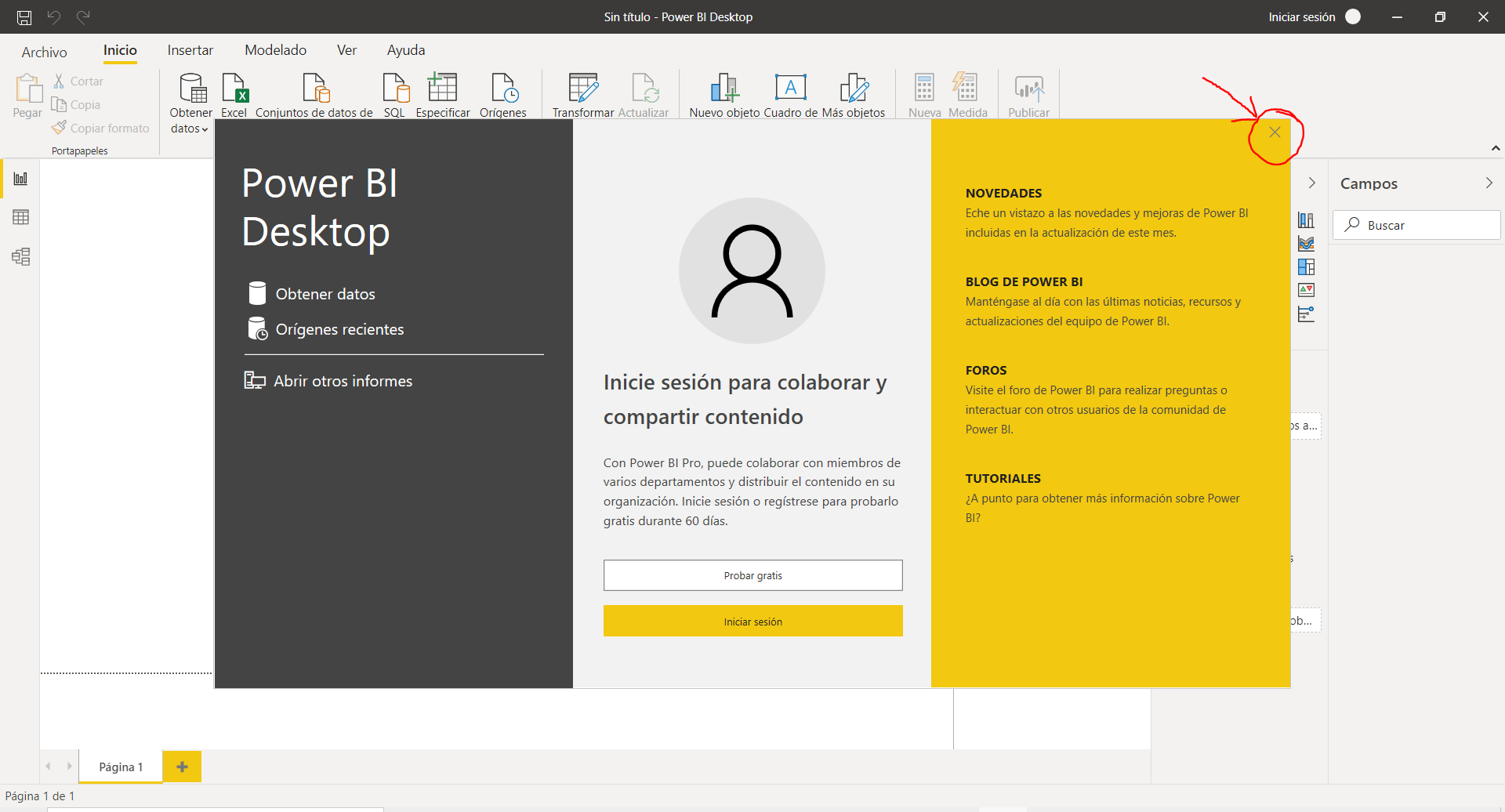
Tras finalizar el proceso de instalación, nos debería saltar la siguiente ventana:



Le damos a “Finalizar” y se nos abrirá el programa, mostrando la siguiente interfaz:



En ella, nos pedirá registrarnos o identificarnos con nuestra cuenta Power BI Pro. Sin embargo, no es necesaria para utilizar el programa de forma gratuita, por lo que pincharemos en la X de arriba a la derecha de la ventanita que nos pide la cuenta y podremos acceder a la funcionalidad del programa.



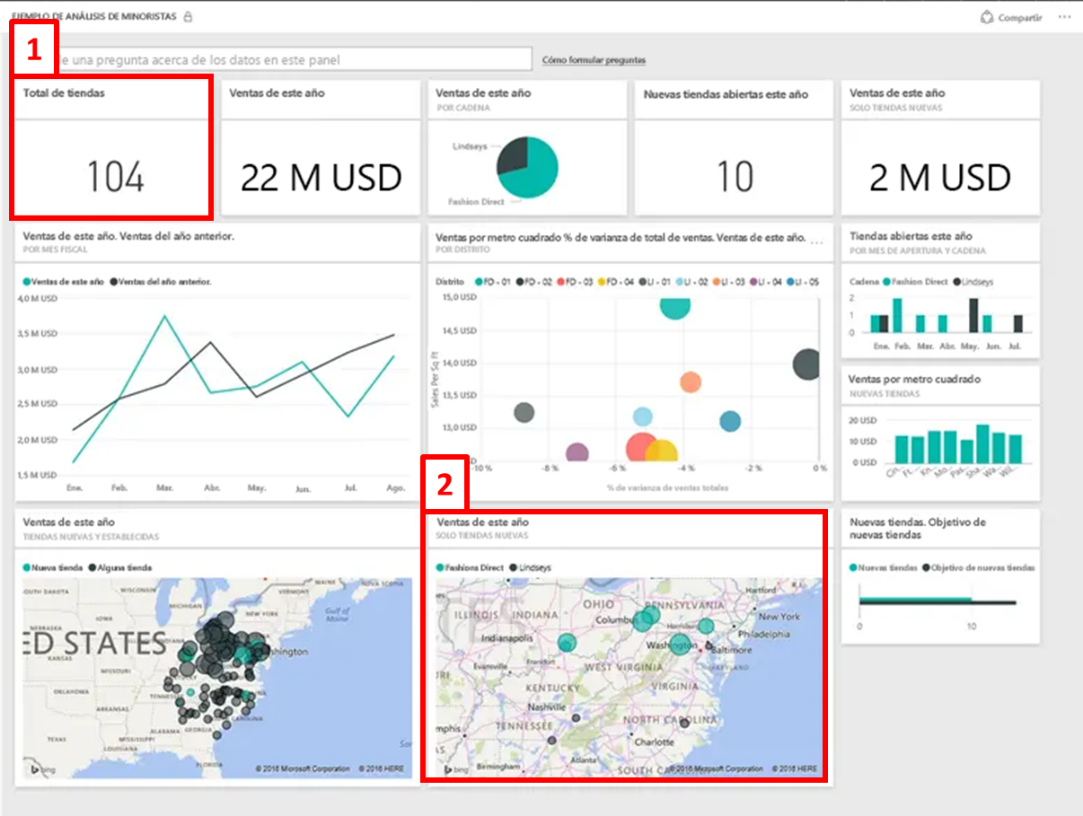
## Elementos básicos Power BI

Ya estaríamos en condiciones de empezar a crear nuestros informes, sin embargo, para terminar este apartado de introducción, se hace necesario definir los elementos más básicos para entender el funcionamiento del programa, que reciben el nombre de bloques de creación básicos.

* **Visualizaciones:**

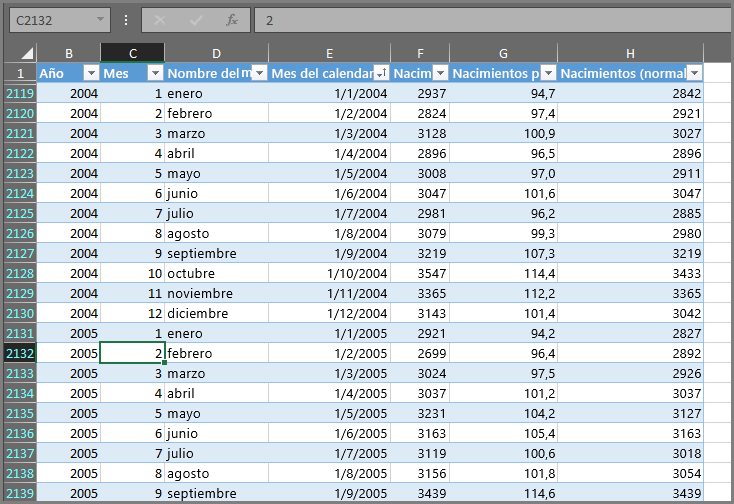
Una visualización, u objeto visual, es una representación visual de los datos. Pueden ser simples, como un simple número que refleje una cantidad (1), o complejas, como un mapa de colores degradados que muestre la opinión del votante sobre un determinado problema o preocupación social (2). El objetivo de la visualización es presentar los datos de forma que ofrezca contexto e información detallada, lo que probablemente resultaría difícil tan solo con una tabla de números o texto sin formato.

A continuación, se recoge un informe con diversas visualizaciones:



* **Conjuntos de datos:**

Un conjunto de datos es una colección de datos que Power BI utiliza para crear las visualizaciones. El conjunto de datos puede ser algo sencillo como una sola tabla de un libro de Microsoft Excel, similar a este.

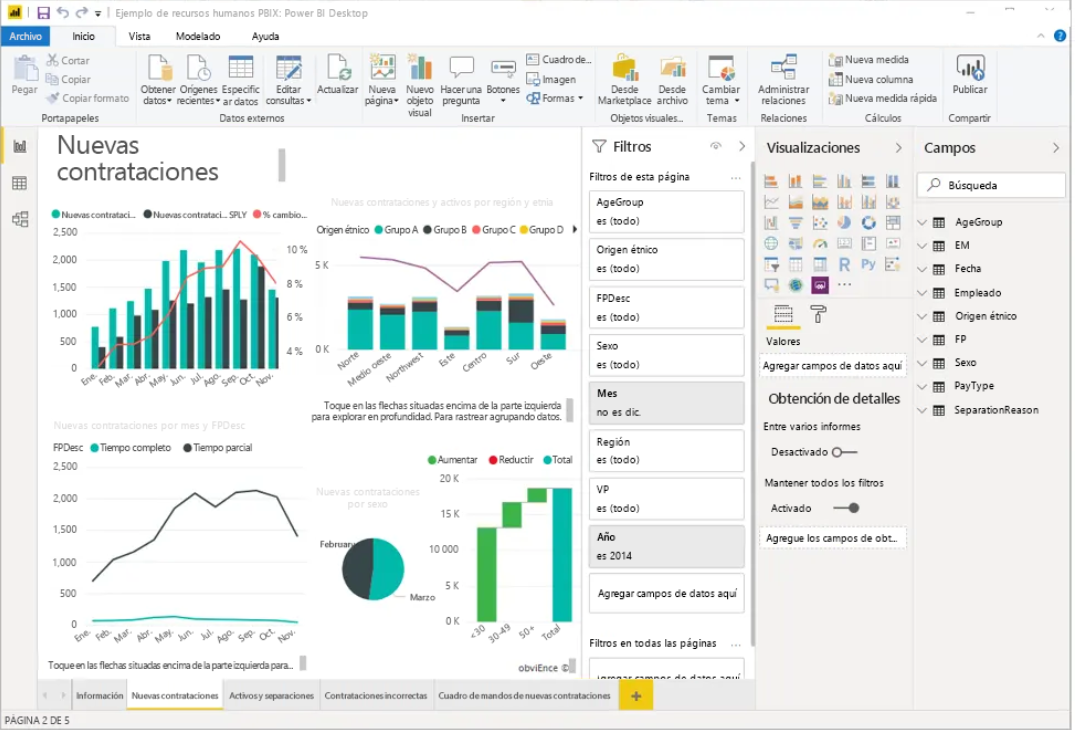


Sin embargo, un conjunto de datos también podría ser algo más complejo, como una combinación de muchos orígenes diferentes, que podremos filtrar y combinar para utilizar en nuestros informes. Será en esta parte donde realizaremos el procesamiento de datos para que lleguen limpios a las visualizaciones, pero no solo filtrar sino también hacer cruces de tablas, cambiar la forma de interpretar los datos por Power BI o seleccionar ciertos campos.

* **Informes:**

Un informe es una colección de visualizaciones que aparecen juntas en una o varias páginas. Es importante remarcar que los elementos de un informe están relacionados entre sí, por lo que podremos hacer que al interactuar con uno de ellos se produzcan cambios en los otros, como seleccionar un mes para que nos aparezcan los datos de ese mes en el resto de visualizaciones. Además, estas relaciones se podrán extender entre diferentes páginas.

A continuación, se recoge un ejemplo del programa donde podemos observar las distintas partes de un informe, como las pestañas para navegar entre páginas o los filtros que podremos aplicar a nuestros informes, que veremos más adelante:



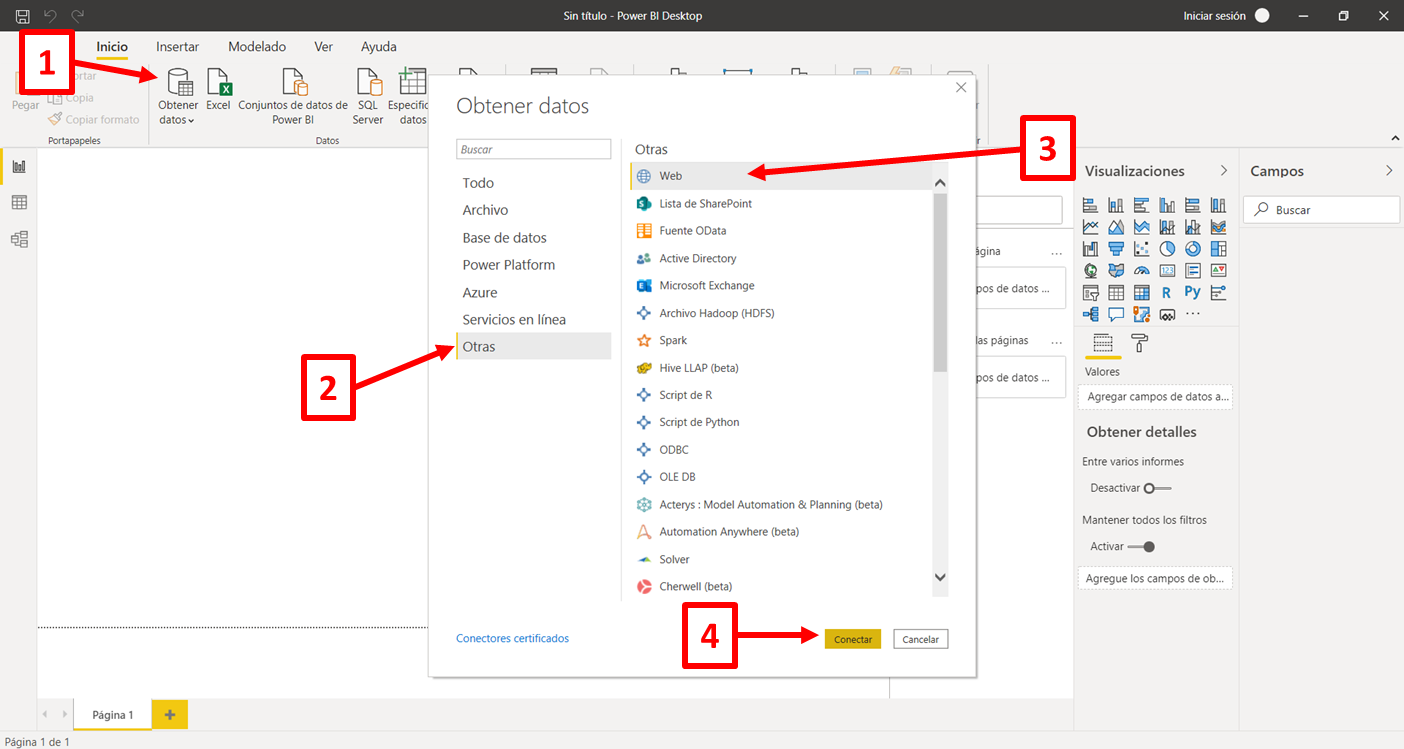
# Importando los datos en Power BI

Una vez conocemos lo más básico de Power BI, podemos comenzar a jugar con los datos. Para ello, veremos cómo creamos paso a paso un reporte del mundial de fútbol de 2014, el cual está estructurado siguiendo los 3 grandes bloques funcionales en el procedimiento genérico de uso de Power BI: procesamiento de datos, modelado del sistema y visualización.

## ¿Cómo leemos los datos?

Como hemos comentado, tenemos varias formas de hacernos con los datos, desde un simple fichero Excel hasta bases de datos no estructurados en la nube. En este caso, vamos a utilizar la funcionalidad de Power BI para poder leer tablas de una dirección web, donde estarán los datos que queremos importar. En concreto, vamos a utilizar los datos de la web: <https://www.mykhel.com/football/fifa-world-cup-2018-team-stats-l4/>

Comenzaremos cargando la tabla de las estadísticas de cada equipo, para lo cual haremos lo que se indica en la imagen:

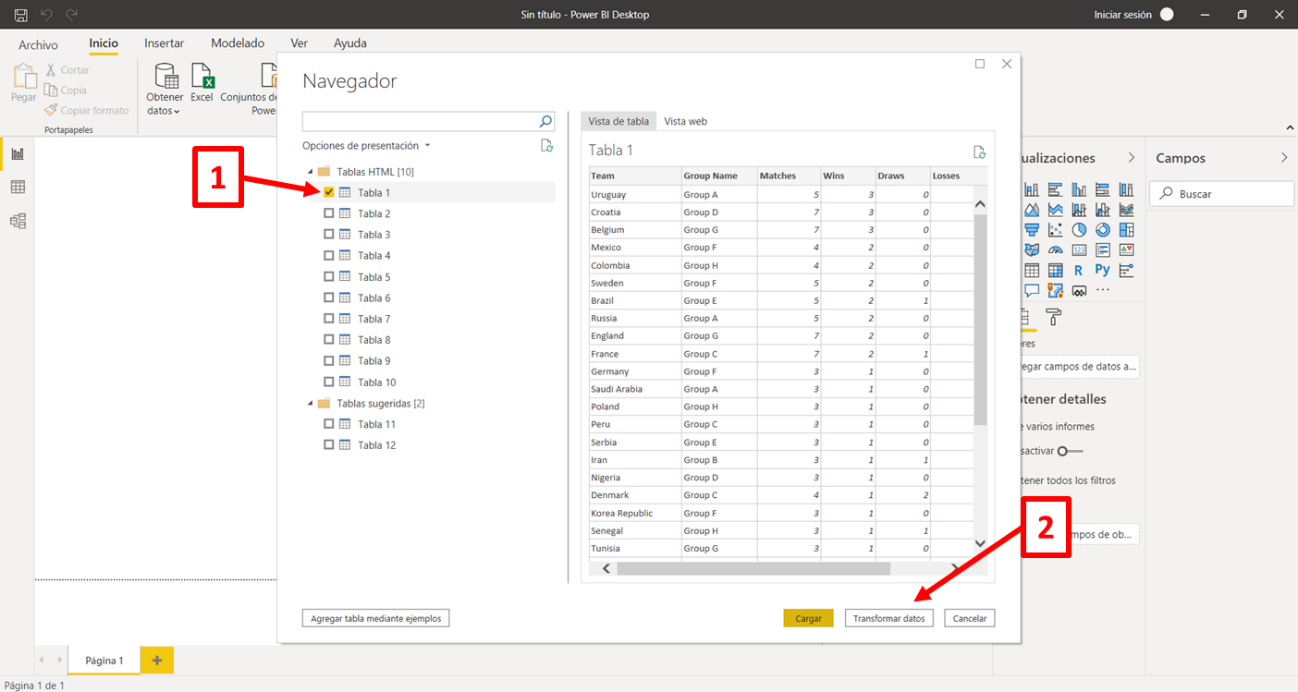


Como podemos observar, hay muchos tipos de fuentes de datos de los que puede beber Power BI. En este caso, utilizaremos Web.

Al seleccionar conectar, tendremos que rellenar los datos de la web a la que nos conectaremos. Introducimos <https://www.mykhel.com/football/fifa-world-cup-2018-team-stats-l4/> y aceptamos (dejando Básico seleccionado):

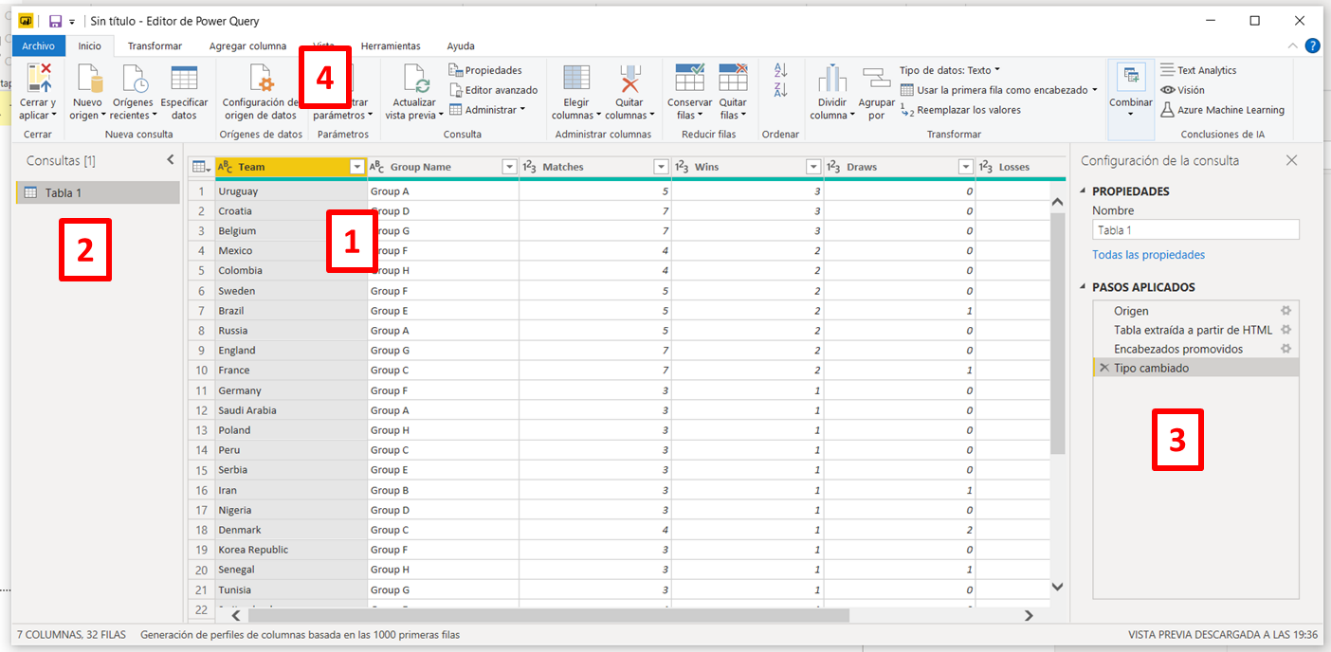


Después de cargar, nos aparecerán todas las tablas que haya detectado en la web. De momento, solo nos interesa la primera, así que nos quedaremos con ella. En este punto tenemos 2 opciones: cargarla directamente (Cargar) o editar alguna cosa de la tabla antes de cargarla (Transformar datos), así que, ya que estamos aprendiendo, vamos a optar por la segunda:

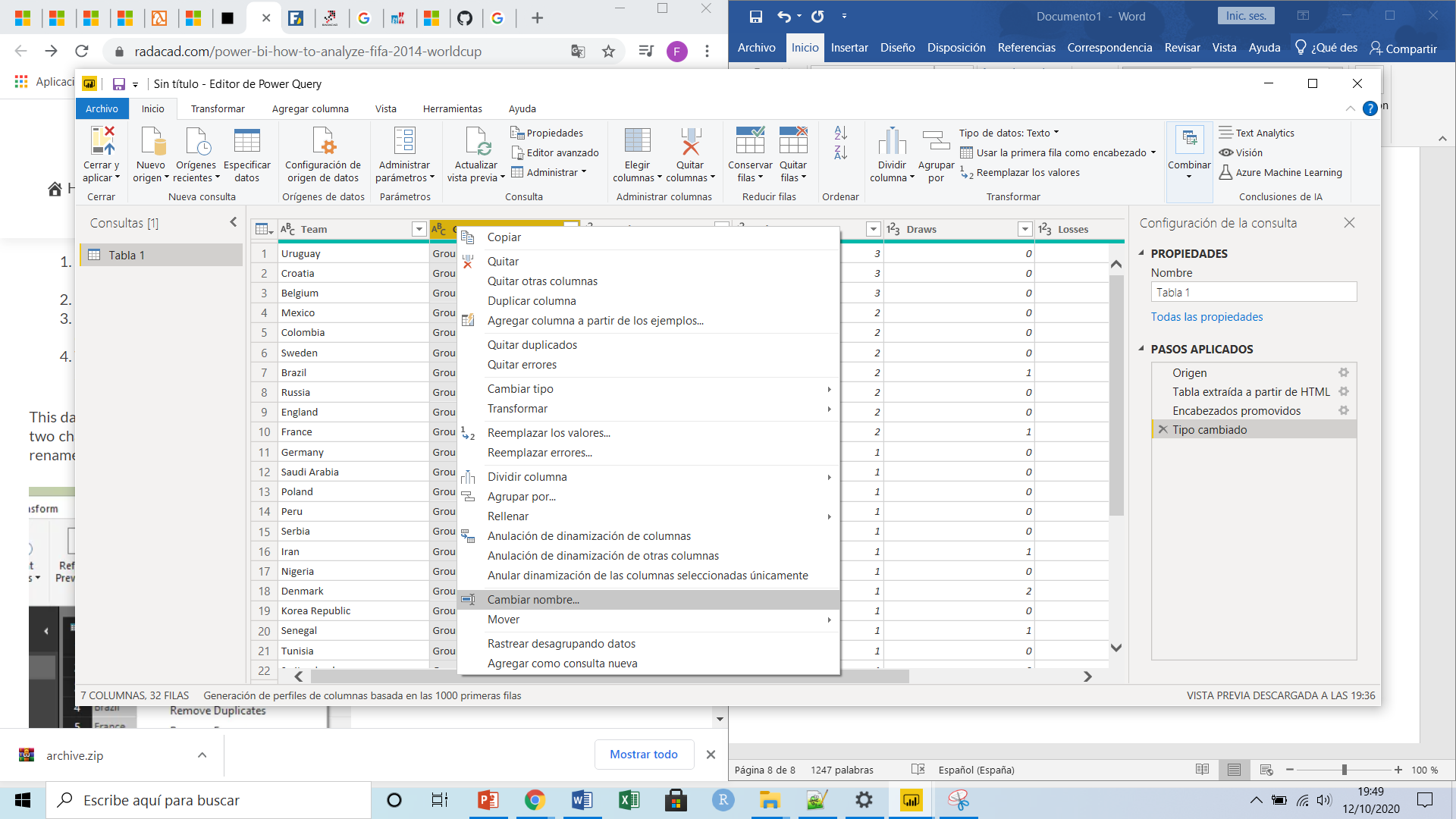


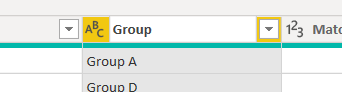
Al seleccionar “Transformar datos” se nos abrirá el editor de consultas que tiene Power BI, llamado Power Query, donde podemos editar nuestro dataset y adaptarlo a nuestro modelo de datos de Power BI. Su estructura es como sigue:

1. Muestra una vista previa de la tabla tras aplicarle las transformaciones que hayamos definido.
2. Indica las consultas que hemos realizado, que se corresponderán con las tablas de nuestro modelo de datos.
3. Lista con cada una de las transformaciones que se han realizado sobre los datos, como la lectura de datos desde su origen o el cambio de nombre de una columna.
4. Panel de transformación, donde podremos seleccionar las modificaciones que queremos realizar a nuestros datos de un modo visual.

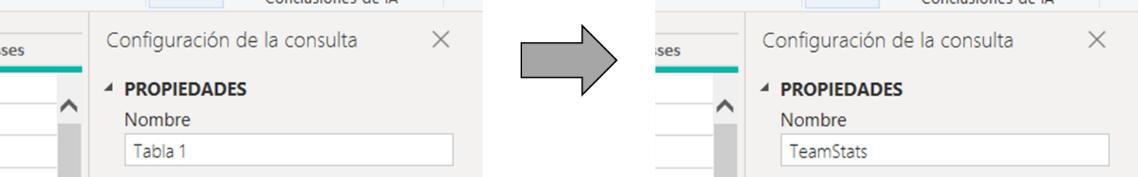


Como podemos ver a simple vista, en esta tabla tenemos los datos limpios para utilizar directamente. Sin embargo, vamos a cambiar el nombre de la columna “Group Name” por “Group”. Para ello, simplemente accederemos a la columna y haremos clic derecho sobre ella. Seleccionamos “Cambiar nombre…”, tecleamos el nuevo nombre y pulsamos Enter:

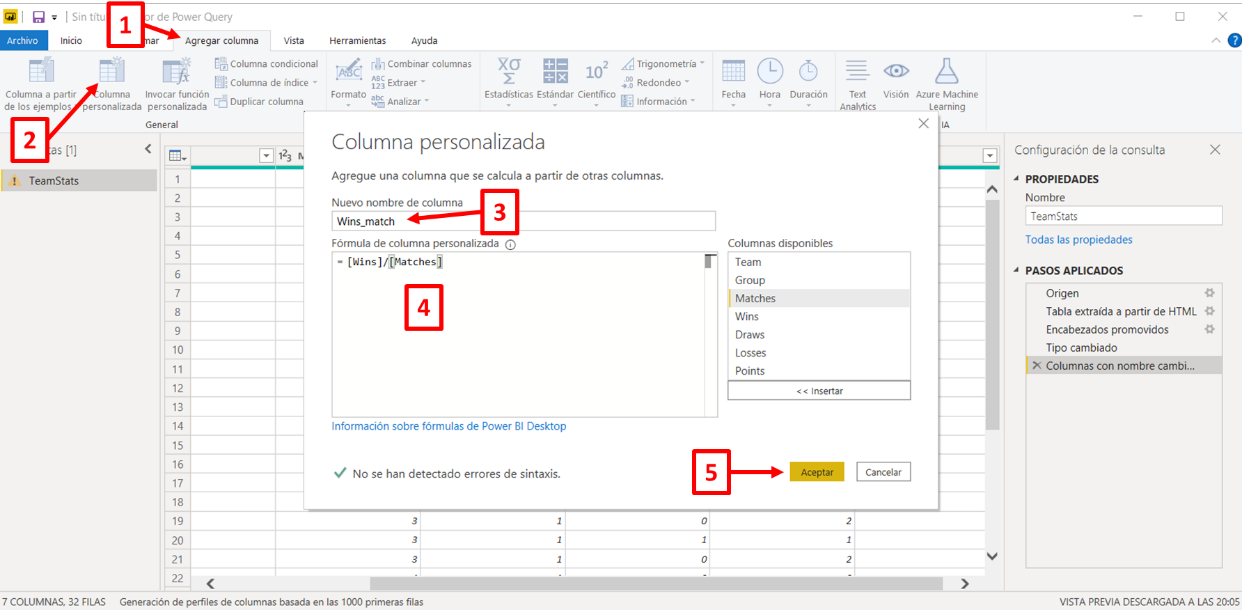




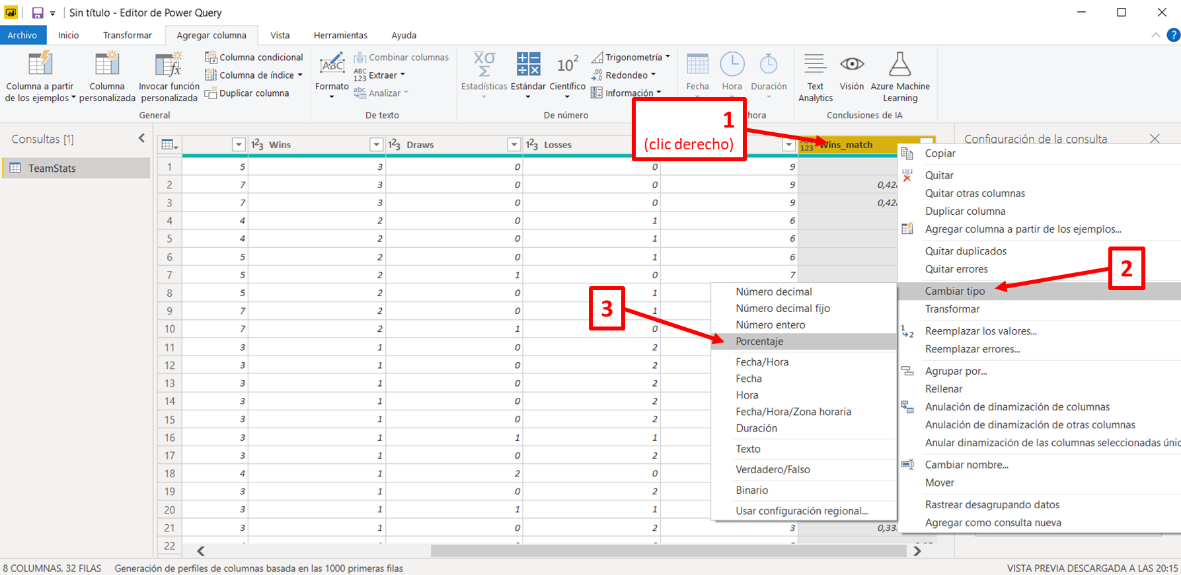
También aprovecharemos para cambiar el nombre de la tabla en la parte de Propiedades de la tabla (a la derecha):



Además, podemos crearnos nuevas columnas con información que no tenemos de forma explícita en la tabla, a partir de otras columnas de la misma, como podría ser calcular la media de victorias por partido. Para ello, crearemos una nueva columna basada en fórmulas, accediendo a “Agregar columna” 🡪 “Columna personalizada”, donde podremos escribir fórmulas para combinar columnas, lo cual es especialmente útil cuando trabajamos con columnas numéricas. En este caso, realizaremos la división de la columna “Wins” entre la columna “Matches” para sacar el ratio victorias/partido, que llamaremos “Wins\_match”. Para insertar una columna, haremos doble clic sobre su nombre.



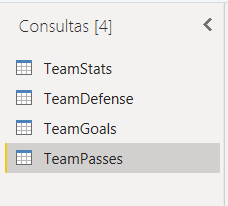
Finalmente, cambiaremos el tipo de datos de la columna que hemos creado. Se ha creado como decimal, pero estamos trabajando con un porcentaje. Power BI trata de forma distinta los porcentajes, por lo que será mejor convertirlo a este tipo de datos. Esto es muy sencillo, tal y como se muestra en la siguiente imagen.



Existen otras muchas funciones predefinidas en Power BI para realizar el procesamiento de datos, como reemplazar textos, agrupaciones de datos, conversión de fechas…, las cuales están perfectamente explicadas en su documentación y recurriremos a ellas en función de lo que necesitemos para cada caso particular. Por ello, en este taller no veremos todas y cada una de ellas, sino una visión genérica de un caso de uso.

## Leyendo más datos

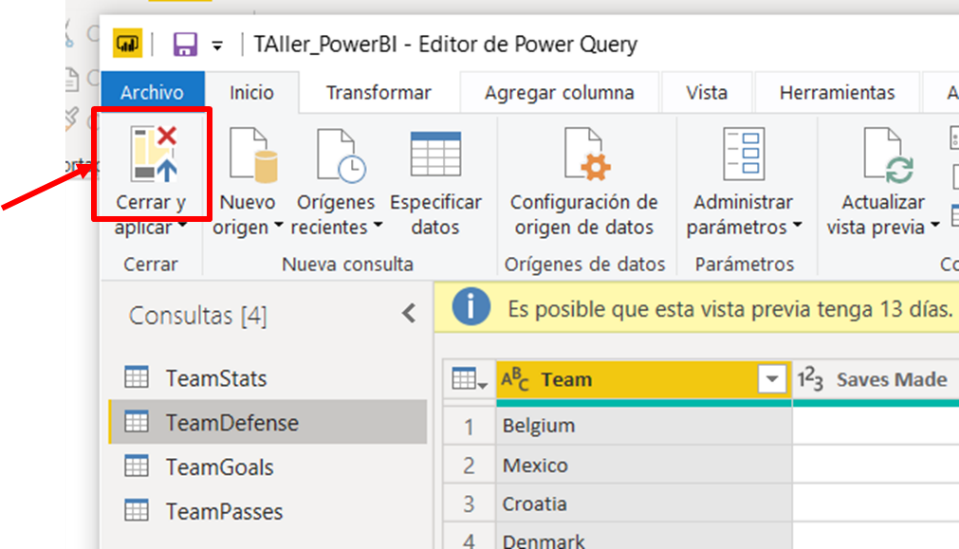
Hemos visto cómo leer datos en una tabla, sin embargo, tenemos mucha más información que queremos utilizar. Por lo tanto, nuestro sistema se compondrá de más tablas. Para cargar estas tablas nuevas podemos hacer como hemos visto antes, para lo no es necesario salir del Query Editor, ya que accederemos al menú anterior haciendo clic en “Inicio” -> “Nuevo Origen”. Desde ahí haremos lo mismo que antes, solo que seleccionaremos las tablas que nos interesen. En este caso, seleccionaremos las tablas 2, 3 y 6, que renombraremos como “TeamDefense”, “TeamGoals” y “TeamPasses”, respectivamente:



# Modelado de datos en Power BI

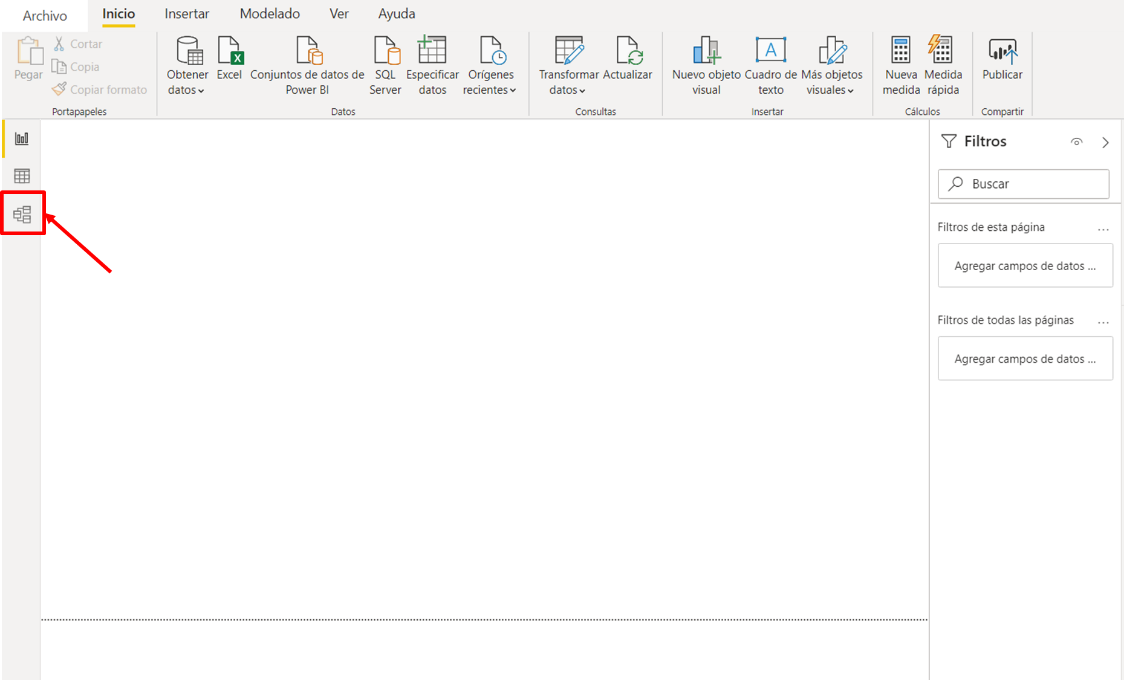
Tras importar los datos, tenemos que establecer la relación entre ellos ya que, si no, no podremos relacionar nuestras tablas.

Comenzaremos aplicando los cambios realizados en el apartado anterior haciendo clic en el siguiente botón:

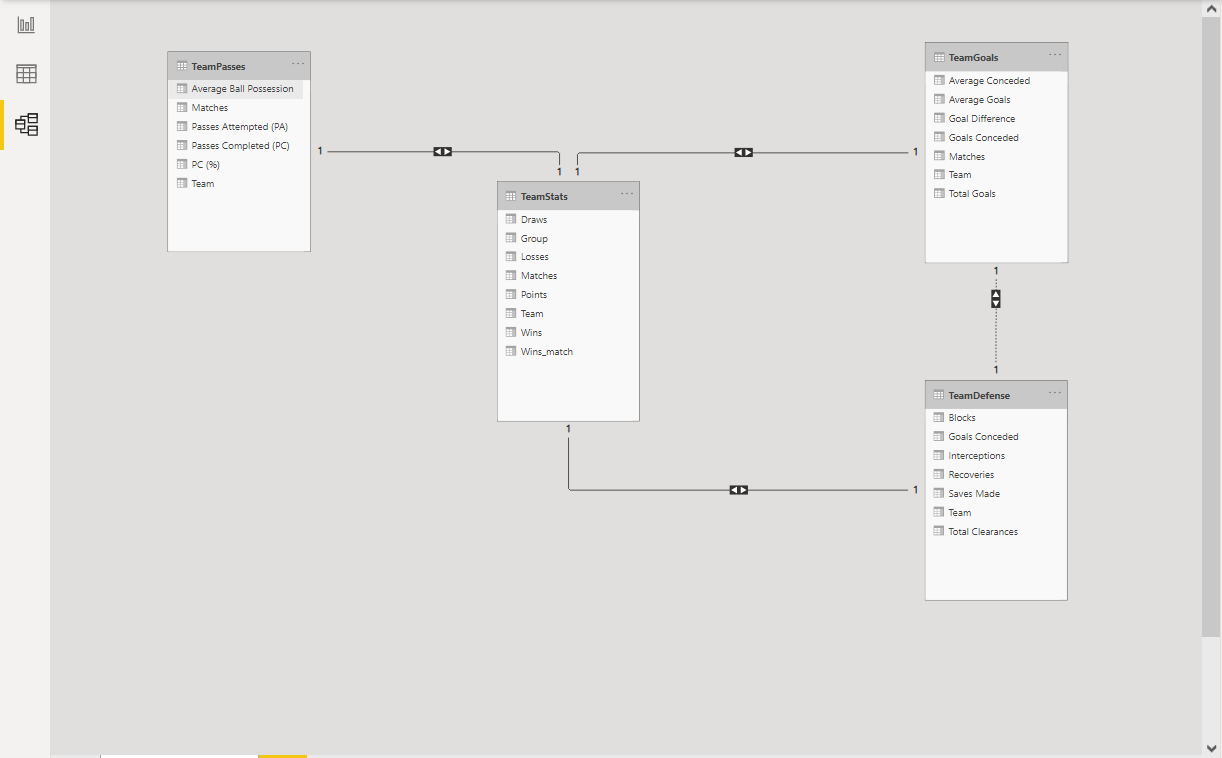


## Relacionando datos

Una vez realizado esto, pasaremos a establecer las relaciones entre nuestros datos. Para ello, accedemos a la pestaña de relaciones:

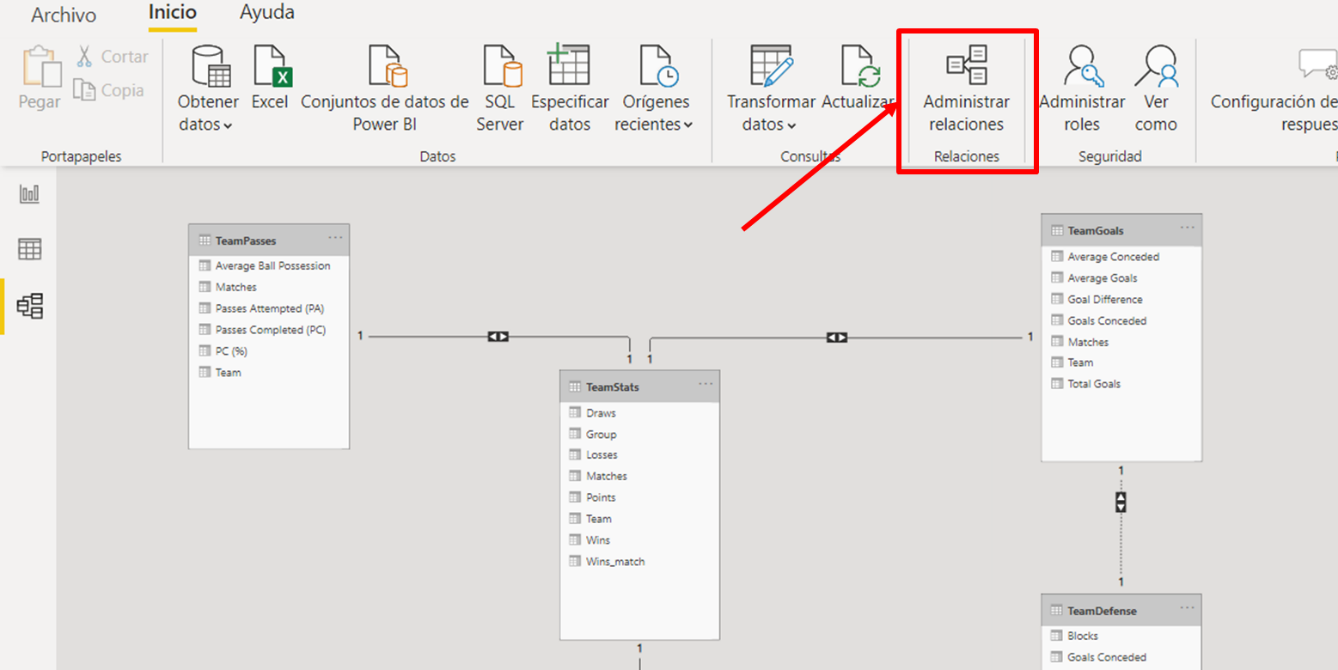


En esta pestaña, veremos cómo se relacionan las tablas. Podremos arrastrarlas de la parte derecha al centro para establecer sus relaciones. Por defecto, tenemos una pestaña con todas las tablas, que será en la que nos centremos en este taller, donde se representa gráficamente todas las relaciones entre las tablas (inicialmente, las tablas estarán repartidas aleatoriamente por el espacio pero las podemos mover para facilitarnos el entendimiento gráficamente):

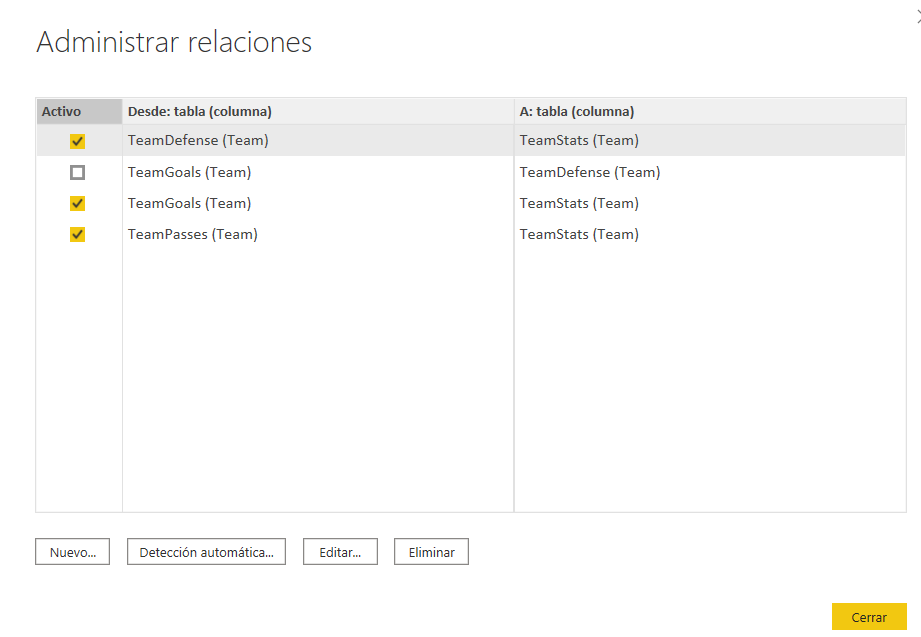


Por defecto, Power BI identifica las posibles relaciones entre tablas, por lo que nos ahorra mucho trabajo. Sin embargo, esta inteligencia para detectar relaciones no siempre funciona tan bien como nosotros quisiéramos, sobre todo cuando hay más de un campo que tiene una estructura semejante y nombres de campo distintos. Por eso, será mejor repasar las relaciones.

Si queremos repasar las relaciones que se están utilizando, podemos acceder al “Administrador de relaciones”:



Ahora, nos aparecerá una tabla donde se identifica cada una de las relaciones:



Como podemos comprobar es bastante intuitivo:

* **Activo**: indica si se está usando esa relación o no. En Power BI podemos indicar más de una relación pero, dependiendo del caso, podríamos necesitar desactivar (o activar) alguna de ellas. Generalmente, suele ocurrir cuando tenemos tablas con relaciones tan claras que hacen que exista una relación implícita entre algún par de tablas. Lo veremos con más detalle en cuanto profundicemos en los cruces.
* **Desde**: indica la tabla y la columna (entre paréntesis) de cruce de la tabla de la izquierda. “Desde”, junto con “A” permitirán establecer el sentido del cruce. En este caso, lo normal será indicar en “Desde” la columna de la tabla con los registros que se van a buscar en los de la columna de la tabla que ponga “A”.
* **A**: De manera análoga a “Desde”, indica la columna de la tabla de la derecha del cruce, es decir, donde se van a buscar los registros para hacer el cruce.

Si quisiéramos editar alguno de los cruces, podemos seleccionar la fila con la relación que queremos editar (no tenemos por qué darle a “Activo”, simplemente hacer clic en cualquier lugar de esa fila) y “Editar”.

Si queremos crear una relación nueva, haremos clic en “Nueva”, que hará que nos aparezca una ventana con el detalle del cruce:



1. Selección de Tabla desde la que se realizará el filtro, es decir, la tabla que contendrá el campo que filtrará los registros de la tabla que indicaremos en 3.
2. Selección de la columna que utilizaremos para hacer este filtrado. La selección de valores de este campo, que afectará a los registros de la tabla indicada en 3.
3. Selección de Tabla que se filtrará en función del cruce. Los datos que se mostrarán en el dashboard se verán afectados por esta relación, que estará gobernada por 2.
4. Selección de la columna que se utilizará para “sufrir” el filtrado.
5. Cardinalidad: indica el tipo de relación esperada:
   1. Varios a 1 (\*: 1) 🡪 1 o más registros de la columna de la tabla “Desde” se relaciona con 1 único registro de la columna de la tabla “A”. Esto implica que no se puede repetir ningún valor en los registros de la columna de la tabla “A”.
   2. Uno a varios (1: \*) 🡪 1 registro de la columna de la tabla “Desde” se relaciona con 1 o más registros de la columna de la tabla “A”. Esto implica que los registros de la columna de la tabla “Desde” tienen que ser únicos.
   3. Uno a uno (1: 1) 🡪 El cruce con el que vamos a tener mayor control. 1 registro de la columna de la tabla “Desde” se relaciona con un único registro de la columna de la tabla “A”. Esto implica que tanto los valores de la columna de la tabla “Desde” como los de la tabla “A” deben ser únicos.

En este ejemplo, todos los cruces van a ser 1:1, ya que, al ser tablas de resumen agrupadas por equipo, no habrá ninguna de ellas que tenga más de un registro para cada equipo.

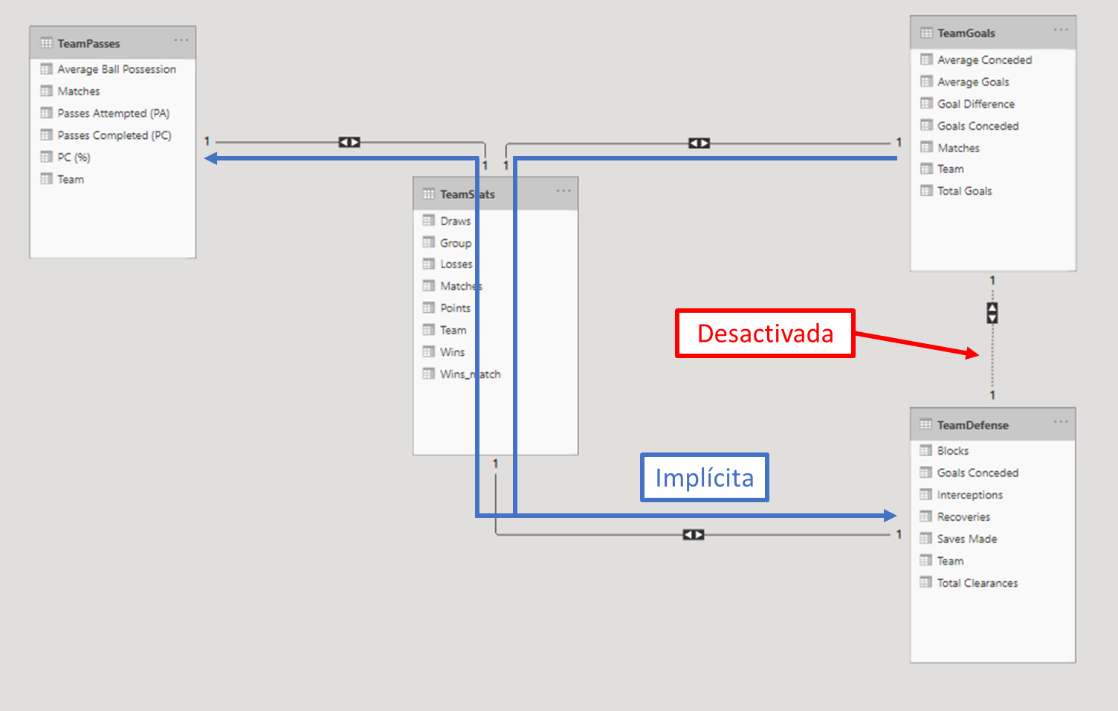
* 1. Varios a varios (\*: \*) 🡪 Cruce sin control. 1 o más registros de la columna de la tabla “Desde” se relaciona/n con 1 o más registros de la columna de la tabla “A.

1. Con la dirección del filtro de cruzado podremos establecer si queremos que la tabla “Desde” filtre a la tabla “A” o que ambas se puedan filtrar indistintamente:
   1. Ambas: Si selecciono una serie de valores de la columna de la tabla “Desde”, filtraré la tabla “A”, del mismo modo que si filtro la tabla “A” seleccionando cualquier combinación de valores de la columna establecida, filtraré “Desde”.
   2. Única: En el caso de que la relación no sea 1:1, habrá que establecer una prioridad en el cruce. Al seleccionar esta opción, se podrá elegir si queremos darle prioridad a la tabla “Desde” o a la tabla “A” (normalmente se suele seguir la convención al definir la relación para darle prioridad a la tabla de la izquierda). Los registros de la tabla “A” se filtrarán en función de los valores seleccionados de la columna de la tabla “Desde”, pero la selección de valores de la tabla “A” no afectará a la tabla “Desde”. Esta opción no se puede utilizar en las relaciones “1 a 1”, ya que la relación está implícita.

Generalmente, se recomienda que no haya relaciones “Varios a varios (\*: \*)”, y que, en el caso de relaciones que no sean “Uno a uno”, la dirección sea única, ya que mejorará tanto el rendimiento del software como el modelo en sí.

Además de estas opciones, podremos hacer la “Detección automática” de cruces, donde Power BI analizará las tablas para establecer las posibles relaciones (tenemos que darle los datos bien preparados para que lo detecte de la mejor manera posible, como este caso). También podremos “Editar” alguna relación que tengamos definida (con la misma interfaz que para la creación de relaciones) o “Eliminar” las que no queramos seguir utilizando.

Volviendo a nuestro modelo, hemos visto que las relaciones han sido detectadas perfectamente, ya que la relación de todas las tablas se hace a través del campo “Team”, que tiene valores únicos en todas y cada una de las tablas. Esta vez se lo hemos puesto fácil a Power BI. Por lo tanto, todas las relaciones serán “Uno a uno” y en ambas direcciones, quedando la relación entre las tablas “TeamGoals” y “TeamDefense” desactivada por estar implícitamente definida por el resto de relaciones establecidas:



## Creando nuevas variables

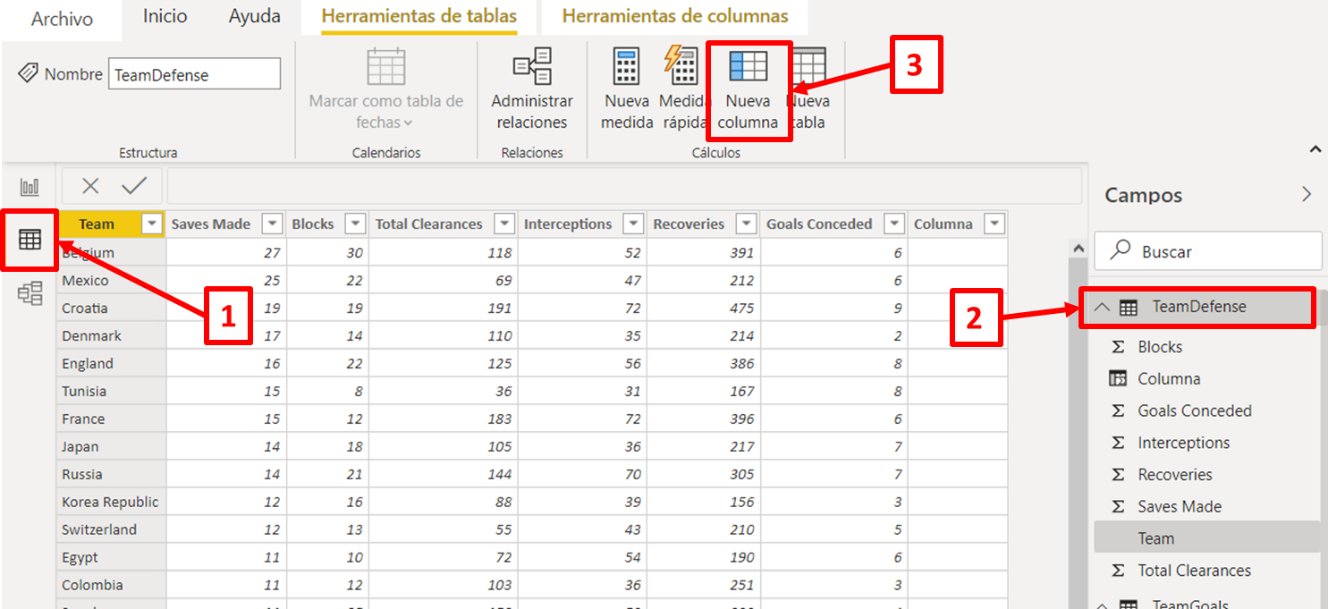
Otro de los aspectos importantes en todo proyecto de BI es la creación de nuevos datos a partir de los que ya tenemos. Normalmente, los datos se preprocesan con algún tipo de software, como podría ser mediante un script de Python, para dejarlo limpios y preparaditos para que la herramienta de BI se encargue principalmente de la visualización. Sin embargo, no siempre se da este caso, puede que se quiera realizar un estudio atacando directamente datos disponibles en algún entorno por diferentes motivos, como independencia de otros elementos o, por ejemplo, crearse medidas extra que no se habían tenido en cuenta en el procesamiento inicial.

Para aumentar la funcionalidad, Power BI nos permite crear:

* Columnas: podemos crear una nueva columna en una tabla con datos del resto de columnas. Un ejemplo muy típico consiste en crear porcentajes a partir de 2 columnas con valores absolutos, como el porcentaje de victorias a partir del total de partidos y los partidos ganados. Es importante destacar que la columna es eso, una columna, por lo que tendrá un valor para cada registro de la columna y funcionará como si de un campo más de la columna se tratase. Los valores de una columna no cambian.
* Medidas: las medidas son como KPIs. Van cambiando en función de los datos seleccionados, realizando la medida que se haya programado para que ejecuten. Por ejemplo, podríamos usar una medida para obtener la suma de todos los goles de los equipos. En el momento en el que filtremos para quedarnos con los equipos del “Grupo A”, por ejemplo, esta medida cambiará. Las medidas, a diferencia de las columnas, cambian dinámicamente y no tienen un valor por cada registro, solo tienen un valor, por lo que no se pueden hacer operaciones a nivel de registro en función de ellas.

En este caso, vamos a crearnos una columna que nos diga el porcentaje de goles que se conceden por intervención del guardameta, es decir, por parada. De este modo podremos valorar el desempeño de los porteros.

Para ello, vamos a ventana de datos, seleccionamos la tabla donde queramos añadir la columna y seleccionamos “Nueva columna”:



Una vez hecho esto, nos aparecerá una línea donde podremos escribir texto, donde la sintaxis es:

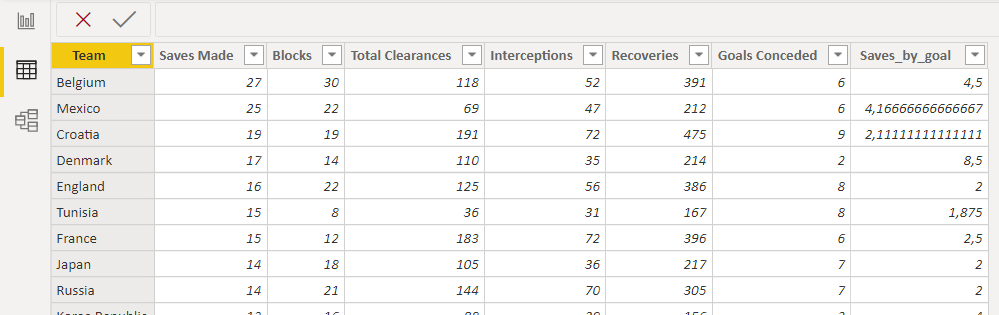


El lenguaje que tiene propio Power BI para crearnos tanto columnas como medidas es DAX, que nos permitirá realizar operaciones más complicadas. Sin embargo, para realizar las operaciones más simples que tenemos en mente en este taller, y que teóricamente son las únicas que se deberían hacer porque los datos ya tendrían que estar pretratados antes de leerlos aquí, no nos será necesario meternos a fondo con el lenguaje (si quieres profundizar, [aquí te dejo información propia de Microsoft](https://support.microsoft.com/es-es/office/expresiones-de-an%C3%A1lisis-de-datos-dax-en-powerpivot-bab3fbe3-2385-485a-980b-5f64d3b0f730#:~:text=DAX%20es%20un%20lenguaje%20de,tambi%C3%A9n%20conocidas%20como%20campos%20calculados).)).

Es este caso, para crearnos la columna que indique las paradas por gol concedido, escribiremos algo muy sencillo, donde accederemos a los campos de cada tabla como “Tabla[campo]”:



Dejando la tabla como:

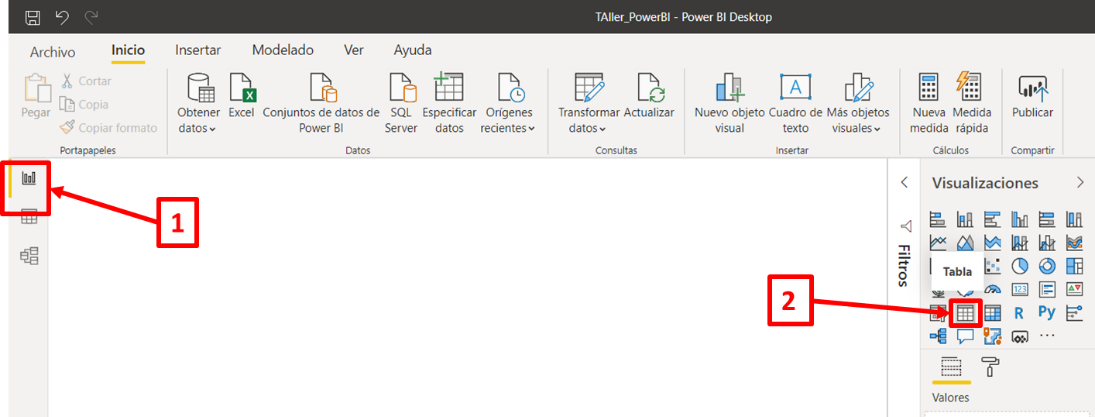


Pues con esto ya podríamos pasar a visualizar cosas 😊.

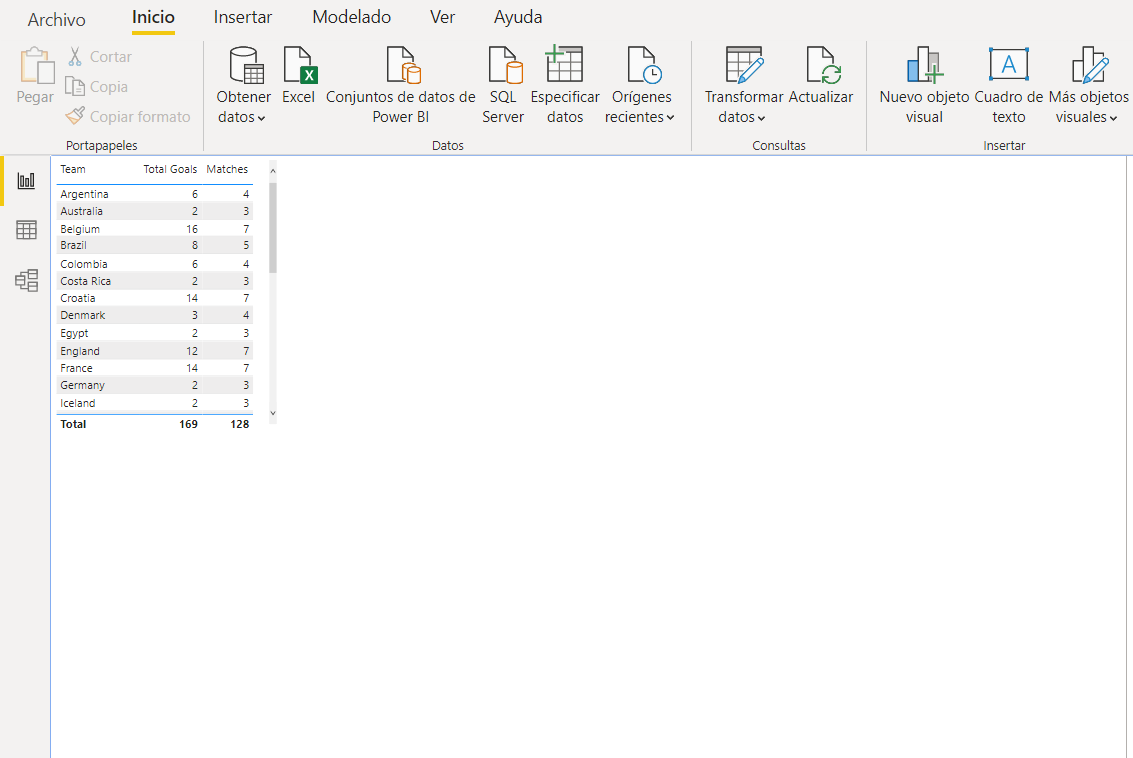
# Visualizando los datos en Power BI

Tras leer los datos y adaptarlos a nuestras necesidades, ya podemos hacer lo que veníamos buscando: visualizar los datos.

Nos dirigimos a la ventana principal del dashboard y empezamos a crear. Para añadir elementos visuales en nuestro informe, nos dirigiremos a la parte de “Visualizaciones”. Por ejemplo, podemos poner una tabla:



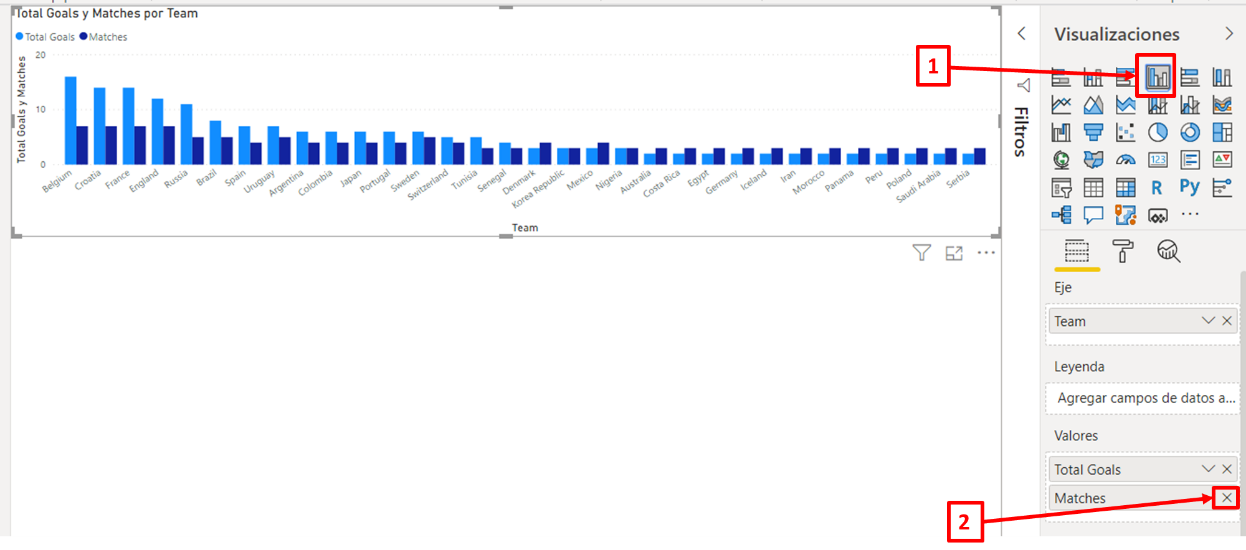
Y añadimos los campos que queramos ver arrastrándolos a la parte de “Valores”. En este caso, vamos a ver el equipo, los goles que han metido y los partidos jugados (campos “Team, “Total Goals” y “Matches”, de la tabla “TeamGoals”):



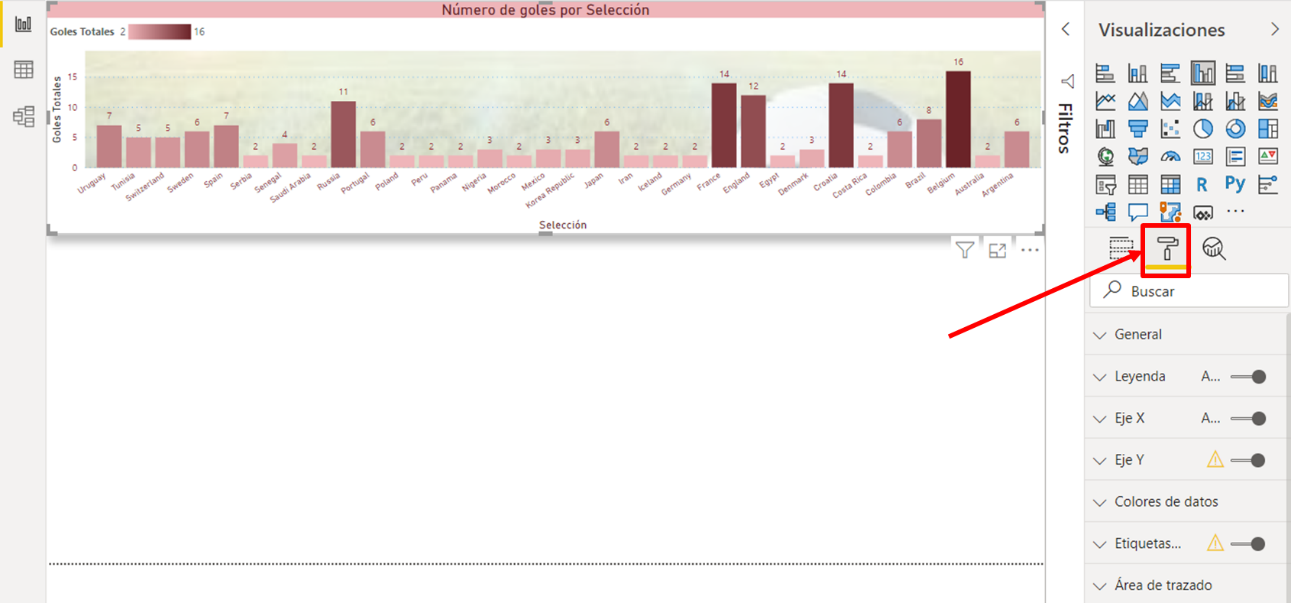
Bueno, ya empezamos a ver algo, pero nos gustaría ver más cosas, ¿verdad? Pues siéntete libre y deja que tu creatividad se alíe con tu capacidad de análisis para investigar y realizar un informe brillante.

En este caso, vamos a proponer un ejemplo con el que vamos a ver diferentes posibilidades.

Tras darle vueltas a la tabla de arriba, nuestro espíritu creativo nos dice que sería mejor otra visualización, como un gráfico de barras. Para ello, seleccionamos nuestra tabla y, en la parte de visualizaciones seleccionamos el icono correspondiente al gráfico de barras. Además, como la información del número de partidos no nos termina de convencer, decidimos quitarla:



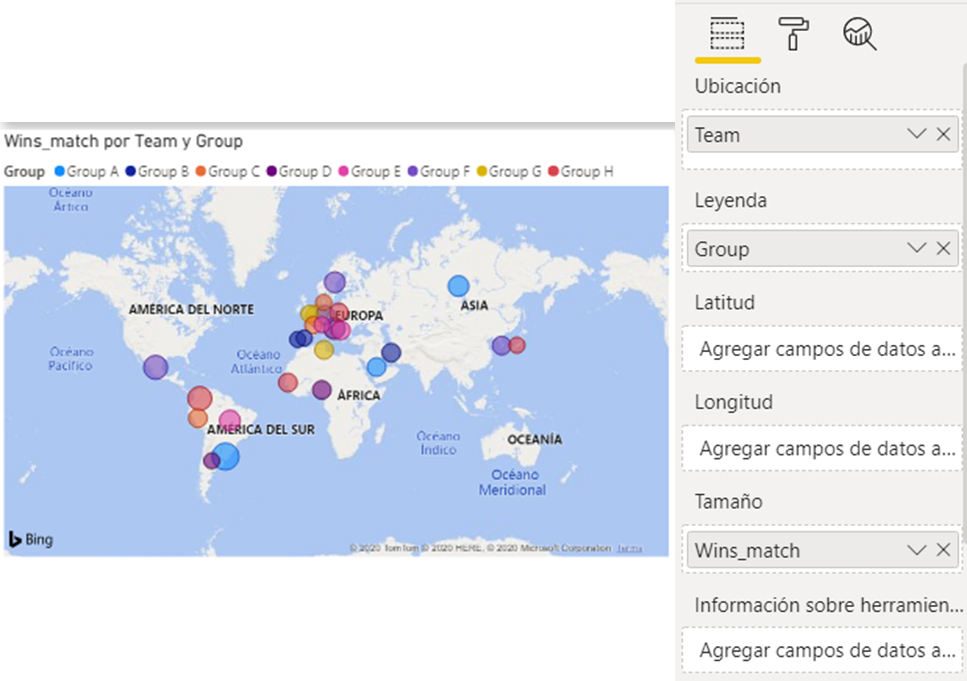
Si queremos cambiar algunas opciones propias del gráfico, lo podemos hacer en la segunda pestañita del marco de visualizaciones:



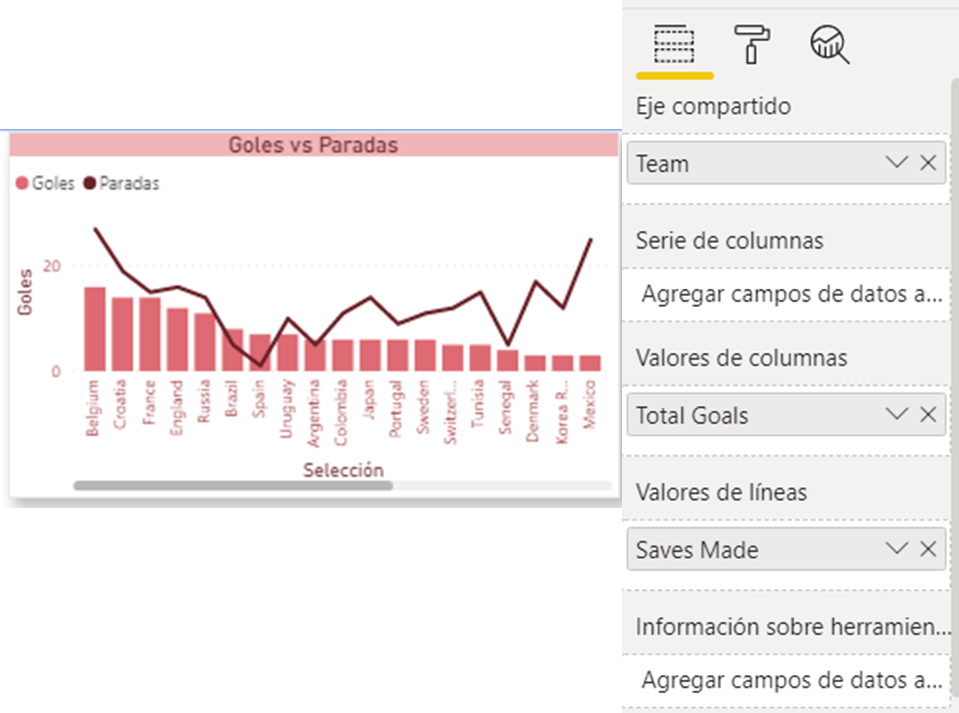
Nuestro informe va cogiendo forma, añadámosle algo realmente interesante que va a subir el nivel de nuestro informe: un mapa que nos diga cuántos partidos ha ganado cada selección.

Para ello, nos dirigimos a las visualizaciones, seleccionamos el mapa y arrastramos los campos que queramos mostrar a las subáreas correctas. En nuestro caso, nos interesará saber qué equipos (campo “Team” en “Ubicación”) han ganado cuántos partidos (campo “Wins” en “Tamaño”). Esto dibujará círculos en el mapa en las posiciones establecidas en “Team” con un tamaño basado en “Wins”. Además, le vamos a añadir otro elemento que nos aporte información de forma visual, que será saber a qué grupo pertenece cada uno. Esto podemos hacerlo creando una leyenda, que nos pintará de diferentes colores los círculos de cada grupo (arrastramos el campo “Group” a “Leyenda”).

Si realizamos lo anteriormente mencionado, nos quedará algo así:



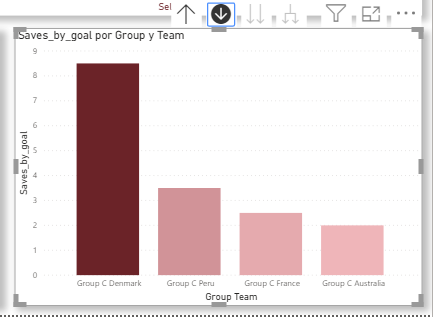
También podemos crear informes que nos relacionen campos de varias tablas. Por ejemplo, vamos a crear un gráfico de barras y líneas que nos compare los goles marcados (“TeamGoals[Total Goals]”) con las paradas realizadas (“TeamDefense[Saves Made]”).



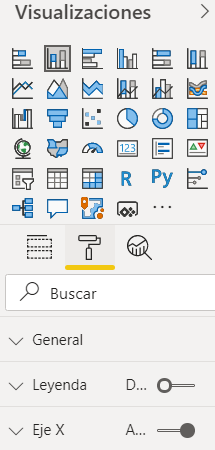
Por último, podríamos realizar gráficos con más de un eje, por ejemplo, en función del grupo al que pertenecen y del equipo. En este caso, podemos hacer algo combinando este aspecto y el anterior, de modo que podamos representar las paradas por gol concedido de cada equipo o en función del grupo:



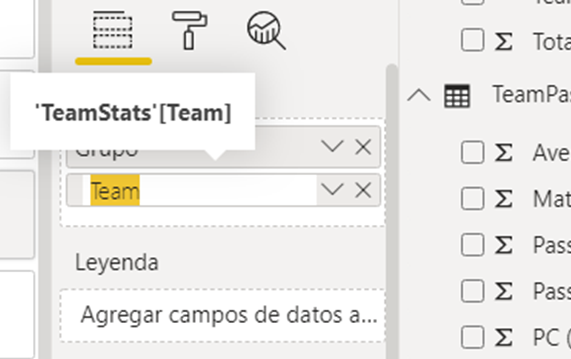
Para navegar en profundidad, tendremos que activar esa opción en el gráfico. Ahora, si seleccionamos un grupo, nos saldrá el desglose a nivel equipo:



Para finalizar, si queremos cambiar el nombre de los ejes, colores o demás particularidades de las visualizaciones, podemos acceder al submenú de “Formato”, como hemos explicado anteriormente:

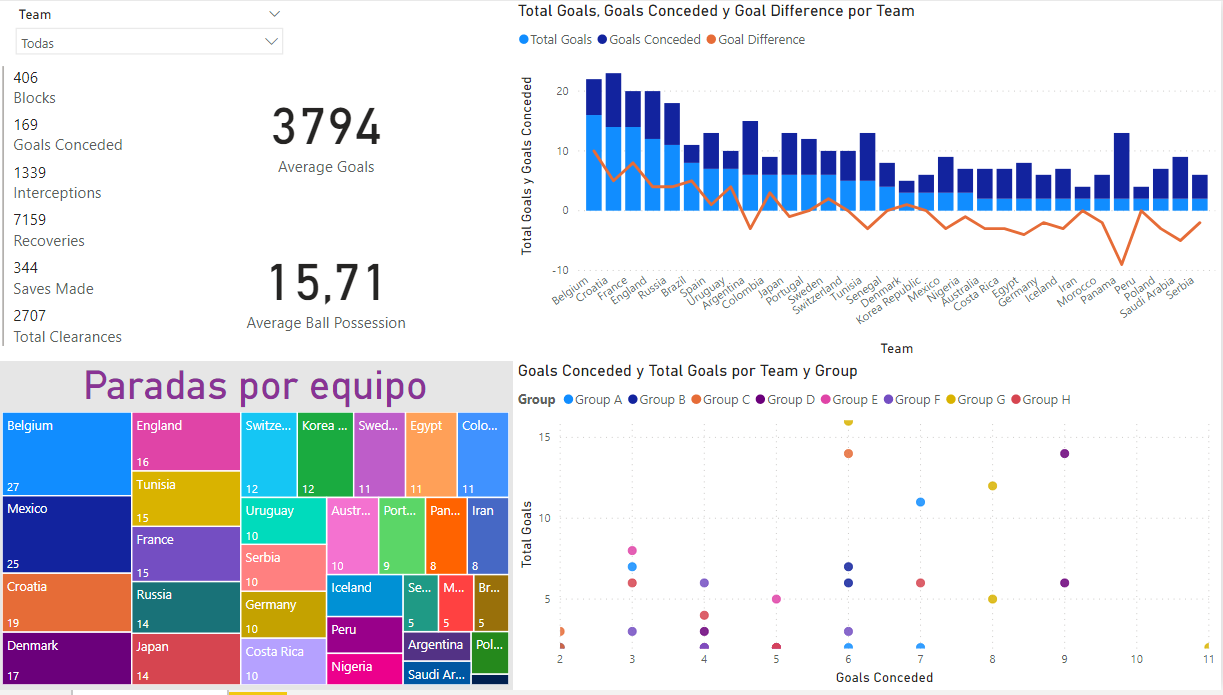


Y para cambiar los nombres de los campos en una visualización, lo podemos hacer con un doble clic sobre el campo que hemos arrastrado a la subárea de la visualización correspondiente:



Y hasta aquí el taller. Con esto ya podréis construir vuestros reportes. A partir de aquí, os animaría a consultar documentación más técnica para brillar con esos informes.

# Prueba a montar tus propias gráficas



1. Crea una hoja nueva (menu de abajo) y ponle el nombre de “Teams Data”

2. Pon un filtro (es una de las visualizaciones). El filtro tiene que ser un selector múltiple desplegable del equipo. En la propia visualización encontrarás una pestaña donde puedes modificar el tipo de filtro a “Menu desplegable”

3. Crea un KPI (Tarjeta), con la media de los goles

4. Crea otro KPI (Tarjeta) con la media de posesión de balón

5. Crea varios KPIs (Tarjeta varias filas) con todas las variables numéricas de la tabla de TeamDefense.

6. Monta un Treemap de las paradas por equipo.

Añade las etiquetas de las paradas

Modifica el título: centrado, color y tamaño

Ponle un fondo gris

7. Diagrama de barras. Crea un gráfico de columnas apiladas y de líneas. Añade Team en el Eje X. En los valores de columnas Total goals y Goals conceded, y en las líneas Goal Difference.

8. Scatter plot: Goals Conceded vs Total Goals, por Team, y la leyenda por Grupo.

9. Configura a tu gusto cómo se vería el dashboard en el móvil (Ver -> Diseño para móviles)

