

INTRODUCCIÓN

El *Análisis de Datos* se considera como el proceso mediante el cual se crea un modelo de datos, como resultado de diversas acciones de limpieza, transformación y depuración de los datos que permiten obtener o rescatar información útil con el objeto de aportar inferencias y servir de apoyo en la toma de decisiones.

El fenómeno *Big Data* en la generación de grandes volúmenes de datos de todo tipo (estructurados y no estructurados), ha propiciado que los *Analistas de Datos* hayan creado diversas herramientas y técnicas de análisis de datos desde las comunes o tradicionales a las sofisticadas y complejas. La tabla 1 describe las principales técnicas de análisis de datos tanto tradicionales como sofisticados.

| Técnicas tradicionales | |
|---|---|
| Técnica | Descripción |
| Media aritmética | Es la suma de una lista de números dividida por el número de elementos en esa lista y se usa para determinar la tendencia general de un conjunto de datos |
| Desviación estándar | Este cálculo es útil para determinar rápidamente la dispersión de puntos de datos. Una desviación estándar alta significa que los datos se distribuyen más ampliamente desde la media, mientras que una baja indica que hay más datos que se alinean con la media |
| Determinación del tamaño de la muestra | A veces no es necesario recopilar información de cada miembro de una población y basta con una muestra la cual debe ser representativa de la población de estudio |
| Regresión | Sirve para determinar tendencias a lo largo del tiempo, ya que modela las relaciones entre variables dependientes y explicativas, que generalmente se grafican en un diagrama de dispersión |
| Prueba de hipótesis | Esta técnica permite evaluar si una premisa es realmente cierta para su conjunto de datos o población. Elimina la posibilidad de que algo sea casual |

| Técnicas sofisticadas | |
|--|---|
| Técnica | Descripción |
| Aprendizaje automático | Esta subespecialidad de la informática se engloba en el campo de la inteligencia artificial y es conocida también como machine learning. Está relacionada con el diseño y desarrollo de algoritmos que permiten a las computadoras promover la acción en base a datos empíricos |
| Redes neuronales | Este tipo de técnicas de análisis de datos consiste en modelos computacionales, inspirados por la estructura y el funcionamiento de redes neuronales biológicas |
| Aprendizaje de reglas de asociación | Se trata de un conjunto de técnicas de análisis de datos empleadas para descubrir relaciones interesantes entre variables en grandes bases de datos |
| Algoritmos genéticos | Tipo de técnicas de análisis de datos inspirada en la naturaleza. En este caso, tiene una vertiente Darwiniana, al basarse en la evolución natural, la supervivencia del más fuerte. Al aplicar esta técnica, las soluciones potenciales se codifican para ser combinadas entre sí e incluso someterse a mutaciones |
| Análisis de series temporales | En este caso, se analizan secuencias de puntos de datos, que representan valores en sucesivos tiempos, para extraer las características más significativas de la información. Podría decirse que se trata del uso de un modelo para predecir valores futuros de una serie temporal en base a valores pasados conocidos de la misma u otras series |

Tabla 1. 10 Técnicas de análisis de datos. Fuente. Adaptado de (PowerData, 2017)¹.

El libro *de análisis de datos*, condensa y adapta distintas prácticas y ejemplos de la guía de documentación de *Power BI Desktop*, de la guía del lenguaje estadístico para análisis de datos con *R - RStudio* y de la guía de documentación de ejemplos y tutoriales de la aplicación *Knime*.

El capítulo 1 titulado *Prácticas con Power BI*, comienza con la descripción y generalidades de la aplicación Power BI Desktop, donde se enfatiza en: Usos de Power BI, conexión a datos, creación de un modelo de datos, creación de objetos visuales, creación de informes, compartir y publicar informes. Luego se describen las distintas prácticas o guías de trabajo de análisis de datos que incluyen: Análisis de datos financieros, Análisis de una página Web (Eurocopa), Combinar datos con Power BI, Creación de medidas propias y Análisis de datos de una supertienda.

¹ <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/10-tecnicas-de-analisis-de-datos-para-estadisticas-de-big-data>

El capítulo 2 nombrado como *Prácticas con R y RStudio*, empieza con las generalidades del lenguaje R, la descripción del entorno de desarrollo integrado IDE de RStudio con las características o generalidades y descripción de las ventanas, introducción al lenguaje R donde se destaca: Tipos de datos (simples, almacenamiento de valores, vectores, matrices, factors, Data frames y listas), carga de datos y a continuación se describen las distintas prácticas de análisis de datos como: Estadística descriptiva de una variable cuantitativa continua, Regresión lineal, Árboles de decisión y Minería de texto.

El capítulo 3 identificado como *Prácticas con Knime*, inicia con la descarga e instalación de Knime Analytics, seguido de la introducción a Knime Analytics donde se destaca: Elementos de la ventana de inicio de Knime, nodos y flujo de trabajo, ventajas y desventajas de Knime y crear un proyecto Knime. Se termina el capítulo con la descripción de las diferentes prácticas que involucran: Ciencia de datos, Modelo de entrenamiento de clasificación de datos y Modelo de predicción de supervivencia del titanic.

Finalmente, se presenta la solución de todos y cada uno de los ejercicios planteados en las distintas prácticas desarrolladas con cada una de las aplicaciones empleadas para el análisis de datos (Power BI, R - RStudio y Knime).

1

PRÁCTICAS CON POWER BI DESKTOP

1.1 GENERALIDADES DE POWER BI

Power BI en su versión **Desktop** es una aplicación de carácter gratuito, se puede instalar en el equipo local y le permite conectarse a los datos, transformarlos y visualizarlos. Con la aplicación de **Power BI Desktop**, puede conectarse a varios orígenes de datos distintos y combinarlos (operación que se conoce como modelado) en un modelo de datos que le facilita crear objetos visuales y colecciones como el que se ilustra en la figura 1, que puede compartir como informes con otras personas dentro de la empresa u organización. La mayoría de los usuarios que trabajan con proyectos de inteligencia empresarial, usan **Power BI Desktop** para crear informes y luego utilizan el **servicio Power BI** para comunicar sus informes con los demás. (Microsoft Power BI, 2021)².

2 <https://docs.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop>

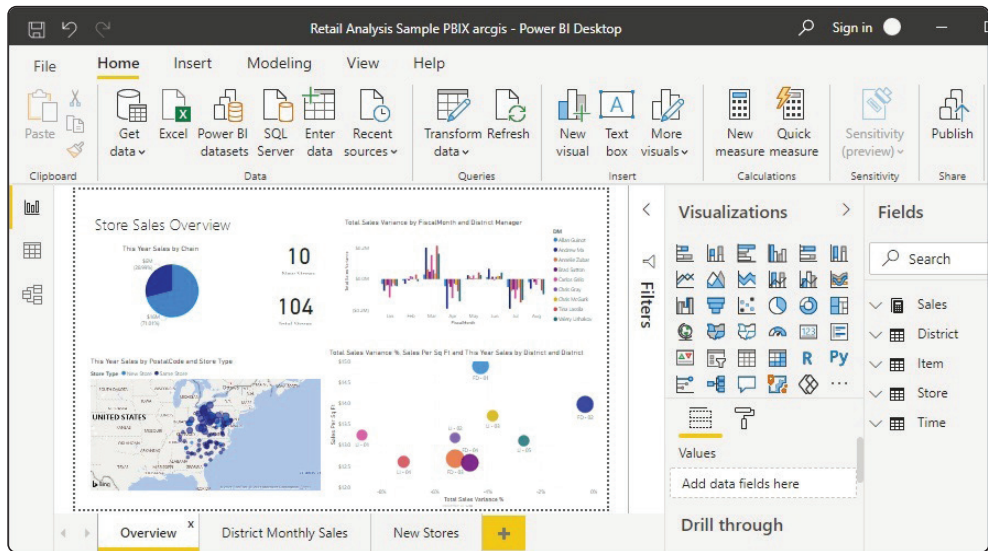


Figura 1. Ejemplo de objetos visuales con Power BI. Fuente. (Microsoft Power BI, 2021)³

1.1.1 Usos de Power BI

La aplicación de *Power BI Desktop*, puede ser empleada para diferentes usos como: modelado, crear objetos visuales, crear informes o compartir informes, tal como se describe en la tabla 2.

| Uso | Descripción |
|-------------------------------|--|
| Conexión | Conexión a varios orígenes de datos |
| Modelado | Realizar la transformación y limpieza de los datos para crear un modelo de datos |
| Crear objetos visuales | Crear objetos como gráficos o grafos, que proporcionan representaciones visuales de los datos |
| Crear informes | Crear colecciones de objetos visuales, en una o varias páginas de informes |
| Compartir | Comunicar los informes con otros usuarios dentro de la organización mediante el servicio de Power BI |

Tabla 2. Usos de Power BI Desktop. Fuente. Adaptado de (Microsoft Power BI, 2021)⁴.

3 <https://docs.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop>

4 <https://docs.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop>

Power BI Desktop, le proporciona herramientas para crear informes completos y visualmente eficientes, al facilitar combinar datos de varios orígenes, en un informe que puede comunicar o compartir con otras personas de la corporación, todo ello depende de la experticia y habilidad de quien elabora el informe.

1.1.2 Conexión a datos

El primer paso para trabajar con **Power BI Desktop** consiste en realizar la conexión a los datos objeto de análisis. Para ello, se da clic en la cinta **Inicio** y a continuación en la opción **Obtener datos** del grupo **Datos externos**, seguidamente en la opción **Más**, con lo cual se despliega la ventana emergente como se ilustra en la figura 2.

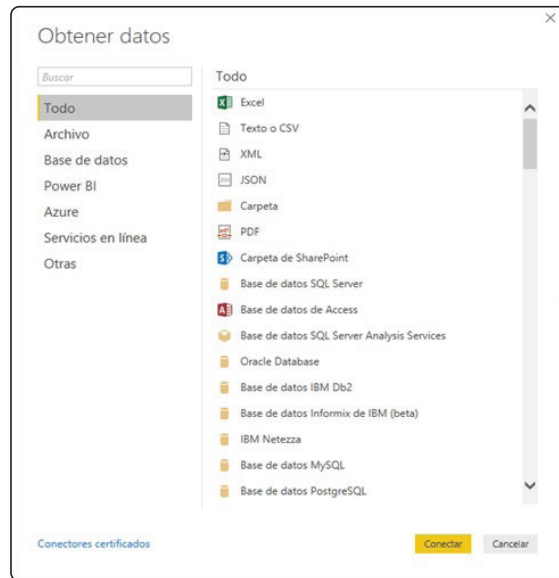


Figura 2. Ventana obtener datos. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

En la ventana emergente que se despliega, se visualizan diferentes categorías de datos que permiten el acceso a distintos orígenes como:

- Todo (Lista de todos los orígenes)
- Archivos (Excel, Texto o CSV, etc.)
- Bases de datos (SQL, Access, etc.)
- Power BI (Conjuntos o flujos de datos)
- Azure (SQL, Tablas, Bases de datos)
- Servicios en línea (SharePoint, CDS)
- Otras (Web, HDFS, ODBC, OLE, etc.)

1.1.3 Creación de un modelo de datos

Una vez realizada la conexión a los orígenes de datos, es conveniente efectuar la *limpieza* o *depuración de los datos* y las *transformaciones* que se requieran para crear el *modelo de análisis*. El *editor de consultas* que provee *Power BI* permite ejecutar las modificaciones a los datos como cambiar tipos de datos, quitar columnas o combinar datos de varios orígenes hasta obtener el modelo de datos deseado. Cada una de las acciones realizadas queda registrada en el editor de consultas, tal como se ilustra en la figura 3.

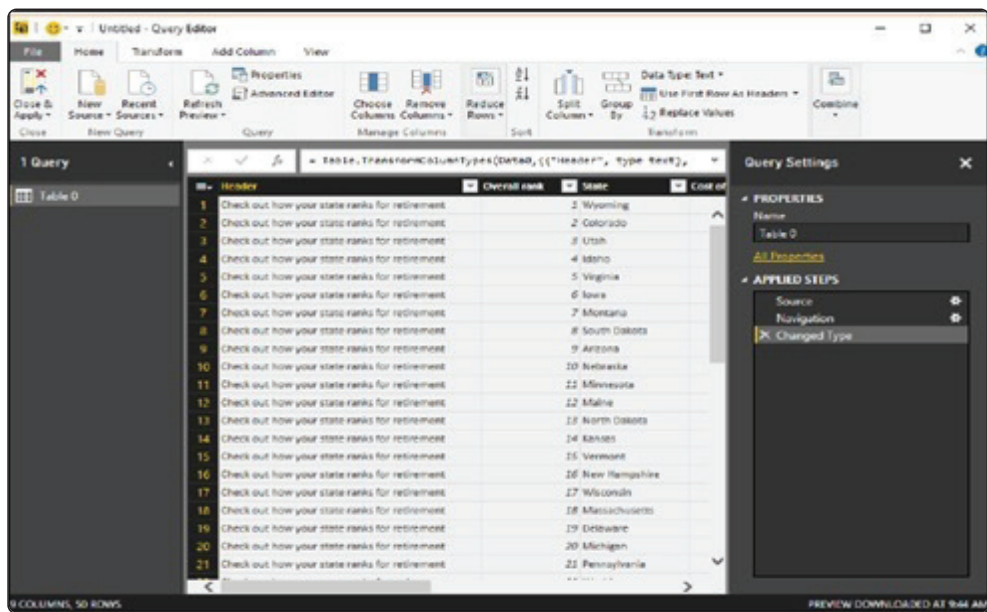


Figura 3. Limpieza y transformación de los datos.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

1.1.4 Creación de objetos visuales

Creado el *modelo de datos* como resultado de la limpieza y transformación de los datos, puede arrastrar *campos* de la *tabla de datos* al *lienzo del informe* para crear *objetos visuales*, entendiendo que un objeto visual es la representación gráfica de los datos del modelo. La figura 4 ilustra los diversos *objetos visuales* que provee *Power BI*.

- En la primera fila de izquierda a derecha se presentan los objetos: Gráficos de barras apiladas, gráfico de columnas apiladas, gráfico de barras agrupadas, gráfico de columnas agrupadas, gráfico de barras 100% apiladas y gráfico de columnas 100% apiladas.
- En la segunda fila: gráfico de líneas, gráficos de áreas, gráfico de áreas apiladas, gráfico de columnas apiladas y de líneas, gráfico de columnas agrupadas y de líneas y gráfico de la barra de herramientas.
- En la tercera fila: gráfico de cascada, gráfico de embudo, gráfico de dispersión, gráfico circular, gráfico de anillos y treemap.
- En la cuarta fila: Mapa, Mapa coroplético, medidor, tarjeta, tarjeta de varias filas y KPI.
- En la quinta fila: Segmentación de datos, tabla, matriz, objeto visual de script de R, objeto visual de Python y elementos influyentes clave.
- En la sexta fila: Esquema jerárquico, preguntas y respuestas, narración inteligente, ArGIS Maps for Power BI, PowerApps for Power BI y obtener más objetos visuales.

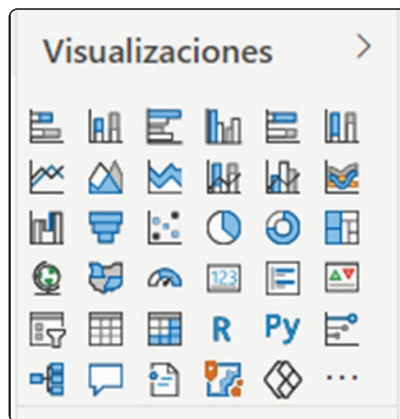


Figura 4. Visualizaciones de Power BI.

Fuente. Obtenido mediante (Microsoft Power BI, 2021).

1.1.5 Creación de informes

En *Power BI Desktop*, la colección de objetos visuales que muestran diversos aspectos de los datos utilizados para crear el modelo de análisis recibe el nombre de

informe. Al igual que un archivo de Excel contiene varias hojas, igual acontece con el informe en Power BI, el cual puede contener varias páginas tal como se ilustra en la figura 5, en la que se vislumbran tres páginas nombradas **Overview**, **District Monthly Sales** y **New Stores**.

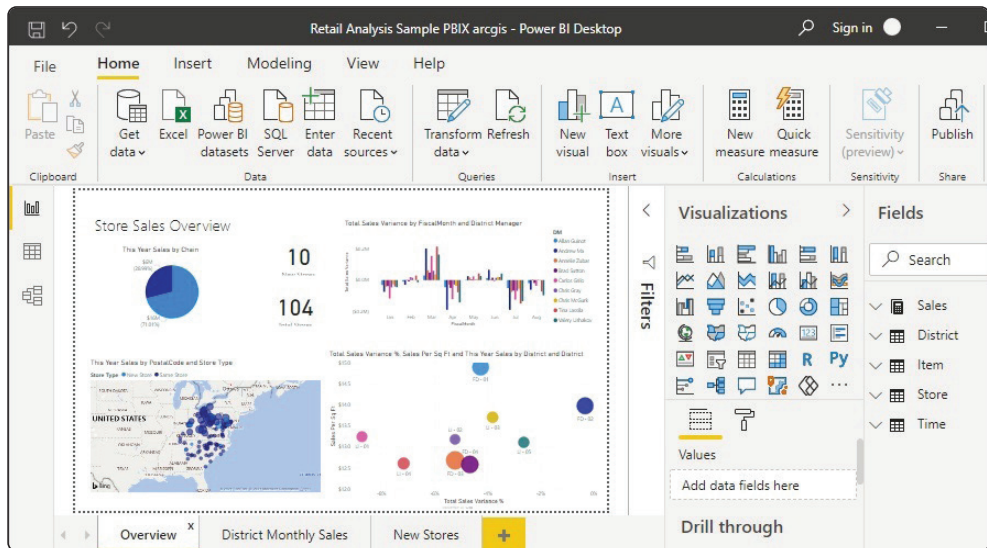


Figura 5. Ejemplo de un informe en Power BI Desktop. Fuente. (Microsoft Power BI, 2021)⁵.

1.1.6 Compartir y publicar informes

Creado y depurado el *informe del modelo de análisis de datos*, puede publicarlo con el *servicio de Power BI* con el objeto de compartirlo y que esté disponible para cualquier miembro de la organización que tenga una licencia de Power BI.

Para publicar el informe debe dar clic en la cinta **Inicio** y luego en el botón **Publish (Publicar)** tal como se ilustra en la figura 6, **Power BI Desktop** se conectará al **servicio Power BI** con su cuenta de Power BI y, a continuación, le solicitará que seleccione en qué lugar del servicio Power BI desea compartir el informe, puede por ejemplo, indicar el *área de trabajo propia*, *un área de trabajo de equipo* o *alguna otra ubicación dentro del servicio Power BI*.

5 <https://docs.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop>

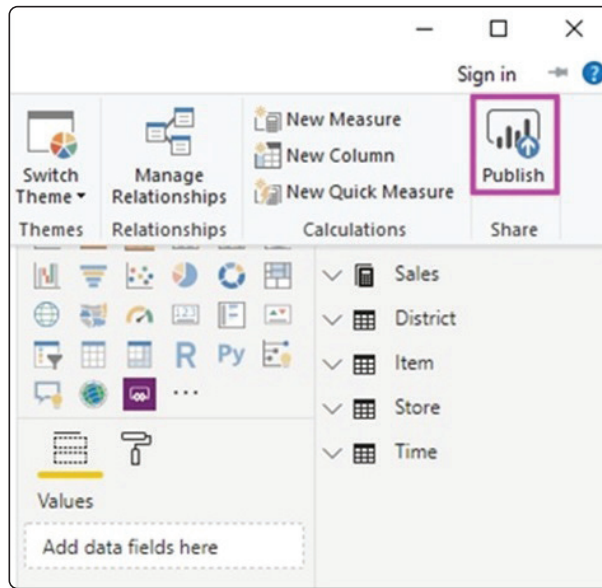


Figura 6. Publicar un informe en Power BI. Fuente. (Microsoft Power BI, 2021)⁶.

1.2 PRÁCTICA ANÁLISIS DE DATOS FINANCIEROS

El objetivo de la práctica radica en analizar un conjunto de datos financieros mediante Power BI Desktop, el primer paso consiste en **cargar los datos** correspondiente a un archivo de Excel que proporciona Power BI, a continuación se efectúa el proceso de **transformación para crear el modelo** y luego se procede a generar el **informe** para dar respuesta a los diferentes interrogantes planteados en el **análisis de los datos**.

1.2.1 Carga de datos

El archivo de trabajo se puede descargar de la dirección <https://go.microsoft.com/fwlink/?LinkID=521962> (Microsoft Power BI, 2021)⁷. Una vez descargado el archivo guárdelo en la ubicación correspondiente. Aplique los siguientes pasos para obtener los datos de análisis.

6 <https://docs.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop>

7 <https://docs.microsoft.com/es-es/power-bi/create-reports/desktop-excel-stunning-report>

- Abra la aplicación de **Power BI Desktop**, de doble clic sobre la pestaña **página1** del informe y asigne el nombre **Análisis_Financiero**, configure las propiedades correspondientes al hacer clic en el botón **rodillo** del **panel central** (defina el **tamaño**: relación 16:9, especifique **color de fondo** con transparencia al 80%) tal como se ilustra en la figura 7.

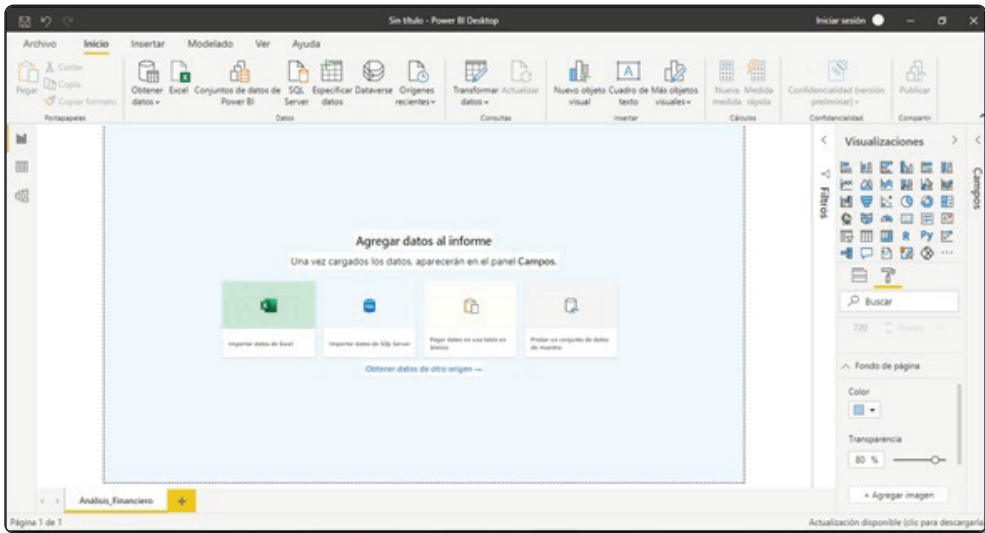


Figura 7. Configuración página del informe **Análisis_Financiero**.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

- Guarde la práctica con el nombre **Ana_Financiero** y a continuación por la pestaña **Inicio**, de clic en el botón **Obtener datos** y seleccione la opción **Excel** tal como se ilustra en la figura 8.
- Ubique la dirección donde guardó el archivo **FinancialSample** y ábralo.
- En la ventana emergente seleccione la tabla **financials** y luego de clic en el botón **Transformar Datos** como se indica en la figura 9.

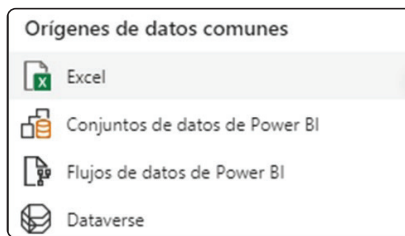


Figura 8. Conexión a un origen de datos de Excel.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

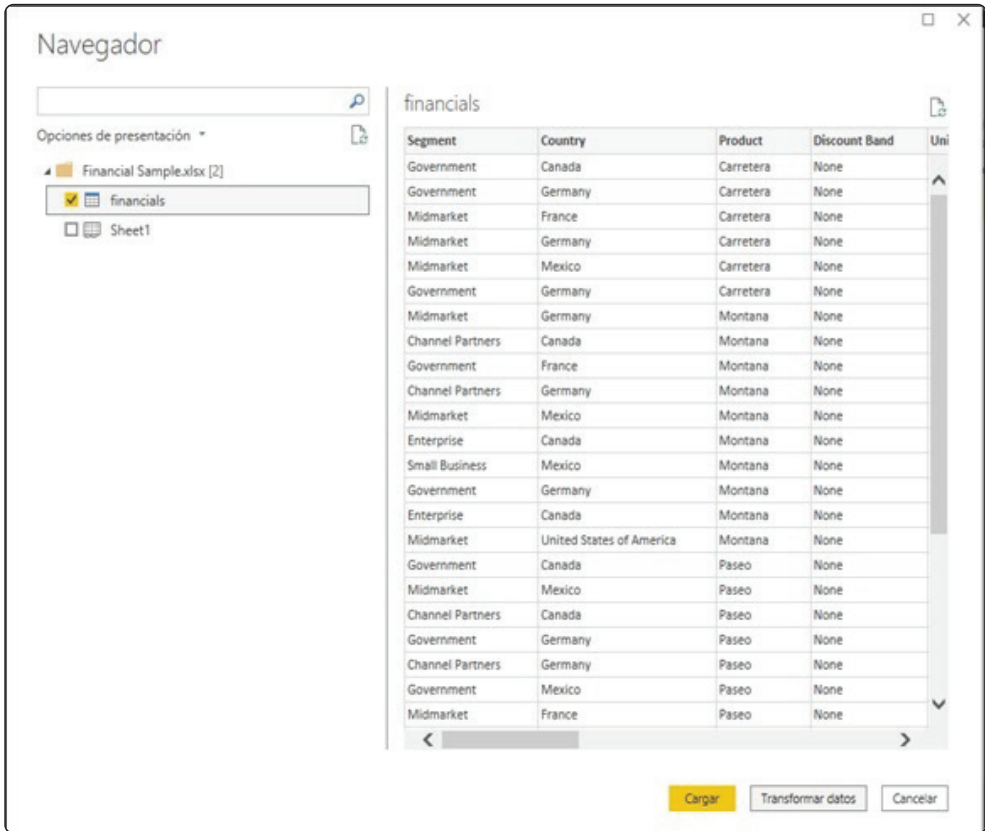


Figura 9. Apertura del archivo Financial Sample.
 Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

El archivo **Financial Sample** se compone de **700 observaciones** y **16 variables** que corresponden a: *Segment (Segmento)*, *Country (País)*, *Product (Producto)*, *Discount Band (Banda de descuento)*, *Units Sold (Unidades vendidas)*, *Manufacturing Price (Precio de manufactura)*, *Sale Price (Precio de venta)*, *Gross Sales (Ventas brutas)*, *Discounts (Descuentos)*, *Sales (Ventas)*, *COGS, Profit (Beneficio)*, *Date (Fecha)*, *Month Number (Número de Mes)*, *Month Name (Nombre del mes)*, *Year (Año)*.