

Análisis de Datos con Power BI, R-RStudio y Knime

Curso Práctico



WWW



Desde www.ra-ma.es podrá
descargar material adicional.

**JOFEBEUS
IRYOPOGU**

Análisis de Datos con Power BI, R-RStudio y Knime

Curso práctico

*JOFEBEUS
IRYOPOGU*



Ra-Ma[®]

edü[®]

Conocimiento a su alcance
BOGOTÁ - MÉXICO, D.F.

Betancourt Uscátegui, Jorge Fernando y Polanco Guzmán, Irma Yolanda

Análisis de Datos con Power BI, R-Studio y Knime / Jorge Fernando Betancourt Uscátegui e Irma Yolanda Polanco Guzmán --. Bogotá: Ediciones de la U, 2022

294 p. ; 24 cm

ISBN 978-958-792-406-0 e-ISBN 978-958-792-407-7

1. Informática 2. Programación 3. Prácticas con Power BI 4. Prácticas con R-Studio
5. Prácticas con Knime I. Tít.
621.39 ed.

*Edición original publicada por © Editorial Ra-ma (España)
Edición autorizada a Ediciones de la U para Colombia*

Área: Informática

Primera edición: Bogotá, Colombia, septiembre de 2022

ISBN. 978-958-792-406-0

- © Jorge Fernando Betancourt Uscátegui e Irma Yolanda Polanco Guzmán
- © Ra-ma Editorial. Calle Jarama, 3-A (Polígono Industrial Igarsa) 28860 Paracuellos de Jarama
www.ra-ma.es y www.ra-ma.com / E-mail: editorial @ra-ma.com
Madrid, España
- © Ediciones de la U - Carrera 27 #27-43 - Tel. (+57-1) 3203510 -3203499
www.edicionesdelau.com - E-mail: editor@edicionesdelau.com
Bogotá, Colombia

Ediciones de la U es una empresa editorial que, con una visión moderna y estratégica de las tecnologías, desarrolla, promueve, distribuye y comercializa contenidos, herramientas de formación, libros técnicos y profesionales, e-books, e-learning o aprendizaje en línea, realizados por autores con amplia experiencia en las diferentes áreas profesionales e investigativas, para brindar a nuestros usuarios soluciones útiles y prácticas que contribuyan al dominio de sus campos de trabajo y a su mejor desempeño en un mundo global, cambiante y cada vez más competitivo.

Coordinación editorial: Adriana Gutiérrez M.

Carátula: Ediciones de la U

Impresión: DGP Editores SAS

Calle 63 #70D-34, Pbx (57+1) 3203510

Impreso y hecho en Colombia

Printed and made in Colombia

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro y otros medios, sin el permiso previo y por escrito de los titulares del Copyright.

*A nuestras queridas familias en cabeza
de nuestros padres por sus enseñanzas,
por los valores impartidos y por su
constante apoyo para la superación
tanto personal como profesional*

*A todos los interesados en acrecentar
su conocimiento en el campo del
análisis de datos con herramientas
como Power BI, R y Knime*

JOFEBEUS - IRYOPOGU

ÍNDICE

INTRODUCCIÓN	11
CAPÍTULO 1. PRÁCTICAS CON POWER BI DESKTOP	15
1.1 GENERALIDADES DE POWER BI.....	15
1.1.1 Usos de Power BI.....	16
1.1.2 Conexión a datos	17
1.1.3 Creación de un modelo de datos.....	18
1.1.4 Creación de objetos visuales	18
1.1.5 Creación de informes.....	19
1.1.6 Compartir y publicar informes	20
1.2 PRÁCTICA ANÁLISIS DE DATOS FINANCIEROS	21
1.2.1 Carga de datos	21
1.2.2 Crear el modelo de datos	24
1.2.3 Crear el informe.....	29
1.2.4 Práctica análisis de datos por año.....	34
1.3 PRÁCTICA ANÁLISIS DE UNA PÁGINA WEB	36
1.3.1 Conexión a un origen de datos	36
1.3.2 Limpieza de datos mediante el editor de Power Query.....	39
1.3.3 Importación de la consulta en la vista de informe.....	41
1.3.4 Creación de una visualización - Dashboard	43
1.4 PRÁCTICA COMBINAR DATOS CON POWER BI	49
1.4.1 Conectarse a un origen de datos	49
1.4.2 Crear el modelo de datos	51
1.4.3 Combinar datos.....	57
1.4.4 Elaboración del Dashboard.....	62
1.5 PRÁCTICA CREACIÓN DE MEDIDAS PROPIAS (EMPRESA CONTOSO).....	67
1.5.1 Lectura y carga de archivo	67
1.5.2 Elaboración del Dashboard.....	69
1.5.3 Creación de medidas propias.....	79

1.6	PRÁCTICA ANÁLISIS DE DATOS DE UNA SUPERTIENDA	88
1.6.1	Conexión y carga del archivo de datos.....	88
1.6.2	Creación del Dashboard General.....	90
1.6.3	Creación del Dashboard Caribe.....	95
1.6.4	Creación del Dashboard Centro.....	102
1.6.5	Creación del Dashboard Norte	103
1.6.6	Creación del Dashboard Sur.....	104
1.6.7	Análisis de datos de la muestra_supertienda.....	108
CAPÍTULO 2. PRÁCTICAS CON R-RSTUDIO.....		115
2.1	GENERALIDADES DEL LENGUAJE R	115
2.2	ENTORNO DE DESARROLLO INTEGRADO (IDE) RSTUDIO.....	116
2.2.1	Características o generalidades de RStudio.....	116
2.2.2	Ventanas del entorno IDE de RStudio	117
2.3	INTRODUCCIÓN AL LENGUAJE R	118
2.3.1	Tipos de datos en R	118
2.3.2	Carga de datos	128
2.4	PRÁCTICA: ESTADÍSTICA DESCRIPTIVA DE UNA VARIABLE CUANTITATIVA CONTINUA	129
2.4.1	Origen de los datos	130
2.4.2	Medidas de tendencia central	131
2.4.3	Tabla de frecuencia e histograma	132
2.4.4	Medidas de variabilidad	132
2.4.5	Medidas de posición	134
2.4.6	Normalidad de los datos	135
2.4.7	Estadística descriptiva de la variable dist.....	137
2.4.8	Estadística descriptiva de la variable accel	138
2.5	PRÁCTICA REGRESIÓN LINEAL.....	138
2.5.1	Correlación Temperatura vs Nivel de Ozono	140
2.5.2	Correlación Nivel de Ozono vs Radiación Solar.....	141
2.5.3	Correlación Temperatura vs Nivel de Radiación Solar	142
2.5.4	Correlación Temperatura vs Velocidad de Viento	143
2.5.5	Correlación entre múltiples variables.....	144
2.5.6	Correlación Nivel de Ozono vs Velocidad del Viento.....	146
2.5.7	Correlación Nivel de Radiación Solar vs Velocidad del Viento.....	147
2.6	PRÁCTICA ÁRBOLES DE DECISIÓN.....	147
2.6.1	Característica de los árboles de decisión	148
2.6.2	Requerimientos.....	148
2.6.3	Importar los datos	148
2.6.4	Generar un set de entrenamiento y prueba	152
2.6.5	Elección del modelo	158
2.6.6	Sistematizando el modelo.....	158
2.6.7	Conclusión.....	160
2.6.8	Ejercicio: Creación y análisis de un árbol de decisión	160

2.7	PRÁCTICA MINERÍA DE TEXTO.....	161
2.7.1	Instalación de los paquetes requeridos	161
2.7.2	Carga de datos	162
2.7.3	Ejercicio: Análisis del texto: Aplicaciones de la inteligencia artificial... ..	167
CAPÍTULO 3. PRÁCTICAS CON KNIME		169
3.1	DESCARGAR E INSTALAR KNIME ANALYTICS	169
3.1.1	Instalar Knime Analytics	169
3.1.2	Actualizar datos	177
3.2	INTRODUCCIÓN A KNIME ANALYTICS	178
3.2.1	Elementos de la ventana de inicio de Knime.....	179
3.2.2	Nodos y flujo de trabajo	180
3.2.3	Ventajas y desventajas de Knime	182
3.2.4	Crear un proyecto Knime	182
3.3	PRÁCTICA CIENCIA DE DATOS.....	190
3.3.1	Concepto y fases.....	190
3.3.2	Crear un flujo de trabajo Workflow	192
3.3.3	Síntesis del análisis de datos del sistema CRM.....	213
3.4	PRÁCTICA MODELO DE ENTRENAMIENTO DE CLASIFICACIÓN DE DATOS	214
3.4.1	Lectura de datos.....	214
3.4.2	Tratamiento y limpieza de los datos	216
3.4.3	Propiedades gráficas	218
3.4.4	Estadísticas descriptivas	219
3.4.5	Partición de datos	223
3.4.6	Entrenamiento del modelo de decisión.....	225
3.4.7	Tabla interactiva	228
3.4.8	Aplicar el modelo	229
3.4.9	Gráfico número de horas vs edad	230
3.4.10	Puntuar.....	232
3.4.11	Flujo de trabajo.....	234
3.4.12	Práctica de resultados del modelo	235
3.5	PRÁCTICA MODELO DE PREDICCIÓN DE SUPERVIVENCIA DEL TITANIC.....	238
3.5.1	Lectura de datos.....	238
3.5.2	Exploración y tratamiento de datos	240
3.5.3	Propiedades gráficas (Titanic)	245
3.5.4	Estadísticas descriptivas (Titanic)	246
3.5.5	Partición de datos (Titanic).....	250
3.5.6	Entrenamiento del modelo de decisión.....	252
3.5.7	Aplicar el modelo	254
3.5.8	Puntuar.....	256
3.5.9	Flujo de trabajo.....	257
3.5.10	Práctica de resultados: modelo de predicción supervivencia del titanic.....	258

SOLUCIÓN A LAS PRÁCTICAS Y EJERCICIOS PROPUESTOS	263
CAPÍTULO 1. PRÁCTICAS CON POWER BI DESKTOP.....	263
1.2 Práctica: Análisis de datos financieros.....	263
1.3 Práctica: Análisis de una página Web	264
1.4 Práctica: Combinar datos con Power BI.....	265
1.5 Práctica: Creación de medidas propias (Empresa Contoso)	265
1.6 Práctica: Análisis de datos de una supertienda	267
CAPÍTULO 2. PRÁCTICAS CON R - RSTUDIO	272
2.4 Práctica: Estadística descriptiva de una variable cuantitativa continua.....	272
2.5 Práctica: Regresión lineal	274
2.6 Práctica: Árboles de decisión.....	275
2.7 Práctica: Minería de texto	276
CAPÍTULO 3. PRÁCTICAS CON KNIME	281
3.3.3 Síntesis del análisis de datos del sistema CRM	281
3.4 Modelo de entrenamiento de clasificación de datos	282
3.5. Práctica de resultados modelo de predicción de supervivencia del titanic .	285
REFERENCIAS.....	289
MATERIAL ADICIONAL.....	293

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>

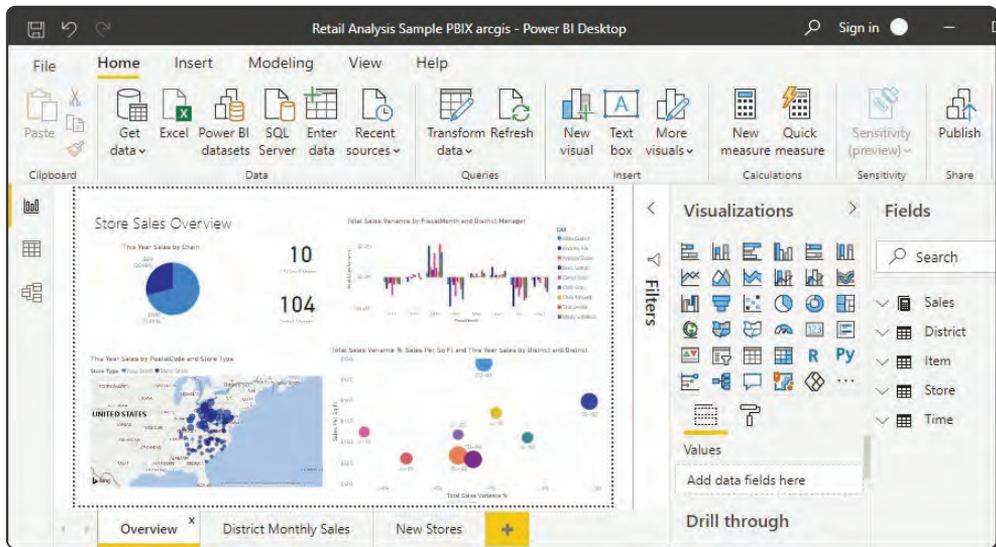


Figura 1. Ejemplo de objetos visuales con Power BI. Fuente. (Microsoft Power BI, 2021)³

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.

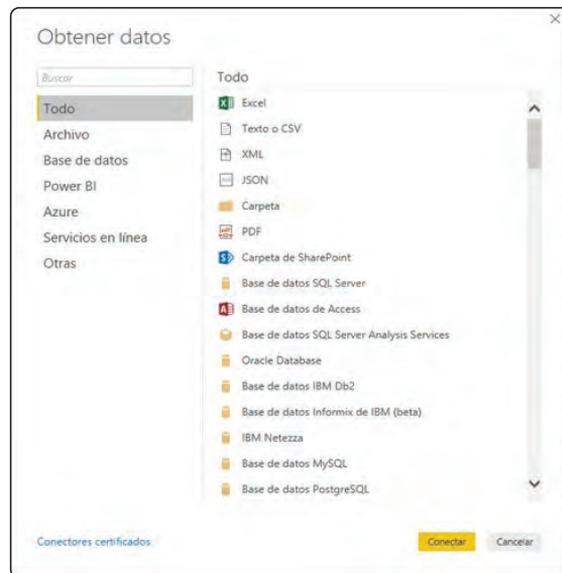


Figura 2. Ventana obtener datos. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

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.

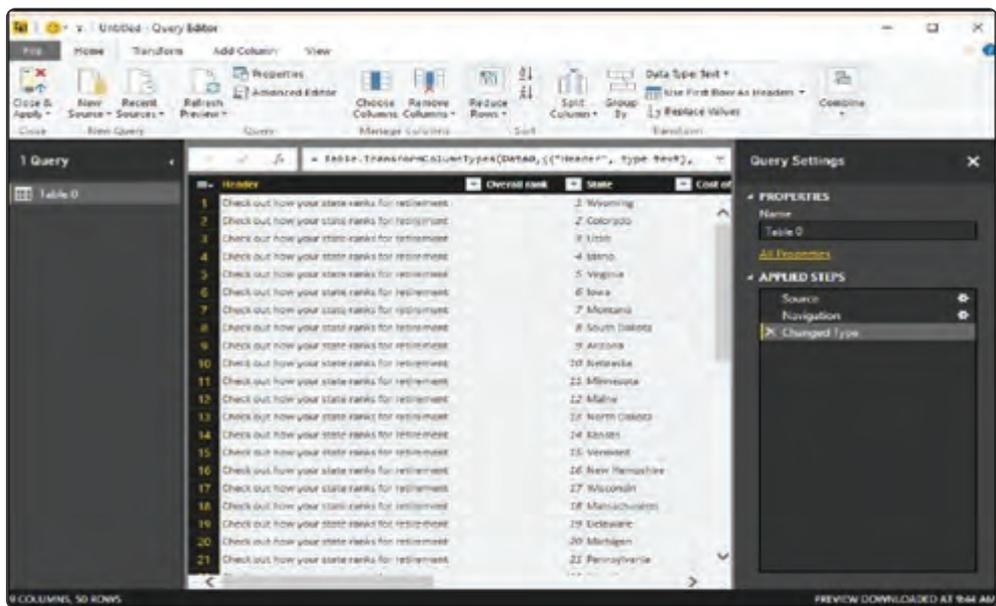


Figura 3. Limpieza y transformación de los datos.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

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.

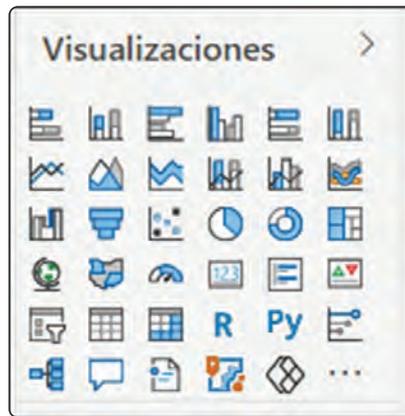


Figura 4. Visualizaciones de Power BI.

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*.

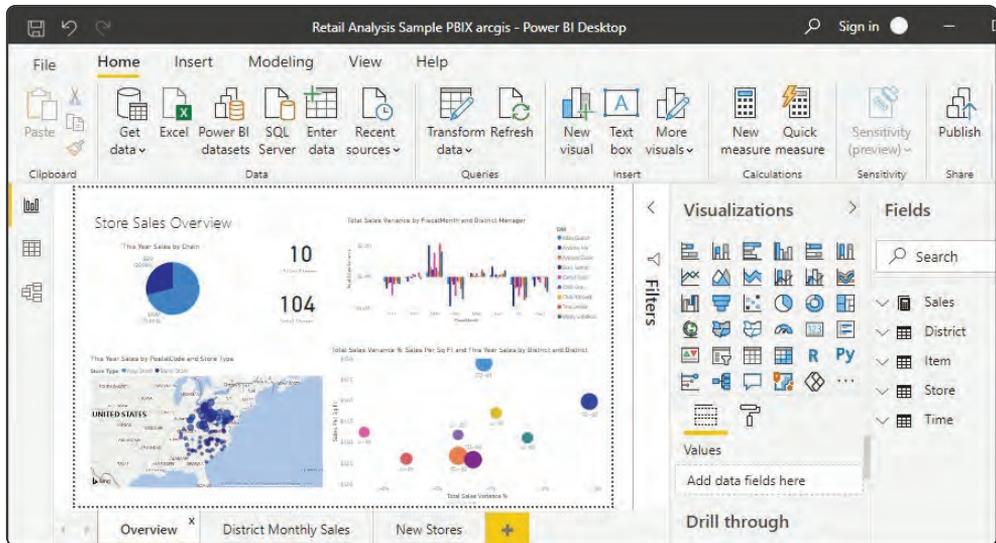


Figura 5. Ejemplo de un informe en Power BI Desktop. Fuente. (Microsoft Power BI, 2021)⁵.

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>

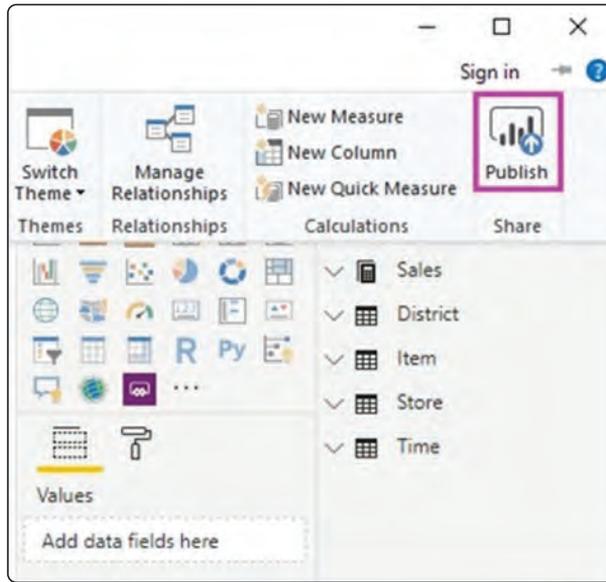


Figura 6. Publicar un informe en Power BI. Fuente. (Microsoft Power BI, 2021)⁶.

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.

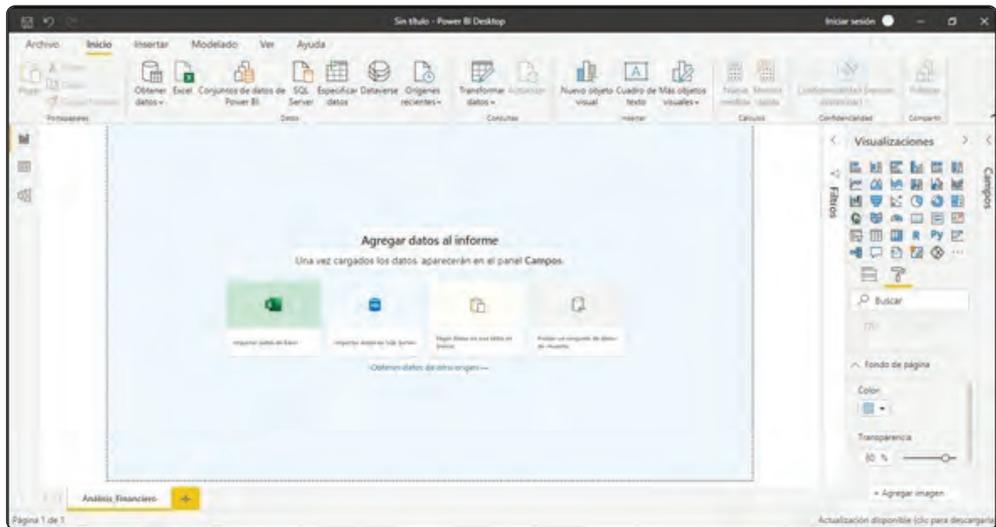


Figura 7. Configuración página del informe **Análisis_Financiero**.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

- 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.

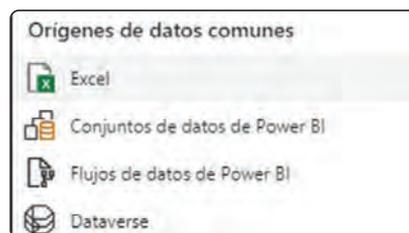


Figura 8. Conexión a un origen de datos de Excel.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

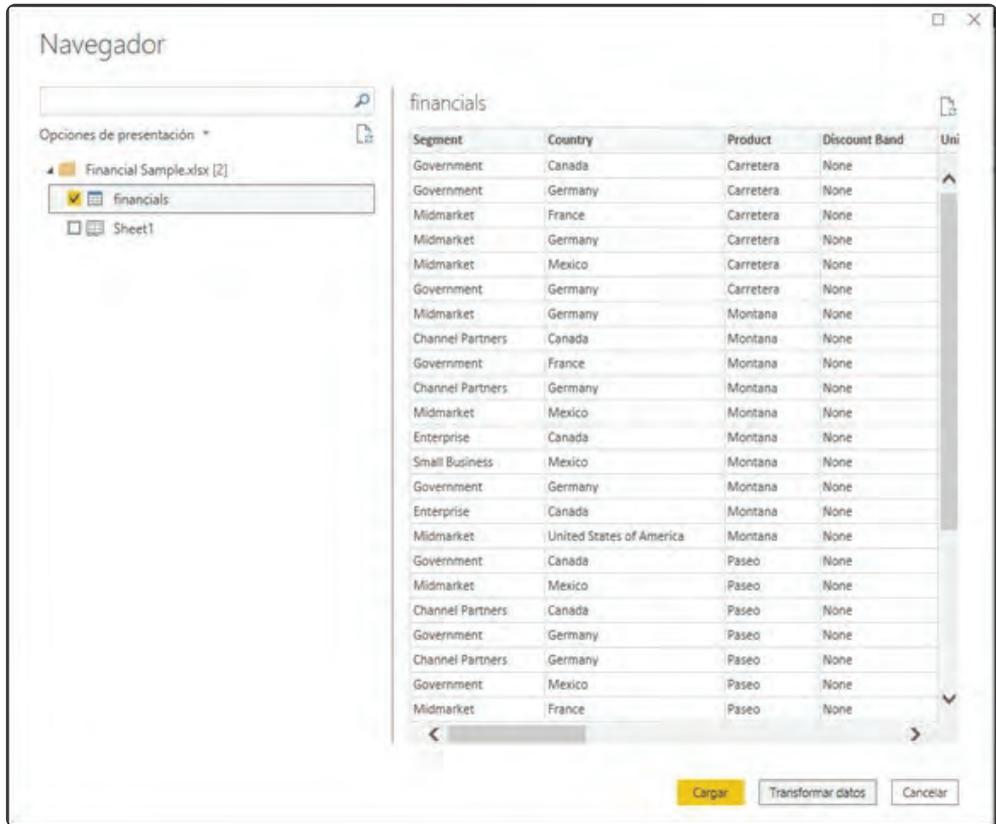


Figura 9. Apertura del archivo Financial Sample.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

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).

1.2.2 Crear el modelo de datos

Una vez cargados los datos es conveniente realizar la *limpieza* y *transformaciones* necesarias para crear el modelo de datos de forma que los datos sean legibles y se pueda efectuar el análisis respectivo.

1.2.2.1 TRANSFORMAR LOS DATOS DE LA VARIABLE UNITS SOLD (UNIDADES VENDIDAS)

La variable *Units Sold (unidades vendidas)* contiene valores decimales, por lo que es necesario convertirlo a valores enteros. Seleccione la columna **Units Sold** y luego seleccione el botón *Tipo de datos* de la pestaña *Inicio* y elija *Número entero* y en la ventana emergente confirme al dar clic en el botón *Sustituir la actual* como se ilustra en la figura 10.

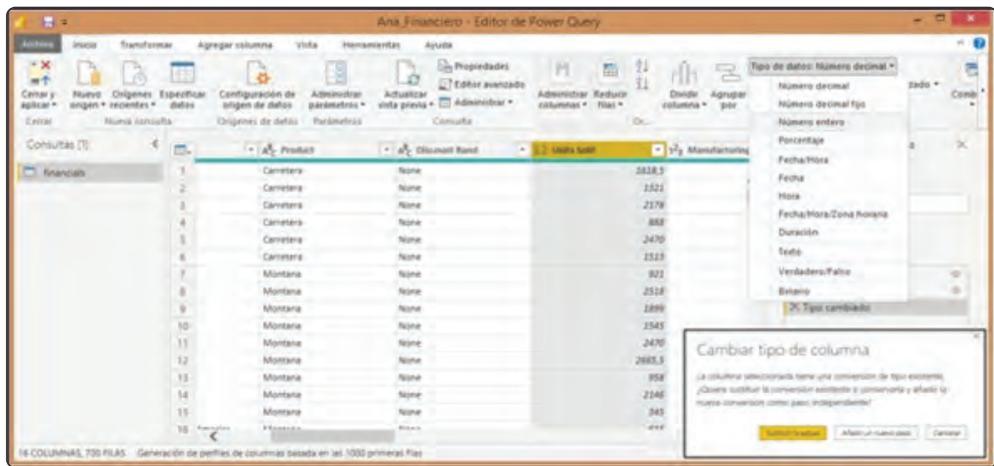


Figura 10. Transformación de los datos de la variable Units Sold (Unidades vendidas).

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

1.2.2.2 MODIFICAR EL FORMATO DE LOS DATOS DE LA VARIABLE SEGMENT (SEGMENTO)

Seleccione la variable *Segment (Segmento)* y de clic en la pestaña *Transformar*, luego en el botón *Columna de texto*, seleccione la opción *Formato* y por último elija *MAYÚSCULAS* tal como se ilustra en la figura 11.

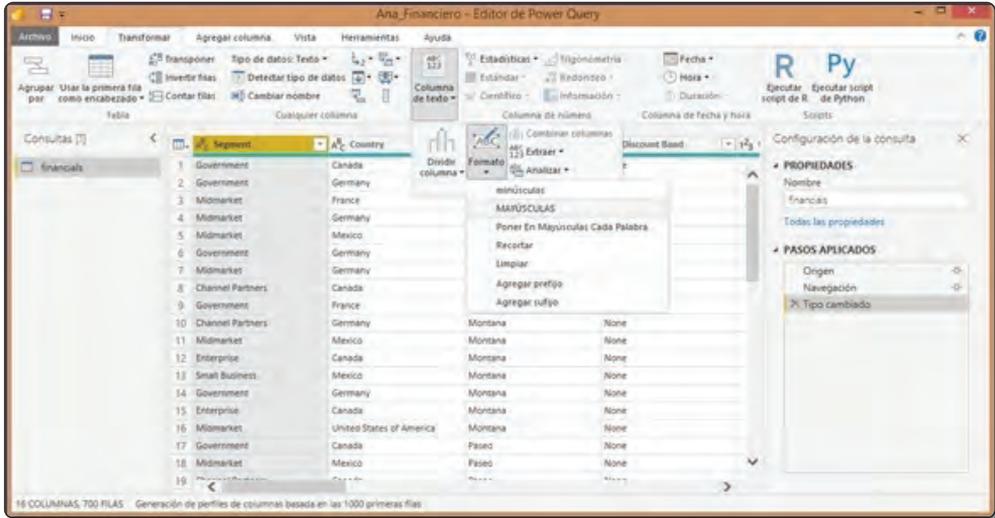


Figura 11. Conversión de los datos de la variable Segment (Segmento) a Mayúsculas. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

1.2.2.3 CAMBIAR EL NOMBRE DE LA VARIABLE MONTH NAME A SÓLO MONTH

De doble clic sobre la variable *Month Name* (*Nombre del mes*) y modifique el nombre a sólo *Month* (*Mes*) tal como se ilustra en la figura 12.

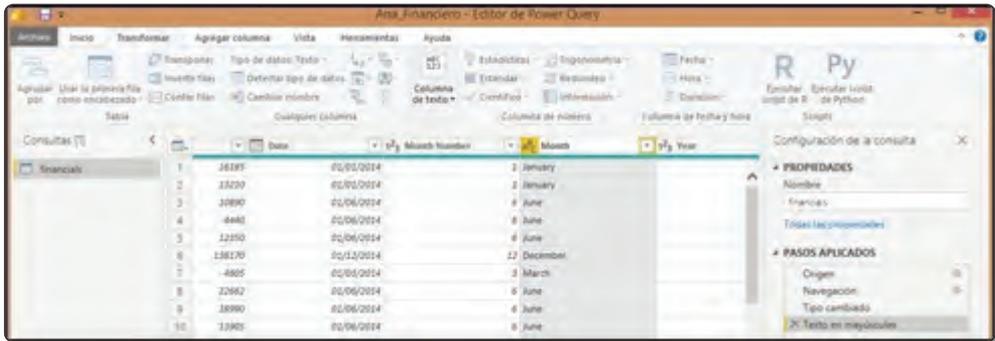


Figura 12. Modificación del nombre de la variable Month Name. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

La figura 13 visualiza el resultado de las acciones realizadas y en la sección *Configuración de la consulta* se ilustra la lista de los *pasos aplicados*.

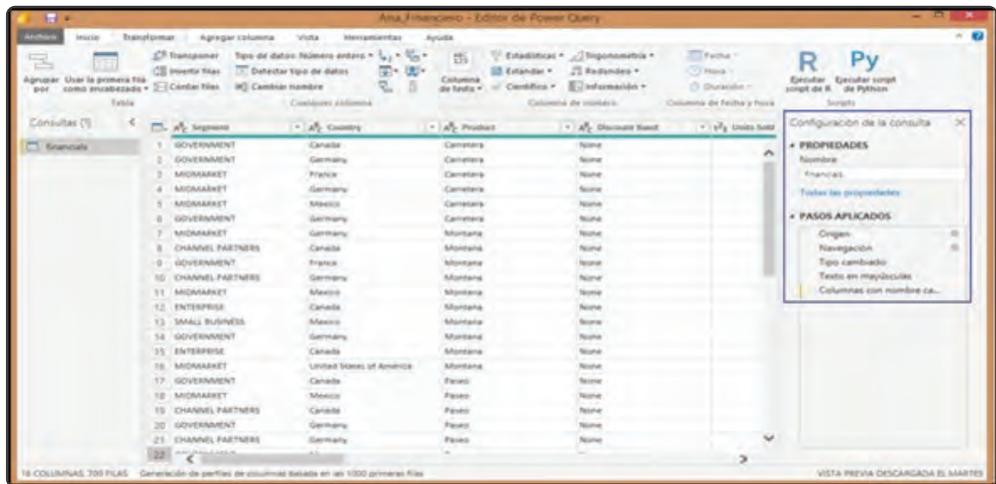


Figura 13. Configuración de la consulta - pasos aplicados.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

De clic en la pestaña **Inicio** del editor de **Power Query** y luego en el botón **Cerrar y aplicar** para disponer los datos para la creación del informe. La figura 14 visualiza la tabla financieras con sus campos respectivos.

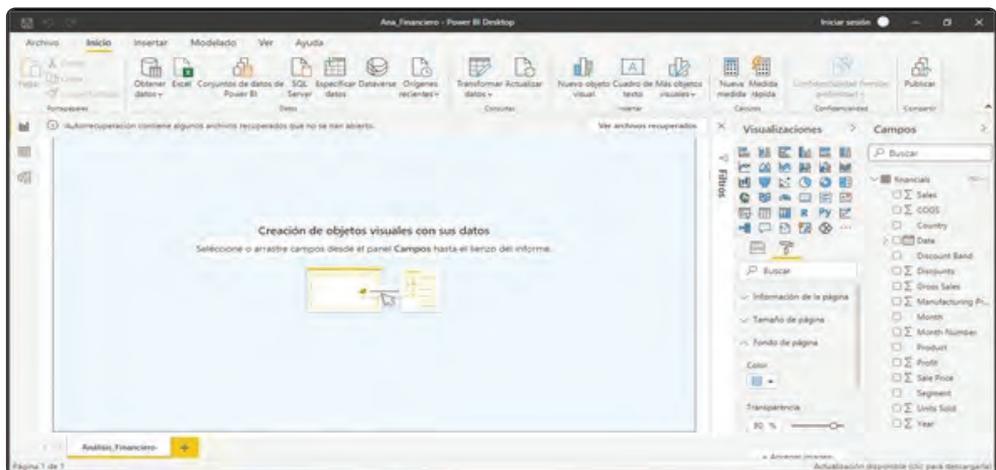


Figura 14. Tabla financieras con sus campos respectivos.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

1.2.2.4 CREACIÓN DE UNA MEDIDA

Para el modelado de datos se crea una medida haciendo uso del lenguaje de fórmulas DAX, de clic en la pestaña *Inicio* y luego en el botón *Nueva medida*. Escriba la expresión $Tot_Units_Sold = SUM(financials[Units Sold])$ y confirme al dar clic en la marca de verificación tal como se indica en la figura 15.

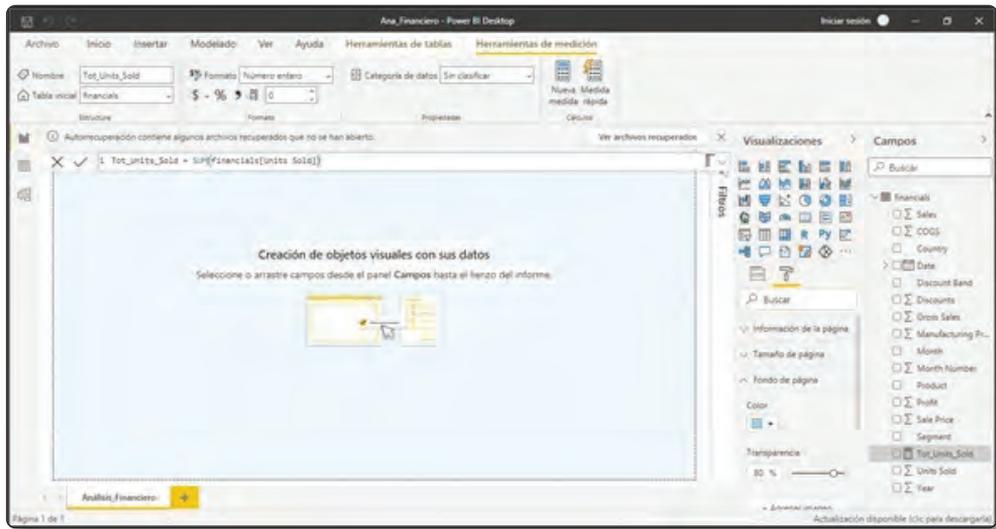


Figura 15. Creación de una medida mediante el lenguaje de fórmulas DAX. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

1.2.2.5 CREACIÓN DE UNA TABLA CALENDARIO

De clic en el botón *Vista de datos* del panel izquierdo, luego en la pestaña *Inicio* y en el botón *Nueva tabla* tal como se indica en la figura 16.

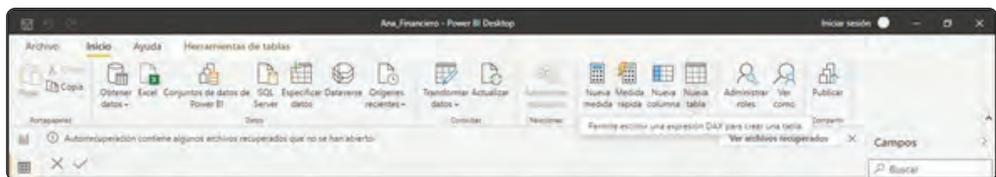


Figura 16. Creación de una nueva tabla. Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

En la barra de fórmula del lenguaje DAX escriba la expresión *Calendario = calendar (DATE (2013, 01, 01), DATE (2014, 12, 31))* y de clic en la marca de verificación para aceptar la expresión. La figura 17 ilustra el resultado obtenido.

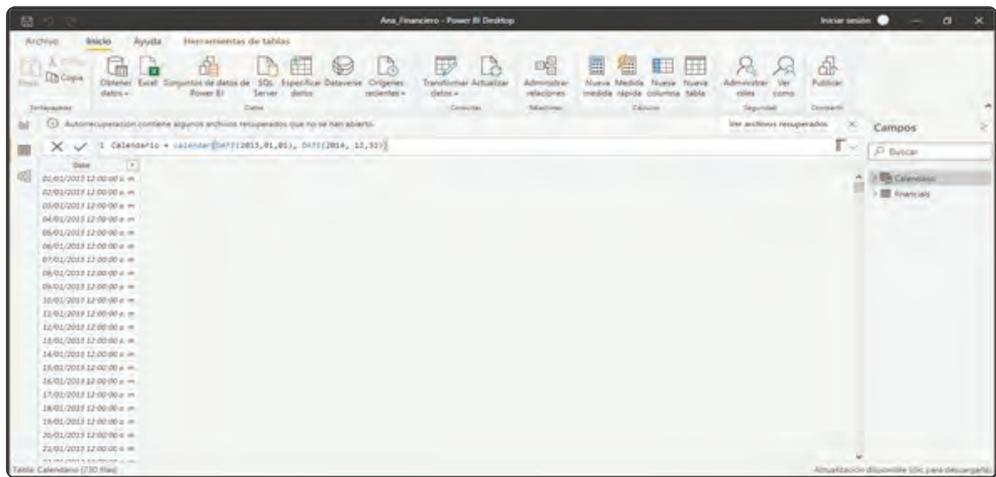


Figura 17. Creación de la tabla calendario para los años 2013 y 2014.

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

De clic en la vista modelo del panel izquierdo, arrastre el campo *Date (Fecha)* de la tabla *financials* a la tabla *calendario* para establecer una relación como se muestra en la figura 18.

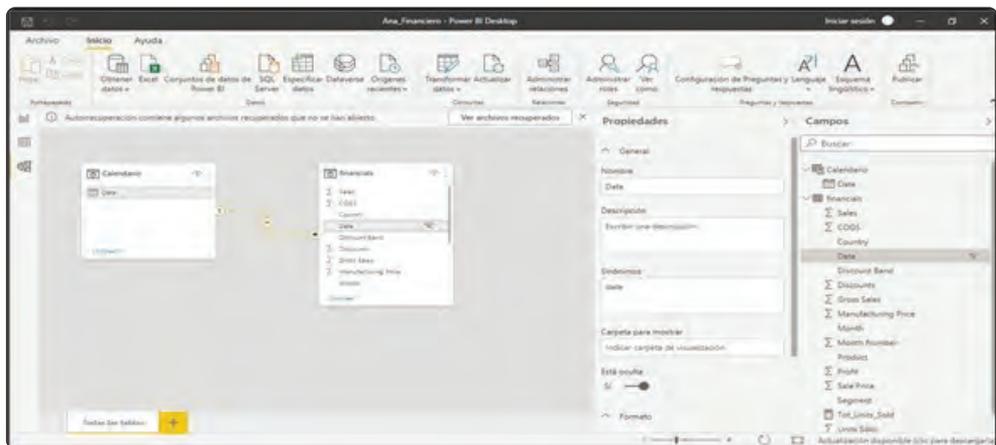


Figura 18. Creación de una relación entre las tablas financieras y calendario por el campo Date (Fecha).

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

1.2.3 Crear el informe

Una vez creado el *modelo de datos* (Carga, limpieza y transformación de los datos), es el momento de crear el informe mediante los diferentes objetos gráficos que provee Power BI Desktop para dar respuesta a los siguientes interrogantes:

- Mes y año en los que se ha conseguido el mayor beneficio
- Ubicaciones donde la empresa tiene más éxito (Top five países)
- Producto en los que la empresa debe definir políticas de expansión
- Principales segmentos donde se obtiene mayor rendimiento (Top three)

1.2.3.1 AGREGAR CUADRO DE TEXTO PARA EL TÍTULO

Para agregar el título al informe, de clic en la pestaña *Insertar* y luego en el botón *Cuadro de texto*, escriba el texto: *Síntesis Ejecutivo: Reporte financiero*. Seleccione el texto y modifique las propiedades según: tamaño 28, fuente Verdana, Color Azul, Negrita, Cursiva, Alineación central como se ilustra en la figura 19.

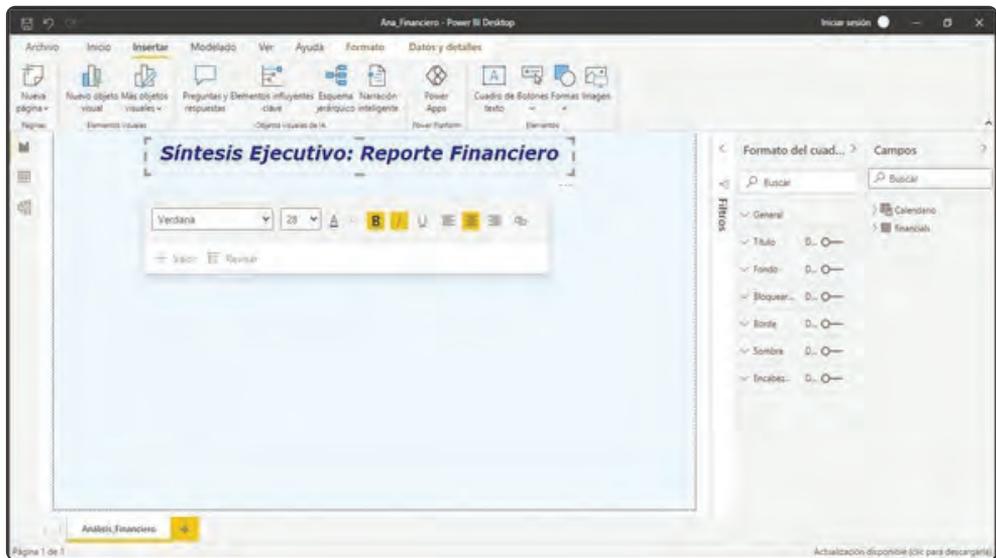


Figura 19. Cuadro de texto para el título.
Fuente. Obtenido mediante software (Microsoft Power BI, 2021).

1.2.3.2 GRÁFICO DE LÍNEAS PARA EL BENEFICIO

Para dar respuesta al interrogante *¿En qué mes y año se ha obtenido el mayor beneficio?*, de clic en el objeto *gráfico de columnas apiladas* del *panel de visualizaciones*, de la tabla *financials* arrastre o de clic en el campo *Profit (Beneficio)* y luego arrastre o de clic en el campo *Date* de la tabla *Calendario* creada. *Power BI* actualiza el gráfico para visualizar el beneficio en cada año de referencia tal como se indica en la figura 20.

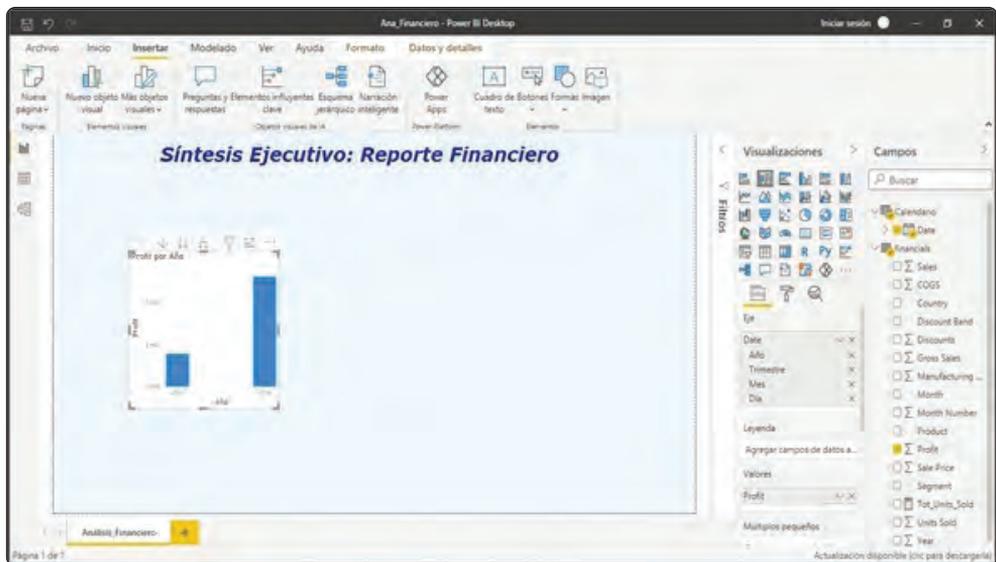


Figura 20. Gráfico de columna apilada que muestra el mayor beneficio por año.

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

Ahora de clic en el botón *Campos* del *panel de visualizaciones*, en la lista desplegable del valor eje de clic en la opción *Date*. *Power BI* actualiza el gráfico para *mostrar el beneficio por mes*, aplique propiedades al objeto gráfico al dar clic en el botón *rodillo* para: agregar un control deslizante de zoom, color y tamaño al título, colores de datos, activar etiquetas de datos, color de fondo con transparencia al 50% y color de borde, cambie el objeto a gráfico de líneas. La figura 21 visualiza el resultado de aplicar las características al objeto gráfico, se infiere que el *mayor beneficio* se obtiene en el *mes de diciembre de 2014* con un valor de **2.025.756,90**