

DASHBOARDS (CUADROS DE MANDO)

- Un proyecto de ciencia de datos es intrínsecamente un proyecto de narración para su audiencia.
- Una forma de presentar un proyecto de ciencia de datos al público es crear un dashboard interactivo.



DASHBOARDS (CUADROS DE MANDO)

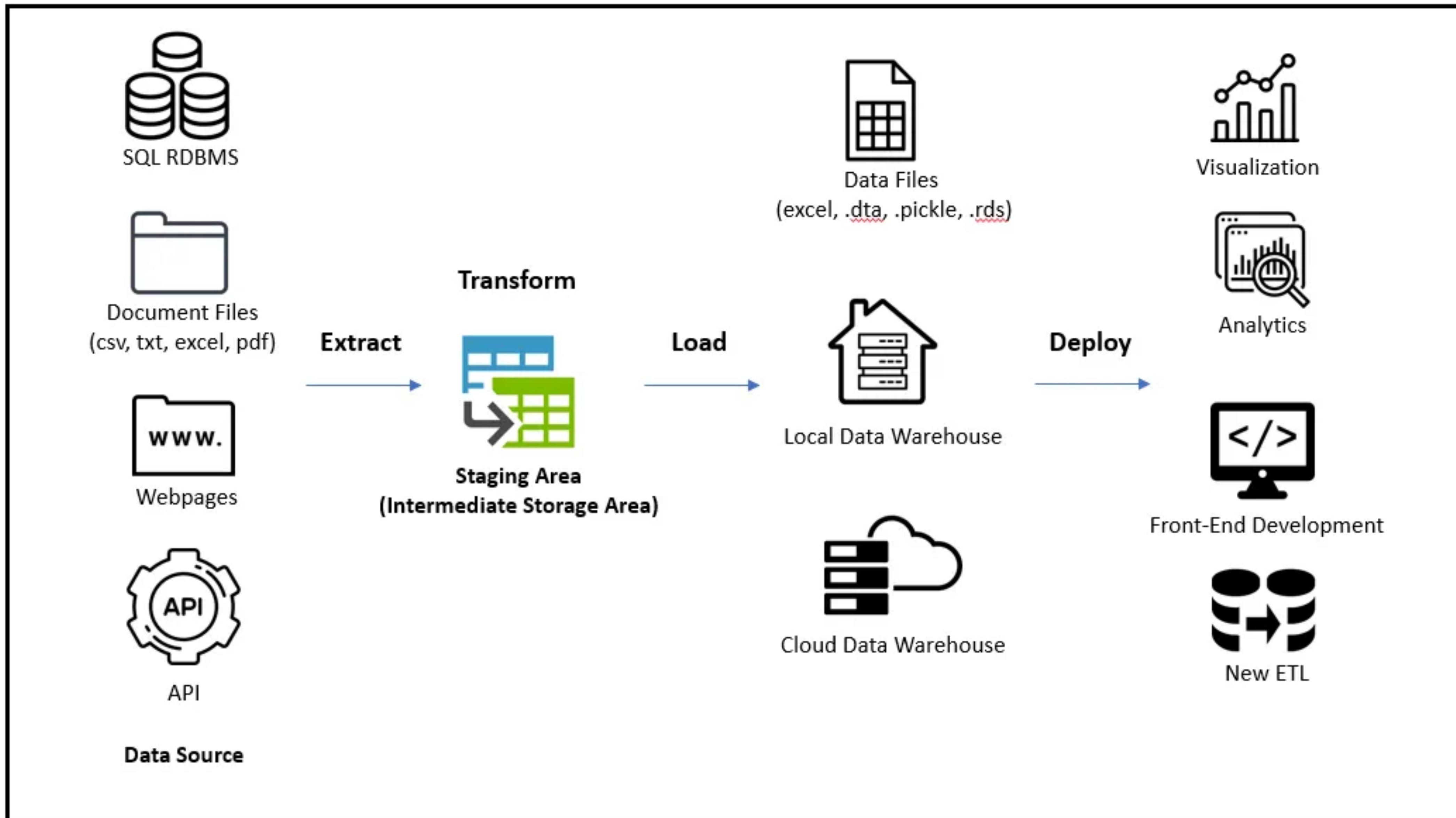
- Los dashboards se conectan a una base de datos y consultan los valores que han sido depositados en ella por otros programas.
- La forma en que se organiza la recogida, limpieza, transformación y presentación de los datos para su análisis está sistematizada y existen numerosas herramientas software que automatizan los flujos de información en la empresa.



PROCESOS ETL

- Extraer, transformar y cargar (ETL) es el proceso de combinar datos de múltiples fuentes en un gran repositorio central denominado almacén de datos (datawarehouse).
- ETL utiliza un conjunto de reglas para limpiar y organizar los datos en bruto y prepararlos para su almacenamiento, análisis de datos y aprendizaje automático.
- Puede abordar necesidades específicas de inteligencia empresarial a través de la analítica de datos (como predecir el resultado de las decisiones empresariales, generar informes y cuadros de mando, reducir la ineficacia operativa, etc.).

ETL Process in Detail



HERRAMIENTAS ETL

- **Informatica PowerCenter** - <https://www.informatica.com/products/data-integration/powercenter.html>
- **Apache Airflow** - <https://airflow.apache.org/>
- **IBM Infosphere Datastage** - <https://www.ibm.com/docs/en/iis/11.7?topic=qualitystage-overview-infosphere-datastage>
- **Oracle Data Integrator** - <https://www.oracle.com/uk/middleware/technologies/data-integrator.html>
- **Microsoft SQL Server Integration Services (SSIS)** - <https://learn.microsoft.com/en-us/sql/integration-services/>
- **Talend Open Studio (TOS)** - <https://www.talend.com/products/talend-open-studio/>
- **Pentaho Data Integration (PDI)** - https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho_Data_Integration
- **Hadoop** - <https://hadoop.apache.org/>
- **AWS Glue** - <https://aws.amazon.com/glue/>
- **AWS Data Pipeline** - <https://aws.amazon.com/datapipeline/>

POR QUÉ ES IMPORTANTE ETL?

- Las organizaciones disponen de datos estructurados y no estructurados procedentes de diversas fuentes, entre las que se incluyen:
 - Datos de clientes procedentes de sistemas de pago en línea y de gestión de relaciones con los clientes (CRM, Customer Relationship Management)
 - Datos de inventario y operaciones procedentes de sistemas de proveedores
 - Datos de sensores de dispositivos del Internet de las cosas (IoT)
 - Datos de marketing de redes sociales y comentarios de clientes
 - Datos de empleados procedentes de sistemas internos de recursos humanos
- Mediante los procesos ETL, los conjuntos de datos individuales sin procesar se pueden preparar en un formato y estructura que sean más consumibles para fines analíticos.
- Por ejemplo, los minoristas en línea pueden analizar los datos de los puntos de venta para prever la demanda y gestionar el inventario. Los equipos de marketing pueden integrar los datos de CRM con los comentarios de los clientes en las redes sociales para estudiar el comportamiento de los consumidores.

CÓMO HA EVOLUCIONADO ETL?

- ETL se originó con la aparición de las bases de datos relacionales que almacenaban datos en forma de tablas para su análisis. Las primeras herramientas ETL intentaban convertir datos de formatos de datos transaccionales a formatos de datos relacionales para su análisis.
- **ETL tradicional**
 - Los datos en bruto solían almacenarse en bases de datos transaccionales que admitían muchas solicitudes de lectura y escritura, pero no se prestaban bien al análisis.
 - Para superar este problema, las herramientas ETL convertían automáticamente estos datos transaccionales en datos relacionales con tablas interconectadas. Los analistas podían utilizar consultas para identificar relaciones entre las tablas, además de patrones y tendencias.
- **ETL moderno**
- **Almacenes de datos (datawarehouses)**
- **Lagos de datos**

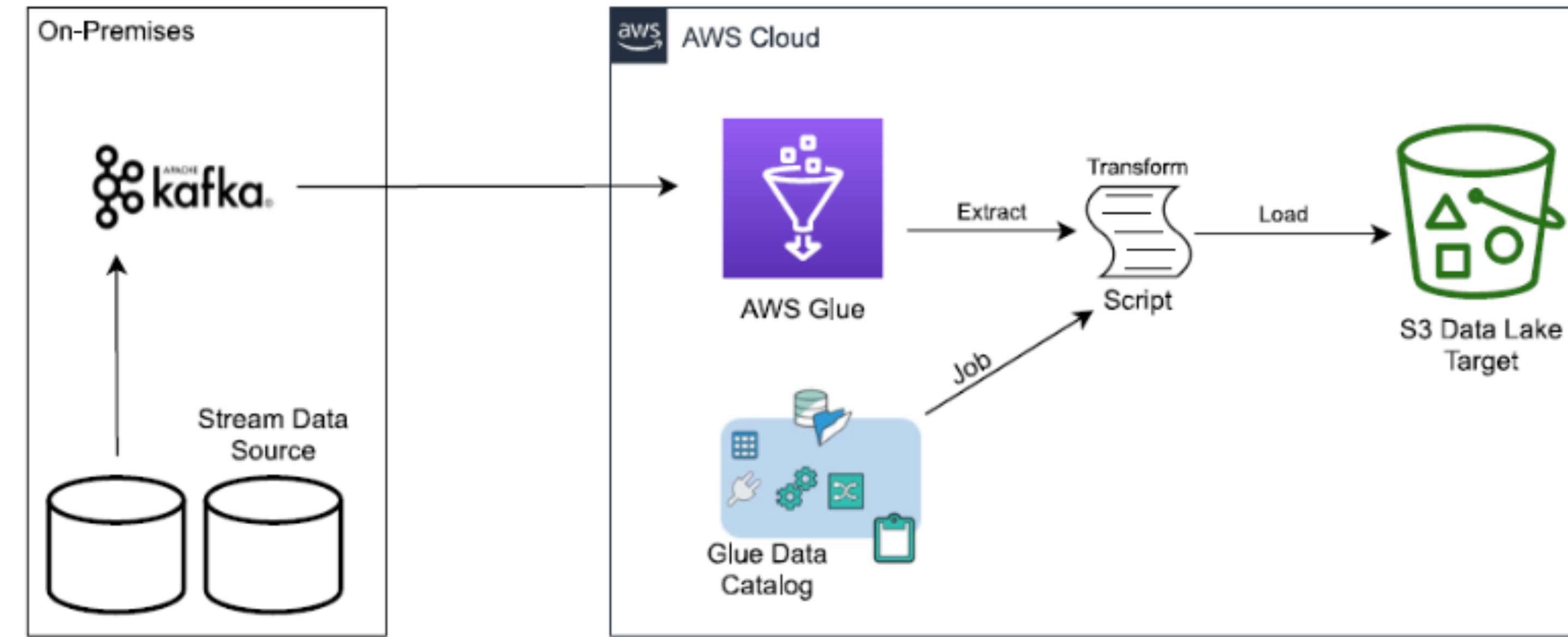
CÓMO HA EVOLUCIONADO ETL?

- ETL tradicional
- ETL moderno
 - A medida que la tecnología ETL evolucionaba, tanto los tipos como las fuentes de datos aumentaban exponencialmente. Surgió la tecnología de nube para crear grandes bases de datos (también llamadas sumideros de datos/data sinks). Estos sumideros de datos pueden recibir datos de múltiples fuentes y disponen de recursos de hardware subyacentes que pueden ampliarse con el tiempo.
- Almacenes de datos (datawarehouses)
- Lagos de datos

CÓMO HA EVOLUCIONADO ETL?

- ETL tradicional
- ETL moderno
- Almacenes de datos (datawarehouses)
 - Un almacén de datos es un repositorio central que puede almacenar múltiples bases de datos. Dentro de cada base de datos, puede organizar los datos en tablas y columnas que describen los tipos de datos de la tabla.
- Lagos de datos
 - En los lagos de datos se almacenan datos estructurados y no estructurados en un repositorio centralizado. Los lagos de datos también le permiten ejecutar diferentes tipos de análisis en sus datos, como consultas SQL, análisis de big data, búsqueda de texto completo, análisis en tiempo real y aprendizaje automático.

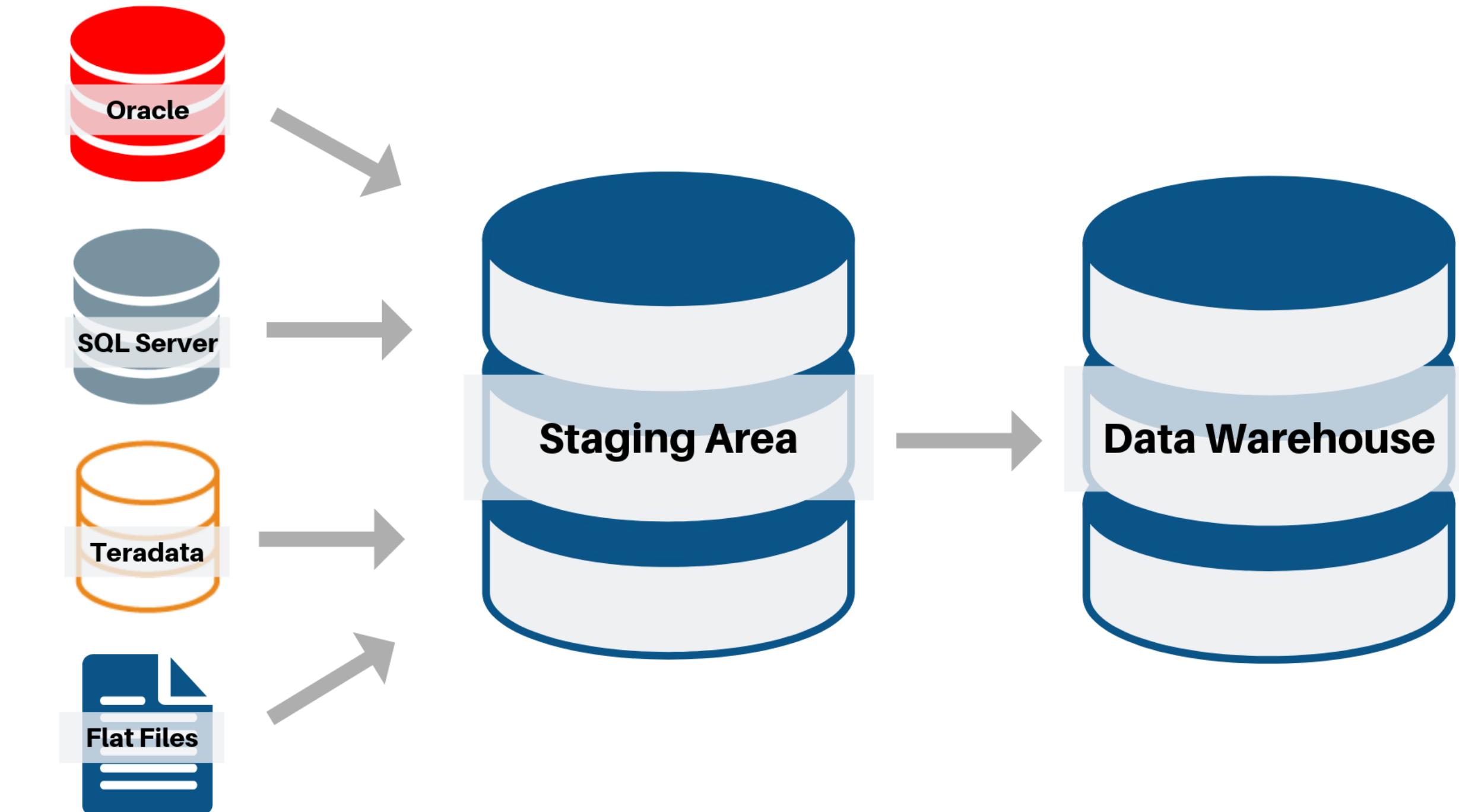
CÓMO FUNCIONA ETL?



- El proceso de extracción, transformación y carga (ETL) traslada datos del sistema de origen al sistema de destino a intervalos periódicos, en tres pasos:
 1. Extraer los datos relevantes de la base de datos de origen
 2. Transformar los datos para que sean más adecuados para el análisis.
 3. Cargar los datos en la base de datos de destino

QUÉ ES LA EXTRACCIÓN DE DATOS?

- En la extracción de datos, las herramientas de extracción, transformación y carga (ETL) extraen o copian datos en bruto de múltiples fuentes y los almacenan en una zona de parada.
- Un área de preparación (o zona de aterrizaje) es un área de almacenamiento intermedio para guardar temporalmente los datos extraídos.
- Las zonas de almacenamiento de datos suelen ser transitorias, lo que significa que su contenido se borra una vez finalizada la extracción de datos.



QUÉ ES LA EXTRACCIÓN DE DATOS?

- La frecuencia con la que el sistema envía datos desde la fuente de datos al almacén de datos de destino depende del mecanismo subyacente de captura de datos de cambios:
 1. **Notificación de actualización:** El sistema de origen notifica cuando cambia un registro de datos. A continuación, puede ejecutar el proceso de extracción para ese cambio. La mayoría de las bases de datos y aplicaciones web proporcionan mecanismos de actualización para soportar este método de integración de datos.
 2. **Extracción incremental:** Algunas fuentes de datos no pueden proporcionar notificaciones de actualización, pero pueden identificar y extraer los datos que se han modificado en un periodo de tiempo determinado. En este caso, el sistema comprueba si se han producido cambios a intervalos periódicos, como una vez a la semana, una vez al mes o al final de una campaña. Sólo es necesario extraer los datos que han cambiado.
 3. **Extracción completa:** Algunos sistemas no pueden identificar los cambios en los datos ni emitir notificaciones, por lo que la única opción es volver a cargar todos los datos. Este método de extracción requiere conservar una copia del último extracto para comprobar qué registros son nuevos. Solo se utiliza con tablas pequeñas.

QUÉ ES LA TRANSFORMACIÓN DE DATOS?

- En la transformación de datos, las herramientas de extracción, transformación y carga (ETL) transforman y consolidan los datos brutos en la zona de preparación para prepararlos para el almacén de datos de destino. La fase de transformación de datos puede implicar los siguientes tipos de cambios en los datos.
- **Transformación básica:** Las transformaciones básicas mejoran la calidad de los datos eliminando errores, vaciando campos de datos o simplificando los datos. Por ejemplo:
 - Limpieza de datos: La limpieza de datos elimina errores y asigna los datos de origen al formato de datos de destino. Por ejemplo, puede asignar campos de datos vacíos al número 0, asignar el valor de datos "Padre" a "P" o asignar "Hijo" a "C".
 - Deduplicación de datos: La deduplicación en la limpieza de datos identifica y elimina los registros duplicados.
 - Revisión del formato de los datos: La revisión del formato convierte datos, como conjuntos de caracteres, unidades de medida y valores de fecha/hora, a un formato coherente. Por ejemplo, una empresa alimentaria puede tener diferentes bases de datos de recetas con ingredientes medidos en kilogramos y libras. ETL lo convertirá todo a libras.
- **Transformación avanzada:**

QUÉ ES LA TRANSFORMACIÓN DE DATOS?

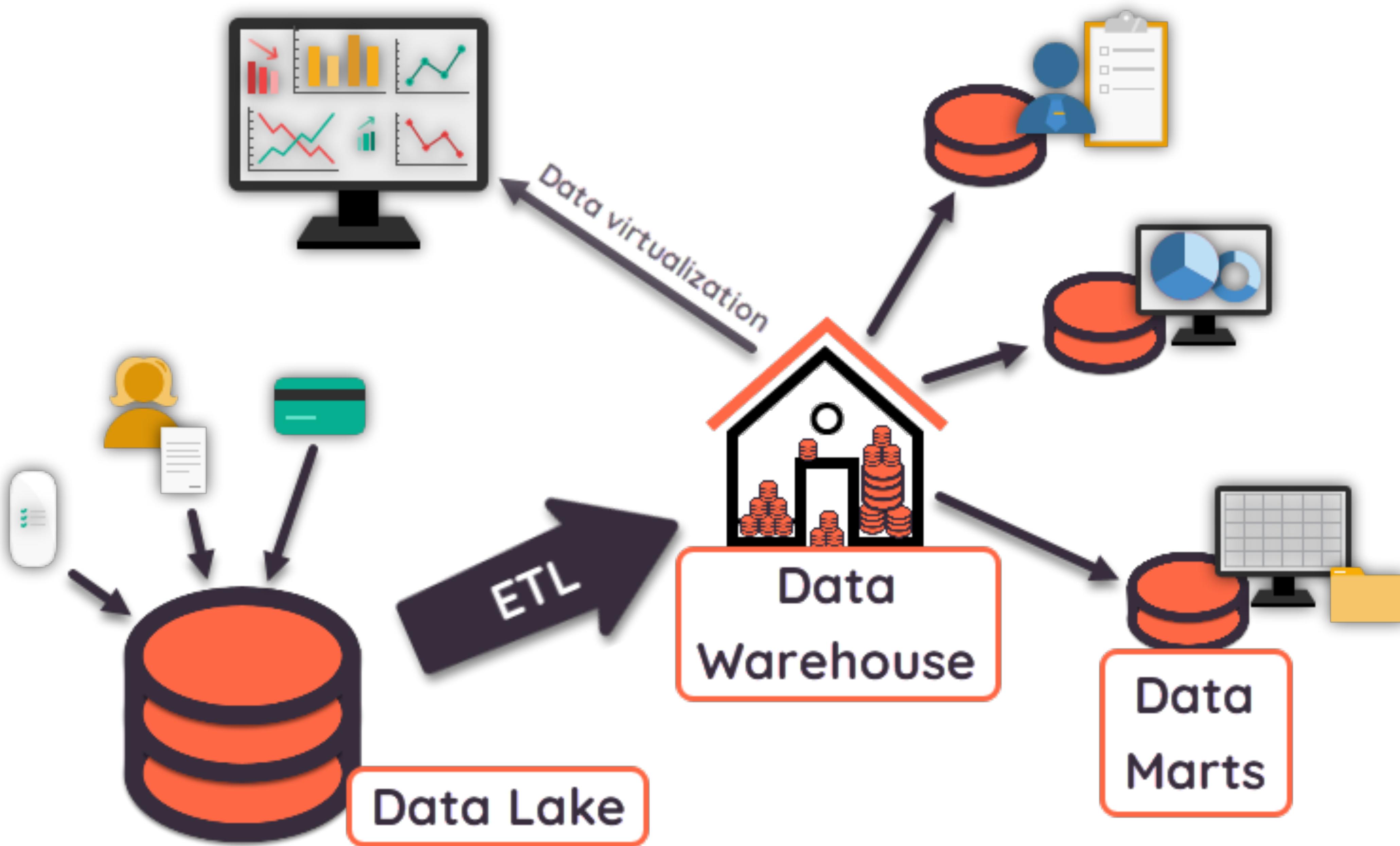
- **Transformación avanzada de datos:** Las transformaciones avanzadas utilizan reglas de negocio para optimizar los datos y facilitar su análisis. Por ejemplo:
 - Derivación: La derivación calcula nuevos valores a partir de valores existentes. Por ejemplo, puede convertir los ingresos en beneficios restando los gastos o calcular el coste total de una compra multiplicando el precio de cada artículo por el número de artículos pedidos.
 - Unión: En la preparación de datos, la unión enlaza los mismos datos de diferentes fuentes de datos. Por ejemplo, puede hallar el coste total de compra de un artículo sumando el valor de compra de diferentes proveedores y almacenando sólo el total final en el sistema de destino.
 - División: Puede dividir una columna o atributo de datos en varias columnas en el sistema de destino. Por ejemplo, si el origen de datos guarda el nombre del cliente como "Jane John Doe", puede dividirlo en nombre, segundo nombre y apellidos.
 - Resumen: La integración mejora la calidad de los datos reduciendo un gran número de valores de datos a un conjunto de datos más pequeño. Por ejemplo, los valores de las facturas de pedidos de clientes pueden tener muchas cantidades pequeñas diferentes. Puede resumir los datos sumándolos a lo largo de un periodo determinado para construir una métrica de valor de vida del cliente (CLV).
 - Cifrado: Puede proteger los datos confidenciales para cumplir con la legislación o la privacidad de los datos añadiendo cifrado antes de que los datos se transmitan a la base de datos de destino.

QUÉ ES LA CARGA DE DATOS?

- En la carga de datos, las herramientas de extracción, transformación y carga (ETL) trasladan los datos transformados de la zona de preparación al almacén de datos de destino. En la mayoría de las organizaciones que utilizan ETL, el proceso está automatizado, bien definido, es continuo y se realiza por lotes. Hay dos tipos de carga:
 - **Carga completa:** Todos los datos de la fuente se transforman y se trasladan al almacén de datos. La carga completa suele tener lugar la primera vez que se cargan datos de un sistema fuente en el almacén de datos.
 - **Carga incremental:** La herramienta ETL carga el delta (o diferencia) entre los sistemas de destino y origen a intervalos regulares. Almacena la fecha de la última extracción para que sólo se carguen los registros añadidos después de esa fecha. Hay dos formas de implementar la carga incremental.
 - Carga incremental en flujo: Si tiene volúmenes de datos pequeños, puede transmitir cambios continuos a través de canalizaciones de datos al almacén de datos de destino. Cuando la velocidad de los datos aumenta a millones de eventos por segundo, puede utilizar el procesamiento de flujos de eventos para supervisar y procesar los flujos de datos con el fin de tomar decisiones más oportunas.
 - Carga incremental por lotes: Si dispone de grandes volúmenes de datos, puede recopilar cambios de datos de carga en lotes periódicamente. Durante este periodo de tiempo establecido, no puede producirse ninguna acción ni en el sistema de origen ni en el de destino mientras se sincronizan los datos.

QUÉ ES LA VIRTUALIZACIÓN DE DATOS?

- La virtualización de datos utiliza una capa de abstracción de software para crear una vista de datos integrada sin extraer, transformar o cargar físicamente los datos.
- Las organizaciones utilizan esta funcionalidad como un repositorio de datos unificado virtual sin el gasto y la complejidad de construir y gestionar plataformas separadas para la fuente y el destino.
- Aunque puede utilizar la virtualización de datos junto con la extracción, transformación y carga (ETL), cada vez se considera más como una alternativa a ETL y otros métodos físicos de integración de datos.



QUÉ ES ELT?

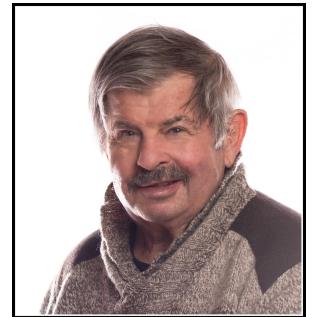
- Extraer, cargar y transformar (ELT) es una extensión de extraer, transformar y cargar (ETL) que invierte el orden de las operaciones.
- Puede cargar datos directamente en el sistema de destino antes de procesarlos. El área de preparación intermedia no es necesaria porque el almacén de datos de destino tiene capacidades de mapeo de datos en su interior. ELT se ha hecho más popular con la adopción de la infraestructura de nube, que proporciona a las bases de datos de destino la potencia de procesamiento que necesitan para las transformaciones.
- ETL comparado con ELT:
 - ELT funciona bien para conjuntos de datos no estructurados de gran volumen que requieren una carga frecuente. También es ideal para big data porque la planificación de la analítica puede realizarse después de la extracción y el almacenamiento de los datos. Deja la mayor parte de las transformaciones para la fase de análisis y se centra en cargar datos brutos mínimamente procesados en el almacén de datos.
 - El proceso ETL requiere más definición al principio. Los analistas deben participar desde el principio para definir los tipos de datos, las estructuras y las relaciones.

ALMACEN DE DATOS (DATAWAREHOUSE)

- Un almacén de datos es un tipo de sistema de gestión de datos diseñado para habilitar y dar soporte a las tareas de inteligencia empresarial (BI), especialmente las analíticas.
- Los data warehouses están diseñados solamente para realizar consultas y tareas de análisis, y suelen contener grandes cantidades de datos históricos.
- A menudo, la información dentro de un data warehouse proviene de una amplia gama de fuentes, como los archivos de registro de aplicaciones o las aplicaciones de transacción.

ALMACEN DE DATOS (DATAWAREHOUSE)

- Un almacén de datos típico suele incluir los siguientes elementos:
 - Una **base de datos relacional** para almacenar y gestionar los datos.
 - Una **solución de extracción, carga y transformación (ELT)** para preparar los datos para el análisis.
 - **Análisis estadísticos**, informes y funciones de extracción de datos.
 - Herramientas de **análisis de clientes** para visualizar y presentar datos a usuarios de negocio.
 - Otras aplicaciones analíticas más sofisticadas generan información procesable mediante la aplicación de algoritmos de ciencia de datos e inteligencia artificial (IA), o gráficos y funciones espaciales que habilitan más tipos de análisis de datos a escala.



VENTAJAS DE UN ALMACÉN DE DATOS

- Los almacenes de datos permiten analizar grandes cantidades de datos y mantener un registro histórico.
- Según William Inmon, los data warehouses cuentan con cuatro características exclusivas:
 - **Orientados a sujetos.** Pueden analizar datos sobre un tema particular o área funcional (como las ventas).
 - **Integrados.** Los data warehouses crean uniformidad entre diferentes tipos de datos de fuentes dispares.
 - **No volátiles.** Cuando la información entra en un data warehouse, es estable y no cambia.
 - **Variante en el tiempo.** La analítica de los data warehouses toma en consideración los cambios que se producen con el tiempo.
- Un almacén de datos bien diseñado realizará las consultas muy rápidamente, ofrecerá un alto rendimiento de datos y proporcionará suficiente flexibilidad para que los usuarios finales puedan segmentar y desglosar o reducir el volumen de datos para un examen más detallado con el fin de satisfacer diversas necesidades, tanto generales como específicas. El data warehouse es la base funcional para los entornos de BI middleware que suministran informes, cuadros de mando y otras interfaces a los usuarios finales.

ARQUITECTURA DE UN DATAWAREHOUSE

- La arquitectura de un data warehouse depende de las necesidades específicas de la organización. Algunas de las arquitecturas más comunes son:
 - **Sencilla.** Todos los data warehouses comparten un diseño básico en el que los metadatos, los datos de resumen y los datos sin procesar se almacenan en el repositorio central del almacén. En uno de los lados, el repositorio se alimenta de fuentes de datos y, en el otro, los usuarios finales acceden para las tareas de análisis, elaboración de informes y extracción.
 - **Sencilla con zona de preparación.** Es necesario limpiar y procesar los datos operativos antes de colocarlos en el almacén. Aunque esto se puede hacer mediante programación, muchos data warehouses añaden una zona de preparación de datos antes de introducirlos en el almacén, a fin de simplificar la preparación.
 - **Radial.** Al añadir data marts entre el repositorio central y los usuarios finales, las organizaciones pueden personalizar su data warehouse para atender a varias áreas de negocio. Cuando los datos ya están listos para el uso, se mueven al data mart correspondiente.
 - **Entorno de pruebas.** Los entornos de pruebas (sandbox) son zonas seguras, privadas y protegidas donde las empresas pueden explorar de forma rápida e informal nuevos conjuntos de datos o nuevas maneras de analizarlos sin tener que cumplir con las reglas y los protocolos formales del data warehouse.

EVOLUCIÓN DE LOS ALMACENES DE DATOS

- Los almacenes de datos han evolucionado desde los análisis básicos de datos hasta la inteligencia artificial y el aprendizaje automático
- Cuando aparecieron por primera vez los data warehouses a finales de los 80, su objetivo era que los datos pasaran de los sistemas operativos a los sistemas de apoyo a las decisiones (DSS). Aquellos primeros data warehouses necesitaban un volumen enorme de redundancias. La mayoría de empresas tenían múltiples entornos DSS para sus diferentes usuarios. Aunque los entornos DSS utilizaban, en gran medida, los mismos datos, la recopilación, la limpieza y la integración de estos se solía replicar para cada entorno.
- A medida que los data warehouses se volvieron más eficientes, pasaron de ser almacenes de información compatibles con plataformas de BI tradicionales a convertirse en infraestructuras analíticas más abiertas que admitían una amplia variedad de aplicaciones, como la analítica operativa o la gestión del rendimiento.

EVOLUCIÓN DE LOS ALMACENES DE DATOS

► El soporte para cada uno de estos cinco pasos requiere de una creciente variedad de conjuntos de datos. Los últimos tres pasos en particular exigen una gama de datos y funciones analíticas más amplia.

Fase	Función	Valor para la empresa
1	Elaboración de informes sobre transacciones	Suministra información relacional para crear instantáneas del rendimiento de la empresa
2	Corte y fragmentación, consultas ad hoc, herramientas de BI	Funciones avanzadas para lograr conocimientos más profundos y un análisis más sólido
3	Predicción de rendimiento en el futuro (minería de datos)	Desarrolla visualizaciones y una inteligencia empresarial progresiva
4	Análisis táctico (espacial, estadístico)	Ofrece escenarios hipotéticos para fundamentar las decisiones prácticas recurriendo a análisis más exhaustivos
5	Almacena meses o años de datos	Solo almacena datos de las últimas semanas o meses

¿QUÉ ES UN DATA WAREHOUSE EN LA NUBE?

- Un data warehouse en la nube utiliza la nube para asimilar y almacenar datos de fuentes de datos dispares.
- Los almacenes de datos originales se crearon en servidores ubicados en entornos locales. Estos almacenes de datos locales siguen ofreciendo ventajas en la actualidad. En muchos casos, brindan una mejor gobernanza, seguridad, soberanía de datos y latencia. Sin embargo, los data warehouses in-situ no son tan elásticos y requieren una previsión compleja para determinar cómo escalar el data warehouse para futuras necesidades. La gestión de estos data warehouses también puede resultar muy compleja.
- Por otro lado, entre las ventajas de los almacenes de datos en la nube se pueden enumerar las siguientes:
 - Soporte flexible y con escalabilidad horizontal para abordar necesidades crecientes de almacenamiento o computación
 - Facilidad de uso y gestión (seguridad ante intrusiones, backups)
 - Ahorro de costes
- Asimismo, la mayoría de los almacenes de datos en la nube siguen un modelo de pago por consumo, que permite a los clientes generar nuevos ahorros de costos.

DATA LAKE (LAGOS DE DATOS)

- Para gestionar grandes volúmenes de datos de diversas fuentes, se utilizan data lakes y data warehouses:
 - Los **data lakes** almacenan una gran cantidad de datos sin filtrar y de todo tipo.
 - En un data lake, los datos de aplicaciones de áreas de negocio, aplicaciones móviles, redes sociales, dispositivos de IoT, etc. se obtienen como datos sin procesar.
 - La estructura, la integridad, la selección y el formato de los diversos conjuntos de datos se adquiere en el momento del análisis por parte de la persona que lo realiza.
 - Cuando las organizaciones necesitan un almacenamiento de bajo costo para datos sin estructura ni formato y de múltiples fuentes (a fin de usarlos en el futuro para algún propósito en particular), los data lakes podrían ser la opción ideal.
 - Los **data warehouses** están diseñados específicamente para analizar datos.
 - Dentro de un almacén de datos, el procesamiento analítico se lleva a cabo con datos previamente preparados para su análisis —recopilados, contextualizados y transformados— con el fin de generar conocimientos basados en análisis.
 - Los data warehouses también son perfectos para gestionar grandes cantidades de datos de diversas fuentes.
 - Cuando las organizaciones necesitan una analítica de datos avanzada o un análisis que se base en datos históricos de múltiples fuentes de toda la empresa, es probable que el almacén de datos sea la mejor opción

DATAWAREHOUSE Y OLTP VS OLAP

- Los almacenes de datos son entornos relacionales que se utilizan para el análisis de datos, sobre todo para datos históricos. Las organizaciones utilizan los data warehouses para descubrir en sus datos patrones y relaciones que se desarrollan con el tiempo.
- En contraste, los entornos transaccionales se utilizan para procesar transacciones de forma continua, y se utilizan comúnmente para la entrada de pedidos y las transacciones financieras y de ventas. No se basan en datos históricos. De hecho, en los entornos OLTP (Procesamiento de Transacciones en Línea), los datos históricos a menudo se archivan o simplemente se eliminan para mejorar el rendimiento.

OLTP VS OLAP

	Almacén de datos	Sistema OLTP
Carga de trabajo	Permite consultas ad-hoc y análisis de datos	Solo permite operaciones predefinidas
Modificaciones de datos	Actualizaciones automáticas de forma regular	Las actualizaciones las realizan los usuarios finales al emitir declaraciones individuales
Diseño de esquemas	Utiliza esquemas parcialmente desnormalizados para optimizar el rendimiento	Utiliza esquemas completamente normalizados para garantizar la uniformidad de los datos
Escaneo de datos	Abarca miles de millones de filas	Solo puede acceder a unos pocos registros al mismo tiempo
Datos históricos	Almacena meses o años de datos	Solo almacena datos de las últimas semanas o meses

DASHBOARDS Y KPIS

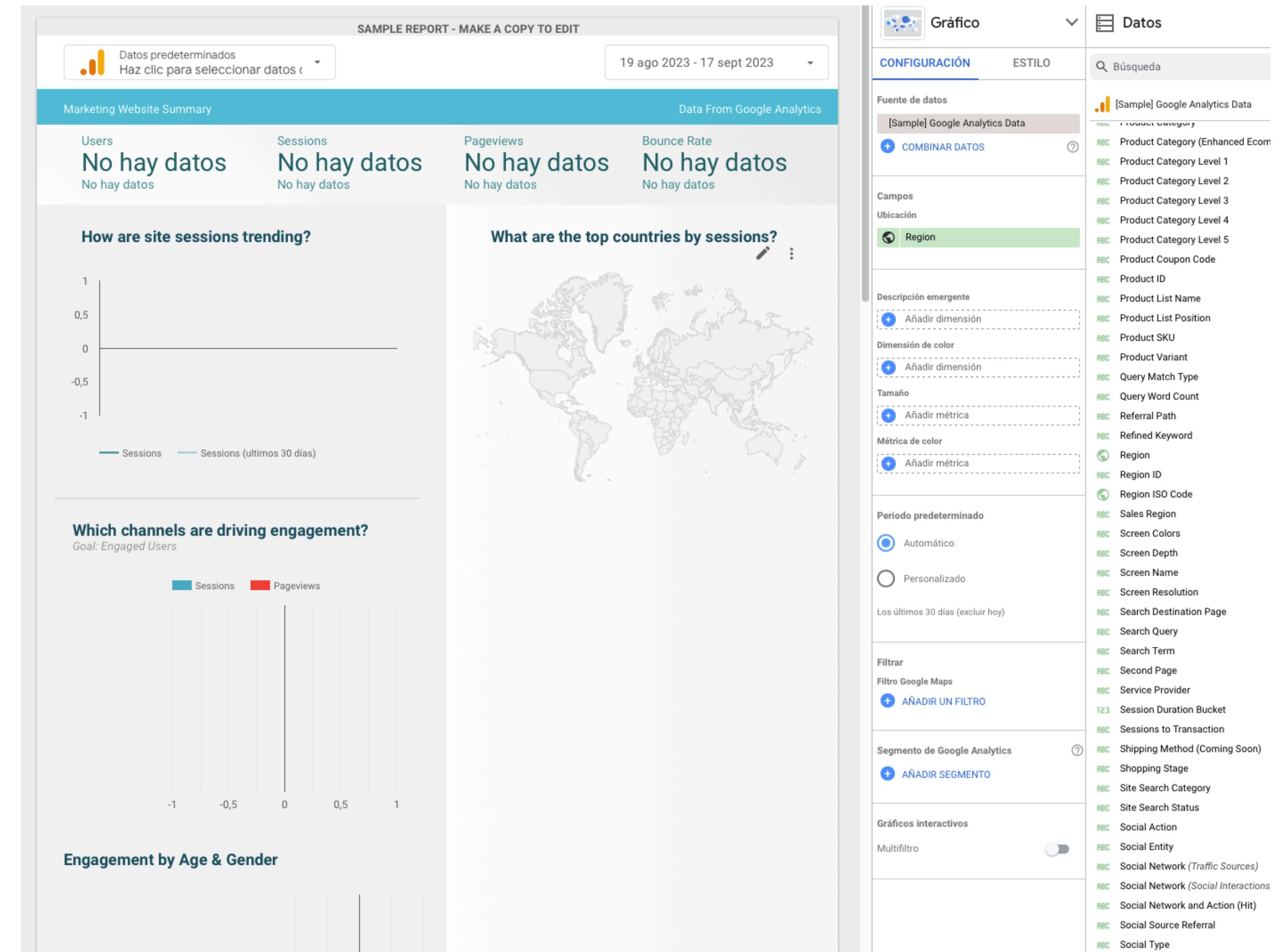
- El término KPI, siglas en inglés, de Key Performance Indicator (Indicador Clave de Desempeño) hace referencia a una serie de métricas que se utilizan para sintetizar la información sobre la eficacia y productividad de las acciones que se lleven a cabo en un negocio con el fin de poder tomar decisiones y determinar aquellas que han sido más efectivas a la hora de cumplir con los objetivos marcados en un proceso o proyecto concreto.
- Un dashboard (cuadro de mando) es una representación gráfica de los principales KPIs que intervienen en la consecución de los objetivos de un negocio.
- Es decir, un dashboard es un conjunto de gráficas que sirven para saber cómo va un negocio de forma rápida y sencilla.
- Este formato de presentación ha incrementado su popularidad en las empresas debido a que ayuda de forma significativa a la toma de decisiones.

HERRAMIENTAS PARA CREACIÓN DE DASHBOARDS

- <https://www.datamation.com/big-data/best-dashboard-software/>
- Best for Built-In Analytics and Support: [Tableau](#)
- Best for Visualizations and Included Library/Templates: [Grafana](#)
- Best for Integrations/Datasources: [Google Looker Studio](#)
- Best for Pricing: [Microsoft Power BI](#)
- Best for Third-party Connectors: [Klipfolio](#)
- Best for Multi-Channel Sharing: [Geckoboard](#)
- Best for Pricing Options: [Databox](#)
- Best for a Wide Range of Use Cases: [Qlik](#)
- Best for Ease-of-Use: [Metabase](#)
- Best for AWS Cloud Infrastructures/Services: [Amazon QuickSight](#)

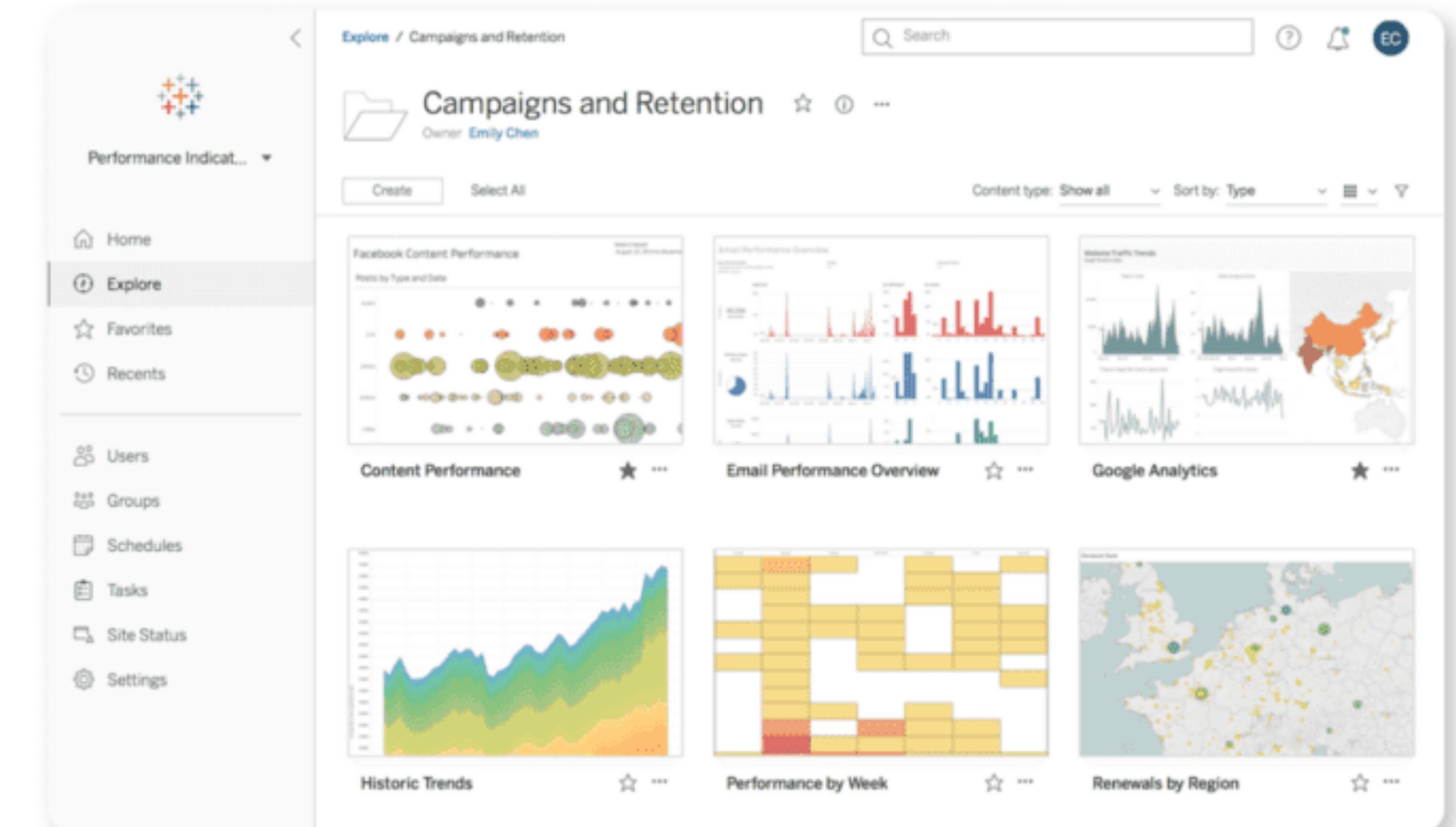
GOOGLE LOOKER STUDIO

- <https://developers.google.com/static/looker-studio/>
- Google Looker Studio es la solución de Google para crear informes gráficos, dashboards y elementos de visualización de datos informativos.



TABLEAU

- Uno de los productos líderes del mercado
- No se necesitan conocimientos previos de programación
- Puede conectarse a varias fuentes de datos
- Permite a los usuarios crear informes uniendo y mezclando diferentes conjuntos de datos
- Tableau Server puede gestionar todas las fuentes de datos publicadas dentro de una organización



KLIPFOLIO

- <https://www.klipfolio.com/>
- Es una aplicación en la nube / SaaS (Software as a Service, https://en.wikipedia.org/wiki/Software_as_a_service)
- Cómodo de usar para analizar redes sociales



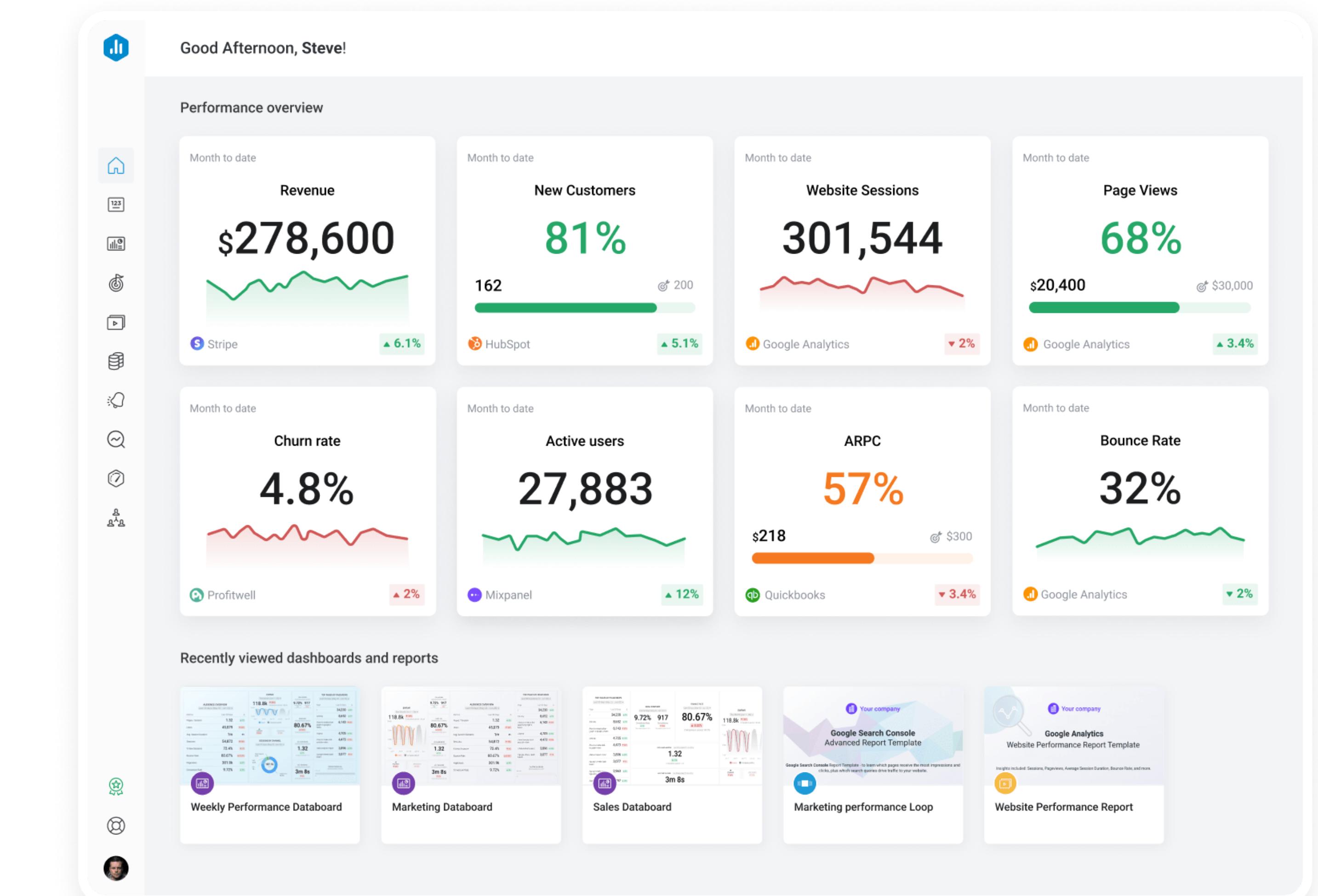
GECKOBOARD

- <https://www.geckoboard.com/>
- Orientado a negocio
- Permite desarrollo rápido de dashboards, sencillo de configurar

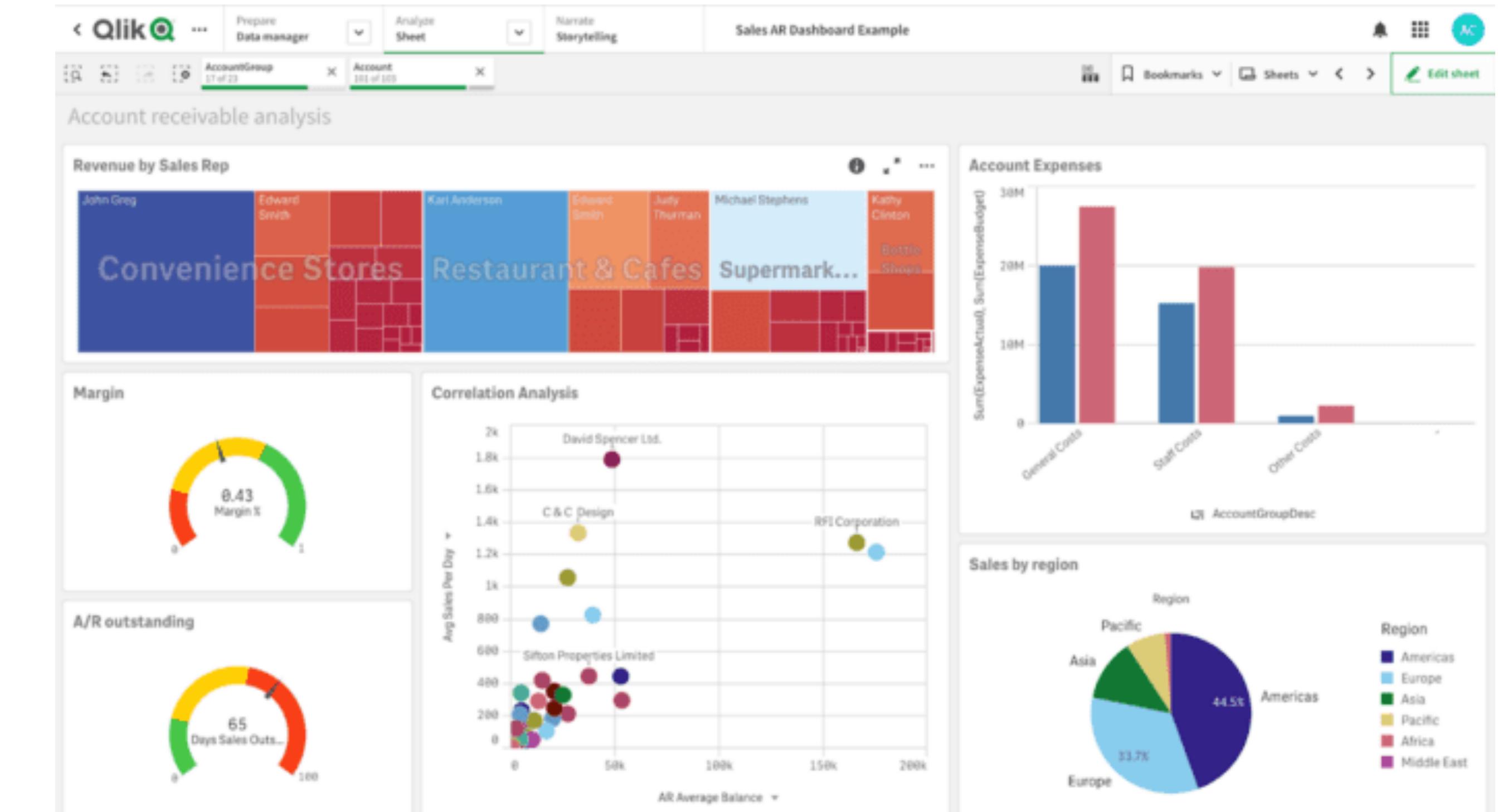


DATABOX

- <https://databox.com/>
- Tiene una versión gratuita, con algunos límites de uso

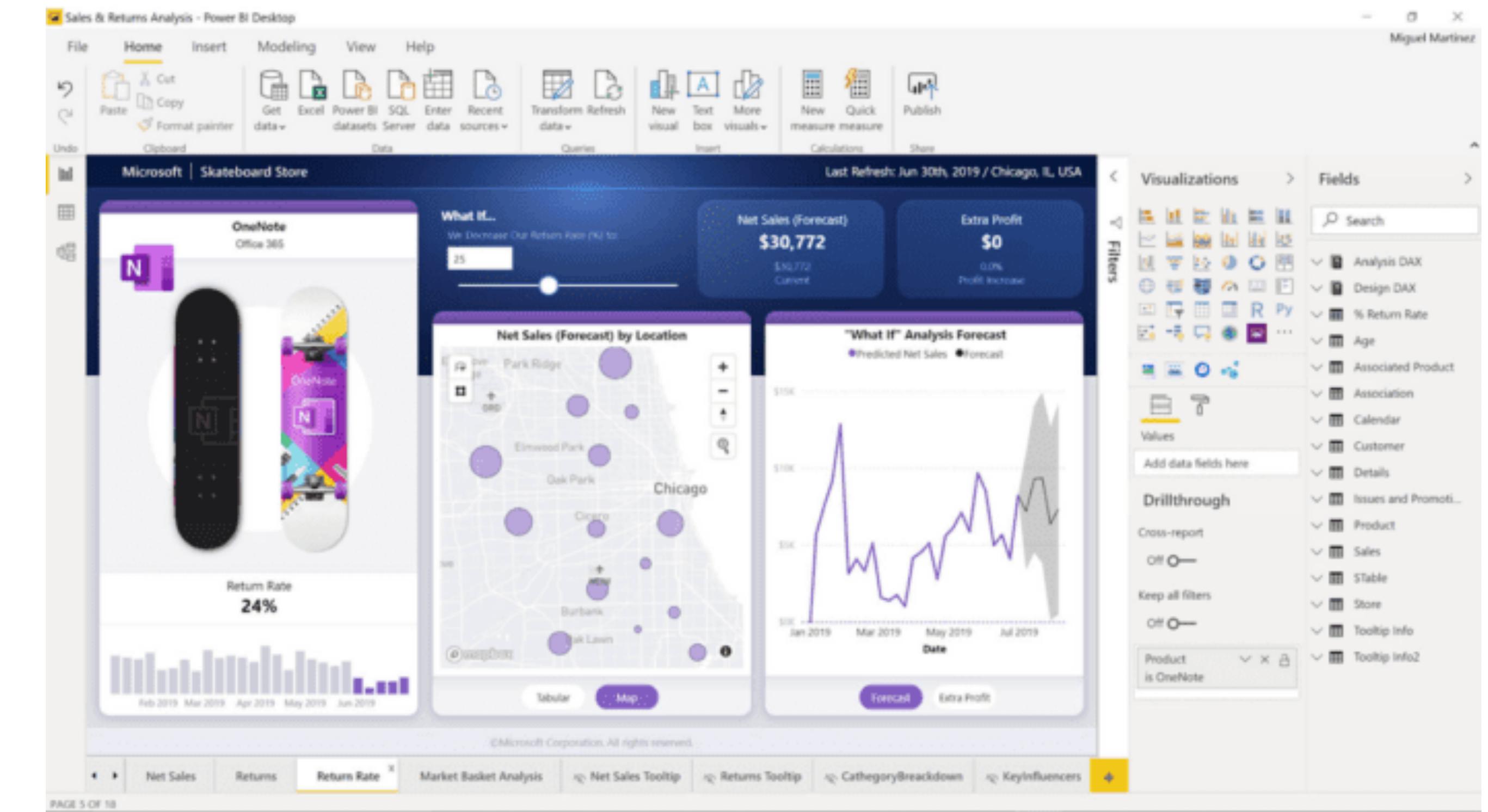


- <https://wwwqlik.com/us/products/qlikview>
- Orientada a la empresa, permite definir procesos ETL



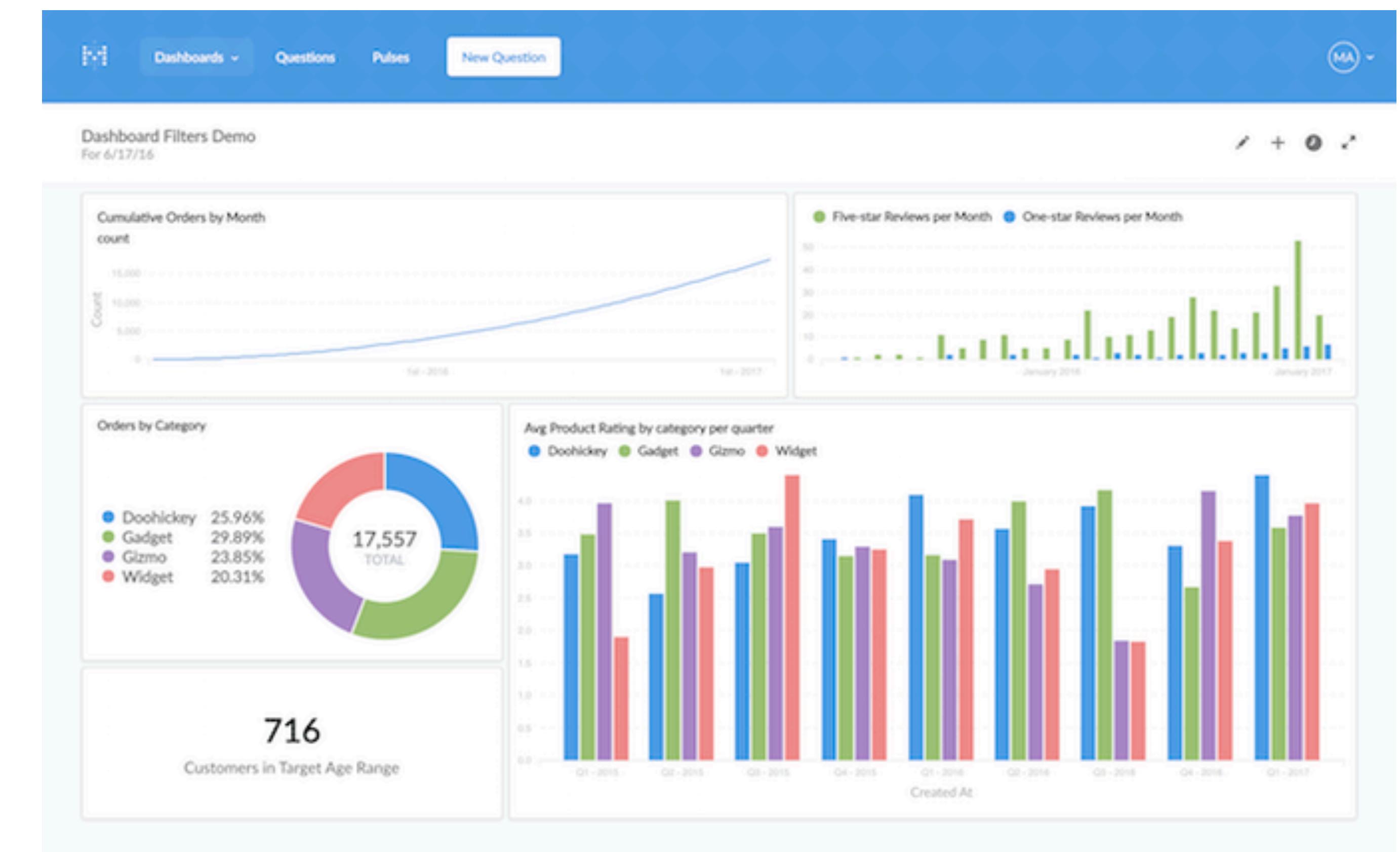
MICROSOFT POWER BI

- <https://powerbi.microsoft.com/en-us/desktop/>
- Otro de los productos líderes del mercado
- Muy buena integración con herramientas de Microsoft
- No es sencillo acceder a algunas fuentes de datos



METABASE

- <https://www.metabase.com/>
- Se originó como un proyecto de código abierto
- Sencillo de configurar, pero con menos opciones que otras alternativas



AWS QUICKSIGHT

- <https://aws.amazon.com/quicksight/>
- La solución que mejor se integra con una empresa que use AWS para almacenar los datos o en sus procesos ETL
- Visualizaciones menos desarrolladas que otras herramientas como Tableau



GRAFANA

- <https://grafana.com/>
- También es una aplicación cloud/SaaS
- Buenas herramientas de visualización



CÓDIGO ABIERTO

- Hay diferentes paquetes Python para crear dashboards. Por ejemplo:
 - Ipywidgets
 - Voila
 - Dash by Plotly
 - Streamlit

IPYWIDGETS

- Ver ejemplo en ipywidgets.ipynb

jupyter ipywidgets Last Checkpoint: 6 minutes ago (unsaved changes) Logout Trusted Python 3 (ipykernel) O

In [1]:

```
1 # ! pip install ipywidgets
2 #
3 # Puede ser necesario
4 # ! pip install seaborn
5 # Desde linea de comandos
6 # jupyter nbextension enable --py widgetsnbextension
```

In [2]:

```
1 import seaborn as sns
2 import matplotlib.pyplot as plt
3 titanic = sns.load_dataset('titanic')
4 titanic.head()
```

Out[2]:

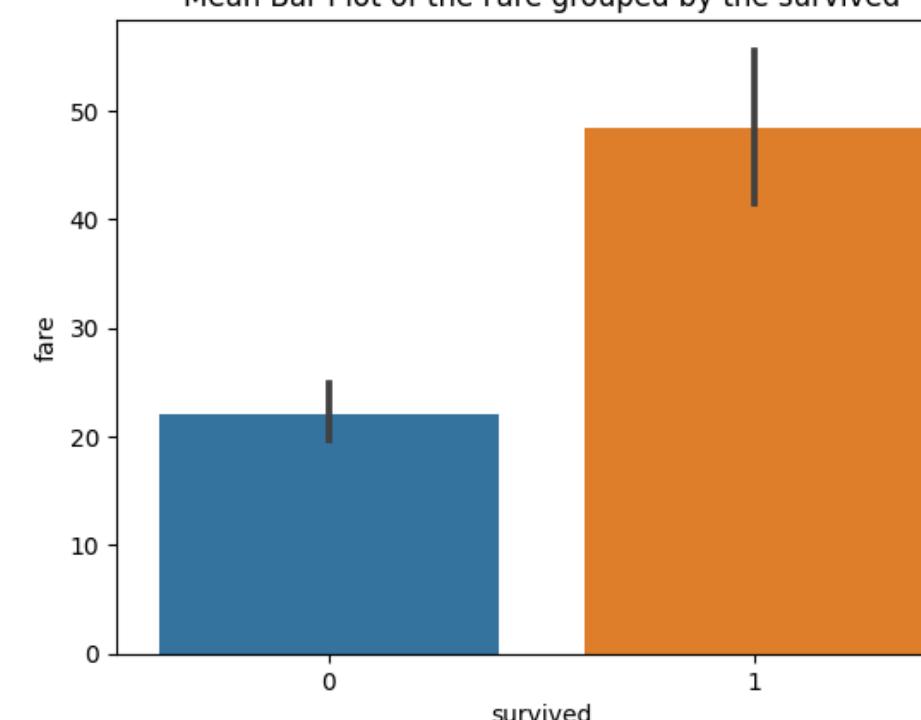
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no

In [3]:

```
1 #Creating the interactive dashboard
2 from ipywidgets import interact
3 @interact
4 def create_fare_plot(col = titanic.drop(['fare', 'age'], axis=1).columns):
5     sns.barplot(data = titanic, x = col, y ='fare')
6     plt.title(f'Mean Bar Plot of the Fare grouped by the {col}')
```

col survived ▾

Mean Bar Plot of the Fare grouped by the survived



The figure is a bar chart titled "Mean Bar Plot of the Fare grouped by the survived". The x-axis is labeled "survived" and has two categories: 0 and 1. The y-axis is labeled "fare" and ranges from 0 to 50 with increments of 10. There are two bars: one for "survived" value 0 with a fare of approximately 22, and one for "survived" value 1 with a fare of approximately 48. Both bars have black vertical error bars extending above and below the top of the bar.

In []: 1

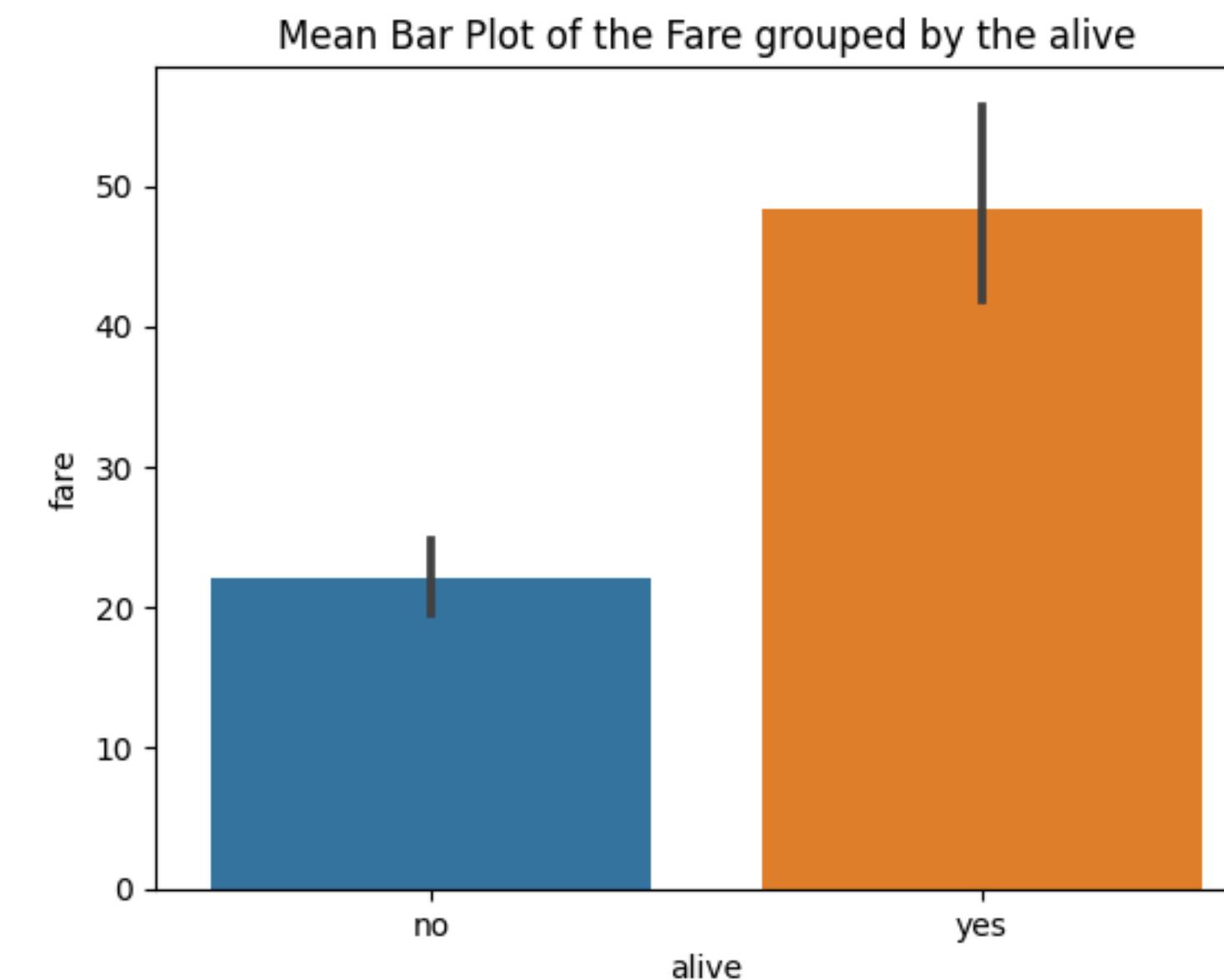
VOILA

- Voila-dashboards (<https://github.com/voila-dashboards/voila>) es un paquete que transforma un notebook de python en un dashboard web
- ejecutar voila ipywidgets.py

Ipywidgets

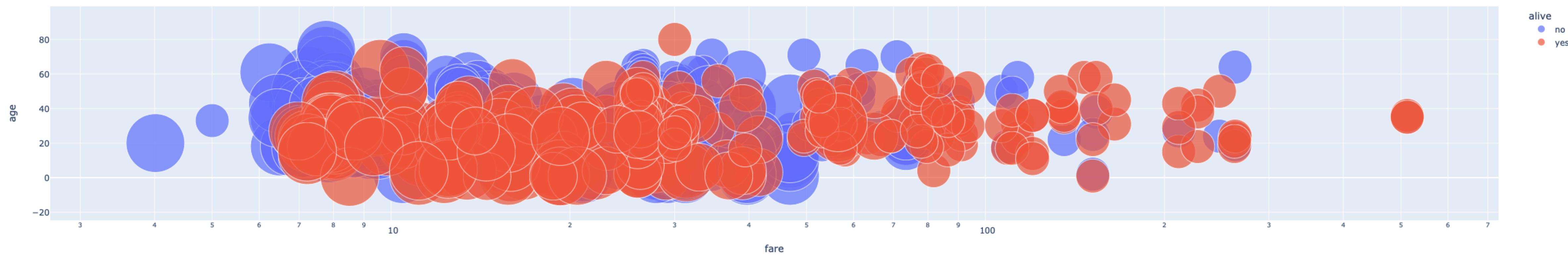
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

col alive ▾



DASH

Titanic Dashboard



- ejecutar ejemplo-dash.py desde línea de comandos o desde Visual Studio Code

STREAMLIT

Titanic Dashboard

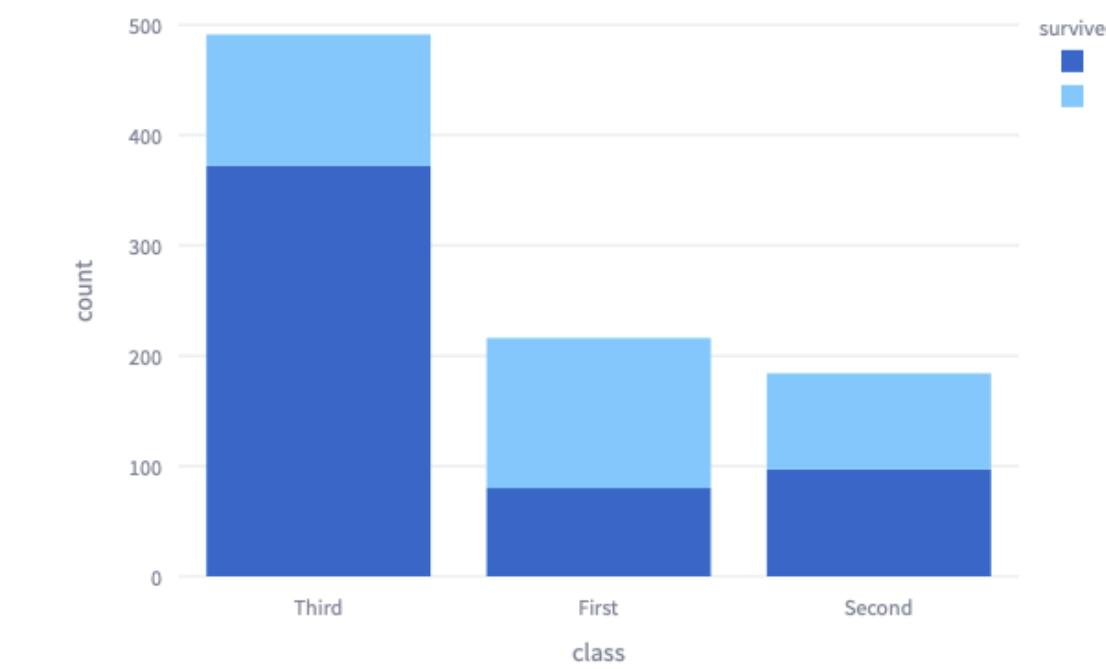
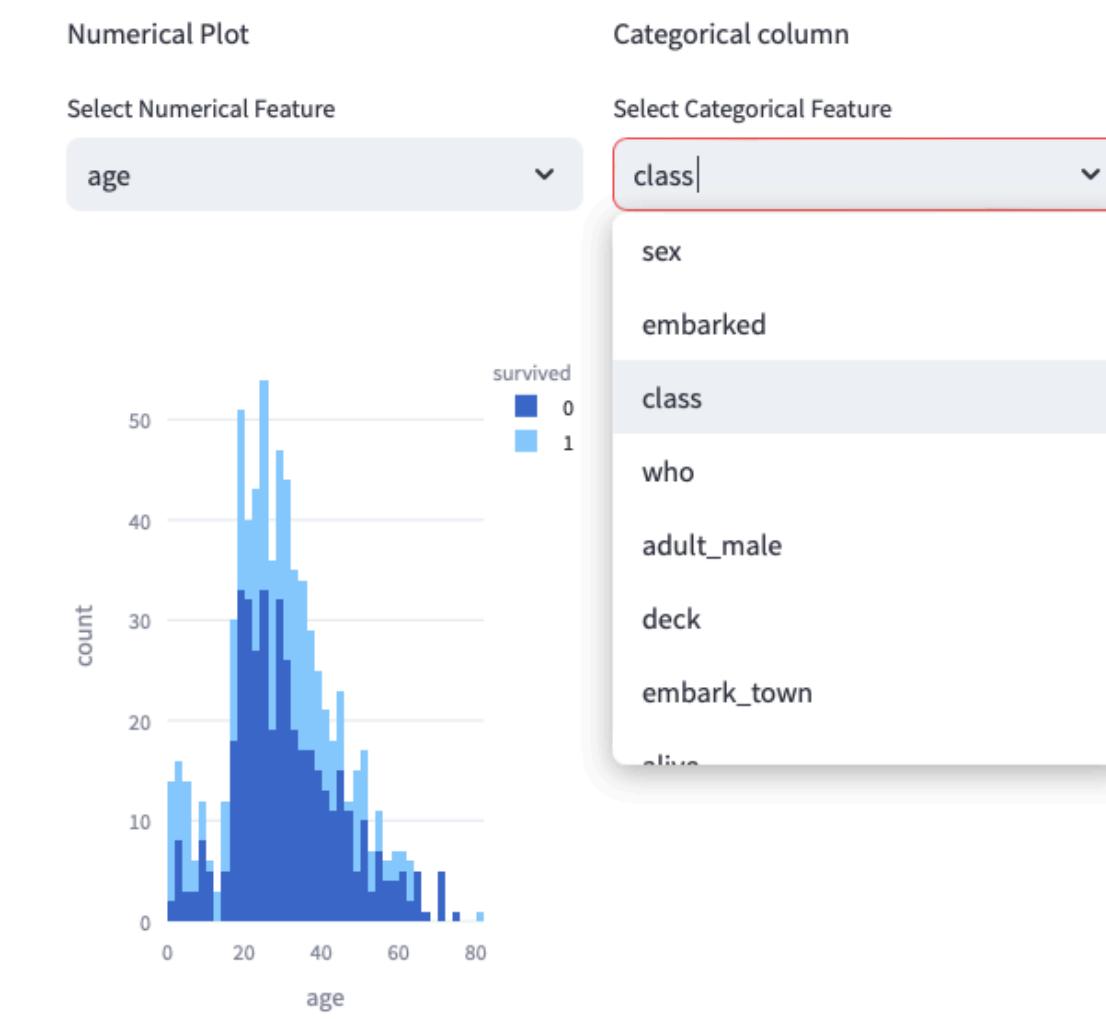
Dataset

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class
0	0	3	male	22	1	0	7.25	S	Third
1	1	1	female	38	1	0	71.2833	C	First
2	1	3	female	26	0	0	7.925	S	Third
3	1	1	female	35	1	0	53.1	S	First
4	0	3	male	35	0	0	8.05	S	Third
5	0	3	male	None	0	0	8.4583	Q	Third
6	0	1	male	54	0	0	51.8625	S	First
7	0	3	male	2	3	1	21.075	S	Third
8	1	3	female	27	0	2	11.1333	S	Third
9	1	2	female	14	1	0	30.0708	C	Second

Data Numerical Statistic

	survived	pclass	age	sibsp	parch	fare
count	891	891	714	891	891	891
mean	0.3838	2.3086	29.6991	0.523	0.3816	32.2042
std	0.4866	0.8361	14.5265	1.1027	0.8061	49.6934
min	0	1	0.42	0	0	0
25%	0	2	20.125	0	0	7.9104
50%	0	3	28	0	0	14.4542
75%	1	3	38	1	0	31
max	1	3	80	8	6	512.3292

Data Visualization with respect to Survived



➤ ejecutar streamlit run ejemplo-streamlit.py