

Guía Práctica: Cómo Navegar en un Data Lake

A person in a boat navigating through a digital landscape with data visualizations. The background is a light blue gradient with faint, glowing data visualizations like bar charts and line graphs. A person in a white shirt and dark pants is standing in a small boat, holding a long wooden oar. The boat is positioned in the center-right of the frame, and the person is looking towards the left. The overall aesthetic is clean and modern, with a focus on data and navigation.

Convierte el caos de información en decisiones estratégicas

¿Te sientes ahogado entre miles de datos? ¿Tienes dashboards, reportes, hojas de Excel... pero no claridad?

No estás solo. Estamos viviendo en la era del "exceso de datos" y no siempre tenemos las herramientas ni la mentalidad para organizarlos.

Esta guía es una **brújula para navegar en tu Data Lake**. Con pasos sencillos y aplicables, aprenderás a **flotar, remar y avanzar con propósito**. Porque no se trata de nadar más rápido... **Se trata de saber hacia dónde vas.**

K por Keisa Avila



⚠️ Mini Diagnóstico: ¿Te estás ahogando en tu Data Lake?

Revisa este checklist inicial y marca lo que aplica en tu caso:

- ☐ Tengo más de 5 fuentes de datos (CRM, ERP, Excel, sistemas locales, nube...) y no tengo un inventario claro de cuáles son confiables.
- ☐ Me solicitan reportes urgentes, pero rara vez sé **qué decisión estratégica** se tomará con ellos.
- ☐ Veo dashboards llenos de métricas, pero no existe un **indicador central** que todos reconozcan como “la verdad”.
- ☐ Paso más tiempo limpiando, uniendo o corrigiendo datos que analizando y generando insights.
- ☐ Diferentes áreas presentan cifras distintas para el mismo KPI (ej. ventas, clientes activos, churn).
- ☐ Los reportes parecen cumplir un requisito, pero no se usan para tomar decisiones reales.
- ☐ Me cuesta transformar números en mensajes claros que conecten con la dirección o con los equipos de negocio.

Si marcaste 2 o más puntos, estás oficialmente en modo “ahogo”.

La buena noticia es que no estás solo/a: a muchos equipos les pasa. La diferencia está en **aprender a remar con una estrategia de datos clara** para llegar a tierra firme. 🚣



Pasos para Empezar a Navegar en tu Data Lake

A continuación te presentamos los cinco pasos fundamentales que te ayudarán a transformar el caos de información en decisiones estratégicas claras. Cada paso está diseñado para guiarte progresivamente desde la limpieza de datos hasta la narración efectiva de historias basadas en ellos.

Beneficios técnicos clave

- **Escalabilidad automática:** desde GB hasta PB sin rediseñar tu arquitectura.
- **Costos optimizados:** pagas por almacenamiento y consultas, no por servidores o licencias fijas.
- **Seguridad de nivel bancario:** cifrado en reposo y en tránsito, IAM granular, auditorías automáticas.
- **Velocidad en el time-to-insight:** en lugar de esperar semanas por infraestructura, puedes levantar consultas en minutos.

Paso 0: Llevar tus Datos a la Nube

Antes de limpiar, medir, visualizar y narrar, existe un paso crítico: **definir cómo migrar y almacenar tus datos en un Data Lake en la nube.**

On-Premise vs. Nube

- **On-Premise**
 - Servidores físicos en la empresa.
 - Ventaja: control total sobre hardware y configuraciones.
 - Desventaja: escalabilidad limitada, altos costos de mantenimiento, riesgo de fallas locales.
- **Nube**
 - Infraestructura administrada por un proveedor (AWS, Azure, GCP).
 - Ventaja: elasticidad (pago por uso), seguridad gestionada, disponibilidad 24/7, despliegue global.
 - Desventaja: dependencia de conectividad y buen gobierno de datos.





Flujo técnico recomendado (ejemplo con AWS)



1. Ingesta de Datos

Migra tus datos a la nube. Usa AWS DMS o Data Wrangler para cargas periódicas (Batch) y Kinesis Data Streams o Kafka en MSK para flujos en tiempo real (Streaming).



2. Almacenamiento en Data Lake (Amazon S3)

Organiza los datos en zonas: Raw (crudo), Cleansed (limpio) y Curated (para análisis). Utiliza formatos eficientes como Parquet u ORC para optimizar costos y rendimiento.



3. Catálogo y Gobernanza

AWS Glue crea un catálogo de metadatos centralizado para consultas SQL. AWS Lake Formation gestiona los permisos de acceso a los datos y columnas específicas.



4. Consulta y Análisis

Usa Amazon Athena para ejecutar SQL directamente sobre S3. Para análisis avanzados o procesamiento masivo, considera Amazon Redshift (almacén columnar) o Amazon EMR (Spark).



5. Visualización

Conecta herramientas como Amazon QuickSight (integración nativa con Athena/Redshift), Power BI o Tableau para crear dashboards interactivos que consumen los datos sin necesidad de moverlos de la nube.

Paso 1: Identifica y prioriza tus fuentes de datos

Antes de analizar, necesitas saber de dónde viene la información y cuáles son confiables. Mapear tus fuentes (ERP, CRM, hojas de cálculo, plataformas en la nube) y definir cuáles tienen mayor impacto en las decisiones de negocio es el punto de partida para no ahogarte en duplicados o datos irrelevantes.

¿Qué datos realmente importan?

🎯 Empieza eliminando ruido. Céntrate en los datos que están conectados a **una pregunta o decisión concreta**.

- ? ¿Este dato ayuda a tomar una decisión?
- ↻ ¿Este dato se actualiza con frecuencia confiable?
- 👤 ¿Es comprendido fácilmente por mi equipo?

📝 Anota 3 datos que sí necesitas:



Paso 2: Limpieza y estandarización

Un Lago de Datos solo es útil si su agua está clara. Corregir errores, estandarizar formatos y eliminar duplicados es una tarea que muchos evitan, pero que se convierte en el cimiento de todo análisis confiable. Sin este paso, cualquier modelo, dashboard o reporte estará construido sobre terreno inestable, llevando a decisiones erróneas y pérdida de recursos.

La limpieza y estandarización de datos son esenciales para transformar el, "ruido" en información valiosa y accionable.

Problemas comunes en tus datos:

Duplicados Registros idénticos que sesgan los recuentos y promedios.		Inconsistencias Misma información presentada de diferentes maneras (ej. "USA", "U.S.", "Estados Unidos").
Valores Nulos Campos vacíos que impiden análisis completos y precisos.		Formatos Incorrectos Fechas como "MM/DD/AAAA" y "AAAA-MM-DD", o números con texto.

Pasos clave para una limpieza efectiva:

	
1. Perfilado de Datos Identifica la calidad, estructura y anomalías. Usa herramientas para visualizar distribuciones y patrones.	2. Estandarización Normaliza formatos (ej. fechas, unidades de medida) y corrige valores inconsistentes.
	
3. Eliminar duplicado Detecta y elimina registros duplicados basándote en criterios clave.	4. Validación Establece reglas para asegurar que los nuevos datos cumplan con los estándares definidos.

🔗 Considera tu herramienta actual:_____ ¿Te ayuda a mantener la claridad en tu lago de datos o te genera más confusión?

Paso 3: Diseña métricas y KPIs accionables

No basta con mostrar números. Define indicadores que respondan a preguntas reales del negocio: crecimiento de clientes, costos evitados, nivel de satisfacción. Cada KPI debe tener un propósito y una audiencia clara, evitando la tentación de medir por medir. Un KPI efectivo no solo informa, sino que impulsa a la acción y ayuda a evaluar el progreso hacia los objetivos estratégicos.

Claves para crear KPIs efectivos: El enfoque S.M.A.R.T.



Específico (Specific)

Debe ser claro y enfocado. ¿Qué exactamente quieres medir y por qué es importante para tu objetivo?



Medible (Measurable)

Necesita cuantificación. ¿Cómo se medirá el progreso y con qué unidad?



Alcanzable (Achievable)

Aunque ambicioso, debe ser realista. ¿Es posible lograr el valor objetivo con los recursos disponibles?



Relevante (Relevant)

Debe estar alineado con los objetivos estratégicos. ¿Contribuye directamente a las metas generales de la empresa?



Con Plazo Definido (Time-bound)

Debe tener un marco de tiempo claro. ¿Cuándo se espera alcanzar este objetivo?

Ejemplos de KPIs por área de negocio:

Marketing

- Costo de Adquisición de Cliente (CAC)
- Retorno de la Inversión en Marketing (ROMI)
- Tasa de Conversión de Leads
- Engagement en Redes Sociales

Ventas

- Valor de Vida del Cliente (LTV)
- Ticket Promedio
- Tasa de Cierre de Ventas
- Ingresos por Vendedor

Producto/Servicio

- Tasa de Retención de Clientes
- Net Promoter Score (NPS)
- Tiempo de Actividad (Uptime)
- Adopción de Nuevas Características

Operaciones

- Costo por Unidad Producida
- Tiempo de Ciclo del Pedido
- Tasa de Defectos
- Eficiencia Energética



Criterios para evaluar un KPI útil

- **Accionable:** ¿La métrica te permite tomar decisiones o cambiar una estrategia?
- **Alineado:** ¿Contribuye directamente a un objetivo estratégico del negocio?
- **Entendible:** ¿Es fácil de comprender por todos los stakeholders relevantes?
- **Oportuno:** ¿Puedes obtener los datos a tiempo para tomar decisiones informadas?
- **Fiable:** ¿Los datos utilizados para calcular el KPI son precisos y consistentes?
- **Equilibrado:** ¿El KPI considera múltiples perspectivas (ej. financieras y no financieras)?

Mejores prácticas para la implementación de KPIs:

1 Involucra a los stakeholders

Asegúrate de que los equipos clave participen en la definición y entendimiento de los KPIs.

2 No te excedas

Limita el número de KPIs a los más críticos para evitar la "parálisis por análisis". Unos pocos KPIs bien definidos son más efectivos que docenas de ellos.

3 Automatiza la recolección

Utiliza herramientas para recopilar y visualizar datos de forma automática, reduciendo el esfuerzo manual y los errores.

4 Revisa y adapta regularmente

Los objetivos del negocio cambian, y tus KPIs también deberían hacerlo. Evalúa su relevancia periódicamente.

5 Comunica y visualiza claramente

Presenta los KPIs de manera que sean fáciles de entender y consumir, utilizando visualizaciones adecuadas para cada tipo de dato.

Paso 4: Construye visualizaciones con intención

Los dashboards no son murales decorativos: deben ser brújulas de navegación. Una visualización efectiva resalta patrones, tendencias y riesgos, facilitando la interpretación. Aquí la clave es traducir datos complejos en representaciones visuales que cualquier persona en la organización pueda entender y usar.

La forma en que visualizas tus datos es tan importante como las métricas que eliges. Un diseño intencional no solo presenta los datos, sino que también guía al observador hacia los insights más relevantes.

Principios Clave del Diseño de Visualizaciones

Jerarquía Visual

Utiliza el tamaño, el color, el contraste y la posición para dirigir la atención del usuario a los elementos más importantes primero. Los datos críticos deben ser prominentes, mientras que la información de soporte puede ser menos destacada.

Psicología del Color

El color es una herramienta poderosa. Úsalo con moderación y significado. Los colores cálidos pueden indicar advertencia o urgencia, mientras que los fríos sugieren calma o menos importancia. Evita el uso excesivo de colores y considera la accesibilidad (daltonismo) usando paletas aptas para todos.

Simplicidad y Claridad

Elimina el desorden innecesario (gráficos 3D, sombras, exceso de texto). Cada elemento debe servir a un propósito. Asegúrate de que las etiquetas, leyendas y títulos sean claros y concisos.

Contexto

Proporciona siempre suficiente contexto para que la visualización sea interpretable. Esto incluye unidades de medida, periodos de tiempo, puntos de referencia o comparaciones con objetivos.

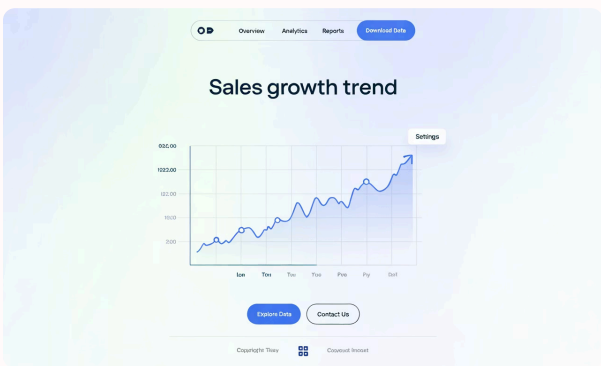
Tipos de Gráficos y Cuándo Usarlos

Elegir el gráfico adecuado es fundamental para comunicar tu mensaje de forma efectiva. Aquí te presentamos algunos de los más comunes y sus aplicaciones:

Tendencias y Evolución Temporal

Para mostrar cómo cambian los datos a lo largo del tiempo.

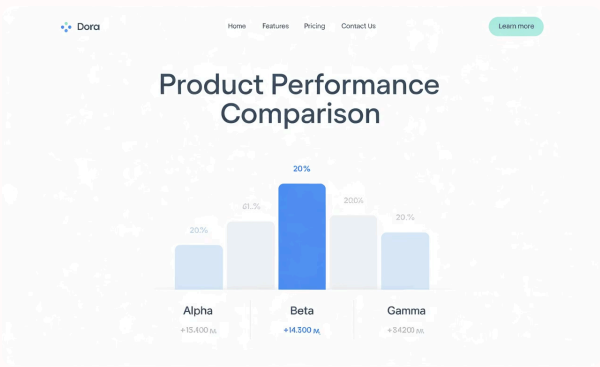
→ **Gráficos de Líneas:** Ideales para visualizar el rendimiento de una o varias métricas a lo largo de un período continuo (ej., ventas mensuales, crecimiento de usuarios).



Comparación de Categorías

Para contrastar valores entre diferentes elementos.

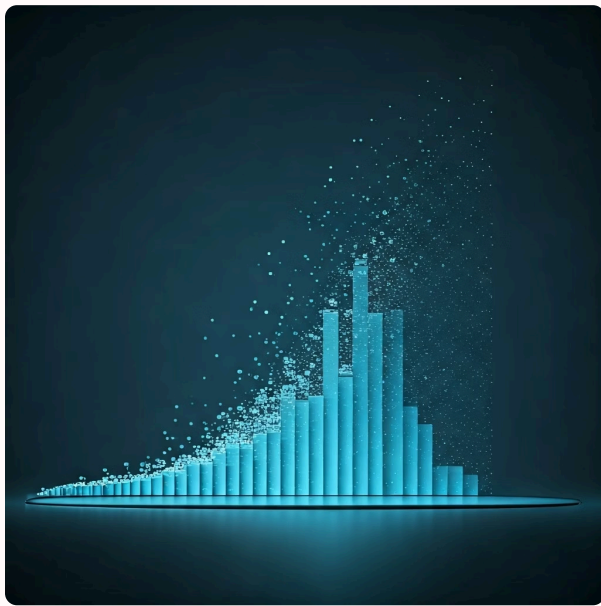
→ **Gráficos de Barras:** Excelentes para comparar cantidades discretas entre categorías (ej., ventas por producto, usuarios por región). Las barras horizontales son útiles para nombres de categorías largos.



Distribución y Relaciones

Para entender cómo se distribuyen los datos o cómo se relacionan variables.

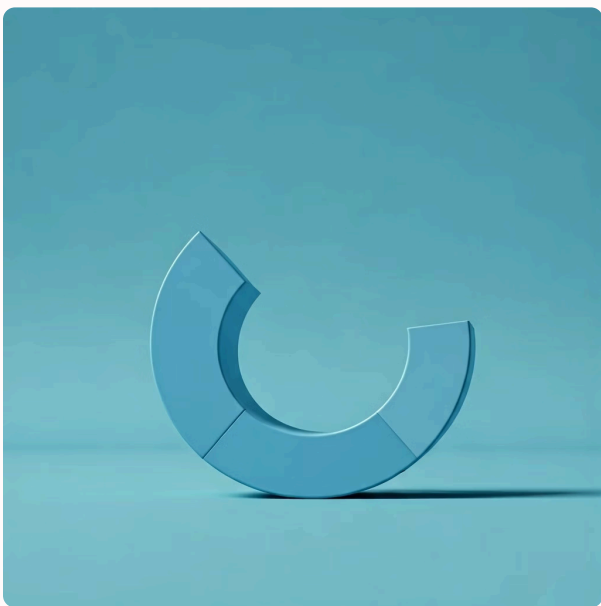
→ **Gráficos de Dispersión (Scatter Plot):** Muestran la relación entre dos variables numéricas, revelando correlaciones o clústeres (ej., gasto en marketing vs. ingresos).



Composición y Proporciones

Para mostrar cómo las partes contribuyen a un todo.

→ **Gráficos de Tarta/Donut:** Útiles para mostrar porcentajes o proporciones de un todo (ej., cuota de mercado por competidor), pero úsalos con precaución y para pocas categorías.



Jerarquía y Magnitud

Para visualizar estructuras jerárquicas y el tamaño relativo de sus componentes.

→ **Treemaps:** Muestran datos jerárquicos como un conjunto de rectángulos anidados, donde el tamaño del rectángulo representa la magnitud (ej., contribución de departamentos a los ingresos totales).



Métricas Únicas y Progreso

Para destacar el estado de una métrica clave en relación con un objetivo.

→ **Gráficos de Velocímetro (Gauge Charts):** Indican el progreso hacia un objetivo o el estado actual de una métrica (ej., progreso de KPI hacia el objetivo mensual).





Mejores Prácticas para Dashboards Efectivos

01

Define el Propósito y la Audiencia

Antes de empezar, ten claro qué preguntas debe responder el dashboard y quién lo utilizará. Esto dictará el nivel de detalle, el tipo de gráficos y la terminología.

02

Cuenta una Historia

Organiza tus visualizaciones de manera lógica para guiar al usuario a través de una narrativa. Empieza con un resumen de alto nivel y permite la exploración de detalles.

03

Diseño Consistente

Mantén una coherencia en el uso de colores, fuentes, iconos y disposición. Un diseño consistente reduce la carga cognitiva y mejora la experiencia del usuario.

04

Hazlo Interactivo

Incorpora filtros, segmentadores y opciones de desglose (drill-down) para que los usuarios puedan explorar los datos según sus necesidades y obtener insights más profundos.

05

Optimiza para el Rendimiento

Asegúrate de que tu dashboard cargue rápidamente y sea responsive en diferentes dispositivos. Un dashboard lento frustrará a los usuarios, independientemente de lo bueno que sea el contenido.

qué visualizaciones usaré para cada uno?

Errores Comunes a Evitar



Sobrecargar el Dashboard

Demasiada información en un solo panel confunde. Prioriza los KPIs más importantes y crea dashboards específicos para diferentes necesidades.



Elegir el Gráfico Incorrecto

Usar un gráfico de tarta para muchas categorías o un gráfico de barras para tendencias temporales puede distorsionar la información y dificultar la interpretación.



Escalas Engañosas

Manipular el eje Y para exagerar o minimizar cambios puede llevar a interpretaciones erróneas. Siempre empieza el eje Y desde cero para comparaciones de magnitud.



Mal Uso del Color

Colores sin significado, combinaciones ilegibles o ignorar la accesibilidad pueden hacer que tu visualización sea ineficaz o incluso ofensiva.

Adaptando Visualizaciones para Diferentes Audiencias

No todos los usuarios necesitan el mismo nivel de detalle o tipo de información. Personaliza tus dashboards para resonar con cada grupo:

Ejecutivos

Necesitan un resumen conciso con KPIs de alto nivel, tendencias clave y alertas. Se enfocan en la visión general y las implicaciones estratégicas. Menos detalles, más insights accionables.

Analistas/Gerentes

Requieren datos más granulares para entender el "por qué" detrás de las tendencias. Necesitan capacidad para explorar, filtrar y comparar diferentes segmentos de datos.

Operaciones/Equipos

Buscan métricas en tiempo real que los ayuden en sus tareas diarias, monitoreando el rendimiento y la eficiencia. El foco está en la acción inmediata.

Consejo: No trates de impresionar con herramientas complejas, **impresiona con claridad, relevancia y la facilidad de tus insights para la toma de decisiones.**

✍️ Mis principales KPIs para [mi área/proyecto] son:_____ ¿Cómo los estoy monitoreando y

Paso 5: Convierte datos en historias que inspiren acción

El viaje culmina cuando los datos dejan de ser cifras y se transforman en narrativas que generan impacto. Contar historias con datos no significa adornar, sino darles un contexto que resuene con la audiencia: qué está pasando, por qué importa y qué decisión se debe tomar.

La Esencia del Storytelling con Datos

El storytelling con datos es el arte de comunicar insights complejos de manera clara y atractiva, utilizando una estructura narrativa que guía a la audiencia desde un punto de partida (la situación actual) hasta un punto de llegada (la acción recomendada). No se trata solo de mostrar gráficos bonitos, sino de crear un viaje cognitivo y emocional que permita a tu audiencia comprender la importancia de tus hallazgos y actuar en consecuencia.

Elementos Clave de una Narrativa de Datos Efectiva

Contexto

Sitúa los datos en un marco comprensible. ¿Cuál es el problema o la oportunidad? ¿Qué antecedentes son relevantes? Ayuda a la audiencia a entender el punto de partida.

Conflicto

Introduce el desafío o el cambio que revelan tus datos. ¿Qué está impidiendo el progreso o qué tendencia inesperada ha surgido? Genera tensión y captura la atención.

Análisis/Revelación

Presenta tus hallazgos clave. Aquí es donde los datos hablan. Usa visualizaciones claras para mostrar lo que has descubierto, las causas y los patrones.

Resolución/Oportunidad

Ofrece una solución o una dirección basada en tus insights. ¿Qué puedes hacer al respecto? ¿Cómo pueden los datos guiar el camino hacia una mejora?

Llamada a la Acción

Indica claramente qué debe hacer la audiencia. ¿Qué decisión se espera? ¿Qué pasos concretos se deben tomar? Es el propósito final de tu historia.



Estructura Narrativa: El Viaje del Héroe de tus Datos

Una historia de datos efectiva a menudo sigue una estructura similar al "Viaje del Héroe", pero adaptada a la información. Considera estos pasos:

01

El Estado Actual

Comienza describiendo el punto de partida: qué sabemos, cómo están las cosas y cuál es la situación inicial (el "mundo ordinario").

03

La Exploración de Datos

El "viaje" o la "búsqueda" donde exploras los datos, realizas análisis, pruebas hipótesis y descubres los insights clave. Aquí es donde los gráficos y visualizaciones brillan.

05

El Nuevo Estado

Muestra cómo esta revelación cambia la perspectiva y qué implica para el futuro. Cómo sería el "mundo después" de aplicar los insights.

02

El Incidente Desencadenante

Introduce el "llamado a la aventura": los datos que revelan una desviación, un problema o una oportunidad. Algo que interrumpe el estado actual.

04

La Gran Revelación

El "clímax" de tu historia: la conclusión más importante que tus datos te permiten hacer. El "ajá" o el momento de la verdad.

06

La Recompensa y la Acción

Termina con la "recompensa": las recomendaciones accionables que derivan de tu análisis, y el impacto que se espera lograr. Qué se debe hacer ahora.

Frameworks para Construir Historias de Datos

Existen diversas metodologías para estructurar tu narrativa. Aquí dos ejemplos:

SCQA (Situation, Complication, Question, Answer)

Un método muy usado en consultoría: **Situación** (el contexto estable), **Complicación** (lo que perturbó la situación), **Pregunta** (la duda que surge de la complicación) y **Respuesta** (tu insight clave y recomendación).

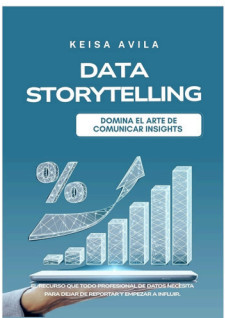
"What? So What? Now What?"

Un enfoque más directo: **¿Qué está sucediendo?** (los datos y hechos), **¿Y qué?** (la importancia o implicación de esos hechos) y **¿Ahora qué?** (las acciones a tomar basadas en esas implicaciones).

Lectura Recomendada

Si quieres llevar tu comunicación con datos al siguiente nivel, en mi libro **Data Storytelling: Domina el arte de comunicar insights** encontrarás:

- Ejemplos reales de storytelling aplicado a negocios.
- Métodos prácticos para estructurar narrativas de datos.
- Ejercicios para transformar métricas en decisiones accionables.



Amazon.com



Data Storytelling: Domina el arte de comunicar insights (Spanish Editio...

Buy Data Storytelling: Domina el arte de comunicar insights (Spanish Edition):
Read Kindle Store Reviews – Amazon.com

Errores Comunes en el Storytelling con Datos



Falta de Pregunta Central

Iniciar sin una pregunta clara a responder hace que la historia carezca de dirección y propósito.



Sobreabundancia de Datos

Incluir demasiados datos irrelevantes o complejos abruma a la audiencia y diluye el mensaje principal.



No Conocer a la Audiencia

No adaptar el nivel de detalle y el tono a quién escuchará la historia. Esto lleva a desconexión y falta de interés.



Omitir la Llamada a la Acción

Presentar insights sin una recomendación clara o un siguiente paso deja a la audiencia sin saber qué hacer.

Del Caos al Propósito

Un Data Lake no es solo un repositorio de información. Es un reflejo de cómo una organización concibe el valor de sus datos.

- Mal diseñado, se convierte en un pantano: opaco, difícil de navegar, lleno de duplicados e inconsistencias.
- Bien gestionado, se transforma en un motor de innovación: accesible, confiable y preparado para responder preguntas que todavía no sabemos que tendremos que hacer.

Navegar un Lago de Datos no se trata de tener la última tecnología, sino de tener una **visión clara**:

- ¿Qué preguntas queremos responder?
- ¿Qué decisiones necesitamos impulsar?
- ¿Cómo aseguramos que cada dato sea útil y confiable?

El verdadero desafío no está en almacenar más, sino en **dar sentido** a lo que almacenamos.

Porque los datos, como el agua, solo generan vida cuando fluyen y se usan con propósito.

Un Data Lake puede hundirte... o puede convertirse en el mapa que te guíe al futuro.

La diferencia está en **cómo decides navegarlo**.





¡Ahora es tu Turno!

Has recorrido los pasos esenciales para navegar un Lago de Datos: identificar, limpiar, medir, visualizar y narrar. Ahora llega el momento de poner en práctica estos aprendizajes y transformar la abundancia de datos en decisiones estratégicas.

Tu reto no es almacenar más información, sino **darle propósito**: que cada dato fluya, se conecte y genere valor real en tu organización.

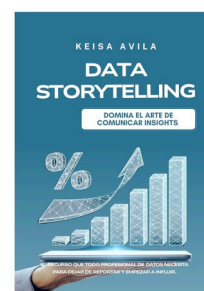
El verdadero poder de un Data Lake no está en su tamaño, sino en la **claridad con la que lo gestionas** y en la confianza que inspira en quienes toman decisiones.

Cada consulta, cada dashboard y cada insight puede ser un remo que te acerque a un destino más claro. La diferencia está en **cómo eliges navegar tu lago de datos a partir de hoy**.



Conecta y Sigue Aprendiendo

Si quieres dar un paso más y aprender cómo contar historias que inspiren a partir de tus hallazgos, te recomiendo mi libro sobre ***Data Storytelling: Domina el arte de comunicar insights***, un complemento ideal para profundizar en este camino.



Amazon.com



Data Storytelling: Domina el art...

Buy Data Storytelling: Domina el arte de comunicar insights (Spanish...

- Visita mi sitio web: keisaavila.net

Si esta guía te ayudó, comparte tus aprendizajes conmigo en LinkedIn. Estoy creando una comunidad de profesionales que transforman datos en decisiones, y me encantaría que seas parte de ella.

- Encuéntrame en LinkedIn: [Keisa Avila](https://www.linkedin.com/in/keisaavila)

Gracias por tu dedicación y compromiso con la excelencia. ¡Juntos podemos construir un futuro más informado y eficiente!