Caso de estudio: Mega-pek internacional

- problema: caída en ingresos de taquilla (5 de 6 películas apenas cubrieron costos)

- Objetivo: realizar un exploratorio de data analysis para replicar exitos pasados

- Metas del EDA

  - Identificar factores de exito en el primer fin de semana

  - relación presupuesto vs ingresos

  - Generar mas rentables

Las 5 habilidades analíticas en acción

1) Curiosity

- Acción: hacer preguntas que el dataset no responde a simple vista.

  - Ejemplo mega-pik: "¿Existe relación entre el presupuesto y la recaudación de la noche de estreno o necesitamos datos adicionales?

  - $\boxed{\text{Clave}}$ impulsa a descubrir información inesperada

Understanding context

- Acción: identificar factores externos que afectan los números

- Factores críticos: temporadas, días festivos, competencia

- Ejemplo: las películas familiares rinden mas en vacaciones pero los fechas de vacaciones varian en cada país

- Impacto: ayuda a filtrar variables y encontrar insights significativos

Technical Mindset

- Acción: Aborda problemas de forma sistemática y lógica.

- Uso: limpieza, organización y preparación de datos

- Enfoque: descompone problemas complejos en partes pequeñas y manejables

Data Design

- Acción: organiza la información para facilitar el análisis

- Ejemplo: Estructuras los datos por revenue y luego por genre para poder ver qué es más rentable

- Propósito: Que la estructura misma de los datos revele patrones

Data strategy

- Acción: gestión de personas, procesos y herramientas

- Selección de herramientas:

  - Google Sheets / Excel: para dashboards simples con pocas columnas

  - Tableau: para dashboards con robustez y actualizaciones en tiempo real

- Asignación de recursos: priorizar análisis que afecten directamente los ingresos del próximo trimestre

SQL

```sql
SELECT
    genre,
    AVG(opening_weekend_revenue -
budget) AS avg_profit,
    COUNT(*) AS total_movies
FROM mega_pik_data
GROUP BY genre
ORDER BY avg_profit DESC;
```

Root Cause Analysis

Tecnica para identificar la razon fundamental de un problema

Objetivo: Eliminar la causa raiz para evitar que el problema se repita

Tecnica: realizar la pregunta ¿por que? de forma reiterativa hasta que la respuesta revele el origen del fallo

Caso: Quejas por entregas deficientes

1 Why #1 ¿por que hay quejas? los productos llegan dañados

2 Why #2 por que llegan dañados

El empaque no es el adecuado

3 Why #3 por que el empaque falla

Los empacadores no estan bien capacitados

4 Why #4 por qué no están capacitados?

el 35% del personal es nuevo y no ha tomado el curso

5 Why #5 por qué no han tomado el curso?

Root cause: HR está rediseñando el programa y solo entregó una guía rápida insuficiente

Solución: Sincronización inmediata de información entre el equipo de software y mantenimiento

Tip: si se llega a una causa y no se puede intervenir retroceder un paso y buscar una donde sí

investigación: no adivinar las respuestas. Reunirse con los stakeholders y revisar los datos

```sql
-- Identificar el pico de errores por fecha para iniciar los
5 Whys
SELECT
    DATE(error_timestamp) AS error_date,
    COUNT(error_id) AS total_defects
FROM pump_quality_logs
GROUP BY error_date
ORDER BY total_defects DESC
LIMIT 5;
```

Analytical skills: Cualidades para usar hechos y resolver problemas.
Analytical thinking : Proceso de identificar, definir y resolver problemas mediante datos de forma organizada [cite: 2025-11-12].

Technical mindset : Capacidad de desglosar problemas en partes pequeñas y trabajar de forma lógica.

Context : La condición o entorno en el que algo sucede. Vital para dar sentido a los

Data: Colección de hechos.

Data analysis: Proceso de recolectar, transformar y organizar datos para sacar conclusiones y predicciones [cite: 2025-12-18].

Data analytics: La ciencia de los datos en su totalidad.

Data science: Campo que usa datos crudos para crear nuevas formas de modelar lo desconocido.

Data-driven decision-making: Uso de hechos reales para guiar la estrategia de negocio.

Data ecosystem: Elementos que interactúan para producir, gestionar, almacenar y compartir datos.

Data design: Cómo se organiza la información (clave para arquitectura de bases de datos) [cite: 2025-11-12].

Data strategy: Gestión de personas, procesos y herramientas en el análisis.

Dataset: Colección de datos que se manipula como una sola unidad.

Gap analysis: Método para evaluar el estado actual vs. el estado futuro deseado.

Root cause: La razón principal por la que ocurre un problema.

Data visualization: Representación gráfica de la información (también llamado Visualization) | 1{

Ciclo de vida del dato

Plan: Que datos se necesitan y quien los maneja

- Capture: recolección de diversas fuentes
- Manage: almacenamiento, mantenimiento y herramientas
- Analyze: Uso de datos para resolver problemas

- analyze: Uso de datos para resolver problemas

- archive: almacenamiento a largo plazo para referencia futura
- Destroy: Eliminación segura de copias y registros

Diferencia

Característica	Data Life Cycle	Data Analysis Process
Objetivo	Gestión del ciclo vital	Resolución de problemas
Enfoque	El dato como "activo"	El dato como "respuesta"
Etapas clave	Archive / Destroy	Share / Act

Versiones por gobiernos y ciencia

- Enfoque: transparencia y preservación histórica

Etapas iniciales: acquire, preserve, publish/share

Actividades transversales: describe, quality management, backup

Instituciones financieras:

Enfoque: auditoría, transformación y reporte

Etapas iniciales

- Quality: validar la calidad del dato

- Transform: Cambiar formatos para utilidad

- Purge: Eliminación definitiva

Analisis:

- Enfoque: profundidad del analisis e interpretacion

- 8 etapas: generacion, collection, processing, storage, management, Analysis, visualizacion interpretacion

```
-- Consultar el "Ciclo de Vida" de una tabla a través de sus
metadatos (PostgreSQL)
SELECT
    table_name,
    create_date,      -- Plan/Capture
    last_update,     -- Manage
    is_archived       -- Archive status
FROM information_schema.tables_metadata
WHERE schema_name = 'logistica_fletes';
```

Regla de oro: Siempre pregunta en qué industria estás burbujeando para ajustar tu ciclo de vida. no es lo mismo borrar datos de un video de tiktok que borrar datos de una transacción bancaria.

Las 6 fases del analisis de datos

este proceso es el estandar que seguiras en cada proyecto y en el Case Study final del programa

1 Ask:

- objetivo: Entender los expectativas y definir el problema
- Acciones clave:
  - Identificar a los stakeholders
  - Definir el estado actual vs el estado ideal
  - Asegurarse de entender el qué, quien, cuando, por qué y cómo

2 prepare

- Objetivo: identificar y localizar los datos necesarios
- Acciones clave

- Decidir que tipos de datos son utiles
- Asegurar la objectivity y evitar el Bias
- Los resultados deben basarse en hechos imparciales

3 process

- Objetivo: Limpiar y refinar los datos
- Acciones Clave
  - Data cleaning
    Elimino errores e inconsistencia
  - Mantener outliers
    que pueden sesgar la información
  - Cambiar datasets
    para tener una vision completa
  - Herramientas: Spreadsheets y SQL

4 Analyze

- Objetivo: Transformar datos en información accionable

  - Accion clave
    - Buscar patrones y tendencias

- Herramientas avanzadas: SQL y el lenguaje de programación R
- Es la fase donde se usan los datos limpios para responder las preguntas iniciales

S Share

Objetivo: interpretar y comunicar los hallazgos

Acciones clave

- Data visualization: El mejor aliado del analista

- Ayudar a los stakeholders a tomar decisiones basadas en datos (Data-driven)
- Usar herramientas como R para crear graficos efectivos

Act

objetivo: poner en practica los insights

Acciones clave:

- Resolver el problema inicial

• En este problema: preparar tu portafolio y realizar el case study para descartar ante reclutadores

```
-- Ejemplo de fase PROCESS: Limpiar
nulos y filtrar outliers en fletes
DELETE FROM logistics_data
WHERE freight_cost IS NULL
   OR freight_cost > (SELECT
AVG(freight_cost) * 5 FROM
logistics_data); -- Filtro simple de
outliers
```

Diferencia clave: > * **Data Life Cycle:** Cómo se gestionan los datos (negocio/IT).

- **Data Analysis Process:** Cómo resuelves un problema específico (tu trabajo diario).

Herramientas clave del analista de datos

Las herramientas se dividen en 3 pilares principales mas lenguajes de programación

1) Spreadsheets
- apps: microsoft excel, google sheets
- Uso principal: recolectar, almacenar, organizar y ordenar información de forma rapida
- Fortalezas
- Identificar patrones iniciales
- Crear visualizaciones rapidas (graficos de barras, lineas)
- ideal para proyectos de datos pequeños o medianos

2 Database y Query Language
- Apps (SQL): My sql, Microsoft SQL server, Big Query
- Definicion: Una database es una colección de datos estructurados en

un sistema computacional

- Funciones de SQL
  - Aislar información específica
  - Seleccionar, crear, ordenar o descargar datos para análisis profundos
  - Manejar volúmenes de datos que harían colapsar a excel

3) Visualization Tools

- Apps: tableau, looker
- propósito: convertir número complejos en una historia visual comprensible para los stakeholders

- Diferencias clave:

  - tableau: famoso por su función drag-and-drop para crear dashboards interactivos
  - looker: se conecta directamente a la base de datos (como big query) para visualizaciones en tiempo real

4) programming language

• R & Python: Se usan para análisis estadístico avanzado, automatizado y visualizaciones complejas que las herramientas estandar no pueden realizar

Herramienta	Cuándo usarla	Volumen de datos
Excel / Sheets	Limpieza rápida y cálculos básicos.	Pequeño / Medio
SQL	Extraer datos de servidores masivos.	Grande / Masivo
Tableau / Looker	Presentar resultados finales a jefes.	N/A (Capa visual)
R / Python	Análisis estadístico y automatización.	Masivo

• SQL: Extraer los costos reales de la base de datos de bigquery

- Excel: Ajustar manualmente unos errores de decimales detectados

- Tableau: Crear un dashboard para que el gerente vea donde se esta perdiendo dinero

```
-- SQL: Extracción rápida en BigQuery
SELECT
    origin_city,
    destination_city,
    SUM(total_cost) AS total_freight_cost
FROM `proyectos_logistica.fletes_2025`
WHERE delivery_status = 'Completed'
GROUP BY 1, 2;
```

Eligiendo la herramienta correcta

El analista no elige una herramienta, sino que las combina según la fase del proyecto

Característica	Spreadsheets [Hojas de Cálculo]	Databases [Bases de Datos]
Acceso	Aplicación de software (Excel/Sheets)	**Query Language** (SQL)
Estructura	Filas y columnas simples	Reglas y relaciones complejas
Organización	Información en celdas	Colecciones complejas (tablas)
Capacidad	Datos limitados (se vuelve lento)	Cantidades masivas de datos
Entrada de Datos	Manual (más propenso a errores)	Estricta y consistente
Usuarios	Generalmente uno a la vez	Múltiples usuarios simultáneos
Control	Controlado por el usuario	**DBMS** [Sistema de Gestión de BD]

Flujo de trabajo híbrido

No es necesario elegir solo una. El analista profesional usa ambos

Importar: si una hoja del cálculo crece demasiado, se migra a una base de datos

Exportar: Se extraen datos maestros de una base de datos vía SQL para analizarlos en una hoja

de calculo o herramientas de BI

Lenguajes de programacion

R y Python: ofrecen un control aun mayor que SQL o excel

uso: Analisis estadistico avanzado, limpieza de datos masivo y visualizacion personalizada que las herramientas estandar no permiten

```
-- SQL: Ejemplo de consulta para extraer datos de la DB
hacia el análisis
SELECT
    carrier_name,
    route_id,
    SUM(cost) AS total_freight_cost
FROM `logistics_db.shipments_2025`
WHERE status = 'Delivered'
GROUP BY 1, 2
HAVING total_freight_cost > 5000; -- Filtrar solo rutas
costosas
```

Guía de inicio: spreadsheets 101

1) Formato y entrada datos

- Data entry: selecciones una celda y escribe. Simple, pero la precisión es clave
- Adjust size: Arrastra el borde derecho de la columna para expandir

  - text wrapping

Proceso de planificación de visualización

Paso 1: Explore the data for patterns

- Acción: revisar registros de ventas y analítica web
- Que buscas: Comportamiento del cliente, geográfico y montos de compra
- Ejemplo: identificas que las ventas son mas fuertes en el noreste ayuda a enfocar el nuevo diseño del sitio web

Paso 2: Plan your visuals

Enfoque:
- Adaptas el grafico a la audiencia (orientada a ventas)
- objetivo comunes:
  - mostrar ventas a lo largo del tiempo
  - Conectar ventas con ubicacion geografica
  - relacionar uso del sitio web con compras

Paso 3: Create your visuals

Selecciona el formato según la historia que quieres contar

1) Line Charts: ideales para rastrear cambios en el tiempo
2) Maps: perfectos para conectar datos con ubicaciones geográficas
3) Donut Charts: Utiles para mostrar segmentos del cliente
4) Bar Charts: Los mejores para comparar categorías

Herramienta	Nivel	Uso Principal
Spreadsheets	Básico/Rápido	Gráficos simples, histogramas y diagramas de embudo (*funnel*).
Tableau	Profesional	Dashboards interactivos, exploración de datos masivos [cite: 2025-12-18].
R / RStudio	Avanzado	Visualización programada, análisis estadístico y personalización total.

Recursos de exploración:

- Tableau public: herramienta gratuita para practicar y ver el "Viz of the day"
- RStudio Cheatsheets: Guías rápidas para programar visualizaciones en R

- R Studio visualizar data primes: introducción visual para principiantes en programación

Roles y descripción en el mundo de datos

Decodificando el título del puesto

no te dejes llevar solo por el nombre lee las tareas. Aquí la comparativa clave

Título	Foco Principal	Herramientas Típicas
Data Analyst	Resolver problemas con datos actuales.	SQL, Tableau, Excel, Dashboards.
Data Scientist	Predicciones y creación de nuevos modelos.	Estadísticas avanzadas, Python/R, ML.
Data Specialist	Gestión y organización de bases de datos.	Data Modeling, Seguridad, Recuperación.
Data Engineer	Preparar y limpiar la "tubería" de datos.	ETL, Integración de fuentes, Big Data.

Especializaciones por industria

Un analista puede "moverse" entre industrias aplicando la misma lógica analítica

- Business Analyst: Foco en procesos y eficiencia operativa
- Marketing Analyst: Evalúa el potencial de ventas y comportamiento del mercado
- Operations Analyst: Analiza flujos

de trabajo (ideal para logica)

- Risk Analyst: Evalua documentos financieros y economia para decidir niveles de riesgos

- Healthcare Analyst: Mejora la parte administrativa de hospitales usando datos medicos

Diferencias de Problem Solving

Es vital entender que espera la empresa de ti segun el rol

Analista: Usa herramientas existentes para ayudar a los stakeholders a decidir mejor hoy

Cientifico: Formula preguntas abiertas y resolviendo nuevos tipos de datos para predecir el mañana

Especialista: Aplica conocimiento profundo de bases de datos para asegurar la especialidad y desastres

Estrategias para un Analisis justo

Mejores Prácticas y Ejemplos

Práctica Recomendada	Explicación	Ejemplo en el Mundo Real
Considerar todos los datos	No ignorar datos que parezcan irrelevantes o contrarios a lo esperado.	Incluir datos del **clima** al analizar tráfico en días festivos para evitar conclusiones incompletas.
Identificar factores circundantes	Entender el contexto que influye en los resultados.	Al planear vacaciones, considerar festivos religiosos/culturales, no solo los bancarios oficiales [cite: 2025-12-18].
Incluir datos auto-reportados	Los participantes proporcionan su propia información.	Usar encuestas de clientes en lugar de dejar que los vendedores "supongan" la demografía del comprador.
Oversampling [Sobre-muestreo]	Aumentar el tamaño de la muestra de grupos minoritarios.	En una app de fitness, recolectar más datos de usuarios >70 años para asegurar que el contenido sea inclusivo [cite: 2025-11-12].

Ciclo de la equidad

1) Recolección: usar sobre-muestreo y auto-reportados

2) Procesamiento / Limpieza: no borrar datos "incomodos" sin justificación técnica.

3) Analisis: Buscar factores ocultos (Contexto)

4) Comunicación: informar a los stakeholders sobre las medidas de equidad

lomodos para que no interpretes
mal los resultados

Preguntas para la toma de decisiones

1) Objetivo general del curso

- Transformar problemas de negocio en preguntas efectivas
- Dominar el structured thinking
- gestionar la comunicación con stakeholders

2) Los 4 pilares del Modulo

- Modulo 1: preguntas efectivas
  Tecnicas para guiar el analisis y evitar anhegadas
- Modulo 2: Decisiones Basadas en datos → uso de reportes spreadsheets como herramienta
- Modulo 3: hojas de calculo
- Modulo: Gestion stakeholders

Conceptos

- Data-driven decision-making: proceso de basar decisiones en hechos y datos, no en instintos

Structured thinking: Marco de trabajo que consiste en desglosar un problema complejo en partes pequeñas y manejables

Stakeholders:

Herramientas y metodología

- Spreadsheets: no solo para calculos sino para organizar el scope of work
- Comunicación: Slabbles canales claros desde el inicio para evitar el scope creep
-

Las 6 Fases del analisis de datos

1) Ask

Objetivo: definir el problema y entender los expectativas de los stakeholders

Focos: Evitar distracciones y ver el contexto completo

Pregunta clave: ¿Como puedo ayudar a resolver el problema real?

2) Prepare

- Objetivo: decidir que datos recolectar y como organizarlos
- Actividades: identificar Metric a medir
- Ubicar datos en la database
- Establecer mediciones de seguridad

3) Process

- Objetivo: limpieza de datos. Los datos sucios llevan a conclusiones erroneas

- Acciones

  - Eliminar duplicados y corregir errores de entrada
  - Identificar Bais en la recolección
  - Uso de SQL para validación

```sql
-- Ejemplo: Eliminar espacios innecesarios en nombres de proveedores
SELECT TRIM(vendor_name) AS clean_vendor_name
FROM logistics_table;
```

4) Analyze

- objetivo pensar analíticamente para encontrar patrones
- Actividades: Realizar calculos, Combinar fuentes de datos y crear tablas resumen
- resumen clave: Que me están contando los datos

5) Share

- objetivo: Comunicar hallazgos mediante data visualizations
- Herramientas: Graficos reportes Dashboards
- Clave: hacer que la informacion sea facil de digerir para quienes no son tecnicos

6) Act

- Objetivo: Convertir insights en recomendaciones de negocio
- Resultado: Tomar decisiones informadas que resuelvan el problema original definido en la fase 1

1 reconocer el problema
2 organizar la info
3 Revelar brechas y oportunidades
4 identificar acciones de accion

Los 6 tipos de problemas en Data Analytics

1 Making predictions

- Usar datos existentes para predecir resultados futuros
- Ejemplo: Determinar cual es el mejor canal publicitario para atraer clientes el próximo trimestre

2 Categorizing things

- Clasificar datos basados en características comunes
- Ej:

  Clasificar llamadas de servicio al cliente por palabras clave para evaluar el desempeño de los agentes

3 Spotting something unusual

- identificar datos que se alejan del comportamiento normal
- un software de en un smart Watch que detecta anomalías en el ritmo cardiaco

4 Identifying themes

- Agrupar categorías en conceptos más amplios para explorar datos cualitativos
- Ej:
  En un estudio de User Experience (UX), agrupo las creencias y necesidades de los usuarios para mejorar un producto

5 Discovering Connections

-
- Encontrar relaciones similares entre fuentes de datos aparentemente distintas
- Ej:
  analizo los tiempos de espera en shipping hubs para ajustar los horarios y mejorar los entregas a tiempo

6 Finding patterns

- Usar datos históricos para entender la frecuencia con la que ocurre algo o las condiciones que lo provocan
- Ej:
  : noto que los fallos mecánicos

se producen cuando el mantenimiento preventivo se retrasa mas de 15 dias

Preguntas SMART: el estándar de oro del Analista

Que es?

Specific: Aborda el problema real? Tiene contexto?

Measurable: Se puede cuantificar? tiene metrica asociada

Action-oriented: La respuesta nos da un plan o ayuda a una decisión

Relevant: Es importante para el problema que entendemos resolver ahora

Time-bound: definir un periodo de estudio claro

2) Tipo de pregunta a evitar

- Leading questions: preguntas que ya inclinan al interlocutor hacia una respuesta
- mal: El costo del flete es alto, verdad?
- Bien: Que opinión tiene sobre la relación costo-Beneficio de nuestro

problema actual

- Closed-ended questions: si responde con un sí o un no, no aportan profundidad

  - Está satisfecho con el dashboard
  - Qué info adicional les gustaría ver en este dashboard para facilitar tu toma de decisiones?

- vague questions -> no tienes contexto ni claridad

  - funciona la herramienta?

- En términos de velocidad de carga, ¿es la nueva base de datos más eficiente que la anterior?

3/ El poder de las open-ended questions

- obtenes respuestas detalladas que permiten calificar o descalificar soluciones
- ideal para la fase de ask y poner

discutirs temas

Tipos de decisiones
Data-driven vs data inspired

Tipos de enfoques

Data-driven decisions:

- Uso de hechos concretos para guiar la estrategia
- Limitado por la cantidad y calidad de los datos disponibles
- Riesgo: ignorar qualitative insights y sesgos en la recolección
- Ejemplo
  A/B testing para layouts en los sitios web

- Data-inspired decisions
  - Combina datos con experiencia sentimientos y comparaciones externas
  - permite explorar una gama mas amplia de ideas
  - Ejemplo: Mejora el CSAT score analiza el numero pero tambien entrevistando al personal

Datos cuantitativos vs cualitativos en negocios

- Cualitativo Data: información numérica que se puede contar o medir. Responde al ¿Qué? ¿Cuanto? ¿Con que frecuencia?

Cualitativo Data: información no numérica como opiniones, descripciones o sentimientos. Responde al ¿por qué?

Herramientas:

Cuantitativo: Encuestas, sondeos, registros de transacción, bases de datos SQL

Cualitativo: Entrevistas, Focus groups, analisis de comentarios en redes

Toolkit de visualización
Spreadsheets vs Tableau

1) Sheets / Excel

- Crear gráficos y tablas de forma rápida

ventajas

- Simplicidad y fácil para datos en un solo lugar

- Interconectividad: vinculación fácil con procesadores de texto y presentaciones

- Automatic updates: los gráficos se actualizan automáticamente si los datos de origen cambian

ideal para reportes simples y análisis ad-hoc que no requieren interactividad profunda

2) tableau

- Creacion de interactive Dashboard y analisis de datos a gran escala

ventajas

- interfaz de drag-and-drop
- Soporta fuentes de datos masivos
- capacidades de advance analysis para explorar tendencias y patrones complejos

- Desventajas: Curva de aprendi: zaje mas alta tomo tiempo dominarlo profesionalmente

Recurso: practicas en tableau public

Caracteristica	Spreadsheet (Excel/Sheets)	Tableau
Volumen de datos	Pequeño / Mediano	Masivo (Big Data)
Interactividad	Limitada	Alta (Filtros dinámicos)
Velocidad de entrega	Muy rápida (minutos)	Media (requiere diseño)
Propósito	Reporte estático / Limpieza	Exploración de datos / Dashboards vivos

Diseño de dashboards de alto impacto

1)

- Es una herramienta visual que organiza multiples datasets en una ubicacion central para monitorear datos en vivo

2) Beneficios para analistas y stakeholders

analistas:
- × una sola fuente
- × actualizacion de datos en tiempo real
- × Cruce de datos en tiempo real
- × Cruce de dataset (varios)
- × vistas dedicadas por proyecto
- × vista completa de objetivos por procesos

stakeholders:
- × Detecciones rapidas de tendencias y patrones
- × Entiende la "historia detras de los numeros
- × posibilidad de hacer drill down

3) proceso de creacion en 5 pasos

- Identificar stakeholders

¿Quien sera los datos?
¿Que problemas se deben resolver?

- Diseño (layout):
  - Encabezados claros
  - prioridad: lo mas importante en la parte de arriba

- Mockups:
  - Esbozos en borrador del dashboard antes de construirlo

- Selecion de visualizaciones
  - Line charts / bar graphs: para cambio en el tiempo
  - Pie / donut charts: para mostrar como partes contribuyen a un todo

- Filters:

herramientas para ocultar datos irrelevantes y centrarse en patrones especificos sin alterar la fuente original

- Regla:
  Si la data structure cambia en la base de datos, el dashboard no se actualiza automaticamente. Primero se deben ajustar el diseño para que coincida con la nueva estructura

- Tips de tableau
  - ideal para huge datasets
  - ofrece interactive visualizations
  - Usa el Viz of the day de tableau para buscar inspiracion en logistica y transporte

Comparativa: Small vs Big Data

Small Data

- Métricas específicas, periodos cortos
- hojas de cálculo / SQL local
- Datos estructurados y organizados
- Simples de recolectar y visualizar

Big Data

- Datasets masivos, periodos extensos
- Bases de datos distribuidas / Data lake
- Desestructurados; requieren procesamiento
- requiere alta capacidad de cómputo y ETLs complejos

(2) Pilares del Big Data

Volumen: Escala de los datos (Teras a Zetta)

Variedad: Formatos diversos (video, JSon, audio)

velocidad: Ritmo de generación y

procesamiento (Batch vs Real Time)

Veracidad: Grado de incertidumbre y calidad del dato (Data Cleaning

③ Desafíos técnicos y estratégicos

- Sobrecarga de información: ruido que oculta los insights críticos

- Accesibilidad: Datos aislados en silos que ralentizan la toma de decisiones rápidas

- Sesgo algorítmico: Resultados injustos derivados de datos históricos con prejuicios o falta de representatividad

④ Beneficios de la toma de decisiones

- Eficiencia operativa: identificación de cuellos de botella en la cadena de suministro

- Patrones de consumo: Prediccion de demandas basada en comportamiento a gran escala

- Proteccion de marca: Monitoreo de reputacion mediante analisis de sentimiento en redes sociales a gran escala

(5) Aplicacion tecnica (Bash/SQL/contex

- En small, un select * con un par de join es suficiente

- En Big data para distribuir la carga que en SQL tradicional no soportaria eficientemente

Tipos de datos y medicion

- Dato cuantitativo: medida objetiva y numerica
- Dato cualitativo: Medida subjetiva o descriptiva
- Metrica: Tipo de dato unico y cuantificable usado para medir
- Meta de metrica: objetivo medible evaluado a través de metricas (KPIs)

2 Estructura y herramientas de visualizacion

- Cuadro de mando: Herramientas de monitoreo de datos en vivo
- Informe (Report): Coleccion estadisticas de datos entregada periodicamente
- Tabla dinamica: herramienta para resumir, agrupar y promediar datos rapidamente
- Grafico dinamico: representacion visual basada en una tabla dinamica

③ Marco de resolución de problemas
los analistas de datos abordan problemas mediante

④ Categorización: clasificar elementos

B) Descubrir Conexiones: relaciones entre variables

C) Búsqueda de patrones:

tendencias repetitivas

D) identificación de temas: agrupamiento de de conceptos cualitativos

E) Realización de predicciones: Modelo de eventos futuros

F) detección de anomalías: identificar algo inusual

④ Concepto de negocio y decisión

- Algoritmo: Reglas lógicas para ejecutar una tarea (Clave para automatización en Python / Bash)

- ROI : Formula de exito (utilidad / inversion) × 100

- ingresos : ingreso total por ventas antes de gastos

- Toma de decisiones inspirada en datos : Explorar multiples fuentes para encontrar puntos comunes y generar estrategias

Data life Cycle

es el framework que garantiza que el dato sea integro desde su nacimiento hasta su eliminacion.

Los lagos de datos no son solo tablas; en este contexto, actuan con interfaz de usuario de los datos, permitiendo una manipulacion rapida antes de escalar procesos mas robustos en SQL o Python

Fundamental

el ciclo es iterativo y normado. No se trata solo de borrar archivos, sino de cumplir con politicas de gobernanza de datos. Se evalua la capacidad para decidir cuando un dato debe ser archivado (por cumplimiento legal/historico) y cuando debe ser destruido (por seguridad/privacidad)

Fases del ciclo de vida

- Planificacion
  - Establecer estandares de organizacion

· objetivo: Consistencia y eficiencia en la colaboración

- Captura

- Conexión a fuentes externas (Encuestas SQL, bases, APIs)

- Data integrity: Automatización para asegurar datos actuales y sin errores

- Gestión:

  - filtrado, ordenamiento y actualización

  - Seguridad: Control de permisos y accesos

- Análisis

  - Uso de fórmulas de agregación

  - tablas dinámicas para visualización rápida de KPI

- Archivado
  - Almacenamiento de datos históricos no activos
  - referencia futura sin saturar los sistemas operativos
- Destrucción
  - Eliminación definitiva por razones de seguridad o cumplimiento legal
  - Uso de copias de seguridad antes de la purga final

② Eficiencia técnica

- Atajos: el dominio del lectorado es crítico

Gestion de archivos y entorno

① Gestion de archivo y entorno

- nomenclatura: usar nombres descriptivos y especificos
- organización: Utilizar Carpetas logicas (Files → Move → New Folder)

para evitar el desorden en el drive / Disco

② Estructuración y formato

- Ajuste de dimensiones
- Seleccionar el vertice para seleccionar todos los filas
- redimensionar columnas para evitar que los datos se corten
- Jerarquia de atributos
- Full color + Bold: diferenciar la fila de encabezados de los datos crudos

- Bordes: Definir límites blanco entre celdas para mejorar la lectura rápida

③ Manipulación de esquema

- insertar columna
- eliminar columna
- congelar filas: View → freeze →

row 1

Comparativa técnica

1

Característica	Microsoft Excel	Google Sheets
Entorno	Local / Escritorio (Powerhouse).	Nube / SaaS (Colaborativo).
Capacidad	Maneja datasets más grandes.	Limitado por celdas (actualmente 10M).
Conectividad	Power Query, VBA.	Apps Script, Conexión nativa con GCP.
Colaboración	Limitada en versión de escritorio.	Nativa, en tiempo real, historial detallado.

2 operaciones críticas de datos

- ordenamiento
  - Jerarquía : (Filtro
  - custom sort
- Filtrado
- Formato

3 Funciones de organización

- Eliminar duplicados : limpieza rápida de registros redundantes
- texto en columnas : útil para separar datos delimitados

- Validación de datos: Restringir entrada en cuadros para asegurar la integridad de los datos

Fundamentos de formulas

- inicio de instrucion : Siempre =
- operaciones basicas : +, -, *, /

② Averages

= (Suma de los Celdas) / N

③ Anatomia de la formula

Relativa vs absoluta

Relativa	A1	cambia la posicion relativa A1 → B1
absoluta	A1	Bloqueado siempre apunta a la mismo Celda
Mixta (Fila)	A$1	la columna cambia pero la fila queda fija
Mixta (columna)	$A1	la fila cambia pero la columna queda fija

③ Formulas con logica integrada

Formulas + Funciones : combinar instru para analisis especifico

- Uso = Countif(D2 : D100, ">500")

(cuenta celdas con valor mayor a 500)

Problemas de prevención

- Reducción de complejidad: filtrar datos para mantener la hoja limpia
- Congelar encabezados mantiene contexto en datasets extensos
- Sintaxis estricta
  - Siempre =
  - Multiplicación *
  - Balance de paréntesis Todo (debe cerrarse con)
- Estructura de trabajo
  - Hoja 1: Datos crudos
  - Hoja 2: Datos procesados

Error	Causa Principal	Ejemplo Logístico
`#DIV/0!`	División por cero o celda vacía.	Costo total / Unidades (donde unidades = 0).
`#ERROR!`	Error de análisis (Parsing).	Olvidar una coma entre rangos de celdas.
`#N/A`	Dato no encontrado.	`VLOOKUP` de un ID de contenedor inexistente.
`#NAME?`	Función mal escrita.	Escribir `=SUMA()` en una versión en inglés (`=SUM()`).
`#NUM!`	Valor numérico inválido.	Cálculo de fechas donde la fecha inicial es posterior a la final.
`#REF!`	Referencia de celda no válida.	Eliminar una columna que alimentaba un cálculo de KPI.
`#VALUE!`	Tipo de dato incorrecto.	Intentar sumar "100 USD" (texto) + 50 (número).

Detección Avanzada con formato condicional

Lógica de la regla

- Función clave =iserror(celda)
- objetivo: resaltar celdas que contienen cualquier error limpieza de datos inmediata

implementación

1) Seleccionamos todo el dataset
2) Menu Formato -> Formato condicional
3) Formula personalizada es (custom formula is)

Formula

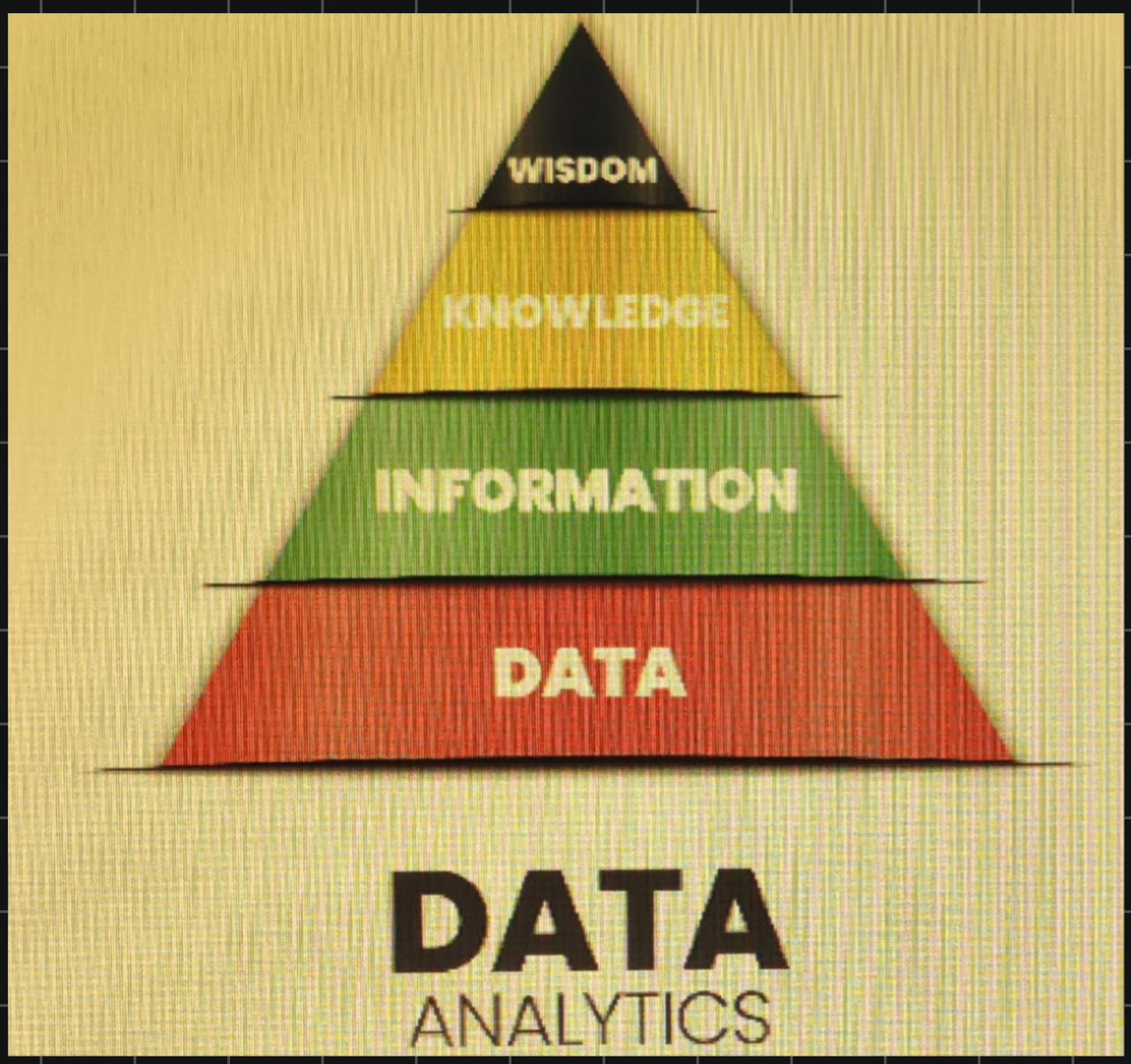
=iserror(A1)

5) Estilo: relleno amarillo o rojo

Contextualización de datos

- Definición: Proceso de dar perspectiva a los datos mediante su definición y delimitación

- Transformación: Datos crudos → contexto → información con con significado



Framework de identificación

- origen: ¿Quién recolectó el dato y qué sistema proviene?

ERP, sensores IOT, logs de Bash

Temporalidad: Crucial para el análisis de series de tiempo

propósito: ¿La recolecta para control de inventarios o para auditoría financiera?

Contexto en el ciclo de vida

1) Recolección

- Entender los procesos de negocio
- Asegurar que los datos responden a la pregunta de investigación

2) Organización

- Uso de convenciones de nombres claros
- identificar relaciones entre variables

3) Presentación

- Proveer el marco de referencia a los partes interesados
- evitar interpretación errónea de los resultados técnicos

(4) Riesgos de la falta de contexto

- Sesgos: Conclusiones erroneas por ignorar la fuente de datos

- Inaccesibilidad: Datos que nadie entiende y, por tanto, nadie usa

- Perdida de valor: datos sin contexto = ruido estadístico

(5) Concepto de negocio y gestión

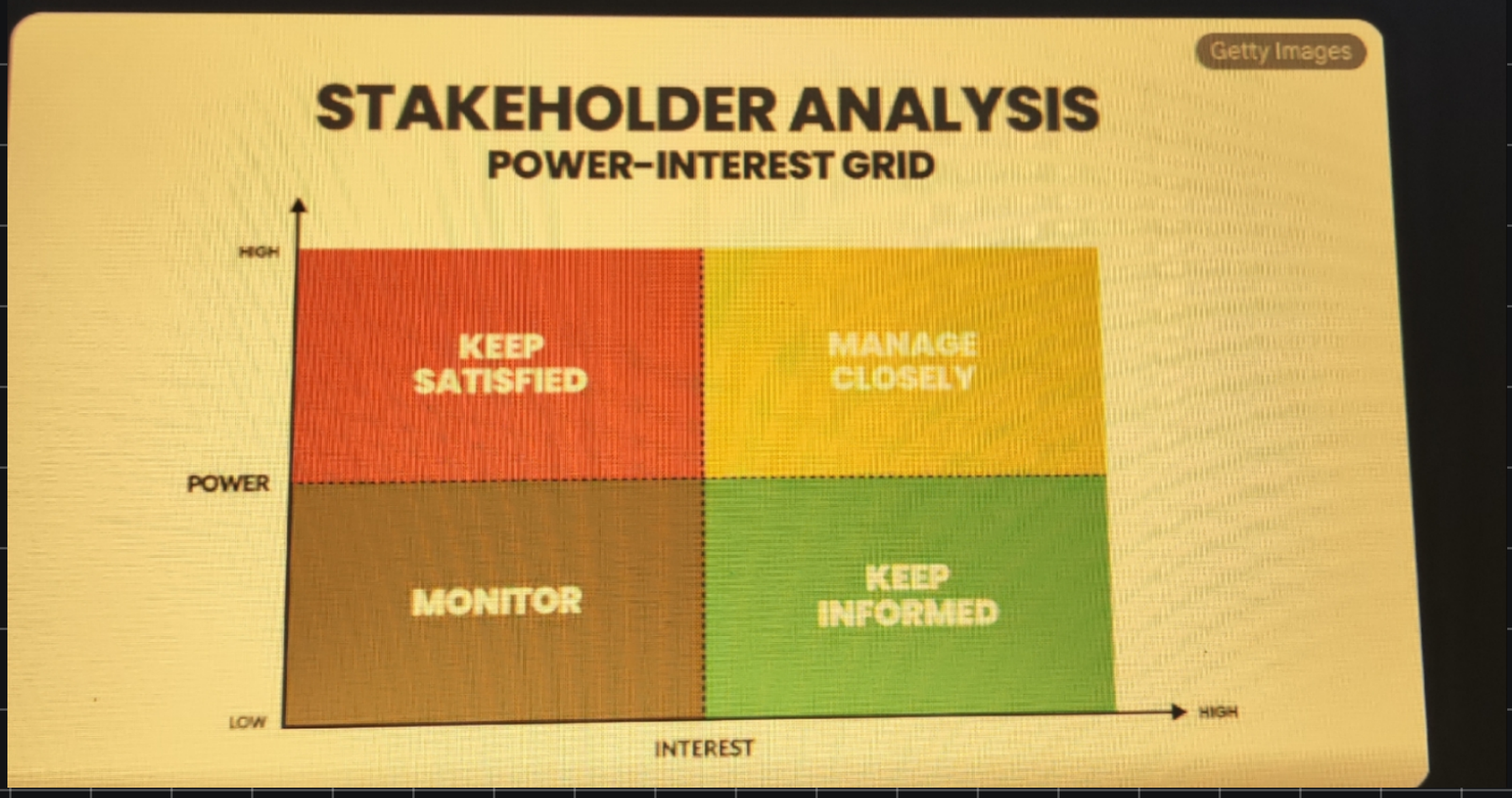
- Retorno de inversión (ROI) = metrica de eficiencia financiera

$$\frac{(\text{Beneficio} - \text{inversion})}{\text{inversion}}$$

Segmentación de Stakeholders

Grupo	Rol / Enfoque	Necesidad de Información
Ejecutivo (Executive)	Liderazgo estratégico (VPs, Directores).	Titulares, impacto en ROI, visión 30,000 ft.
Cliente (Customer-facing)	Ventas, Soporte, Marketing.	Feedback del cliente, viabilidad de producto.
Data Science Team	Analistas, Ingenieros, Científicos.	Metodología, limpieza de datos, predicciones.

Export to Sheets



Discutir objetivos: No ocupes productos de valor sin entender el porqué

Empoderamiento del "NO": priorizar calidad sobre velocidad

- Explicar limitaciones técnicas con contexto de negocio
- ofrecer alternativas viables

en lugar de una negativa rotunda

- Patrocinio lo inesperado: añadir buffers de tiempo para resolución de problemas

- Evitar la ilusión de acuerdo

mécico con la descripción verbal + bocetos visuales rápidos

- Validar constante "Es esto lo que esperas ver"

Herramientas de gestión

- Registro de cambios: Documento cronológico de modificaciones al proyecto. Es tu seguro de vida laboral

- Apéndice: Juegos para el código SQL complejo o métricas de errores de Bash que el ejecutivo quiere ver en el slid principal

- Documentación del proyecto: Registro centralizado que permita la reproducibilidad

Reglas de oro

Vulnerabilidad: Los datos deben contar la historia no cada cuenta stakeholders que queremos que el dato conforme su interpretación respecto

Comunicación frecuente: Compartir hitos y obstáculos inmediatamente

- Conocimiento del dominio: Entender como tu análisis de flujos o inventarios se conecta con la rentabilidad total de la empresa