

*data by milo*

DATA. PERO CUTE.

PRIMERA EDICIÓN

· 2026 ·



200



Diccionario

*Anti-Ferga*

de Datos

*traducido al humano*

200 términos técnicos con su traducción al español, una explicación en lenguaje cotidiano, y un ejemplo de la vida real. para las girls que nunca creyeron que eran de datos.



ANTES DE EMPEZAR

# esto es *para vos* si alguna vez...



- *abriste Excel y te dio ansiedad existencial*
- *asentiste en una reunión sin entender qué es un KPI*
- *googleaste "qué significa cohort" en modo incógnito*
- *las girls no somos de datos? cooked. perdón, sí somos.*

cada término está traducido del inglés técnico al español que hablamos, con una explicación corta y un ejemplo que podrías vivir un martes cualquiera. está ordenado en nueve módulos, de lo más básico a lo más nerdo. lo podés leer de corrido o ir a buscar el término que te dijeron en la reunión de la mañana.

---

*no te mintieron cuando te dijeron que esto es difícil. pero tampoco te dijeron que hay forma de aprender sin llorar.*

*con cariño, milo*

# índice del diccionario

9 módulos · 200 términos

<b>01</b>	<b>Análisis y Métricas Básicas</b> <i>los números que todo el mundo nombra en la reunión y vos hacés que entendés.</i>	<b>20 TÉRMINOS</b>	<b>05</b>
<b>02</b>	<b>Estadística</b> <i>prometo que esto no se parece a la profe que te traumó en el secundario.</i>	<b>24 TÉRMINOS</b>	<b>11</b>
<b>03</b>	<b>Machine Learning e Inteligencia Artificial</b> <i>el tema del que habla todo el mundo sin saber bien qué es. acá te cuento.</i>	<b>40 TÉRMINOS</b>	<b>18</b>
<b>04</b>	<b>Bases de Datos</b> <i>donde viven los datos cuando nadie los mira. spoiler: en tablas.</i>	<b>24 TÉRMINOS</b>	<b>29</b>
<b>05</b>	<b>Ingeniería de Datos</b> <i>las plomerías del mundo data. sin esto, no llega agua a la canilla.</i>	<b>20 TÉRMINOS</b>	<b>36</b>
<b>06</b>	<b>Visualización de Datos</b> <i>donde los números se ponen guapos y por fin se dejan entender.</i>	<b>20 TÉRMINOS</b>	<b>42</b>
<b>07</b>	<b>Business Intelligence y Negocio</b> <i>las siglas que se tiran en reuniones estratégicas. ahora las tirás vos.</i>	<b>23 TÉRMINOS</b>	<b>48</b>
<b>08</b>	<b>Conceptos Avanzados y Calidad de Datos</b> <i>cuando ya sabés lo básico y empezás a preguntar por qué se rompen las cosas.</i>	<b>20 TÉRMINOS</b>	<b>55</b>
<b>09</b>	<b>Términos Frecuentes en Reuniones</b> <i>guardá esto. esto es lo que se dice y lo que significa de verdad.</i>	<b>9 TÉRMINOS</b>	<b>61</b>

*guardá esto*

# Análisis y Métricas Básicas

*los números que todo el mundo nombra en la reunión y vos hacés que entendés.*

*20 términos*

## Dashboard

n° 001

*Panel de control / tablero*

**QUÉ ES** Pantalla visual que muestra los datos más importantes de un vistazo.

**EJEMPLO** “ Como el tablero de un auto: velocidad, gasolina y temperatura en un lugar.”

## KPI

n° 002

*Indicador Clave de Rendimiento*

**QUÉ ES** Número que mide si se está cumpliendo un objetivo importante.

**EJEMPLO** “ Si la meta es vender 100 unidades al mes, el KPI dice cuántas llevas hoy.”

## Métrica

n° 003

*Medida de seguimiento*

**QUÉ ES** Cualquier valor que se mide y registra con el tiempo.

**EJEMPLO** “ El número de pasos que cuenta tu celular cada día es una métrica.”

## Benchmark

n° 004

*Punto de referencia*

**QUÉ ES** Valor promedio de la industria usado para compararse.

**EJEMPLO** “ Saber que el restaurante promedio tiene 4.2 estrellas te dice si tus 3.8 están bien o mal.”

## Insight

n° 005

*Hallazgo accionable*

**QUÉ ES** Una conclusión útil que surge de analizar datos, no es solo un número.

**EJEMPLO** “ No es 'las ventas bajaron 20%', sino 'bajaron porque llueve los lunes'.”

## Trend

n° 006

*Tendencia*

**QUÉ ES** Dirección general que siguen los datos a lo largo del tiempo.

**EJEMPLO** “ Si cada mes gastás un poco más en delivery, eso es una tendencia al alza.”

## Outlier

n° 007

*Valor atípico / dato anómalo*

**QUÉ ES** Dato que se aleja mucho del resto y llama la atención.

**EJEMPLO** “ Si todos ganan entre 2000 y 5000 y uno gana 50000, ese es un outlier.”

## Data point

n° 008

*Dato individual / observación*

**QUÉ ES** Un solo valor dentro de un conjunto de datos.

**EJEMPLO** “ Tu compra del martes a las 3pm en el súper es un data point en su sistema.”

## Sample

n° 009

### Muestra

**QUÉ ES** Un subconjunto representativo del total de datos.

**EJEMPLO** “Revisar 200 celulares de una fábrica en vez de los 10,000 fabricados.”

## Aggregate

n° 010

### Valor agregado / resumen numérico

**QUÉ ES** Resultado de combinar muchos datos en un solo número.

**EJEMPLO** “El total de ventas del mes es un agregado de cientos de ventas individuales.”

## Filter

n° 011

### Filtro

**QUÉ ES** Condición que limita los datos mostrados a solo los que cumplen un criterio.

**EJEMPLO** “Ver solo las ventas de Argentina es aplicar un filtro por país.”

## Segment

n° 012

### Segmento / grupo

**QUÉ ES** Subconjunto de personas que comparten una característica común.

**EJEMPLO** “Clientes que compraron más de 3 veces en el último mes forman un segmento.”

## Cohort

n° 013

### Cohorte / grupo de seguimiento

**QUÉ ES** Grupo de personas que vivieron el mismo evento en el mismo período.

**EJEMPLO** “Todas las usuarias que se registraron en enero, seguidas mes a mes.”

## Funnel

n° 014

### Embudo de conversión

**QUÉ ES** Secuencia de pasos que sigue un usuario y en cada paso algunos se van.

**EJEMPLO** “De 1000 que ven un producto, 300 lo agregan al carrito y 80 compran.”

## Conversion rate

n° 015

### Tasa de conversión

**QUÉ ES** Porcentaje de personas que completaron la acción deseada.

**EJEMPLO** “Si 100 visitan tu tienda y 5 compran, la tasa de conversión es del 5%.”

## Churn

n° 016

### Tasa de abandono

**QUÉ ES** Porcentaje de clientes que dejaron de usar el servicio en un período.

**EJEMPLO** “Si 10 de tus 100 suscriptores cancelaron este mes, el churn es del 10%.”

## Retention

n° 017

### Retención

**QUÉ ES** Porcentaje de usuarios que siguen activos después de un período.

**EJEMPLO** “Si el 70% de los que instalaron tu app hace un mes la siguen usando, tenés 70% de retención.”

## Engagement

n° 018

### Nivel de interacción

**QUÉ ES** Qué tanto interactúan los usuarios con tu producto o contenido.

**EJEMPLO** “Cuántos likes, comentarios o minutos de lectura genera un post.”

## Attribution

n° 019

### Atribución

**QUÉ ES** Determinar qué canal o acción fue responsable de una conversión.

**EJEMPLO** “Saber si el cliente compró gracias al email, al anuncio de Instagram o a Google.”

## ROI

n° 020

### Retorno sobre la inversión

**QUÉ ES** Cuánto ganaste comparado con lo que gastaste, en porcentaje.

**EJEMPLO** “Si invertiste 100 y ganaste 150, tu ROI es del 50%.”



*respirá. es solo math*

# Estadística

*prometo que esto no se parece a la  
profe que te traumó en el  
secundario.*

*24 términos*

## Mean

n° 021

*Media / promedio*

**QUÉ ES** Suma de todos los valores dividida entre la cantidad de valores.

**EJEMPLO** “El promedio de notas 7, 8 y 9 es  $(7+8+9)/3 = 8$ .”

## Median

n° 022

*Mediana*

**QUÉ ES** El valor que queda justo en el centro cuando ordenás los datos.

**EJEMPLO** “En los sueldos [1000, 1200, 5000], la mediana es 1200, más representativa que el promedio de 2400.”

## Mode

n° 023

*Moda*

**QUÉ ES** El valor que aparece con más frecuencia en un conjunto de datos.

**EJEMPLO** “Si 5 de 10 clientes piden pizza de jamón, la moda del pedido es esa pizza.”

## Standard deviation

n° 024

*Desviación estándar*

**QUÉ ES** Medida de qué tan dispersos están los datos alrededor del promedio.

**EJEMPLO** “Si los tiempos de entrega siempre son 30 min, la desviación es baja. Si varían entre 10 y 90, es alta.”

## Distribution

n° 025

*Distribución*

**QUÉ ES** Cómo se reparten los valores de un conjunto de datos.

**EJEMPLO** “Ver cuántos productos se venden a cada precio: muchos baratos, pocos caros.”

## Normal distribution

n° 026

*Distribución normal / campana de Gauss*

**QUÉ ES** Distribución en forma de campana: la mayoría de valores están en el centro.

**EJEMPLO** “Las alturas de las personas: la mayoría mide entre 1.60 y 1.80, pocos son extremos.”

## Skewness

n° 027

*Asimetría*

**QUÉ ES** Medida de qué tan inclinada está una distribución hacia un lado.

**EJEMPLO** “Los sueldos suelen estar sesgados: muchos ganan poco, pocos ganan muchísimo.”

## Percentile

n° 028

*Percentil*

**QUÉ ES** Posición relativa de un valor dentro de una distribución, del 0 al 100.

**EJEMPLO** “Estar en el percentil 90 significa que te fue mejor que al 90% de los participantes.”

## Quartile

n° 029

### Cuartil

**QUÉ ES** División de los datos en cuatro partes iguales (Q1, Q2, Q3, Q4).

**EJEMPLO** “Q1 es el 25% inferior, Q2 es la mediana, Q3 es el 75% superior.”

## Confidence interval

n° 030

### Intervalo de confianza

**QUÉ ES** Rango en el que probablemente cae el valor real, con cierta certeza.

**EJEMPLO** “El candidato tiene entre 42% y 48% de intención de voto con 95% de confianza.”

## P-value

n° 031

### Valor p

**QUÉ ES** Probabilidad de obtener ese resultado por puro azar. Si es bajo, el resultado es significativo.

**EJEMPLO** “Un p-valor de 0.02 significa solo 2% de probabilidad de que el resultado sea casualidad.”

## Hypothesis testing

n° 032

### Prueba de hipótesis

**QUÉ ES** Método estadístico para decidir si un efecto observado es real o producto del azar.

**EJEMPLO** “Probar si el nuevo botón de compra realmente aumenta ventas o fue suerte.”

## Statistical significance

n° 033

### Significancia estadística

**QUÉ ES** Un resultado es significativo cuando es poco probable que sea por azar.

**EJEMPLO** “Si el medicamento baja la presión en todos los pacientes del estudio, es significativo.”

## Correlation

n° 034

### Correlación

**QUÉ ES** Medida de cuánto se mueven juntas dos variables. Va de -1 a 1.

**EJEMPLO** “El helado y el calor están correlacionados: cuando hace calor, se vende más helado.”

## Causation

n° 035

### Causalidad

**QUÉ ES** Una variable causa directamente el cambio en otra. Más fuerte que correlación.

**EJEMPLO** “Fumar causa cáncer de pulmón. No solo están correlacionados, uno provoca al otro.”

## Regression

n° 036

### Regresión

**QUÉ ES** Técnica para predecir un valor numérico basándose en otras variables.

**EJEMPLO** “Predecir el precio de una casa según sus metros cuadrados y ubicación.”

## R-squared

n° 037

*Coefficiente de determinación*

**QUÉ ES** Cuánto explica el modelo la variación de los datos. De 0 a 1.

**EJEMPLO** “Un  $R^2$  de 0.85 significa que el modelo explica el 85% de por qué los precios varían.”

## Sample size

n° 038

*Tamaño de muestra*

**QUÉ ES** Cuántos casos o personas se incluyen en un estudio o análisis.

**EJEMPLO** “Encuestar a 50 personas no da la misma confianza que encuestar a 5000.”

## Bias

n° 039

*Sesgo*

**QUÉ ES** Error sistemático que inclina los resultados hacia un lado de forma no aleatoria.

**EJEMPLO** “Si solo encuestás a tus clientes más fieles, los resultados tendrán sesgo positivo.”

## A/B testing

n° 040

*Prueba A/B*

**QUÉ ES** Experimento donde se comparan dos versiones de algo para ver cuál funciona mejor.

**EJEMPLO** “Mostrar el botón verde al 50% y el rojo al otro 50%, y medir cuál genera más clics.”

## Null hypothesis

n° 041

*Hipótesis nula*

**QUÉ ES** La suposición de que no hay efecto o diferencia. Se intenta refutar.

**EJEMPLO** “ $H_0$ : este nuevo tratamiento no tiene efecto. Si el p-valor es bajo, se rechaza.”

## Population

n° 042

*Población*

**QUÉ ES** El conjunto completo de individuos o datos que se quiere estudiar.

**EJEMPLO** “Todos los clientes son la población; los 500 que encuestaste son la muestra.”

## Random sample

n° 043

*Muestra aleatoria*

**QUÉ ES** Subconjunto de datos elegido al azar para que sea representativo.

**EJEMPLO** “Sacar 100 nombres al azar de una lista de 10,000 para una encuesta.”

## Multicollinearity

n° 044

*Multicolinealidad*

**QUÉ ES** Cuando dos o más variables predictoras están muy correlacionadas entre sí.

**EJEMPLO** “Usar 'temperatura' y 'si hace calor (sí/no)' en el mismo modelo causa multicolinealidad.”

*La IA no es magia*

# Machine Learning e Inteligencia Artificial

*el tema del que habla todo el mundo sin saber bien qué es. acá te cuento.*

*40 términos*

## Model

nº 045

### Modelo

**QUÉ ES** Un programa entrenado con datos para hacer predicciones o tomar decisiones.

**EJEMPLO** “Como una receta aprendida: el modelo aprendió de ejemplos pasados y predice casos nuevos.”

## Algorithm

nº 046

### Algoritmo

**QUÉ ES** Conjunto de instrucciones que el modelo sigue para aprender de los datos.

**EJEMPLO** “La receta en sí, no el cocinero. Define cómo se procesan los ingredientes (datos).”

## Training data

nº 047

### Datos de entrenamiento

**QUÉ ES** Los ejemplos históricos con los que el modelo aprende.

**EJEMPLO** “Mostrarle al modelo miles de fotos de gatos y perros etiquetadas para que aprenda.”

## Test data

nº 048

### Datos de prueba

**QUÉ ES** Datos que el modelo nunca vio, usados para evaluar qué tan bien aprendió.

**EJEMPLO** “El examen final después de estudiar con los apuntes de entrenamiento.”

## Validation

nº 049

### Validación

**QUÉ ES** Proceso de medir el rendimiento del modelo durante el entrenamiento para ajustarlo.

**EJEMPLO** “Como los exámenes parciales que dicen si vas bien antes del final.”

## Feature

nº 050

### Variable / característica

**QUÉ ES** Cada columna o dato de entrada que el modelo usa para hacer predicciones.

**EJEMPLO** “Edad, sueldo e historial de pagos son features para predecir si alguien pagará un préstamo.”

## Label

nº 051

### Etiqueta

**QUÉ ES** La respuesta correcta que el modelo intenta predecir.

**EJEMPLO** “En un filtro de spam, la etiqueta es 'spam' o 'no spam' para cada email de entrenamiento.”

## Target variable

nº 052

### Variable objetivo

**QUÉ ES** Lo que el modelo intenta predecir. Es la columna 'respuesta' en los datos.

**EJEMPLO** “Si predecís el precio de una casa, el precio es la variable objetivo.”

## Prediction

n° 053

### Predicción

**QUÉ ES** El resultado que el modelo genera para un nuevo dato.

**EJEMPLO** “El modelo dice '85% de probabilidad de que este cliente se vaya'. Esa es la predicción.”

## Classification

n° 054

### Clasificación

**QUÉ ES** Tarea de ML donde el resultado es una categoría, no un número.

**EJEMPLO** “Decidir si un email es spam o no, o si una imagen muestra un gato o un perro.”

## Clustering

n° 055

### Agrupamiento

**QUÉ ES** Técnica para agrupar datos similares sin etiquetas previas.

**EJEMPLO** “Agrupar clientes por comportamiento sin haberles dicho de antemano en qué grupo van.”

## Supervised learning

n° 056

### Aprendizaje supervisado

**QUÉ ES** El modelo aprende de ejemplos que ya tienen la respuesta correcta.

**EJEMPLO** “Como estudiar con un libro de ejercicios que tiene las respuestas al final.”

## Unsupervised learning

n° 057

### Aprendizaje no supervisado

**QUÉ ES** El modelo encuentra patrones en datos sin etiquetas ni respuestas correctas.

**EJEMPLO** “Como ordenar tu ropa sin que nadie te diga cómo: el modelo crea sus propias categorías.”

## Reinforcement learning

n° 058

### Aprendizaje por refuerzo

**QUÉ ES** El modelo aprende probando acciones y recibiendo recompensas o penalizaciones.

**EJEMPLO** “Como entrenar a un perro: si hace algo bien, recibe un premio; si no, no.”

## Neural network

n° 059

### Red neuronal

**QUÉ ES** Modelo inspirado en el cerebro humano, con capas de nodos interconectados.

**EJEMPLO** “Como una cadena de personas pasando información y transformándola en cada paso.”

## Deep learning

n° 060

### Aprendizaje profundo

**QUÉ ES** Redes neuronales con muchas capas que aprenden patrones muy complejos.

**EJEMPLO** “Lo que hay detrás del reconocimiento de voz del celular y las recomendaciones de Netflix.”

## Overfitting

n° 061

### Sobreajuste

**QUÉ ES** El modelo memoriza los datos de entrenamiento pero falla con datos nuevos.

**EJEMPLO** “ Como memorizar las respuestas del libro sin entender el concepto: fallás en el examen real.”

## Underfitting

n° 062

### Subajuste

**QUÉ ES** El modelo es demasiado simple y no logra aprender los patrones básicos.

**EJEMPLO** “ Como estudiar 5 minutos para un examen de 3 horas: claramente insuficiente.”

## Accuracy

n° 063

### Exactitud

**QUÉ ES** Porcentaje de predicciones correctas sobre el total.

**EJEMPLO** “ Si el modelo acierta 90 de cada 100 predicciones, su accuracy es del 90%.”

## Precision

n° 064

### Precisión en clasificación

**QUÉ ES** De todos los casos que el modelo dijo que eran positivos, cuántos realmente lo eran.

**EJEMPLO** “ Si el modelo dijo 'spam' 100 veces y 90 realmente lo eran, precision = 90%.”

## Recall

n° 065

### Sensibilidad / Cobertura

**QUÉ ES** De todos los casos positivos reales, cuántos logró identificar el modelo.

**EJEMPLO** “ Si había 100 spams reales y el modelo detectó 80, recall = 80%.”

## F1 score

n° 066

### Puntuación F1

**QUÉ ES** Promedio armónico de precision y recall. Útil cuando ambas medidas importan.

**EJEMPLO** “ Si precision es alta pero recall bajo (o viceversa), el F1 equilibra los dos.”

## Confusion matrix

n° 067

### Matriz de confusión

**QUÉ ES** Tabla que muestra cuántas predicciones fueron correctas e incorrectas por categoría.

**EJEMPLO** “ Un reporte que dice: de los 100 spam, acerté 95 y me equivoqué en 5.”

## Cross-validation

n° 068

### Validación cruzada

**QUÉ ES** Evaluar el modelo dividiendo los datos en partes y rotando cuál se usa para prueba.

**EJEMPLO** “ Como hacerse exámenes con distintos grupos de preguntas para no memorizarlas.”

## Hyperparameter

n° 069

### Hiperparámetro

**QUÉ ES** Configuración del modelo definida antes del entrenamiento, no se aprende de los datos.

**EJEMPLO** “Cuántas capas tiene la red neuronal o qué tan rápido aprende son hiperparámetros.”

## Gradient descent

n° 070

### Descenso de gradiente

**QUÉ ES** Algoritmo que ajusta el modelo minimizando el error poco a poco.

**EJEMPLO** “Como bajar una montaña a ciegas dando pequeños pasos hacia donde el terreno baja más.”

## Loss function

n° 071

### Función de pérdida

**QUÉ ES** Medida del error del modelo: cuánto se equivoca en sus predicciones.

**EJEMPLO** “Si predijo precio 100 y era 120, la pérdida mide esa diferencia de 20.”

## Epoch

n° 072

### Época de entrenamiento

**QUÉ ES** Una pasada completa del modelo por todos los datos de entrenamiento.

**EJEMPLO** “Como leer el mismo libro de estudio completo una vez. Con 10 épocas, lo lees 10 veces.”

## Batch

n° 073

### Lote de datos

**QUÉ ES** Subconjunto de datos procesados de una vez durante el entrenamiento.

**EJEMPLO** “En vez de ver todas las fotos de una vez, el modelo las procesa de a 32 o 64.”

## Tokenization

n° 074

### Tokenización

**QUÉ ES** Dividir texto en unidades más pequeñas para procesarlo.

**EJEMPLO** “'Hola mundo' se tokeniza en ['Hola', 'mundo'] o incluso en sílabas.”

## Embedding

n° 075

### Representación vectorial

**QUÉ ES** Convertir palabras o categorías en listas de números que el modelo puede procesar.

**EJEMPLO** “La palabra 'rey' se convierte en [0.2, 0.8, -0.3, ...] para que la matemática funcione.”

## NLP

n° 076

### Procesamiento de Lenguaje Natural

**QUÉ ES** Área de la IA que trabaja con texto y lenguaje humano.

**EJEMPLO** “Lo que permite que los chatbots entiendan preguntas o que el corrector del celular funcione.”

## Transfer learning

nº 077

### Aprendizaje transferido

**QUÉ ES** Usar un modelo ya entrenado en una tarea y adaptarlo a otra con pocos datos nuevos.

**EJEMPLO** “Como aprender a conducir una camioneta si ya sabés manejar un auto.”

## Fine-tuning

nº 078

### Ajuste fino

**QUÉ ES** Tomar un modelo preentrenado y entrenarlo un poco más con datos específicos.

**EJEMPLO** “Coger un modelo de lenguaje y entrenarlo con tus emails internos para que hable como tu empresa.”

## Hallucination AI

nº 079

### Alucinación de IA

**QUÉ ES** Cuando un modelo genera información falsa con total confianza.

**EJEMPLO** “Le preguntás por un libro y te inventa el título, el autor y hasta citas que nunca existieron.”

## RAG

nº 080

### Generación Aumentada por Recuperación

**QUÉ ES** Técnica que conecta el modelo con una base de conocimiento externa antes de responder.

**EJEMPLO** “En vez de inventar la respuesta, el modelo primero busca en tus documentos y luego responde.”

## LLM

nº 081

### Modelo de Lenguaje Grande

**QUÉ ES** Modelo de IA entrenado con enormes cantidades de texto para entender y generar lenguaje.

**EJEMPLO** “GPT, Claude y Gemini son ejemplos de LLMs.”

## Feature engineering

nº 082

### Ingeniería de características

**QUÉ ES** Crear o transformar variables para que el modelo aprenda mejor.

**EJEMPLO** “En vez de dar la fecha de nacimiento, calcular la edad actual.”

## Feature importance

nº 083

### Importancia de variables

**QUÉ ES** Medida de cuánto influye cada variable en las predicciones del modelo.

**EJEMPLO** “Saber que el historial de pagos explica el 60% de si alguien pagará su préstamo.”

## Anomaly detection

nº 084

### Detección de anomalías

**QUÉ ES** Identificar datos que se comportan de manera inusual respecto al patrón normal.

**EJEMPLO** “Detectar una transacción rara en tu tarjeta a las 3am en otro país.”

*donde viven los datos*



# Bases de Datos

*donde viven los datos cuando  
nadie los mira. spoiler: en tablas.*

*24 términos*

## Database

n° 085

*Base de datos*

**QUÉ ES** Sistema organizado para almacenar, gestionar y recuperar información.

**EJEMPLO** “La libreta de contactos de tu celular es una base de datos simple.”

## Table

n° 086

*Tabla*

**QUÉ ES** Estructura de filas y columnas donde se almacenan los datos, como una hoja de cálculo.

**EJEMPLO** “Una tabla de clientes tiene columnas como nombre, email y ciudad.”

## Row

n° 087

*Fila / registro*

**QUÉ ES** Un registro individual en una tabla, con todos sus campos.

**EJEMPLO** “La fila de Juan Pérez con su email y ciudad es un registro en la tabla de clientes.”

## Column

n° 088

*Columna / campo*

**QUÉ ES** Un atributo o tipo de dato en una tabla.

**EJEMPLO** “Email, nombre y teléfono son columnas en la tabla de clientes.”

## Query

n° 089

*Consulta*

**QUÉ ES** Pregunta o instrucción que se le hace a la base de datos para obtener o modificar datos.

**EJEMPLO** “Pedir 'dame todos los clientes de Buenos Aires' es una consulta.”

## SQL

n° 090

*Lenguaje de Consulta Estructurado*

**QUÉ ES** Lenguaje estándar para hablar con bases de datos relacionales.

**EJEMPLO** “SELECT, FROM y WHERE son comandos de SQL que cualquier analista usa.”

## Join

n° 091

*Unión de tablas*

**QUÉ ES** Operación que combina filas de dos tablas basándose en una columna en común.

**EJEMPLO** “Unir la tabla de pedidos con la de clientes usando el ID de cliente.”

## Primary key

n° 092

*Clave primaria*

**QUÉ ES** Columna que identifica de forma única a cada registro en una tabla.

**EJEMPLO** “El DNI en una tabla de personas: no puede repetirse ni estar vacío.”

## Foreign key

n° 093

### Clave foránea

**QUÉ ES** Columna que apunta a la clave primaria de otra tabla, creando una relación.

**EJEMPLO** “El ID de cliente en la tabla de pedidos conecta con la tabla de clientes.”

## Index

n° 094

### Índice

**QUÉ ES** Estructura que acelera las búsquedas en una tabla, como el índice de un libro.

**EJEMPLO** “Sin índice, buscar un email en millones de registros es muy lento.”

## Schema

n° 095

### Esquema

**QUÉ ES** La estructura y organización de una base de datos: qué tablas hay y cómo se relacionan.

**EJEMPLO** “El plano de la base de datos: qué columnas tiene cada tabla y de qué tipo son.”

## NULL

n° 096

### Valor vacío / ausente

**QUÉ ES** Indica que un campo no tiene valor. Diferente a cero o a cadena vacía.

**EJEMPLO** “Si un cliente no tiene teléfono registrado, ese campo es NULL, no es cero.”

## CRUD

n° 097

### Crear, Leer, Actualizar, Eliminar

**QUÉ ES** Las cuatro operaciones básicas sobre cualquier dato.

**EJEMPLO** “Registrarte (Create), ver tu perfil (Read), cambiar tu foto (Update), borrar tu cuenta (Delete).”

## Relational database

n° 098

### Base de datos relacional

**QUÉ ES** Base de datos que organiza la información en tablas relacionadas entre sí.

**EJEMPLO** “PostgreSQL, MySQL y SQLite son bases de datos relacionales.”

## NoSQL

n° 099

### Base de datos no relacional

**QUÉ ES** Base de datos que no usa tablas tradicionales; puede guardar documentos o pares clave-valor.

**EJEMPLO** “MongoDB guarda datos como documentos JSON, sin necesidad de columnas fijas.”

## Data warehouse

n° 100

### Almacén de datos

**QUÉ ES** Base de datos diseñada para análisis histórico y reportes, no para operaciones diarias.

**EJEMPLO** “Donde se guardan 5 años de ventas para analizar tendencias sin afectar el sistema de ventas.”

**Data lake**

nº 101

*Lago de datos*

**QUÉ ES** Repositorio que guarda datos en su formato original, estructurados y no estructurados.

**EJEMPLO** “Un disco duro gigante donde se guarda todo: emails, logs, fotos, tablas, sin organizar primero.”

**ETL**

nº 102

*Extraer, Transformar, Cargar*

**QUÉ ES** Proceso de mover datos de una fuente, limpiarlos y cargarlos en otro sistema.

**EJEMPLO** “Sacar ventas del sistema de la tienda, limpiarlas y meterlas en el almacén de análisis.”

**ELT**

nº 103

*Extraer, Cargar, Transformar*

**QUÉ ES** Variante donde primero se carga y luego se transforma dentro del destino.

**EJEMPLO** “Meter los datos crudos en el almacén primero y luego transformarlos con SQL.”

**View**

nº 104

*Vista*

**QUÉ ES** Consulta guardada que se comporta como una tabla virtual.

**EJEMPLO** “Una vista de ‘ventas del mes’ que siempre muestra los datos actualizados sin copiarlos.”

**Normalization**

nº 105

*Normalización de base de datos*

**QUÉ ES** Organizar la base de datos para reducir la redundancia de datos.

**EJEMPLO** “En vez de repetir el nombre del cliente en cada pedido, guardarlo una sola vez.”

**Transaction**

nº 106

*Transacción*

**QUÉ ES** Conjunto de operaciones que se ejecutan todas juntas o ninguna.

**EJEMPLO** “Transferir dinero: restar de una cuenta y sumar en otra deben pasar juntas o no pasar.”

**Stored procedure**

nº 107

*Procedimiento almacenado*

**QUÉ ES** Bloque de código SQL guardado en la base de datos que se puede ejecutar cuando se necesite.

**EJEMPLO** “Como una macro de Excel pero en la base de datos.”

**Data mart**

nº 108

*Subconjunto de datos por área*

**QUÉ ES** Versión reducida de un data warehouse enfocada en un área específica del negocio.

**EJEMPLO** “El data mart de ventas solo tiene los datos que necesita el equipo comercial.”

*las plomerías*

# Ingeniería de Datos

*las plomerías del mundo data. sin esto, no llega agua a la canilla.*

*20 términos*

## Pipeline

nº 109

*Flujo de procesamiento de datos*

**QUÉ ES** Serie de pasos automatizados por donde pasan los datos desde el origen hasta el destino.

**EJEMPLO** “ Como una línea de ensamblaje: los datos entran crudos y salen listos para analizar.”

## Batch processing

nº 110

*Procesamiento por lotes*

**QUÉ ES** Procesar un gran volumen de datos acumulados en intervalos de tiempo.

**EJEMPLO** “ Generar los reportes de ventas del día anterior cada mañana a las 6am.”

## Streaming

nº 111

*Procesamiento en tiempo real*

**QUÉ ES** Procesar datos a medida que llegan, sin esperar a acumularlos.

**EJEMPLO** “ Detectar una transacción fraudulenta en segundos mientras el cliente paga.”

## API

nº 112

*Interfaz de programación de aplicaciones*

**QUÉ ES** Puerta de acceso estandarizada para que dos sistemas intercambien datos.

**EJEMPLO** “ Cuando una app del clima consulta datos de temperatura en un servidor externo, usa una API.”

## Webhook

nº 113

*Notificación automática entre sistemas*

**QUÉ ES** Mensaje que un sistema envía automáticamente a otro cuando ocurre un evento.

**EJEMPLO** “ Cuando alguien paga en tu web, el sistema de pagos envía un webhook a tu app para avisar.”

## Data ingestion

nº 114

*Ingesta de datos*

**QUÉ ES** Proceso de recopilar y cargar datos de múltiples fuentes en un sistema central.

**EJEMPLO** “ Juntar datos de ventas, marketing y soporte en un solo lugar para analizarlos.”

## Data cleaning

nº 115

*Limpieza de datos*

**QUÉ ES** Proceso de corregir o eliminar datos incorrectos, duplicados o incompletos.

**EJEMPLO** “ Eliminar filas duplicadas, corregir fechas mal escritas y rellenar campos vacíos.”

## Data wrangling

nº 116

*Preparación de datos*

**QUÉ ES** Transformar datos crudos y desordenados en un formato útil para el análisis.

**EJEMPLO** “ Convertir una columna de fechas en distintos formatos a un formato único consistente.”

## Deduplication

nº 117

*Eliminación de duplicados*

**QUÉ ES** Identificar y eliminar registros repetidos en un conjunto de datos.

**EJEMPLO** “Si el mismo cliente aparece dos veces con distinto email, deduplicar los uno en uno.”

## Orchestration

nº 118

*Orquestación de datos*

**QUÉ ES** Gestionar y coordinar la ejecución de múltiples pipelines en el orden correcto.

**EJEMPLO** “Asegurarse de que primero se limpian, luego se transforman y luego se cargan.”

## Latency

nº 119

*Latencia*

**QUÉ ES** Tiempo que tarda un dato desde que se genera hasta que está disponible para usar.

**EJEMPLO** “Si los datos de ventas de hoy aparecen en el reporte mañana, la latencia es de un día.”

## Scalability

nº 120

*Escalabilidad*

**QUÉ ES** Capacidad de un sistema para manejar más datos o usuarios sin degradarse.

**EJEMPLO** “Un sistema escalable funciona igual de bien con 1000 usuarios que con 1 millón.”

## Cloud computing

nº 121

*Computación en la nube*

**QUÉ ES** Usar servidores y recursos computacionales remotos a través de internet.

**EJEMPLO** “En vez de tener tu propio servidor, alquilas capacidad en AWS, Google Cloud o Azure.”

## Data governance

nº 122

*Gobernanza de datos*

**QUÉ ES** Políticas y procesos para asegurar la calidad, seguridad y uso correcto de los datos.

**EJEMPLO** “Definir quién puede ver los datos de clientes, cómo se almacenan y cuánto tiempo se guardan.”

## Metadata

nº 123

*Metadatos*

**QUÉ ES** Datos que describen otros datos: cuándo se crearon, quién los creó, de dónde vienen.

**EJEMPLO** “El nombre del archivo, la fecha de creación y el tamaño son metadatos del archivo.”

## Data lineage

nº 124

*Linaje de datos*

**QUÉ ES** Registro del recorrido de un dato desde su origen hasta su uso final.

**EJEMPLO** “Saber que este número de ventas vino del sistema de caja, pasó por ETL y llegó al reporte.”

## Version control

nº 125

*Control de versiones*

**QUÉ ES** Sistema que registra los cambios en archivos o código a lo largo del tiempo.

**EJEMPLO** “Git: si algo falla, podés volver a la versión anterior del código.”

## CI/CD

nº 126

*Integración y entrega continua*

**QUÉ ES** Prácticas que automatizan las pruebas y el despliegue de cambios en el sistema.

**EJEMPLO** “Cada vez que alguien actualiza el código, el sistema lo prueba y lo publica si está bien.”

## Data catalog

nº 127

*Catálogo de datos*

**QUÉ ES** Inventario de todos los datos disponibles en la organización, con descripción y ubicación.

**EJEMPLO** “Como la biblioteca de datos de la empresa: buscás 'ventas' y te dice dónde están.”

## Data enrichment

nº 128

*Enriquecimiento de datos*

**QUÉ ES** Agregar información de fuentes externas para completar los datos existentes.

**EJEMPLO** “Agregar el código postal a una tabla de clientes que solo tiene el nombre y email.”



*que se pongan guapos*

# Visualización de Datos

*donde los números se ponen  
guapos y por fin se dejan entender.*

*20 términos*

## Chart

nº 129

### Gráfico

**QUÉ ES** Representación visual de datos para facilitar su comprensión.

**EJEMPLO** “Un gráfico de barras de ventas mensuales es más fácil de leer que una tabla de números.”

## Bar chart

nº 130

### Gráfico de barras

**QUÉ ES** Gráfico que usa barras horizontales o verticales para comparar categorías.

**EJEMPLO** “Comparar las ventas de cada producto en un mes con una barra por producto.”

## Line chart

nº 131

### Gráfico de líneas

**QUÉ ES** Gráfico que conecta puntos de datos con líneas para mostrar cambios en el tiempo.

**EJEMPLO** “Ver cómo subieron o bajaron las visitas a la web semana a semana.”

## Pie chart

nº 132

### Gráfico de pastel / circular

**QUÉ ES** Círculo dividido en sectores que muestran la proporción de cada categoría.

**EJEMPLO** “Ver qué porcentaje de ventas vino de cada región del país.”

## Scatter plot

nº 133

### Diagrama de dispersión

**QUÉ ES** Gráfico que muestra la relación entre dos variables usando puntos.

**EJEMPLO** “En el eje X el precio de un producto y en el Y sus ventas para ver si hay relación.”

## Heatmap

nº 134

### Mapa de calor

**QUÉ ES** Tabla donde los colores representan la magnitud de los valores.

**EJEMPLO** “Ver de qué hora y qué día de la semana hay más tráfico en el sitio web.”

## Histogram

nº 135

### Histograma

**QUÉ ES** Gráfico de barras que muestra la frecuencia de valores dentro de rangos.

**EJEMPLO** “Ver cuántos clientes tienen entre 0-10, 10-20, 20-30 compras en el año.”

## Treemap

nº 136

### Mapa de árbol

**QUÉ ES** Gráfico que muestra jerarquías y proporciones usando rectángulos anidados.

**EJEMPLO** “Ver cuánto representa cada categoría de producto dentro del total de ventas.”

## Waterfall chart

nº 137

### Gráfico cascada

**QUÉ ES** Gráfico que muestra cómo los valores positivos y negativos suman al total.

**EJEMPLO** “ Ver cómo los ingresos, gastos y descuentos llegaron a la ganancia final del mes.”

## Box plot

nº 138

### Diagrama de caja

**QUÉ ES** Gráfico que muestra la distribución de datos con la mediana, cuartiles y outliers.

**EJEMPLO** “ Ver rápidamente si los tiempos de entrega son consistentes o hay muchos casos atípicos.”

## Time series

nº 139

### Serie temporal

**QUÉ ES** Secuencia de datos medidos en intervalos de tiempo regulares.

**EJEMPLO** “ El precio del dólar cada día durante un año es una serie temporal.”

## Drill-down

nº 140

### Profundizar en datos

**QUÉ ES** Navegar desde un resumen a un nivel más detallado de información.

**EJEMPLO** “ Clic en 'ventas de Argentina' para ver las ventas por provincia, y luego por ciudad.”

## Granularity

nº 141

### Granularidad

**QUÉ ES** Nivel de detalle de los datos: por día, por hora, por minuto.

**EJEMPLO** “ Datos por hora tienen mayor granularidad que datos por mes.”

## Tooltip

nº 142

### Información emergente

**QUÉ ES** Texto que aparece al pasar el cursor sobre un elemento del gráfico.

**EJEMPLO** “ Pasar el mouse sobre una barra y ver el valor exacto: 'Ventas: 45,230'.”

## Annotation

nº 143

### Anotación en gráfico

**QUÉ ES** Texto o marca agregada a un gráfico para explicar algo relevante.

**EJEMPLO** “ Una flecha en el gráfico que dice 'acá lanzamos el nuevo producto'.”

## Stacked chart

nº 144

### Gráfico apilado

**QUÉ ES** Variante del gráfico de barras donde se apilan categorías dentro de cada barra.

**EJEMPLO** “ Cada barra mensual apila ventas online, en tienda y por teléfono.”

## Legend

nº 145

### Leyenda

**QUÉ ES** Clave que explica qué representa cada color o forma en un gráfico.

**EJEMPLO** “La cajita que dice 'azul = ventas 2024, rojo = ventas 2023.’”

## Axis

nº 146

### Eje del gráfico

**QUÉ ES** Línea de referencia del gráfico. El eje X es horizontal, el Y es vertical.

**EJEMPLO** “En un gráfico de ventas por mes: X tiene los meses, Y tiene los montos.”

## Data label

nº 147

### Etiqueta de dato

**QUÉ ES** Valor numérico mostrado directamente sobre o dentro de una barra o punto.

**EJEMPLO** “El número '3,450' escrito encima de una barra para no tener que ir al eje.”

## Funnel chart

nº 148

### Gráfico de embudo

**QUÉ ES** Gráfico que muestra la reducción de usuarios en cada etapa de un proceso.

**EJEMPLO** “Ver visualmente cómo de 1000 visitas se llega a 80 compras pasando por cada paso.”



*siglas en reunión*

# Business Intelligence y Negocio

*las siglas que se tiran en reuniones  
estratégicas. ahora las tirás vos.*

*23 términos*

**Business Intelligence (BI)**

nº 149

*Inteligencia de negocios***QUÉ ES** Procesos y herramientas para transformar datos en información útil para decidir.**EJEMPLO** “ Usar datos de ventas, marketing y finanzas para decidir dónde invertir el próximo trimestre. ”**Report**

nº 150

*Informe***QUÉ ES** Documento o vista que presenta datos organizados para comunicar una situación.**EJEMPLO** “ El reporte mensual de ventas que el equipo revisa cada primer lunes del mes. ”**Scorecard**

nº 151

*Tarjeta de puntuación***QUÉ ES** Vista compacta de los indicadores clave con su estado (verde/amarillo/rojo).**EJEMPLO** “ Una pantalla que muestra si cada KPI está en meta, cerca o lejos del objetivo. ”**OKR**

nº 152

*Objetivos y Resultados Clave***QUÉ ES** Marco para definir metas ambiciosas y medir su progreso con indicadores específicos.**EJEMPLO** “ Objetivo: ser la marca más reconocida. Resultado clave: 60% de awareness. ”**SLA**

nº 153

*Acuerdo de Nivel de Servicio***QUÉ ES** Compromiso formal sobre el nivel de calidad o tiempo de respuesta de un servicio.**EJEMPLO** “ Responder todos los tickets de soporte en menos de 24 horas es un SLA. ”**Forecast**

nº 154

*Pronóstico***QUÉ ES** Estimación de valores futuros basada en datos históricos y tendencias.**EJEMPLO** “ Predecir cuántas unidades se venderán el próximo trimestre para ordenar el inventario. ”**CAC**

nº 155

*Costo de Adquisición de Cliente***QUÉ ES** Cuánto gasta una empresa en promedio para conseguir un nuevo cliente.**EJEMPLO** “ Si gastaste 10,000 en marketing y conseguiste 100 clientes nuevos, el CAC es 100. ”**LTV / CLV**

nº 156

*Valor de Vida del Cliente***QUÉ ES** Cuánto dinero genera un cliente durante toda su relación con la empresa.**EJEMPLO** “ Si un cliente gasta 50 por mes durante 3 años, su LTV es 1,800. ”

**MRR**

n° 157

*Ingreso Mensual Recurrente*

**QUÉ ES** Total de ingresos que la empresa espera recibir cada mes de manera predecible.

**EJEMPLO** “ Si tenés 1000 suscriptores que pagan 20 por mes, el MRR es 20,000.”

**ARR**

n° 158

*Ingreso Anual Recurrente*

**QUÉ ES** Versión anualizada del MRR. Se multiplica el MRR por 12.

**EJEMPLO** “ Un MRR de 20,000 equivale a un ARR de 240,000.”

**NPS**

n° 159

*Puntuación Neta del Promotor*

**QUÉ ES** Medida de la lealtad del cliente: cuántos te recomendarían vs. cuántos no.

**EJEMPLO** “ Pregunta del 0 al 10 cuánto nos recomendarías. Un NPS alto indica clientes muy leales.”

**CSAT**

n° 160

*Satisfacción del Cliente*

**QUÉ ES** Medición directa de qué tan satisfecho está un cliente con una interacción específica.

**EJEMPLO** “ La encuesta de 1 a 5 estrellas que aparece al cerrar un ticket de soporte.”

**DAU / MAU**

n° 161

*Usuarios Activos Diarios y Mensuales*

**QUÉ ES** Cantidad de usuarios únicos que usan el producto en un día o en un mes.

**EJEMPLO** “ Si 10,000 personas abren tu app hoy, el DAU es 10,000.”

**Burn rate**

n° 162

*Tasa de consumo de efectivo*

**QUÉ ES** Velocidad a la que una empresa gasta su dinero, generalmente por mes.

**EJEMPLO** “ Si tenés 1 millón en el banco y gastás 100 mil por mes, el burn rate es 100 mil.”

**Runway**

n° 163

*Tiempo de vida financiero*

**QUÉ ES** Cuánto tiempo le queda a la empresa antes de quedarse sin dinero.

**EJEMPLO** “ Con 1M en el banco y burn rate de 100K, tenés 10 meses de runway.”

**Market share**

n° 164

*Cuota de mercado*

**QUÉ ES** Porcentaje del mercado total que controla una empresa.

**EJEMPLO** “ Si el mercado mueve 100M y vos vendés 15M, tenés 15% de cuota de mercado.”

## North Star Metric

n° 165

*Métrica estrella del norte*

**QUÉ ES** El número que mejor captura el valor que genera el producto para sus usuarios.

**EJEMPLO** “Para Spotify son los minutos de música escuchados. Todo lo demás sirve a esa métrica.”

## Vanity metric

n° 166

*Métrica de ego*

**QUÉ ES** Número que parece impresionante pero no dice nada sobre el estado real del negocio.

**EJEMPLO** “Tener 100,000 seguidores pero 0 ventas. Los seguidores son la vanity metric.”

## Leading indicator

n° 167

*Indicador anticipado*

**QUÉ ES** Métrica que predice resultados futuros; cambia antes de que cambie el resultado final.

**EJEMPLO** “El número de demos agendadas esta semana predice las ventas del mes que viene.”

## Lagging indicator

n° 168

*Indicador rezagado*

**QUÉ ES** Métrica que confirma lo que ya pasó; cambia después del resultado.

**EJEMPLO** “Los ingresos del mes son un indicador rezagado: confirman el resultado cuando ya terminó.”

## Pivot

n° 169

*Cambio de estrategia*

**QUÉ ES** Cambio significativo en la dirección del producto o negocio basado en los datos.

**EJEMPLO** “Pasarse de vender a empresas a vender a consumidores finales porque los datos lo sugieren.”

## GMV

n° 170

*Volumen Total de Transacciones*

**QUÉ ES** Valor total de las ventas procesadas a través de una plataforma en un período.

**EJEMPLO** “Si en tu marketplace se vendieron productos por 5M en el mes, el GMV es 5M.”

## ARPU

n° 171

*Ingreso Promedio por Usuario*

**QUÉ ES** Cuánto ingresa la empresa en promedio por cada usuario activo.

**EJEMPLO** “Si tenés 1000 usuarios y generás 5000 al mes, tu ARPU es 5 por usuario.”

A large, light gray, stylized number '08' is centered in the upper half of the page. A small, white, rounded rectangular tag with a thin purple border is attached to the right side of the '8', containing the text 'para nerdas' in a handwritten-style font.

*para nerdas*

# Conceptos Avanzados y Calidad de Datos

*cuando ya sabés lo básico y  
empezás a preguntar por qué se  
rompen las cosas.*

*20 términos*

## Data drift

nº 172

*Deriva de datos*

**QUÉ ES** Cuando la distribución de los datos de entrada cambia con el tiempo y el modelo se desactualiza.

**EJEMPLO** “Un modelo de crédito entrenado en 2019 puede fallar en 2024 porque los hábitos cambiaron.”

## Concept drift

nº 173

*Deriva de concepto*

**QUÉ ES** Cuando la relación entre las variables y el resultado cambia con el tiempo.

**EJEMPLO** “Lo que significaba 'cliente en riesgo' en 2020 puede ser diferente en 2025.”

## Explainability (XAI)

nº 174

*Explicabilidad de la IA*

**QUÉ ES** Capacidad de entender y comunicar por qué un modelo tomó una decisión.

**EJEMPLO** “Que el banco pueda decirte por qué tu préstamo fue rechazado, no solo que fue rechazado.”

## Fairness AI

nº 175

*Equidad en IA*

**QUÉ ES** Asegurarse de que el modelo no discrimina injustamente a grupos de personas.

**EJEMPLO** “Un modelo de selección de empleados no debería penalizar a candidatos por su género o edad.”

## Data augmentation

nº 176

*Aumento de datos*

**QUÉ ES** Generar nuevas variaciones de datos existentes para entrenar mejor el modelo.

**EJEMPLO** “Rotar y voltear fotos de gatos para que el modelo aprenda en cualquier ángulo.”

## Dimensionality reduction

nº 177

*Reducción de dimensionalidad*

**QUÉ ES** Reducir el número de variables manteniendo la mayor información posible.

**EJEMPLO** “Pasarse de 100 preguntas de encuesta a 5 factores principales que explican las respuestas.”

## PCA

nº 178

*Análisis de Componentes Principales*

**QUÉ ES** Técnica de reducción de dimensionalidad que combina variables correlacionadas.

**EJEMPLO** “Combinar altura, peso y edad en un 'índice de condición física' que resume los tres.”

## Imputation

nº 179

*Imputación de datos faltantes*

**QUÉ ES** Rellenar valores ausentes en un dataset con estimaciones razonables.

**EJEMPLO** “Rellenar los ingresos faltantes con el promedio del grupo de edad correspondiente.”

## One-hot encoding

n° 180

*Codificación binaria por categoría*

**QUÉ ES** Convertir una variable categórica en múltiples columnas de 0 y 1.

**EJEMPLO** “La columna 'país' con valores ARG, MEX, COL se convierte en tres columnas de 0 y 1.”

## Data quality

n° 181

*Calidad de datos*

**QUÉ ES** Medida de qué tan precisos, completos y consistentes son los datos.

**EJEMPLO** “Datos de alta calidad no tienen duplicados, tienen fechas correctas y están completos.”

## Sampling bias

n° 182

*Sesgo de muestreo*

**QUÉ ES** Error que ocurre cuando la muestra no representa bien a la población total.

**EJEMPLO** “Encuestar solo a usuarios de iPhone sobre preferencias de celular da resultados sesgados.”

## Microservices

n° 183

*Microservicios*

**QUÉ ES** Arquitectura donde cada función del sistema es un servicio independiente.

**EJEMPLO** “En vez de una app gigante, tenés servicios separados para pagos, usuarios y notificaciones.”

## Containerization

n° 184

*Contenedorización*

**QUÉ ES** Empaquetar una aplicación con todo lo que necesita para correr en cualquier entorno.

**EJEMPLO** “Docker: como poner tu app en una caja estandarizada que funciona en cualquier computadora.”

## A/B/n testing

n° 185

*Prueba multivariante*

**QUÉ ES** Extensión del A/B testing donde se prueban tres o más versiones simultáneamente.

**EJEMPLO** “Probar cuatro diseños distintos de portada al mismo tiempo para ver cuál convierte más.”

## Labeling

n° 186

*Etiquetado de datos*

**QUÉ ES** Proceso de agregar etiquetas a datos crudos para usarlos en ML supervisado.

**EJEMPLO** “Personas que marcan manualmente si cada email es spam o no para entrenar un filtro.”

## Ground truth

n° 187

*Verdad de referencia*

**QUÉ ES** Los datos correctos y verificados con los que se comparan las predicciones del modelo.

**EJEMPLO** “Fotos etiquetadas a mano por humanos son la ground truth para un modelo de visión.”

## Synthetic data

nº 188

*Datos sintéticos*

**QUÉ ES** Datos generados artificialmente que imitan datos reales.

**EJEMPLO** “Generar rayos X falsos para entrenar un modelo médico cuando hay pocos reales.”

## Recommendation system

nº 189

*Sistema de recomendación*

**QUÉ ES** Modelo que predice qué productos o contenidos le gustarán a un usuario.

**EJEMPLO** “Lo que hace Netflix cuando te dice 'porque viste X, quizás te guste Y'.”

## Time series forecasting

nº 190

*Pronóstico de series temporales*

**QUÉ ES** Predecir valores futuros basándose en patrones históricos de datos en el tiempo.

**EJEMPLO** “Predecir cuánto se venderá en diciembre basándose en los últimos 5 diciembrees.”

## Normalization (stats)

nº 191

*Normalización de escala*

**QUÉ ES** Escalar los valores de una variable para que estén en un rango estándar como 0 a 1.

**EJEMPLO** “Si una variable va de 0 a 1000 y otra de 0 a 1, normalizarlas evita que la primera domine el modelo.”

*glosario de reuniones*

# Términos Frecuentes en Reuniones

*guardá esto. esto es lo que se dice y lo que significa de verdad.*

*9 términos*

**Data-driven**

nº 192

*Basado en datos***QUÉ ES** Tomar decisiones apoyadas en datos en lugar de intuición o suposiciones.**EJEMPLO** “En vez de decidir qué producto lanzar por corazonada, analizar qué piden los clientes.”**Actionable**

nº 193

*Accionable / útil para actuar***QUÉ ES** Información que lleva directamente a una acción concreta.**EJEMPLO** “Saber que el 60% abandona el carrito en el paso de pago es accionable: podés mejorarlo.”**Single source of truth**

nº 194

*Fuente única de verdad***QUÉ ES** Un solo sistema o dataset que todos en la empresa usan como referencia oficial.**EJEMPLO** “Si marketing tiene un número de ventas y finanzas tiene otro, no hay fuente única de verdad.”**Data silos**

nº 195

*Silos de datos***QUÉ ES** Cuando los datos de distintos equipos están aislados y no se comparten.**EJEMPLO** “Ventas tiene sus datos en Excel, marketing en otra herramienta y no se pueden cruzar.”**Real-time**

nº 196

*En tiempo real***QUÉ ES** Datos o procesos que se actualizan y procesan de inmediato, sin demora.**EJEMPLO** “Ver cuántas personas están en tu web ahora mismo es un dato en tiempo real.”**Baseline**

nº 197

*Línea base***QUÉ ES** El valor de referencia antes de hacer un cambio, contra el que se compara el resultado.**EJEMPLO** “Si la tasa de conversión era del 2% antes del experimento, ese 2% es la línea base.”**Iteration**

nº 198

*Iteración***QUÉ ES** Una ronda de mejora sobre algo ya existente, basada en los resultados obtenidos.**EJEMPLO** “Probar el diseño A, ver los resultados, mejorarlo y probar el diseño B: eso es iterar.”**Proxy metric**

nº 199

*Métrica sustituta***QUÉ ES** Una métrica que se usa para estimar otra que es difícil de medir directamente.**EJEMPLO** “Medir el tiempo en la app como proxy de satisfacción cuando no se puede preguntar directamente.”

## Volatility

nº 200

### Volatilidad

**QUÉ ES** Medida de qué tanto varía un valor a lo largo del tiempo. Alta volatilidad = mucha variación.

**EJEMPLO** “Una acción que sube 10% un día y baja 12% al siguiente tiene alta volatilidad.”

· FIN DEL VOLUMEN ·

ahora sí  
sos *de datos*.

*fin.*

*no te lo dijeron ,  
pero siempre lo  
fuiste.*

CURADURÍA

milo · @databymilo  
diccionario anti-jerga · vol. 01

EDICIÓN

primera edición · 2026  
200 términos · 9 módulos