

Introducción al mundo de los datos

Del dato a la decisión

Manuel Fernández San Martín
Juan Ignacio da Torre

Computational Economics
Universidad de Buenos Aires

Segundo Cuatrimestre 2025

Objetivo: Comprender qué es la DATA y conceptos claves de este mundo

Objetivo de la charla

¿Qué vamos a ver hoy?

- Entender qué son los datos hoy en día.
- Conocer conceptos clave del ecosistema data.
- Ver qué hace realmente un técnico o analista de datos.
- Conocer hacia dónde va el mundo de los datos (Tendencias).

Datos en la vida cotidiana

¿Qué tipo de datos podemos obtener de estas situaciones?



Personas consumiendo contenido en streaming



Uso de un reloj inteligente

La Economía del Dato

Del dato bruto al valor económico

Vivimos en la economía del dato

Hoy todo genera datos:

- Transacciones bancarias o Billeteras virtuales
- Sensores
- Libros virtuales y ¿físicos?
- E-commerce & Marketplaces
- Música, Podcast - Entretenimiento General
- Vuelos
- Diarios y Noticias
- Plataformas digitales (redes sociales)
- Registros y logs de servidores
- ¿ad infinitum?

Distinción clave

Datos \neq Información

El dato por sí solo es ruido. El valor real está en el **análisis**.

Big Data: Contexto

El concepto se define tradicionalmente por las "3 V":

- ❶ **Volumen:** La cantidad masiva de datos generados.
- ❷ **Velocidad:** La rapidez con la que entran y deben procesarse.
- ❸ **Variedad:** Datos estructurados (tablas) y no estructurados (video, audio, texto).

La nueva frontera digital (MIT Technology Review, 2016)

"Las redes digitales dan paso a un nuevo subsuelo repleto de recursos por extraer: desde videovigilancia, transacciones bancarias, hasta dispositivos IoT (termostatos, contadores inteligentes, refrigeradores conectados). El que llega primero y toma su control obtiene los recursos que encubren; en este caso, sus riquezas en datos."

Conclusión

Big Data no es solo "muchos datos", es **escala + complejidad**.

Del dato a la decisión

Datos \rightarrow Información \rightarrow Análisis \rightarrow
Decisiones

El valor no está en acumular datos, sino en saber transformarlos.

El Dato como Recurso Estratégico

La potencia de los Big Data (Cédric Durand)

"La potencia de los Big Data radica en un efecto de lupa donde cada algoritmo que utiliza captura de los datos alimenta los algoritmos, y a cambio estos vienen a guiar las conductas, reforzando esos datos."

El ciclo de retroalimentación:

- 1 Datos alimentan algoritmos
- 2 Algoritmos guían conductas
- 3 Conductas generan más datos
- 4 El ciclo se refuerza

Implicancia

Quien controla los datos, controla el ciclo completo de generación de valor

¿Dónde se guarda y cómo? : Data Types

Definición

Los tipos de datos definen cómo se almacena un dato en la memoria y qué operaciones se pueden realizar con él.

Impacto:

- Determinan el espacio en disco.
- Definen la precisión de los cálculos.
- Impactan directamente en análisis y decisiones posteriores.

Tipos de datos más comunes

Numéricos (Numeric)

- **Enteros (Integers):** Cantidad de hijos, edad.
- **Decimales (Floats):** Precio, temperatura, peso.

Categoricos (Categorical)

- **Texto (Strings):** Nombres, descripciones.
- **Clases (Labels):** “Alto/Bajo”, “Aprobado/Desaprobado”.
- **Objetos (object – Pandas):** Texto o mezclas de tipos.

Temporales (Temporal)

- **Fechas (Date):** DD/MM/AAAA.
- **Marcas de tiempo (Timestamps):** Fecha + hora exacta.

¿Por qué los data types importan?

Definir bien el tipo de dato desde el principio permite:

- Evitar errores silenciosos en el código.
- Mejorar la velocidad de los cálculos.
- Mejorar las visualizaciones automáticas.
- Optimizar el almacenamiento.

Atención

Un mal data type genera malas conclusiones.

(Ej: Tratar un código postal como un número y sumarlo).

Gobernanza de datos

¿Qué es?

Es el conjunto de reglas, políticas y procesos para gestionar los datos de una organización.

¿Qué busca la gobernanza?

- **Calidad:** Que el dato sea real.
- **Seguridad:** Quién accede a qué.
- **Trazabilidad:** De dónde vino el dato.
- **Cumplimiento legal:** (GDPR, etc.)
- **Uso responsable.**

► **No todo dato puede usarse libremente.**

Origen y Recolección

Formalismo de cómo nacen y llegan los datos

¿De dónde vienen los datos?

Fuentes Internas

Datos generados por la propia organización:

- Sistemas ERP/CRM.
- Transacciones de venta.
- Logs de servidores.

Fuentes Externas

Datos obtenidos de terceros:

- APIs públicas o privadas.
- Web (Redes sociales, noticias).
- Datasets gubernamentales (Datos públicos).

Formas de recolección de datos

Dependiendo de la fuente y la necesidad:

- **Sistemas transaccionales:** Bases de datos SQL clásicas.
- **Archivos planos:** CSV, Excel, JSON.
- **Streaming:** Datos que llegan continuamente (IoT, clicks).
- **APIs:** Conexión directa entre aplicaciones.
- **Web scraping:** Extracción desde sitios web.

¿Qué es una API?

Definición

Una API (Interfaz de Programación de Aplicaciones) es una interfaz que permite que dos sistemas se comuniquen de forma estructurada.

- Forma estándar de acceso a datos.
- Más estable que el scraping.
- Definida por "contratos" (documentación).

La función del mozo: Una analogía de API

¿Cuál es la función del mozo?



El mozo actúa como una API: toma pedidos (requests) y entrega respuestas (responses) sin que necesites entrar a la cocina

Batch vs Stream

Batch (Por lotes)

- Reportes mensuales/diarios.
- Procesos periódicos nocturnos.
- Análisis histórico.
- *"Foto del ayer"*.

Stream (Tiempo real)

- Eventos en el momento exacto.
- Detección de fraude.
- Monitoreo de servidores.
- *"Película del ahora"*.

Pipeline de datos

El flujo de trabajo estándar en datos:



► El técnico de datos trabaja en varias de estas etapas.

Web Scraping

¿Qué es?

Extracción automatizada de datos desde la web (HTML) simulando ser un usuario.

¿Cuándo se usa?

- Cuando no existe una API oficial.
- Para recolectar datos públicos.

Responsabilidad y Ética

- **Legalidad:** Revisar términos de uso.
 - **Privacidad:** No extraer datos sensibles.
 - **Carga:** No saturar el servidor ajeno.
- ▶ *No todo lo que se puede scrapear, se debe scrapear.*

Automatización y Roles

Qué se hace con los datos en la práctica

Automatización vs Machine Learning

Automatización

- Basada en **reglas** fijas.
- "Si pasa X, hacé Y".
- Ejemplo: n8n, Zapier.
- Elimina tareas repetitivas.

Machine Learning

- Sistemas que **aprenden** de los datos.
- Detectan patrones no explícitos.
- Usos: Predicción, clasificación, recomendación.

Diferencia: La automatización sigue reglas; el ML aprende (o infiere) las reglas.

¿Qué hace hoy un Técnico / Analista de Datos?

Contrario a la creencia popular, no pasamos todo el día haciendo modelos complejos.

- 1 **Limpieza de datos** y Validación.
- 2 Consultas a bases de datos (SQL).
- 3 Automatización de flujos.
- 4 Generación de reportes y documentación.

La regla del 80/20

El 80 % del trabajo es preparación de datos.

Dashboards (Tableros de Control)

Herramientas como PowerBI, Tableau o Streamlit.

- Visualizan información resumida.
- Ayudan a tomar decisiones rápidas.
- **No muestran datos crudos**, muestran métricas (KPIs).

Ejemplos de uso:

- Monitoreo de ventas en tiempo real.
- Estado de operaciones logísticas.
- Matrices de riesgo financiero.

Automatización + IA + Datos

El ecosistema moderno conecta todo:

- 1 Los **Datos** alimentan los procesos.
- 2 Los **Procesos** generan información para decidir.
- 3 La **IA** escala el análisis donde el humano no llega.

Todo está conectado: el dato es el combustible de la IA.

Futuro y Tendencias

Hacia dónde va el mundo de los datos

La Carrera por los Datos

Metáfora de la tierra digital

"Desde el punto de vista de la producción de datos, las actividades son como tierras que esperan ser descubiertas. El que llega primero y toma su control obtiene los recursos que encubren; en este caso, sus riquezas en datos."

— MIT Technology Review (2016)

El nuevo territorio digital incluye:

- Imágenes de videovigilancia
- Registros de transacciones (tickets, pagos)
- Datos de dispositivos IoT conectados
- Interacciones en redes sociales
- Formularios y servicios en línea

El futuro: IA Avanzada y AGI

Hacia dónde vamos:

- Más automatización inteligente.
- Más responsabilidad en el manejo de datos.

Sobre la AGI (Inteligencia Artificial General)

- Se busca una IA cada vez más autónoma.
- La AGI (igualar a la mente humana en todo) aún no existe.
- Hay mucho *hype*, pero poca certeza de fechas.

Hoy usamos IA especializada (estrecha) para tareas concretas.

Computación Cuántica

- Promete acelerar exponencialmente ciertos problemas matemáticos.
- **Complementa** a la computación clásica, no la borra.
- Aún está en fase experimental y de desarrollo.

Realidad

La computación cuántica resolverá problemas complejos de optimización, pero **no reemplaza tu Excel ni tus dashboards** en el corto plazo.

Blockchain y Smart Contracts

Blockchain

- Registro distribuido e inmutable.
- Aporta **trazabilidad** y **confianza**.
- No es para velocidad, es para seguridad.

Smart Contracts

- Reglas de negocio escritas en código.
- Ejecución automática sin intermediarios.

No reemplaza a las bases de datos, resuelve el problema de la confianza.

Qué significa esto para vos

Entorno cambiante

Las herramientas (Python, Pandas, PowerBI, ChatGPT) cambian constantemente.

Lo que permanece:

- Los fundamentos de la lógica.
- La capacidad de hacerse preguntas sobre los datos.
- El pensamiento crítico.

Datos → Información → Decisiones

- El futuro no es aprender todas las herramientas de memoria.
- Es entender los datos y saber qué preguntarles.

Preguntas

¿Preguntas?

No hay preguntas malas, solo datos que aún no miramos.

Fuentes y bibliografía

- Google Cloud — Data Engineering & Data Types Documentation
- AWS — What is Big Data?
- IBM — Data Governance Overview
- Apache Foundation — Data Pipelines & Streaming
- Netflix Tech Blog — Data, Analytics & Recommendations
- Martin Kleppmann — *Designing Data-Intensive Applications*
- Provost & Fawcett — *Data Science for Business*

Las imágenes utilizadas son ilustrativas y con fines educativos.