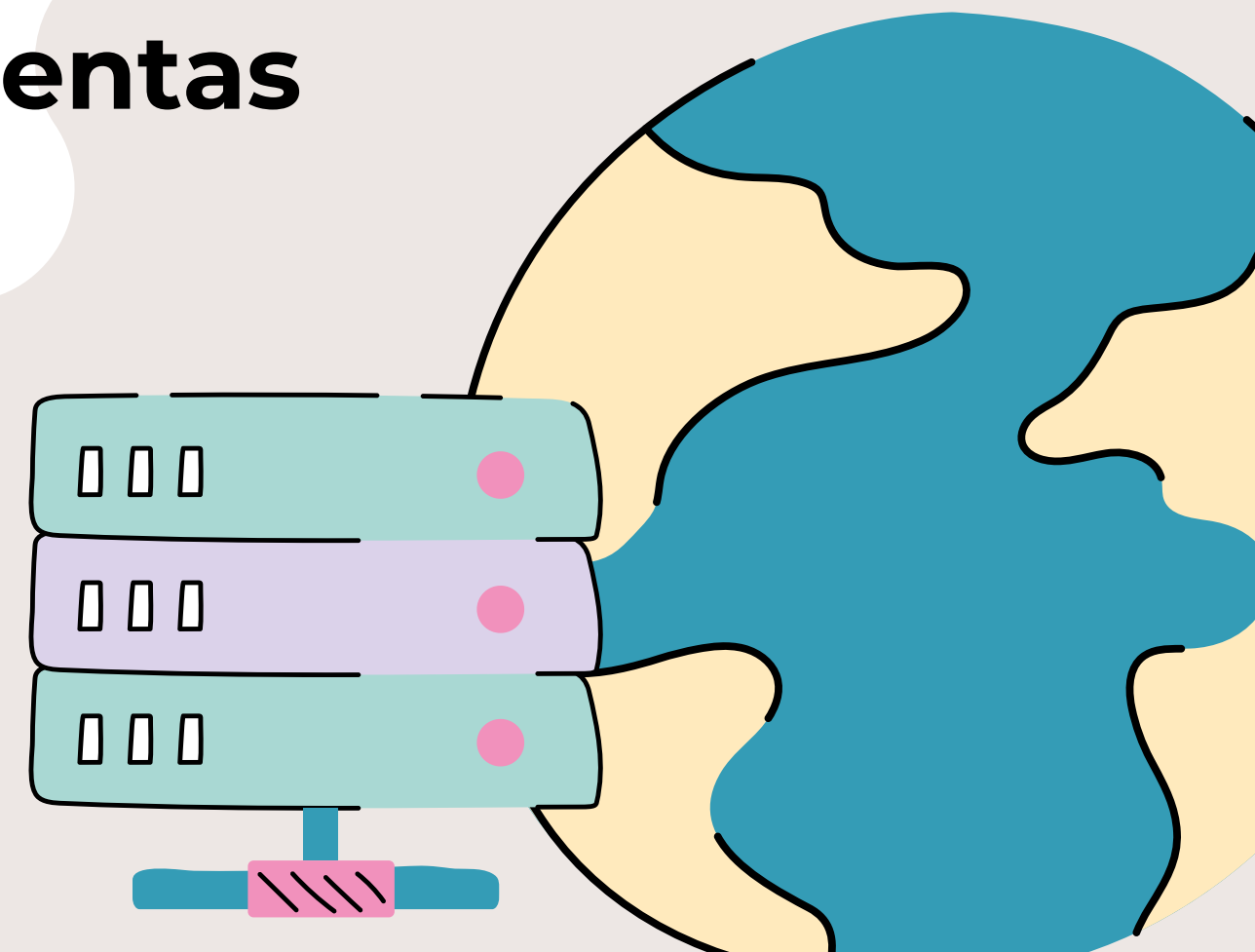
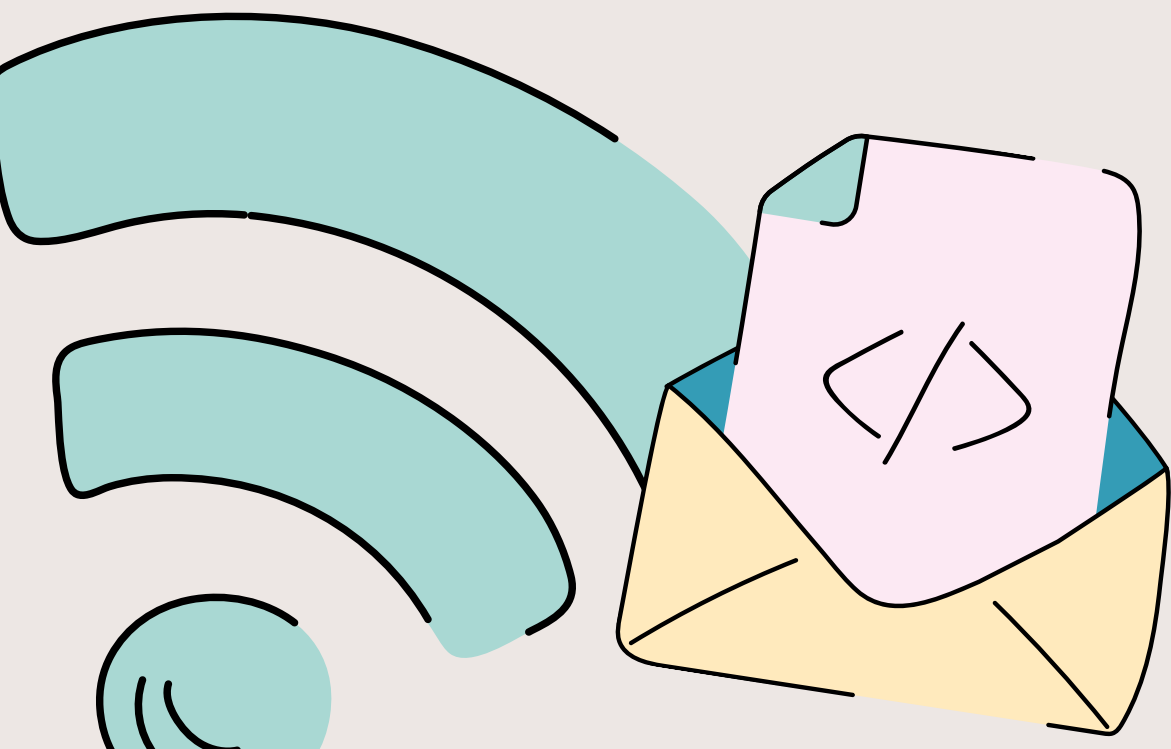
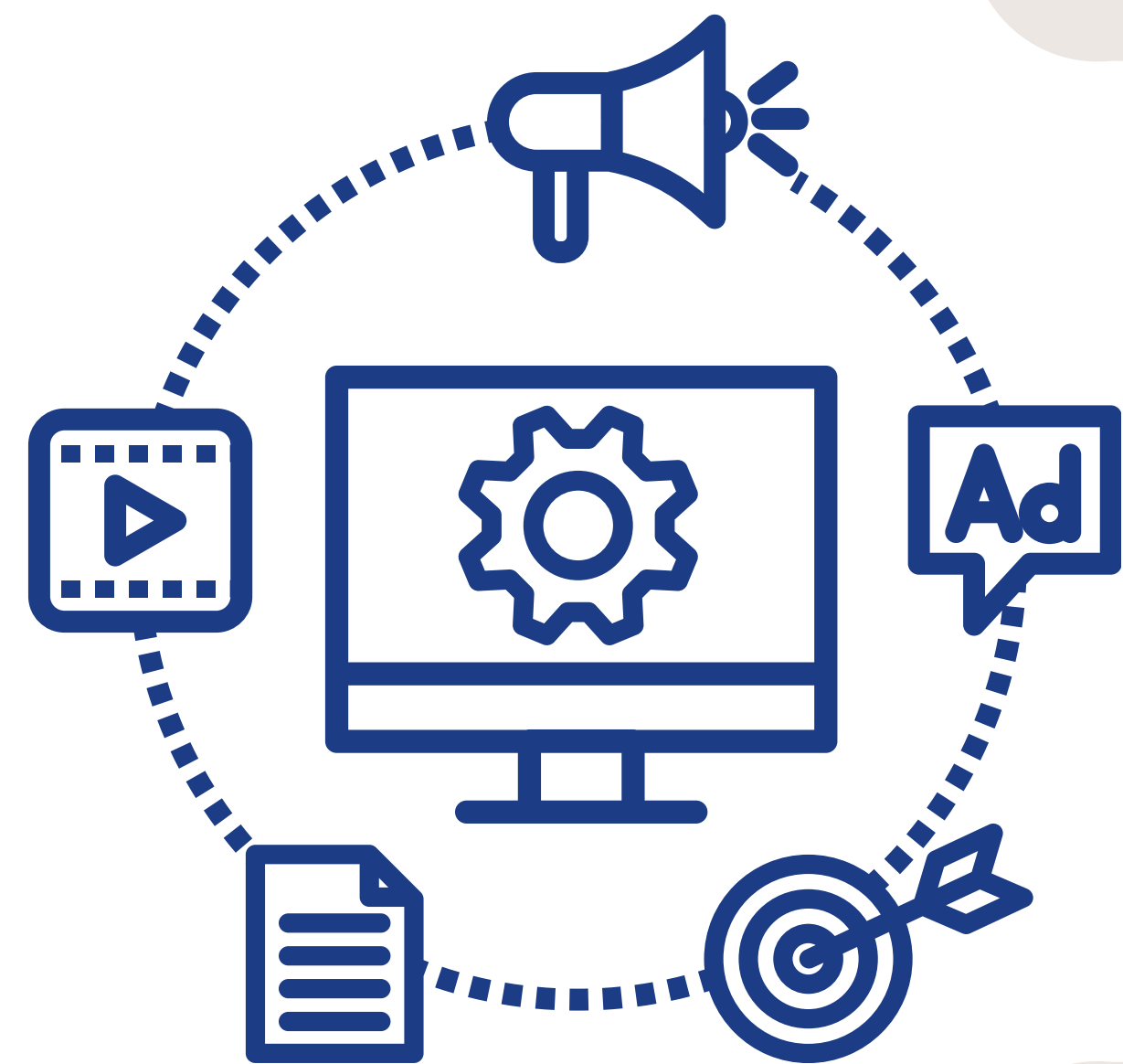


Obtención de datos en Data

Fuentes, métodos y herramientas

Por Juan Duran





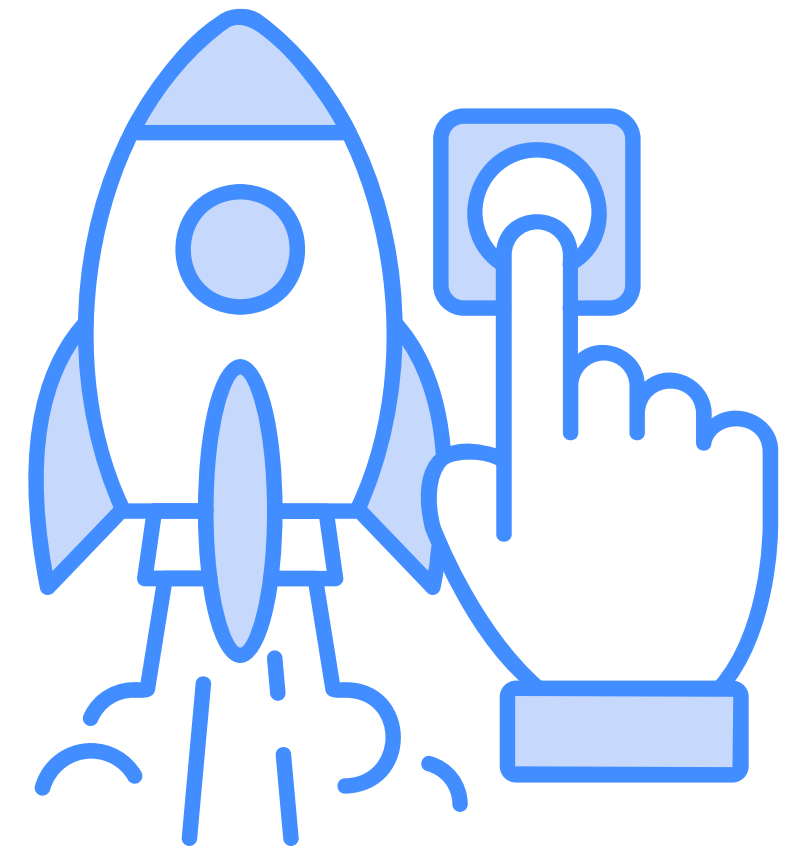
Contenidos



- Introducción 📖
- Puntos clave 📌
- Importancia de la obtención de datos ⭐
- Fuentes de datos 🌐
- Métodos de obtención de datos 🛠️
- Herramientas para la obtención de datos 🧰
- Buenas prácticas en la obtención de datos ✅
- Conclusiones 🏁

Introducción

La **obtención** de datos es un proceso fundamental en el mundo de la data. En esta presentación, exploraremos la importancia de los datos como **activo** estratégico, las diversas **fuentes** de donde se pueden obtener, los **métodos** utilizados para recolectarlos y las herramientas que facilitan este proceso. Además, discutiremos las buenas prácticas que deben seguirse para asegurar la **calidad** y la **ética** en la obtención de datos. Al final, tendrás una visión clara y completa de cómo manejar los datos de manera efectiva y responsable.



Puntos clave



Fuentes de datos

Las fuentes de datos pueden ser **internas**, como bases de datos empresariales y sistemas de gestión (ERP, CRM), y **externas**, como APIs públicas, redes sociales y datos abiertos (Open Data). Estas fuentes proporcionan una amplia **variedad** de datos que pueden ser utilizados para análisis y toma de decisiones.



Métodos de recolección

La obtención de datos puede realizarse mediante métodos **manuales**, como encuestas y entrevistas, y **automáticos**, como web scraping y la integración de APIs. Estos métodos permiten **recolectar** datos de manera eficiente y adaptada a las necesidades específicas de cada proyecto.



Herramientas especializadas

Existen diversas **herramientas** que facilitan la obtención de datos, como BeautifulSoup y Scrapy para **web scraping**, y Postman para la integración de **APIs**. Estas herramientas permiten extraer, gestionar y analizar datos de manera estructurada y eficiente.

Importancia de la obtención de datos

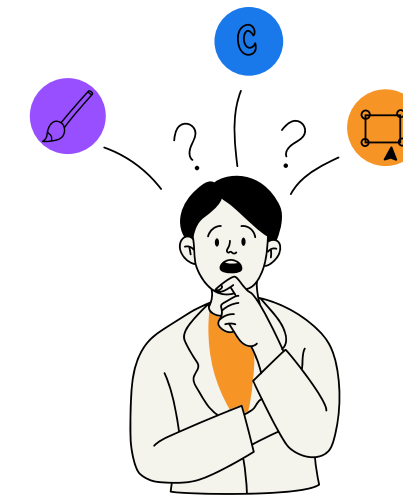
Los datos se han convertido en un **activo estratégico** esencial para las organizaciones modernas. La capacidad de recolectar, analizar y utilizar datos permite a las empresas tomar **decisiones informadas**, mejorar procesos y productos, y descubrir nuevas oportunidades de negocio. Sin datos precisos y relevantes, las organizaciones estarían navegando a ciegas en un entorno competitivo y en constante cambio. La correcta gestión de datos puede marcar la diferencia entre el **éxito** y el fracaso empresarial.

Los datos como activo estratégico



Toma de decisiones informadas

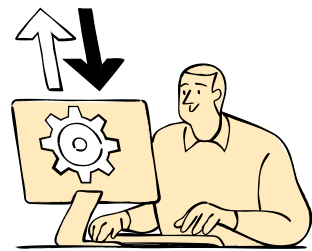
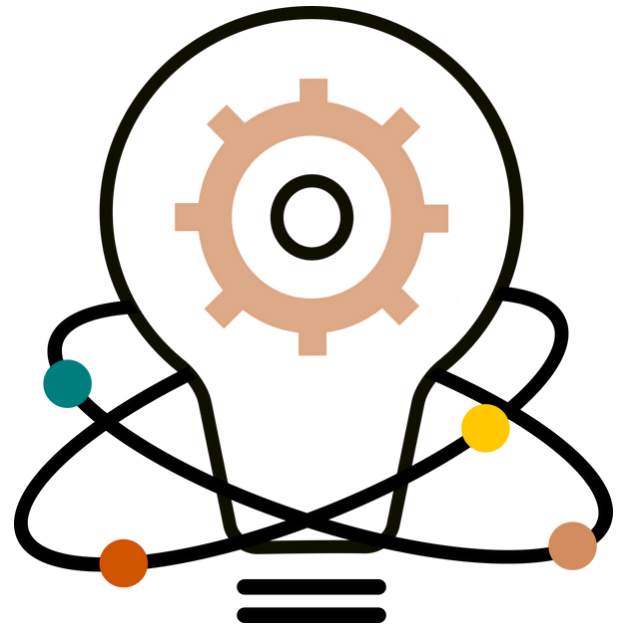
Mejora de procesos y productos



Descubrimiento de oportunidades de negocio



Fuentes de datos

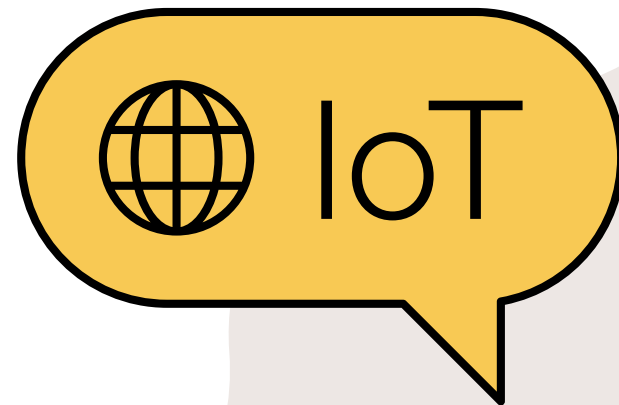


Existen múltiples fuentes de datos que pueden ser aprovechadas para obtener **información** valiosa. Las fuentes internas incluyen **bases de datos** empresariales y sistemas de gestión como **ERP** y **CRM**, que contienen datos operativos y transaccionales. Por otro lado, las fuentes externas abarcan **APIs** públicas, redes sociales y datos abiertos (**Open Data**) proporcionados por gobiernos y organizaciones. Estas fuentes externas pueden complementar los datos internos y proporcionar una visión más completa del entorno. La combinación de estas fuentes permite una perspectiva más rica y detallada.

Aspectos clave

- **Fuentes internas:** bases de datos empresariales, sistemas de gestión (ERP, CRM).
- **Fuentes externas:** APIs públicas, redes sociales, datos abiertos (Open Data).
- Complementariedad de fuentes internas y externas.

Métodos de obtención de datos

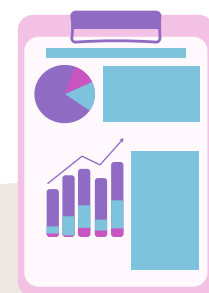
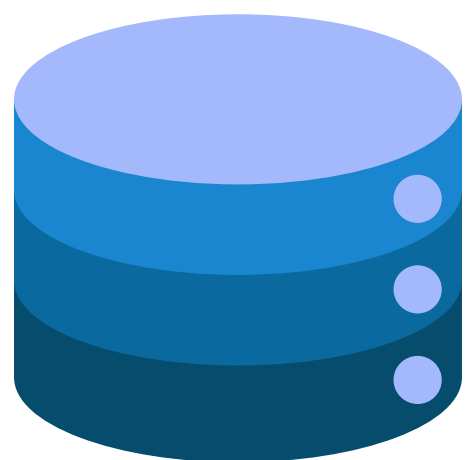


La **obtención** de datos puede realizarse mediante métodos **manuales** y **automáticos**. La extracción manual implica la recolección de datos a través de encuestas, formularios y entrevistas, lo cual puede ser útil para obtener información cualitativa y específica. La extracción automática, por otro lado, utiliza técnicas como el **web scraping**, la integración de **APIs** y el uso de sensores y dispositivos IoT para recolectar grandes volúmenes de datos de manera eficiente y en tiempo real. Estos métodos permiten una **recolección** de datos más rápida y precisa, adaptada a las necesidades específicas de cada proyecto.

Aspectos clave

- **Extracción manual:** encuestas, formularios, entrevistas.
- **Extracción automática:** web scraping, integración de APIs, sensores y dispositivos IoT.
- **Eficiencia** y **recolección** en tiempo real.

Herramientas para la obtención de datos



Existen diversas **herramientas** que facilitan la obtención de datos. Para el web scraping, herramientas como **BeautifulSoup** y **Scrapy** en Python son ampliamente utilizadas. Estas herramientas permiten extraer datos de sitios web de manera estructurada. Para la integración de APIs, **Postman** e **Insomnia** son herramientas populares que facilitan la interacción con APIs y la recolección de datos. Además, herramientas de gestión de datos como **SQL** y **Pandas** en Python son esenciales para almacenar, manipular y analizar los datos recolectados. El uso adecuado de estas herramientas puede optimizar significativamente el proceso de obtención de datos.

Aspectos clave

- Herramientas de **web scraping**: BeautifulSoup, Scrapy.
- Herramientas de **integración** de APIs: Postman, Insomnia.
- Herramientas de **gestión de datos**: SQL, Pandas.

Buenas prácticas en la obtención de datos



La **calidad** de los datos es crucial para asegurar resultados **precisos** y **confiables**. Es importante verificar y limpiar los datos recolectados para eliminar errores y inconsistencias. Además, la **ética** y la **privacidad** deben ser consideradas en todo momento. Cumplir con regulaciones como el **GDPR** y obtener el consentimiento informado de los individuos cuyos datos se recolectan son prácticas esenciales para proteger la privacidad y los derechos de las personas. Estas prácticas no solo garantizan la legalidad, sino también la **confianza** y la **integridad** en el manejo de datos.

Aspectos clave

- Calidad de los datos: verificación y limpieza.
- Ética y privacidad: cumplimiento de regulaciones (GDPR).
- Consentimiento informado.



Conclusiones



Importancia de los datos

Los datos son un **activo** estratégico esencial para la toma de **decisiones** informadas y la mejora de procesos y productos.

Herramientas especializadas

Existen herramientas específicas para la obtención de datos, como BeautifulSoup y Scrapy para **web scraping**, y Postman para integración de **APIs**.

Fuentes internas y externas

Las fuentes de datos pueden ser **internas**, como bases de datos empresariales, y **externas**, como APIs públicas y redes sociales.

Calidad y ética

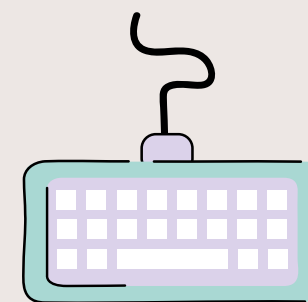
Es crucial **verificar** y **limpiar** los datos recolectados, así como cumplir con regulaciones de privacidad y obtener el consentimiento informado.

Métodos de recolección

La obtención de datos puede realizarse mediante métodos **manuales**, como encuestas, y **automáticos**, como web scraping y APIs.

Aplicación práctica

La correcta **obtención** y **gestión** de datos permite a las organizaciones descubrir nuevas **oportunidades** y mantenerse competitivas en el mercado.



Gracias



Por Juan Duran

“Coding, Gaming and Leveling Up”