

guayerd

Fundamentos IA

Introducción IA y datos

Clase 2

En colaboración con
IBM **SkillsBuild**





¡Bienvenidos!

¿Nos presentamos?

- ¿Qué recuerdan de la clase anterior?
- ¿Qué esperan aprender?
- ¿Tienen alguna pregunta?

Contenidos

Por temas

01

- Introducción IA

03

- Introducción Python

02

- Fundamentos del dato
- Pensamiento computacional

04

- Introducción a Python

Objetivos de la clase




- Dato, información e insights
- Tipos de datos
- Ciclo de vida del dato
- Escalas de medida
- Estructuras
- Secuencias, condiciones y bucles
- Descomposición de problemas
- Relación con la automatización

Introducción a la IA y los datos

Fundamentos del dato

Plataforma Skill Build: Fundamentos del dato



Data Fundamentals
IBM SkillsBuild

Plan de formación

Fundamentos de los Datos (¡Consigue una...

⌚ Aproximadamente 7 horas

📖 En curso

0%



Dato, información e insight

¿Son lo mismo? ¿Porqué?

Fundamentos de IA: De Datos a Insights

La Escalera del Conocimiento en Inteligencia Artificial

Objetivo

Comprender cómo la IA transforma datos brutos en conocimiento accionable a través de la progresión:

Dato → **Información** → **Insight**



DATOS: La Materia Prima

Un dato es un hecho crudo, sin procesar ni interpretar

Ejemplos cotidianos:

- 23°C (temperatura en Buenos Aires)
- María García (nombre de una cliente)
- \$1,250 MXN (precio de un producto)
- 15:30 (hora de una transacción)
- 👍 (like en una red social)

Características de los Datos:

- **Objetivos:** No tienen interpretación
- **Fragmentados:** Cada uno existe independientemente
- **Sin contexto:** No explican el "por qué" o "qué significa"

Dato = característica

*Piensa en 3 datos que tu organización recolecta diariamente.
¿Qué representan por sí solos?*

INFORMACIÓN: Datos con Contexto

Información = Datos + Contexto + Estructura

Transformación: Dato → Información

Dato Crudo	+ Contexto	→ Información
23°C	Temperatura promedio en marzo	"Buenos Aires tuvo 23°C promedio en marzo"
\$1,250 MXN	Precio de smartphone, categoría media	"Los smartphones de gama media cuestan \$1,250 MXN"
15:30	Hora pico de compras online	"Las compras online se concentran a las 15:30"

INFORMACIÓN: Datos con Contexto

Información = Datos + Contexto + Estructura

Ejemplos de Información Procesada:

- **Reporte:** "Las ventas de e-commerce en Colombia crecieron 40% en 2024"
- **Dashboard:** "Usuarios activos por país: México (45%), Brasil (30%), Argentina (15%)"
- **Estadística:** "El 68% de jóvenes entre 18-25 años usa apps de delivery"

¿Cómo la IA procesa datos en información?

- **Agregación:** Suma, promedia, cuenta
- **Clasificación:** Categoriza y etiqueta
- **Filtrado:** Selecciona datos relevantes
- **Correlación:** Encuentra relaciones entre variables

INSIGHT: El Conocimiento Accionable

Un insight es una comprensión profunda que revela patrones ocultos y permite tomar decisiones estratégicas

Transformación: Información → Insight

Características de un Insight Valioso:

- **Revelador:** Descubre algo no evidente
- **Accionable:** Permite tomar decisiones concretas
- **Relevante:** Impacta directamente en objetivos
- **Oportuno:** Llega en el momento adecuado

INSIGHT: El Conocimiento Accionable

Un insight es una comprensión profunda que revela patrones ocultos y permite tomar decisiones estratégicas

Ejemplo 1: E-commerce Regional

- **Información:** "Las ventas caen 20% los viernes en México"
- **Insight:** "Los consumidores mexicanos evitan compras grandes antes del fin de semana por cultura de gasto semanal. Estrategia: promociones de productos pequeños los viernes"

Ejemplo 2: App de Delivery

- **Información:** ""Los pedidos aumentan 300% cuando llueve""
- **Insight:** "Los usuarios valoran más la conveniencia que el precio durante días lluviosos. Estrategia: aumentar tarifas de delivery en 15% los días de lluvia sin afectar demanda"

Cómo la IA Potencia cada Nivel

En DATOS:

- **Recolección automatizada:** Sensores IoT, APIs, scraping
- **Limpieza:** Elimina duplicados, corrige errores
- **Estandarización:** Formatos uniformes

En INFORMACIÓN:

- **Análisis descriptivo:** ¿Qué pasó?
- **Visualización inteligente:** Dashboards adaptativos
- **Reportes automáticos:** Actualizaciones en tiempo real

En INSIGHTS:

- **Análisis predictivo:** ¿Qué va a pasar?
- **Análisis prescriptivo:** ¿Qué deberíamos hacer?
- **Detección de anomalías:** Identifica oportunidades ocultas
- **Personalización:** Insights específicos por segmento

CASO PRÁCTICO: Mercado Libre

DATOS (millones diarios):

- Búsquedas: "iPhone", "laptop", "zapatillas"
- Clics, tiempo en página, abandonos
- Ubicaciones, horarios, dispositivos
- Calificaciones, comentarios, quejas

INFORMACIÓN (reportes semanales):

- "Las búsquedas de electrónicos suben 60% en noviembre"
- "Los usuarios de Brasil prefieren pago en cuotas"
- "Las quejas de delivery se concentran en zonas rurales"

INSIGHTS (decisiones estratégicas):

- **Insight comercial:** "Los usuarios buscan productos premium los domingos pero compran versiones económicas los lunes. Estrategia: mostrar opciones premium los domingos con recordatorios para el lunes"
- **Insight logístico:** "Las quejas rurales correlacionan con distancia > 50km del centro urbano. Estrategia: alianza con comercios locales para pick-up points"

La Realidad de los Datos Globales

La Realidad de los Datos Globales:

El volumen global de datos crecerá hasta 394 zettabytes para 2028, con 181 zettabytes proyectados para 2025:

(Se necesitan 1000 millones de discos de 1TB para almacenar un solo zettabyte)



Composición de los datos:

- 80% de todos los datos serán no estructurados para 2025 (videos, audios, texto libre, imágenes)
- Solo 20% son datos estructurados (tablas, bases de datos tradicionales)



Velocidad de creación:

- 90% de todos los datos del mundo fueron creados en los últimos dos años
- Crecimiento del 192.68% en creación y consumo de datos entre 2019-2023



La paradoja del desperdicio:

- 68% de los datos empresariales nunca se utilizan
- IBM estima que 90% de los datos de sensores nunca se usan
- Hasta 80% de todos los datos actuales se consideran "datos oscuros" sin valor asignado

Reflexión Final

"Los datos son el nuevo petróleo, pero como el petróleo, necesitan refinarse para ser útiles. La IA es nuestra refinería moderna."

Nadamos en un océano de datos, pero morimos de sed de insights.

La oportunidad está en transformar ese 68% de datos desperdiciados en conocimiento accionable.

¿Cuál creen que es el mayor desafío en LATAM para convertir datos en insights accionables?

Datos

Materia prima

- **Describen las características de una entidad**
- Representación simbólica o general de un atributo o variable
- Pueden ser cuantitativos o cualitativos
- Existen de tipo valor numérico, texto y fórmulas



Dato = característica

Datos

Materia prima

Cuantitativo

- **Representan cantidades**
- Referidos a números
- Se pueden transformar en estadísticas utilizables para medir y probar

Ejemplo: Edad, altura, precio, etc.

Cualitativo

- **Representan categorías**
- Referidos a cualidades
- Se pueden identificar con preguntas como ¿cuál? o ¿cuáles?

Ejemplo: Nombre, país, sexo, etc.

Datos

Materia prima

Según estructura

- **Estructurados:** tabulares, organizados en filas y columnas. Ej: Excel, SQL.
- **No estructurados:** texto libre, imágenes, audio, video. Ej: comentarios, fotos, grabaciones.
- **Semiestructurados:** tienen estructura flexible. Ej: JSON, XML, formularios web.

Según origen

- **Primarios:** recolectados directamente por el analista.
- **Secundarios:** recopilados por otros (fuentes abiertas, sistemas previos).

Datos

Ejemplos de datos “semiestructurados”

Tipo de dato	Ejemplo concreto
JSON	Respuesta de una API con datos de usuarios: {"nombre": "Luis", "edad": 42}
XML	Archivos de configuración o facturas electrónicas con etiquetas jerárquicas
Logs de servidores	Registros con timestamp, IP, evento, pero sin formato tabular
Correos electrónicos	Tienen campos estructurados (remitente, asunto) y cuerpo libre
Datos de sensores IoT	Lecturas con timestamp y valores, pero con estructura flexible
Redes sociales	Posts con hashtags, menciones, emojis, y metadatos como ubicación o timestamp

¿Cómo determinar la calidad de los datos?

Mediante las siguientes **dimensiones**:

- **Exactitud:** Datos correctos
- **Compleitud:** Sin valores faltantes
- **Consistencia:** Mismo formato y codificación
- **Actualidad:** Datos vigentes
- **Relevancia:** Adecuados al objetivo

Información



Conjunto de datos procesados y organizados que proporcionan significado y contexto. Son fundamental para la toma de decisiones, el análisis y la comunicación.

Insight

Descubrimientos o comprensiones profundas y reveladoras obtenidas a partir del análisis de datos e información. Estos revelan patrones, tendencias o verdades ocultas que no son inmediatamente obvios.



Ciclo de vida del dato



Ciclo de vida del dato

Etapa	Objetivo	Ejemplos
Captura	Recoger datos	Formularios, sistemas, redes
Almacenamiento	Guardar datos	Bases de datos, archivos
Preparación	Limpiar, transformar y organizar	Python, Azure Data
Análisis	Explorar patrones y relaciones	
Comunicación	Presentar hallazgos	Power BI, Looker Studio
Decisión	Accionar en base a información	
Retroalimentación	Medir y ajustar resultados	

Ciclo de vida del dato - Aplicación

Etapa	¿Qué ocurre?	Ejemplo práctico (HR / Educación / Legal)
1. Captura	Se recolectan datos desde formularios, sensores, encuestas o sistemas.	Candidato llena solicitud de empleo; padre completa formulario legal.
2. Almacenamiento	Se guarda en BD, hojas de cálculo, o plataformas en la nube.	Datos en Excel, MySQL, OneNote, Power BI, o MongoDB.
3. Preparación	Limpieza, normalización, codificación, validación	Se corrigen errores, se estandarizan nombres y fechas.
4. Análisis	Se aplican estadísticas, modelos, visualizaciones o segmentaciones.	Se detectan patrones de ausentismo o brechas salariales.
5. Comunicación	Se presentan los hallazgos con dashboards, informes o visuales.	Se genera un flyer con insights, una presentación, para padres o gerentes.
6. Decisión	Se toman acciones basadas en los datos: políticas, intervenciones, ajustes.	Se ajusta el proceso de selección o se negocia una pensión.
7. Retroalimentación	Se evalúan los resultados y se mejora el proceso de captura y análisis.	Se rediseña el formulario o se mejora el modelo predictivo.

Escala de medición

Escala	Características ¿Qué mide?	¿Qué operaciones permite?	Ejemplos
Nominal	Categorías sin orden	Solo contar o clasificar	Género, tipo de producto, nacionalidad, tipo de contrato
Ordinal	Categorías con orden	Comparar rangos, no diferencias exactas	Nivel de satisfacción, nivel educativo
Intervalo	Numérica sin cero real (absoluto)	Suma y resta	Temperatura, fechas del calendario
Razón	Numérica con cero absoluto	Todas las operaciones matemáticas	Ingresos, cantidad de ventas, edad, peso, duración



Saber la **escala** de cada dato es clave para aplicar el **análisis correcto** y evitar errores de interpretación

Estructuras básicas de datos

- **Campos (columnas):** Representan características individuales de una entidad
- **Registros (filas):** Instancias únicas u observaciones de la entidad
- **Tabla:** Estructura que agrupa múltiples registros bajo un mismo conjunto de campos

Identificadores

- **Clave primaria (Primary Key - PK):** Campo/s que garantizan la unicidad de cada registro
- **Clave foránea (Foreign Key - FK):** Campo que establece una relación lógica con la clave primaria

¿Qué tabla es correcta?

¿Porqué?



A

Producto	Enero	Febrero	Marzo
Avena	50	45	60
Granola	30	40	50

B

Producto	Mes	Ventas
Avena	Enero	50
Avena	Febrero	45
Avena	Marzo	60
Granola	Enero	30
Granola	Febrero	40
Granola	Marzo	50

Formato de Tabla Larga o “Tidy”

“Tidy” = ordenado, limpio y estructurado de forma lógica (Hadley Wickhman)

¿Qué es un conjunto de datos “tidy”?

Un dataset está en formato **tidy** cuando cumple estas tres reglas:

1. Cada **variable** está en su **propia columna**.
2. Cada **observación** está en su **propia fila**.
3. Cada **tipo de unidad observacional** está en su **propia tabla**.



Producto	Enero	Febrero	Marzo
Avena	50	45	60
Granola	30	40	50



¿Por qué es útil?

- Compatible con herramientas como Power BI, R (tidyverse), Python (pandas).
- Facilita filtros, agrupamientos, visualizaciones y modelos estadísticos.
- Evita errores comunes en análisis y reportes.



Producto	Mes	Ventas
Avena	Enero	50
Avena	Febrero	45
Avena	Marzo	60
Granola	Enero	30
Granola	Febrero	40
Granola	Marzo	50

Introducción a la IA y los datos

Pensamiento computacional

Pensamiento Computacional

Una forma de estructurar el razonamiento lógico que permite resolver problemas complejos con claridad y eficiencia

Conjunto de **habilidades para plantear problemas y definir soluciones** de manera que puedan ser comprendidas y ejecutadas tanto por personas como por sistemas informáticos.

- **"El pensamiento computacional no es solo para programadores—es una forma de estructurar el razonamiento lógico que permite resolver problemas complejos con claridad y eficiencia."**
- **"Antes de entrenar una IA, necesitamos entender cómo descomponer un problema en partes manejables. Eso es pensamiento computacional en acción."**
- **"La IA no 'piensa' como los humanos, pero el pensamiento computacional nos ayuda a diseñar algoritmos que simulan decisiones inteligentes."**



Secuencias, condiciones y bucles

Secuencias: Pasos que se ejecutan en orden para alcanzar un objetivo.

- Cada paso depende del anterior
- El orden afecta el resultado

Condiciones: Reglas que activan acciones según un criterio.

- Relación tipo “si... entonces...”
- Permiten decisiones en el flujo
- Deben cubrir todos los casos posibles

Bucles: Repiten instrucciones mientras se cumpla una condición.

- Requieren de inicio, condición de parada y actualización
- Usados para repetir procesos automáticamente
- Bucles infinitos si no están bien definidos

Descomposición de problemas

¿Qué es?

Descomponer es **separar un problema complejo en partes** más simples, abordables y analizables.



Descomponer = separar

Descomposición de problemas

¿Por qué es importante?

- Permite entender mejor el problema
- Ayuda a planificar el análisis
- Facilita la colaboración en equipos
- Permite aplicar herramientas digitales por etapas

Descomposición de problemas

Pasos

1. Definir el objetivo con un verbo
2. Listar subtareas y dependencias
3. Definir entradas y salidas de cada subtask
4. Estimar reglas y excepciones
5. Priorizar lo crítico

Lupa logística

¿Cómo resolverías este problema?



A partir de la tabla, determina cuántos pedidos fueron enviados.

Tareas

- Indicar la cantidad de pedidos enviados
- Detallar la secuencia, condición y bucle existente

ID	Pago	Stock
O-502	Aprobado	Sí
O-506	Pendiente	No
O-501	Aprobado	No
O-505	Anulado	Sí
O-503	Pendiente	Sí
O-504	Aprobado	No
O-508	Aprobado	No
O-507	Aprobado	Sí

Pseudocódigo

Definición estructurada de algoritmos en un formato independiente de cualquier lenguaje de programación.

Debe ser fácil de leer para las personas y sencillo de convertir en código ejecutable.

```
1  INICIO
2  |   LEER precio
3  |   SI precio > 1000 ENTONCES
4  |   |   MOSTRAR "producto caro"
5  |   SINO
6  |   |   MOSTRAR "Producto accesible"
7  FIN
```

Sintaxis

- Boques con sangría
- Palabras clave (INICIO/FIN, SI/ENTONCES/SINO)
- Entrada, Proceso, Salida
- Variables con nombres descriptivos
- Comentarios breves

Diagrama de flujo

- **Permiten visualizar procesos y detectar huecos**
- Facilitan la comunicación de reglas con equipos
- Transforman pseudocódigos en símbolos conectados

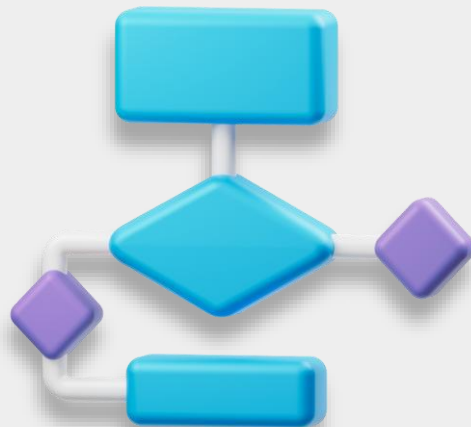


Diagrama de flujo

Símbolos

Inicio/Fin



Forma: Elipses

Proceso



Forma: Rectángulo

Decisión



Forma: Rombos

Entrada/Salida



Forma: Paralelogramo

Conector



Forma: Líneas

Diagrama de flujo

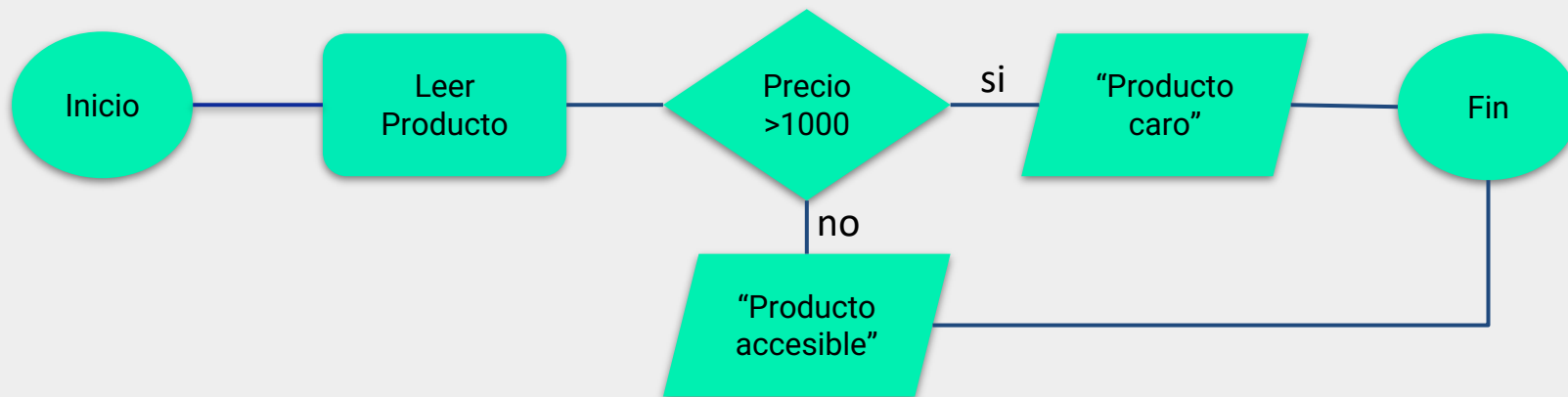
Reglas

1. Una decisión = dos salidas (Sí/No)
2. Evitar cruces de flechas, usar conectores
3. Titulos breves por bloque
4. Mantener granularidad consistente

Granularidad es el nivel de detalle con que se describe una información

Diagrama de flujo

Ejemplo



¿Qué es la automatización?

Es el proceso de hacer que **tareas repetitivas o lógicas se realicen de forma automática, sin intervención manual.**

Esto se logra a partir de **reglas, condiciones o flujos** predefinidos que las herramientas tecnológicas pueden ejecutar por sí mismas.



¿Qué es la automatización?

Ejemplos



Python

Escribir un script que limpie un conjunto de datos y lo deje listo para el análisis.

Power BI

Programar la actualización automática de un dashboard para mantener la información siempre al día.

¿Qué se puede automatizar en la IA?

- Limpieza y procesamiento de datos
- Aplicación de reglas y condiciones
- Ejecución de cálculos
- Predicciones y proyecciones
- Generación de reportes o resultados



IA y automatización se complementan para
mejorar procesos y decisiones

Despacho inteligente



Se debe calcular el recuento de pedidos por estado y método de envío, realiza para ello un pseudocódigo y diagrama de flujo y ten en consideración las siguientes reglas:

- Si Pago = Anulado >> Estado = Anulado
- Si Pago \neq Aprobado >> Estado = Pendiente
- Si Pago = Aprobado y Stock = No >> Estado = Enviado
- Si Pago = Aprobado y Stock = Sí >>

Método de envío:

- Moto si Destino = Capital y Peso ≤ 5
- Correo si Destino = Interior y Peso ≤ 10
- Expreso en cualquier otro caso

ID	Pago	Stock	Destino	Peso
O-702	Pendiente	Sí	Interior	7
O-708	Aprobado	Sí	Interior	10
O-705	Aprobado	No	Capital	2
O-701	Aprobado	Sí	Capital	3
O-703	Aprobado	Sí	Interior	8
O-707	Aprobado	Sí	Capital	6
O-704	Aprobado	Sí	Interior	12
O-706	Anulado	Sí	Capital	1



Retro

¿Cómo nos vamos?

- ¿Qué fue lo más útil de la clase?
- ¿Qué parte te costó más?
- ¿Qué te gustaría repasar o reforzar?