



# Sistemas organizacionales y gerenciales 2[N]

## José Luis Reynoso Tiú

Día, Fecha:	Sábado, 22/02/2025
Hora de inicio:	08:00AM



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

# Clase 4

# Información de contacto



## FOROS DE UEDI

Semanalmente se creará un foro en UEDI para resolver dudas.

**2655896221401@ingenieria.usac.edu.gt**

[SOG2]Asunto

# Datos estructurados

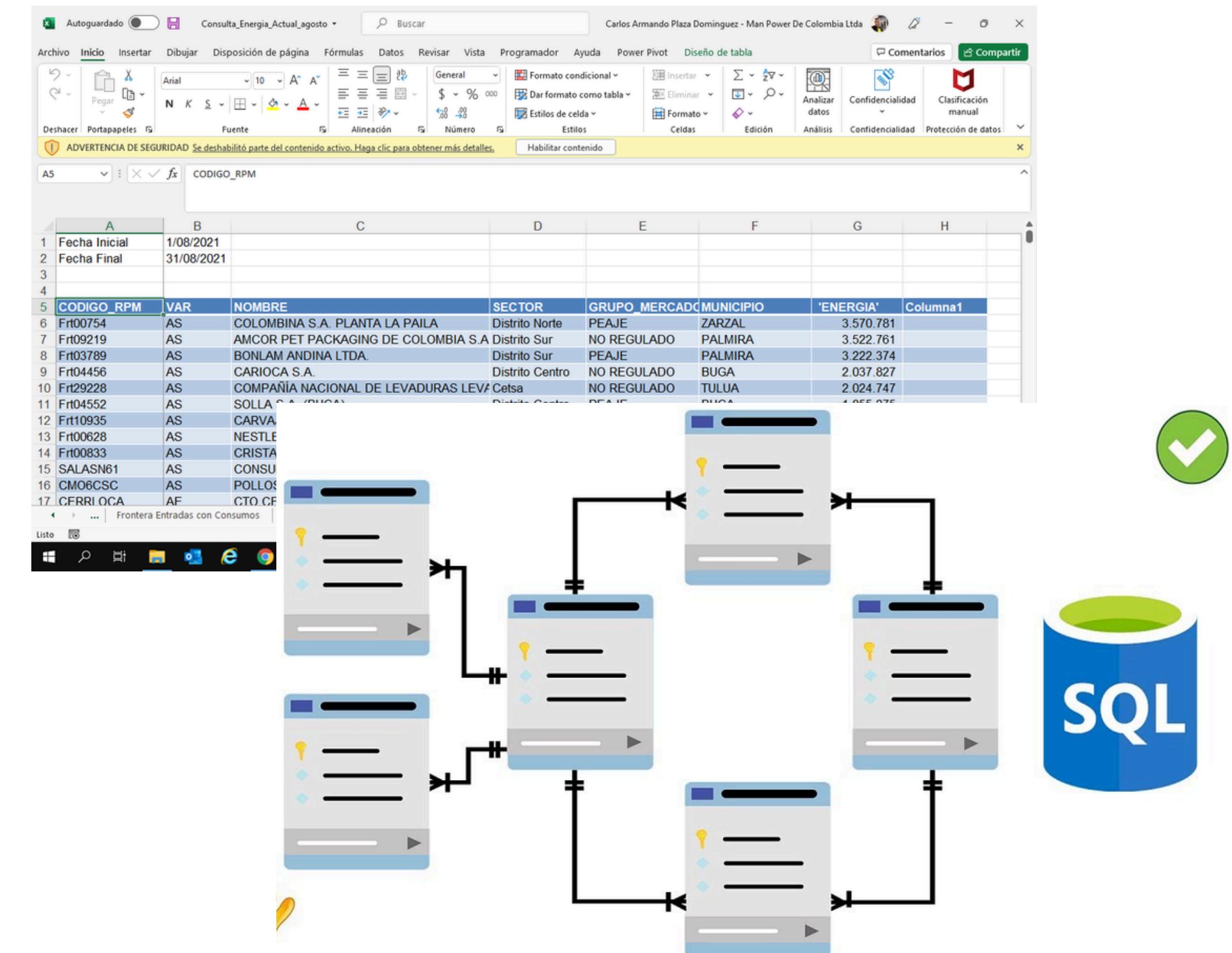
- Tienen un formato estandarizado
- Por lo general, se trata de datos tabulares con filas y columnas que definen claramente sus atributos. Las computadoras pueden procesar eficazmente los datos estructurados en busca de información dado que se trata de información cuantitativa.
- **Ejemplo:** una tabla estructurada de datos de clientes que contenga columnas (nombre, dirección y número de teléfono) puede proporcionar información, como la cantidad total de clientes, así como el lugar con el mayor número de clientes. Por el contrario, los datos no estructurados, como una lista de publicaciones en redes sociales, resultan más difíciles de analizar.

# Características de los datos estructurados

- Atributos definibles: Los datos estructurados tienen los mismos atributos para todos los valores de datos.
- Atributos relacionales: Tienen valores comunes que vinculan entre sí diferentes conjuntos de datos.
- Datos cuantitativos: Son apropiados para el análisis matemático.
- Almacenamiento: Se pueden guardar datos estructurados en bases de datos relacionales.

# Ejemplos de datos estructurados

- Archivos de Excel
- Bases de datos SQL
- Resultados de formularios web
- Control de inventario



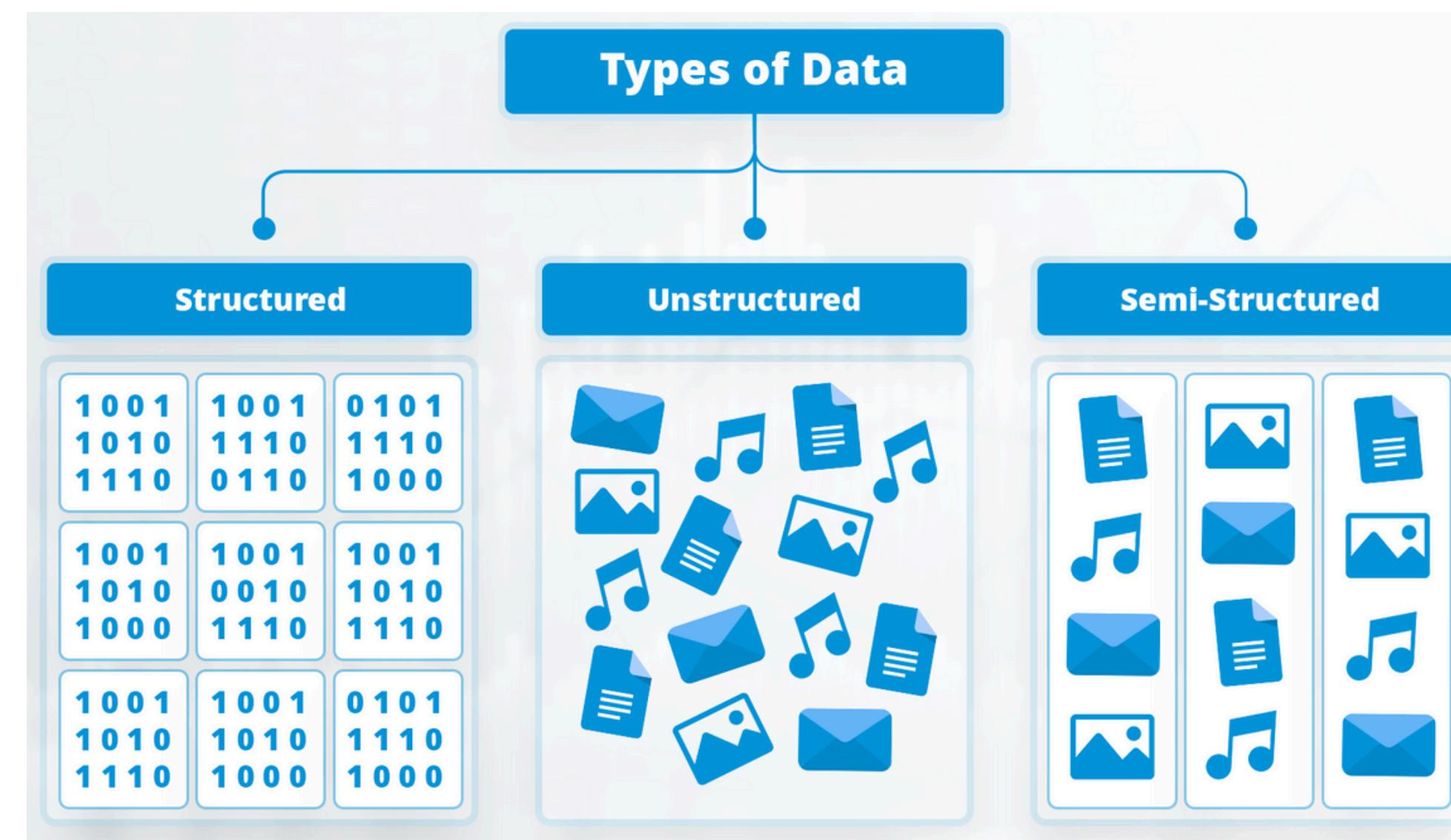
# Datos no estructurados

Los datos no estructurados son información sin un modelo de datos establecido o son datos que no están ordenados de manera predefinida, por lo regular son datos cualitativos.

Ejemplos de datos no estructurados:

- Archivos de texto
- Archivos de video
- ?
- Imágenes

- Las empresas están creando datos a un ritmo exponencial y entre un **80%** y **90%** son datos no estructurados.
- Para los datos no estructurados se necesitan diferentes tecnologías y estrategias para llevar a cabo un análisis eficiente de los mismos.



# Datos semiestructurados

- Son una categoría intermedia entre los datos estructurados y los no estructurados.
- No pueden considerarse datos totalmente estructurados porque carecen de un modelo de datos relacional o tabular específico. No obstante incluyen metadatos que se pueden analizar, como etiquetas y otros marcadores.
- Es más fácil obtener información de los datos semiestructurados que de los no estructurados.
- No se adhieren a un modelo de datos predefinido del modo en que lo hacen los datos estructurados.

# Ejemplos de datos semiestructurados

- XML
- JSON



{JSON}

JavaScript Object Notation

# Datos estructurados vs no estructurados

	Datos estructurados	Datos no estructurados
¿Qué es?	Datos que se ajustan a un esquema o modelo de datos predefinido.	Datos sin un modelo subyacente para discernir los atributos.
Ejemplo básico	Una tabla de Excel.	Una colección de archivos de video.
Ideales para	Una colección asociada de valores numéricos y de texto discretos, cortos y no continuos.	Una colección asociada de datos, objetos o archivos en la que los atributos cambian o se desconocen.
Tipos de almacenamiento	Bases de datos relacionales, bases de datos de gráficos, bases de datos espaciales, cubos OLAP y más.	Sistemas de archivos, sistemas DAM, CMS, sistemas de control de versiones y más.
Mayor beneficio	Son más fáciles de organizar, limpiar, buscar y analizar.	Pueden analizar datos que no se pueden convertir fácilmente en datos estructurados.
Mayor desafío	Todos los datos deben ajustarse al modelo de datos prescrito.	Pueden ser difíciles de analizar.
Técnica de análisis principal	Consultas SQL.	Varía.



You should definitely know this when working with Structured and Unstructured data...



Share

# DATA STRUCTURED VS. UNSTRUCTURED?



Watch on YouTube

# Big data

Se refiere a un conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías o herramientas convencionales.

La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas.

En la mayoría de los casos, con el fin de utilizar eficazmente el Big Data, debe combinarse con datos estructurados (normalmente de una base de datos relacional) de una aplicación comercial más convencional, como un ERP (Enterprise Resource Planning) o un CRM (Customer Relationship Management).

# Las 3 Vs del Big data

- **Volumen:** Se refiere a la cantidad de datos recopilados, va desde terabytes hasta petabytes.
- **Variedad:** Los datos provienen de diferentes fuentes y en diferentes formatos. Los datos pueden ser estructurados, semiestructurados o no estructurados.
- **Velocidad:** Cada vez más las empresas tienen requisitos exigentes desde el momento en que se generan los datos al momento que se entrega información procesable a los usuarios. Por lo tanto, es necesario recopilar, almacenar, procesar y analizar los datos en períodos relativamente cortos de tiempo.



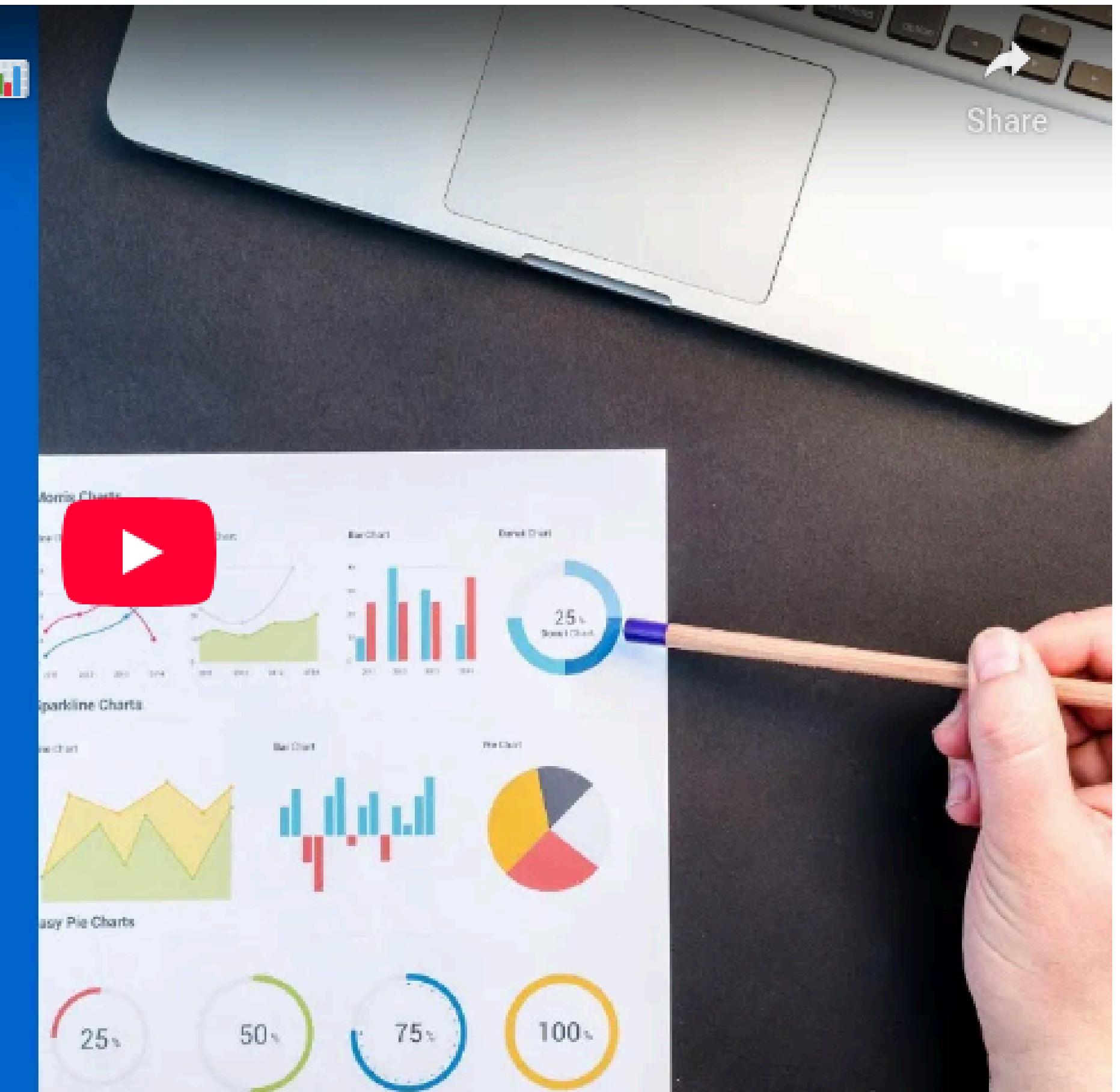
Ventajas de la Tecnología Big Data



# Big Data en la era Digital

## VENTAJAS

Watch on YouTube



# Ciencia de datos

La ciencia de datos es el estudio de la extracción de conocimientos útiles a partir de datos utilizando métodos científicos, técnicas estadísticas y algoritmos computacionales.

La ciencia de datos implica el proceso de recopilar, analizar e interpretar grandes cantidades de datos para encontrar patrones, conocimientos y tendencias. Luego, los conocimientos se utilizan para tomar decisiones basadas en datos.

Combina principios y prácticas de:

- Matemáticas
- Estadística
- Programación
- Inteligencia artificial

Este análisis permite extraer información significativa para una empresa u organización. Además permite que los científicos de datos planteen y respondan a preguntas como:

- ¿Qué paso?
- ¿Por que paso?
- ¿Qué pasará?
- ¿Qué se puede hacer con los resultados



**Insigth:** descubrimiento o hallazgo dentro de los datos que es útil para tomar decisiones.

# ¿Por qué es importante la ciencia de datos?

Es importante porque combina herramientas, métodos y tecnología para generar significado a partir de los datos. Las organizaciones modernas están inundadas de datos; hay una proliferación de dispositivos que pueden recopilar y almacenar información de manera automática. Los sistemas en línea y los portales de pago capturan más datos en los campos del comercio electrónico, la medicina, las finanzas y cualquier otro aspecto de la vida humana. Disponemos de grandes cantidades de datos de texto, audio, video e imágenes.

# ¿Para qué se utiliza la ciencia de datos?

Utiliza los datos de cuatro maneras principales:

- Análisis descriptivo
- Análisis de diagnóstico
- Análisis predictivo
- Análisis prescriptivo

# Análisis descriptivo

El análisis descriptivo examina los datos para obtener información sobre lo que ha ocurrido u ocurre en el entorno de datos. Se caracteriza por las visualizaciones de datos, como los gráficos circulares, de barras o líneas, las tablas o las narraciones generadas.

## Ejemplo

Un servicio de reserva de vuelos registra datos como el número de billetes reservados cada día. El análisis descriptivo revelará los picos y las caídas de las reservas, así como los meses de alto rendimiento del servicio.

# Análisis de diagnóstico

El análisis de diagnóstico es un examen profundo o detallado de datos para entender por qué ha ocurrido algo. Se caracteriza por técnicas como el análisis detallado, el descubrimiento y la minería de datos o las correlaciones. Se pueden llevar a cabo varias operaciones y transformaciones de datos en un determinado conjunto con el fin de descubrir patrones únicos en cada una de estas técnicas.

## Ejemplo

El servicio de vuelos podría hacer el análisis detallado de un mes con un rendimiento particularmente alto para entender mejor el pico de reservas. Esto puede revelar que muchos clientes visitan una determinada ciudad para asistir a un evento deportivo mensual.

# Análisis predictivo

El análisis predictivo utiliza los datos históricos para hacer previsiones precisas sobre los patrones de datos que pueden producirse en el futuro. Se caracteriza por técnicas como el machine learning, la previsión, la coincidencia de patrones y el modelado predictivo. En cada una de estas técnicas, se entrena a las computadoras para aplicar ingeniería inversa a las conexiones de causalidad en los datos.

## Ejemplo:

El equipo de servicios de vuelo podría utilizar la ciencia de datos para predecir los patrones de reserva de vuelos del año siguiente al inicio de cada año. El programa o algoritmo de la computadora pueden examinar datos anteriores y predecir picos de reservas de determinados destinos en mayo. Al anticiparse a las futuras necesidades de viaje de los clientes, la empresa podría empezar desde febrero a hacer publicidad específica para esas ciudades.

# Análisis prescriptivo

El análisis prescriptivo lleva los datos predictivos al siguiente nivel. No solo predice lo que es probable que ocurra, sino que sugiere una respuesta óptima para ese resultado. Puede analizar las posibles implicaciones de las diferentes alternativas y recomendar el mejor curso de acción. Utiliza el análisis de gráficos, la simulación, el procesamiento de eventos complejos, las redes neuronales y los motores de recomendación del machine learning.

## Ejemplo:

De la reserva de vuelos, el análisis prescriptivo podría examinar las campañas de marketing históricas para maximizar la ventaja del próximo pico de reservas. Un científico de datos podría proyectar los resultados de las reservas de diferentes niveles de gasto en varios canales de marketing. Estas previsiones de datos dan a la empresa de reserva de vuelos una mayor confianza en sus decisiones de marketing.

# Beneficios de la ciencia de datos para las empresas

- **Descubrir patrones desconocidos de transformación:** La ciencia de datos permite a las empresas descubrir nuevos patrones y relaciones con el potencial de transformar la organización. Puede revelar cambios de bajo coste en la administración de recursos para obtener el máximo impacto en los márgenes de beneficio.
- **Innovar con nuevos productos y soluciones:** La ciencia de datos puede revelar lagunas y problemas que de otro modo pasarían desapercibidos. Mejor información sobre las decisiones de compra, los comentarios de los clientes y los procesos empresariales puede impulsar la innovación en las operaciones internas y las soluciones externas.

# Beneficios de la ciencia de datos para las empresas

- **Optimización en tiempo real:** Para las empresas, en especial las grandes, es un gran reto responder en tiempo real a las condiciones cambiantes. Esto puede causar importantes pérdidas o interrupciones en la actividad empresarial. La ciencia de datos puede ayudar a las empresas a predecir los cambios y reaccionar de forma óptima ante las distintas circunstancias.

# Diferencia entre ciencia de datos y análisis de datos

- Aunque los términos se pueden usar de manera indistinta, el análisis de datos es un subconjunto de la ciencia de datos.
- La ciencia de datos es un término que engloba todos los aspectos del procesamiento de datos, desde la recopilación hasta el modelado y la información.
- El análisis de datos se ocupa principalmente la estadística, las matemáticas y el análisis estadístico. Solo se centra en el análisis de datos, mientras que la ciencia de datos está relacionada con el panorama general de los datos de la organización.

# Coeficiente de correlación

- Coeficiente de correlación positivo: indica que si una variable aumenta la otra también lo hace
- Coeficiente de correlación negativo: si una variable aumenta la otra disminuye y viceversa
- Coeficiente de correlación 0: indica que no hay correlación lineal entre las dos variables

# Ciclo de vida de la ciencia de datos

# Ingestión

El primer paso en el proceso de análisis de datos es la ingestión de datos. Este término se refiere a la recolección de datos de diversas fuentes disponibles. Estas fuentes pueden incluir web scraping, Internet de las cosas, transmisión de datos en tiempo real de aplicaciones propias o de terceros, dispositivos, archivos CSV, recolección manual, redes sociales, entre otros. Los datos recopilados pueden ser estructurados o no estructurados.

Algunas herramientas utilizadas en este etapa son:

- Apache Kafka
- Apacha Flume
- Amazon Kinesis
- Azure Data Factory
- Cloud Data Fusion

# Almacenamiento y procesamiento

Los datos pueden provenir de diversas fuentes y presentarse en formas estructuradas o no estructuradas. Esta variedad de datos requiere de sistemas de almacenamiento especializados que puedan manejar eficientemente sus características únicas.

En el proceso de preparación de los datos, se llevan a cabo varias operaciones críticas. Estas incluyen la limpieza de los datos, la eliminación de duplicados, la transformación de los datos a formatos más útiles, y la combinación de diferentes conjuntos de datos. Para realizar estas tareas, se utilizan sistemas de ETL (Extracción, Transformación y Carga) u otras tecnologías de integración de datos.

# Almacenamiento y procesamiento

Herramientas de ETL:

- Amazon Glue
- Apache Spark
- Azure Data Factory

Herramientas de almacenamiento

- Amazon Redshift
- Azure Synapse Analytics
- Snowflake

# Análisis de datos

En esta etapa el científico de datos conduce diferentes tipos de análisis, como puede ser un análisis exploratorio de datos (EDA), para obtener patrones, sesgos, rangos y distribuciones que sean útiles. Este análisis permite obtener hipótesis para poder realizar pruebas de concepto en el futuro. También permite verificar la relevancia que estos datos tienen para su uso en modelos predictivos o análisis futuros.

Herramientas de análisis de datos:

- Excel
- Python
- R
- Apache Spark

# Visualización

Los insights, o conocimientos obtenidos a partir de los datos, deben ser comunicados de manera efectiva para que puedan tener un impacto real en el negocio. Esto se logra a través de reportes y visualizaciones de datos que no solo presentan la información de manera clara y comprensible, sino que también destacan cómo estos datos pueden afectar al negocio. Al hacerlo, se facilita la toma de decisiones basadas en datos y se permite una comprensión más sencilla de la información.

Herramientas de visualización de datos:

- Tableau
- Power Bi
- Google Looker
- Amazon QuickSight
- SQL Server Reporting Services (SSRS)



<b>Nombre de la actividad:</b>	Corto 2
<b>Cantidad de participantes:</b>	0
<b>Doy fe que esta actividad está planificada en dtt (Sí/No):</b>	Sí

<b>Hora de inicio:</b>	09:15
<b>Hora de fin:</b>	09:40
<b>Duración (min):</b>	25

**Participantes: llenar las siguientes cajas de texto (tomar información del chat del meet)**

