



# Bienvenidos a **Recopilación e Integración de Datos**

Cursado 2023  
Instituto Superior Tecnológico Empresarial Argentino  
Ing. Carla Pezzone



**ISTEA**

## AGENDA DE LA CLASE

- ✓ Nos presentamos
- ✓ Alcances de este curso
- ✓ Materiales del curso
- ✓ Tips para estudiar
- ✓ Introducción a la ciencias de datos
- ✓ Actividad



# NOS PRESENTAMOS



# ALCANCES DE ESTE CURSO

## Carrera

Ciencias de datos e Inteligencia Artificial

## Materia

Recopilación e Integración de Datos

- ✓ Análisis computacional de datos y pensamiento analítico
- ✓ El mundo de los datos. Configura tu caja de herramientas
- ✓ Preguntas efectivas. Decisiones basadas en datos
- ✓ Tipos y estructuras de datos
- ✓ Conceptos básicos sobre hojas de cálculo. Stakeholders
- ✓ Sesgo, credibilidad, privacidad, ética y acceso. Organizar y proteger tus datos
- ✓ La importancia de la integridad. Datos impecables
- ✓ Limpieza de datos con SQL
- ✓ Visualización de datos
- ✓ Verificar e informar los resultados de tu limpieza de datos
- ✓ Análisis exploratorio de datos
- ✓ Distribuciones de datos y muestreo

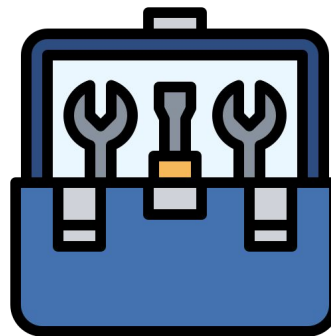
# 01 - Análisis computacional de datos y pensamiento analítico

- ? Datos y análisis de datos
- ? Ciclo de vida de los datos
- ? Utilidad en las organizaciones
- ? Decisiones basadas en datos
- ? Pensamiento analítico
- ? Herramientas para el análisis de datos
- ? Causa raíz y análisis GAP
- ? Extraer, Transformar y Cargar (ETL)



## 02 - El mundo de los datos. Configura tu caja de herramientas

- ? Hojas de cálculos
- ? SQL
- ? Visualización de datos
- ? Selección de herramientas



## 03 - Preguntas efectivas. Decisiones basadas en datos

- ? Preguntas en el Proceso Analítico de Datos y Pensamiento estructurado
- ? Reconocer el problema: Tipos de problemas
- ? Preguntas efectivas... y algunas no tanto
- ? Preguntas SMART



## 04 - Tipos y estructuras de datos

- ? Decisiones y datos
- ? Formatos de datos
- ? Tipo de datos
- ? Factores a considerar cuando se recolectan datos
- ? Transformación de datos
- ? Como informa los datos (Informes y Dashboard)
- ? Datos y métricas
- ? Pensamiento matemático y Big y Small Data





## 05 - Conceptos básicos sobre hojas de cálculo. Stakeholders

- ? Hojas de cálculo y el Ciclo de Vida de los datos
- ? Fórmulas en Hojas de Cálculos
- ? Funciones en Hojas de Cálculos
- ? Errores comunes en Hojas de Cálculos
- ? Dominio y alcance de los problemas
- ? Contexto de los Conjuntos de Datos



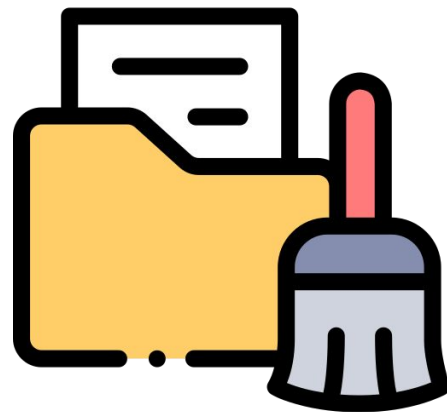
## 06 - Sesgo, credibilidad, privacidad, ética y acceso. Organizar y proteger tus datos

- ? Análisis de sesgo en datos
- ? Credibilidad de datos
- ? Ética de los datos
- ? Anonimización de datos
- ? Organización de los datos
- ? Protección de datos



## 07 - La importancia de la integridad. Datos impecables

- ? Integridad de los Datos
- ? Datos Insuficientes
- ? Limpieza de datos con hojas de cálculo
- ? Herramientas para la limpieza de datos en hojas de cálculos



## 08 - Limpieza de datos con SQL

- ? Bases de datos en el análisis computacional de datos
- ? SQL Vs Hojas de Cálculos
- ? Comandos SQL para limpieza de datos



## 09 - Visualización de datos

- ? Historia e importancia de la visualización
- ? Frameworks para organizar tus ideas sobre la visualización
- ? Atributos preatencionales: marcas y canales
- ? Elementos del arte
- ? Principios de diseño
- ? Evitar gráficos confusos o engañosos
- ? Tipos de gráficos ¿Qué utilizar en cada caso?
- ? Pensamiento de diseño (Design Thinking)
- ? Visualizaciones en Tableau
- ? Paneles (Dashboard) e información estática



## 10 - Verificar e informar los resultados de tu limpieza de datos

- ? Verificar e informar los resultados
- ? Herramientas para la verificación y limpieza final
- ? Capturar los cambios de la limpieza



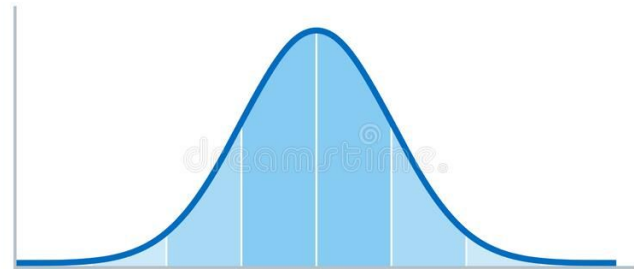
# 11 - Análisis exploratorio de datos

- ? La importancia del EDA
- ? Elementos de datos estructurados
- ? Estimación de localización
- ? Estimación de variabilidad
- ? Exploración de la distribución de datos
- ? Exploración de datos binarios y categóricos
- ? Correlación
- ? Probabilidad



## 12 - Distribuciones de datos y muestreo

- ? Muestreo aleatorio y sesgo de la muestra
- ? Tamaño y calidad de los datos
- ? La distribución muestral
- ? Teorema del límite central
- ? Error estándar
- ? Bootstrap
- ? Intervalo de confianza
- ? Distribución normal
- ? Tamaño de la muestra





# MATERIALES DEL CURSO

Recopilación e Integración de Datos

1 2 3 4 5 6 7 8 9 10 11 12

Modulo 1 - Análisis computacional de datos y pensamiento analítico  
Datos y Análisis de datos

¿Qué tienen en común las empresas de comercio electrónico, entretenimiento, atención médica, fabricación, marketing, finanzas, tecnología y cientos de otras industrias? ¡Adivínaste! Todas utilizan datos. Las organizaciones de todo tipo necesitan analistas de datos para mejorar sus procesos, identificar oportunidades y tendencias, lanzar nuevos productos, proporcionar un excelente servicio al cliente y tomar decisiones bien pensadas.

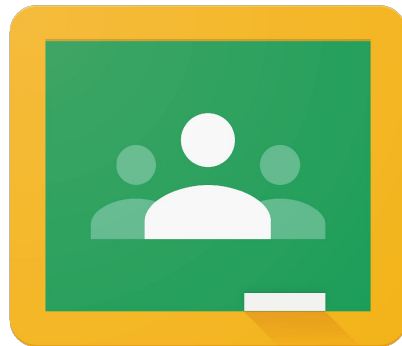
Pero primero, ¿qué son exactamente los datos? Bueno, los **datos** son una colección de hechos. Esta colección puede incluir números, imágenes, videos, palabras, mediciones, observaciones y más. Una vez que tienes los datos, se los analiza mediante el análisis. El **análisis de datos** es la recopilación, transformación y organización de los datos para sacar conclusiones, hacer predicciones e impulsar una toma de decisiones fundamentada. Y eso no es todo. Los datos evolucionan con el tiempo, lo que significa que este análisis o la analítica, como lo llamamos, puede darnos nueva información a lo largo de todo el ciclo de vida de los datos. Los datos están en todas partes. Usas y creas datos todos los días.

¿Alguna vez has leído las reseñas de un producto antes de decidir si comprarlo o no? Eso es el análisis de datos. O tal vez monitorees tu actividad física para contar tus pasos a fin de mantenerte activo durante todo el día. Eso es el análisis de datos. Pero no solo usas datos. También creas enormes cantidades de datos todos los días. Cada vez que usas el teléfono, buscas algo en línea, haces streaming de música, compras con tarjeta de crédito, publicas en redes sociales o usas el GPS para mapear una ruta, estás creando datos. Nuestro mundo digital y los millones de dispositivos inteligentes que hay allí hacen que la cantidad de datos disponible realmente sea asombrosa.

Módulo 1









Contenido

Actividades



# Google Classroom

# TIPS PARA ESTUDIAR

-  Enfocarse en una sola ruta de estudio
-  Poner objetivos cumplibles en función del tiempo disponible
-  Tener un cuaderno para tomar notas de clases y resumir materiales
-  Hacer mapas conceptuales de los temas
-  Reproducir los ejemplos dados en clases
-  Participar en las clases y foro
-  Tener curiosidad
-  Practicar, practicar, practicar



## TIPS PARA ESTUDIAR



Conocer las palabras claves en inglés



Usar traductores (Google o DeepL Translator)



Instalar las aplicaciones del traductor preferido



Acostumbrar el oído al idioma



Los mejores materiales están en inglés



Los lenguajes de programación o de consulta tienen sintaxis en inglés



Aprender inglés

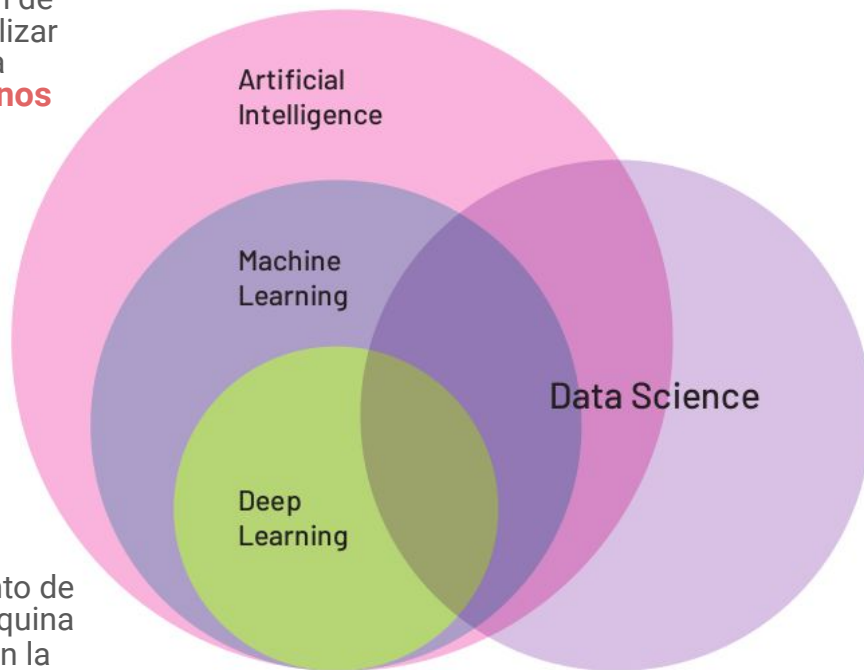


# INTRODUCCIÓN A LA CIENCIA DE DATOS

**Inteligencia Artificial** es la disciplina enfocada en la creación de **máquinas** programadas para realizar tareas que requieren la misma **inteligencia** de los **seres humanos**

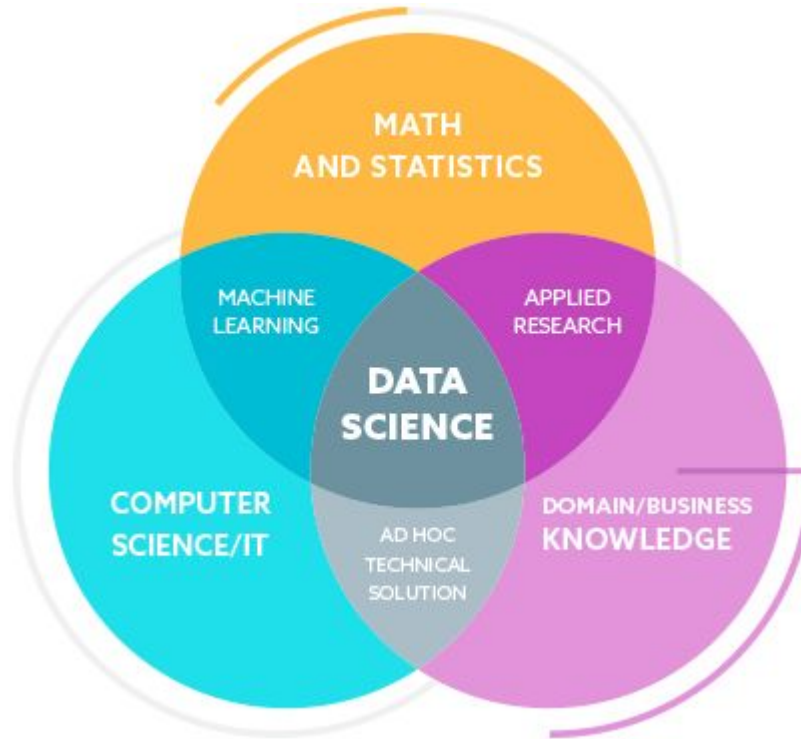
**Machine Learning** es un subconjunto de la IA donde las personas **entrenan** a las máquinas para reconocer **patrones** basados en datos y hacer sus **predicciones**

**Deep Learning** es un subconjunto de Machine Learning en el que la máquina utiliza **algoritmos** inspirados en la estructura y funcionalidades del **cerebro humano**



**Ciencia de Datos** es un área de trabajo **interdisciplinar** que incluye **procesos** de recopilación, preparación, análisis, visualización y modelización de **datos** para generar **conocimiento útil**

# INTRODUCCIÓN A LA CIENCIA DE DATOS



# ACTIVIDAD



Listar los conceptos que involucran la Ciencia de Datos



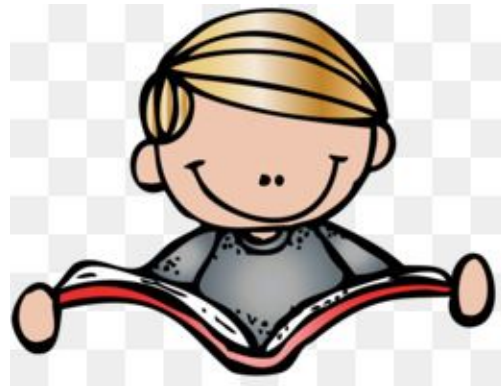
Definir cada concepto



Elegir un diagrama de Venn que crea que mejor explica la relación entre los distintos términos



Compartí en el Class Room tu producción



¿PREGUNTAS?





**¡Muchas gracias!**