



Institución  
**Universitaria**  
Reacreditada en Alta Calidad

# INTRODUCCIÓN A LA CIENCIA DE DATOS

INTRODUCCIÓN A LA CIENCIA DE DATOS

530202018

2026-1

[www.itm.edu.co](http://www.itm.edu.co)

**Docente:** Esteban Gonzalez Valencia

**Contacto:**

[estebangonzalez4302@correo.itm.edu.co](mailto:estebangonzalez4302@correo.itm.edu.co) (correo, MS Teams)

Oficina M-212

**Horario:**

Jueves: 18:00-19:59

**Modalidad Alternante**

- Laboratorio N-102 (Sistemas de comunicaciones), Sede Fraternidad
- MS. Teams (Virtual)

## Importante

1. La asignatura **NO** es habilitable.
2. El registro de faltas de asistencia y el ingreso de notas se hará tal cual se explica en el reglamento estudiantil vigente.
3. Las clases se desarrollarán en modalidad presencial, a menos que el docente indique que se mediara alguna sesión bajo la virtualidad

## Requisitos

**1. Cuenta del ITM:** Se requiere una cuenta de ITM activa para acceder TEAMS y a las licencias de la institución.



**2. Conexión a Internet:** Conexión estable para acceder y trabajar en el curso.

## ¡Cuidado!

1. La materia se puede **reprobar** por inasistencias (20%), **4 clases**.
2. Si el estudiante tiene eventos personales programados (citas médicas, entrevistas de trabajo, entre otros) que impidan su asistencia a clase, lo debe anunciar al docente para que este le establezca una alternativa de no atraso en contenidos de la asignatura.
3. En caso de incapacidad médica solo será admitida la expedida por la EPS. **Las incapacidades no eliminan las faltas de asistencia.**

## Honestidad académica

1. Utilizar herramientas de Inteligencia Artificial, como ChatGPT, Gemini, etc, para crear o editar su trabajo y presentarlo como propio.
2. No hacer copia de respuestas de sus compañeros del curso.
3. No Emplear o permitir que otra persona altere o revise su trabajo y luego presentarlo como propio

## ¿Qué es la Ciencia de datos?

La ciencia de datos es el estudio de datos con el fin de **extraer información significativa** para empresas.

Es un enfoque multidisciplinario que combina principios y prácticas del campo de las **matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación** para analizar grandes cantidades de datos.

Este análisis permite que los científicos de datos planteen y respondan a preguntas como “qué pasó”, “por qué pasó”, “qué pasará” y “qué se puede hacer con los resultados”.

## ¿Por qué es importante la Ciencia de datos?

Las organizaciones modernas están **inundadas de datos**; hay una proliferación de **dispositivos** que recopilan y almacenan información de forma automática y **los sistemas en línea** capturan más datos en los campos del comercio electrónico, la medicina, las finanzas y cualquier otro aspecto de la vida cotidiana.

Cada vez, disponemos de mayores cantidades de datos de texto, audio, imágenes y video.

La ciencia de datos es importante porque combina herramientas, métodos y tecnología para **generar significado** a partir de los datos.

## ¿Para que se utiliza la ciencia de datos?

1. Análisis descriptivo



2. Análisis de diagnóstico



3. Análisis predictivo



4. Análisis prescriptivo



## ¿Para que se utiliza la ciencia de datos?

### Diferencias entre los distintos tipos de modelos de análisis

Insights sobre lo que ha pasado



**Descriptivos**

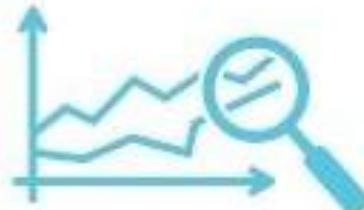
Enfoque en eventos pasados

Razones por las que pasaron



**De Diagnóstico**

¿Cuáles son los posibles resultados?



**Predictivos**

Enfoque en eventos futuros

¿Cómo obtener estos resultados?



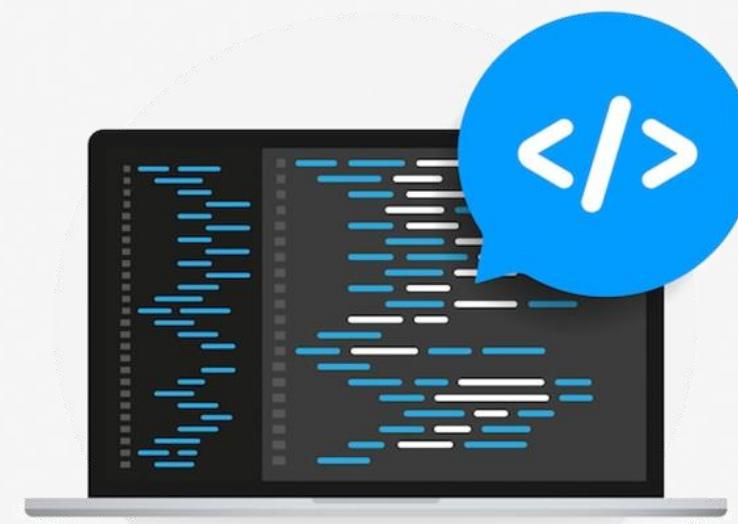
**Prescriptivos**

## Componentes claves en la ciencia de datos

### 1. Estadística



### 2. Programación



### 3. Conocimiento del área de interés



## Tipos de datos



### Datos Estructurados

A menudo, números o etiquetas, almacenados en un marco estructurado de columnas y filas relacionadas con parámetros preestablecidos.



Identificadores en bases de datos



Datos numéricos en Google sheets



Reviews en calificaciones de la Play Store



### Datos Semiestructurados

Organizados libremente en categorías utilizando etiquetas meta.



Correos en inbox, enviados y en borrador



Tweets organizados por hashtags



Folders organizados por temas



### Datos No Estructurados

Información con mucho texto que no está organizada en un marco o modelo claramente definido.



Media posts, correos, reseñas



Videos, imágenes y otro contenido multimedia



Charlas, sonidos



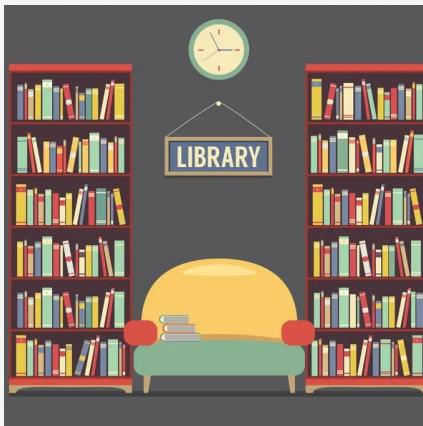
## Bases de datos

### ¿Qué son las bases de datos?

Conjunto de datos organizados y relacionados que se almacena de manera sistemática para ser consultado, gestionado y actualizado de forma eficiente.



Colección de datos  
organizada



Base de datos física

	A	B	C	D	E
1	No. Empleado	Departamento	Nombre	Apellido	
2	4000307	Recursos Humanos	Andrea	Mendoza	
3	4000308	Finanzas	Juan Carlos	Hernández	
4	4000309	Informática	Arturo	Becerril	
5	4000310	Marketing	Claudia	Camacho	
6	4000311	Dirección General	Jacob	Sarmiento	
7	4000312	Finanzas	Isidro	Jiménez	
8	4000313	Finanzas	Consuelo	Palacios	
9	4000314	Marketing	Fernanda	Fonseca	
10	4000315	Recursos Humanos	Jorge	Villarreal	
11					

Base de datos en  
términos informáticos



Institución  
**Universitaria**  
Reacreditada en Alta Calidad

# ¡MUCHAS GRACIAS!



A decorative graphic consisting of three horizontal bars of different colors: yellow, teal, and orange. The bars are slightly curved and overlap each other, positioned below the main title.

Material de apoyo usado para las clases de Introducción a la Ciencia de Datos

Elaborado por:

**Esteban Gonzalez Valencia**  
[estebangonzalez@itm.edu.co](mailto:estebangonzalez@itm.edu.co)