1. Definición de Ciencia de Datos

La Ciencia de Datos es la disciplina que combina programación, estadística y conocimiento del negocio para extraer información valiosa de los datos. Su objetivo es generar conocimiento, predicciones y apoyar la toma de decisiones basadas en evidencia.

2. Roles en la Ciencia de Datos

-Científico de datos: desarrolla modelos predictivos y analíticos.

-Ingeniero de datos: diseña la infraestructura para capturar y procesar datos.

-Analista de datos: crea reportes y visualizaciones para la toma de decisiones.

-Especialista en ML/IA: entrena modelos de machine learning.

-Business Analyst: conecta los hallazgos técnicos con el negocio.

3. ¿Qué son los dataset?

Un dataset es un conjunto estructurado o no estructurado de datos que sirve como base para análisis, visualización o entrenamiento de modelos. Puede estar en archivos `.csv`, bases de datos SQL, o incluso colecciones de imágenes y textos.