

Análisis exploratorio de datos

El Análisis Exploratorio de Datos (EDA - Exploratory Data Analysis) es una técnica usada por científicos de datos para analizar e investigar conjuntos de datos, resumiendo sus características principales mediante visualización y estadísticas. Sirve para descubrir patrones, detectar anomalías, probar hipótesis, entender relaciones entre información o evaluar la adecuación de técnicas estadísticas o de modelado.

Tipos de EDA	Funciones y técnicas con EDA	Lenguajes de análisis de datos exploratorio
<ul style="list-style-type: none"> • Univariante no gráfica: Analiza una sola variable para describir sus características básicas y detectar patrones, sin evaluar las relaciones con otras variables. • Gráfico univariante: Representa visualmente una variable para mostrar su distribución, mediante herramientas como diagramas de tallo y hojas, histogramas y diagramas de caja. • Multivariante no gráfica: Examina dos o más variables, mediante estadísticas o tablas de contingencia que permiten analizar la relación entre dos o más variables categóricas, organizándose en una tabla para identificar relaciones, pero sin usar gráficos. • Gráfico multivariante: Utiliza visualizaciones, para mostrar cómo interactúan varias variables. Incluye diagramas de barras agrupadas, dispersión, gráficos de ejecución, burbujas y mapas de calor. 	<ul style="list-style-type: none"> • Técnicas de agrupamiento de datos y reducción de dimensiones, que ayudan a crear visualizaciones gráficas de datos de alta dimensión, con muchas variables. • Visualización univariante de cada campo del conjunto de datos sin procesar, con estadísticas de resumen. • Visualizaciones bivariantes y estadísticas de resumen, que permiten evaluar la relación entre cada variable del conjunto de datos y la variable de destino que se está viendo. • Visualizaciones multivariantes, para mapear y comprender las interacciones entre los diferentes campos de los datos. • K-medias es un método para agrupar datos que se parecen entre sí. Funciona dividiendo los datos en K grupos (clústeres). Cada grupo tiene un punto central llamado centroide, y cada dato se asigna al grupo cuyo centroide esté más cerca. En otras palabras, los datos que queden más próximos entre sí, formarán el mismo grupo. Este método se usa en tareas como segmentar clientes, identificar patrones o reducir la información en imágenes. • Los modelos predictivos, como la regresión lineal, utilizan estadísticas y datos para predecir resultados. 	<ul style="list-style-type: none"> • Python: Es un lenguaje que se ejecuta mediante un intérprete y se basa en el modelo de programación orientada a objetos, utilizando tipos y comportamientos que se determinan de forma dinámica. Su amplia variedad de estructuras y bibliotecas facilita el desarrollo rápido y el análisis de datos. En EDA se usa para detectar valores faltantes y preparar datos para modelos de machine learning. • R: Lenguaje de código abierto especializado en estadística y visualización. Es ampliamente utilizado por científicos de datos para realizar análisis estadísticos profundos y generar gráficos avanzados.

Referencias:

- IBM. (s.f) *¿Qué es el análisis exploratorio de datos?* Recuperado el 4 de diciembre de 2025, de <https://www.ibm.com/mx-es/think/topics/exploratory-data-analysis>

Elaboró contenido: Dr. Humberto Marín Vega



Transformación Digital
Agencia de Transformación Digital y Telecomunicaciones



TECNOLOGICO
NACIONAL DE MEXICO