

# Actividad 3

## Recolección y Preprocesamiento de Datos

# Recolección y Preprocesamiento de Datos



## Objetivo:

Los estudiantes deberán recolectar una base de datos adecuada para su PoC y realizar el preprocesamiento necesario para su análisis.

## Pasos a seguir:

### Paso 1 - Elección la base de datos:

Los estudiantes deben seleccionar una base de datos relevante para su tema de PoC. Pueden utilizar plataformas como Kaggle, UCI Machine Learning Repository, o cualquier otra fuente confiable de datos.

### Paso 2 - Descarga de datos:

Una vez seleccionada la base de datos, los estudiantes deben descargarla y guardarla en un formato adecuado para su análisis (por ejemplo, CSV, Excel, etc.).

# Recolección y Preprocesamiento de Datos



## Paso 3 - Exploración inicial de datos:

Los estudiantes deben realizar una exploración inicial de los datos para comprender su estructura y contenido. Pueden utilizar herramientas como Pandas en Python para cargar y visualizar los datos.

## Paso 4 - Limpieza de datos:

Los estudiantes deben identificar y manejar los valores atípicos, faltantes o erróneos en la base de datos. Esto puede incluir la eliminación de filas o columnas con datos incompletos, la imputación de valores faltantes, o la corrección de errores en los datos.



# Recolección y Preprocesamiento de Datos



## Paso 5 - Transformación de datos:

Si es necesario, los estudiantes deben realizar transformaciones en los datos, como la codificación de variables categóricas, la normalización de variables numéricas, o la creación de nuevas características a partir de las existentes.

## Paso 6 - Visualización de datos:

Los estudiantes deben realizar visualizaciones exploratorias de los datos para comprender mejor sus características y patrones. Pueden utilizar gráficos como histogramas, diagramas de dispersión, o gráficos de líneas, según el tipo de datos que estén analizando.

Al completar este ejercicio, los estudiantes estarán preparados para utilizar datos de calidad y representativos en su PoC, lo que contribuirá a obtener resultados confiables y significativos en su proyecto final.



**TIC**

▶ **TALENTO**  
**TECH**

**AZ** | **PROYECTOS**  
**EDUCATIVOS**

