

Universidad Nacional de General Sarmiento

Machine Learning y la Preparación de Datos



Grupo 10: Perez Giannina y Prieto Lucas

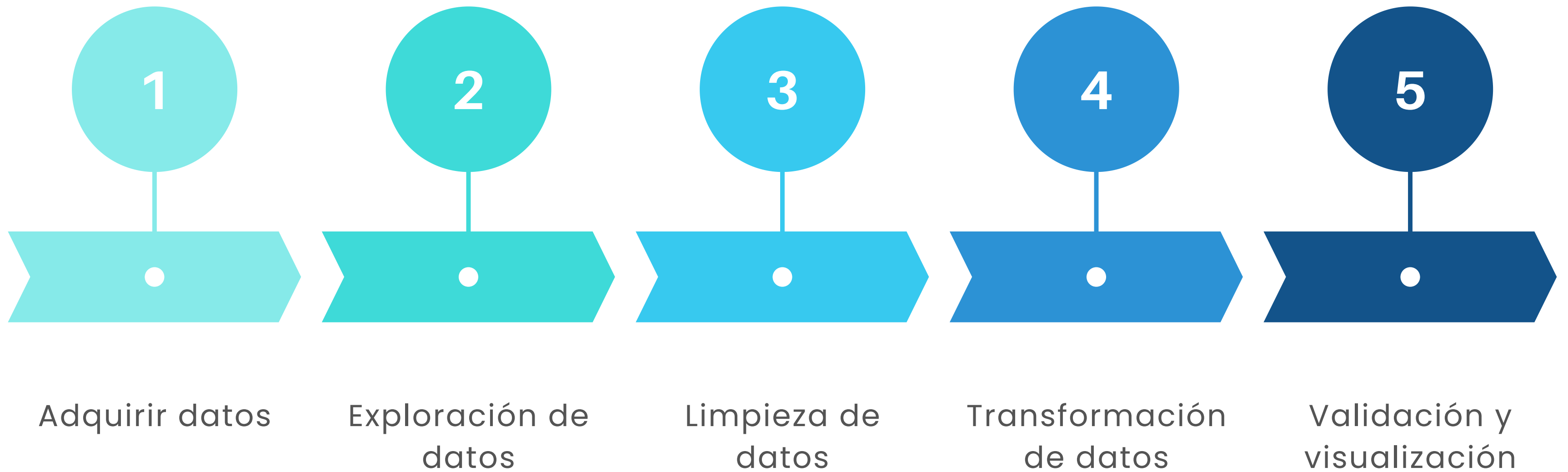
Preparación de datos

¿Por qué es importante?



- Calidad de resultados
- Eficiencia en el análisis
- Toma de decisiones informadas
- Desarrollo de modelos precisos

Pasos a seguir





Paso 1: Adquirir datos



El primer paso es adquirir los datos necesarios para el ML. Nuestros **datasets** elegidos fueron sobre:

- Accidentes viales fatales ocurridos en Argentina.
- Accidentes viales fatales y no fatales ocurridos en Gran Bretaña.



Paso 2:

Exploración de datos

En este paso se buscan posibles **errores**.
Explorar nos sirvió para:

- Encontrar valores nulos.
- Homogeneizar la información de ambos datasets.

Paso 3:

Limpieza de datos



En este paso se organizan y **limpian** los datos y se **corrigen** los errores.

- En el dataset de **Argentina**, eliminamos columnas innecesarias (*fuentes_datos, numero_victima, municipio_id, municipio_nombre, fecha_hecho, hora_hecho*).
- En el dataset de **Gran Bretaña**, nos quedamos sólo con algunas columnas (*Number of Vehicles, Lighting & Weather Conditions, Casualty Class, Casualty Severity, Sex, Age, Type of Vehicle*).

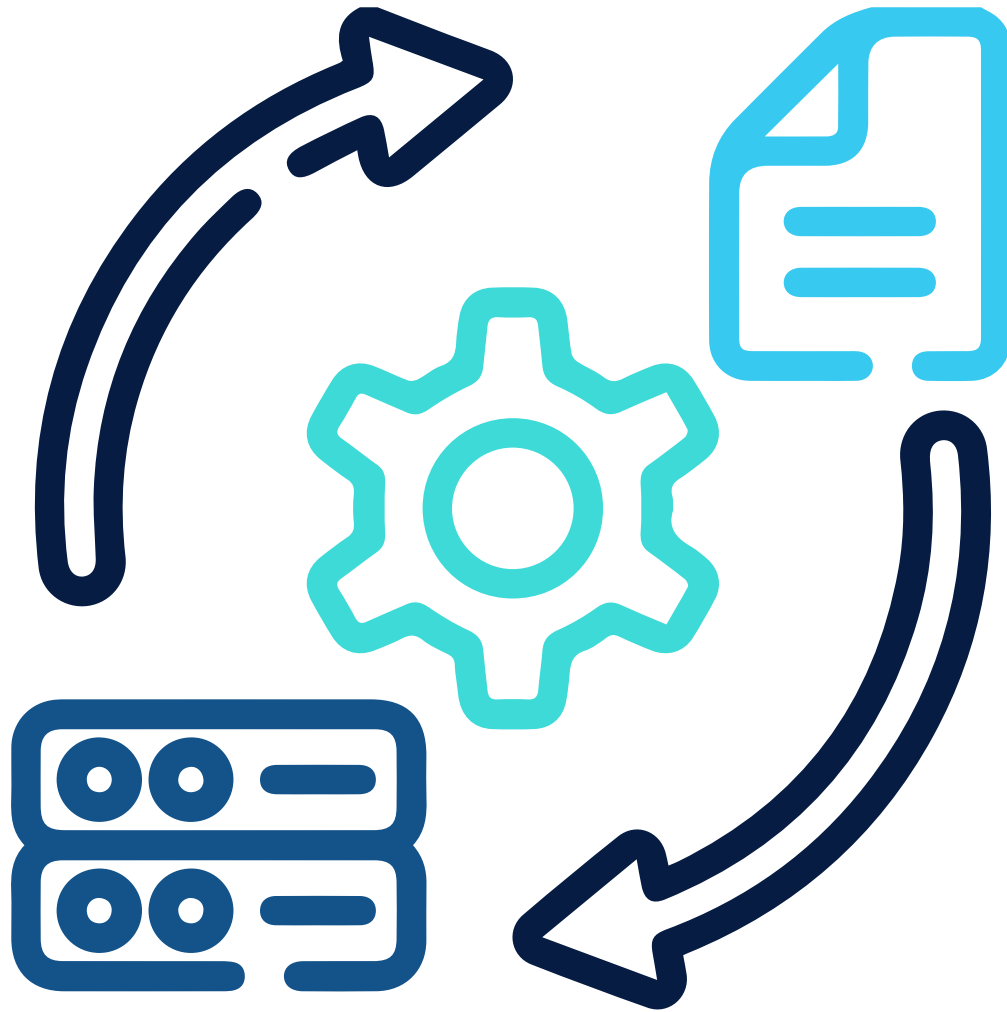
Paso 4:

Transformación de datos



Se garantiza que los datos se encuentren en un **formato** o una **estructura adecuada**.

- Agregamos la columna "Fatal" que indica si la víctima del incidente falleció (1) o no (0).
- La columna Casualty Severity paso a ser "Fatal", cuyos valores "Slight" y "Serious" se convirtieron en 0 y los valores "Fatal" en 1.
- Traducción de los datos del dataset de Gran Bretaña.
- Transformación de la columna Number of Vehicles, para clasificar el tipo de incidente.





Paso 5: Validación y visualización



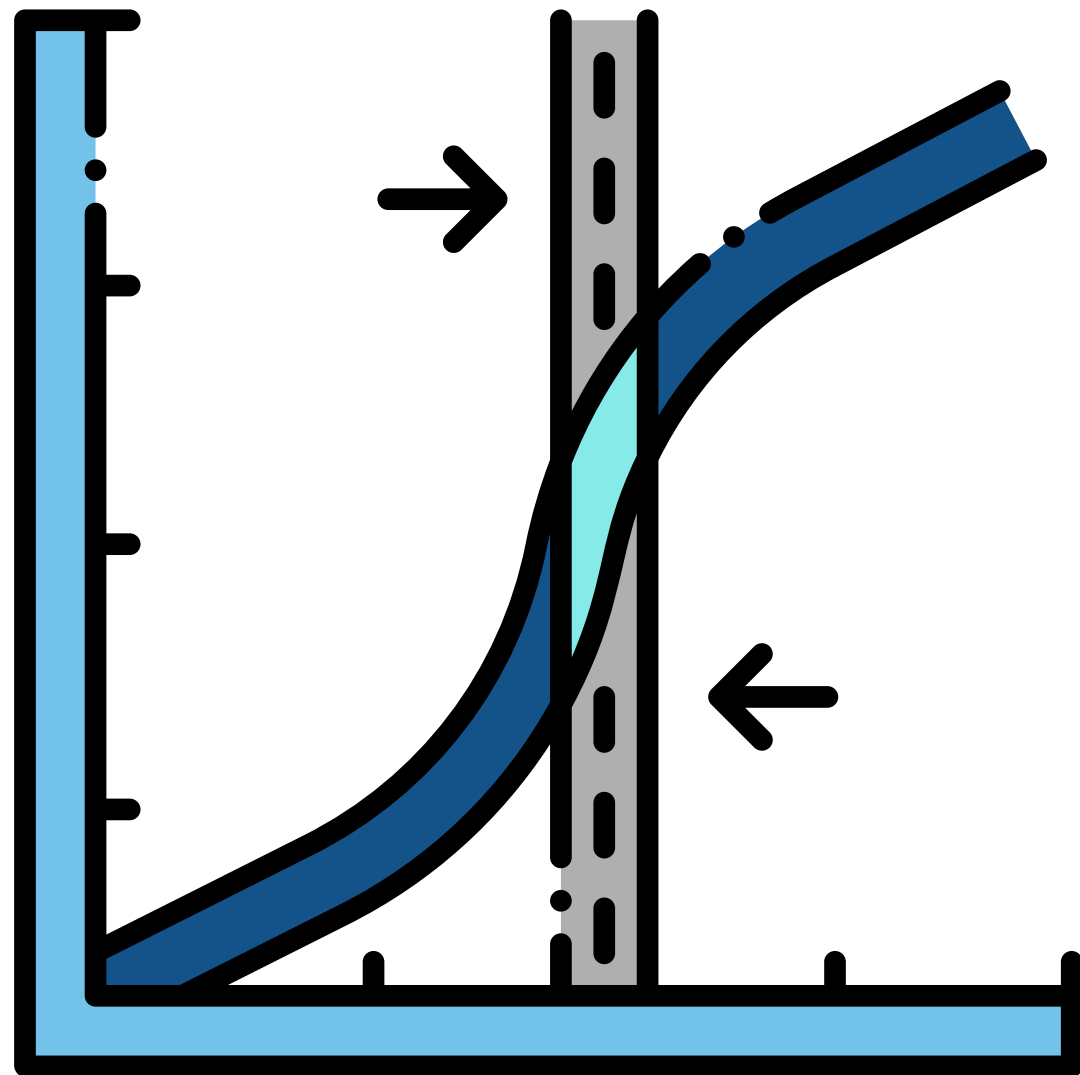
En la última etapa los datos están **limpios** y **etiquetados**.

Paso 5: Validación y visualización



momento_hecho	edad	sexo	clase_vic	vehiculo_vic	tipo_via	tipo_incidente	clima	fatal
Diurno	51	Femenino	Pasajero	Colectivo	Calle	Colisión vehículo / Objeto-Vuelco	Bueno	0
Diurno	26	Femenino	Conductor	Automóvil	Calle	Colisión vehículo / vehículo	Bueno	0
Diurno	20	Masculino	Conductor	Motocicleta	Calle	Colisión vehículo / vehículo	Bueno	1
Nocturno	57	Masculino	Conductor	Automóvil	Ruta Nacional	Colisiín vehículo / Objeto-Vuelco	Bueno	1

Entrenamiento del modelo



- Accidentes fatales y no fatales
- Modelo: Regresión logística
- Clasificación de accidentes entre los que tienen más probabilidades de tener víctimas fatales y los que no

Entrenamiento del modelo

- Transformamos las variables categóricas a través de la codificación "one-hot".
- Manejamos los datos de forma aleatoria.



Original

Gender
Male
Female
Male
Male



One-hot encoded

Gender	Male	Female
Male	1	0
Female	0	1
Male	1	0
Male	1	0