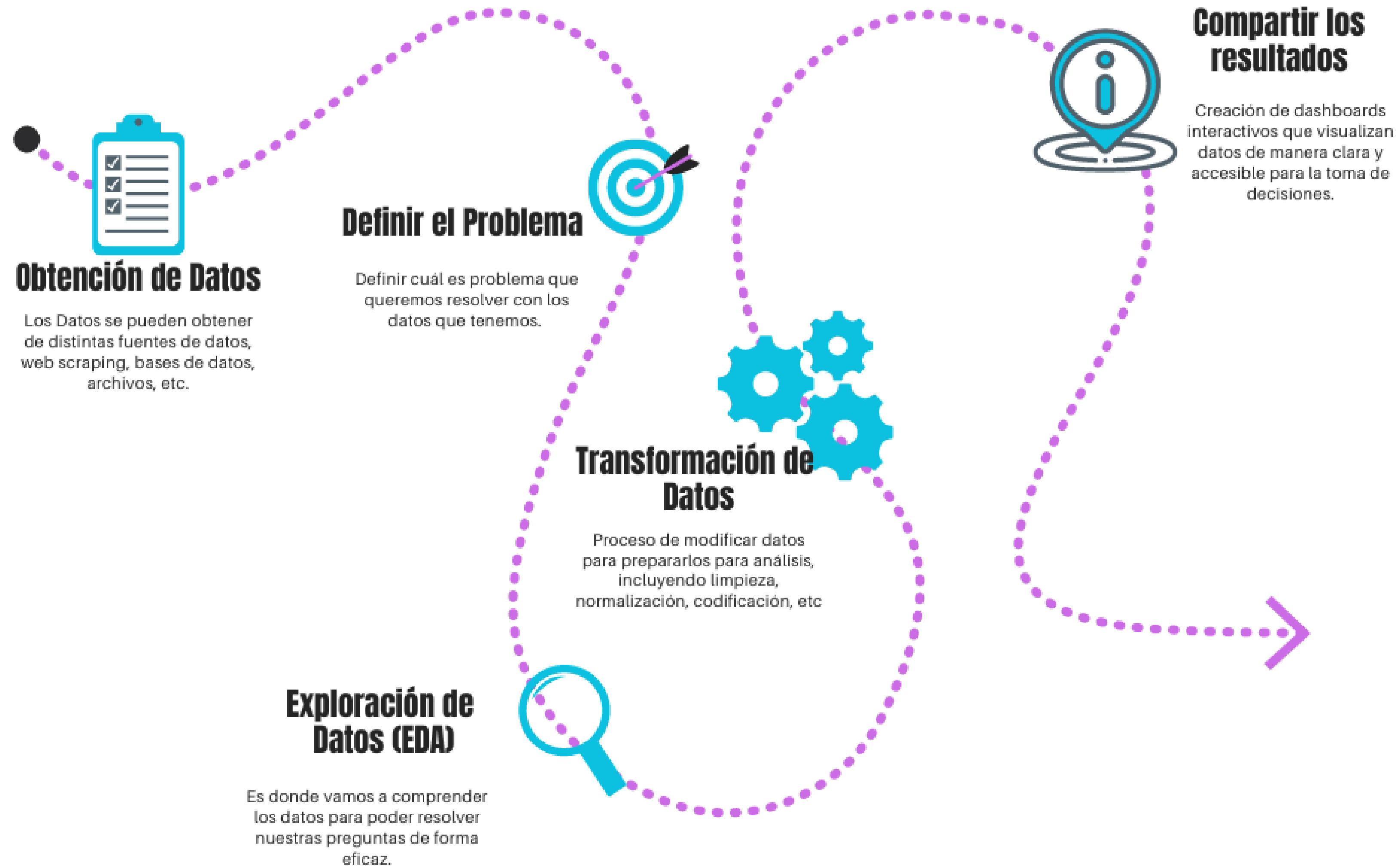


# Transformación de Datos

Preparando los datos para poder predecir

## CLAVES PARA TRABAJAR CON DATOS





## Principales Fases:

- Imputación de Nulos
- Imputación de Outliers
- Selección de Variables
- **Codificación de variables categóricas (Encoding)**
- **Estandarización**

# Codificación VC

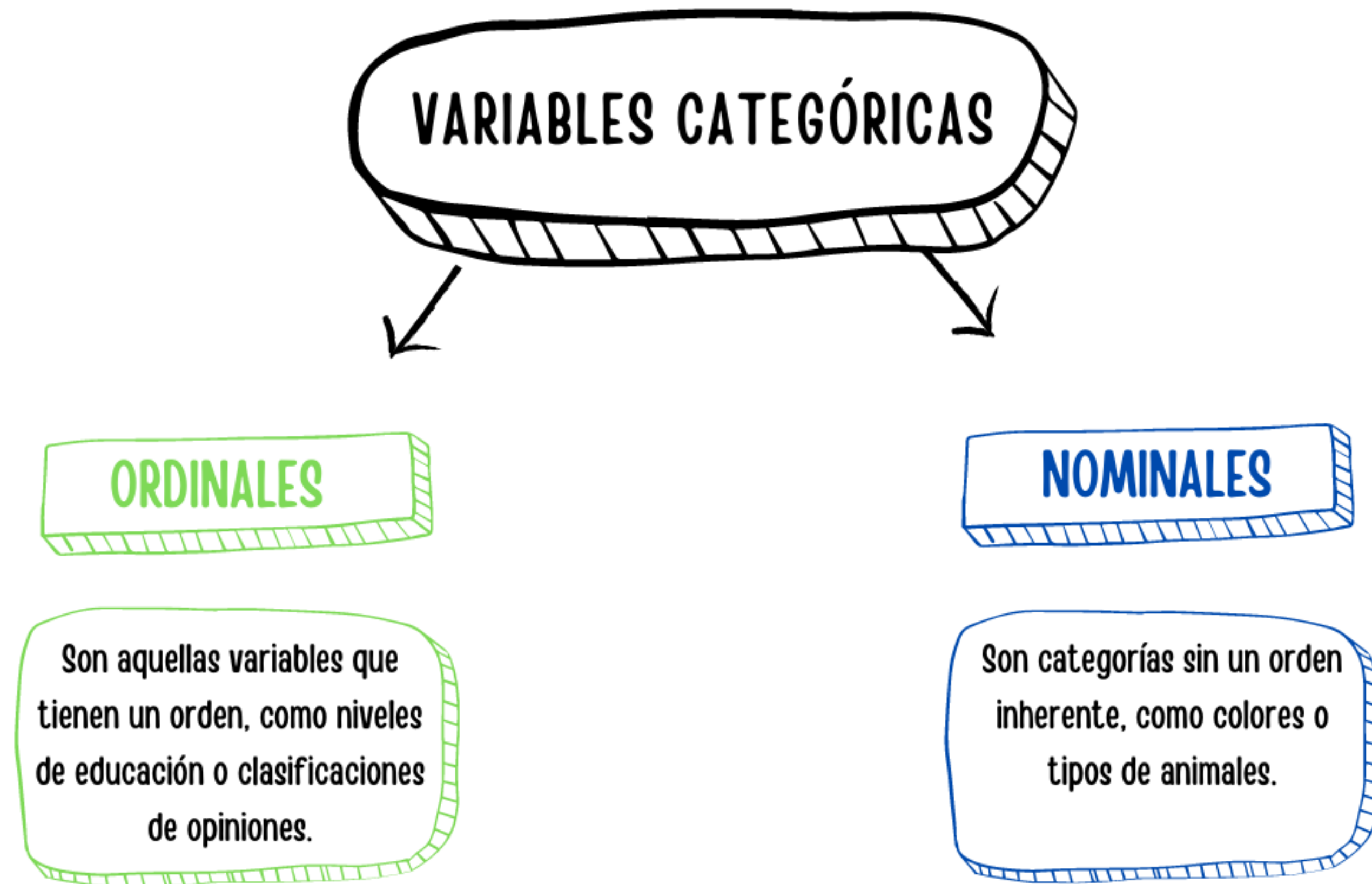
El encoding es crucial en ML porque los algoritmos requieren datos numéricos, mientras que muchas características son categóricas.

En esta fase vamos a transformar las columnas categóricas a formatos numéricos comprensibles, permitiendo que los modelos aprendan relaciones y patrones precisos para tomar decisiones efectivas.

Para esto primero tendremos que tipo de variables categóricas tenemos en el conjunto de datos.

# Tipos Variables Categorías

(basado en si tienen orden o no)



$$Z = \frac{x_i - \bar{X}}{\sigma}$$

Donde:

- $x_i$ : cada valor de la columna
- $\bar{x}$ : es la media de la columna
- $\sigma$ : la desviación estándar

## Estandarización VN

La estandarización es el proceso de transformar los datos para que tengan una media de cero y una desviación estándar de uno.

Se realiza para igualar la escala de las variables predictoras/independientes, lo que facilita la comparación y mejora el rendimiento de los algoritmos.

# Web para hacer el modelo