# Ejercicios para practicar: Depuración

El conjunto de datos DatosVino contiene información sobre ciertas características de vinos, junto con las ventas de los mismos.

Las variables contenidas en el fichero son:

| Variable | Descripción |
|---|---|
| Id | Código identificativo del tipo de vino |
| Beneficio (objetivo) | Beneficio obtenido por la venta de ese tipo de vino |
| Compra (objetivo) | Variable dicotómica que toma valor 1 si se ha realizado algún pedido de ese tipo de vino, y 0, en caso contrario |
| Acidez | Características químicas de los distintos tipos de vino:<br>• Densidad y azúcar: sólo valores positivos.<br>• pH: entre 4 y 10.<br>• Restantes: rango ilimitado de valores |
| Acidocitrico | |
| Azucar | |
| Clorurosodico | |
| Densidad | |
| Ph | |
| Sulfatos | |
| Alcohol | Contenido de alcohol en % (debe situarse entre 0 y 100) |
| Etiqueta | Percepción del diseño de la etiqueta (MM=muy malo, M=malo, R=regular, B=bueno, MB=muy bueno) |
| CalifProductor | Calificación (entre 0 y 9) del vino según el productor. |
| Clasificacion | Clasificación obtenida por un equipo de expertos (4 * = excelente, 1 * = pobre) |
| Region | Región de la que proviene (toma 3 valores distintos) |
| PrecioBotella | Precio por botella |

**El ejercicio consta de las siguientes partes:**

1) Crear un proyecto, instala las librerías necesarias y cárgalas.
2) Importar el conjunto de datos y asegúrate de que todas las variables sean del tipo adecuado.
3) Realiza un análisis descriptivo para determinar si existen errores en las variables (valores mal codificados, valores fuera de rango, categorías con poca representación, simetría y curtosis de las variables cuantitativas, etc.).
4) Realiza las correcciones pertinentes, teniendo en cuenta los errores detectados con anterioridad y verifica que los cambios se hayan realizado correctamente.
5) Analiza la existencia de valores atípicos para las variables cuantitativas según el método más apropiado y transformarlos a datos faltantes si lo consideras oportuno.
6) Crea una variable que guarde información sobre la proporción de valores perdidos por observación.
7) Analiza si existe alguna observación y/o variable con demasiados datos faltantes y, de ser así, elimínala.
8) Para las variables con datos faltantes, decide cómo tratarlos, justificando la respuesta. En caso de imputar los datos ausentes. Hazlo con la media para las variables cuantitativas y de modo aleatorio con las variables cualitativas.