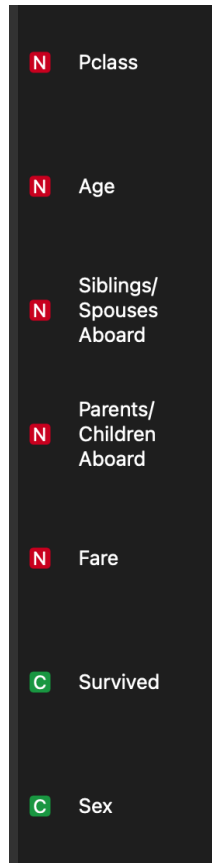


## 1. Investigar sobre las variables y dominio

El dominio es de 887 filas y 8 columnas. Con 3 variables categóricas y 5 numéricas

Las variables en cuestión son las siguientes



N	Pclass
N	Age
N	Siblings/ Spouses Aboard
N	Parents/ Children Aboard
N	Fare
C	Survived
C	Sex

Las variables numéricas son las siguientes:

**PClass**, significa clase del pasajero, representa en este caso el nivel socioeconómico, siendo **1** = Primera clase, **2** = Segunda clase y **3** = Tercera clase

**Age**, es la edad del pasajero (en años)

**Siblings/Spouses Aboard**, indica el número de **hermanos/as o cónyuge a bordo**.

- **0** significa que viajaba solo en ese sentido.
- Ejemplo: **1** si viajaba con un hermano o pareja, **2** si viajaba con 2 hermanos o 1 hermano y su pareja, y así...

**Parents/Children Aboard**, indica el número de **padres/madres o hijos/as a bordo**.

- Ejemplo: 2 si viajaba con sus padres o hijos.

**Fare**, representa el precio que pagó por el pasaje (en libras).

- Es una variable continua.
- Ejemplo: 7.25, 71.28

y las siguientes son las variables categóricas:

**Survived**, es un Indicador de si la persona sobrevivió o no:

- 0 = No sobrevivió
- 1 = Sobrevivió

**Sex**, nos da información del Sexo del pasajero:

- male o female

Y por último, tenemos la variable categórica name, que es de formato **string**

2.

La edad media de la tripulación era de unos 29 años, de los cuales el 65% eran hombres y el restante 35% mujeres.

3.Verificar tipos de variables y realizar ajustes si es necesario

#### 4. Identificar outliers y plantear que conviene hacer en este caso

Hay un caso de outliers, con la variable siblings/spouses aboard, en este caso conviene reemplazarla por cero.

La moda y la mediana son ambas cero para esta variable, y la media de 0,53. El valor outlier es de 8, y termina sesgando positivamente la media, por eso es conveniente reemplazarla por cero.

Exactamente lo mismo ocurre con la variable Parents/Children aboard

Y por último tenemos el caso de Fare, con dos outliers, uno de cero, y otro de 512. En este caso calcularía una nueva media sin tener en cuenta esos valores, y reemplazaría ambos outliers por esa media insesgada

#### 5. Realizar las correlaciones

