

Informe de Análisis Exploratorio de **mtcars**

Melani Forsythe Matos
Daniela Guerrero Álvarez
Rubén Martínez Rojas

El conjunto de datos **Titanic** es un conjunto clásico que contiene información sobre los pasajeros del Titanic, y se utiliza frecuentemente en análisis de supervivencia. A continuación, se describen las variables incluidas en este conjunto de datos, su significado, tipo de escala, y si son discretas o continuas:

1. **PassengerId**

- **Descripción:** ID único para cada pasajero.
- **Escala:** Cuantitativa Discreta.
- **Significado:** Un identificador numérico para cada pasajero en la base de datos.

2. **Class**

- **Descripción:** Clase del pasajero (1st = Primera, 2nd = Segunda, 3rd = Tercera, Crew = Trabajadores).
- **Escala:** Cualitativa Ordinal.
- **Significado:** Representa la clase socioeconómica del pasajero, con 1 siendo la clase más alta y 3 la más baja, y los trabajadores.

3. **Age**

- **Descripción:** Si el pasajero es Adulto o Niño.
- **Escala:** La edad normalmente es una variable cuantitativa, pero en este caso está dividido por grupos etáreos, Adulto y Niño, por tanto se clasifica como Cualitativa Nominal.
- **Significado:**

4. **Survived**

- **Descripción:** Indicador de supervivencia (No, Sí).
- **Escala:** La variable es Cualitativa Nominal, pero podría convertirse a Cuantitativa Discreta (Binaria), asignándole 1 a Sí y 0 a no, o viceversa.
- **Significado:** Indica si el pasajero sobrevivió al hundimiento del Titanic.

5. **Sex**

- **Descripción:** Género del pasajero.
- **Escala:** Cualitativa Nominal.
- **Significado:** Indica el sexo del pasajero (hombre o mujer).

6. **Freq**

- **Descripción:** Número de pasajeros.
- **Escala:** Cuantitativa Discreta.
- **Significado:** Indica la **frecuencia** o el **número de pasajeros** que caen en cada combinación de las otras variables (Class, Sex, Age, Survived).

Ejemplo:

- **Class:** La clase del pasajero (1st, 2nd, 3rd, Crew).
- **Sex:** El género del pasajero (Male, Female).
- **Age:** El grupo de edad del pasajero (Child, Adult).
- **Survived:** Si el pasajero sobrevivió o no (No, Yes).
- **Freq:** El número de pasajeros que pertenecen a esa combinación de Class, Sex, Age, y Survived.

Supongamos que tienes la siguiente fila en el dataset:

Class	Sex	Age	Survived	Freq
1st	Male	Adult	Yes	57

Esto significa que **57 pasajeros adultos de sexo masculino en primera clase sobrevivieron**.

La columna **Freq** te permite ver cuántas personas están representadas por cada combinación de las otras variables, en lugar de tener una fila separada para cada pasajero individual.



Figure 1: Cantidad de Niños Varones Sobrevivientes por Clase

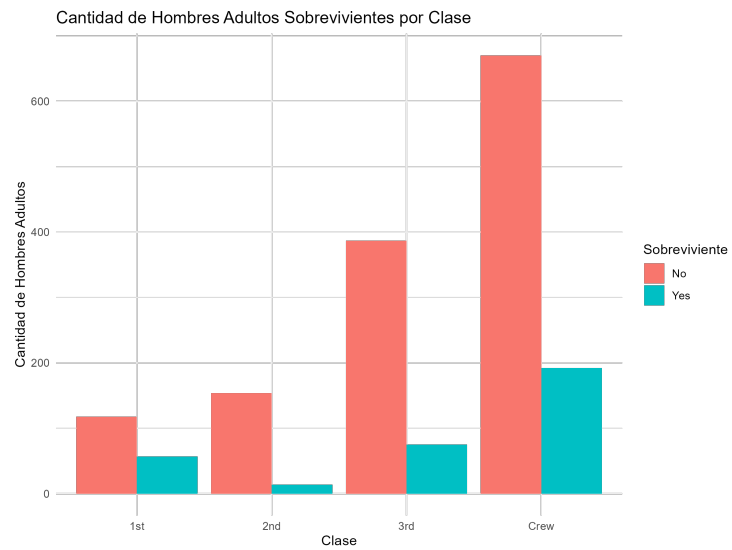


Figure 2: Cantidad de Hombres Adultos Sobrevivientes por Clase

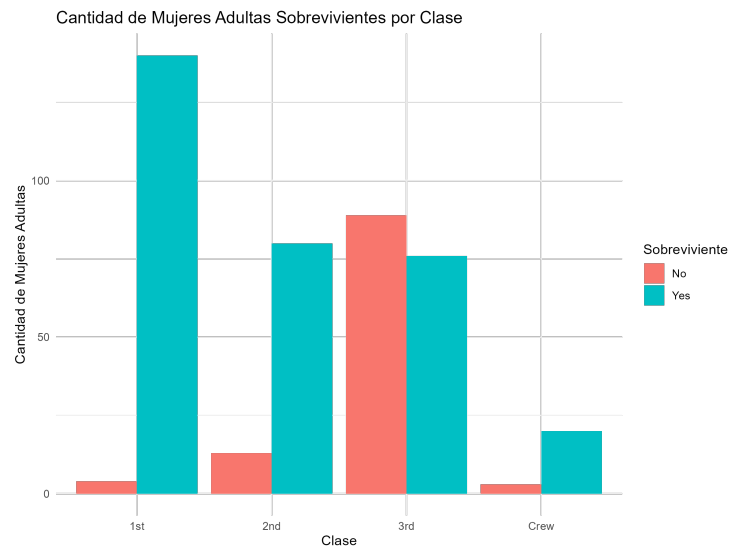


Figure 3: Cantidad de Mujeres Adultas Sobrevivientes por Clase

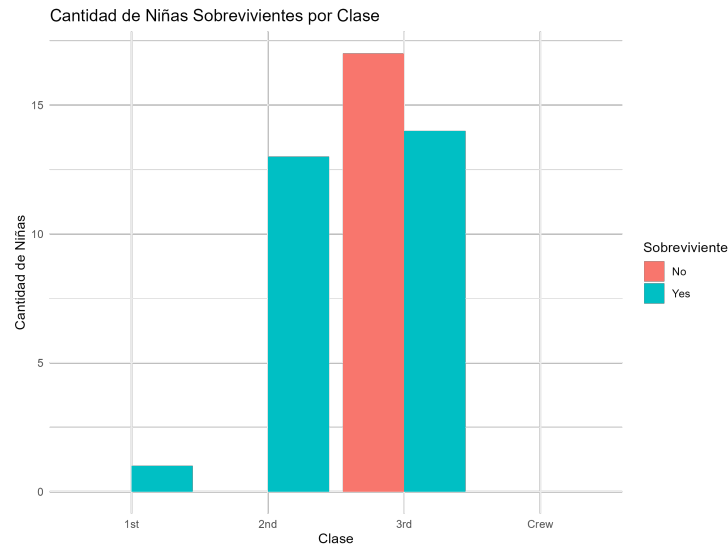


Figure 4: Cantidad de Niñas Sobrevivientes por Clase

1 Matriz de Correlación

La matriz de correlación muestra la relación entre diferentes variables categóricas convertidas a numéricas en el dataset del Titanic. Algunos hallazgos clave son:

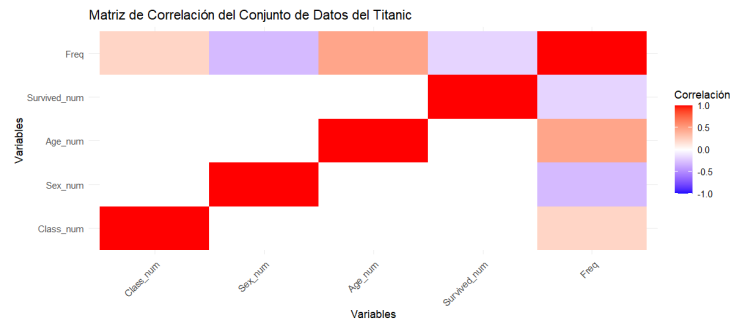


Figure 5: Grafico de Calor , Correlaciones

1.1 Clase y Supervivencia

Es probable que se observe una correlación positiva entre **Class_num** y **Survived_num**, lo que sugiere que los pasajeros de primera clase tenían más probabilidades de sobrevivir en comparación con los de tercera clase o la tripulación.

1.2 Sexo y Supervivencia

Podría haber una correlación positiva entre `Sex_num` y `Survived_num`, indicando que las mujeres tenían una tasa de supervivencia más alta que los hombres.

1.3 Edad y Supervivencia

La correlación entre `Age_num` y `Survived_num` podría indicar que los niños (Child) tenían una mayor tasa de supervivencia en comparación con los adultos.

2 Prueba de Chi-Cuadrado

La prueba de chi-cuadrado se utilizó para determinar si existe una asociación significativa entre la clase de los pasajeros y su supervivencia. Si el valor p es menor que 0.05, se concluye que hay una relación significativa entre estas variables.

Un valor p menor que 0.05 implica que ciertos grupos, como los pasajeros de primera clase, tenían más probabilidades de sobrevivir en comparación con aquellos de tercera clase o la tripulación. Esto sugiere que factores como la clase social influyeron en las tasas de supervivencia durante el desastre del Titanic.

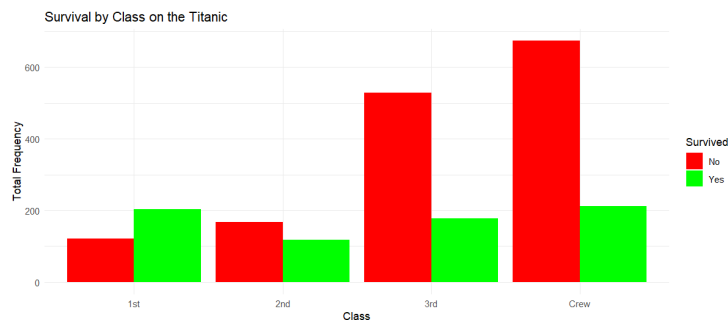


Figure 6: Chi -Cuadrado