



Centro de Investigación
en Computación
Instituto Politécnico Nacional

Instituto Politécnico Nacional

Julio 2023

Profesor: Alan Badillo Salas

Alumno: Samuel Axel Castro Juárez

INTRODUCCIÓN

El hundimiento del Titanic en 1912 fue un desastre marítimo de gran magnitud que causó la pérdida de numerosas vidas. El estudio de los datos disponibles sobre los pasajeros del Titanic puede brindarnos información valiosa sobre los factores que influyeron en las tasas de supervivencia y las características demográficas de los supervivientes. En este contexto, se ha detectado la necesidad de analizar una serie de puntos clave en la base de datos del Titanic para comprender mejor las probabilidades de supervivencia de diferentes grupos de pasajeros.

JUSTIFICACIÓN

La motivación para resolver este problema radica en la importancia de comprender las circunstancias que rodearon el desastre del Titanic y cómo influyeron en las tasas de supervivencia de los pasajeros..

La obtención de datos sobre el total de supervivientes, tanto en general como desglosados por género y grupos de edad, nos permitirá comprender mejor qué grupos tuvieron mayores tasas de supervivencia y si existieron diferencias significativas en la probabilidad de sobrevivir en función de estas variables.

Además, investigar las tasas de supervivencia por clase socioeconómica nos brindará información sobre cómo la posición social pudo haber influido en las posibilidades de sobrevivir al desastre. Este análisis nos permitirá evaluar si existieron desigualdades en la distribución de los recursos de salvamento durante el hundimiento del Titanic.

El uso de la regla de Bayes para calcular la probabilidad de sobrevivir dado el género o la clase socioeconómica nos permitirá obtener estimaciones más precisas sobre las posibilidades de supervivencia en diferentes escenarios. Este enfoque nos ayudará a comprender cómo estas variables se relacionan con la probabilidad de sobrevivir y cómo interactúan entre sí.

En resumen, la resolución de este problema nos permitirá obtener una visión más clara de los factores que afectaron las tasas de supervivencia en el desastre del Titanic, lo que a su vez puede tener implicaciones para la toma de decisiones en situaciones similares en el futuro y para mejorar la seguridad en el transporte marítimo.

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

1. Cargar el DataFrame

```
[2]: data = pd.read_csv("https://raw.githubusercontent.com/dragonnomada/pycien-2023/main/datasets/titanic.csv")

data.columns = ["ID", "SOBREVIVE", "CLASE", "NOMBRE", "SEXO", "EDAD",
                "HERMANOS", "PADRES", "TICKET", "TARIFA", "CABINA", "MUELLE"]
```

2. Limpieza de Datos

Vamos a rectificar los datos para evitar valores malformados.

```
[3]: data["SEXO"] = data["SEXO"].map({
    "female": "MUJER",
    "male": "HOMBRE"
})

def generar_edad_aleatoria(edad):
    if np.isnan(edad):
        return np.random.normal(data["EDAD"].mean(), data["EDAD"].std())
    else:
        return edad
data["EDAD"] = data["EDAD"].map(generar_edad_aleatoria).map(int)
## Para sustituir los datos NULL por "X"
data["CABINA"] = data["CABINA"].fillna("X")

data.sample(10)
```

	ID	SOBREVIVE	CLASE		NOMBRE	SEXO	EDAD	HERMANOS	PADRES	TICKET	TARIFA	CABINA	MUELLE
711	712	0	1		Klaber, Mr. Herman	HOMBRE	40	0	0	113028	26.5500	C124	S
695	696	0	2		Chapman, Mr. Charles Henry	HOMBRE	52	0	0	248731	13.5000	X	S
600	601	1	2	Jacobsohn, Mrs. Sidney Samuel (Amy Frances Chr...	MUJER	24	2	1	243847	27.0000	X	S	
643	644	1	3		Foo, Mr. Choong	HOMBRE	32	0	0	1601	56.4958	X	S
797	798	1	3		Osman, Mrs. Mara	MUJER	31	0	0	349244	8.6833	X	S
470	471	0	3		Keefe, Mr. Arthur	HOMBRE	58	0	0	323592	7.2500	X	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	MUJER	35	1	0	113803	53.1000	C123	S	
468	469	0	3		Scanlan, Mr. James	HOMBRE	39	0	0	36209	7.7250	X	Q
682	683	0	3		Olsvigen, Mr. Thor Anderson	HOMBRE	20	0	0	6563	9.2250	X	S

Total de sobrevivientes

```
[4]: sobrevivio = data["SOBREVIVE"]
sobrevivientes= data[sobrevivio==1]["SOBREVIVE"].count()
total= sobrevivio.count()

print("El total de sobrevivientes es {} de un total de {}".format(sobrevivientes, total))
```

El total de sobrevivientes es 342 de un total de 891

▼ Total de supervivientes mujeres

```
[5]: sexo= data["SEXO"]
sobrevivio= data["SOBREVIVE"]
mujerSobrevivio= data[(sexo=="MUJER") & (sobrevivio==1)]["SEXO"].count()
print("El total de mujeres que sobrevivieron es de {}".format(mujerSobrevivio))
```

El total de mujeres que sobrevivieron es de 233

Total de supervivientes hombres

```
[6]: sexo= data["SEXO"]
sobrevivio= data["SOBREVIVE"]
hombreSobrevivio= data[(sexo=="HOMBRE") & (sobrevivio==1)]["SEXO"].count()
print("El total de hombres que sobrevivieron es de {}".format(hombreSobrevivio))
```

El total de hombres que sobrevivieron es de 109

Total de supervivientes mayores a 18 años

```
[7]: edad= data["EDAD"]
sobrevivio= data["SOBREVIVE"]
sobrevivio18= data[(sobrevivio==1) & (edad>18)]["SOBREVIVE"].count()
print("El total de sobrevivieron mayores a 18 años es de {}".format(sobrevivio18))
```

El total de sobrevivieron mayores a 18 años es de 259

Total de supervivientes menores a 18 años

```
[8]: edad= data["EDAD"]
sobrevivio= data["SOBREVIVE"]
sobrevivioMenor18= data[(sobrevivio==1) & (edad<18)]["SOBREVIVE"].count()
print("El total de sobrevivieron menores a 18 años es de {}".format(sobrevivioMenor18))
```

El total de sobrevivieron menores a 18 años es de 73

Total de supervivientes mayores a 50 años

```
[9]: edad= data["EDAD"]
sobrevivio= data["SOBREVIVE"]
sobrevivio50= data[(sobrevivio==1) & (edad>50)]["SOBREVIVE"].count()
print("El total de sobrevivieron mayores a 50 años es de {}".format(sobrevivio50))
```

El total de sobrevivieron mayores a 50 años es de 28

Total de muertos

```
[10]: sobrevivio= data["SOBREVIVE"]
muertos= data[sobrevivio==0]["SOBREVIVE"].count()
print("El total de muertos es de {}".format(muertos))
```

El total de muertos es de 549

Total de muertos mujeres

```
[11]: sexo= data["SEXO"]
sobrevivio= data["SOBREVIVE"]
mujerMuertas= data[(sexo=="MUJER") & (sobrevivio==0)]["SEXO"].count()
print("El total de mujeres que murieron es de {} de un total de muertos de {}".format(mujerMuertas, muertos))
```

El total de mujeres que murieron es de 81 de un total de muertos de 549

Total de muertos hombres

```
[12]: sexo= data["SEXO"]
sobrevivio= data["SOBREVIVE"]
mujerHombres= data[(sexo=="HOMBRE") & (sobrevivio==0)]["SEXO"].count()
print("El total de hombres que murieron es de {} de un total de muertos de {}".format(mujerHombres, muertos))
```

El total de hombres que murieron es de 468 de un total de muertos de 549

Tasa de supervivencia de la clase 1

```
[13]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
clase1Sobrevivio= data[(clase==1) & (sobrevivio==1)]["SOBREVIVE"].count()
clase1= data[clase==1]["CLASE"].count()
tasaSobreviveClase1= clase1Sobrevivio/clase1
print("La tasa de supervivencia de la clase 1 es de {:.2f}% del total de la clase 1 que es {}".format(tasaSobreviveClase1, clase1))
```

La tasa de supervivencia de la clase 1 es de 0.63% del total de la clase 1 que es 216

▼ Tasa de supervivencia de la clase 2

```
[14]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
clase2Sobrevivio= data[(clase==2) & (sobrevivio==1)]["SOBREVIVE"].count()
clase2= data[clase==2]["CLASE"].count()
tasaSobreviveClase2= clase2Sobrevivio/clase2
print("La tasa de supervivencia de la clase 2 es de {:.2f}% del total de la clase 1 que es {}".format(tasaSobreviveClase2, clase1))
```

La tasa de supervivencia de la clase 2 es de 0.47% del total de la clase 1 que es 184

Tasa de supervivencia de la clase 3

```
[15]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
clase3Sobrevivio= data[(clase==3) & (sobrevivio==1)][("SOBREVIVE")].count()
clase3= data[clase==3][("CLASE")].count()
tasaSobreviveClase3= clase3Sobrevivio/clase3
print("La tasa de supervivencia de la clase 2 es de {:.2f}% del total de la clase 1 que es {}".format(tasaSobreviveClase3, clase3))
```

La tasa de supervivencia de la clase 2 es de 0.24% del total de la clase 1 que es 491

Tasa de supervivencia de la clase 1 siendo mujer

```
[16]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
clase1SobrevivioMujer= data[(clase==1) & (sobrevivio==1) & (sexo=="MUJER")][("SOBREVIVE")].count()
clase1Mujer= data[(clase==1) & (sexo=="MUJER")][("SOBREVIVE")].count()
tasaSobreviveClase1= clase1SobrevivioMujer/clase1Mujer
print("La tasa de supervivencia de la clase 1 siendo mujer es de {:.2f}% del total de la clase 1 de mujeres que es {}".format(tasaSobreviveClase1, clase1Mujer))
```

La tasa de supervivencia de la clase 1 siendo mujer es de 0.97% del total de la clase 1 de mujeres que es 94

Tasa de supervivencia de la clase 2 siendo mujer

```
[17]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
clase2SobrevivioMujer= data[(clase==2) & (sobrevivio==1) & (sexo=="MUJER")][("SOBREVIVE")].count()
clase2Mujer= data[(clase==2) & (sexo=="MUJER")][("SOBREVIVE")].count()
tasaSobreviveClase1= clase2SobrevivioMujer/clase2Mujer
print("La tasa de supervivencia de la clase 2 siendo mujer es de {:.2f}% del total de la clase 2 de mujeres que es {}".format(tasaSobreviveClase1, clase2Mujer))
```

La tasa de supervivencia de la clase 2 siendo mujer es de 0.47% del total de la clase 2 de mujeres que es 76

▼ Tasa de supervivencia de la clase 2 siendo mujer

```
[18]: clase= data["CLASE"]
sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
clase3SobrevivioMujer= data[(clase==3) & (sobrevivio==1) & (sexo=="MUJER")][("SOBREVIVE")].count()
clase3Mujer= data[(clase==3) & (sexo=="MUJER")][("SOBREVIVE")].count()
tasaSobreviveClase3= clase3SobrevivioMujer/clase3Mujer
print("La tasa de supervivencia de la clase 3 siendo mujer es de {:.2f}% del total de la clase 3 de mujeres que es {}".format(tasaSobreviveClase3, clase3Mujer))
```

La tasa de supervivencia de la clase 3 siendo mujer es de 0.50% del total de la clase 3 de mujeres que es 144

Probabilidad de sobrevivir dado que se es mujer (usar regla de Bayes)

```
[19]: sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
P_B_A=(data[(sobrevivio==1) & (sexo=="MUJER")][("SEXO")].count())/data[sobrevivio==1][("SOBREVIVE")].count()
total=data[("SOBREVIVE")].count()

probabilidadSobrevivir=(data[sobrevivio==1][("SOBREVIVE")].count())/total
probabilidadMujer=data[sexo=="MUJER"][("SEXO")].count()/total

P_A_B= (P_B_A * probabilidadSobrevivir)/probabilidadMujer

print("la probabilidad de sobrevivir dado que se es mujer es de {:.2f}% ".format(P_A_B))
```

la probabilidad de sobrevivir dado que se es mujer es de 0.74%

Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes)

```
[20]: sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
P_B_A=(data[(sobrevivio==1) & (sexo=="HOMBRE")]["SEXO"].count())/data[sobrevivio==1]["SOBREVIVE"].count()
total=data["SOBREVIVE"].count()

probabilidadSobrevivir=(data[sobrevivio==1]["SOBREVIVE"].count())/total
probabilidaHombre=data[sexo=="HOMBRE"]["SEXO"].count()/total

P_A_B= (P_B_A* probabilidadSobrevivir)/probabilidaHombre

print("la probabilidad de sobrevivir dado que se es hombre es de {:.2f}% ".format(P_A_B))
```

la probabilidad de sobrevivir dado que se es hombre es de 0.19%

Probabilidad de sobrevivir dado que se es mujer y está en la clase 1 (usar regla de Bayes compuesta)

```
[21]: sobrevivio= data["SOBREVIVE"]
sexo= data["SEXO"]
clase= data["CLASE"]
P_B_A=(data[(sobrevivio==1) & (sexo=="MUJER") & (clase==1)]["SEXO"].count())/data[sobrevivio==1]["SOBREVIVE"].count()
total=data["SOBREVIVE"].count()

probabilidadSobrevivir=(data[sobrevivio==1 ]["SOBREVIVE"].count())/total
probabilidadMujerClase1=data[(sexo=="MUJER") & (clase==1)]["SEXO"].count()/total

P_A_B= (P_B_A* probabilidadSobrevivir)/probabilidadMujerClase1

print("la probabilidad de sobrevivir dado que se es mujer y está en la clase 1 es de {:.2f}% ".format(P_A_B))
```

la probabilidad de sobrevivir dado que se es mujer y está en la clase 1 es de 0.97%

CONCLUSIONES:

La tragedia con el titanic es un evento que impacto al mundo (aun hoy en día), ya que de un total de 891 personas se obtiene que solo 342 personas sobrevivieron en este trágico evento, esto indica que solo el 38% sobrevivió, una cifra francamente baja.

De este total de sobrevivientes 233 fueron mujeres y 109 fueron hombres. Esto nos dice que el 68% de los que sobrevivieron fueron mujeres y un 32% fueron hombres.

Esto indica que claramente las mujeres fueron la prioridad al momento de buscar salvar una vida.

También obtuvo que, de un total de 342 sobrevivientes, el 259 eran mayores de 18 años, mientras que solo 73 personas eran menores de 18 años y solo 28 sobrevivientes eran mayores de 50 años.

Mostrando así que el 29% de los sobrevivientes fueron niños y adultos de la tercera edad, indicando que la prioridad para la salvar las vidas fue hacia los adultos (mayores de 18 años, pero menores a 50 años).

En cuanto a la tasa de supervivencia se dividieron por clases.

Tenemos que para la clase 1 la tasa de supervivencia es del 63%, para la clase 2 es de 47% mientras que para la clase 3 es de 24%.

Esto muestra claramente que la prioridad al momento de salvar una vida fueron los de la clase 1 después la clase 2 y por ultimo la clase 3. Se observa un gran descenso en cuanto a la tasa de supervivencia de la clase 1 a la 3.

Pero si se analiza más a fondo tenemos que la tasa de supervivencia por clase para las mujeres muestra datos interesantes.

La tasa de supervivencia de la clase 1 siendo mujer es del 97%, para la clase 2 siendo mujer es del 47% y para la clase 3 es del 50%. Mostrando así que es casi indiferente el sobrevivir si eres hombre y mujer para la clase 2 y 3, pero si se es mujer de la clase 1 tiene un fuerte peso para sobrevivir.

Para calcular la probabilidad que se tiene de sobrevivir siendo hombre o mujer se hace uso de la regla de bayes la cual es un principio fundamental en teoría de la probabilidad y estadística que permite actualizar nuestras creencias o estimaciones iniciales en función de nueva evidencia o información observada.

Con la regla de bayes se obtiene que la posibilidad de sobrevivir para las mujeres es del 74% mientras que para los hombres es del 19%.

Esto indica claramente que si se es mujer se tiene una gran posibilidad de sobrevivir (Respecto a los hombres) en un evento catastrófico como fue el del titanic.

Pero la probabilidad de sobrevivir siendo mujer aumenta muy considerablemente cuando se está en la primera clase, ya que su posibilidad de sobrevivir es del 97%.

Con esto se concluye que las mujeres fueron la prioridad al momento de salvar una vida y si ocurre un evento similar a este las mujeres de la clase 1 tienen una alta probabilidad de sobrevivir, son las que mayor probabilidad de sobrevivir tienen respecto a todos los demás pasajeros.