

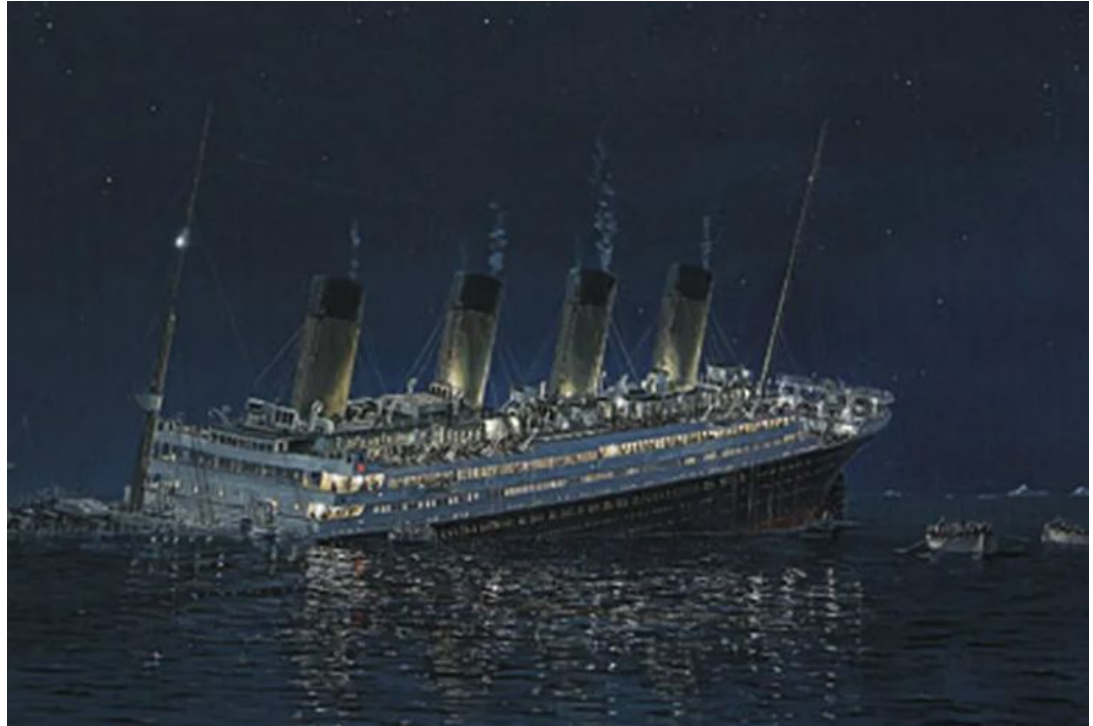
Titanic



Introduccion

El RMS Titanic fue un transatlántico británico, el mayor barco de pasajeros del mundo al finalizar su construcción, que naufragó en las aguas del océano Atlántico durante la noche del 14 y la madrugada del 15 de abril de 1912, tras chocar con un iceberg mientras realizaba su viaje inaugural desde Southampton a Nueva York.

En el hundimiento murieron 1496 personas de las 2208 que iban a bordo, lo que convierte a esta catástrofe en uno de los mayores naufragios de la historia ocurridos en tiempos de paz.



Objetivo del trabajo

Gran parte de mis conocimientos del evento surgen a partir de la película dirigida por James Cameron en 1997. Mi objetivo fue poner a prueba varios de los preconceptos reflejado en la película al analizar que variables influenciaron la probabilidad de sobrevivir de cada pasajero a bordo.

La información utilizada se extrajo a competencia abierta dentro de la plataforma de Kaggle donde a partir del listado de pasajeros y sus características individuales se busca predecir si sobrevivió el hundimiento del Titanic.

Materiales

Los datasets provistos para el desarrollo del modelo predictivo son el Train set y el Test set.

Mientras que en el Train set viene incluida una columna que indica si el pasajero sobrevivió o no (1:Si; 0:No), permitiendo evaluación del modelo predictivo, el test set sera utilizado para aplicar dicho modelo y sus resultados subidos a la plataforma de Kaggle para su posterior evaluación.

Los datos contenidos son:

survival: sobrevivo al accidente? 1:si, 0:No

pclass: Clase en la que abordo, 1 (primera clase), 2 (segunda clase), 3 (tercera clase)

sex : Sexo de la persona

Age : Edad en años

sibsp : Número de hermanos presentes

parch : Número de padres/hijo abordo

ticket: Numero de Ticket

fare: Costo del pasaje que abono

cabin: Numero de cabina

embarked: Puerto donde embarcó

Conceptos a investigar

En el final de la película vemos como fue la evacuación y que variables podrían haber influenciado sobre la posibilidad de sobrevivir de los pasajeros. Algunas de estas eran:

- Pasajero de clase alta fueron priorizados al momento de abandonar el barco
- Las mujeres y los niños tuvieron prioridad
- Los hombres se hundieron con el barco

Hipotesis

Dentro de las hipótesis a probar en el presente trabajo son:

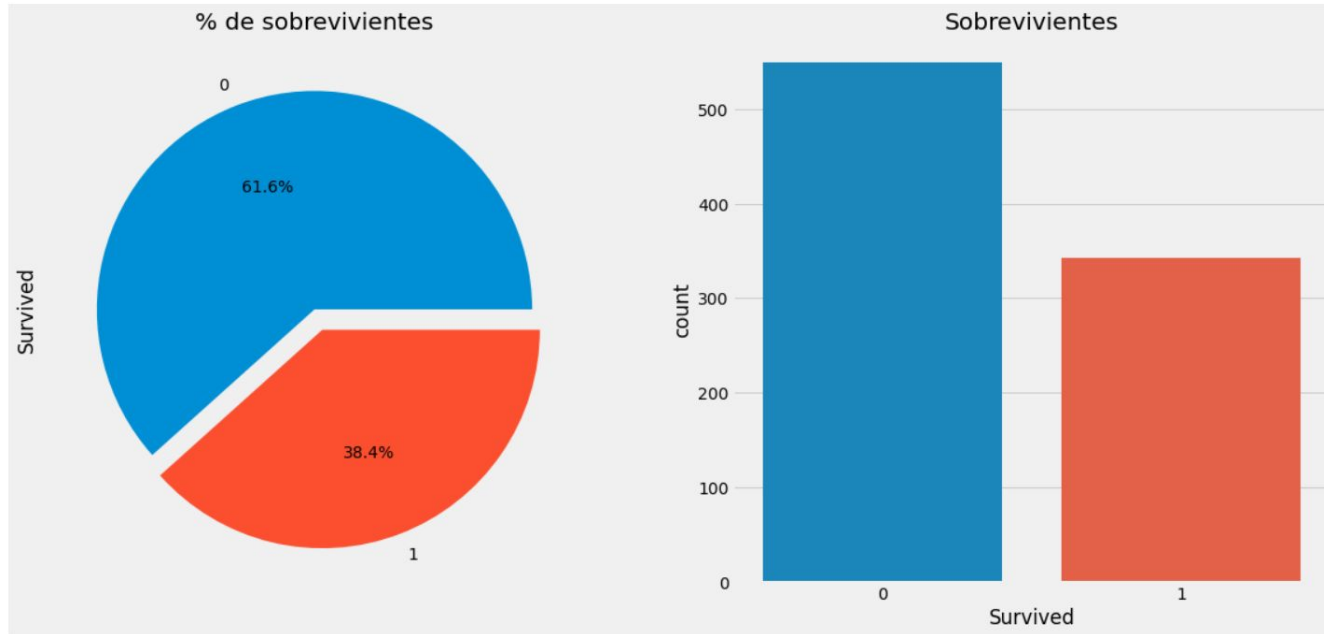
- Las mujeres adineradas tuvieron la mayor tasa de supervivencia.
- Los pasajeros más jóvenes, como niños menores de 15, tuvieron mayor probabilidad de sobrevivir.
- Los hombres más pobres tuvieron la menor probabilidad de sobrevivir
- Los hombres de primera clase tuvieron mayor probabilidad de sobrevivir

Estructura del Dataset:

]:

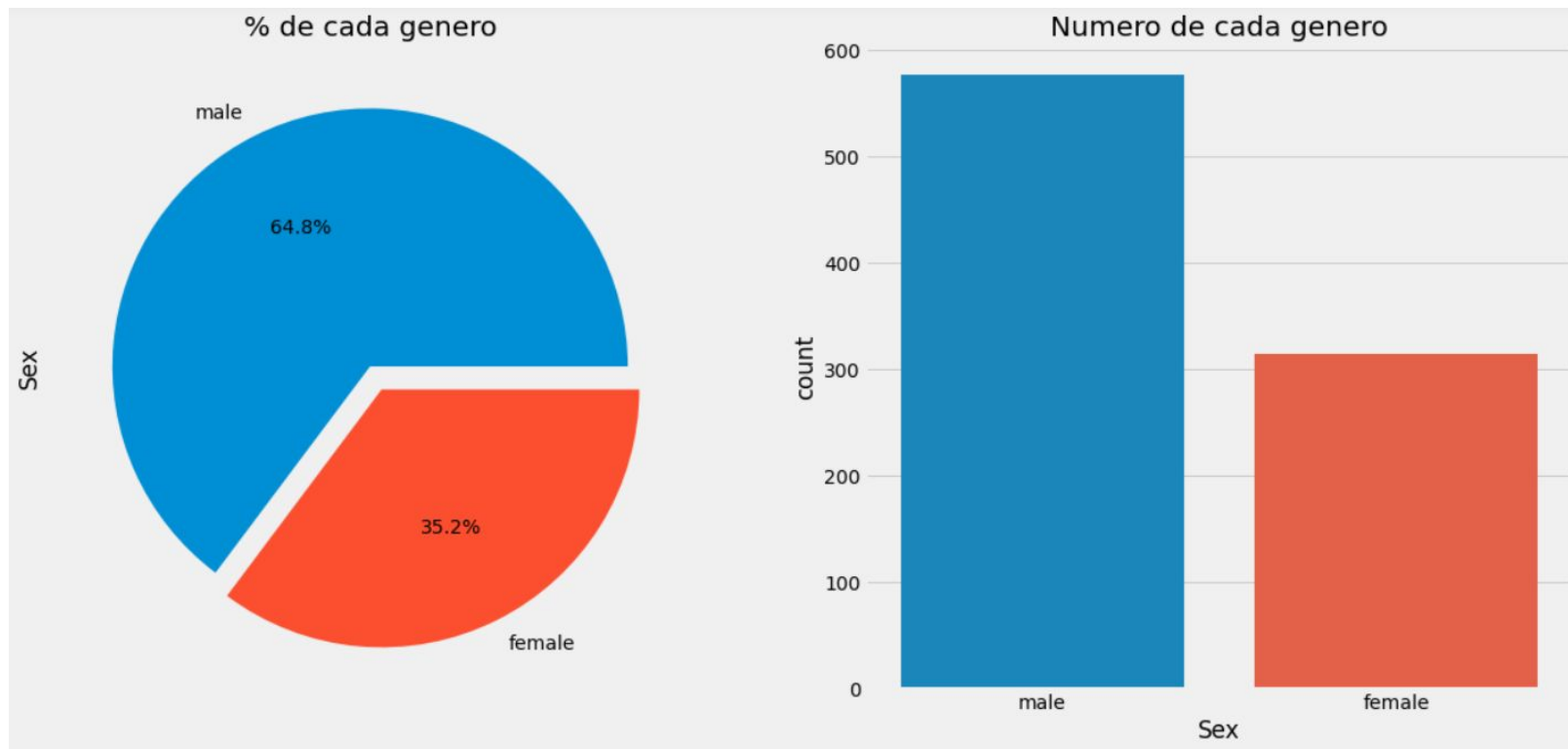
| PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked | |
|-------------|----------|--------|------|---|--------|-------|-------|--------|------------------|---------|----------|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN | S |

Cuántas personas sobrevivieron en el Train Set



Esto podría deberse a que hay más mujeres o más personas a bordo de primera clase en este dataset.

Hombres Vs Mujeres



Opuesto a lo propuesto por la filmina anteriore, en este dataset hay mas hombres que mujeres.

Contexto:

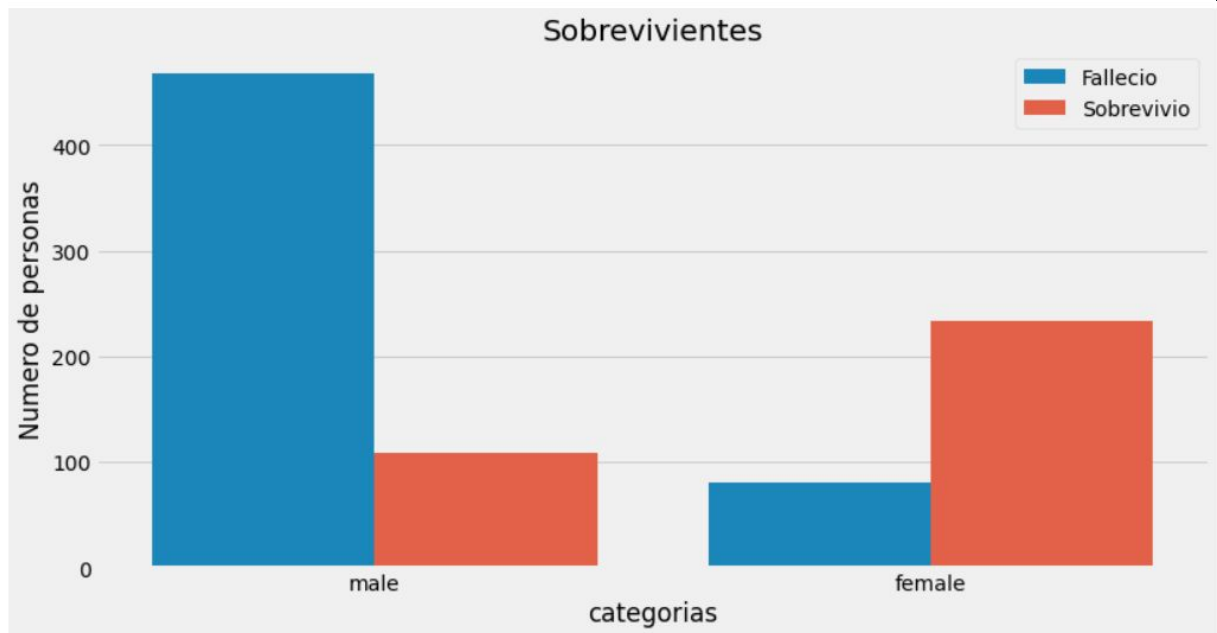
Un número muy elevado de hombres perecieron debido al estricto protocolo de salvamento que se siguió en el proceso de evacuación, conocido como “Las mujeres y los niños primero”.

A las 2:17 del 15 de abril, el barco se partió en dos y se hundió con cientos de personas todavía a bordo. La mayoría de los que quedaron flotando en la superficie fallecieron de hipotermia, aunque algunos consiguieron ser rescatados por los botes salvavidas.

Para evaluar el impacto de este protocolo se hicieron las siguientes visualizaciones:

Mujeres y niños primero!

```
Sex    Survived
female 0         81
       1        233
male   0        468
       1        109
Name: Survived, dtype: int64
```



Pese a haber mas hombres en el dataset, la tasa de supervivencia de ese grupo fue mucho menor al de las mujeres lo que refuerza lo plasmado en la pelicula donde las mujeres tuvieron prioridad al abordar los botes salvavidas.

Evacuacion:

A las 00:25, se ordenó embarcar primero a las mujeres y los niños en los botes salvavidas. Al mismo tiempo, la orquesta comenzó a tocar aparentemente en la parte delantera de la cubierta de botes a las órdenes del capitán Smith, quien quería así evitar el pánico.

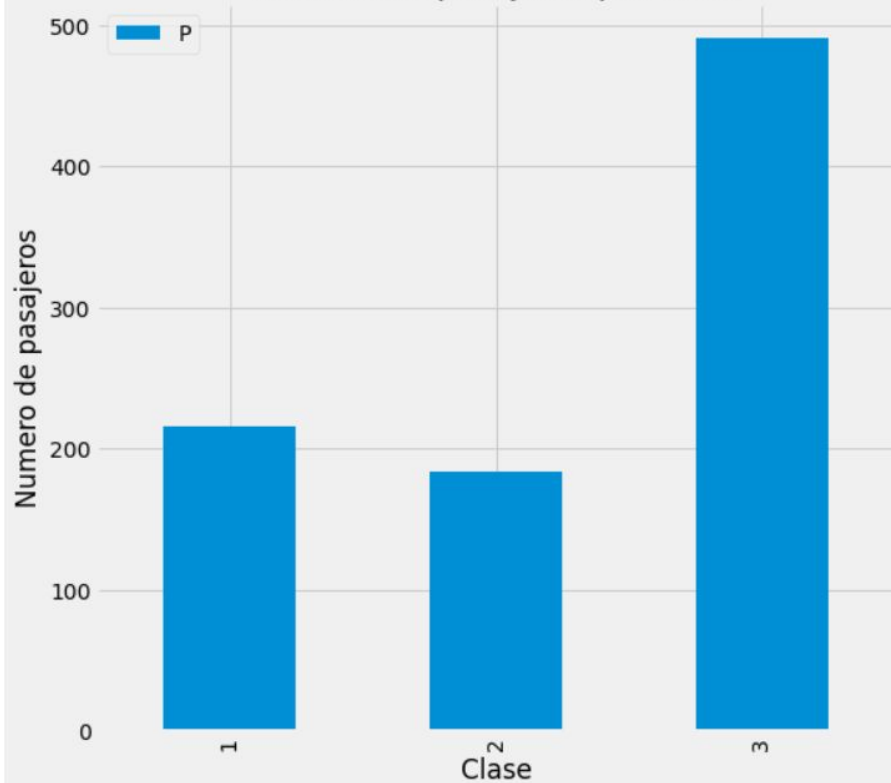
A pesar de esto, la evacuación se hizo muy lenta, ya que la mayoría de los pasajeros rechazaban el embarque en los botes, por no creer en la realidad del hundimiento. La primera clase tuvo obviamente una ventaja para conseguir un asiento en los botes.

El Titanic disponía de 20 botes salvavidas para un total de 1178 personas contra las 2208 que viajaban en el buque.

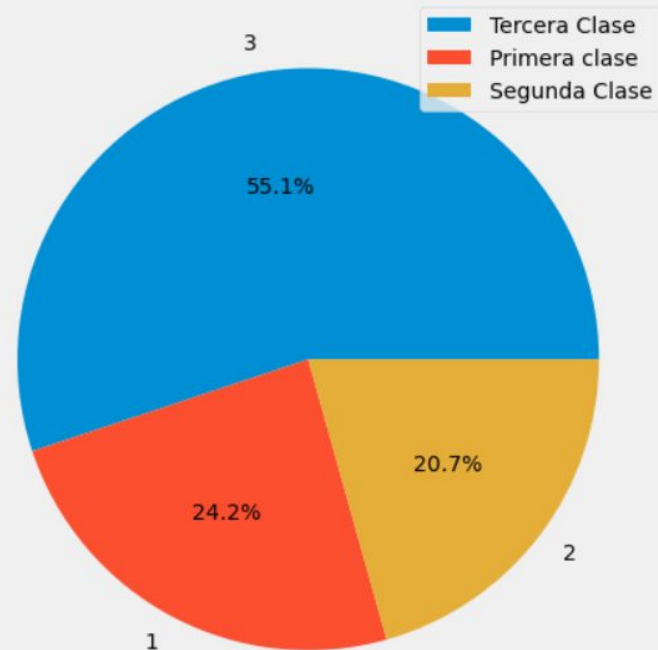
Categoría del pasajero

En la película puede verse cómo los pasajeros de primera clase tuvieron prioridad al abordar los botes salvavidas mientras que los de tercera clase no se les permitía subir a cubierta para evacuar.

Numero de pasajeros por clase

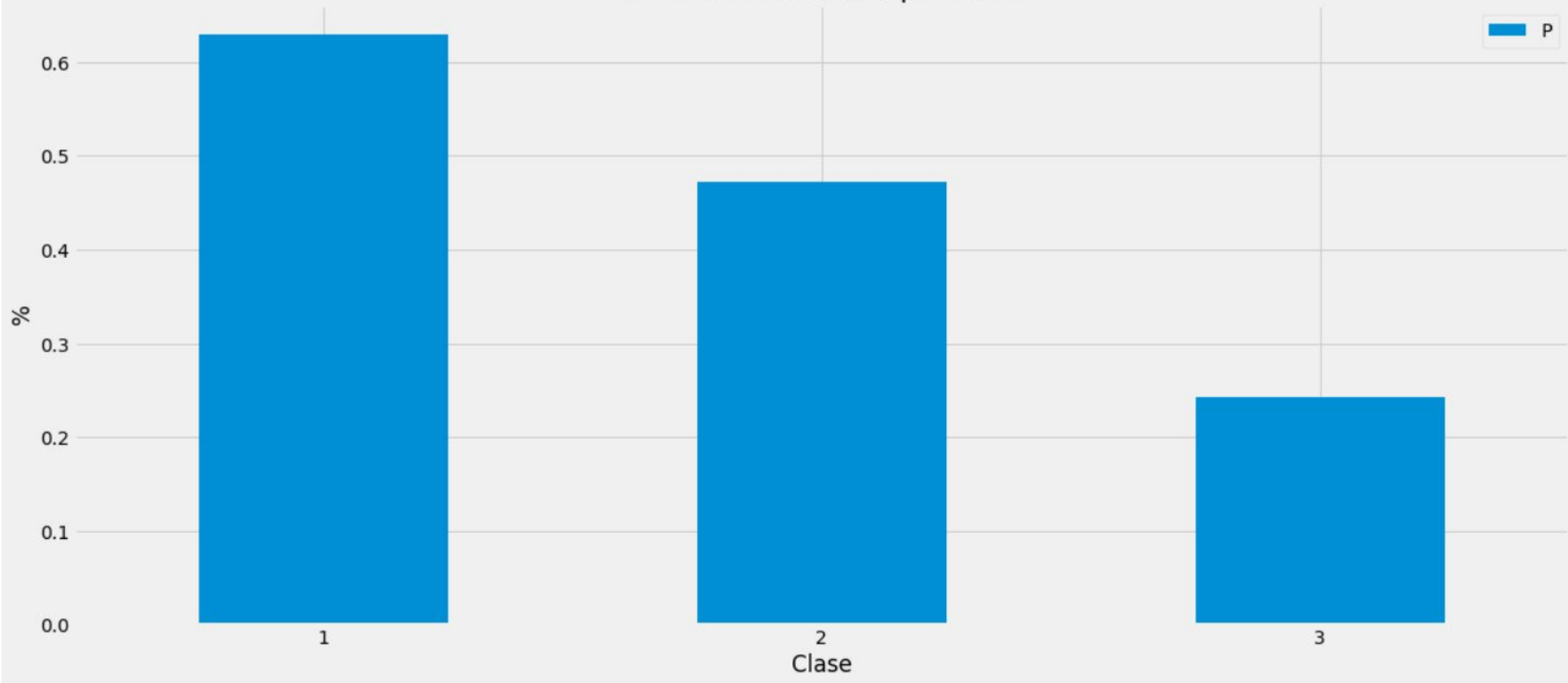


% de cada Clase

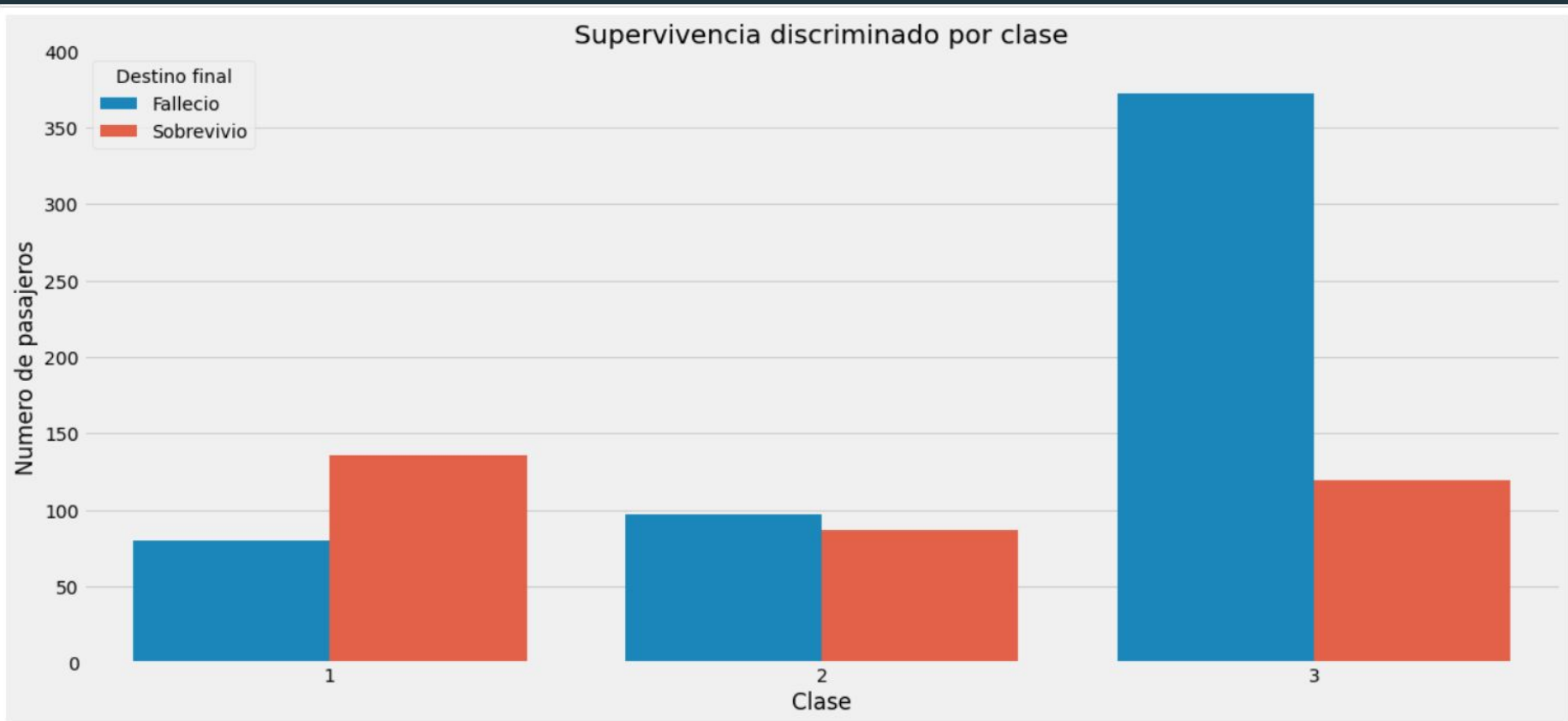


La mayor parte de los 2200 pasajeros a bordo al momento de naufragio viajaban en la tercera clase

% de sobrevivientes por clase

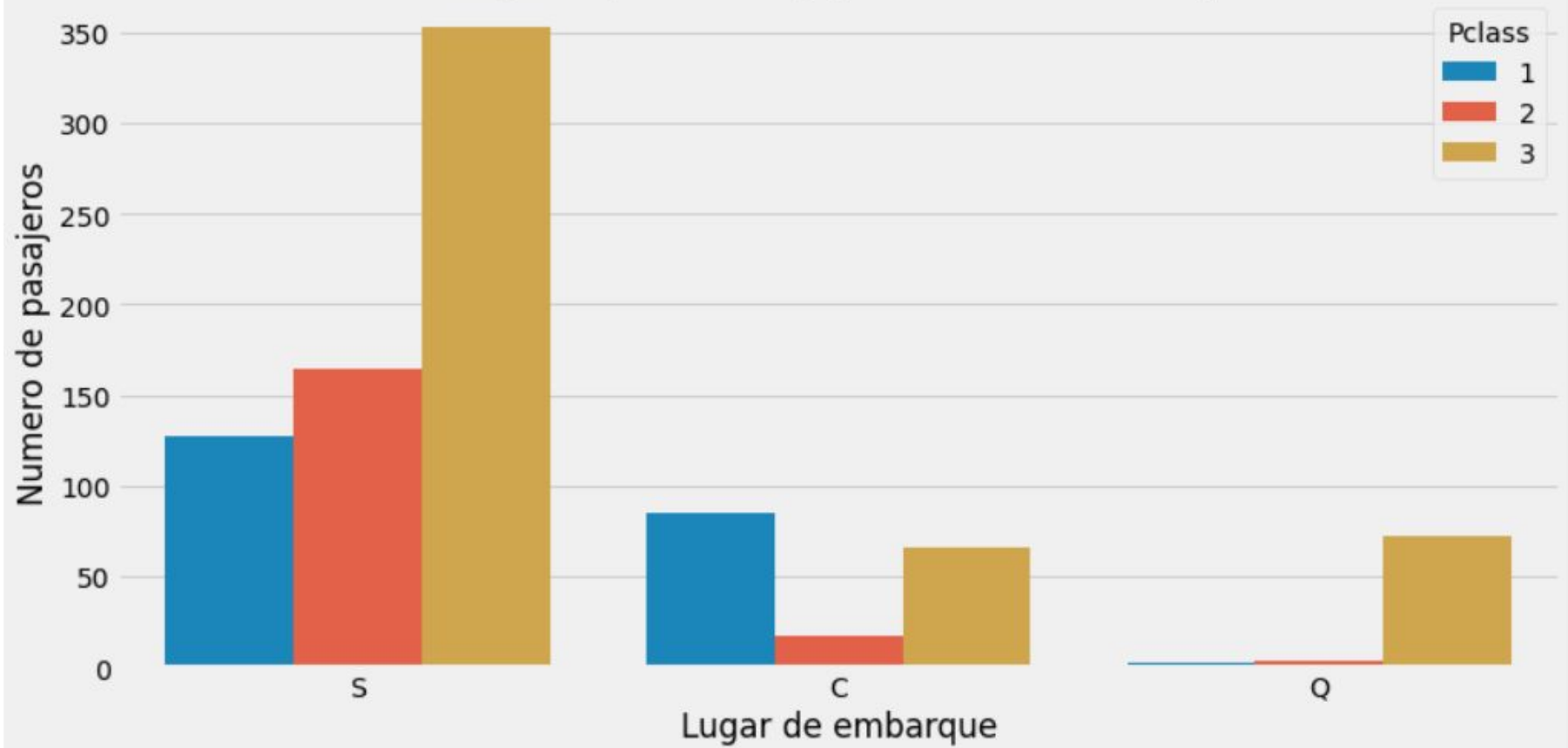


Sin embargo, pese a ser la segunda minoría, los pasajeros de la primera clase tuvieron mayores probabilidades de sobrevivir, mientras que los de la tercera la menor.



Paralelamente, las personas a bordo de la tercera clase tuvieron la menor probabilidad de sobrevivir. Existe una alta posibilidad que la mayor parte de esos pasajeros haya sido hombres jovenes sin hijos.

Pasajeros por clase y por sitio de embarque



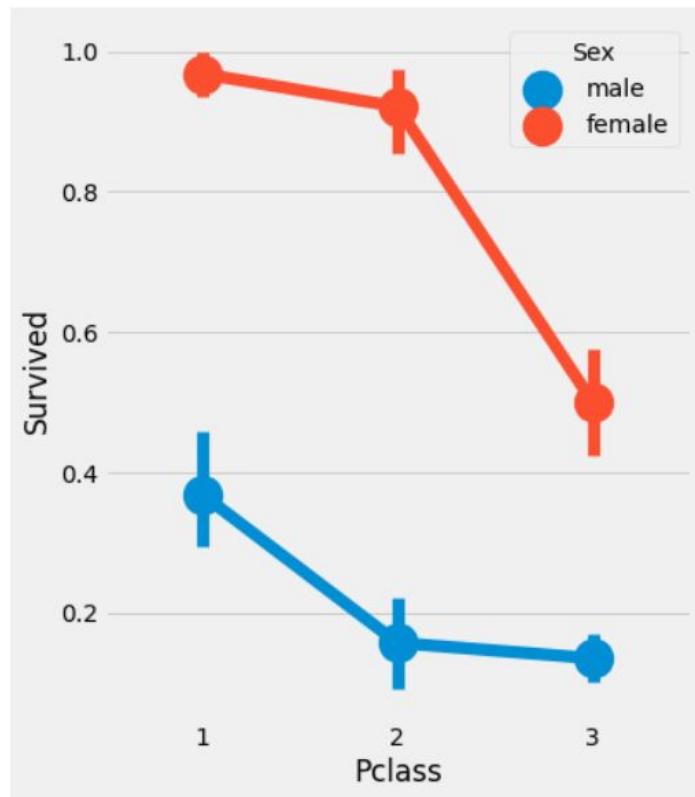
Conclusiones parciales

- Pese a haber más pasajeros en la tercera clase, su tasa de supervivencia fue menor a la de la primera clase.
- Pese a haber más hombres que mujeres, la tasa de supervivencia de este último grupo fue mayor.
- Basándonos en ambas observaciones, se puede especular que las mujeres de primera clase tuvieron la mayor chance de sobrevivir que cualquier otro grupo a bordo.

Resumen de las observaciones

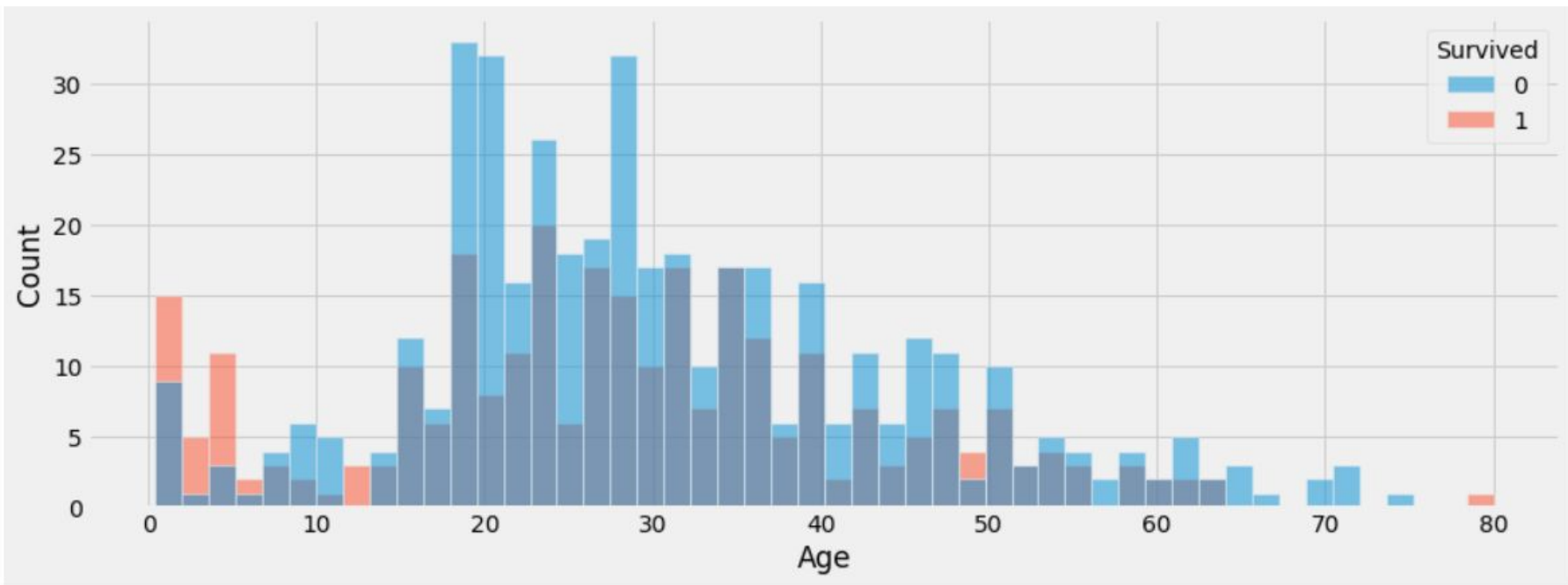
En contexto:

- Los hombres de tercera clase tuvieron la menor tasa de supervivencia, seguido por los hombres de segunda clase, los hombres de primera clase.
- Paralelamente, el grupo poblacional con más probabilidad de sobrevivir fueron las mujeres de primer, segunda y tercer clase en orden decreciente.



“Las mujeres y los niños primero”

Se evalúa si la edad del pasajero influyó sobre su tasa de supervivencia.



Como puede verse en el gráfico gran parte de las personas a bordo tenían entre 20 y 30 años, sin embargo, este grupo etario parece tener la menor tasa de supervivencia.

Paralelamente los pasajeros más jóvenes tuvieron la mayor tasa de supervivencia lo que implica que fueron priorizados en la evacuación.

| | |
|--------------|-----|
| Mr | 517 |
| Miss | 182 |
| Mrs | 125 |
| Master | 40 |
| Dr | 7 |
| Rev | 6 |
| Col | 2 |
| Mlle | 2 |
| Major | 2 |
| Ms | 1 |
| Don | 1 |
| Jonkheer | 1 |
| Capt | 1 |
| Lady | 1 |
| Mme | 1 |
| Sir | 1 |
| the Countess | 1 |

Name: Titulo, dtype: int64

```
df.groupby(["Titulo"])["Age"].mean()
```

| Titulo | |
|--------------|-----------|
| Capt | 70.000000 |
| Col | 58.000000 |
| Don | 40.000000 |
| Dr | 42.000000 |
| Jonkheer | 38.000000 |
| Lady | 48.000000 |
| Major | 48.500000 |
| Master | 4.574167 |
| Miss | 21.773973 |
| Mlle | 24.000000 |
| Mme | 24.000000 |
| Mr | 32.368090 |
| Mrs | 35.898148 |
| Ms | 28.000000 |
| Rev | 43.166667 |
| Sir | 49.000000 |
| the Countess | 33.000000 |

Name: Age, dtype: float64

A la izquierda se muestra el número de personas con cada título y a la derecha la edad promedio de cada una de esas categorías con el objetivo de asignarle ese valor a las filas con null values en la columna de la edad.