

# UNIVERSIDAD AMERICANA

Facultad de Ingeniería y Arquitectura

---



## INTELIGENCIA DE NEGOCIOS

---

### *Exploración y Análisis*

---

**Estudiante:**

Javier Ignacio León Mejía

**Docente:**

Arlen Jeannette López

Managua, Nicaragua

13 de septiembre del 2024

# ABSTRACTO

El contexto socio histórico del hundimiento del Titanic dificulta la recopilación de información contextual verificable y valiosa en cuanto a las características que llegaron a radicar en mayor medida en la supervivencia de los pasajeros del barco. Mediante la exploración y análisis de datos relativos al trasfondo demográfico y socioeconómico de 887 pasajeros, es posible indagar conclusiones que contextualizan esta tragedia desde una nueva perspectiva, por medio de la identificación, esquematización y graficación de patrones y relaciones de diversas índoles entre las diferentes características de cada individuo, haciendo uso de herramientas de visualización que generan una variedad de diagramas en dependencia del contexto de los datos analizados, tales como diagramas de barra, caja y de dispersión. Las principales conclusiones generadas en base al previamente mencionado análisis fueron que existe una relación ligera e inversamente proporcional entre la edad de un pasajero y sus posibilidades de supervivencia, y que en el caso de los sobrevivientes un porcentaje similar de hombres y mujeres lograron superar la tragedia, mientras que en el caso de aquellos que perdieron la vida, una cantidad significativamente mayor de hombres sufrieron dicho fatídico final. De una u otra forma, los resultados obtenidos demostraron que el aspecto de mayor importancia para predecir la posibilidad de supervivencia de una persona es la clase a la que pertenecía su boleto y consecuentemente, el precio que pagó por él; los pasajeros de primera y segunda clase tenían un grado de supervivencia radicalmente alto en comparación a aquellos pertenecientes a la tercera clase.

# INTRODUCCIÓN

El hundimiento del RMS Titanic es una de las tragedias más conocidas del siglo 20; aproximadamente 1500 de los más de 2200 pasajeros a bordo del barco perdieron la vida. El trasfondo social de estas personas variaba en gran medida. El objetivo principal de la presente investigación es identificar las relaciones existentes entre la tasa de mortalidad de los pasajeros y su demográfica por medio de la existencia de patrones verificables. Generar estas conclusiones llegaría a contextualizar la

significativa pérdida de vida que significó esta tragedia, presentando una idea de la relevancia de ciertas características en cuanto a su influencia para su supervivencia.

## RECOPIACIÓN Y PREPROCESAMIENTO DE DATOS

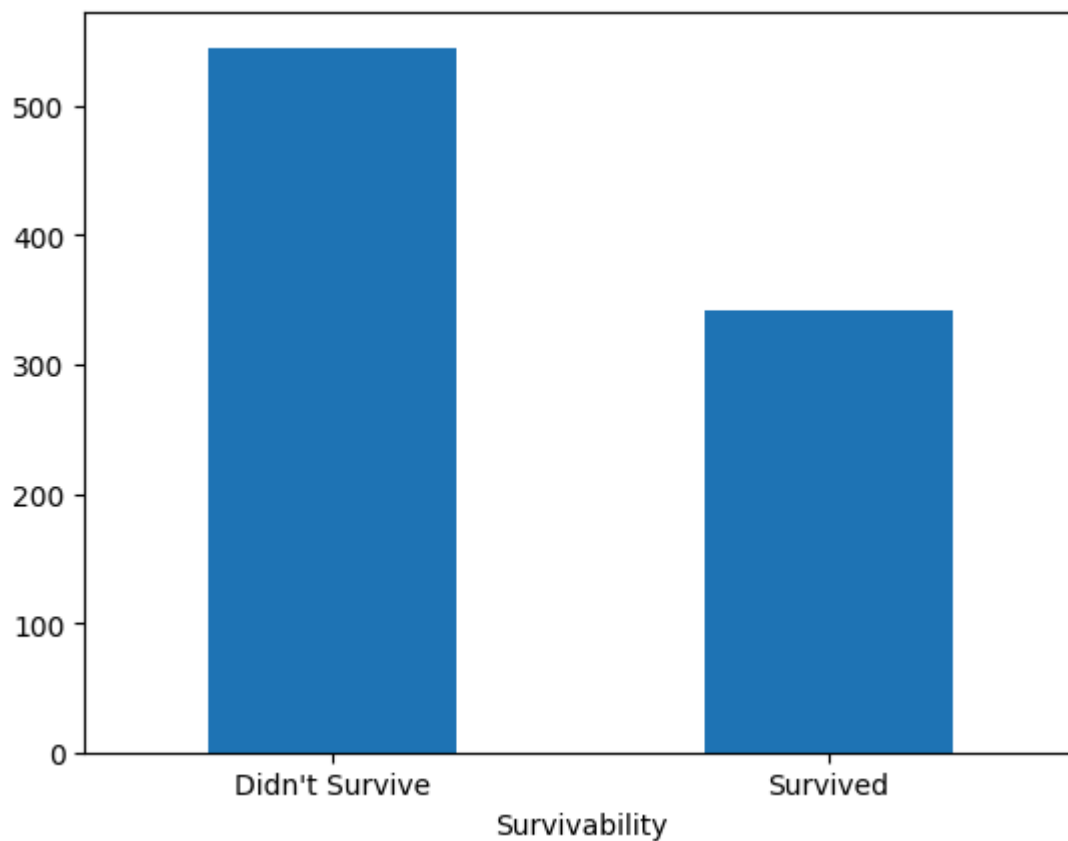
La fuente de datos del dataset del Titanic fue el repositorio de información de la Universidad de Stanford, que a su vez lo implementó en base a los estudios recopilados en el sitio web *Encyclopedia Titanica*. Para poder acceder a la información del dataset desde el cuaderno de Jupyter, se almacenó dicho archivo en un repositorio público de GitHub, cuyo enlace de datos en bruto fue cargado dentro de un Pandas Dataframe.

La transformación de los datos consistió en generar nuevas columnas contextuales que segmentan a los registros existentes en base a su supervivencia para estudiar ambos grupos por separado. También se generaron nuevas columnas para segmentar a los pasajeros en base a la clase de su boleto. Estas columnas contextuales cumplen la función de añadir información valiosa al conjunto de datos y transformar datos previamente textuales en numéricos.

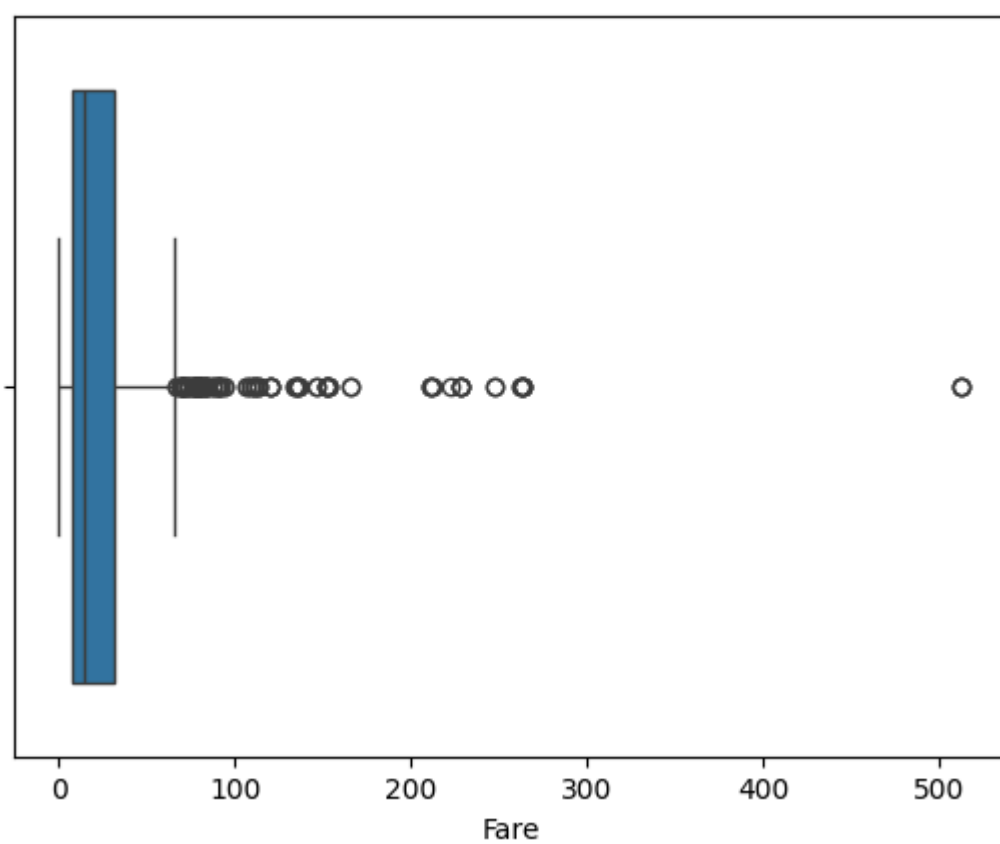
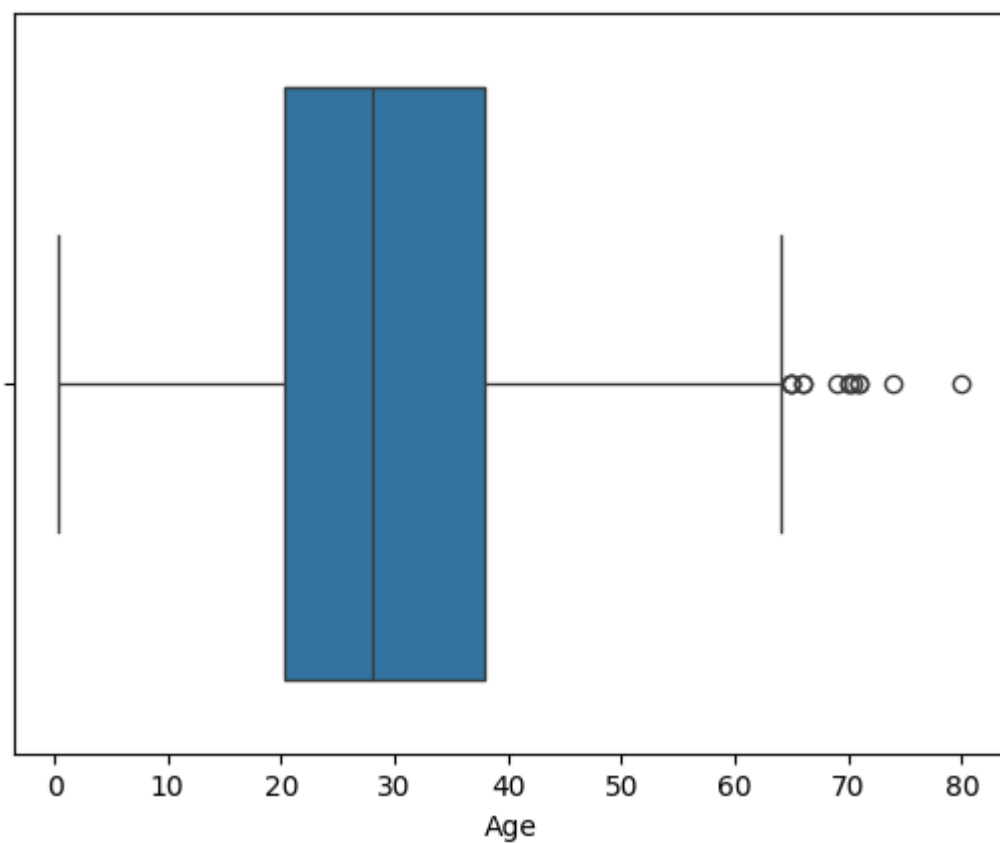
## METODOLOGÍA

Luego de visualizar la estructura general del dataset, así como las estadísticas básicas de las columnas numéricas, se comenzó por graficar con un diagrama de barras

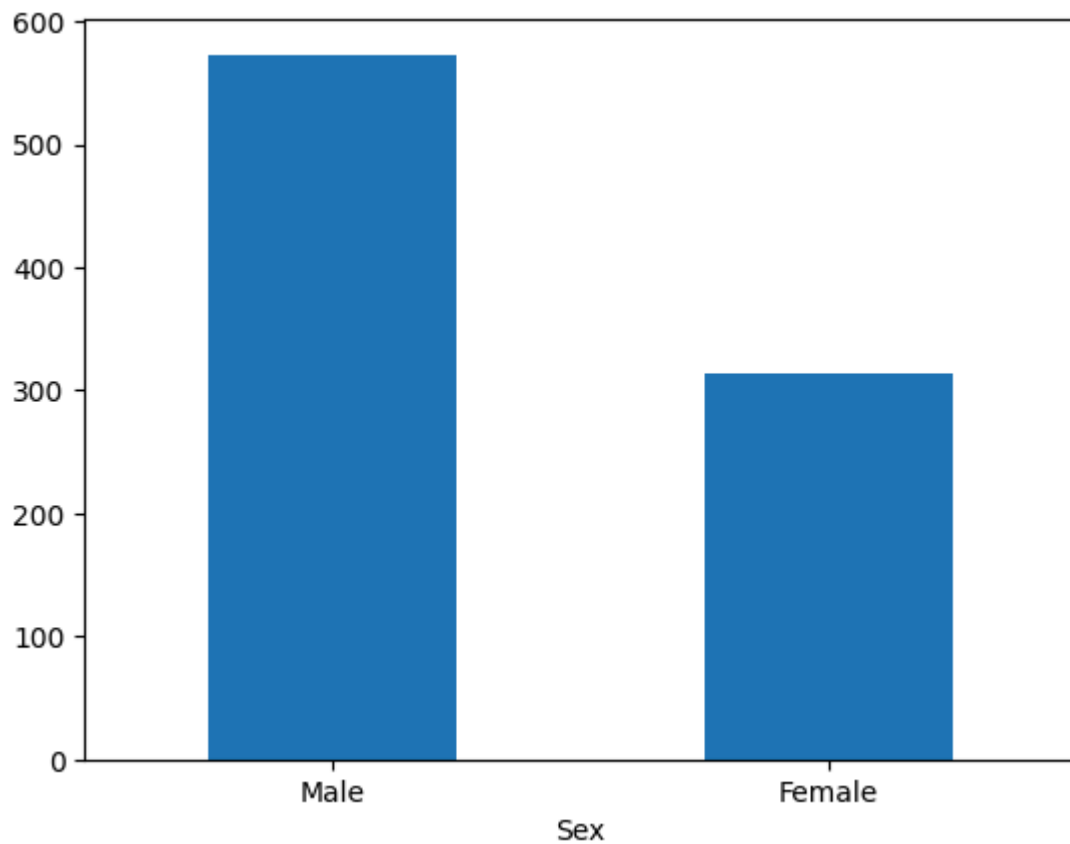
los porcentajes de supervivencia de los pasajeros.



Después se generaron boxplots para visualizar la frecuencia de la edad de los pasajeros y su pago realizado por el boleto. Se decidió implementar este tipo de diagrama por la naturaleza no categórica de estos valores.

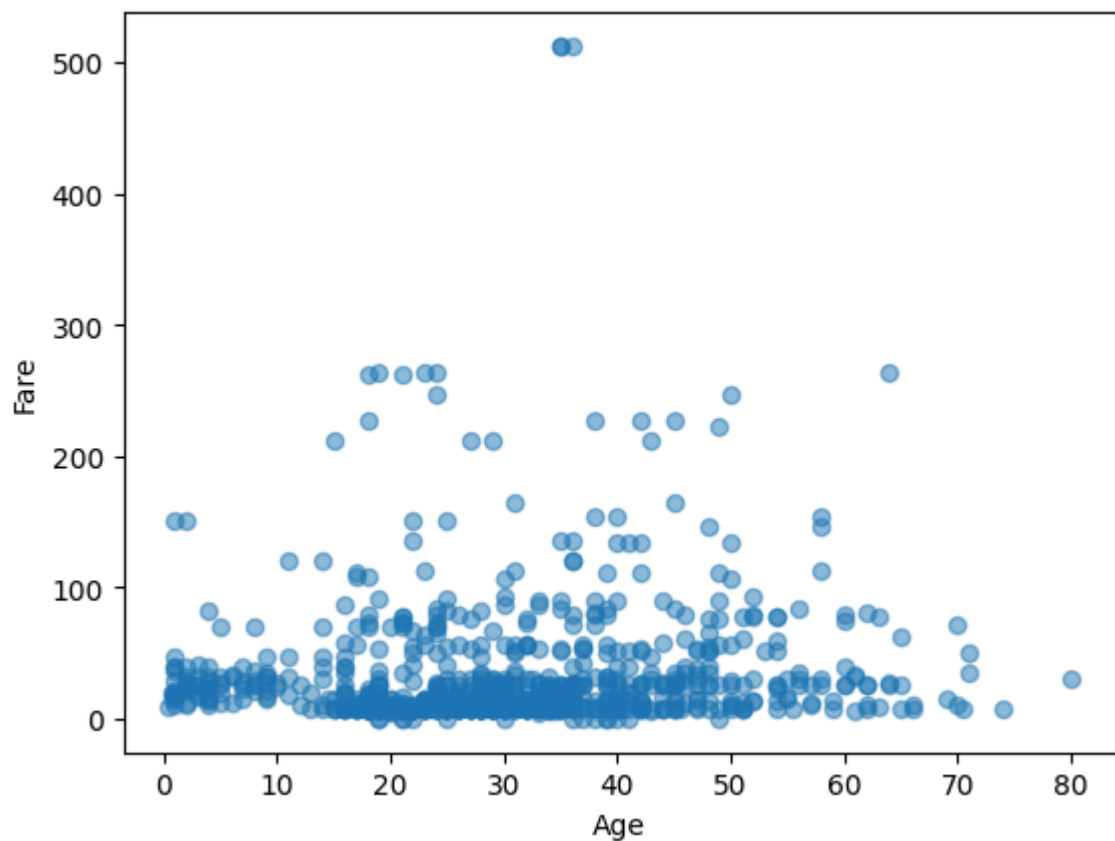


También se transformaron los datos de la columna de género en valores numéricos y se graficaron sus porcentajes.



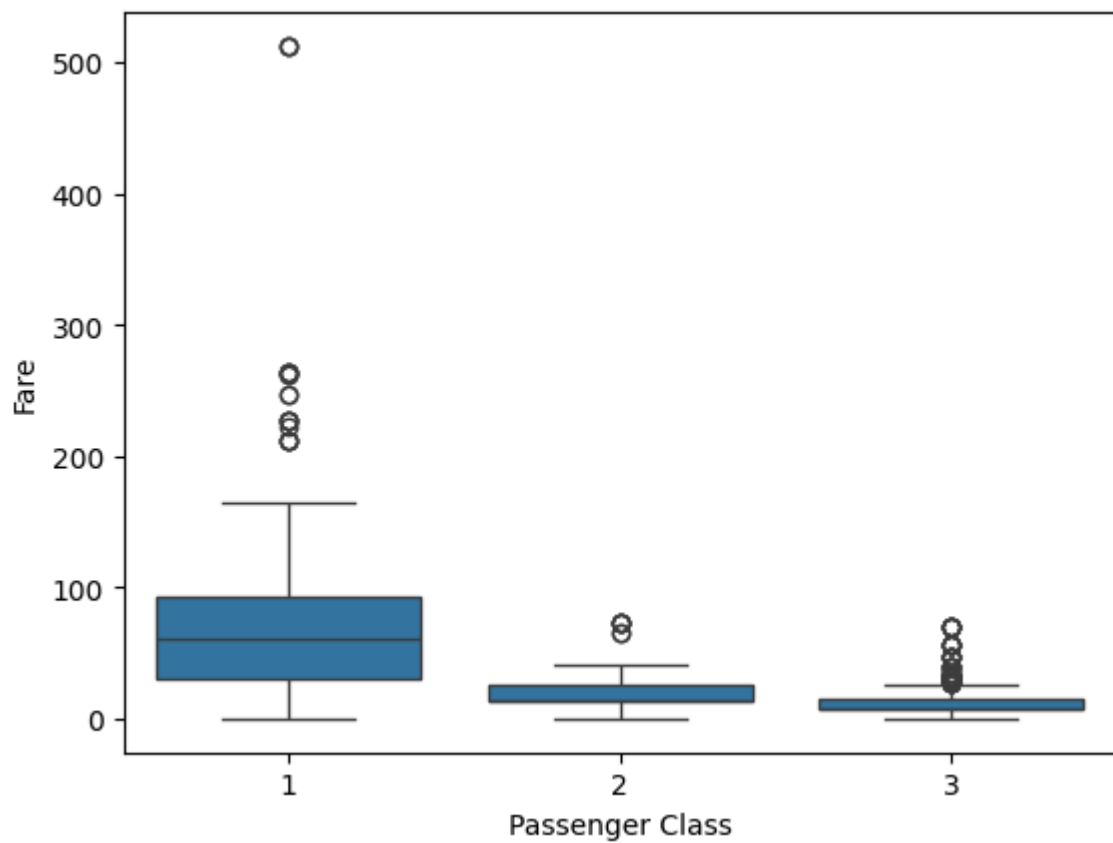
Luego de haber generado estos gráficos iniciales, se procedió a graficar las relaciones existentes entre las distintas variables del dataset, como fue el caso del siguiente diagrama de dispersión que señala la edad de los pasajeros y el costo de su boleto. Un gráfico como este puede mostrar de manera efectiva la conglomeración

segmentada de precios y grupos de personas a bordo del barco.



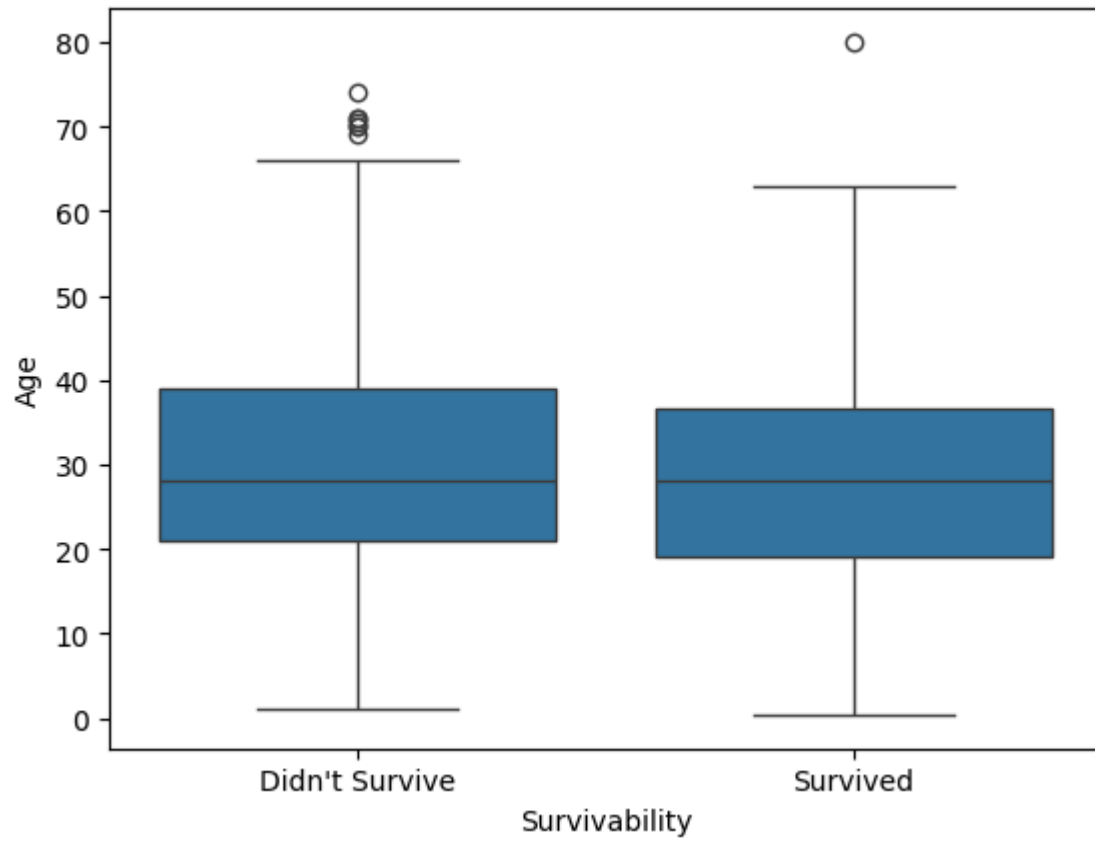
Se generó un boxplot múltiple para señalar la diferencia en precios en dependencia de la clase a la que pertenecía cada pasajero. Inicialmente se asumió que no existiría ninguna variación entre los precios de los boletos de una clase en

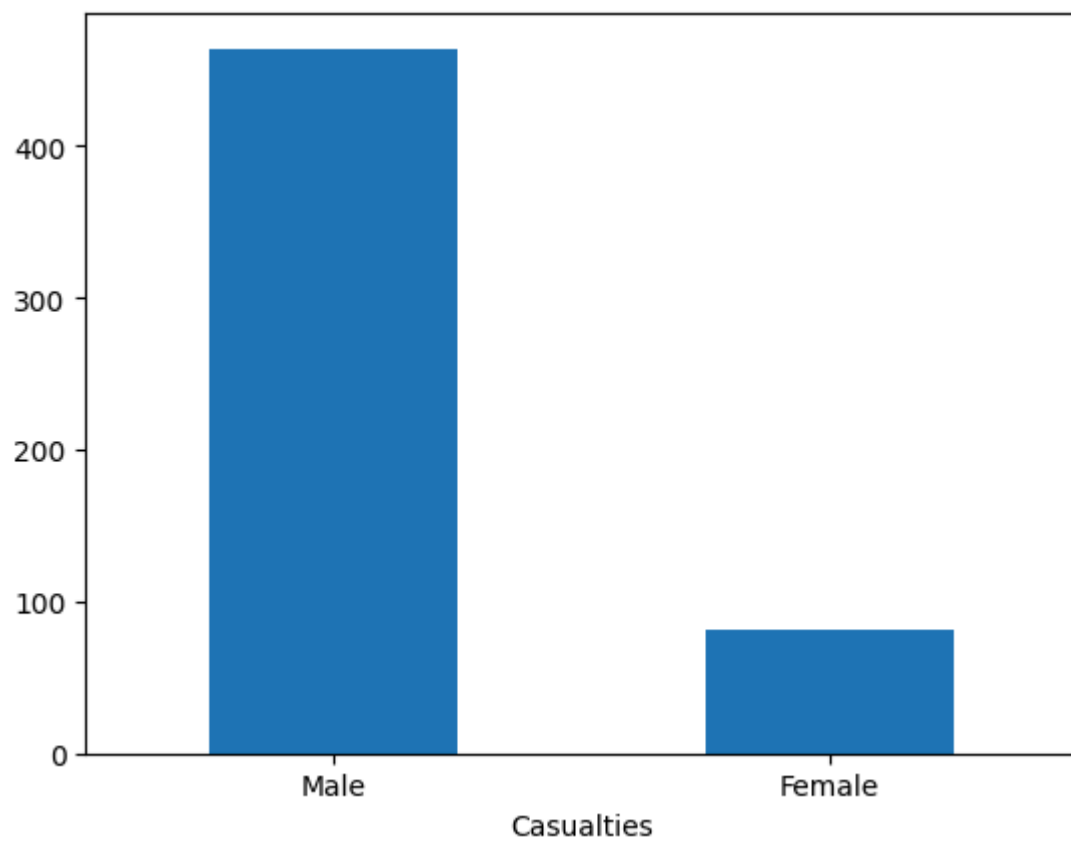
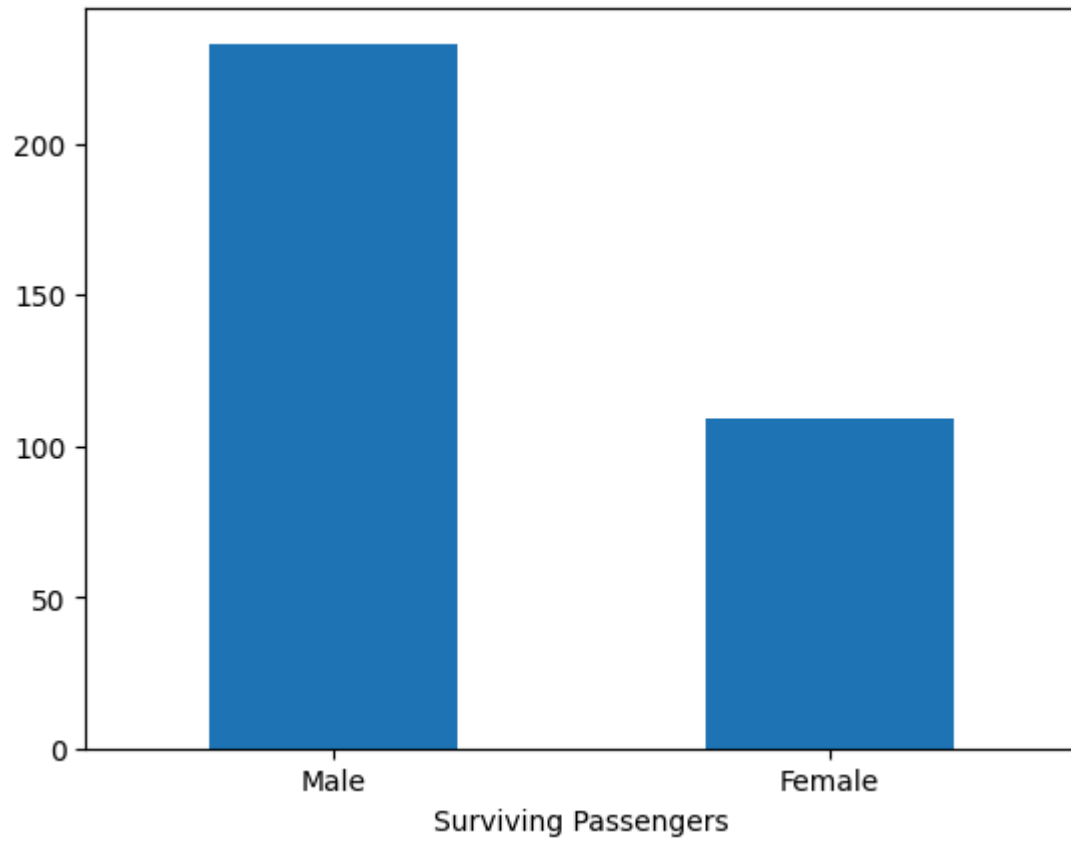
particular.



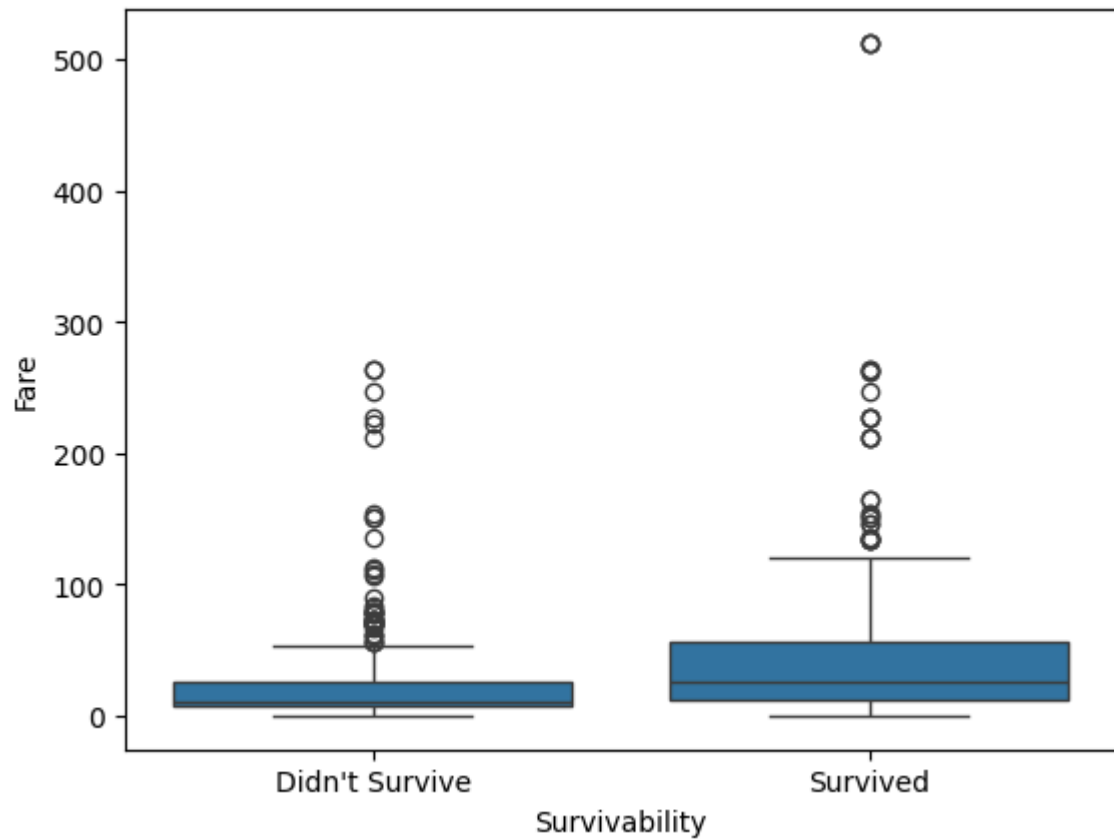
Posteriormente, se elaboraron diagramas para mostrar la relación entre la edad de un pasajero y su supervivencia, así como una cantidad total de muertes y personas que sobrevivieron según su género.



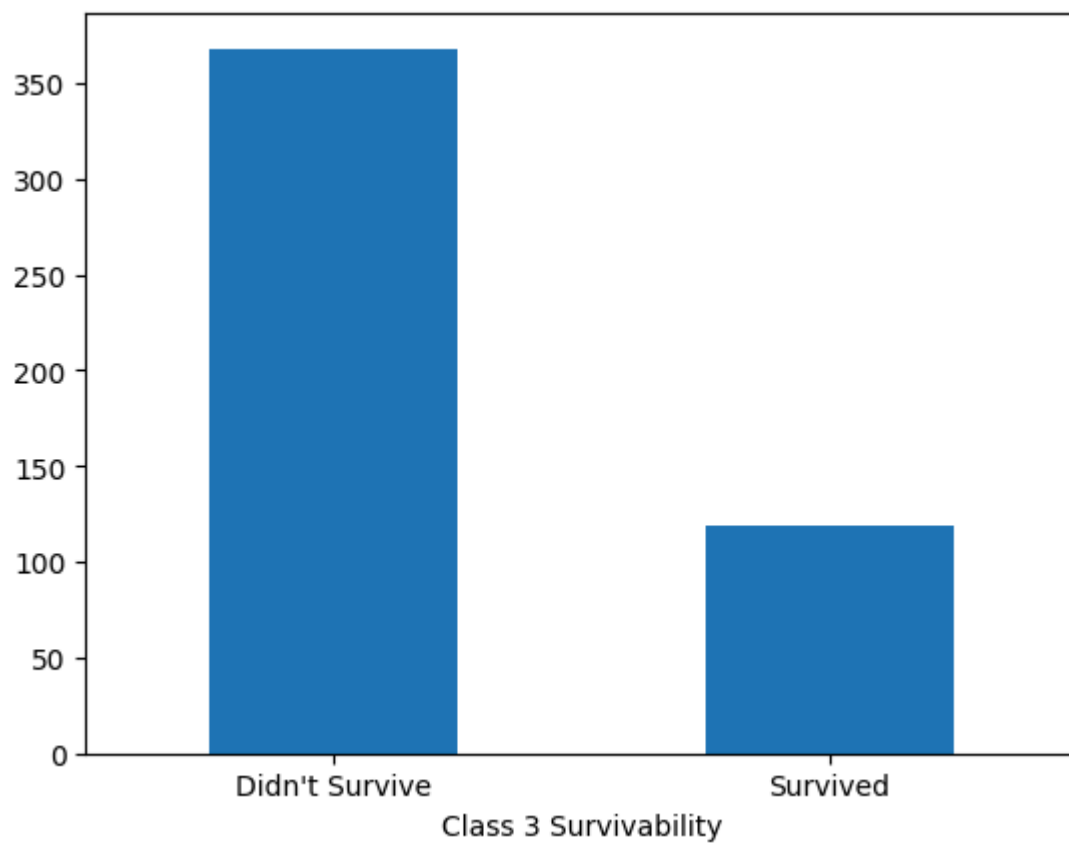


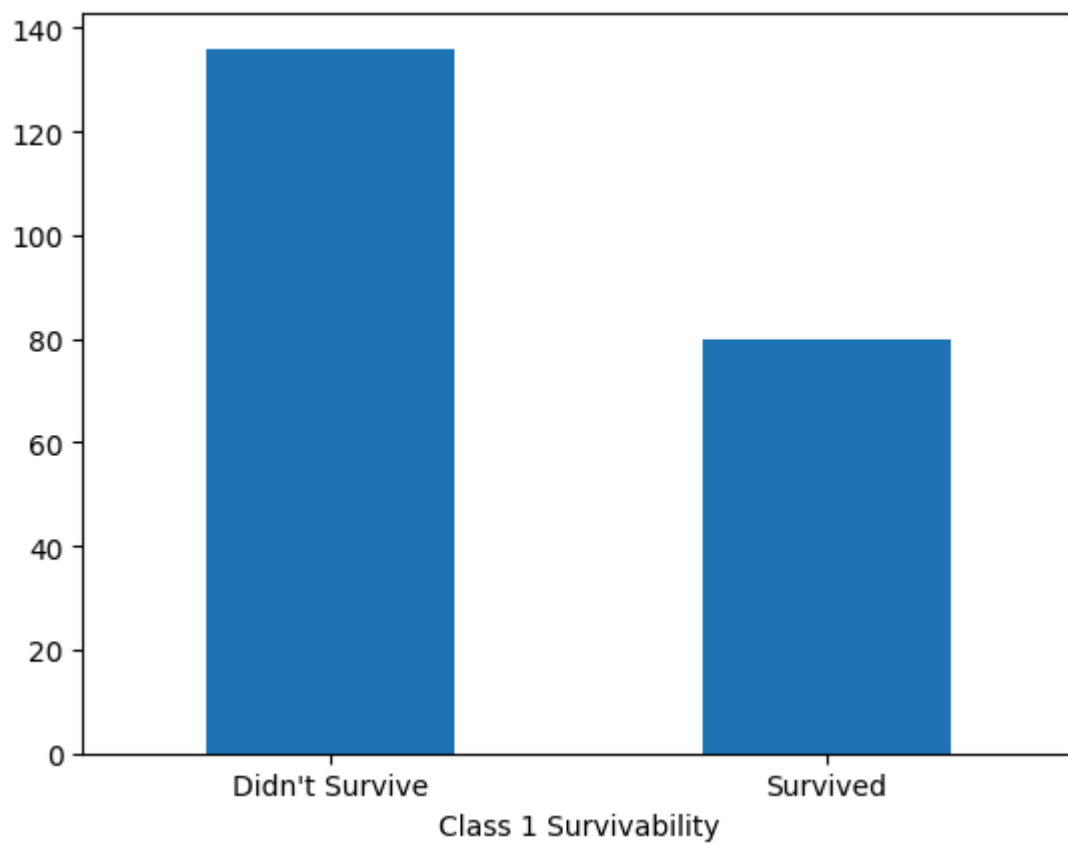
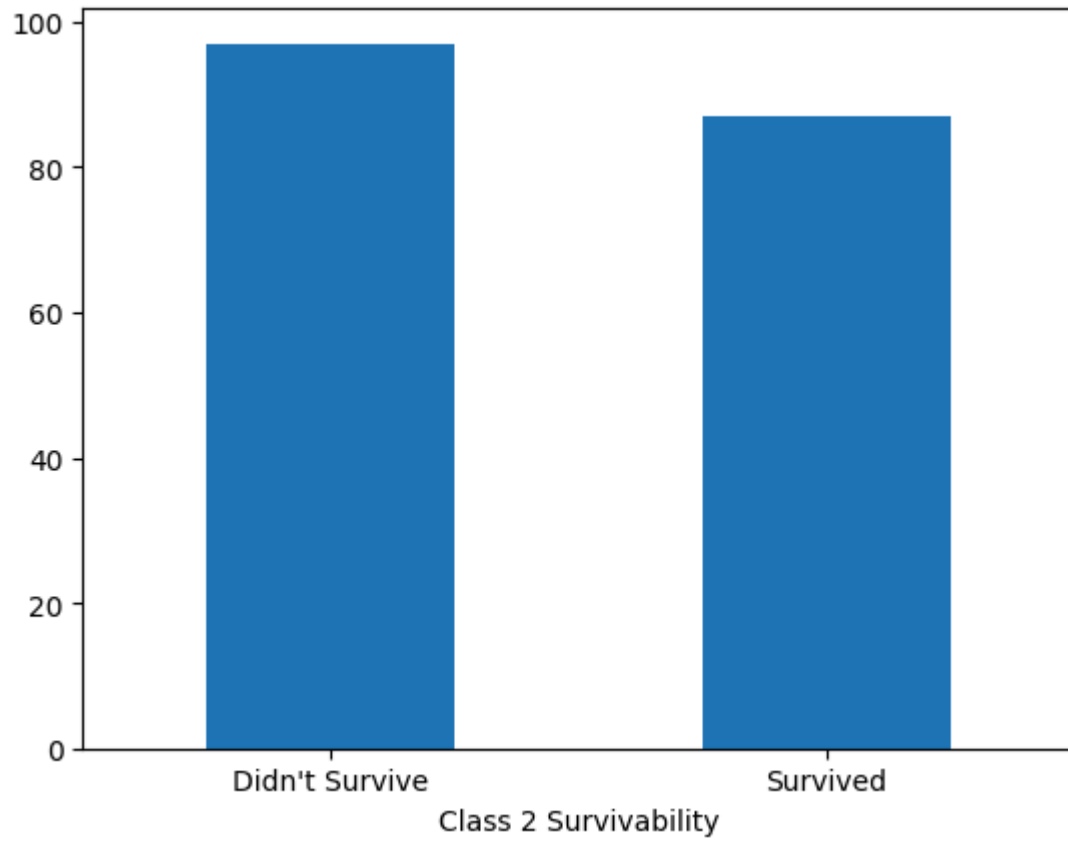


También se realizó un gráfico similar para la relación entre las muertes y el pago realizado por pasajero.



Finalmente, se realizaron 3 gráficos para visualizar la supervivencia de los pasajeros según su clase.





# RESULTADOS

- Menos del 45% de los pasajeros sobrevivieron a la tragedia
- La extrema mayoría de los pasajeros tenían entre 20 y 40 años, con solo unos cuantos pasajeros mayores a los 65 años
- Casi  $\frac{3}{4}$  de los pasajeros eran hombres
- El grueso de los pasajeros pagó menos de 80 dólares por su boleto, que serían US\$2500 aproximadamente en la actualidad
- La gran mayoría de los boletos de primera clase no difieren en gran medida del costo de aquellos de segunda clase, pero sus valores atípicos llegan a cifras monumentales
- Los sobrevivientes tienden a ser ligeramente menores en edad que aquellos que perdieron la vida; la mayoría de las personas de la tercera edad perecieron en el incidente
- El porcentaje de hombres y mujeres que sobrevivieron es proporcional a la cantidad total de hombres y mujeres en el barco, mientras que en el caso de las personas que perecieron, era mucho más posible que muriese un hombre que una mujer
- La cantidad del pago por el boleto afectaba en gran medida la oportunidad de supervivencia de los pasajeros de forma directamente proporcional
- La clase que más oportunidades de sobrevivir tuvo fue la segunda clase, seguida muy de cerca por la primera clase y por último, la tercera muy por detrás

# DISCUSIÓN

En base a los resultados generados por la investigación, es preciso afirmar que al momento de la tragedia, la tripulación de la nave realizaron ciertas distinciones con la finalidad de priorizar la supervivencia de ciertos pasajeros; se le cedió un mayor nivel prioritario a las personas del sexo femenino. Además, las personas ligeramente más jóvenes tuvieron mayores oportunidades de supervivencia.

Sin embargo, la característica que indubitadamente radicó en mayor medida en si un pasajero sobrevivía o no, era la clase a la cual pertenecía y consecuentemente, su estatus económico como podía verse ejemplificado en el precio que pagó por su boleto.

## CONCLUSIONES

Los resultados obtenidos proveen una perspectiva nueva de los valores de la sociedad inglesa a inicios del siglo 20, en base a las características valoradas en mayor medida al momento de una tragedia donde una figura de autoridad habría de decidir entre lo malo y lo peor. Las conclusiones que fueron generadas en base a la investigación fueron que existe una relación inversamente proporcional entre la edad de un pasajero y sus posibilidades de supervivencia, y que en el caso de aquellos que perdieron la vida, una cantidad significativamente mayor de hombres sufrieron dicho fatídico final. Los resultados obtenidos también demostraron que el aspecto de mayor importancia para predecir la posibilidad de supervivencia de una persona es la clase a la que pertenecía su boleto; los pasajeros de primera y segunda clase tenían un grado de supervivencia radicalmente alto en comparación a aquellos pertenecientes a la tercera clase.

## REFERENCIAS

- CS109. (n.d.).  
<https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html>