

UNIVERSIDAD AMERICANA (UAM)

FACULTAD DE INGENIERÍA Y ARQUITECTURA

Inteligencia de Negocios

Tarea U3T2

Estudiante:

Carlos Diego Toruño Espinales

Docente:

Arlen Jeannette López

Fecha:

6 de octubre de 2024



Resumen

Esta evaluación se enfoca en el análisis de factores claves determinantes para la supervivencia de los pasajeros del Titanic, este trabajo tiene especial enfoque en variables tales como el género, las distintas clases socioeconómicas, edades, entre otras. El principal objetivo fue indagar cómo características demográficas y factores externos influyeron en las estadísticas y probabilidades de la supervivencia de los pasajeros. Utilizando el conjunto de datos recopilados del Titanic se utilizaron técnicas de estadística para la descripción de los datos numéricos necesarios para el análisis a profundidad. Se calcularon tasas de supervivencia y porcentajes basados en los factores previamente mencionados. El análisis reveló un gran contraste entre las tasas de supervivencia basadas en el género; mujeres y niños presentaron una tasa de supervivencia mayor a comparación de los hombres y ancianos. También se demostró cómo pasajeros de primera clase presentaron una mejor tasa de supervivencia con respecto a los pasajeros de segunda y tercera clase, siendo esto evidencia de que los pasajeros con un mejor boleto tuvieron preferencia en áreas de evacuación con respecto a los demás. La evaluación concluye con un análisis de cómo factores sociales y estructuras familiares influyeron en las probabilidades de los pasajeros, proporcionando una visión para la mejora de estrategias equitativas y eficientes en situaciones críticas.

1. Introducción

El presente informe lleva a cabo la exploración del conjunto de datos histórico del Titanic, el cual muestra información acerca de los pasajeros a bordo, tales como género, edades, las distintas clases de pasajes y su supervivencia. El análisis exploratorio permite brindar un contexto y la identificación de patrones a partir de la comprensión de los datos, facilitando el desarrollo de conclusiones fundamentadas.

2. Revisión de Literatura

El presente análisis tiene como objetivo abordar la comprensión de los factores que influenciaron la supervivencia de los pasajeros en el Titanic. Esta revisión explora diferentes estudios basados en la aplicación de modelos predictivos y análisis estadístico. **Gupta, Sharma, y Bouza (2018)** llevaron a cabo un análisis sociológico utilizando un modelo de *machine learning*. En el estudio *Sobreviviendo a la tragedia del Titanic: Un estudio sociológico utilizando modelos de aprendizaje automático*, los autores emplearon técnicas como la regresión logística para la predicción de variables. Otro enfoque se presenta en el artículo *Ensemble Approach for Titanic Survival Predictor*, donde los autores **Gupta y Saurabh (2023)** destacaron una visión optimista del aprendizaje automático con variables enfocadas en ciencias sociales. Argumentaron la importancia de descubrir patrones en los datos para mejorar la comprensión de las dinámicas sociales. El enfoque principal del análisis es expandir lo estudiado proponiendo nuevos resultados e hipótesis sobre los factores que determinaron la supervivencia en el Titanic, contribuyendo con una perspectiva holística para una mejor comprensión del comportamiento humano.

3. Exploración Inicial del Conjunto de Datos

El primer paso a realizar es una inspección del conjunto de datos, en este se mostraron las primeras columnas para observar cómo está estructurada la información, el tipo de las variables y una breve descripción de cada una.

```
# Exploracion inicial de los datos
print(data.head())
print(data.shape)
print(data.info())
print(data.describe())
print(data.isnull().sum())
```

Figura 1: Estructura inicial del conjunto de datos

Previa a este reconocimiento de los datos, se calcularon estadísticas básicas para la descripción de las variables numéricas. Utilizamos las columnnas **Age** y **Fare**, que representan la edad de los pasajeros y precio de los boletos que cada pasajero pagó para el viaje.

```
Estadística basica de columnas numericas
count      887.000000
mean       29.471443
std        14.121908
min         0.420000
25%        20.250000
50%        28.000000
75%        38.000000
max        80.000000
Name: Age, dtype: float64
count      887.00000
mean       32.30542
std        49.78204
min         0.00000
25%         7.92500
50%        14.45420
75%        31.13750
max       512.32920
Name: Fare, dtype: float64
```

Figura 2: Estadísticas básicas de las variables numéricas

4. Metodología

Dentro de la metodología fueron utilizados métodos estadísticos y visualización de datos para una mejor comprensión de las variables socioeconómicas y demográficas.

4.1. Análisis Descriptivo

Se realizó un análisis exploratorio de los datos para identificar distribuciones, valores atípicos y patrones generales. Esto incluyó:

- Cálculo de medias, medianas y distribuciones para las variables numéricas, como la edad y la tarifa del boleto.
- Cálculo de tasas de supervivencia de los pasajeros en dependencia de los grupos.
- Uso de gráficos de barras, gráficos de dispersión y diagramas de caja para visualizar las relaciones entre las variables y las probabilidades de supervivencia.

4.2. Algoritmos y Modelos

- **Modelos de agrupación por categoría:** Se agruparon las variables categóricas, como el género, la clase y la edad, en rangos.
- **Tasas de supervivencia:** Las tasas de supervivencia se calcularon como el promedio de columnas.

Este enfoque metodológico permitió comprender claramente cómo las variables de género, clase social y edad afectaron las probabilidades de supervivencia de los pasajeros del Titanic.

5. Visualización de Datos

Posterior al análisis del conjunto de datos, se crearon variables para facilitar el análisis. Se decidió crear la cantidad total de **hombres** y **mujeres** a partir de la columna **Sex**, de igual manera se hizo una división de la cantidad de pasajeros por clase de boleto, con el fin de iniciar el análisis de los datos.

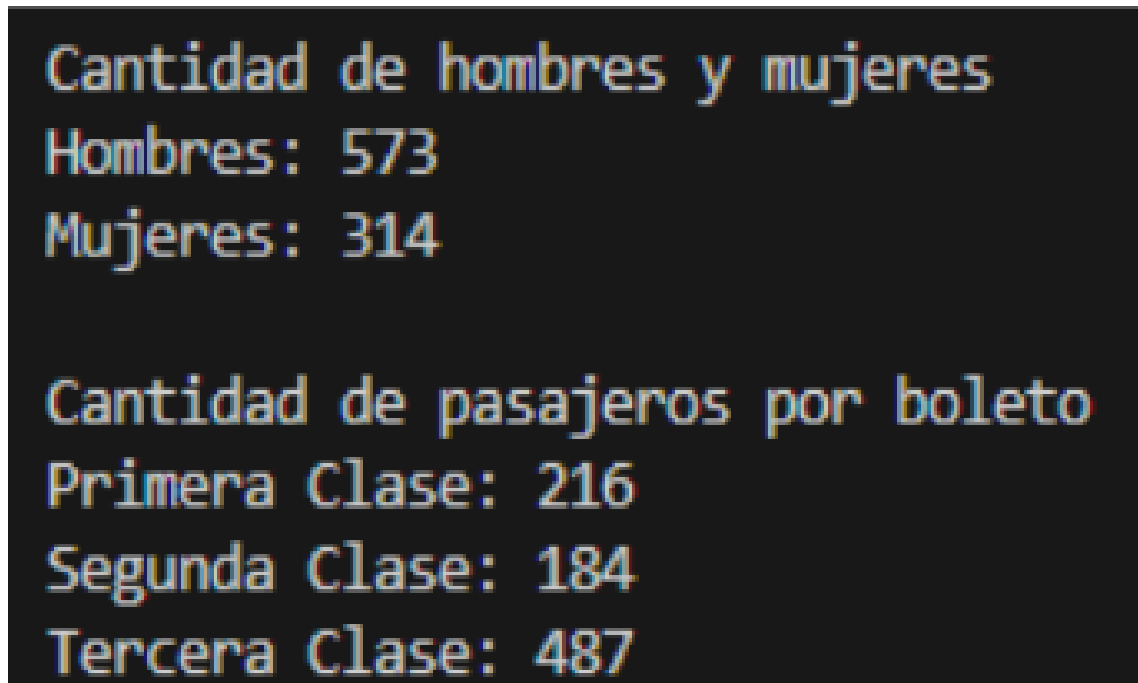


Figura 3: Distribución de pasajeros por clase y género

Para una exploración más gráfica se creó un histograma con la distribución de las edades y el número de pasajeros a bordo.

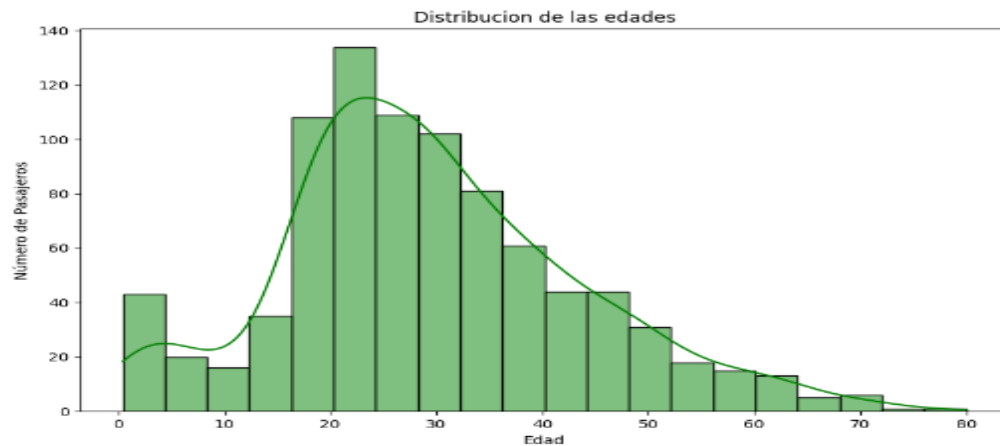


Figura 4: Histograma de distribución de edades

6. Análisis de Supervivencia

6.1. Supervivencia por Sexo

Dentro del análisis de supervivencia se muestran patrones y estadísticas significativas del conjunto de datos, donde ciertas características tales como género, edad y la clase de pasaje tuvieron un impacto en las probabilidades de sobrevivir. La primera estadística que se presentará será el porcentaje de sobrevivientes por sexo, donde observamos que las mujeres presentaron una tasa de supervivencia considerablemente mayor a la de los hombres, reflejando las políticas de evacuación de la época, posiblemente demostrando la frase "Mujeres y niños primero".

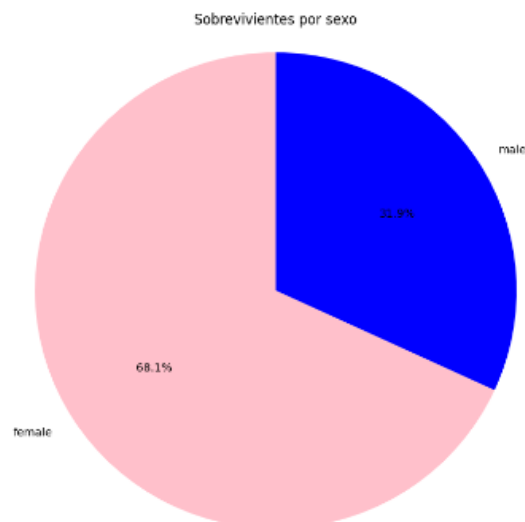


Figura 5: Supervivencia por sexo

6.2. Supervivencia por Categoría de Boleto

La siguiente estadística que se mostrará será el porcentaje de sobrevivientes por la clase de boleto que fue comprado. En este gráfico se puede ver la proporción de sobrevivientes según la clase de boleto con la que viajaban; donde los pasajeros con boletos de primera clase tuvieron una mayor probabilidad de sobrevivir. Esto indica un acceso preferencial en las áreas de evacuación en comparación a las demás clases.

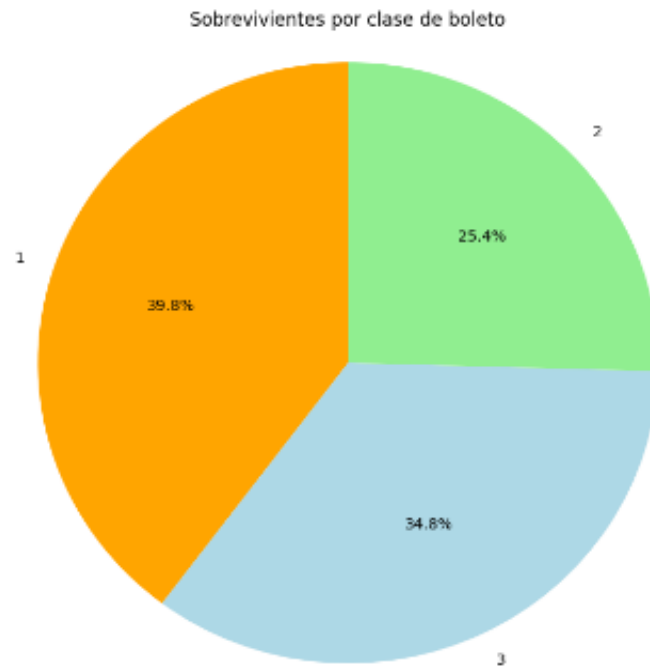


Figura 6: Supervivencia por clase de boleto

6.3. Supervivencia por Grupo de Edades

El siguiente factor a analizar será la tasa de supervivencia por edad, para facilitar el manejo de los datos se dividió los grupos de edades en 4 categorías: **Niño**, **Adolescente**, **Adulto**, y **Anciano**. Las edades por categoría están divididas de la siguiente manera:

- **Niños** → De 0 a 12 años.
- **Adolescentes** → De 12 a 18 años.
- **Adultos** → De 18 a 65 años.
- **Ancianos** → De 65 a 100 años.

Cantidad de personas por grupo de edad	
GrupoEdad	
Adulto	711
Adolescente	87
Niño	79
Anciano	10

Figura 7: Cantidad de pasajeros por grupo de edad

El gráfico muestra cómo los niños tenían una mayor probabilidad de sobrevivir, demostrando que fueron priorizados durante la evacuación, mientras que a medida que la edad aumentaba, las posibilidades de supervivencia disminuían.

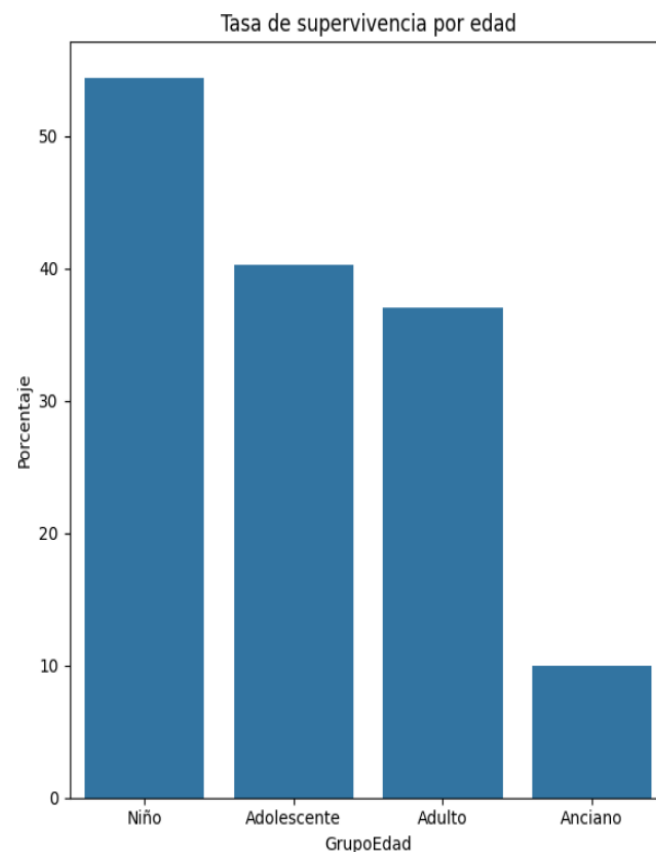


Figura 8: Tasa de supervivencia por grupo de edad

7. Discusión

7.1. Interpretación de los Resultados

El objetivo del análisis de este conjunto de datos era identificar factores, patrones y relaciones en la supervivencia de los pasajeros del Titanic, a través de técnicas estadísticas, una exploración inicial y distribuciones de gráficos que permitieron visualizar la dispersión de variables como **edades**, **géneros**, etc.

7.2. Limitaciones e Incertidumbres

A pesar de que se logró la captura exitosa de patrones relevantes para el análisis, existen ciertas limitaciones con este conjunto de datos. Un ejemplo puede ser que los datos no presentan información relevante como la salud de los pasajeros o la ubicación de estos, lo cual podría tener un impacto en la tasa de supervivencia. Además, la naturaleza impredecible del evento limita la capacidad de obtener conclusiones definitivas sobre decisiones individuales.

8. Conclusiones

8.1. Patrones Generales

El análisis mostró cómo características como el género, las clases sociales y la edad estuvieron directamente relacionadas con la probabilidad de sobrevivir, y cómo otros factores tales como las normas de evacuación y los accesos a las rutas de evacuación fueron también significativos. Además, se destacó el rol de la estructura familiar, ya que los pasajeros que viajaban con familiares mostraron una ligera ventaja en las tasas de supervivencia, probablemente debido a la organización conjunta durante el proceso de evacuación.

Referencias

- Stolz, J., Lindemann, A., & Antonietti, J. (2018). Sociological explanation and mixed methods: the example of the Titanic. *Quality & Quantity*, 53(3), 1623–1643. <https://doi.org/10.1007/s11135-018-00830-0>
- Gupta, S., & Saurabh, S. (2023). Ensemble Approach for Titanic Survival Predictor. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4663312>
- Gupta, K., Sharma, P., & Bouza, C. N. (2018, April 24). Surviving the Titanic Tragedy: A Sociological Study Using Machine Learning Models. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3417433