

UNIVERSIDAD AMERICANA

FACULTAD DE INGENIERÍA Y ARQUITECTURA

Inteligencia de negocios

Tarea U3T2

Estudiante: Carlos Diego Toruño Espinales

1 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. Se demostró de igual manera cómo pasajeros de primera clase presentaron una mejor tasa de supervivencia con respecto a los pasajeros de segunda y tercera clase, esto es 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 en la mejora de estrategias equitativas y eficientes en situaciones críticas.

Índice

1.	Introducción	3
2.	Revisión de literatura	3
3.	Exploración inicial del conjunto de datos	3
4.	Visualización de datos	4
5.	Análisis de supervivencia	5
5.1.	Supervivencia por sexo	5
5.2.	Supervivencia por categoría de boleto	6
6.	Resultados	7
6.1.	Tasa de supervivencia por género	7
6.2.	Tasa de supervivencia por clase de boleto	8
6.3.	Tasa de supervivencia por grupo de edades	8
7.	Discusión	10
7.1.	Interpretación de los resultados	10
7.2.	Limitaciones e incertidumbres	10
8.	Conclusiones	11
8.1.	Patrones Generales	11
9.	Referencias	11

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 en la supervivencia de los pasajeros en el Titanic. Esta revisión explora diferentes estudios basados en la aplicación de modelos predictivos y estadística analítica. **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 regresión logística para la predicción de variables.

Otro enfoque el artículo *Ensemble Approach for Titanic Survival Predictor* los autores **Swati Gupta, Sonal Saurab (2023)**, destacan una visión optimista del aprendizaje automatizado con variables enfocadas en ciencias sociales, argumentando la importancia de descubrir patrones en los datos para mejorar la comprensión de dinámicas sociales.

El enfoque principal del análisis es expandir lo estudiado al proponer nuevos resultados e hipótesis sobre los factores que determinaron la supervivencia en el Titanic, contribuyendo con un enfoque holístico 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 mostraremos 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())
```

Previo a este reconocimiento de los datos, calcularemos estadísticas básicas para la descripción de las variables numéricas. Utilizaremos las columnas “Age” y “Fare”, las cuales representan la edad de los pasajeros y precio de los boletos que cada pasajero pagó para el viaje.

```
Estadística básica de columnas numericas
count      887.000000
mean       29.471443
std        14.121908
min         0.420000
25%        29.250000
50%        28.000000
75%        38.000000
max        80.000000
Name: Age, dtype: float64
count      887.000000
mean       32.30542
std        49.78264
min         0.000000
25%         7.92500
50%        14.45420
75%        31.13750
max        512.32920
Name: Fare, dtype: float64
```

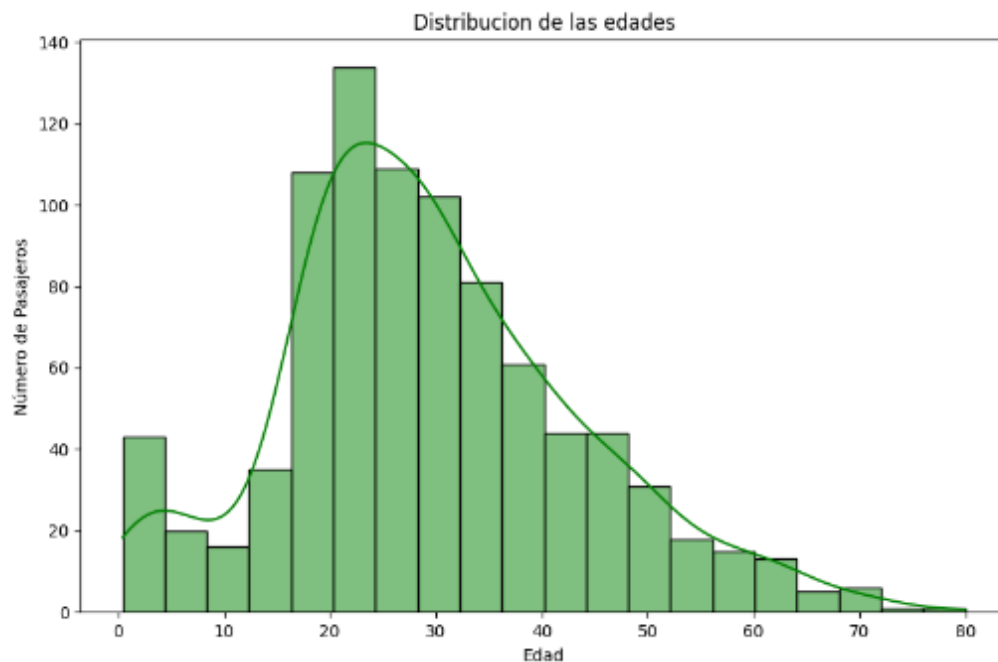
4. Visualización de datos

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

```
Cantidad de hombres y mujeres
Hombres: 573
Mujeres: 314

Cantidad de pasajeros por boleto
Primera Clase: 216
Segunda Clase: 184
Tercera Clase: 487
```

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

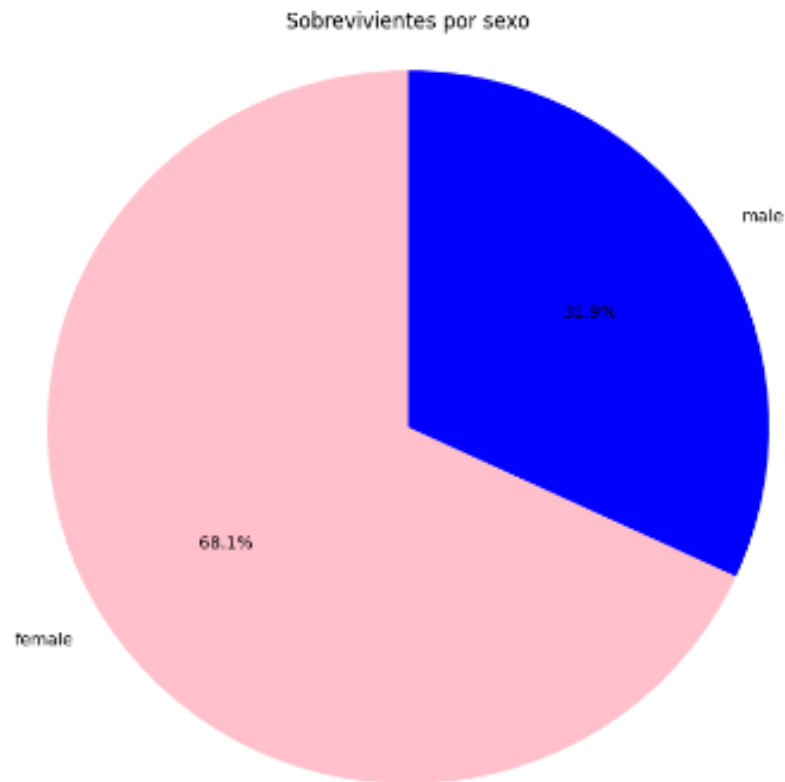


Como podemos observar en el histograma, este indica que la mayoría de los pasajeros contaba con edades comprendidas entre 20 y 30 años. También podemos observar que las personas mayores y niños eran menos frecuentes, lo cual será de **relevancia posteriormente**.

5. Análisis de supervivencia

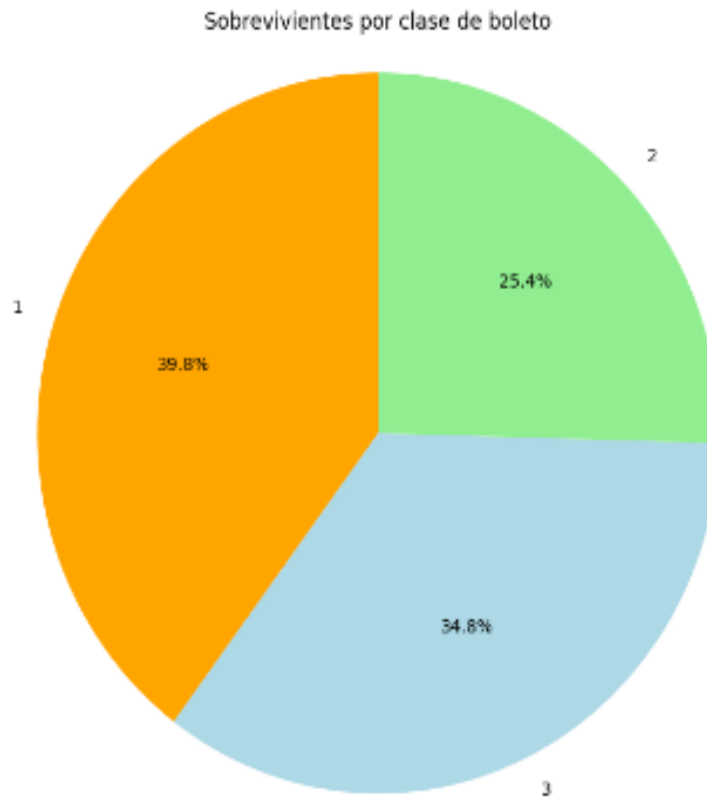
5.1. Supervivencia por sexo

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 primer 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, este análisis nos refleja como eran las políticas de evacuación de la época, posiblemente demostrando la frase “Mujeres y niños primero”.



5.2. 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 viajaba donde los pasajeros con boletos de primera clase tuvieron una mayor probabilidad de sobrevivir, con esto se puede deducir un acceso preferencial en las áreas de evacuación en comparación a las demás clases. Un dato a tomar en cuenta de este gráfico es que en esta representación no se toma en cuenta el número total de pasajeros por clase, sino la distribución de sobrevivientes en cada clase, la relación de supervivencia por el total de pasajeros será mostrada más adelante.



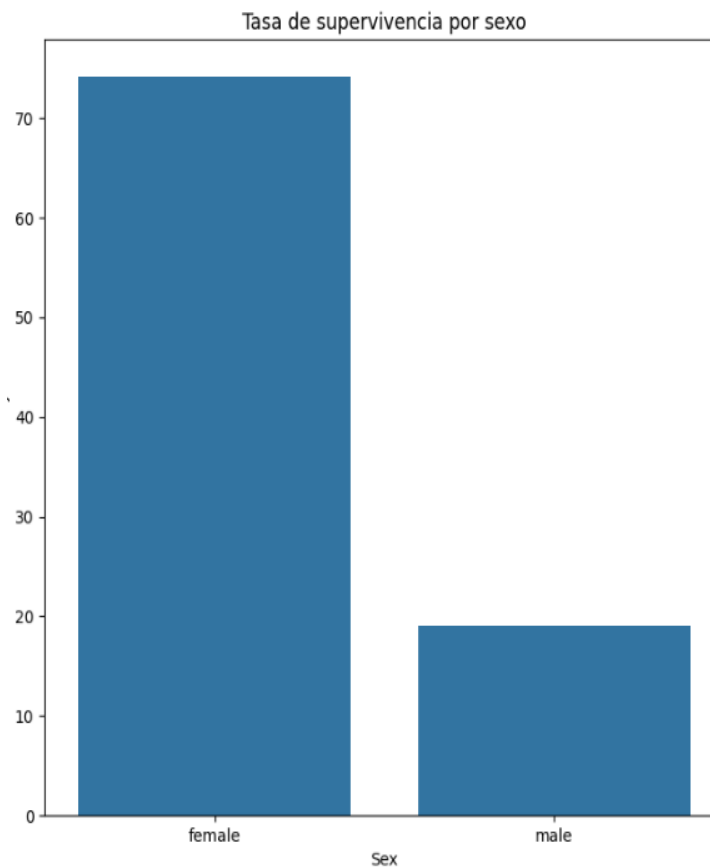
6. Resultados

Las tasas de supervivencia muestran la proporción de individuos que lograron superar una situación con respecto a un número de afectados. Como primer análisis realizaremos un gráfico para mostrar la tasa de supervivencia con respecto al sexo de los pasajeros. Esta tasa será calculada con las columnas “**Sex**” y “**Survived**” de la siguiente manera:

```
# Analisis de tasa de supervivencia por sexo
porcentaje_genero = data.groupby('Sex')['Survived'].mean() * 100
print(porcentaje_genero)
```

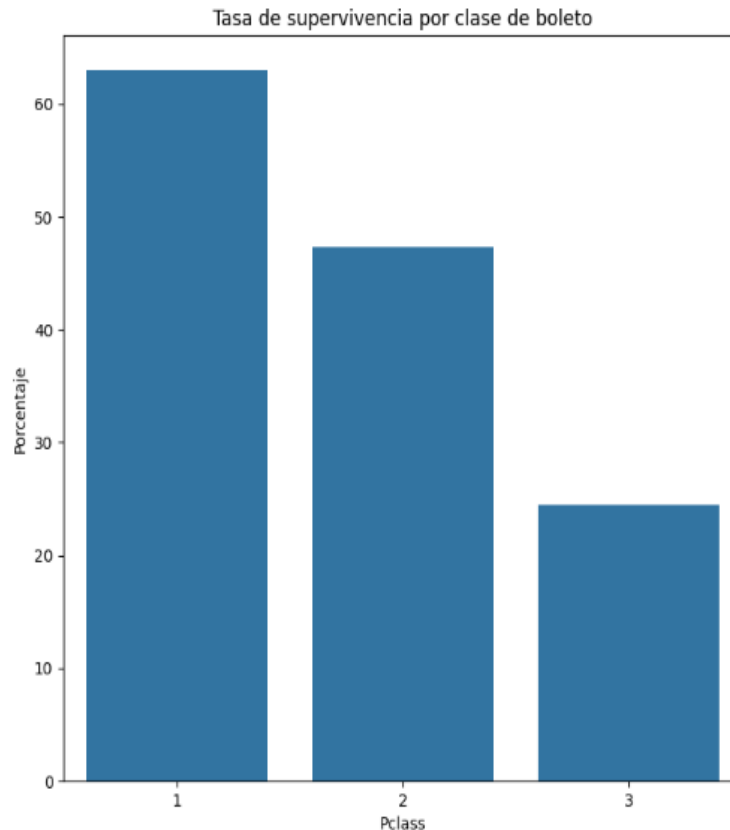
6.1. Tasa de supervivencia por género

Con el siguiente gráfico de barras se representa la tasa de supervivencia, donde se puede observar que las mujeres tienen una tasa superior al 70 %, por el contrario de los hombres con una tasa considerablemente menor de alrededor del 20 %. Este gráfico representa cómo el género fue un factor determinante para la supervivencia.



6.2. Tasa de supervivencia por clase de boleto

El siguiente factor a analizar será la tasa de supervivencia con respecto a la clase de boleto que fue comprado. Podemos observar en el gráfico como los pasajeros con un boleto de primera tenían una probabilidad considerablemente mayor con respecto a los de segunda y tercera clase, podemos asumir que otro factor que jugó un papel importante en la supervivencia fueron las clases sociales.



6.3. Tasa de supervivencia por grupo de edades

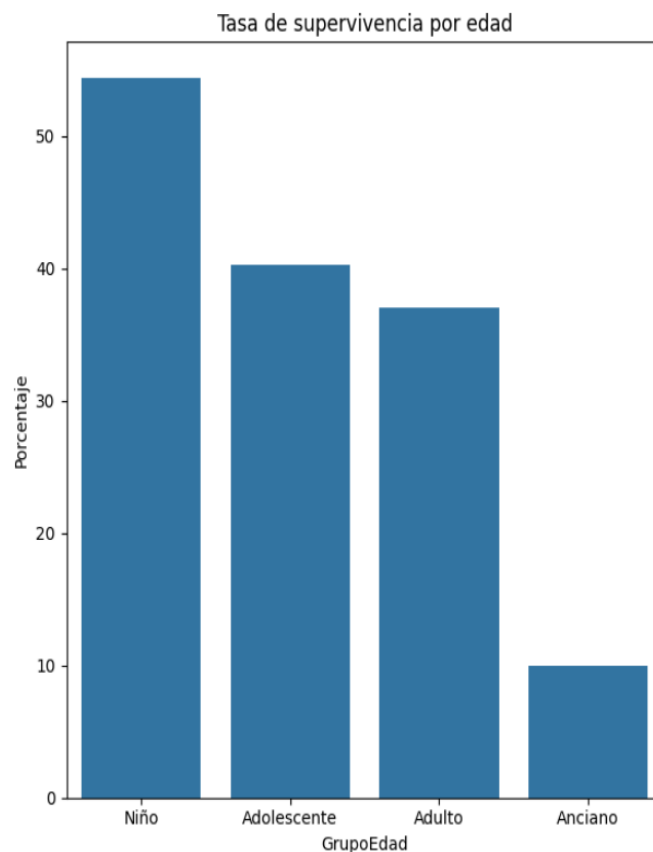
Como último factor a analizar será la tasa de supervivencia por edad, para facilitar el manejo de los datos se ha dividido los grupos de edades en 4 categorías, las cuales son: “Niño”, “Adolescente”, “Adulto”, “Anciano”. Las edades por categoría están divididas de la siguiente manera:

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

Primero observaremos la cantidad de personas que hay por grupo de edades, ya que esto nos dará una visión de la distribución de las edades. Esto nos permitirá observar la predominancia de los grupos de edades.

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

Una vez tenemos la cantidad de personas por el grupo de edad podemos observar que la mayoría de personas estaban entre los 18 y 65 años, pero esto no es un factor clave para la supervivencia, esto lo veremos a **continuación**:



El gráfico muestra cómo los niños tenían una mayor probabilidad de sobrevivir, demostrando que fueron priorizados durante la evacuación, a medida que la edad aumentaba las posibilidades de supervivencia considerablemente disminuían. Demostrando que la edad fue uno de los factores más importantes para la supervivencia en el Titanic.

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 los cuales permitieron visualizar la dispersión de variables tales 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 de esto puede ser que los datos no presentan datos relevantes como la salud de los pasajeros o la ubicación de estos, dichos datos pueden tener un impacto en la tasa de supervivencia y la cantidad total de sobrevivientes.

Para finalizar, la naturaleza impredecible de un evento como este, a pesar de que los patrones y relaciones puedan proporcionar una visión general de factores influyentes, estos se ven limitados a una captura masiva y no pueden capturar por completos decisiones individuales y circunstancias únicas de cada pasajero.

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 de igual manera partes significativas en las tasas de supervivencia de los pasajeros.

Otro factor a destacar es el rol de la estructura familiar debido a que los pasajeros que viajaban con familiares mostraron una ligera ventaja en las tasas de supervivencia, probablemente esto es debido a una organización conjunta durante el proceso de evacuación, aumentando sus probabilidades de ser rescatados.

9. 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 (Sobreviviendo a la tragedia del Titanic: un estudio sociológico utilizando modelos de aprendizaje automático). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3417433