

## Proyecto final Pandas

Diego Felipe Herrán Contreras

Paula Catalina Rodríguez Torres

Juan Andrés Zambrano Mora

Programación 2

Prof. José Jorge Sierra Molina

Universidad Externado de Colombia

Departamento de Matemáticas

Bogotá D.C.

2025

## **Introducción**

Este proyecto consiste en un análisis de datos exploratorio y predictivo enfocado en el histórico desastre del Titanic. Utilizando la librería Pandas y sus herramientas en un conjunto de datos con el objetivo de analizar y determinar los factores clave que influyeron en la supervivencia de los pasajeros.

## **Principales hallazgos**

- El análisis de supervivencia del Titanic reveló que el género y el estatus socioeconómico (clase) fueron los predictores más fuertes para la probabilidad de rescate. Las mujeres disfrutaron de una tasa de supervivencia significativamente superior a la de los hombres, reflejando la política de "mujeres y niños primero". Los pasajeros de Primera Clase tuvieron, con diferencia, la mejor probabilidad de sobrevivir, lo que subraya el privilegio de la clase social durante la evacuación.
- La posesión de una cabina fue un factor decisivo, actuando como un fuerte indicador de la clase social. Por ejemplo, las mujeres con cabina (Adultos y Mayores) lograron una tasa de supervivencia del 100%. En contraste, los hombres sin cabina (clase baja) tuvieron las peores probabilidades, con los Adultos sobreviviendo solo en un 17.8%.
- Un hallazgo crucial fue la importancia de la variable Familiares (suma de SibSp y Parch), que clasifica a los pasajeros en grupos de riesgo y es un mejor predictor que las variables separadas. El análisis de la distribución de cabinas mostró que las Familias Grandes tenían la proporción más baja de registros de cabina (11.4%) , lo que sugiere una fuerte correlación con la Tercera Clase y valida la idea de que el tamaño de la familia estaba ligado al estatus socioeconómico.
- Aunque los menores de edad fueron priorizados, la política de "niños primero" fue limitada por la clase social. Los menores de edad sin cabina (típicamente de clase baja) tuvieron solo el 21.5% de supervivencia, lo cual demuestra que el estatus social fue el predictor más fuerte de rescate, incluso sobre la edad.

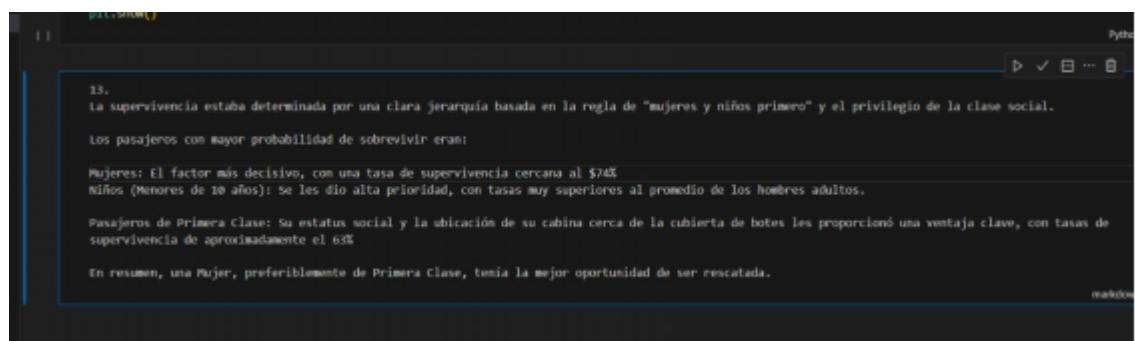
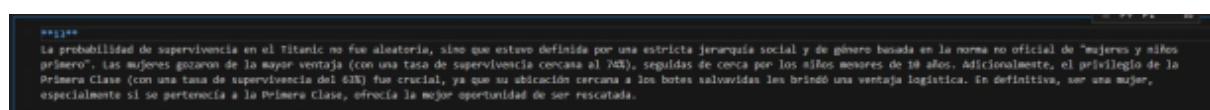
## Proceso colaborativo

La realización de este proyecto fue un claro ejemplo de trabajo colaborativo eficaz, donde la distribución de roles y la comunicación constante fueron fundamentales para alcanzar los objetivos. La coordinación entre los miembros del equipo permitió abordar simultáneamente distintas secciones del análisis de datos.

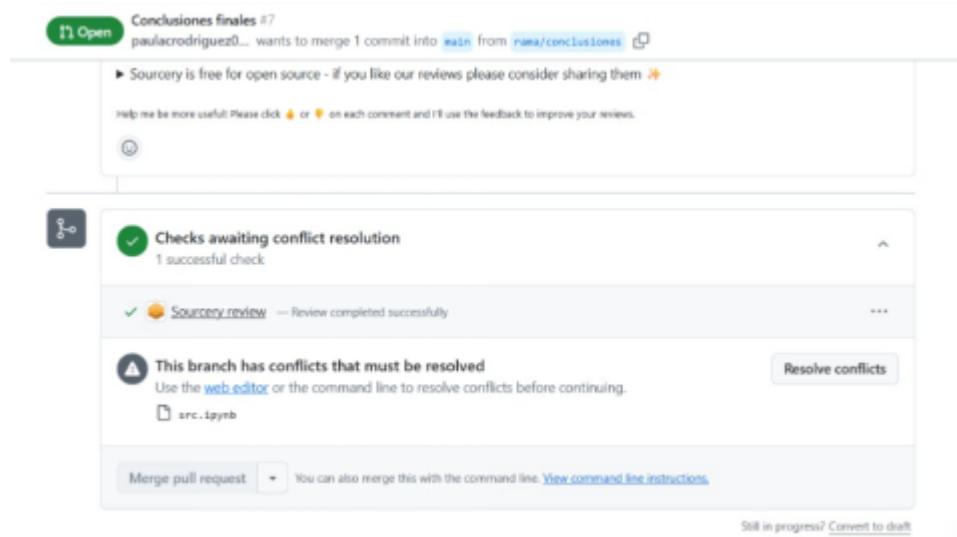
Un elemento clave para el éxito fue la integración de GitHub Desktop en nuestro trabajo, lo que permitió una gestión de versiones fluida y organizada. Al utilizar esta herramienta pudimos sincronizar los avances de manera eficiente, asegurando que todos los colaboradores tuvieran acceso a la versión más reciente del código y los análisis en todo momento. Este uso de Git no solo facilitó la revisión cruzada y la integración de contribuciones de manera segura, sino que también aseguró una sólida justificación de las conclusiones, como el análisis de la representatividad de los datos y las interpretaciones finales sobre la supervivencia, optimizando así el tiempo de ejecución.

A continuación, veremos una parte del proceso resolviendo un conflicto:

1. Aquí dos de los integrantes querían mandar la respuesta del punto 13.



2. Entrando a github podemos ver como ya nos aparece el conflicto por tratar de editar la misma línea con contenido diferente.



3. Para resolver el conflicto debemos escoger la versión que queremos mantener, dependiendo de la elección le damos a “Accept current change” or “Accept incoming change”, en este caso elegimos “Accept current change” porque es la versión perteneciente a la rama “rama/conclusiones”.

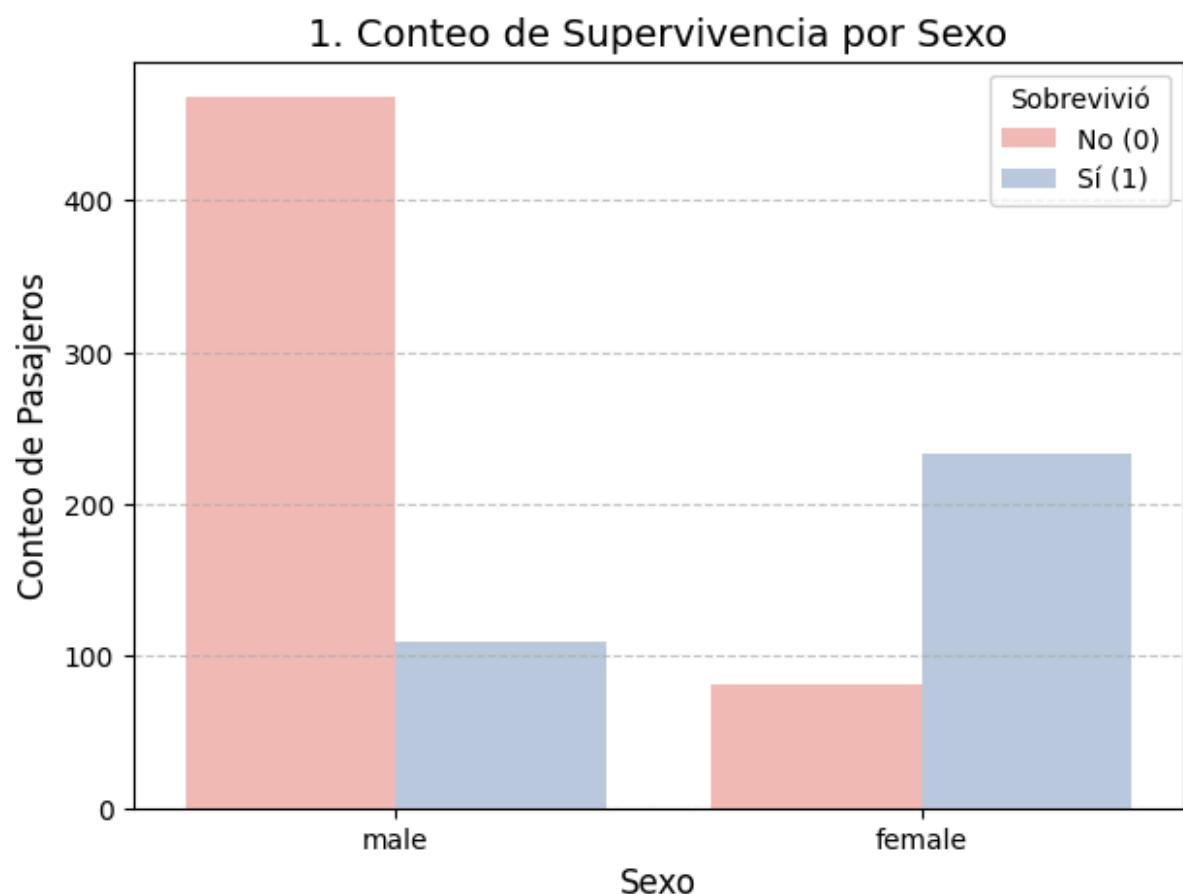
cell\_type : code ,  
Accept current change | Accept incoming change | Accept both changes  
.....

4. Continuamos dándole clic a “merge Branch” para hacer un push a la rama principal.

5. Finalmente, desde github Desktop actualizamos con pull origin y ya queda solucionado el conflicto.

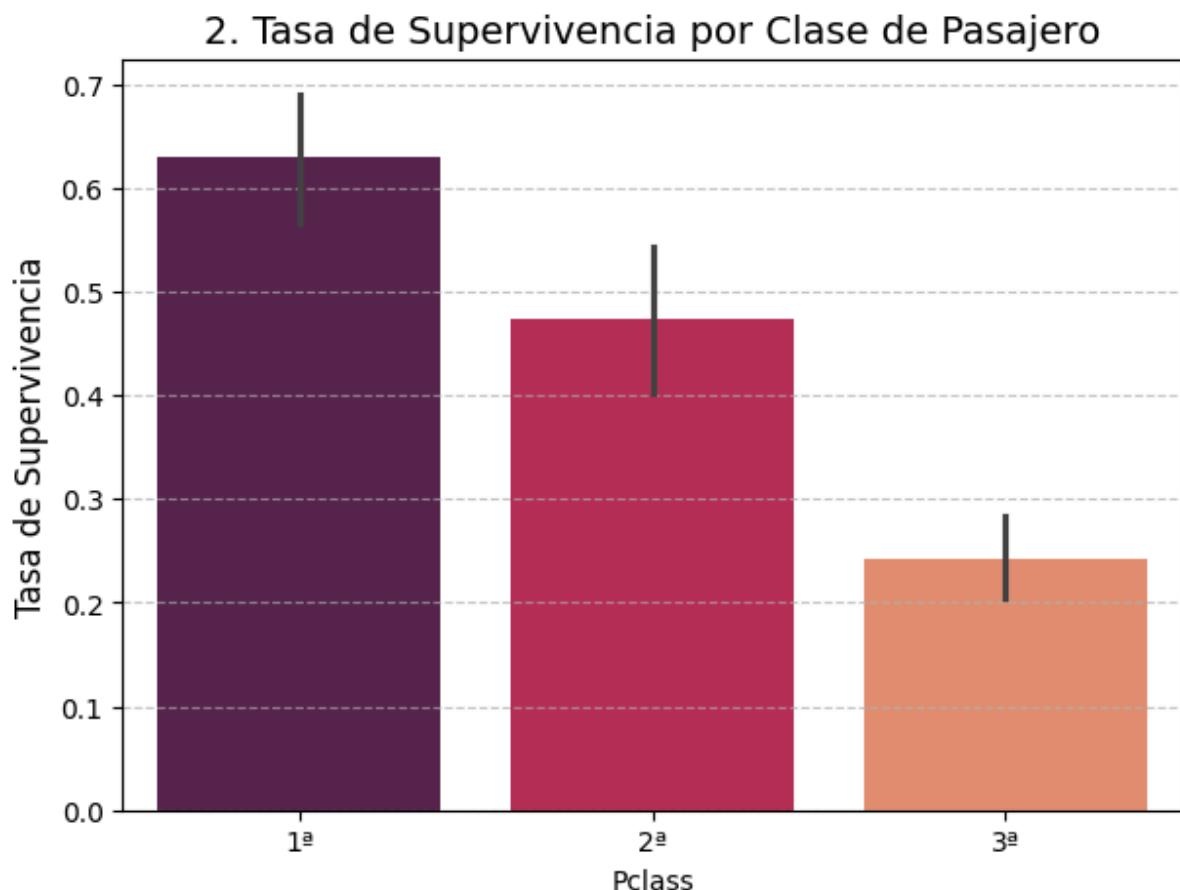
## Evidencia Visual

### 1. Supervivencia por Sexo (Grafico de Barras)



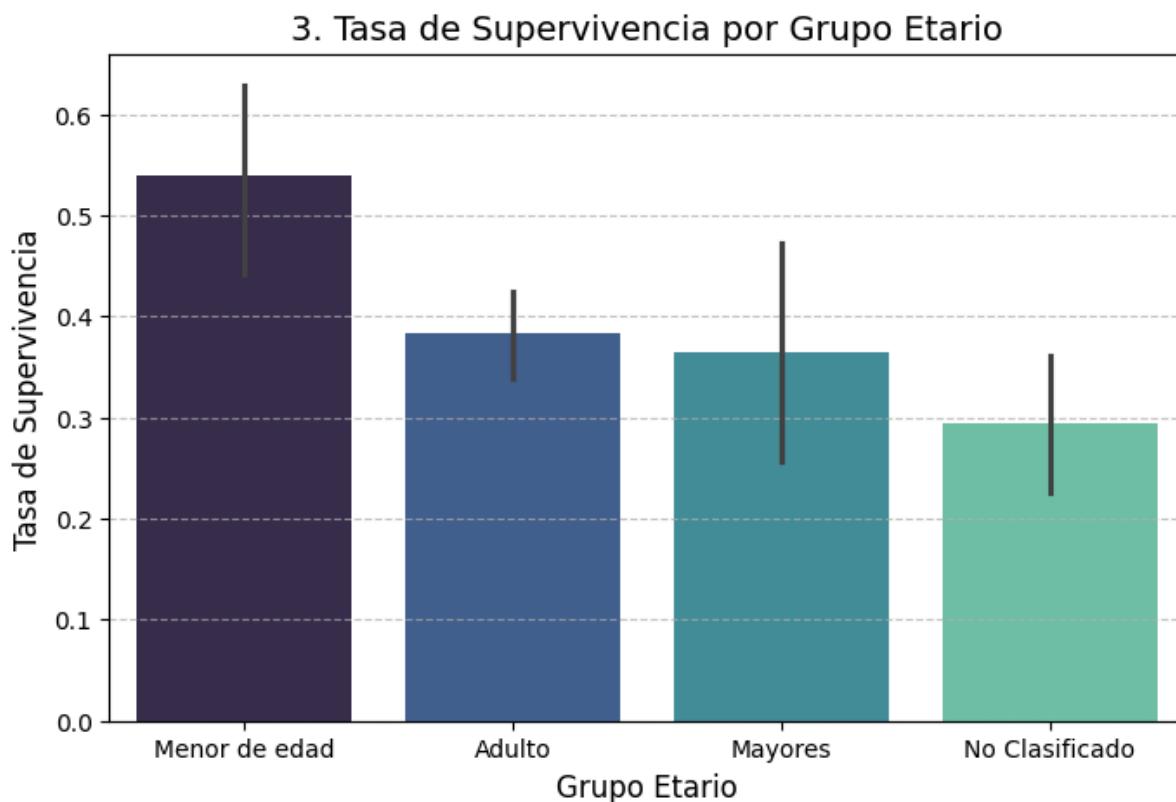
La visualización confirma que, aunque el número total de hombres era mayor, la gran mayoría de los hombres no sobrevivió, mientras que la mayoría de las mujeres sí lo hizo, verificando que el género fue el principal criterio de selección durante la evacuación.

## 2. Tasa de Supervivencia por Clase (Grafico de proporcion)



La visualización confirma que, además del género, la Clase de Pasajero fue un predictor fundamental. Los pasajeros de Primera Clase tuvieron una ventaja de supervivencia significativamente superior a la de los pasajeros de Segunda y, especialmente, de Tercera Clase.

### 3. Tasa de Supervivencia por grupo Etario (Grafico de proporcion)



La edad fue un factor positivo solo en la infancia. El gráfico valida que la regla de "niños primero" se cumplió, pero el riesgo de muerte se incrementó tan pronto como los pasajeros alcanzaron la edad adulta.

## **Conclusiones**

- Gracias a Pandas, fue posible descubrir patrones reales y no evidentes a simple vista, como la relación entre supervivencia y nivel socioeconómico, el impacto del tamaño familiar en el rescate y la influencia indirecta de la cabina como indicador de clase social.
- La ejecución del proyecto se benefició significativamente del trabajo colaborativo eficaz y el uso de herramientas como GitHub Desktop, asegurando una gestión de versiones organizada y una integración fluida de los análisis de cada miembro del equipo.
- El factor más decisivo en la probabilidad de rescate fue el género (con las mujeres sobreviviendo en una tasa cercana al 74%), seguido por el estatus socioeconómico. Los pasajeros de Primera Clase tuvieron las mejores probabilidades (aproximadamente 63%).