

# **INFORME FINAL DEL PROYECTO**

## ***Análisis del Titanic con Pandas y GitHub***

Valery Plaza – Líder del repositorio

Santiago Palacios – Líder de datos

Sofía Argüello – Líder del informe

**Asignatura:** *Programación 2*

*12 de Noviembre de 2025*

**UNIVERSIDAD EXTERNADO DE COLOMBIA**

## Objetivo general del proyecto

El objetivo principal de este proyecto fue aplicar herramientas de análisis de datos en Python, utilizando la librería Pandas, para explorar y comprender el famoso conjunto de datos del *Titanic*.

Además, se buscó fortalecer el trabajo colaborativo en GitHub, aprendiendo sobre control de versiones, manejo de ramas y resolución de conflictos, tal como ocurre en entornos reales de ciencia de datos.

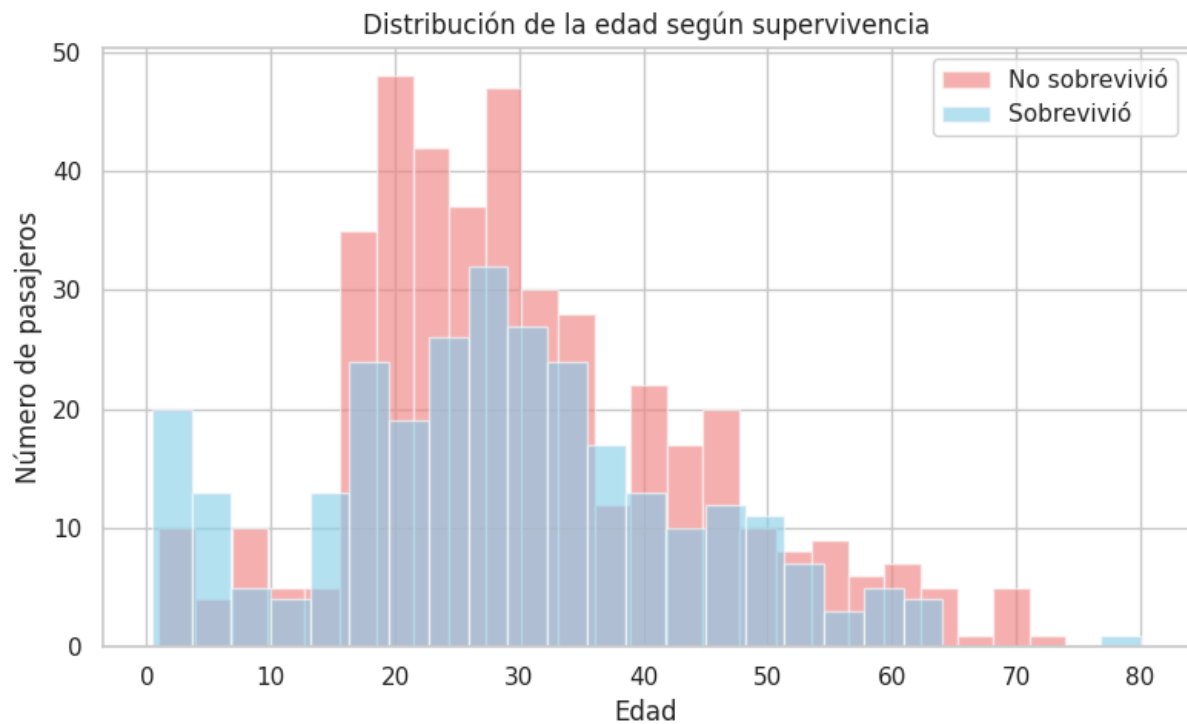
## Principales hallazgos obtenidos a partir del análisis con Pandas

Durante el análisis realizado con Pandas, el grupo identificó varios patrones relevantes sobre la supervivencia de los pasajeros del Titanic:

- El sexo fue un factor determinante: las mujeres tuvieron una tasa de supervivencia significativamente más alta que los hombres.
- La clase social también influyó: los pasajeros de primera clase contaban con mayores probabilidades de sobrevivir, seguidos por los de segunda y tercera clase.
- La edad presentó una tendencia clara: los niños y jóvenes sobrevivieron con mayor frecuencia que los adultos mayores.
- El puerto de embarque tuvo una ligera relación con la supervivencia, destacando que los pasajeros que abordaron en el puerto de Cherburgo tuvieron mejores resultados.
- El tamaño de la familia (suma de hermanos y padres/hijos a bordo) también mostró influencia: quienes viajaban solos o con familias muy numerosas tendían a tener menos posibilidades de sobrevivir.

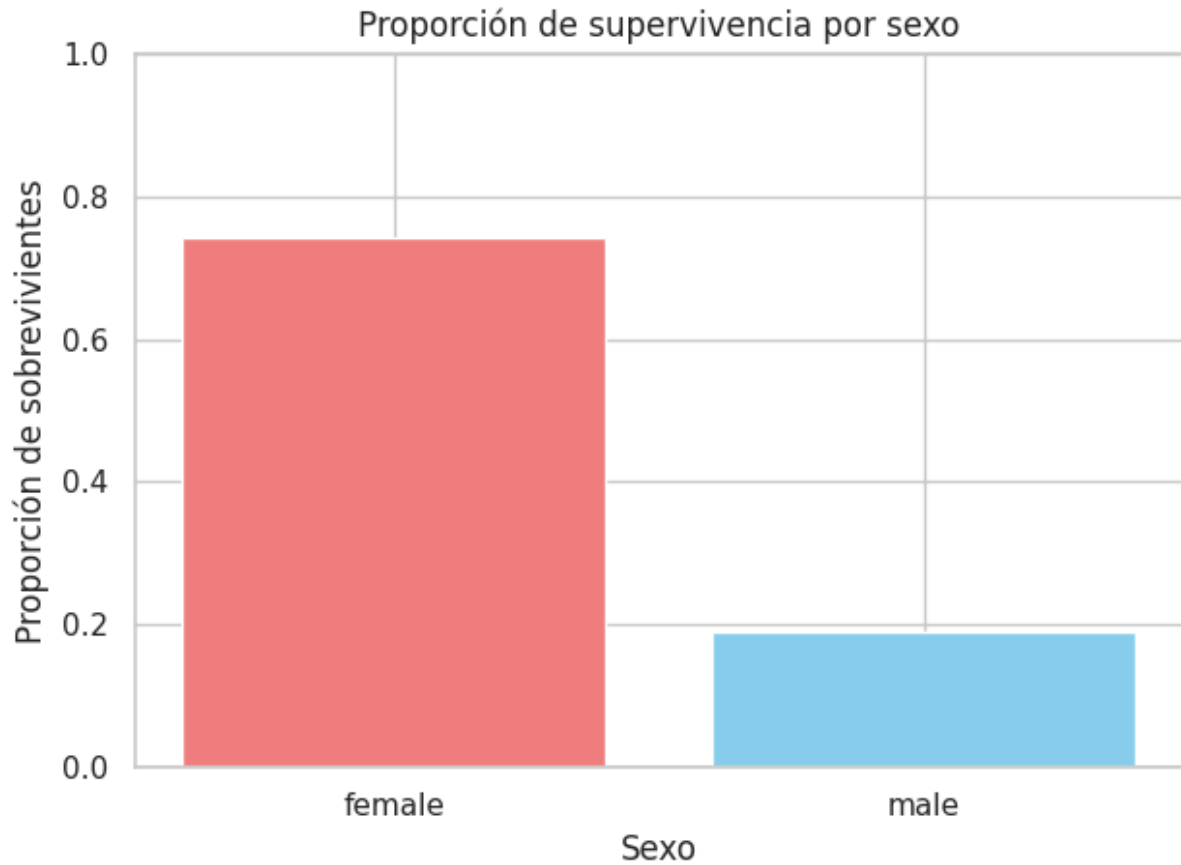
Estos hallazgos permitieron comprender que la supervivencia no fue un evento aleatorio, sino que dependió de una combinación de factores sociales y demográficos.

## Gráficos



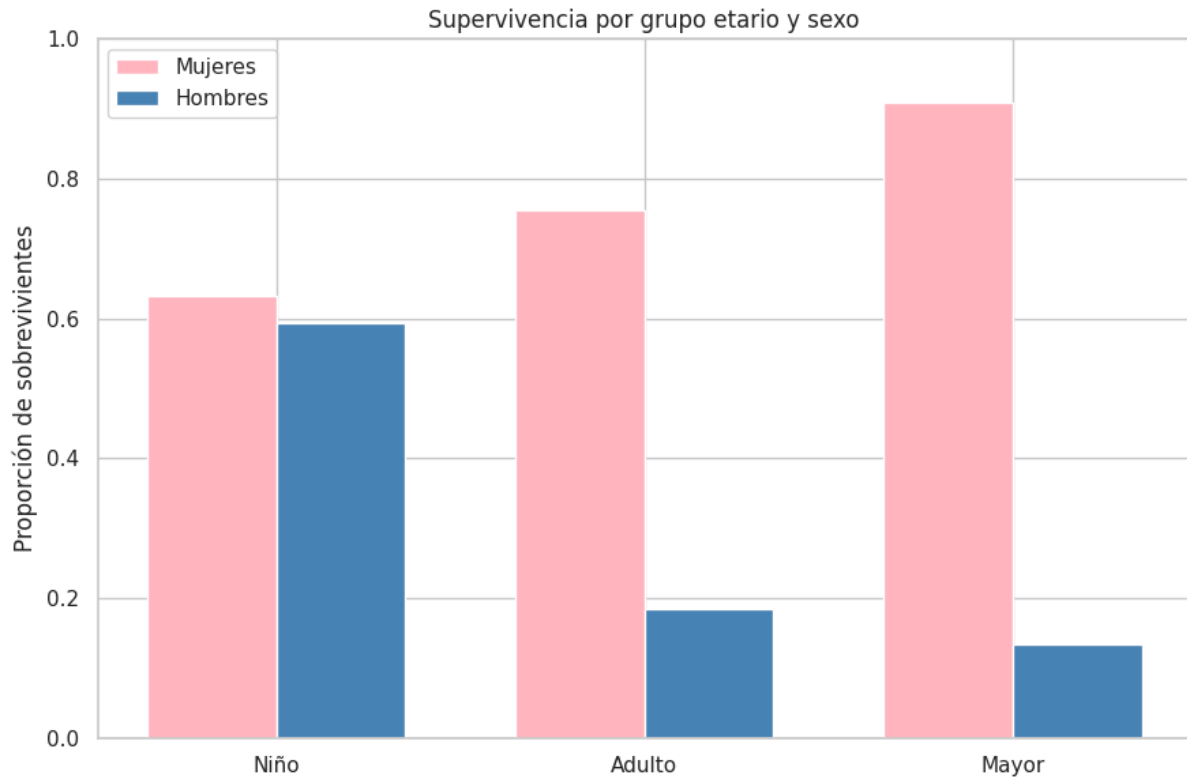
### Histograma de Edad vs Supervivencia

- El histograma evidencia que la **mayor mortalidad se concentró en adultos jóvenes (20–40 años)**, mientras que los **niños pequeños presentaron una supervivencia proporcionalmente más alta**.
- Esto confirma que la edad fue un factor determinante, coherente con la política de rescate del Titanic basada en la prioridad de **“mujeres y niños primero”**.



### Gráfico de barras de Sexo vs Supervivencia

- El gráfico de barras evidencia que el **sexo** fue una variable decisiva en la supervivencia del Titanic.
- Las **mujeres** presentaron una probabilidad de supervivencia mucho mayor ( $\approx 74\%$ ) que los **hombres** ( $\approx 19\%$ ),
- lo cual coincide con los reportes históricos de evacuación prioritaria hacia mujeres y niños.

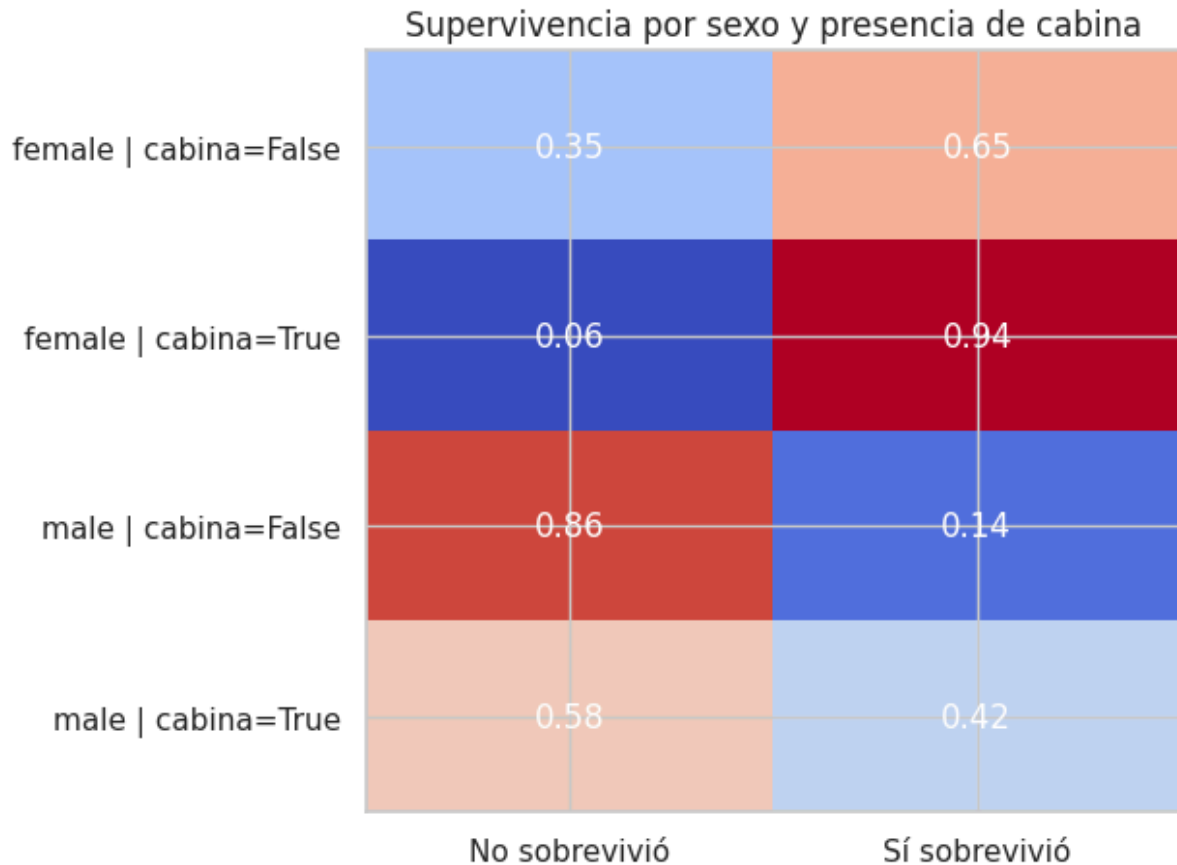


## Gráfico agrupado de Grupo Etario y Sexo

El análisis combinado de **grupo etario y sexo** demuestra que:

- Las **mujeres**, sin importar su edad, tuvieron **muchas más probabilidades de sobrevivir** que los hombres.
- Los **niños pequeños** también fueron un grupo protegido, con altas tasas de rescate.
- Los **hombres adultos y mayores** fueron los más vulnerables.

En conjunto, se confirma que **el sexo (mujer)** y **la edad (niñez)** fueron factores determinantes de supervivencia en el Titanic, coherentes con la política de evacuación **“women and children first”**.



### Heatmap (Sexo × Cabina × Supervivencia)

- El heatmap demuestra una **relación directa entre el sexo, la clase social (tener cabina) y la probabilidad de sobrevivir**.  
Las **mujeres con cabina** fueron el grupo con **mayores tasas de supervivencia**, mientras que los **hombres sin cabina** presentaron las más bajas.
- En resumen, **ser mujer y pertenecer a una clase alta** aumentó significativamente la probabilidad de sobrevivir al desastre, confirmando que **el género y el estatus social fueron los factores más determinantes** en la tragedia del Titanic.

### Conclusiones finales del equipo

El trabajo permitió reforzar tanto los conocimientos técnicos como las habilidades de trabajo en grupo.

Entre las principales conclusiones del equipo se destacan:

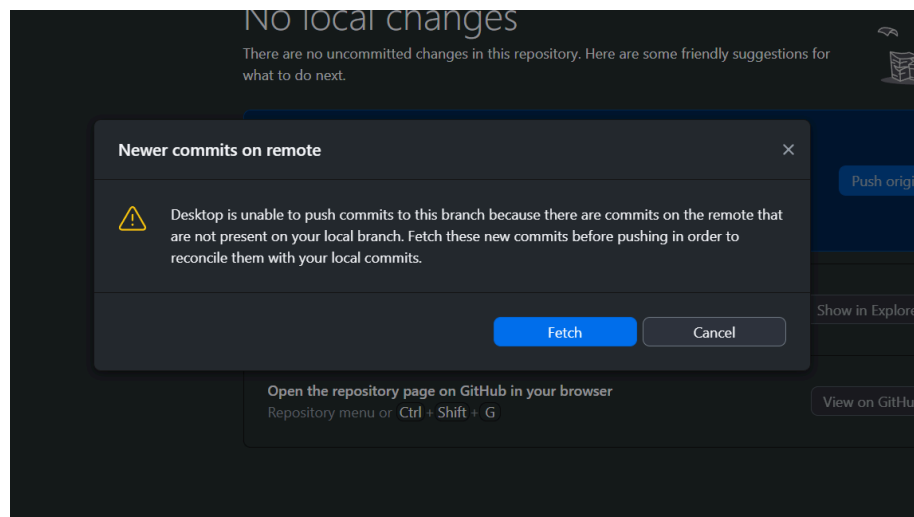
- El valor del trabajo colaborativo: GitHub facilitó la comunicación y la organización del grupo, permitiendo combinar aportes individuales en un proyecto común.
- La importancia de la limpieza de datos: antes de analizar, fue necesario depurar la información para garantizar resultados coherentes.

- El uso de Pandas como herramienta clave: permitió realizar cálculos, agrupar datos y generar resúmenes de manera sencilla y potente.
- El aprendizaje práctico: la experiencia simuló un entorno real de trabajo, donde se integran programación, análisis y colaboración.

En conjunto, el proyecto no solo fortaleció las habilidades técnicas, sino también la organización, la planificación y la comunicación dentro del equipo.

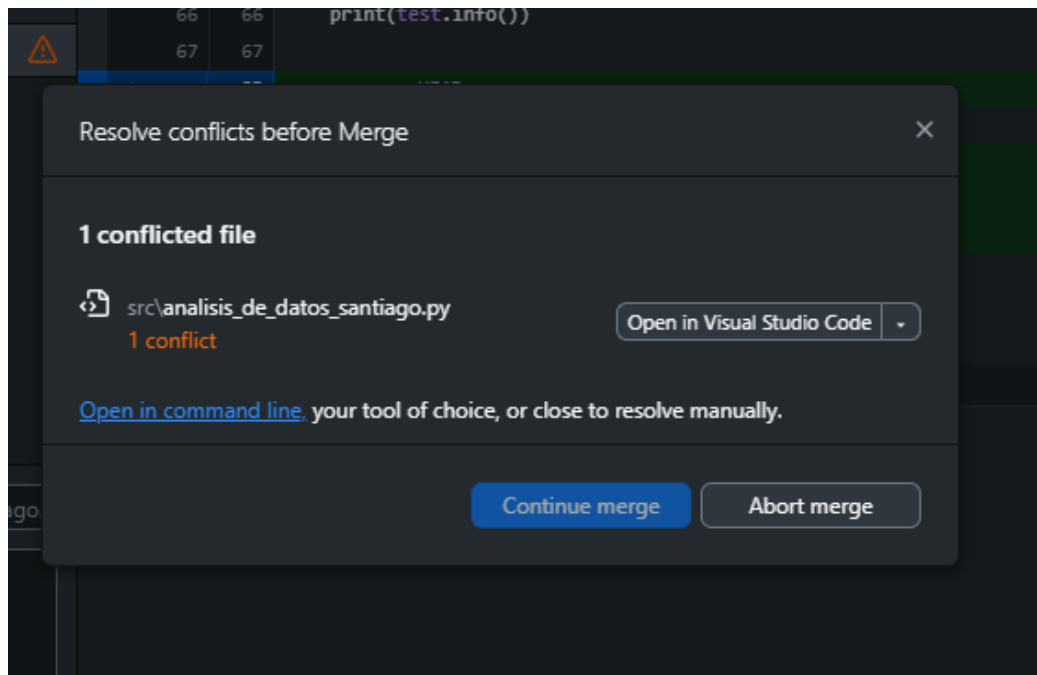
## Descripción de los conflictos resueltos en GitHub Desktop

1. Desde cuentas diferentes modificamos un archivo .py del repositorio desde Visual Studio y guardamos los cambios
2. Hacemos el commit desde la 2 cuentas y creamos el pull request
3. Desde la cuenta del dueño del repositorio al hacer push, apareció el siguiente mensaje:



decía que había *commits más recientes en el remoto*, lo cual indicaba que alguien del equipo había hecho modificaciones en la misma rama antes.

4. Para resolverlo, primero realizamos un **Fetch** para traer los cambios del repositorio remoto. Luego hicimos **Pull** para fusionar esos cambios con nuestra versión local.
5. Al hacerlo, GitHub Desktop nos mostró qué archivos estaban en conflicto



6. Abrimos Visual Studio y revisamos las diferencias:

```
Analisis de datos_santiago.py
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73

* Solo el conjunto train tiene Survived que corresponde a la variable objetivo, indicando si el pasajero
* En las estadísticas numéricas, se puede observar que las variables tienen rangos similares en ambos conjuntos
"""
print("Estadísticas del DataFrame TRAIN")
print(train.describe(include='all'))
print("\nInformación general de TRAIN:")
print(train.info())

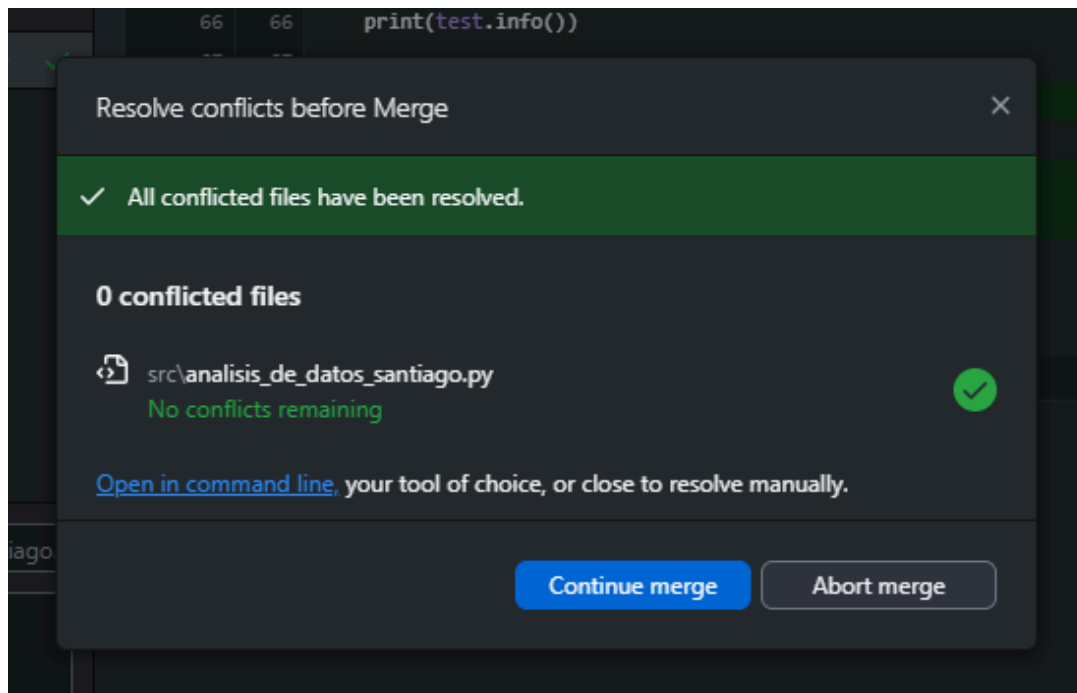
print("\nEstadísticas del DataFrame TEST")
print(test.describe(include='all'))
print("\nInformación general de TEST:")
print(test.info())

Accept Current Change | Accept Incoming Change | Accept Both Changes | Compare Changes
<<<<<< HEAD (Current Change)
# info datasets ...
=====
# Estadísticas
>>>>>> d2d78c44a2252aea177db4c3516c5ce49d15c4b4 (Incoming Change)
```

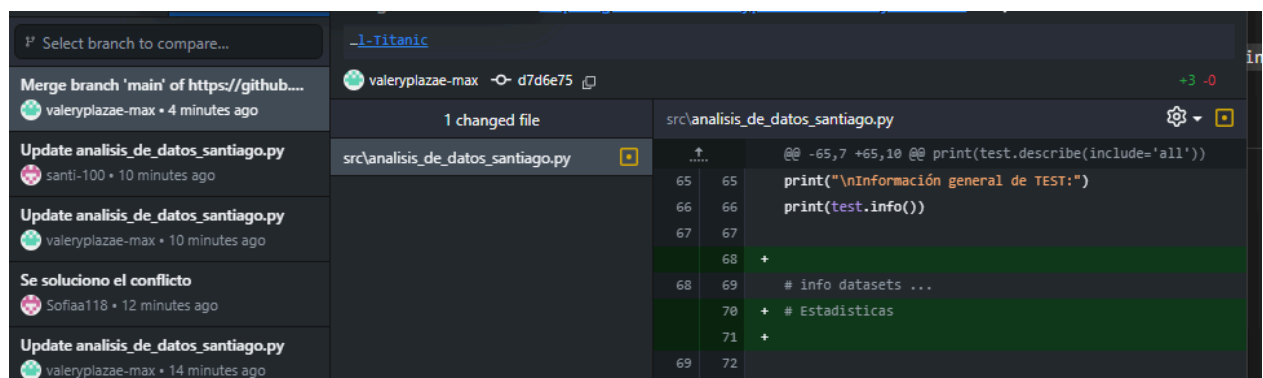
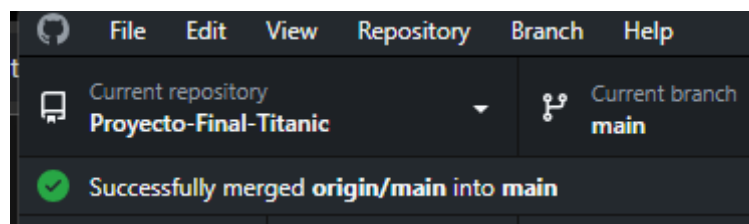
Nuestra solución fue combinar las partes necesarias de ambos códigos.

7. Después de resolver los conflictos, guardamos los cambios y nos aparece el siguiente mensaje:





8. Realizamos el **Continue merge** y finalmente hicimos **Push** para subir la versión final sin errores.



Se subió correctamente la versión final, sin conflictos

# Historial del Repositorio

### Summary

Excluding merges, **3 authors** have pushed **50 commits** to main and **50 commits** to all branches.

On main, **0 files** have changed and there have been **0 additions** and **0 deletions**

### Top Committers

Committer	Commits
[Profile Icon]	35
[Profile Icon]	20
[Profile Icon]	5

**Conflict**

#8 merged 53 minutes ago

**Update graficos.py**

#7 merged 1 hour ago

**Conflict**

#6 merged 2 hours ago

**conflicto**

#5 merged 3 hours ago

**Commit**

#4 merged 2 days ago

**Commit 1 Valery**

#3 merged 2 days ago

**Commit 1 Sofia**

#2 merged 2 days ago

**Commit 1 Santiago**

#1 merged 2 days ago

8 pull requests merged by 3 people