

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.

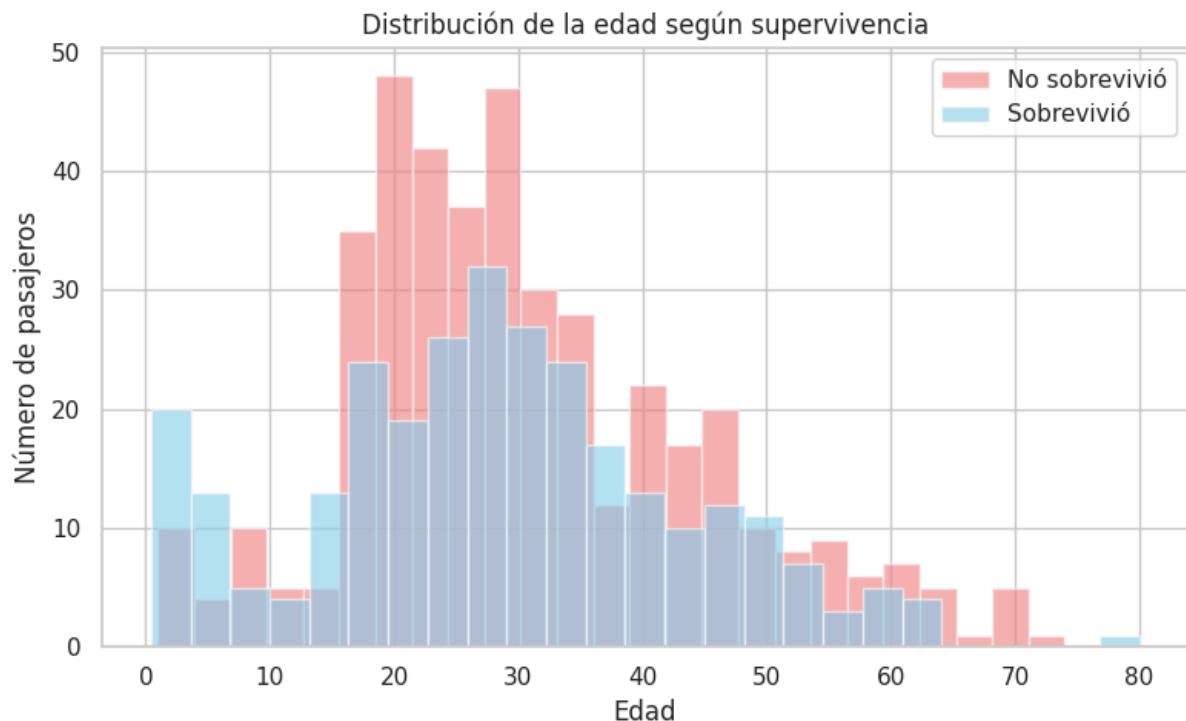
Principales hallazgos obtenidos a partir del análisis con Pandas

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**”.

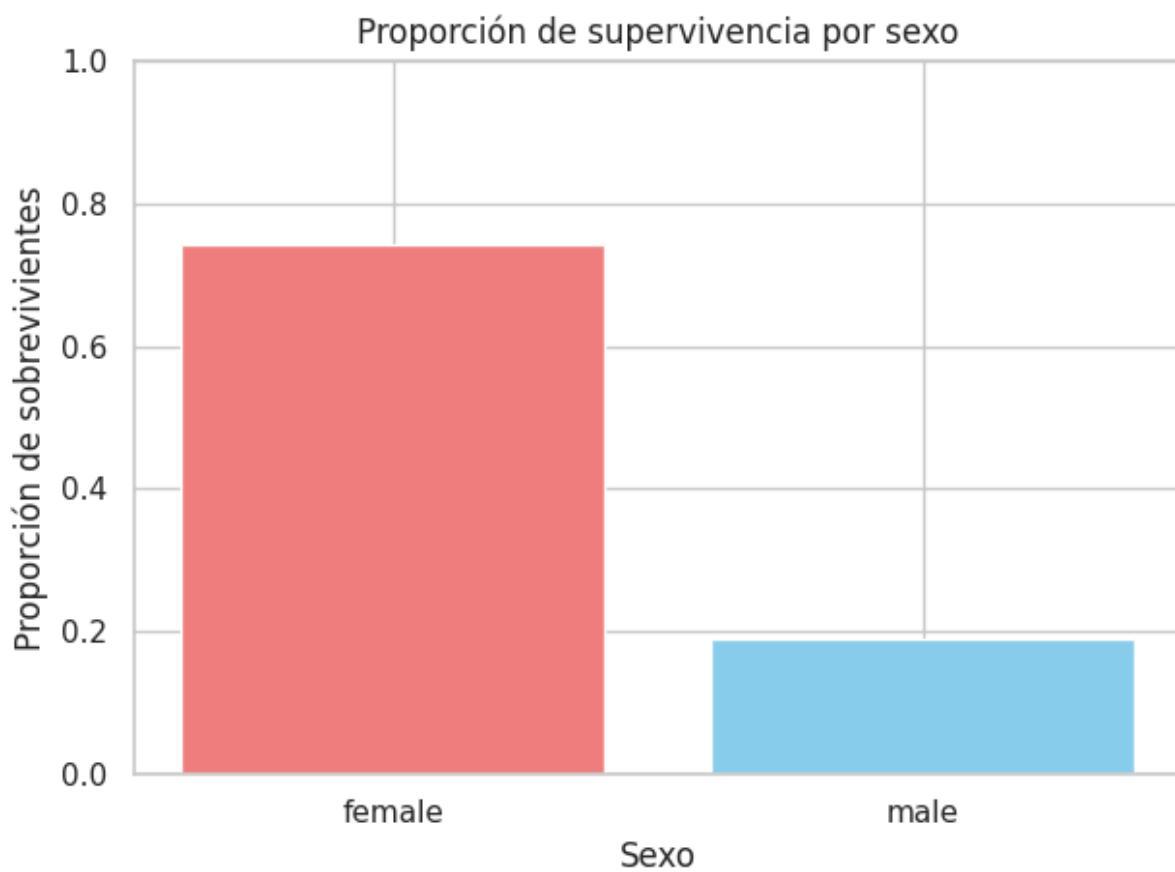


Gráfico de barras de Sexo vs Supervivencia

- 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.

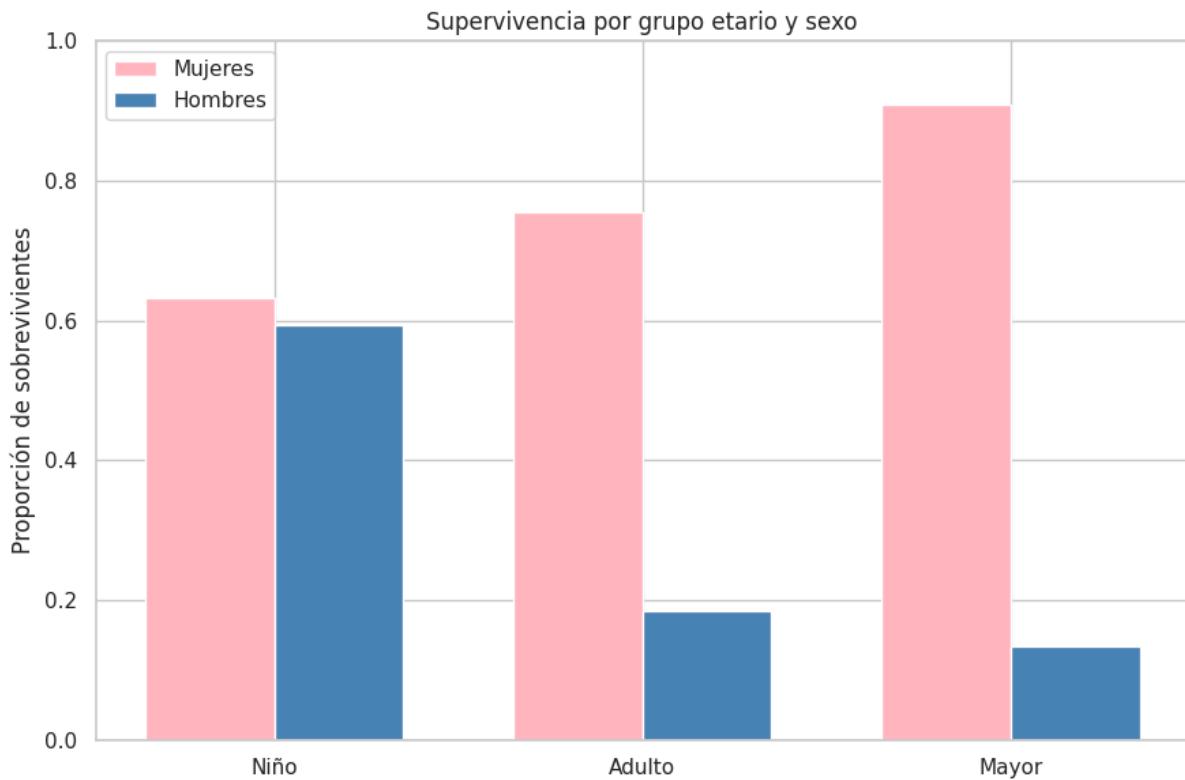
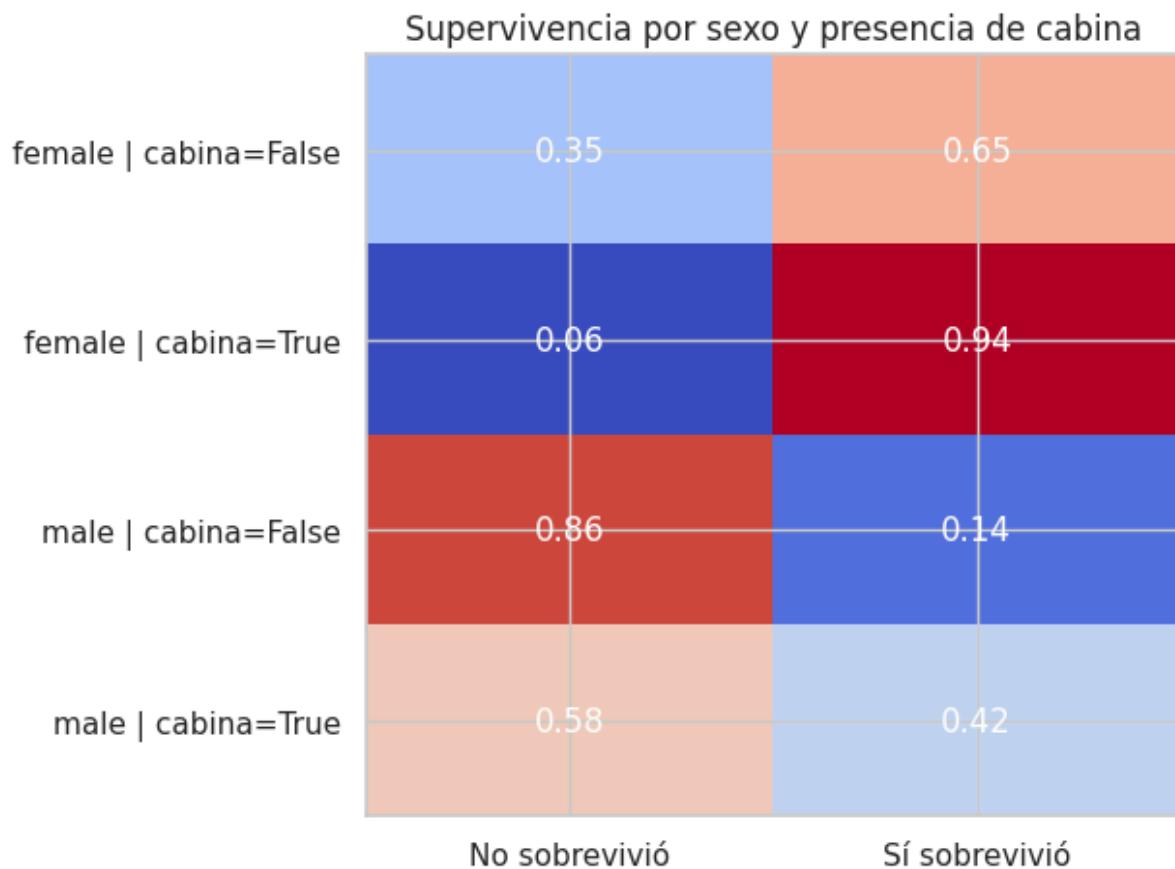


Gráfico agrupado de Grupo Etario y Sexo

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:

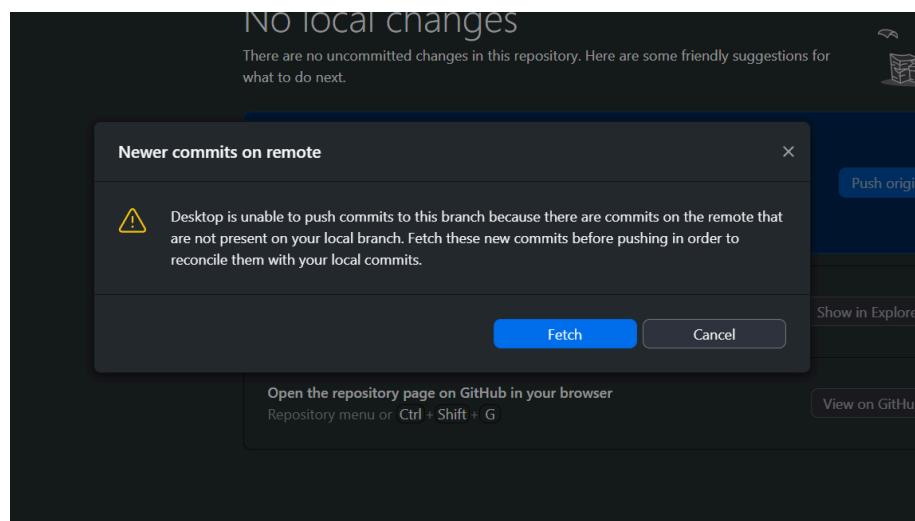
- El valor del trabajo colaborativo: GitHub facilitó la comunicación y la organización del grupo, permitiendo combinar aportes individuales en un proyecto común.
- La importancia de la limpieza de datos: antes de analizar, fue necesario depurar la información para garantizar resultados coherentes.

- 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.

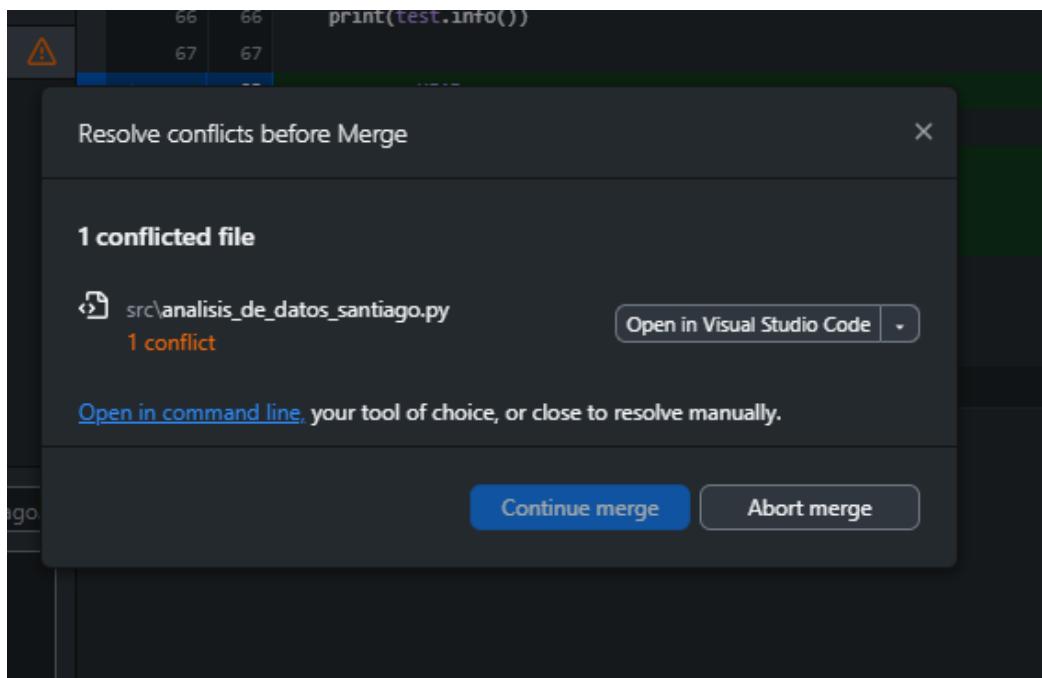
Descripción de los conflictos resueltos en GitHub Desktop

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

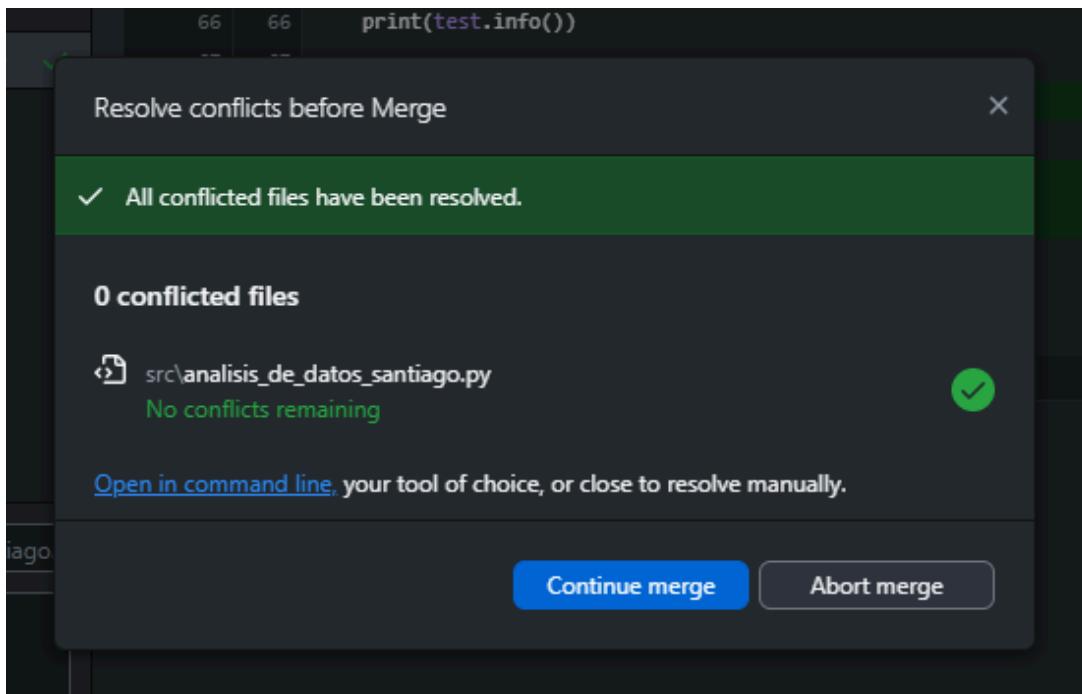


6. Abrimos Visual Studio y revisamos las diferencias:

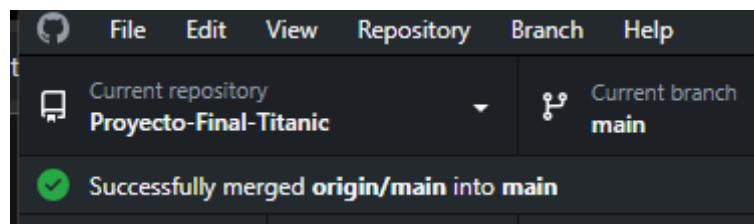
```
1 #Análisis de datos_SantiagoGvipynd
2
3 * Solo el conjunto train tiene Survived que corresponde a la variable objetivo, indicando si el pasajero
4
5 * En las estadísticas numéricas, se puede observar que las variables tienen rangos similares en ambos co
6 """
7
8 print("Estadísticas del DataFrame TRAIN")
9 print(train.describe(include='all'))
10 print("\nInformación general de TRAIN:")
11 print(train.info())
12
13 print("\nEstadísticas del DataFrame TEST")
14 print(test.describe(include='all'))
15 print("\nInformación general de TEST:")
16 print(test.info())
17
18 Accept Current Change | Accept Incoming Change | Accept Both Changes | Compare Changes
19 <<<< HEAD (Current Change)
20 # info datasets ...
21 =====
22 # Estadísticas
23 >>>> d2d78c44a2252aea177db4c3516c5ce49d15c4b4 (Incoming Change)
24
```

Nuestra solución fue combinar las partes necesarias de ambos códigos.

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



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



Select branch to compare...

Merge branch 'main' of https://github.com/valeryplazae-max/Proyecto-Final-Titanic into -1-Titanic

valeryplazae-max • 4 minutes ago

Update analysis_de_datos_santiago.py
santi-100 • 10 minutes ago

Update analysis_de_datos_santiago.py
valeryplazae-max • 10 minutes ago

Se solucionó el conflicto
Sofia118 • 12 minutos ago

Update analysis_de_datos_santiago.py
valeryplazae-max • 14 minutes ago

1 changed file: src\analisis_de_datos_santiago.py

Line	Line	Content
65	65	@@ -65,7 +65,10 @@ print(test.describe(include='all'))
66	66	print("Información general de TEST:")
67	67	print(test.info())
68	68	+
69	69	# info datasets ...
70	70	+ # Estadísticas
71	71	+
69	72	

Se subió correctamente la versión final, sin conflictos

Historial del Repositorio

