

Análisis de Datos del Titanic utilizando Python y Pandas

Por:

Daza P. Diego J

Castillo O. Salomé

Universidad Externado de Colombia

Programación 2

12 de noviembre de 2025

Proyecto de investigación programación 2

El análisis de los datos del Titanic se realizó utilizando la librería Pandas, con el propósito de identificar las variables más relevantes, obtener estadísticas descriptivas y comprender qué factores influyeron en la probabilidad de supervivencia de los pasajeros. Para ello, se generaron diversas representaciones gráficas y comparaciones entre variables, lo que permitió extraer conclusiones más precisas sobre los patrones.

Inicialmente se miraron las variables, tales como: la edad, genero, sobrevivientes, familiares y tarifas. Gracias a eso se obtuvo diferentes hallazgos como los siguientes:

Distribución por edad y supervivencia

Se observó que los pasajeros menores de edad tuvieron una mayor tasa de supervivencia, especialmente los niños pequeños. En cambio, los adultos en su mayoría hombres presentaron menor probabilidad de sobrevivir (Véase anexo 2.2).

Influencia del género

Las mujeres mostraron una tasa de supervivencia significativamente más alta que los hombres, evidenciando la prioridad de rescate hacia mujeres y niños durante el hundimiento (Véase anexo 2.3).

Clases sociales y tarifas

Los pasajeros de primera clase, que pagaron tarifas más altas, tuvieron más oportunidades de sobrevivir. Por el contrario, los de tercera clase presentaron las tasas más bajas, lo que sugiere diferencias en el acceso a zonas seguras (Véase anexo 2.1).

Familiares a bordo

Los pasajeros que viajaban con familiares (hermanos, cónyuges o padres) tendieron a sobrevivir más, posiblemente porque se apoyaban durante la evacuación. Quienes viajaban solos tuvieron peores resultados.

Puerto de embarque

Los pasajeros embarcados en Cherbourg, donde predominaban viajeros de clase alta, presentaron mayor supervivencia que los embarcados en Southampton o Queenstown, donde predominaban pasajeros de clases trabajadoras.

La problemática que se produjo durante el proyecto fue la llamada “merge error” que fue causada por modificar la misma parte de un archivo, en este caso un archivo .py, en ramas

distintas creadas por ambos usuarios. Para su solución, primero se buscó la rama que producía el error y se presionaba el botón “Resolve conflict” (Véase anexo 1.1), eso nos llevó a una nueva pestaña (Véase anexo 1.2), donde aparece el error y se selecciona el botón “accept both changes” (Véase anexo 1.3), que lo que hizo fue aceptar ambos cambios para mantener las aportaciones de ambos miembros, después se marca el botón de “mark resolved” (Véase anexo 1.4), para finalmente fusionar la rama completa (Véase anexo 1.5), sin conflictos, y así solucionar el error.

Al analizar los datos del Titanic con Pandas y diferentes gráficos, se pudieron ver claramente los factores que más influyeron en la supervivencia de los pasajeros. Se notó que la edad, el género, la clase en la que viajaban y si tenían familiares a bordo fueron aspectos muy importantes.

Los resultados muestran que las mujeres y los niños fueron quienes más lograron sobrevivir, mientras que los pasajeros de primera clase tuvieron más posibilidades que los de tercera. También se observó que quienes viajaban con familiares tendieron a sobrevivir más, probablemente porque se ayudaban entre ellos.

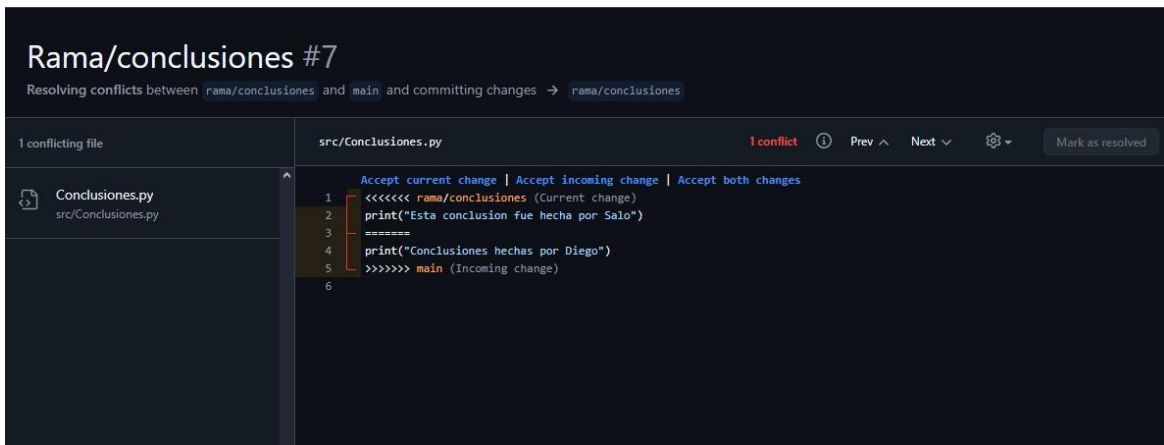
En general, el análisis permitió entender mejor cómo las condiciones sociales y personales de cada pasajero afectaron sus probabilidades de sobrevivir, y demuestra cómo el uso de herramientas de análisis de datos puede ayudarnos a sacar conclusiones claras incluso de hechos históricos.

ANEXOS

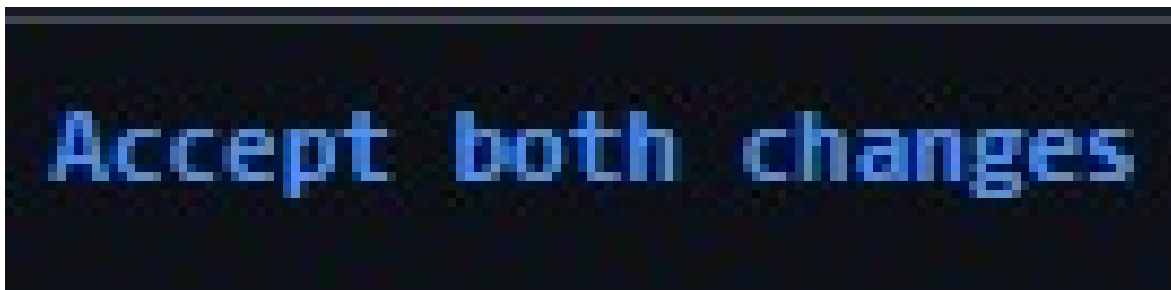
1.1



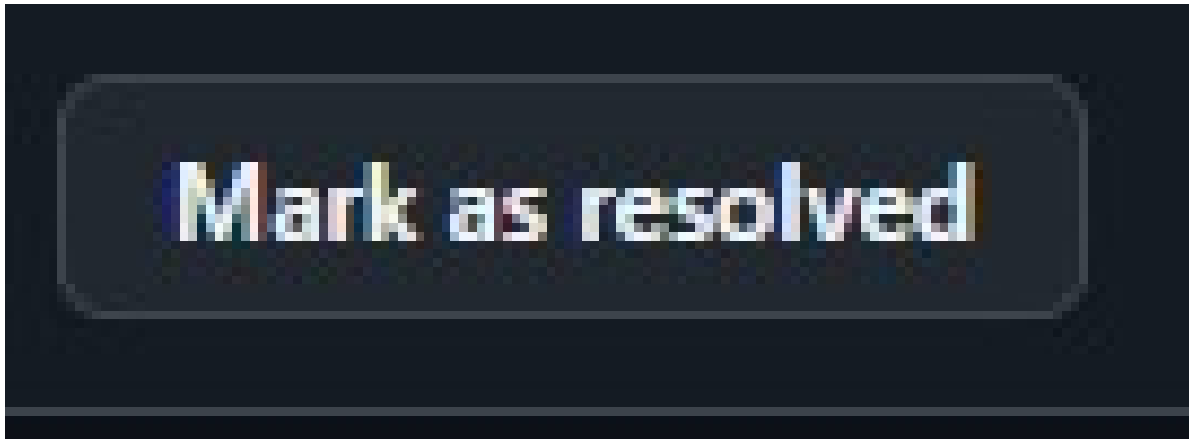
1.2



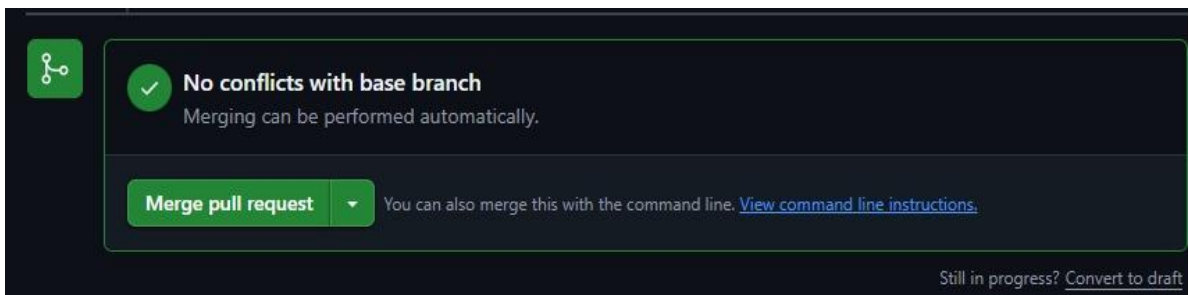
1.3



1.4

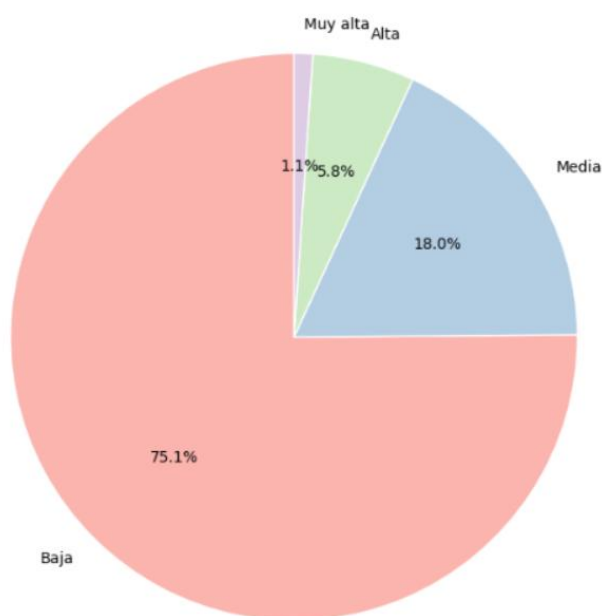


1.5



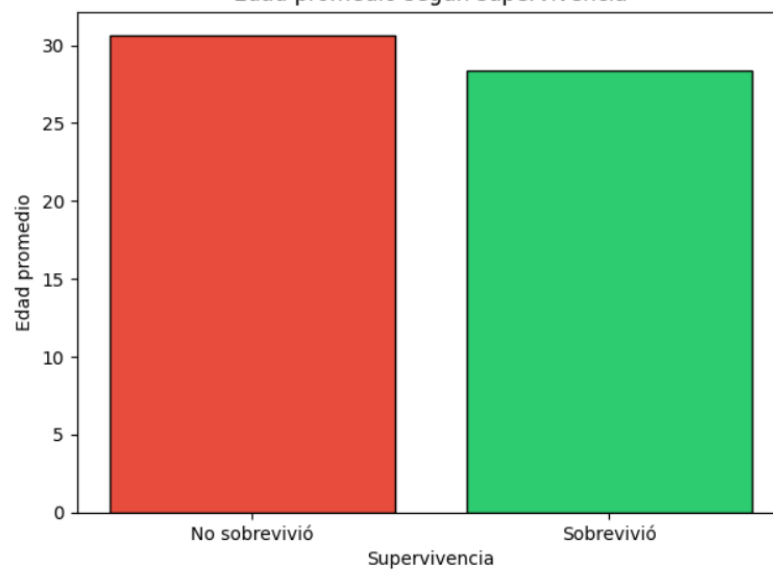
2.1

Distribución de sobrevivientes según la tarifa pagada



2.2

Edad promedio según supervivencia



2.3

