

## REPORTE: ANALISIS EXPLORATORIO DE DATOS

Datos: Nombre de la data <p> </p> Autor: Nombre del autor <p> </p>

Este reporte fue generado por Anthea, una asistente de análisis de datos creada en Marte para apoyar a todos los analistas y no analistas de todo el universo que deseen obtener conocimiento de sus datos.

Esta base de datos parece estar relacionada con los videojuegos, incluyendo información sobre el nombre del juego, la plataforma en la que se lanzó, el año de lanzamiento, el género, el editor, las ventas en diferentes regiones (Norteamérica, Europa, Japón y otras), las ventas globales, la puntuación de los críticos, el recuento de críticos, la puntuación de los usuarios, el recuento de usuarios, el desarrollador y la clasificación del juego. Sin embargo, parece que hay un problema con la importación de los datos, ya que muchas columnas contienen valores 'nan', lo que indica que faltan datos. Además, la columna 'index,Name' parece contener dos tipos de información, el índice y el nombre del juego, que normalmente se esperaría que estuvieran en columnas separadas.

Los datos parecen estar desordenados y mal formateados. Por ejemplo, la columna 'index,Name' parece contener dos tipos de información: un índice y un nombre, que deberían ser dos columnas separadas. Además, hay muchos valores 'nan' (Not a Number), lo que indica que faltan datos.

Además, las columnas 'Unnamed: 16' y 'Unnamed: 17' parecen no contener ninguna información útil, ya que solo contienen valores 'nan'.

Las columnas 'Platform', 'Year\_of\_Release', 'Genre', 'Publisher', 'NA\_Sales', 'EU\_Sales', 'JP\_Sales', 'Other\_Sales', 'Global\_Sales', 'Critic\_Score', 'Critic\_Count', 'User\_Score', 'User\_Count', 'Developer', 'Rating' podrían contener información útil, pero es difícil decirlo con certeza debido a la cantidad de datos faltantes y al formato desordenado de los datos.

Para obtener una visión más clara de los datos, sería útil limpiar y reorganizar el DataFrame. Para limpiar este DataFrame, podrías seguir los siguientes pasos:

1. **\*\*Separar la columna 'index,Name'\*\*: Parece que la columna 'index,Name' contiene dos tipos de información: un índice y un nombre, separados por una coma. Podrías dividir esta columna en dos columnas separadas 'index' y 'Name' utilizando la función `split()` de pandas.**
2. **\*\*Eliminar las columnas innecesarias\*\*\*\*: Las columnas 'Unnamed: 16' y 'Unnamed: 17' parecen no contener información útil, ya que solo contienen valores 'nan'. Podrías eliminar estas columnas utilizando la función `drop()` de pandas.**
3. **\*\*Manejar los valores faltantes\*\*\*\*: Hay varias columnas que contienen valores 'nan', que representan datos faltantes. Dependiendo del análisis que quieras hacer, podrías reemplazar estos valores con ceros utilizando la función `fillna()`, eliminar las filas que contienen estos valores utilizando la función `dropna()`, o reemplazarlos con la media o la mediana de los valores de la columna.**
4. **\*\*Corregir los tipos de datos\*\*\*\*: Algunas columnas parecen tener el tipo de datos incorrecto. Por ejemplo, la columna 'Platform' debería ser de tipo string, pero parece ser de tipo numérico. Podrías convertir el tipo de datos de estas columnas utilizando la función `astype()` de pandas.**
5. **\*\*Eliminar las filas duplicadas\*\*\*\*: Si hay filas duplicadas en el DataFrame, podrías eliminarlas utilizando la función `drop_duplicates()` de pandas.**
6. **\*\*Normalizar los textos\*\*\*\*: En las columnas de texto, podrías convertir todos los textos a minúsculas y eliminar los espacios en blanco al principio y al final de los textos utilizando las funciones `lower()` y `strip()` de pandas.**
7. **\*\*Verificar la consistencia de los datos\*\*\*\*: Por último, podrías verificar si los datos son consistentes. Por ejemplo, si la columna 'Global\_Sales' debería ser la suma de 'NA\_Sales', 'EU\_Sales', 'JP\_Sales' y 'Other\_Sales', podrías verificar si esto es cierto para todas las filas. Si no es así, podrías corregir los datos de manera adecuada.**