What is the verbose parameter while loading a data set in jupyter notebook ?

The verbose parameter is often used in various functions and methods within data science libraries to control the amount of information displayed during the execution of a task. When loading a dataset in a Jupyter Notebook, the verbose parameter, if available, controls how much feedback you get about the loading process.

For instance, in some functions, setting verbose=True may provide detailed information about each step being performed, such as the progress of loading data, the shape of the data, and any issues encountered. Conversely, setting verbose=False will suppress these details, leading to a quieter output.

Tokenization?

Tokenization is the process of breaking down text into smaller units called tokens. These tokens can be words, subwords, characters, or other units of meaning. Tokenization is a common preprocessing step in NLP tasks.

Skiprows in pandas

The skiprows parameter takes an integer or a list of integers, indicating the line numbers to skip.

How to deal with the null values in the data set?

1. **Remove Rows or Columns with Null Values** 
   1. df.dropna(axis=0, inplace=True)
   2. df.dropna(axis=1, inplace=True)
2. **Impute Missing Values**
   1. df['column'].fillna(df['column'].mean(), inplace=True)
   2. df['column'].fillna(df['column'].median(), inplace=True)
   3. df['column'].fillna(df['column'].mode()[0], inplace=True)
3. **Forward/Backward Fill**
   1. df.fillna(method='ffill', inplace=True) # Forward fill
   2. df.fillna(method='bfill', inplace=True) # Backward fill
4. **Using Interpolation**
   1. df.interpolate(method='linear', inplace=True)

**How to count the number of item in the series**

1. Use the len() function to get the total number of elements in the Series, including NaN values.
2. Use the count() method to count the number of non-null (non-NaN) values.
3. Use the value\_counts() method to count occurrences of each unique value.
4. se the value\_counts() method combined with .get() to count the occurrences of a specific item.
   1. specific\_count = s.value\_counts().get(2, 0)

**SimpleImputer**

replaces missing values in a dataset with a specific value or statistic, such as the mean, median, or most frequent value. This is essential for preprocessing data before applying machine learning models, as most algorithms require complete data.

