Data Cleaning Methods: Interpolation, Forward Fill, Backward Fill

This document explains three major methods used to handle missing values in datasets: Interpolation, Forward Fill, and Backward Fill. These techniques ensure continuity and usability of the data for analysis and modeling.

1. Interpolation
Interpolation estimates missing values by assuming a smooth, continuous trend between existing data points. It mathematically calculates the missing value based on surrounding values.

Advantages:
- Produces realistic estimates
- Maintains natural trends

Disadvantages:
- Not suitable for categorical data
- Can be misleading if data is irregular

2. Forward Fill (ffill)
Forward fill replaces missing values with the last known valid value.

Advantages:
- Simple, fast, and logical for many time-series

Disadvantages:
- Can artificially repeat old values
- Not ideal when rapid changes occur

3. Backward Fill (bfill)
Backward fill replaces missing values with the next available valid value.

Advantages:
- Useful when future values are more stable

Disadvantages:
- Illogical for many time-dependent datasets
- Can distort early data segments

Conclusion
Each method is useful depending on the dataset and analysis. Interpolation is ideal for smooth data, forward fill suits persistent-state data, and backward fill works when future measurements represent missing points better.