Pandas Methods and NaN Behavior

Methods That Automatically Fill in NaN

Method	Description
merge() (outer)	Adds NaN where there's no match between DataFrames
join() (left join)	Keeps left rows, fills NaN for missing matches
concat(axis=1)	Adds NaN if row indices don't align
pivot_table()	Missing value combinations result in NaN
reindex()	Adds NaN for new labels not in the index
groupby().agg()	Can return NaN for empty groups
unstack()/pivot()	Missing reshaped values become NaN
crosstab()	Missing combinations default to NaN or 0 depending on options

Methods That Do NOT Fill in NaN by Default

Method	Description
value_counts()	Drops NaNs unless dropna=False
groupby().size()	Counts rows, excluding NaN in group keys
dropna()	Explicitly removes NaNs
fillna()	Replaces NaN with given value
mean()/sum()	Ignores NaN by default
count()	Counts non-NaN values only
apply()	Returns NaN only if your function does
combine_first()	Fills NaN from one DataFrame with another

Rule of Thumb

- Join, merge, and reshape operations (like pivot, reindex, join) often introduce NaN for missing structure.
- Aggregation and summary functions (like mean, sum, count) usually ignore NaN values by default.
- Counting methods typically exclude NaNs unless specified (e.g., value_counts with dropna=False).