

Why missing values are a problem

1. Most machine learning algorithms don't natively handle missing values leading to errors.
2. Removing samples with missing values introduces bias.

Mean-Mode Imputation

X1	X2
1	'yes'
nan	'yes'
3	nan
2	'no'

$$X1: 1+2+3 / 3 = 6/3 = 2$$



X2: 'yes' : 2 , 'no' : 1
→ 'yes'

X1	X2
1	'yes'
2	'yes'
3	'yes'
2	'no'