

Ways Data Can get Dirty



Mis-Entries



Typos, misspelling,
basically wrong
entries

Duplicates



Replicated data or
records

Missing Data



NaNs, Nulls and
Empty records

Combining Data Sources



Combining multiple
data sources into one

Outliers



Data points which
differ significantly
from the others

Quality Checks



Making sure the
data makes sense