Types of Values

Rare Value

Can create bias in factor analysis and other analysis, by appearing more important than they are.

What to do:

Remove

Replace with a more frequent value.

Missing Value

Gaps in the data.

What to do:

<10% data not excluded: **Remove** corresponding observations

Mean substitution

Aberrant Value

Erroneous value corresponding to incorrect measurement, a calculation error, or a false declaration.

***Incorrect dates**: Unknown DOB replaced by 'round numbers', subscription dates before customer's DOB, dates of last updates in the year 2050, 29/Feb in non-leap year.*

Customers declared as 'private' when they are 'business'.

Amount input as cents when it should be in Euros.

What to do:

Delete if not too numerous and if their distribution is suitably random.

Replace with statistically imputed value.

Extreme Value

Observations in a sample so far separated in value from the remainder as to suggest they may be from a different population, or the result of an error in measurement.

What to do:

1-2% data not excluded: Exclude outliers

Neutralize: Divide continuous values into classes

Winsorizing: Replace values of the variable beyond 99th percentile with this percentile.

Missing Value

Name	Income	Job	name	children
Alice	8000000	trust fund retiree	Ethan	2
Bob	40000	rideshare driver		2
Charlie	1	racecar driver	Refuse to answer	1
Danielle	90000	marketing mgr	Gerald	2

Extreme Value, Aberrant Value, Rare Value

Age	Gender	Hair	Eye	Weight	Salary
14	F	Blue	Blue	143	12500
28	F	Brown	Brown	9	32150
22	M	Blue	Brown	215	34200
46	F	Brown	Orange	190	53200
75	M	Gray	Green	187	28040

Aberrant Values