

Imputation – what is it?

• Imputation consists in replacing missing data with statistical estimates of the missing values.

• The goal is to produce a **complete dataset** that can be used to train machine learning models.



More imputation methods

Methods

Complete case analysis

End of tail imputation

Random sample imputation

Mean / median per group



More imputation methods

Methods

Complete case analysis

End of tail imputation

Random sample imputation

Mean / median per group

Remove observations with nan

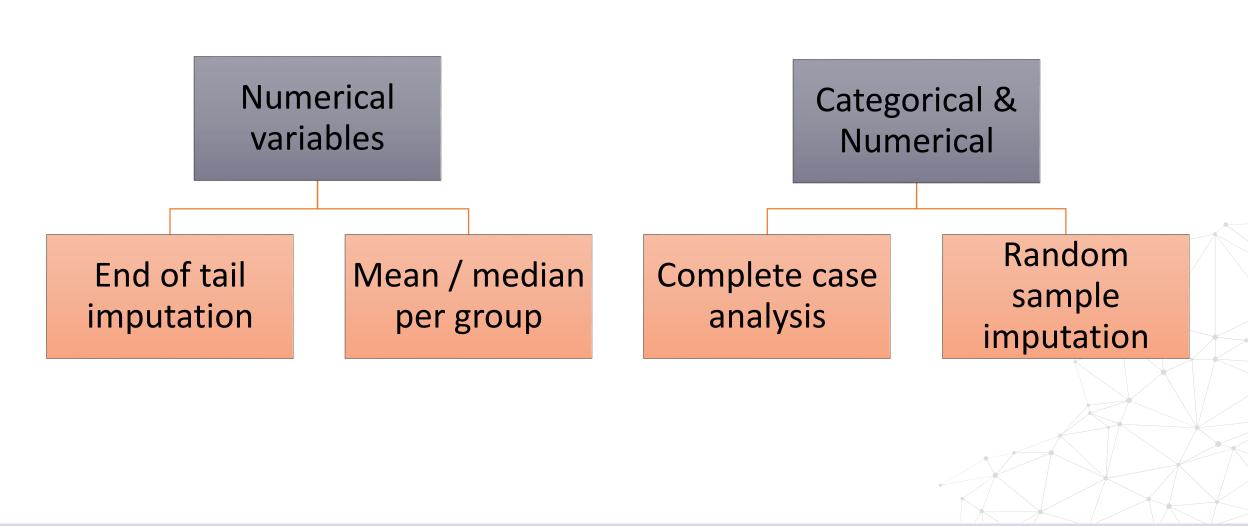
Automate arbitrary number imputation

Alternative to mean, median and mode imputation

Different mean / median values per group



More imputation methods





Imputation: Open Source Libraries











Feature-engine



Objectives





Implement them in pandas and Feature-engine





THANK YOU

www.trainindata.com