

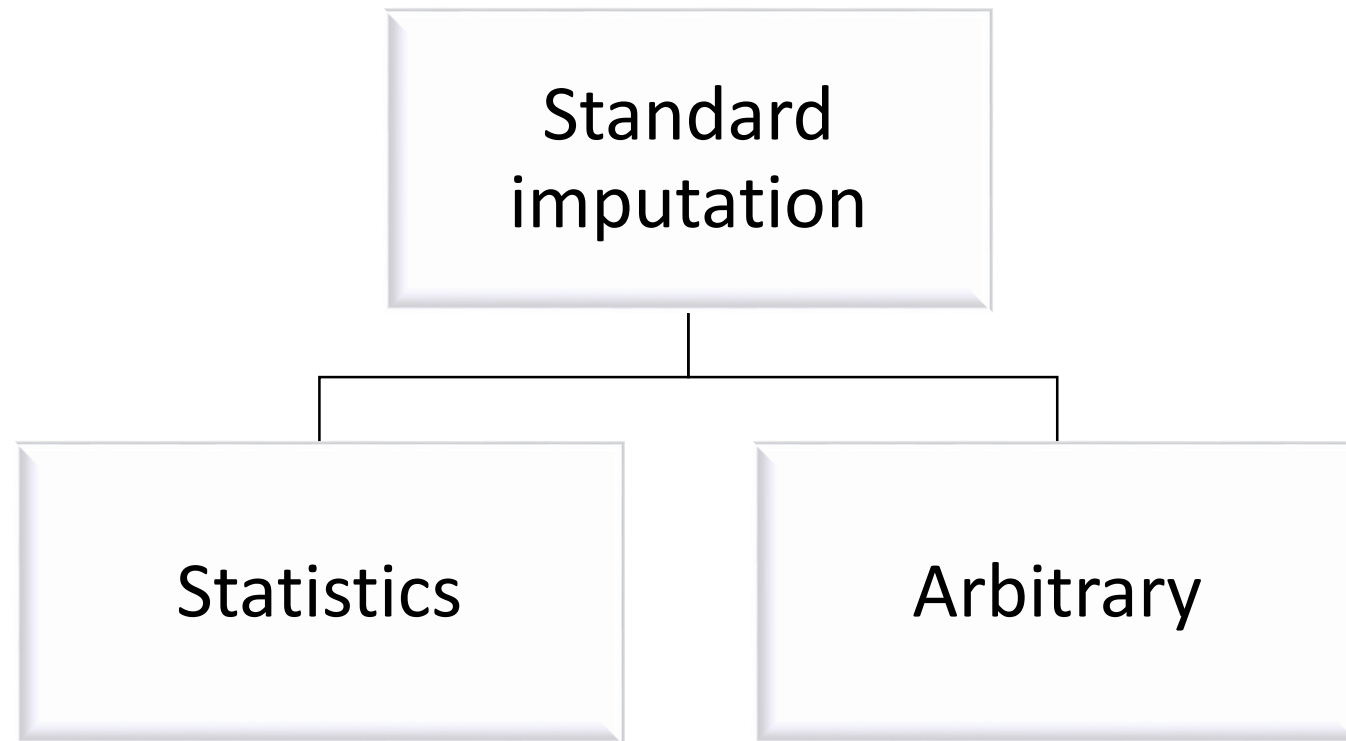


Missing Data Imputation

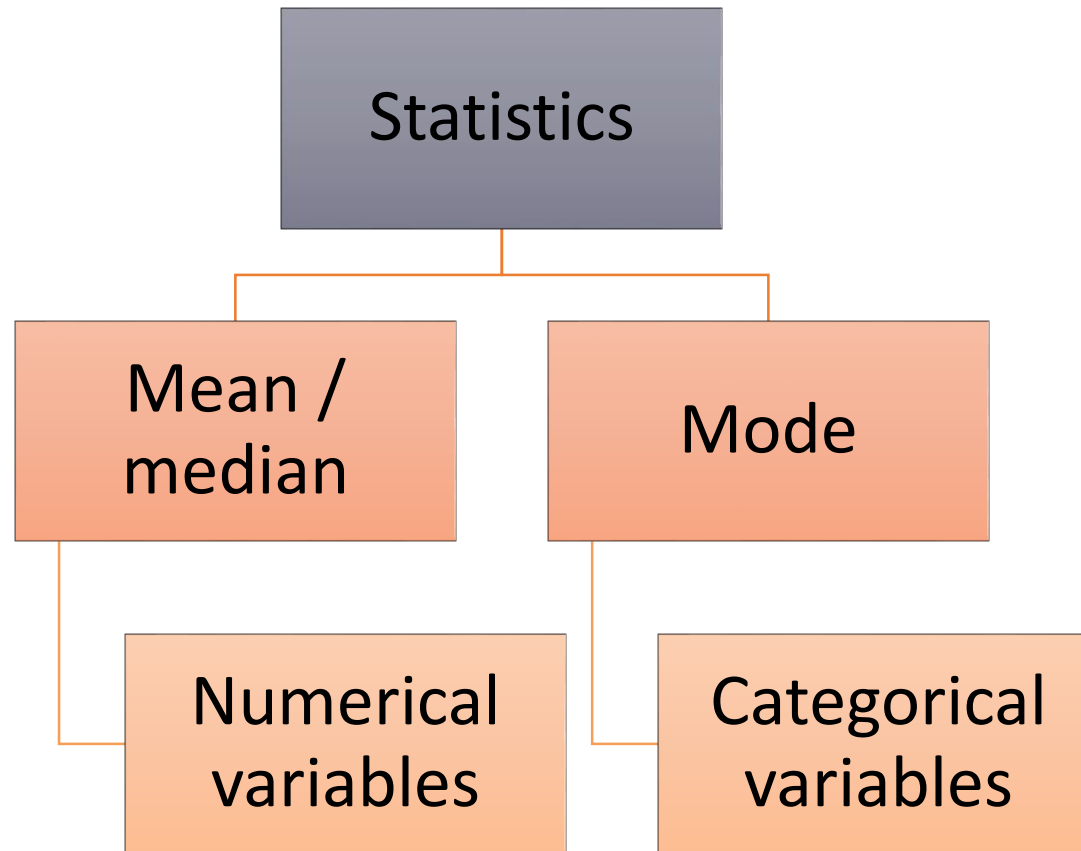
Missing Data Imputation

- Imputation is the act of replacing missing data with statistical estimates of the missing values.
- The goal of any imputation technique is to produce a **complete dataset** that can be used to train machine learning models.

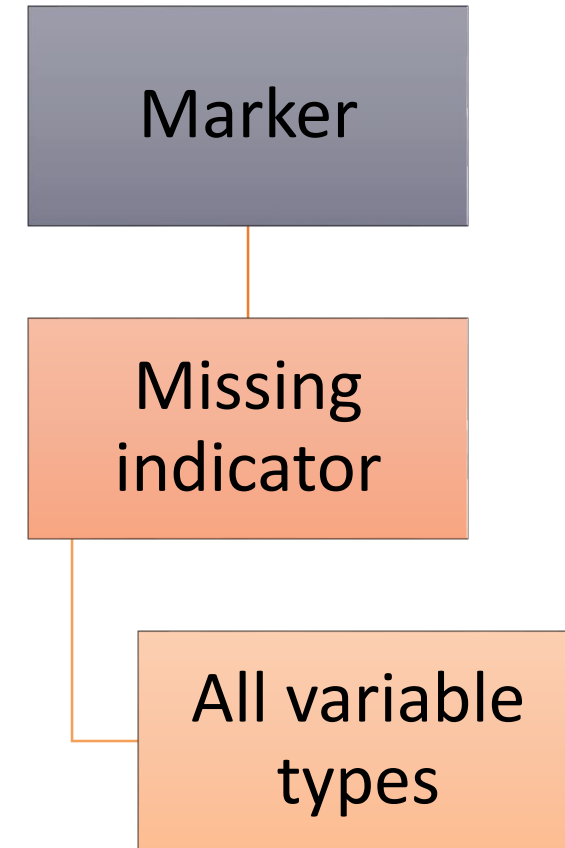
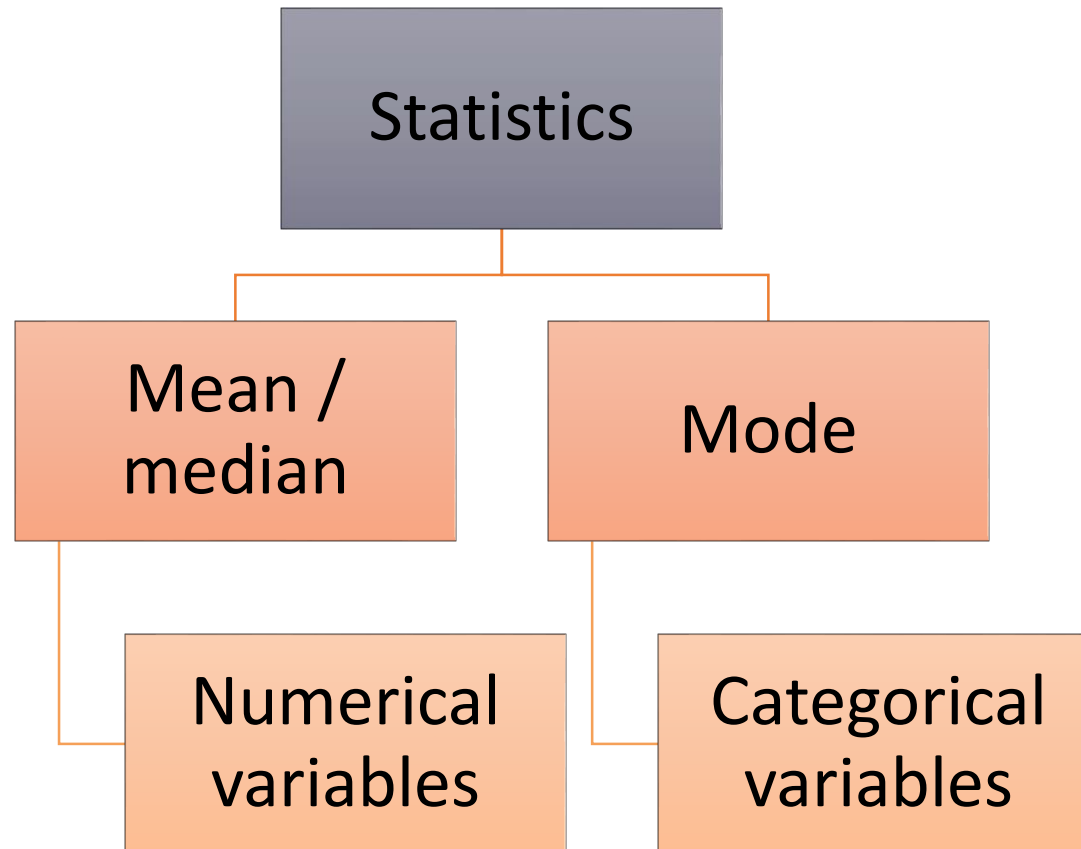
Standard Imputation Methods



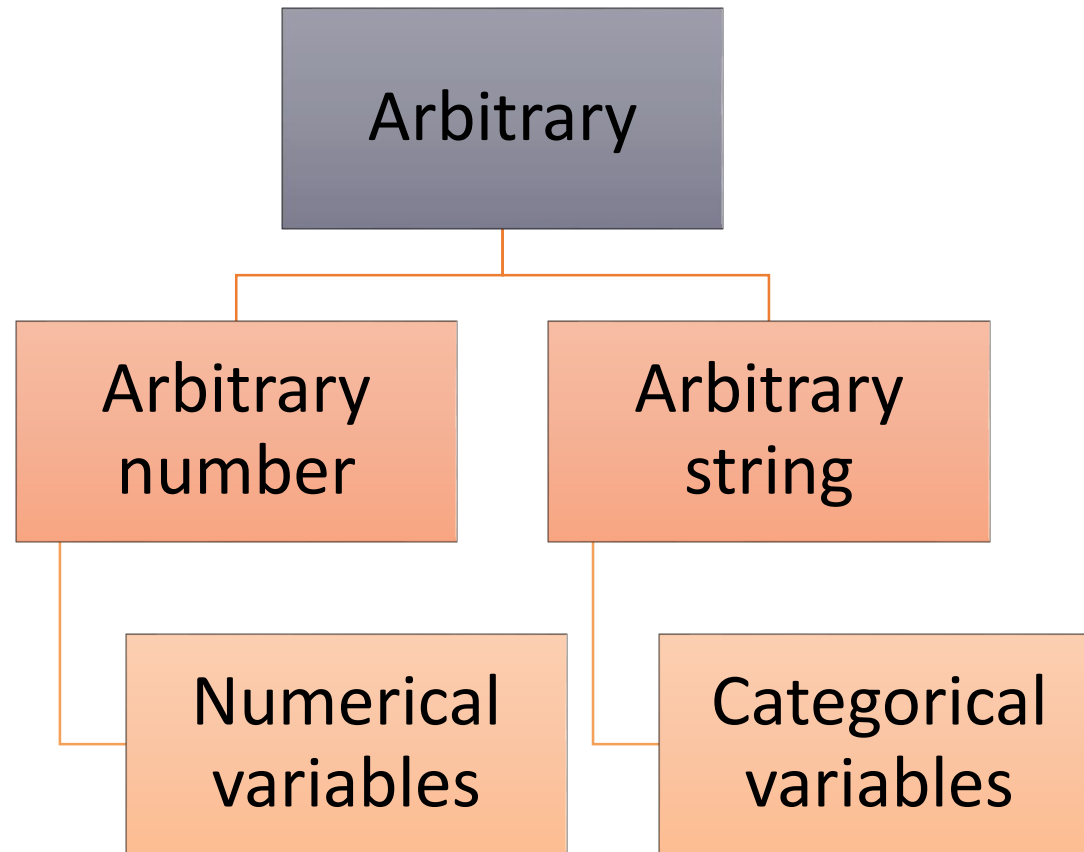
Statistics Based Methods



Statistics Based Methods



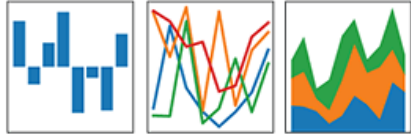
Arbitrary Values



Imputation: Open Source Libraries

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Feature-engine



Objectives



Learn basic imputation techniques



Understand their impact on the variable and the machine learning model



Implement them in pandas, Scikit-learn and Feature-engine

THANK YOU

www.trainindata.com