

a) Open data is data that is available for anyone to access. An example of open data that I produce that others can use are comments and posts on social media

b1) Human Error, Computation Error

b2) Omission: We are assuming the missing data is random, and that deleting it won't introduce bias. Risks include reducing the size of the dataset.

Imputation: We are assuming that the missing values are similar to the rest of the data points. Risks include reducing variance of data.