

Incomplete Dataset T_d

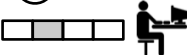
Complete Dataset T_c

① Human Imputation 



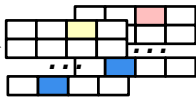
High cost

③ CoreClean

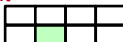


highest utility tuple

add



represent



greedy method

High accuracy, Low cost

Coreset Selection

Low accuracy



②



Low cost

\neq

High accuracy

Coreset Selection



Coreset C

Coreset C'

Coreset C^*