

Dirty Dataset D_d

Clean Dataset D_c



Human Cleaning

High cost



Coreset4DC
High accuracy

*Coreset
Selection
& Clean*

**Low
accuracy**

Coreset C'

Coreset C



*Human
Cleaning*

Low cost

Coreset C^*

*Coreset
Selection*