

Compute the utility

Dirty Dataset D

	Name	Gender	Depart.	Age	Working Years
t_1	Nancy	F	Sales	24	2
t_2	Amy		Finance	42	10
t_3	Lei	M		35	
t_4	Lily	F	Market		8
t_5	James	M	Sales	38	2
t_6	Tom	M		36	6

Sample $h=3$ tuples

Greedly construct one-by-one

Greedy Coreset Construction

c_1	James	M	Sales	38	2
c_2	Amy		Finance	42	10

t_6	Tom	M		36	6	c_3
-------	-----	---	--	----	---	-------

t_3					
t_4					
t_6					

Pick the one with the largest utility