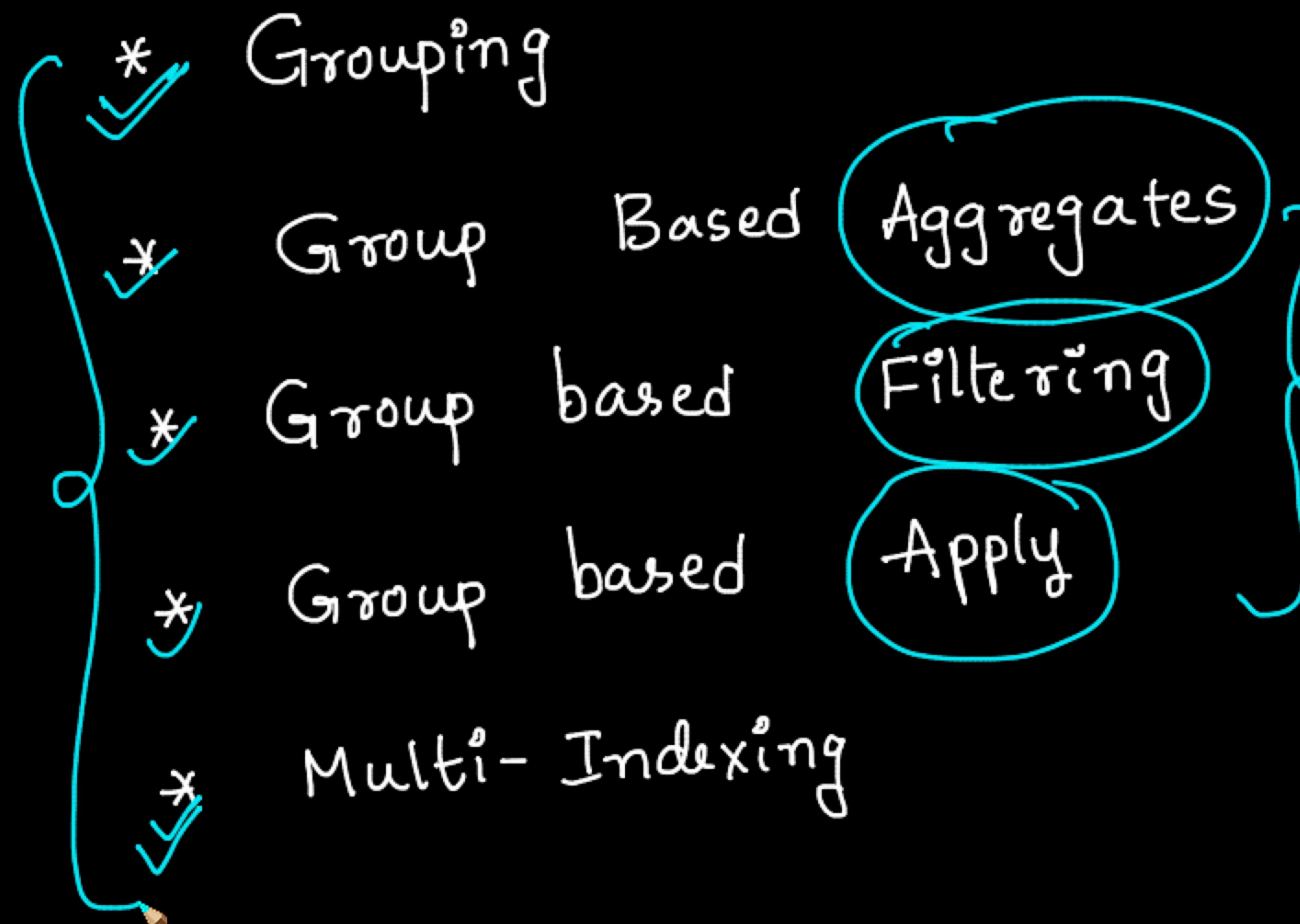
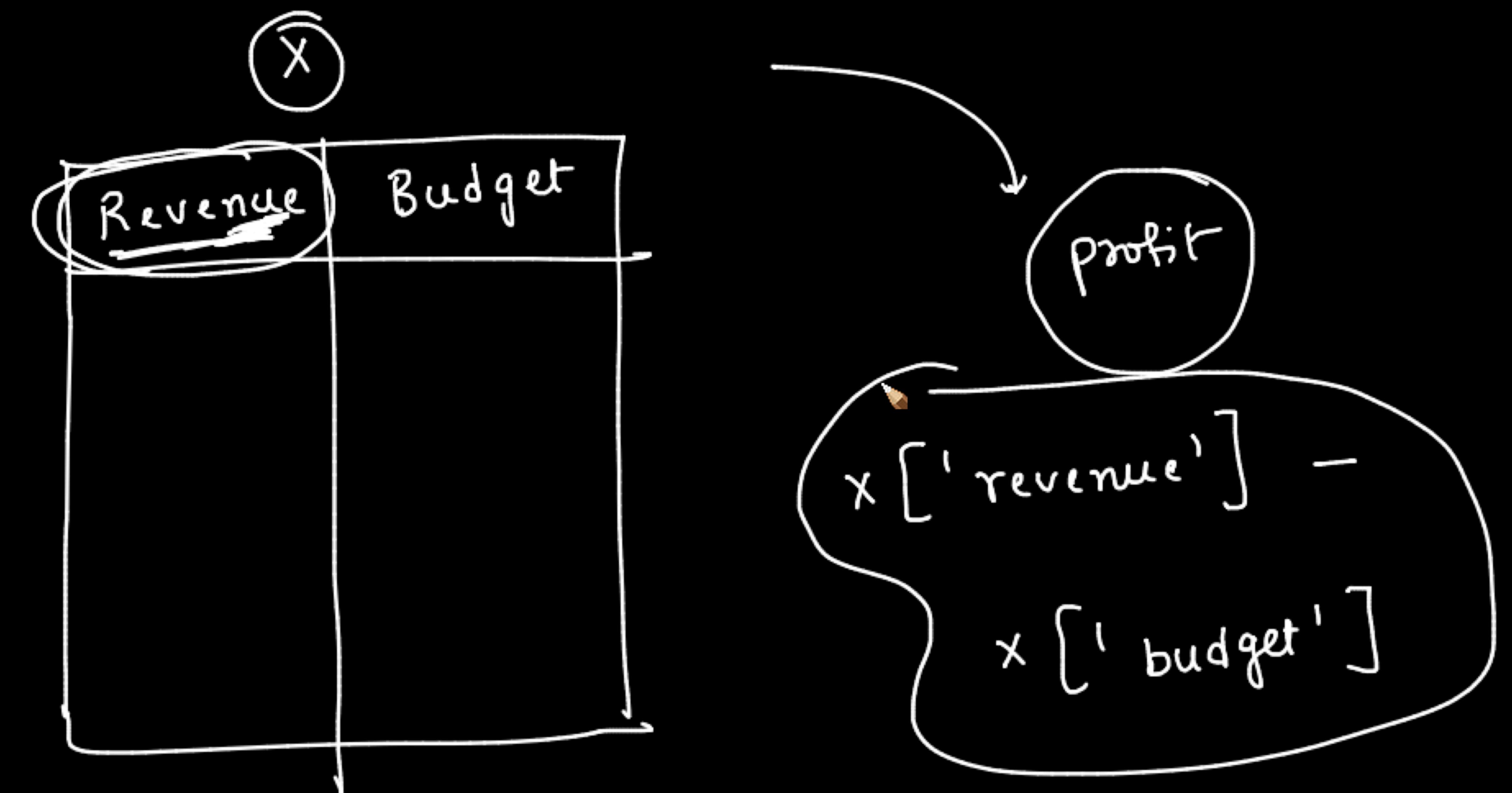


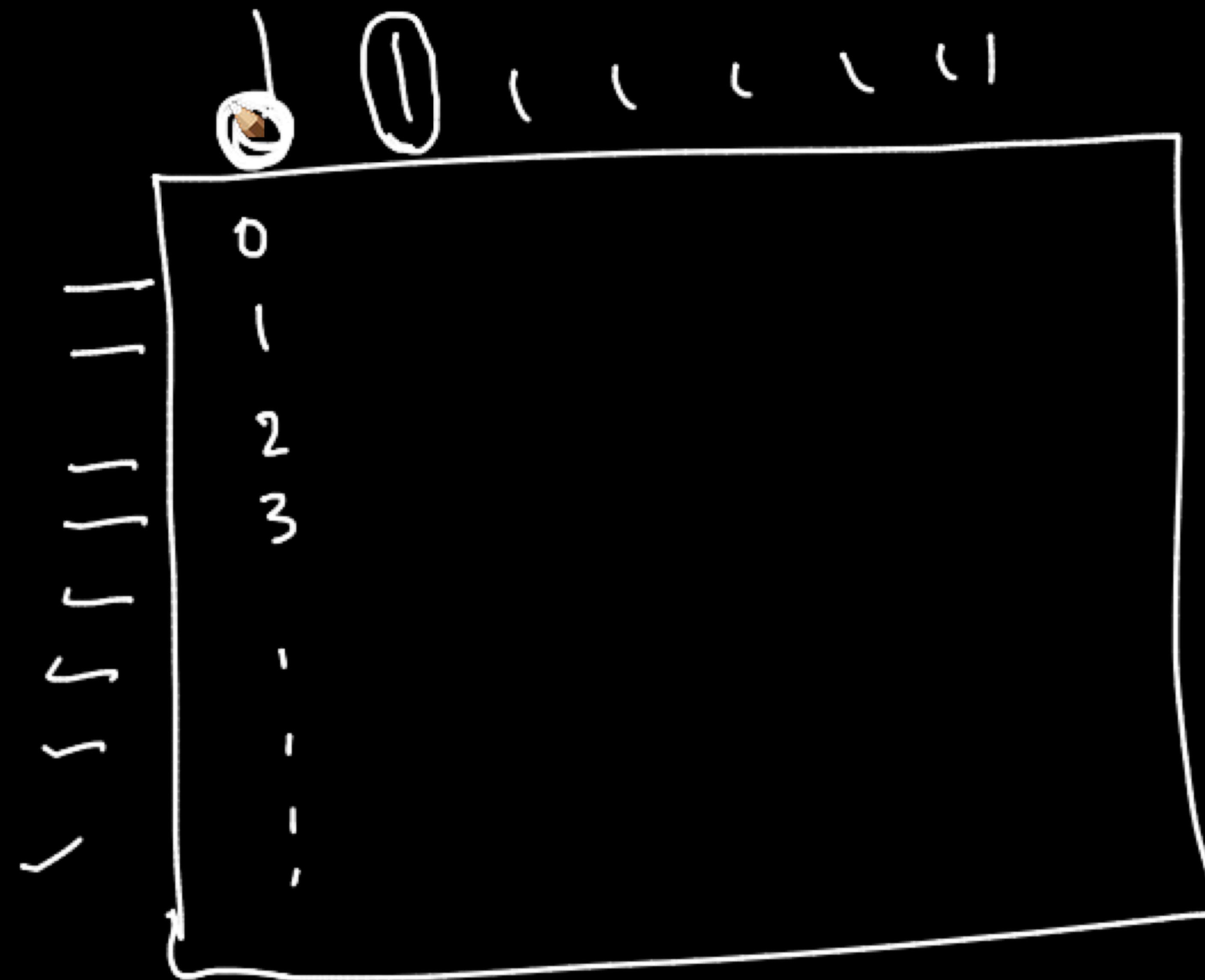
* Agenda *



Groupby



Grouping



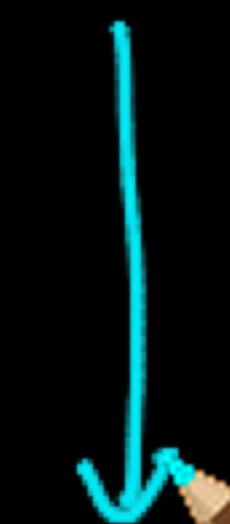
index_col = p

(1) Find the

no of movies released

by particular director

christopher nolan ?



What is Grouping?

Key	Data
A	1
B	2
C	3
A	4
B	5
C	6

K	D
A	1
A	4

K	D
B	2
B	5

K	D
C	3
C	6

split, apply, combine
(sum)

K	D
A	5

K	D
B	7

K	D
C	9

K	D
A	5
B	7
C	9

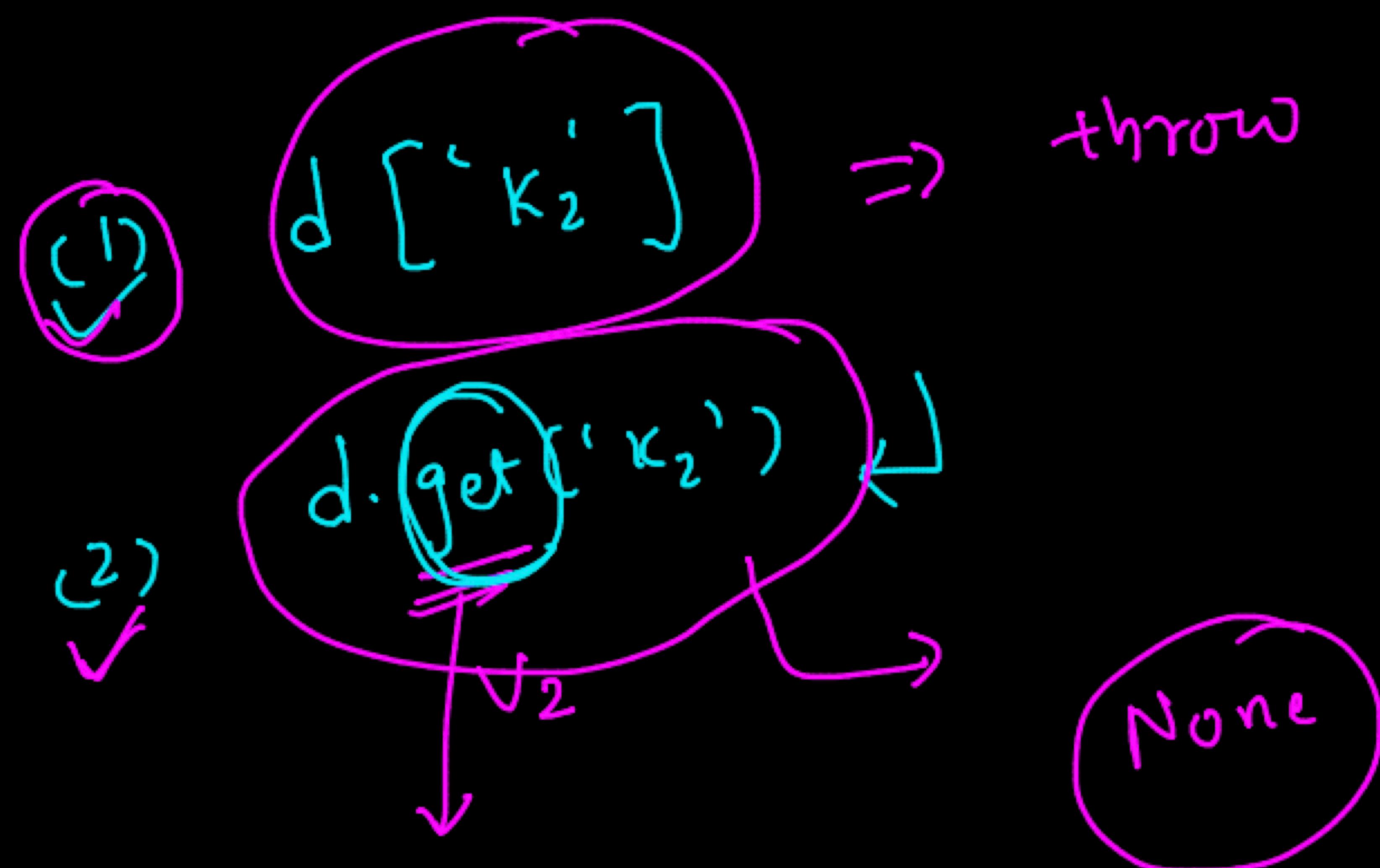
Aggregate
filter
apply



Group based

Aggregate

$d = \{ (k_1 : v_1, k_2 : v_2, k_3 : v_3) \}$



get_group

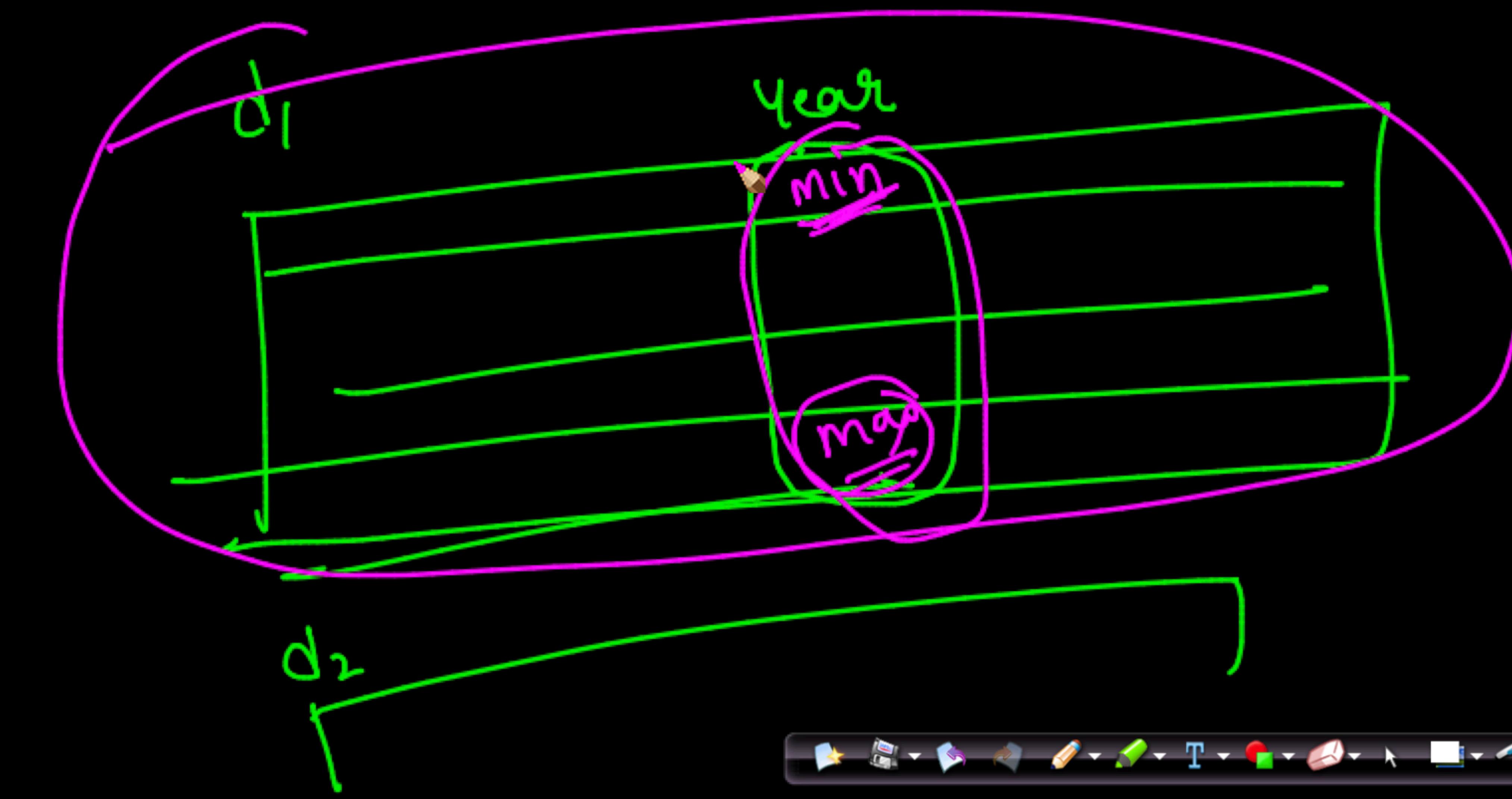
θ

d_1

d_2

d_3

first year min = last year max

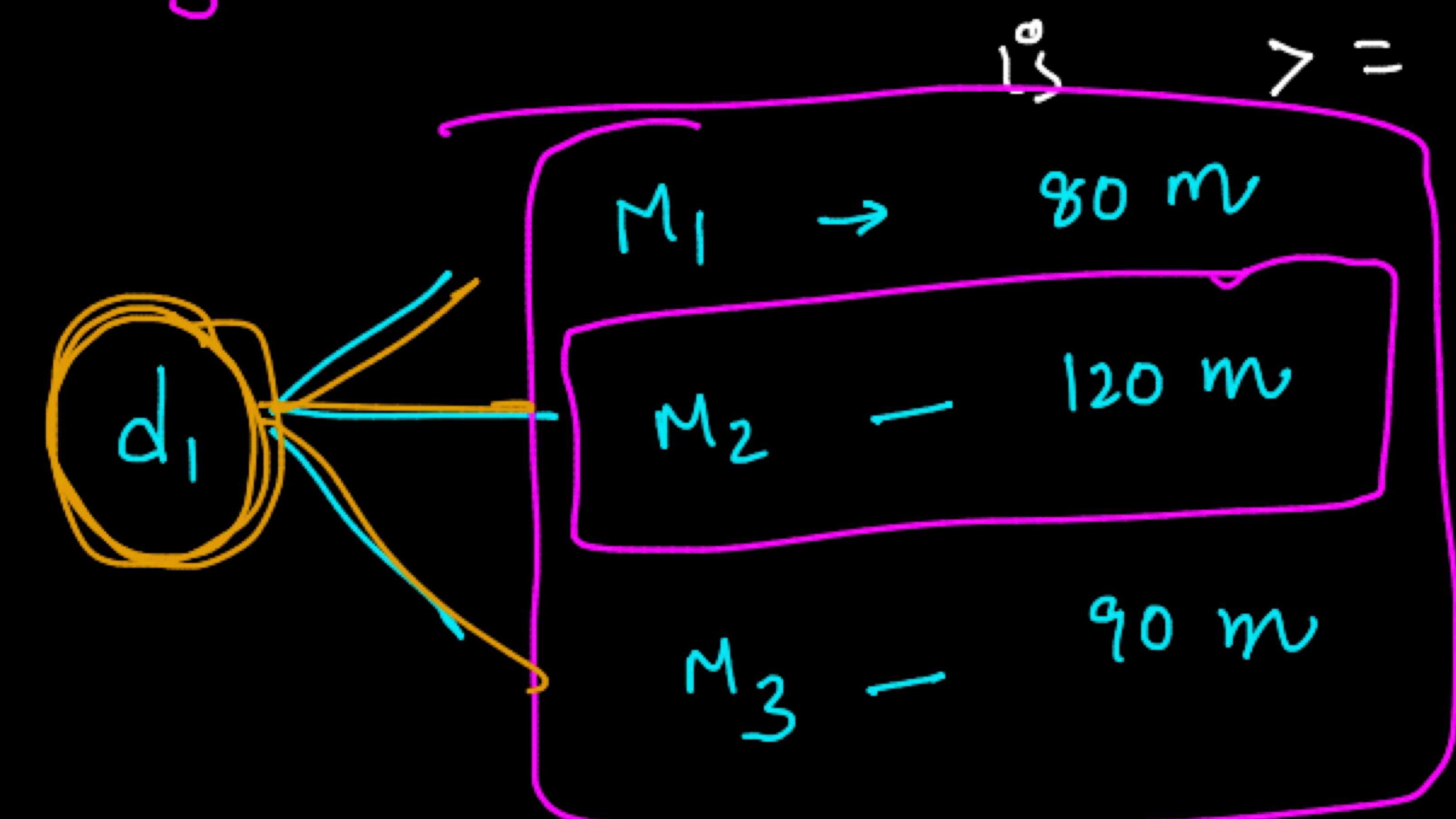


Group based Filtering

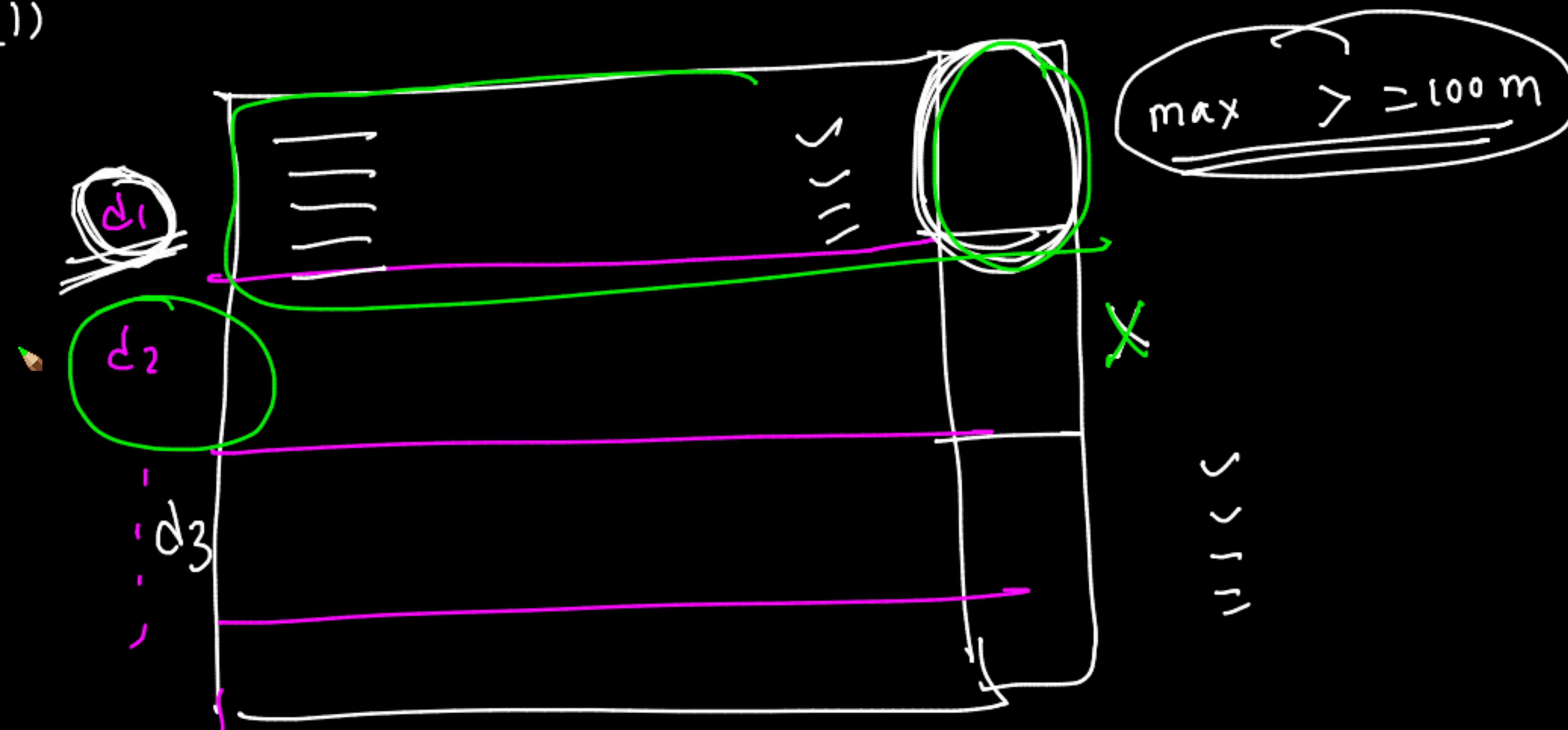


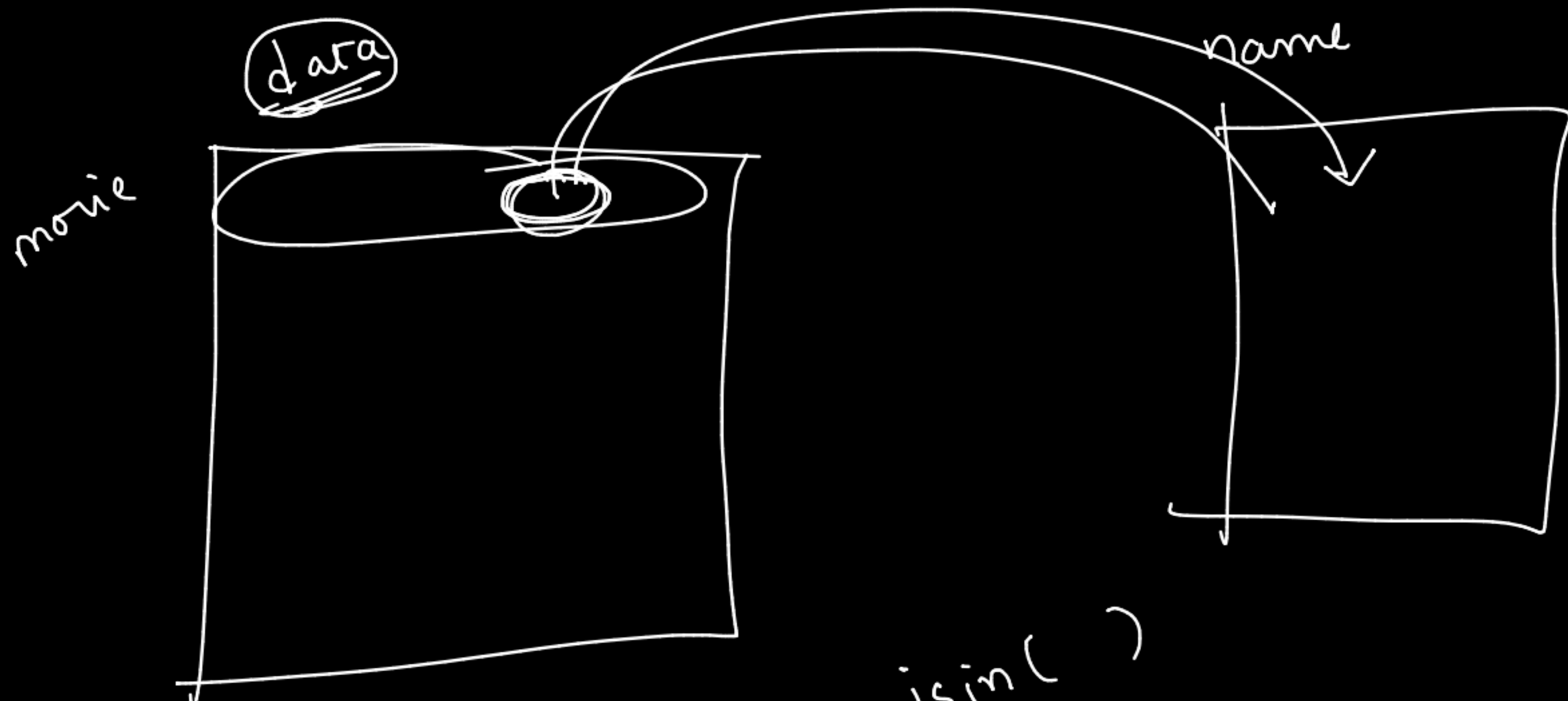
* Find details of the movies by high budget directors ?

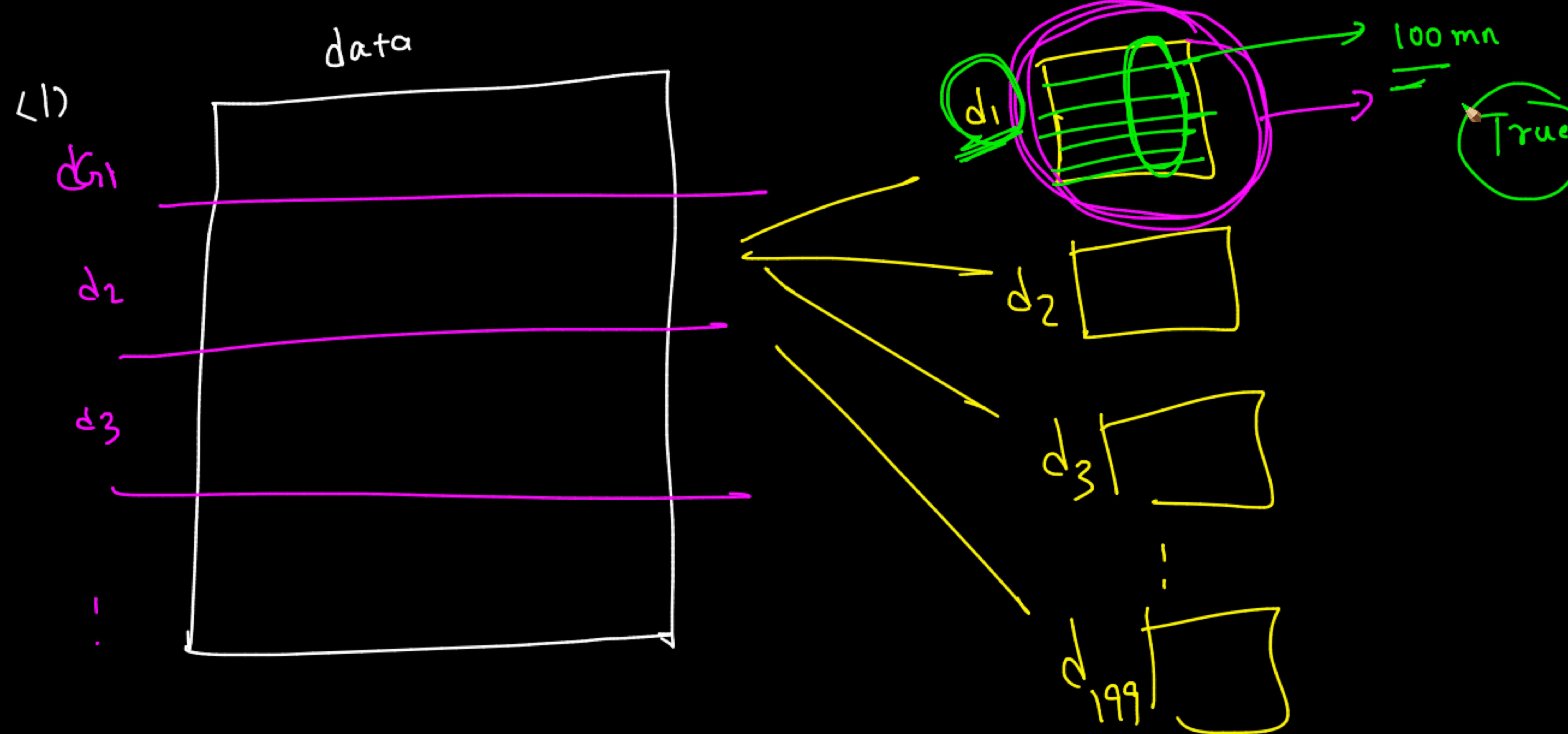
high budget : atleast one movie budget



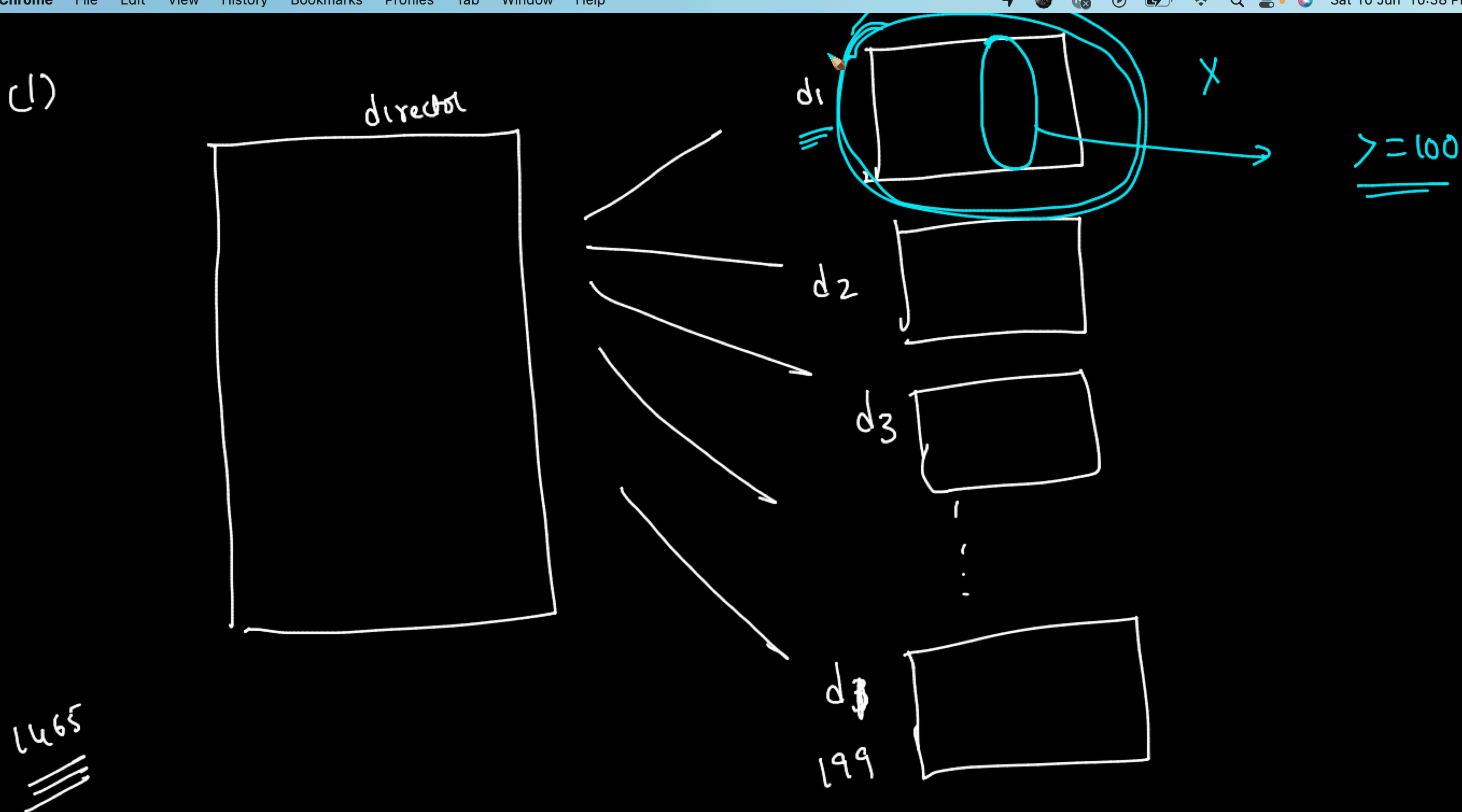
(1)



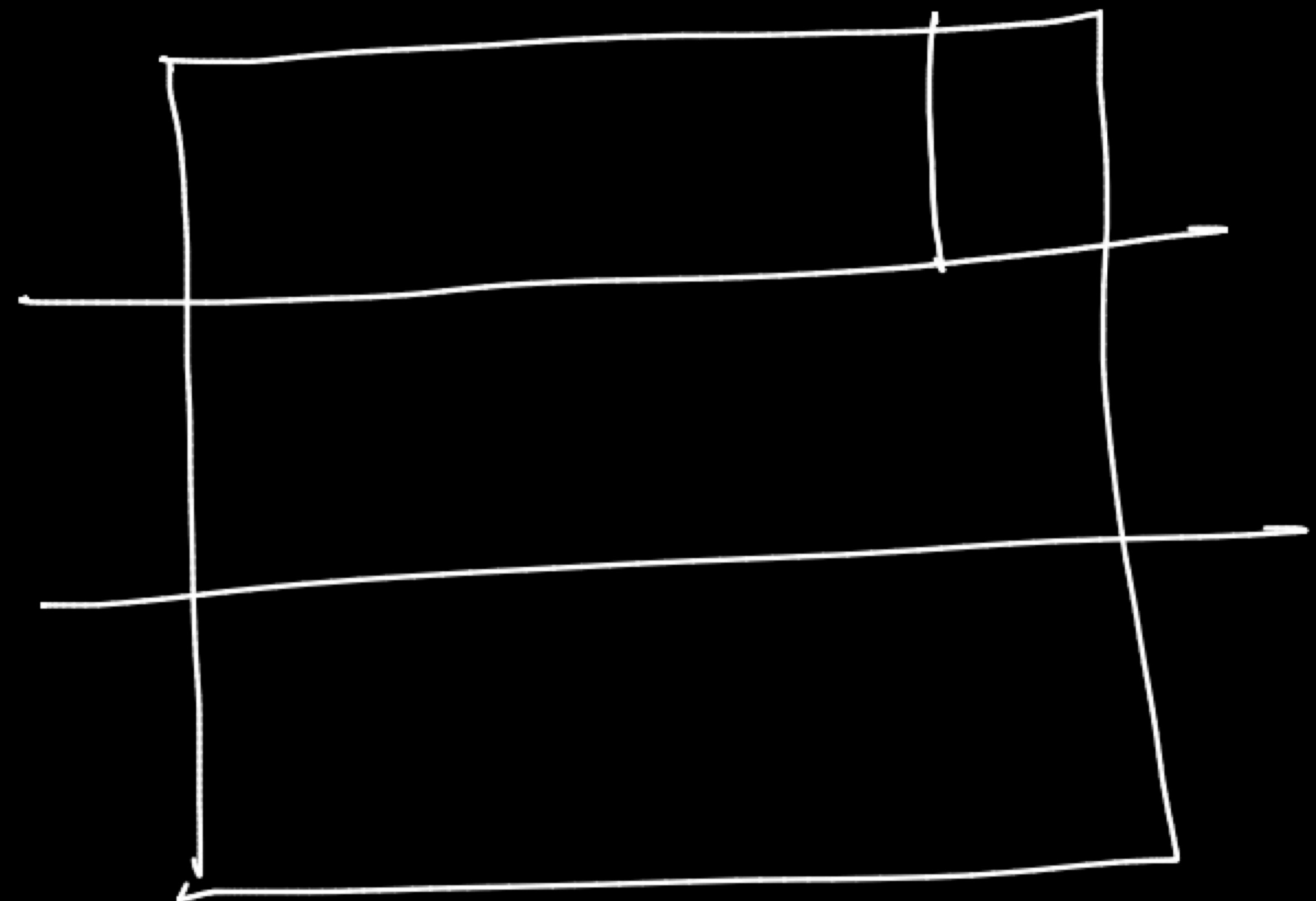




c)



Group based Apply

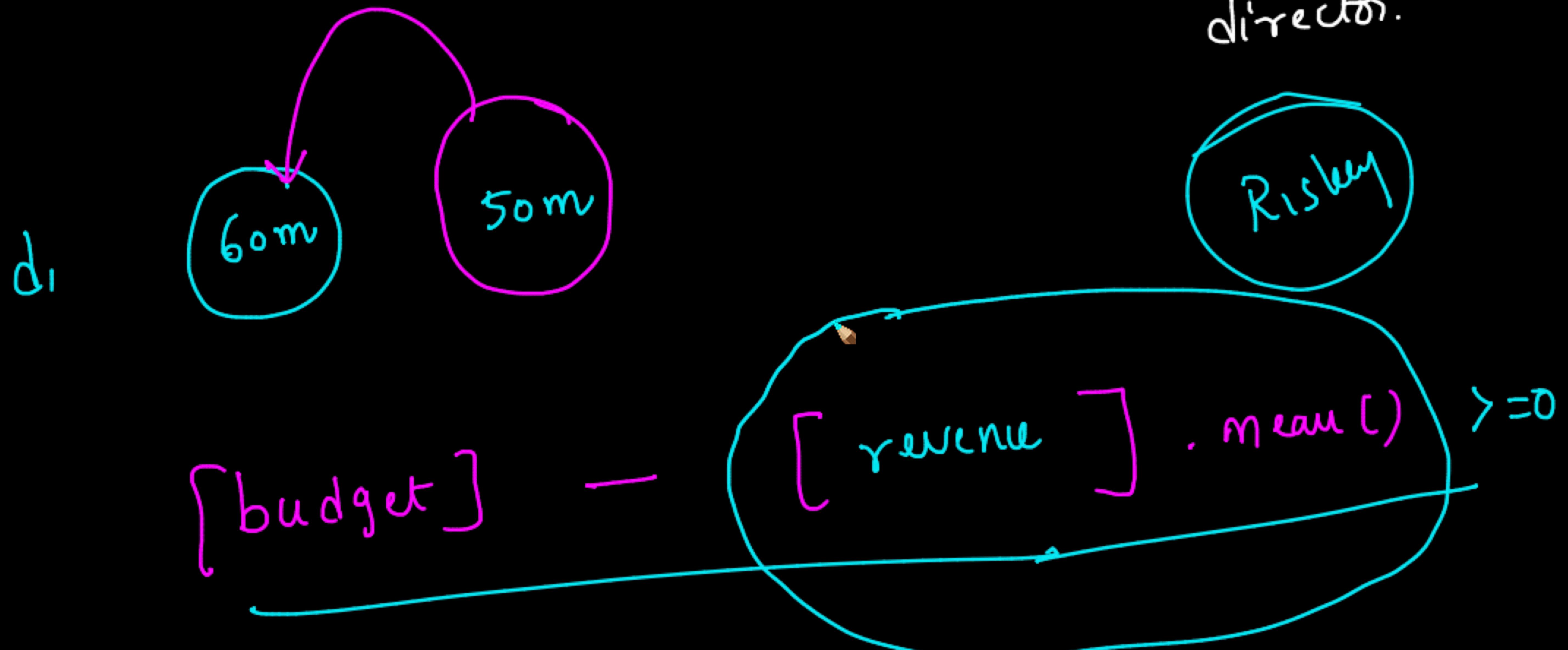


Apply

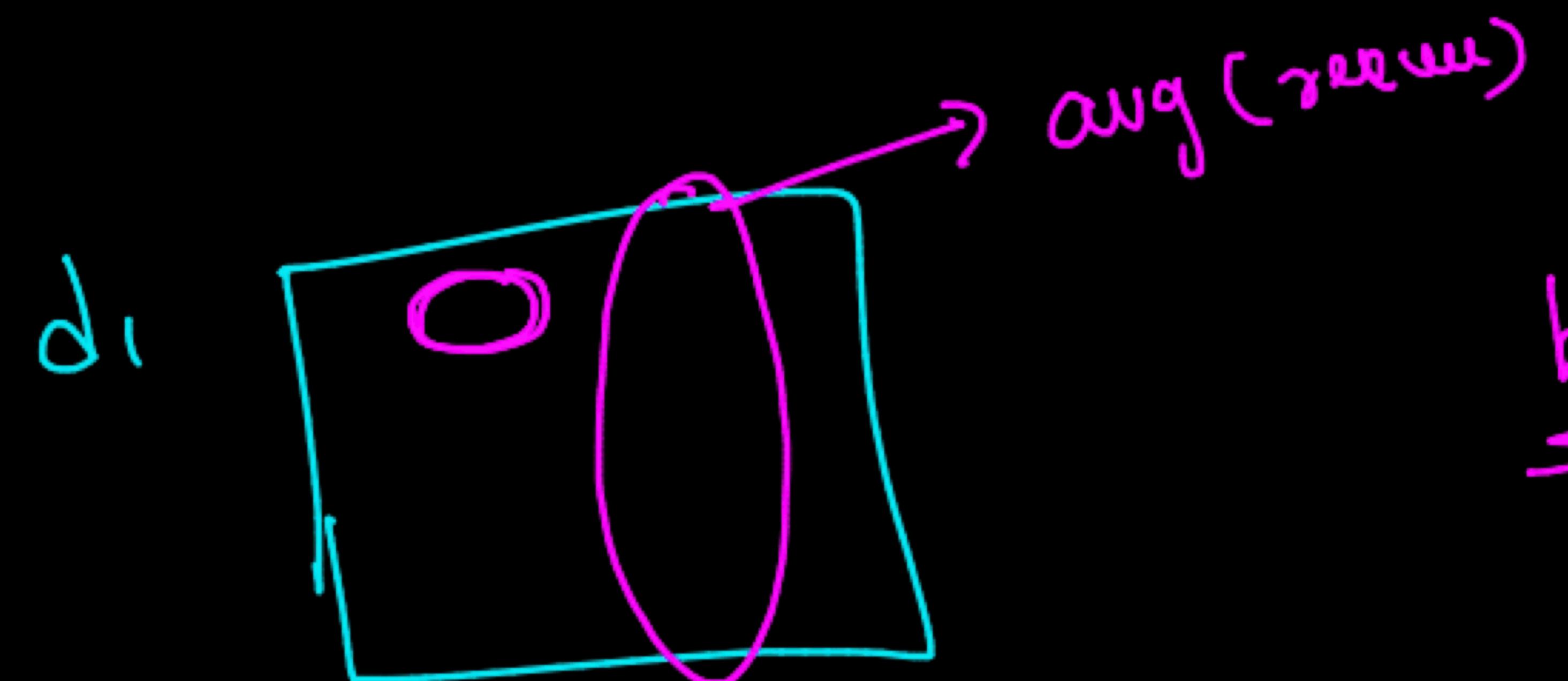
* is movie risky ?



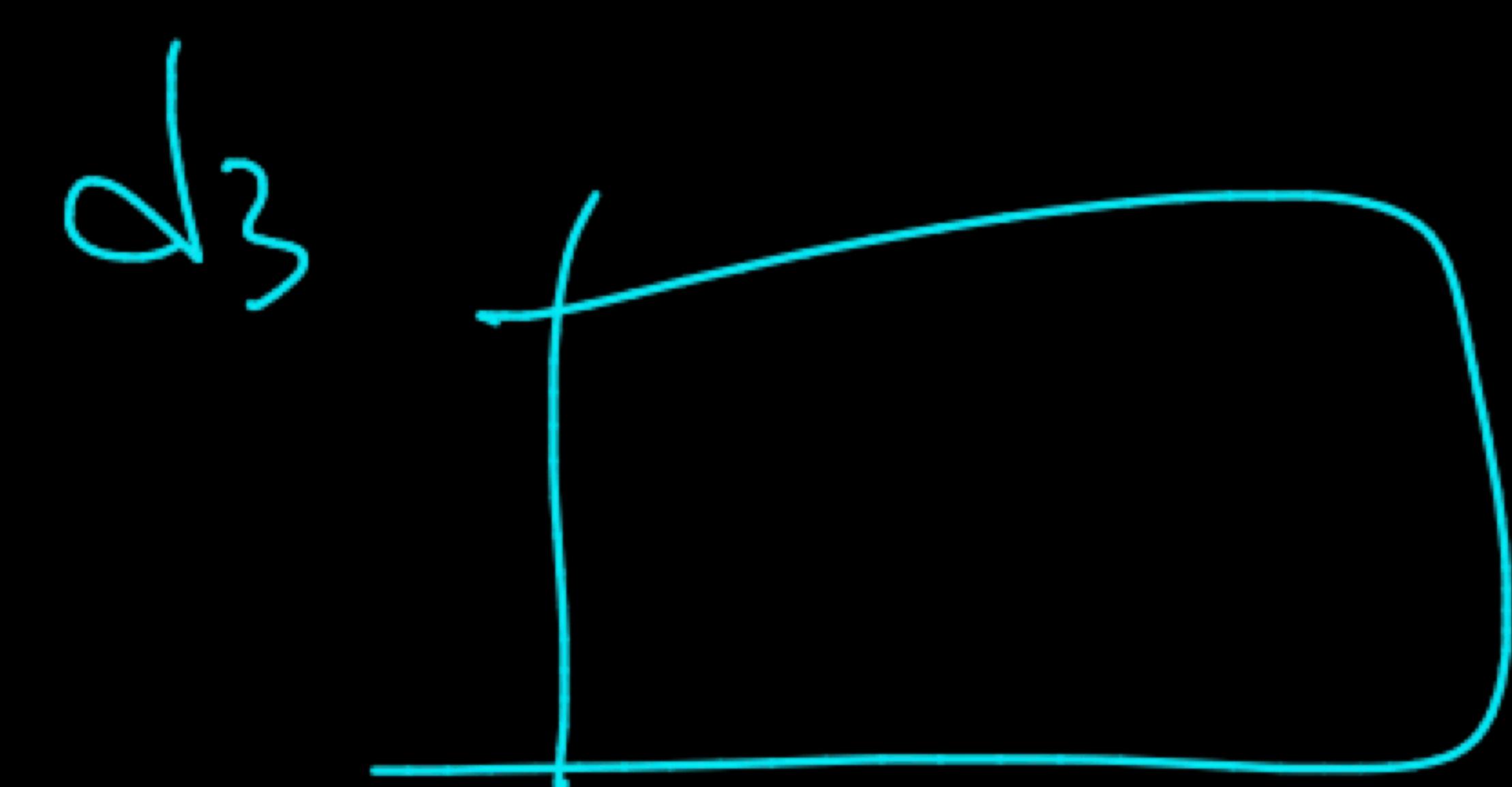
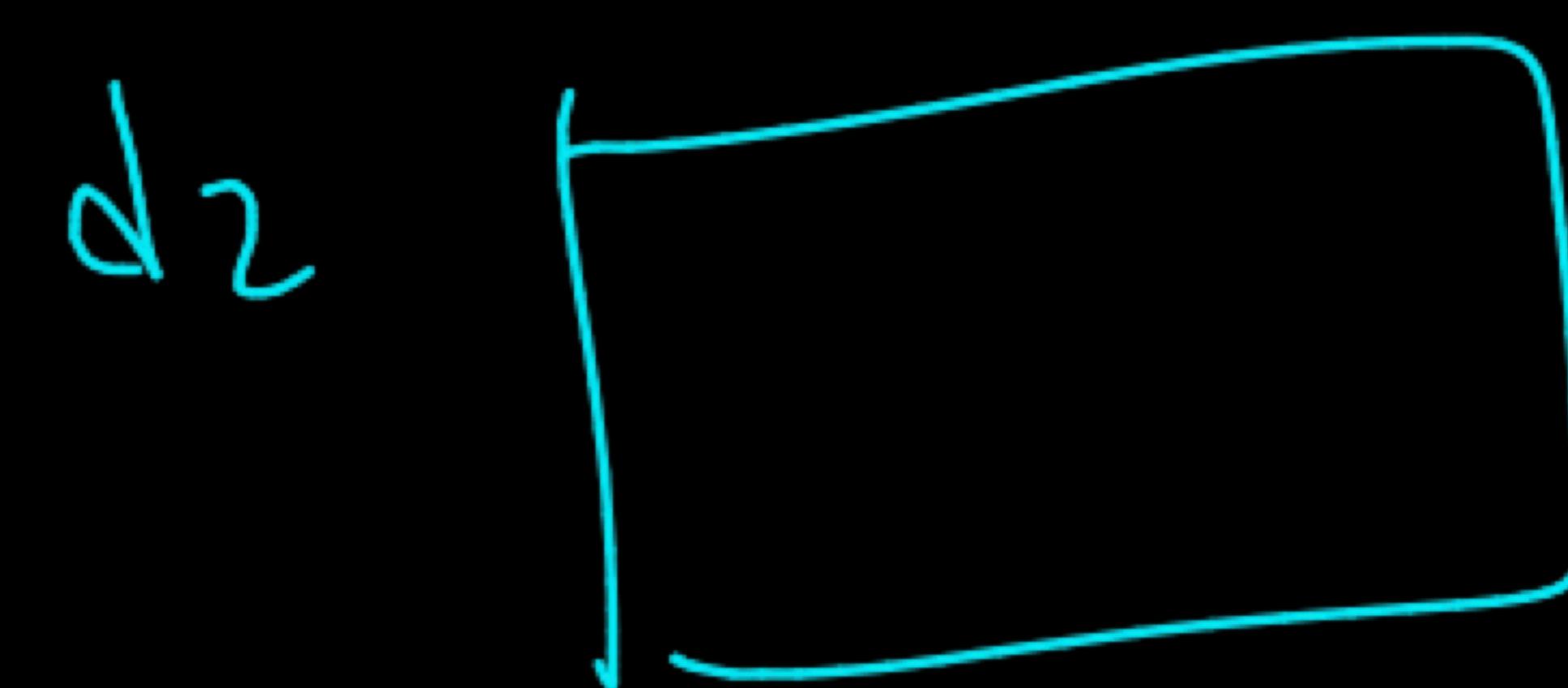
budget is higher than overall revenue of a director.



d)



budget — avg (rem) >= 0



Multi Indexing

Most Productive director



Q no of movies

Actor

AKSHY

u-5

100 CR

