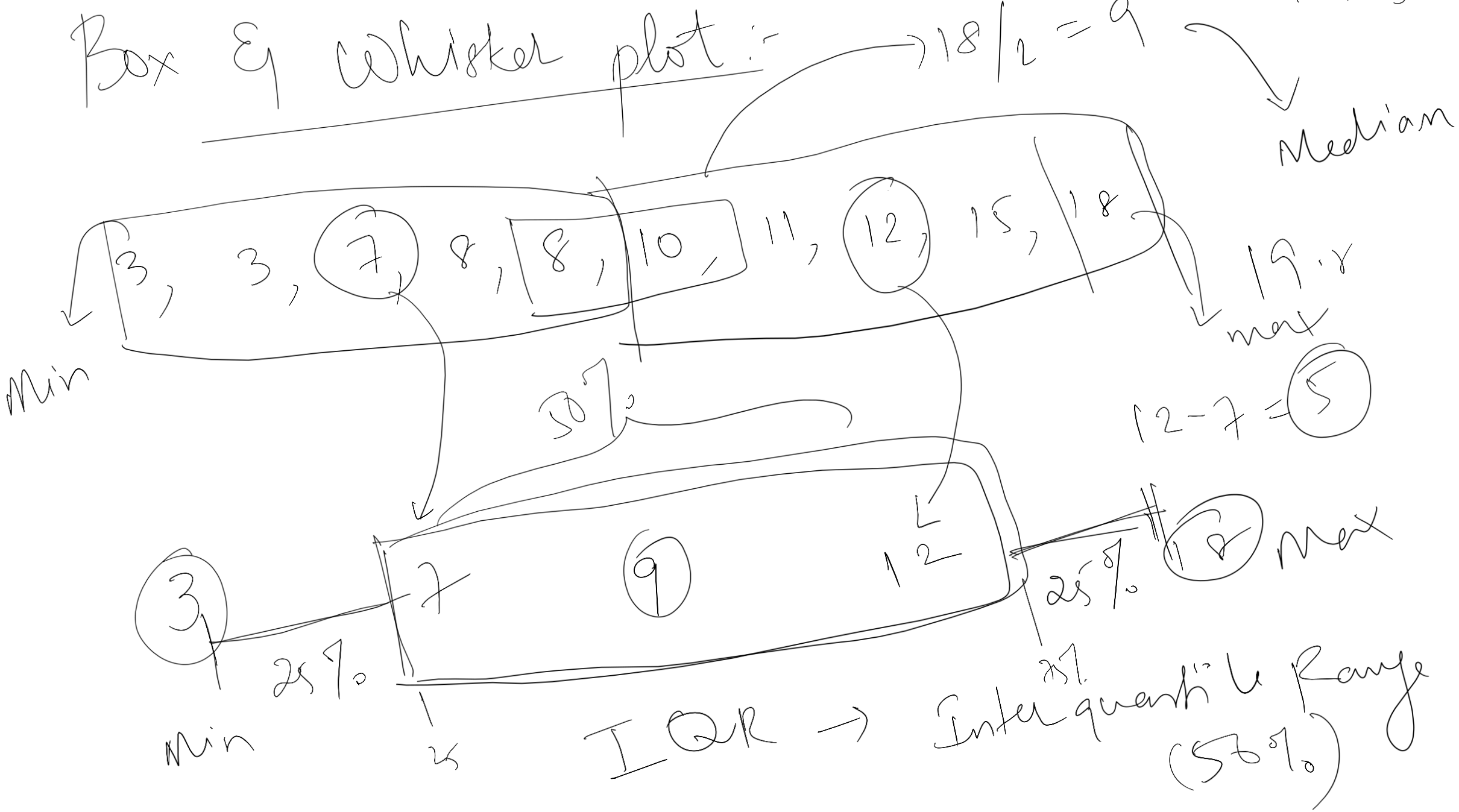
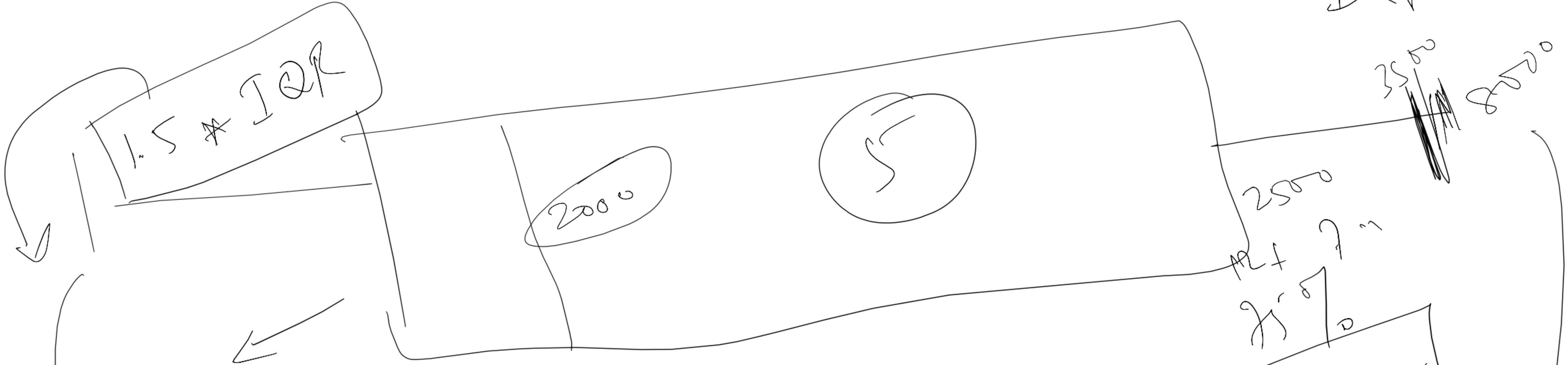
12/15

Box & Whisker plot :-



Median = 2000

75% − 25% = IQR



Outlier

25%

5 × 1.5 = 7.5

IQR (50%)

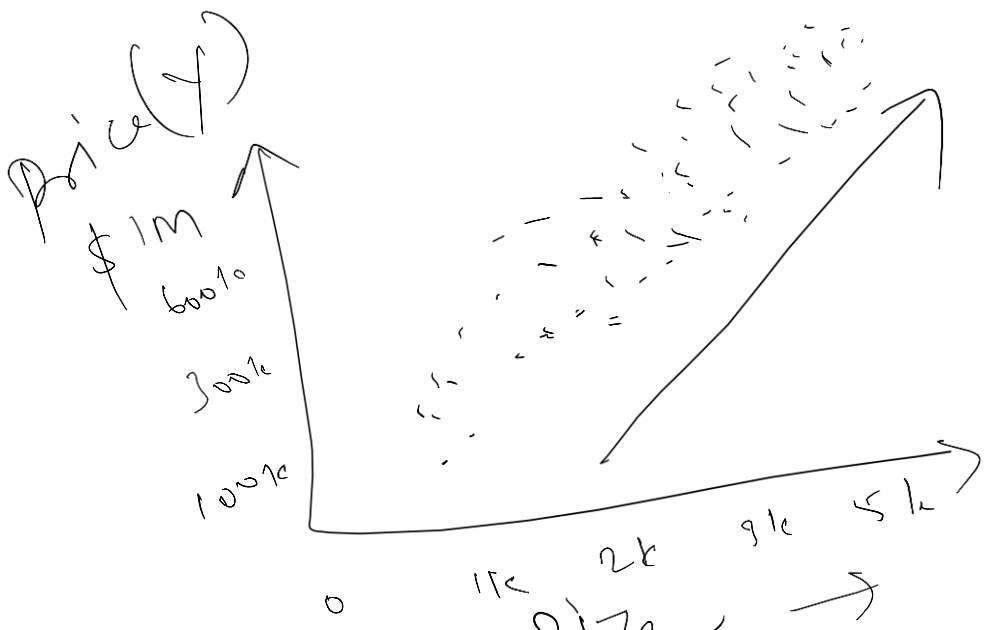
12 + 7.5 = 19.5

7 + 7.5

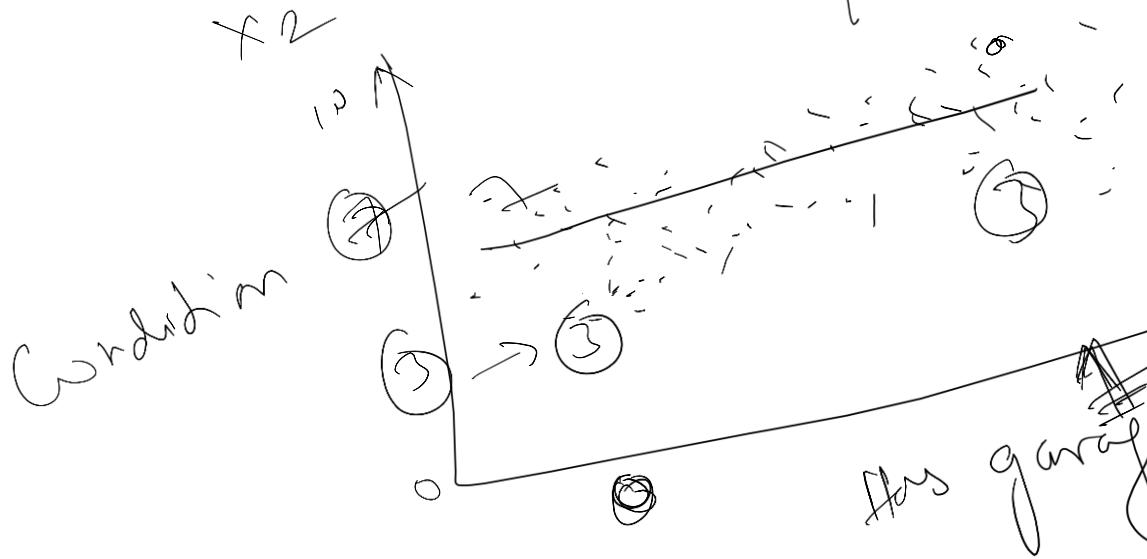
Outlier

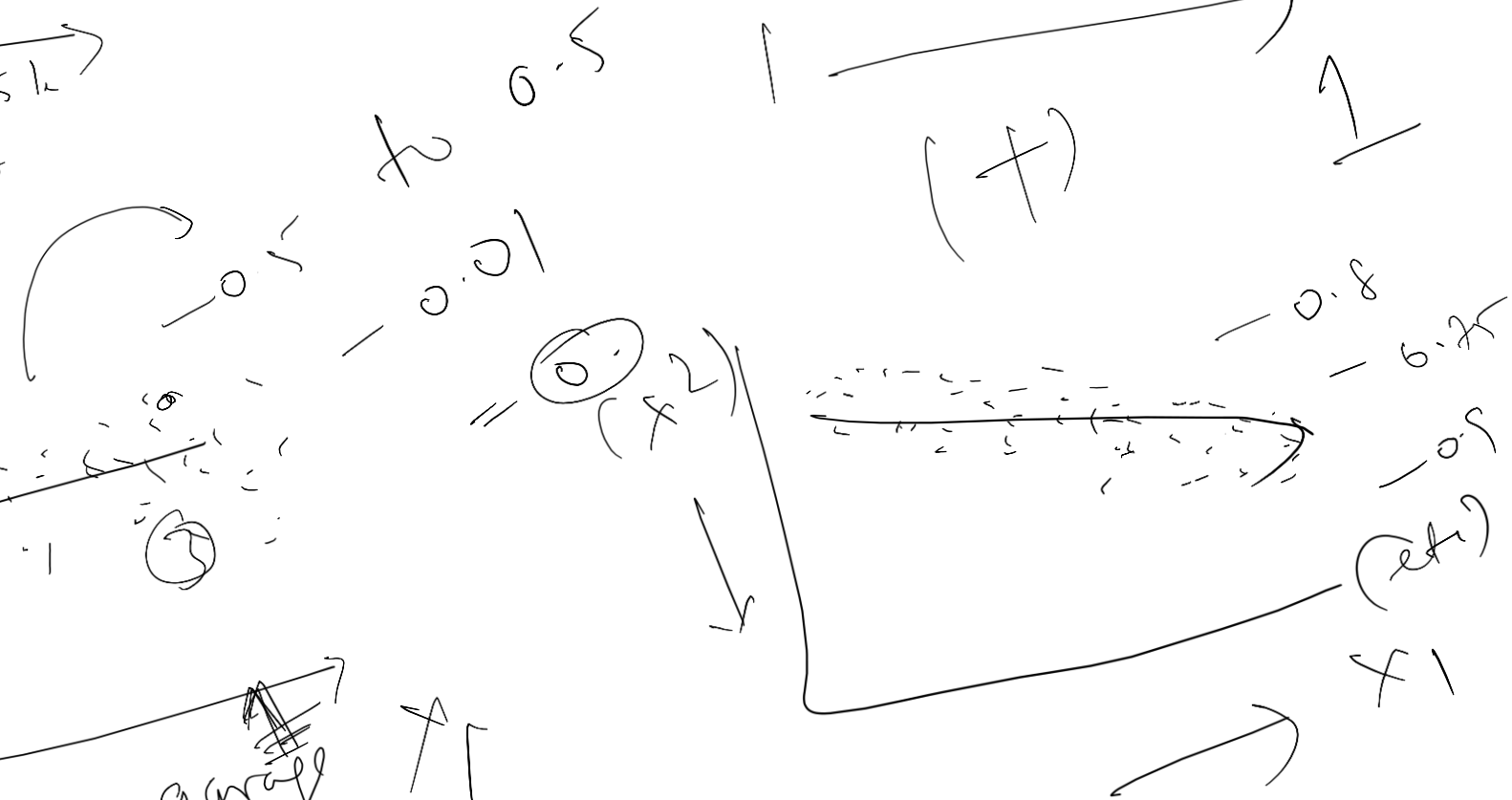
-0.5

2 3 4 5 6 7



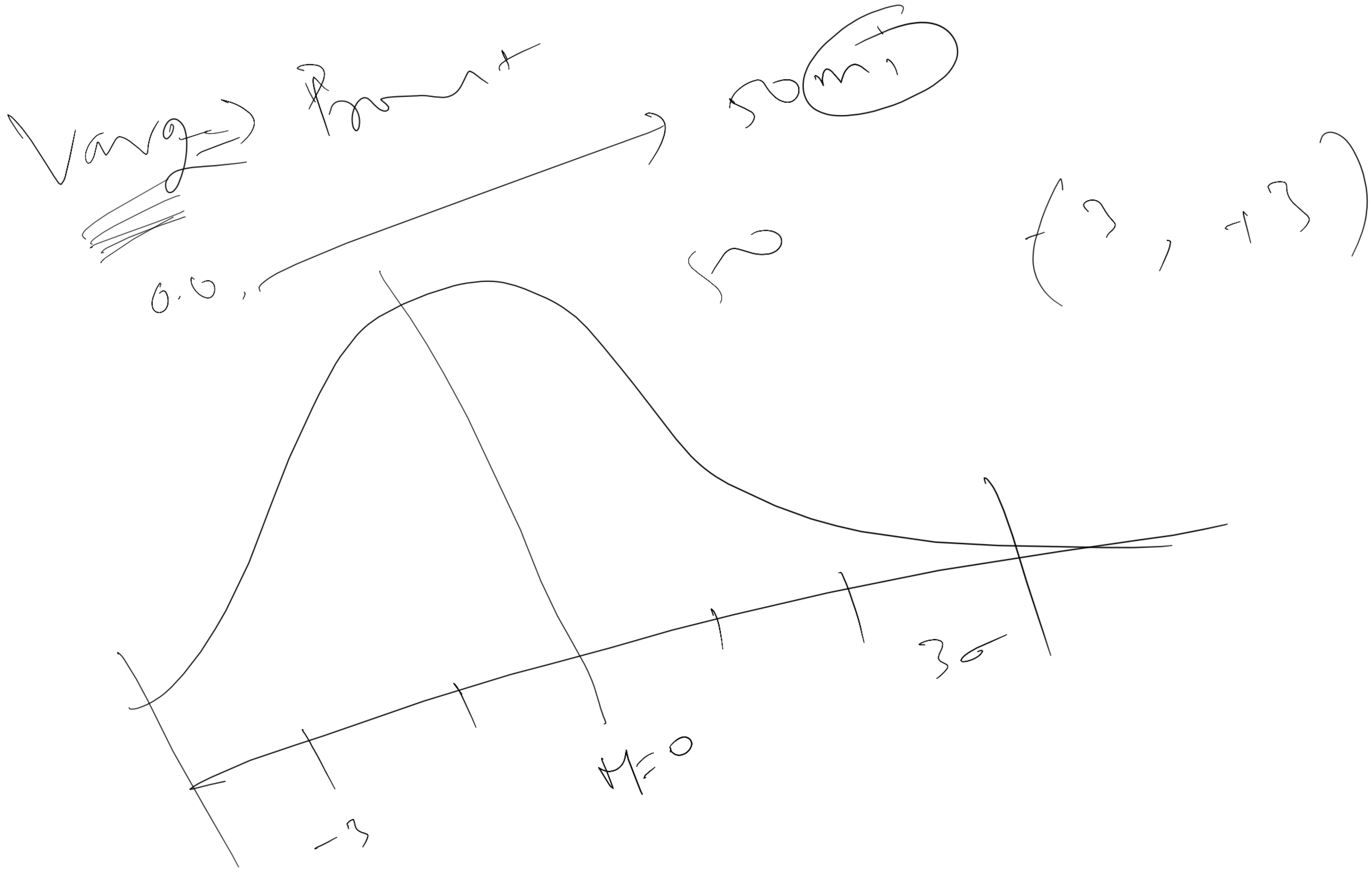
Corr b/w size & price = 0.95

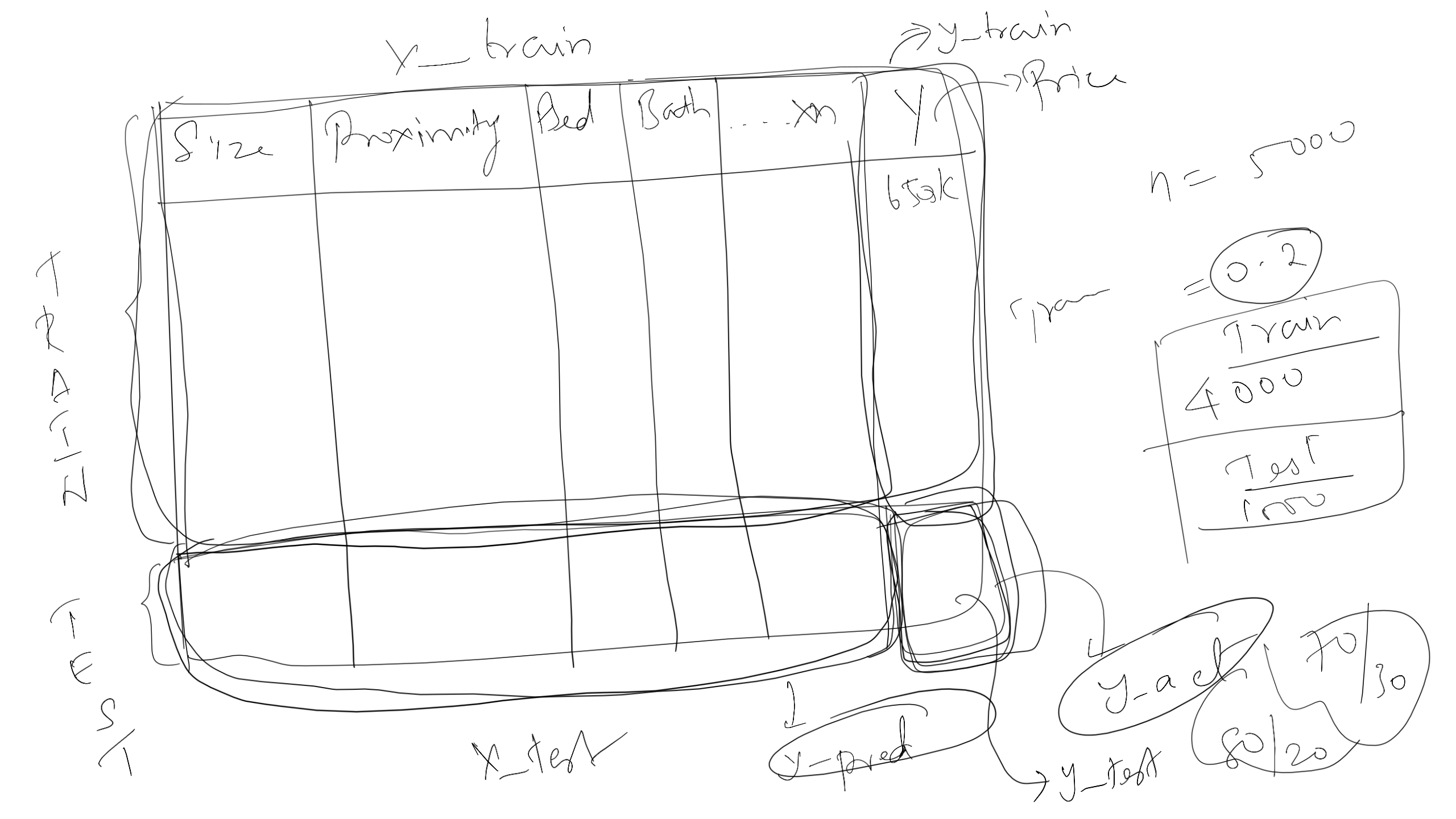




(Let's get back at 7:46 pm IST)

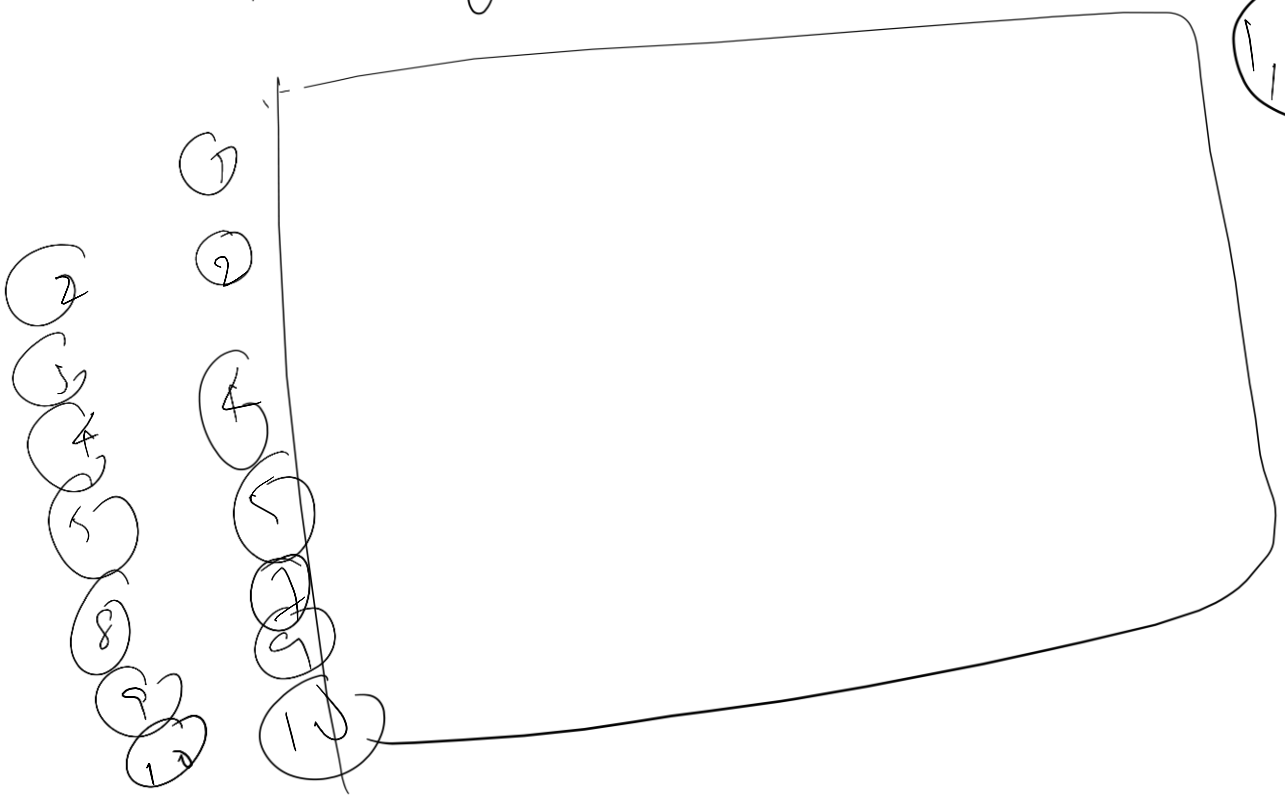

Var 1
Size
Sq ft
200
250
300
-5
-3
2σ
-1σ
μ=0
1σ
2σ
3σ


0.0
-3
μ=0
3σ


X_train
y_train
Price
Size
Proximity
Bed
Bath
...
Xn
Y
65k
n = 5000
TRAIN
TEST
r_train = 0.2
Train
4000
Test
1000
y_act
70/30
60/20
y_pred
X_test
y_test

random_state = 99

↳ The model has its own calculation in picking up the rows for train & test split



n = 10

(0.3)

1, 2, 4, 5, 7, 9, 10

70

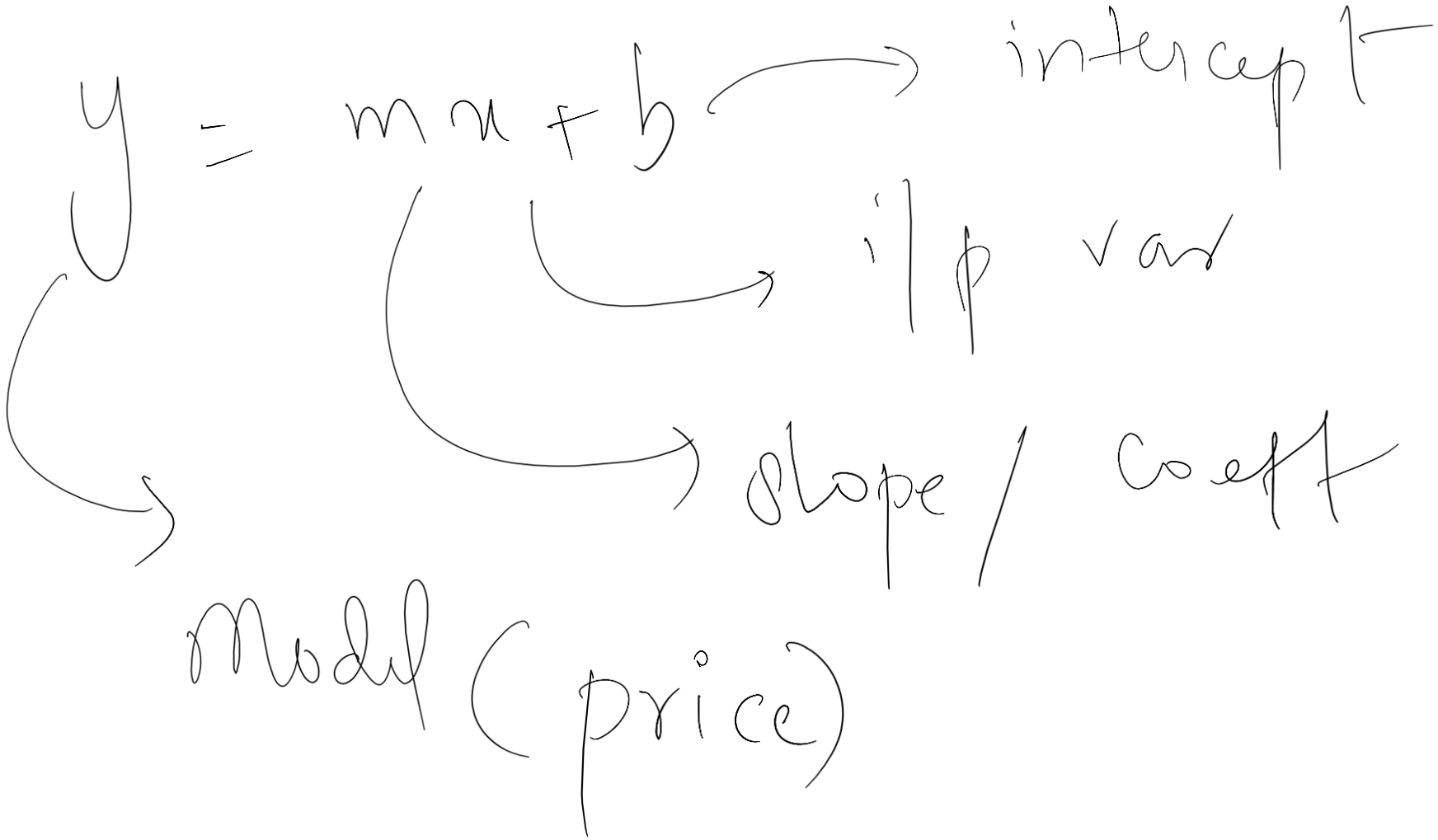
Train

7

Test

3

3, 6, 8


y = mx + b
intercept
i/p var
slope / coeff
model (price)

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 \ldots\ldots m_n x_n + b$$

$y \rightarrow$ Price

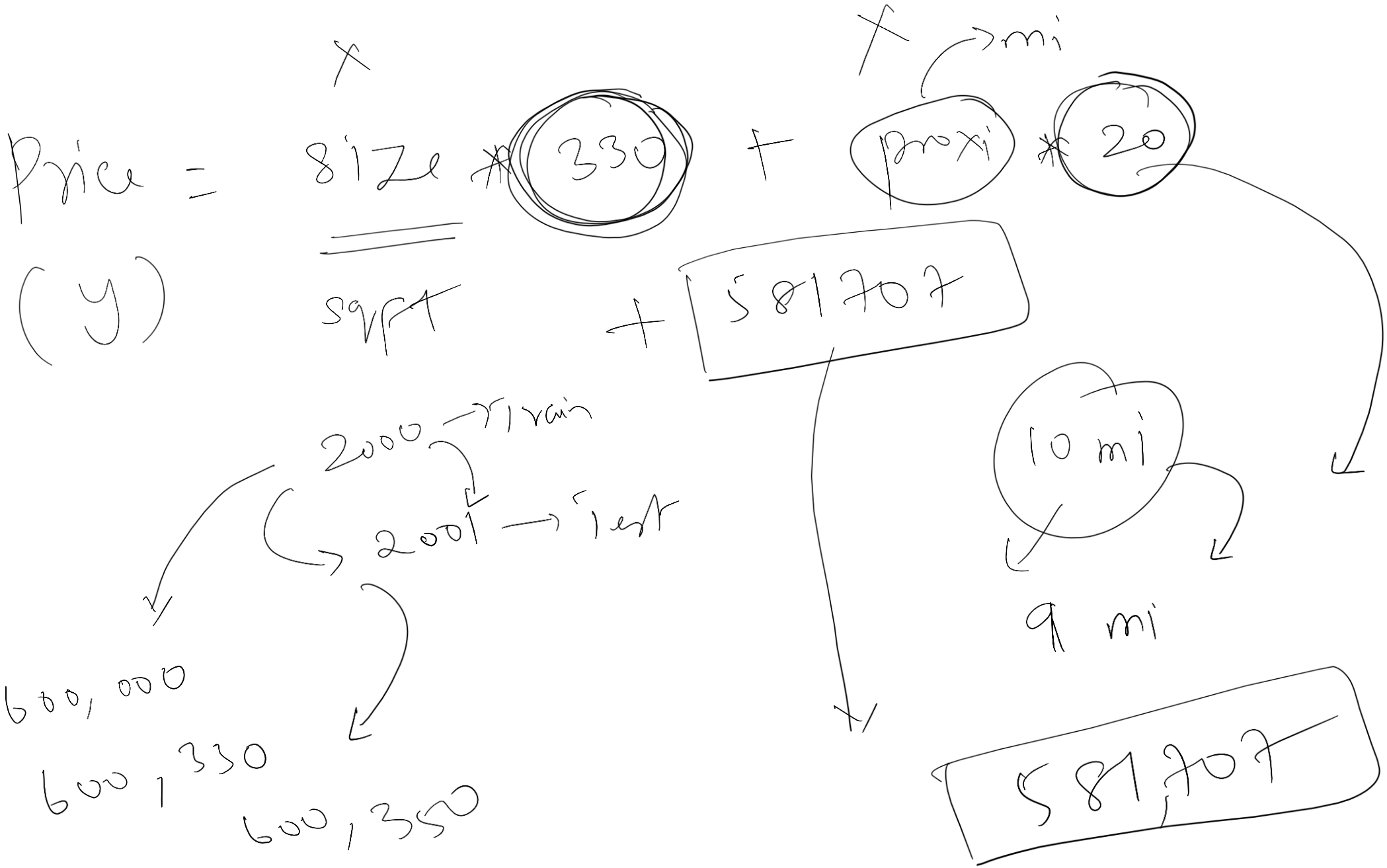
$x_1, x_2 \ldots \rightarrow \sum_{i}^{n} \rightarrow$ Size, Bed ---- x_i

$b \rightarrow$ Intercept

$m_1, m_2, m_3 \ldots m_n \rightarrow$ "Coeff"

$$\text{Price} = (74553 * \text{Size}) + (31801 * \text{Bedroom}) + (23667 * \text{Bathroom}) + \ldots + (\text{Neighborhood} * -0.00012) + 58170$$

Price


Price =
(y)
size
sqft
330
+
proxi
mi
20
+
581707
2000 → train
2001 → test
600,000
600,330
600,350
10 mi
9 mi
581707

$$\frac{\sum (y - \hat{y})^2}{n} = MSE$$

100,201
90,005
60,006

$$\frac{\sum |y - \hat{y}|}{n} = MAE$$

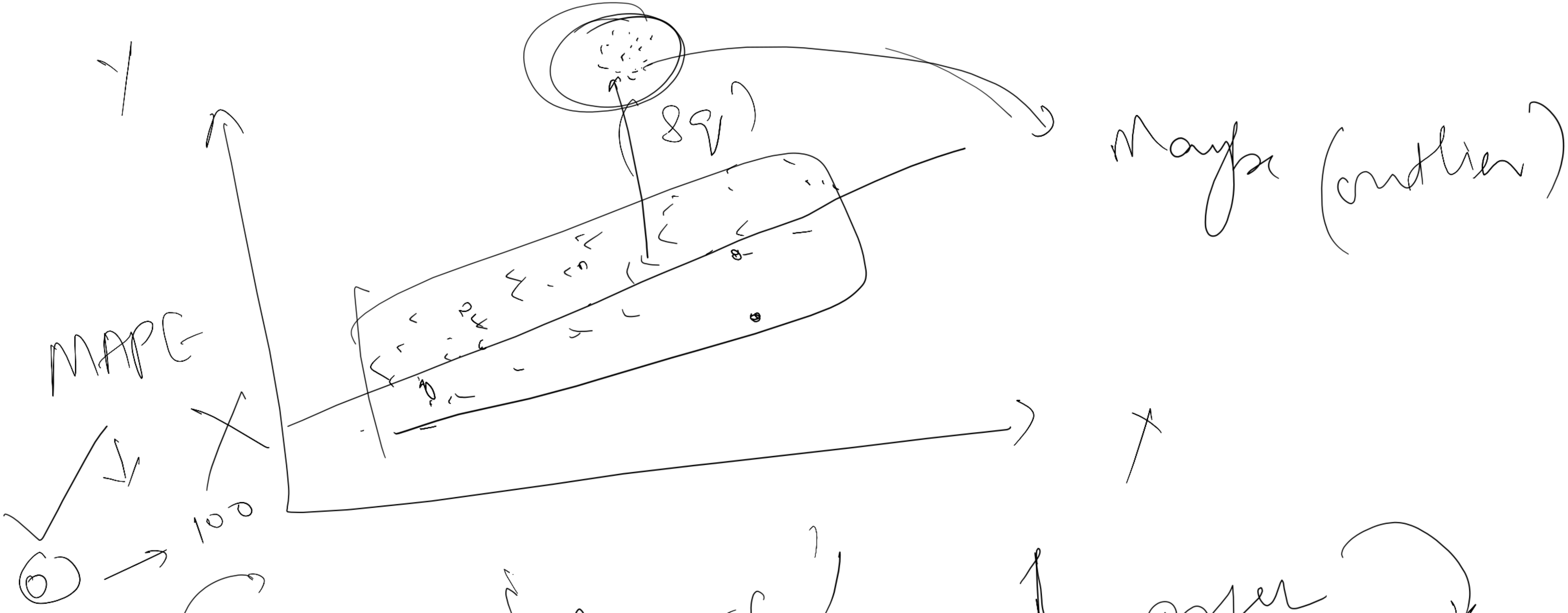
MAPE → <15
<20

0 — 100

MAPE →

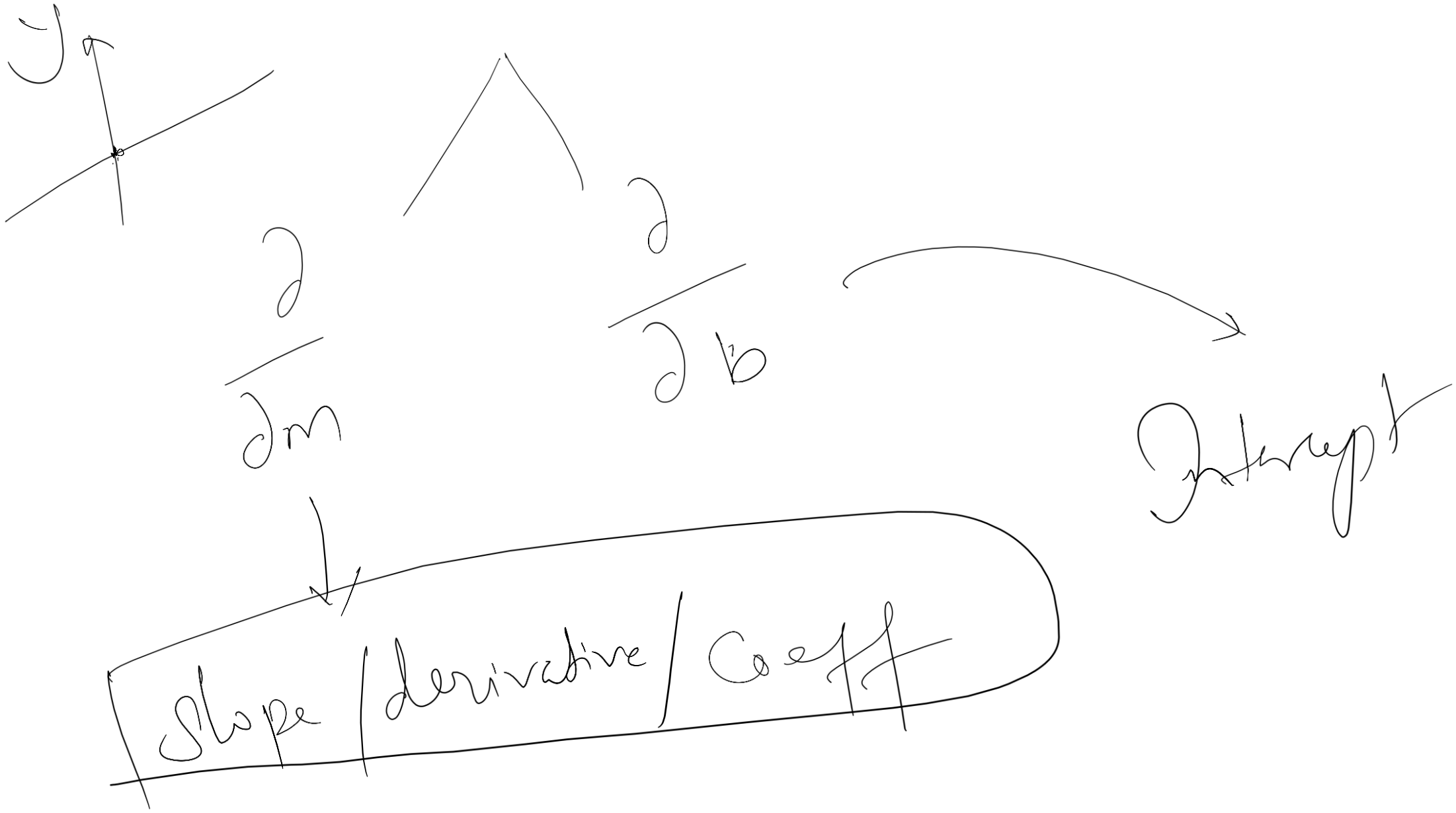
$$\frac{\sum (|y - \hat{y}| / y)}{n} * 100$$

MAPE → 20,8



MSE → 'Magnify' = Larger 6 digit

MAE → | | → < digit


y
∂/∂m
∂/∂b
slope / derivative / coeff
Intercept