



# UNIT 3

# LINEAR REGRESSION

Content Curated by Pollux M. Rey

# FOR THIS UNIT...

01

What is Linear Regression?

02

How does it work?

03

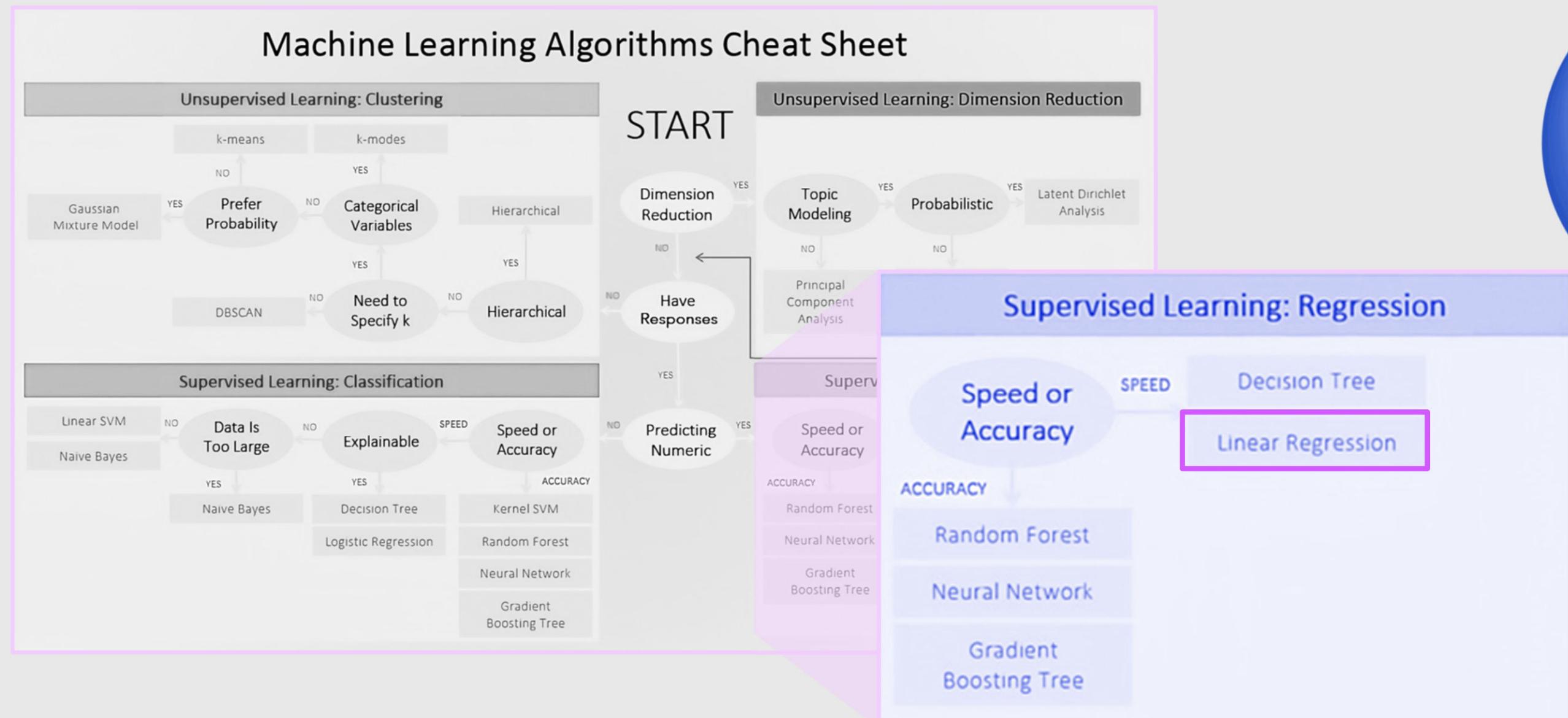
Performance Metrics

04

Lab Exercise

Content Curated by Pollux M. Rey

# Linear Regression



Content Curated by Pollux M. Rev

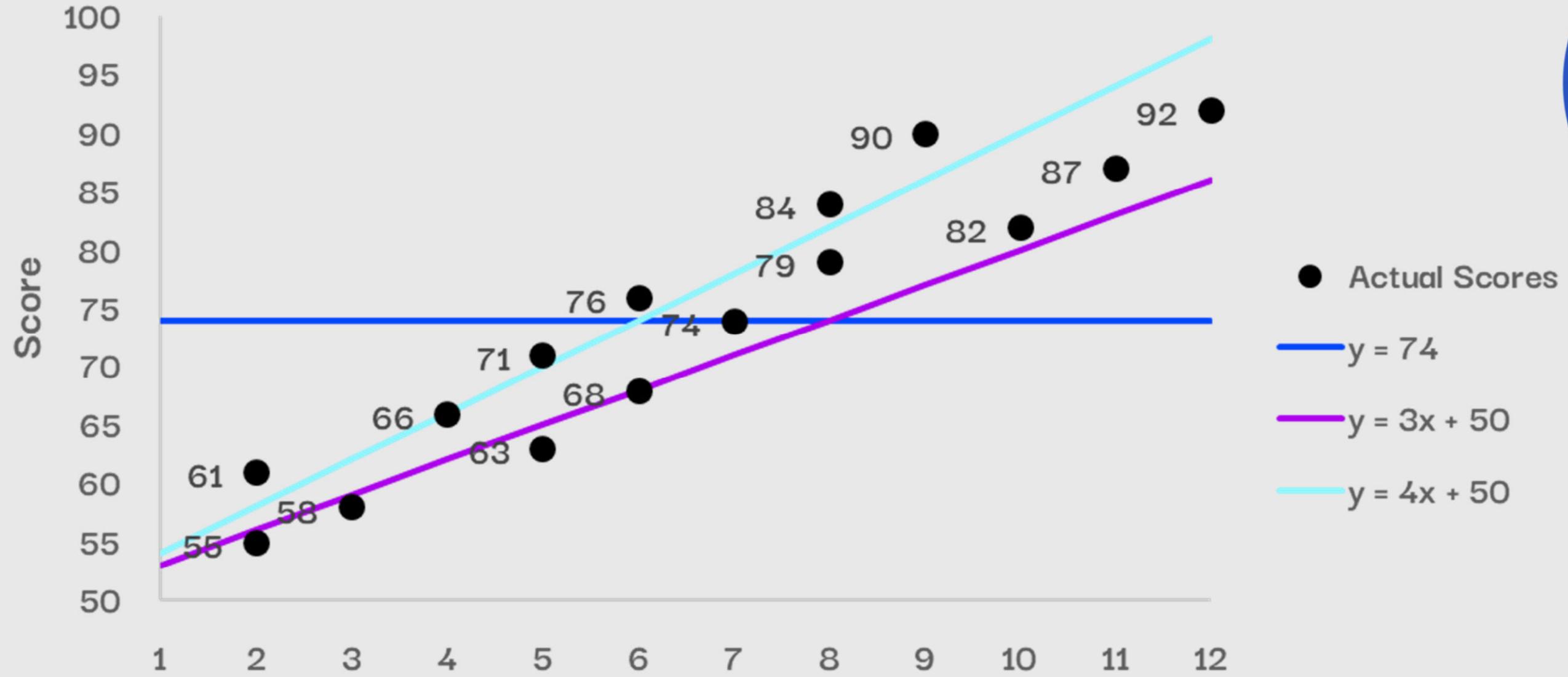
<https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use>

# Linear Regression

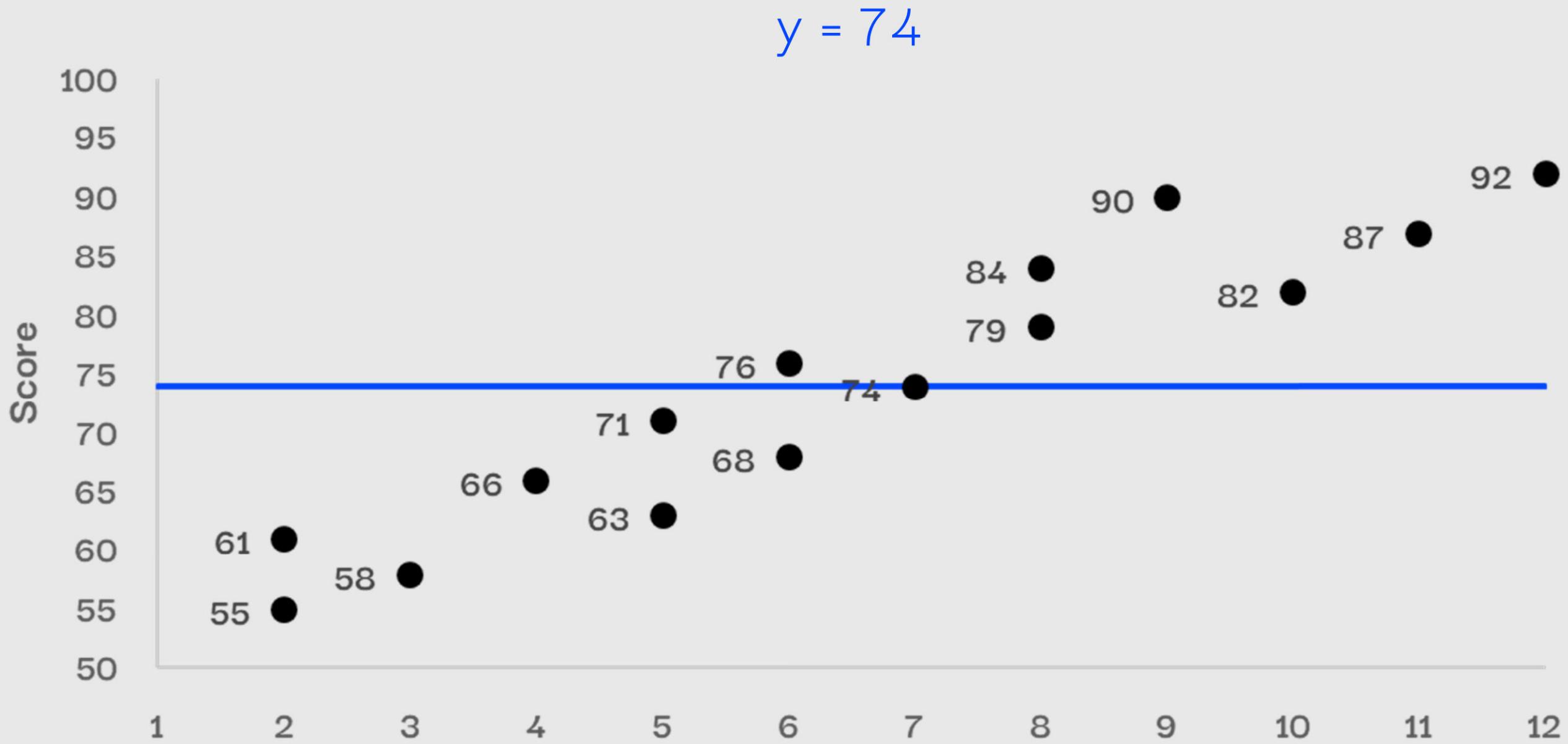
A supervised learning algorithm that models the relationship between features and a label to predict a continuous value by finding the best-fit line (or surface).

# Linear Regression - Activity

Which of these lines best represents the data?

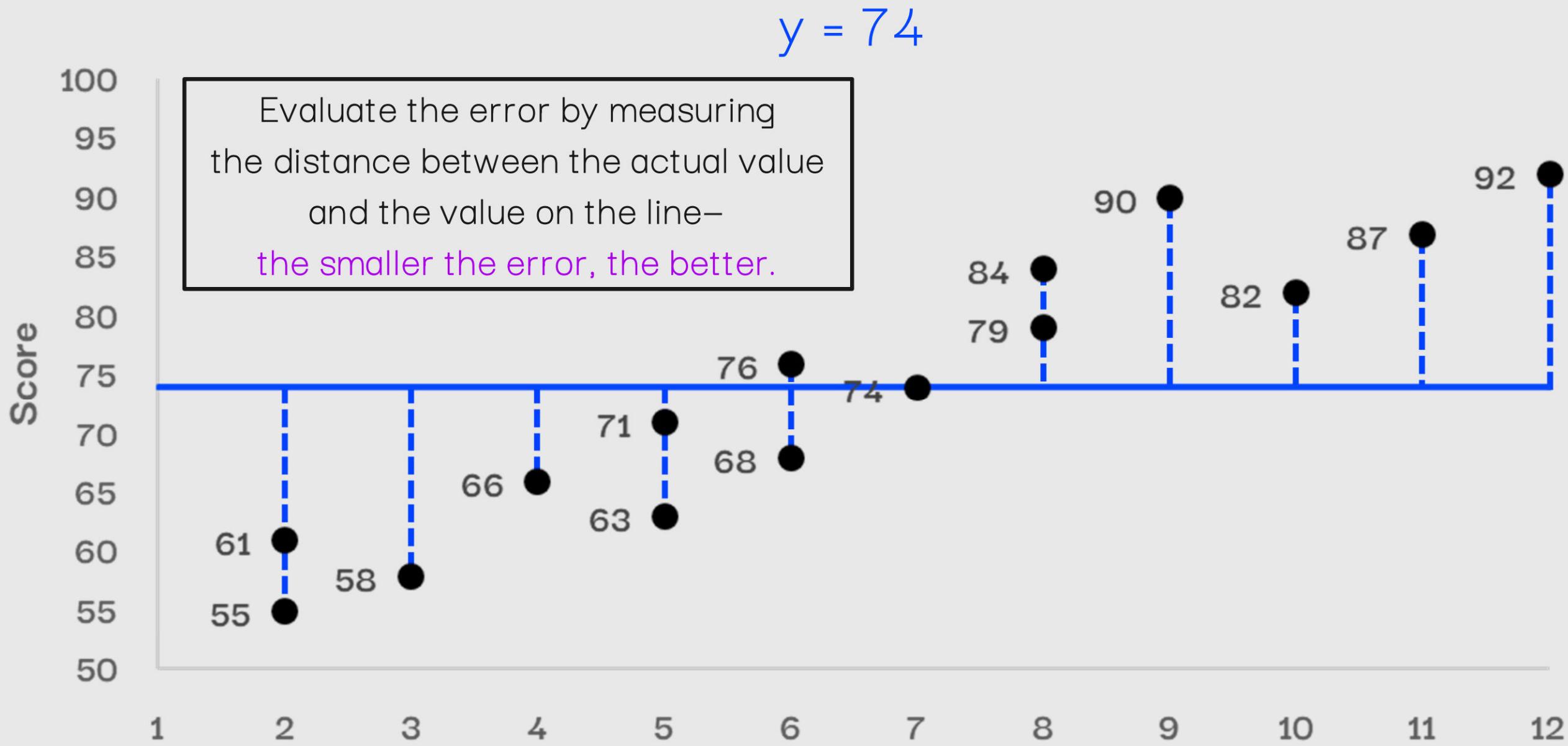


# Linear Regression - Activity

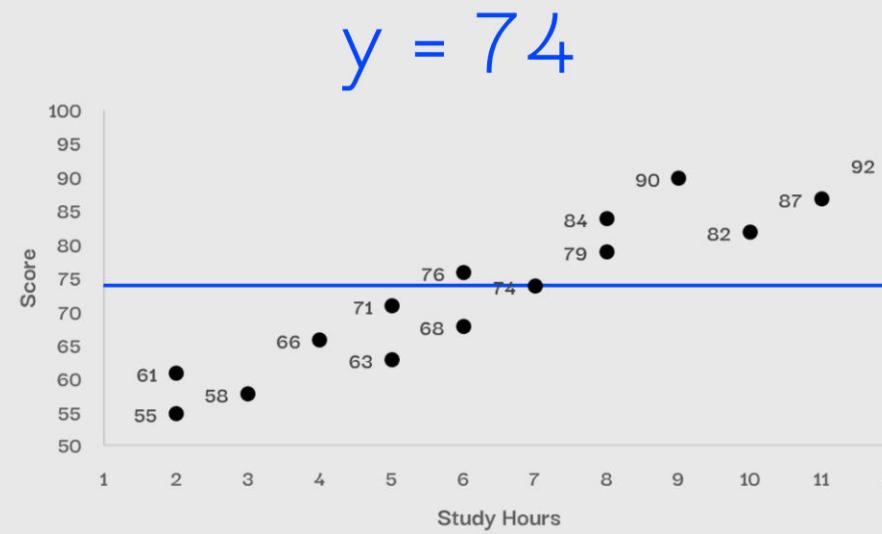


Content Curated by Pollux M. Rey

# Linear Regression - Activity

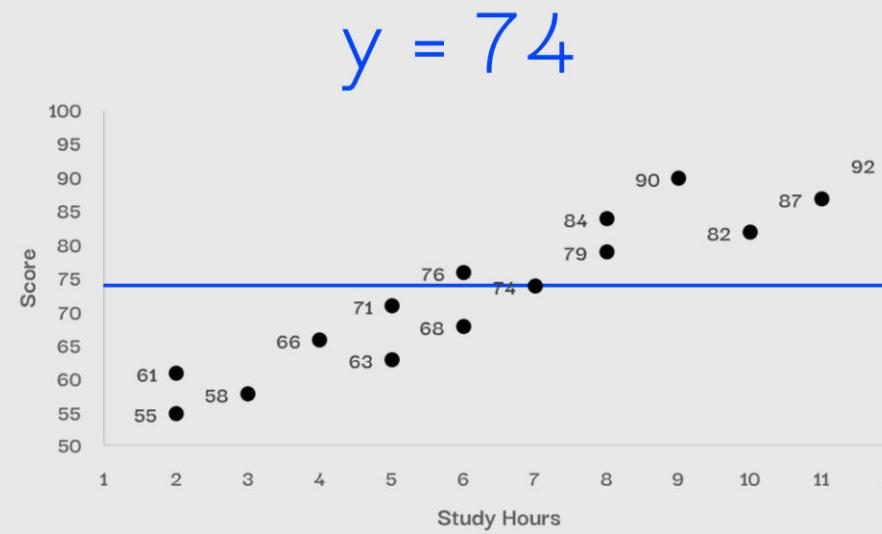


# Linear Regression - Activity



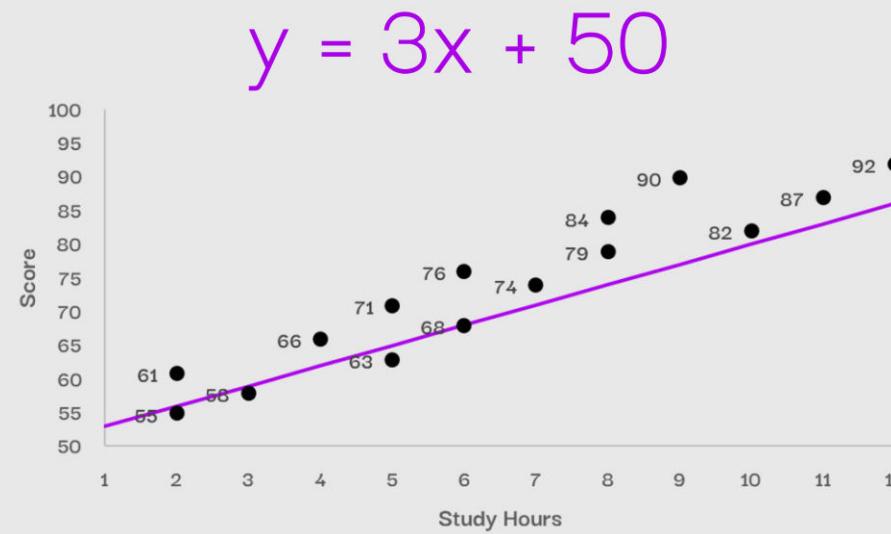
x	y (Actual value)	$y' = 74$	$ y - y' $	$( y - y' )^2$
2	55	74	19	361
2	61	74	13	169
3	58	74	16	256
4	66	74	8	64
5	63	74	11	121
5	71	74	3	9
6	68	74	6	36
6	76	74	2	4
7	74	74	0	0
8	79	74	5	25
8	84	74	10	100
9	90	74	16	256
10	82	74	8	64
11	87	74	13	169
12	92	74	18	324
Sum of $( y - y' )^2$				1296
Mean of sum of $( y - y' )^2$				324

# Linear Regression - Activity



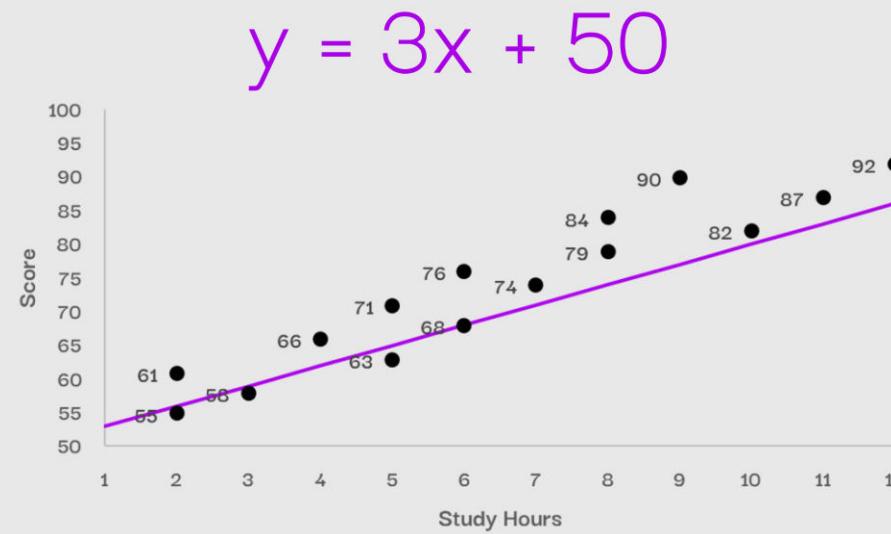
x	y (Actual value)	$y' = 74$	$ y - y' $	$( y - y' )^2$
2	55	74	19	361
2	61	74	13	169
3	58	74	16	256
4	66	74	8	64
5	63	74	3	9
5	71	74	11	121
6	68	74	2	4
6	76	74	6	36
7	74	74	0	0
8	79	74	5	25
8	84	74	10	100
9	90	74	16	256
10	82	74	8	64
11	87	74	13	169
12	92	74	18	324
Sum of $( y - y' )^2$				1958
Mean of sum of $( y - y' )^2$				130.53

# Linear Regression - Activity



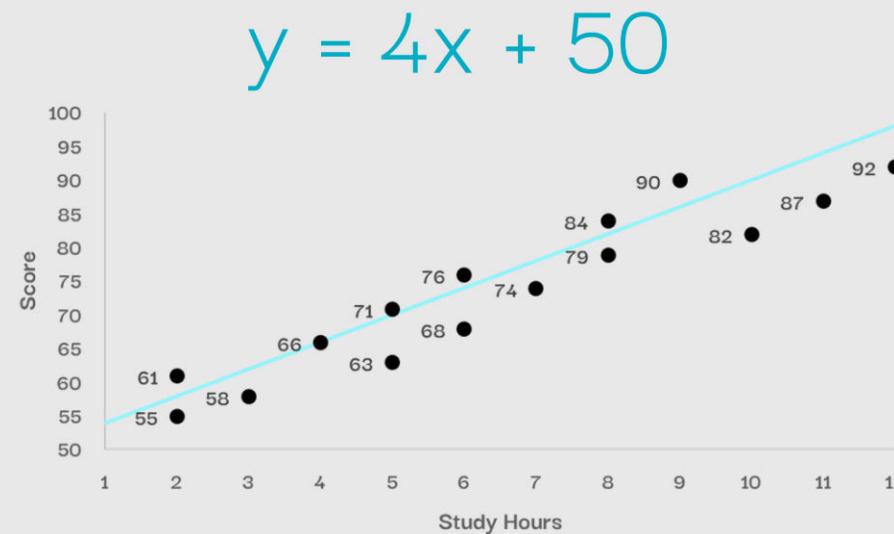
x	y (Actual value)	$y' = 3x + 50$	$ y - y' $	$( y - y' )^2$
2	55	56		
2	61	56		
3	58	59		
4	66	62		
5	63	65		
5	71	65		
6	68	68		
6	76	68		
7	74	71		
8	79	74		
8	84	74		
9	90	77		
10	82	80		
11	87	83		
12	92	86		
Sum of $( y - y' )^2$				1000
Mean of sum of $( y - y' )^2$				250

# Linear Regression - Activity



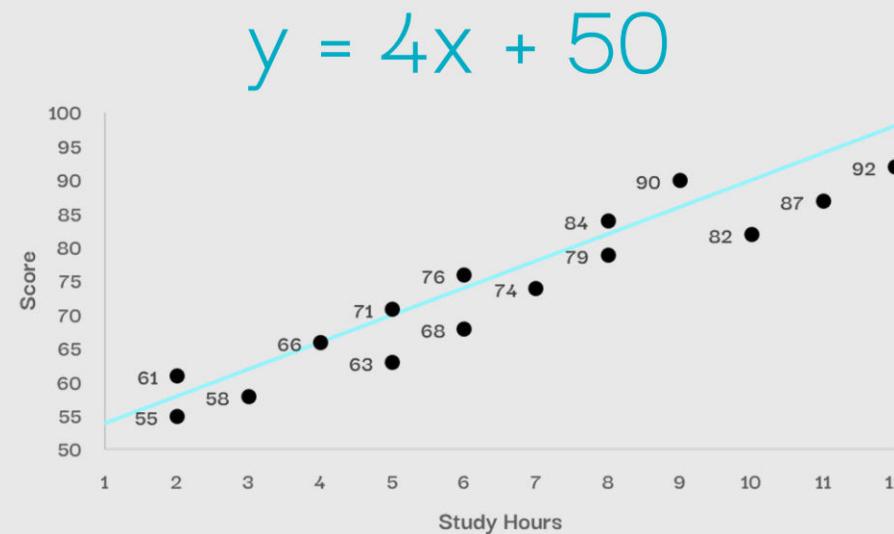
x	y (Actual value)	$y' = 3x + 50$	$ y - y' $	$( y - y' )^2$
2	55	56	1	1
2	61	56	5	25
3	58	59	1	1
4	66	62	4	16
5	63	65	6	36
5	71	65	2	4
6	68	68	8	64
6	76	68	0	0
7	74	71	3	9
8	79	74	5	25
8	84	74	10	100
9	90	77	13	169
10	82	80	2	4
11	87	83	4	16
12	92	86	6	36
Sum of $( y - y' )^2$				506
Mean of sum of $( y - y' )^2$				33.73

# Linear Regression - Activity



x	y (Actual value)	$y' = 4x + 50$	$ y - y' $	$( y - y' )^2$
2	55	58	3	9
2	61	58	3	9
3	58	62	4	16
4	66	66	0	0
5	63	70	7	49
5	71	70	1	1
6	68	74	6	36
6	76	74	2	4
7	74	78	4	16
8	79	82	3	9
8	84	82	2	4
9	90	86	4	16
10	82	90	8	64
11	87	94	7	49
12	92	98	6	36
Sum of $( y - y' )^2$				300
Mean of sum of $( y - y' )^2$				25

# Linear Regression - Activity

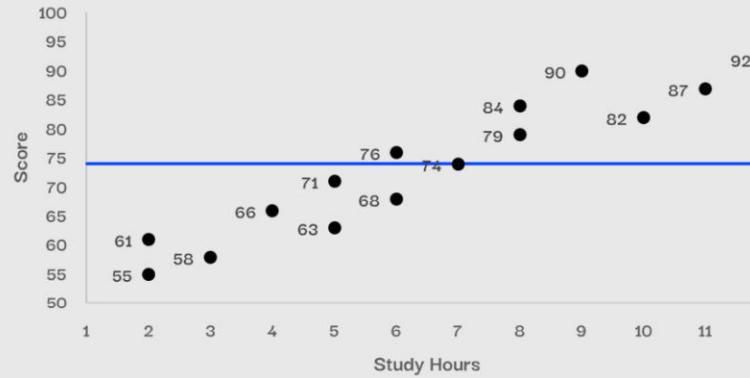


x	y (Actual value)	$y' = 4x + 50$	$ y - y' $	$( y - y' )^2$
2	55	58	3	9
2	61	58	3	9
3	58	62	4	16
4	66	66	0	0
5	63	70	1	1
5	71	70	7	49
6	68	74	2	4
6	76	74	6	36
7	74	78	4	16
8	79	82	3	9
8	84	82	2	4
9	90	86	4	16
10	82	90	8	64
11	87	94	7	49
12	92	98	6	36
Sum of $( y - y' )^2$				318
Mean of sum of $( y - y' )^2$				21.20

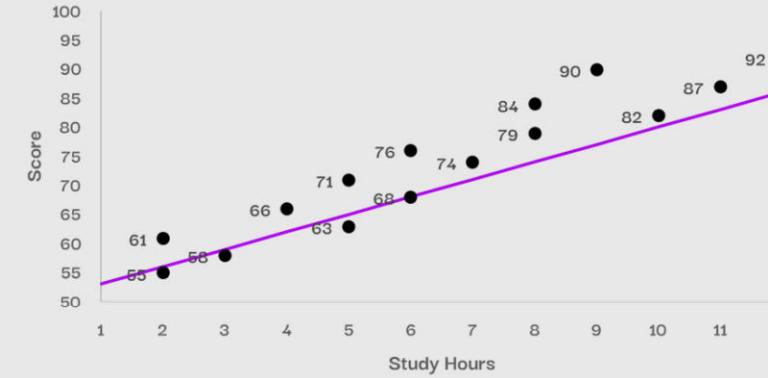
# Linear Regression - Activity

Which of these lines best represents the data?

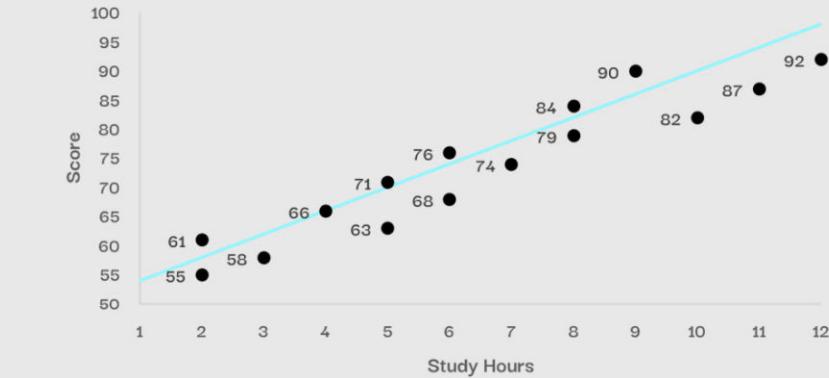
$$y = 74$$



$$y = 3x + 50$$



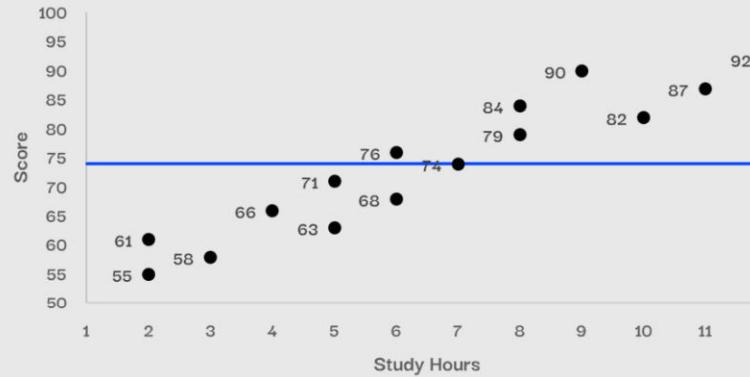
$$y = 4x + 50$$



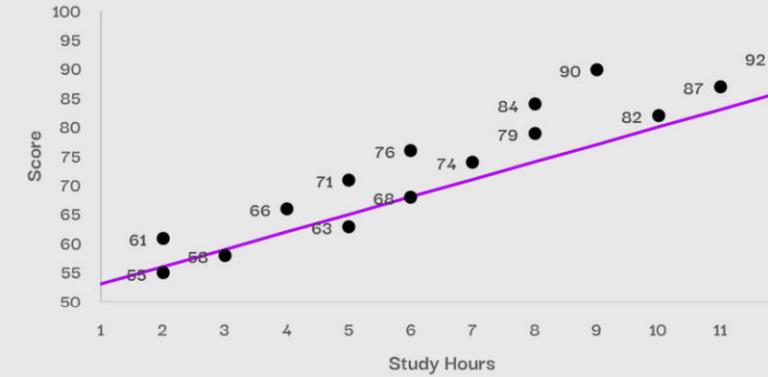
# Linear Regression - Activity

Which of these lines best represents the data?

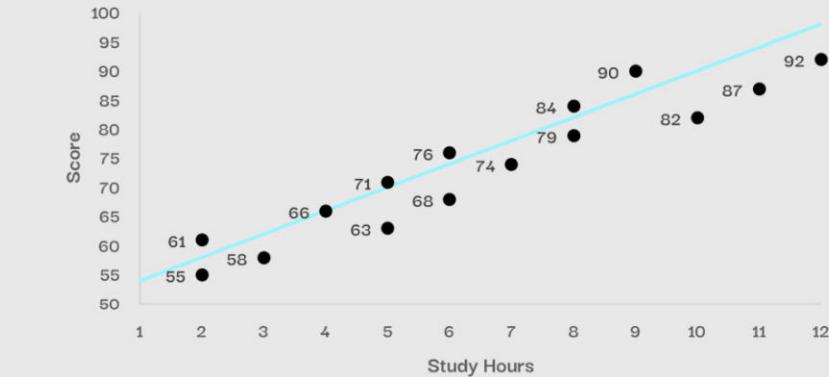
$$y = 74$$



$$y = 3x + 50$$



$$y = 4x + 50$$



130.53

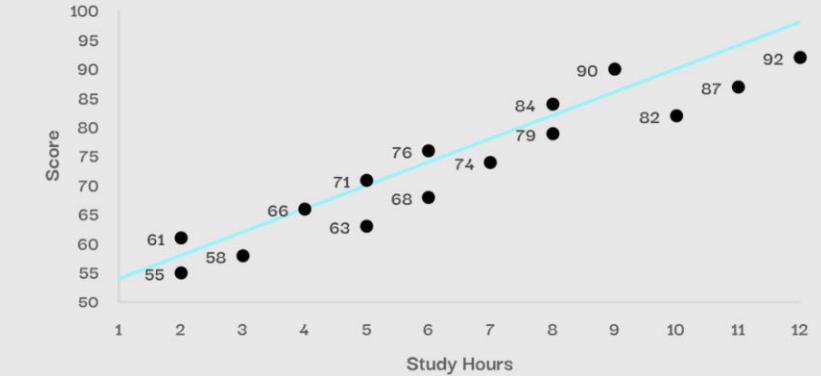
33.73

21.20

# Linear Regression - Activity

Which of these lines best represents the data?

$$y = 4x + 50$$



21.20

# Linear Regression

The goal is to find a line that **best fits all the data points**.

The line is represented by the equation:

$$y = m * x + b$$

where  $m$  is the slope,  $b$  is the y-intercept,  $x$  is the input, and  $y$  is the output.

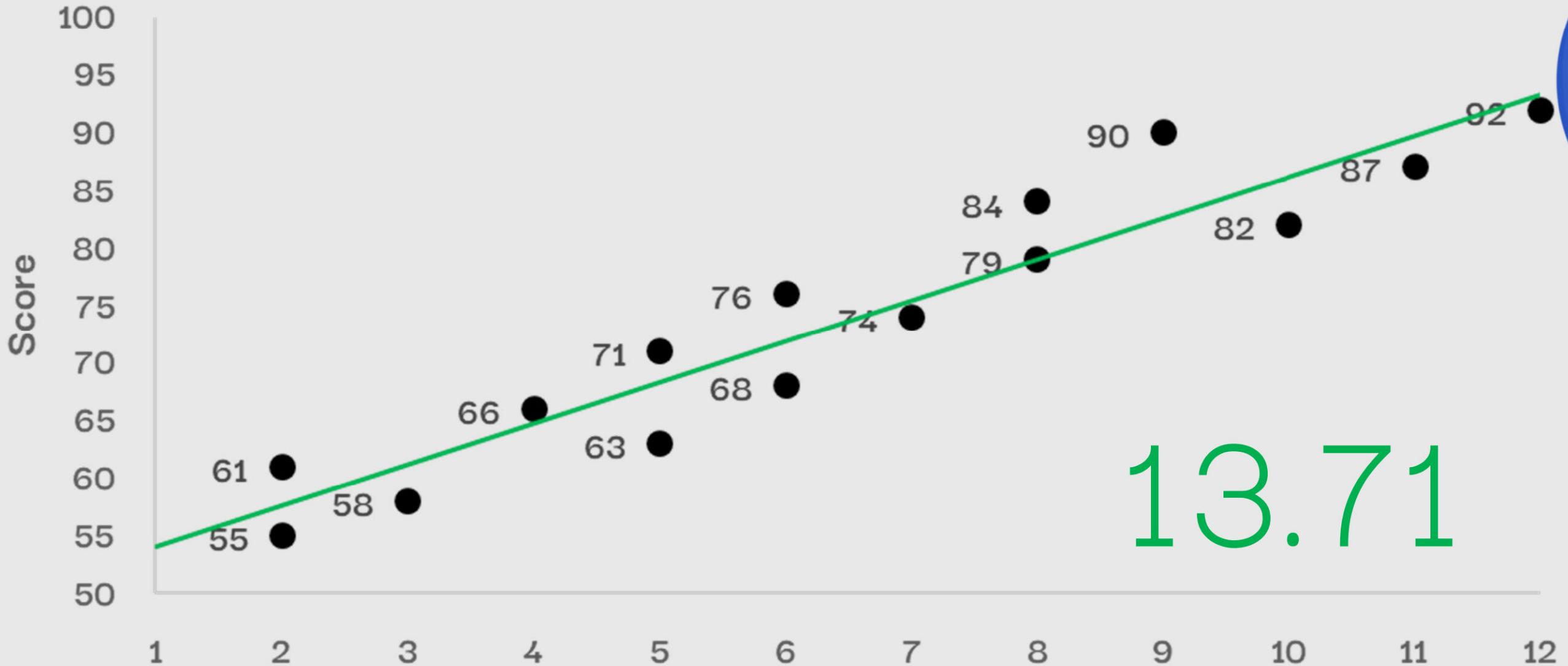
# Linear Regression

$$y = m * x + b$$

We need determine  $m$  and  $b$  to minimize  
the difference between the actual  
data points and the line.

# Linear Regression - Activity

$$y = 3.57x + 50.44$$



13.71

Content Curated by Pollux M. Rey

# Linear Regression Application



Content Curated by Pollux M. Rev

<https://www.youtube.com/watch?v=qxo8p8PtFeA>

# Linear Regression Evaluation Metrics



## MAE

Mean Absolute Error

$$\frac{1}{n} \sum_{i=1}^n |\text{Actual}_i - \text{Predicted}_i|$$

# Linear Regression Evaluation Metrics

## MSE

Mean  
Squared Error

$$\frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2$$

## RMSE

Root Mean  
Squared Error

---

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2}$$

# Linear Regression Evaluation Metrics

**R<sup>2</sup>**

## Coefficient of Discrimination

Measures how well a regression model predicts an outcome

Coefficient of Discrimination (R <sup>2</sup> )	Interpretation
0	The model <b>does not</b> predict the outcome.
Between 0 and 1	The model <b>partially</b> predicts the outcome.
1	The model <b>perfectly</b> predicts the outcome.

# THANK YOU!

Content Curated by Pollux M. Rey