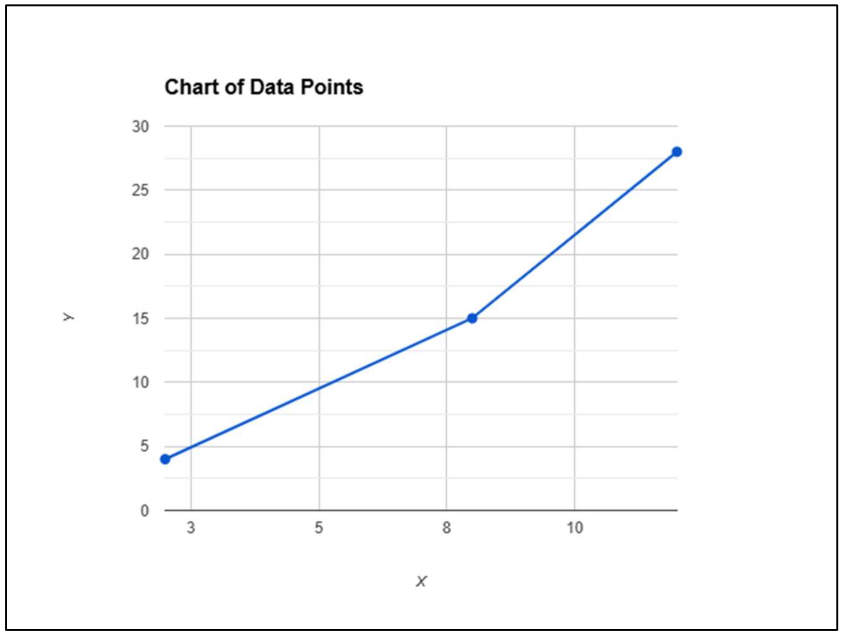
Linear Regression case study:

Input data:

i	x_i	y_i
0	2	4
1	8	15
2	12	28



Sum of squared Residuals (SSR):

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

Mean squared error (MSE):

$$MSE = \frac{1}{n} SSR = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

Linear equation:

$$\hat{y}_i = k.x_i + q \tag{3}$$

Combine equations = cost function:

$$J(k, q) = \frac{1}{n} \sum_{i=1}^{n} [y_i - (k.x_i + q)]^2 \tag{4}$$

Find global minimum of the function:

$$\frac{\partial J}{\partial q} = 0$$

$$\frac{\partial J}{\partial k} = 0 \tag{5}$$

- derivatives:

$$\frac{\partial J}{\partial q} = \frac{2}{n}\sum_{i=1}^{n}(y_i - k.x_i - q).(-x_i)$$

$$\frac{\partial J}{\partial k} = \frac{2}{n}\sum_{i=1}^{n}(y_i - k.x_i - q).(-1) \qquad (6)$$

- after plugging the numbers into the equations, we get:

$$k \approx \mathbf{2.355}$$
$$q \approx -\mathbf{1.605} \qquad (7)$$

The best fitting line has equation:

$$\hat{y}_i = k.x_i + q = \mathbf{2.355}x_i - \mathbf{1.605} \qquad (8)$$