Before getting into linear Regression, firstly let us understand the perspective in which data can be understood.

1. *Statistical Perspective of data*
   output=funtion(input)
   Or
   DependentVariable=f(IndependentVariable)
2. *Algorithmic view*
   Model=Algorithm(Data)
   (Model: Specific Patterns Learned from Data
   Algorithm: The process of learning/understanding Data)

So, here our model is Linear Regression and we write a certain algorithm from scratch which predicts outcomes for given specific inputs.

LINEAR REGRESSION:

1. This examines the relationship between two variables by determining the line of best fit
2. The essential requirements for this fit are
   - Slope: m
   - Intercept: c

The equation of linear regression can be written as such,

$$y = m.x + c$$

(where n= number of values in x or y)

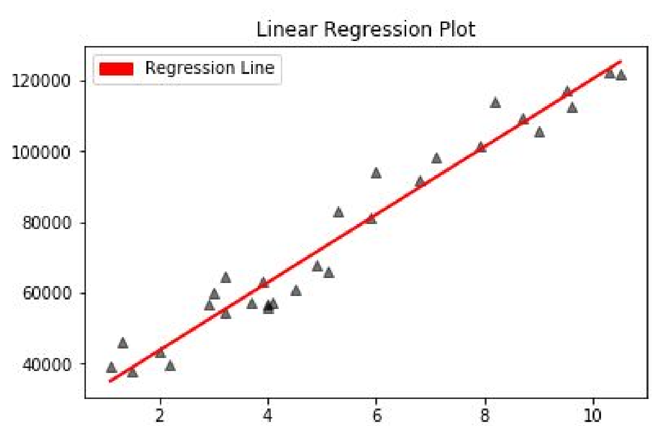
The slope of the given regression line can be determined by the formula,

$$m = (n.\sum_i^n x.y - \sum_i^n x.\sum_i^n y)/(n.\sum_i^n x^2 - (\sum_i^n x)^2)$$

(where n = number of values in x or y)

The intercept of the given regression line can be determined by the formula,

$$c = \overline{y} - m.\overline{x}$$



So, after the prediction, we need to measure the goodness of fit and for that, we use regression metric named coefficient of determination or R-squared measure.

R-Squared(Coefficient of determination):

- If our dataset has n values marked $y1, y2, ..., yi$ and for that values resulted in certain predicted values $f1, f2, ..., fi$

  So,

  $$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

  Let us calculate the sum of squares,

  $$SS_{tot} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

  Next, calculate the regression sum of squares,

  $$SS_{reg} = \sum_{i=1}^{n} (f_i - \overline{y})^2$$

  Next, calculate the residual sum of squares,

  $$SS_{res} = \sum_{i=1}^{n} (y_i - f_i)^2$$

  Finally, calculate the coefficient of determination,

  $$R^2 = 1 - \left(\frac{SS_{reg}}{SS_{tot}}\right)$$

It provides a measure of how well-observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.