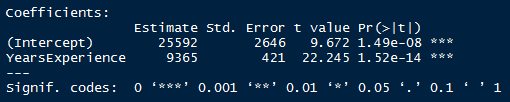
**Business problem:** Identify if the years of experience have any impact on the employee’s salary.

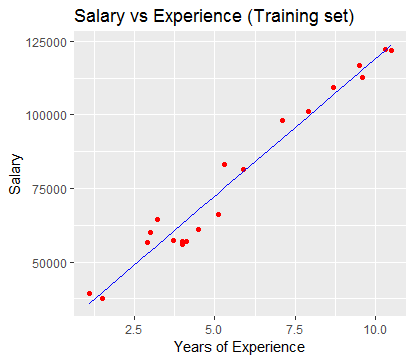
**Explaining the model**

1. Import the dataset, which is in a csv format.
2. After installing the package caTools, I’ve decided to split the sets on a 2/3 – 1/3 proportion for the training and test sets, respectively. Since we only have 30 entries on our dataset, I believe 20 entries should be enough to train our model well enough.
3. After splitting, I could have transformed the categorical values to numerical, as the model cannot work with categorical values. However, the package I’ll choose to fit the Regression Model will do just that, so it is not necessary to do it manually.
4. Next, the regressor is created. We want to see how the salary is impacted by the years of experience of the employees, so that means our dependent variable is Salary, and the independent is Years of Experience. The data we want to use in our regressor is our Training Set, because we’re training our model.
5. When we analyze our regressor, we can see that the coefficient of our independent variable is well below 0.05, which allows us to conclude (as suspected) that it has a strong effect on the dependent variable.

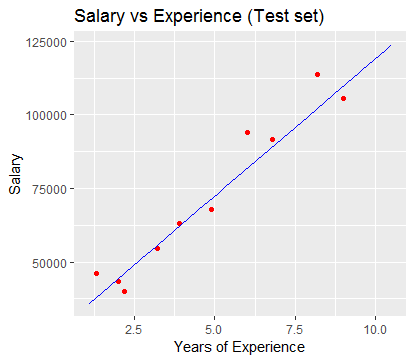


1. It’s now time to test our model; the first argument is the regressor we’ve created, and the second is the Test set.

**Analyzing the results**



Being the red dots the real observations of the training set and the line our regressor, we can see that it’s a good model. According to it, there are around 7 individuals whose salary should be a little higher than it is, and around 8 whose salary should drop down a little bit.



Now testing our model on the test set, we can clearly see it’s pretty accurate.