Simple Linear Regression

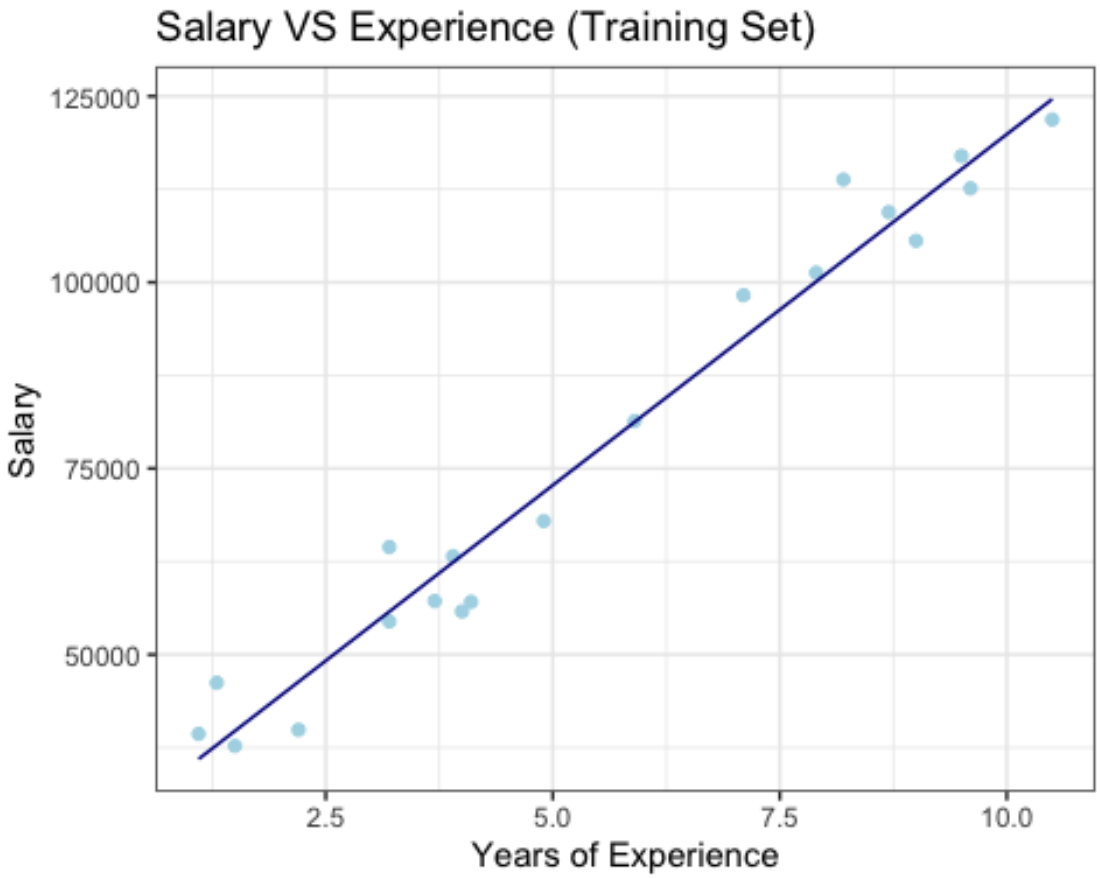
Author : Rose Ellison

I will be building a simple linear regression model based off the 'Salary_Data' data. In this dataset there are two columns, $Salary$, and $YearsExperience$. The YearsExperience is our dependent variable while the Salary is our independent variable. We want to determine if there are any correlations between profit and experience. Additionally, we want to determine if there is a linear dependency.

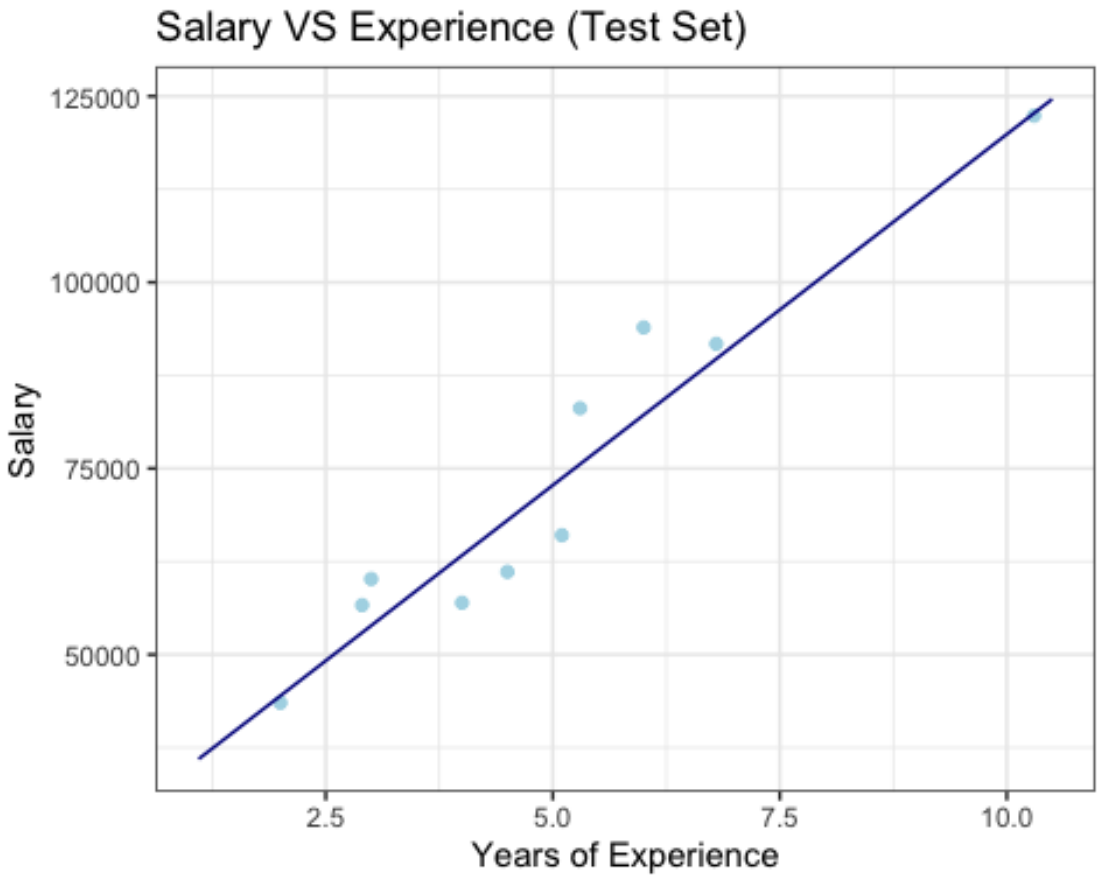
$$SimpleLinearRegressionFormula:$$

$$y = b_0 + b_1 x_1$$

```
# Set seed
set.seed(1)

# Importing salary data
salary <- read.csv('../../data/Salary_Data.csv')

# Splitting the dataset into the training set and test set
split <- sample.split(salary$Salary, SplitRatio = 2/3)
training.set <- subset(salary, split == TRUE)
test.set <- subset(salary, split == FALSE)

# Fitting simple linear regression to the training set
regressor <- lm(formula = Salary ~ YearsExperience,
                data = training.set)

# Predicting the test set results
y.pred <- predict(regressor, newdata = test.set)

# Visualizing the training set results
ggplot() +
  geom_point(aes(x = training.set$YearsExperience, y = training.set$Salary),
col = 'lightblue') +
  geom_line(aes(x = training.set$YearsExperience, y = predict(regressor,
newdata = training.set)), col = 'darkblue') +
  theme_bw() +
  ggtitle('Salary VS Experience (Training Set)') +
  xlab('Years of Experience') +
  ylab('Salary')
```




```
# Visualizing the test set results
ggplot() +
  geom_point(aes(x = test.set$YearsExperience, y = test.set$Salary), col =
'lightblue') +
  geom_line(aes(x = training.set$YearsExperience, y = predict(regressor,
newdata = training.set)), col = 'darkblue') +
  theme_bw() +
  ggtitle('Salary VS Experience (Test Set)') +
  xlab('Years of Experience') +
  ylab('Salary')
```



#

Conclusion

```
summary(regressor)
```

```
##
## Call:
## lm(formula = Salary ~ YearsExperience, data = training.set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7522.2 -3584.8  -598.9  2187.8 10888.3
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      25594.7     2546.6   10.05 8.26e-09 ***
## YearsExperience   9430.4      407.7   23.13 7.71e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5483 on 18 degrees of freedom
## Multiple R-squared:  0.9675, Adjusted R-squared:  0.9656
## F-statistic: 535.1 on 1 and 18 DF,  p-value: 7.707e-15
```