

**Tecnológico Nacional De México**

**Instituto Tecnológico De Tijuana**

**Subdirección Académica**

**Departamento de Sistemas y Computación**

**Semestre Enero - Junio 2022**

**Ingeniería Informática**

**Mineria De Datos**

**Práctica 1 - Unidad 3**

**Unidad 3**

**Perez Ortega Victoria Valeria No.18210718**

**Díaz Ruiz Uriel No.18210839**

**JOSE CHRISTIAN ROMERO HERNANDEZ**

**Tijuana, B.C. a 17 de Mayo de 2022.**

Instalamos nuestra paqueteria de ggplot y catools para poder generar las graficaciones para nuestra práctica:

| install.packages("ggplot2") install.packages("caTools")  library(caTools) library(ggplot2) |
| --- |

Creamos un dataset para que nos pueda buscar y leer nuestro archivo .CSV :

| dataset <- read.csv(file.choose()) dataset <- read.csv('Salary\_Data.csv') |
| --- |

Debemos de importar la librería caTools para que su función. Al igual debemos de poner la semilla para poder crear los números. En cual debemos de asignar los valores requeridos:

| library(caTools) set.seed(123) split <- sample.split(dataset$Salary, SplitRatio = 2/3) training\_set <- subset(dataset, split == TRUE) test\_set <- subset(dataset, split == FALSE) |
| --- |

Debemos de crear la variable regressor para que nos permita guardar los datos del dataset con el cual el Salary y Years of Experience para que después nos deje imprimir nuestro Summary:

| regressor = lm(formula = Salary ~ YearsExperience,  data = dataset) summary(regressor) |
| --- |

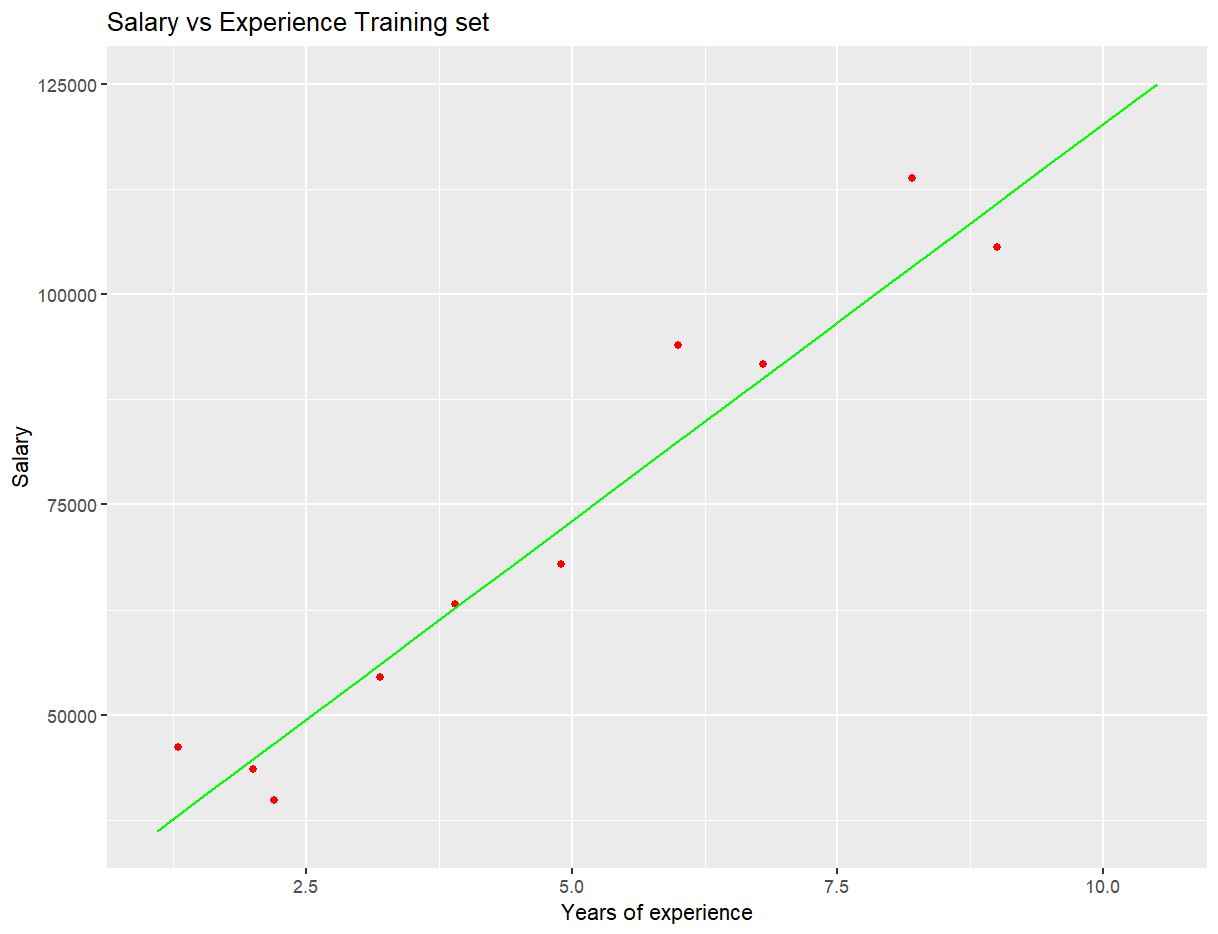
Variable de predicción de datos:

| y\_pred = predict(regressor, newdata = test\_set) |
| --- |

Importación de nuestar libreria de ggplot para nuestra graficación de los datos de Salary vs Experience de Training Set y de Test Set:

Salary vs Experience de Training Set:

| ggplot() + geom\_point(aes(x=test\_set$YearsExperience, y=test\_set$Salary),  color = 'red') + geom\_line(aes(x = training\_set$YearsExperience, y = predict(regressor,newdata = training\_set)),  color = 'green') + ggtitle('Salary vs Experience Training set ') + xlab('Years of experience') + ylab('Salary') |
| --- |



Salary vs Experience de Test Set:

| ggplot() + geom\_point(aes(x=test\_set$YearsExperience, y=test\_set$Salary), color = 'red') + geom\_line(aes(x = training\_set$YearsExperience, y = predict(regressor,newdata = training\_set)),  color = 'black') + ggtitle('Salary vs Experience Test Set ') + xlab('Years of experience') + ylab('Salary') |
| --- |

