 **TIJUANA INSTITUTE OF TECHNOLOGY**

**ACADEMIC**

**DEPARTMENT OF SYSTEMS AND COMPUTATION COMPUTER**

**SYSTEMS ENGINEERING**

**SEMESTER**

FEBRUARY- JULY 2022

**SUBJECT**

BDD-1703SC9C Data Mining

**Activity**

Practice 1

**Teacher**

MC JOSE CHRISTIAN ROMERO HERNANDEZ

**Student**

17212327 -Alderete Delgado Angel Esteban

17210659 - Villegas Carmona Damaris

Tijuana, BC May 17, 2022

getwd()

setwd("/home/chris/Documents/itt/Enero\_Junio\_2020/Mineria\_de\_datos/DataMining/MachineLearning/SimpleLinearRegression")

getwd()

# Importing the dataset

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

dataset <- read.csv(file.choose())

# Splitting the dataset into the Training set and Test set

# Install.packages('caTools')

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)

# Fitting Simple Linear Regression to the Training set

regressor = lm(formula = Salary ~ YearsExperience,

data = dataset)

summary(regressor)

# Predicting the Test set results

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

# Visualising the Training set results

library(ggplot2)

ggplot() +

geom\_point(aes(x=training\_set$YearsExperience, y=training\_set$Salary),

color = 'red') +

geom\_line(aes(x = training\_set$YearsExperience, y = predict(regressor, newdata = training\_set)),

color = 'blue') +

ggtitle('Salary vs Experience (Training Set)') +

xlab('Years of experience') +

ylab('Salary')

# Visualising the Test set results

ggplot() +

geom\_point(aes(x=test\_set$YearsExperience, y=test\_set$Salary),

color = 'red') +

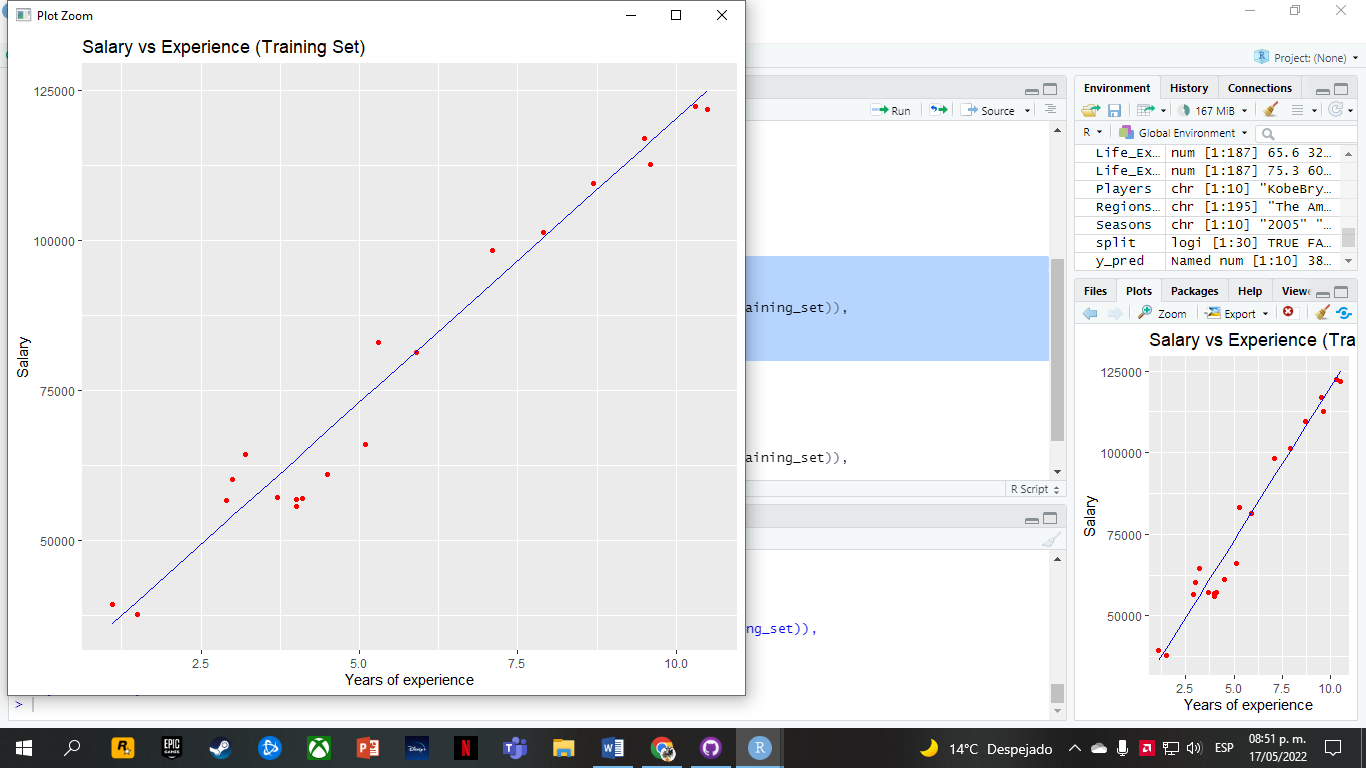
geom\_line(aes(x = training\_set$YearsExperience, y = predict(regressor, newdata = training\_set)),

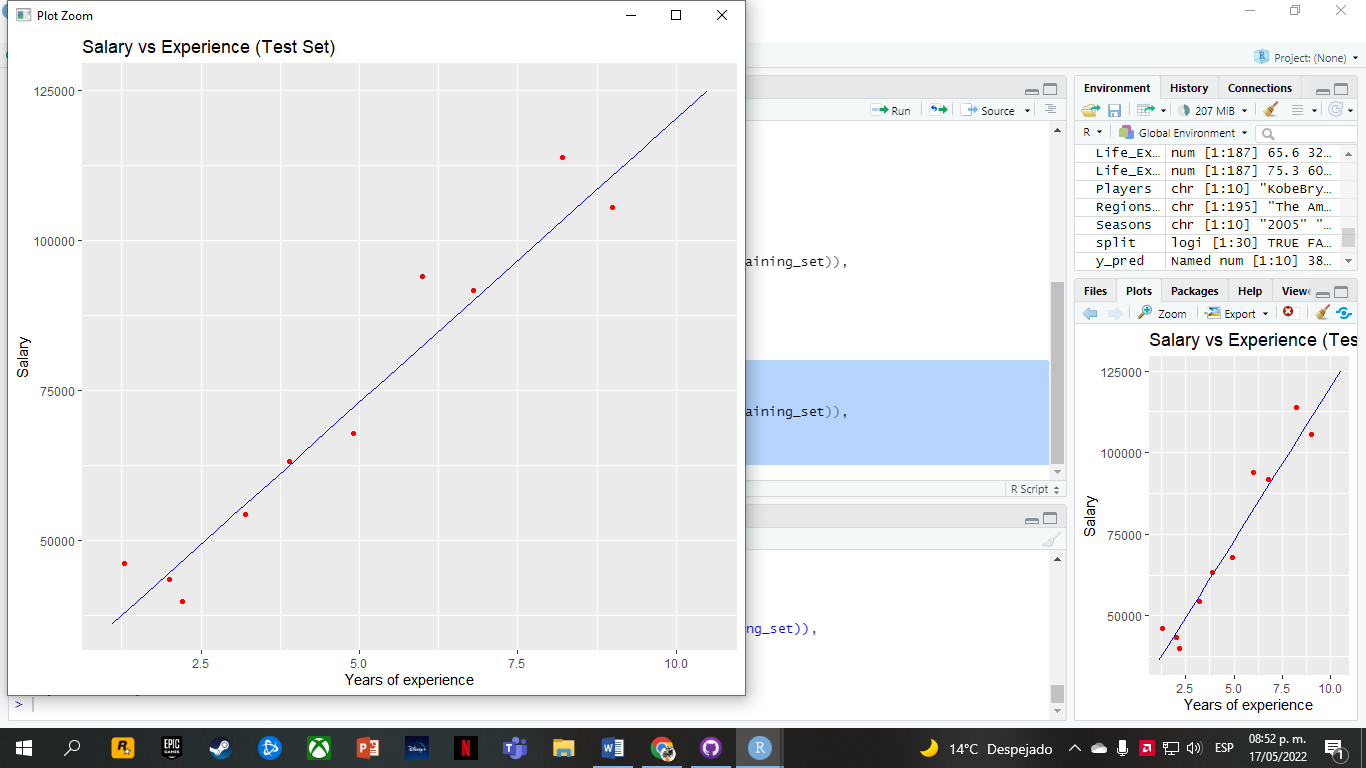
color = 'blue') +

ggtitle('Salary vs Experience (Test Set)') +

xlab('Years of experience') +

ylab('Salary')





Conclusion:

One of the observations that we can give is that this method is one of the simplest to implement, in terms of its syntax and the easy understanding of the language.

One of the problems that we presented was the result obtained from the representation, since we do not know much of the subject and it was a little complex to understand.

As a conclusion, thanks to the inversion on the representation it was possible to understand it in a better way. We were able to observe that some columns were related and this was evidenced in one of the graphs.