DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
Discover. Learn. Empower.

Experiment 9

Student Name: Praduman Kumar
UID: 20BCS9446
Branch: CSE
Section/Group: 714/A
Semester: 6th
Subject Name: Data Mining Lab
Subject Code: 20CSP-376

---

1. **Aim:**
   Study of Regression Analysis using R Programming.

2. **Apparatus / Simulation Used:**
   - Windows 7 or above
   - R Studio

3. **Objective:**
   - Demonstration of the Regression using R.
   - Performing the Regression Analysis using R.

4. **Theory and Output:**

Linear Regression: It is a commonly used type of predictive analysis. It is a statistical approach for modeling the relationship between a dependent variable and a given set of independent variables.

There are two types of linear regression.

- Simple Linear Regression
- Multiple Linear Regression

➢ **Simple Linear Regression:**

It is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables. One variable denoted x is regarded as an independent variable and the other one denoted y is regarded as a dependent variable. It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value as accurately as possible as a function of the feature or independent variable(x).

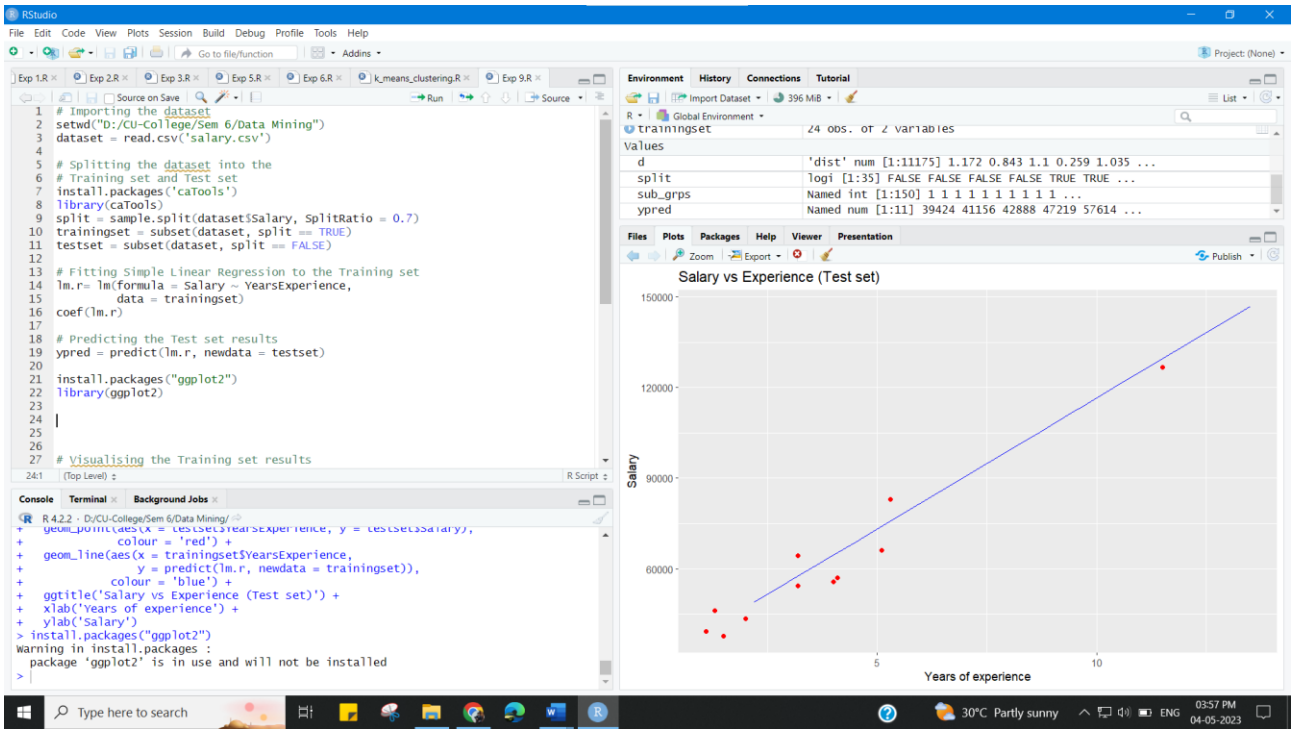
➢ **Multiple linear regression**

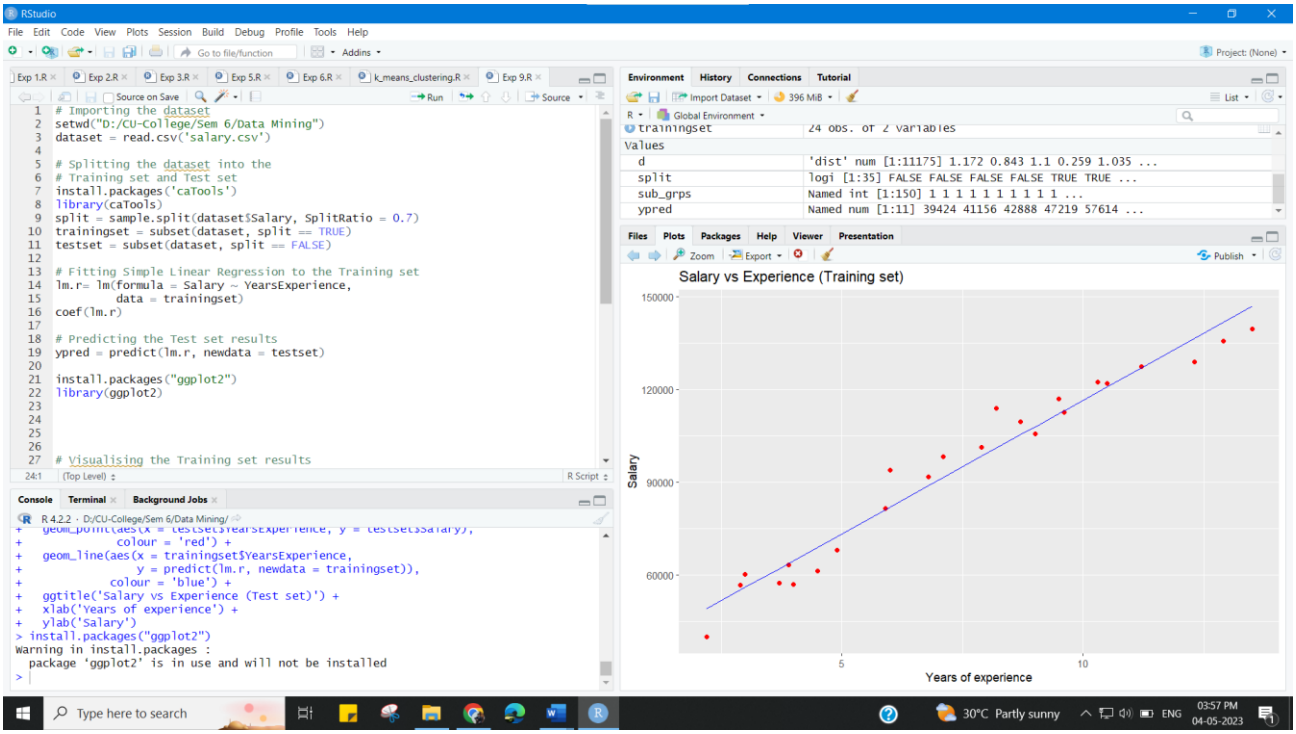
Multiple linear regression is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line.

5. **Code:**

```
# Importing the dataset
setwd("D:/CU-College/Sem 6/Data Mining")
dataset = read.csv('salary.csv')

# Splitting the dataset into the
# Training set and Test set
install.packages('caTools')
library(caTools)
split = sample.split(dataset$Salary, SplitRatio = 0.7)
trainingset = subset(dataset, split == TRUE)
testset = subset(dataset, split == FALSE)

# Fitting Simple Linear Regression to the Training set
lm.r= lm(formula = Salary ~ YearsExperience,
         data = trainingset)
coef(lm.r)

# Predicting the Test set results
ypred = predict(lm.r, newdata = testset)

install.packages("ggplot2")
library(ggplot2)


# Visualising the Training set results
ggplot() + geom_point(aes(x = trainingset$YearsExperience,
                          y = trainingset$Salary), colour = 'red') +
  geom_line(aes(x = trainingset$YearsExperience,
                y = predict(lm.r, newdata = trainingset)), colour = 'blue') +

  ggtitle('Salary vs Experience (Training set)') +
  xlab('Years of experience') +
  ylab('Salary')

# Visualising the Test set results
ggplot() +
  geom_point(aes(x = testset$YearsExperience, y = testset$Salary),
             colour = 'red') +
  geom_line(aes(x = trainingset$YearsExperience,
                y = predict(lm.r, newdata = trainingset)),
            colour = 'blue') +
  ggtitle('Salary vs Experience (Test set)') +
  xlab('Years of experience') +
  ylab('Salary')
```

6. Output:





Learning outcomes (What I have learnt):

- Demonstration of the Regression using R.
- Performing the Regression Analysis using R.