```
# The Sparks Foundation GRIP March 2021
# Data Science and Business Analytics Internship
#Name- Usha Kumari
# TASK 1- Prediction using Supervised ML

#reading the data from the csv file
dataset=read.csv(file="spark data.csv",header =T);dataset;
#extracting the study hours as the independent variable
x=dataset$Hours;x

#extracting the scores as the dependent variable
y=dataset$Scores;y

#plotting the data as a scatterplot
library(ggplot2)
g=ggplot(dataset,aes(x,y)) + geom_point(col="#000099",size=2) + labs(title="Scatterplot of Score vs
g

# regression of y on x
reg=lm(y~x)
summary(reg)
beta=reg$coefficients;beta;
# to predict the score if a student studies 9.25 hours/day
beta[1]+(beta[2]*9.25)

#hence a student will score approximately 93 if he/she studies 9.25 hrs/day

g1=ggplot(dataset,aes(x,y)) + geom_point(col="#000099",size=2) + labs(title="Scatterplot of Score vs
g1
y_pred=reg$fitted.values;
d=data.frame(x,y,y_pred)
#plotting both observed scores and predicted scores to compare
g2=ggplot(d)+geom_point(aes(x,y=y),col="#cc0066",size=2)+geom_point(aes(x,y_pred),col="#009900",size
g2 #the green dots are the predicted values and the pink dots are the observed values
# Thank You
```