##########Data Science and Business Analytics##########

########Task-1:"Prediction Using Supervised Machine Learning"########

######Author: Yashaswa Verma######

###PROBLEM: Predict the score of a student studying 9.25 hrs/day###

####### Reading Data From CSV file and Predicting the Score with a function #######

###Extracting the data

rm(list=ls())

setwd("D:/")

s\_data=read.csv('Sparks1.csv',header=T)

###Plotting the data in a Scatterplot

x=s\_data$Hours

y=s\_data$Scores

library(ggplot2)

ggplot(s\_data,aes(x=Hours,y=Scores))+

geom\_point(color="green")+

labs(title="Scatterplot of Hours v/s Score")+theme\_classic()

###Fitting a linear regression model with an itercept term

fit1=lm(y~x)

summary(fit1)

###Fitting a linear regression model with no intercept term

fit2=lm(y~x-1)

summary(fit2)

###Residual Plot

resi=resid(fit2)

residualplot=data.frame(resi,x)

ggplot(residualplot,aes(x=x,y=resi))+

geom\_point()

####Fitted Line PLot and Prediction

PredictScr=function(tm)

{

s\_data=read.csv('Sparks1.csv',header=T)

x=s\_data$Hours

y=s\_data$Scores

b=coefficients(lm(y~x-1))

pred=b[1]\*tm

plot(x,y,main="Regression Line Plot ",xlab="Hours",ylab="Scores","p",col="green")

abline(lm(y~x-1),col='red')

cat('The Predicted Score for hour of study : ',tm,' is : ')

return(as.numeric(pred))

}

PredictScr(9.25)