Name - Aman Raj

GRIP : The Sparks Foundation

Data Science and Business Analytics Internship

Task_1 : Prediction using supervised Machine Learning

Importing data from csv file and storing it in Score

```
Score <- read.csv("D:/SparksFoundation/LinearRegression.csv")
```

```
Score
```

```
##    Hours Scores
## 1    2.5     21
## 2    5.1     47
## 3    3.2     27
## 4    8.5     75
## 5    3.5     30
## 6    1.5     20
## 7    9.2     88
## 8    5.5     60
## 9    8.3     81
## 10   2.7     25
## 11   7.7     85
## 12   5.9     62
## 13   4.5     41
## 14   3.3     42
## 15   1.1     17
## 16   8.9     95
## 17   2.5     30
## 18   1.9     24
## 19   6.1     67
## 20   7.4     69
## 21   2.7     30
## 22   4.8     54
## 23   3.8     35
## 24   6.9     76
## 25   7.8     86
```

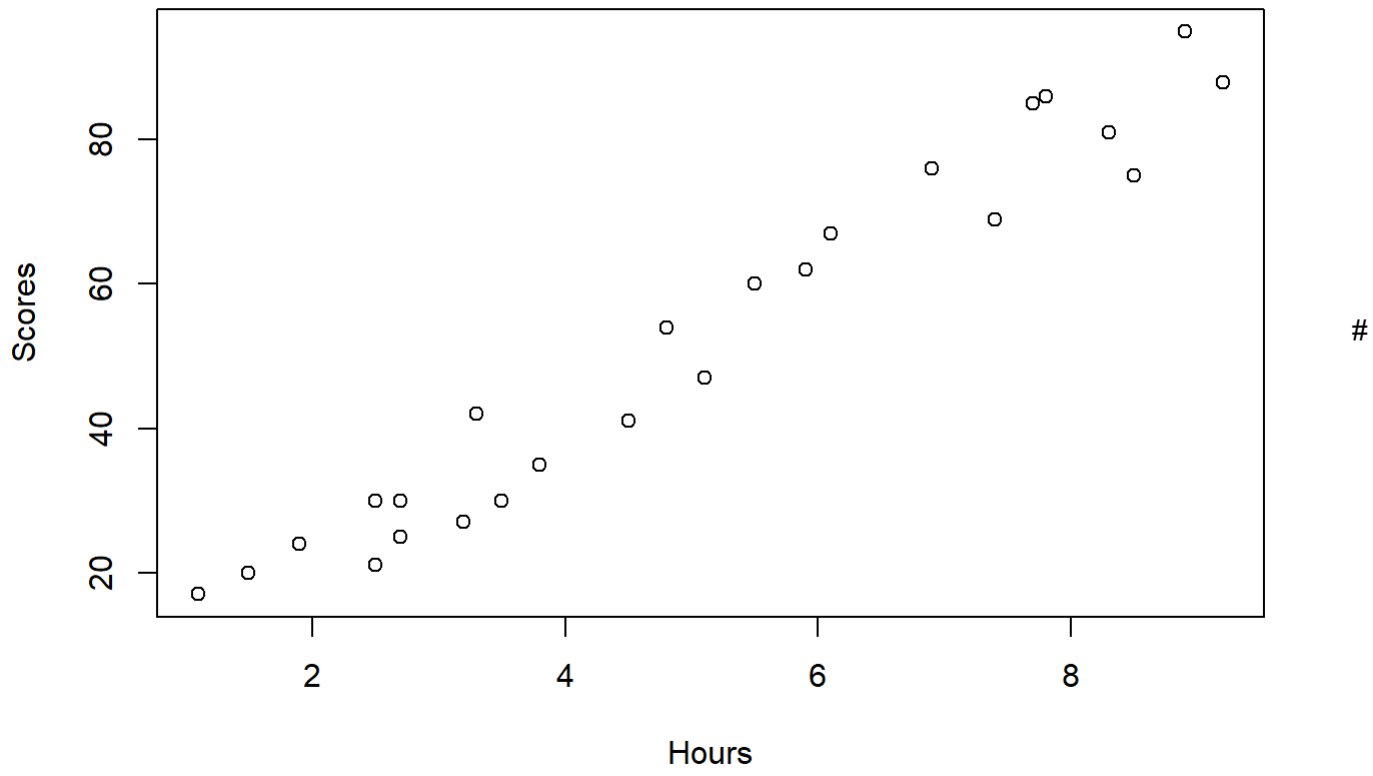
#summary of the data

```
summary(Score)
```

```
##      Hours           Scores     
##  Min.   :1.100   Min.   :17.00  
##  1st Qu.:2.700   1st Qu.:30.00  
##  Median :4.800   Median :47.00  
##  Mean   :5.012   Mean   :51.48  
##  3rd Qu.:7.400   3rd Qu.:75.00  
##  Max.   :9.200   Max.   :95.00
```

#Scatter plot of the data

```
plot(Score)
```



#

Splitting the data into training and test set

```
library(caTools)
```

```
split <- sample.split(Score$Scores, SplitRatio=0.8)
split
```

```
##  [1]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE
## [25] FALSE
```

```
train <- subset(Score, split==TRUE)
test <- subset(Score, split==FALSE)
```

```
train
```

```
##    Hours Scores
## 1    2.5     21
## 2    5.1     47
## 4    8.5     75
## 5    3.5     30
## 6    1.5     20
## 7    9.2     88
## 8    5.5     60
## 9    8.3     81
## 10   2.7     25
## 11   7.7     85
## 13   4.5     41
## 14   3.3     42
## 15   1.1     17
## 16   8.9     95
## 17   2.5     30
## 18   1.9     24
## 19   6.1     67
## 20   7.4     69
## 22   4.8     54
## 24   6.9     76
```

```
test
```

```
##    Hours Scores
## 3    3.2     27
## 12   5.9     62
## 21   2.7     30
## 23   3.8     35
## 25   7.8     86
```

#Regression Model

```
Model <- lm(Scores ~., data=train)
```

```
summary(Model)
```

```
##
## Call:
## lm(formula = Scores ~ ., data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.691 -5.282  1.896  4.615  7.907
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.9568     2.8173   1.404    0.177
## Hours         9.4982     0.4929  19.270 1.83e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.711 on 18 degrees of freedom
## Multiple R-squared:  0.9538, Adjusted R-squared:  0.9512
## F-statistic: 371.3 on 1 and 18 DF,  p-value: 1.828e-13
```

#Accuracy of ~ 94.5% obtained from Regression model (as given by the R-squared value)

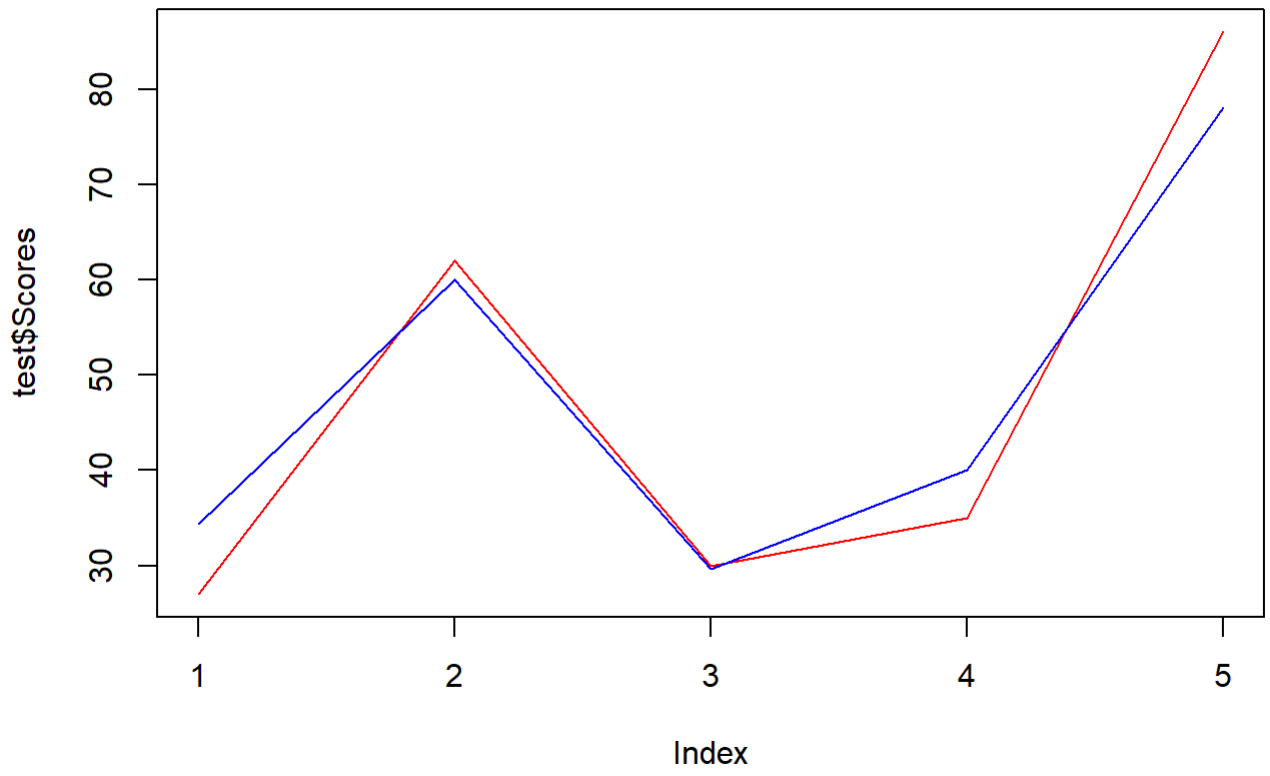
#Prediction on the test set

```
pred <- predict(Model, test)
pred
```

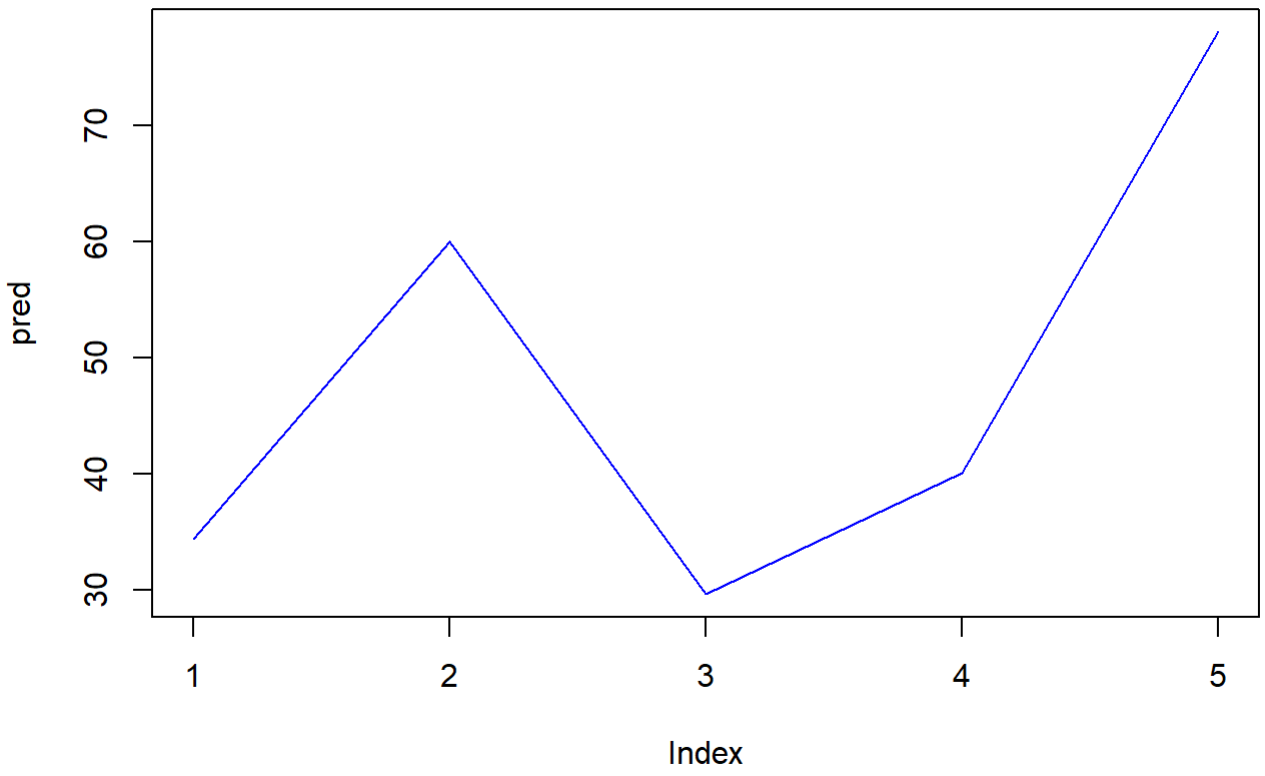
```
##        3       12       21       23       25
## 34.35096 59.99603 29.60188 40.04987 78.04256
```

#comparing predcted vs actual values for the test set

```
plot(test$Scores,type="l", lty=1.8, col="red") #plot of test set values
lines(pred, type="l", col="blue")
```



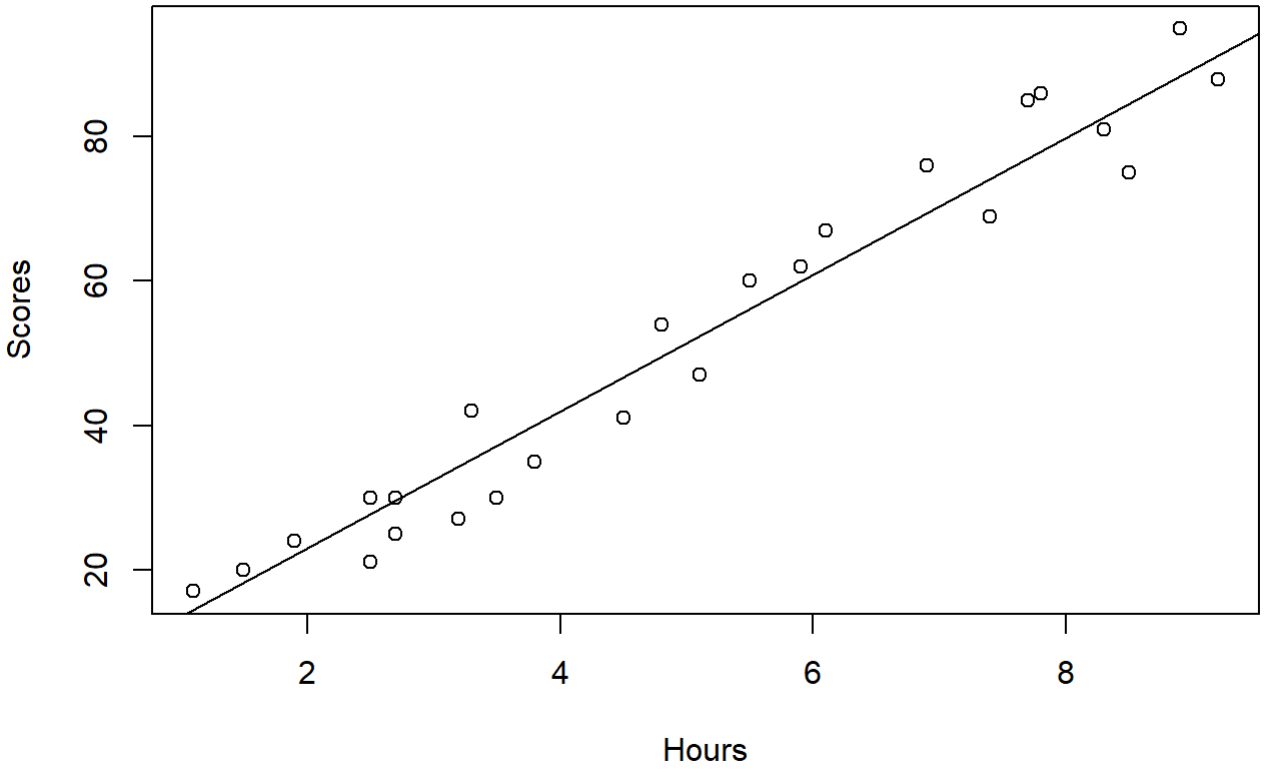

```
plot(pred,type="l", lty=1.8, col="blue") #plot of predicted values
```



#The

Regression line with the scatter plot of given data

```
plot(Score)
abline(Model)
```


Scores
20
40
60
80
2
4
6
8
Hours