**Exercise – 2**

**Predict customer credit risk for credit card dataset using Linear Regression**

**Program:**

library(caret)

library(ggplot2)

data <- read.csv("credit\_card\_data.csv")

head(data)

str(data)

target <- "Credit\_Score"

predictors <- setdiff(names(data), target)

set.seed(123)

trainIndex <- createDataPartition(data[[target]], p = 0.8, list = FALSE)

trainData <- data[trainIndex, ]

testData <- data[-trainIndex, ]

cat("Training set size:", dim(trainData), "\n")

cat("Testing set size:", dim(testData), "\n")

model <- lm(Credit\_Score ~ ., data = trainData)

summary(model)

predictions <- predict(model, newdata = testData)

actuals <- testData$Credit\_Score

mse <- mean((predictions - actuals)^2)

rmse <- sqrt(mse)

cat("Mean Squared Error (MSE):", mse, "\n")

cat("Root Mean Squared Error (RMSE):", rmse, "\n")

results <- data.frame(Actual = actuals, Predicted = predictions)

print(head(results))

ggplot(results, aes(x = Actual, y = Predicted)) +

geom\_point(color = "blue") +

geom\_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed") +

labs(title = "Actual vs. Predicted Credit Score",

x = "Actual Credit Score",

y = "Predicted Credit Score") +

theme\_minimal()

**Output:**



