**R codes**

# step 1 importing data set

data(cars)

cars

# step 2 Graphical analysis

# Scatter plot code

scatter.smooth(x=cars$speed, y=cars$dist, main="Dist ~ Speed")

par(mfrow=c(1, 2)) # divide graph area in 2 columns

# Box plot code

boxplot(cars$speed, main="Speed")

boxplot(cars$dist, main="Distance")

# Density plot code

library(e1071)

par(mfrow=c(1, 2))

plot(density(cars$speed), main="Density Plot: Speed", ylab="Frequency")

polygon(density(cars$speed), col="blue")

plot(density(cars$dist), main="Density Plot: Distance", ylab="Frequency")

polygon(density(cars$dist), col="blue")

# Step 3 Checking correlation

cor(cars$speed, cars$dist)

# Step 4 *build* linear regression model on full data

linearMod <- **lm**(dist ~ speed, data=cars)

**print**(linearMod)

# Step 5 Checking for statistical significance

# Step 6 Create the training (development) and test (validation) data samples from original data.

**set.seed**(100)

trainingRowIndex <- **sample**(1:**nrow**(cars), 0.8\***nrow**(cars))

trainingData <- cars[trainingRowIndex, ]

**testData <- cars[-trainingRowIndex, ]**

#Step 7: Develop the model on the training data and use it to predict the distance on test data

lmMod <- **lm**(dist ~ speed, data=trainingData)

distPred <- **predict**(lmMod, testData)

Step 8: Review diagnostic measures.

**summary** (lmMod)

# Step 8: Calculate prediction accuracy and error rates

actuals\_preds <- **data.frame**(**cbind**(actuals=testData$dist, predicteds=distPred))

correlation\_accuracy <- **cor**(actuals\_preds)