1. Perform the below given activities:

a. Take a sample data set of your choice

b. Apply random forest, logistic regression using Spark R

c. Predict for new dataset.

**Answer:**

**#Spark-R Integration**

**library(sparklyr)**

**spark\_install(version = "2.3.0")**

**sc <- spark\_connect(master = "local")**

**library(dplyr)**

**iris\_tbl <- copy\_to(sc, iris)**

**iris\_tbl**

**# Machine Learning using Apache Spark**

**# copy mtcars into spark**

**mtcars\_tbl <- copy\_to(sc, mtcars)**

**mtcars\_tbl**

**# transform our data set, and then partition into 'training', 'test'**

**partitions <- mtcars\_tbl %>%**

**filter(hp >= 100) %>%**

**mutate(cyl8 = cyl == 8) %>%**

**sdf\_partition(training = 0.5, test = 0.5, seed = 1099)**

**partition**

**#Spark-R Integration**

**library(sparklyr)**

**spark\_install(version = "2.3.0")**

**sc <- spark\_connect(master = "local")**

**library(dplyr)**

**iris\_tbl <- copy\_to(sc, iris)**

**iris\_tbl**

**# Machine Learning using Apache Spark**

**# copy mtcars into spark**

**mtcars\_tbl <- copy\_to(sc, mtcars)**

**mtcars\_tbl**

**# transform our data set, and then partition into 'training', 'test'**

**partitions <- mtcars\_tbl %>%**

**filter(hp >= 100) %>%**

**mutate(cyl8 = cyl == 8) %>%**

**sdf\_partition(training = 0.5, test = 0.5, seed = 1099)**

**partition**



