head(airquality)

mean(airquality)

mean(airquality$Solar.R, na.rm = TRUE)

New\_df = airquality

New\_df$Ozone = ifelse(is.na(New\_df$Ozone),

median(New\_df$Ozone,

na.rm = TRUE),

New\_df$Ozone)

head(New\_df)

##create excel file with two columns roll,marks keep few marks blank.save it as csv file

dt=read.csv("C:/Users/DELL/Desktop/titanic.csv")

head(dt)

dt$marks = ifelse(is.na(dt$marks),

mean(dt$marks,

na.rm = TRUE),

dt$marks)

#Removing outliers using boxplot

data <- iris[,2]

length(data)

boxplot(data)

boxplot(data, plot = FALSE)$out

outliers <- boxplot(data, plot = FALSE)$out

data\_no\_outlier <- data[-which(data %in% outliers)]

boxplot(data\_no\_outlier, plot = FALSE)$out

length(data\_no\_outlier)

boxplot(data\_no\_outlier)

# Load the mtcars dataset

data(mtcars)

# Calculate the Pearson's correlation coefficient between mpg and hp

correlation = cor(mtcars$mpg, mtcars$hp, method = "pearson")

# Print the correlation coefficient

print(correlation)

library(MASS)

print(str(survey))

stu\_data = data.frame(survey$Smoke,survey$Exer)

stu\_data = table(survey$Smoke,survey$Exer)

print(stu\_data)

print(chisq.test(stu\_data))

> # Get the input values.

> input <- mtcars[, c('wt', 'mpg')]

> # Plot the chart for cars with

> # weight between 1.5 to 4 and

> # mileage between 10 and 25.

> plot(x = input$wt, y = input$mpg,

+ xlab = "Weight",

+ ylab = "Milage",

+ xlim = c(1.5, 4),

+ ylim = c(10, 25),

+ main = "Weight vs Milage"

+ )

> # Create the data for the chart

> A <- c(20, 40, 25, 50, 10)

> # Plot the bar chart

> barplot(A, horiz = TRUE, xlab = "X-axis",

+ ylab = "Y-axis", main ="Bar-Chart")

install.packages("partykit")

install.packages('caret')

install.packages("pROC")

install.packages('rattle')

install.packages('rpart.plot')

install.packages('RColorBrewer')

titanic<-read.csv(file.choose(),header = T,sep=",")

summary(titanic)

names(titanic)

library(partykit)

titanic$Survived<-as.factor(titanic$Survived)#convert to categorical

summary(titanic$Pclass)

names(titanic)

set.seed(1234)

pd<-sample(2,nrow(titanic),replace = TRUE, prob=c(0.8,0.2))

#two samples with distribution 0.8 and 0.2

trainingset<-titanic[pd==1,]#first partition

head(trainingset)

validationset<-titanic[pd==2,]#second partition

class(titanic$PassengerId)

class(titanic$Survived)

class(titanic$Sex)

class(titanic$Pclass)

class(titanic$Name)

class(titanic$Age)

class(titanic$SibSp)

class(titanic$Parch)

class(titanic$Ticket)

class(titanic$Fare)

class(titanic$Cabin)

class(titanic$Embarked)

tree<-ctree(formula = Survived ~ Pclass + Age + SibSp + Parch + Fare ,data=trainingset)

class(titanic$Survived)

plot(tree)

#Prunning

tree<-ctree(formula = Survived ~ Pclass + Age + SibSp + Parch +

Fare ,data=trainingset,control=ctree\_control(mincriterion =

0.99,minsplit = 500))

plot(tree)

pred<-predict(tree,validationset,type="prob")

pred

pred<-predict(tree,validationset)

pred

library(caret)

confusionMatrix(pred,validationset$Survived)

pred<-predict(tree,validationset,type="prob")

pred

library(pROC)

plot(roc(validationset$Survived,pred[ ,2]))

library(rpart)

fit <- rpart(Survived ~ Pclass+ Age + SibSp + Parch +

Fare,data=titanic,method="class")

plot(fit)

text(fit)

library(rattle)

library(rpart.plot)

library(RColorBrewer)

fancyRpartPlot(fit)

Prediction <- predict(fit, titanic, type = "class")

Prediction