# week 3 Assignment

# 1. Create a function which takes one argument (a data frame containing any number of categorical variables), and returns a data frame containing the dummies of each level of each categorical variable. The column names should be renamed in the following form: Variable Name\_Level.

# Example: For a categorical variable Gender, with the levels Male and Female, the dummy data having columns of 1s and 0s, with the names Gender\_Male and Gender\_Female.

data(iris)

create\_dummy = function(x) {

for (i in 1:ncol(iris)) {

if(class(iris[,i])=="factor"){

column\_name = colnames(iris)[i]

level <- levels(iris[,i])

for(j in level) {

iris[,paste0(column\_name,"\_",j)]=ifelse(iris[,i]==j,1,0)

}

}

}

return(iris)

}

iris = create\_dummy(iris)

# 2. Use IRIS dataset that is already there in R environment.

# Create a function that takes value of Species column as input and returns the mean value of all other columns in the dataset.

mean\_iris = function(x) {

data(iris)

iris\_1 = subset(iris,tolower(iris$Species)==tolower(x))

lapply(iris,mean)

}

mean\_iris('setosa')

# 3. Build a function, which would create buckets based on the quartile values having customized labels.

# Then add the bucketed field using an apply function which in turn uses the user defined function.

quantile\_buckets = function(x) {

if(class(x)=="numeric") {

input\_1=as.numeric(readline("Give the first label: "))

input\_2=as.numeric(readline("Give the Second label: "))

input\_3=as.numeric(readline("Give the Third label: "))

input\_4=as.numeric(readline("Give the fourth label: "))

iris$x\_quantile <- ifelse(x <= quantile(x,probs=0.25),input\_1,

ifelse(x <= quantile(x,probs=0.5) &

x > quantile(x,probs=0.25),input\_2,

ifelse(x <= quantile(x,probs=0.75) &

x >= quantile(x,probs=0.5),input\_3,input\_4)))

return(iris)

}

else{print("Not a numeric variable")}

}

quantile\_buckets(iris$Sepal.Length)

lapply(iris,quantile\_buckets)