SOURCE CODES FOR THE PROJECT-

**HEALTH CARE COST ANALYSIS**

R PROGRAMMING

PRESENTER-BANASHREE SAIKIA

1. Importing the file ‘health care data set.csv’

getwd()

data1 <- read.csv("filename.csv", header=TRUE, sep=",")

data1

1. DATA EXPLORATION

View(data1)

str(data1)

dim(data1)

names(data1)

nrow(data1)

ncol(data1)

1. Splitting data

set.seed(1)

sample<-sample.split(data1$TOTCHG,SplitRatio=0.70)

sample

train\_data <- subset(data1, sample==TRUE)

test\_data <- subset(data1, sample==FALSE)

1. MODEL BUILDING

model1 <-lm(TOTCHG~.,data=train\_data)

summary(model1)

model2<-lm(TOTCHG~AGE+LOS+APRDRG, data=train\_data)

summary(model2)

1. #Predicting the model

prediction <- predict(model2,test\_data)

head(prediction)

1. #to attach with the data frame

prediction1<-data.frame(prediction)

1. #to bind the original data with the predicted data set by cbind function

final\_data<-cbind(test\_data,prediction1)

View(final\_data)

1. #to export the data set

write.csv(final\_data,"logistic\_output.csv")