***Codes of the Project 9***

getwd()

#Step 1: install the package to read xl files

install.packages("readxl")

library("readxl")

#Step 2: upload the data import the data

data1 = read\_excel("/home/vinusahoo\_gmail/hospitalcosts.xlsx")

View(data1)

#Step 3: Data Expolration

head(data1)

tail(data1)

data2= head(data1,100)

data3= tail(data1,50)

nrow(data1)

ncol(data1)

dim(data1)

str(data1)

summary(data1$TOTCHG)

levels (data1$TOTCHG)

table (data1$TOTCHG)

mean(data1$TOTCHG)

median(data1$TOTCHG)

min(data1$TOTCHG)

max(data1$TOTCHG)

var(data1$TOTCHG)

sd(data1$TOTCHG)

range(data1$TOTCHG)

class(data1)

#Step 4 : Split the data into training and testing

install.packages("caTools")

library(caTools)

sample = sample.split(data1$TOTCHG,SplitRatio=0.80)

sample

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

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

#Step 5 : Building model using train\_data

model = lm(TOTCHG~.,data= train\_data)

summary(model)

age = 5.05e-16

format(age,sci=FALSE)

los=2e-16

format(los,sci=FALSE)

aprdrg=2e-16

format(aprdrg,sci=FALSE)

final\_model = lm(TOTCHG~ AGE + LOS + APRDRG, data = train\_data)

summary(final\_model)

predtest<- predict(final\_model,test\_data)

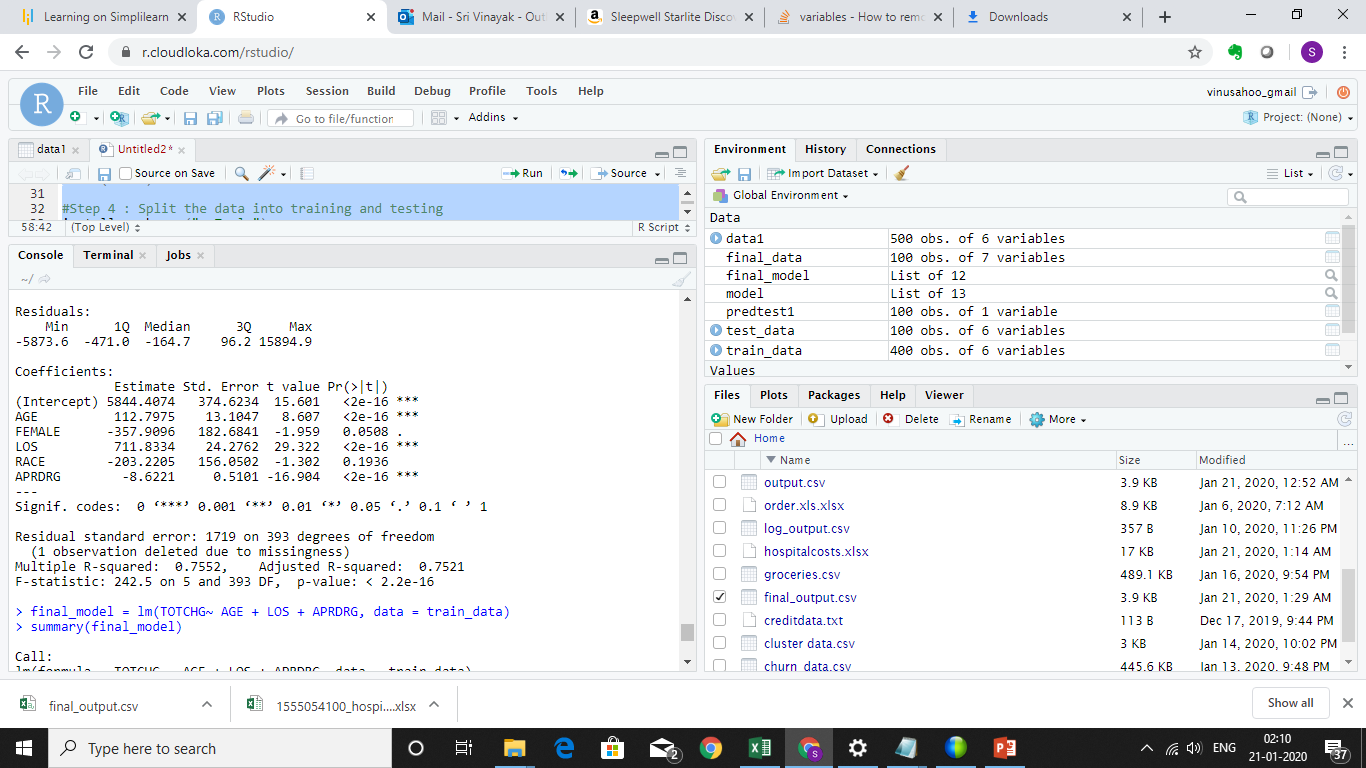
head(predtest)

predtest1<- data.frame(predtest)

final\_data<- cbind(test\_data,predtest1)

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

Screen shot of Coefficient Table of First Model



Screen shot of Coefficient Table of Final Model

