library(dplyr)

library(readxl)

library(writexl)

library(ggmap)

library(tm)

library(RColorBrewer)

library(lubridate)

library(ggpubr)

library(ggplot2)

library(stringi)

library(GlmSimulatoR)

library(correlation)

setwd("E:/Assignment/project/7th project")

getwd()

hosp <- read\_excel("E:/Assignment/project/7th project/1555054100\_hospitalcosts.xlsx")

class(hosp)

#Data should be in fata frame formate

hosp <- data.frame(hosp)

class(hosp)

str(hosp)

summary(hosp)

attach(hosp)

hosp

AGE <- as.factor(AGE)

levels(AGE)

summary(AGE)

#Cost analysis

tapply(TOTCHG,AGE,sum)

which.max(tapply(TOTCHG,AGE,sum))

APRDRG <- as.factor(APRDRG)

APRDRG

summary(APRDRG)

tapply(TOTCHG,APRDRG,sum)

which.max(tapply(TOTCHG,APRDRG,sum))

max(tapply(TOTCHG,APRDRG,sum))

class(RACE)

RACE <- as.factor(RACE)

class(RACE)

str(RACE)

summary(AGE)

hosp <- na.omit(hosp)

sum(is.na(hosp))

summary(hosp)

#H0 shows the RACE opf the patient is related to the Hospital cost &

#H1 shows there is no relation between RACE and Hospital cost

ANOVA <- aov(hosp$LOS~hosp$AGE+hosp$FEMALE+hosp$RACE)

ANOVA

summary(ANOVA)

#P-value of LOS is high so LOS doesnt depend over age,gender and RACE.

#Reject the Null hypothesis

# making model on which on which hospital cost depends

model1 <- lm(hosp$TOTCHG~ .,data = hosp)

model1

summary(model1)

summarise(group\_by(hosp$AGE,hosp$LOS,hosp$TOTCHG))