**Programming Codes:**

**#Reading Hospital cost data**

library(dplyr)

hc <- read.csv('E:\\Simplilearn\\Data Science with R\\Samriddhi Data\\HealthCare\\HospitalCosts.csv')

head(hc)

#Changing RACE & FEMALE, APRDRG into factor

hc$RACE <- as.factor(hc$RACE)

hc$FEMALE <- as.factor(hc$FEMALE)

hc$APRDRG\_Factor <- as.factor(hc$APRDRG)

**#To record the patient statistics, the agency wants to find the age category of people who frequent the hospital and has the maximum expenditure.**

hist(hc$AGE,main = "Frequency of patients",col = "orange",xlab = "Age") # hisogram

summary(as.factor(hc$AGE)) # summary of age data

df <- aggregate(TOTCHG ~ AGE, FUN = sum, data = hc)

df[(df$TOTCHG == max(df$TOTCHG)),]

# Creating age categories age\_cat

hc$age\_cat <- ifelse((hc$AGE < 1), "infant",

ifelse(hc$AGE < 3, 'toddler',

ifelse(hc$AGE < 11, 'child',

'adolescent')))

hc$age\_cat <- as.factor(hc$age\_cat)

df <- aggregate(TOTCHG ~ age\_cat, FUN = sum, data = hc)

df[(df$TOTCHG == max(df$TOTCHG)),]

barplot(table(hc$age\_cat), xlab = "age\_categories",ylab = "Frequency",col="green",border="red", main = "Frequency vs. Age Categories")

**#In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis related group that has maximum hospitalization and expenditure.**

barplot(table(hc$APRDRG), xlab = "Treatment Codes",ylab = "Frequency",col="blue",border="pink", main = "Frequency vs. Age Categories")

x <- aggregate(TOTCHG ~ APRDRG, FUN = sum, data = hc)

x[(x$TOTCHG == max(x$TOTCHG)),]

z<-aggregate(LOS ~ APRDRG, FUN = sum, data = hc)

z[(z$LOS == max(z$LOS)),]

#or using this

hc%>%group\_by(hc$APRDRG)%>%summarise(LOS=sum(LOS),EXP=sum(TOTCHG))->res

as.data.frame(res)->res

res[which(res$LOS==max(res$LOS)),]->los\_totalcost

los\_totalcost

y<-aggregate(TOTCHG ~ LOS, FUN = sum, data = hc)

y[(y$LOS == max(y$LOS)),]

s<-aggregate(LOS ~ APRDRG, FUN = max, data = hc)

s[(s$LOS == max(s$LOS)),]

#or using this

hc%>%group\_by(hc$APRDRG)%>%summarise(LOS=max(LOS),EXP=sum(TOTCHG))->maxlos

as.data.frame(maxlos)->maxlos

maxlos[which(maxlos$LOS==max(maxlos$LOS)),]->maxlos\_exp

maxlos\_exp

**#To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.**

model <- aov(hc$TOTCHG ~ hc$RACE, data = hc) #numerical ~ categorical variable

summary(model)

summary(hc$RACE)

**#To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for proper allocation of resources.**

model1 = aov(TOTCHG ~ FEMALE + AGE, data = hc)

summary(model1)

**#Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.**

model2 <-lm(LOS ~ AGE +FEMALE +RACE, data = hc)

summary(model2)

**#To perform a complete analysis, the agency wants to find the variable that mainly affects the hospital costs.**

hc <- read.csv("HospitalCosts.csv")

model3 <- lm(TOTCHG ~ ., data = hc)

summary(model3)

-------------------------------------------------------------------The End-------------------------------------------------------------------