**Loading the file**

data=read.csv("D:/shamrin/Heart disease/correct/Heart.csv")

str(data)

head(data)

str(data)

data[data == "?"] <- NA

data[data$sex == 0,]$sex <- "F"

data[data$sex == 1,]$sex <- "M"

data$sex <- as.factor(data$sex)

data$cp <- as.factor(data$cp)

data$fbs <- as.factor(data$fbs)

data$restecg <- as.factor(data$restecg)

data$exang <- as.factor(data$exang)

data$slope <- as.factor(data$slope)

data$ca <- as.integer(data$ca)

data$ca <- as.factor(data$ca

data$thal <- as.integer(data$thal)

data$thal <- as.factor(data$thal)

data$hd <- ifelse(test=data$hd == 0, yes="Healthy",no="Unhealthy")

data$hd <- as.factor(data$hd)

str(data)

nrow(data[is.na(data$ca) | is.na(data$thal),])

data[is.na(data$ca) | is.na(data$thal),]

nrow(data)

data <- data[!(is.na(data$ca) | is.na(data$thal)),]

nrow(data)

attach(data)

**Packages used**

library(Amelia)

library(mlbench)

library(corrplot)

library (car)

library(ggplot2)

library(generalhoslem)

library(pROC)

library(ROCR)

**Missing Values Map**

missmap(data, col=c("blue", "red"), legend=FALSE)

**Corrplot**

Correlations=cor(data[,c("age","trestbps","chol","thalach","oldpeak")])

corrplot(correlations, method="square")

**Scatterplotmatrix**

pairs(~age+trestbps+chol+thalach+oldpeak,data=data,main=" Scatterplot Matrix",col="darkred")

**TO create contigency tables**

xtabs(~ hd + age, data=data)

xtabs(~ hd + sex, data=data)

xtabs(~ hd + cp, data=data)

xtabs(~ hd + fbs, data=data)

xtabs(~ hd + restecg, data=data)

xtabs(~ hd + exang, data=data)

xtabs(~ hd + slope, data=data)

xtabs(~ hd + ca, data=data)

xtabs(~ hd + thal, data=data)

xtabs(~ hd + trestbps, data=data)

xtabs(~ hd + oldpeak, data=data)

xtabs(~ hd + chol, data=data)

xtabs(~ hd + thalach, data=data)

**Correlation Analysis**

**Age v/s resting BP**

cor.test(age,trestbps, method=c("pearson"))

ggplot(data, aes(age,trestbps)) +

geom\_point()+geom\_smooth()+facet\_grid(hd~ .)+ theme(axis.title=element\_text(face="bold",size="14", color="brown"))

**Age v/s chol**

cor.test(age,chol, method=c("pearson"))

ggplot(data, aes(age,chol)) +

geom\_point()+geom\_smooth()+facet\_grid(hd~ .)+ theme(axis.title=element\_text(face="bold",size="14", color="blue"))

**Age v/s thalach**

cor.test(age,thalach, method=c("pearson"))

ggplot(data, aes(age,thalach)) +

geom\_point()+geom\_smooth()+facet\_grid(hd~ .)+ theme(axis.title=element\_text(face="bold",size="14", color="red"))

**AGE**

hist(age,main="Histogram of Age",col=c("skyblue","white"),xlab="Age",ylab="Frequency",xlim=c(20,90))

ggplot(data,aes(x=hd,y=age),xlab="HeartDisease")+geom\_boxplot(col="red")+labs(title="Age v/s Heart disease",xlab="Heart Disease",ylab="Age")+ theme(axis.title=element\_text(face="bold",size="14", color="red"))

age<- glm(hd ~ age, data=data, family="binomial")

summary(age)

logitgof(data$hd,fitted(age))

**SEX**

ggplot(data,aes(x=sex))+geom\_bar(fill="darkred")+ theme(axis.title=element\_text(face="bold",size="14", color="blue"))

ggplot(data,aes(x=sex))+geom\_bar(fill="purple")+facet\_wrap(~hd)+ theme(axis.title=element\_text(face="bold",size="14", color="blue"))

sex<- glm(hd ~ sex, data=data, family="binomial")

summary(sex)

logitgof(data$hd,fitted(sex))

**CP**

x=table(data$cp)

barplot(x,names=c("type1","type2","type3","type4"),xlab="Chest pain",ylab="frequency",ylim=c(0,160),xlim=c(0,5))

df=table(data$hd,data$cp)

barplot(df,beside=T,col=c("salmon","skyblue"),names=c("type1","type2","type3","type4"),xlab="Chest pain type",ylab="frequency",ylim=c(0,130),legend=rownames(df))

cp <- glm(hd ~ cp, data=data, family="binomial")

summary(cp)

logitgof(data$hd,fitted(cp))

**Trestbps**

hist(trestbps,xlim=c(90,210),ylim=c(0,80),col=c("red","lightgreen"),xlab="Resting Blood Pressure",beside=T)

d=density(data$trestbps)

plot(d,main="density plot of resting BP")

polygon(d,col="grey",border="blue")

rug(trestbps,col="brown")

ggplot(data, aes(x=trestbps, fill=hd)) +geom\_histogram(binwidth=8,position="dodge")

ggplot(data, aes(x=trestbps, fill=hd)) + geom\_density(alpha=.3)

trestbps <- glm(hd ~ trestbps, data=data, family="binomial")

summary(trestbps)

logitgof(data$hd,fitted(trestbps))

**CHOL**

hist(chol,xlim=c(0,620),ylim=c(0,140),col=c("lightgreen","yellow"),main="Histogram of Cholestrol in mg/dl",xlab="cholestrol")

ggplot(data, aes(x=chol, fill=hd)) +geom\_histogram(binwidth=30,position="dodge")+facet\_wrap(~hd)

chol<- glm(hd ~ chol, data=data, family="binomial")

summary(chol)

logitgof(data$hd,fitted(chol))

**FBS**

data$fbs <- ifelse(test=data$fbs == 0, yes="No", no="Yes")

data$fbs <- as.factor(data$fbs)

attach(data)

x=table(fbs,hd)

x

barplot(x,horiz=T,col=c("lightpink","yellow"),xlim=c(0,200),ylim=c(0,3),legend=rownames(x))

fbs <- glm(hd ~ fbs, data=data, family="binomial")

summary(fbs)

logitgof(data$hd,fitted(fbs))

**Restecg**

library(wesanderson)

ggplot(data, aes(x=restecg)) + geom\_bar( color="darkred")+scale\_color\_grey() + theme\_classic()+facet\_wrap(~hd)

restecg<- glm(hd ~ restecg, data=data, family="binomial")

summary(restecg)

logitgof(data$hd,fitted(restecg))

**THalach**

**>** hist(thalach,main="Histogram of max Heart rate during Stress test",ylim=c(0,100),col=c("pink","steelblue","magenta"),xlab="max Heart rate during Stress Test")

ggplot(data, aes(x=hd, y=thalach)) + geom\_boxplot(color="blue",fill=c("yellow","white"))

thalach<- glm(hd ~ thalach, data=data, family="binomial")

summary(thalach)

logitgof(data$hd,fitted(thalach))

ggplot(data, aes(x=thalch, fill=hd)) +geom\_histogram(binwidth=8)

**Exang**

data$exang <- ifelse(test=data$exang == 0, yes="No", no="Yes")

data$exang <- as.factor(data$exang)

df=table(exang,hd)

barplot(df,beside=T,col=c("navy","skyblue"),names=c(),xlab="Exercise induced Angina",ylab="frequency",ylim=c(0,180),legend=rownames(df),main="Exercise induced Angina pain")

data$exang <- ifelse(test=data$exang == 0, yes="no", no="yes")

data$exang <- as.factor(data$exang)

pie = ggplot(data, aes(x="", y=hd, fill=exang)) + geom\_bar(stat="identity", width=1)+facet\_wrap(~hd)+ labs(title = "Exercise induced Angina v/s Heart Disease")

pie = pie + coord\_polar("y", start=0,direction=-1)+theme\_void()

pie = pie + scale\_fill\_manual(values=c("#55DDE0", "#33658A", "#2F4858", "#F6AE2D", "#F26419", "#999999"))

geom\_text(aes(y = factor(exang)/3 + c(0, cumsum(value)[-length(value)]), label = percent(factor(exang)/100)), size=5)

exang<- glm(hd ~ exang, data=data, family="binomial")

summary(exang)

logitgof(data$hd,fitted(exang))

**old peak**

qqnorm(data$oldpeak, pch = 1, frame = FALSE,main="Old Peak",ylim=c(0,10))

> qqline(data$oldpeak, col = "steelblue", lwd = 2)

ggplot(data, aes(x=oldpeak, fill=hd)) +geom\_histogram(binwidth=8,position="dodge")+scale\_fill\_manual(values=c("#55DDE0", "#33658A", "#2F4858", "#F6AE2D", "#F26419", "#999999"))

logistic <- glm(hd ~ oldpeak, data=data, family="binomial")

summary(logistic)

logitgof(data$hd,fitted(logistic))

**Slope**

levels(data$slope) <- c("Upsloping","Flat","Downsloping")

levels(data$slope)

df=data.frame(Make=c("Upsloping","Flat","Downsloping"),Cnt=c(139,137,21))

head(df)

dfc=df %>%

group\_by(Make) %>%

summarise(volume = sum(Cnt)) %>%

mutate(share=volume/sum(volume)\*100.0) %>%

arrange(desc(volume))

bp <- ggplot(dfc[c(1:10),], aes(x="", y= share, fill=Make)) +

geom\_bar(width = 1, stat = "identity")

pie <- bp + coord\_polar("y")

pie

dfc$Make <- factor(dfc$Make, levels = rev(as.character(dfc$Make)))

ggplot(dfc[1:10, ], aes("", share, fill = Make)) +

geom\_bar(width = 1, size = 1, color = "white", stat = "identity") +

coord\_polar("y") +

geom\_text(aes(label = paste0(round(share), "%")),

position = position\_stack(vjust = 0.5)) +

labs(x = NULL, y = NULL, fill = NULL,

title = "Slope") +

guides(fill = guide\_legend(reverse = TRUE)) +

scale\_fill\_manual(values = c("red","salmon","lightblue")) +

theme\_classic() +

theme(axis.line = element\_blank(),

axis.text = element\_blank(),

axis.ticks = element\_blank(),

plot.title = element\_text(hjust = 0.5, color = "#666666"))

g=ggplot(data,aes(slope))

g + geom\_bar(aes(slope = ,fill=slope))+facet\_wrap(~hd)+labs(title="Barchart of Slope")

logistic <- glm(hd~slope, data=data, family="binomial")

summary(logistic)

logitgof(data$hd,fitted(logistic))

**Ca**

qqnorm(data$ca, pch = 1, frame = FALSE,main="Number of major vessels coloured by flouroscopy ",ylim=c(0,10))

qqline(data$ca, col = "red", lwd = 2)

ggplot(data, aes(x=ca, fill=hd)) +geom\_histogram(binwidth=8,position="dodge")+scale\_fill\_manual(values=c("#2F4858", "#F6AE2D", "#F26419", "#999999"))+labs(title="Number of vesssels coloured by flouroscopy")

logistic <- glm(hd~ca, data=data, family="binomial")

summary(logistic)

logitgof(data$hd,fitted(logistic))

**Thal**

levels(data$thal) <- c("N/A", "Normal", "Fixed D","Reversible D")

levels(data$thal)

g=ggplot(data,aes(thal))

g + geom\_bar(aes(thal = ,fill=thal))+facet\_wrap(~hd)+labs(title="Barchart of Thallium Stress Test ")

logistic <- glm(hd~thal, data=data, family="binomial")

summary(logistic)

logitgof(data$hd,fitted(logistic))

**Heart Disease**

df=data.frame(Make=c("Healthy","Unhealthy"),Cnt=c(160,137))

head(df)

dfc <- df %>%

group\_by(Make) %>%

summarise(volume = sum(Cnt)) %>%

mutate(share=volume/sum(volume)\*100.0) %>%

arrange(desc(volume))

bp <- ggplot(dfc[c(1:10),], aes(x="", y= share, fill=Make)) +

geom\_bar(width = 1, stat = "identity")

pie <- bp + coord\_polar("y")

pie

dfc$Make <- factor(dfc$Make, levels = rev(as.character(dfc$Make)))

ggplot(dfc[1:10, ], aes("", share, fill = Make)) +

geom\_bar(width = 1, size = 1, color = "white", stat = "identity") +

coord\_polar("y") +

geom\_text(aes(label = paste0(round(share), "%")),

position = position\_stack(vjust = 0.5)) +

labs(x = NULL, y = NULL, fill = NULL,

title = "Heart Disease") +

guides(fill = guide\_legend(reverse = TRUE)) +

scale\_fill\_manual(values = c("#ffd700","#254290")) +

theme\_classic() +

theme(axis.line = element\_blank(),

axis.text = element\_blank(),

axis.ticks = element\_blank(),

plot.title = element\_text(hjust = 0.5, color = "#666666"))

**ROC CURVE**

library(nnet)

mymodel=multinom(hd~.,data=data)

p=predict(mymodel,data)

tab=table(p,data$hd)

tab

table(data$hd)

pred=predict(mymodel,data,type='prob')

pred=prediction(pred,data$hd)

eval=performance(pred,"acc")

plot(eval)

max=which.max(slot(eval,"y.values")[[1]])

acc=slot(eval,"y.values")[[1]][max]

acc

cut=slot(eval,"x.values")[[1]][max]

print(c(Accuracy=acc,Cutoff=cut))

roc=performance(pred,"tpr","fpr")

plot(roc,colorize=T,main="ROC Curve of Heart disease Prediction",ylab="Sensitivity",xlab="1-specificity")

abline(a=0,b=1)

auc=performance(pred,"auc")

auc=unlist(slot(auc,"y.values"))

auc=round(auc,4)

legend(0.6,0.2,auc,title="AUC",cex=1.2)

**Stepwise Regression**

L=step(lm(hd~.,data=data),direction="backward")

M=step(lm(hd~1,data=data),direction="forward",scope=~age+sex+cp+trestbps+chol+fbs+restecg+thalach+exang+oldpeak+slope+ca+thal)

step(lm(hd~.,data=data),direction="both")