**Credit Risk Modeling**

**Code-**

# Load required libraries

library(dplyr)

library(ggplot2)

library(caret)

library(pROC)

# Read the dataset

df <- read.csv("bank-loan.csv")

# STEP 1: Clean and Explore the Data

df <- df %>%

mutate(default = as.factor(default)) # Convert target to factor

# Basic visualization

ggplot(df, aes(x = income, fill = default)) +

geom\_histogram(bins = 30, position = "identity", alpha = 0.6) +

labs(title = "Income Distribution by Loan Default")

# STEP 2: Split into training and test sets

set.seed(123)

train\_index <- createDataPartition(df$default, p = 0.7, list = FALSE)

train\_data <- df[train\_index, ]

test\_data <- df[-train\_index, ]

# STEP 3: Build Logistic Regression Model

model <- glm(default ~ ., data = train\_data, family = binomial)

# STEP 4: Predict probabilities and classify

probabilities <- predict(model, test\_data, type = "response")

predicted\_class <- ifelse(probabilities > 0.5, 1, 0)

predicted\_class <- factor(predicted\_class, levels = c(0, 1))

# STEP 5: Evaluate Model

conf\_matrix <- confusionMatrix(predicted\_class, test\_data$default)

print(conf\_matrix)

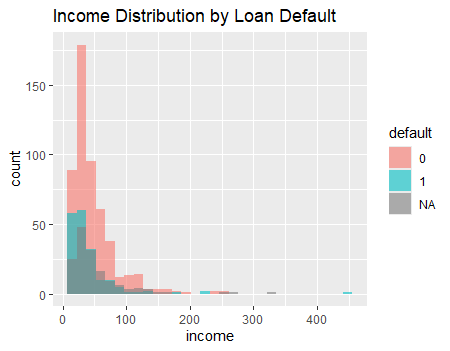
# ROC Curve and AUC

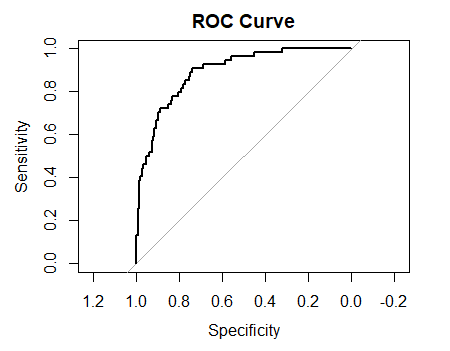
roc\_obj <- roc(as.numeric(as.character(test\_data$default)), probabilities)

plot(roc\_obj, main = "ROC Curve")

cat("AUC =", auc(roc\_obj), "\n")

**GRAPHS**

****

****