# Set seed for reproducibility

set.seed(123)

# Simulating dataset with features affecting ad clicks

n <- 500 # Number of observations

ad\_data <- data.frame(

ad\_position = sample(1:5, n, replace = TRUE), # Position of the ad on the webpage (1 = top, 5 = bottom)

ad\_duration = runif(n, min = 1, max = 10), # Duration in seconds the ad was shown

user\_engagement = runif(n, min = 0, max = 1), # User engagement score (0-1)

past\_clicks = rpois(n, lambda = 2), # Number of past ad clicks

device\_type = sample(c(0, 1), n, replace = TRUE), # 0 = Desktop, 1 = Mobile

clicked = sample(c(0, 1), n, replace = TRUE, prob = c(0.7, 0.3)) # Target variable (0 = Not Clicked, 1 = Clicked)

)

# ✅ Display first few rows of dataset

print(head(ad\_data))

# ✅ Split data into training (70%) and testing (30%)

train\_index <- sample(1:n, size = 0.7 \* n, replace = FALSE)

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

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

# ✅ Train Logistic Regression Model (Binary Classification)

model <- glm(clicked ~ ad\_position + ad\_duration + user\_engagement + past\_clicks + device\_type,

data = train\_data, family = binomial)

# ✅ Model Summary

summary(model)

# ✅ Make predictions on test data

test\_data$predicted <- predict(model, newdata = test\_data, type = "response")

# ✅ Convert probabilities to binary classes (0 or 1)

test\_data$predicted\_class <- ifelse(test\_data$predicted > 0.5, 1, 0)

# ✅ Display first few predicted values

print(head(test\_data))

# ✅ Compute Accuracy

accuracy <- sum(test\_data$predicted\_class == test\_data$clicked) / nrow(test\_data)

cat("Model Accuracy:", accuracy, "\n")