Program:

# ✅ STEP 1: Import Libraries

library(dplyr)

library(ggplot2)

# ✅ STEP 2: Create a Sample Dataset

dataset <- data.frame(

Name = c("Alice", "Bob", "Charlie", "David", "Eve", "Frank", "Grace", "Alice"),

Age = c(25, 30, 35, 40, NA, 30, 29, 25),

Salary = c(50000, 60000, 55000, 70000, 62000, NA, 52000, 50000),

Department = c("HR", "Finance", "IT", "HR", "Finance", "IT", NA, "HR"),

stringsAsFactors = FALSE

)

# ✅ STEP 3: Explore Dataset

cat("Original Dataset:\n")

print(dataset)

cat("\nDataset Structure:\n")

str(dataset)

cat("\nSummary Statistics:\n")

summary(dataset)

# ✅ STEP 4: Data Cleaning

# Remove rows with missing values

clean\_data <- na.omit(dataset)

# Remove duplicate rows

clean\_data <- distinct(clean\_data)

cat("\nCleaned Dataset:\n")

print(clean\_data)

# ✅ STEP 5: Variable Filtering

filtered\_columns <- clean\_data %>% select(Name, Age, Salary)

cat("\nFiltered Columns (Name, Age, Salary):\n")

print(filtered\_columns)

# ✅ STEP 6: Row Filtering

filtered\_rows <- filtered\_columns %>% filter(Age > 30)

cat("\nFiltered Rows (Age > 30):\n")

print(filtered\_rows)

# ✅ STEP 7: Data Visualization (Age vs Salary)

ggplot(clean\_data, aes(x = Age, y = Salary)) +

geom\_point(color = "blue", size = 3) +

ggtitle("Age vs. Salary") +

theme\_minimal() +

xlab("Age") +

ylab("Salary")

Output:





