**PROGRAM:**

if (!require(dplyr)) install.packages("dplyr")

library(dplyr)

df <- data.frame(

id = 1:6,

name = c("Alice", "Bob", "Carol", "Dave", NA, "Eve"),

age = c(25, 17, 30, 22, NA, 40),

income = c(50000, 45000, NA, 60000, 52000, 55000),

gender = c("Female", "Male", "Female", "Male", "Female", NA),

constant\_col = c("A", "A", "A", "A", "A", "A")

)

clean\_df <- df %>%

select(where(~ mean(is.na(.)) < 0.5)) %>%

select(where(~ n\_distinct(., na.rm = TRUE) > 1)) %>%

select(-any\_of(c("id", "constant\_col"))) %>%

filter(!is.na(age), !is.na(income), !is.na(gender)) %>%

filter(age >= 18) %>%

distinct() %>%

filter(

income >= quantile(income, 0.25, na.rm = TRUE) - 1.5 \* IQR(income, na.rm = TRUE),

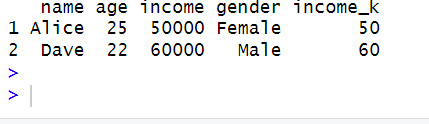
income <= quantile(income, 0.75, na.rm = TRUE) + 1.5 \* IQR(income, na.rm = TRUE)

) %>%

mutate(income\_k = income / 1000)

print(clean\_df)

**OUTPUT:**

**A screenshot of a computer

AI-generated content may be incorrect.**