install.packages('dplyr')

car\_data <- read.csv(file.choose())

car\_data

#a

colSums(is.na(car\_data))

car\_data <- data.frame(car\_data)

km\_driven\_median <- median(car\_data$Kilometers\_Driven, na.rm = TRUE)

km\_driven\_median

car\_data$Kilometers\_Driven <- ifelse(is.na(car\_data$Kilometers\_Driven), km\_driven\_median, car\_data$Kilometers\_Driven)

car\_data$Seats

seats\_median <- median(car\_data$Seats, na.rm = TRUE)

car\_data$Seats <- ifelse(is.na(car\_data$Seats), seats\_median, car\_data$Seats)

car\_data

#b

car\_data$Mileage <- as.character(car\_data$Mileage)

car\_data$Power <- as.character(car\_data$Power)

car\_data$Engine <- as.character(car\_data$Engine)

car\_data$New\_Price <- as.character(car\_data$New\_Price)

car\_data$Mileage

car\_data$Mileage <- sapply(strsplit(car\_data$Mileage, " "), function(x) x[1])

car\_data$Power <- sapply(strsplit(car\_data$Power, " "), function(x) x[1])

car\_data$Engine <- sapply(strsplit(car\_data$Engine, " "), function(x) x[1])

car\_data$New\_Price <- sapply(strsplit(car\_data$New\_Price, " "), function(x) x[1])

car\_data

#c

install.packages(caret)

library(caret)

# Sample dataframe

df\_fuel\_type <- data.frame(car\_data$Fuel\_Type)

df\_transmission <- data.frame(car\_data$Transmission)

# Create dummy variables

dummies\_fuel\_type <- dummyVars("~ .", data = df\_fuel\_type)

dummies\_transmission <- dummyVars("~ .", data = df\_transmission)

# Apply to dataframe

df\_encoded\_fuel\_type <- predict(dummies\_fuel\_type, newdata = df\_fuel\_type)

df\_encoded\_transmission <- predict(dummies\_transmission, newdata = df\_transmission)

# Convert to a dataframe

df\_final\_fuel\_type <- as.data.frame(df\_encoded\_fuel\_type)

df\_final\_transmission <- as.data.frame(df\_encoded\_transmission)

# Combine with original data

df\_final <- cbind(df\_fuel\_type, df\_final\_fuel\_type, df\_transmission, df\_final\_transmission)

# Print the final dataframe

print(df\_final)

car\_data\_latest <- cbind(car\_data, df\_final)

car\_data\_latest

head(car\_data\_latest)

car\_data

#d

library(dplyr)

car\_data <- mutate(car\_data, Current\_Age = as.numeric(format(Sys.Date(), "%Y")) - Year)

car\_data

#select

selected\_df <- select(car\_data, Model\_Name, Year, Selling\_Price, Location, Owner\_Type)

#filter

filtered\_df <- filter(car\_data, Year == 2018, Seats == 5)

#rename

renamed\_df <- rename(car\_data, Updated\_Price = New\_Price)

#mutate has already been used

car\_data <- mutate(car\_data, Current\_Age = as.numeric(format(Sys.Date(), "%Y")) - Year)

car\_data

#arrange

arranged\_data <- arrange(car\_data\_latest, Year)

#summary

summary\_data <- car\_data\_latest %>%

group\_by(Transmission) %>%

summarize(Avg\_Price = mean(Selling\_Price), Max\_Year = max(Year))









