Used-Car-Price-Prediction-ML

Kyle Su

2024-12-10

K-Prototype

```
library(factoextra)
library(cluster)
library(ggplot2)
library(dplyr)

if (!requireNamespace("clustMixType", quietly = TRUE)) {
  stop("Package 'clustMixType' is required. Please install it before knitting.")
}
library(clustMixType)

data <- read.csv("final_data_car.csv")

# Convert character columns to factors
data <- data %>%
  mutate(across(where(is.character), as.factor))

# Separate numeric and categorical data with aligned rows
numeric_data <- data %>% select(where(is.numeric))
complete_idx <- complete.cases(numeric_data)

data <- data[complete_idx, ]
numeric_data <- numeric_data[complete_idx, ]

categorical_data <- data %>% select(where(is.factor))

# Scale the numeric data
scaled_numeric_data <- scale(numeric_data)

# Combine scaled numeric data with categorical data
combined_data <- cbind(as.data.frame(scaled_numeric_data), categorical_data)
```

```
# Function to calculate WSS for different numbers of clusters
wss_values <- numeric(20)
for (k in 1:20) {
  set.seed(42)
  kproto_model <- kproto(combined_data, k = k, nstart = 5)
  wss_values[k] <- kproto_model$tot.withinss
}
```

```
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
```

```
##                 0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
```

```
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
```

```
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
```

```
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
```

```
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
```

```
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
```

```
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
```

```
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
```

```
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
```

```
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
```

```
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                    0
##              brand                model            fuel_type               engine
##                  0                    0                    0                    0
##       transmission              ext_col              int_col             accident
##                  0                    0                    0                    0
##        clean_title
```

```
##                    0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
##        transmission              ext_col              int_col            accident
##                   0                    0                    0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage                price                  age Engine_Displacement
##                   0                    0                    0                   0
##               brand                model            fuel_type              engine
##                   0                    0                    0                   0
```

```
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
```

```
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
```

```
##
## # NAs in variables:
##               milage                price                  age Engine_Displacement
##                    0                    0                    0                    0
##                brand                model            fuel_type               engine
##                    0                    0                    0                    0
##         transmission              ext_col              int_col             accident
##                    0                    0                    0                    0
##          clean_title
##                    0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##               milage                price                  age Engine_Displacement
##                    0                    0                    0                    0
##                brand                model            fuel_type               engine
##                    0                    0                    0                    0
##         transmission              ext_col              int_col             accident
##                    0                    0                    0                    0
##          clean_title
##                    0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##               milage                price                  age Engine_Displacement
##                    0                    0                    0                    0
##                brand                model            fuel_type               engine
##                    0                    0                    0                    0
##         transmission              ext_col              int_col             accident
##                    0                    0                    0                    0
##          clean_title
##                    0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##               milage                price                  age Engine_Displacement
##                    0                    0                    0                    0
##                brand                model            fuel_type               engine
##                    0                    0                    0                    0
##         transmission              ext_col              int_col             accident
##                    0                    0                    0                    0
##          clean_title
##                    0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##               milage                price                  age Engine_Displacement
##                    0                    0                    0                    0
##                brand                model            fuel_type               engine
##                    0                    0                    0                    0
##         transmission              ext_col              int_col             accident
##                    0                    0                    0                    0
##          clean_title
##                    0
```

```
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##              milage                 price                   age Engine_Displacement
##                   0                     0                     0                   0
##               brand                 model             fuel_type              engine
##                   0                     0                     0                   0
##        transmission               ext_col               int_col            accident
##                   0                     0                     0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                 price                   age Engine_Displacement
##                   0                     0                     0                   0
##               brand                 model             fuel_type              engine
##                   0                     0                     0                   0
##        transmission               ext_col               int_col            accident
##                   0                     0                     0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                 price                   age Engine_Displacement
##                   0                     0                     0                   0
##               brand                 model             fuel_type              engine
##                   0                     0                     0                   0
##        transmission               ext_col               int_col            accident
##                   0                     0                     0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                 price                   age Engine_Displacement
##                   0                     0                     0                   0
##               brand                 model             fuel_type              engine
##                   0                     0                     0                   0
##        transmission               ext_col               int_col            accident
##                   0                     0                     0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage                 price                   age Engine_Displacement
##                   0                     0                     0                   0
##               brand                 model             fuel_type              engine
##                   0                     0                     0                   0
##        transmission               ext_col               int_col            accident
```

```
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
```

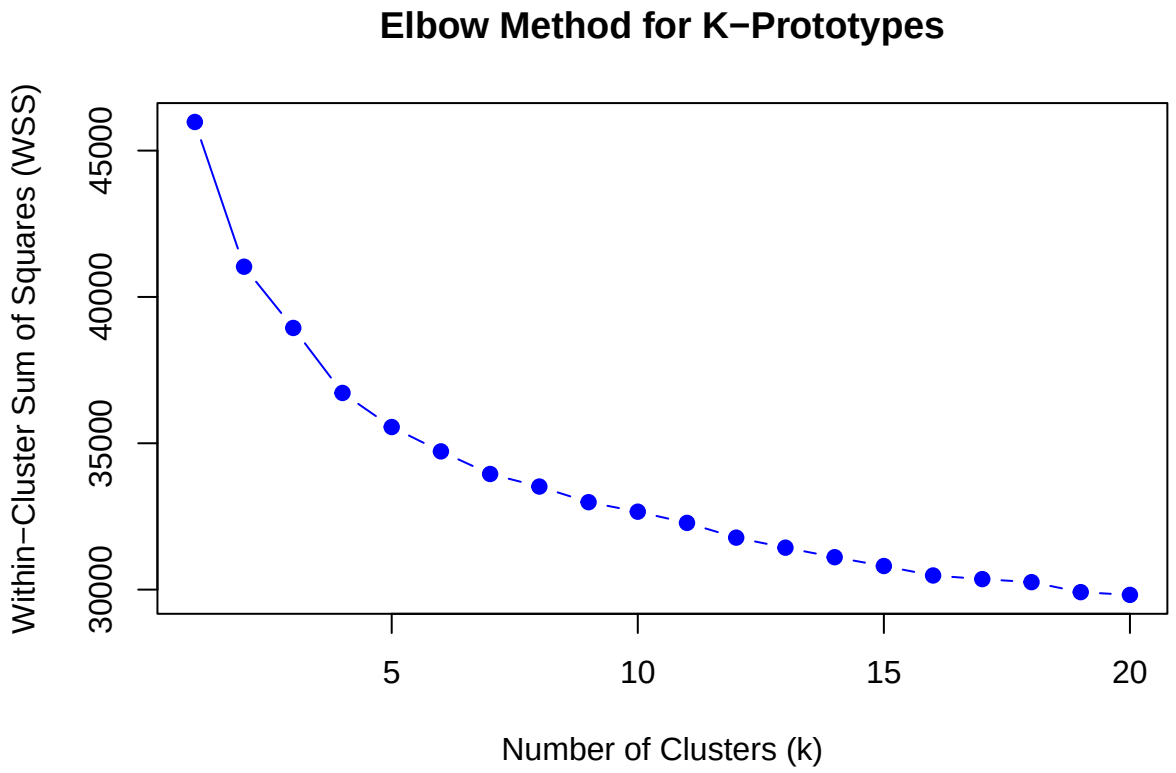
```
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage              price                age Engine_Displacement
##                  0                  0                  0                  0
##              brand              model          fuel_type             engine
##                  0                  0                  0                  0
##       transmission            ext_col            int_col           accident
##                  0                  0                  0                  0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
```

```
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage               price                 age Engine_Displacement
##                  0                   0                   0                   0
##              brand               model           fuel_type              engine
##                  0                   0                   0                   0
##       transmission             ext_col             int_col            accident
##                  0                   0                   0                   0
##        clean_title
```

```
##                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                   0
##              brand                model            fuel_type              engine
##                  0                    0                    0                   0
##       transmission              ext_col              int_col            accident
##                  0                    0                    0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.601261
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                   0
##              brand                model            fuel_type              engine
##                  0                    0                    0                   0
##       transmission              ext_col              int_col            accident
##                  0                    0                    0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                   0
##              brand                model            fuel_type              engine
##                  0                    0                    0                   0
##       transmission              ext_col              int_col            accident
##                  0                    0                    0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                   0
##              brand                model            fuel_type              engine
##                  0                    0                    0                   0
##       transmission              ext_col              int_col            accident
##                  0                    0                    0                   0
##        clean_title
##                  0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##             milage                price                  age Engine_Displacement
##                  0                    0                    0                   0
##              brand                model            fuel_type              engine
##                  0                    0                    0                   0
```

```
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
```

```
# Plot the elbow chart
plot(1:20, wss_values, type = "b", pch = 19, col = "blue",
     xlab = "Number of Clusters (k)", ylab = "Within-Cluster Sum of Squares (WSS)",
     main = "Elbow Method for K-Prototypes")
```




```
set.seed(42)
optimal_clusters <- 8
kproto_result <- kproto(combined_data, k = optimal_clusters)
```

```
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title
##                   0
## 0 observation(s) with NAs.
```

```
##
## Estimated lambda: 1.601261
```

```
# Add cluster labels to the dataset
combined_data$cluster <- as.factor(kproto_result$cluster)

# Print the clustering results
print(kproto_result)
```

```
## Distance type: huang
##
## Numeric predictors: 4
## Categorical predictors: 9
## Lambda: 1.601261
##
## Number of Clusters: 8
## Cluster sizes: 512 641 3 441 430 976 381 559
## Within cluster error: 4668.222 5513.099 189.4875 3620.632 4100.679 7623.599 3577.803 4616.786
##
## Cluster prototypes:
##       milage       price        age Engine_Displacement     brand
## 1 -0.8576325  0.58824031 -0.8641147          -0.3750681      Audi
## 2 -0.5214682  0.13167649 -0.5367646           1.3775452 Chevrolet
## 3 -0.9652731 27.55286705  1.4804376           0.3491387   Bugatti
## 4 -0.6978377  0.14597566 -0.7672086          -0.2742626       BMW
## 5  0.5538241 -0.35084336  1.6560107           0.1439026 Chevrolet
## 6 -0.1184169 -0.09041049 -0.1658993          -0.6702820       BMW
## 7  1.9179963 -0.40956777  1.2620644          -0.4461131 Chevrolet
## 8  0.4126747 -0.24591938  0.1598874           0.3420696      Ford
##             model fuel_type                                        engine
## 1    1500 Laramie  Gasoline                             3.6L V6 24V MPFI DOHC
## 2      Camaro 2SS  Gasoline        455.0HP 6.2L 8 Cylinder Engine Gasoline Fuel
## 3 Carrera GT Base  Gasoline        394.0HP 4.2L 8 Cylinder Engine Gasoline Fuel
## 4       X6 M Base  Gasoline                       2.0L I4 16V GDI DOHC Turbo
## 5   Corvette Base  Gasoline        302.0HP 5.0L 8 Cylinder Engine Gasoline Fuel
## 6         M4 Base  Gasoline        240.0HP 2.0L 4 Cylinder Engine Gasoline Fuel
## 7         M3 Base  Gasoline 320.0HP 5.3L 8 Cylinder Engine Flex Fuel Capability
## 8    F-250 Lariat  Gasoline        355.0HP 5.3L 8 Cylinder Engine Gasoline Fuel
##   transmission ext_col int_col                            accident
## 1    Automatic   black   black                       None reported
## 2    Automatic    gray   black                       None reported
## 3    Automatic    gray   black                       None reported
## 4    Automatic    blue   other                       None reported
## 5    Automatic    blue   other                       None reported
## 6    Automatic   white   black                       None reported
## 7    Automatic   black   black At least 1 accident or damage reported
## 8    Automatic   black   black At least 1 accident or damage reported
##   clean_title
## 1          No
## 2         Yes
## 3         Yes
## 4         Yes
## 5         Yes
## 6         Yes
```

```
## 7         Yes
## 8         Yes
```

```
# Summarize cluster means for scaled numeric variables
numeric_summary <- data %>%
  select_if(is.numeric) %>%
  mutate(cluster = kproto_result$cluster) %>%
  group_by(cluster) %>%
  summarize(across(everything(), list(mean = mean, sd = sd), .names = "{.col}_{.fn}"))

print("Numeric Cluster Summary:")
```

```
## [1] "Numeric Cluster Summary:"
```

```
print(numeric_summary)
```

```
## # A tibble: 8 x 9
##   cluster milage_mean milage_sd price_mean price_sd age_mean age_sd
##     <int>       <dbl>     <dbl>      <dbl>    <dbl>    <dbl>  <dbl>
## 1       1      19857.    18129.      89438.   91529.     3.30   2.24
## 2       2      37391.    29121.      54243.   43505.     5.22   3.24
## 3       3      14243.    15408.    2168026   703128.    17      3.46
## 4       4      28192.    21451.      55345.   33324.     3.87   2.33
## 5       5      93474.    41505.      17048.   13707.    18.0    4.14
## 6       6      58412.    32919.      37123.   24181.     7.38   3.45
## 7       7     164624.    45793.      12521.    9398.    15.7    4.37
## 8       8      86112.    34370.      25136.   14652.     9.29   3.23
## # i 2 more variables: Engine_Displacement_mean <dbl>,
## #   Engine_Displacement_sd <dbl>
```

```
data$cluster <- kproto_result$cluster
```

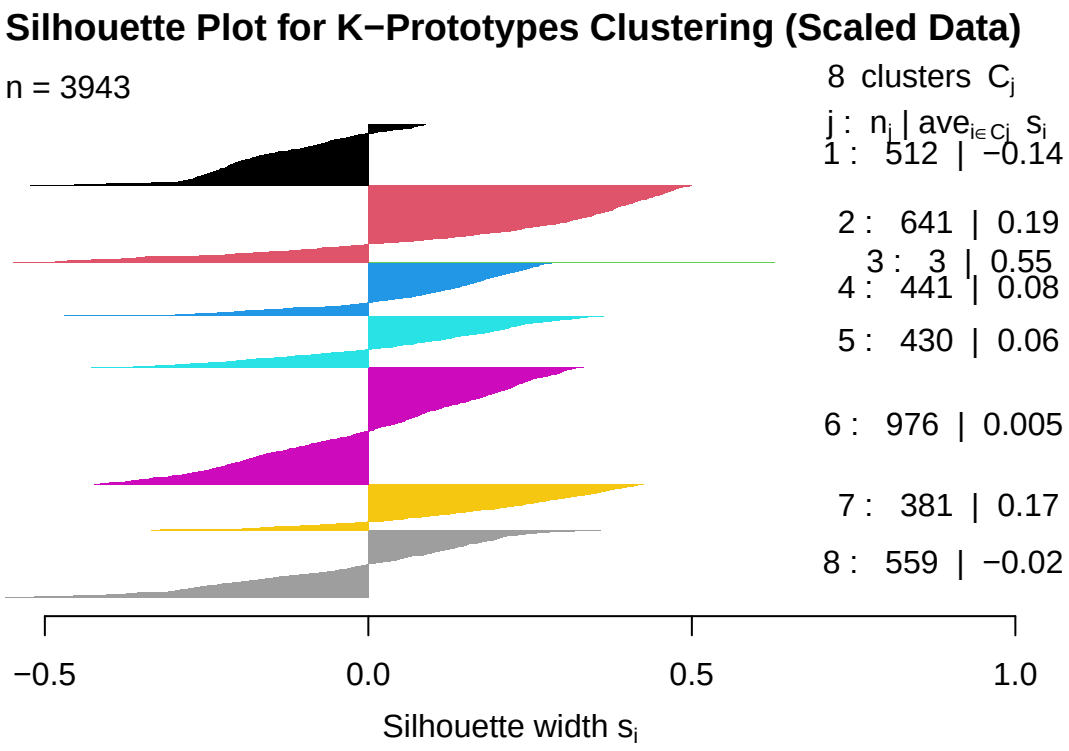
```
# Assess WSS (within-cluster sum of squares)
cat("Within-cluster sum of squares:", kproto_result$tot.withinss, "\n")
```

```
## Within-cluster sum of squares: 33910.31
```

```
# Compute silhouette score using scaled numeric data
silhouette_scores <- silhouette(kproto_result$cluster, dist(scaled_numeric_data))
avg_silhouette <- mean(silhouette_scores[, 3])
cat("Average Silhouette Score:", avg_silhouette, "\n")
```

```
## Average Silhouette Score: 0.04263012
```

```
# Plot silhouette scores
plot(silhouette_scores, main = "Silhouette Plot for K-Prototypes Clustering (Scaled Data)", col = 1:opti
```




```
optimal_clusters <- 8  # Use the optimal k from the elbow chart
set.seed(42)
kproto_result <- kproto(combined_data, k = optimal_clusters, nstart = 5)
```

```
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title             cluster
##                   0                   0
## 0 observation(s) with NAs.
##
## Estimated lambda: 1.547467
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title             cluster
##                   0                   0
```

```
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title             cluster
##                   0                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title             cluster
##                   0                   0
## 0 observation(s) with NAs.
##
## # NAs in variables:
##              milage               price                 age Engine_Displacement
##                   0                   0                   0                   0
##               brand               model           fuel_type              engine
##                   0                   0                   0                   0
##        transmission             ext_col             int_col            accident
##                   0                   0                   0                   0
##         clean_title             cluster
##                   0                   0
## 0 observation(s) with NAs.
```

```r
# Add cluster labels to the dataset
data$cluster <- kproto_result$cluster

# Print clustering results
print(kproto_result)
```

```
## Distance type: huang
##
## Numeric predictors: 4
## Categorical predictors: 10
## Lambda: 1.547467
##
## Number of Clusters: 8
## Cluster sizes: 562 3 378 646 429 429 684 812
## Within cluster error: 4533.351 188.7881 3452.385 5453.864 3997.072 3580.687 6482.699 6244.075
##
## Cluster prototypes:
##        milage       price        age Engine_Displacement         brand
## 1  0.40244177 -0.24495975  0.1584074           0.3513912          Ford
```

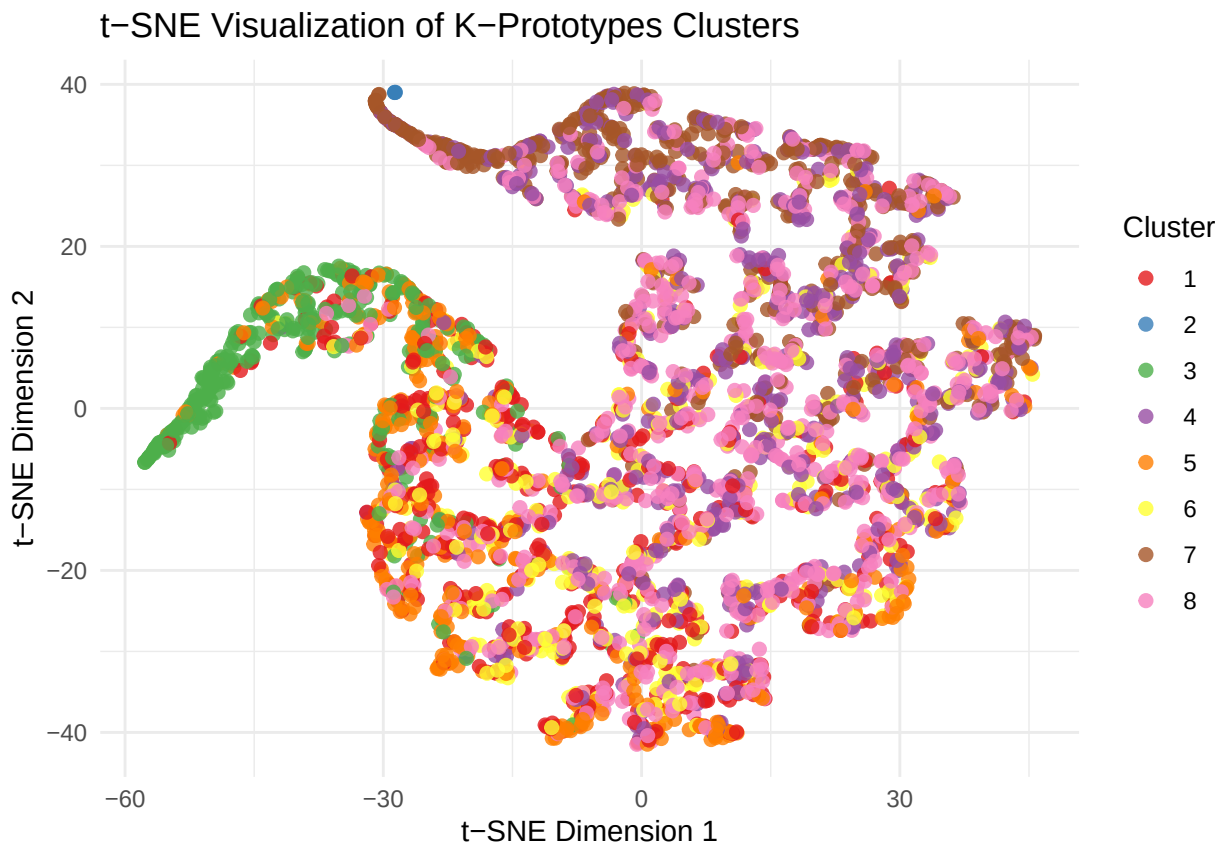
```
## 2 -0.96527306 27.55286705  1.4804376           0.3491387       Bugatti
## 3  1.92310006 -0.41028628  1.2707478          -0.4407567     Chevrolet
## 4 -0.51982933  0.12393313 -0.5375819           1.4005691          Ford
## 5  0.55456428 -0.34946135  1.6663963           0.1499409     Chevrolet
## 6  0.01723221 -0.15805768 -0.1461028          -0.5642620           BMW
## 7 -0.90616277  0.54620680 -0.9142730          -0.3257000     Chevrolet
## 8 -0.29542351 -0.03182709 -0.3120372          -0.6603058 Mercedes-Benz
##                model fuel_type
## 1       F-250 Lariat  Gasoline
## 2    Carrera GT Base  Gasoline
## 3            M3 Base  Gasoline
## 4         Camaro 2SS  Gasoline
## 5      Corvette Base  Gasoline
## 6            M3 Base  Gasoline
## 7       1500 Laramie  Gasoline
## 8 Model Y Long Range  Gasoline
##                                                engine transmission ext_col
## 1        355.0HP 5.3L 8 Cylinder Engine Gasoline Fuel    Automatic   black
## 2        394.0HP 4.2L 8 Cylinder Engine Gasoline Fuel    Automatic    gray
## 3 320.0HP 5.3L 8 Cylinder Engine Flex Fuel Capability    Automatic   black
## 4        420.0HP 6.2L 8 Cylinder Engine Gasoline Fuel    Automatic    gray
## 5        302.0HP 5.0L 8 Cylinder Engine Gasoline Fuel    Automatic   white
## 6        240.0HP 2.0L 4 Cylinder Engine Gasoline Fuel    Automatic   black
## 7                                3.6L V6 24V MPFI DOHC    Automatic   black
## 8       285.0HP 3.6L V6 Cylinder Engine Gasoline Fuel    Automatic   white
##   int_col                                accident clean_title cluster
## 1   black At least 1 accident or damage reported         Yes       8
## 2   black                           None reported         Yes       3
## 3   black At least 1 accident or damage reported         Yes       7
## 4   black                           None reported         Yes       2
## 5   other                           None reported         Yes       5
## 6   other                           None reported         Yes       6
## 7   black                           None reported         Yes       1
## 8   black                           None reported         Yes       6
```

```
if (!requireNamespace("Rtsne", quietly = TRUE)) {
  stop("Package 'Rtsne' is required. Please install it before knitting.")
}
library(Rtsne)

# Create a mixed dataset with dummy variables for categorical data
data_mixed <- model.matrix(~ . - 1, data = data)  # Convert categorical variables into dummy variables

# Run t-SNE on the mixed dataset
set.seed(42)
tsne_result <- Rtsne(data_mixed, dims = 2, perplexity = 30, verbose = TRUE)
```

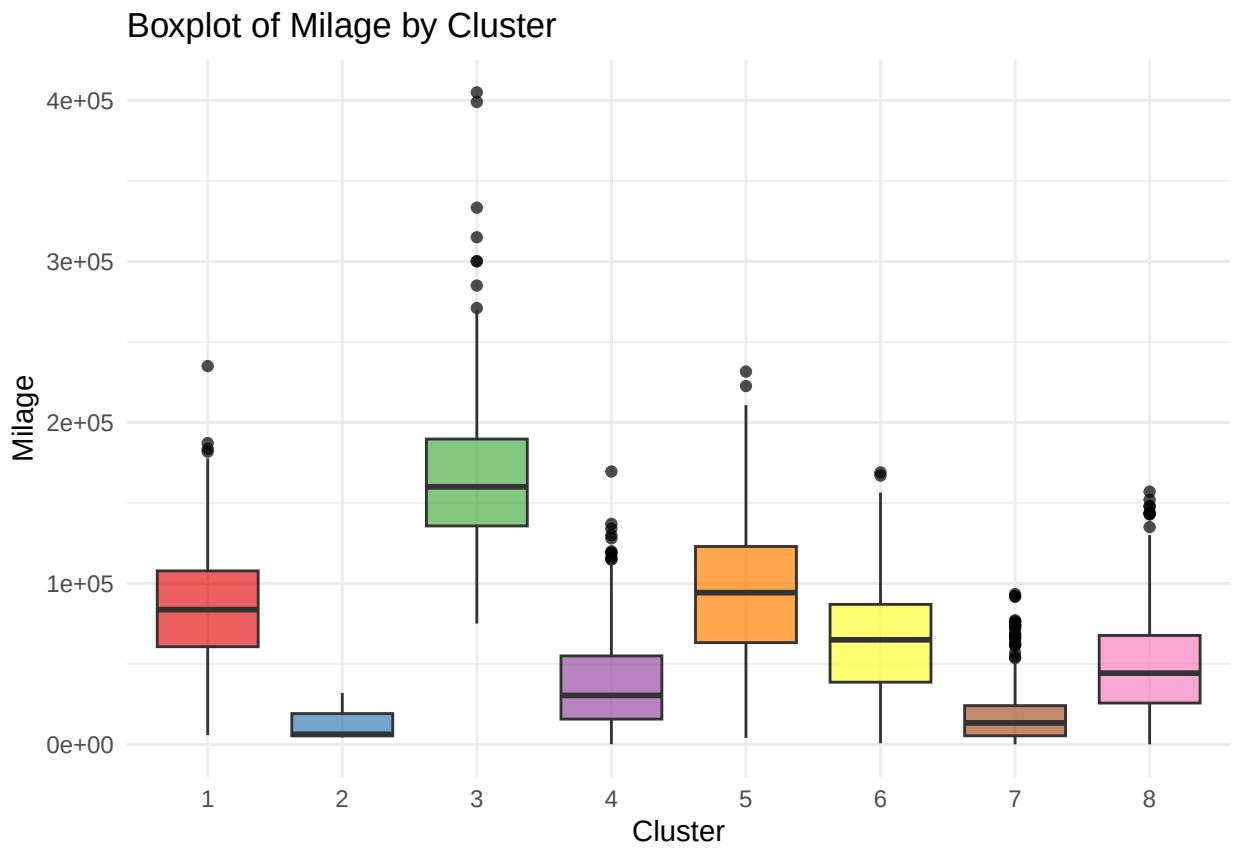
```
## Performing PCA
## Consider setting partial_pca=TRUE for large matrices
## Read the 3943 x 50 data matrix successfully!
## Using no_dims = 2, perplexity = 30.000000, and theta = 0.500000
## Computing input similarities...
## Building tree...
## Done in 0.21 seconds (sparsity = 0.026726)!
```

```
## Learning embedding...
## Iteration 50: error is 82.674273 (50 iterations in 0.35 seconds)
## Iteration 100: error is 67.088356 (50 iterations in 0.31 seconds)
## Iteration 150: error is 64.103425 (50 iterations in 0.32 seconds)
## Iteration 200: error is 62.419329 (50 iterations in 0.33 seconds)
## Iteration 250: error is 61.728791 (50 iterations in 0.34 seconds)
## Iteration 300: error is 1.572250 (50 iterations in 0.32 seconds)
## Iteration 350: error is 1.152453 (50 iterations in 0.30 seconds)
## Iteration 400: error is 0.954027 (50 iterations in 0.30 seconds)
## Iteration 450: error is 0.843008 (50 iterations in 0.31 seconds)
## Iteration 500: error is 0.783158 (50 iterations in 0.31 seconds)
## Iteration 550: error is 0.753416 (50 iterations in 0.32 seconds)
## Iteration 600: error is 0.736638 (50 iterations in 0.32 seconds)
## Iteration 650: error is 0.726362 (50 iterations in 0.33 seconds)
## Iteration 700: error is 0.717879 (50 iterations in 0.33 seconds)
## Iteration 750: error is 0.709007 (50 iterations in 0.34 seconds)
## Iteration 800: error is 0.700722 (50 iterations in 0.34 seconds)
## Iteration 850: error is 0.692262 (50 iterations in 0.34 seconds)
## Iteration 900: error is 0.683776 (50 iterations in 0.34 seconds)
## Iteration 950: error is 0.675546 (50 iterations in 0.34 seconds)
## Iteration 1000: error is 0.668145 (50 iterations in 0.34 seconds)
## Fitting performed in 6.50 seconds.
```

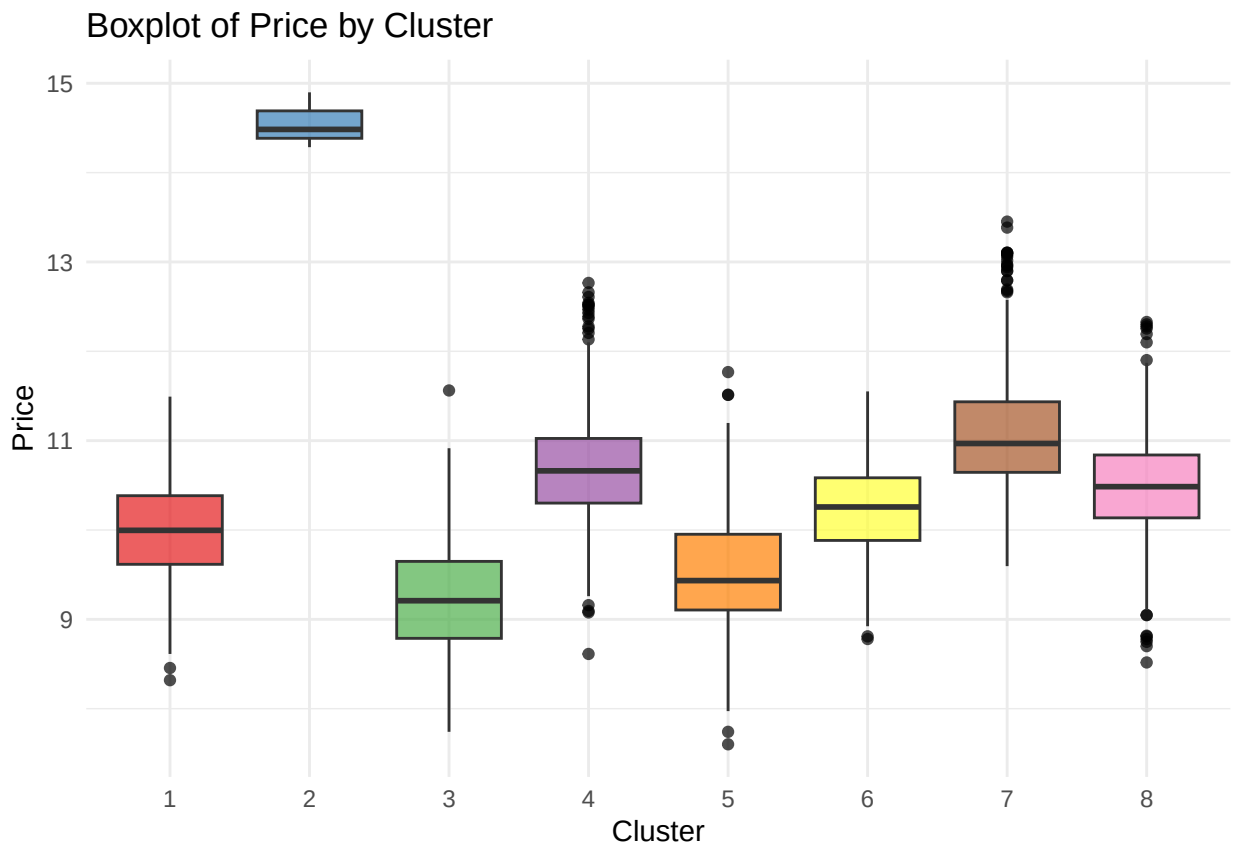
```r
# Create a data frame with t-SNE results and cluster assignments
tsne_df <- as.data.frame(tsne_result$Y)
colnames(tsne_df) <- c("Dim1", "Dim2")
tsne_df$Cluster <- as.factor(data$cluster)

# Plot t-SNE results
ggplot(tsne_df, aes(x = Dim1, y = Dim2, color = Cluster)) +
  geom_point(size = 2, alpha = 0.8) +
  labs(
    title = "t-SNE Visualization of K-Prototypes Clusters",
    x = "t-SNE Dimension 1",
    y = "t-SNE Dimension 2"
  ) +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```



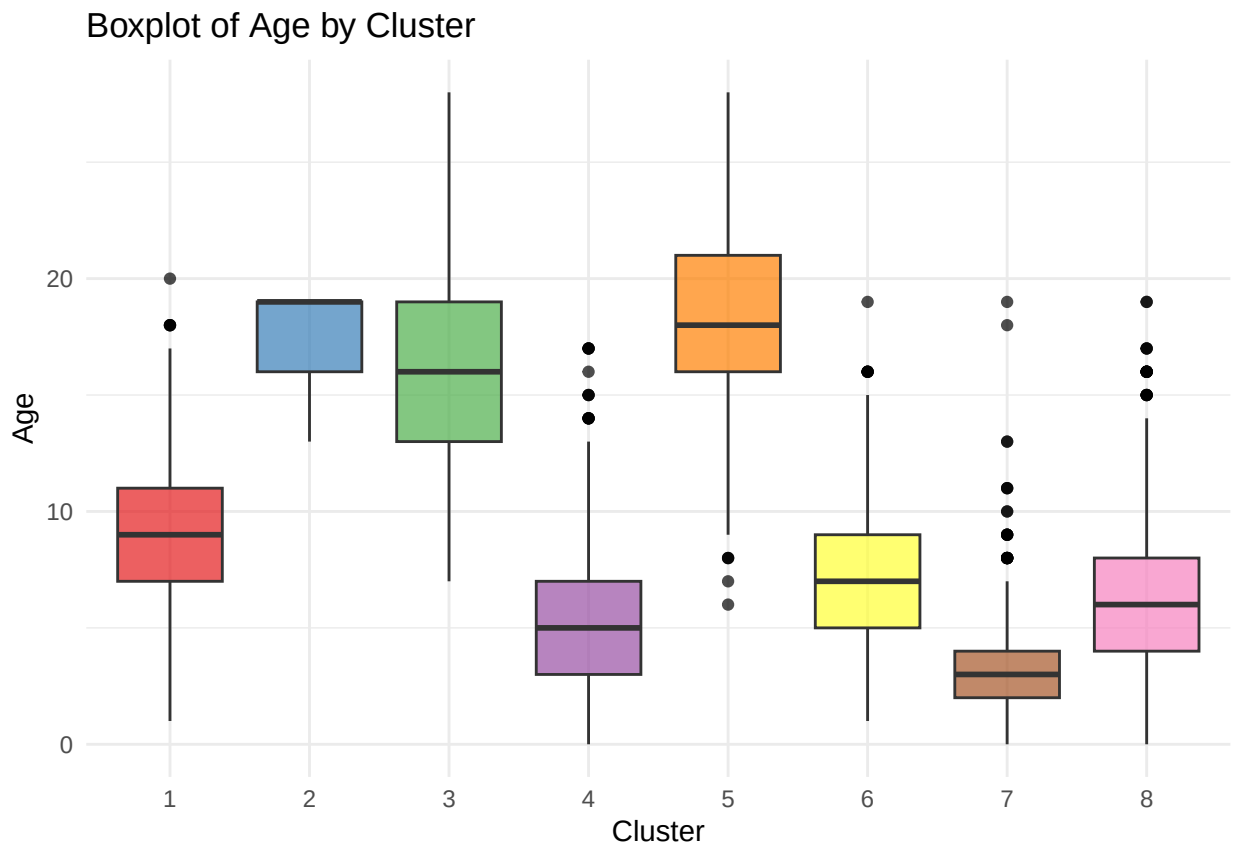

```r
# Ensure that cluster labels are properly factored and numeric data is grouped correctly
data$cluster <- factor(data$cluster)  # Convert cluster to factor for proper labeling
# Boxplot of Milage by Cluster
ggplot(data, aes(x = cluster, y = milage, fill = cluster)) +
  geom_boxplot(alpha = 0.7, outlier.color = "black", outlier.size = 1.5) +
  labs(
    title = "Boxplot of Milage by Cluster",
    x = "Cluster",
    y = "Milage"
  ) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position = "none")  # Remove redundant legend if clusters are on the x-axis
```



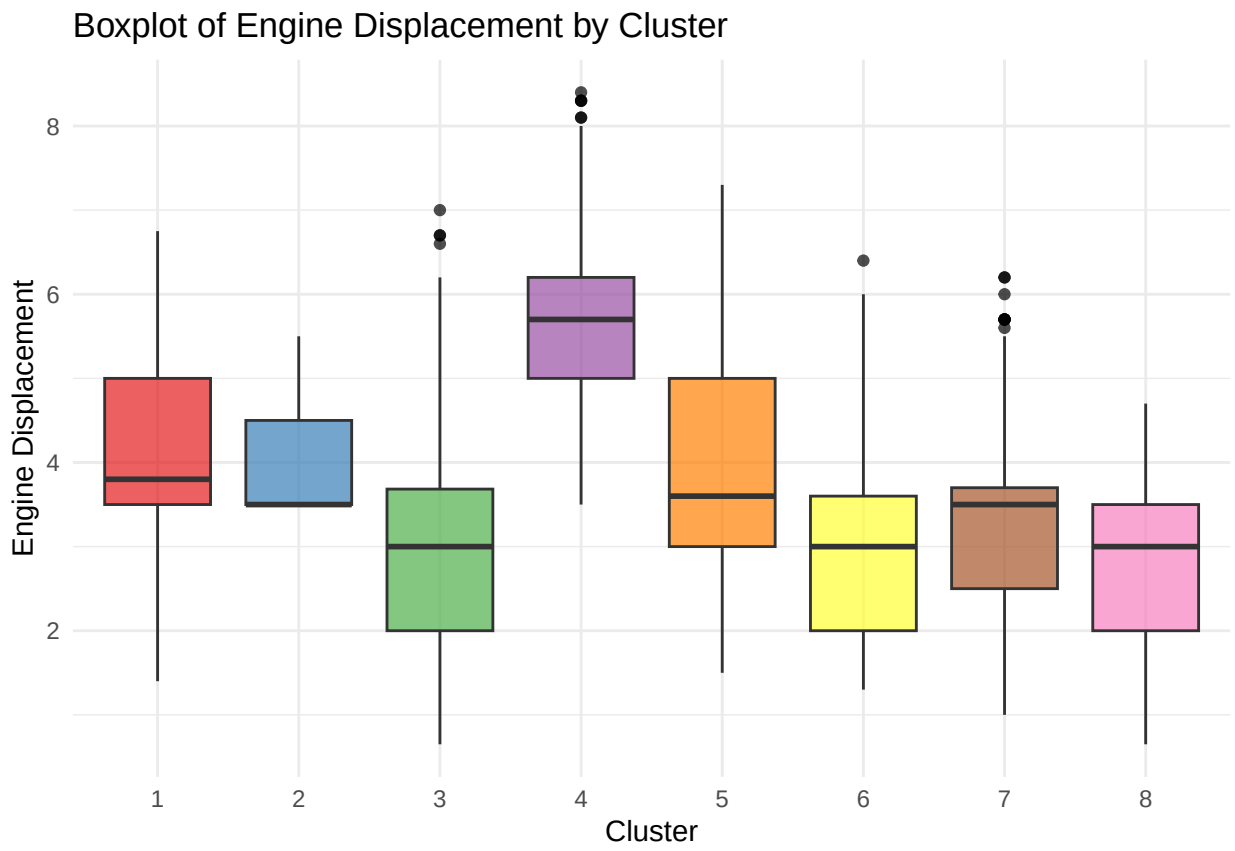

```r
data$log_price <- log(data$price + 1)

# Boxplot of Price by Cluster
ggplot(data, aes(x = cluster, y = log_price, fill = cluster)) +
  geom_boxplot(alpha = 0.7, outlier.color = "black", outlier.size = 1.5) +
  labs(
    title = "Boxplot of Price by Cluster",
    x = "Cluster",
    y = "Price"
  ) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position = "none")
```




```r
# Boxplot of Age by Cluster
ggplot(data, aes(x = cluster, y = age, fill = cluster)) +
  geom_boxplot(alpha = 0.7, outlier.color = "black", outlier.size = 1.5) +
  labs(
    title = "Boxplot of Age by Cluster",
    x = "Cluster",
    y = "Age"
  ) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position = "none")
```




```r
# Boxplot of Engine Displacement by Cluster
ggplot(data, aes(x = cluster, y = Engine_Displacement, fill = cluster)) +
  geom_boxplot(alpha = 0.7, outlier.color = "black", outlier.size = 1.5) +
  labs(
    title = "Boxplot of Engine Displacement by Cluster",
    x = "Cluster",
    y = "Engine Displacement"
  ) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position = "none")
```




```
library(broom)

# Fit separate linear regression models for each cluster
cluster_models <- data %>%
  group_by(cluster) %>%
  group_map(~ {
    model <- lm(price ~ milage + age + Engine_Displacement, data = .x)
    tidy(model) %>% mutate(cluster = unique(.x$cluster))
  }) %>%
  bind_rows()

# Print regression summaries
print("Cluster-Specific Regression Results:")
```

```
## [1] "Cluster-Specific Regression Results:"
```

```
print(cluster_models)
```

```
## # A tibble: 32 x 5
##    term                    estimate std.error statistic    p.value
##    <chr>                      <dbl>     <dbl>     <dbl>      <dbl>
##  1 (Intercept)           51364.     2467.        20.8     1.09e-71
##  2 milage                   -0.133     0.0160    -8.29    8.28e-16
##  3 age                   -1634.      167.        -9.78    6.12e-21
```

```
##  4 Engine_Displacement      86.7     483.         0.179   8.58e- 1
##  5 (Intercept)         2197557.      NaN        NaN     NaN
##  6 milage                   49.1     NaN        NaN     NaN
##  7 age                 -42873.      NaN        NaN     NaN
##  8 Engine_Displacement      NA        NA         NA      NA
##  9 (Intercept)           23760.     2494.         9.53    2.12e-19
## 10 milage                   -0.0331    0.0106    -3.11    1.98e- 3
## # i 22 more rows
```

```
descriptive_stats <- data %>%
  group_by(cluster) %>%
  summarize(
    count = n(),
    mean_price = mean(price, na.rm = TRUE),
    median_price = median(price, na.rm = TRUE),
    sd_price = sd(price, na.rm = TRUE),
    mean_milage = mean(milage, na.rm = TRUE),
    median_milage = median(milage, na.rm = TRUE),
    sd_milage = sd(milage, na.rm = TRUE),
    mean_age = mean(age, na.rm = TRUE),
    median_age = median(age, na.rm = TRUE),
    sd_age = sd(age, na.rm = TRUE),
    mean_engine_displacement = mean(Engine_Displacement, na.rm = TRUE),
    median_engine_displacement = median(Engine_Displacement, na.rm = TRUE),
    sd_engine_displacement = sd(Engine_Displacement, na.rm = TRUE)
  )

print("Descriptive Statistics by Cluster:")
```

```
## [1] "Descriptive Statistics by Cluster:"
```

```
print(descriptive_stats)
```

```
## # A tibble: 8 x 14
##   cluster count mean_price median_price sd_price mean_milage median_milage
##   <fct>   <int>      <dbl>        <dbl>    <dbl>       <dbl>         <dbl>
## 1 1         562     25210.       21945    14607.      85578.        83732.
## 2 2           3   2168026      1950995   703128.      14243.         6330
## 3 3         378     12466.        9980     9386.     164890.       160000
## 4 4         646     53646.       42708.   42926.      37476.        30450
## 5 5         429     17154.       12500    14238.      93512.        94291
## 6 6         429     31909.       28499    16493.      65487.        65000
## 7 7         684     86198.       57998    81705.      17326.        13391
## 8 8         812     41639.       35775    26465.      49180.        44300
## # i 7 more variables: sd_milage <dbl>, mean_age <dbl>, median_age <dbl>,
## #   sd_age <dbl>, mean_engine_displacement <dbl>,
## #   median_engine_displacement <dbl>, sd_engine_displacement <dbl>
```