Lab07Group08TummuruShah

R Code:

```
# Vihari Reddy Tummuru & Maurvin Gaurav Shah
# MIS 545-001
# Lab07Group08TummuruShah.R
# Import csv files, assign data types, subset datasets, calculate summary
# Manipulate tibbles and prepare histogram and analyze the data using KNN Model

# Install tidyverse package
install.packages("tidyverse")

# Load tidyverse package
library(tidyverse)

# Install corrplot package
install.packages("corrplot")

# Load corrplot package
library(corrplot)

# Install class package
install.packages("class")

# Load class package
library(class)

# Set working directory to Lab06 folder
setwd("C:/Users/ual-laptop/Documents/MIS545/Lab07")

# Read SedanSize.csv file into a tibble and defining column types
sedanSize <- read_csv(file = "SedanSize.csv",
                      col_types = "cfnii",
                      col_names = TRUE
```

```
                    )
# Display the sedanSize tibble in the console
print(sedanSize)

# Display the structure of the tibble
print(str(sedanSize))

# Display the summary of the mobilephone tibble in the console
print(summary(sedanSize))

# Removing MakeModel attribute from the tibble dataset
sedanSize <- sedanSize %>% select(-MakeModel)

# Separating the tibble in Lables tibble and values tibble
sedanSizeLabel <- sedanSize %>% select(SedanSize)
sedanSize <- sedanSize %>% select(-SedanSize)

# Creating a function to display histogram bu converting all columns to numeric
displayAllHistograms <- function(tibbleDataSet) {
  tibbleDataSet %>%
    keep(is.numeric) %>%
    gather() %>%
    ggplot() + geom_histogram(mapping = aes(x=value,fill=key),
                              color = "black") +
    facet_wrap(~ key, scales="free") +
    theme_minimal()
}

# Display the histogram of the sedanSize tibble
displayAllHistograms(sedanSize)

# Setting the seed as 517
set.seed(517)
```

```
# Splitting mobilephone data into training dataset and test data set based on IQR
sampleSet <- sample(nrow(sedanSize),
                    round(nrow(sedanSize)*.75),
                    replace = FALSE)

# Initilizing the training datasets
sedanSizeTraining <- sedanSize[sampleSet,]
sedanSizeLabelTraining <- sedanSizeLabel[sampleSet,]

# Initilizing the testing datasets
sedanSizeTesting <- sedanSize[-sampleSet,]
sedanSizeLabelTesting <- sedanSizeLabel[-sampleSet,]

# Using the KNN model for prediction of sedan size
sedanSizePrediction <- knn(train = sedanSizeTraining,
                           test = sedanSizeTesting,
                           cl = sedanSizeLabelTraining$SedanSize,
                           k = 7)
# Display the prediction of knn model on console
print(sedanSizePrediction)

# Display the summary of prediction of knn model on console
print(summary(sedanSizePrediction))

# Generating a sedan size confusion matrix
sedanSizeConfusionMatrix <- table(sedanSizeLabelTesting$SedanSize,
                                  sedanSizePrediction)

# Display confusion matrix (sedanSizeConfusionMatrix)
print(sedanSizeConfusionMatrix)

# Use the model to predict outcomes in the testing dataset
sedanSizeAccuracy <- sum(diag(sedanSizeConfusionMatrix)) /
  nrow(sedanSizeTesting)
```

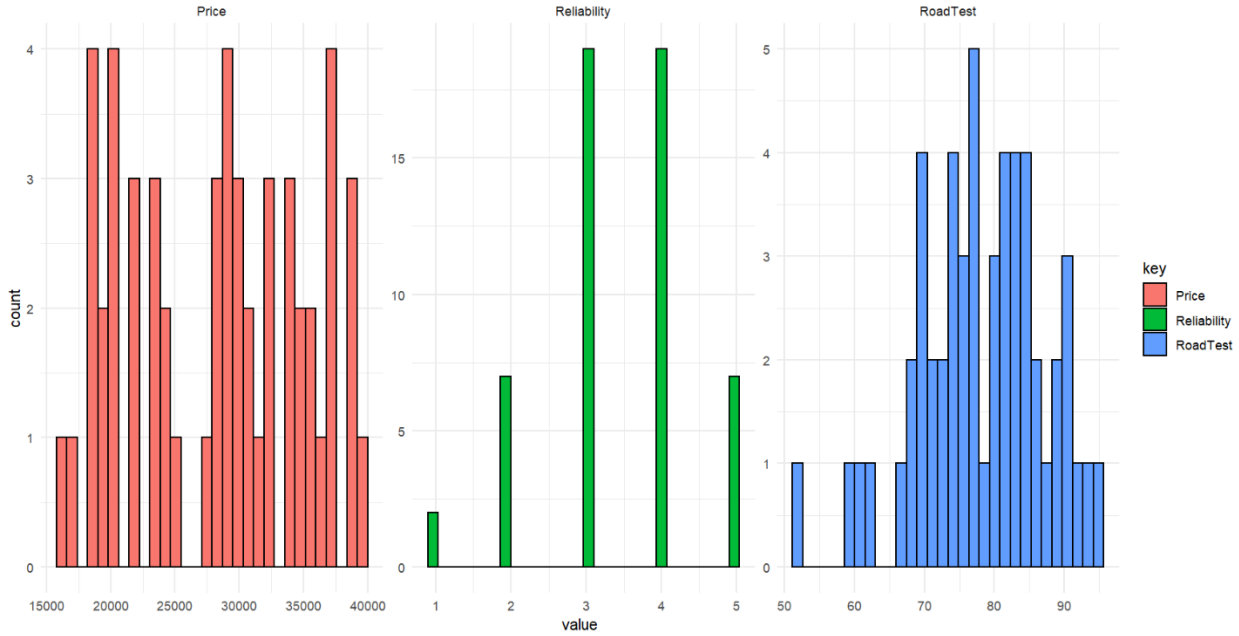
```
# Display the test model
print(sedanSizeAccuracy)

# New matrix of k values with the predictive accuracy
kValueMatrix <- matrix(data = NA,
                       nrow = 0,
                       ncol = 2)

# Setting the column names as "k value" and "Predictive Accuracy"
colnames(kValueMatrix) <- c("k value","Predictive Accuracy")

# Storing the k value and the predictive accuracy in the training dataset
for (kValue in 1:nrow(sedanSizeTraining)) {
  if (kValue %% 2 != 0) {
    # Using the KNN model to predict sedan size
    sedanSizePrediction <- knn(train = sedanSizeTraining,
                               test = sedanSizeTesting,
                               cl = sedanSizeLabelTraining$SedanSize,
                               k = kValue)
    # Creating a confusion matrix
    sedanSizeConfusionMatrix <- table(sedanSizeLabelTesting$SedanSize,
                                      sedanSizePrediction)
    # The predictive accuracy of the model
    sedanSizeAccuracy <- sum(diag(sedanSizeConfusionMatrix)) /
      nrow(sedanSizeTesting)
    # Adding a new row to the K value matrix
    kValueMatrix = rbind(kValueMatrix, c(kValue,sedanSizeAccuracy))
  }
}
# Display the k value matrix on the console
print(kValueMatrix)
```

Histograms:



K Value Matrix:

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Source

Console | Terminal | Background Jobs

R 4.2.1 · ~/MIS545/Lab07/

```
+     # The predictive accuracy of our model
+     sedanSizeAccuracy <- sum(diag(sedanSizeConfusionMatrix)) /
+       nrow(sedanSizeTesting)
+     # Adding a new row to the matrix
+     kValueMatrix = rbind(kValueMatrix, c(kValue,sedanSizeAccuracy))
+   }
+ }
> # Display the k value matrix on the console
> print(kValueMatrix)
      k value Predictive Accuracy
 [1,]       1           0.8571429
 [2,]       3           0.8571429
 [3,]       5           0.8571429
 [4,]       7           0.8571429
 [5,]       9           0.8571429
 [6,]      11           0.8571429
 [7,]      13           0.7857143
 [8,]      15           0.8571429
 [9,]      17           0.5714286
[10,]      19           0.5714286
[11,]      21           0.5714286
[12,]      23           0.5714286
[13,]      25           0.5714286
[14,]      27           0.5714286
[15,]      29           0.5714286
[16,]      31           0.5714286
[17,]      33           0.2857143
[18,]      35           0.5000000
[19,]      37           0.4285714
[20,]      39           0.2857143
>
```

Question 2: What would be the best value for k given this data? Why?

Answer: Best value of K is nearest integer less than the square root of the sample size. Therefore 7 is the best value of K and it has highest accuracy.

Question 3: How could an automobile manufacturer take advantage of this model?

Answer: Automobile manufacturers can assess their model offerings based on the model by determining the size of sedan based on price, road test and reliability