R script code:

# Aadhithya Dinesh

# MIS 545 Section 02

# Lab02DineshA.R

# Import and prepare a dataset of automobile tire usage and perform data

# preprocessing tasks like imputing missing data, identifying outliers,

# normalizing features, discretizing features and dummy coding.

# intsall.packages("tidyverse")

# install.packages("dummies")

library(tidyverse)

library(dummies)

library(scales)

# set the working directory

setwd("~/MIS/Classes/MIS545/Assignments/Lab04")

# read the csv file with column types specified

tireTread1 <- read\_csv(file = "TireTread.csv",

col\_types = "cfnni",

col\_names = TRUE)

# print the tire Tread data along with summary

print(tireTread1)

str(tireTread1)

print(summary(tireTread1))

# Impute missing values with the mean value

tireTread2 <- tireTread1 %>%

mutate(UsageMonths = ifelse(is.na(UsageMonths), mean(UsageMonths, na.rm = TRUE), UsageMonths))

# printe the summary

print(summary(tireTread2))

# outliers are separately stored in treadDepthOutliers

outlierMin <- quantile(tireTread2$TreadDepth, .25) -

(IQR(tireTread2$TreadDepth) \* 1.5)

outlierMax <- quantile(tireTread2$TreadDepth, .75) +

(IQR(tireTread2$TreadDepth) \* 1.5)

treadDepthOutliers <- tireTread2 %>%

filter(TreadDepth < outlierMin | TreadDepth > outlierMax)

# normalize the data by taking the log

tireTread3 <- tireTread2 %>%

mutate(LogUsageMonths = log(UsageMonths))

# discretization by setting values to true and false based on TreadDepth

tireTread4 <- tireTread3 %>%

mutate(NeedsReplacing = TreadDepth <=1.6)

tireTread4DataFrame <- data.frame(tireTread4)

# storing the values in a tibble

tireTread5 <- as\_tibble(dummy.data.frame(data = tireTread4DataFrame,

names = "Position"))

# creating a scatter plot visualization

scatterPlotMilesTreadDepth <- ggplot(data = tireTread5,

aes(x = Miles,

y = TreadDepth

))

scatterPlotMilesTreadDepth + geom\_point(color = "dark gray") +

scale\_y\_continuous() +

geom\_smooth(method = lm,

level = 0,

color = "red") +

ggtitle("Tire Miles and Tread Depth Scatter Plot.")

Chart, scatter chart

Description automatically generated

Rapid Miner Process and Result screenshots:

Graphical user interface, diagram, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedA picture containing text, screenshot, indoor

Description automatically generated

Yes, a correlation exists as the data points follow a linear regression model except for the outliers.