

**MIS 545 LAB 04**

Group 11

Anjana Pillai, Vihari Reddy Tummuru

09/19/2022

```

# Vihari Reddy Tummuru, Anjana Pillai
# MIS 545 -001
# Lab04Group11PillaiTummuru.R
# The following code processes data from TireTread.csv
# Calculates outliers min,max, normalizes certain columns,
# Discretizes and plots a scatterplot to visualize data

# Set the current working directory
setwd("C:/Users/anjan/Downloads/Lab04")

# Install dummies package
install.packages("dummies", repos = NULL, type="source")

# Load dummies
library(dummies)

#Load tidyverse
library(tidyverse)

# Read tireTread1 as a tibble from the csv
tireTread1 <- read_csv(file="TireTread.csv",
                        col_types="cfnni",
                        col_names=TRUE)

# Display the first 20 rows
print(tireTread1)

# Display the structure
str(tireTread1)

# Display the Summary
summary(tireTread1)

# Impute missing data for UsageMonths with the mean
# Read into tireThread2 tibble
tireTread2 <- tireTread1 %>%
  mutate(UsageMonths = ifelse(is.na(UsageMonths),
                              mean(UsageMonths, na.rm = TRUE), UsageMonths))

# Display summary Of Tiretread2
summary(tireTread2)

# Determine outliers in the TreadDepth feature
# Calculate outlier min and max and store into variables called
# outlierMin and outlierMax.
outlierMin <- quantile(tireTread2$TreadDepth, .25) -
  (IQR(tireTread2$TreadDepth) * 1.5)

outlierMax <- quantile(tireTread2$TreadDepth, .75) +
  (IQR(tireTread2$TreadDepth) * 1.5)

# Add the outliers to their own tibble called treadDepthOutliers

```

```

treadDepthOutliers <- tireTread2 %>%
  filter(TreadDepth < outlierMin | TreadDepth > outlierMax)

# Normalize the UsageMonths feature by taking the log of UsageMonths into a
# New feature called LogUsageMonths and store the additional column in a tibble
# Called tireTread3
tireTread3<- tireTread2%>%
  mutate(LogUsageMonths = log(UsageMonths))

# Discretize TreadDepth into NeedsReplacing and store into tireTread4 tibble
tireTread4 <- tireTread3%>%
  mutate(NeedsReplacing = TreadDepth <= 1.6)

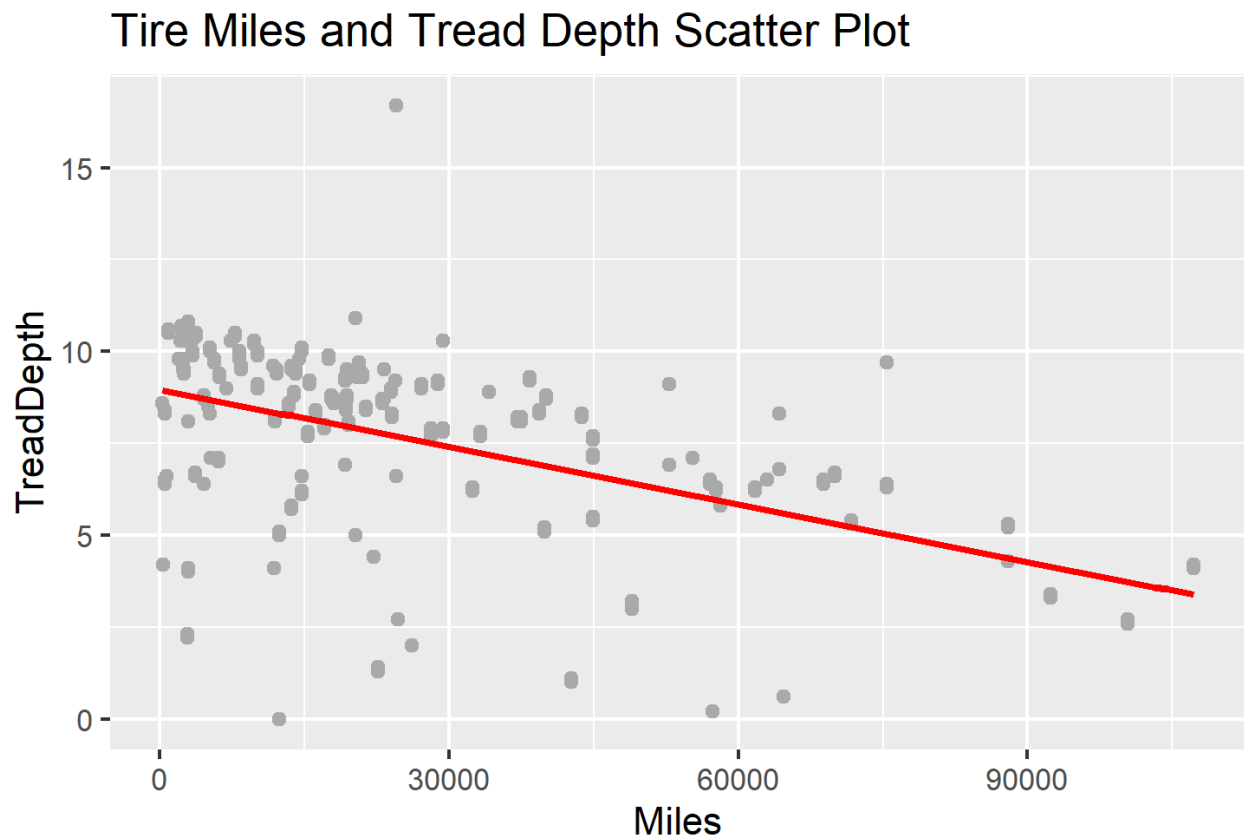
# Convert tireTread4 into data frame
tireTread4DataFrame <- data.frame(tireTread4)

# Dummy code the Position using dummy.data.frame(), convert it back into
# a tibble, and store the result into a new tireTread5 tibble
tireTread5 <- as_tibble(dummy.data.frame(data = tireTread4DataFrame,
                                         names = "Position"))

# Build a scatter plot of Miles (x) with TreadDepth (y)
# Change point colour to dark grey,
# Title is added using ggtitle()
# Linear best fit using geom_smooth() and changing the color to red
scatterplotMilesTreadDepth <- ggplot(data = tireTread5,
                                     aes(x = Miles,
                                         y = TreadDepth))
scatterplotMilesTreadDepth + geom_point(color = "dark grey") +
  ggtitle("Tire Miles and Tread Depth Scatter Plot")+
  geom_smooth(method = lm,
             level = 0,
             color = "red")

```

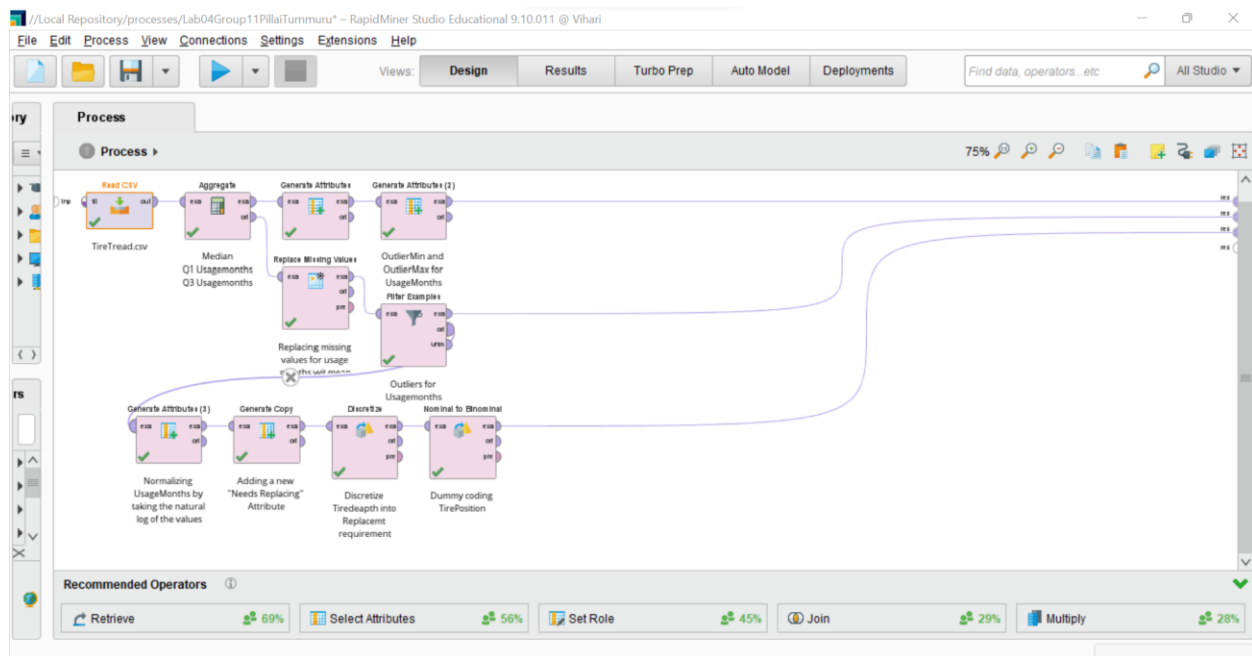
Scatterplot:



Analysis:

Visually we can say that there is a negative correlation between two features as the line of regression is with a negative slope from the scatter plot above.

## Rapid Miner Process:



## Rapid Miner Identified Outliers:

The screenshot shows the Results view of Rapid Miner Studio. The 'ExampleSet (Filter Examples)' is selected, displaying a table of 17 examples. The table has columns: Row No., UsageMonths, TireID, Position, TreadDepth, and Miles. The data is filtered to show 17 examples. The 'Repository' panel on the right shows the data source as 'Local Repository (Local)'.

Row No.	UsageMonths	TireID	Position	TreadDepth	Miles
1	69	10395904	RF	0.600	64742
2	69	10395904	LR	0.600	64742
3	73.300	3574739	RR	0.200	57313
4	64.100	21817569	RR	1.300	22618
5	66.500	16691383	LF	1.100	42632
6	69	10395904	RR	0.600	64742
7	66.500	16691383	LR	1	42632
8	66.500	16691383	RR	1	42632
9	69	10395904	LF	0.600	64742
10	64.100	21817569	RF	1.400	22618
11	111	2122934	LR	9.300	21000
12	64.100	21817569	LR	1.300	22618

ExampleSet (17 examples, 0 special attributes, 5 regular attributes)

## Rapid Miner Results:

Local Repository/processes/Lab04Group11PillaiTummuru - RapidMiner Studio Educational 9.10.011 @ Vihari

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments Find data, operators... etc All Studio

Result History ExampleSet (Nominal to Binominal) ExampleSet (Filter Examples) ExampleSet (Generate Attributes (2))

Open in Turbo Prep Auto Model Filter (456 / 456 examples): all

Row No.	Position = LR	Position = RR	Position = LF	Position = RF	NeedsRepla...	UsageMonths	TireID	TreadDepth	Miles	LogUsageM...
1	true	false	false	false	false	58.400	13391487	2.200	2805	4.067
2	true	false	false	false	false	17.300	21678308	8.300	39371	2.851
3	false	true	false	false	false	16.500	18414311	8.600	13367	2.803
4	false	true	false	false	false	8.200	19778103	9.800	1931	2.104
5	false	true	false	false	false	13.700	16355454	8.900	23992	2.617
6	true	false	false	false	false	52.800	8952817	3	48961	3.967
7	false	true	false	false	false	14.700	6559652	8.800	4585	2.688
8	true	false	false	false	false	6.200	16289814	10.100	5221	1.825
9	false	false	true	false	false	60.100	8696859	2	26129	4.096
10	false	false	true	false	false	11.700	8921457	9.200	19237	2.460
11	false	false	true	false	false	10.900	7505467	9.300	6211	2.389
12	false	true	false	false	false	7.700	18307885	9.600	11708	2.041

ExampleSet (456 examples, 0 special attributes, 10 regular attributes)

## Rapid Miner Scatterplot:

