UNIVERSITI TEKNOLOGI PETRONAS

LAB WEEK 11

SEP 2025 SEMESTER

‘AFFAN NAJIY BIN RUSDI

22010453

BACHELOR OF COMPUTER SCIENCE

DATA SCIENCE

TEB2164

Contents

Code

Activity 1

```
#Load Dataset
data("ToothGrowth")

#Convert Categorical to Numerical
df <- ToothGrowth
#Assign Auto
df$supp <- as.numeric(df$supp) #VC vitamin C, OJ orange juice

#Assign Manually
#df$supp <- ifelse(ToothGrowth == "OJ", 1, 0)

#Calc Corr Analysis
corr_mat <- round(cor(df), 2)

#Reshape
library(reshape2)

melted_corr <- melt(corr_mat)

#Plot
library(ggplot2)
ggplot(melted_corr, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  geom_text(aes(label = value), color = "white", size = 5) +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0) +
  theme_minimal()
```

Activity 2

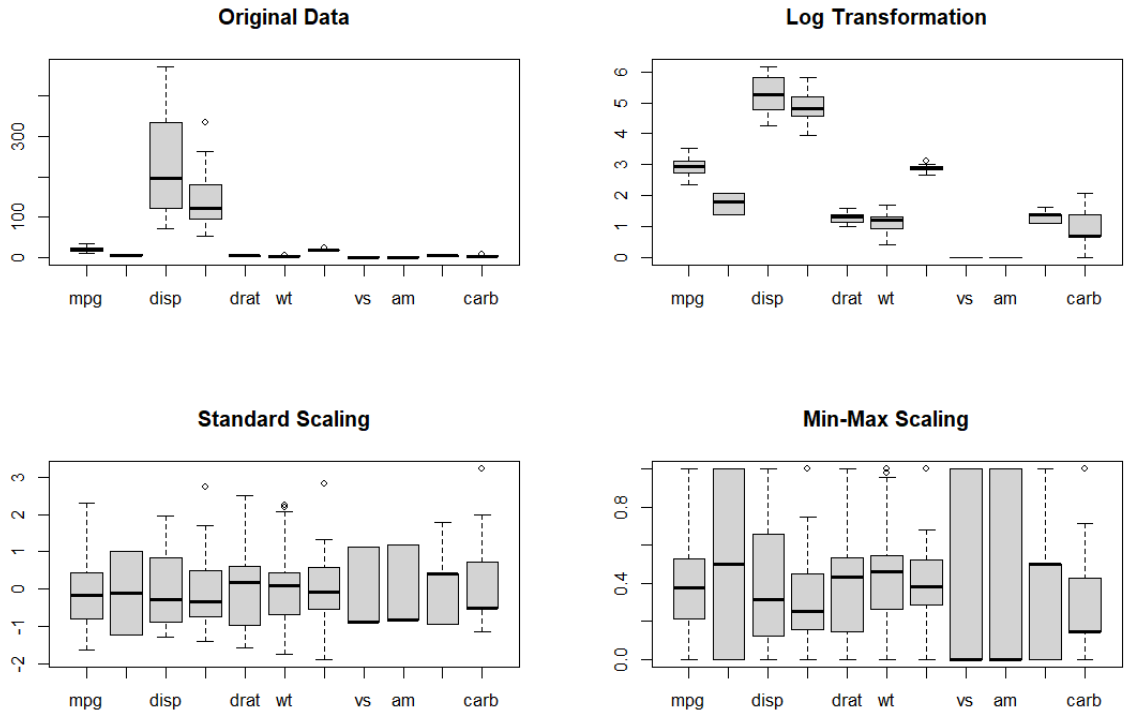
```
#Load Dataset
data(mtcars)
df <- mtcars

#Lib
library(caret)

#Log Transformation
df_log <- log(df)

#Standard Scaling
df_standard <- as.data.frame(scale(df))

#Min-Max Scaling
minmax <- preProcess(as.data.frame(df), method = c("range"))
df_minmax <- predict(minmax, as.data.frame(df))

#Summary
summary(df)
summary(df_log)
summary(df_standard)
summary(df_minmax)

#Plot
par(mfrow = c(2, 2)) #grid 2 by 2 for side by side
boxplot(df, main = "Original Data")
boxplot(df_log, main = "Log Transformation")
boxplot(df_standard, main = "Standard Scaling")
boxplot(df_minmax, main = "Min-Max Scaling")
```

Visualisation

Activity 1

Theres a very strong correlation between dose and len which account to 0.8. While dose and supp have a 0 and len with supp have a negative correlatio of -0.24.

Activity 2



Log transformation compresses the high values, standard scaling centres data around a mean of 0 and standard deviation of 1, and min-max scaling bounds values in a range of [0, 1]