Code\_Description

# Data Preprocessing

Load data and libraries

library(dplyr)

Warning: package 'dplyr' was built under R version 4.4.1

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

library(readxl)

Warning: package 'readxl' was built under R version 4.4.1

library(writexl)

Warning: package 'writexl' was built under R version 4.4.1

library(ggplot2)

Warning: package 'ggplot2' was built under R version 4.4.1

data <- read\_xlsx("response\_data.xlsx")  
summary(data)

timestamp Email Address Score   
 Min. :2024-09-23 17:27:53.17 Length:261 Min. :0.000   
 1st Qu.:2024-09-28 15:11:56.16 Class :character 1st Qu.:4.000   
 Median :2024-09-28 20:00:16.75 Mode :character Median :5.000   
 Mean :2024-09-28 20:27:04.65 Mean :4.866   
 3rd Qu.:2024-09-29 19:44:38.65 3rd Qu.:6.000   
 Max. :2024-09-30 23:05:46.82 Max. :7.000   
 BITS ID Mobile Number Gender   
 Length:261 Length:261 Length:261   
 Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character   
   
   
   
 Which degree are you pursuing? What is your specialization?  
 Length:261 Length:261   
 Class :character Class :character   
 Mode :character Mode :character   
   
   
   
 Which year of study are you currently in?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 Do you have any prior knowledge about Machine Learning?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 Which of the following best describes the function and structure of a neural network in the context of artificial intelligence?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 How does a neural network typically process input data?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 What is the primary purpose of training a neural network?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 What role does backpropagation play in the training of neural networks?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 What is one advantage of using neural networks over traditional computational methods?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 What is an emerging trend in the field of neural networks?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 Which of the following statements does not hold true?  
 Length:261   
 Class :character   
 Mode :character   
   
   
   
 Did the teaching method used in the video capture your interest?  
 Min. :1.000   
 1st Qu.:5.000   
 Median :7.000   
 Mean :6.295   
 3rd Qu.:8.000   
 Max. :9.000   
 Were your thoughts/focus fully on this video while watching?  
 Min. :1.000   
 1st Qu.:5.000   
 Median :7.000   
 Mean :6.138   
 3rd Qu.:8.000   
 Max. :9.000   
 Did this video made you feel excited about the subject?  
 Min. :1.000   
 1st Qu.:5.000   
 Median :7.000   
 Mean :6.326   
 3rd Qu.:8.000   
 Max. :9.000

Selecting columns required for ANOVA testing.

selected\_data <- data |> select(4,9,10,3,18,19,20)  
selected\_data

# A tibble: 261 × 7  
 `BITS ID` Which year of study are you curr…¹ Do you have any prio…² Score  
 <chr> <chr> <chr> <dbl>  
 1 2024H1120194P 1st year Yes 6  
 2 2024H1030080P 1st year Yes 7  
 3 2024H1540858P 1st year No 6  
 4 2021B4A72749 4th year Yes 7  
 5 2024ABPS0687P 1st year No 7  
 6 2024A1PS1207P 1st year No 7  
 7 2022B2A21350P 3rd year No 1  
 8 2024H1540806P 1st year No 5  
 9 2024H1540862P 1st year Yes 6  
10 2024H1540819P 1st year No 6  
# ℹ 251 more rows  
# ℹ abbreviated names: ¹​`Which year of study are you currently in?`,  
# ²​`Do you have any prior knowledge about Machine Learning?`  
# ℹ 3 more variables:  
# `Did the teaching method used in the video capture your interest?` <dbl>,  
# `Were your thoughts/focus fully on this video while watching?` <dbl>,  
# `Did this video made you feel excited about the subject?` <dbl>

Checking for null values.

sum(is.na(data))

[1] 0

Renaming columns.

selected\_data <- selected\_data |> rename("First Year or Not?"="Which year of study are you currently in?","Prior\_knowledge"="Do you have any prior knowledge about Machine Learning?","Retention\_Score"="Score")  
selected\_data

# A tibble: 261 × 7  
 `BITS ID` `First Year or Not?` Prior\_knowledge Retention\_Score  
 <chr> <chr> <chr> <dbl>  
 1 2024H1120194P 1st year Yes 6  
 2 2024H1030080P 1st year Yes 7  
 3 2024H1540858P 1st year No 6  
 4 2021B4A72749 4th year Yes 7  
 5 2024ABPS0687P 1st year No 7  
 6 2024A1PS1207P 1st year No 7  
 7 2022B2A21350P 3rd year No 1  
 8 2024H1540806P 1st year No 5  
 9 2024H1540862P 1st year Yes 6  
10 2024H1540819P 1st year No 6  
# ℹ 251 more rows  
# ℹ 3 more variables:  
# `Did the teaching method used in the video capture your interest?` <dbl>,  
# `Were your thoughts/focus fully on this video while watching?` <dbl>,  
# `Did this video made you feel excited about the subject?` <dbl>

selected\_data$Engaging\_Score <- rowMeans(selected\_data[, c(5,6,7)], na.rm = TRUE)  
View(selected\_data)

Converting Categorical values to Numerical Values.

selected\_data <- selected\_data %>%  
 mutate(`First Year or Not?` = if\_else(`First Year or Not?` == "1st year", 1, 0))  
selected\_data <- selected\_data %>%  
 mutate(`Prior\_knowledge` = if\_else(`Prior\_knowledge` == "Yes", 1, 0))  
selected\_data <- selected\_data |> select(1:4,8)  
selected\_data <- selected\_data |> mutate(`Engaging/Dull`=if\_else(row\_number()>=1 & row\_number()<=132,1,0))  
View(selected\_data)

Reordering the columns in Meaningful order.

final\_data <- selected\_data |> select(1,6,3,2,4,5)  
View(final\_data)

Exporting the final data into an Excel file.

#write\_xlsx(final\_data, "Cleaned\_Data.xlsx")

# Running ANOVA Tests

Doing the tests for both dependent variables.

final\_data <- read\_xlsx("Cleaned\_Data.xlsx")  
anova\_result <- aov(`Retention\_Score` ~ `Engaging/Dull` \* `Prior\_knowledge` \* `First Year or Not?`,data = final\_data)  
coefficients(anova\_result)

(Intercept)   
 4.0333333   
 `Engaging/Dull`   
 0.6666667   
 Prior\_knowledge   
 0.3666667   
 `First Year or Not?`   
 0.1508772   
 `Engaging/Dull`:Prior\_knowledge   
 0.4666667   
 `Engaging/Dull`:`First Year or Not?`   
 0.6369277   
 Prior\_knowledge:`First Year or Not?`   
 0.4168647   
`Engaging/Dull`:Prior\_knowledge:`First Year or Not?`   
 -1.1896159

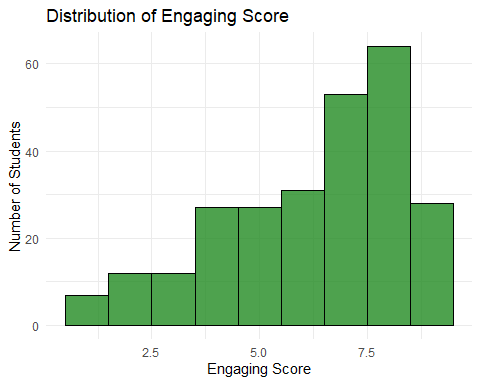
anova2\_result <- aov(`Engaging\_Score` ~ `Engaging/Dull` \* `Prior\_knowledge` \* `First Year or Not?`,data = final\_data)  
coefficients(anova\_result)

(Intercept)   
 4.0333333   
 `Engaging/Dull`   
 0.6666667   
 Prior\_knowledge   
 0.3666667   
 `First Year or Not?`   
 0.1508772   
 `Engaging/Dull`:Prior\_knowledge   
 0.4666667   
 `Engaging/Dull`:`First Year or Not?`   
 0.6369277   
 Prior\_knowledge:`First Year or Not?`   
 0.4168647   
`Engaging/Dull`:Prior\_knowledge:`First Year or Not?`   
 -1.1896159

# Visualizations

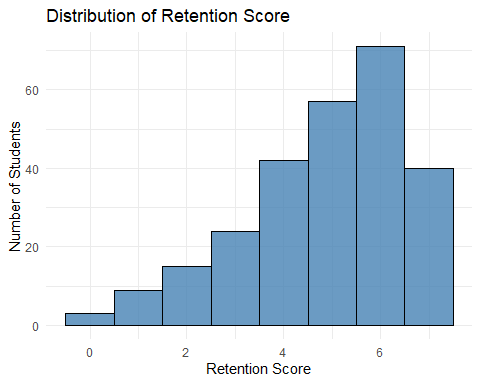
Displaying the distribution for engaging score.

ggplot(final\_data, aes(x = Engaging\_Score)) +  
 geom\_histogram(binwidth = 1, fill = "forestgreen", color = "black", alpha = 0.8) +  
 theme\_minimal() +  
 labs(  
 title = "Distribution of Engaging Score",  
 x = "Engaging Score",  
 y = "Number of Students"  
 )



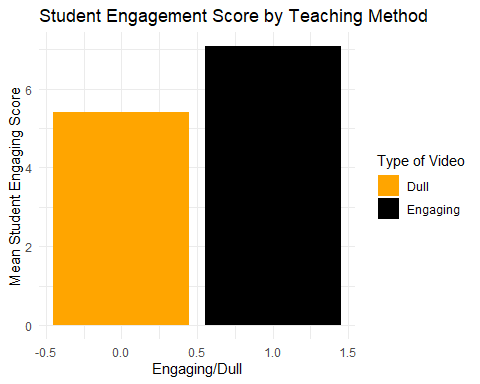
Displaying the distribution for retention score.

ggplot(final\_data, aes(x = Retention\_Score)) +  
 geom\_histogram(binwidth = 1, fill = "steelblue", color = "black", alpha = 0.8) +  
 theme\_minimal() +  
 labs(  
 title = "Distribution of Retention Score",  
 x = "Retention Score",  
 y = "Number of Students"  
 )



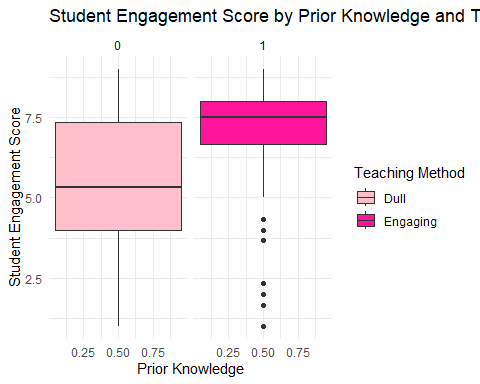
Distribution of Student Engagement scores by teaching method

ggplot(final\_data, aes(x = `Engaging/Dull`, y = Engaging\_Score, fill = factor(`Engaging/Dull`))) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Student Engagement Score by Teaching Method",   
 x = "Engaging/Dull",   
 y = "Mean Student Engaging Score",  
 fill = "Type of Video") + # Custom legend title  
 scale\_fill\_manual(values = c("0" = "orange", "1" = "black"), # Customize colors  
 labels = c("0" = "Dull", "1" = "Engaging")) +  
 theme\_minimal()



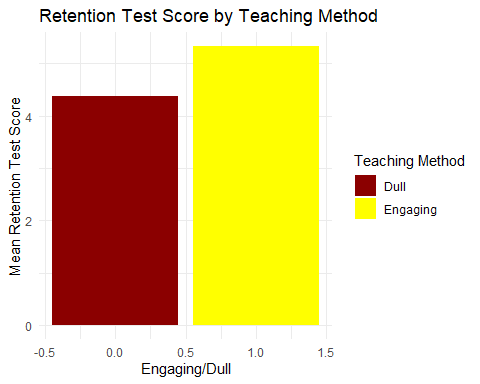
Boxplot for Engaging/Dull and Prior\_knowledge interaction

ggplot(final\_data, aes(x = Prior\_knowledge, y = `Engaging\_Score`, fill = factor(`Engaging/Dull`))) +  
 geom\_boxplot() +  
 labs(  
 title = "Student Engagement Score by Prior Knowledge and Teaching Method",   
 x = "Prior Knowledge",   
 y = "Student Engagement Score",  
 fill = "Teaching Method" # Custom legend title  
 ) +  
 scale\_fill\_manual(  
 values = c("0" = "#FFC0CB", "1" = "#FF1599"), # Customize colors  
 labels = c("0" = "Dull", "1" = "Engaging") # Custom labels  
 ) +  
 facet\_wrap(~ `Engaging/Dull`) +  
 theme\_minimal()



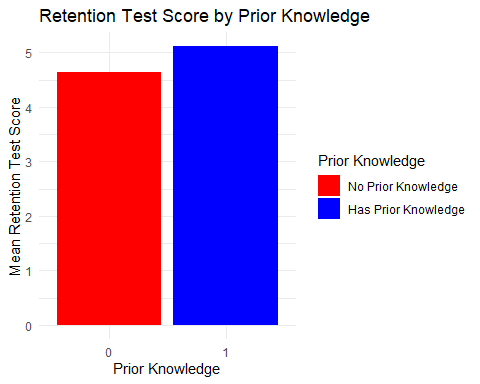
Distribution of Retention Test scores by Teaching Method

ggplot(final\_data, aes(x = `Engaging/Dull`, y = Retention\_Score, fill = factor(`Engaging/Dull`))) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Retention Test Score by Teaching Method",   
 x = "Engaging/Dull",   
 y = "Mean Retention Test Score",  
 fill = "Teaching Method") + # Custom legend title  
 scale\_fill\_manual(values = c("0" = "darkred", "1" = "yellow"), # Customize colors  
 labels = c("0" = "Dull", "1" = "Engaging")) +  
 theme\_minimal()



Distribution of Retention Test scores by Prior Knowledge

ggplot(final\_data, aes(x = factor(Prior\_knowledge), y = Retention\_Score, fill = factor(Prior\_knowledge))) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Retention Test Score by Prior Knowledge",   
 x = "Prior Knowledge",   
 y = "Mean Retention Test Score",  
 fill = "Prior Knowledge") + # Custom legend title  
 scale\_fill\_manual(values = c("0" = "red", "1" = "blue"), # Customize colors  
 labels = c("0" = "No Prior Knowledge", "1" = "Has Prior Knowledge")) +  
 theme\_minimal()



Distribution of Retention Test scores by Year of Study

ggplot(final\_data, aes(x = factor(`First Year or Not?`), y = Retention\_Score, fill = factor(`First Year or Not?`))) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Retention Test Score by Student's Year of Study",   
 x = "First Year or Senior?",   
 y = "Mean Retention Test Score",  
 fill = "Year of Study") + # Update legend title  
 scale\_fill\_manual(values = c("0" = "blue", "1" = "green"), # Customize colors if needed  
 labels = c("0" = "Not First Year", "1" = "First Year")) +  
 theme\_minimal()

