# Import library

library(ggplot2)

getwd()

setwd("C:/Andat")

# Formulate your question

# Pertanyaan: "Apakah jumlah video yang ditonton (Number Of Videos Watched) memengaruhi tingkat penyelesaian kursus (Completion Rate) pada pengguna platform e-learning?"

# Read in your data

data<- read.csv("DataAnalisisDataB4.CSV", sep=",", header= TRUE)

data

# Check the packaging

str(data)

head(data)

tail(data)

# Check your “n”s

dim(data)

# Validate with at least one external data source

# (Penjelasan dilakukan di laporan)

# Make a plot

# Scatter plot untuk hubungan antara Number Of Video Watched dan Completion Rate

ggplot(data, aes(NumberOfVideosWatched, CompletionRate)) +

geom\_boxplot(col = "black", fill = "green")+

labs(title = "Hubungan antara Number Of Video Watched dan Completion Rate",

x = "Number Of Video Watched",

y = "Completion Rate")

# Analisis regresi linier sederhana

model <- lm(CompletionRate ~ NumberOfVideosWatched, data = data)

model

summary(model)

# Membuat histogram variabel dependen (y)

hist(data$CompletionRate,

main = "Histogram Variabel Dependen (y)",

xlab = "y",

col = "yellow",

border = "black",

probability = TRUE)

# Menambahkan kurva distribusi normal pada histogram

curve(dnorm(x, mean = mean(data$CompletionRate, na.rm = TRUE),

sd = sd(data$CompletionRate, na.rm = TRUE)),

col = "red",

lwd = 2,

add = TRUE)

# Membuat histogram data yang berdistribusi normal untuk perbandingan

set.seed(123)

normal\_data <- rnorm(n = length(data$CompletionRate),

mean = mean(data$CompletionRate, na.rm = TRUE),

sd = sd(data$CompletionRate, na.rm = TRUE))

hist(normal\_data,

main = "Histogram Data Berdistribusi Normal",

xlab = "Normal Data",

col = "lightblue",

border = "black",

probability = TRUE)

# Menambahkan kurva distribusi normal pada histogram perbandingan

curve(dnorm(x, mean = mean(normal\_data),

sd = sd(normal\_data)),

col = "red",

lwd = 2,

add = TRUE)

# Menentukan apakah perlu perbaikan model ekspektasi

cat("Apakah perlu perbaikan model ekspektasi?\n")

if (summary(model)$r.squared < 0.1) {

cat("Ya, model ekspektasi perlu diperbaiki karena R-squared rendah, yang menunjukkan model tidak cocok.\n")

cat("Dapat menggunakan model yang lebih kompleks (misalnya, regresi polinomial atau model non-linear) atau menggunakan fitur tambahan yang lebih relevan.\n")

} else {

cat("Tidak, model ekspektasi sudah cukup baik berdasarkan R-squared yang memadai.\n")

}