

```

# Loading the dataset
datf <- read.csv("N:/HR Absenteeism data.csv")

# Data exploration
head(datf)
summary(datf)
str(datf)

# Descriptive analysis
mean_absent_hours <- mean(datf$AbsentHours)
median_absent_hours <- median(datf$AbsentHours)
sd_absent_hours <- sd(datf$AbsentHours)

# Data visualisation
## Creating a histogram of AbsentHours
library(ggplot2)
ggplot(datf, aes(x = AbsentHours)) +
  geom_histogram(binwidth = 10, fill = "pink", color = "turquoise") +
  labs(x = "Absent Hours", y = "Frequency", title = "Distribution of Absent Hours")
## Creating a boxplot of AbsentHours
ggplot(datf, aes(y = AbsentHours)) +
  geom_boxplot(fill = "lightblue") +
  labs(y = "Absent Hours", title = "Box Plot of Absent Hours")
## Creating a scatterplot of AbsentHours
ggplot(datf, aes(x = Age, y = AbsentHours)) +
  geom_point(color = "purple") +
  labs(x = "Age", y = "Absent Hours", title = "Scatter Plot of Age vs. Absent Hours")

# Correlation analysis
correlation_matrix <- cor(datf[c("AbsentHours", "Age")])

# Predictive Analytics (Regression)
model <- lm(AbsentHours ~ Age, data = datf)

# Displaying the regression model summary
summary(model)

```