Naiya Patel

**Summary**

set.seed(10923)

# Number of students  
num\_students <- 500

# Enter code below:

# Simulate study hours (ranging from 1 to 20 hours)  
study\_hours <- sample(1:20, num\_students, replace = TRUE)

# Simulate quiz scores (ranging from 0 to 100)  
quiz\_scores <- sample(0:100, num\_students, replace = TRUE)

# Simulate forum participation (ranging from 0 to 50 posts)  
forum\_posts <- sample(0:50, num\_students, replace = TRUE)

# Simulate previous grades (ranging from 0 to 100)  
previous\_grades <- sample(0:100, num\_students, replace = TRUE)

# Simulate final grades (ranging from 0 to 100)  
final\_grades <- 0.3 \* study\_hours + 0.4 \* quiz\_scores + 0.2 \* forum\_posts + 0.1 \* previous\_grades + rnorm(num\_students, mean = 0, sd = 5) + 25

# Create a data frame  
student\_data <- data.frame(StudyHours = study\_hours, QuizScores = quiz\_scores, ForumPosts = forum\_posts, PreviousGrades = previous\_grades, FinalGrades = final\_grades)

# View the first few rows of the generated data  
head(student\_data)

1. **Set Seed:**
   * **set.seed(10923) ensures the randomness is reproducible. All subsequent random data generation will be consistent if the code is run again.**
2. **Define Number of Students:**
   * **num\_students <- 500 sets the dataset size to 500 students.**
3. **Simulate Study Hours:**
   * **study\_hours <- sample(1:20, num\_students, replace = TRUE) simulates study hours per student ranging from 1 to 20 hours.**
4. **Simulate Quiz Scores:**
   * **quiz\_scores <- sample(0:100, num\_students, replace = TRUE) simulates quiz scores ranging from 0 to 100.**
5. **Simulate Forum Participation:**
   * **forum\_posts <- sample(0:50, num\_students, replace = TRUE) generates forum participation with posts ranging from 0 to 50.**
6. **Simulate Previous Grades:**
   * **previous\_grades <- sample(0:100, num\_students, replace = TRUE) simulates previous grades ranging from 0 to 100.**
7. **Simulate Final Grades:**
   * **final\_grades <- 0.3 \* study\_hours + 0.4 \* quiz\_scores + 0.2 \* forum\_posts + 0.1 \* previous\_grades + rnorm(num\_students, mean = 0, sd = 5) + 25**
     + **This formula calculates a synthetic final grade using weights for each input variable:**
       - **Study Hours: 0.3 weight**
       - **Quiz Scores: 0.4 weight**
       - **Forum Posts: 0.2 weight**
       - **Previous Grades: 0.1 weight**
     + **rnorm(num\_students, mean = 0, sd = 5) adds random Gaussian noise to each student's final grade to simulate real-world variability.**
     + **A baseline value of +25 is added to ensure grades are within a reasonable range.**
8. **Create Data Frame:**
   * **student\_data <- data.frame(StudyHours = study\_hours, QuizScores = quiz\_scores, ForumPosts = forum\_posts, PreviousGrades = previous\_grades, FinalGrades = final\_grades) creates a data frame with columns for each input variable and the calculated final grade.**
9. **View the First Few Rows:**
   * **head(student\_data) displays the first few rows of the dataset for verification.**

**Summary**

**The generated dataset mimics realistic student performance data by simulating four features: study hours, quiz scores, forum participation, and previous grades. These features are combined using specific weights to predict final grades. Gaussian noise is added to replicate real-world unpredictability in student outcomes.**

**Top of Form**