

Assignment 3

Data set

penguins dataset in the palmerpenguins package.

```
library(palmerpenguins)
data(penguins)
```

Training and test set preparation

Take 80% of the observations for the training dataset. When splitting the data, use the last three digits of your index number to set the seed.

```
library(tidyverse)
data(penguins)
df <- penguins %>% mutate(id = row_number())
set.seed(xxx) # use last three digits.
train <- df %>% sample_frac(.80)
dim(train)
```

Task

Develop a model to predict the body mass of penguins using species, island, bill length, flipper length, and gender variables.

You should perform exploratory data analysis before developing the model.

Document your analysis using .Rmd and the output file should be HTML. You should submit both .Rmd file and HTML file.

Due date: 30 September 2022