

Lab Two

MSDS 598 Spring 2022

Directions

- Submit a .ipynb notebook to Canvas.
- The Notebook should begin with a Markdown cell with your **Name** and the title of the Assignment, **Lab 2**. Failure to do so will result in points lost.
- Use Markdown Cells to **clearly** indicate which code answers which question and to answer short answer questions. Failure to do so will result in points lost.
- The filename for your notebook should be formatted like

FirstName_LastName_AssignmentName.ipynb.

Failure to do so will result in points lost.

- This is due on February 14th at Midnight Pacific time.

Use the penguins dataset for the following questions.

0. Recreate the following set-up which we did in class.

- (a) Create the variable `is_male` which is 0 when the penguin is female and 1 when the penguin is male.
- (b) Find the line of best fit to predict body mass from flipper length and whether the penguin is male or female.

1. Create a scatter plot with flipper length on the x-axis and body mass on the y-axis. Color the points according to the penguin's sex.

2. Find the line of best fit for the penguins dataset to predict body mass from flipper length. Plot this on the plot from Q1.

3. Find two lines of best fit: one for the male and one for the female penguins. Plot this on the plot from Q1.

4. Take the line of best fit from above which used both flipper length and `is_male`. Plot this line with $X_{is_male} = 0$ and with $X_{is_male} = 1$. What do you notice?