

Homework

Raoul Grouls

Explore the penguins

Explore the penguins dataset you cleaned the previous lesson. Integrate this new code into your main.py file you have created. Add a README.md file, where you put instructions how to use the code.

Also, make your code so that you don't need to run the preprocessing every time: make it optional with a flag with `click`.

initial exploration

For the initial exploration, you can focus on questions like: - what types of data do you have? - how many unique values do you have for the categorical data? - what are the ranges of the floating point data? - are there outliers? Anything weird?

Separating

After that, you can focus on the question how to separate the data.

- If you look at the species, how do they differ? How do they overlap?
- Plot distributions for the floating point data, with the species as hue.

Interpret what you see. Will it be easy to predict the species of a penguin with just the floating point data? How many variables do you need?

Now do the same for sex. Can you separate male and female?

Implement

Main.py

Pick the plots that make sense for telling a story about the data. You will not need every plot you made; just the plots that illustrate your conclusion.

Implement these visualisations in your main.py script. Add a flag with `click` to run your visualisations. save the plots in `/report/img`

Make a report

Use [quarto](#) to generate a report / presentation. Use your images from `/report/img`