

Brianna Hodge

SI 201: Barbara Ericson

Project 1 Checkpoint

- the name of the dataset you are using,
- the columns you will be working with,
- the calculations you will be performing,
- the function decomposition diagram, and
- the names of collaborators (if any)

Dataset: Penguins

- [penguins](#): A dataset containing detailed information about penguins gathered from Dr. Kristen Gorman and the Palmer Station, Antarctica LTER. Uploaded to Kaggle by Data Science Sean.

Nested dictionary example →

```
{1: {species: Adelie, island: Dream, bill_length_mm: 39.5, bill_depth_mm: 16.7, flipper_length_mm: 178, body_mass_g: 3250, sex: female, year: 2007}}
```

Columns being used:

Columns used for calculating average bill length for a specific species:

Row Number

species

bill_length_mm

Columns used for calculating percentage of female penguins on different islands:

Row Number

island

sex

Calculations:

1. The average value of bill_length_mm for a specific species type

Calculation example: What is the average bill length (bill_length_mm) for Gentoo penguins?

2. The percentage of a dataset that contains a specific value for a categorical variable

Calculation example: What percentage of penguins on Dream island are female?

Names of collaborators: N/A, I am working alone on my project

Function Decomposition Diagram:

