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

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

<u>Calculation example:</u> 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

<u>Calculation example:</u> What percentage of penguins on Dream island are female?

N/A, I am working alone on my project

Function Decomposition Diagram:

main()
Runs the program and calls the functions in a sequential order

Input: None
Output: None

read_penguins_data(csv_file)

Reads in csv file (penguins data) and transform to python
Input: csv_file
Output: penguins (nested dictionaries)
Where the keys are the variables and values are the data related

get_penguin_info(penguins)

Takes penguin dictionary and removes unnecessary info
Input: penguins (nested dictionaries)
Output: returns new dictionary with variables being used for functions

Didn't use the helper function in my actual project!

average_bill_length(penguins)

Calculates average bill length for a specified species
Input: penguins (nested dictionaries)
Output: average lengths for all species

island_female_percentage(penguins)

Calculates percentage of female penguins for different islands

Input: penguins (nested dictionaries)
Output: percentage of females on each island

generate_results()

Takes in output/results from functions and puts into text file
Input: results of average and percentage
Output: (outputs to a text file)