

Assignment #5

Assignment Overview

In this assignment you will create a simple program for using dictionaries to store and process the contents of a very popular dataset, the Iris flower dataset.

Background

The Iris flower dataset is one of the most popular datasets in human history. The dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant: setosa, virginica, or versicolor. For each sample, 4 attributes are stored: petal length, petal width, sepal length, and sepal width.

See: <http://archive.ics.uci.edu/ml/datasets/Iris/> and <https://www.kaggle.com/uciml/iris> and https://en.wikipedia.org/wiki/Iris_flower_data_set for more.

Project Specification

In this assignment you will implement the functionality to read the CSV file containing the dataset, building a dictionary, and producing summary statistics and plots per class.

Your program should:

1. Print a brief (2-5 lines) message explaining its purpose.
2. Read in the data from the `iris.csv` file.
3. Create a dictionary with the key as the species name and the values as (placeholders for) the averages of each of the four attributes/features of each data point: petal length, petal width, sepal length, and sepal width.
4. Compute the averages of each attribute for each species.
5. ("Pretty") print the results.

Requirements

You are required to:

1. Use `import csv` and `csv.reader()` (or `csv.DictReader()`) to read the CSV files.
2. Design and implement **a function that computes the averages** that takes suitable (number of) parameters.
3. Design and implement **another function to handle the "pretty" printing** for this particular type of dictionary.
4. Keep the file handling functionality outside of the functions above.

Deliverables

You must submit (via Canvas):

- The file **a5_username.py** (where "username" is your FAU username; in my case the file would be called **a5_omarques.py**)
 - o This is your source code solution; be sure to include your name, date, assignment number and comments describing your code.
- A **README.md** file with "project notes" (describing what my TA and I cannot see by looking at your source code and/or running your program).
 - o Examples: design decisions, documented limitations, future improvements, etc.

- A **screenshot of the results** produced by your code (one screenshot is enough, since everything is hardcoded and predictable).

It might look like this:

This program computes summary statistics for the Iris Dataset

Species:	Setosa	Versicolor	Virginica
Attributes (cm):			
Avg petal length:	1.87	5.67	6.88
Avg petal width:	6.87	3.67	6.18
Avg sepal length:	1.47	5.22	3.80
Avg sepal width:	7.99	1.64	3.33

Notes and Hints:

- Follow the “cardinal rules” of programming in Python (as per the textbook).
- Start by breaking the program down into parts and solve smaller problems before producing the final solution.
- Try to handle special cases and prevent runtime errors to the best of your knowledge.
- Don't overdo it!

Bonus opportunities:

This is an odd-numbered assignment.

There are no bonus opportunities (unless, of course, you guess my zoom background in related lectures). 😊