

Makenson Noel

A5: Iris Dataset and dictionaries

Code:

```
a5_mnoel2018.py X
python > A5 > a5_mnoel2018.py > ...
1 '''Makenson Noel
2 | Iris Dataset and dictionaries
3 | ...
4 |
5 | def welcome():
6 |     print("\nCreating a simple program for using dictionaries to store and process the contents of a very popular dataset, the Iris flower dataset.\n")
7 |
8 |     print("The Iris flower dataset is one of the most popular datasets in human history. \nThe dataset contains 3 classes of 50 instances each, where each class refers to a type of Iris
9 |
10 | import os
11 | import csv
12 | import pprint
13 | import numpy as np
14 | filename = (r"C:\Users\maken\OneDrive\Documents\apps\python\A5\iris.csv")
15 |
16 | def calculate_average(data_lines):
17 |     # declare sum values for each species
18 |     # first is petal length
19 |     # second is petal width
20 |     # third is sepal length
21 |     # fourth is sepal width
22 |     sum_attributes = {
23 |         "setosa": [0, 0, 0, 0],
24 |         "versicolor": [0, 0, 0, 0],
25 |         "virginica": [0, 0, 0, 0]
26 |     }
27 |
28 |     counts = {
29 |         "setosa": 0,
30 |         "versicolor": 0,
31 |         "virginica": 0
32 |     }
33 |
34 |     # read lines from data_lines
35 |     for row in data_lines:
36 |         species = row[-1] # get the species
37 |         # add the columns
38 |         sum_attributes[species][0] += float(row[2])
39 |         sum_attributes[species][1] += float(row[3])
40 |         sum_attributes[species][2] += float(row[0])
41 |         sum_attributes[species][3] += float(row[1])
42 |         # increase count
43 |         counts[species] += 1
44 |
45 |     # calculate and return a dict of averages
46 |     averages = {
47 |         species: [x / counts[species] for x in sum_attributes[species]] for species in sum_attributes
48 |     }
49 |
50 |     # return averages
51 |     return averages
52 |
53 | def pretty_print(averages):
54 |     # pretty print function
55 |     welcome()
56 |     print("-----")
57 |     print(f'{"Species":<20} {"Setosa":>10} {"Versicolor":>20} {"Virginica":>20}')
58 |     print("-----")
59 |     print("Attributes (cm):")
60 |     print(f'{" Avg petal length":<20} {averages["setosa"][0]:>10.2f} {averages["versicolor"][0]:>20.2f} {averages["virginica"][0]:>20.2f}')
61 |     print(f'{" Avg petal width":<20} {averages["setosa"][1]:>10.2f} {averages["versicolor"][1]:>20.2f} {averages["virginica"][1]:>20.2f}')
62 |     print(f'{" Avg sepal length":<20} {averages["setosa"][2]:>10.2f} {averages["versicolor"][2]:>20.2f} {averages["virginica"][2]:>20.2f}')
63 |     print(f'{" Avg sepal width":<20} {averages["setosa"][3]:>10.2f} {averages["versicolor"][3]:>20.2f} {averages["virginica"][3]:>20.2f}')
64 |     print("\n\n")
65 |
66 | # main script
67 | # open file for reading
68 | with open(filename, newline='') as datafile:
69 |     reader = csv.reader(datafile)
70 |     # skip the header line
71 |     next(reader)
72 |     data = []
73 |     # add all data
74 |     for row in reader:
75 |         data.append(row)
76 |
77 | # calculate averages
78 | averages = calculate_average(data)
79 | pretty_print(averages)
```

Output:

Makenson Noel

A5: Iris Dataset and dictionaries

Creating a simple program for using dictionaries to store and process the contents of a very popular dataset, the Iris flower dataset.

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.

Species	Setosa	Versicolor	Virginica
Attributes (cm):			
Avg petal length:	1.46	4.26	5.55
Avg petal width:	0.24	1.33	2.03
Avg sepal length:	5.01	5.94	6.59
Avg sepal width:	3.42	2.77	2.97

PS C:\Users\maken\OneDrive\Documents\apps>