Flower Classification Data Appendix

This data appendix offers an overview of each variable in **Final_Dataset.csv**. The dataset itself was compiled by merging 3 University of Oxford datasets: a JPG folder of 800+ flowers, a dataset with Image ID and its index, and a final dataset that paired the index with the actual flower classification. For each variable, this data appendix describes the variable itself and provides summary statistics and visualizations, when applicable. The unit of observation is the image itself, which is illustrated as pixel values.

Filename: The name of the image file in the JPG folder, expressed as image_00001.jpg, for example.

	filename
count	8189
unique	8189
top	image_08189.jpg
freq	1

Path: The file path, expressed as /content/jpg/image_00001.jpg, for example. The JPG folder is local–and not pulling from GitHub–because it is too large for GitHub.

	path
count	8189
unique	8189
top	/content/jpg/image_08189.jpg
freq	1

Image: The flower image expressed as pixel values.

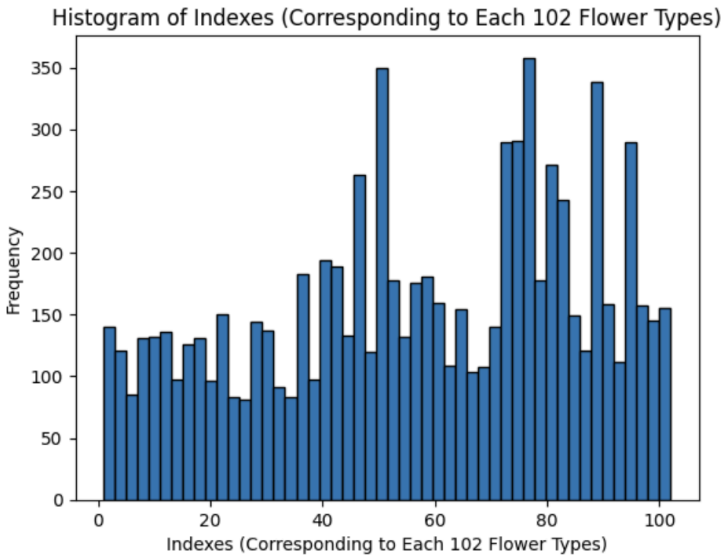
	image
count	8189
unique	8152
top	[[[0 0 0]\n [0 0 0]\n [0 0 0]\n ...\n [0 0...
freq	30

Image_id: A unique numerical value corresponding to each individual image, from 1 to 8,189.

	Image_id
count	8189.00000
mean	4095.00000
std	2364.10501
min	1.00000
25%	2048.00000
50%	4095.00000
75%	6142.00000
max	8189.00000

Index: A numerical value corresponding to the flower classification. There are 102 possible flower classifications total, so the range of possible values is 1 to 102.

	Index
count	8189.000000
mean	56.991696
std	27.730698
min	1.000000
25%	37.000000
50%	59.000000
75%	80.000000
max	102.000000



Name: The name/classification of the flower itself, which corresponds to the index.

	Name
count	8189
unique	102
top	petunia
freq	258