Training set image directory

- Unit of observation: each image corresponds to one snake
- The data is divided into training and testing sets, both of which were sourced from imagescv.com and organized into folders by class before being zipped into a file titled snake_data.zip.

Within the training set, we have:

➢ 1,581 non_venomous images, which include:
  - *Garter_snake*
  - *Green_snake*

➢ 1,357 venomous images, which include:
  - *Diamondback*
  - *Sidewinder*

Within the testing set, we have:

➢ 766 non_venomous images
  - *garter_snake*
  - *green_snake*

➢ 466 venomous images
  - *diamondback*
  - *sidewinder*

➢ **Stage 1: Image Classification Dataset**

- Unit of observation is image
- We first used the zip file package in Google Colab to extract this dataset and the tensorflow library to preprocess the images using ImageDataGenerator.
- Our dataset is composed of labeled snake images used for binary classification: identifying whether a snake is nonvenomous [0] or venomous [1].

Column	**Description**	**Example Data**
image	The image file of the snake	JPG file (e.g., Z0W94369LDU6.jpg)
Classification	Indicates whether the snake is venomous or non-venomous	[venemous], [non_venemous]

- EDA plots:
  - Sampler of images from the training data set

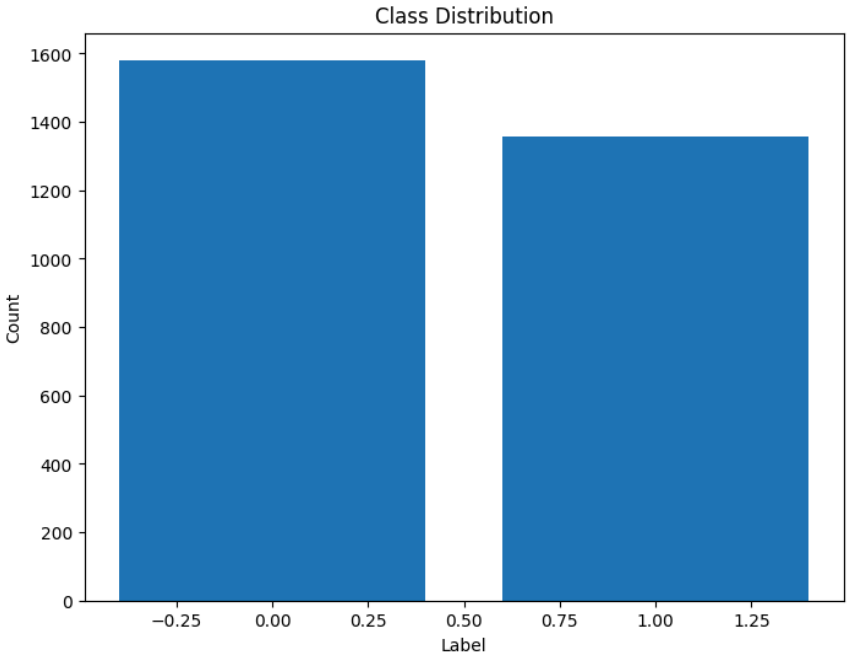
- 

➢ **Stage 2: Flattened Dataset for Traditional Models / EDA**

- The unit of observation is image
- Using numpy, pandas, and matplotlib, we transformed the image tensors into lists of pixel values and created a dataframe where each image corresponds to a binary class label (0 or 1). This structured data was then exported as a CSV file for further analysis and visualization.

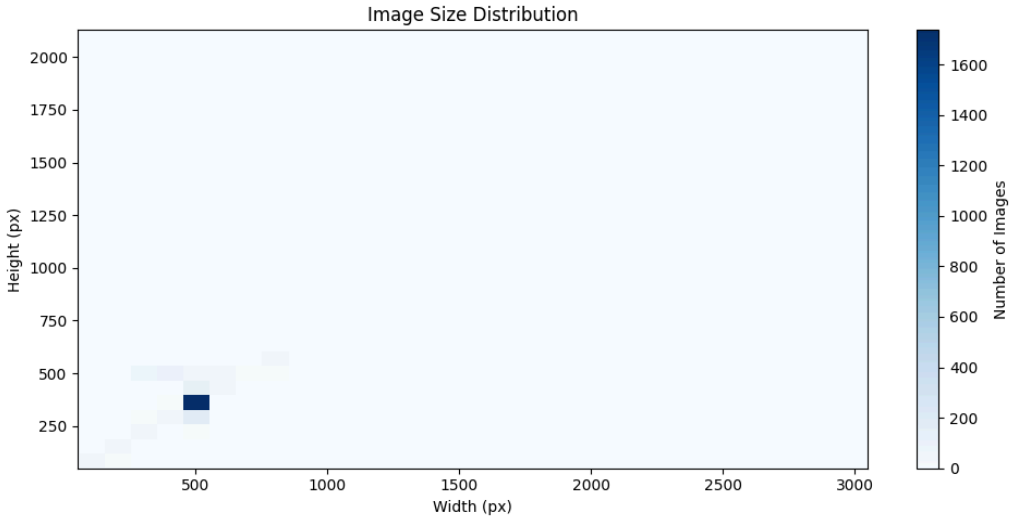
Column	**Description**	**Example Data**
label	Binary variable that describes the class of the image where 0 is non-venomous and 1 is venomous	0
image	A list of scaled pixels associated with each image. Values range from 1 to -1.	[-0.458236, -0.423578, …]

- **EDA Plots**
  - Plot of number of images within each class (0 and 1)



-

- Graph of image sizes within training set



-