

# Plant Edibility Data

## Dataset: Pl@ntNet-300K (Test Set)

The Pl@ntNet-300K dataset contains 306,146 plant images spanning 1,081 species. The dataset exhibits strong class imbalance—a small number of species have a disproportionately high number of images. For example, *Cirsium vulgare* has 648 images, while *Barringtonia acutangula* has only 1 image.

This appendix focuses on the test split of the dataset, which consists of 31,112 images. These test images are used to evaluate the performance of pre-trained classification models from Pl@ntNet.

## Unit of Observation

Each observation in the dataset represents a single labeled image of a plant species.

## Dataset Content

- Total images in test set: 31,112
- Total species covered: 1,081
- Species with the most images: *Daucus carota* (902 images)
- Species with only one image: 296 species
- Species with only two images: 151 species

Each species' images are stored in separate folders within the test dataset. These folders are named using 7-digit numeric IDs, which map to the species' scientific names using a JSON file. For example: folder 1568914 contains images of *Acacia saligna* (golden wreath wattle).

## Variables

There are no explicit variables in this dataset aside from:

- Image content (JPG files)
- Folder ID (used as a species identifier)

## Summary Statistics

Following statistics show image resolution information.

Average width	569.72 pixels
---------------	---------------

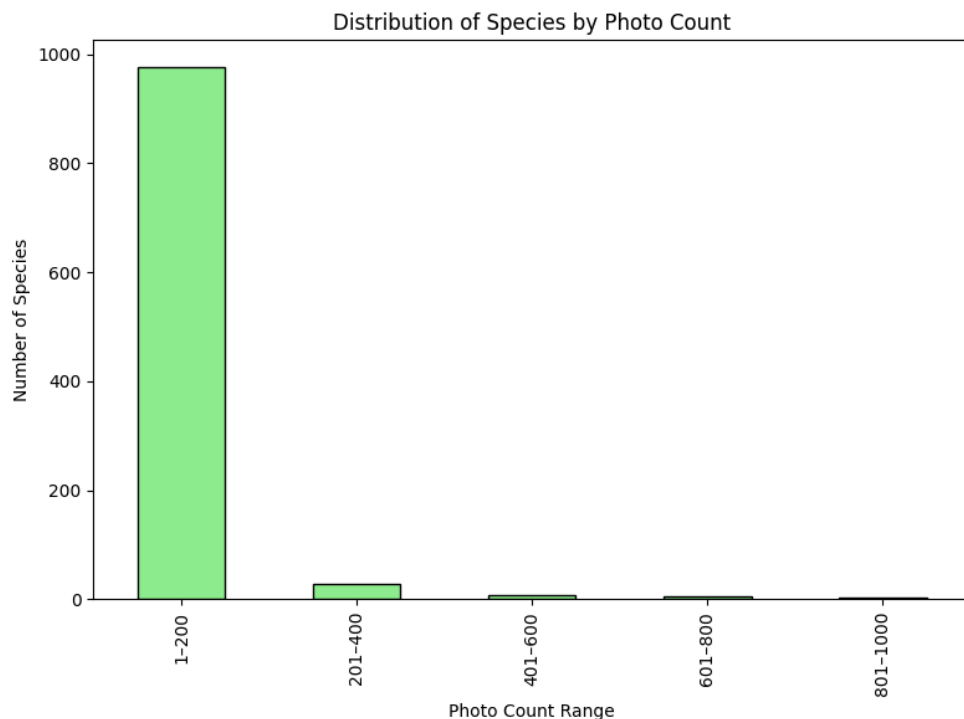
Average height	569.92 pixels
Average total pixels per image	328006.98 pixels

## Histogram

The histogram below shows the distribution of species photo counts:

- X-axis: photo count ranges (e.g., 1–200, 201–400, etc.)
- Y-axis: number of species within each range

This visual emphasizes the class imbalance—a few species have many images, while most species have very few.



## Mapping file: plantnet300K\_species\_names.json

### Unit of Observation

Each entry (key-value pair) maps a 7-digit numeric label (folder ID) to a scientific plant species name.

### Variables

- Key: 7-digit species label (e.g., "1568914")
- Value: Full Latin name of the species (e.g., "Acacia saligna")