

Omdena (Algeria Chapter)  
Green Algeria Project

# Task - 2

EDA for diseases and pest detection



# Task Overview

This task is aimed to perform Exploratory Data Analysis (EDA) on the selected Plant Village dataset.

This would help us to:

- > Check the consistency w.r.t **number of images** in the Plant Village dataset.
- > Check the consistency w.r.t **size of images** in the Plant Village dataset.
- > Should we proceed with the selected dataset or we resort to merging multiple datasets?

# EDA

Analyzing the EDA shared by members, The following traits about the **Plant Village Dataset** were brought into light:

- > The dataset is highly imbalanced. Number of Images per classes are inconsistent.
- > Inconsistent Image Sizes (Aspect Ratio). Multiple classes were found containing images with inconsistent dimensions.

# The Merged Dataset

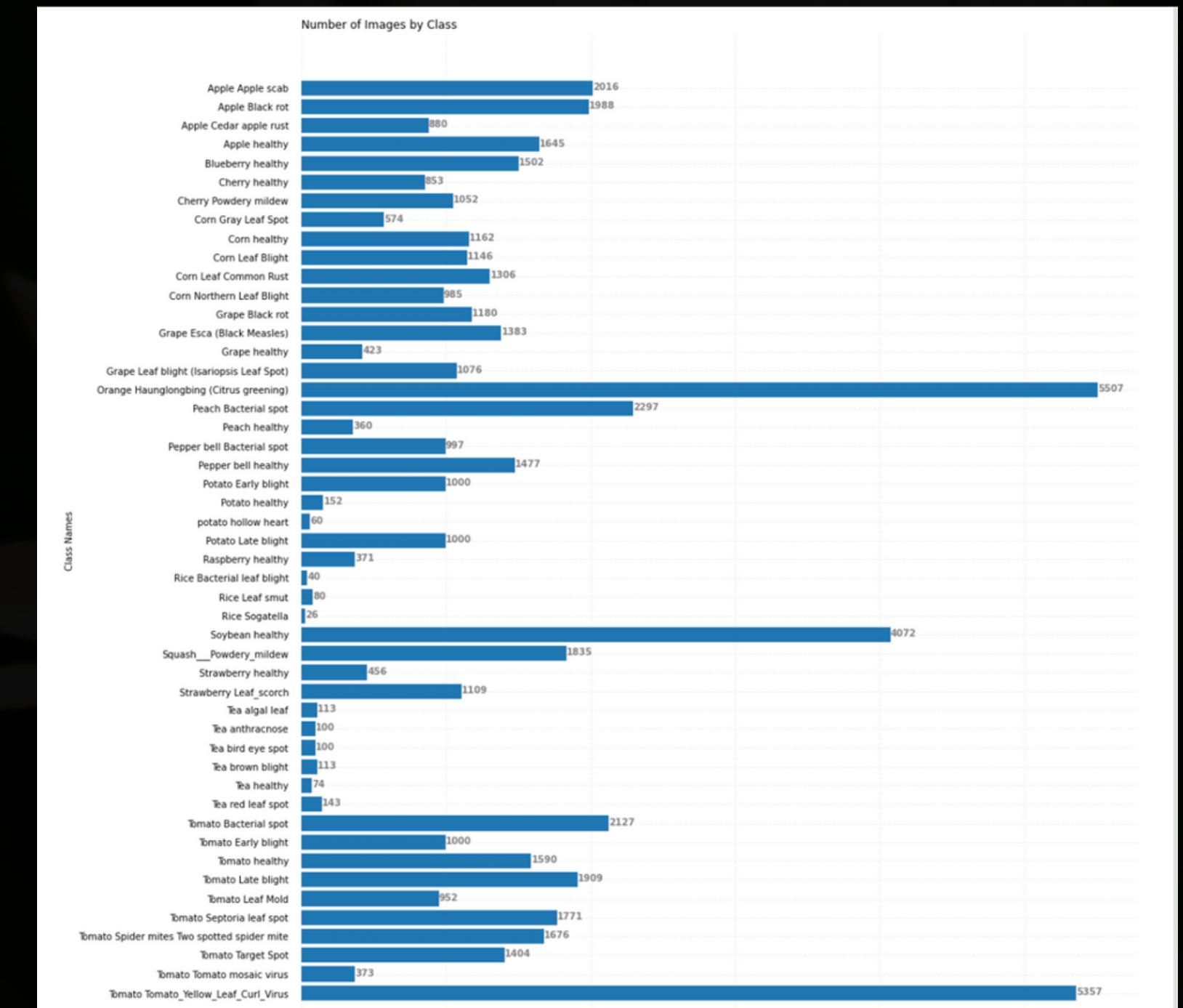
Plant Species : 16  
Total Classes : 49

The following datasets are to be merged:

- Plant Village
- Plant Disease Expert
- Image database

We can perform the following actions to balance the merged dataset:

- > Remove the Unwanted Classes.
- > Perform Data Augmentation in classes with less images.
- > Undersampling the classes with excess images.



# How do we proceed?

> Start with **Pre-Processing** the merged dataset.

Approaches for model building after Pre-Processing:

- **One-Shot Classification** (Karan Talwar)
- **Anomaly Detection** (Satyaki Bhattcharjee)
- **Transfer Learning** (Vikash Sahni)'

All 3 models can be used together to achieve a higher accuracy score.



Thank You

# List of Contributors

## Contributors

Karan Talwar

Satyaki Bhattcharjee

Safia Labdi

Vineet Sharma

Vikash Sahni

Intekhab Alam

Ahilya Kale