# Weed Detection Project – Report – Week-1

**Basic Summary:**

The dataset contains 17509 unique images, 9 labels (identifiers for species), 9 corresponding species.
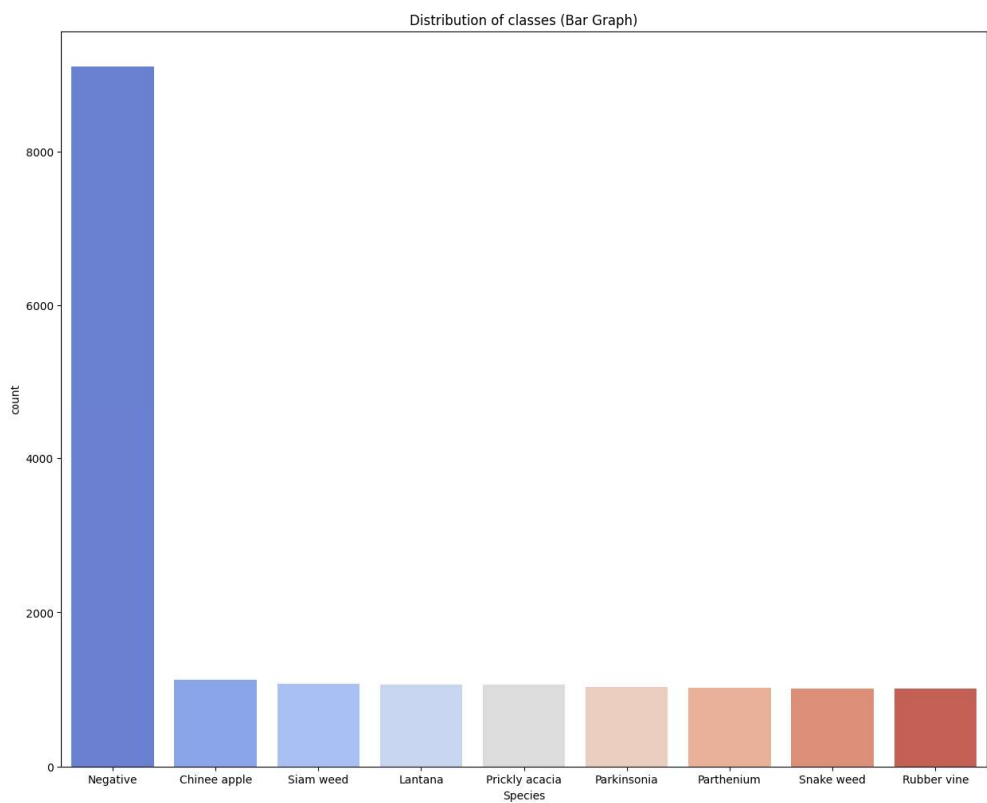
Label-to-Species mapping:

- **Label 0** → Chinee apple

- **Label 1** → Lantana

- **Label 2** → Parkinsonia

- **Label 3** → Parthenium

- **Label 4** → Prickly acacia

- **Label 5** → Rubber vine

- **Label 6** → Siam weed

- **Label 7** → Snake weed

- **Label 8** → Negative (no weed present)

**Specific Details:**

Images per class:

```
Species
Negative          9106
Chinee apple      1125
Siam weed         1074
Lantana           1064
Prickly acacia    1062
Parkinsonia       1031
Parthenium        1022
Snake weed        1016
Rubber vine       1009
Name: count, dtype: int64
```

## Class Distribution Graph:



## 20 Random samples:

Preview of DataFrame (df.head()):

| | Filename | Label | Species |
|---|---|---|---|
| 0 | 20160928-140314-0.jpg | 0 | Chinee apple |
| 1 | 20160928-140337-0.jpg | 0 | Chinee apple |
| 2 | 20160928-140731-0.jpg | 0 | Chinee apple |
| 3 | 20160928-140747-0.jpg | 0 | Chinee apple |
| 4 | 20160928-141107-0.jpg | 0 | Chinee apple |

Problem of Class Imbalance:

This can be resolved by data augmentation, where we balance the positive 8 classes and the single negative class.

Ideally, we could augment each image in 7-8 different ways in order to come close to balanced distribution, else there will be a problem in accuracy as the model will learn a lot of negative examples and fewer positive ones. And we need positive examples to be identified since we have to classify negative v/s positive as well as inter-class classification.