
# Analytics Exercise

Henrique Zaidan

---

## Assumptions

Two assumptions were proposed along with this document:

1. Each row represents an issue;
2. The report focused on plants flagged with health problems because they are costly to the client. As a result, solving this problem increases the client's ROI.

## 1 - Overview

| | |
|---|---|
| **Period** | November, 2021 - January, 2023 |
| **Issues** | 232,652 |
| **Plants with health problems** | 505,005 |

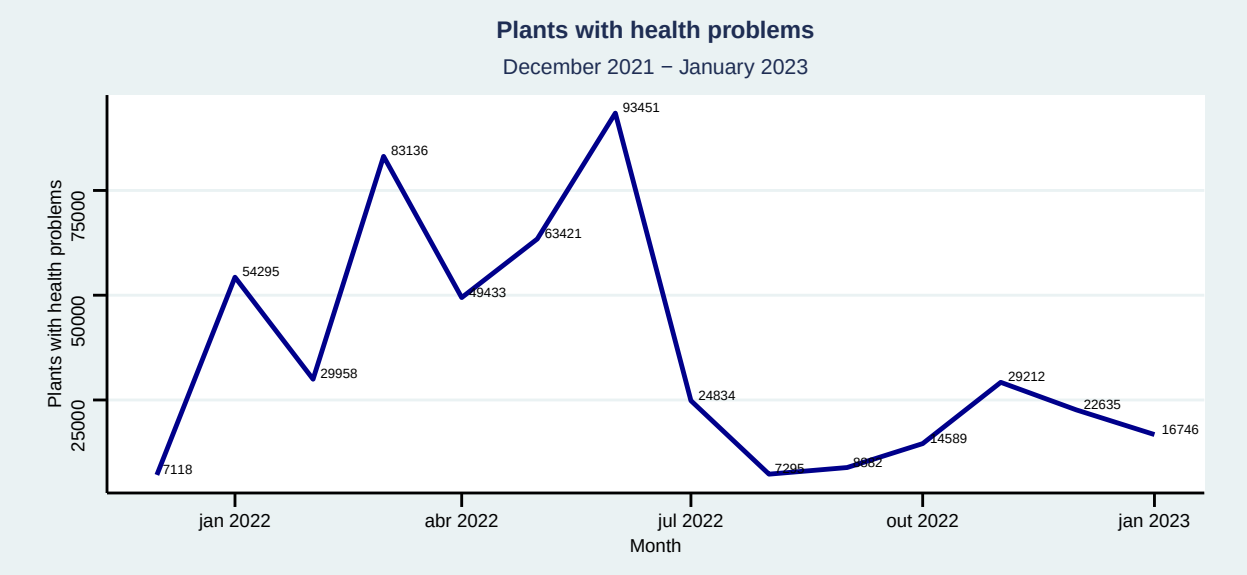
- 15% of the issues showed at least one plant with health problems.

### Critical scenario

Based on the data, five features are associated with plants with health problems (more frequency):

| | |
|---|---|
| **Room** | 7, 8 |
| **Croop week** | 8, 9, 10 |
| **Cultivar** | PE, HCC, ZMNTS, HG |
| **Colored severity** | Yellow |
| **Issue** | Viral anomaly, Nutrient Def |

# 2 - Problem Statement




Based on the plot above, it is possible to see that June was the month with more plants with health problems. The reason for that is associated with four points:

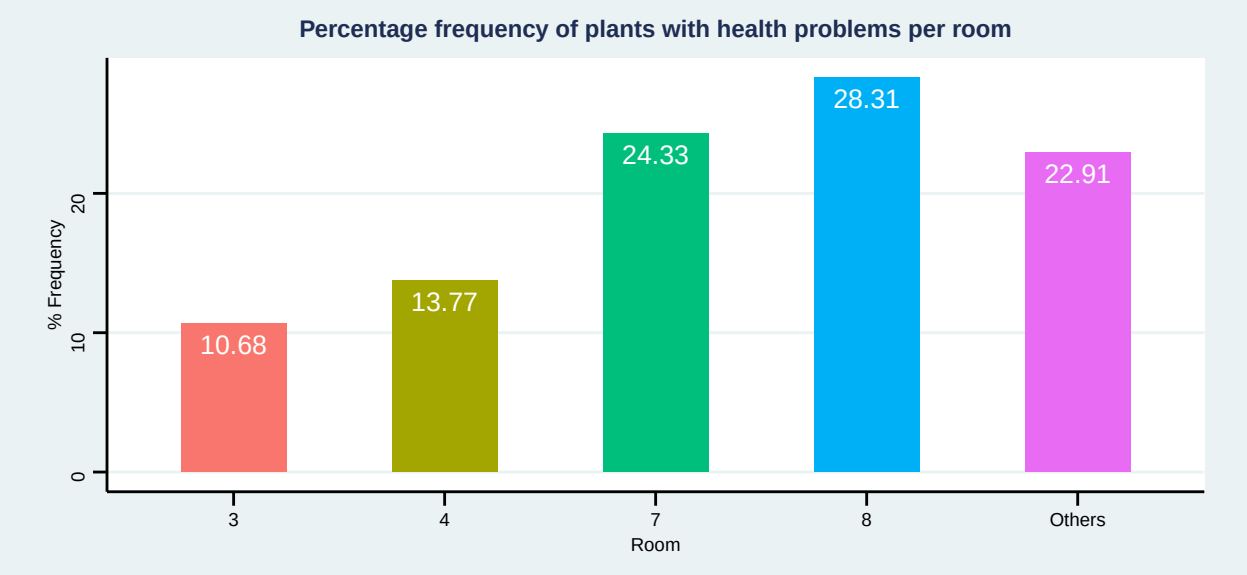
1. **Issue** - Viral anomaly;
2. **Cultivar** - PE;
3. **Colored Severity** - Yellow;
4. **Room** - 8.

This problem started in the middle of May and extended until the end of June. From July, the frequency of plants with issues decreased considerably. Finally, comparing the first and second semesters, the second one presented a lower frequency of ill plants.

The average of plants with health problems per month was 36072.

## Room

**Rooms 7 and 8 concentrated 53% of the plants with health problems**. In total, there are eight types of rooms.

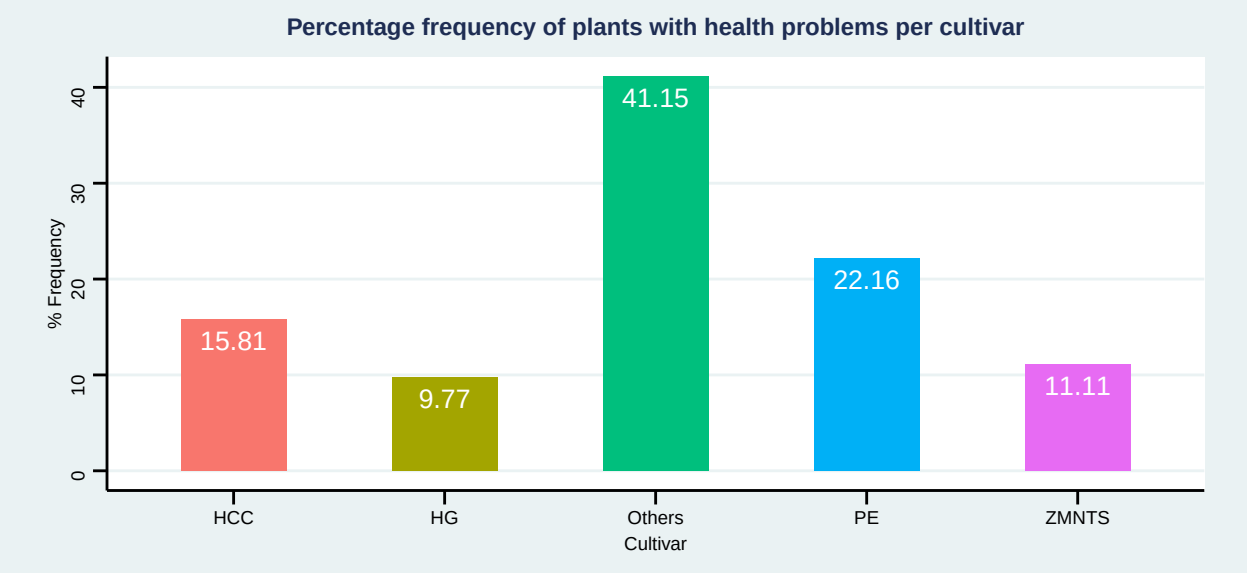



## Crop Week

**Between the harvest weeks 5 and 11, the frequency of plants with health problems was 92%**. Weeks 8, 9, and 10 present a higher frequency than others. These three concentrated around 50% of ill plants. In total, there are 13 crop weeks.

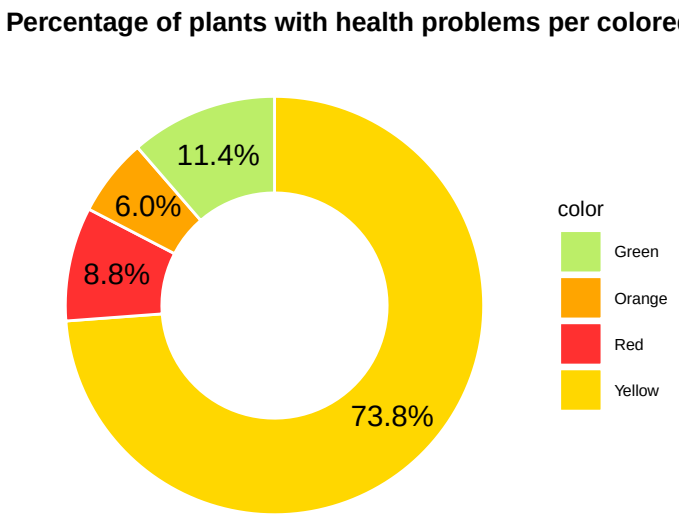
## Cultivar

**Four types of cultivars concentrated 59% of plants with health problems** - PE, HCC, ZMNTS, and HG. In total, there are 52 types of cultivars.



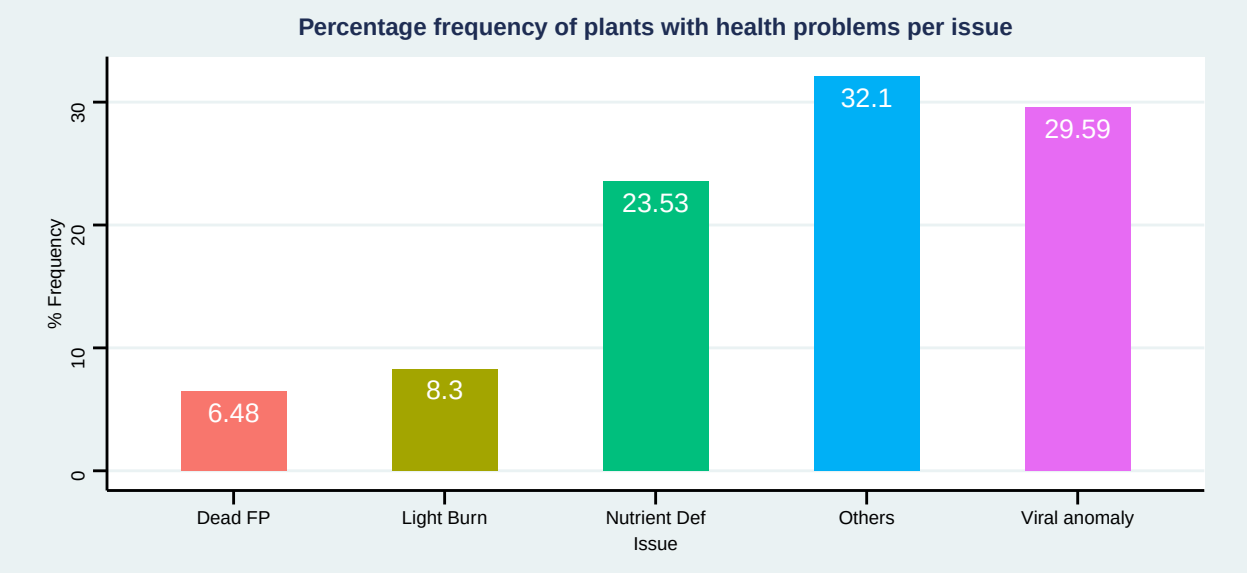

## Colored Severity

**Yellow was the color present in 74% of the plants with health problems**. In total, there are 4 types of colored severity.




## Issue

**Four types of issues concentrated 68% of the plants with health problems** - Viral anomaly, Nutrient Def, Light Burn, and Dead FP. In total, there are 40 types of them.




- **Note** - Comparing the variable week and plants with health problems did not show a pattern. Thus, the chronological timeline did not reflect a trend -a higher (or lower) quantity of plants with health problems along with time.

# 3 - Recommendations

The problem statement focused on the find a pattern related to the plants with health problems. The main idea is to check which attributes affect the ill plants more. In broad words, some levels of the features Room, Crop week, Cultivar, Colored severity, and Issue concentrated more ill plants.

## Strategy 1

It is possible to develop a machine-learning model of the classification to check if a plant can be healthy or not. Thus, if the probability is higher than the threshold, the plant can be classified as ill, and treatment is applied.

## Strategy 2

A contingency plan can be designed considering the levels of the features of the critical scenario. An action plan for the crop week can be planned, for example. The highest frequency of the plants with health problems was concentrated in the middle of the cycle. Therefore, the monitoring can be done considering weeks between 8 and 10. Another action is checking rooms 7 and 8 to know because more plants are ill.