**Explanation File:**

Data Loading and Description:

- We observed that the columns, except for Y1, contain String type data.

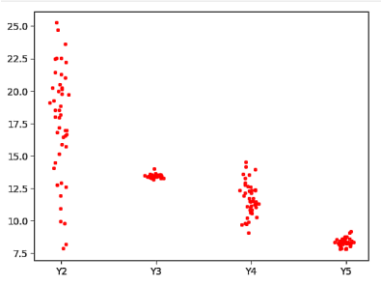
- Additionally, some rows have missing values represented by '[]', resulting in entire rows being affected.

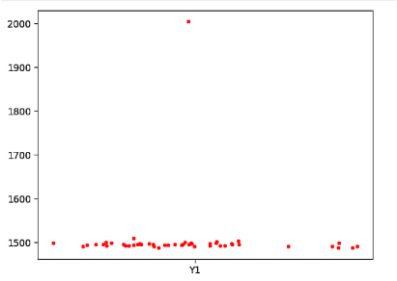
- To address this, we performed data cleaning to remove the missing values.

Visualization and Explanation:

- Our data displays an uneven distribution, with Y5 showing a more concentrated pattern and Y2 displaying a scattered distribution.

- To mitigate this issue, we normalized the data, ensuring that one variable's values do not overpower others.




Based on our observation, we found that a specific row with Y1 equal to 2004 significantly impacts the data. Thus, we decided to exclude this row.

Preprocessing:

- The identified row was removed from the dataset, and the data was further normalized as previously described.

Data Exploring:

- For ranking new devices based on their probability of belonging to the sample group, we utilized Mahalanobis distance calculation.

- We also normalized the data in the file 'new_device.csv' for consistency.

Bonus:

- In this section, we sought to find the threshold using cross-validation.

- By extracting samples from the train set and validating them, we calculated the distances and determined the maximum distance to serve as the threshold for evaluating device failure.