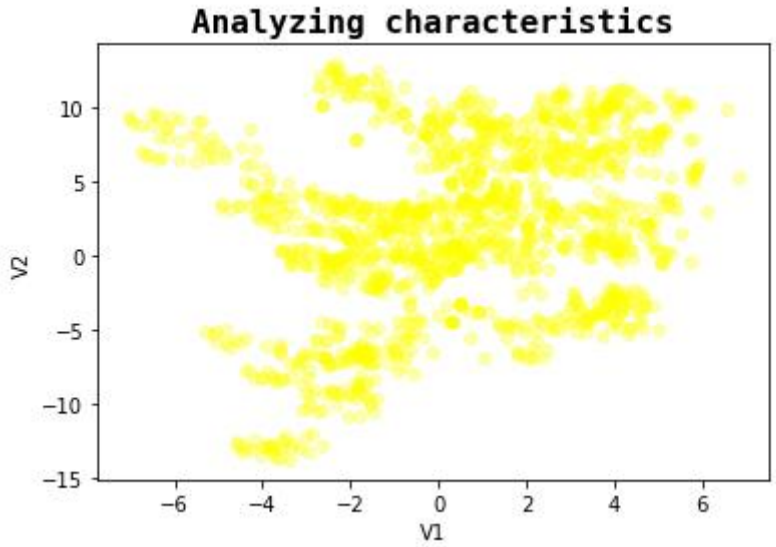
by Joao Pedro

Data evaluation in relation to the veracity of banknotes

1) We will import some data from a spreadsheet (.csv) and check if these values representing each cell correspond to a real banknote or a fake banknote. In this case, we must separate the data into 2 groups.

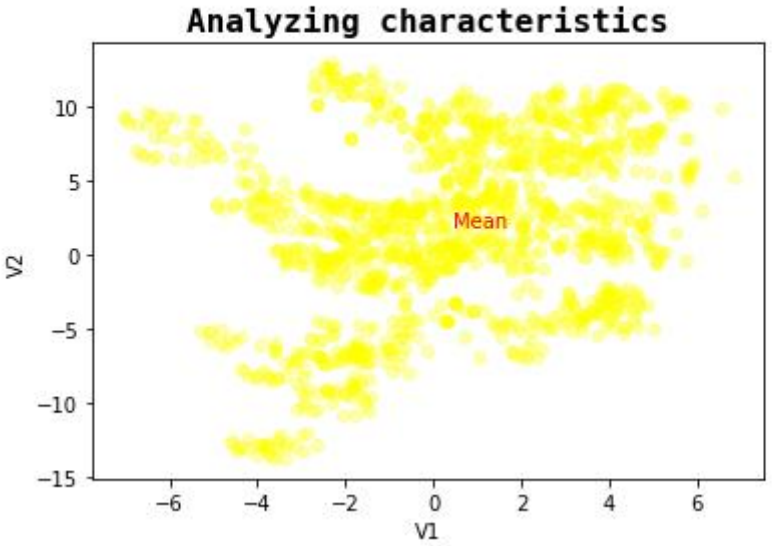
2) Initially, we must verify if the data that will answer our doubts are numerical or characteristic data. So, as it is numerical data, at this point, we must plot the points (data) on a graph.



3) Now, we can see that there is the possibility of using some basic statistical measures (such as the mean and the standard deviation).

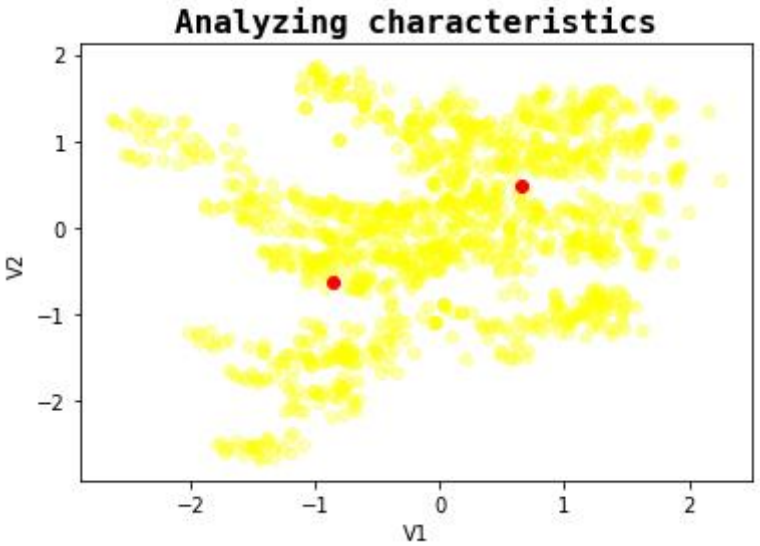
```
Mean: ( 0.43373525728862977 ; 1.9223531209912539 )

Standard deviation: ( 2.841726405206097 ; 5.866907488271993 )
```

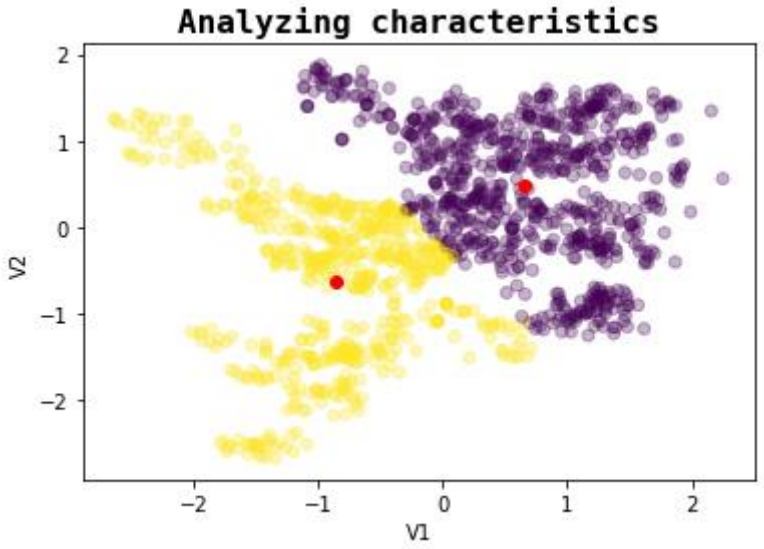


4) Next, we can use a method called "standardization" and a method "k-means clustering". The first method will help to "refine" the discrepancy of one data in relation to another. The second method, can separate the data from each cell into 2 large groups.

To develop this, we need to find the 2 centroids that will be the base of the 2 big groups.



5) Now, we must separate the points into 2 groups. We will divide these groups into two colors.



6) In the end, we can see that the lower dots (the group with the yellow color) correspond to the group of real banknotes. The top dots (the purple group) correspond to the group of counterfeit banknotes.

7) Recommendation: You can now use these results in a machine that processes both real and fake banknotes. Use the same procedure in this report to orient a machine.