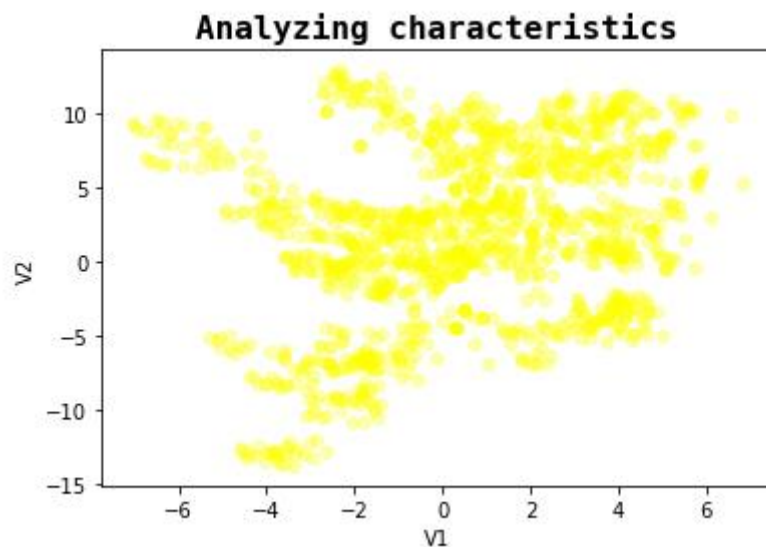


Avaliação de dados em relação à veracidade de cédulas

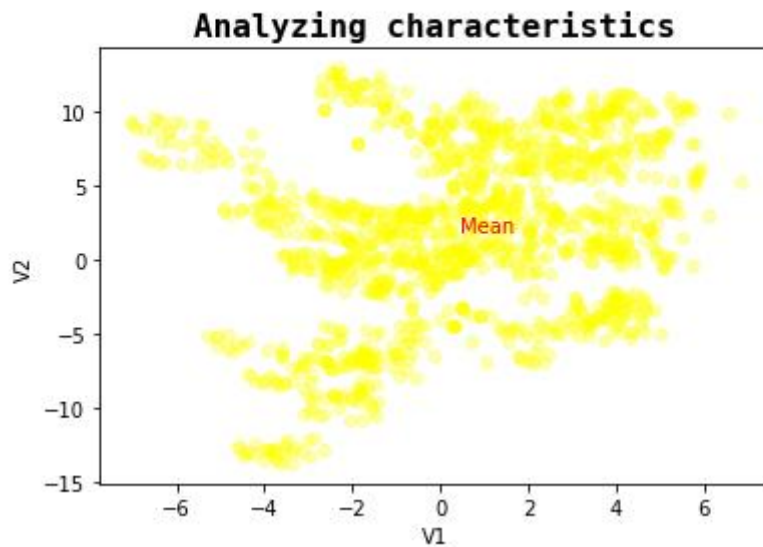
- 1) Iremos importar alguns dados de uma planilha (.csv) e verificar se esses valores que representam cada célula correspondem a uma cédula real ou uma cédula falsa. Nesse caso, devemos separar os dados em 2 grupos.
- 2) Inicialmente, devemos verificar se os dados que responderão às nossas dúvidas são dados numéricos ou característicos. Então, como são dados numéricos, neste momento, devemos plotar os pontos (dados) num gráfico.



- 3) Agora, podemos notar que há a possibilidade de utilizar algumas medidas estatísticas básicas (como a média e o desvio padrão).

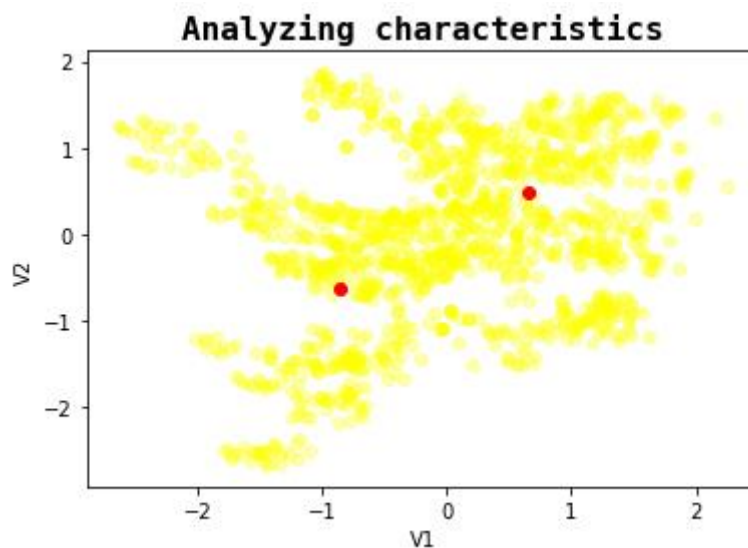
Mean: (0.43373525728862977 ; 1.9223531209912539)

Standard deviation: (2.841726405206097 ; 5.866907488271993)

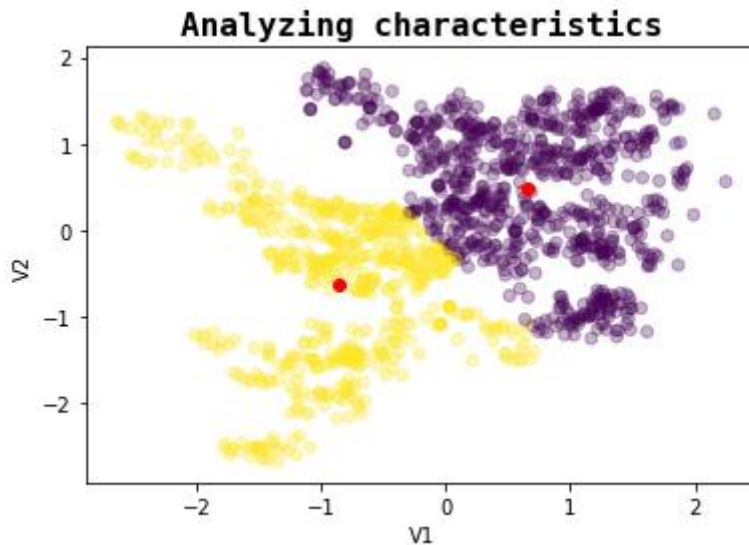


4) Em seguida, podemos utilizar um método chamado de "padronização" e um método "k-means clustering". O primeiro método, vai ajudar a "refinar" a discrepância de um dado em relação a outro. O segundo método, pode separar os dados de cada célula em 2 grandes grupos.

Para desenvolver isso, precisamos encontrar os 2 centróides que serão base dos 2 grandes grupos.



5) Agora, devemos separar os pontos em 2 grupos. Iremos dividir estes grupos em duas cores.



6) No fim, nós podemos perceber que os pontos de baixo (o grupo de cor amarela) correspondem ao grupo de cédulas verdadeiras. Já os pontos de cima (o grupo de cor roxa) correspondem ao grupo de cédulas falsas.

7) Recomendação: agora, você pode utilizar esses resultados numa máquina que processe cédulas verdadeiras e cédulas falsas. Utilize o mesmo procedimento deste relatório para orientar uma máquina.