

```
In [12]: import pandas
import numpy
import matplotlib.pyplot as plt

In [15]: from sklearn.model_selection import train_test_split

In [5]: dados = pandas.read_csv(".\DropsBrawl\Starr Drops.txt", delimiter="|")

In [6]: print(dados)

      data data_modificacao      tipo  raridade
0  27/06/2023      30/12/2023    Diário      Raro
1  27/06/2023      30/12/2023    Diário      Raro
2  27/06/2023      30/12/2023    Diário  Lendário
3  28/06/2023      30/12/2023    Diário      Raro
4  28/06/2023      30/12/2023    Diário  Super Raro
..      ...
879 15/01/2024      15/01/2024    Diário      Raro
880 15/01/2024      15/01/2024    Diário  Super Raro
881 15/01/2024      15/01/2024    Diário  Super Raro
882 15/01/2024      15/01/2024  Promocional      Raro
883 15/01/2024      15/01/2024  Promocional  Lendário

[884 rows x 4 columns]

In [8]: dados['tipo'].value_counts()

Out[8]:
tipo
Diário      676
Mega Pig    119
Promocional   89
Name: count, dtype: int64

In [9]: dados['raridade'].value_counts()

Out[9]:
raridade
Raro      439
Super Raro 245
Épico     130
Mítico     43
Lendário   27
Name: count, dtype: int64

In [17]: dados.iloc[:,3]

Out[17]:
0      Raro
1      Raro
2  Lendário
3      Raro
4  Super Raro
...
879      Raro
880  Super Raro
881  Super Raro
882      Raro
883  Lendário
Name: raridade, Length: 884, dtype: object

In [38]: #teste de stratificação
x, _, y, _ = train_test_split(dados.iloc[:,[0,1,3]], dados.iloc[:,2], test_size=0.5, stratify = dados.iloc[:,2])

In [39]: #testei mudar a coluna para ver o comportamento
tx, _, ty, _ = train_test_split(dados.iloc[:,[0,1,3]], dados.iloc[:,2], test_size=0.5, stratify = dados.iloc[:,3])

In [54]: #base original
dados["raridade"].value_counts()

Out[54]:
raridade
Raro      439
Super Raro 245
Épico     130
Mítico     43
Lendário   27
Name: count, dtype: int64

In [55]: #estratificado baseado no tipo
x["raridade"].value_counts()

Out[55]:
raridade
Raro      221
Super Raro 123
Épico      56
Mítico      26
Lendário    16
Name: count, dtype: int64

In [56]: #estratificado baseado na raridade
tx["raridade"].value_counts()

Out[56]:
raridade
Raro      220
Super Raro 122
Épico      65
Mítico      21
Lendário    14
Name: count, dtype: int64

In [57]: #base original
dados["tipo"].value_counts()

Out[57]:
tipo
Diário      676
Mega Pig    119
Promocional   89
Name: count, dtype: int64

In [58]: #estratificado baseado no tipo
y.value_counts()

Out[58]:
tipo
Diário      338
Mega Pig     60
Promocional  44
Name: count, dtype: int64

In [59]: #estratificado baseado na raridade
ty.value_counts()

Out[59]:
tipo
Diário      337
Mega Pig     62
Promocional  43
Name: count, dtype: int64

In [60]: x

Out[60]:
      data  data_modificacao  raridade
793  31/12/2023      14/01/2024  Super Raro
630  04/12/2023      14/01/2024      Raro
514  16/11/2023      14/01/2024      Épico
130  09/08/2023      07/01/2024      Épico
552  22/11/2023      14/01/2024  Super Raro
...      ...
450  04/11/2023      14/01/2024  Super Raro
144  14/08/2023      07/01/2024      Épico
737  18/12/2023      14/01/2024  Super Raro
427  29/10/2023      14/01/2024      Raro
286  01/10/2023      14/01/2024  Super Raro

442 rows x 3 columns

In [61]: y

Out[61]:
793      Diário
630      Mega Pig
514      Diário
130      Diário
552      Diário
...
450  Promocional
144      Diário
737      Mega Pig
427      Diário
286      Diário
Name: tipo, Length: 442, dtype: object

In [ ]: #from sklearn.model_selection import train_test_split
#a, _, b, _ = train_test_split(dados.iloc[l_b,c_a], dados.iloc[l_b,c_b], test_size=tam, stratify = dados.iloc[l_c,c_c])
#l_a são as linhas da tabela que vão para a, c_a são as colunas
#l_b são as linhas da tabela que vão para b, c_b são as colunas
#tam é quanto quer dividir, ex em 40%, teria que colocar tam=0.6, o 1(100%) - 0.4(40% desejado)
#l_c são as linhas que serão baseadas para dividir a colunas que serão usadas de base para dividir no tamanho desejado(com base nos valores da linha
#se tiver 10 valores 1 e 20 valores 2 e dividir em 40%, ele irá separar 4 valores 1 e 8 valores 2.), c_c são as colunas

In [63]: a, _, b, _ = train_test_split(dados.iloc[:,[0,1,3]], dados.iloc[:,2], test_size=0.6, stratify = dados.iloc[:,2])

In [64]: a

Out[64]:
      data  data_modificacao  raridade
158  19/08/2023      07/01/2024      Raro
684  09/12/2023      14/01/2024      Raro
444  03/11/2023      14/01/2024      Raro
834  08/01/2024      14/01/2024      Épico
602  28/11/2023      14/01/2024      Raro
...      ...
189  29/08/2023      07/01/2024      Raro
164  21/08/2023      07/01/2024  Super Raro
808  05/01/2024      14/01/2024      Raro
592  28/11/2023      14/01/2024  Super Raro
224  10/09/2023      14/01/2024      Mítico

353 rows x 3 columns

In [65]: a["raridade"].value_counts()

Out[65]:
raridade
Raro      163
Super Raro 101
Épico      59
Mítico      19
Lendário    11
Name: count, dtype: int64

In [66]: b.value_counts()
```

```
Out[66]: tipo
Diário      270
Mega Pig    47
Promocional 36
Name: count, dtype: int64
```