

In [2]:

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv('Downloads/data_passwords.csv',error_bad_lines=False)
```

b'Skipping line 2810: expected 2 fields, saw 5\nSkipping line 4641: expected 2 fields, saw 5\nSkipping line 7171: expected 2 fields, saw 5\nSkipping line 11220: expected 2 fields, saw 5\nSkipping line 13809: expected 2 fields, saw 5\nSkipping line 14132: expected 2 fields, saw 5\nSkipping line 14293: expected 2 fields, saw 5\nSkipping line 14865: expected 2 fields, saw 5\nSkipping line 17419: expected 2 fields, saw 5\nSkipping line 22801: expected 2 fields, saw 5\nSkipping line 25001: expected 2 fields, saw 5\nSkipping line 26603: expected 2 fields, saw 5\nSkipping line 26742: expected 2 fields, saw 5\nSkipping line 29702: expected 2 fields, saw 5\nSkipping line 32767: expected 2 fields, saw 5\nSkipping line 32878: expected 2 fields, saw 5\nSkipping line 35643: expected 2 fields, saw 5\nSkipping line 36550: expected 2 fields, saw 5\nSkipping line 38732: expected 2 fields, saw 5\nSkipping line 40567: expected 2 fields, saw 5\nSkipping line 40576: expected 2 fields, saw 5\nSkipping line 41864: expected 2 fields, saw 5\nSkipping line 46861: expected 2 fields, saw 5\nSkipping line 47939: expected 2 fields, saw 5\nSkipping line 48628: expected 2 fields, saw 5\nSkipping line 48908: expected 2 fields, saw 5\nSkipping line 57582: expected 2 fields, saw 5\nSkipping line 58782: expected 2 fields, saw 5\nSkipping line 58984: expected 2 fields, saw 5\nSkipping line 61518: expected 2 fields, saw 5\nSkipping line 63451: expected 2 fields, saw 5\nSkipping line 68141: expected 2 fields, saw 5\nSkipping line 72083: expected 2 fields, saw 5\nSkipping line 74027: expected 2 fields, saw 5\nSkipping line 77811: expected 2 fields, saw 5\nSkipping line 83958: expected 2 fields, saw 5\nSkipping line 85295: expected 2 fields, saw 5\nSkipping line 88665: expected 2 fields, saw 5\nSkipping line 89198: expected 2 fields, saw 5\nSkipping line 92499: expected 2 fields, saw 5\nSkipping line 92751: expected 2 fields, saw 5\nSkipping line 93689: expected 2 fields, saw 5\nSkipping line 94776: expected 2 fields, saw 5\nSkipping line 97334: expected 2 fields, saw 5\nSkipping line 102316: expected 2 fields, saw 5\nSkipping line 103421: expected 2 fields, saw 5\nSkipping line 106872: expected 2 fields, saw 5\nSkipping line 109363: expected 2 fields, saw 5\nSkipping line 110117: expected 2 fields, saw 5\nSkipping line 110465: expected 2 fields, saw 5\nSkipping line 113843: expected 2 fields, saw 5\nSkipping line 115634: expected 2 fields, saw 5\nSkipping line 121518: expected 2 fields, saw 5\nSkipping line 123692: expected 2 fields, saw 5\nSkipping line 124708: expected 2 fields, saw 5\nSkipping line 129608: expected 2 fields, saw 5\nSkipping line 133176: expected 2 fields, saw 5\nSkipping line 135532: expected 2 fields, saw 5\nSkipping line 138042: expected 2 fields, saw 5\nSkipping line 139485: expected 2 fields, saw 5\nSkipping line 140401: expected 2 fields, saw 5\nSkipping line 144093: expected 2 fields, saw 5\nSkipping line 149850: expected 2 fields, saw 5\nSkipping line 151831: expected 2 fields, saw 5\nSkipping line 158014: expected 2 fields, saw 5\nSkipping line 162047: expected 2 fields, saw 5\nSkipping line 164515: expected 2 fields, saw 5\nSkipping line 170313: expected 2 fields, saw 5\nSkipping line 171325: expected 2 fields, saw 5\nSkipping line 171424: expected 2 fields, saw 5\nSkipping line 175920: expected 2 fields, saw 5\nSkipping line 176210: expected 2 fields, saw 5\nSkipping line 183603: expected 2 fields, saw 5\nSkipping line 190264: expected 2 fields, saw 5\nSkipping line 191683: expected 2 fields, saw 5\nSkipping line 191988: expected 2 fields, saw 5\nSkipping line 195450: expected 2 fields, saw 5\nSkipping line 195754: expected 2 fields, saw 5\nSkipping line 197124: expected 2 fields, saw 5\nSkipping line 199263: expected 2 fields, saw 5\nSkipping line 202603: expected 2 fields, saw 5\nSkipping line 209960: expected 2 fields, saw 5\nSkipping line 213218: expected 2 fields, saw 5\nSkipping line 217060: expected 2 fields, saw 5\nSkipping line 220121: expected 2 fields, saw 5\nSkipping line 223518: expected 2 fields, saw 5\nSkipping line 226293: expected 2 fields, saw 5\nSkipping line 227035: expected 2 fields, saw 7\nSkipping line 227341: expected 2 fields, saw 5\nSkipping line 227808: expected 2 fields, saw 5\nSkipping line 228516: expected 2 fields, saw 5\nSkipping line 228733: expected 2 fields, saw 5\nSkipping line 232043: expected 2 fields, saw 5\nSkipping line 232426: expected 2 fields, saw 5\nSkipping line 234490: expected 2 fields, saw 5\nSkipping line

239626: expected 2 fields, saw 5\nSkipping line 240461: expected 2 fields, saw 5\nSkipping line 244518: expected 2 fields, saw 5\nSkipping line 245395: expected 2 fields, saw 5\nSkipping line 246168: expected 2 fields, saw 5\nSkipping line 246655: expected 2 fields, saw 5\nSkipping line 246752: expected 2 fields, saw 5\nSkipping line 247189: expected 2 fields, saw 5\nSkipping line 250276: expected 2 fields, saw 5\nSkipping line 255327: expected 2 fields, saw 5\nSkipping line 257094: expected 2 fields, saw 5\nSkipping line 264626: expected 2 fields, saw 5\nSkipping line 265028: expected 2 fields, saw 5\nSkipping line 269150: expected 2 fields, saw 5\nSkipping line 271360: expected 2 fields, saw 5\nSkipping line 273975: expected 2 fields, saw 5\nSkipping line 274742: expected 2 fields, saw 5\nSkipping line 276227: expected 2 fields, saw 5\nSkipping line 279807: expected 2 fields, saw 5\nSkipping line 283425: expected 2 fields, saw 5\nSkipping line 287468: expected 2 fields, saw 5\nSkipping line 292995: expected 2 fields, saw 5\nSkipping line 293496: expected 2 fields, saw 5\nSkipping line 293735: expected 2 fields, saw 5\nSkipping line 295060: expected 2 fields, saw 5\nSkipping line 296643: expected 2 fields, saw 5\nSkipping line 296848: expected 2 fields, saw 5\nSkipping line 308926: expected 2 fields, saw 5\nSkipping line 310360: expected 2 fields, saw 5\nSkipping line 317004: expected 2 fields, saw 5\nSkipping line 318207: expected 2 fields, saw 5\nSkipping line 331783: expected 2 fields, saw 5\nSkipping line 333864: expected 2 fields, saw 5\nSkipping line 335958: expected 2 fields, saw 5\nSkipping line 336290: expected 2 fields, saw 5\nSkipping line 343526: expected 2 fields, saw 5\nSkipping line 343857: expected 2 fields, saw 5\nSkipping line 344059: expected 2 fields, saw 5\nSkipping line 348691: expected 2 fields, saw 5\nSkipping line 353446: expected 2 fields, saw 5\nSkipping line 357073: expected 2 fields, saw 5\nSkipping line 359753: expected 2 fields, saw 5\nSkipping line 359974: expected 2 fields, saw 5\nSkipping line 366534: expected 2 fields, saw 5\nSkipping line 369514: expected 2 fields, saw 5\nSkipping line 377759: expected 2 fields, saw 5\nSkipping line 379327: expected 2 fields, saw 5\nSkipping line 380769: expected 2 fields, saw 5\nSkipping line 381073: expected 2 fields, saw 5\nSkipping line 381489: expected 2 fields, saw 5\nSkipping line 386304: expected 2 fields, saw 5\nSkipping line 387635: expected 2 fields, saw 5\nSkipping line 389613: expected 2 fields, saw 5\nSkipping line 392604: expected 2 fields, saw 5\nSkipping line 393184: expected 2 fields, saw 5\nSkipping line 395530: expected 2 fields, saw 5\nSkipping line 396939: expected 2 fields, saw 5\nSkipping line 397385: expected 2 fields, saw 5\nSkipping line 397509: expected 2 fields, saw 5\nSkipping line 402902: expected 2 fields, saw 5\nSkipping line 405187: expected 2 fields, saw 5\nSkipping line 408412: expected 2 fields, saw 5\nSkipping line 419423: expected 2 fields, saw 5\nSkipping line 420962: expected 2 fields, saw 5\nSkipping line 425965: expected 2 fields, saw 5\nSkipping line 427496: expected 2 fields, saw 5\nSkipping line 438881: expected 2 fields, saw 5\nSkipping line 439776: expected 2 fields, saw 5\nSkipping line 440345: expected 2 fields, saw 5\nSkipping line 445507: expected 2 fields, saw 5\nSkipping line 445548: expected 2 fields, saw 5\nSkipping line 447184: expected 2 fields, saw 5\nSkipping line 448603: expected 2 fields, saw 5\nSkipping line 451732: expected 2 fields, saw 5\nSkipping line 458249: expected 2 fields, saw 5\nSkipping line 460274: expected 2 fields, saw 5\nSkipping line 467630: expected 2 fields, saw 5\nSkipping line 473961: expected 2 fields, saw 5\nSkipping line 476281: expected 2 fields, saw 5\nSkipping line 478010: expected 2 fields, saw 5\nSkipping line 478322: expected 2 fields, saw 5\nSkipping line 479999: expected 2 fields, saw 5\nSkipping line 480898: expected 2 fields, saw 5\nSkipping line 481688: expected 2 fields, saw 5\nSkipping line 485193: expected 2 fields, saw 5\nSkipping line 485519: expected 2 fields, saw 5\nSkipping line 486000: expected 2 fields, saw 5\nSkipping line 489063: expected 2 fields, saw 5\nSkipping line 494525: expected 2 fields, saw 5\nSkipping line 495009: expected 2 fields, saw 5\nSkipping line 501954: expected 2 fields, saw 5\nSkipping line 508035: expected 2 fields, saw 5\nSkipping line 508828: expected 2 fields, saw 5\nSkipping line 509833: expected 2 fields, saw 5\nSkipping line 51041

```
0: expected 2 fields, saw 5\nSkipping line 518229: expected 2 fields, saw
5\nSkipping line 520302: expected 2 fields, saw 5\nSkipping line 520340: e
xpected 2 fields, saw 5\n'
b'Skipping line 525174: expected 2 fields, saw 5\nSkipping line 526251: ex
pected 2 fields, saw 5\nSkipping line 529611: expected 2 fields, saw 5\nSk
ipping line 531398: expected 2 fields, saw 5\nSkipping line 534146: expect
ed 2 fields, saw 5\nSkipping line 544954: expected 2 fields, saw 5\nSkippi
ng line 553002: expected 2 fields, saw 5\nSkipping line 553883: expected 2
fields, saw 5\nSkipping line 553887: expected 2 fields, saw 5\nSkipping li
ne 553915: expected 2 fields, saw 5\nSkipping line 554172: expected 2 fiel
ds, saw 5\nSkipping line 563534: expected 2 fields, saw 5\nSkipping line 5
65191: expected 2 fields, saw 5\nSkipping line 574108: expected 2 fields,
saw 5\nSkipping line 574412: expected 2 fields, saw 5\nSkipping line 57598
5: expected 2 fields, saw 5\nSkipping line 580091: expected 2 fields, saw
5\nSkipping line 582682: expected 2 fields, saw 5\nSkipping line 585885: e
xpected 2 fields, saw 5\nSkipping line 590171: expected 2 fields, saw 5\nS
kipping line 591924: expected 2 fields, saw 5\nSkipping line 592515: expec
ted 2 fields, saw 5\nSkipping line 593888: expected 2 fields, saw 5\nSkipp
ing line 596245: expected 2 fields, saw 5\nSkipping line 607344: expected
2 fields, saw 5\nSkipping line 607633: expected 2 fields, saw 5\nSkipping
line 610939: expected 2 fields, saw 5\nSkipping line 613638: expected 2 fi
elds, saw 5\nSkipping line 615643: expected 2 fields, saw 5\nSkipping line
615901: expected 2 fields, saw 5\nSkipping line 617389: expected 2 fields,
saw 5\nSkipping line 634641: expected 2 fields, saw 5\nSkipping line 63575
5: expected 2 fields, saw 5\nSkipping line 646243: expected 2 fields, saw
5\nSkipping line 647165: expected 2 fields, saw 5\nSkipping line 648610: e
xpected 2 fields, saw 5\nSkipping line 648772: expected 2 fields, saw 5\nS
kipping line 651833: expected 2 fields, saw 5\nSkipping line 653663: expec
ted 2 fields, saw 5\nSkipping line 656233: expected 2 fields, saw 5\nSkipp
ing line 656694: expected 2 fields, saw 5\nSkipping line 659783: expected
2 fields, saw 5\nSkipping line 660478: expected 2 fields, saw 5\nSkipping
line 661133: expected 2 fields, saw 5\nSkipping line 661736: expected 2 fi
elds, saw 5\nSkipping line 669827: expected 2 fields, saw 5\n'
```

In [3]:

```
data.head()
```

Out[3]:

	password	strength
0	kzde5577	1
1	kino3434	1
2	visi7k1yr	1
3	megzy123	1
4	lamborghini1	1

In [4]:

```
data.shape
```

Out[4]:

```
(669640, 2)
```

In [5]:

```
data.strength.unique()
```

Out[5]:

```
array([1, 2, 0], dtype=int64)
```

In [6]:

```
data.isnull().sum()
```

Out[6]:

```
password    1  
strength    0  
dtype: int64
```

In [7]:

```
data[data['password'].isnull()]
```

Out[7]:

	password	strength
367579	NaN	0

In [8]:

```
data.dropna(inplace=True)
```

In [9]:

```
data.isnull().sum()
```

Out[9]:

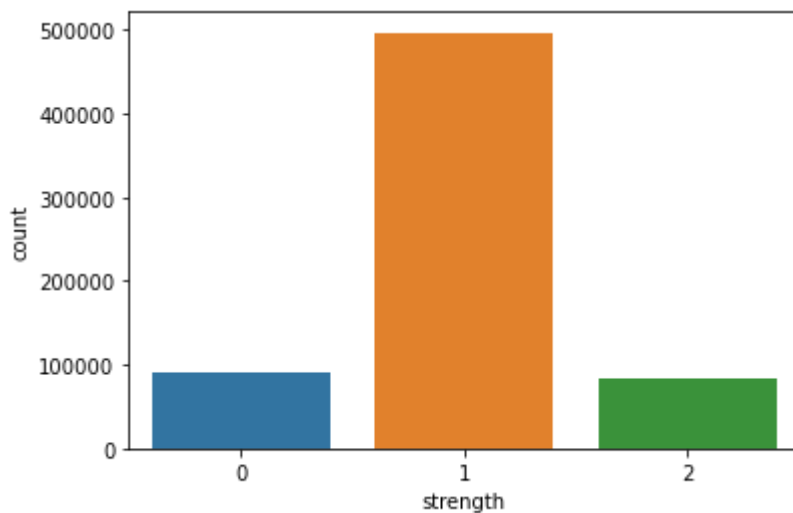
```
password    0  
strength    0  
dtype: int64
```

In [10]:

```
sns.countplot(data['strength'])
```

Out[10]:

<matplotlib.axes._subplots.AxesSubplot at 0x1e0dadd9a60>



In [11]:

```
pass_tup=np.array(data)
```

In [12]:

```
pass_tup
```

Out[12]:

```
array([[ 'kzde5577', 1],  
       [ 'kino3434', 1],  
       [ 'visi7k1yr', 1],  
       ...,  
       [ '184520socram', 1],  
       [ 'marken22a', 1],  
       [ 'fxx4pw4g', 1]], dtype=object)
```

In [13]:

```
import random  
random.shuffle(pass_tup)  
x=[labels[0] for labels in pass_tup]  
y=[labels[1] for labels in pass_tup]
```

In [14]:

```
def word_divider(words):  
    char=[]  
    for i in words:  
        char.append(i)  
    return char
```

In [15]:

```
word_divider('nikhila1005')
```

Out[15]:

```
['n', 'i', 'k', 'h', 'i', 'l', 'a', '1', '0', '0', '5']
```

In [16]:

```
from sklearn.feature_extraction.text import TfidfVectorizer  
vectorizer=TfidfVectorizer(tokenizer=word_divider)  
X=vectorizer.fit_transform(x)  
X.shape
```

Out[16]:

```
(669639, 129)
```

In []:

In [17]:

```
first_document_vector=X[0]  
first_document_vector
```

Out[17]:

```
<1x129 sparse matrix of type '<class 'numpy.float64'>'  
    with 6 stored elements in Compressed Sparse Row format>
```

In []:

In [18]:

```
df=pd.DataFrame(first_document_vector.T.todense(),index=vectorizer.get_feature_names(),
columns=['TF-IDF'])
df.sort_values(by=['TF-IDF'],ascending=False)
```

Out[18]:

	TF-IDF
7	0.591221
5	0.567444
z	0.335731
k	0.291896
d	0.285657
...	...
>	0.000000
=	0.000000
<	0.000000
;	0.000000
™	0.000000

129 rows × 1 columns

In [19]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2)
X_train.shape
```

Out[19]:

(535711, 129)

In [20]:

```
from sklearn.ensemble import RandomForestClassifier
rn=RandomForestClassifier(n_estimators=10,criterion='entropy')
rn.fit(X_train,y_train)
```

Out[20]:

RandomForestClassifier(criterion='entropy', n_estimators=10)

In [21]:

```

y_pred=rn.predict(X_test)
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
cm=confusion_matrix(y_test,y_pred)
print(cm)
print(accuracy_score(y_test,y_pred))

```

```

[[16755  1276    0]
 [  478 98517   198]
 [    25   763 15916]]
0.9795412460426498

```

In [22]:

```
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.97	0.93	0.95	18031
1	0.98	0.99	0.99	99193
2	0.99	0.95	0.97	16704
accuracy			0.98	133928
macro avg	0.98	0.96	0.97	133928
weighted avg	0.98	0.98	0.98	133928

In []: