About Dataset

10000 Most Common Passwords

If your password is on this list of 10,000 most common passwords, you need a new password. A hacker can use or generate files like this, which may readily be compiled from breaches of sites such as Ashley Madison. Usually, passwords are not tried one-by-one against a system's secure server online; instead, a hacker might manage to gain access to a shadowed password file protected by a one-way encryption algorithm, then test each entry in a file like this to see whether it encrypted form matches what the server has on record. The passwords may then be tried against any account online that can be linked to the first, to test for passwords reused on other sites.

Acknowledgements

The dataset was procured by SecLists. SecLists is the security tester's companion. It's a collection of multiple types of lists used during security assessments, collected in one place. List types include usernames, passwords, URLs, sensitive data patterns, fuzzing payloads, web shells, and many more. The goal is to enable a security tester to pull this repository onto a new testing box and have access to every type of list that may be needed.

Analysis

1. There are no null values
2. There are 2 duplicate passwords

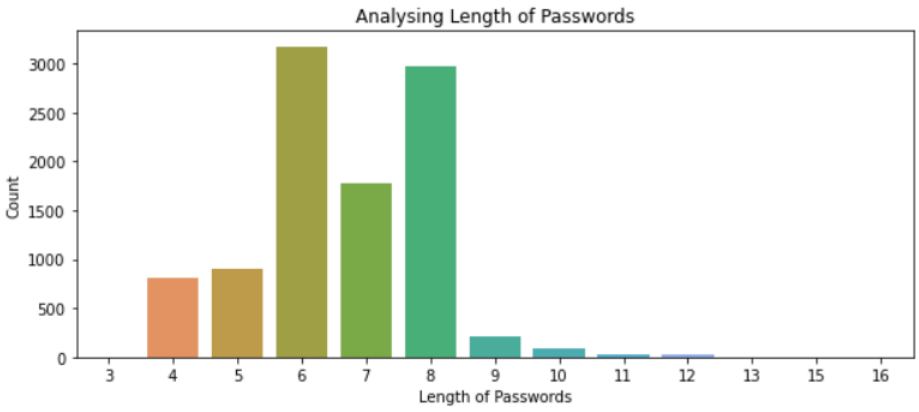
	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	num_syllables
484	abcdef	6	6	0	0	6	0	2	1
3507	easy	4	4	0	0	4	0	2	1

Below, we see that there were only 2 passwords that were repeated more than once:

- "easy"
- "abcdef"

3. Results of the describe() function;
    a. There are 10000 datapoints in the dataset
    b. Of the leaked password:
        i. mean length ~= 6.65
        ii. mean num_chars ~= 5.03
        iii. mean num_digits ~= 1.62
        iv. mean num_upper ~= 0.03
        v. mean num_lower ~= 5.005
        vi. mean num_special ~= 0.003
        vii. mean num_vowels ~= 1.81
        viii. mean num_syllables ~= 1.61
    c. minimum password length = 3, an maximum password length = 16
    d. On Average there are 5 characters, 2 digits, 3 lowercase, 1 vowel in the leaked password
    e. There are lengthy passwords with upto 16 characters or digits but still got leaked

4. Majority of the password are of length 6 or 8. There are passwords that are even of length 3 and large as 13/15/16



5. top 10 longest passwords

password	length
PolniyPizdec0211	16
Mailcreated5240	15
Sojdlg123aljg	13
wrongpassword	13
123456qwerty	12
123456789qwe	12
masterbating	12
123qweasdzxc	12
q1w2e3r4t5y6	12
ghjcnjgfhjkm	12

6. top 10 shortest passwords

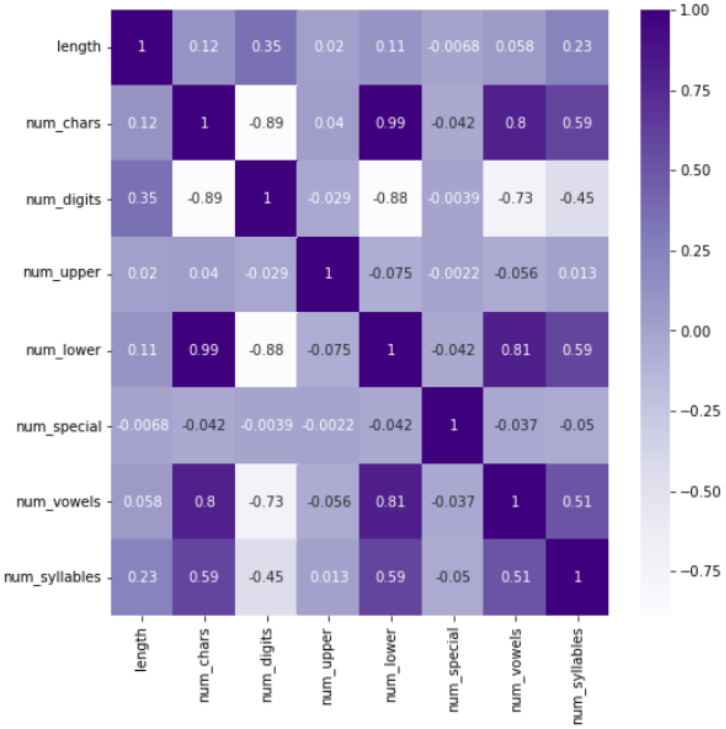
password	length
sex	3
mad	3
123	3
shao	4
1980	4
hung	4
beta	4
junk	4
qiao	4
1949	4

7. top 10 passwords with the most syllables;

password	num_syllables
cccccccc	6
llllllll	6
mmmmmmmm	6
hhhhhhhh	6
mmmmmmm	5
fktrcfylhf	5
sonyericsson	5
qwerasdfzxcv	5
rjycnfynby	5
qwertyqwerty	5

8. There is high correlation among;
    a. num_chars and num_lower
    b. num_vowels and num_lower
    c. num_vowels and num_chars

9. There is low correlation among;
    a. num_lower and num_digits
    b. num_vowels and num_digits
    c. num_chars and num_digits
    d.



10. Most Secure password is considered at least 8 characters long.

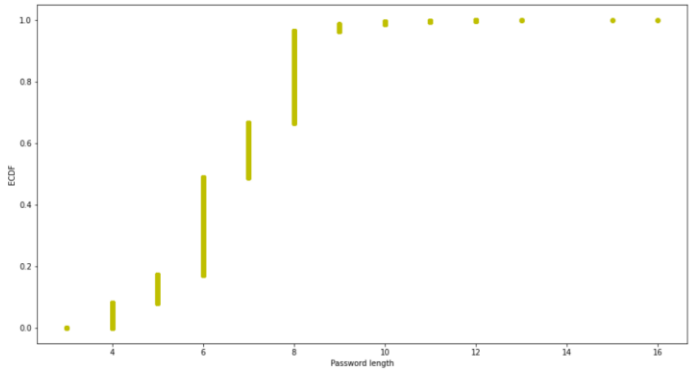
```
Percentage of Secured Password : 33.37%
Percentage of Insecured Password : 66.63%
```

11. As we can see, more than 66% of the most popular passwords are not long enough. Now let's check those with symbols, numbers and capital letters. Total 24 records are there.

13. As we can see in the DataFrame, only 0.24% of the most popular password would be secure before checking criteria number 3, which is not using obvious words. We can see that among those passwords, some of them are popular words with added just one integer (most with "1" at the end, and "Passw0rd" being obvious one to drop, since it is easy to hack). Let's clear these.

```
Usuckballz1
Soso123aljg
Mailcreated5240
Passw0rd
8J4yE3Uz
Password1
Turkey50
1Passwor
Sojdlg123aljg
Passwor1
PolniyPizdec0211
7uGd5HIp2J
vSjasnel12
Michael1
Good123654
Kordell1
Misfit99
Letmein1
Password123
Trustno1
Welcome1
5Wr2i7H8
Jordan23
Mustang1
```

14. The number of secure passwords out of 10000 most popular is 13. This easily explains why most people's accounts are easily hacked.



15. We have created a column that tells the strength of the password on the basis;
    a. Length minimum of 8 characters (+4 Strength)
    b. Use of upper case characters (+1 Strength)
    c. Use of lower case characters (+1 Strength)
    d. Use of digits (+1 Strength)
    e. Use of Special Characters (+3 Strength)
    f. Use of Spaces (Invalid Password)

```
1 df.head()
```

	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	num_syllables	Remarks
0	123456	6	0	6	0	0	0	0	1	Poor
1	password	8	8	0	0	8	0	2	2	Average
2	12345678	8	0	8	0	0	0	0	1	Average
3	qwerty	6	6	0	0	6	0	1	3	Poor
4	123456789	9	0	9	0	0	0	0	1	Average

```
1 df.Remarks.value_counts()
```

```
Poor         6663
Average      3309
Excellent      28
Name: Remarks, dtype: int64
```

Conclusion:

Almost all of the most popular passwords are not secure. As long as people keep using easy to break passwords, they will be exposed to being hacked.