In [1]:
```
from bs4 import BeautifulSoup
import requests
import re
import pandas as pd
```

In [2]:
```
url = "http://genki.japantimes.co.jp/self/genki-kanji-list-linked-to-wwkanji"
page = requests.get(url)
soup = BeautifulSoup(page.content, "html.parser")
```

```
Some characters could not be decoded, and were replaced with REPLACEMENT CHARACT
ER.
```

In [3]:
```
things = soup.find_all("td")
lesson = []
kanji = []
curr = 0
for i in range(1,len(things)):
    l = re.search(r'Lesson (.*):', things[i].getText())
    if not l == None:
        curr = l[1]
    else:
        if not things[i].getText() == '\xa0':
            kanji.append(things[i].getText())
            lesson.append(curr)
```

In [4]:
```
lesson
```

Out[4]:
```
['3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '3',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '4',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '5',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '6',
 '7',
 '7',
 '7',
 '7',
 '7',
```

In [5]:
```
df = pd.DataFrame({'kanji':kanji, 'Genki_Level': lesson})
```

In [6]:
```
df
```

Out[6]:

	kanji	Genki_Level
0	一	3
1	二	3
2	三	3
3	四	3
4	五	3
...	...	...
312	調	23
313	査	23
314	果	23
315	感	23
316	答	23

317 rows × 2 columns

In [7]:
```
df.to_csv("genki.csv", index=False)
```

In [ ]:
```

```