NAME = "Xudong Guo"

Description: This file is used to clean up extra links not needed.

Input: 'combined_with_frequency.csv' Output: 'cleaned_link.csv'

```python
import pandas as pd
import ast
```

```python
file = pd.read_csv('combined_with_frequency.csv')
```

In [6]: `file`

Out[6]:

	kanji	strokes	frequency	grade	jlpt	parts	radicals	on_readings	kun_readings	on_read
0	亜	7.0	1509.0	junior high	N1	{'一': 'http:////jisho.org/search/%E4%B8%80%20%...	{'二': 'two'}	{'ア': 'http:////jisho.org/search/%E4%BA%9C%20%...	{'つ.ぐ': 'http:////jisho.org/search/%E4%BA%9C%2...	['亜 (indica
1	哀	9.0	1715.0	junior high	N1	{'亠': 'http:////jisho.org/search/%E4%BA%A0%20%...	{'口': 'mouth, opening'}	{'アイ': 'http:////jisho.org/search/%E5%93%80%20...	{'あわ.れ': 'http:////jisho.org/search/%E5%93%80%...	[' condoler
2	挨	10.0	2258.0	junior high	NaN	{'ム': 'http:////jisho.org/search/%E5%8E%B6%20%...	{'手 (扌龵)': 'hand'}	{'アイ': 'http:////jisho.org/search/%E6%8C%A8%20...	{'ひら.く': 'http:////jisho.org/search/%E6%8C%A8%...	[' g
3	愛	13.0	640.0	grade 4	N3	{'冖': 'http:////jisho.org/search/%E5%86%96%20%...	{'心 (忄, 㣺)': 'heart'}	{'アイ': 'http:////jisho.org/search/%E6%84%9B%20...	{'いと.しい': 'http:////jisho.org/search/%E6%84%9B...	affection
4	曖	17.0	NaN	junior high	NaN	{'冖': 'http:////jisho.org/search/%E5%86%96%20%...	{'日': 'sun, day'}	{'アイ': 'http:////jisho.org/search/%E6%9B%96%20...	{'くら.い': 'http:////jisho.org/search/%E6%9B%96%...	['曖昧 【 ar
...	...	...	...	...	...	...	...	...	...	
2131	脇	10.0	1806.0	junior high	NaN	{'力': 'http:////jisho.org/search/%E5%8A%9B%20%...	{'肉 (月)': 'meat'}	{'キョウ': 'http:////jisho.org/search/%E8%84%87%2...	{'わき': 'http:////jisho.org/search/%E8%84%87%20...	['脇侍 image (e
2132	惑	12.0	777.0	junior high	N1	{'口': 'http:////jisho.org/search/%E5%8F%A3%20%...	{'心 (忄, 㣺)': 'heart'}	{'ワク': 'http:////jisho.org/search/%E6%83%91%20...	{'まど.う': 'http:////jisho.org/search/%E6%83%91%...	[' planet', '
2133	枠	8.0	922.0	junior high	N1	{'十': 'http:////jisho.org/search/%E5%8D%81%20%...	{'木': 'tree'}	NaN	{'わく': 'http:////jisho.org/search/%E6%9E%A0%20...	['栾 frame
2134	湾	12.0	545.0	junior high	N2	{'亠': 'http:////jisho.org/search/%E4%BA%A0%20%...	{'水 (氵, 氺)': 'water'}	{'ワン': 'http:////jisho.org/search/%E6%B9%BE%20...	{'いりえ': 'http:////jisho.org/search/%E6%B9%BE%2...	['湾 inlet', '
2135	腕	12.0	1163.0	junior high	N2	{'卩': 'http:////jisho.org/search/%E5%8D%A9%20%...	{'肉 (月)': 'meat'}	{'ワン': 'http:////jisho.org/search/%E8%85%95%20...	{'うで': 'http:////jisho.org/search/%E8%85%95%20...	['腕 phys

2136 rows × 25 columns

In [7]: `d = ast.literal_eval(file['parts'][0])`

In [8]:
```python
newcol = [None] * len(file['parts'])
for i in range(0,len(file['parts'])):
    newcol[i] = list(ast.literal_eval(file['parts'][i]).keys())
```

In [9]:
```python
len(newcol)
```

Out[9]: 2136

In [10]:
```python
file['parts'] = newcol
```

In [11]: `file`

Out[11]:

	kanji	strokes	frequency	grade	jlpt	parts	radicals	on_readings	kun_readings	on_readings_com
0	亜	7.0	1509.0	junior high	N1	[一, ｜, 口]	{'二': 'two'}	{'ア': 'http:////jisho.org/search/%E4%BA%9C%20%...'	{'つ.ぐ': 'http:////jisho.org/search/%E4%BA%9C%2...'	['亜 【ア】 su (indicating a low
1	哀	9.0	1715.0	junior high	N1	[亠, 口, 衣]	{'口': 'mouth, opening'}	{'アイ': 'http:////jisho.org/search/%E5%93%80%20%...'	{'あわ.れ': 'http:////jisho.org/search/%E5%93%80%...'	['哀悼 【アイ condolence, regret
2	挨	10.0	2258.0	junior high	NaN	[厶, 扎, 矢, 乞]	{'手 (扌 㐄)': 'hand'}	{'アイ': 'http:////jisho.org/search/%E6%8C%A8%20%...'	{'ひら.く': 'http:////jisho.org/search/%E6%8C%A8%...'	['挨拶 【アイ greeting, gr salu
3	愛	13.0	640.0	grade 4	N3	[冖, 夂, 心, 爪]	{'心 (忄 㣺)': 'heart'}	{'アイ': 'http:////jisho.org/search/%E6%84%9B%20%...'	{'いと.しい': 'http:////jisho.org/search/%E6%84%9B%...'	['愛 【アイ affection, care, atta
4	曖	17.0	NaN	junior high	NaN	[冖, 夂, 心, 日, 爪]	{'日': 'sun, day'}	{'アイ': 'http:////jisho.org/search/%E6%9B%96%20%...'	{'くら.い': 'http:////jisho.org/search/%E6%9B%96%...'	['曖昧 【アイマイ】 ambiguous,
...	...	...	...	...	...	...	...	...	...	...
2131	脇	10.0	1806.0	junior high	NaN	[力, 月]	{'肉 (月)': 'meat'}	{'キョウ': 'http:////jisho.org/search/%E8%84%87%2...'	{'わき': 'http:////jisho.org/search/%E8%84%87%20...'	['脇侍 【ワキジ】 image (e.g. in a Bu
2132	惑	12.0	777.0	junior high	N1	[口, 心, 戈]	{'心 (忄 㣺)': 'heart'}	{'ワク': 'http:////jisho.org/search/%E6%83%91%20...'	{'まど.う': 'http:////jisho.org/search/%E6%83%91%...'	['惑星 【ワ〈 planet', '惑星科学 イカガク】
2133	枠	8.0	922.0	junior high	N1	[十, 九, 木]	{'木': 'tree'}	NaN	{'わく': 'http:////jisho.org/search/%E6%9E%A0%20...'	['枠 【わく】 framework, bord
2134	湾	12.0	545.0	junior high	N2	[亠, 弓, 汁]	{'水 (氵 氺)': 'water'}	{'ワン': 'http:////jisho.org/search/%E6%B9%BE%20...'	{'いりえ': 'http:////jisho.org/search/%E6%B9%BE%2...'	['湾 【ワン】 b inlet', '湾岸 【ワ〉
2135	腕	12.0	1163.0	junior high	N2	[卩, 夕, 宀, 月]	{'肉 (月)': 'meat'}	{'ワン': 'http:////jisho.org/search/%E8%85%95%20...'	{'うで': 'http:////jisho.org/search/%E8%85%95%20...'	['腕力 【ワン physical streng

2136 rows × 25 columns

In [12]:
```python
newcol = [None] * len(file['on_readings'])
for i in range(0,len(file['on_readings'])):
    try:
        newcol[i] = list(ast.literal_eval(file['on_readings'][i]).keys())
    except:
        continue
len(newcol)
```

Out[12]: 2136

In [13]:
```python
file['on_readings']  = newcol
```

In [12]:
```python
newcol = [None] * len(file['on_readings'])
for i in range(0,len(file['on_readings'])):
    try:
```

In [14]: `file`

Out[14]:

	kanji	strokes	frequency	grade	jlpt	parts	radicals	on_readings	kun_readings	on_readings_compo
0	亜	7.0	1509.0	junior high	N1	[一, |, 口]	{'二': 'two'}	[ア]	{'つ.ぐ': 'http:////jisho.org/search/%E4%BA%9C%2...' }	['亜 【ア】 sub-(indicating a low oxi
1	哀	9.0	1715.0	junior high	N1	[亠, 口, 衣]	{'口': 'mouth, opening'}	[アイ]	{'あわ.れ': 'http:////jisho.org/search/%E5%93%80%...' }	['哀悼 【アイ⊦ condolence, regret, tr
2	挨	10.0	2258.0	junior high	NaN	[厶, 扎, 矢, 乞]	{'手 (扌龵)': 'hand'}	[アイ]	{'ひら.く': 'http:////jisho.org/search/%E6%8C%A8%...' }	['挨拶 【アイザ greeting, gree salutat
3	愛	13.0	640.0	grade 4	N3	[冖, 夂, 心, 爪]	{'心 (忄, 㣺)': 'heart'}	[アイ]	{'いと.しい': 'http:////jisho.org/search/%E6%84%9B%...' }	['愛 【アイ】 affection, care, attach
4	曖	17.0	NaN	junior high	NaN	[冖, 夂, 心, 日, 爪]	{'日': 'sun, day'}	[アイ]	{'くら.い': 'http:////jisho.org/search/%E6%9B%96%...' }	['曖昧 【アイマイ】 v ambiguous, un fu
...	...	...	...	...	...	...	...	...	...	...
2131	脇	10.0	1806.0	junior high	NaN	[力, 月]	{'肉 (月)': 'meat'}	[キョウ]	{'わき': 'http:////jisho.org/search/%E8%84%87%20...' }	['脇侍 【ワキジ】 fla image (e.g. in a Budd
2132	惑	12.0	777.0	junior high	N1	[口, 心, 戈]	{'心 (忄, 㣺)': 'heart'}	[ワク]	{'まど.う': 'http:////jisho.org/search/%E6%83%91%...' }	['惑星 【ワクセ planet', '惑星科学 【ワ イカガク】 pl
2133	枠	8.0	922.0	junior high	N1	[十, 九, 木]	{'木': 'tree'}	None	{'わく': 'http:////jisho.org/search/%E6%9E%A0%20...' }	['枠 【わく】 f framework, border
2134	湾	12.0	545.0	junior high	N2	[亠, 弓, 汁]	{'水 (氵, 氺)': 'water'}	[ワン]	{'いりえ': 'http:////jisho.org/search/%E6%B9%BE%2...' }	['湾 【ワン】 bay inlet', '湾岸 【ワンガ
2135	腕	12.0	1163.0	junior high	N2	[卩, 夕, 宀, 月]	{'肉 (月)': 'meat'}	[ワン]	{'うで': 'http:////jisho.org/search/%E8%85%95%20...' }	['腕力 【ワンリョ physical strength, str

2136 rows × 25 columns

In [15]:
```python
newcol = [None] * len(file['kun_readings'])
for i in range(0,len(file['kun_readings'])):
    try:
        newcol[i] = list(ast.literal_eval(file['kun_readings'][i]).keys())
    except:
        continue
len(newcol)
```

Out[15]: 2136

In [16]:
```python
file['kun_readings'] = newcol
```

In [17]:
```python
file
```

Out[17]:

	kanji	strokes	frequency	grade	jlpt	parts	radicals	on_readings	kun_readings	on_readings_compou
0	亜	7.0	1509.0	junior high	N1	[一, 丨, 口]	{'二': 'two'}	[ア]	[つ.ぐ]	['亜 【ア】 sub-, (indicating a low oxid
1	哀	9.0	1715.0	junior high	N1	[亠, 口, 衣]	{'口': 'mouth, opening'}	[アイ]	[あわ.れ, あわ.れむ, かな.しい]	['哀悼 【アイト condolence, regret, trib !
2	挨	10.0	2258.0	junior high	NaN	[厶, 扎, 矢, 乞]	{'手 (扌 龵)': 'hand'}	[アイ]	[ひら.く]	['挨拶 【アイサ greeting, greeti salutatic
3	愛	13.0	640.0	grade 4	N3	[冖, 夂, 心, 爪]	{'心 (忄 㣺)': 'heart'}	[アイ]	[いと.しい, か な.しい, め.で る, お.しむ, ま な]	['愛 【アイ】 affection, care, attachn
4	曖	17.0	NaN	junior high	NaN	[冖, 夂, 心, 日, 爪]	{'日': 'sun, day'}	[アイ]	[くら.い]	['曖昧 【アイマイ】 va ambiguous, und fu
...	...	...	...	...	...	...	...	...	...	
2131	脇	10.0	1806.0	junior high	NaN	[力, 月]	{'肉 (月)': 'meat'}	[キョウ]	[わき, わけ]	['脇侍 【ワキジ】 flan image (e.g. in a Buddh
2132	惑	12.0	777.0	junior high	N1	[口, 心, 戈]	{'心 (忄 㣺)': 'heart'}	[ワク]	[まど.う]	['惑星 【ワクセ planet', '惑星科学 【ワ イカガク】 pla
2133	枠	8.0	922.0	junior high	N1	[十, 九, 木]	{'木': 'tree'}	None	[わく]	['枠 【わく】 fra framework, border, li
2134	湾	12.0	545.0	junior high	N2	[亠, 弓, 氵]	{'水 (氵 氺)': 'water'}	[ワン]	[いりえ]	['湾 【ワン】 bay, inlet', '湾岸 【ワンガ !
2135	腕	12.0	1163.0	junior high	N2	[卩, 夕, 宀, 月]	{'肉 (月)': 'meat'}	[ワン]	[うで]	['腕力 【ワンリョ physical strength, b stre

2136 rows × 25 columns

In [18]:
```python
file.to_csv('cleaned_link.csv', index=False)
```

In [ ]: