

Part 1

Counting

```
from collections import Counter
count = Counter([ url.split("/")[-1] for url in urls ])
for i,v in enumerate(count.most_common(3)):
    print("{} {}".format(v[0],v[1]))
```

```
In [77]: 1 from collections import Counter
          2 count = Counter([ url.split("/")[-1] for url in urls ])
          3 for i,v in enumerate(count.most_common(3)):
          4     print("{} {}".format(v[0],v[1]))

a.txt 3
c.jpg 2
b.txt 2
```

Integration

```
def anonymous(x):
    return i%3 ==0 or i%5 ==0 # 只回傳 整除3 整除4 的list
def integrate(fun, start, end):
    step = 1
    intercept =start
    area = 0
    while intercept < end:
        if fun(intercept) :
            area += intercept
            intercept += step
    return area
print(integrate(anonymous, 0, 1000))
```

Part 2 爬蟲 ptt

```
import requests
from bs4 import BeautifulSoup
import lxml
import random
import collections
# twice 1-308
page = random.randint(1,308)
url = "https://www.ptt.cc/bbs/TWICE/index{}.html".format(str(page))
base = "https://www.ptt.cc"
r = requests.get(url)
soup = BeautifulSoup(r.text,"lxml")
divs = soup.find_all(class_ = "r-ent")
dicts = []
for div in divs:
```

```

content = div.find(class_="title").find("a")["href"]
req = requests.get(base+content)
if req.status_code == 200 :
    result = collections.defaultdict()
    soup2 = BeautifulSoup(req.text,"lxml")
    tag = soup2.find_all(class_="article-meta-value")
    if tag : #防止空集合
        result["author"] = tag[0].text
        result["clf"] = tag[1].text

```

```

[defaultdict(None, {'author': 'kk9430167 (NJMSJMDCT)', 'clf': 'TWICE', 'title': '[LIVE] 170530 MBC FM4U Tei的夢想電台', 'date': 'Tue May 30 20:02:33 2017'}), defaultdict(None, {'author': 'venus40123 (嬌妍)', 'clf': 'TWICE', 'title': '[分享] 第一次買新專輯之被本命鎖定真好', 'date': 'Tue May 30 22:01:57 2017'}), defaultdict(None, {'author': 'culiu327 ()', 'clf': 'TWICE', 'title': '[社群] 170530 TWICE IG/twitter更新', 'date': 'Tue May 30 22:16:00 2017'}), defaultdict(None, {'author': 'sichen1029 (chen)', 'clf': 'TWICE', 'title': '[ONCE] sichen1029', 'date': 'Wed May 31 02:01:52 2017'}), defaultdict(None, {'author': 'PMTzaiyo (哉悠)', 'clf': 'TWICE', 'title': '[ONCE] PMTzaiyo', 'date': 'Wed May 31 07:00:46 2017'}), defaultdict(None, {'author': 'elvisssu (won)', 'clf': 'TWICE', 'title': '[分享] 遲來的開箱文之終於遇到忙內了!!!!!!!', 'date': 'Wed May 31 11:35:08 2017'}), defaultdict(None, {'author': 'kk9430167 (NJMSJMDCT)', 'clf': 'TWICE', 'title': '[LIVE] 170531 KBS1,MBC,SBS"U Have a Dream"', 'date': 'Wed May 31 16:30:27 2017'}), defaultdict(None, {'author': 'x80680613 (龍神天劍)', 'clf': 'TWICE', 'title': '[LIVE] 170531 MBC MUSIC Show Champion', 'date': 'Wed May 31 16:47:50 2017'}), defaultdict(None, {'author': 'okoko2232000 (社會邊緣人)', 'clf': 'TWICE', 'title': '[分享] 遲來的開箱文', 'date': 'Wed May 31 16:57:10 2017'}), defaultdict(None, {'author': 'kk9430167 (NJMSJMDCT)', 'clf': 'TWICE', 'title': '[影音] 170531 TWICE TV5 EP.05', 'date': 'Wed May 31 17:00:04 2017'}), defaultdict(None, {'author': 'x80680613 (龍神天劍)', 'clf': 'TWICE', 'title': '[影音] 170531 MBC MUSIC Show Champion', 'date': 'Wed May 31 19:01:56 2017'}), defaultdict(None, {'author': 'ohnet3214 (藤咲秋內)', 'clf': 'TWICE', 'title': '[影音] 170531 KBS1,MBC,SBS,UHD U Have a Dream', 'date': 'Wed May 31 19:13:37 2017'}), defaultdict(None, {'author': 'tycousin (Nong)', 'clf': 'TWICE', 'title': '[LIVE] 日本デビュー直前特番 Part.3', 'date': 'Wed May 31 19:33:57 2017'}), defaultdict(None, {'author': 'hsh131 (shib)', 'clf': 'TWICE', 'title': '[LIVE] 170531 TWICE TV5 EP.05', 'date': 'Wed May 31 19:33:57 2017'})]

```

```

result["title"] = tag[2].text
result["date"] = tag[3].text
dicts.append(result)

```

```
print(dicts)
```

我看板內文未完成