

```

1 from bs4 import BeautifulSoup #pip install bs4
2 import requests
3 import re
4 import pandas as pd
5 import datetime
6
7 category = ['Finance', 'Securities', 'Industry', 'Venture_company', 'Real_estate', '
Global_economy', 'Living_economy', 'General_economy']
8 sections = [259, 258, 261, 771, 260, 262, 310, 263]
9 df_titles = pd.DataFrame()
10 re_title = re.compile('[^가-힣|a-z|A-Z]')
11 headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (
KHTML, like Gecko) Chrome/120.0.0.0 Safari/537.36'}
12
13 for i, section in enumerate(sections):
14     url = 'https://news.naver.com/breakingnews/section/101/{}'.format(section)
15     resp = requests.get(url, headers=headers)
16     soup = BeautifulSoup(resp.text, 'html.parser')
17     title_tags = soup.select('.sa_text_strong')
18     titles = []
19     for title_tag in title_tags:
20         titles.append(re_title.sub(' ', title_tag.text))
21     df_section_titles = pd.DataFrame(titles, columns=['titles'])
22     df_section_titles['category'] = category[i]
23     df_titles = pd.concat([df_titles, df_section_titles], axis='rows',
24                           ignore_index=True)
25 print(df_titles.head())
26 df_titles.info()
27 print(df_titles['category'].value_counts())
28 df_titles.to_csv('./crawling_data/naver_test_headline_news_{}.csv'.format(
29     datetime.datetime.now().strftime('%Y%m%d')
30 ), index=False)

```