

1. 코스피 / 코스닥 종목 가져와 출력하기

- code sample

```
def get_item_kospi(self):
    df_stockMkt = pd.read_excel(os.path.join(BASE_DIR,
                                              '상장법인목록_stockMkt.xlsx'))

    print('코스피 종목 수: ', len(df_stockMkt))
    print(df_stockMkt[['회사명', '종목코드']])
```

```
코스피 종목 수: 824
   회사명   종목코드
0      DL      210
1  DRB동일   4840
2      DSR  155660
3       GS   78930
4  GS글로벌   1250
..
819  현대코퍼레이션   11760
820  현대코퍼레이션홀딩스  227840
821      화신   10690
822   흥국화재    540
823   흥아해운   3280

[824 rows x 2 columns]
코스닥 종목 수: 1528
   회사명   종목코드
0       3S   60310
1    EDGC  245620
2   ES큐브   50120
3     ISC   95340
4   J₩신약   67290
..
1523 한컴위드   54920
1524 현대사료   16790
1525 홈센타홀딩스   60560
1526 휴맥스홀딩스   28080
1527   힘스  238490

[1528 rows x 2 columns]
```

[이슈]

1. pd.read_excel 로 xls 파일을 읽을 수 없는 문제

해결: xls 파일을 xlsx 로 변환 (<https://blog.naver.com/jin31200/221355318078>)

2. Pandas 기본 엔진이었던 xlrd 가 업데이트되면서 xlsx 를 지원하지 않는 문제

해결: Pandas 기본 엔진을 openpyxl 로 변경 (pip install openpyxl)

(<https://steadiness-dev-invest.tistory.com/141>)

2. 심화과정(크롤링)

사용 기술

- Python
- Selenium
- BeautifulSoup
- ChromeDriver
- CSV

구현 방법

1. 웹 개발자 도구에서 class name 을 활용해 테이블을 구성하는 element 에 접근

- 웹 개발자 도구

```
▼<table class="list type-00 tmt30" summary="회사명, 업종, 주요제품, 상장일, 결산월, 대표자명, 홈페이지, 지역"> == $0
  <caption>목록</caption>
  <colgroup>...</colgroup>
  <thead>...</thead>
  <tbody>
    <tr class="first">
      <td class="first" title="케이티비네트웍(주)">...</td>
```

- code sample

```
soup = BeautifulSoup(browser.page_source, "html.parser")
data_rows = soup.find("table", "list type-00tmt30")
               .find("tbody").find_all("tr")
```

2. 각 row 마다 element 의 text 값을 가져와 리스트 형태로 전달

```
for row in data_rows:
    columns = row.find_all("td")

    # 현재 행의 text 값을 가져온다
    data = [column.get_text().strip() for column in columns]
```

3. 리스트를 넘겨받아 csv 파일에 기록

```
now = datetime.datetime.now().strftime('%Y-%m-%d %H %M')
filename = BASE_DIR + '/상장법인목록(' + now + ').csv'
f = open(filename, 'w', encoding='utf-8-sig', newline='')
writer = csv.writer(f)

...

writer.writerow(data)
```

4. 다음 페이지 버튼으로 페이지를 이동하고 1~3 을 반복 (마지막 페이지까지 도달하면 종료)

```
while True:

    ...

    curr_page = soup.find("div", "info type-00").find("strong")
                .get_text().split()
    # 마지막 페이지면 종료
    if int(curr_page[0]) == int(full_page):
        break;
    # 다음 페이지
    element = browser.find_element_by_class_name("next")
    browser.execute_script("arguments[0].click();", element)
```